PhD Thesis

BIOLOGICAL CLOCK MEASURES:

Assessing the association between the circadian and epigenetic clock as predictors of migration phenology and biological aging in wildlife

by Louis Stephanus Le Clercq

UNIVERSITY OF THE FREE STATE
UNIVERSITEIT VAN DIE VRYSTAAT
YUNIVESITHI YA FREISTATA
UFS UV

NATURAL AND AGRICULTURAL SCIENCES
UFS

Biological clock measures: Assessing the association between the circadian and epigenetic clock as predictors of migration phenology and biological aging in wildlife

By

Louis Stephanus Le Clercq

A thesis by publication submitted in fulfilment of the requirements for the **Doctorate** degree in **Genetics**

in the

DEPARTMENT OF GENETICS
FACULTY OF NATURAL AND AGRICULATURAL SCIENCES
UNIVERSITY OF THE FREE STATE
BLOEMFONTEIN

2024

SUPERVISOR: Dr. D.L. Dalton
CO – SUPERVISOR: Prof. J. P. Grobler and Prof. A. Kotzé

COPYRIGHT

ORCID iD: 0000-0002-8713-8920

ISNI: 0000-0005-0270-4324

VIAF: 4721 6266 4553 6550 0259 7

Published by the **University of the Free State**.

University of the Free State
205 Nelson Mandela Drive,
Park West,
Bloemfontein, 9301
https://www.ufs.ac.za/

A catalogue record for this thesis is available from the **University of the Free State** library, along with an electronic version on **KOVSIE Scholar**.

ISBN 978-1-77633-216-8 (print)
ISBN 978-1-77633-217-5 (electronic)
DOI: http://dx.doi.org/10.13140/RG.2.2.26697.47203/2

DECLARATION

To whom it may concern,

I, Louis Stephanus Le Clercq, declare that the research thesis I submit at the University of the Free State, is my independent work and that I have not previously submitted it for qualification at another institution of higher education.

I also declare that this research was conducted in accordance with the code of conduct, stipulated in the *Natural Scientific Professions Act (Act 27 of 2003),* for registered scientists.

Experiments were recorded in the following laboratory books:

- Res18/2020 Le Clercq 001

Sincerely,

2024-08-02

Louis S. Le Clercq **Date**

B.Sc., B.Sc. (Hons), M.Sc. (UP), Ph.D. (UFS)
Can.Sci.Nat (SACNASP)

ACKNOWLEDGEMENTS

Firstly, I would like to acknowledge and thank my supervisors, Dr Desiré Dalton, Prof Paul Grobler and Prof Antoinette Kotzé for their continued mentorship and support during this project. More specifically, I would like to express my gratitude to Dr Dalton for providing me of the opportunity to do my degree under her stewardship on a project that encompasses so many of my research interests. Collaborating on this project has been an immensely positive experience. The encouragement and trust that I've received, along with the freedom to always remain authentic to myself, has revigorated my own confidence and trust in my knowledge and skills. I also want to thank Prof Grobler and Prof Kotzé for their sage advice and for creating a collaborative environment with mutual respect, even though I am still an early career researcher with a few things left to learn. After all, I am just a Ph.D. student with D.Sc. dreams. I can honestly say that good collaborators are often difficult to come by for postgraduates, however, I was truly lucky to have had such an excellent support network.

To my international collaborators, from Italy (Dr Bazzi, Dr Cecere, Prof Rubolini, and Prof Gianfranceschi), Germany (Prof Liedvogel), and Spain (Dr Ferrer Obiol, Prof Riutort, and Prof González-Solís), my sincere gratitude for the willingness to contribute data and serve as co-authors. Your feedback was always kind, prompt, and helpful in deepening my understanding of avian research and migration studies. It has been a great honour and privilege to collaborate on two articles and to be published alongside international researchers this early in my career. It is my hope that this will be but the first of many collaborations.

At SANBI, I would like to thank Sonia Kropff, Veronica Phetla, Prudent Mokgokong, and Kim Labuschagne, for research support provided during this project, specifically in navigating the lab, acquiring samples, and dealing with finances. At UFS, I would like to thank research librarian Carmen Nel and Cornelle Scheltema for assisting me with the sourcing of difficult to find literature as well as open access funding for my publications. At Inqaba biotech., I would like to thank Dr Aron Abera for assistance in double-checking the EpiTYPER assay designed and used in this study, and assistance in processing the samples and raw mass array data. This work is based on the research supported wholly/in part by the National Research Foundation of South Africa (Grant Number: 112062).

SUMMARY

This thesis explores the use of biological clocks, studied at the molecular level, in understanding and predicting animal traits that change over time. In particular, the thesis focuses on two biological clocks: the circadian clock and the epigenetic clock. The study aimed to shed light on how these clocks, along with specific genes associated with them, influence traits such as migration patterns in birds and age in animals. In the investigation of circadian clock genes, this thesis presents a detailed review of existing literature, presenting both supporting and conflicting results on the association between clock gene polymorphisms and migration patterns. This review summarized the central hypotheses tested in these studies, identified several candidate genes that have been used, revealed distinct patterns in terms of the taxonomy and phylogeny of studied species, and present new insights into why conflicting results happen as well as what future research is needed. Furthermore, the review emphasizes the importance of considering molecular clock differences between lineages when studying multiple species. The two most studied candidate genes were then tested for associations to phenology in an intra-African migrant, the Diederik cuckoo, and provides the first evidence that phenotypic correlates identified in Eurasian passerine species is conserved for these species. The second part of the thesis delves into epigenetic clocks, examining two widely used methods: methylation and telomere length. A systematic review and meta-analysis were performed to compare the performance of these methods across vertebrate classes. Methylation was found to outperform telomere length in predictive power, with both methods showing promise as biomarkers for age determination in animals. To demonstrate the practical application of epigenetic clocks, the study included the development of a new age determination model on the African cheetah. Six candidate genes were identified, and a model using CpG methylation levels was created using machine learning techniques to refine the model, resulting in accurate age predictions. This approach offers a less invasive means of age estimation for population monitoring. The research also introduces two new PYTHON tools, PAReTT, to incorporate phylogenetic and molecular clock data into ecological and evolutionary reviews, and ABCal, to address potential biases in systematic reviews in evolution and ecology. The thesis concludes by highlighting the broader implications of the study, emphasizing the utility of biological clocks in understanding

temporal traits, from annual life events in birds to lifelong aging in mammals. The generated datasets and tools contribute to ecological systematic reviews and individual studies, expanding our knowledge of biological clocks and guiding future research endeavours.

Key terms: molecular marker**,** biological clocks, circadian, migration, birds, epigenetic, aging, methylation, telomeres, animals.

The sample and study data were deposited, with accession numbers and links, to BioProject accession number **PRJNA737185** in the NCBI database.
Available at: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA737185

ETHICS STATEMENT

Research was conducted according to strict animal research ethics regulations in line with institutional policies, the *National Environmental Management: Biodiversity Act (Act 10 of 2004),* and section 20 of the *Animal Diseases Act 1984 (Act 35 of 1984)*. Relevant committee approval and permits were obtained as follows:

Protocol approval:

Committee name:	**Approval number:**
Department of Genetics Research Committee, UFS	Res18/2020

Ethics approval:

Committee name:	**Approval number:**
Interfaculty Animal Ethics Committee, UFS	UFS-AED2020/0015/1709
SANBI NZG Animal Research Ethics and Scientific Committee	SANBI/RES/P2020/30

Research permits:

Authority name:	**Permit type:**	**Permit number:**
Internal (UFS/SANBI):		
Department of Agriculture, Land Reform, and Rural Development (RSA)	Section 20	12/11/1/1/18(1824JD)
Department of Environment, Forestry, and Fisheries (Oceans and Coasts, RSA)	TOPS	S-03175
Department of Environment, Forestry, and Fisheries (RSA)	TOPS	O-52903
External (Inqaba):		
Department of Environment, Forestry, and Fisheries (RSA)	TOPS	O-03336

Sample requisitions:

Institution name:	**Reference number:**	**Material Transfer Agreement:**
SANBI Biobank (RSA)	SANBI/BB/2021/P2020/30	SANBI/BB/2021/P2020/30

FUNDING STATEMENT

The following applications were made for this project as well as student financial support, those indicated with as asterisk (*) were successful.

Bursary applications:

Institution/Bursar name:	**Number:**
Doctoral Innovation, Free Standing and Scarce Skills (NRF)*	112062
Postgraduate Tuition Fee Bursary (UFS)	2006026690

Grant applications:

Institution/Grantee name:	**Number:**
African Bird Club Conservation Fund (2020)	n. a.
African Bird Club Conservation Fund (2021)	n. a.
Club300 Bird Protection	n. a.
Inqaba Biotech Industries, SEED-It Grant*	n. a.
Rufford Foundation 1st Small Grant (2020)	31514-1
Rufford Foundation 1st Small Grant (2021)	35796-1
South African Veterinary Foundation	n. a.
The MBZ Species Conservation Fund	212527624
UFS Library Open Access Fund*	13560591

RESEARCH OUTPUT

Presentations:

Le Clercq, L.S, Dalton, D.L., Kotzé, A., Grobler, J.P. **Biological clock measures: Assessing the association between the circadian and epigenetic clock as predictors of migration phenology and biological aging in wildlife**. Protocol presentation, July 2020. Department of Genetics, UFS.
URL: https://www.youtube.com/watch?v=Qmz0YSoOPMA

Le Clercq, L.S. **What's your number? Animal age estimation for the 21st century**. Flash Facts competition, July 2021. Department of Genetics, UFS. (Winner in Student Category). URL: https://www.youtube.com/watch?v=MJXbvrWCPhM

Le Clercq, L.S, Dalton, D.L., Kotzé, A., Grobler, J.P. **Biological clock measures: Assessing the association between the circadian and epigenetic clock as predictors of migration phenology and biological aging in wildlife**. Project presentation 1, July 2021. Department of Genetics, UFS. (same as below)

Le Clercq, L.S. L.S, Dalton, D.L., Kotzé, A., Grobler, J.P**. What's your number? Animal age estimation for the 21st century**. Presentation, 16 September 2021. SASBi-SC/SAGS Student Symposium 2021.
URL: https://www.youtube.com/watch?v=rD1rdLtY1uU

Le Clercq, L.S, Dalton, D.L., Kotzé, A., Grobler, J.P. **Biological clock measures: Assessing the association between the circadian and epigenetic clock as predictors of migration phenology and biological aging in wildlife**. Project presentation 2, June 2022. Department of Genetics, UFS.
URL: https://www.youtube.com/watch?v=RpS5DrdkPHU

Le Clercq, L.S, Dalton, D.L., Kotzé, A., Grobler, J.P. **Biological clock measures: Assessing the association between the circadian and epigenetic clock as predictors of migration phenology and biological aging in wildlife**. Project presentation 3, October 2023. Department of Genetics, UFS.
URL: https://www.youtube.com/watch?v=B0pSAe6g_x8

Posters:

Le Clercq, L.S, Dalton, D.L., Kotzé, A. **Molecular age estimation based on promotor CpG methylation using Methylation Sensitive PCR**. Poster, June 2018. NZG, SANBI.
DOI: http://doi.org/10.13140/RG.2.2.11352.21768

Publications:

Le Clercq, L.-S., Bazzi, G., Cecere, J.G., Gianfranceschi, L., Grobler, J.P., Kotzé, A., Rubolini, D., Liedvogel, M. and Dalton, D.L. (2023), **Time trees and clock genes: a systematic review and comparative analysis of contemporary avian migration genetics**. *Biological Reviews*. https://doi.org/10.1111/brv.12943

Le Clercq, L.-S., Bazzi, G., Ferrer Obiol, J., Cecere, J., Gianfranceschi, L., Grobler, J.P., Kotze, A., Riutort, M., González-Solís, J., Rubolini, D., et al. (2023), **Birds of a feather flock together: a dataset for *Clock* and *Adcyap1* genes from migration genetics studies**. *Scientific Data.* https://doi.org/10.1038/s41597-023-02717-8.

Le Clercq, L.-S., Kotzé, A., Grobler, J.P. and Dalton, D.L. (2023), **PAReTT: a Python package for the automated retrieval and management of divergence time data from the TimeTree resource for downstream analyses**. *Journal of Molecular Evolution.* https://doi.org/10.1007/s00239-023-10106-3

Le Clercq, L.-S., Kotzé, A., Grobler, J.P., and Dalton, D.L. (2023), **Biological clocks as age estimation markers in animals: a systematic review and meta-analysis**. *Biological Reviews*. https://doi.org/10.1111/brv.12992

Le Clercq, L.-S., Kotzé, A., Grobler, J.P., and Dalton, D.L. (2023), **Dataset generated in a systematic review and meta-analysis of biological clocks as age estimation markers in animal ecology**. *Data in Brief*. https://doi.org/10.1016/j.dib.2023.109615

Le Clercq, L.-S. (2023), **ABCal: a Python package for Author Bias Computation and Scientometric Plotting for Reviews and Meta-Analyses**. *Scientometrics*. https://doi.org/10.1007/s11192-023-04880-6.

Le Clercq, L.-S., Kotzé, A., Grobler, J.P., and Dalton, D.L. (2024), **Phenotypic correlates between candidate genes for migration among races of Diederik cuckoos, *Chrysococcyx caprius***. *Ecology and Evolution.* https://doi.org/10.1002/ece3.70117

Le Clercq, L.-S., Kotzé, A., Grobler, J.P., and Dalton, D.L. (2024), **Methylation-based markers for the estimation of age in African Cheetah, *Acinonyx jubatus***. *Molecular Ecology Resources.* https://doi.org/10.1111/1755-0998.13940.

Protocols:

Le Clercq, L.-S, Dalton, D.L., Kotzé, A. & Grobler, J.P. (2023), **PCR Amplification of *Clock* and *Adcyap1* genes with EmeraldAmp® GT PCR Master Mix in Avian species for polymorphism elucidation**. *protocols.io.* https://dx.doi.org/10.17504/protocols.io.6qpvrdwk3gmk/v1

Le Clercq, L.-S., Dalton, D.L., Kotzé, A. & Grobler, J.P. (2023), **ABI Sanger Sequencing of Avian Clock genes to elucidate markers for Migration Phenology**. *protocols.io.* https://dx.doi.org/10.17504/protocols.io.3byl4k6zrvo5/v1

Le Clercq, L.-S., Dalton, D.L., Kotzé, A. & Grobler, J.P. (2023), **Designing an EpiTYPER bisulfite sequencing assay for age estimation in *Acinonyx jubatus* based on human orthologues**. *protocols.io*. https://dx.doi.org/10.17504/protocols.io.j8nlk4yk1g5r/v1

Le Clercq, L.-S., Dalton, D.L., Kotzé, A. & Grobler, J.P. (2023), **DNA extraction protocol for animal blood samples using the E.Z.N.A blood mini kit**. *protocols.io*. https://dx.doi.org/10.17504/protocols.io.ewov141xpvr2/v1

GitHub Repositories:

Le Clercq, L.-S. (2023**). Python Automated Retrieval of TimeTree data (PAReTT)**, *version 1.0.2*. https://github.com/LSLeClercq/PAReTT

Le Clercq, L.-S. (2023). **Dataset of Clock and Adcyap1 alleles for Birds (AvianClocksData)**, *version 1.0.2*. https://github.com/LSLeClercq/AvianClocksData

Le Clercq, L.-S. (2023). **Author Bias Computation and Scientometric Plotting (ABCal)**, *version 1.0.2*. https://github.com/LSLeClercq/ABCal

QUOTE

“All the work of the [Scientists] serves only to demonstrate that there is only variety everywhere where they suppose uniformity ... that in nature there is nothing absolute, nothing perfectly regular.”

— **Georges-Louis Leclerc, Count of Buffon**
(1707–1788)

TABLE OF CONTENTS

LIST OF APPENDICES

CHAPTER 1

Introduction

1.1 MOLECULAR STUDIES OF BIOLOGICAL CLOCKS

Numerous biological attributes in animals have been studied at the biochemical and molecular level, leading to the development of specialised tools that are useful in ecological studies and conservation efforts (Carroll *et al.*, 2018; Riekkola *et al.*, 2018). Such tools are useful because they provide important demographic data used in simulation studies and wildlife management. Studies have included methods for sex identification (Rovatsos & Kratochvíl, 2017), parentage determination (Olson *et al.*, 2008), species or population assignment (Helyar *et al.*, 2011; Kim & Sappington, 2013), and hybrid speciation events (Väli *et al.*, 2018). A common feature of these attributes is that they are static events or immutable traits that do not show progressive change with time. There are, however, several biological attributes that vary temporally such as the molecular clock (Thorpe, 1982; Li, Tanimura, & Sharp, 1987; Robinson & Robinson, 2001) and biological clocks (Horvath & Raj, 2018; Lincoln, 2019).

The molecular clock hypothesis was first conceived in the 1960's through work that characterised the rate of changes in haemoglobin proteins of closely related primate species (Zuckerkandl, 1960; Zuckerkandl & Pauling, 1962, 1965). These studies illustrated that fairly constant patterns of molecular changes occur in conserved genes and that the extent of such changes when comparing related species reflects the relative time since these species separated (Zuckerkandl, 1987). This discovery was pivotal in advancing evolutionary studies and became the keystone concept for the development of many evolutionary concepts such as time-calibrated species trees (Kumar & Hedges, 2016) and related divergence time (Rannala & Yang, 2003), diversification rate (Lanfear *et al.*, 2010), speciation rate (Etienne & Apol, 2009), and phylogenetics (Voelker, Bowie, & Klicka, 2013). While the molecular clock typically deals with changes that occur over multiple generations, and often millions of years, biological clocks are tied to changes that occur within the lifetime of contemporary populations. Biological clocks may refer one of several attributes including: lifelong changes in fertility (Cooper *et al.*, 2021), the circadian- (Aguilar-Roblero, Díaz-Muñoz, & Fanjul-Moles, 2015) and circannual

clocks (Lincoln, 2019), or the epigenetic clocks (Horvath & Raj, 2018). Several of these biological clocks being tied to key attributes of animal ecology including migration (Kumar *et al.*, 2010; Merlin & Liedvogel, 2019) and aging (Jarman *et al.*, 2015; Horvath & Raj, 2018).

1.1.1 Biological Clocks and Migration

Each year, flocks of birds take to the sky to make their annual journey from non-breeding to breeding grounds. This event happens like clockwork: carefully timed to ensure an optimal flight to reach their destination. The repeated occurrence of this characteristic behaviour has sparked much interest in the field of chronobiology with emphasis on both intrinsic as well as environmental cues, called *Zeitgeber*, contributing to timekeeping (Åkesson *et al.*, 2017). Understanding how a biological clock is able to anticipate these changes and adapt to them is especially important for conservation given the potential effect of climate change on migratory behaviour and phenology (Carey, 2009), requiring shifts in spring staging site (Verhoeven *et al.*, 2018), as well as range expansion (Greig, Wood, & Bonter, 2017) and habitat changes (LeClerc & Cristol, 2005) due to urbanization.

Several studies have aimed at answering these questions by comparing migratory phenology to putative variation within clock genes (Caprioli *et al.*, 2012; Saino *et al.*, 2015a; Delmore *et al.*, 2016). Two promising clock genes have been identified: Circadian Locomoter Output Cycles Protein Kaput (*Clock*) and Adenylate Cyclase Activating Polypeptide 1 (*Adcyap1*; (Johnsen *et al.*, 2007; Bazzi *et al.*, 2016). The former showed variability within a series of poly-glutamine residues which differs both between and within species; the latter shows variation within the 3'-UTR of the gene (Contina *et al.*, 2018). A significant correlation between variability and migration patterns has been illustrated in multiple species, including swallows (Caprioli *et al.*, 2012; Bazzi *et al.*, 2015), tits (Liedvogel *et al.*, 2009), and flycatchers (Kuhn *et al.*, 2013); whilst a number of studies have failed to show a significant correlation for other species (Contina *et al.*, 2018). Most studies focussed on a relationship between clock gene diversity, and inheritance, and migration patterns of groups within a species that showed variable timing for departure and arrival; although the possibility of an epigenetic effect has also been proposed (Saino *et al.*, 2017; Merlin & Liedvogel, 2019). An inter-species study found there to be no clear, generalized, genetic basis for divergent migration phenology between

migratory and resident bird species (Lugo Ramos, Delmore, & Liedvogel, 2017); thus, migratory phenology seems to vary in a species specific manner, where some species are fine-tuned to environmental cues to initiate migration whilst others are not, and some species have significant variability in key clock genes that partially contribute to such differences in migration phenology.

A key factor in whether selection of these genes drives divergent migratory behaviour may be the taxonomic and phylogenetic context within which they were studied. Some species, such as the Streaky-breasted flufftail, *Sarothrura boehmi* (Taylor & Kirwan, 2020) and European bee-eater, *Merops apiaster* (Fry & Boesman, 2020), show two distinct conspecific populations of resident and migratory birds despite being classified as monotypic. On the other hand, migration studies on congeneric spotted eagle hybrids (*Clanga clanga* and *Clanga pomarine*) have shown that hybridization between species with divergent movement patterns results in a new population with migration strategies that differ from their progenitors (Väli *et al.*, 2018). The exact mechanisms responsible for such differences in migratory phenology within a species are still poorly understood (Newton, 2007). Further studies, particularly in previously unstudied species, are thus warranted to assess the degree of variability in candidate genes, their relation to migration phenology, and utility in answering important questions regarding breeding phenology such as persistent gene flow (Williams & Lindell, 2018).

1.1.2 Biological Clocks and Age

Another key biological attribute in animal ecology is age, as it can be used to assign animals to important functional classes ranging from immature young to old adults (Lambert *et al.*, 2018; Singleton *et al.*, 2018). Each class differs substantially in terms of degree of development and reproductive maturity (Noren & Edwards, 2007), reproductive potential (Charpentier *et al.*, 2008), reproductive success (Bradley & Safran, 2014), future reproductive potential (Jones *et al.*, 2014), disease burden (Starr & Saito, 2014), and mortality rates (Pérez-Barberìa *et al.*, 2014). These factors are usually considered when modelling population dynamics (Tarnita *et al.*, 2009; Allen & Tarnita, 2014) to determine the conservation status of species and predict future fitness or recovery (Stalmans *et al.*, 2019).

Several methods have been applied to determine age in wildlife. The most common method employs physical measures that are usually compared to an

established chart. This includes measures of the skull (Smuts, Anderson, & Austin, 1978), the body (Marker & Dickman, 2003), scales (Isermann, Wolter, & Breeggemann, 2010), wingspan (Erdem *et al.*, 2021), or weight. While this approach is particularly utile in younger individuals that experience significant growth or species with continuous life-long growth (e.g., some fish and reptiles), it generally becomes less useful in adults of varying age classes for species where growth arrests at adulthood. As such, age determination from physical measures has been extended to other physical attributes such as gumline recession (Hiller & Tyre, 2014) or more general body condition (Marker & Dickman, 2003). A second approach is the use of a “mark-recapture” or “identify-resighting” approach that is useful in long term studies of species confined to a specific habitat. This approach can use natural markings, like the distinct pigmentation patterns on the flukes of a whale (Polanowski *et al.*, 2014), or artificial markings, including uniquely numbered bands (Sherley *et al.*, 2014), geolocators (Saino *et al.*, 2015b), or tracking collars (Marker *et al.*, 2003). Either approach has several limitations and varied degrees of accuracy which has necessitated the development of laboratory-based methods for more accurate age determination.

Two biological clocks, measuring histological features of internal structures such as tooth cementum annulation (Takken Beijersbergen, 2019) in toothed vertebrates and otolith annulation in fish (Campana, 2001), are currently used to provide accurate age estimation of animals. While both these clocks are accurate their invasiveness makes age determination in living animals more cumbersome. Seeing as many aspects of the natural aging process are under genetic control as part of epigenetic clocks, and therefore a programmed process (Horvath and Raj, 2018), molecular methods to determine age in wildlife are possible. Epigenetics is a collective term for mechanisms that modify DNA and DNA packaging independent of genetic sequence. One widely studied epigenetic feature is DNA methylation: a process that adds a methyl group to the 5’ cytosine of Cytosine-Guanine pairs (CpG’s). Studies have revealed that within genes, nearly a third of all CpG sites are influenced by age and given its consistency, the epigenetic clock is a promising avenue of chronological age prediction (Hong *et al.*, 2019). Considering, however, that this has been illustrated most extensively in human studies (Bocklandt *et al.*, 2011; Bekaert *et al.*, 2015; Giuliani *et al.*, 2016), more studies are needed to identify

those genes with conserved age-dependent methylation patterns across taxa and develop new assays to screen and model them for age determination in animals.

This study will explore the use of two biological clocks, the circadian clock and the epigenetic clock, to answer key questions regarding the differential migration patterns of African birds as well as the utility of epigenetic clocks to model and predict age in African mammals of high conservation concern due to their population status.

1.2 AIMS

To use genetic methods to investigate two biological clock measures: (I) migration phenology and (II) age, by:

- Investigating the relationship between genetic variance in circadian clock genes and their potential role in differential migration phenology in African birds.
- Investigating the potential of age-related changes in the methylation of genes under epigenetic clock control to be used to create an accurate age estimation model in African mammals.

1.3 OBJECTIVES

The aims of this study will be achieved through focussed studies on selected clock measures in a range of species of conservation concern. Specific objectives are to:

- Perform a systematic review and standardised comparative analysis to synthesize the existing studies and data for the use of clock gene polymorphisms in studying annual synchronicity in life events such as breeding and migration.
- Explore or design methods of incorporating phylogenetic, taxonomic, and divergence time data into ecological and evolutionary systematic reviews.
- Sample different populations of Diederik cuckoo (*Chrysococcyx caprius*) that show diverse migratory phenology such as resident versus migrant, migration to two different sites, as well as early migration and nesting birds rather than those arriving later.

- Amplify and sequence a region of the poly-Q repeat residue of the *Clock* gene and the 3'-UTR of the *Adcyap1* gene.
- Determine if a correlation exists between gene variation and migration phenology (migratory distance, time of departure/arrival, and breeding latitude) in this species.
- Perform a systematic review and meta-analysis to summarise and evaluate the existing studies that used epigenetic clocks as biomarkers for modelling animal age and identify the most useful biomarkers.
- Explore potential sources of heterogeneity and bias and possible methods to account for bias in ecological and evolutionary systematic reviews.
- Select an optimal cohort of samples for cheetah (*Axinonyx jubatus*) with distinct, relevant, age classes across their lifespan.
- Find the best candidate genes based on known age associations and a high degree of conserved human-cheetah CpG's from the literature to design a cheetah specific assay.
- Screen for differential CpG methylation with significant age-related changes in methylation for cheetah, using mass array technology, to develop and validate a model with machine learning for age estimation.

1.4 THESIS ORGANISATION

The thesis is structured as chapters, as per the table of contents, and thesis pagination is indicated in the middle of the footer (as opposed to article or journal page numbers). The first chapter presents a general introduction and the global aims and objectives of the study as well as the thesis organisation. Hereafter, each chapter starts with a title page to detail the publications that constitute each chapter. The second and third chapters represent published literature reviews, with related data and method companion articles, on biological clocks and studies that have used those clocks to answer questions pertinent to the aims and objectives of the present study. Subsequent chapters represent scientific publications presented together based on thematic content and follow the standard article format. Due to the scope of species covered in this study, each chapter contains the species-specific details such as taxonomy, conservation concern, and distribution as well as the chapter-specific objectives and methods. The final chapter summarizes the

broader conclusions, contextualizing the main findings of the study with the reviewed literature and aims of the study. This thesis also includes several appendices that collate the supplementary materials and methods for individual publications.

REFERENCES

AGUILAR-ROBLERO, R., DÍAZ-MUÑOZ, M. & FANJUL-MOLES, M.L. (2015) *Mechanisms of circadian systems in animals and their clinical relevance*. Springer International Publishing.

ÅKESSON, S., ILIEVA, M., KARAGICHEVA, J., RAKHIMBERDIEV, E., TOMOTANI, B. & HELM, B. (2017) Timing avian long-distance migration: from internal clock mechanisms to global flights. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372**, 20160252.

ALLEN, B. & TARNITA, C.E. (2014) Measures of success in a class of evolutionary models with fixed population size and structure. *Journal of Mathematical Biology* **68**, 109–143.

BAZZI, G., AMBROSINI, R., CAPRIOLI, M., COSTANZO, A., LIECHTI, F., GATTI, E., GIANFRANCESCHI, L., PODOFILLINI, S., ROMANO, A., ROMANO, M., SCANDOLARA, C., SAINO, N. & RUBOLINI, D. (2015) Clock gene polymorphism and scheduling of migration: A geolocator study of the barn swallow *Hirundo rustica*. *Scientific Reports* **5**, 12443.

BAZZI, G., CECERE, J.G., CAPRIOLI, M., GATTI, E., GIANFRANCESCHI, L., PODOFILLINI, S., POSSENTI, C.D., AMBROSINI, R., SAINO, N., SPINA, F. & RUBOLINI, D. (2016) Clock gene polymorphism, migratory behaviour and geographic distribution: a comparative study of trans-Saharan migratory birds. *Molecular Ecology* **25**, 6077–6091.

BEKAERT, B., KAMALANDUA, A., ZAPICO, S.C., VAN DE VOORDE, W. & DECORTE, R. (2015) Improved age determination of blood and teeth samples using a selected set of DNA methylation markers. *Epigenetics* **10**, 922–930.

BOCKLANDT, S., LIN, W., SEHL, M.E., SÁNCHEZ, F.J., SINSHEIMER, J.S., HORVATH, S. & VILAIN, E. (2011) Epigenetic Predictor of Age. *PLoS One* **6**, e14821.

BRADLEY, R.J. & SAFRAN, R.J. (2014) Conceptual revision and synthesis of proximate factors associated with age-related improvement in reproduction. *Ethology* **120**, 411–426.

CAMPANA, S.E. (2001) Accuracy, precision and quality control in age determination, including a review of the use and abuse of age validation methods. *Journal of Fish Biology* **59**, 197–242.

CAPRIOLI, M., AMBROSINI, R., BONCORAGLIO, G., GATTI, E., ROMANO, A., ROMANO, M., RUBOLINI, D., GIANFRANCESCHI, L. & SAINO, N. (2012) Clock gene variation is associated with breeding phenology and maybe under directional selection in the migratory barn swallow. *PLoS One* **7**, e35140.

CAREY, C. (2009) The impacts of climate change on the annual cycles of birds.

Philosophical Transactions of the Royal Society B: Biological Sciences **364**, 3321–3330.

CARROLL, E.L., BRUFORD, M.W., DEWOODY, J.A., LEROY, G., STRAND, A., WAITS, L. & WANG, J. (2018) Genetic and genomic monitoring with minimally invasive sampling methods. *Evolutionary Applications* **11**, 1094–1119.

CHARPENTIER, M.J.E., TUNG, J., ALTMANN, J. & ALBERTS, S.C. (2008) Age at maturity in wild baboons: Genetic, environmental and demographic influences. *Molecular Ecology* **17**, 2026–2040.

CONTINA, A., BRIDGE, E.S., ROSS, J.D., SHIPLEY, J.R. & KELLY, J.F. (2018) Examination of Clock and Adcyap1 gene variation in a neotropical migratory passerine. *Plos One* **13**, e0190859.

COOPER, E.B., BONNET, T., OSMOND, H.L., COCKBURN, A. & KRUUK, L.E.B. (2021) Aging and senescence across reproductive traits and survival in superb fairy-wrens (*Malurus cyaneus*). *American Naturalist* **197**, 111–127.

DELMORE, K.E., TOEWS, D.P.L., GERMAIN, R.R., OWENS, G.L. & IRWIN, D.E. (2016) The Genetics of Seasonal Migration and Plumage Color. *Current Biology* **26**, 2167–2173.

ERDEM, E., ÖZBAŞER, F.T., GÜRCAN, E.K. & SOYSAL, M.I. (2021) The morphological and morphometric characteristics of Alabadem pigeons. *Turkish Journal of Veterinary and Animal Sciences* **45**, 372–379.

ETIENNE, R.S. & APOL, M.E.F. (2009) Estimating speciation and extinction rates from diversity data and the fossil record. *Evolution* **63**, 244–255.

FRY, H. & BOESMAN, P.F.D. (2020) European Bee-eater (*Merops apiaster*), version 1.0. In *Birds of the World* (eds J. DEL HOYO, A. ELLIOTT, J. SARGATAL, D.A. CHRISTIE & E. DE JUANA), Cornell Lab of Ornithology, Ithaca, NY, USA.

GIULIANI, C., CILLI, E., BACALINI, M.G., PIRAZZINI, C., SAZZINI, M., GRUPPIONI, G., FRANCESCHI, C., GARAGNANI, P. & LUISELLI, D. (2016) Inferring chronological age from DNA methylation patterns of human teeth. *American Journal of Physical Anthropology* **159**, 585–595.

GREIG, E.I., WOOD, E.M. & BONTER, D.N. (2017) Winter range expansion of a hummingbird is associated with urbanization and supplementary feeding. *Proceedings of the Royal Society B: Biological Sciences* **284**, 20170256.

HELYAR, S.J., HEMMER-HANSEN, J., BEKKEVOLD, D., TAYLOR, M.I., OGDEN, R., LIMBORG, M.T., CARIANI, A., MAES, G.E., DIOPERE, E., CARVALHO, G.R. & NIELSEN, E.E. (2011) Application of SNPs for population genetics of nonmodel organisms: New opportunities and challenges. *Molecular Ecology Resources* **11**, 123–136.

HILLER, T.L. & TYRE, A.J. (2014) Comparison of Two Age-Estimation Techniques for Cougars. *Northwestern Naturalist* **95**, 77–82.

HONG, S.R., SHIN, K.J., JUNG, S.E., LEE, E.H. & LEE, H.Y. (2019) Platform-independent models for age prediction using DNA methylation data. *Forensic Science International: Genetics* **38**, 39–47.

HORVATH, S. & RAJ, K. (2018) DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nature Reviews Genetics* **19**, 371–384.

ISERMANN, D.A., WOLTER, M.H. & BREEGGEMANN, J.J. (2010) Estimating Black Crappie Age: An Assessment of Dorsal Spines and Scales as Nonlethal Alternatives to Otoliths. *North American Journal of Fisheries Management* **30**, 1591–1598.

JARMAN, S.N., POLANOWSKI, A.M., FAUX, C.E., ROBBINS, J., DE PAOLI-ISEPPI, R., BRAVINGTON, M. & DEAGLE, B.E. (2015) Molecular biomarkers for chronological age in animal ecology. *Molecular Ecology* **24**, 4826–4847.

JOHNSEN, A., FIDLER, A.E., KUHN, S., CARTER, K.L., HOFFMANN, A., BARR, I.R., BIARD, C., CHARMANTIER, A., EENS, M., KORSTEN, P., SIITARI, H., TOMIUK, J. & KEMPENAERS, B. (2007) Avian Clock gene polymorphism: Evidence for a latitudinal cline in allele frequencies. *Molecular Ecology* **16**, 4867–4880.

JONES, O.R., SCHEUERLEIN, A., SALGUERO-GÓMEZ, R., CAMARDA, C.G., SCHAIBLE, R., CASPER, B.B., DAHLGREN, J.P., EHRLÉN, J., GARCÍA, M.B., MENGES, E.S., QUINTANA-ASCENCIO, P.F., CASWELL, H., BAUDISCH, A. & VAUPEL, J.W. (2014) Diversity of ageing across the tree of life. *Nature* **505**, 169–173.

KIM, K.S. & SAPPINGTON, T.W. (2013) Microsatellite data analysis for population genetics. *Methods in Molecular Biology* **1006**, 271–295.

KUHN, K., SCHWENK, K., BOTH, C., CANAL, D., JOHANSSON, U.S., VAN DER MIJE, S., TÖPFER, T. & PÄCKERT, M. (2013) Differentiation in neutral genes and a candidate gene in the pied flycatcher: Using biological archives to track global climate change. *Ecology and Evolution* **3**, 4799–4814.

KUMAR, S. & HEDGES, S.B. (2016) Advances in time estimation methods for molecular data. *Molecular Biology and Evolution* **33**, 863–869.

KUMAR, V., WINGFIELD, J.C., DAWSON, A., RAMENOFSKY, M., RANI, S. & BARTELL, P. (2010) Biological Clocks and Regulation of Seasonal Reproduction and Migration in Birds. *Physiological and Biochemical Zoology* **83**, 827–835.

LAMBERT, S., BLONDIN, P., VIGNEAULT, C., LABRECQUE, R., DUFORT, I. & SIRARD, M.-A. (2018) Spermatozoa DNA methylation patterns differ due to peripubertal age in bulls. *Theriogenology* **106**, 21–29.

LANFEAR, R., HO, S.Y.W., LOVE, D. & BROMHAM, L. (2010) Mutation rate is linked to diversification in birds. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 20423–20428.

LECLERC, J.E. & CRISTOL, D.A. (2005) Are golf courses providing habitat for birds of conservation concern in Virginia? *Wildlife Society Bulletin* **33**, 463–470.

LI, W.H., TANIMURA, M. & SHARP, P.M. (1987) An evaluation of the molecular clock hypothesis using mammalian DNA sequences. *Journal of Molecular Evolution* **25**, 330–342.

LIEDVOGEL, M., SZULKIN, M., KNOWLES, S.C.L., WOOD, M.J. & SHELDON, B.C. (2009) Phenotypic correlates of clock gene variation in a wild blue tit population: Evidence for a role in seasonal timing of reproduction. *Molecular Ecology* **18**, 2444–2456.

LINCOLN, G. (2019) A Brief History of Circannual Time. *Journal of Neuroendocrinology*, e12694.

LUGO RAMOS, J.S., DELMORE, K.E. & LIEDVOGEL, M. (2017) Candidate genes for

migration do not distinguish migratory and non-migratory birds. *Journal of Comparative Physiology A* **203**, 383–397.

MARKER, L.L. & DICKMAN, A.J. (2003) Morphology, Physical condition, and Growth of the Cheetah (*Acinonyx jubatus jubatus*). *Journal of Mammalogy* **84**, 840–850.

MARKER, L.L., DICKMAN, A.J., JEO, R.M., MILLS, M.G.L. & MACDONALD, D.W. (2003) Demography of the Namibian cheetah, *Acinonyx jubatus jubatus*. *Biological Conservation* **114**, 413–425.

MERLIN, C. & LIEDVOGEL, M. (2019) The genetics and epigenetics of animal migration and orientation: birds, butterflies and beyond. *The Journal of experimental biology* **222**, jeb191890.

NEWTON, I. (2007) *The Migration Ecology of Birds*. 1st edition. Elsevier Publishing.

NOREN, S.R. & EDWARDS, E.F. (2007) Physiological and behavioral development in delphinid calves: Implications for calf separation and mortality due to tuna purse-seine sets. *Marine Mammal Science* **23**, 15–29.

OLSON, D.E., FITZGERALD, G., HAND, D.M., ARDREN, W.R., BAUMSTEIGER, J. & SPATEHOLTS, R. (2008) Use of Parentage Analysis to Determine Reproductive Success of Hatchery-Origin Spring Chinook Salmon Outplanted into Shitike Creek, Oregon. *North American Journal of Fisheries Management* **28**, 1472–1485.

PÉREZ-BARBERÌA, F.J., DUFF, E.I., BREWER, M.J., GUINNESS, F.E., PÏREZ-BARBER??A, F.J., DUFF, E.I., BREWER, M.J., GUINNESS, F.E., PÉREZ-BARBERÌA, F.J., DUFF, E.I., BREWER, M.J., GUINNESS, F.E., PÉREZ-BARBERÍA, F.J., DUFF, E.I., BREWER, M.J., ET AL. (2014) Evaluation of methods to age Scottish red deer: The balance between accuracy and practicality. *Journal of Zoology* **294**, 180–189.

POLANOWSKI, A.M., ROBBINS, J., CHANDLER, D. & JARMAN, S.N. (2014) Epigenetic estimation of age in humpback whales. *Molecular ecology resources* **14**, 976–987.

RANNALA, B. & YANG, Z. (2003) Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics* **164**, 1645–1656.

RIEKKOLA, L., ZERBINI, A.N., ANDREWS, O., ANDREWS-GOFF, V., BAKER, C.S., CHANDLER, D., CHILDERHOUSE, S., CLAPHAM, P., DODÉMONT, R., DONNELLY, D., FRIEDLAENDER, A., GALLEGO, R., GARRIGUE, C., IVASHCHENKO, Y., JARMAN, S., ET AL. (2018) Application of a multi-disciplinary approach to reveal population structure and Southern Ocean feeding grounds of humpback whales. *Ecological Indicators* **89**, 455–465.

ROBINSON, N.E. & ROBINSON, A.B. (2001) Molecular clocks. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 944–949.

ROVATSOS, M. & KRATOCHVÍL, L. (2017) Molecular sexing applicable in 4000 species of lizards and snakes? From dream to real possibility. *Methods in Ecology and Evolution* **8**, 902–906.

SAINO, N., AMBROSINI, R., ALBETTI, B., CAPRIOLI, M., DE GIORGIO, B., GATTI, E., LIECHTI, F., PAROLINI, M., ROMANO, A., ROMANO, M., SCANDOLARA, C.,

GIANFRANCESCHI, L., BOLLATI, V. & RUBOLINI, D. (2017) Migration phenology and breeding success are predicted by methylation of a photoperiodic gene in the barn swallow. *Scientific Reports* **7**, 45412.

SAINO, N., BAZZI, G., GATTI, E., CAPRIOLI, M., CECERE, J.G., POSSENTI, C.D., GALIMBERTI, A., ORIOLI, V., BANI, L., RUBOLINI, D., GIANFRANCESCHI, L. & SPINA, F. (2015a) Polymorphism at the Clock gene predicts phenology of long-distance migration in birds. *Molecular Ecology* **24**, 1758–1773.

SAINO, N., RUBOLINI, D., AMBROSINI, R., ROMANO, M., SCANDOLARA, C., FAIRHURST, G.D., CAPRIOLI, M., ROMANO, A., SICURELLA, B. & LIECHTI, F. (2015b) Light-level geolocators reveal covariation between winter plumage molt and phenology in a trans-Saharan migratory bird. *Oecologia* **78**, 1105–1112.

SHERLEY, R.B., ABADI, F., LUDYNIA, K., BARHAM, B.J., CLARK, A.E. & ALTWEGG, R. (2014) Age-specific survival and movement among major African Penguin *Spheniscus demersus* colonies. *Ibis* **156**, 716–728.

SINGLETON, C.L., SAUTHER, M.L., CUOZZO, F.P. & JACKY, I.A.Y. (2018) Age-Related Changes in Hematology and Blood Biochemistry Values in endangered, Wild Ring-Tailed Lemurs (*Lemur Catta*) at the Bezà Mahafaly Special Reserve, Madagascar. *Journal of Zoo and Wildlife Medicine* **49**, 30–47.

SMUTS, G.L., ANDERSON, J.L. & AUSTIN, J.C. (1978) Age determination of the African lion (*Panthera leo*). *Journal of Zoology* **185**, 115–146.

STALMANS, M.E., MASSAD, T.J., PEEL, M.J.S., TARNITA, C.E. & PRINGLE, R.M. (2019) War-induced collapse and asymmetric recovery of large-mammal populations in Gorongosa National Park, Mozambique. *PLoS One* **14**, e0212864.

STARR, M.E. & SAITO, H. (2014) Sepsis in old age: Review of human and animal studies. *Aging and Disease* **5**, 126–136.

TAKKEN BEIJERSBERGEN, L.M. (2019) Determining Age and Season of Death by Use of Incremental Lines in Norwegian Reindeer Tooth Cementum. *Environmental Archaeology* **24**, 49–60.

TARNITA, C.E., ANTAL, T., OHTSUKI, H. & NOWAK, M.A. (2009) Evolutionary dynamics in set structured populations. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 8601–8604.

TAYLOR, B. & KIRWAN, G.M. (2020) Streaky-breasted Flufftail (*Sarothrura boehmi*), version 1.0. In *Birds of the World* (eds J. DEL HOYO, A. ELLIOTT, J. SARGATAL, D.A. CHRISTIE & E. DE JUANA), Cornell Lab of Ornithology, Ithaca, NY, USA.

THORPE, J.P. (1982) The molecular clock hypothesis: biochemical evolution, genetic differentiation and systematics. *Annual review of ecology and systematics* **13**, 139–168.

VÄLI, Ü., MIRSKI, P., SELLIS, U., DAGYS, M. & MACIOROWSKI, G. (2018) Genetic determination of migration strategies in large soaring birds: evidence from hybrid eagles. *Proceedings of the Royal Society B: Biological Sciences* **285**, 20180855.

VERHOEVEN, M.A., LOONSTRA, A.H.J., HOOIJMEIJER, J.C.E.W., MASERO, J.A., PIERSMA, T. & SENNER, N.R. (2018) Generational shift in spring staging site use by a long-distance migratory bird. *Biology Letters* **14**, 20170663.

VOELKER, G., BOWIE, R.C.K. & KLICKA, J. (2013) Gene trees, species trees and Earth history combine to shed light on the evolution of migration in a model avian system. *Molecular Ecology* **22**, 3333–3344.

WILLIAMS, S.M. & LINDELL, C.A. (2018) Nuclear Species in Peruvian Amazonian Mixed-species Flocks are Differentially Attractive to Transient Species and to Each Other. *The Wilson Journal of Ornithology* **130**, 1–9.

ZUCKERKANDL, E. (1960) A comparison of animal hemoglobins by tryptic peptide pattern analysis. *National Acad Sciences* **46**, 85.

ZUCKERKANDL, E. (1987) On the molecular evolutionary clock. *Journal of Molecular Evolution* **26**, 34–46.

ZUCKERKANDL, E. & PAULING, L. (1962) *Molecular Disease, Evolution, and Genic Heterogeneity*. In *Horizons in biochemistry*. (eds. M. Kasha & B. Pullman), pp. 189. Academic Press, New York.

ZUCKERKANDL, E. & PAULING, L. (1965) Evolutionary Divergence and Convergence in Proteins. In *Evolving Genes and Proteins* (eds V. BRYSON & H.J. VOGEL), pp. 97–166. Academic Press, New York.

CHAPTER 2

Literature Review – Biological Clocks and Migration

This chapter represents the literature review for the first biological clock of the study which focuses on the circadian and circannual clocks and comprises three articles. The first article is a systematic review (Le Clercq *et al.*, 2023*a*) that summarizes studies exploring a putative association between polymorphisms in candidate genes, such as *Clock* and *Adcyap1*, in relation to seasonally programmed behaviour like breeding and migration in birds. The review is structured based on taxonomic concepts and further includes a comparative re-appraisal of the available allele data within the context of time-calibrated phylogenies. The second article (Le Clercq *et al.*, 2023*b*) provides further details on the methods used, including information following the preferred reporting items for a systematic review and meta-analyses (PRISMA) and data gathered (Le Clercq, 2023*a*) as part of the review. The third article (Le Clercq *et al.*, 2023*c*) details a custom PYTHON program (Le Clercq, 2023*b*) created to assist in the retrieval of data pertinent to contextualising the literature and performing the comparative analysis, including time trees and divergence times. Supplementary materials and methods are included in Appendix A and B.

Articles:

LE CLERCQ, L.-S., BAZZI, G., CECERE, J.G., GIANFRANCESCHI, L., GROBLER, J.P., KOTZÉ, A., RUBOLINI, D., LIEDVOGEL, M. & DALTON, D.L. (2023*a*) Time trees and clock genes: a systematic review and comparative analysis of contemporary avian migration genetics. *Biological Reviews* **98**, 1051–1080.

LE CLERCQ, L., BAZZI, G., CECERE, J.G., GIANFRANCESCHI, L., GROBLER, J.P., KOTZÉ, A., RUBOLINI, D., LIEDVOGEL, M. & DALTON, D.L. (2023*b*) Birds of a feather flock together: a dataset for Clock and Adcyap1 genes from migration genetics studies. *Scientific Data* **10**, 787.

LE CLERCQ, L., KOTZÉ, A., GROBLER, J.P. & DALTON, D.L. (2023*c*) PAReTT: A Python Package for the Automated Retrieval and Management of Divergence Time Data from the TimeTree Resource for Downstream Analyses. *Journal of molecular evolution* **91**, 502–513.

GitHub Repositories:

LE CLERCQ, L.S. (2023*a*) AvianClocksData: Dataset of Clock and Adcyap1 alleles for Birds. GitHub. Available from: https://github.com/LSLeClercq/AvianClocksData.

LE CLERCQ, L. (2023*b*) PAReTT: Python Automated Retrieval of TimeTree data. GitHub. Available from: https://github.com/LSLeClercq/PAReTT.

BIOLOGICAL REVIEWS
Cambridge Philosophical Society
Biol. Rev. (2023), **98**, pp. 1051–1080.
doi: 10.1111/brv.12943

Time trees and clock genes: a systematic review and comparative analysis of contemporary avian migration genetics

Louis-Stéphane Le Clercq[1,2,*], Gaia Bazzi[3], Jacopo G. Cecere[3], Luca Gianfranceschi[4], Johannes Paul Grobler[2], Antoinette Kotzé[1,2], Diego Rubolini[5,6], Miriam Liedvogel[7,8] and Desiré Lee Dalton[9]

[1]*South African National Biodiversity Institute, P.O. Box 754, Pretoria, 0001, South Africa*
[2]*Department of Genetics, University of the Free State, PO Box 339, Bloemfontein, 9300, South Africa*
[3]*Area Avifauna Migratrice, Istituto Superiore per la Protezione e la Ricerca Ambientale, via Ca' Fornacetta 9, Ozzano Emilia (BO), I-40064, Italy*
[4]*Dipartimento di Bioscienze, Università degli Studi di Milano, via Celoria 26, Milan, I-20133, Italy*
[5]*Dipartimento di Scienze e Politiche Ambientali, Università degli Studi di Milano, via Celoria 26, Milan, I-20133, Italy*
[6]*Istituto di Ricerca sulle Acque, IRSA-CNR, Via del Mulino 19, Brugherio (MB), I-20861, Italy*
[7]*Max Planck Research Group Behavioral Genomics, Max Planck Institute for Evolutionary Biology, Plön, 24306, Germany*
[8]*Institute of Avian Research, An der Vogelwarte 21, Wilhelmshaven, 26386, Germany*
[9]*School of Health and Life Sciences, Teesside University, Middlesbrough, TS1 3BA, UK*

ABSTRACT

Timing is a crucial aspect for survival and reproduction in seasonal environments leading to carefully scheduled annual programs of migration in many species. But what are the exact mechanisms through which birds (class: Aves) can keep track of time, anticipate seasonal changes, and adapt their behaviour? One proposed mechanism regulating annual behaviour is the circadian clock, controlled by a highly conserved set of genes, collectively called 'clock genes' which are well established in controlling the daily rhythmicity of physiology and behaviour. Due to diverse migration patterns observed within and among species, in a seemingly endogenously programmed manner, the field of migration genetics has sought and tested several candidate genes within the clock circuitry that may underlie the observed differences in breeding and migration behaviour. Among others, length polymorphisms within genes such as *Clock* and *Adcyap1* have been hypothesised to play a putative role, although association and fitness studies in various species have yielded mixed results. To contextualise the existing body of data, here we conducted a systematic review of all published studies relating polymorphisms in clock genes to seasonality in a phylogenetically and taxonomically informed manner. This was complemented by a standardised comparative re-analysis of candidate gene polymorphisms of 76 bird species, of which 58 are migrants and 18 are residents, along with population genetics analyses for 40 species with available allele data. We tested genetic diversity estimates, used Mantel tests for spatial genetic analyses, and evaluated relationships between candidate gene allele length and population averages for geographic range (breeding- and non-breeding latitude), migration distance, timing of migration, taxonomic relationships, and divergence times. Our combined analysis provided evidence (*i*) of a putative association between *Clock* gene variation and autumn migration as well as a putative association between *Adcyap1* gene variation and spring migration in migratory species; (*ii*) that these candidate genes are not diagnostic markers to distinguish migratory from sedentary birds; and (*iii*) of correlated variability in both genes with divergence time, potentially reflecting ancestrally inherited genotypes rather than contemporary changes driven by selection. These findings highlight a tentative association between these candidate genes and migration attributes as well as genetic constraints on evolutionary adaptation.

* Author for correspondence (Tel.: +2782 551 5895; E-mail: leclercq.l.s@gmail.com).
Assistant Editor: Alison Cooper, Editor: Shinichi Nakagawa.

Key words: migration, birds, circadian, *Clock*, *Adcyap1*, candidate genes, phylogenetic, time trees, divergence times, ornithology.

CONTENTS

I. INTRODUCTION

Each year, billions of birds take to the sky to make their annual journey from non-breeding to breeding grounds, often migrating at night with several key stop-over sites to refuel and rest along well-established flyways. This event occurs like clockwork, carefully timed to ensure an optimal flight to reach their destination. These annual flight plans are orchestrated to balance day length (Sockman & Hurlbert, 2020), night-time visibility (Brown & Mewaldt, 1968; Pyle *et al.*, 1993; Norevik *et al.*, 2019), and time spent at staging sites (Roques *et al.*, 2022). The repeated occurrence of this characteristic behaviour on an annual basis with seemingly little deviation has sparked interest in the field of chronobiology, with

emphasis on both intrinsic as well as environmental cues contributing to timekeeping in birds (Åkesson *et al.*, 2017).

To maintain their schedules, annual life events such as moult, fattening, migratory behaviour, and breeding are tied to many extrinsic or environmental factors (*Zeitgebers*) including changes in photoperiod (length of sunlight; Leclerc *et al.*, 2010), changing temperatures (Jenni & Kéry, 2003; Pancerasa *et al.*, 2018), and availability of food sources (Hau & Gwinner, 1996; Stephan, 2002; Scheuerlein & Gwinner, 2002). For many migratory bird species, however, it has been demonstrated that they persistently exhibit seasonally appropriate migration-related behaviours, such as migratory restlessness (*Zugunruhe*) and moulting phenology, even when kept under constant conditions in captivity (Newton, 2007; Aguilar-Roblero, Díaz-Muñoz & Fanjul-Moles, 2015). Therefore, a possible intrinsic mechanism of timekeeping may exist that can function in isolation from environmental factors and may be under genetic control, although involvement of an epigenetic effect has also been proposed (Saino *et al.*, 2017; Merlin & Liedvogel, 2019).

One mechanism of intrinsic annual timekeeping is the circadian clock, which regulates daily activity in almost every organism from bacteria to mammals (Aguilar-Roblero, 2015). The circadian clock comprises several genes which can be defined as: '… genes that interact with each other to make up an auto-regulatory feedback loop, in which its activation and repression cycle takes about one day' (Albrecht & Ripperger, 2009, p. 759). Circadian clock genes therefore have a central axis with a positive feedback loop, which promotes transcription, and a negative feedback loop, which prevents transcription (Fig. 1), and the expression levels of these key elements fluctuate throughout the day in different tissues (Albrecht & Ripperger, 2009; Aguilar-Roblero, 2015) in response to signal transduction of light exposure. For example, as the photoperiod changes due to the orbit of the Earth and seasonality, with days becoming shorter/longer and nights becoming longer/shorter (see Figs 2 and 3), the internal phase of the circadian clock adjusts in order to maintain appropriate sleep–wake cycles in a process known as entrainment (Albrecht & Ripperger, 2009; Robart, McGuire & Watts, 2018). Previous observations of length polymorphisms within these genes in several model species (Tauber & Kyriacou, 2005; O'Malley & Banks, 2008) have found evidence of selection for specific variants along a latitudinal gradient. As photoperiod is tied to latitude, this partially explains how environmental changes modulate the circadian clock on a circannual basis possibly driving the timing and duration of migration (see Fig. 3).

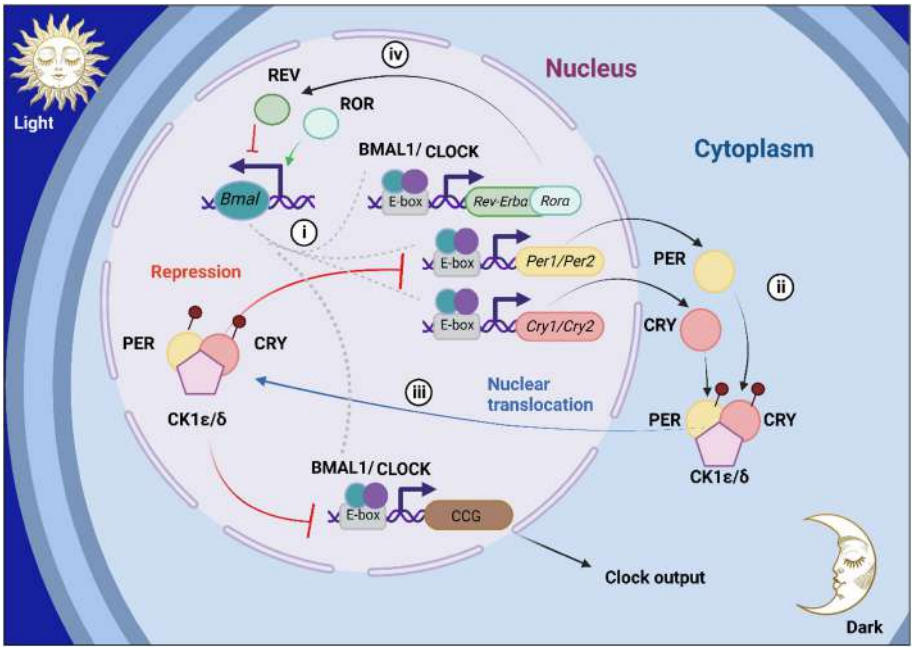


Fig. 1. Diagrammatic representation of the regulation of the circadian clock. (i) Brain and muscle ARNT-like protein 1 (BMAL1, blue circles) forms a dimer with circadian locomotor output cycles protein kaput (CLOCK, purple circles) in the nucleus, which binds to the enhancer box (E-box, grey) region of nuclear receptors (*Rev-Erbα*, green) and retinoic acid-related orphan nuclear receptors (*Rorα*), as well as the *Period 1* or *2* (Per1/2, yellow), *Cryptochrome 1* or *2* (Cry1/2, orange), and other circadian clock genes (CCGs, brown). (ii) When the dimer binds to the *Per1/Per2* and *Cry1/Cry2* E-box, PER and CRY proteins are expressed and, following phosphorylation, form a complex with casein kinase 1 isoform epsilon or delta (CK1ε/δ) in the cytoplasm. (iii) The PER/CRY/CK1ε/δ complex then undergoes nuclear translocation (blue arrow) where it downregulates *Per1/Per2*, *Cry1/Cry2*, and other CCG transcription by inhibiting the binding of BMAL1/CLOCK complexes to E-box regions (red lines). (iv) Concurrently, the binding of BMAL1/CLOCK complexes to the enhancer elements of *Rev-Erbα* and *Rorα* results in the expression of the REV and ROR proteins which act on the transcription elements of *Bmal1*, with REV acting as an inhibitor and ROR acting as an inducer. (Created with BioRender.com).

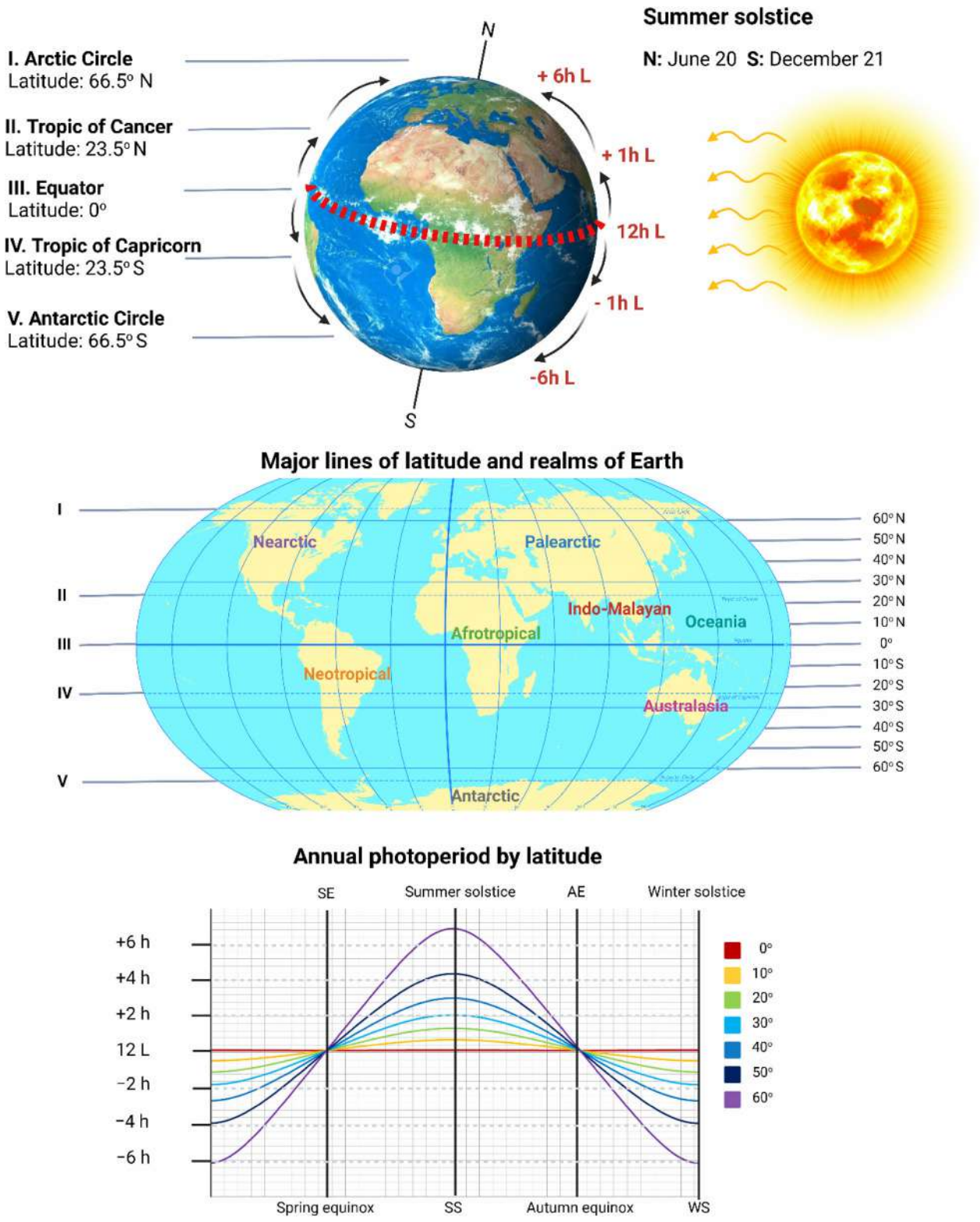


Fig. 2. Diagrammatic representation of the major lines of latitude of Earth and the relative photoperiod at each latitude between the four key dates for changing photoperiods. In the top panel, the orientation of the Earth at the summer solstice is indicated with the difference in day length between the equator (broken line) and the two major tropics and the polar circles. In the centre panel, these lines are indicated on a map, along with the intermediate lines of latitude. In the bottom panel, the annual variation in photoperiod is plotted for each 10° increment in latitude across the year. (Created with BioRender.com).

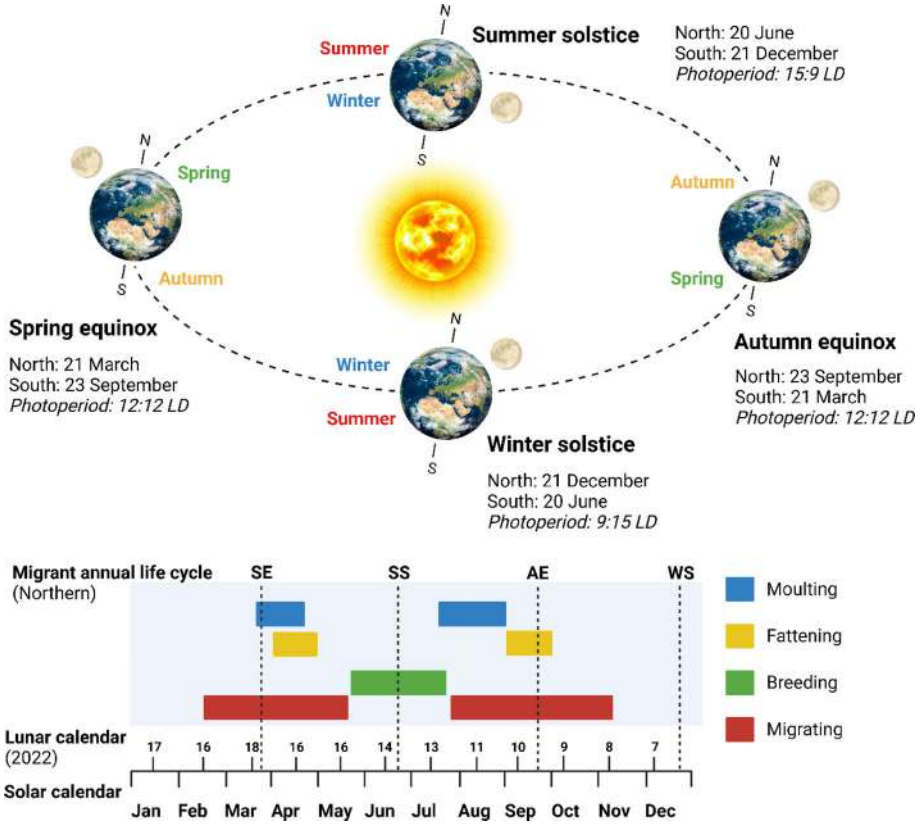


Fig. 3. Diagrammatic illustration of the seasonal cycles of the year indicating the four key dates for changing photoperiods of the year: the summer solstice, the winter solstice, the spring equinox, and the autumn equinox. The equinoxes correspond to near-equal day and night [12 h light:12 h dark; (12:12 LD)] while the summer solstice is the longest day (15:9 LD) and the winter solstice is the shortest day (9:15 LD). The dates of the solstices and equinoxes fall on the exact opposite days in the northern and southern hemispheres each year. The bottom panel indicates the annual timing of major life events during the year for migrants of the northern hemisphere, which encompasses most migratory species in this review. Dates are adapted from the literature and are shown relative to the solar and lunar calendars. (Created with BioRender.com).

One confounding observation for the argument that migration and annual life events are genetically programmed, with birds of the same species expected to have similar genotypes and therefore similar migration strategies resulting from these genotypes, is the occurrence of differential migration patterns within a single species. Differential migration describes the observation that specific populations or groups within a single species follow divergent migratory strategies. These patterns of differential migration include differences in the timing, direction, duration, or distance of migration as well as the occurrence of migration behaviour itself: many birds do not exhibit migratory behaviour but instead form resident populations within their range (Spina & Volponi, 2008, 2009; Billerman *et al.*, 2022). The current view regarding resident populations of bird species is that their ancestors were migratory and that subpopulations adapted to be resident on an individual basis (de Zoeten & Pulido, 2020).

For example, species such as the streaky-breasted flufftail, *Sarothrura boehmi* (Reichenow) (Taylor & Kirwan, 2020), and European bee-eater, *Merops apiaster* (Linnaeus) (Fry & Boesman, 2020), show distinct subpopulations of resident and migratory birds. Migration studies on congeneric spotted eagle hybrids between *Clanga clanga* (Pallas) and *Clanga pomarine* (Brehm) have shown that hybridisation between species with divergent movement patterns has resulted in a new population with migration strategies that differ from their progenitors (Väli *et al.*, 2018). Scandinavian populations of the common chaffinch, *Fringilla coelebs* (Linnaeus) were originally described by Linnaeus as a 'bachelor bird' (Linnaeus, 1758) due to the earlier timing and longer migration of females whereas males appeared nearly resident year-round. By contrast, in most species the males are often the first to depart or migrate further (Dierschke, Mendel & Schmaljohann, 2005; Briedis *et al.*, 2019). This raises the question as to how differential responses to environmental stimuli, and resulting variable migration patterns, are established and maintained between individuals within a species or between closely related subspecies if migration behaviour is a genetically programmed trait.

Several approaches have identified putative polymorphic repeats within clock-regulating genes for use in behaviour

Table 1. Summary of candidate gene studies relating polymorphisms in clock genes to migration-related behaviour. Studies are grouped based on the groupings of species used in Section III. The number of known subspecies (Ssp.) is indicated along with the total alleles and most frequent allele, the Poly-Q allele for *Clock* and *NPAS*, the fragment length for *Adcyap1* and *CREB1*, and the presence of a single nucleotide polymorphism (SNP) for *DRD4*. Studies used latitude/longitude/spatial analyses (a), migratory restlessness (b), timing of egg laying/breeding (c), clutch size (d), timing of migration (e), moult rate (f), urbanisation (g) or exploratory behaviour (h). Symbols against species names indicate migration status: M, migrant; PM, partial migrant; S, sedentary. Species with an asterisk were part of a cross-species shared flyway study that did not include species-level analyses.

Species	Ssp.	Reference	*Clock*	*Adcyap1*	*CREB1*	*NPAS*	*DRD4*	Study method
1(a) Tits:								
Blue tit (*Cyanistes caeruleus*)^{PM}	9	Johnsen *et al.* (2007)	9 (Q_{12})					a
Blue tit (*Cyanistes caeruleus*)^{PM}	9	Liedvogel *et al.* (2009, 2012)	6 (Q_{12})					a, c, d
Blue tit (*Cyanistes caeruleus*)^{PM}	9	Steinmeyer *et al.* (2009)	5 (Q_{12})	7 (162)	7 (548)	5 (Q_{12})		–
Great tit (*Parus major*)^{PM}	15	Liedvogel & Sheldon (2010)	5 (Q_{14})					c
Great tit (*Parus major*)^{PM}	15	Korsten *et al.* (2010)					SNP	h
Great tit (*Parus major*)^{PM}	15	Mueller *et al.* (2013*a*)					SNP	h
1(b) Warblers:								
Blackpoll warbler (*Setophaga striata*)^{M}	1	Ralston *et al.* (2019)	4 (Q_{6})	16 (189)				a, e
Common whitethroat (*Curruca communis*)*,^{M}	4	Bazzi *et al.* (2016*a*)	9 (Q_{10})	13 (172)				a, e
Eastern subalpine warbler (*Curruca cantillans*)*,^{M}	2	Bazzi *et al.* (2016*a*)	4 (Q_{9})	7 (168)				a, e
Eurasian blackcap (*Sylvia atricapilla*)^{PM}	5	Mueller *et al.* (2011)	8 (Q_{8})	13 (161)	10 (532)	2 (Q_{8})	SNP	a, b
Eurasian blackcap (*Sylvia atricapilla*)^{PM}	5	Mettler *et al.* (2015)		11 (161)				e
Eurasian reed warbler (*Acrocephalus scirpaceus*)*,^{M}	4	Bazzi *et al.* (2016*a*)	1 (Q_{11})	10 (169)				a, e
Garden warbler (*Sylvia borin*)*,^{M}	2	Bazzi *et al.* (2016*a*)	6 (Q_{11})	6 (169)				a, e
Great reed warbler (*Acrocephalus arundinaceus*)*,^{M}	2	Bazzi *et al.* (2016*a*)	2 (Q_{12})	2 (163)				a, e
Icterine warbler (*Hippolais icterina*)*,^{M}	1	Bazzi *et al.* (2016*a*)	2 (Q_{8})	7 (169)				a, e
Sedge warbler (*Acrocephalus schoenobaenus*)*,^{M}	1	Bazzi *et al.* (2016*a*)	1 (Q_{11})	5 (163)				a, e
Seychelles warbler (*Acrocephalus sechellensis*)^{PM}	1	Edwards *et al.* (2015)					SNP	h
Willow warbler (*Phylloscopus trochilus*)*,^{M}	3	Bazzi *et al.* (2016*a*)	5 (Q_{9})	10 (174)				a, e
Willow warbler (*Phylloscopus trochilus*)^{M}	3	Bazzi *et al.* (2017)	5 (Q_{9})	10 (174)	4 (529)	5 (Q_{10})		e, f
Wilson's warbler (*Cardellina pusilla*)^{M}	3	Bazzi *et al.* (2016*b*)	2 (Q_{9})	9 (158)				a, e
Wood warbler (*Phylloscopus sibilatrix*)*,^{M}	1	Bazzi *et al.* (2016*a*)	3 (Q_{11})	5 (162)				a, e
1(c) Swallows:								
Barn swallow (*Hirundo rustica*)^{M}	7	Dor *et al.* (2011)	3 (Q_{7})					a, c
Barn swallow (*Hirundo rustica*)^{M}	7	Caprioli *et al.* (2012)	3 (Q_{7})					c
Barn swallow (*Hirundo rustica*)^{M}	7	Bazzi *et al.* (2015)	3 (Q_{7})					e
Chilean swallow (*Tachycineta meyeni*)^{M}	1	Dor *et al.* (2012)	3 (Q_{8})					a, c, d
Mangrove swallow (*Tachycineta albilinea*)^{S}	1	Dor *et al.* (2012)	2 (Q_{8})					a, c, d
Tree swallow (*Tachycineta bicolor*)^{M}	1	Dor *et al.* (2012)	4 (Q_{8})					a, c, d
Tree swallow (*Tachycineta bicolor*)^{M}	1	Bourret & Garant (2015)	4 (Q_{8})	13 (173)	3 (518)	7 (Q_{11})		a, c
Violet-green swallow (*Tachycineta thalassina*)^{M}	2	Dor *et al.* (2012)	4 (Q_{8})					a, c, d

(Continues on next page)

Table 1. *(Cont.)*

Species	Ssp.	Reference	*Clock*	*Adcyap1*	*CREB1*	*NPAS*	*DRD4*	Study method
White-rumped swallow (*Tachycineta leucorrhoa*)[M]	1	Dor *et al.* (2012)	3 (Q_7)					a, c, d
1(d) Larks:								
Asian short-toed lark (*Alaudala cheleensis*)[PM]	4	Zhang *et al.* (2017)	6 (Q_9)					c
1(e) Sparrows, juncos, and buntings:								
Dark-eyed junco (*Junco hyemalis*)[PM]	14	Peterson *et al.* (2013)	7 (Q_{11})	16 (161)				a, b
Yellow-eyed junco (*Junco phaeonotus*)[S]	4	Peterson *et al.* (2013)	5 (Q_{11})	11 (161)				a, b
Song sparrow (*Melospiza melodia*)[M]	25	Posliff (2020)					SNP	a, h
1(f) Cardinals:								
Painted bunting (*Passerina ciris*) [M]	2	Contina *et al.* (2018)	6 (Q_{11})	4 (169)				e
1(g) Flycatchers and chats:								
African stonechat (*Saxicola torquatus*)[S]	16	Justen *et al.* (2022)	6 (Q_{13})					a, e
Bluethroat (*Cyanecula svecica*)[M]	12	Johnsen *et al.* (2007)	7 (Q_{13})					a
Canary island stonechat (*Saxicola dacotiae*)[S]	1	Justen *et al.* (2022)	3 (Q_{14})					a, e
Collared flycatcher (*Ficedula albicollis*)[M]	1	Krist *et al.* (2021)	4 (Q_{12})	6 (182)	9 (534)	4 (Q_{11})		e
Common nightingale (*Luscinia megarhynchos*)[M]	3	Saino *et al.* (2015*a*)	5 (Q_{12})	7 (151)				e
Common nightingale (*Luscinia megarhynchos*)*,[M]	3	Bazzi *et al.* (2016*a*)	5 (Q_{12})	7 (151)				a, e
Common redstart (*Phoenicurus phoenicurus*)*,[M]	2	Bazzi *et al.* (2016*a*)	4 (Q_{14})	13 (169)				a, e
European pied flycatcher (*Ficedula hypoleuca*)[M]	3	Kuhn *et al.* (2013)	5 (Q_{12})					a, c
European pied flycatcher (*Ficedula hypoleuca*)[M]	3	Saino *et al.* (2015*a*)	5 (Q_{12})	11 (180)				e
European pied flycatcher (*Ficedula hypoleuca*)*,[M]	3	Bazzi *et al.* (2016*a*)	5 (Q_{12})	11 (180)				a, e
European stonechat (*Saxicola rubicola*)[M]	2	Justen *et al.* (2022)	7 (Q_{14})					a, e
Northern wheatear (*Oenanthe oenanthe*)*,[M]	4	Bazzi *et al.* (2016*a*)	5 (Q_{14})	6 (167)				a, e
Siberian stonechat (*Saxicola maurus*)[M]	5	Justen *et al.* (2022)	5 (Q_{13})					a, e
Spotted flycatcher (*Muscicapa striata*)*,[M]	7	Bazzi *et al.* (2016*a*)	2 (Q_9)	5 (162)				a, e
Whinchat (*Saxicola rubetra*)[M]	1	Saino *et al.* (2015*a*)	7 (Q_{14})	13 (169)				e
Whinchat (*Saxicola rubetra*)*,[M]	1	Bazzi *et al.* (2016*a*)	7 (Q_{14})	13 (169)				a, e
1(h) Pipits:								
Tree pipit (*Anthus trivialis*)[M]	2	Saino *et al.* (2015*a*)	5 (Q_9)	12 (170)				e
Tree pipit (*Anthus trivialis*)*,[M]	2	Bazzi *et al.* (2016*a*)	5 (Q_9)	12 (170)				a, e
1(i) Thrushes:								
Eurasian blackbird (*Turdus merula*)[PM]	7	Mueller *et al.* (2013*b*)	2 (Q_7)	20 (165)	2 (532)	3 (Q_{10})	SNP	g
Mountain bluebird (*Sialia currucoides*)[M]	1	Sauve *et al.* (2021)		7 (169)			SNP	a
Western bluebird (*Sialia mexicana*)[PM]	6	Sauve *et al.* (2021)		7 (170)			SNP	a
1(j) Shrikes and orioles:								
Eurasian golden oriole (*Oriolus oriolus*)*,[M]	1	Bazzi *et al.* (2016*a*)	2 (Q_6)	7 (163)				a, e
Woodchat shrike (*Lanius senator*)*,[M]	3	Bazzi *et al.* (2016*a*)	3 (Q_6)	8 (176)				a, e
2(a) Buzzards, hawks, and kites:								
Eurasian buzzard (*Buteo buteo*)[S]	6	Chakarov *et al.* (2013)	1 (Q_8)	3 (152)	3 (533)	2 (Q_9)		a, c, e
Northern goshawk (*Accipiter gentilis*)[S]	10	Chakarov *et al.* (2013)	2 (Q_{11})		1 (534)	2 (Q_{11})		a, c, e

(Continues on next page)

Table 1. *(Cont.)*

Species	Ssp.	Reference	*Clock*	*Adcyap1*	*CREB1*	*NPAS*	*DRD4*	Study method
Red kite (*Milvus milvus*)^{S}	2	Chakarov *et al.* (2013)	2 (Q_8)	2 (139)	2 (534)	2 (Q_8)		a, c, e
2(b) Hoopoes:								
Eurasian hoopoe (*Upupa epops*)*,^{M}	7	Bazzi *et al.* (2016*a*)	3 (Q_8)	3 (157)				a, e
2(c) Wrynecks:								
Eurasian wryneck (*Jynx torquilla*)*,^{M}	6	Bazzi *et al.* (2016*a*)	4 (Q_8)	5 (135)				a, e
2(d) Bee-eaters:								
European bee-eater (*Merops apiaster*)*,^{M}	1	Bazzi *et al.* (2016*a*)	1 (Q_4)	6 (163)				a, e
2(e) Nightjars:								
European nightjar (*Caprimulgus europaeus*)*,^{M}	6	Bazzi *et al.* (2016*a*)	2 (Q_8)	7 (154)				a, e
2(f) Doves:								
European turtle dove (*Streptopelia turtur*)*,^{M}	4	Bazzi *et al.* (2016*a*)	2 (Q_7)	5 (150)				a, e
2(g) Shorebirds:								
Bar-tailed godwit (*Limosa lapponica baueri*)^{M}	6	Parody-Merino *et al.* (2019)	6 (Q_9)					a, e
Collared plover (*Charadrius collaris*)^{S}	1	de Almeida Miranda *et al.* (2022)		6 (172)				a
Semipalmated plover (*Charadrius semipalmatus*)^{M}	1	de Almeida Miranda *et al.* (2022)		5 (178)				a
Semipalmated sandpiper (*Calidris pusilla*)^{M}	1	de Almeida Miranda *et al.* (2022)		6 (188)				a
Spotted sandpiper (*Actitis macularius*)^{M}	1	de Almeida Miranda *et al.* (2022)		4 (196)				a
Yellow-legged gull (*Larus michahellis*)^{PM}	2	Romano *et al.* (2018)	2 (Q_5)	4 (162)		1 (Q_7)		b, c
Other:								
Black swan (Anseriformes: Anatidae, *Cygnus atratus*)^{S}	1	Van Dongen *et al.* (2015)					SNP	g, h

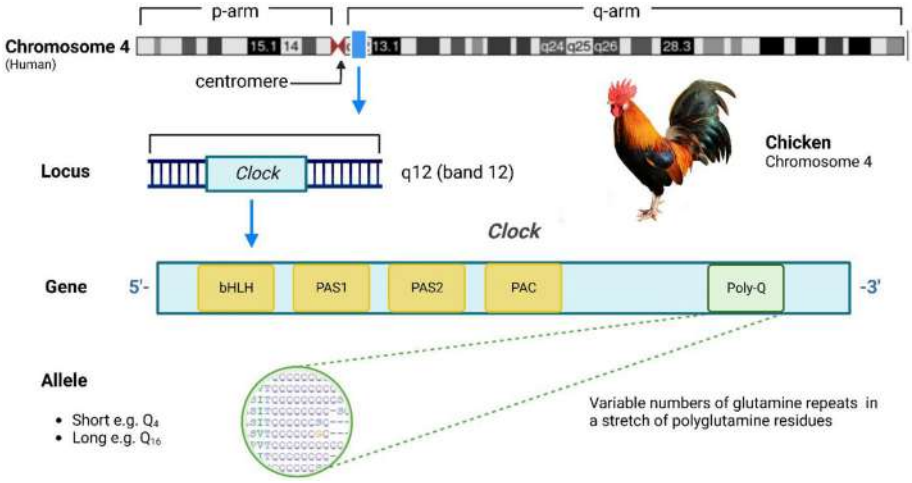


Fig. 4. Depiction of the *Clock* gene (NCBI Gene ID: 9575) and its variable polyglutamine (Poly-Q) repeat region associated with migration phenology. At the top of the figure, the location of *Clock* in the human genome on chromosome 4 at position 12.0 on the q-arm is shown. Below is the gene transcript with its four primary domains in yellow: basic helix–loop–helix (bHLH), period-ah receptor nuclear translocator (ARNT)–single minded protein (PAS1/PAS2), and PAS-associated C-terminal (PAC). The Poly-Q region is indicated in green; this is a region of glutamine (Q) residues that varies in length within and among species. (Created with BioRender.com).

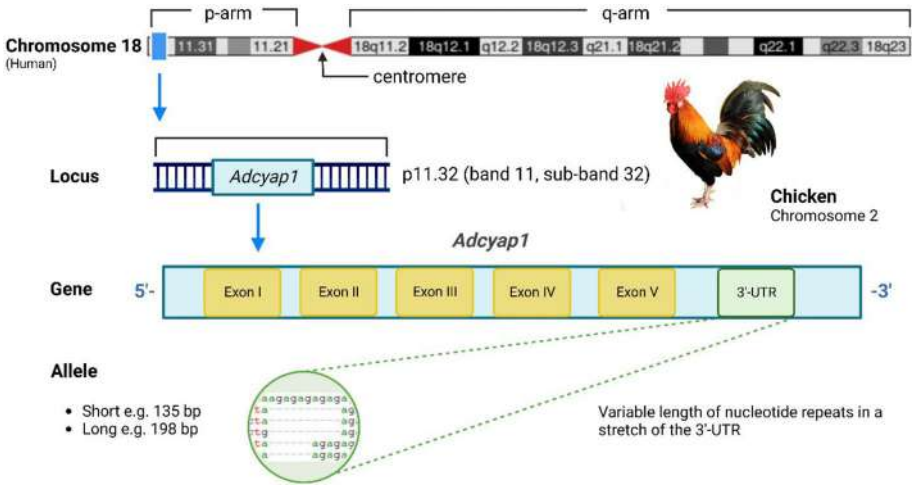


Fig. 5. Depiction of the *Adcyap1* gene (NCBI Gene ID: 116) and its variable 3′-untranslated region (UTR) associated with migration phenology. At the top of the figure, the location of *Adcyap1* in the human genome on chromosome 18 at position 11.32 on the p-arm of the chromosome is shown. Below this is the gene transcript with the five exons indicated in orange. The 3′-UTR is indicated in green; this region contains a stretch of nucleic acid repeats of adenine and guanine that vary in number within and among species. (Created with BioRender.com).

association studies of birds (Steinmeyer, Mueller & Kempenaers, 2009). Steinmeyer *et al.* (2009) described polymorphic genes in the blue tit [recently separated into the Eurasian blue tit, *Cyanistes caeruleus* (Linnaeus), and African blue tit, *Cyanistes teneriffae* (Lesson)]. They also developed methods to assay the identified polymorphisms in Eurasian blackcap *Sylvia atricapilla* (Linnaeus) populations. Several candidate genes (summarised in Table 1) have since been tested, including *Clock*, *Adcyap1*, *CREB1*, *NPAS2*, and *DRD4*. Of these genes, *Clock* and *Adcyap1* have been studied most extensively.

The *Clock* gene (Fig. 4) is located on chromosome 4 at the 12th band of the q-arm in humans and on chromosome 4 in chickens, *Gallus domesticus* (Linnaeus). Towards the 3′-end there is a polyglutamine repeat (Poly-Q) region that can vary in length among and within species, resulting in shorter and longer alleles. For this gene, changes in entrainment resulting from length polymorphisms have been hypothesised to result from changes in protein folding and binding entropies but this has not yet been confirmed (Johnsen *et al.*, 2007). A second gene, *Adcyap1* (Fig. 5), encodes the pituitary-adenylate cyclase-activating polypeptide (PACAP) protein which stimulates the production of melatonin in the pineal gland, thereby conveying light information from the retina to the brain and regulating the circadian rhythm (Simonneaux, Ouichou & Pévet, 1993; Hannibal *et al.*, 1998). *Adcyap1* is located on chromosome 18 of the human genome, at band 11.32 of the p-arm and on chromosome 2 in chickens. The 3′-untranslated region (3′-UTR) contains a homopolymer run of adenine (A) and guanine (G) which can vary in length within and among species, resulting in shorter and longer alleles. For this gene the putative differences in entrainment could be related to altered post-transcriptional regulation of its messenger RNA (mRNA), due to increased instability or metabolic changes (Puga *et al.*, 2005; Steri *et al.*, 2018).

Studies have investigated questions surrounding differential migratory behaviour by comparing migratory attributes (or related breeding phenology) of individuals or populations to putative variation within candidate clock genes (Caprioli *et al.*, 2012; Saino *et al.*, 2015*a*; Delmore *et al.*, 2016). The central hypothesis in these candidate gene studies, including those on birds (Johnsen *et al.*, 2007; Steinmeyer *et al.*, 2009) is that variation, in the form of length polymorphisms within genes associated with the circadian clock, may result in differential responses to environmental cues due to delayed or enhanced entrainment, resulting in differences in migration behaviour which may drive speciation through a migratory divide that establishes selection for specific genotypes based on the photoperiod/latitude of breeding and non-breeding ranges.

A significant association between candidate gene variability and factors contributing to migration patterns, annual synchronicity in life events, or geographical processes has been illustrated in multiple species-specific studies across several lineages within the order Passeriformes, including Palaearctic and Nearctic warblers (Bazzi *et al.*, 2017; Ralston *et al.*, 2019), swallows (Caprioli *et al.*, 2012; Bazzi *et al.*, 2015), tits (Liedvogel *et al.*, 2009), chats (Justen *et al.*, 2022), and flycatchers (Kuhn *et al.*, 2013); whilst associations were not clear or absent for other lineages (Contina *et al.*, 2018; Parody-Merino *et al.*, 2019). Additionally, although some heritability was observed in a study comparing migratory species from several lineages that share the same trans-Saharan migratory flyway (Bazzi *et al.*, 2016*a*), a cross-species comparative study on a subset of candidate genes failed to detect a clear, generalised relationship between clock gene diversity and divergence between migratory *versus* resident bird species (Lugo Ramos, Delmore & Liedvogel, 2017). This defies the expectation that genes that strongly influence and are selected for

in shaping migratory behaviour will be conserved among species that exhibit similar migratory behaviour, although other factors such as size differences (Mettler, Segelbacher & Schaefer, 2015), diet (Stephan, 2002), and habitat preference (Väli *et al.*, 2018) may alter the degree of influence of genetic effects that use only the photoperiod as a *zeitgeber*.

Furthermore, these studies often only considered contemporary ranges and did not consider the role of palaeogeography in shaping ranges throughout the evolutionary history of these species (Meert, 2012; Voelker, Bowie & Klicka, 2013). This is critical for two reasons: firstly, selection is postulated to be stronger during speciation events at the time of divergence and diminish over time (Nosil, Harmon & Seehausen, 2009), and secondly, only parts of the geographic regions within their range represent likely historical ranges before speciation while others represent more recent colonisation events after speciation (Olson *et al.*, 2001; Le *et al.*, 2022). This includes periods where currently separated continents would have been connected, such as Laurasia, the supercontinent that included parts of Africa, North America, and Eurasia, as well as periods of high fragmentation such as the sea formerly separating Europe and Asia as well as the North American inland sea (Western interior seaway) that separated the western and eastern parts of modern North America (Kauffman, 1984). A further case exists for the colonisation of islands including insular India (Prasad & Parmar, 2022), after it separated from Madagascar but before it merged with Asia, as well as islands that formed more recently such as the Cape Verde islands that formed ∼40–50 million years ago (MYA) while the Canary islands formed ∼20 MYA (Schmincke, 1976).

Thus, the current literature on the putative role of these polymorphisms in shaping differential migration among bird species provides conflicting evidence that needs to be clarified to reframe our understanding of migration genetics. Furthermore, the context within which the data are interpreted may need reappraisal due to frequent taxonomic revisions that occur in bird classification and disparities that exist between phylogeny and taxonomy for species concepts (Sangster, 2014), particularly if closely related taxa inherited a specific subset of genotypes that potentially restrict plasticity in future behavioural adaptation and therefore fitness. This is of particular importance as understanding how a biological clock is able to anticipate environmental cues and adapt to them, along with any potential genetic constraints, is of relevance to conservation given the potential effects of climate change and habitat erosion on migratory behaviour (Carey, 2009).

The aim of this review is to (*i*) systematically synthesise and review the existing literature on polymorphisms in clock genes, primarily within the context of migration, breeding, and annual life events, to identify existing patterns, (*ii*) perform a comparative analysis of the existing data to test for an association between shared clock gene polymorphisms and similarities in attributes of migration such as latitude, distance, and timing, and (*iii*) contextualise the resulting evidence in a taxonomically, phylogenetically, and palaeogeographically informed manner to compare similarities within and among lineages with shared evolutionary histories.

II. METHODS

(1) Literature search and systematic review

A systematic approach was used to search for and synthesise the available literature. Literature was searched on the *Scopus* (www.scopus.com) and *Dimensions* (www.dimensions.ai) databases using the following Boolean search string: ('Clock genes' OR 'Clock' OR 'Adcyap1') AND ('Birds' OR 'Avian') AND ('Migration' OR 'Flying'). Search results were exported in the comma separated value format and the literature was subsequently summarised, guided by citation networks visualised using *CitNetExplorer* 1.0.0 and *VOSviewer* 1.6.16 (van Eck & Waltman, 2017). The results retrieved from *Scopus* were converted into the appropriate format using the R package *Scopus2CitNet* 0.1.0.0 in *RStudio* 1.4.1106 (RStudio Team, 2021), running R version 4.0.5 (R Core Team, 2020). Due to the diverse array of species in which these studies were conducted, the literature was organised by year of publication followed by species, with taxonomic grouping based firstly on order (passerine or non-passerine), followed by family. Families were grouped, based on higher taxonomic classifications, into superfamilies and parvorders for the sake of a concise and cohesive comparison (see Section III); warblers, which is a paraphyletic group, are discussed together as they share significant overlaps in morphology.

(2) Species

Species for comparative analysis were selected based on the existing literature for either the *Clock* or *Adcyap1* gene in relation to migration phenology and/or for which genomic or transcriptomic studies have been conducted. This included unpublished data from eight species: American redstart, *Setophaga ruticilla* (Linnaeus); common chiffchaff, *Phylloscopus collybita* (Vieillot); common yellowthroat, *Geothlypis trichas* (Linnaeus); hermit thrush, *Catharus guttatus* (Pallas), magnolia warbler, *Setophaga magnolia* (Wilson); Swainson's thrush, *Catharus ustulatus* (Nuttall), and white-throated sparrow, *Zonotrichia albicollis* (Gmelin). As most studies thus far have focused on Palearctic and Nearctic birds, species were further complemented with data (see Section II.3) for migrant and resident bird species from other locations to have a globally distributed data set. This included the addition of several species of manakins, resident birds found in equatorial regions of the Neotropics, the endangered Elfin woods warbler, *Setophaga angelae* (Kepler & Parkes), endemic to Puerto Rico, and the Australasian superb fairy-wren, *Malurus cyaneus* (Ellis). Our final species list included 76 species (76 for *Clock* and 71 for *Adcyap1*) of which 58 were classified as migrants and 18 were classified as residents. Migrants were defined

as species with complete or partial migratory behaviour (i.e. species with both resident and migratory populations); species with a single resident population that was not sampled were treated as migratory. Residents were defined as species that do not follow an annual cycle of migration (although they may follow a pattern of altitudinal migration within their resident range); species with a single migratory population that was not sampled were treated as resident.

Inclusion in subsequent analyses was based on data availability in two categories: (*i*) the availability of genetic data as either within-population-level data on individual alleles for *Clock* or *Adcyap1* or records of the most common allele for the species from the literature or online databases, as well as species availability for the phylogenetic tree from the Bird Tree website (see Section II.3); (*ii*) the availability of migration data including migration dates for each migratory season as well as shapefile data for GIS to compute different parameters related to migration (see Section II.4). As far as possible, species were only included if they share one of the major migratory flyways with more than one other included species and therefore experience similar *Zeitgebers* for annual synchronicity.

(3) Genetic data

Population-level allele data were retrieved for 40 species (39 for *Clock* and 37 for *Adcyap1*) from either the supplementary material of the article or the online data repositories *Dryad* (www.datadryad.org), *Figshare* (www.figshare.com), or directly from the article authors (see Section VII, Acknowledgements). *Clock* data were transformed to represent the actual number of poly-Q repeats, as different studies used different primers resulting in variable lengths in the raw data. *Adcyap1* was consistently amplified and sequenced using the same primer set or region, facilitating between study comparisons. Sequence data were not available for all species, so the total reported allele length of *Adcyap1* was used rather than the length of the AG-repeat region alone. However, this will not alter our central hypothesis that any length difference could result in altered gene regulation resulting in different migration-related traits.

Data summarised from the literature included species, number of alleles, most common allele, and observed heterozygosity when available. Additional information such as the number of extant and presently recognised subspecies was retrieved from *Birds of the World* (Billerman *et al.*, 2020). *Clock* and *Adcyap1* data for additional species (see Section II.2), from a wider geographic distribution and including additional resident birds, was retrieved from the National Centre for Biotechnology Information (NCBI) website using Basic Local Alignment and Search Tool (BLAST) searches (Altschul *et al.*, 1990) against reference genomes and available databases, including *PopSet* and *Nucleotide* (Agarwala *et al.*, 2018); where no sequence data were available a further BLAST search was carried out against the Sequence Read Archive (SRA) for specific species. The tree used for the phylogeny was generated from the 'Ericson' phylogeny (Jetz *et al.*, 2012) by sampling 5000 trees from the Bird Tree website (www.birdtree.org) for available species from the list of 76. The trees were summarised to a 60% consensus tree by maximum clade credibility using *TreeAnnotator* 2.6.3, part of BEAST 2.6.3 (Bouckaert *et al.*, 2014), with a 10% burn-in.

(4) Migration and range data estimates

Migration data were computed using *QGIS* 3.16.15 (QGIS Development Team, 2022) from shapefiles extracted from the geodatabase of distribution maps compiled by BirdLife International, version 2021.1 (BirdLife International & Handbook of the Birds of the World, 2021), supplemented with shapefile data from eBird (Fink *et al.*, 2021). Centroids were computed for the non-breeding and breeding ranges to determine the average coordinates by latitude and longitude (in degrees) for each. The average migration distance between these centroids was calculated (in metres). Geographic distance matrices for the breeding and non-breeding coordinates of each species were generated with the java application *Geographic Distance Matrix Generator* 1.2.3 (Ersts, 2012) in degrees. Species classified as partial migrants, for which breeding and non-breeding range data were available, were treated as migrants, while data-deficient partial migrants were treated as resident birds. For resident species, the calculations were based on their full range.

When population and species-specific migration dates were not stated in the source publications, this information was retrieved from eBird (Fink *et al.*, 2021) or, for species sampled in Italy, from the Italian bird migration atlas (Spina & Volponi, 2008, 2009) or, for Asian buntings and larks, from Ali, Ripley & Roberts (1999). As seasons and migration dates vary according to hemisphere, dates were normalised to a standard reference point (Sockman & Hurlbert, 2020) that roughly corresponds to the photoperiod and temperature (see Figs 2 and 3). The difference in days was calculated between the start, middle and end dates for spring migration and the summer solstice [approximately 15 h light:9 h dark (15:9 LD)] and spring equinox (12:12 LD), while autumn migration dates were normalised with reference to the winter solstice (∼9:15 LD) and autumn equinox (12:12 LD), for each respective hemisphere depending on the species range.

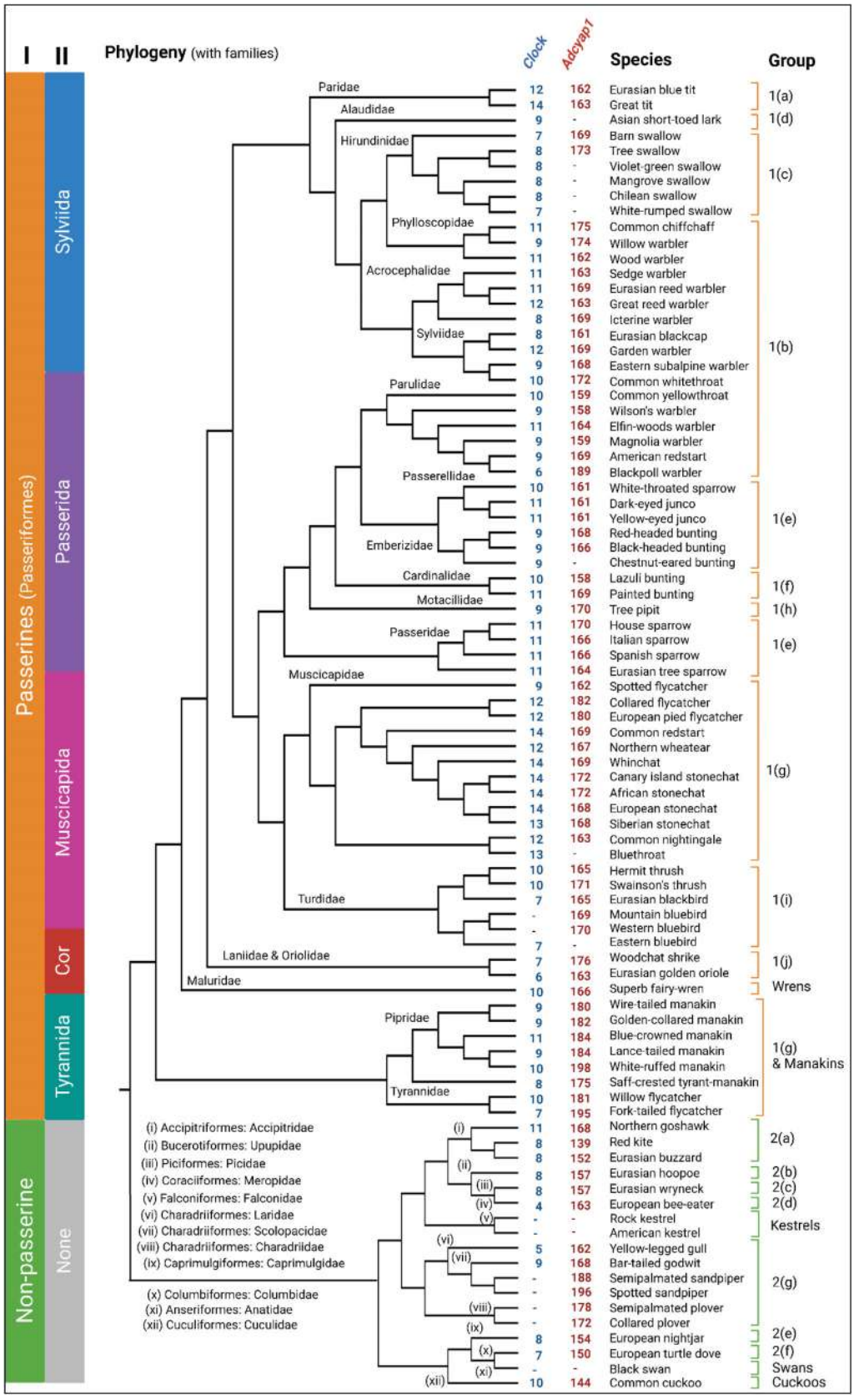
(5) Population genetics

Population genetics analyses were done for the 40 species for which individual allele data was available. POPGENE 1.32 (Yeh *et al.*, 1997) was used to test for Hardy–Weinberg (Hardy, 1908; Weinberg, 1908) equilibrium using chi-squared (χ^2) tests (with significance measured at $\alpha = 0.02$), to calculate the observed (H_o) and expected (H_e) heterozygosity, and to create a genetic distance matrix using fixation index (F_{ST}) values. Python for Population Genetics (*PyPop*) version 0.7.0 (Lancaster *et al.*, 2007) was used to test for selection and neutrality as well as linkage disequilibrium. Neutrality was assessed using Slatkin's implementation

Table 2. Summary of studies that used non-candidate gene approaches to identify genetic regions in birds that either co-vary with migration phenology or are expressed differentially on a circannual basis. Ssp. lists the number of known subspecies. Study methods were transcriptomics (a), genomics (b), or epigenetics (c). Letters against species names indicate migration status: M, migrant; PM, partial migrant; S, sedentary.

Species	Ssp.	Reference	Study method
1(a) Tits:			
Great tit (*Parus major*)^{PM}	15	Laine *et al.* (2019)	a
Great tit (*Parus major*)^{PM}	15	Mäkinen *et al.* (2019)	c
Great tit (*Parus major*)^{PM}	15	Viitaniemi *et al.* (2019)	c
1(b) Warblers:			
Eurasian blackcap (*Sylvia atricapilla*)^{PM}	5	Delmore *et al.* (2020*a*,*b*)	b
Willow warbler (*Phylloscopus trochilus*)^{M}	3	Lundberg *et al.* (2013, 2017)	a, b
Willow warbler (*Phylloscopus trochilus*)^{M}	3	Boss *et al.* (2015)	a
Wilson's warbler (*Cardellina pusilla*)^{M}	3	Ruegg *et al.* (2014*b*)	b
1(c) Swallows:			
Barn swallow (*Hirundo rustica*)^{M}	7	Saino *et al.* (2017)	c
Tree swallows (*Tachycineta bicolor*)^{M}	1	Brown (2019)	a
1(e) Sparrows, juncos, and buntings:			
Black-headed bunting (*Emberiza melanocephala*)^{M}	1	Singh *et al.* (2015)	a
Black-headed bunting (*Emberiza melanocephala*)^{M}	1	Mishra *et al.* (2017)	a
Black-headed bunting (*Emberiza melanocephala*)^{M}	1	Sharma *et al.* (2018*a*)	a
House sparrow (*Passer domesticus*)^{PM}	14	Guldvog (2015)	a
Italian sparrow (*Passer italiae*)^{S}	1	Guldvog (2015)	a
Red-headed bunting (*Emberiza bruniceps*)^{M}	1	Singh *et al.* (2013)	a
Red-headed bunting (*Emberiza bruniceps*)^{M}	1	Sharma *et al.* (2018*b*)	a, c
Red-headed bunting (*Emberiza bruniceps*)^{M}	1	Trivedi *et al.* (2019)	a
Spanish sparrow (*Passer hispaniolensis*)^{PM}	2	Guldvog (2015)	a
White-crowned sparrow (*Zonotrichia leucophrys*)^{PM}	5	Jones *et al.* (2008*a*,*b*)	a
1(f) Cardinals:			
Painted bunting (*Passerina ciris*)^{M}	2	Contina *et al.* (2019)	b
1(g) Flycatchers and chats:			
Northern wheatear (*Oenanthe oenanthe*)^{M}	4	Frias-Soler *et al.* (2020, 2021)	a
1(i) Thrushes:			
Bicknell's thrush (*Catharus bicknelli*)^{M}	1	Voelker *et al.* (2013)	b
Black-billed nightingale-thrush (*Catharus gracilirostris*)^{S}	2	Voelker *et al.* (2013)	b
Black-headed nightingale-thrush (*Catharus mexicanus*)^{S}	3	Voelker *et al.* (2013)	b
Eurasian blackbird (*Turdus merula*)^{PM}	7	Franchini *et al.* (2017)	a
Gray-cheeked thrush (*Catharus minimus*)^{M}	2	Voelker *et al.* (2013)	b
Hermit thrush (*Catharus guttatus*)^{PM}	9	Voelker *et al.* (2013)	b
Orange-billed nightingale-thrush (*Catharus aurantiirostris*)^{PM}	14	Voelker *et al.* (2013)	b
Ruddy-capped nightingale-thrush (*Catharus frantzii*)^{S}	7	Voelker *et al.* (2013)	b
Russet nightingale-thrush (*Catharus occidentalis*)^{S}	4	Voelker *et al.* (2013)	b
Slaty-backed nightingale-thrush (*Catharus fuscater*)^{S}	7	Voelker *et al.* (2013)	b
Swainson's thrush (*Catharus ustulatus*)^{M}	6	Voelker *et al.* (2013)	b
Swainson's thrush (*Catharus ustulatus*)^{M}	6	Ruegg *et al.* (2014*a*)	b
Swainson's thrush (*Catharus ustulatus*)^{M}	6	Delmore *et al.* (2015, 2016)	b
Swainson's thrush (*Catharus ustulatus*)^{M}	6	Johnston *et al.* (2016)	b
Veery (*Catharus fuscescens*)^{M}	4	Voelker *et al.* (2013)	b
Wood thrush (*Hylocichla mustelina*)^{M}	1	Voelker *et al.* (2013)	b
Yellow-throated nightingale-thrush (*Catharus dryas*)^{S}	2	Voelker *et al.* (2013)	b
Other:			
Passerines			
Chestnut-crowned babbler (Pomatostomidae, *Pomatostomus ruficeps*)^{S}	1	Liebl *et al.* (2021)	c
Gray catbird (Mimidae, *Dumetella carolinensis*)^{M}	1	DeMoranville *et al.* (2019)	a
Non-passerines			
American kestrel (Falconiformes: Falconidae, *Falco sparverius*)^{PM}	17	Bossu *et al.* (2022)	b

(Slatkin, 1994) of the Ewens–Watterson (Ewens, 1972; Watterson, 1977) test, with the probability values expressed as the relative degree at which the observed *F*-value occurs in a sample distribution for a simulation run with 10,000 repeats. Linkage disequilibrium was assessed using two measures: the overall linkage disequilibrium, D' (Hedrick, 1987), and Cramer's V Statistic, W_n (Cramer, 1946). $P < 0.05$ is indicative of significant linkage disequilibrium.



(Figure 6 legend continues on next page.)

(6) Mantel tests

Mantel tests (Mantel, 1967) were conducted on the 40 species for which allele data was available using the *Mantel* 2.1.0 (Carr, 2021) package in Python 3.9 (Python Team, 2021). Tests compared the genetic distance within two candidate genes and attributes of migration including the distance between latitudes of breeding and non-breeding ranges among species, as well as the relationship between the genetic distance and taxonomic distance and divergence times as measures of evolutionary distance, to assess the strength of heritability of genotypes within lineages.

Test were run between the genetic distance matrices, generated with *CONVERT* 1.31, the geographic distance matrices, generated with *Geographic Distance Matrix Generator* 1.2.3 (for both breeding and non-breeding coordinates), the taxonomic distance matrices, generated using the R package *vegan* 2.5-7 (Oksanen *et al.*, 2020), and the divergence times matrix. Divergence times between pairs of species were retrieved from the Time Tree resource (Kumar *et al.*, 2017) website (www.timetree.org) using Python Automated Retrieval of Time Trees (*PAReTT* version 1.0.1) and exported as a vectorised matrix. $P < 0.02$ and $\tilde{Z} > 1.96$ (or < -1.96) is considered significant.

(7) Phylogenetic generalised least squares analysis

Phylogenetic generalised least square (PGLS) models were fitted for the full list of 76 species for which the most common allele and migration data were available. Each test was run independently using the R package *caper* 1.0.1 (Orme *et al.*, 2018) to avoid error from repeat sampling. PGLS was used to relate both *Clock* and *Adcyap1* length to breeding and non-breeding latitude (as distance from the equator in degrees), as well as to total migration distance between regions and to the normalised dates for the beginning, middle, and end of spring and autumn migration, assuming Brownian motion and a lambda (λ) = 1.0. The phylogenetic tree used as input for PGLS was retrieved from the Bird Tree website (www.birdtree.org). The phylogenetic signal for each gene was measured using the R package *phytools* 0.7-90 (Revell, 2012) to compute both lambda and kappa for the gene and tree data and verify the presence of Brownian motion.

(8) Time trees and palaeogeography

Trees were downloaded from Bird Tree website (www.birdtree.org) as described in Section II.3. Time trees were computed from calibrated divergence time estimates using the Time Tree resource (Kumar *et al.*, 2017) to visualise the evolutionary history and relatedness of study species in terms of shared common ancestry and the length of time individual lineages have been evolving independently. The relevant topography of Earth for each time period was reconstructed in *GPlates* 2.3.0 (Müller *et al.*, 2018) with the PALEOMAP paleoAtlas (Scotese, 2016). This was used to visualise relevant barriers to gene flow, and potential differences in selective forces between modern and historical geography for each time period that likely would have contributed to selection and speciation across the genomes of the study species.

III. SYSTEMATIC REVIEW OF PUBLISHED STUDIES

The results of studies assaying candidate gene polymorphisms and comparing them to relevant attributes of seasonal phenology are summarised in Table 1. Non-candidate-gene studies, conducted at the genome, epigenome, or transcriptome level, are summarised in Table 2; such studies have frequently complemented candidate gene studies in terms of species coverage. Because many biological traits among birds are highly heritable within lineages (Silva *et al.*, 2017; Lamichhaney *et al.*, 2018; Cava, Perlut & Travis, 2019), our systematic review is structured according to taxonomic group (Fig. 6). The estimated divergence times of families and species for the most pertinent lineages are illustrated in Fig. 7 with the non-passerine orders as an outgroup. Two studies, included in the following comprehensive review, represented cross-species analyses. Bazzi *et al.* (2016*a*), reported individual species clock genes and diversity measures that are highlighted below, however, this study was a cross-species analysis of 23 species that share a single trans-Saharan migratory flyway and did not provide detailed within-species analyses. Lugo Ramos *et al.* (2017) explored general patterns within clock genes to delineate resident and migratory species. Primary findings from these studies are discussed where applicable below.

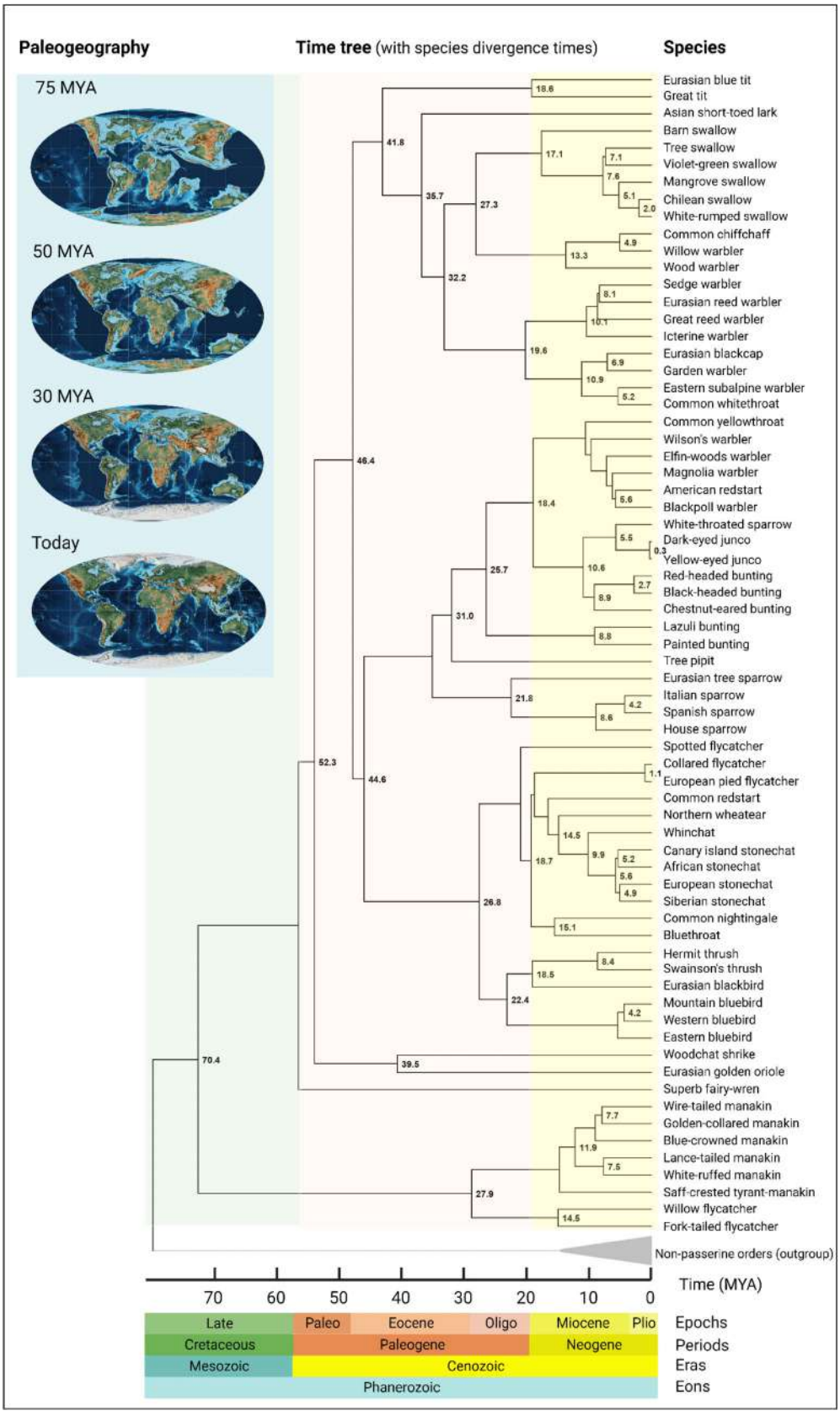
(1) Passerine birds

(*a*) *Tits (Family: Paridae)*

Johnsen *et al.* (2007) were the first to investigate if polymorphic differences in *Clock* exist within avian populations based on evolutionary and functional studies of allelic variants of

(Figure legend continued from previous page.)

Fig. 6. Phylogenetic tree of species in class Aves for which clock genes have been assayed in previous studies, as well as the additional species included in the phylogenetic generalised least squares (PGLS) analysis. This tree was used for both the taxonomic and phylogenetic groupings for comparing studies in the review section, as well as for the input tree used in the PGLS analysis. The tree shows the major taxonomic divisions among species by order (I) and parvorder (II). Families are indicated at the relevant clades of the tree while the main groupings used in Section III are indicated to the right. The most common *Clock* allele (number of Poly-Q repeats) and the length of the most common *Adcyap1* allele (in base pairs) is indicated for each species. Cor, Corvida. (Created with BioRender.com).



(Figure 7 legend continues on next page.)

fruit fly (*Drosophila melanogaster*) circadian clock genes (Tauber & Kyriacou, 2005). They characterised *Clock* variation in Poly-Q regions in the context of breeding phenology using 14 blue tit populations, including subspecies, across a latitudinal gradient. They found evidence for significant relationships between latitudinal clines in habitat and *Clock* lengths among 13 of the studied populations, as well as significant levels of allelic differentiation across subspecies. The most common allele was Q_{12} which was found in at least 44% of genotypes, and in all individuals in the Italian subpopulation. They concluded that, at least in the blue tit populations studied, longer *Clock* alleles were correlated with greater clines in habitat, partially distinguished subspecies, and revealed genetic processes not associated with other microsatellite markers (Johnsen *et al.*, 2007). Studies on a single blue tit population in Oxfordshire woods in the UK with records of timing of breeding (Liedvogel *et al.*, 2009) and clutch size (Liedvogel, Cornwallis & Sheldon, 2012) across a two-year period found that shorter *Clock* alleles corresponded to earlier breeding, shortened incubation period, and higher fledging rates among females; allele frequencies were homogenously distributed. This supported the conclusion that *Clock* polymorphisms can be linked to adaptation to local environments, which may explain population variation in this allele in blue tits. Similar phenotypic correlates were not present in the sympatric great tit, *Parus major* (Linnaeus), suggesting that correlations between behaviour and *Clock* polymorphisms may not be universal to all passerines (Liedvogel & Sheldon, 2010), although studies on circadian period length within this species have identified heritable differences in their migratory behaviour (Helm & Visser, 2010).

(b) *Warblers (Families: Acrocephalidae, Parulidae, Phylloscopidae, Sylviidae)*

(i) 'Old World warblers'. In the family Sylvidae, *Clock*, *Adcyap1*, *NPAS2*, and *CREB1* were studied in several populations (representing most recognised subspecies) of Eurasian blackcap (Mueller, Pulido & Kempenaers, 2011) including more recent radiations to the Cape Verde islands, in a comparison of length polymorphisms to individual calculated migratory distance. A fifth gene, *DRD4*, which has been linked to 'exploratory behaviour' in great tit populations (Korsten *et al.*, 2010; Mueller *et al.*, 2013*a*) was also analysed. Only *Adcyap1* showed a significant relationship, with longer alleles positively associated with a higher breeding latitude and shorter alleles with a higher degree of migratory restlessness. A study on the Seychelles warbler, *Acrocephalus sechellensis* (Oustalet) (Acrocephalidae) also failed to establish a role for *DRD4* (Edwards *et al.*, 2015).

Bazzi *et al.* (2016*a*) analysed clock genes in common whitethroat, *Curruca* (*Sylvia*) *communis* (Latham), eastern subalpine warbler, *Curruca* (*Sylvia*) *cantillans* (Pallas), and garden warbler, *Sylvia borin* (Boddaert), identifying significant heterozygosity for *Clock* (0.68–0.76) and *Adcyap1* (0.66–0.88).

In the Acrocephalidae, Bazzi *et al.* (2016*a*) studied four species: Eurasian reed warbler, *Acrocephalus scirpaceus* (Hermann), great reed warbler, *Acrocephalus arundinaceus* (Linnaeus), Icterine warbler, *Hippolais icterina* (Vieillot) and sedge warbler, *Acrocephalus schoenobaenus* (Linnaeus). Low diversity was observed within the *Clock* gene in this family with nearly all individuals homozygous for the same allele, while heterozygosity for the *Adcyap1* allele varied from 0.18 to 0.81. Overall, the greatest diversity was observed for great reed warbler which has two known subspecies. No indication was given as to whether sampling included these known subspecies but given the single sampling site on a Mediterranean island it is possible that both subspecies were not included.

In the Phylloscopidae, Bazzi *et al.* (2016*a*) studied two species, willow warbler, *Phylloscopus trochilus* (Linnaeus) and wood warbler, *Phylloscopus sibilatrix* (Bechstein) and found greater *Clock* allele diversity as compared to Acrocephalidae with heterozygosity scores of 0.63 and 0.43 respectively and a similar trend for *Adcyap1*. A later study on willow warbler (Bazzi *et al.*, 2017) identified a high degree of homozygosity for the Q_9 allele, which was present in 98% of all individuals, but no significant relationship was found in relation to migration phenology. *NPAS2* and *CREB1* were also assayed; *NPAS2* positively predicted the date of migration while *CREB1* positively predicted moult speed; both effects were sex-specific and only present in males (Bazzi *et al.*, 2017). In a non-candidate gene approach, mRNA expression levels were measured in willow warblers using transcriptomics to elucidate potential gene candidates (Lundberg *et al.*, 2013; Boss *et al.*, 2015) and whole-genome sequencing (Lundberg *et al.*, 2017). These studies identified distinct haplotype distributions between migratory phenotypes despite low allele variability; with only a fraction of the tested regions showing measurable variations.

(ii) 'New World warblers'. For the Parulidae, data are available for Wilson's warbler, *Cardellina pusilla* (Wilson) (Bazzi *et al.*, 2016*b*) and blackpoll warbler, *Setophaga striata* (Forster) (Ralston *et al.*, 2019). Wilson's warbler showed a high degree

(Figure legend continued from previous page.)

Fig. 7. Time tree of study species indicating the relative divergence times (millions of years ago, MYA), of the major lineages for the species discussed or analysed in this review. The tree was rooted and calibrated with non-passerine orders as an outgroup and therefore reflects below order-level divergence times of the more closely related passerine families, genera, and species. Colour coding is according to the geological periods given below the figure. The inset panel on the top left shows the corresponding palaeogeography from before the continents were fully assembled (75 MYA) to when most continents had assumed their contemporary shape, approximately 30 MYA, and finally assumed their modern-day positions in terms of latitude and longitude (today). (Created with BioRender.com).

of homozygosity for the Q_9 allele and there was no correlation with migration phenology. By contrast, Ralston *et al.* (2019) found a correlation between spring and autumn migration and allele length diversity for both *Clock* and *Adcyap1*.

In a non-candidate gene approach, a panel of single nucleotide polymorphisms (SNPs) in Wilson's warblers was used to assess potential genomic regions driving differences in geographic distribution, with several genomic regions showing evidence for differentiation including many genes related to circadian circuitry (Ruegg *et al.*, 2014*b*).

(c) *Swallows (Family: Hirundinidae)*

Dor *et al.* (2011) analysed the *Clock* Poly-Q region in five populations (including three subspecies) of barn swallow, *Hirundo rustica* (Linnaeus), and found low levels of variability despite evident population structure. Nearly 98% of all allelic diversity was accounted for by the Q_7 allele (Dor *et al.*, 2011). The authors suggested that social cues may take precedence over other environmental cues for migration which could result in negative selection for *Clock* polymorphism. Later studies on European populations of this species revealed a similar pattern, with the Q_7 allele present in 96% of the tested genotypes in Italy (Caprioli *et al.*, 2012) and 91% in Switzerland (Bazzi *et al.*, 2015). Caprioli *et al.* (2012) found an association between *Clock* alleles and breeding phenology in females, while rarer genotypes occurred in males, and Bazzi *et al.* (2015) reported correlations with departure and arrival times. Correlations between genotype and timing of moulting have also been identified (Saino *et al.*, 2013), and annual timing of moulting is known to co-vary with migratory phenology (Saino *et al.*, 2015*b*). The apparently low genetic variability may be due to these genes currently being under directional selection, with this species showing changes in migratory phenology in response to climate change (Altwegg *et al.*, 2012).

Dor *et al.* (2012) studied *Clock* length polymorphisms in five species of *Tachycineta* swallows (Table 1). Tree swallow, *T. bicolor* (Vieillot), violet-green swallow, *T. thalassina* (Swainson) and mangrove swallow, *T. albilinea* (Lawrence) displayed a characteristically higher proportion of the Q_8 allele (64%), whilst white-rumped swallow, *T. leucorrhoa* (Vieillot) had a higher prevalence of the Q_7 allele. Chilean swallow *T. meyeni* (Cabanis) had a near equal distribution of both the Q_7 and Q_8 alleles (Dor *et al.*, 2012). Although the study species came from a wide range of breeding latitudes, no significant correlation was found between breeding latitude and *Clock* mean allele length, although there was a small effect in some species on breeding/laying time in females (Dor *et al.*, 2012). A major limiting factor in this study was that each species was sampled from only one location, meaning that allele diversity and evidence for selection within a species could not be assessed accurately. Bourret & Garant (2015) evaluated polymorphism in four genes in tree swallow in relation to two phenological traits related to migration: laying date and incubation duration. For the *Clock* gene, the Q_8 allele was the most abundant (61.6%). *Adcyap1* was found to be highly polymorphic with 13 alleles (as in Eurasian blackcaps, see Table 1). Significant correlations were observed for all genes measured with environmental variables such as latitude, temperature, and breeding density; in particular, longer alleles were associated with more eastern breeding latitudes (Bourret & Garant, 2015). In a transcriptomics study, several regions were identified that putatively co-varied with migration phenology in tree swallows, although none corresponded to previously identified candidate genes (Brown, 2019).

(d) *Larks (Family: Alaudidae)*

In Asian short-toed lark, *Alaudala* (*Calandrella*) *cheleensis* (Swinhoe), a near-equal distribution of the *Clock* Q_9 and Q_{11} alleles was reported. There was a significant correlation with the timing of egg laying and the seasonal endocrine response that initiates breeding (Zhang *et al.*, 2017). Note that this species is currently under taxonomic review and is not universally recognised as separate from lesser short-toed lark (de Juana & Suárez, 2020; BirdLife International, 2021; HBW and BirdLife International, 2021), although speciation may have occurred during the period that Europe and Asia were separated by a body of water.

(e) *Sparrows, juncos, and buntings (Families: Passerellidae, Emberizidae)*

The Italian sparrow, *Passer italiae* (Vieillot) is a hybrid of the house sparrow *Passer domesticus* (Linnaeus) and Spanish sparrow *Passer hispaniolensis* (Temminck). Despite likely differences in migration behaviour (other closely related sympatric sparrow species are known to show different migratory phenology; Borowske, Gjerdrum & Elphick, 2017), all three species were found to be homozygous for the Q_{11} *Clock* allele (Guldvog, 2015). Higher diversity without evident patterning, but able to partition migrating and sedentary populations, was observed for *Adcyap1*. In white-crowned sparrow, *Zonotrichia leucophrys* (Forster), using transcriptomic approaches (Jones *et al.*, 2008*b*), significant variation was reported in the expression of one clock-related gene, *Copine 4* (*CPNE4*), which has also been observed in other bird species (Ruegg *et al.*, 2014*b*; Delmore *et al.*, 2015; Bossu *et al.*, 2022). A study in song sparrow, *Melospiza melodia* (Wilson), found that exploratory behaviour and migration distance were both correlated with SNPs in *DRD4* among migrants (Posliff, 2020).

A study of 15 populations of the highly divergent junco species complex, including eight subspecies of the dark-eyed junco, *Junco hyemalis* (Linnaeus), and two subspecies of the yellow-eyed junco, *Junco phaeonotus* (Wagler), found no consistent relationship across all congeneric species for *Clock* or *Adcyap1* length polymorphisms and the measured phenological traits. They did, however, find longer *Clock* alleles in two subspecies known to migrate longer distances as well as a relationship between *Adcyap1* length and migratory restlessness in one of two captive populations (Peterson *et al.*, 2013).

Several transcriptomic studies have been carried out on migratory species of the family Emberizidae. Singh, Rani &

Kumar (2013) measured expression levels in captive red-headed bunting, *Emberiza bruniceps* (Brandt), under varying photoperiods, for several genes including *Adcyap1*, *Bmal1*, *Clock*, *Cry1/2*, and *Per2*. A strong correlation between photoperiod and differential gene expression was observed in all the assayed tissue types, providing evidence of light-attuned oscillations within the circadian circuitry of migratory birds (Singh *et al.*, 2013; Leclerc *et al.*, 2010). Higher expression levels were observed under photoperiods consistent with triggers for autumn migration. Spring photoperiods are known to lead to many metabolic and endocrine changes (Sharma *et al.*, 2018*b*). Interestingly, Singh *et al.* (2013) also reported differential expression of the genes for DNA methyltransferase 3 (DNMT3) and Tet methylcytosine dioxygenase 2 (TET2), two enzymes involved in methylation, indicating a possible role for epigenetic control over migratory and non-migratory states as suggested previously for barn swallows (Saino *et al.*, 2017). Transcriptomic studies in the congenic black-headed bunting, *E. melanocephala* (Scopoli), revealed a similar pattern of differential expression in neural and peripheral tissues that oscillated with photoperiod (Singh *et al.*, 2015). These studies did not assess potential inter-individual variation in the measured responses, perhaps because both species of bunting are considered monotypic. These buntings represent species with part of their annual range including India and may represent more recent colonisation of this area after the landmasses merged about 20–40 MYA.

(*f*) *Cardinals (Family: Cardinalidae)*

The painted bunting, *Passerina ciris* (Linnaeus), and the lazuli bunting, *Passerina amoena* (Say), are classified in the family Cardinalidae (Lowther *et al.*, 2020). These species are located on the modern North American continent with early speciation possibly influenced by the presence of the North American interior seaway millions of years ago. Allelic diversity of three painted bunting populations, including both subspecies, was compared for both *Clock* and *Adcyap1*; migration phenology was only analysed in the western population for which geolocator data were available (Contina *et al.*, 2018). Greater allelic diversity was observed for both genes in the western population, with the most common allele Q_{11}; in the eastern population Q_{12} was more abundant. No significant correlation was detected with the initiation or duration of autumn migration, and it was concluded that individual allele studies may have limited resolution for this species. The same authors later assessed differences in migration behaviour between subspecies using a panel of SNPs similar to those used in thrushes (Ruegg *et al.*, 2014*a*) and warblers (Ruegg *et al.*, 2014*b*) and detected at least three distinct breeding populations across the continental USA that may form pertinent conservation units (Contina *et al.*, 2019).

(*g*) *Flycatchers and chats (Family: Muscicapidae)*

Johnsen *et al.* (2007) investigated variation in *Clock* Poly-Q regions in 12 populations of the bluethroat, *Cyanecula* (*Luscinia*) *svecica* (Linnaeus), an Old-World chat. No significant correlation between allelic diversity and latitudinal clines was found, with the Q_{13} allele present in 85.4% of genotypes; all individuals from the Italian subpopulation were homozygous for this allele (Johnsen *et al.*, 2007). A study on the common nightingale, *Luscinia megarhynchos* (Brehm), found a near-equal prevalence of the Q_{11} and Q_{12} alleles with a significant correlation between longer alleles and a later migration date (Saino *et al.*, 2015*a*). In the Northern wheatear, *Oenanthe Oenanthe* (Linnaeus), the most common allele is Q_{14} (Bazzi *et al.*, 2016*a*). Transcriptomic studies on this species detected seasonal differences in the expression of key clock elements in birds with different migration phenologies (Frias-Soler *et al.*, 2020, 2021).

Studies on the European pied flycatcher, *Ficedula hypoleuca* (Pallas), found five *Clock* Poly-Q alleles in contemporary (Saino *et al.*, 2015*a*) and historical samples (Kuhn *et al.*, 2013), with the Q_{12} allele present in 70% of individuals (Saino *et al.*, 2015*a*). Neither study identified a significant association between migration timing and polymorphisms in *Clock* (Kuhn *et al.*, 2013; Saino *et al.*, 2015*a*) or *Adcyap1* (Saino *et al.*, 2015*a*), however, Saino *et al.* (2015*a*) did report a slight relationship between timing and sex. A similar sex-dependent pattern of *Clock* and *Adcyap1* lengths with timing was found in the related spotted flycatcher, *Muscicapa striata* (Pallas) (Bazzi *et al.*, 2016*a*) and whinchat, *Saxicola rubetra* (Linnaeus) (Saino *et al.*, 2015*a*). A recent study in flycatchers evaluated polymorphisms in four *Clock* genes in the collared flycatcher, *Ficedula albicollis* (Temminck), but found no evidence for a relationship between genetic diversity and migration phenology over a four-year period (Krist *et al.*, 2021).

A study on the stonechat species complex (Justen *et al.*, 2022) found no clear correlates between phenological attributes such as breeding latitude and *Clock* genotypes but did detect a relationship with timing of autumn migration for the migratory species. This study included subspecies of the African stonechat, *Saxicola torquatus* (Linnaeus), resident in central Africa but not those of the resident population in South Africa, Canary Island stonechat, *Saxicola dacotiae* (Meade-Waldo), European stonechat, *Saxicola rubicola* (Linnaeus), and Siberian stonechat, *Saxicola maurus* (Pallas), and identified a variety of *Clock* alleles (Q_8–Q_{15}), with the most common alleles being Q_{13} and Q_{14}. The Q_{14} allele similarly was the most common allele observed in the congeneric whinchat (Saino *et al.*, 2015*a*). The Canary island stonechat, which was nearly homozygous for Q_{14}, likely represents a more recent radiation and colonization as the Canary islands were formed approximately 20 MYA.

This illustrates that, at least within the family Muscicapidae, migration phenology may vary independently of genetic polymorphisms but there is evidence for differential expression of circadian clock genes (Frias-Soler *et al.*, 2020).

(*h*) *Pipits (Family: Motacillidae)*

Tree pipits, *Anthus trivialis* (Linnaeus) were found to have relatively short *Clock* alleles (range Q_6–Q_{10}), with the most

prevalent being Q_9 which is found in 85% of the tested genotypes (Saino *et al.*, 2015*a*). Saino *et al.* (2015*a*) reported a significant relationship between migration date and longer alleles in females, but not males. Sex dependence of migration phenology on *Clock* photoperiodic responses remains to be documented for this species. It is possible that such dependence could be related to time of egg laying and hatching, which would be more strongly related to the timing of migration in females as recorded in blue tits (Liedvogel *et al.*, 2009).

(*i*) *Thrushes (Family: Turdidae)*

High-resolution detection of genetic markers was used in the Swainson's thrush, *Catharus ustulatus*, to identify genomic regions related to migration. Several genes were found to have higher levels of differentiation than expected, including genes previously associated with the circadian clock and migratory behaviour such as *Adcyap1*, *CREB1*, *NPAS2*, and *Per3*, between the two subspecies (Ruegg *et al.*, 2014*a*). The authors concluded that future work should include sampling from hybrid zones to assess the potential effect of the observed differentiation on maintaining barriers to gene flow between subspecies (Ruegg *et al.*, 2014*a*). Subsequent studies by Delmore *et al.* (2015) in a hybrid zone between coastal and inland populations of Swainson's thrush used next generation sequencing (NGS) of the whole genome and targeted their analyses to gene regions previously associated with migratory phenology, including *Clock*, *Adcyap1*, *NPAS2* and *DRD4*. Between-group analyses revealed high levels of heterogeneity in diversity estimates, while within-group diversity was lower in areas of higher speciation. Further analyses of the same data set (Delmore *et al.*, 2016) compared known differences in migration phenology and found that three genes co-vary with phenology, one of which was *Clock*. The complex interplay between genes, species concepts, and palaeogeography has also been studied in detail for *Catharus* thrushes (Voelker *et al.*, 2013) using genomic and mitochondrial markers. This study inferred an ancestral range north of Mexico and identified the adaptation of sedentary behaviour within *Catharus* to be correlated with the first speciation events that temporally coincided with range expansion to the newly formed Central America and the closing of the North American inland seaway. This was followed by recolonisation of land connecting the west and east of the continent in the late Pleistocene.

Genetic analysis of clock genes in Eurasian blackbirds, *Turdus merula* (Linnaeus), in relation to urbanisation assayed six established candidate genes *Clock*, *Adcyap1*, *NPAS2*, *CREB1*, *DRD4* and *SERT* (Mueller *et al.*, 2013*b*). The authors identified two *Clock* alleles, with the most common genotype being homozygous Q_7, while *Adcyap1* was considerably more diverse with almost 20 detected alleles. Transcriptomic analyses of 12 Eurasian blackbirds displaying differential resident *versus* migratory behaviour found several clusters of differentially expressed genes, however, none were circadian clock genes; one was associated with moult rate, which may influence migration timing (Franchini *et al.*, 2017). Although this study detected differential gene expression between two morphs within a sympatric population, the absence of relevant phenological functions for the identified genes makes interpretation difficult.

A candidate gene approach was applied to two congeneric species of bluebird: mountain bluebird, *Sialia currucoides* (Bechstein) and western bluebird, *Sialia mexicana* (Swainson) (Sauve *et al.*, 2021). *Adcyap1* and *DRD4* were assayed and correlations with migration phenology assessed; western bluebirds are partial migrants, often switching between strategies, while mountain bluebirds are obligate migrants. The analyses revealed a potential role for *DRD4*, a gene linked previously to exploratory behaviour in tits (Korsten *et al.*, 2010; Mueller *et al.*, 2013*a*) and swans (Van Dongen *et al.*, 2015), but absent in some warblers (Mueller *et al.*, 2011; Edwards *et al.*, 2015). No role was evident for *Adcyap1* (Sauve *et al.*, 2021) although less allelic diversity was found compared to studies which did find such a correlation (e.g. Eurasian blackcaps; Mueller *et al.*, 2011).

(*j*) *Shrikes and orioles (Families: Laniidae, Oriolidae)*

The Eurasian golden oriole, *Oriolus oriolus* (Linnaeus), and woodchat shrike, *Lanius senator* (Linnaeus), were included in a larger study evaluating inter-species differences in clock length polymorphisms (Bazzi *et al.*, 2016*a*) and three elements of migration phenology: migration date, migration distance, and latitude of breeding. Both orioles and shrikes had relatively short Poly-Q repeat regions with the most common allele being Q_6 and Q_9 respectively. How intra-species variation in the allelic diversity of these species is related to migratory strategy remains to be investigated.

(2) Non-passerine birds

(*a*) *Buzzards, hawks, and kites (Order: Accipitriformes, Family: Accipitridae)*

The Eurasian buzzard, *Buteo buteo* (Linnaeus), is monoallelic for the *Clock* allele, with all genotyped individuals homozygous for Q_8 (Chakarov *et al.*, 2013). The same study also assayed *Adcyap1*, *NPAS2*, and *CREB1* and found significant correlates with the observed polymorphisms in juveniles for timing of dispersal and dispersal distance. Interestingly no major differences were detected between the German ($N = 976$) and Bulgarian ($N = 23$) populations, indicating that population structure for gene polymorphisms is present even in smaller samples and that more extensive sampling may not significantly improve resolution. This study also included the red kite, *Milvus milvus* (Linnaeus), and northern goshawk, *Accipiter gentilis* (Linnaeus), but found no significant correlations with migration behaviour.

(*b*) *Hoopoes (Order: Bucerotiformes, Family: Upupidae)*

In Eurasian hoopoe, *Upupa epops* (Linnaeus), the most frequent allele for *Clock* was Q_8, although Q_7 was almost as common; calculated heterozygosity was 0.50. *Adcyap1* showed

comparable diversity with three alleles ranging in size from 157 to 159 bp and a higher calculated heterozygosity of 0.56 (Bazzi *et al.*, 2016*a*).

(*c*) *Wrynecks (Order: Piciformes, Family: Picidae)*

In the Eurasian wryneck, *Jynx torquilla* (Linnaeus), the most frequent allele size for *Clock* was the Q_8 allele. Four alleles were present ranging in size from Q_6 to Q_{10}; calculated heterozygosity was 0.44. Five *Adcyap1* alleles were identified, ranging in size from 131 to 137 bp, considerably shorter than those reported for most other non-passerines (Table 1). Calculated heterozygosity was considerably higher at an estimated 0.68 (Bazzi *et al.*, 2016*a*).

(*d*) *Bee-eaters (Order: Coraciiformes, Family: Meropidae)*

The European bee-eater was mono-allelic for *Clock* (Q_4 allele), considerably shorter than in most other non-passerines. For *Adcyap1* three alleles were identified, ranging in size from 155 bp to 169 bp. There was considerable homozygosity for the most common allele (163 bp), and an estimated diversity of only 0.19 (Bazzi *et al.*, 2016*a*). This study only included individuals from migratory European bee-eaters in Italy and did not include samples from the resident population found in South Africa.

(*e*) *Nightjars (Order: Caprimulgiformes, Family: Caprimulgidae)*

In the European nightjar, *Caprimulgus europaeus* (Linnaeus), the most common *Clock* allele was the Q_8 allele and two alleles were identified; calculated heterozygosity was 0.35 and most individuals were homozygous. A larger number of *Adcyap1* alleles were identified than in other non-passerine species (nine alleles) and a calculated heterozygosity of ∼0.81 (Bazzi *et al.*, 2016*a*).

(*f*) *Pigeons and doves (Order: Columbiformes, Family: Columbidae)*

In European turtle dove, *Streptopelia turtur* (Linnaeus), the most common allele size for *Clock* was the Q_7 allele, with a minimum of two alleles, but with a heterozygosity of 0.03 (i.e. they were nearly monoallelic). *Adcyap1* was more polymorphic with five alleles detected ranging in size from 148 to 152 bp. Overall diversity was relatively low with a calculated heterozygosity of 0.30 and many individuals homozygous for the most common 150 bp allele (Bazzi *et al.*, 2016*a*).

(*g*) *Gulls and shorebirds (Order: Charadriiformes, Families: Laridae, Scolopacidae, Charadriidae)*

Correlations of length polymorphisms in two regions of *Clock*, *Adcyap1* and *NPAS2* with laying date were studied in females of a large breeding colony of the largely resident yellow-legged gull, *Larus michahellis* (Naumann) (Romano *et al.*, 2018). Similar distributions of all assayed alleles were found between early- and late-laying birds, with the authors concluding that selection due to photoperiod on clock gene polymorphisms may not be universal in birds (Romano *et al.*, 2018). It should be noted that this study included only one of the two recognised subspecies, *L. m. michahellis*. Polymorphisms are likely to be more evident in early speciation (Rolland *et al.*, 2014), between ecological niches (Gómez *et al.*, 2016), or along lateral gradients (Linck, Freeman & Dumbacher, 2019).

The bar-tailed godwit subspecies *Limosa lapponica baueri* (Linnaeus) makes one of the most arduous annual journeys, breeding in Siberia and Alaska and overwintering in New Zealand. Parody-Merino *et al.* (2019) reported a high degree of *Clock* variability, with three nearly equally distributed alleles Q_9–Q_{11}. Statistical analyses revealed no clear relationship between allele size and timing of migration, but a slight latitudinal cline was observed with longer alleles tending to be found in individuals that travel further north. The authors concluded that clock gene polymorphisms are 'not a strong candidate for driving migration timing in migratory birds generally.' (p. 843), however, the fact that this species displayed significantly more heterozygosity than shorter range migrants should not be overlooked. Furthermore, this study only tested one distinct subspecies and may therefore have missed potential diversification among subspecies.

A recent study (de Almeida Miranda *et al.*, 2022) assessed diversity within *Adcyap1* in several species of shorebirds: the collared plover, *Charadrius collaris* (Vieillot), semipalmated sandpiper, *Calidris pusilla* (Linnaeus), semipalmated plover, *Charadrius semipalmatus* (Bonaparte), and spotted sandpiper, *Actitis macularius* (Linnaeus). They found several alleles for each species but the study was limited in its resolution by small sample sizes. Shorebirds represent a group with unique challenges in studying any underlying migration genetics as in many species their ranges are restricted to narrow strips along the coast of continents, with several occurring on both western and eastern coasts, in addition to several species following a distinct migration pattern along a flyway between Australasia and parts of the Nearctic. As such they are better suited to studies that track individuals using geolocators rather than studies using a more general approach with population-level averages for their range.

IV. CROSS-SPECIES COMPARATIVE ANALYSIS

(1) Population genetics

The results of the population genetics analyses are summarised in Table 3, with more detailed results available as online supporting information in Tables S1–S3. Across all 40 species included in this analysis, a total of 13 Poly-Q alleles was identified, ranging in size from Q_4 in the European bee-eater to Q_{16} in the blue tit, common redstart, and whinchat. Far more *Adcyap1* alleles were reported: 51 alleles ranging in size from 131 to 189 bp. The Hardy–Weinberg test showed that the *Clock* and *Adcyap1* alleles were in equilibrium for nearly all species, with only five species failing equilibrium assumptions in each case (Tables 3 and S1).

Table 3. Summary of results of population genetics comparative analysis testing of individual species for which allele data were available for *Clock* and *Adcyap1*. *, $P < 0.10$; **, $P < 0.05$; ***, $P < 0.02$.

Hardy–Weinberg	*Clock*			*Adcyap1*		
	N	Total	P	N	Total	P
In equilibrium	34	39	<0.02	32	37	<0.02***
Not in equilibrium	5			5		

Heterozygosity	*Clock*			*Adcyap1*		
	N	Total	H_o/H_e	N	Total	H_o/H_e
Observed (H_o) > Expected (H_e)	17	39	0.325/0.33	19	37	0.678/0.69
Expected (H_e) > Observed (H_o)	15			18		
Observed (H_o) = Expected (H_e)	3			–		
Single allele	4			–		

Neutrality (Ewens–Watterson)	*Clock*			*Adcyap1*		
	N	Total	P	N	Total	P
Neutral	32	39	>0.05	32	37	<0.05**
Selection	3			5		
Single allele	4			–		

Linkage disequilibrium	*Clock*	
	N	Total
Clock vs *Adcyap1*		
In equilibrium	33	36
Not in equilibrium	3	

The Hardy–Weinberg test on the entire data set also showed a lack of equilibrium ($P < 0.02$), as anticipated for species without known gene flow between them. Three species had equal observed (H_o) and expected (H_e) heterozygosity for the *Clock* allele while four species were mono-allelic (Table S1). For both alleles H_o was slightly higher than H_e in nearly half of the tested species: 17/39 for *Clock* and 19/37 for *Adcyap1*; mean H_o for the entire data set (Table 3) was 0.325 for *Clock* and 0.678 for *Adcyap1*, which was lower than the calculated expected heterozygosity (0.33 and 0.69, respectively).

The neutrality tests (Tables 3 and S2) detected significant evidence for deviation from neutrality in three out of 39 species for the *Clock* allele and for the *Adcyap1* allele. Overall tests for selection among species detected no evidence for deviation from neutrality in *Clock* but did detect deviation from neutrality for *Adcyap1*. For the 36 species for which data on both alleles were available, linkage disequilibrium was detected in only three species; the remainder appeared to be in equilibrium (Table S3).

(2) Mantel tests

A significant correlation was detected between the genetic distance (F_{ST}) of the *Clock* alleles and breeding latitude ($P = 0.099$), non-breeding latitude ($P = 0.018$), taxonomic distance ($P = 0.013$) and divergence time ($P = 0.051$) (Tables 4 and 5). Apart from divergence times, these correlations were positive. By contrast, no significant relationship was detected between F_{ST} for *Adcyap1* alleles and geographic attributes of migration phenology or taxonomic distance, but a significant negative correlation was found with divergence time ($P = 0.010$).

(3) Phylogenetic generalised least squares analyses

PGLS analyses did not find significant relationships between the average latitudes of breeding and non-breeding ranges and the allele length of the most common *Clock* or *Adcyap1* allele (Tables 4 and 5). The only significant relationship was between *Clock* allele length and total migration distance. Some correlations (at $P < 0.01$) were detected between *Clock* allele length and the start and middle dates of autumn migration when these dates were calculated relative to either the summer solstice or autumn equinox. A significant correlation was detected between *Adcyap1* and only the middle and end dates of spring migration and only when these were calculated in relation to the date of the winter solstice. The strength of the overall phylogenetic signal, assessed by estimating lambda and kappa, was strongly correlated with allele length of the most common alleles for both genes. A phylogenetic ANOVA between migratory and resident species

Table 4. Summary statistics for data used in the cross-species comparative analyses of the *Clock* and *Adcyap1* alleles (see Table 5). MYA, million years ago; PGLS, phylogenetic generalised least squares.

	Clock					*Adcyap1*				
Variable (x)	N	Mean	Std deviation	Minimum	Maximum	N	Mean	Std deviation	Minimum	Maximum
Mantel tests:										
Genetic distance (F_{ST})	39	1.54	1.98	0	9.42	39	37	1.96	0	2.23
Geographic distance (°):										
Breeding latitude	39	58.129	48.351	0	156.573	39	37	57.436	0	48.853
Non-breeding latitude	39	71.046	63.941	0	178.575	39	37	71.872	0	64.245
Taxonomic distance:	39	81.12	21.69	0	100.00	39	37	80.51	0	22.09
Evolutionary distance:										
Mean divergence times (MYA)	39	39.86	19.19	0	75.00	39	37	39.63	0	19.71
PGLS models:										
Migrants										
Allele size (*Clock* as Poly-Q, *Adcyap1* in bp)	58	9	2.22	4	14	58	54	166	135	9.60
Observed heterozygosity (H_o)	48	0.325	0.249	<0.001	0.839	48	39	0.670	0.191	0.179
Breeding latitude (°)	58	17.987	13.039	0.837	56.049	58	54	17.964	0.838	13.486
Non-breeding latitude (°)	58	48.429	9.741	4.669	66.762	58	54	47.549	4.669	14.537
Migration distance (m)	58	4,573,175	2,232,252	875,528	12,388,217	58	54	4,669,914	875,528	2,263,790
Migration distance (°)	58	41.018	19.934	8.250	111.290	58	54	41.873	8.250	20.240
Residents										
Allele size	18	10	1.94	8	14	18	17	170	139	13.55
Observed heterozygosity (H_o)	8	0.131	0.154	<0.001	0.343	8	2	0.538	0.312	0.319
Latitude (°)	18	23.876	17.396	2.419	55.384	18	17	24.457	2.419	17.751
Range (m)	18	691,024	727,105	10,531	2,657,515	18	17	715,938	10,531	741,522

Table 5. Summary of results of Mantel tests and phylogenetic generalised least squares (PGLS) regression analysis for *Clock* and *Adcyap1* alleles in the studied avian species (see Table 4 for summary of data used in the analysis). For Mantel tests, results are reported for tests for comparisons of the genetic distance (F_{ST}, see Table 4) between species with migration-related variables, taxonomic distance between species, and mean divergence times. For PGLS, the tests compared the length of the most common allele for each species to the migration-related variables shown. The strength of the phylogenetic signal for each gene was also assessed by estimating lambda and kappa. *, $P < 0.10$; **, $P < 0.05$; ***, $P < 0.02$.

Mantel tests		*Clock*			*Adcyap1*		
		Z-value	*P*-value	*R*-value	*Z*-value	*P*-value	*R*-value
Geography							
Breeding latitude		0.086	0.099*	1.638	0.036	0.499	0.657
Non-breeding latitude		0.124	0.018***	2.426	0.033	0.520	0.631
Taxonomy							
Taxonomic distance		0.116	0.013***	2.479	0.058	0.239	1.185
Evolutionary history							
Mean divergence time		−0.139	0.051*	−1.901	−0.190	0.010***	−2.573
PGLS			***P***	R^2	**DF**	***P***	R^2	**DF**
Geography								
Breeding latitude			0.275	0.02	74	0.300	0.02	69
Non-breeding latitude			0.220	0.03	56	0.472	0.01	52
Migration distance			0.093*	0.04	74	0.342	0.01	69
Timing								
Spring migration *versus* winter solstice (9:15 LD)	Start		0.343	0.02	56	0.273	0.02	52
	Mid		0.325	0.02	56	0.096*	0.05	52
	End		0.406	0.01	56	0.070*	0.06	52
Spring migration *versus* spring equinox (12:12 LD)	Start		0.638	0.00	56	0.450	0.01	52
	Mid		0.454	0.01	56	0.480	0.01	52
	End		0.389	0.01	56	0.532	0.01	52
Autumn migration *versus* summer solstice (15:9 LD)	Start		0.077*	0.05	56	0.905	0.00	52
	Mid		0.058*	0.06	56	0.643	0.00	52
	End		0.221	0.03	56	0.443	0.01	52
Autumn migration *versus* autumn equinox (12:12 LD)	Start		0.097*	0.05	56	0.950	0.00	52
	Mid		0.070*	0.06	56	0.602	0.01	52
	End		0.236	0.03	56	0.420	0.01	52
Phylogenetic signal								
Lambda (λ)			0.000***	0.92	74	0.000***	0.75	69
Kappa (κ)			0.001***	0.73	74	0.001***	0.61	69

revealed no significant partitioning based on *Clock* or *Adcyap1* allele length (data not shown) while the largely sedentary species complex of manakins had clock alleles ranging in size from Q_8 to Q_{11}, overlapping with those observed in migratory species.

(4) Time tree and palaeogeography

Divergence times for the main lineages of the species included in our analysis are represented in the reconstructed time tree (Fig. 7). The primary speciation event dividing the order Passeriformes from non-passerine birds took place around 85 MYA during the Late Cretaceous. The mean divergence time for the study species was ∼35 MYA with the most recent divisions taking place during the Miocene and Pliocene; the most recent speciation for this data set was at less than 1 MYA between two junco species.

According to palaeogeographic reconstructions (Fig. 7 and Video S1), the primary division between passerines and non-passerines occurred prior to the formation of most continents, and would likely have required several range shifts, which would only have occurred ∼50 MYA with a substantial amount of continental drift continuing until about 21 MYA when most continents started to assume their contemporary positions. There was still, however, a continuation of tectonic plate movement resulting in significant geographic remodelling on most continents, shaping the landscape within which range selection and speciation occurred.

V. DISCUSSION

Our systematic review of the existing evidence for a potential role of diversity within genes associated with the circadian clock machinery in regulating or shaping migration phenology clearly identifies conflicting evidence. Several studies have reported such relationships (Liedvogel *et al.*, 2009;

Caprioli *et al.*, 2012; Kuhn *et al.*, 2013) whilst others found no evidence for an association (Peterson *et al.*, 2013; Contina *et al.*, 2018; Parody-Merino *et al.*, 2019). Additionally, several transcriptomic studies on related species have detected no differences in expression levels of key clock genes (Jones *et al.*, 2008*a*; Franchini *et al.*, 2017; Brown, 2019) used in candidate gene association studies.

It should be noted that pitfalls within study design or data analyses could confound the results. For example, most early studies relied on the Mantel test (Mantel, 1967) to identify correlations between genetic distance and geographic distance, as is a common practice in studies of spatial genetics and ecology. More recently, however, several authors have questioned the suitability of the Mantel test in such applications (Guillot & Rousset, 2013; Legendre, Fortin & Borcard, 2015) as this test is designed to compare two distance matrices that both measure differences rather than physical distance and argued that geographic distance expressed in kilometres may exaggerate the relationship. Other potential confounding issues include the *post-hoc* grouping of study individuals based on observed outlier status (Bazzi *et al.*, 2015) or sex (Dor *et al.*, 2011) rather than predetermined grouping, small sample sizes (de Almeida Miranda *et al.*, 2022), and the averaging of alleles in heterozygotes to analyse mean allele length (Zhang *et al.*, 2017).

Analysing individuals by mean allele length makes several assumptions about the underlying biology of the system that have yet to be validated. Firstly, it assumes equal bi-allelic expression from a single locus, an absence of parental imprinting (Jang *et al.*, 2013), and identical regulation of both alleles. This contradicts the central hypothesis of altered entrainment cycles in the presence of a length polymorphism as well as some existing evidence for heritable patterns of methylation in birds (Romano *et al.*, 2017; Saino *et al.*, 2019). Secondly it assumes knowledge of the copy number variation for the studied genes (Skinner *et al.*, 2014), which may be particularly complex in birds given that different karyotypes where chromosome numbers ranging between 78 and 82 can be found in most species (Degrandi *et al.*, 2020). Initial studies grouped individuals based on homozygosity, presence of the most common allele, and heterozygosity with a longer or shorter allele to model likely gene effects without assuming equal bi-allelic expression from a single locus (Johnsen *et al.*, 2007).

Some studies may also have deviated by analysing total fragment length and including alleles that differed by only one base pair in the data set (Contina *et al.*, 2018), thereby not implicitly measuring only the identified length polymorphism, as well as including an excessively narrow cohort in the analysis by focusing on a single subspecies (Parody-Merino *et al.*, 2019). The central hypothesis tested in such studies is that genetic differences establish a migratory divide that leads to speciation; how variation within a single subspecies fits into this hypothesis is unclear as speciation is unlikely to be documented below the subspecies level. Additionally, in many of the studied species for which known subspecies exist, no clear effort was made to assign individuals to specific subspecies, but rather the assumption was that all individuals at the same study site would belong to a single subspecies. The importance of accounting for subspecies has been highlighted by the increasing number of documented cases of hybrid speciation among birds, including eagles (Väli *et al.*, 2018), finches (Lamichhaney *et al.*, 2018), manakins (Barrera-Guzmán *et al.*, 2017), sparrows (Hermansen *et al.*, 2011), warblers (Brelsford, Milá & Irwin, 2011; Ralston, Ermacor & Kirchman, 2015) and tits (Janas *et al.*, 2021).

Most studies published to date were conducted at the species level and have focused on European species of birds within the order Passeriformes. The name 'Passeriformes' was derived from the Latin *passer*, meaning sparrow, while the French verb *passer* refers to movement (Vieillot, 1816). As many French naturalists contributed to early ornithological works, Vieillot (1816) noted that some scientists may have had a mistaken sense that passerines were migratory while non-passerines were sedentary by conflating these words in early ornithological works like the *Histoire Naturelle* (Leclerc, 1780). This may have established a persistent subconscious bias considering most of the early works were written in either Latin or French. Other potential reasons for a focus on passerines in migration research could include their suggested higher diversification, and possibly speciation rates (Jetz *et al.*, 2012; Prum *et al.*, 2015), differences in appropriate trapping methods, or a perception that they have more complex migration strategies. However, mist-net trapping can be used effectively used in many non-passerine species including kingfishers (Dalton *et al.*, 2022), cuckoos (Chaisi *et al.*, 2019), fluff-tails (Dalton *et al.*, 2016), and hornbills (Theron *et al.*, 2013), and some exceptional long-distance migrations are found in non-passerine species (Parody-Merino *et al.*, 2019).

The full data set for *Clock* revealed a limited number of genotypes with a small number of alleles and widespread homozygosity while *Adcyap1* was considerably more diverse with many more genotypes per species and more heterozygotes. A general trend was found for higher observed than expected heterozygosity among nearly half of the species (Table S1) while the whole data set analysis showed only marginally less heterozygosity (Table 3). Population genetics analyses found most tested species to be in Hardy–Weinberg equilibrium, with only five species failing equilibrium assumptions. For the willow warbler and common chiffchaff, this was anticipated as their data included individuals of known subspecies. For the remainder, it is possible that a similar population substructure exists or that there are heterozygote deficiencies or advantages potentially arising from a variety of factors such as inbreeding, hybridisation or recent population bottlenecks (Hedrick, 1987; Lade *et al.*, 1996; Luikart & Cornuet, 1998).

Our comparative analysis found a significant correlation between *Clock* alleles and both breeding and non-breeding latitudes using Mantel tests, but not PGLS (Freckleton, Harvey & Pagel, 2002), with concerns regarding false positives arising from use of the Mantel test in spatial genetics potentially an explanation (Guillot & Rousset, 2013; Legendre *et al.*, 2015). No correlations with these variables

were found for *Adcyap1* using either method, although this may be due to the difficulty in comparing this gene across taxa in the absence of sequence data, as our use of the full length of the allele may result in inclusion of length variation in non-focal regions (Bazzi *et al.*, 2016*a*). Additionally, no evidence was found for partitioning migratory and sedentary species based on either candidate gene, consistent with previous findings using phylogenetic approaches (Lugo Ramos *et al.*, 2017). When specifically analysing the relationship between alleles and the timing of migration among migratory species, the results weakly supported a correlation between *Clock* alleles and the timing of autumn migration (Ralston *et al.*, 2019; Justen *et al.*, 2022) while for *Adcyap1*, there was weak evidence for a relationship with the timing and staging of spring migration (Bazzi *et al.*, 2016*a*; Ralston *et al.*, 2019).

Interestingly, an analysis of genetic distance in relation to divergence times found a significant correlation for both *Clock* and *Adcyap1* alleles, as well as a relationship between taxonomic distance and *Clock* alleles, indicating that contemporary genotypes may still resemble the ancestral genotypes inherited millions of years ago. This was further supported by a strong phylogenetic signal for both genes, indicating that closely related species have similar allele lengths, and that variation is best explained by lineage, similar to previous findings (Bazzi *et al.*, 2016*a*). This also supports findings from European pied flycatcher studies that identified similar alleles in contemporary (Saino *et al.*, 2015*a*) and historical samples (Kuhn *et al.*, 2013). Considering the palaeogeographic remodelling which coincided with the divergence of these lineages over the past 85 million years, along with global climatic changes, selective sweeps and convergent evolution would have abrogated this relationship if these genes were under strong selection in terms of adapting migration strategies in response to geographic or environmental changes (Stern, 2013). In the comparative analysis, the tests for deviation from assumptions of neutrality and evidence of selection failed to detect selection in either gene for most species.

A substantial amount of genetic research in ornithology from the 1990 s (Vos *et al.*, 1995) to shortly after the turn of the century (Bensch & Åkesson, 2005) focused on identifying genes with length polymorphisms that co-vary with latitude (Bensch, Åkesson & Irwin, 2002) and could be used as molecular markers in population assignment or barcoding (Ottvall *et al.*, 2005). This included the identification of several markers that have been used successfully for population assignment in species such as willow warbler (Bensch *et al.*, 2002), house wren, *Troglodytes aedon* (Vieillot) (Arguedas & Parker, 2000), superb fairy-wren (Double *et al.*, 1997), and long-tailed manakin, *Chiroxiphia linearis* (Bonaparte) (McDonald & Potts, 1994). Therefore, although some evidence exists that polymorphisms within the tested candidate genes *Clock* and *Adcyap1* co-vary with migration behaviour, considering similar relationships (such as latitudinal clines in allele length within ranges and an influence on timing of breeding) was also illustrated in several sedentary species (Johnsen *et al.*, 2007; Liedvogel *et al.*, 2012), caution should be applied when translating these findings into a causal relationship.

Future studies are currently needed in the field of migration genetics to address current gaps in our understanding of the regulation of the circadian machinery. Such studies should ideally expand the breadth of species for which data are available, include transcription data comparing expression levels and dominance of alleles in heterozygous individuals, compare expression levels among species with known polymorphic length variation in candidate genes, determine the copy number variation of circadian genes in avian species, and use information on epigenetics to address any potential roles in regulating these genes. Future studies should, ideally, be designed to avoid the limitations identified above, including appropriate sample sizes, sampling to capture as much variation in migration strategies as possible, controlling for subspecies status in sample populations, the use of correct taxonomy, the inclusion of non-passerines, and data analysis procedures that make the fewest assumptions about the regulation and expression of heterozygous alleles.

VI. CONCLUSIONS

(1) Some evidence exists that polymorphisms within the poly-Q region of the *Clock* gene are related to geo-spatial differences in the range of migrating birds, as well as the timing of Autumn migration.
(2) Little evidence exists that polymorphism expressed as total allele length of *Adcyap1* has a geo-spatial pattern, however there is weak evidence for a relationship with the timing of spring migration.
(3) Both *Clock* and *Adcyap1* have a strong phylogenetic signal indicative that alleles and genotypes are highly heritable within lineages.
(4) For both *Clock* and *Adcyap1* the observed patterning is well correlated to divergence times and may therefore still reflect the ancestrally inherited genotypes rather than recently acquired changes.
(5) No clear evidence exists that either candidate gene can be used to distinguish migratory from sedentary birds across all taxa.

VII. ACKNOWLEDGEMENTS

The authors would like to thank Chris Guglielmo, Yolanda Morbey and Ivan Maggini for support with field data collection at Long Point Bird Observatory (Canada), and Andrea Galimberti, Cristina Possenti, Fernando Spina and the late Nicola Saino for support with genetic analyses and field data collection in Ventotene Island (Italy). We would like to thank BirdLife International for supplying the species range data used to calculate migration phenology traits. Images were created in BioRender.com. This work is based on the research supported wholly/in part by the National Research Foundation of South Africa (Grant Number: 112062).

VIII. DATA AVAILABILITY STATEMENT

Data used in this study are available for download from the Zenodo repository at: https://doi.org/10.5281/zenodo.6637839. The custom-designed python script PAReTT version 1.0.1 is available for download for installation from source code on GitHub (https://github.com/LSLeClercq/PAReTT).

IX. REFERENCES

Agarwala, R., Barrett, T., Beck, J., Benson, D. A., Bollin, C., Bolton, E., Bourexis, D., Brister, J. R., Bryant, S. H., Canese, K., Cavanaugh, M., Charowhas, C., Clark, K., Dondoshansky, I., Feolo, M., *et al.* (2018). Database resources of the National Center for biotechnology information. *Nucleic Acids Research* **46**, D8–D13.

Aguilar-Roblero, R. (2015). Introduction to circadian rhythms, clocks, and its genes. In *Mechanisms of Circadian Systems in Animals and Their Clinical Relevance*, pp. 1–12. Springer International Publishing, Switzerland.

Aguilar-Roblero, R., Díaz-Muñoz, M. & Fanjul-Moles, M. L. (2015). *Mechanisms of Circadian Systems in Animals and their Clinical Relevance*. Springer International Publishing, Switzerland.

Åkesson, S., Ilieva, M., Karagicheva, J., Rakhimberdiev, E., Tomotani, B. & Helm, B. (2017). Timing avian long-distance migration: from internal clock mechanisms to global flights. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372**, 20160252.

Albrecht, U. & Ripperger, J. A. (2009). Clock genes. In *Encyclopedia of Neuroscience* (Volume **1**), pp. 759–762. Academic Press, Spain.

Ali, S., Ripley, S. D. & Roberts, T. J. (1999). *Handbook of the Birds of India and Pakistan: Together with those of Bangladesh, Nepal, Sikkim, Bhutan and Sri Lanka*, 2nd Edition. Bombay Natural History Society, Oxford University Press, Delhi.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410.

Altwegg, R., Broms, K., Erni, B., Barnard, P., Midgley, G. F. & Underhill, L. G. (2012). Novel methods reveal shifts in migration phenology of barn swallows in South Africa. *Proceedings of the Royal Society B: Biological Sciences* **279**, 1485–1490.

Arguedas, N. & Parker, P. G. (2000). Seasonal migration and genetic population structure in house wrens. *The Condor* **102**, 517–528.

Barrera-Guzmán, A. O., Aleixo, A., Shawkey, M. D. & Weir, J. T. (2017). Hybrid speciation leads to novel male secondary sexual ornamentation of an Amazonian bird. *Proceedings of the National Academy of Sciences of the United States of America* **115**, E218–E225.

Bazzi, G., Ambrosini, R., Caprioli, M., Costanzo, A., Liechti, F., Gatti, E., Gianfranceschi, L., Podofillini, S., Romano, A., Romano, M., Scandolara, C., Saino, N. & Rubolini, D. (2015). Clock gene polymorphism and scheduling of migration: a geolocator study of the barn swallow *Hirundo rustica*. *Scientific Reports* **5**, 12443.

Bazzi, G., Cecere, J. G., Caprioli, M., Gatti, E., Gianfranceschi, L., Podofillini, S., Possenti, C. D., Ambrosini, R., Saino, N., Spina, F. & Rubolini, D. (2016*a*). Clock gene polymorphism, migratory behaviour and geographic distribution: a comparative study of trans-Saharan migratory birds. *Molecular Ecology* **25**, 6077–6091.

Bazzi, G., Galimberti, A., Hays, Q. R., Bruni, I., Cecere, J. G., Gianfranceschi, L., Hobson, K. A., Morbey, Y. E., Saino, N., Guglielmo, C. G. & Rubolini, D. (2016*b*). Adcyap1 polymorphism covaries with breeding latitude in a Nearctic migratory songbird, the Wilson's warbler (*Cardellina pusilla*). *Ecology and Evolution* **6**, 3226–3239.

Bazzi, G., Podofillini, S., Gatti, E., Gianfranceschi, L., Cecere, J. G., Spina, F., Saino, N. & Rubolini, D. (2017). Candidate genes have sex-specific effects on timing of spring migration and moult speed in a long-distance migratory bird. *Current Zoology* **63**, 479–486.

Bensch, S. & Åkesson, M. (2005). Ten years of AFLP in ecology and evolution: why so few animals? *Molecular Ecology* **14**, 2899–2914.

Bensch, S., Åkesson, S. & Irwin, D. E. (2002). The use of AFLP to find an informative SNP: genetic differences across a migratory divide in willow warblers. *Molecular Ecology* **11**, 2359–2366.

Billerman, S. M., Keeney, B. K., Rodewald, P. G. & Schulenberg, T. S. (2020). *Birds of the World*. Cornell Laboratory of Ornithology, Ithaca, NY https://birdsoftheworld.org/bow/home.

Billerman, S. M., Keeney, B. K., Rodewald, P. G. & Schulenberg, T. S. (2022). *Birds of the World*. Cornell Laboratory of Ornithology, Ithaca, NY https://birdsoftheworld.org/bow/home.

BirdLife International & Handbook of the Birds of the World (2021). Bird species distribution maps of the world. http://datazone.birdlife.org/species/requestdis.

BirdLife International (2021) Species factsheet: *Alaudala cheleensis*. Downloaded from http://www.birdlife.org.

Borowske, A., Gjerdrum, C. & Elphick, C. (2017). Timing of migration and prebasic molt in tidal marsh sparrows with different breeding strategies: comparisons among sexes and species. *Auk* **134**, 51–64.

Boss, J., Liedvogel, M., Lundberg, M., Olsson, P., Reischke, N., Naurin, S., Åkesson, S., Hasselquist, D., Wright, A., Grahn, M. & Bensch, S. (2015). Gene expression in the brain of a migratory songbird during breeding and migration. *Movement Ecology* **4**, 1–11.

Bossu, C. M., Heath, J. A., Kaltenecker, G. S., Helm, B. & Ruegg, K. C. (2022). Clock-linked genes underlie seasonal migratory timing in a diurnal raptor. *Proceedings of the Royal Society B: Biological Sciences* **289**, 20212507.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., Suchard, M. A., Rambaut, A. & Drummond, A. J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* **10**, e1003537.

Bourret, A. & Garant, D. (2015). Candidate gene-environment interactions and their relationships with timing of breeding in a wild bird population. *Ecology and Evolution* **5**, 3628–3641.

Brelsford, A., Milá, B. & Irwin, D. E. (2011). Hybrid origin of Audubon's warbler. *Molecular Ecology* **20**, 2380–2389.

Briedis, M., Bauer, S., Adamík, P., Alves, J. A., Costa, J. S., Emmenegger, T., Gustafsson, L., Kolecek, J., Liechti, F., Meier, C. M., Procházka, P. & Hahn, S. (2019). A full annual perspective on sex-biased migration timing in long-distance migratory birds. *Proceedings of the Royal Society B* **286**, 20182821.

Brown, C. (2019). Phylogeny and Evolution of Swallows (*Hirundinidae*) with a Transcriptomic Perspective on Seasonal Migration. Doctoral thesis: Louisiana State University and Agricultural & Mechanical College. Louisiana, USA. https://digitalcommons.lsu.edu/gradschool_dissertations/4823.

Brown, I. L. & Mewaldt, L. R. (1968). Behavior of sparrows of the genus *Zonotrichia*, in orientation cages during the lunar cycle. *Zeitschrift für Tierpsychologie* **25**, 668–700.

Caprioli, M., Ambrosini, R., Boncoraglio, G., Gatti, E., Romano, A., Romano, M., Rubolini, D., Gianfranceschi, L. & Saino, N. (2012). Clock gene variation is associated with breeding phenology and maybe under directional selection in the migratory barn swallow. *PLoS One* **7**, e35140.

Carey, C. (2009). The impacts of climate change on the annual cycles of birds. *Philosophical Transactions of the Royal Society B: Biological Sciences* **364**, 3321–3330.

Carr, J. (2021). Mantel version 2.1.0. https://github.com/jwcarr/mantel.

Cava, J. A., Perlut, N. G. & Travis, S. E. (2019). Heritability and evolvability of morphological traits of Savannah sparrows (*Passerculus sandwichensis*) breeding in agricultural grasslands. *PLoS One* **14**, e0210472.

Chaisi, M. E., Osinubi, S. T., Dalton, D. L. & Suleman, E. (2019). Occurrence and diversity of avian haemosporidia in Afrotropical landbirds. *International Journal for Parasitology: Parasites and Wildlife* **8**, 36–44.

Chakarov, N., Jonker, R. M., Boerner, M., Hoffman, J. I. & Krüger, O. (2013). Variation at phenological candidate genes correlates with timing of dispersal and plumage morph in a sedentary bird of prey. *Molecular Ecology* **22**, 5430–5440.

Contina, A., Alcantara, J. L., Bridge, E. S., Ross, J. D., Oakley, W. F., Kelly, J. F. & Ruegg, K. C. (2019). Genetic structure of the painted bunting and its implications for conservation of migratory populations. *Ibis* **161**, 372–386.

Contina, A., Bridge, E. S., Ross, J. D., Shipley, J. R. & Kelly, J. F. (2018). Examination of clock and Adcyap1 gene variation in a neotropical migratory passerine. *PLoS One* **13**, e0190859.

Cramer, H. (1946). *Mathematical Models of Statistics*. Princeton University Press, Princeton, NJ.

Dalton, D. L., Nupen, L. J., Mwale, M., Pretorius, C., Kropff, A. S., Monchusi, B. A., Labuschagne, K. & Osinubi, S. T. (2022). First steps to success: identification of divergence among the northern and the southern lineages of African pygmy kingfisher (*Ispidina picta*) (Coraciiformes: *Alcedinidae*). *Journal of Ornithology* **163**, 931–943.

Dalton, D. L., Vermaak, E., Smit-Robinson, H. A. & Kotze, A. (2016). Lack of diversity at innate immunity toll-like receptor genes in the critically endangered white-winged Flufftail (*Sarothrura ayresi*). *Scientific Reports* **6**, 1–8.

de Almeida Miranda, D., Araripe, J., de Morais Magalhaes, N. G., de Siqueira, L. S., de Abreu, C. C., Pereira, P. D. C., Henrique, E. P., da Silva Chira, P. A. C., de Melo, M. A. D., do Rêgo, P. S., Diniz, D. G., Sherry, D. F., Diniz, C. W. P. & Guerreiro-Diniz, C. (2022). Shorebirds' longer migratory distances are associated with larger ADCYAP1 microsatellites and greater morphological complexity of hippocampal astrocytes. *Frontiers in Psychology* **12**, 6679.

de Juana, E. & Suárez, F. (2020). Asian short-toed lark (*Alaudala cheleensis*), version 1.0. In *Birds of the World*. Cornell Lab of Ornithology, Ithaca, NY. https://doi.org/10.2173/bow.lstlar2.01.

de Zoeten, T. & Pulido, F. (2020). How migratory populations become resident. *Proceedings of the Royal Society B: Biological Sciences* **287**, 20193011.

Degrandi, T. M., Barcellos, S. A., Costa, A. L., Garnero, A. D. V., Hass, I. & Gunski, R. J. (2020). Introducing the bird chromosome database: an overview of cytogenetic studies in birds. *Cytogenetic and Genome Research* **160**, 199–205.

Delmore, K., Illera, J. C., Pérez-Tris, J., Segelbacher, G., Ramos, J. S. L., Durieux, G., Ishigohoka, J. & Liedvogel, M. (2020*a*). The evolutionary history and genomics of European blackcap migration. *eLife* **9**, e54462.

Delmore, K. E., Hübner, S., Kane, N. C., Schuster, R., Andrew, R. L., Câmara, F., Guigó, R. & Irwin, D. E. (2015). Genomic analysis of a migratory divide reveals candidate genes for migration and implicates selective sweeps in generating islands of differentiation. *Molecular Ecology* **24**, 1873–1888.

Delmore, K. E., Toews, D. P. L., Germain, R. R., Owens, G. L. & Irwin, D. E. (2016). The genetics of seasonal migration and plumage color. *Current Biology* **26**, 2167–2173.

Delmore, K. E., Van Doren, B. M., Conway, G. J., Curk, T., Garrido-Garduño, T., Germain, R. R., Hasselmann, T., Hiemer, D., Van Der Jeugd, H. P., Justen, H., Lugo Ramos, J. S., Maggini, I., Meyer, B. S., Phillips, R. J., Remisiewicz, M., *et al.* (2020*b*). Individual variability and versatility in an eco-evolutionary model of avian migration: migratory strategies of blackcaps. *Proceedings of the Royal Society B: Biological Sciences* **287**, 20201339.

DeMoranville, K. J., Corder, K. R., Hamilton, A., Russell, D. E., Huss, J. M. & Schaeffer, P. J. (2019). PPAR expression, muscle size and metabolic rates across the gray catbird's annual cycle are greatest in preparation for fall migration. *Journal of Experimental Biology* **222**, jeb198028.

Dierschke, V., Mendel, B. & Schmaljohann, H. (2005). Differential timing of spring migration in northern wheatears *Oenanthe oenanthe*: hurried males or weak females? *Behavioral Ecology and Sociobiology* **57**, 470–480.

Dor, R., Cooper, C. B., Lovette, I. J., Massoni, V., Bulit, F., Liljesthrom, M. & Winkler, D. W. (2012). Clock gene variation in *Tachycineta* swallows. *Ecology and Evolution* **2**, 95–105.

Dor, R., Lovette, I. J., Safran, R. J., Billerman, S. M., Huber, G. H., Vortman, Y., Lotem, A., McGowan, A., Evans, M. R., Cooper, C. B. & Winkler, D. W. (2011). Low variation in the polymorphic clock gene poly-Q region despite population genetic structure across barn swallow (*Hirundo rustica*) populations. *PLoS One* **6**, e28843.

Double, M. C., Dawson, D., Burke, T. & Cockburn, A. (1997). Finding the fathers in the least faithful bird: a microsatellite-based genotyping system for the superb fairy-wren *Malurus cyaneus*. *Molecular Ecology* **6**, 691–693.

Edwards, H. A., Hajduk, G. K., Durieux, G., Burke, T. & Dugdale, H. L. (2015). No association between personality and candidate gene polymorphisms in a wild bird population. *PLoS One* **10**, e0138439.

Ersts, P. J. (2012). *Geographic Distance Matrix Generator Version 1.23*. Center for Blodiversity and Conservation, American Museum of Natural History, http://biodiversityinformatics.amnh.org/open_source/gdmg.

Ewens, W. J. (1972). The sampling theory of selectively neutral alleles. *Theoretical Population Biology* **3**, 87–112.

Fink, D., Auer, T., Johnston, A., Strimas-Mackey, M., Robinson, O., Ligocki, S., Hochachka, W., Jaromczyk, L., Wood, C., Davies, I., Iliff, M. & Seitz, L. (2021). *eBird status and trends, data version: 2021; released: 2022*. Cornell Lab of Ornithology, Ithaca, NY. https://doi.org/10.2173/ebirdst.2021.

Franchini, P., Irisarri, I., Fudickar, A., Schmidt, A., Meyer, A., Wikelski, M. & Partecke, J. (2017). Animal tracking meets migration genomics: transcriptomic analysis of a partially migratory bird species. *Molecular Ecology* **26**, 3204–3216.

Freckleton, R. P., Harvey, P. H. & Pagel, M. (2002). Phylogenetic analysis and comparative data: a test and review of evidence. *American Naturalist* **160**, 712–726.

Frias-Soler, R. C., Pildaín, L. V., Pârâu, L. G., Wink, M. & Bairlein, F. (2020). Transcriptome signatures in the brain of a migratory songbird. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics* **34**, 100681.

Frias-Soler, R. C., Pildaín, L. V., Wink, M. & Bairlein, F. (2021). A revised and improved version of the northern wheatear (*Oenanthe oenanthe*) transcriptome. *Diversity* **13**, 151.

Fry, H. & Boesman, P. F. D. (2020). European bee-eater (*Merops apiaster*), version 1.0. In *Birds of the World* (eds J. del Hoyo, A. Elliott, J. Sargatal, D. A. Christie and E. de Juana). Cornell Lab of Ornithology, Ithaca, NY, USA. https://doi.org/10.2173/bow.eubeat1.02.

Gómez, C., Tenorio, E. A., Montoya, P. & Cadena, C. D. (2016). Niche-tracking migrants and nicheswitching residents: evolution of climatic niches in new world warblers (*Parulidae*). *Proceedings of the Royal Society B: Biological Sciences* **283**, 20152458.

Guillot, G. & Rousset, F. (2013). Dismantling the Mantel tests. *Methods in Ecology and Evolution* **4**, 336–344.

Guldvog, C.Ø. (2015). Clock genes and their role in migratory phenotype among Passer sparrows. Masters dissertation: University of Oslo. Oslo, Norway. http://urn.nb.no/URN:NBN:no-49731.

Hannibal, J., Ding, J. M., Chen, D., Fahrenkrug, J., Larsen, P. J., Gillette, M. U. & Mikkelsen, J. D. (1998). Pituitary adenylate cyclase activating peptide (PACAP) in the retinohypothalamic tract: a daytime regulator of the biological clock. *Annals of the New York Academy of Sciences* **865**, 197–206.

Hardy, G. H. (1908). Mendelian proportions in a mixed population. *Science* **28**, 49–50.

Hau, M. & Gwinner, E. (1996). Food as a circadian zeitgeber for house sparrows: the effect of different food access durations. *Journal of Biological Rhythms* **11**, 196–207.

HBW and BirdLife International (2021). Handbook of the birds of the world and BirdLife international digital checklist of the birds of the world. Version 6. http://datazone.birdlife.org/userfiles/file/Species/Taxonomy/HBW-BirdLife_Checklist_v6_Dec21.zip%0D.

Hedrick, P. W. (1987). Gametic disequilibrium measures: proceed with caution. *Genetics* **117**, 331–341.

Helm, B. & Visser, M. E. (2010). Heritable circadian period length in a wild bird population. *Proceedings of the Royal Society B: Biological Sciences* **277**, 3335–3342.

Hermansen, J. S., Sæther, S. A., Elgvin, T. O., Borge, T., Hjelle, E. & Sætre, G. P. (2011). Hybrid speciation in sparrows I: phenotypic intermediacy, genetic admixture and barriers to gene flow. *Molecular Ecology* **20**, 3812–3822.

Janas, K., Gaweł, P., Łatkiewicz, A., Lutyk, D., Gustafsson, L., Cichoń, M. & Drobniak, S. M. (2021). Sexual dichromatism, size dimorphism, and microscale anatomy of white wing stripe in blue tits. *Current Zoology* **67**, 585–596.

Jang, H. J., Seo, H. W., Lee, B. R., Yoo, M., Womack, J. E. & Han, J. Y. (2013). Gene expression and DNA methylation status of chicken primordial germ cells. *Molecular Biotechnology* **54**, 177–186.

Jenni, L. & Kéry, M. (2003). Timing of autumn bird migration under climate change: advances in long-distance migrants, delays in short-distance migrants. *Proceedings of the Royal Society B: Biological Sciences* **270**, 1467–1471.

Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K. & Mooers, A. O. (2012). The global diversity of birds in space and time. *Nature* **491**, 444–448.

Johnsen, A., Fidler, A. E., Kuhn, S., Carter, K. L., Hoffmann, A., Barr, I. R., Biard, C., Charmantier, A., Eens, M., Korsten, P., Siitari, H., Tomiuk, J. & Kempenaers, B. (2007). Avian clock gene polymorphism: evidence for a latitudinal cline in allele frequencies. *Molecular Ecology* **16**, 4867–4880.

Johnston, R. A., Paxton, K. L., Moore, F. R., Wayne, R. K. & Smith, T. B. (2016). Seasonal gene expression in a migratory songbird. *Molecular Ecology* **25**, 5680–5691.

Jones, S., Pfister-Genskow, M., Benca, R. M. & Cirelli, C. (2008*a*). Molecular correlates of sleep and wakefulness in the brain of the white-crowned sparrow. *Journal of Neurochemistry* **105**, 46–62.

Jones, S., Pfister-Genskow, M., Cirelli, C. & Benca, R. M. (2008*b*). Changes in brain gene expression during migration in the white-crowned sparrow. *Brain Research Bulletin* **76**, 536–544.

Justen, H., Hasselmann, T., Illera, J. C., Delmore, K. E., Serrano, D., Flinks, H., Senzaki, M., Kawamura, K., Helm, B. & Liedvogel, M. (2022). Population-specific association of clock gene polymorphism with annual cycle timing in stonechats. *Scientific Reports* **12**, 1–13.

Kauffman, E. G. (1984). Paleobiogeography and evolutionary response dynamic in the cretaceous Western interior seaway of North America. *Geological Association of Canada Special Paper* **27**, 273–306.

Korsten, P., Mueller, J. C., Hermannstädter, C., Bouwman, K. M., Dingemanse, N. J., Drent, P. J., Liedvogel, M., Matthysen, E., Van Oers, K., Van Overveld, T., Patrick, S. C., Quinn, J. L., Sheldon, B. C., Tinbergen, J. M. & Kempenaers, B. (2010). Association between DRD4 gene polymorphism and personality variation in great tits: a test across four wild populations. *Molecular Ecology* **19**, 832–843.

Krist, M., Munclinger, P., Briedis, M. & Adamík, P. (2021). The genetic regulation of avian migration timing: combining candidate genes and quantitative genetic approaches in a long-distance migrant. *Oecologia* **196**, 373–387.

Kuhn, K., Schwenk, K., Both, C., Canal, D., Johansson, U. S., van der Mije, S., Töpfer, T. & Päckert, M. (2013). Differentiation in neutral genes and a candidate gene in the pied flycatcher: using biological archives to track global climate change. *Ecology and Evolution* **3**, 4799–4814.

Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. (2017). TimeTree: a resource for timelines, Timetrees, and divergence times. *Molecular Biology and Evolution* **34**, 1812–1819.

Lade, J. A., Murray, N. D., Marks, C. A. & Robinson, N. A. (1996). Microsatellite differentiation between Phillip Island and mainland Australian populations of the red fox *Vulpes vulpes*. *Molecular Ecology* **5**, 81–87.

Laine, V. N., Atema, E., Vlaming, P., Verhagen, I., Mateman, C., Ramakers, J. J. C., van Oers, K., Spoelstra, K. & Visser, M. E. (2019). The genomics of circadian timing in a wild bird, the great tit (*Parus major*). *Frontiers in Ecology and Evolution* **7**, 152.

Lamichhaney, S., Han, F., Webster, M. T., Andersson, L., Grant, B. R. & Grant, P. R. (2018). Rapid hybrid speciation in Darwin's finches. *Science* **359**, 224–228.
Lancaster, A. K., Single, R. M., Solberg, O. D., Nelson, M. P. & Thomson, G. (2007). PyPop update – a software pipeline for large-scale multilocus population genomics. *Tissue Antigens* **69**, 192–197.
Le, C. T., Do, T. B., Pham, T. M. A., Nguyen, V. D., Nguyen, S. K., Nguyen, V. H., Cao, P. B. & Omollo, W. O. (2022). Reconstruction of the evolutionary biogeography reveals the origins of *Alpinia Roxb*. (Zingiberaceae): a case of "out-of-Asia" migration to the southern hemisphere. *Acta Botanica Brasilica* **36**, e2021abb0255.
Leclerc, B., Kang, S. W., Mauro, L. J., Kosonsiriluk, S., Chaiseha, Y. & El Halawani, M. E. (2010). Photoperiodic modulation of clock gene expression in the avian premammillary nucleus. *Journal of Neuroendocrinology* **22**, 119–128.
Leclerc, G.-L. (1780). *Histoire Naturelle des Oiseaux*. L'imprimerie Royale, Paris, France. https://doi.org/10.5962/bhl.title.66490.
Legendre, P., Fortin, M. J. & Borcard, D. (2015). Should the Mantel test be used in spatial analysis? *Methods in Ecology and Evolution* **6**, 1239–1247.
Liebl, A. L., Wesner, J. S., Russell, A. F. & Schrey, A. W. (2021). Methylation patterns at fledging predict delayed dispersal in a cooperatively breeding bird. *PLoS One* **16**, e0252227.
Liedvogel, M. & Sheldon, B. C. (2010). Low variability and absence of phenotypic correlates of clock gene variation in a great tit *Parus major* population. *Journal of Avian Biology* **41**, 543–550.
Liedvogel, M., Cornwallis, C. K. & Sheldon, B. C. (2012). Integrating candidate gene and quantitative genetic approaches to understand variation in timing of breeding in wild tit populations. *Journal of Evolutionary Biology* **25**, 813–823.
Liedvogel, M., Szulkin, M., Knowles, S. C. L., Wood, M. J. & Sheldon, B. C. (2009). Phenotypic correlates of clock gene variation in a wild blue tit population: evidence for a role in seasonal timing of reproduction. *Molecular Ecology* **18**, 2444–2456.
Linck, E., Freeman, B. G. & Dumbacher, J. P. (2019). Speciation and gene flow across an elevational gradient in New Guinea kingfishers. *Evolutionary Biology* **33**, 1643–1652.
Linnaeus, C. (1758). *Systema Naturae per Regna Tria Naturae Secundum Classes, Ordines, Genera, Species Cum Characteribus Differentiis Synonymis, Locis*, 10th Edition. Impensis Direct. Laurentii Salvii, Holmiae, Sweden. https://doi.org/10.5962/bhl.title.542.
Lowther, P. E., Lanyon, S. M., Thompson, C. W. & Schulenberg, T. S. (2020). Painted bunting (*Passerina ciris*), version 1.0. In *Birds of the World* (ed. S. M. Billerman). Cornell Lab of Ornithology, Ithaca, NY. https://doi.org/10.2173/bow.paibun.01.
Lugo Ramos, J. S., Delmore, K. E. & Liedvogel, M. (2017). Candidate genes for migration do not distinguish migratory and non-migratory birds. *Journal of Comparative Physiology A* **203**, 383–397.
Luikart, G. & Cornuet, J. M. (1998). Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conservation Biology* **12**, 228–237.
Lundberg, M., Boss, J., Canbäck, B., Liedvogel, M., Larson, K. W., Grahn, M., Åkesson, S., Bensch, S. & Wright, A. (2013). Characterisation of a transcriptome to find sequence differences between two differentially migrating subspecies of the willow warbler *Phylloscopus trochilus*. *BMC Genomics* **14**, 1–11.
Lundberg, M., Liedvogel, M., Larson, K., Sigeman, H., Grahn, M., Wright, A., Åkesson, S. & Bensch, S. (2017). Genetic differences between willow warbler migratory phenotypes are few and cluster in large haplotype blocks. *Evolution Letters* **1**, 155–168.
Mäkinen, H., Viitaniemi, H. M., Visser, M. E., Verhagen, I., van Oers, K. & Husby, A. (2019). Temporally replicated DNA methylation patterns in great tit using reduced representation bisulfite sequencing. *Scientific Data* **6**, 1–7.
Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research* **27**, 209–220.
McDonald, D. B. & Potts, W. K. (1994). Cooperative display and relatedness among males in a lek-mating bird. *Science* **266**, 1030–1032.
Meert, J. G. (2012). The (Paleo)geography of evolution: making sense of changing biology and changing continents. *Evolution: Education and Outreach* **5**, 547–554.
Merlin, C. & Liedvogel, M. (2019). The genetics and epigenetics of animal migration and orientation: birds, butterflies and beyond. *The Journal of Experimental Biology* **222**, jeb191890.
Mettler, R., Segelbacher, G. & Schaefer, H. M. (2015). Interactions between a candidate gene for migration (ADCYAP1), morphology and sex predict spring arrival in blackcap populations. *PLoS One* **10**, e0144587.
Mishra, I., Singh, D. & Kumar, V. (2017). Seasonal alterations in the daily rhythms in hypothalamic expression of genes involved in the photoperiodic transduction and neurosteroid-dependent processes in migratory blackheaded buntings. *Journal of Neuroendocrinology* **29**, 1–10.
Mueller, J. C., Korsten, P., Hermannstaedter, C., Feulner, T., Dingemanse, N. J., Matthysen, E., Van Oers, K., Van Overveld, T., Patrick, S. C., Quinn, J. L., Riemenschneider, M., Tinbergen, J. M. & Kempenaers, B. (2013*a*). Haplotype structure, adaptive history and associations with exploratory behaviour of the DRD4 gene region in four great tit (*Parus major*) populations. *Molecular Ecology* **22**, 2797–2809.
Mueller, J. C., Partecke, J., Hatchwell, B. J., Gaston, K. J. & Evans, K. L. (2013*b*). Candidate gene polymorphisms for behavioural adaptations during urbanization in blackbirds. *Molecular Ecology* **22**, 3629–3637.
Mueller, J. C., Pulido, F. & Kempenaers, B. (2011). Identification of a gene associated with avian migratory behaviour. *Proceedings of the Royal Society B: Biological Sciences* **278**, 2848–2856.
Müller, R. D., Cannon, J., Qin, X., Watson, R. J., Gurnis, M., Williams, S., Pfaffelmoser, T., Seton, M., Russell, S. H. J. & Zahirovic, S. (2018). GPlates: building a virtual earth through deep time. *Geochemistry, Geophysics, Geosystems* **19**, 2243–2261.
Newton, I. (2007). *The Migration Ecology of Birds*. Academic Press, Great Britain. https://doi.org/10.1016/B978-0-12-517367-4.X5000-1.
Norevik, G., Åkesson, S., Andersson, A., Bäckman, J. & Hedenström, A. (2019). The lunar cycle drives migration of a nocturnal bird. *PLoS Biology* **17**, e3000456.
Nosil, P., Harmon, L. J. & Seehausen, O. (2009). Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution* **24**, 145–156.
O'Malley, K. G. & Banks, M. A. (2008). A latitudinal cline in the Chinook salmon (*Oncorhynchus tshawytscha*) clock gene: evidence for selection on PolyQ length variants. *Proceedings of the Royal Society B: Biological Sciences* **275**, 2813–2821.
Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E. & Wagner, H. (2020). vegan: community ecology package. https://cran.r-project.org/package=vegan.
Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., D'Amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., *et al.* (2001). Terrestrial ecoregions of the world: a new map of life on earth: a new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *Bioscience* **51**, 933–938.
Orme, D., Freckleton, R., Thomas, G., Petzoldt, T., Fritz, S., Isaac, N. & Pearse, W. (2018). caper: Comparative Analyses of Phylogenetics and Evolution in R. https://cran.r-project.org/package=caper.
Ottvall, R., Höglund, J., Bensch, S. & Larsson, K. (2005). Population differentiation in the redshank (*Tringa totanus*) as revealed by mitochondrial DNA and amplified fragment length polymorphism markers. *Conservation Genetics* **6**, 321–331.
Pancerasa, M., Ambrosini, R., Saino, N. & Casagrandi, R. (2018). Barn swallows long-distance migration occurs between significantly temperature-correlated areas. *Scientific Reports* **8**, 12359.
Parody-Merino, Á. M., Battley, P. F., Conklin, J. R. & Fidler, A. E. (2019). No evidence for an association between Clock gene allelic variation and migration timing in a long-distance migratory shorebird (*Limosa lapponica bauer*i). *Oecologia* **191**, 843–859.
Peterson, M. P., Abolins-Abols, M., Atwell, J. W., Rice, R. J., Milá, B. & Ketterson, E. D. (2013). Variation in candidate genes CLOCK and ADCYAP1 does not consistently predict differences in migratory behavior in the songbird genus *Junco*. *F1000Research* **2**, 115.
Posliff, C.D. (2020). *Correlations in Movement Behaviour over Large and Small Geographic Scales in Song Sparrows (*Melospiza melodia*)*. Masters dissertation: Western University of Ontario. Ontario, Canada. https://ir.lib.uwo.ca/etd/7027.
Prasad, G. & Parmar, V. (2022). Cretaceous insular India–Consequences for biological evolution and faunal interchanges. *Journal of the Palaeontological Society of India* **67**, 56–71.
Prum, R. O., Berv, J. S., Dornburg, A., Field, D. J., Townsend, J. P., Lemmon, E. M. & Lemmon, A. R. (2015). A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* **526**, 569–573.
Puga, I., Lainez, B., Fernández-Real, J. M., Buxadé, M., Broch, M., Vendrell, J. & Espel, E. (2005). A polymorphism in the 3′ untranslated region of the gene for tumor necrosis factor receptor 2 modulates reporter gene expression. *Endocrinology* **146**, 2210–2220.
Pyle, P., Nur, N., Henderson, R. P. & DeSante, D. F. (1993). The effects of weather and lunar cycle on nocturnal migration of landbirds at southeast Farallon Island, California. *The Condor* **95**, 343.
Python Team (2021). *Python Programming Language*. Wilmington, Delaware, USA https://www.python.org/.
QGIS Development Team (2022). QGIS Geographic Information System. https://www.qgis.org.
R Core Team (2020). *R: A Language and Environment for Statisitical Computing*. R Foundation for Statistical Computing, Vienna https://www.r-project.org/.
Ralston, J., Ermacor, D. A. & Kirchman, J. J. (2015). A new wood warbler hybrid (*Oreothlypis celata* x *O. ruficapilla*) from the Adirondack mountains, New York. *The Wilson Journal of Ornithology* **127**, 21–28.

Ralston, J., Lorenc, L., Montes, M., DeLuca, W. V., Kirchman, J. J., Woodworth, B. K., Mackenzie, S. A., Newman, A., Cooke, H. A., Freeman, N. E., Sutton, A. O., Tauzer, L. & Norris, D. R. (2019). Length polymorphisms at two candidate genes explain variation of migratory behaviors in blackpoll warblers (*Setophaga striata*). *Ecology and Evolution* **9**, 8840–8855.

Revell, L. J. (2012). Phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* **3**, 217–223.

Robart, A. R., McGuire, M. M. K. & Watts, H. E. (2018). Increasing photoperiod stimulates the initiation of spring migratory behaviour and physiology in a facultative migrant, the pine siskin. *Royal Society Open Science* **5**, 180876.

Rolland, J., Jiguet, F., Jønsson, K. A., Condamine, F. L. & Morlon, H. (2014). Settling down of seasonal migrants promotes bird diversification. *Proceedings of the Royal Society B: Biological Sciences* **281**, 20140473.

Romano, A., De Giorgio, B., Parolini, M., Favero, C., Possenti, C. D., Iodice, S., Caprioli, M., Rubolini, D., Ambrosini, R., Gianfranceschi, L., Saino, N. & Bollati, V. (2017). Methylation of the circadian Clock gene in the offspring of a free-living passerine bird increases with maternal and individual exposure to PM10. *Environmental Pollution* **220**, 29–37.

Romano, A., Possenti, C. D., Caprioli, M., Gatti, E., Gianfranceschi, L., Rubolini, D., Saino, N. & Parolini, M. (2018). Circadian genes polymorphism and breeding phenology in a resident bird, the yellow-legged gull. *Journal of Zoology* **304**, 117–123.

Roques, S., Henry, P.-Y., Guyot, G., Bargain, B., Cam, E. & Pradel, R. (2022). When to Depart from a Stopover Site? Time since arrival matters more than current weather conditions. *Ornithology* **139**, 1–13.

RStudio Team (2021). *RStudio: Integrated Development Environment for R*. Boston, MA http://www.rstudio.com/.

Ruegg, K., Anderson, E. C., Boone, J., Pouls, J. & Smith, T. B. (2014*a*). A role for migration-linked genes and genomic islands in divergence of a songbird. *Molecular Ecology* **23**, 4757–4769.

Ruegg, K., Anderson, E. C., Paxton, K. L., Apkenas, V., Lao, S., Siegel, R. B., Desante, D. F., Moore, F. & Smith, T. B. (2014*b*). Mapping migration in a songbird using high-resolution genetic markers. *Molecular Ecology* **23**, 5726–5739.

Saino, N., Albetti, B., Ambrosini, R., Caprioli, M., Costanzo, A., Mariani, J., Parolini, M., Romano, A., Rubolini, D., Formenti, G., Gianfranceschi, L. & Bollati, V. (2019). Inter-generational resemblance of methylation levels at circadian genes and associations with phenology in the barn swallow. *Scientific Reports* **9**, 1–16.

Saino, N., Ambrosini, R., Albetti, B., Caprioli, M., De Giorgio, B., Gatti, E., Liechti, F., Parolini, M., Romano, A., Romano, M., Scandolara, C., Gianfranceschi, L., Bollati, V. & Rubolini, D. (2017). Migration phenology and breeding success are predicted by methylation of a photoperiodic gene in the barn swallow. *Scientific Reports* **7**, 45412.

Saino, N., Bazzi, G., Gatti, E., Caprioli, M., Cecere, J. G., Possenti, C. D., Galimberti, A., Orioli, V., Bani, L., Rubolini, D., Gianfranceschi, L. & Spina, F. (2015*a*). Polymorphism at the Clock gene predicts phenology of long-distance migration in birds. *Molecular Ecology* **24**, 1758–1773.

Saino, N., Romano, M., Caprioli, M., Fasola, M., Lardelli, R., Micheloni, P., Scandolara, C., Rubolini, D. & Gianfranceschi, L. (2013). Timing of molt of barn swallows is delayed in a rare Clock genotype. *PeerJ* **2013**, e17.

Saino, N., Rubolini, D., Ambrosini, R., Romano, M., Scandolara, C., Fairhurst, G. D., Caprioli, M., Romano, A., Sicurella, B. & Liechti, F. (2015*b*). Light-level geolocators reveal covariation between winter plumage molt and phenology in a trans-Saharan migratory bird. *Oecologia* **178**, 1105–1112.

Sangster, G. (2014). The application of species criteria in avian taxonomy and its implications for the debate over species concepts. *Biological Reviews* **89**, 199–214.

Sauve, D., Dale, C. A., Tigano, A., Ratcliffe, L. M. & Friesen, V. L. (2021). Do candidate genes for migration and behavior explain migratory variation in bluebirds (*Sialia spp.*)? *The Wilson Journal of Ornithology* **132**, 820–829.

Scheuerlein, A. & Gwinner, E. (2002). Is food availability a circannual zeitgeber in tropical birds? A field experiment on stonechats in tropical Africa. *Journal of Biological Rhythms* **17**, 171–180.

Schmincke, H.-U. (1976). The geology of the Canary Islands. In *Biogeography and Ecology in the Canary Islands* (Volume **30**, ed. G. Kunkel), pp. 67–184. Monographiae Biologicae, Dordrecht.

Scotese, C.R. (2016). PALEOMAP PaleoAtlas for GPlates and the PaleoData plotter program PALEOMAP Project. http://www.earthbyte.org/paleomap---%0Apaleoatlas---for---gplates/.

Sharma, A., Singh, D., Das, S. & Kumar, V. (2018*a*). Hypothalamic and liver transcriptome from two crucial life-history stages in a migratory songbird. *Experimental Physiology* **103**, 559–569.

Sharma, A., Singh, D., Malik, S., Gupta, N. J., Rani, S. & Kumar, V. (2018*b*). Difference in control between spring and autumn migration in birds: insight from seasonal changes in hypothalamic gene expression in captive buntings. *Proceedings of the Royal Society B: Biological Sciences* **285**, 2018153.

Silva, C. N. S., McFarlane, S. E., Hagen, I. J., Rönnegård, L., Billing, A. M., Kvalnes, T., Kemppainen, P., Rønning, B., Ringsby, T. H., Sæther, B. E., Qvarnström, A., Ellegren, H., Jensen, H. & Husby, A. (2017). Insights into the genetic architecture of morphological traits in two passerine bird species. *Heredity* **119**, 197–205.

Simonneaux, V., Ouichou, A. & Pévet, P. (1993). Pituitary adenylate cyclase-activating polypeptide (PACAP) stimulates melatonin synthesis from rat pineal gland. *Brain Research* **603**, 148–152.

Singh, D., Rani, S. & Kumar, V. (2013). Daily expression of six clock genes in central and peripheral tissues of a night-migratory songbird: evidence for tissue-specific circadian timing. *Chronobiology International* **30**, 1208–1217.

Singh, D., Trivedi, A. K., Rani, S., Panda, S. & Kumar, V. (2015). Circadian timing in central and peripheral tissues in a migratory songbird: dependence on annual life-history states. *FASEB Journal* **29**, 4248–4255.

Skinner, B. M., Al Mutery, A., Smith, D., Völker, M., Hojjat, N., Raja, S., Trim, S., Houde, P., Boecklen, W. J. & Griffin, D. K. (2014). Global patterns of apparent copy number variation in birds revealed by cross-species comparative genomic hybridization. *Chromosome Research* **22**, 59–70.

Slatkin, M. (1994). An exact test for neutrality based on the Ewens sampling distribution. *Genetical Research* **64**, 71–74.

Sockman, K. W. & Hurlbert, A. H. (2020). How the effects of latitude on daylight availability may have influenced the evolution of migration and photoperiodism. *Functional Ecology* **34**, 1752–1766.

Spina, F. & Volponi, S. (2008). Atlante della Migrazione degli Uccelli in Italia. In *Non-Passeriformi* (Volume **1**). ISPRA – MATTM, Rome.

Spina, F. & Volponi, S. (2009). Atlante della Migrazione degli Uccelli in Italia. In *Passeriformi* (Volume **2**). ISPRA–MATTM, Rome.

Steinmeyer, C., Mueller, J. C. & Kempenaers, B. (2009). Search for informative polymorphisms in candidate genes: Clock genes and circadian behaviour in blue tits. *Genetica* **136**, 109–117.

Stephan, F. K. (2002). The 'other' circadian system: food as a zeitgeber. *Journal of Biological Rhythms* **17**, 284–292.

Steri, M., Idda, M. L., Whalen, M. B. & Orrù, V. (2018). Genetic variants in mRNA untranslated regions. *Wiley Interdisciplinary Reviews: RNA* **9**, e1474.

Stern, D. L. (2013). The genetic causes of convergent evolution. *Nature Reviews Genetics* **14**, 751–764.

Tauber, E. & Kyriacou, C. P. (2005). Molecular evolution and population genetics of circadian clock genes. *Methods in Enzymology* **393**, 797–817.

Taylor, B. & Kirwan, G. M. (2020). Streaky-breasted Flufftail (*Sarothrura boehmi*), version 1.0. In *Birds of the World* (eds J. del Hoyo, A. Elliott, J. Sargatal, D. A. Christie and E. de Juana). Cornell Lab of Ornithology, Ithaca, NY. https://doi.org/10.2173/bow.stbflu1.01.

Theron, N., Dalton, D., Grobler, J. P., Jansen, R. & Kotze, A. (2013). Molecular insights on the re-colonization of the Limpopo Valley, South Africa, by southern ground-hornbills. *Journal of Ornithology* **154**, 727–737.

Trivedi, A. K., Sur, S., Sharma, A., Taufique, S. T., Gupta, N. J. & Kumar, V. (2019). Temperature alters the hypothalamic transcription of photoperiod responsive genes in induction of seasonal response in migratory redheaded buntings. *Molecular and Cellular Endocrinology* **493**, 110454.

Väli, Ü., Mirski, P., Sellis, U., Dagys, M. & Maciorowski, G. (2018). Genetic determination of migration strategies in large soaring birds: evidence from hybrid eagles. *Proceedings of the Royal Society B: Biological Sciences* **285**, 20180855.

Van Dongen, W. F. D., Robinson, R. W., Weston, M. A., Mulder, R. A. & Guay, P. J. (2015). Variation at the DRD4 locus is associated with wariness and local site selection in urban black swans. *BMC Evolutionary Biology* **15**, 1–11.

van Eck, N. J. & Waltman, L. (2017). Citation-based clustering of publications using CitNetExplorer and VOSviewer. *Scientometrics* **111**, 1053–1070.

Vieillot, L.-P. (1816). *Vieillot's Analyse d'une nouvelle ornithologie élémentaire*. A. Belin, Paris. https://doi.org/10.5962/bhl.title.12613.

Viitaniemi, H. M., Verhagen, I., Visser, M. E., Honkela, A., Van Oers, K., Husby, A. & Meyer, M. (2019). Seasonal variation in genome-wide DNA methylation patterns and the onset of seasonal timing of reproduction in great tits. *Genome Biology and Evolution* **11**, 970–983.

Voelker, G., Bowie, R. C. K. & Klicka, J. (2013). Gene trees, species trees and earth history combine to shed light on the evolution of migration in a model avian system. *Molecular Ecology* **22**, 3333–3344.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Friters, A., Pot, J., Paleman, J., Kuiper, M. & Zabeau, M. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* **23**, 4407–4414.

Watterson, G. A. (1977). Heterosis or neutrality? *Genetics* **85**, 789–814.

Weinberg, W. (1908). On the demonstration of heredity in man. (1963). In *Papers on Human Genetics*. Prentice-Hall, Hoboken.

Yeh, F. C., Yang, R. C., Boyle, T. B. J., Ye, Z. H. & Mao, J. X. (1997). *POPGENE, the User-friendly Shareware for Population Genetic Analysis*. Molecular biology and biotechnology centre, University of Alberta, Canada, https://sites.ualberta.ca/~fyeh/popgene_download.html.

Zhang, S., Xu, X., Wang, W., Yang, W. & Liang, W. (2017). Clock gene is associated with individual variation in the activation of reproductive endocrine and behavior of Asian short toed lark. *Scientific Reports* **7**, 1–8.

X. SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. Heterozygosity and Hardy–Weinberg equilibrium test results for *Clock* and *Adcyap1* alleles for 40 avian species.

Table S2. Fixation index and Ewers–Waterson test results for *Clock* and *Adcyap1* alleles for 40 avian species.

Table S3. Linkage disequilibrium test results comparing *Clock versus Adcyap1* alleles for 36 avian species.

Video S1. Reconstruction of paleogeography relevant to the major periods of divergence indicating the differences between contemporary and historical landscapes of evolution and selection.

(*Received 6 September 2022; revised 16 February 2023; accepted 20 February 2023; published online 6 March 2023*)

scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Birds of a feather flock together: a dataset for *Clock* and *Adcyap1* genes from migration genetics studies

Louis-Stéphane Le Clercq[1,2] ✉, Gaia Bazzi[3], Joan Ferrer Obiol[4,5,6], Jacopo G. Cecere[3], Luca Gianfranceschi[7], J. Paul Grobler[2], Antoinette Kotzé[1,2], Marta Riutort León[4,5], Jacob González-Solís[5,8], Diego Rubolini[6,9], Miriam Liedvogel[10,11] & Desiré Lee Dalton[12]

Birds in seasonal habitats rely on intricate strategies for optimal timing of migrations. This is governed by environmental cues, including photoperiod. Genetic factors affecting intrinsic timekeeping mechanisms, such as circadian clock genes, have been explored, yielding inconsistent findings with potential lineage-dependency. To clarify this evidence, a systematic review and phylogenetic reanalysis was done. This descriptor outlines the methodology for sourcing, screening, and processing relevant literature and data. PRISMA guidelines were followed, ultimately including 66 studies, with 34 focusing on candidate genes at the genotype-phenotype interface. Studies were clustered using bibliographic coupling and citation network analysis, alongside scientometric analyses by publication year and location. Data was retrieved for allele data from databases, article supplements, and direct author communications. The dataset, version 1.0.2, encompasses data from 52 species, with 46 species for the *Clock* gene and 43 for the *Adcyap1* gene. This dataset, featuring data from over 8000 birds, constitutes the most extensive cross-species collection for these candidate genes, used in studies investigating gene polymorphisms and seasonal bird migration.

Background & Summary

Birds occupy nearly every habitat and ecoregion on Earth, however, many of these habitats experience large seasonal shifts in key ecological attributes such as length of day[1], temperature[2], rainfall[3,4], and associated food and nesting material availability[5]. This has necessitated the adaptive evolution of complex strategies to maximise survival through seasonal migrations between breeding and wintering ranges. Migrations are carefully timed events, scheduled in such a manner that birds can optimise hours of daylight[6], nighttime visibility[7,8], and time spent at stop-over sites[9] along their migration route to ensure timely arrivals for optimal habitat use. While most of the ecological attributes play some role in the timing of migration, one of the best studied attributes that serve as a trigger to initiate migration is the length of day or photoperiod. The photoperiod is primarily responsible for daily oscillations within the regulatory feedback loops of the circadian clock, which differentially expresses genes during light or dark phases to maintain sleep-wake cycles in most organisms[10].

[1]South African National Biodiversity Institute, P.O. Box 754, Pretoria, 0001, South Africa. [2]Department of Genetics, University of the Free State, P.O. Box 339, Bloemfontein, 9300, South Africa. [3]Area Avifauna Migratrice, Istituto Superiore per la Protezione e la Ricerca Ambientale, via Ca' Fornacetta 9, I-40064, Ozzano Emilia, BO, Italy. [4]Departament de Genètica, Universitat de Barcelona, Gran Via de les Corts Catalanes, 585, 08007, Barcelona, Spain. [5]Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Gran Via de les Corts Catalanes, 585, 08007, Barcelona, Spain. [6]Dipartimento di Scienze e Politiche Ambientali, Università degli Studi di Milano, via Celoria 26, Milan, I-20133, Italy. [7]Dipartimento di Bioscienze, Università degli Studi di Milano, via Celoria 26, Milan, I-20133, Italy. [8]Departament de Biologia Evolutiva, Universitat de Barcelona, Gran Via de les Corts Catalanes, 585, 08007, Barcelona, Spain. [9]Istituto di Ricerca sulle Acque, IRSA-CNR, Via del Mulino 19, I-20861, Brugherio, (MB), Italy. [10]Max Planck Research Group Behavioural Genomics, Max Planck Institute for Evolutionary Biology, 24306, Plön, Germany. [11]Institute of Avian Research, An der Vogelwarte 21, 26386, Wilhelmshaven, Germany. [12]School of Health and Life Sciences, Teesside University, Middlesbrough, TS1 3BA, UK. ✉e-mail: leclercq.l.s@gmail.com

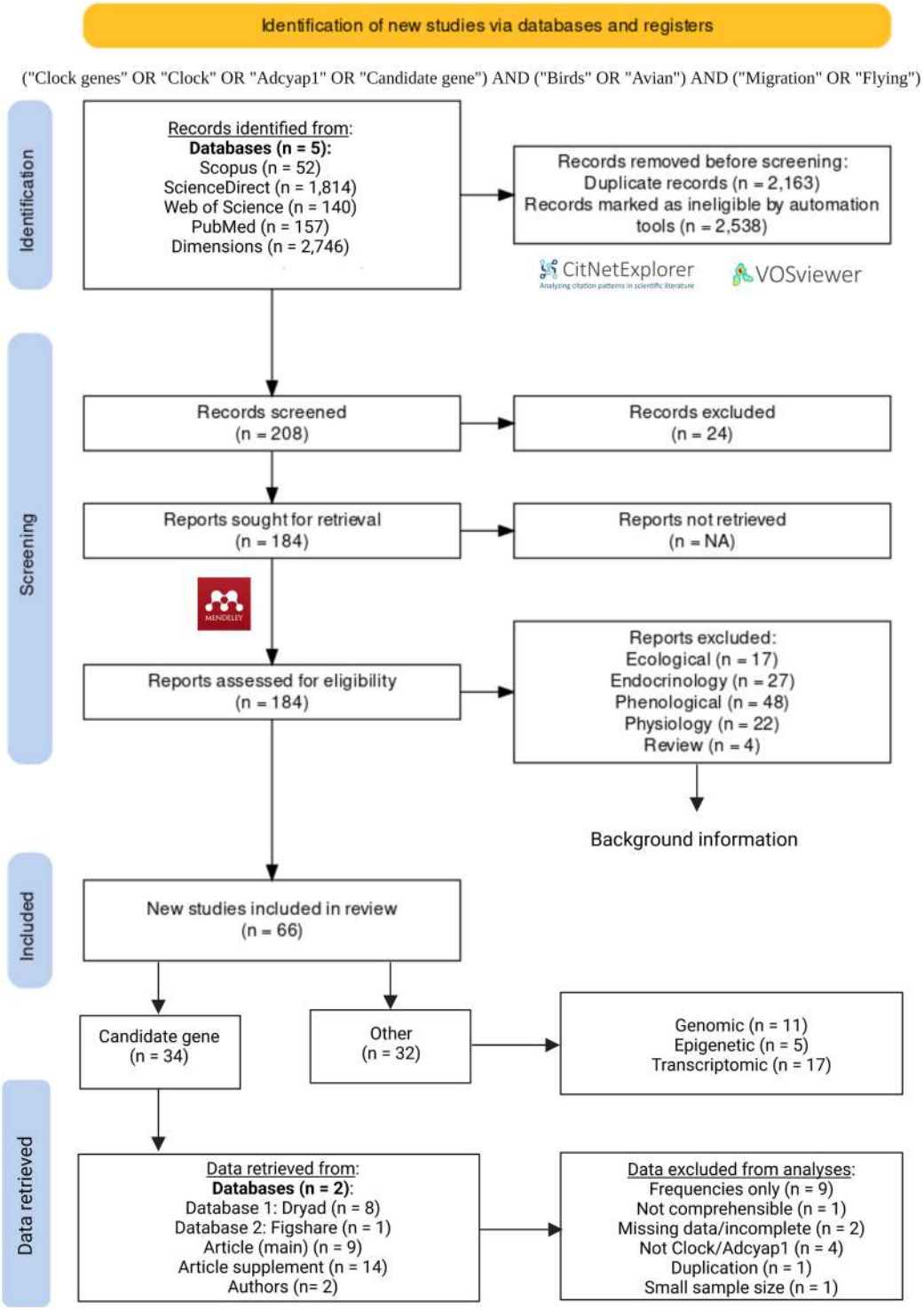


Fig. 1 PRISMA statement for the systematic approach used to identify studies that measured clock gene polymorphisms in relation to annual synchronicity of live events such as breeding and migration in birds. Further details are also provided for the retrieval of allele data for individual studies from various sources as well as reasons for exclusion of studies. (image edited in BioRender.com).

One conundrum regarding migration in birds is how differential migration patterns are established and maintained within singular species, even in the absence of extrinsic environmental triggers. For example, several species within the order Coraciiformes have distinct populations that are either year-round residents, with minimal altitudinal movement, or long-distance migrants. This includes such species as the Lilac-breasted roller[11] (*Coracias caudatus*) and Woodland kingfisher[12] (*Halcyon senegalensis*), both having subspecies that are delineated by differential migration, as well as the European bee-eater[13] (*Merops apiaster*), which is considered

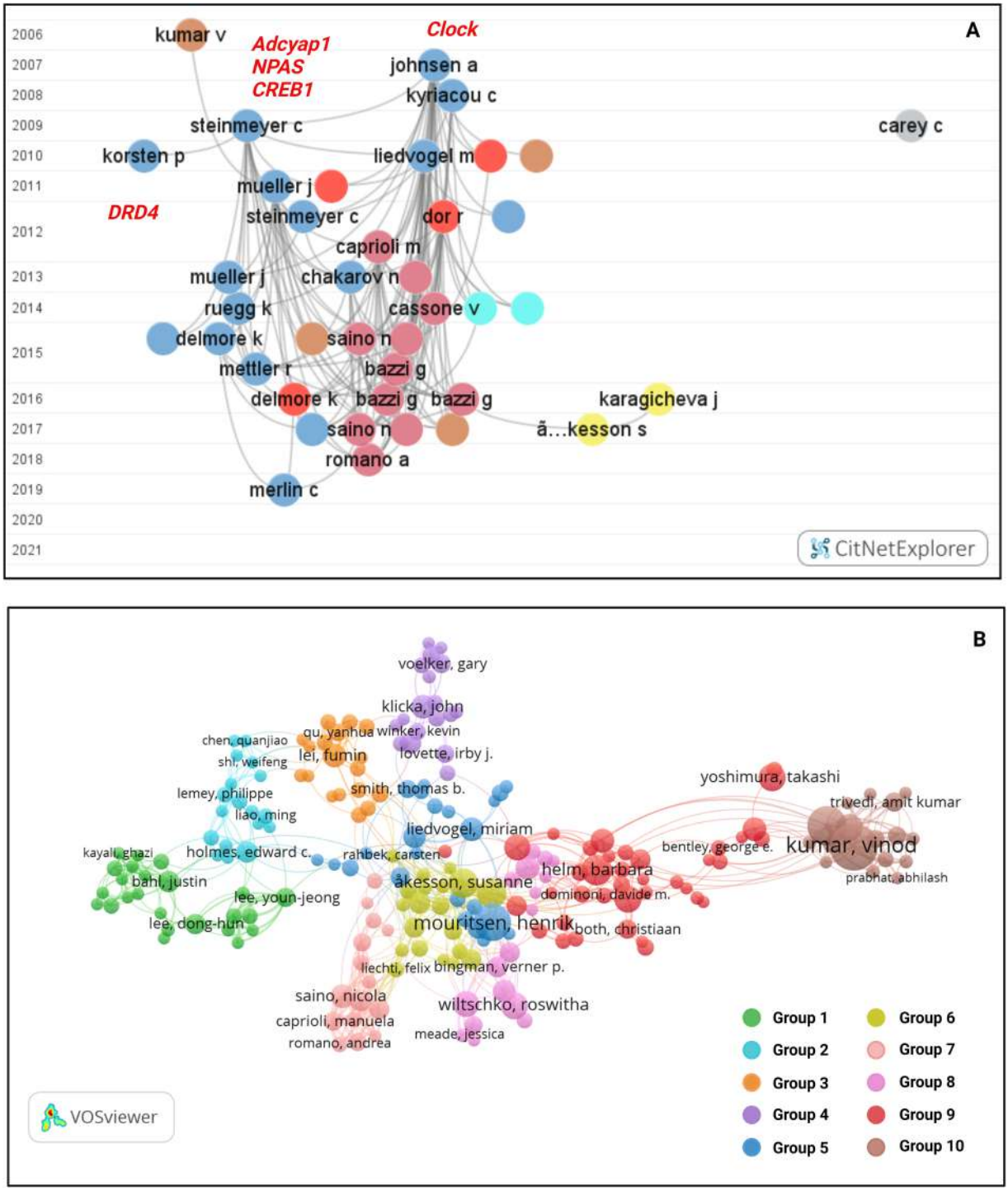


Fig. 2 Visualised citation network for studies identified in literature searches. (**A**) Citation network of the Scopus and PubMed database in CitNetExplorer. Publications are organized by year (2006–2021) with the name and first initial of the first author indicating individual studies. The relationship between studies by virtue of co-citations in the reference lists are indicated by grey lines. Subgroup analyses identified several key groups, indicated by the colour code from VOSviewer. Key candidate genes are indicated in red italics and show studies that assayed polymorphisms in the *Clock*, *Adcyap1*, *CREB1*, *NPAS*, and *DRD4* genes. (**B**) Citation network for studies identified in literature searches of the Dimensions and ScienceDirect database in VOSviewer. First authors are labelled by surname and first name. Automated group analyses identified ten clusters of related studies of which the studies identified from Scopus formed part of five groups, indicated as groups 2, 5, 6, 7, 9, and 10. This network shows the larger field of migration studies including non-candidate gene studies such as transcriptomic studies (group 10). (image edited in BioRender.com).

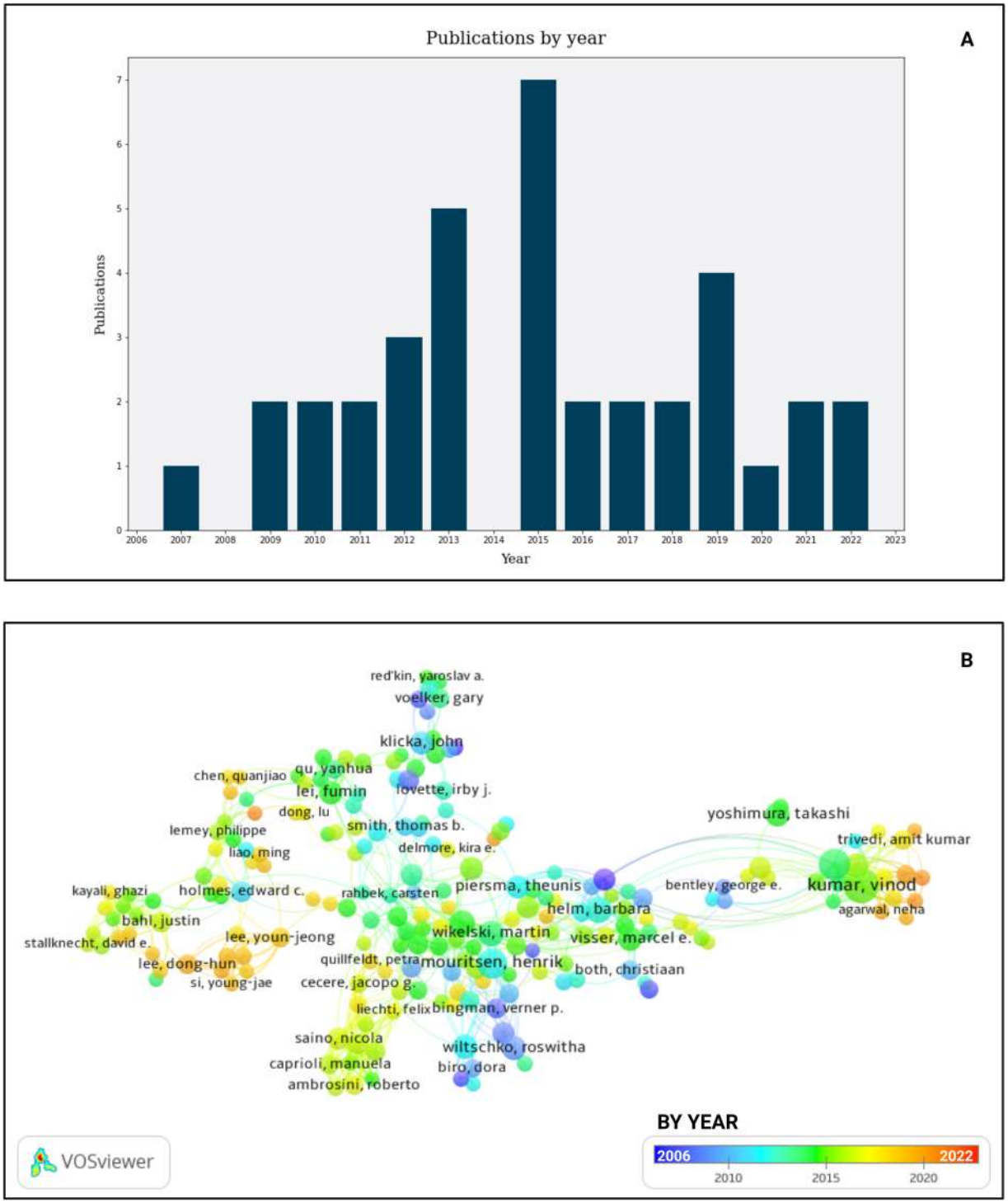


Fig. 3 Plots indicating the distribution for publications by year. (**A**) Histogram for publications by year indicating the first publications starting in 2007 up to more recent publications in 2022, with the largest number of publications between 2013–2015 and in 2019. (**B**) Density gradient display of studies in VOSviewer based on year of publication, indicated most studies were published between 2006 (blue) and 2022 (red) with a high number of publications emanating from 2013–2016 (green to orange). (image edited in BioRender.com).

monotypic but has a distinct resident population in Southern Africa. Understanding how differential migration is established and maintained between such species is key to assessing connectivity[14], speciation at a subspecies level[15], and potential population fitness[16]. This is particularly pertinent with regards to the plasticity or ability to switch between behaviours[17,18] should environmental conditions change considerably due to climate change[19–21] or anthropogenic activity[22–25].

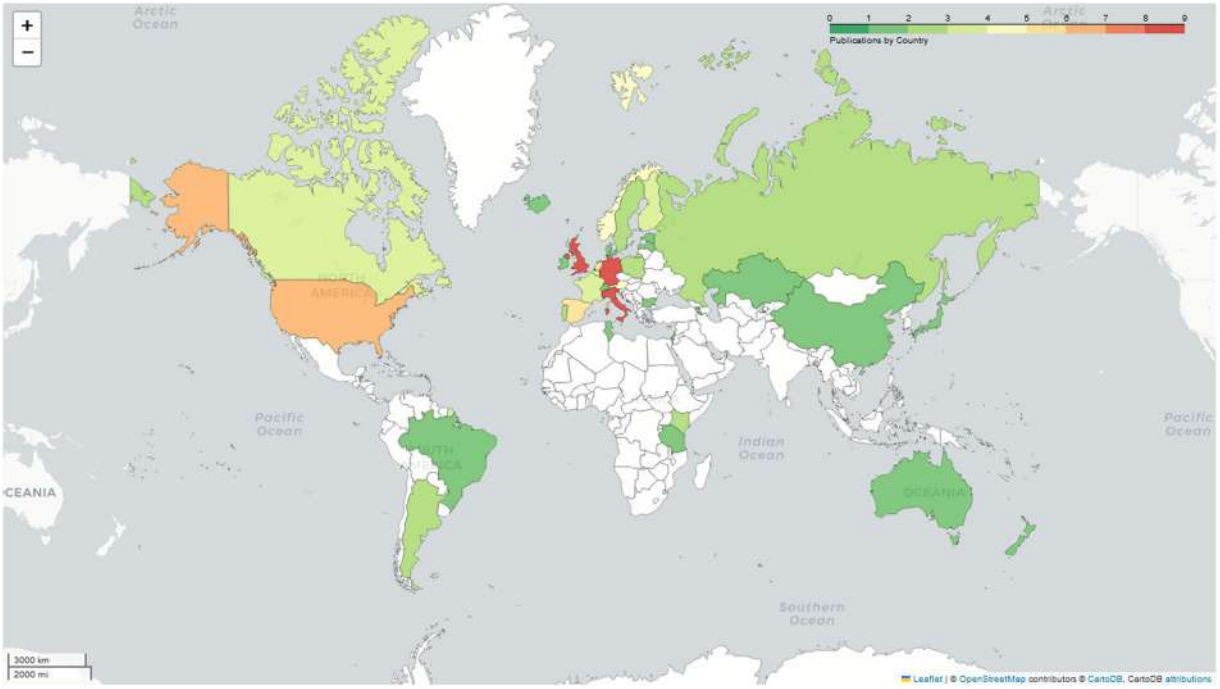


Fig. 4 Geographic distribution of candidate gene studies included in the final review dataset (N = 34) based on sampling locations. Related migration studies (N = 32), such as transcriptomic or epigenetic studies, were excluded. The density gradient plots the number of studies per country ranging from one study (green) to more than eight studies (red); countries in white are data deficient. The overall plot indicates that most studies emanated from sampling locations in Europe and North America, with only a small number of studies including sampling from parts of Africa and South America.

Several studies have explored the possible genetic components that affect intrinsic time keeping mechanisms and migration. Although variable methods have been used, including genomic[26], epigenetic[27], and transcriptomic approaches[28], most studies sought to identify genes or gene regions that show variation in either the sequence itself or the gene expression that can be correlated to divergent migratory behaviour. The key, however, is identifying variation that is linked to processes that interface with annual life events. Thus, variation that is either connected to the endocrine or metabolic changes[29], in preparation for migration and breeding, or intrinsic time-keeping mechanisms, such as the rhythmic expression of circadian genes; particularly those that interface with environmental changes that my serve as cues such as photoperiod, temperature, lunar cycles, and food availability[30]. This is needed to exclude variants that co-vary with migration phenotypes but are not actively involved in shaping them. It is therefore no surprise that many candidate gene studies have explored variation within the network of genes of the circadian clock. Several associated candidate genes have been suggested, with length polymorphisms within short repeats of the *Clock* and *Adcyap1* genes being the focus of many studies[31–33].

To clarify the role of these genes in migratory phenotypes, a systematic review (Fig. 1) was conducted to identify, synthesise, and provide a reappraisal of the available evidence[34]. Structured searches of the literature with an optimised Boolean search string were done in five scientific databases. Search results were exported in formats compatible with citation network analysis software[35]. After duplicate entries were removed, citation network analyses were used for the automated screening of database results to identify the central literature on the topic. Publications identified from the citation network analyses were subjected to manual screening of the title, abstract, and key words to assess the potential eligibility for inclusion in the review. The final list of most eligible publications was sought for full text retrieval. A total of 66 studies were included in the final review of which 34 were candidate gene studies and 32 were other, migration-related, studies. These included latitude/longitude/spatial analyses, timing of migration, and timing of egg laying/breeding. Most of the studies using a candidate gene approach were used for data retrieval. For these studies, datasets were retrieved as either diploid allele data of individuals or allele frequencies. Data sources included the main text of articles, supplementary materials, databases such as Dryad (https://datadryad.org/) or Figshare (https://figshare.com/), data extraction, or data received directly from authors. Unpublished data for an additional 12 species were also included. The dataset included individual level allele data from 52 species of which data was available for 46 species for the *Clock* gene and 43 species for the *Adcyap1* gene. This dataset represents the largest collection of cross species allele data for two candidate genes used to test a putative association between clock gene polymorphisms and divergent migration in birds, which enables the testing for patterns of inheritance, evolutionary selection, relation to divergence times, and associations across a globally distributed dataset.

This data descriptor summarises both the methodology used to screen the literature as well as to compile the data concisely and presents the resulting data used in prior analyses in an easy-to-understand format. At present, none of the scientific databases that collect genetic variation data is suitable for the deposit of this specific type

Common name	Latin binomial	Study	Data	Type	Location	N
Barn swallow*	*Hirundo rustica*	51,62,63	51	C_A	Switzerland, Italy	64
Bar-tailed godwit	*Limosa lapponica baueri*	64	64	C_A	New Zealand	135
Blackpoll warbler*	*Setophaga striata*	65	43	C_A, A_A	USA	72
Blue tit*	*Cyanistes caeruleus*	31,32,66,67	31,42,45	C_A	Europe	950
Collared flycatcher*	*Ficedula albicollis*	68	47	C_A, A_A	Czechia	406
Collared plover	*Charadrius collaris*	69	69	A_A	Brazil	14
Common buzzard	*Buteo buteo*	70	48	A_A	Germany	978
Common nightingale*	*Luscinia megarhynchos*	54,71	44, Authors	C_A, A_A	Italy	150
Common redstart*	*Phoenicurus phoenicurus*	54	Authors	C_A, A_A	Italy	43
Common whitethroat*	*Sylvia communis*	54	Authors	C_A, A_A	Italy	25
Dark-eyed junco*	*Junco hyemalis*	72	72	C_A, A_A	USA	36
Eastern subalpine warbler*	*Curruca cantillans*	54	Authors	C_A, A_A	Italy	31
Eurasian blackbird*	*Turdus merula*	22	22	C_A, A_A	Europe, Tunisia	792
Eurasian blackcap*	*Sylvia atricapilla*	73,74	50,73	A_A	Europe	936
Eurasian golden oriole*	*Oriolus oriolus*	54	Authors	C_A, A_A	Italy	30
Eurasian hoopoe*	*Upupa epops*	54	Authors	C_A, A_A	Italy	25
Eurasian reed warbler*	*Acrocephalus scirpaceus*	54	Authors	C_A, A_A	Italy	24
Eurasian wryneck*	*Jynx torquilla*	54	Authors	C_A, A_A	Italy	30
European bee-eater*	*Merops apiaster*	54	Authors	C_A, A_A	Italy	35
European nightjar*	*Caprimulgus europaeus*	54	Authors	C_A, A_A	Italy	39
European pied flycatcher*	*Ficedula hypoleuca*	71,75,76	44,49, Authors	C_A, A_A	Italy	226
European turtle dove*	*Streptopelia turtur*	54	Authors	C_A, A_A	Italy	29
Garden warbler*	*Sylvia borin*	54	Authors	C_A, A_A	Italy	31
Great reed warbler*	*Acrocephalus arundinaceus*	54	Authors	C_A, A_A	Italy	20
Icterine warbler*	*Hippolais icterina*	54	Authors	C_A, A_A	Italy	29
Northern wheatear*	*Oenanthe oenanthe*	54	Authors	C_A, A_A	Italy	30
Painted bunting*	*Passerina ciris*	77	77	C_A, A_A	USA	60
Sedge warbler*	*Acrocephalus schoenobaenus*	54	Authors	C_A, A_A	Italy	30
Semipalmated plover	*Charadrius semipalmatus*	69	69	A_A	Brazil	13
Semipalmated sandpiper	*Calidris pusilla*	69	69	A_A	Brazil	14
Spotted flycatcher*	*Muscicapa striata*	54	Authors	C_A, A_A	Italy	29
Spotted sandpiper	*Actitis macularius*	69	69	A_A	Brazil	12
Tree pipit*	*Anthus trivialis*	54,71	44, Authors	C_A, A_A	Italy	153
Tree swallow*	*Tachycineta bicolor*	16,78	46	C_A, A_A	Canada	921
Whinchat*	*Saxicola rubetra*	54,71	44, Authors	C_A, A_A	Italy	208
Willow warbler*	*Phylloscopus trochilus*	54–56	Authors	C_A, A_A	Italy	495
Wilson's warbler*	*Cardellina pusilla*	76	Authors	C_A, A_A	USA	102
Wood warbler*	*Phylloscopus sibilatrix*	54	Authors	C_A, A_A	Italy	30
Woodchat shrike*	*Lanius senator*	54	Authors	C_A, A_A	Italy	20
Yellow-legged gull	*Larus michahellis*	52	52	C_A, A_A	Italy	64

Table 1. List of species for which published allele data was collected and/or included in the review and data article. Species indicated with an asterisk (*) were included in the allele dataset for population genetics analyses[34]. The primary study, specific data source, location of the study sites and the sample size (N) is given. C_A: *Clock* gene alleles, A_A: *Adcyap1* gene alleles.

of data. The barcode of life data system (BOLD, https://boldsystems.org/), which does accept length polymorphism data from microsatellite markers, currently only accepts data for markers used in barcoding or population assignment experiments and does not specifically store data for markers used in behavioural or phenotype associated studies. The European variant archive (EVA, https://www.ebi.ac.uk/eva/), which also accepts variant data that includes length polymorphisms, currently only accepts data for species with reference genomes, which is still unavailable for most avian species. To overcome this, we have endeavoured to create a central compilation of the available data in two standard formats which is archived in parallel to this data descriptor; with an additional online version on GitHub[36] (https://github.com/LSLeClercq/AvianClocksData) that will be maintained and updated over time as more data is made available. This may greatly facilitate the reuse of the data where it may be applicable to other forms of analyses within migration genetics and beyond.

Common name	Latin binomial	Study	Data	Type	Location	N
African stonechat	*Saxicola torquatus*	[79]	[79]	C_F	Kenya, Tanzania	172
Asian short-toed lark	*Alaudala cheleensis*	[80]	[80]	C_F	China	257
Black swan	*Cygnus atratus*	[81]	[81]	Non-CA	Australia	100
Bluethroat	*Luscinia svecica*	[31]	[31]	C_F	Europe	369
Blue-winged warbler	*Vermivora cyanoptera*	[82]	[82]	Non-CA	USA	24
Canary Island stonechat	*Saxicola dacotiae*	[79]	[79]	C_F	Canary Islands	61
Chilean swallow	*Tachycineta meyeni*	[16]	[16]	C_F	Argentina	88
Common buzzard	*Buteo buteo*	[70]	[48]	C_F	Germany	978
Eurasian blackcap	*Sylvia atricapilla*	[73,74]	[50,73]	C_F	Europe	936
European roller	*Coracias garrulus*	[14]	[14]	Non-CA	Europe	32
European stonechat	*Saxicola rubicola*	[79]	[79]	C_F	Europe	382
Golden winged warbler	*Vermivora chrysoptera*	[82]	[82]	Non-CA	USA	42
Great tit	*Parus major*	[83–85]	[83]	C_F	UK	225
Mangrove swallow	*Tachycineta albilinea*	[16]	[16]	C_F	Belize	163
Mountain bluebird	*Sialia currucoides*	[18]	[18]	NA	Canada	11
Northern goshawk	*Accipiter gentilis*	[70]	[48]	C_F	Germany	15
Red kite	*Milvus milvus*	[70]	[48]	C_F, A_F	Germany	20
Seychelles warbler	*Acrocephalus sechellensis*	[86]	[86]	Non-CA	Seychelles	57
Siberian stonechat	*Saxicola maurus*	[79]	[79]	C_F	Kazakhstan, Japan	101
Song sparrow	*Melospiza melodia*	[87]	[87]	Non-CA	Canada	78
Violet-green swallow	*Tachycineta thalassina*	[16]	[16]	C_F	USA	48
Western bluebird	*Sialia mexicana*	[18]	[18]	NA	Canada	127
White-rumped swallow	*Tachycineta leucorrhoa*	[16]	[16]	C_F	Argentina	169
Yellow-eyed junco	*Junco phaeonotus*	[72]	[72]	C_F, A_F	USA	178

Table 2. List of species for which other published data was collected and/or included in the review and data article. The primary study, specific data source, location of the study sites and the sample size (N) is given. C_F: *Clock* gene frequencies, A_F: *Adcyap1* gene frequencies, Non-CA: Non clock gene study, NA: Not Available.

Methods

Literature search and automated screening. Literature was searched using systematic review methods, in line with PRISMA Ecology and Evolution guidelines[37], to identify and synthesize relevant sources. The overall approach is depicted in the PRISMA statement[38] in Fig. 1 that was supplemented with further information on the data retrieval and screening process. Literature was searched between January and September of 2022 on five databases: Scopus (N = 52, www.scopus.com), ScienceDirect (N = 1814, www.sciencedirect.com), Web of Science (N = 140, https://clarivate.com/), PubMed (N = 157, https://pubmed.ncbi.nlm.nih.gov/), and Dimensions (N = 2746, www.dimensions.ai). Databases were searched using an optimized Boolean search string derived from the PICO terms for the aim and objectives of the review. The final search string was as follows: ("Birds" OR "Avian") AND ("Clock genes" OR "Clock" OR "Adcyap1" OR "Candidate gene") AND ("Migration" OR "Flying"). As needed, this was complemented by ancillary 'free term' searches based on citations in articles or to include other relevant aspects such as "Breeding", "Moult", "Genomics", "Transcriptomics" or "Photoperiod". For the Scopus and Dimensions database searches, the results were exported in the comma separated value (CSV) format, while the results from the ScienceDirect, Web of Science, and PubMed database search were exported in the research information systems (RIS) format.

Automated screening for inclusion was done through citation network analyses. For the Scopus database, the results were merged and reformatted with the R package 'Scopus2CitNet 0.1.0.0' (https://github.com/MichaelBoireau/Scopus2CitNet) in RStudio 1.4.1106[39], running R 4.0.5[40]. The results were subsequently visualized by year in CitNetExplorer 1.0.0., keeping only those papers that overlapped in terms of references cited and the largest connected set (Fig. 2a). The results from the search on the Dimensions and ScienceDirect databases were visualized in VOSviewer 1.6.16[35] by group as well as by year, keeping only those papers that are connected by citations and reference lists (Fig. 2b). The size of bubbles corresponds to citations and the number of cross-links between studies.

Manual title-abstract screening and full text retrieval. Sources identified from the citation networks were imported (citation and abstract) into Mendeley citation manager (www.mendeley.com) for further screening. Several types of studies relating to migration genetics were included in preliminary screening such as candidate gene studies, genomic studies, transcriptomic studies, and epigenetic studies. Studies with a focus on endocrine systems, physiology, or telomeres were excluded. Studies on migration phenology, without an evident genetic link, were also excluded. The inclusion criteria of candidate gene studies were confined to studies that primarily measure *Clock* or *Adcyap1* gene polymorphisms (as well as other candidate genes studied in parallel e.g., *NPAS*, *CREB1*, and *DRD4*: indicated on Fig. 2[34]) within bird populations to compare putative variation to the annual synchronicity in life events and differential migration. These included latitude/longitude/spatial analyses,

Common name	Latin binomial	Study	Data	Type	Location	N
American redstart*	*Setophaga ruticilla*	[34]	Authors	C_A, A_A	Canada	26
Barolo shearwater	*Puffinus baroli*	[88]	Authors	C_A	Portugal	15
Boyd's shearwater	*Puffinus boydi*	[88]	Authors	C_A	Cape Verde	25
Common chiffchaff*	*Phylloscopus collybita*	[34]	Authors	C_A	Sweden, Kazakhstan	55
Common yellowthroat*	*Geothlypis trichas*	[34]	Authors	C_A, A_A	Canada	31
Great shearwater	*Ardenna gravis*	[88]	Authors	C_A	UK	25
Hermit thrush*	*Catharus guttatus*	[34]	Authors	C_A, A_A	Canada	30
Magnolia warbler*	*Setophaga magnolia*	[34]	Authors	C_A, A_A	Canada	33
Manx shearwater	*Puffinus puffinus*	[88]	Authors	C_A	Iceland	23
Swainson's thrush*	*Catharus ustulatus*	[34]	Authors	C_A, A_A	Canada	29
White-throated sparrow*	*Zonotrichia albicollis*	[34]	Authors	C_A, A_A	Canada	32
Yelkouan shearwater	*Puffinus yelkouan*	[88]	Authors	C_A	France	15

Table 3. List of species for which unpublished data was collected and/or included in the review and data article. Species indicated with an asterisk (*) were included in the allele dataset for population genetics analyses[34]. The primary study, specific data source, location of the study sites and the sample size (N) is given. C_A: *Clock* gene alleles, A_A: *Adcyap1* gene alleles.

timing of migration, migratory restlessness, timing of egg laying/breeding, clutch size, moult, urbanisation, and exploratory behaviour. The final set of studies that passed preliminary screening were sought during full text retrieval and added to the imported reference if it wasn't already included. A total of 66 studies were included in the final review of which 34 were candidate gene studies and 32 were other, migration related, studies using genetic methods. Some basic scientometric assessments of the final set of studies, including the plotting of publications per year (Fig. 3) as well as the geographic distribution (Fig. 4) of studies, was conducted using ABCal version 1.0.2[41] (https://github.com/LSLeClercq/ABCal).

Published datasets. A total of 34 studies were identified that used a candidate gene approach for which data retrieval was done. Data was retrieved from either the main text, supplementary material of the article, online data repositories such as Dryad[42–49] and Figshare[50], or additional data received directly from authors. Data types varied from allele frequencies to individual level diploid allele data. Allele data for the Barn swallow[51] was retrieved from the text while data for the Yellow-legged gull[52] was extracted from images using WebPlotDigitizer version 4.6[53]. Allele data was generally derived from a single source with the exception of the European pied flycatcher[44,49] and Willow warbler[54–56]. The species, data sources, and data types are summarized in Table 1 along with the sampling location and sample sizes. Frequency data was available for most published studies, with the exception of the bluebird species[18], and those species for which allele data was unavailable are summarised in Table 2. This includes species for which only frequency data was reported, species for which a non-clock gene approach was used, and studies for which only data summaries without frequencies were reported.

Unpublished datasets. This study included unpublished data for twelve species in total, summarised in Table 3. The six North American species were sampled at Long Point Old Cut, Ontario, Canada, and included the American redstart (N = 26), Common yellowthroat (N = 31), Hermit thrush (N = 30), Magnolia warbler (N = 33), Swainson's thrush (N = 29), and White-throated sparrow (N = 32). The six European species included the Common chiffchaff (N = 55) and five species of shearwaters: Barolo shearwater (N = 15), Boyd's shearwater (N = 25), Great shearwater (N = 25), Manx shearwater (N = 23), and Yelkouan shearwater (N = 15). The Common chiffchaff was sampled from several locations in Sweden (N = 30, subspecies *abietinus*) and Kazakhstan (N = 25, spp. *tristis*). Blood samples were taken from the brachial vein and stored in SET buffer at −80 °C. Shearwaters were sampled from several locations in Europe including France and Portugal while several species were sampled from islands such as Iceland, Cape Verde, and territories of the United Kingdom such as Gough Island. A 1 ml blood sample was taken from the tarsal or the brachial vein during geolocator retrieval. Samples were collected in 1.5 ml plastic tubes containing 70% ethanol and stored at −20 °C until further analysis.

Samples were genotyped using established methods[54]. Briefly, samples of North American species were preserved in a buffer at room temperature until extraction with the ArchivePure DNA purification kit (5 PRIME, Hilden, Germany). Then, polymorphism at *Clock* and *Adcyap1* 3′-UTR was determined as before[54], with PCR products labelled with HEX (*Clock*), 6-FAM (*Clock* and *Adcyap1*) or TAMRA (*Adcyap1*) dyes. For the Common chiffchaff, genomic DNA was extracted using a standard ammonium acetate protocol. All 55 samples were successfully genotyped and analysed for length polymorphism in the poly-Q repeat of the *Clock* gene following previously published protocols[31]. For Shearwater samples, total genomic DNA was extracted from blood samples using the Speedtools® Tissue DNA Extraction kit (Biotools, Madrid, Spain) following the manufacturer's instructions. Genotyping was subsequently performed with methods adapted from the literature[31]. Briefly, PCR products were generated with shearwater specific primers for the *Clock* gene labelled with 6-FAM or HEX, followed by fragment analysis as in[54] to determine the size of the poly-Q repeat.

Field name	Data
General (Index):	
Species	Common name in English for species
Clock	Logical binary for data availability of *Clock* gene e.g., "Yes" or "No"
Adcyap1	Logical binary for data availability of *Adcyap1* gene e.g., "Yes" or "No"
Code	Abbreviation used for species tabs
Sample (N)	Size (N) of the total individuals for which data are available
Taxonomy (Index):	
Genus	Latin name for genus e.g., "Hirundo"
Species	Latin name for species e.g., "rustica"
Family	Latin name for family e.g., "Hirundinidae"
Superfamily	Latin name for superfamily e.g., "Locustelloidea"
Parvorder	Latin name for parvorder e.g., "Sylviida"
Order	Latin name for order e.g., "Passeriformes"
Species sheet:	
Species	Common name in English for species
Sample ID	Sample ID used in raw data for individuals
Clock 1	1st diploid allele for *Clock* gene (individual) as Q_N
Clock 2	2nd diploid allele for *Clock* gene (individual) as Q_N
Sum	Sum of two alleles for *Clock* gene as Q_N
Mean	Mean value of diploid alleles for *Clock* gene as Q_N
Adcyap 1	1st diploid allele for *Adcyap1* gene (individual) in base pairs (bp)
Adcyap 2	2nd diploid allele for *Adcyap1* gene (individual) in base pairs (bp)
Sum	Sum of two alleles for *Adcyap1* gene in base pairs (bp)
Mean	Mean value of diploid alleles for *Adcyap1* gene in base pairs (bp)

Table 4. Description of field names and data for workbook and CSV files.

Data Records

The data collated during the systematic review and meta-analysis were made available to via the Zenodo repository at the time of publication. Additional inclusion and exclusion criteria were applied and a final set of 40 species (indicated by asterisk in Tables 1, 3) were included in the comparative analyses using mantel and phylogenetic generalised least squares methods to test for an association between migratory phenotypes and candidate gene genotypes[34,57]. This data are available on Zenodo[57], and includes a workbook with the allele data as well as a results workbook with various population genetics measures including allele frequencies, Homozygosity (H_o), Heterozygosity (H_e), Hardy-Weinberg equilibrium[58,59], and Ewens-Watterson[60] results. The complete dataset was reformatted for distribution with this data descriptor and is available from two sources, from the Figshare[61] depository, as submitted with this article, and from a maintained repository with version histories on GitHub[36].

Data (version 1.0.2) are available as a spreadsheet workbook, labelled "Avian Clock Gene Dataset" with multiple sheets. The first sheet of the workbook, labelled "Index", contains the table of contents which has several columns (Table 4) that list species by common names, indicates data availability for *Clock* and *Adcyap1*, and total sample size (N). Furthermore, the taxonomic classifications including genus, species, family, superfamily, parvorder, and order are also given. The species codes are hyperlinked to the allele data for individual species, contained in separate sheets within the same workbook. Individual sheets for species contain several columns including the species name, sample ID, and diploid alleles for *Clock* and/or *Adcyap1* genes. Alleles are expressed as the number of polyglutamine repeats (Q_N) for *Clock* while the *Adcyap1* alleles represent the amplified fragment length in base pairs (bp). The sum and average of alleles is also provided, and missing data is labelled as NA. For the purpose of individual species analyses, the species sheets from the workbook are also provided as individual comma separated value (CSV) files. The same data is also available on GitHub with the workbooks available in the root directory while the individual CSV files are available in a subfolder with the title "CSV". The repository also contains a "README" file which provides some basic background and details on the data. Both the workbook as well as CSV files can be read by Microsoft® Office (https://www.office.com/) as well as StarOffice™ (https://www.staroffice.com/), OpenOffice™ (https://www.openoffice.org/), and LibreOffice™ (https://www.libreoffice.org/).

Technical Validation

Allele data comprises the heterozygous or homozygous diploid allele for one or both studied clock genes as well as the sum and average of allele sizes. The data for *Clock* was normalized according to the poly-glutamine repeat size (Q_N) by subtracting the conserved non-repeat size (L_C) in base pairs from the total fragment size (L_T) and dividing by codon size, following Eq. 1.

$$Q_N = (L_T - L_C)/3 \quad (1)$$

Data for *Adcyap1* was generated using the same published primers and was kept as the total fragment size.

Code availability

The custom R code used to convert data retrieved from Scopus to the appropriate format for visualisation in CitNetExplorer is available from GitHub (https://github.com/MichaelBoireau/Scopus2CitNet). The custom PYTHON script used for plotting the scientometric aspects of the included literature is also available from GitHub (https://github.com/LSLeClercq/ABCal).

Received: 7 July 2023; Accepted: 1 November 2023;
Published online: 09 November 2023

References

1. Leclerc, B. *et al*. Photoperiodic modulation of clock gene expression in the avian premammillary nucleus. *J. Neuroendocrinol.* **22**, 119–128 (2010).
2. Pancerasa, M., Ambrosini, R., Saino, N. & Casagrandi, R. Barn swallows long-distance migration occurs between significantly temperature-correlated areas. *Sci. Rep.* **8**, 12359 (2018).
3. Saino, N. *et al*. Temperature and rainfall anomalies in Africa predict timing of spring migration in trans-Saharan migratory birds. *Clim. Res.* **35**, 123–134 (2007).
4. Studds, C. E. & Marra, P. P. Linking fluctuations in rainfall to nonbreeding season performance in a long-distance migratory bird, *Setophaga ruticilla*. *Clim. Res.* **35**, 115–122 (2007).
5. Hau, M. & Gwinner, E. Food as a circadian zeitgeber for house sparrows: The effect of different food access durations. *J. Biol. Rhythms* **11**, 196–207 (1996).
6. Sockman, K. W. & Hurlbert, A. H. How the effects of latitude on daylight availability may have influenced the evolution of migration and photoperiodism. *Funct. Ecol.* **34**, 1752–1766 (2020).
7. Norevik, G., Åkesson, S., Andersson, A., Bäckman, J. & Hedenström, A. The lunar cycle drives migration of a nocturnal bird. *PLoS Biol.* **17**, e3000456 (2019).
8. Pyle, P., Nur, N., Henderson, R. P. & DeSante, D. F. The Effects of Weather and Lunar Cycle on Nocturnal Migration of Landbirds at Southeast Farallon Island, California. *Condor* **95**, 343 (1993).
9. Roques, S. *et al*. When to depart from a stopover site? Time since arrival matters more than current weather conditions. *Ornithology* **139**, 1–13 (2022).
10. Aguilar-Roblero, R., Díaz-Muñoz, M. & Fanjul-Moles, M. L. *Mechanisms of circadian systems in animals and their clinical relevance*. *Mechanisms of Circadian Systems in Animals and Their Clinical Relevance*. https://doi.org/10.1007/978-3-319-08945-4 (Springer International Publishing, 2015).
11. Fry, H., Kirwan, G. M. & Boesman, P. F. D. Lilac-breasted Roller (Coracias caudatus). in *Birds of the World* (eds. del Hoyo, J., Elliott, A., Sargatal, J., Christie, D. & de Juana, E.) https://doi.org/10.2173/bow.librol2.01.1 (Cornell Lab of Ornithology, 2021).
12. Woodall, P. F. Woodland Kingfisher (Halcyon senegalensis), version 1.0. in *Birds of the World* (eds. *et al*). https://doi.org/10.2173/BOW.WOOKIN1.01 (Cornell Lab of Ornithology, 2020).
13. Fry, H. & Boesman, P. F. D. European Bee-eater (Merops apiaster), version 1.0. in *Birds of the World* (eds. del Hoyo, J., Elliott, A., Sargatal, J., Christie, D. A. & de Juana, E.). https://doi.org/10.2173/BOW.EUBEAT1.01 (Cornell Lab of Ornithology, 2020).
14. Finch, T. *et al*. A pan-European, multipopulation assessment of migratory connectivity in a near-threatened migrant bird. *Divers. Distrib.* **21**, 1051–1062 (2015).
15. Mwale, M. *et al*. Genetic and morphological variation of Woodland Kingfisher Halcyon senegalensis reveals cryptic mitochondrial lineages and patterns of mitochondrial–nuclear discordance. *Ostrich* **93**, 192–207 (2022).
16. Dor, R. *et al*. Clock gene variation in tachycineta swallows. *Ecol. Evol.* **2**, 95–105 (2012).
17. Hegemann, A., Fudickar, A. M. & Nilsson, J. Å. A physiological perspective on the ecology and evolution of partial migration. *J. Ornithol. 2019 1603* **160**, 893–905 (2019).
18. Sauve, D., Dale, C. A., Tigano, A., Ratcliffe, L. M. & Friesen, V. L. Do candidate genes for migration and behavior explain migratory variation in bluebirds (*Sialia* spp.)? *Wilson J. Ornithol.* **132**, 820–829 (2021).
19. Carey, C. The impacts of climate change on the annual cycles of birds. *Philos. Trans. R. Soc. B Biol. Sci.* **364**, 3321–3330 (2009).
20. Saino, N. *et al*. Climate change effects on migration phenology may mismatch brood parasitic cuckoos and their hosts. *Biol. Lett.* **5**, 539–541 (2009).
21. Beresford, A. E. *et al*. Phenology and climate change in Africa and the decline of Afro-Palearctic migratory bird populations. *Remote Sens. Ecol. Conserv.* **5**(1), 55–69 (2019).
22. Mueller, J. C., Partecke, J., Hatchwell, B. J., Gaston, K. J. & Evans, K. L. Candidate gene polymorphisms for behavioural adaptations during urbanization in blackbirds. *Mol. Ecol.* **22**, 3629–3637 (2013).
23. Güneralp, B., Lwasa, S., Masundire, H., Parnell, S. & Seto, K. C. Urbanization in Africa: Challenges and opportunities for conservation. *Environ. Res. Lett.* **13**, 015002 (2018).
24. Navarro, L. M. & Pereira, H. M. Rewilding abandoned landscapes in Europe. in *Rewilding European Landscapes* 3–23. https://doi.org/10.1007/978-3-319-12039-3_1 (2015).
25. Hoogendoorn, G., Meintjes, D., Kelso, C. & Fitchett, J. Tourism as an incentive for rewilding: the conversion from cattle to game farms in Limpopo province, South Africa. *J. Ecotourism* **18**, 309–315 (2019).
26. Delmore, K. E. *et al*. Genomic analysis of a migratory divide reveals candidate genes for migration and implicates selective sweeps in generating islands of differentiation. *Mol. Ecol.* **24**, 1873–1888 (2015).
27. Saino, N. *et al*. Inter-generational resemblance of methylation levels at circadian genes and associations with phenology in the barn swallow. *Sci. Rep.* **9**, 1–16 (2019).
28. Frias-Soler, R. C., Pildaín, L. V., Pårâu, L. G., Wink, M. & Bairlein, F. Transcriptome signatures in the brain of a migratory songbird. *Comp. Biochem. Physiol. - Part D Genomics Proteomics* **34**, 100681 (2020).
29. Sharma, A., Tripathi, V. & Kumar, V. Control and adaptability of seasonal changes in behavior and physiology of latitudinal avian migrants: Insights from laboratory studies in Palearctic-Indian migratory buntings. *J. Exp. Zool. Part A Ecol. Integr. Physiol.* **337**, 902–918 (2022).
30. Åkesson, S. *et al*. Timing avian long-distance migration: from internal clock mechanisms to global flights. *Philos. Trans. R. Soc. B Biol. Sci.* **372**, 20160252 (2017).
31. Johnsen, A. *et al*. Avian Clock gene polymorphism: Evidence for a latitudinal cline in allele frequencies. *Mol. Ecol.* **16**, 4867–4880 (2007).
32. Steinmeyer, C., Mueller, J. C. & Kempenaers, B. Search for informative polymorphisms in candidate genes: Clock genes and circadian behaviour in blue tits. *Genetica* **136**, 109–117 (2009).
33. Steinmeyer, C., Kempenaers, B. & Mueller, J. C. Testing for associations between candidate genes for circadian rhythms and individual variation in sleep behaviour in blue tits. *Genetica* **140**, 219–228 (2012).
34. Le Clercq, L. *et al*. Time trees and clock genes: a systematic review and comparative analysis of contemporary avian migration genetics. *Biol. Rev.* **98**, 1051–1080 (2023).

35. van Eck, N. J. & Waltman, L. Citation-based clustering of publications using CitNetExplorer and VOSviewer. *Scientometrics* **111**, 1053–1070 (2017).
36. Le Clercq, L. S. AvianClocksData: Dataset of Clock and Adcyap1 alleles for Birds. Version 1.0.2. https://github.com/LSLeClercq/AvianClocksData (2023).
37. O'Dea, R. E. *et al*. Preferred reporting items for systematic reviews and meta-analyses in ecology and evolutionary biology: a PRISMA extension. *Biol. Rev.* **96**, 1695–1722 (2021).
38. Haddaway, N. R., Page, M. J., Pritchard, C. C. & McGuinness, L. A. PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. *Campbell Syst. Rev.* **18**, e1230 (2022).
39. RStudio Team. RStudio: Integrated Development Environment for R. Version 4.0.5. http://www.rstudio.com/ (2021).
40. R Core Team. R: A Language and Environment for Statisitical Computing. Version 1.4.1106. https://www.r-project.org/ (2020).
41. Le Clercq, L. S. ABCal: Author Bias Computation and Scientometric Plotting. Version 1.0.2. https://github.com/LSLeClercq/ABCal (2023).
42. Liedvogel, M., Cornwallis, C. K. & Sheldon, B. C. Data from: Integrating candidate gene and quantitative genetic approaches to understand variation in timing of breeding in wild tit populations. *Dryad Dataset*. https://doi.org/10.5061/dryad.6n3d6109 (2012).
43. Ralston, J. *et al*. Data from: Length polymorphisms at two candidate genes explain variation of migratory behaviors in blackpoll warblers (Setophaga striata). *Dryad Dataset*. https://doi.org/10.5061/dryad.d10qb58 (2019).
44. Saino, N. *et al*. Data from: Polymorphism at the Clock gene predicts phenology of long-distance migration in birds. *Dryad Dataset* https://doi.org/10.5061/dryad.j22j1 (2015).
45. Liedvogel, M., Szulkin, M., Knowles, S., Wood, M. & Sheldon, B. Data from: Phenotypic correlates of Clock gene variation in a wild blue tit population: evidence for a role in seasonal timing of reproduction. *Dryad Dataset*. https://doi.org/10.5061/dryad.1303 (2010).
46. Bourret, A. & Garant, D. Data from: Candidate gene-environment interactions and their relationships with timing of breeding in a wild bird population. *Dryad Dataset* https://doi.org/10.5061/dryad.f7t25 (2015).
47. Krist, M., Munclinger, P., Briedis, M. & Adamík, P. Data from: The genetic regulation of avian migration timing: combining candidate genes and quantitative genetic approaches in a long-distance migrant. *Dryad Dataset*. https://doi.org/10.5061/dryad.gb5mkkwpr (2021).
48. Chakarov, N., Jonker, R. M., Boerner, M., Hoffman, J. I. & Krüger, O. Data from: Variation at phenological candidate genes correlates with timing of dispersal and plumage morph in a sedentary bird of prey. *Dryad Dataset* https://doi.org/10.5061/dryad.r058p (2013).
49. Kuhn, K. *et al*. Data from: Differentiation in neutral genes and a candidate gene in the pied flycatcher: using biological archives to track global climate change. *Dryad Dataset* https://doi.org/10.5061/dryad.6dc6k (2014).
50. Mettler, R., Segelbacher, G. & Martin Schaefer, H. Fileset: PONE-D-15-12289. *Figshare Dataset*. https://doi.org/10.6084/m9.figshare.1492930.v1 (2016).
51. Bazzi, G. *et al*. *Clock* gene polymorphism and scheduling of migration: A geolocator study of the barn swallow *Hirundo rustica*. *Sci. Rep.* **5**, 12443 (2015).
52. Romano, A. *et al*. Circadian genes polymorphism and breeding phenology in a resident bird, the yellow-legged gull. *J. Zool.* **304**, 117–123 (2018).
53. Rohatgi, A. WebPlotDigitizer: Version 4.6. https://automeris.io/WebPlotDigitizer (2022).
54. Bazzi, G. *et al*. Clock gene polymorphism, migratory behaviour and geographic distribution: a comparative study of trans-Saharan migratory birds. *Mol. Ecol.* **25**, 6077–6091 (2016).
55. Bazzi, G. *et al*. Candidate genes have sex-specific effects on timing of spring migration and moult speed in a long-distance migratory bird. *Curr. Zool.* **63**, 479–486 (2017).
56. Sokolovskis, K. *et al*. Phenotypic and genetic characterization of the East Siberian Willow Warbler (*Phylloscopus trochilus yakutensis* Ticehurst, 1935) in relation to the European subspecies. *J. Ornithol.* **160**, 721–731 (2019).
57. Le Clercq, L. S. *et al*. Time trees and Clock genes: a Systematic Review and Comparative Analysis of Contemporary Avian Migration Genetics (Dataset). *Zenodo*. https://doi.org/10.5281/ZENODO.6637839 (2022).
58. Hardy, G. H. Mendelian proportions in a mixed population. *Science*. **28**, 49–50 (1908).
59. Weinberg, W. On the demonstration of heredity in man. *Pap. Hum. Genet*. (1963).
60. Slatkin, M. An exact test for neutrality based on the Ewens sampling distribution. *Genet. Res.* **64**, 71–74 (1994).
61. Le Clercq, L. S. *et al*. Birds of a feather flock together—A compilation of data for two candidate genes for migration genetics. *Figshare Dataset*. https://doi.org/10.6084/m9.figshare.23623935.v1 (2023).
62. Dor, R. *et al*. Low variation in the polymorphic *Clock* gene poly-Q region despite population genetic structure across Barn Swallow (*Hirundo rustica*) populations. *PLoS One* **6**, e0028843 (2011).
63. Caprioli, M. *et al*. *Clock* gene variation is associated with breeding phenology and maybe under directional selection in the migratory barn swallow. *PLoS One* **7**, e0035140 (2012).
64. Parody-Merino, Á. M., Battley, P. F., Conklin, J. R. & Fidler, A. E. No evidence for an association between *Clock* gene allelic variation and migration timing in a long-distance migratory shorebird (*Limosa lapponica baueri*). *Oecologia* **191**, 843–859 (2019).
65. Ralston, J. *et al*. Length polymorphisms at two candidate genes explain variation of migratory behaviors in blackpoll warblers (*Setophaga striata*). *Ecol. Evol.* **9**, 8840–8855 (2019).
66. Liedvogel, M., Szulkin, M., Knowles, S. C. L., Wood, M. J. & Sheldon, B. C. Phenotypic correlates of *clock* gene variation in a wild blue tit population: Evidence for a role in seasonal timing of reproduction. *Mol. Ecol.* **18**, 2444–2456 (2009).
67. Liedvogel, M., Cornwallis, C. K. & Sheldon, B. C. Integrating candidate gene and quantitative genetic approaches to understand variation in timing of breeding in wild tit populations. *J. Evol. Biol.* **25**, 813–823 (2012).
68. Krist, M., Munclinger, P., Briedis, M. & Adamík, P. The genetic regulation of avian migration timing: combining candidate genes and quantitative genetic approaches in a long-distance migrant. *Oecologia* **196**, 373–387 (2021).
69. de Almeida Miranda, D. *et al*. Shorebirds' Longer Migratory Distances Are Associated With Larger ADCYAP1 Microsatellites and Greater Morphological Complexity of Hippocampal Astrocytes. *Front. Psychol.* **12**, 6679 (2022).
70. Chakarov, N., Jonker, R. M., Boerner, M., Hoffman, J. I. & Krüger, O. Variation at phenological candidate genes correlates with timing of dispersal and plumage morph in a sedentary bird of prey. *Mol. Ecol.* **22**, 5430–5440 (2013).
71. Saino, N. *et al*. Polymorphism at the Clock gene predicts phenology of long-distance migration in birds. *Mol. Ecol.* **24**, 1758–1773 (2015).
72. Peterson, M. P. *et al*. Variation in candidate genes CLOCK and ADCYAP1 does not consistently predict differences in migratory behavior in the songbird genus Junco. *F1000Research* **2**, 115 (2013).
73. Mueller, J. C., Pulido, F. & Kempenaers, B. Identification of a gene associated with avian migratory behaviour. *Proc. R. Soc. B Biol. Sci.* **278**, 2848–2856 (2011).
74. Mettler, R., Segelbacher, G. & Schaefer, H. M. Interactions between a Candidate Gene for Migration (ADCYAP1), Morphology and Sex Predict Spring Arrival in Blackcap Populations. *PLoS One* **10**, e0144587 (2015).
75. Kuhn, K. *et al*. Differentiation in neutral genes and a candidate gene in the pied flycatcher: Using biological archives to track global climate change. *Ecol. Evol.* **3**, 4799–4814 (2013).
76. Bazzi, G. *et al*. *Adcyap1* polymorphism covaries with breeding latitude in a Nearctic migratory songbird, the Wilson's warbler (*Cardellina pusilla*). *Ecol. Evol.* **6**, 3226–3239 (2016).

77. Contina, A., Bridge, E. S., Ross, J. D., Shipley, J. R. & Kelly, J. F. Examination of *Clock* and *Adcyap1* gene variation in a neotropical migratory passerine. *PLoS One* **13**, e0190859 (2018).
78. Bourret, A. & Garant, D. Candidate gene-environment interactions and their relationships with timing of breeding in a wild bird population. *Ecol. Evol.* **5**, 3628–3641 (2015).
79. Justen, H. *et al.* Population-specific association of *Clock* gene polymorphism with annual cycle timing in stonechats. *Sci. Rep.* **12**, 1–13 (2022).
80. Zhang, S., Xu, X., Wang, W., Yang, W. & Liang, W. Clock gene is associated with individual variation in the activation of reproductive endocrine and behavior of Asian short toed lark. *Sci. Rep.* **7**, 1–8 (2017).
81. Van Dongen, W. F. D., Robinson, R. W., Weston, M. A., Mulder, R. A. & Guay, P. J. Variation at the DRD4 locus is associated with wariness and local site selection in urban black swans. *BMC Evol. Biol.* **15**, 253 (2015).
82. Toews, D. P. L., Taylor, S. A., Streby, H. M., Kramer, G. R. & Lovette, I. J. Selection on VPS13A linked to migration in a songbird. *Proc. Natl. Acad. Sci. USA* **116**, 18272–18274 (2019).
83. Liedvogel, M. & Sheldon, B. C. Low variability and absence of phenotypic correlates of *Clock* gene variation in a great tit *Parus major* population. *J. Avian Biol.* **41**, 543–550 (2010).
84. Korsten, P. *et al.* Association between DRD4 gene polymorphism and personality variation in great tits: A test across four wild populations. *Mol. Ecol.* **19**, 832–843 (2010).
85. Mueller, J. C. *et al.* Haplotype structure, adaptive history and associations with exploratory behaviour of the DRD4 gene region in four great tit (*Parus major*) populations. *Mol. Ecol.* **22**, 2797–2809 (2013).
86. Edwards, H. A., Hajduk, G. K., Durieux, G., Burke, T. & Dugdale, H. L. No association between personality and candidate gene polymorphisms in a wild bird population. *PLoS One* **10**, e0138439 (2015).
87. Posliff, C. D. Correlations in movement behaviour over large and small geographic scales in song sparrows (*Melospiza melodia*). (Western University of Ontario, 2020).
88. Ferrer Obiol, J. New insights into the genetic base of bird migration: a population genetics study based on candidate genes. (University of Barcelona, 2015).

Acknowledgements

Open access provided by the library of University of the Free State (Grant number: 13560591). The authors would like to thank Chris Guglielmo, Yolanda Morbey, and Ivan Maggini for support with field data collection at Long Point Bird Observatory (Canada), Staffan Bensch for field data collection in both Sweden and Kazakhstan, and Andrea Galimberti, Cristina Possenti, Fernando Spina and the late Nicola Saino for support with genetic analyses and field data collection in Ventotene Island (Italy). Images were created/edited in BioRender.com. This work is based on the research supported wholly/in part by the National Research Foundation of South Africa (Grant Number: 112062).

Author contributions

Le Clercq, L.S.: Conceptualisation, Methodology, Data Curation, Formal Analysis, Writing – Original Draft, Writing – Review & Editing, Funding acquisition. Bazzi, G.: Resources, Data Curation. Ferrer Obiol, J.: Resources, Data Curation, Writing - Review & Editing. Cecere, J.G.: Resources, Data Curation, Funding acquisition. Gianfranceschi, L.: Resources, Data Curation. Grobler, J.P.: Writing – Review & Editing, Resources, Supervision. Kotzé, A.: Writing – Review & Editing, Resources, Supervision. Riutort León, M.: Resources, Data Curation. González-Solís, J.: Resources, Data Curation. Rubolini, D.: Resources, Data Curation, Writing - Review & Editing, Funding acquisition. Liedvogel, M.: Data Curation, Writing - Review & Editing. Dalton, D.L.: Conceptualisation, Writing – Review & Editing, Resources, Supervision.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Additional information

Correspondence and requests for materials should be addressed to L.-S.L.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Journal of Molecular Evolution (2023) 91:502–513
https://doi.org/10.1007/s00239-023-10106-3

ORIGINAL ARTICLE

PAReTT: A Python Package for the Automated Retrieval and Management of Divergence Time Data from the TimeTree Resource for Downstream Analyses

Louis-Stéphane Le Clercq[1,2] · Antoinette Kotzé[1,2] · J. Paul Grobler[2] · Desiré Lee Dalton[3]

Received: 10 January 2023 / Accepted: 24 March 2023 / Published online: 20 April 2023

Abstract
Evolutionary processes happen gradually over time and are, thus, considered time dependent. In addition, several evolutionary processes are either adaptations to local habitats or changing habitats, otherwise restricted thereby. Since evolutionary processes driving speciation take place within the landscape of environmental and temporal bounds, several published studies have aimed at providing accurate, fossil-calibrated, estimates of the divergence times of both extant and extinct species. Correct calibration is critical towards attributing evolutionary adaptations and speciation both to the time and paleogeography that contributed to it. Data from more than 4000 studies and nearly 1,50,000 species are available from a central TimeTree resource and provide opportunities of retrieving divergence times, evolutionary timelines, and time trees in various formats for most vertebrates. These data greatly enhance the ability of researchers to investigate evolution. However, there is limited functionality when studying lists of species that require batch retrieval. To overcome this, a PYTHON package termed Python-Automated Retrieval of TimeTree data (PAReTT) was created to facilitate a biologist-friendly interaction with the TimeTree resource. Here, we illustrate the use of the package through three examples that includes the use of timeline data, time tree data, and divergence time data. Furthermore, PAReTT was previously used in a meta-analysis of candidate genes to illustrate the relationship between divergence times and candidate genes of migration. The PAReTT package is available for download from GitHub or as a pre-compiled Windows executable, with extensive documentation on the package available on GitHub wiki pages regarding dependencies, installation, and implementation of the various functions.

Keywords PAReTT · PYTHON · Time trees · Divergence time · Timelines · Diversification rate

Introduction

Evolutionary processes are linked to time, be it diversification within a lineage which may lead to the emergence of a new species, or via subtle molecular changes over several generations steadily driving phenotypic variation (Wagner 2018; Francisco Henao Diaz et al. 2019), leading to speciation at the subspecies or ecotype level. For example, some primary divisions between entire taxonomic orders of birds arose approximately 75 million years ago while more recent divisions between subspecies occurred as recently as 1 million years ago (Prum et al. 2015; Mwale et al. 2022).

Most evolutionary processes are context dependent and are, thus, studied as part of ecological (Olson et al. 2001) and geographic changes (Linck et al. 2020). Both shape the landscape within which selection, adaptation, and extinction take place (Wu et al. 2022), including the historical geography or paleogeography at the time of divergence and speciation (Scotese 2016; Müller et al. 2018). This includes the factorization of well-established phenomena that preceded contemporary geography such as continental drift and known major periods of glaciation (Fig. 1), which can only be factored into the evolutionary history of a lineage or species if the time periods for diversification are known. For example, some species considered to be conspecific based

Handling editor: Arndt von Haesele.

✉ Louis-Stéphane Le Clercq
leclercq.l.s@gmail.com

[1] South African National Biodiversity Institute, Pretoria 0001, South Africa

[2] Department of Genetics, University of the Free State, Bloemfontein 9300, South Africa

[3] School of Health and Life Sciences, Teesside University, Middlesbrough TS1 3BA, UK

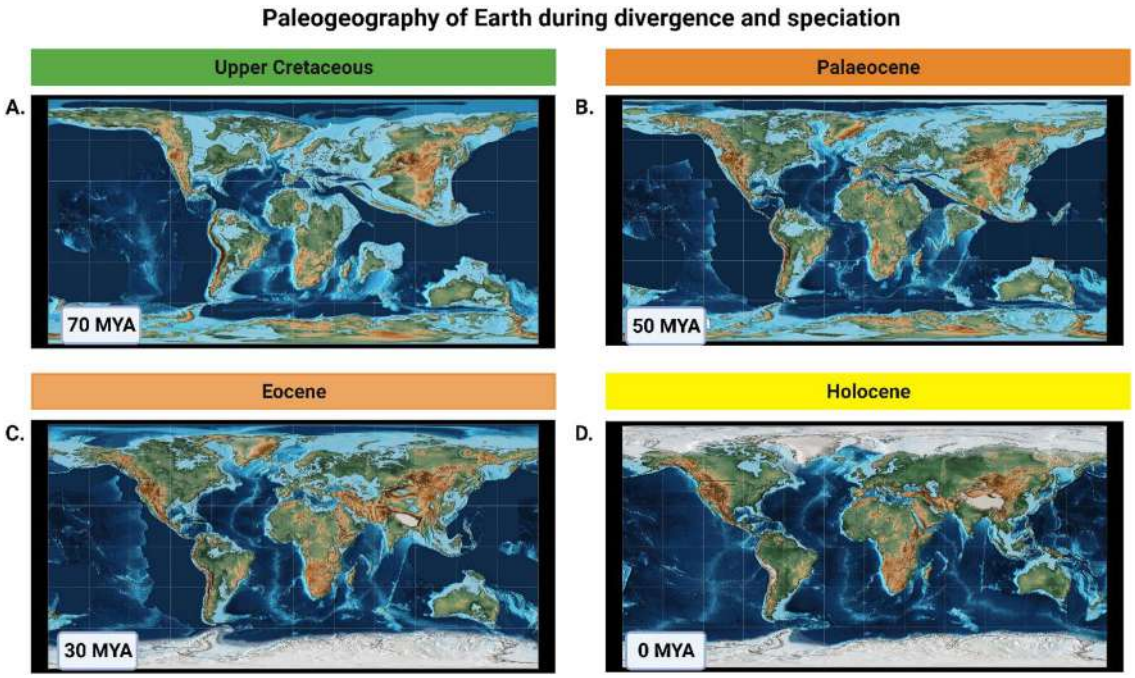


Fig. 1 Paleogeographic reconstructions of Earth for the past 70 million years to illustrate environmental differences during the period of divergence using PALEOMAP. **a** Positions of the continents approximately 70 million years ago during the Upper Cretaceous before West Africa had merged with the main continent and when India was still an island. **b** Continents approximately 50 million years ago during the Palaeocene after the African continent formed but before the Americas were connected and shortly before India merged with Asia or the polar caps formed. **c** Geography of Earth during the Eocene, 30 million years ago, by which time most continents had formed but central America did not connect the North and South yet and much of Europe was still under water. **d** Modern day geography of Earth in the current Holocene. (Image created in BioRender.com)

on continuous distribution maps may have experienced historical barriers to gene flow that drove speciation millions of years ago. Conversely, some habitats may have previously been unsuitable for a species and have only become suitable for recolonization more recently (Olson et al. 2008; Le et al. 2022).

In studying evolutionary rates, accurate estimates of diversification rates within lineages are also dependent on comparing the temporal range within which the species, and subspecies, of a particular lineage are formed (Jetz et al. 2012). This is pertinent when comparing evolutionary rates between lineages to determine diversification rates, e.g., passerine bird species and non-passerines (Jetz et al. 2012); identifying lineages of interest in the study of speciation due to higher diversification rates (O'Connell et al. 2019); or in studying complex historical speciation events (Lamichhaney et al. 2018). It is, therefore, crucial that studies on evolutionary processes are contextualized within relevant time frames.

Over the past few decades, a plethora of molecular studies have been published using variable methods from fossil-calibrated Bayesian inference (Rannala and Yang 2003; Kumar and Hedges 2016) to comparable relative time approaches (Yang and Yoder 2003; Tamura et al. 2012) to establish the timeline for the emergence and diversification. This includes both living and extinct species within several lineages of mammals (Nyakatura and Bininda-Emonds 2012; Springer et al. 2018), reptiles (Tucker et al. 2017), and birds (Barker et al. 2015; Prum et al. 2015). These studies have greatly advanced our understanding of evolutionary processes within the context of environmental changes and the time constraints that they occur in (Scholl and Wiens 2016). They have also helped clarify many of the questions we have with regard to the taxonomy and phylogeny of species, which have frequently been at odds with each other (Sangster 2014; Springer et al. 2018). More specifically, time trees and gene trees do not always have identical topology as gene trees represent the variable substitution and mutation rates in lineages (Lanfear et al. 2010) which can obscure true species divergence times if not calibrated with more accurate divergence times (Tiley et al. 2020).

The result is a compendium of thousands of studies that has culminated in a central TimeTree resource (Hedges et al. 2006) that collects and compiles divergence time estimates and time trees from published and peer-reviewed studies. From this resource, estimates of divergence times, related timelines, and time trees are available online (Kumar et al. 2017). This also includes the ability to apply the use of TimeTree in MEGA (Mello 2018) and a mobile phone app

(Kumar and Hedges 2011). Collectively this provides access to divergence time estimates for use in calibrating phylogenetic trees according to time, comparing clades in phylogenetic trees to know clades of shared common ancestry, as well as comparing genetic distance between species to their temporal or evolutionary distance. There is, however, still limited functionality in retrieving divergence times from the resource when dealing with species lists rather than individual pairs of species for downstream analyses or incorporation in molecular studies. As evolutionary studies frequently focus on multiple species, and even multiple lineages, at a time, this presents a significant barrier towards streamlining the integration of divergence time data into larger studies.

Previous attempts at automation (Krah 2015) to facilitate batch retrieval provided limited utilities and were poorly maintained, resulting in the removal of the package from the CRAN repository; although archived versions may still be available. The TimeTree resource has continued to develop and expand with the latest release of TimeTree 5, and the need for such capabilities is eminent in the ever-expanding field of evolutionary biology. While an application programming interface (API) has recently been created (Kumar et al. 2022), the use of APIs for the interaction with a database is not very biologist-friendly and requires additional computational skills. To bridge the current need, we have endeavored to create an easily accessible and freely available algorithm to retrieve relevant data on evolutionary histories from the TimeTree site for the seamless integration of divergence time data in molecular studies. In this paper we report on the PAReTT (Python-Automated Retrieval of TimeTree data), a menu-driven and user-friendly PYTHON package to automate interaction with the TimeTree resource for retrieving batch data with lists of species. We further provide several case examples to illustrate data retrieval and incorporation of data into scientific studies. The package is freely available on GitHub or as a stand-alone Windows executable.

Methods

Implementation

PAReTT (version 1.0.2) was scripted in the Spyder 5 IDE in pure PYTHON and is compatible with versions 3.6 and upward. A full list of dependencies is provided on the GitHub wiki, along with details for download and installation. This includes the use of several well-established PYTHON-based libraries such as NumPy (version 1.20.1) and pandas (version 1.2.4) for ease of input and high-order molding of data structures with relative ease (McKinney 2010; Harris et al. 2020), as well as Bio (version 1.3.9) for handling trees in the Newick format. PAReTT further uses the headless browser functionality implemented in Splinter (version 0.17.0) with the Selenium (version 4.1.5) extensions for the Firefox browser to submit user-specified web data to the TimeTree website (www.timetree.org) and retrieve the relevant results. Results are printed in real time to the shell, while list results are first stored to a 'dataframe' object which is written to user-specified output file. Some functionality had also been provided to validate data for any errors that may have occurred and preview basic tree files. Furthermore, the Magallon–Sanderson equation (Magallón and Sanderson 2001) was used to provide an option to calculate the diversification rate (r) of a lineage, genus, or species complex for which the number of extant species and divergence times are known. The *extinction ratio* (ε) is the fraction of the *extinction rate* (μ) divided by *speciation rate* (λ) according to the following equation:

$$\text{Extinction ratio } (\varepsilon) = {}^{\mu}/_{\lambda}. \tag{1}$$

The extinction rate represents the number of species formed that goes extinct in a given time period while the speciation rate represents the number of species that form within a given time period independent of survival outcomes.

When the *extinction ratio* is not known, the equation, with crown age, is given by (n is *total known species* and Δt is *divergence time*):

$$\text{Diversification rate}(r) = {}^{\log n - \log 2}/_{\Delta t}. \tag{2}$$

When the extinction rate is known, the equation, with crown age, is given by

$$r = \frac{1}{\Delta t} \times \left(\log \left(\frac{n(1-\varepsilon^2)}{2} + 2\varepsilon + \frac{(1-\varepsilon)}{2} \times \sqrt{n(n\varepsilon^2 - 8\varepsilon + 2n\varepsilon + n)} \right) - \log 2 \right). \tag{3}$$

The diversification rate calculations make use of the math modules that form part of the standard PYTHON releases. The PAReTT script was benchmarked to test for time and memory consumption of individual functions using memory-profiler version 0.16.0.

Input and Output File Formats

PAReTT uses two primary forms of input that users specify when the user is prompted to provide the name of the input file by the interactive menus. The first is a basic text file, indicated as '.txt,' while the second is a standard comma-separated value (CSV) file, indicated as '.csv.' While it is preferred to specify the full name of the file (e.g., 'Species.txt'), PAReTT was scripted with checkpoints to ensure the proper format of input files, even when only a name is given, and should read the input files as long as they are in the current working directory. If the files are not stored in the current working directory, then the full file path name is

preferred. The specific requirements for the content of each input file are detailed under usage in Section 'Usage.'

Output files differ based on the specific data being retrieved. For example, timelines are retrieved as figures in the JPEG format while time trees, for both complete taxa as well as lists of species, are retrieved in the Newick format. Any species in the provided list for which the divergence time could not be resolved in the tree or were substituted by a similar species are stored as a table for review. Trees can be visualized and edited by most phylogenetic software programs including FigTree (Rambaut 2017) and MEGA (Kumar et al. 2018); however, the basic topology of these trees can also be viewed directly in PAReTT. Divergence time data, based on the median time from all studies from which divergence times estimates are derived, are printed on screen for pairs (only two species). Divergence times for lists of species will iterate through every possible combination of pairs in the lists and store data in 'dataframe' objects. These objects can then be stored in the output file as a vectorized three column matrix in a comma-separated value file. Output files are stored in the active working directory.

Usage

The main menu presents the initial options to verify the availability of data for a species, determine divergence times, a timeline, a time tree, validate the data, print the citation for TimeTree, or calculate divergence times as summarized in Fig. 2. The data availability option brings up a submenu to specify if availability should be verified for a single species or species list. When checking a single species, the availability is printed to the screen as the species name followed by either 'Available' or 'Not Available.' This function was designed to ensure data availability prior to further data retrieval steps. For a list of species, the species

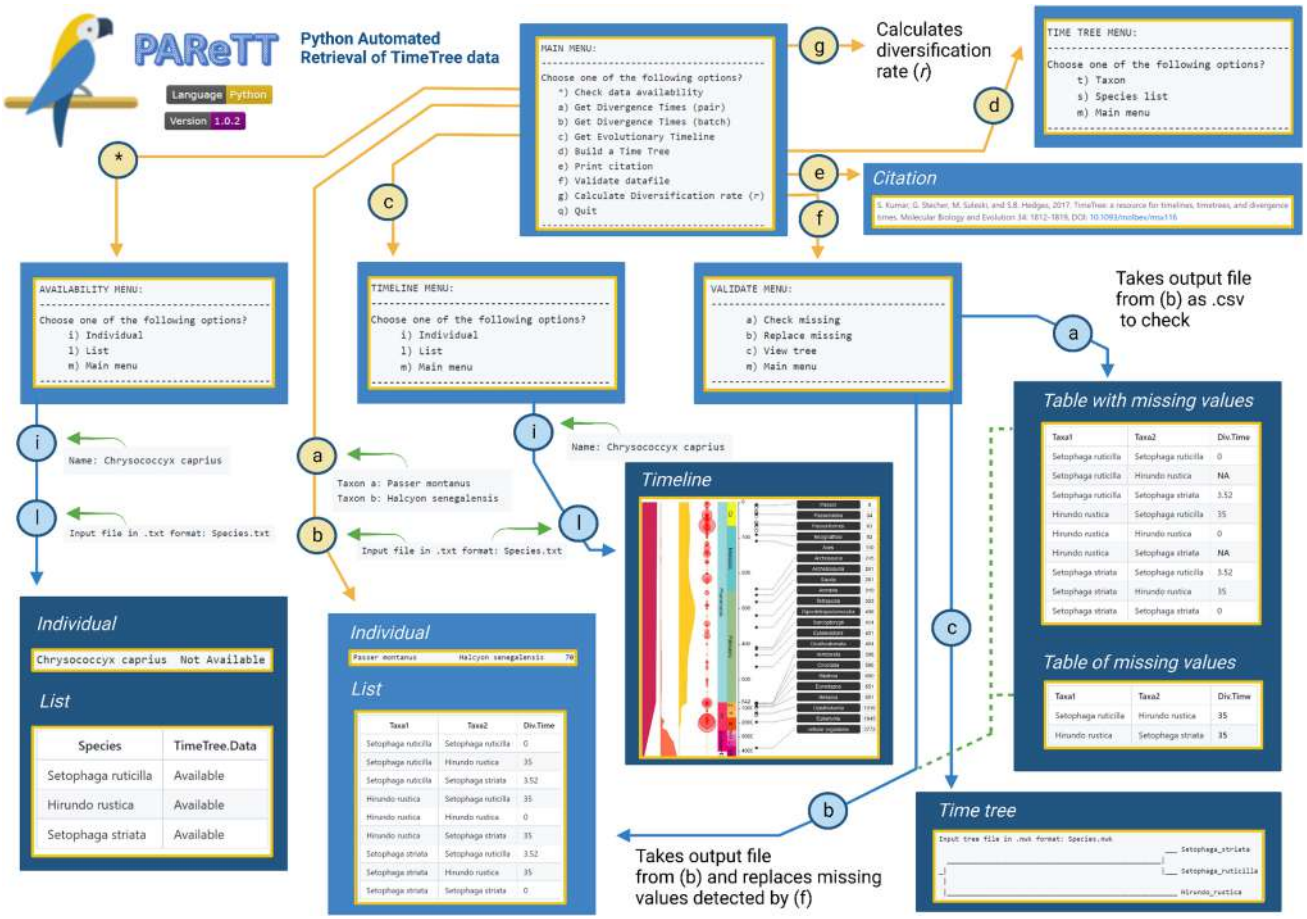


Fig. 2 Graphical summary of menu options and submenus for PAReTT. The main menu options are: *, to verify the availability of data; **a** or **b** to resolve divergence times (pair or batch); **c** retrieve a timeline; **d** retrieve a time tree; **e**, print the TimeTree citation; **f** validate data; or **q** to exit. The data availability, timeline, and time tree options bring up a submenu, indicated by yellow arrows, to retrieve information for an individual species/taxon or a list. Output generated for lists are exported as a table (CSV), images (JPEG), or trees (Newick). The validation option brings up the choice of finding or replacing missing values for divergence times data (options a or b) as well as to view the tree topology (option c), for output files. (Image created in BioRender.com)

name and availability are printed on screen and stored to export results in a CSV file.

The timeline and time tree options will bring up a similar submenu that provides the option to retrieve information for an individual species, individual taxon, or a list of species. As before, individual species will result in a single output. The timeline retrieval function will retrieve an image in the JPEG format that illustrates the evolutionary history for the species highlighting the major time points where a kingdom, order, class, genus, and species first emerged. As an example, a timeline was retrieved for the Lazuli bunting, *Passerina amoena* [Say, 1823] in Example A. Input given as a list will retrieve individual images for each species in the list. For this example, we used the knowledge of divergence times from the timeline to compliment the timeline data with other relevant information such as the paleogeography (Sampson et al. 2010) and range data from *Birds of the World* (Greene et al. 2020).

The time tree option has similar functionality except that input can either be a larger taxonomic group, such as family or genus containing several congenic species, or a list of species as this output provides a time-calibrated tree displaying both the interrelatedness of species as well as the timescale along which the branches diverged. In Example B, two individual time trees were generated for the Nearctic thrushes (genus: *Catharus*) and Australasian fairywrens (genus: *Malurus*), respectively. The individual diversification rate (r) was calculated for each genus in PAReTT with the Magallon–Sanderson equation (Magallón and Sanderson 2001). More detailed speciation, extinction, and diversification rates (Etienne and Apol 2009) were computed with PyRate version 3.1.1 to further illustrate the use of diversification rates in evolutionary studies (Silvestro et al. 2014).

The Divergence time submenu (Example C) provides the option to check for the divergence time between two species e.g., the Neotropical Swainson's thrush, *Catharus ustulatus* [Nuttall, 1840], and the Australasian Superb fairywren, *Malurus cyaneus* [Ellis, 1782]. The result will be printed to the screen as the name of 'Taxon a' and 'Taxon b' followed by the mean divergence time in millions of years ago (MYA) e.g., 35 MYA. The list option takes a text file with a list of species names and iterates through the list to determine the divergence time between each species in the list (Supplementary Fig. 1). The results are stored in a 'dataframe' object which can be exported as a CSV file in the form of a three column vectorized matrix with the names for 'Taxa 1' in the first column, the names of 'Taxa 2' in the second column, and the divergence times between them in the third column. Additional functionality is provided in the main menu to validate data. This option can be used to check output files for missing values, as well as retrieve and replace such values, in case a server error occurred. The table generated for species lists, in the form of a three column vectorized matrix, can then be converted to a full matrix to use divergence times as a measure of separation or differentness in Mantel tests (Mantel 1967; Carr 2021) as applied to landscape or spatial genetics. This was done to compare the distribution of genetic distance, as measured by fixation index (F_{ST}) calculated with POPGENE version 1.32 (Yeh et al. 1997), for two circadian clock gene polymorphisms as they relate to evolutionary histories among forty bird species (Le Clercq et al. 2023).

Results

Example A—Timelines and Paleogeography

The timeline (Fig. 3) was retrieved for the Neotropical bird, the Lazuli bunting. The left panel indicates the major geologic timescales in Eons and Eras while the right indicates the main divergence times when specific taxa emerged. This includes the current Phanerozoic Eon as well as the subdivisions from the Paleozoic Era, which lasted until approximately 250 MYA, and the Mesozoic which lasted until 65 MYA illustrated in the first panel of Fig. 1. This period was marked by the emergence of several major lineages including the first birds (class: Aves) approximately 110 MYA as well as the taxonomic order *Passeriformes,* which includes most of the songbirds, approximately 63 MYA. The specific species, a bunting in the family Cardinalidae, emerged from a common ancestor shared with other cardinals in the Cenozoic approximately 11 MYA, while species of the genus *Passerina* first emerged approximately 4 MYA and subsequently diverged to form species. The range of the Lazuli bunting is located on the Western half of North America (Greene et al. 2020) which was physically separated from the Eastern half by the North American inland sea; the modern continent only formed after the Upper Cretaceous period around the Paleocene (Fig. 1).

Example B—Time Trees and Diversification Rates

Time trees (Fig. 4) were generated for two genera of birds that each contain several congenic species (11–12 species) but have different evolutionary histories in terms of the divergence times within their genus. The first represents the genus *Catharus* for which 12 species are included in the time tree. These species diverged over a period spanning approximately 4.73 million years and have an absolute diversification rate of 0.38 ($N = 12$). By comparison, the second tree for the genus *Malurus* includes a similar number of species; however, these species diverged over a period of approximately 9 million years and have a diversification rate ($r = 0.19$, $N = 11$) of nearly half the rate observed in *Catharus*. More detailed speciation and

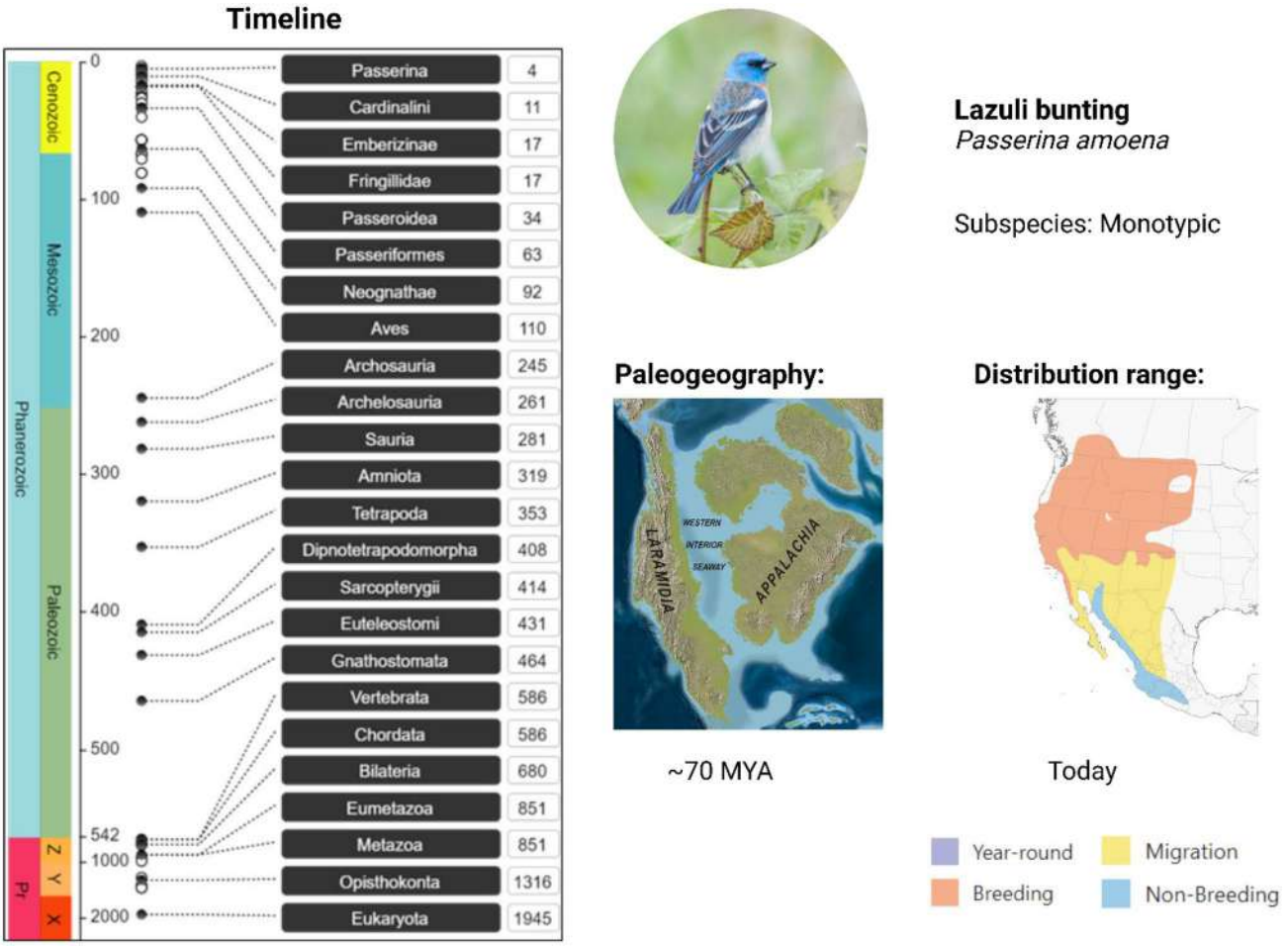


Fig. 3 Example of an evolutionary timeline (left) retrieved using PAReTT for the Lazuli bunting. The left panel of the timeline indicates the major geologic timescale of the past 2000 MYA including the prevailing Phanerozoic Eon as well as the subdivisions by era Paleozoic which lasted until approximately 250 MYA when the Mesozoic started which lasted until 65 MYA. As is illustrated, this period was marked by the emergence of the first birds in the class Aves approximately 110 MYA. The species currently recognized in the family Cardinalidae emerged in the Cenozoic era approximately 11 MYA, with most species of the genus *Passerina* first emerging 4 MYA. On the right and in the middle, the paleogeography of the North American continent during this period is illustrated, clearly showing the separation of the Western and Eastern parts by the Western interior seaway (Sampson et al. 2010). Next to this is the contemporary range map for the species showing their breeding and wintering grounds that still encompass most of the Western half of the continent (Greene et al. 2020). (Image created in BioRender.com)

diversification rate estimates, for each species evolutionary timeline, are indicated in Fig. 5. Both genera displayed comparable trends in both speciation rate (λ) and diversification rate (*r*); however, these processes happened on different timescales, since approximately 10 MYA for *Malurus* species and 6 MYA for *Catharus* species. Overall, the rates for *Catharus* species were slightly higher. As extinct species for these lineages are unknown, the estimates for extinction rates (μ) were close to zero, resulting in highly similar speciation and diversification plots.

Example C—Divergence Times and Genetic Distance

As an example, the divergence times retrieved for four species are given in Table 1: Lazuli bunting, Malachite kingfisher (*Corythornis cristatus* [Pallas, 1764]), Amethyst sunbird (*Chalcomitra amethystine* [Shaw, 1812]), and Emerald cuckoo (*Chrysococcyx cupreus* [Shaw, 1792]). The results clearly illustrate the more recent divergence time (25.4–38.1 MYA) between Lazuli bunting and Amethyst sunbird, both members of the same order (order: Passeriformes), as well

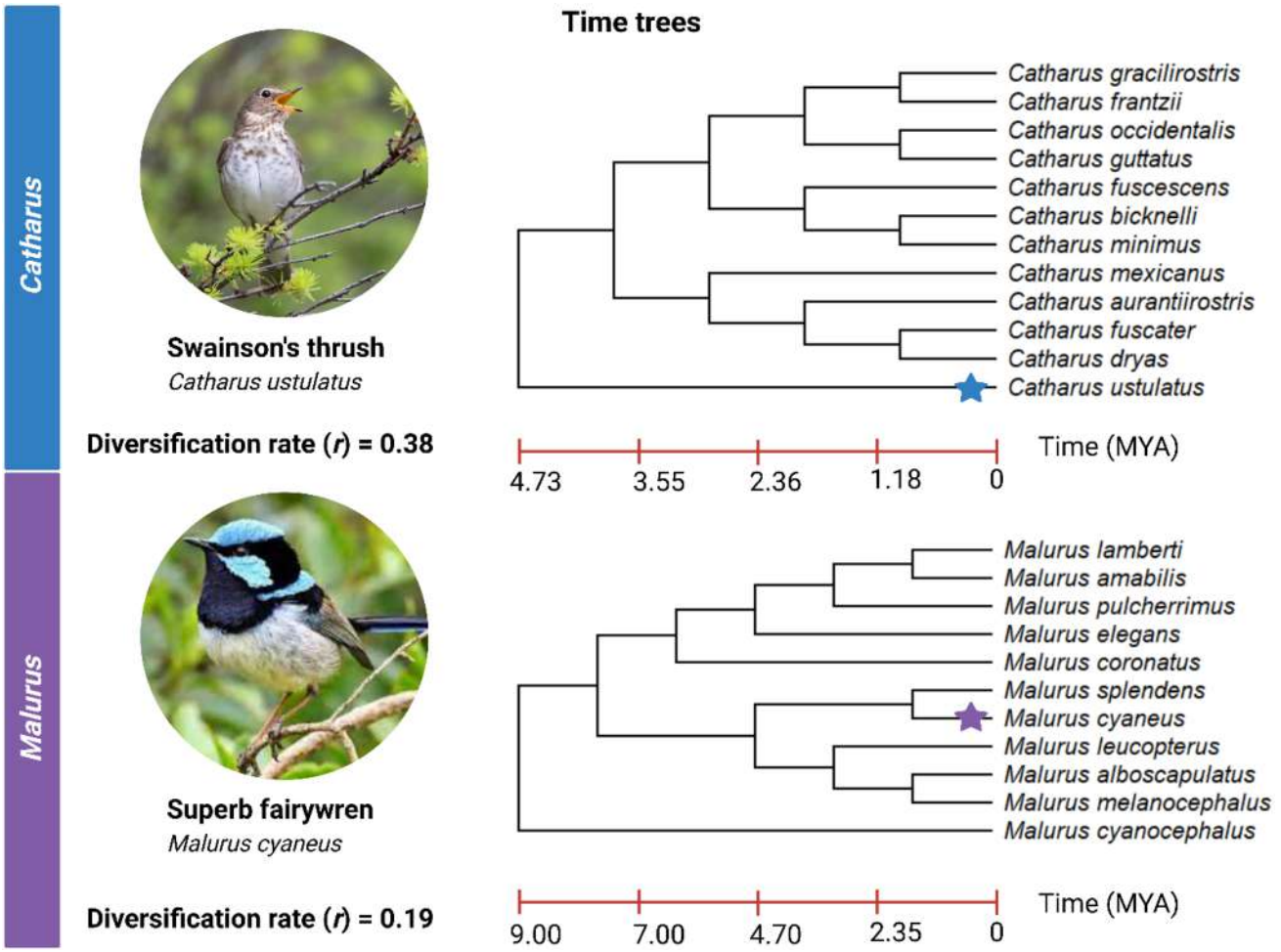
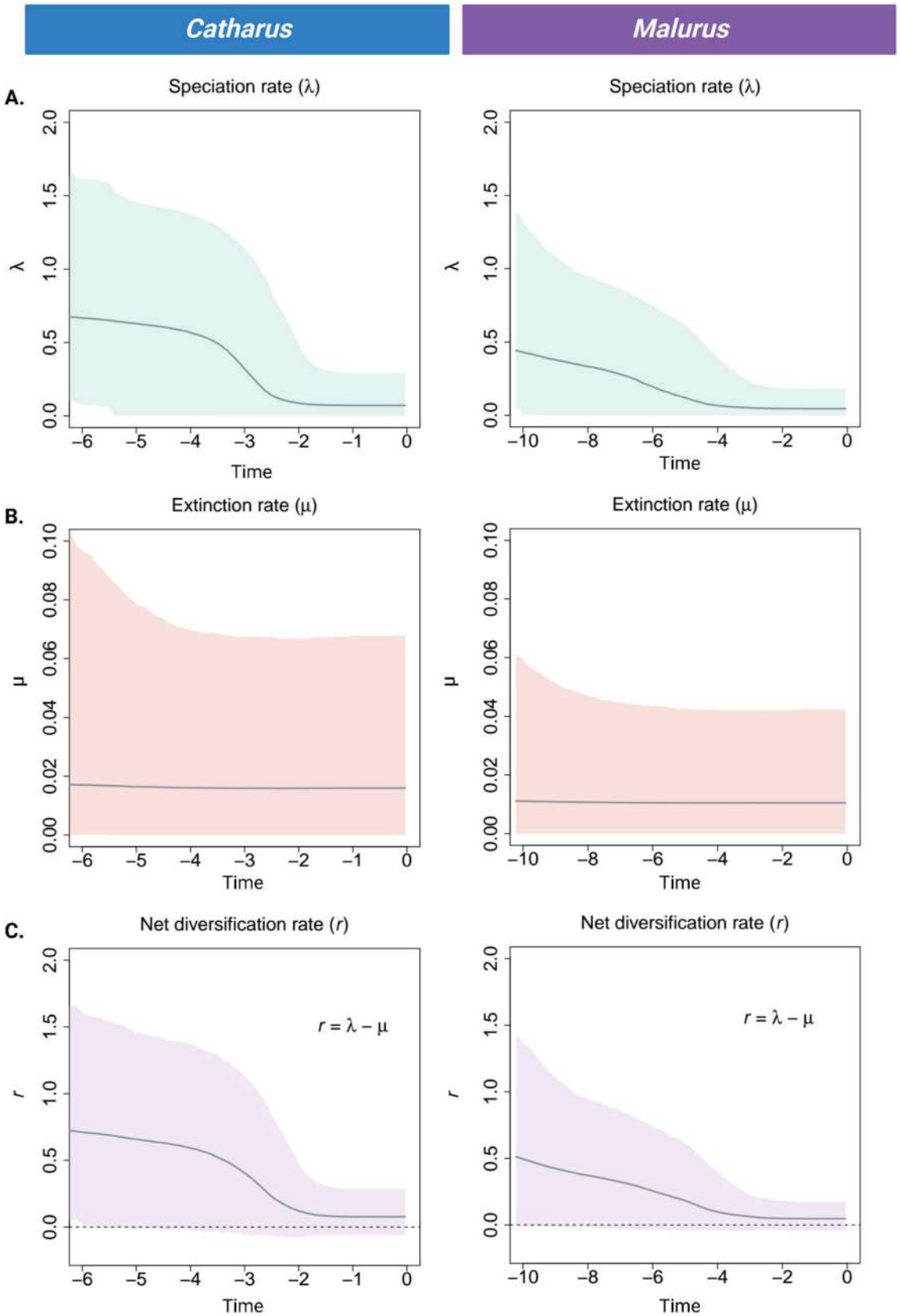


Fig. 4 Comparison of time trees retrieved for two genera using PAReTT, one for several species in the *Catharus* genus of Neotropical thrushes and another of species in the *Malurus* genus of Australasian fairywrens. The *Catharus* genus has 12 species, including Swainson's thrush, which diverged over a period of 4.74 MYA with a diversification rate of 0.38. The *Malurus* genus has 11 species, including the Superb fairywren, which diverged over a period of 9 MYA with a diversification rate of 0.19. This illustrates how two lineages may have similar evolutionary processes happening but have different diversification rates due to the difference in time scales. (Image created in BioRender.com)

as the far-older divergence times (73.7–97.3 MYA) between birds from different orders such as the Malachite kingfisher (order: Coraciiformes) and Emerald cuckoo (order: Cuculiformes). Furthermore, PAReTT was used to assess data availability and retrieve divergence times between species for a list of forty bird species in a recent review and meta-analysis on clock genes as candidate genes in migration studies, published in *Biological Reviews* (Le Clercq et al. 2023). This study found a significant relationship between divergence times and the observed genetic distance as measured by fixation index (F_{ST}), for allele data of two candidate genes, using Mantel tests for matrix comparison (Fig. 6). Thus, the alleles and their distributions potentially still reflect the ancestral state rather than contemporary changes due to selection. The results from the benchmarking tests, performed in tandem, showed an average time for submitting data and retrieving the result ranged from fifteen to thirty seconds and memory consumption remained low, ranging from < 100 to 400 Megabytes (Supplementary Table 1).

Discussion

In this paper, we used three examples to illustrate the types of data that can be retrieved using PAReTT as well as how these data can be incorporated into evolutionary studies. The first involved a detailed timeline of the evolutionary history for Lazuli buntings which enabled a direct comparison of divergence times to the relative paleogeography for their range. The observation that their modern range corresponds to geographic patterns during their divergence time is indicative of speciation happening during former geographic


Catharus
Malurus
A.
Speciation rate (λ)
λ
Time
B.
Extinction rate (μ)
μ
Time
C.
Net diversification rate (r)
r = λ − μ
r
Time

◂**Fig. 5** Plots from the PyRate analysis indicating the **A** speciation, **B** extinction, and **C** diversification rates over time for two genera of birds based on node age data from TimeTree. The left panel is the results for the *Catharus* thrush species complex while the right indicated the results for the *Malurus* fairywren species complex. Both genera followed similar patterns of speciation and diversification; however, the timescales over which these processes occurred are different and the *Catharus* species has slightly higher rates. As no extinct species could be included, the extinction rates (μ) are zero, resulting in highly similar speciation (λ) and diversification (*r*) plots. (Image created in BioRender.com)

isolation due to the North American inland sea (Western interior seaway), without substantial range expansion after the continent formed. This may have contributed to the division between the Lazuli bunting and other Neotropical buntings such as the Painted bunting, *Passerina ciris* [Linnaeus, 1758], which has a range that is restricted to the South-East of North America (Lowther et al. 2020) and Indigo bunting, *Passerina cyanea* [Linnaeus, 1766], which has a North-Eastern range (Payne 2020), although the primary division between these lineages is only dated to approximately 4.48 MYA. Similar observations have been made in North American *Catharus* species of thrushes (Voelker et al. 2013).

This example clearly illustrates the added value of divergence data and accurate timelines in reconstructing and understanding evolutionary processes. While we did use paleogeography (Sampson et al. 2010) from the literature and range data from the *Birds of the World* species records (Greene et al. 2020), this should be achievable in most species. Paleogeography can be reconstructed for nearly any period using PALEOMAP (Scotese 2016), as illustrated in Fig. 1, while range data can be retrieved for most birds (BirdLife International and Handbook of the Birds of the World 2021) as well as many mammals, reptiles, amphibians, and fishes (International Union for Conservation of Nature (IUCN) 2022) for use in range mapping analyses.

The second example illustrated the retrieval of time trees at the species level for two specific genera for the purpose of visually representing the timescale during which speciation and diversification happen in different lineages. Diversification rate calculations within PAReTT are based on user input rather than tree input as accurate measures would require a completely sampled phylogeny which is often unavailable. The diversification rate, based solely on the species for which data were available on the TimeTree resource, differed substantially between the two genera: having both different node ages for each genus as well as differing in the number of species, while having similar phylogenetic tree topology. More detailed analyses of speciation and diversification over time showed higher initial speciation with an eventual decline, possibly indicating a rate shift, followed by a plateau of lower rates. This could be due to the analyses only evaluating speciation at the species level and not the subspecies level, as several species within both lineages are partitioned into subspecies. For both genera, the TimeTree data were fairly complete as only two additional *Catharus* species, and one additional *Malurus* species, are currently recognized. In this example speciation and diversification

Table 1 Examples of divergence times retrieved using the batch retrieval option in PAReTT

Taxa 1	Taxa 2	Divergence time	CI*	Studies*
Lazuli bunting	Lazuli bunting	0.0 MYA	n. a	n. a
Lazuli bunting	Malachite kingfisher	70.0 MYA	64.5–80.0 MYA	16
Lazuli bunting	Amethyst sunbird	27.6 MYA	25.4–38.1 MYA	5
Lazuli bunting	Emerald cuckoo	80.0 MYA	73.7–97.3 MYA	3
Malachite kingfisher	Lazuli bunting	70.0 MYA	64.5–80.0 MYA	16
Malachite kingfisher	Malachite kingfisher	0.0 MYA	n. a	n. a
Malachite kingfisher	Amethyst sunbird	70.0 MYA	64.5–80.0 MYA	16
Malachite kingfisher	Emerald cuckoo	80.0 MYA	73.7–97.3 MYA	3
Amethyst sunbird	Lazuli bunting	27.6 MYA	25.4–38.1 MYA	5
Amethyst sunbird	Malachite kingfisher	70.0 MYA	64.5–80.0 MYA	16
Amethyst sunbird	Amethyst sunbird	0.0 MYA	n. a	n. a
Amethyst sunbird	Emerald cuckoo	80.0 MYA	73.7–97.3 MYA	3
Emerald cuckoo	Lazuli bunting	80.0 MYA	73.7–97.3 MYA	3
Emerald cuckoo	Malachite kingfisher	80.0 MYA	73.7–97.3 MYA	3
Emerald cuckoo	Amethyst sunbird	80.0 MYA	73.7–97.3 MYA	3
Emerald cuckoo	Emerald cuckoo	0.0 MYA	n. a	n. a

Data were retrieved for a list of four species to illustrinput file are detailed under usage in Sectate the vectorized matrix format from PAReTT indicating the two taxa that were compared as well as the divergence time as either the median or adjusted median value. The list included the Lazuli bunting, Malachite kingfisher, Amethyst sunbird, and Emerald cuckoo. Additional data, indicated with an asterisk (*), such as the confidence interval (CI) of the estimate as well as the number of studies used to derive the values, were retrieved from the TimeTree website (www.timetree.org)

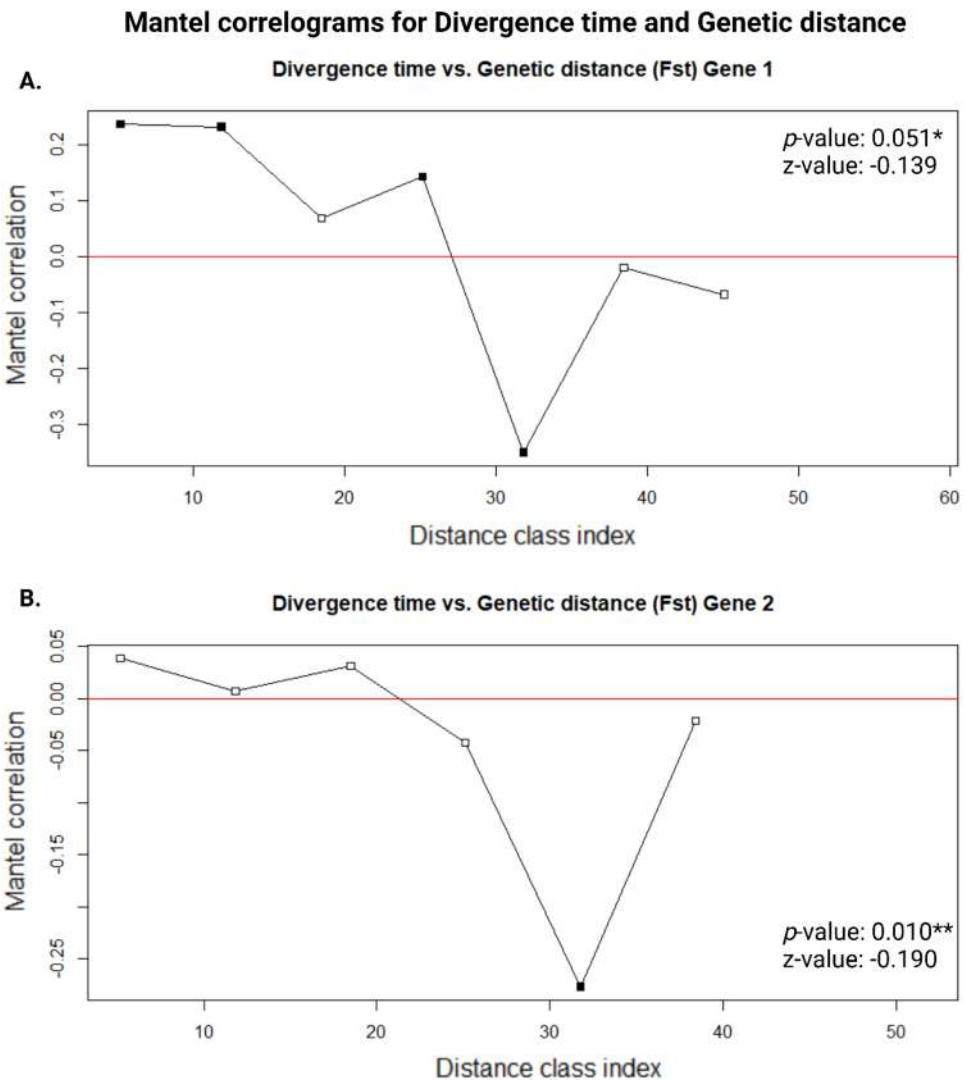


Fig. 6 Mantel correlograms from the comparison of distance as measured by divergence times to genetic distance for gene 1 (**A**) and gene 2 (**B**). Divergence times were retrieved for forty bird species using PAReTT while genetic distance was measured by computing the fixation index (F_{ST}) for two individual genes. For both genes, a significant correlation was found between the genetic distance and divergence times where increases in genetic distance corresponded to older divergence times. (*p-value < 0.1, **p-value < 0.05)

rate plots were nearly identical due to the lack of known extinct species. This is contrary to the results that could be obtained for genera such as *Acrocephalus* warblers that includes 42 species of which 6 insular forms, including the Pagan reed warbler (*Acrocephalus yamashinae* [Taka-Tsukasa, 1931]), are known to be extinct.

The study of speciation, extinction, and diversification rates are critical in identifying lineages that rapidly for evolutionary studies to elucidate key factors that confer evolutionary advantages or disadvantages (Jetz et al. 2012; O'Connell et al. 2019). In addition, identifying lineages with slower speciation are informative for conservation practices, considering such species may experience higher difficulty to adapt to changes in their environment and both speciation and diversification rates have been tied to range changes, a critical attribute used in assessing species viability (Castiglione et al. 2017).

The third example focused on the retrieval of divergence times between lists of species to enable the construction of a matrix where divergence time serves as a proxy for temporal evolutionary distance. This enabled the comparison of genetic distance to evolutionary distance using Mantel tests (Mantel 1967; Carr 2021). The significant correlation between the measures illustrated the high heritability of genotypes within lineages which, combined with and a lack of selection, portends the possibility of these genes reflecting ancestral states (Le Clercq et al. 2023). This illustrates the significance in studying divergence time data in relation to genetic data in molecular studies. The TimeTree resource does, however, still lack options to facilitate batch retrieval of divergence times for lists of species.

Several key differences still exist in the types and formats of data that can be retrieved using PAReTT versus the main website. When retrieving timelines, PAReTT will retrieve the full timeline including other relevant ecological

attributes; if only the major epochs and period are required for the image, then this can be specified on the online version. When retrieving time trees, PAReTT will automatically retrieve trees at the species level while trees above the genus level can be retrieved for a family or order using the website. Additionally, as PAReTT is designed to retrieve data for downstream analyses, the time trees are exported as editable Newick format trees rather than images intended to be used as is. Special features of PAReTT include the ability to retrieve batch divergence times for a list of species as a vectorized matrix that is well suited to downstream data analyses of divergence times as matrix data, although further details such as confidence intervals and the number of included studies is best retrieved from the website as needed. PAReTT can also calculate diversification rates based on retrieved data to ascertain if different rates exist between lineages.

Here, we illustrated the use of a newly scripted PYTHON package called PAReTT that provides an added layer of functionality to the TimeTree resource that can enhance molecular and evolutionary studies. Future updates will include the ability to switch between scientific names and common names for species as well as the ability to calculate diversification rates for a table of multiple lineages.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00239-023-10106-3.

Acknowledgements All images were created or edited in BioRender.com. This work is based on the research supported wholly/in part by the National Research Foundation of South Africa (Grant Number: 112062).

Author Contributions LSLC scripted the PAReTT program, created the images, and wrote the draft manuscripts. DLD, AK, and JPG provided research support and edited the final draft.

Funding Open access funding provided by University of the Free State. This study was supported by the National Research Foundation (Grant No. 112062).

Data Availability The custom Python script for PAReTT version 1.0.2 is available for download for installation from source code on GitHub (https://github.com/LSLeClercq/PAReTT), or as a Windows executable of version 1.0.1 from the Zenodo online depository (https://zenodo.org/record/6653321#.Y6MnNHZBw2w), and includes example files used for testing. Data used in this paper were deposited online (https://zenodo.org/record/7496660#.Y7A-aHZBw2x).

Declarations

Conlict of interest The authors have no conflicts of interest to declare.

References

Barker K, Burns KJ, Klicka J et al (2015) New insights into new world biogeography: an integrated view from the phylogeny of blackbirds, cardinals, sparrows, tanagers, warblers, and allies. Auk 132:333–348. https://doi.org/10.1642/AUK-14-110.1

BirdLife International, Handbook of the Birds of the World (2021) Bird species distribution maps of the world. Version 2021.1. http://datazone.birdlife.org/species/requestdis. Accessed 9 Dec 2021

Carr J (2021) Mantel. PYTHON package version 2.1.0. https://github.com/jwcarr/mantel. Accessed 21 Apr 2022

Castiglione S, Mondanaro A, Melchionna M et al (2017) Diversification rates and the evolution of species range size frequency distribution. Front Ecol Evol 5:147. https://doi.org/10.3389/FEVO.2017.00147/BIBTEX

Etienne RS, Apol MEF (2009) Estimating speciation and extinction rates from diversity data and the fossil record. Evolution 63:244–255. https://doi.org/10.1111/j.1558-5646.2008.00537.x

Francisco Henao Diaz L, Harmon LJ, Sugawara MTC et al (2019) Macroevolutionary diversification rates show time dependency. Proc Natl Acad Sci U S A 116:7403–7408. https://doi.org/10.1073/pnas.1818058116

Greene E, Muehter VR, Davison W (2020) Lazuli Bunting (*Passerina amoena*). In: Poole AF (ed) Birds of the World. Cornell Lab of Ornithology, Ithaca, NY, USA. https://doi.org/10.2173/bow.lazbun.01

Harris CR, Millman KJ, van der Walt SJ et al (2020) Array programming with NumPy. Nature 585:357–362. https://doi.org/10.1038/s41586-020-2649-2

Hedges SB, Dudley J, Kumar S (2006) TimeTree: a public knowledge-base of divergence times among organisms. Bioinformatics 22:2971–2972. https://doi.org/10.1093/bioinformatics/btl505

International Union for Conservation of Nature (IUCN) (2022) Digital distribution maps on *The IUCN Red List of Threatened Species*. Version 6.2. https://www.iucnredlist.org/resources/spatial-data-download. Accessed 23 Nov 2022

Jetz W, Thomas GH, Joy JB et al (2012) The global diversity of birds in space and time. Nature 491:444–448. https://doi.org/10.1038/nature11631

Krah F-S (2015) Timetree: interface to the timetree of life webpage. R package version 1.0. https://cran.r-project.org/web/packages/timetree/index.html

Kumar S, Hedges SB (2011) Timetree2: species divergence times on the iPhone. Bioinformatics 27:2023–2024. https://doi.org/10.1093/bioinformatics/btr315

Kumar S, Hedges SB (2016) Advances in time estimation methods for molecular data. Mol Biol Evol 33:863–869. https://doi.org/10.1093/molbev/msw026

Kumar S, Stecher G, Suleski M, Hedges SB (2017) TimeTree: a resource for timelines, timetrees, and divergence times. Mol Biol Evol 34:1812–1819. https://doi.org/10.1093/MOLBEV/MSX116

Kumar S, Stecher G, Li M et al (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547–1549. https://doi.org/10.1093/molbev/msy096

Kumar S, Suleski M, Craig JM et al (2022) TimeTree 5: an expanded resource for species divergence times. Mol Biol Evol 39:msac174. https://doi.org/10.1093/molbev/msac174

Lamichhaney S, Han F, Webster MT et al (2018) Rapid hybrid speciation in Darwin's finches. Science 359:224–228. https://doi.org/10.1126/science.aao4593

Lanfear R, Ho SYW, Love D, Bromham L (2010) Mutation rate is linked to diversification in birds. Proc Natl Acad Sci U S A 107:20423–20428. https://doi.org/10.1073/pnas.1007888107

Le CT, Do TB, Pham TMA et al (2022) Reconstruction of the evolutionary biogeography reveals the origins of *Alpinia* Roxb. (Zingiberaceae): a case of out-of-Asia migration to the Southern Hemisphere. Acta Bot Brasilica 36:e2021abb0255. https://doi.org/10.1590/0102-33062021abb0255

Le Clercq LS, Bazzi G, Cecere JG et al (2023) Time trees and clock genes: a systematic review and comparative analysis of contemporary avian migration genetics. Biol Rev. https://doi.org/10.1111/brv.12943

Linck E, Freeman BG, Dumbacher JP (2020) Speciation and gene flow across an elevational gradient in New Guinea kingfishers. J Evol Biol 33:1643–1652. https://doi.org/10.1111/jeb.13698

Lowther PE, Lanyon SM, Thompson CW, Schulenberg TS (2020) Painted bunting (Passerina ciris), version 1.0. Birds World. https://doi.org/10.2173/BOW.PAIBUN.01

Magallón S, Sanderson MJ (2001) Absolute diversification rates in angiosperm clades. Evolution (N Y) 55:1762–1780. https://doi.org/10.1111/j.0014-3820.2001.tb00826.x

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220

McKinney W (2010) Data structures for statistical computing in python. In: Proc. of the 9th Python in Science Conf. (SciPy 2010) pp 56–61

Mello B (2018) Estimating timetrees with MEGA and the timetree resource. Mol Biol Evol 35:2334–2342. https://doi.org/10.1093/molbev/msy133

Müller RD, Cannon J, Qin X et al (2018) GPlates: building a virtual earth through deep time. Geochem, Geophys Geosyst 19:2243–2261. https://doi.org/10.1029/2018GC007584

Mwale M, Dalton DL, Kropff AS et al (2022) Genetic and morphological variation of Woodland Kingfisher Halcyon senegalensis reveals cryptic mitochondrial lineages and patterns of mitochondrial nuclear discordance. Ostrich 93:192–207. https://doi.org/10.2989/00306525.2022.2066215

Nyakatura K, Bininda-Emonds ORP (2012) Updating the evolutionary history of Carnivora (Mammalia): a new species-level supertree complete with divergence time estimates. BMC Biol 10:12. https://doi.org/10.1186/1741-7007-10-12

O'Connell DP, Kelly DJ, Lawless N et al (2019) Diversification of a great speciator in the Wallacea region: differing responses of closely related resident and migratory kingfisher species (Aves: Alcedinidae: Todiramphus). Ibis 161:806–823. https://doi.org/10.1111/ibi.12688

Olson DM, Dinerstein E, Wikramanayake ED et al (2001) Terrestrial ecoregions of the world: a new map of life on earth: a new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. Bioscience 51:933–938. https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2

Olson DE, FitzGerald G, Hand DM et al (2008) Use of parentage analysis to determine reproductive success of hatchery-origin spring chinook salmon outplanted into Shitike Creek, Oregon. North Am J Fish Manag 28:1472–1485. https://doi.org/10.1577/m07-195.1

Payne RB (2020) Indigo bunting (*Passerina cyanea*). In: Poole AF (ed) Birds of the World. Cornell Lab of Ornithology, Ithaca, NY USA. https://doi.org/10.2173/bow.indbun.01

Prum RO, Berv JS, Dornburg A et al (2015) A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. Nature 526:569–573. https://doi.org/10.1038/nature15697

Rambaut A (2017) FigTree-version 1.4.3, a graphical viewer of phylogenetic trees. https://github.com/rambaut/figtree/releases. Accessed 8 Jul 2022

Rannala B, Yang Z (2003) Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. Genetics 164:1645–1656. https://doi.org/10.1093/genetics/164.4.1645

Sampson SD, Loewen MA, Farke AA et al (2010) New horned dinosaurs from Utah provide evidence for intracontinental dinosaur endemism. PLoS ONE 5:1–12. https://doi.org/10.1371/journal.pone.0012292

Sangster G (2014) The application of species criteria in avian taxonomy and its implications for the debate over species concepts. Biol Rev 89:199–214. https://doi.org/10.1111/BRV.12051

Scholl JP, Wiens JJ (2016) Diversification rates and species richness across the Tree of Life. Proc R Soc B Biol Sci. https://doi.org/10.1098/RSPB.2016.1334

Scotese CR (2016) PALEOMAP PaleoAtlas for GPlates and the PaleoData Plotter Program. http://www.earthbyte.org/paleomap-%0Apaleoatlas-for-gplates/. Accessed 21 Apr 2022

Silvestro D, Salamin N, Schnitzler J (2014) PyRate: a new program to estimate speciation and extinction rates from incomplete fossil data. Methods Ecol Evol 5:1126–1131. https://doi.org/10.1111/2041-210X.12263

Springer MS, Murphy WJ, Roca AL (2018) Appropriate fossil calibrations and tree constraints uphold the mesozoic divergence of solenodons from other extant mammals. Mol Phylogenet Evol 121:158–165. https://doi.org/10.1016/j.ympev.2018.01.007

Tamura K, Battistuzzi FU, Billing-Ross P et al (2012) Estimating divergence times in large molecular phylogenies. Proc Natl Acad Sci U S A 109:19333–19338. https://doi.org/10.1073/pnas.1213199109

Tiley GP, Poelstra JW, dos Reis M et al (2020) Molecular clocks without rocks: new solutions for old problems. Trends Genet 36:845–856. https://doi.org/10.1016/j.tig.2020.06.002

Tucker DB, Hedges SB, Colli GR et al (2017) Genomic timetree and historical biogeography of Caribbean island ameiva lizards (Pholidoscelis: Teiidae). Ecol Evol 7:7080–7090. https://doi.org/10.1002/ece3.3157

Voelker G, Bowie RCK, Klicka J (2013) Gene trees, species trees and earth history combine to shed light on the evolution of migration in a model avian system. Mol Ecol 22:3333–3344. https://doi.org/10.1111/mec.12305

Wagner CE (2018) Improbable big birds Darwin's finches prove a mechanism for the rapid formation of new species. Science 359:157–159. https://doi.org/10.1126/science.aar4796

Wu MY, Lau CJ, Ng EYX et al (2022) Genomes from historic dna unveil massive hidden extinction and terminal endangerment in a tropical asian songbird radiation. Mol Biol Evol. https://doi.org/10.1093/MOLBEV/MSAC189

Yang Z, Yoder AD (2003) Comparison of likelihood and Bayesian methods for estimating divergence times using multiple gene loci and calibration points, with application to a radiation of cute-looking mouse lemur species. Syst Biol 52:705–716. https://doi.org/10.10635150390235557

Yeh FC, Yang RC, Boyle TBJ, et al (1997) POPGENE, the user-friendly shareware for population genetic analysis. Mol Biol Biotechnol centre, Univ Alberta, Canada. https://sites.ualberta.ca/~fyeh/popgene.html. Accessed 13 Jun 2022

CHAPTER 3

Literature Review – Biological Clocks and Age

This chapter represents the literature review for the second biological clock of the study which focuses on the association between epigenetic clocks and programmed cellular aging to construct molecular age determination models for animals. The chapter is comprised of three published articles for which the first is a systematic review and meta-analysis (Le Clercq *et al.*, 2023*d*) of the existing literature and associated data for studies that used methylation or telomeres to model age in animals and is structured according to taxonomy. The second article provides further details on the methods used and data gathered as part of the review (Le Clercq *et al.*, 2023*e*). The third article (Le Clercq, 2024) details a custom PYTHON program (Le Clercq, 2023*c*) created to assist in the calculation of potential author bias for included studies in meta-analyses as well as contextualising the literature and performing scientometric analyses. This is critical to reviews in ecology where author contributions and study distributions are more likely to vary, as compared to medical systematic reviews, and could possibly affect overall interpretations. Supplementary materials and methods are included in Appendix C and D.

Articles:

LE CLERCQ, L., KOTZÉ, A., GROBLER, J.P. & DALTON, D.L. (2023*d*) Biological clocks as age estimation markers in animals: a systematic review and meta-analysis. *Biological Reviews* **98**, 1972–2011.

LE CLERCQ, L.-S., GROBLER, J.P., KOTZÉ, A. & DALTON, D.L. (2023*e*) Dataset generated in a systematic review and meta-analysis of biological clocks as age estimation markers in animal ecology. *Data in Brief* **51**, 109615.

LE CLERCQ, L. (2024) ABCal: a Python package for Author Bias Computation and Scientometric Plotting for Reviews and Meta-Analyses. *Scientometrics* **129**, 581–600.

GitHub Repositories:

LE CLERCQ, L.S. (2023*c*) ABCal: Author Bias Computation and Scientometric Plotting. GitHub. Available from: https://github.com/LSLeClercq/ABCal.

BIOLOGICAL REVIEWS
Cambridge Philosophical Society
Biol. Rev. (2023), **98**, pp. 1972–2011.
doi: 10.1111/brv.12992

Biological clocks as age estimation markers in animals: a systematic review and meta-analysis

Louis-Stéphane Le Clercq[1,2,*], Antoinette Kotzé[1,2], J. Paul Grobler[2] and Desiré Lee Dalton[3]

[1] *South African National Biodiversity Institute, P.O. Box 754, Pretoria 0001, South Africa*
[2] *Department of Genetics, University of the Free State, P.O. Box 339, Bloemfontein 9300, South Africa*
[3] *School of Health and Life Sciences, Teesside University, Middlesbrough, TS1 3BA, UK*

ABSTRACT

Various biological attributes associated with individual fitness in animals change predictably over the lifespan of an organism. Therefore, the study of animal ecology and the work of conservationists frequently relies upon the ability to assign animals to functionally relevant age classes to model population fitness. Several approaches have been applied to determining individual age and, while these methods have proved useful, they are not without limitations and often lack standardisation or are only applicable to specific species. For these reasons, scientists have explored the potential use of biological clocks towards creating a universal age-determination method. Two biological clocks, tooth layer annulation and otolith layering have found universal appeal. Both methods are highly invasive and most appropriate for *post-mortem* age-at-death estimation. More recently, attributes of cellular ageing previously explored in humans have been adapted to studying ageing in animals for the use of less-invasive molecular methods for determining age. Here, we review two such methods, assessment of methylation and telomere length, describing (*i*) what they are, (*ii*) how they change with age, and providing (*iii*) a summary and meta-analysis of studies that have explored their utility in animal age determination. We found that both attributes have been studied across multiple vertebrate classes, however, telomere studies were used before methylation studies and telomere length has been modelled in nearly twice as many studies. Telomere length studies included in the review often related changes to stress responses and illustrated that telomere length is sensitive to environmental and social stressors and, in the absence of repair mechanisms such as telomerase or alternative lengthening modes, lacks the ability to recover. Methylation studies, however, while also detecting sensitivity to stressors and toxins, illustrated the ability to recover from such stresses after a period of accelerated ageing, likely due to constitutive expression or reactivation of repair enzymes such as DNA methyl transferases. We also found that both studied attributes have parentally heritable features, but the mode of inheritance differs among taxa and may relate to heterogamy. Our meta-analysis included more than 40 species in common for methylation and telomere length, although both analyses included at least 60 age-estimation models. We found that methylation outperforms telomere length in terms of predictive power evidenced from effect sizes (more than double that observed for telomeres) and smaller prediction intervals. Both methods produced age correlation models using similar sample sizes and were able to classify individuals into young, middle, or old age classes with high accuracy. Our review and meta-analysis illustrate that both methods are well suited to studying age in animals and do not suffer significantly from variation due to differences in the lifespan of the species, genome size, karyotype, or tissue type but rather that quantitative method, patterns of inheritance, and environmental factors should be the main considerations. Thus, provided that complex factors affecting the measured trait can be accounted for, both methylation and telomere length are promising targets to develop as biomarkers for age determination in animals.

Key words: animals, biological clocks, tooth layers, otoliths, epigenetics, methylation, telomeres, biomarker, age determination, meta-analysis.

* Author for correspondence (Tel.: +2782 551 5895; E-mail: leclercq.l.s@gmail.com).

CONTENTS

I. INTRODUCTION

Age is a key factor in the study of animal populations, as many critical biological attributes change over time. Age classes are often linked to biological functionality in terms of developmental stage (Noren & Edwards, 2007), reproductive maturity (Charpentier *et al.*, 2008), future reproductive potential (Jones *et al.*, 2014), reproductive success

(Bradley & Safran, 2014), disease burden (Starr & Saito, 2014), and risk of mortality (Pérez-Barberìa *et al.*, 2014). These attributes typically differ between discreet age classes and the fraction of a population contained within the different age classes therefore assists conservation biologists to estimate their future fitness (Pérez-Pereira, López-Cortegano & García-Dorado, 2022). Multiple approaches have been used with varying degrees of success to determine the age of animals and often rely upon either external features that change predictably over time, a mark–recapture approach (Clutton-Brock & Sheldon, 2010), or biochemical attributes.

The first, and most common method, uses morphometrics or physical measurements of the skull (Smuts, Anderson & Austin, 1978), body (Marker & Dickman, 2003), scales (Isermann, Wolter & Breeggemann, 2010), wingspan (Erdem *et al.*, 2021), or mass to determine age by comparing measures to an established chart. While this method is particularly useful for juveniles and young adults it generally loses resolution when applied to adults of varying age classes in species where growth arrests at adulthood. To overcome this, morphometrics have also been complemented with other physical attributes such as degree of gumline recession (Hiller & Tyre, 2014) or the condition of the pelage (Marker & Dickman, 2003). Drawbacks of this approach are (*i*) that it requires the capture and prolonged immobilisation of the animal to record the measures, (*ii*) few and sparsely detailed validation studies exist (which may be difficult to conduct), and (*iii*) there is high inter-observer error, particularly in cases where a reasonable amount of expertise is needed to discern age based on observed traits.

The mark–recapture approach either relies upon distinct natural markings and external features that are documented, or artificial markings applied to young individuals upon first encounter that are used to verify the identity and age of the individual upon subsequent encounters. Natural markings that can be used to identify individuals include, for example, pigmentation patterns on the flukes of humpback whales [*Megaptera novaeangliae* (Borowski)] (Polanowski *et al.*, 2014). Artificial markings, on the other hand, include manmade tools of identification such as uniquely numbered bands (Sherley *et al.*, 2014) and geolocators (Saino *et al.*, 2015) attached to birds, tracking collars (Marker *et al.*, 2003) attached to cheetahs [*Acinonyx jubatus* (Schreber)], or teracycline marking (Cailliet, 1990) in fishes. A major pitfall in the use of these methods is that they rely upon the eventual re-encounter of the same individual and require the attachment or injection of a marker to a live animal. In addition, the lifespan of many animals far outreaches the average lifespan and career of conservationists, making long-term tracking unsuitable in some cases.

For these reasons, conservationists have frequently sought a biomarker or method of age determination that is universally applicable to most species, provides a direct measure of age, and does not require extensive experience or detailed individual validation studies. While several biochemical, physiological, and haematological attributes have been suggested (Short, Williams & Bowden, 1987; Cailliet, 1990; Chaney *et al.*, 2003), the most promising method for achieving this is to identify measurable features of biological clocks that change on an annual basis. A biological clock can be defined as: an innate mechanism that controls the physiological activities of an organism which change on a daily, seasonal, yearly, or other regular cycle. Two biological clocks have found widespread applications in the estimation of animal age. In fish, histological features of internal structures such as growth rings in otoliths, small stones found within the auditory system, have been used to determine age by counting stria that form from each annual deposition of a new layer (Campana, 2001). Similarly, the annual cycle of depositing a new layer of dentin (Lockyer, 1993; Lockyer *et al.*, 2007) or cementum (Wittwer-Backofen, Gampe & Vaupel, 2004) on teeth can be used to determine the age of mammals. Tooth cementum annulation (TCA; reviewed in Viciano, López-Lázaro & Tanga, 2022) was first documented across diverse species of mammals and reptiles in the 19th century (Owen, 1840) and has been a reliable method of age estimation in animals for nearly 75 years. TCA has been used to determine the age of more than a hundred species in the class Mammalia alone: across 10 orders and 38 families (see online Supporting Information, Table S1). There are, however, still several shortcomings to these methods, e.g. otoliths are only effective in fish, tooth histology is only accurate in mammals, and both methods are extremely invasive, either requiring the extraction of a tooth or only being appropriate in *post-mortem* age-at-death determination. Birds, on the other hand, represent a unique conundrum as they lack both teeth and otoliths, requiring the exploration of alternatives such as the measurement of pentosidine accumulation in the skin of lilac-breasted rollers, *Coracias caudatus* (Linnaeus), among others (Chaney *et al.*, 2003).

With the dawn of the molecular revolution many methods have been developed that use DNA to ascertain and study critical individual characteristics as well as population attributes within most species. Such attributes include sex (Morin *et al.*, 2005), parentage (Double *et al.*, 1997), population assignment (Kim & Sappington, 2013), hybridization (Brelsford, Milá & Irwin, 2011), speciation (Linck, Freeman & Dumbacher, 2020), and migration (Merlin & Liedvogel, 2019). As natural ageing is in many respects a programmed process under genetic control (Horvath, 2013), or results from unrepaired environmental damage to DNA (Kujoth *et al.*, 2005), there is great potential to exploit these changes to determine age. Some age-related changes, summarised in Fig. 1, include fluctuating levels of transcripts (Cook *et al.*, 2006), changes to the DNA sequence itself, e.g. mitochondrial mutations (Zapico & Ubelaker, 2016), and epigenetic modifications (Jones, Goodman & Kobor, 2015). Epigenetics is a collective term for mechanisms that modify DNA and DNA packaging, independent of changes to the genetic sequence (Bird, 2007).

Studies seeking to establish a genetic method for age estimation have identified several potential targets among the epigenetic regulatory elements including methylation

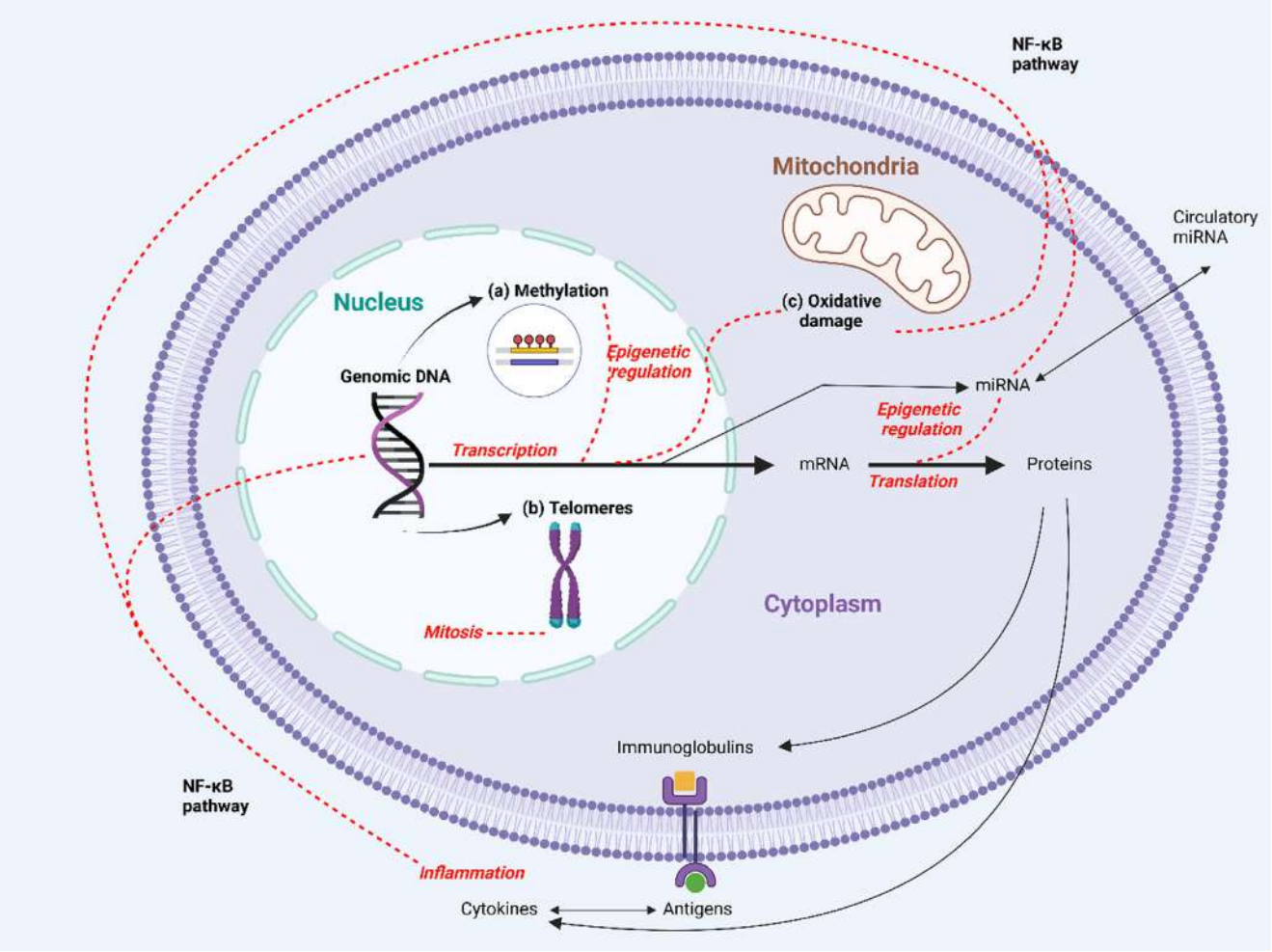


Fig. 1. An animal cell indicating the mechanisms of cellular ageing at the molecular level that are currently used as molecular biomarkers for age. Molecules are labelled in plain text and processes in red italics. Movement of molecules is indicated by lines with arrows; regulatory effects on molecular pathways are indicated by broken red lines. The main pathway of gene expression is shown by thicker arrows in the centre, and comprises the two phases transcription and translation, that turn information stored as DNA into functional molecules including proteins and various types of RNA. (a) DNA methylation levels at several sites change with age. (b) Telomeres have roles in many age-related cellular processes and shorten over time and with age. (c) Oxidative damage due to chronic inflammation through the constitutive activation of the nuclear factor-kappa B (NF-κB) pathway leads to mitochondrial mutations that accumulate over time. miRNA, microRNA; mRNA, messenger RNA. Image created in BioRender.com.

(Johansson, Enroth & Gyllensten, 2013) and telomere shortening (Thomas & Zapico, 2017). Both methylation and telomere shortening were initially tested in humans (Aubert & Lansdorp, 2008; Gopalan *et al.*, 2017) but have since been evaluated as valid methods of less-invasive age determination (Carroll *et al.*, 2018) that is transferable across most species (Jarman *et al.*, 2015; De Paoli-Iseppi *et al.*, 2017*a*). This review provides a synopsis and meta-analysis of two genetic biological clocks for age estimation, methylation and telomeres, in terms of (*i*) what they are, (*ii*) what changes they undergo throughout a lifespan, and (*iii*) the potential use of these features as applied to biological age determination in wildlife, evidenced by a meta-analysis of the literature.

II. METHODS

(1) Literature search and review

A systematic approach was used to search for and synthesise the available literature, as depicted in the PRISMA flow diagrams in Figs S1 and S2 (Haddaway *et al.*, 2022). Literature for methylation-based studies was searched for in the *Scopus* (www.scopus.com) and *Dimensions* (www.dimensions.ai) databases using the following Boolean search string: ('Epigenetics' OR 'Methylation') AND ('age' OR 'aging') AND ('determination' OR 'model') AND ('Animals' OR 'wild'). This process was repeated for telomere-based studies by substituting 'Telomere' for 'Methylation'. Search results

were exported in the comma separated value format and the literature was subsequently summarised, guided by citation networks visualised using CitNetExplorer 1.0.0 and VOSviewer 1.6.16 (van Eck & Waltman, 2017). Results retrieved from *Scopus* were converted to the appropriate format with the R package *Scopus2CitNet 0.1.0.0* in RStudio 1.4.1106 (RStudio Team, 2021), running R version 4.0.5 (R Core Team, 2020). Further studies were identified from the reference lists of articles and free term 'citation searches' on *Google Scholar*. Other background information beyond the studies for systematic review was retrieved with free term searches using 'DNA methylation', 'telomere length', or 'biological clocks for age estimation' and are not included in the PRISMA statements. Studies were summarised based on species covered, publication year, method used, number of sites assayed (methylation), and sample size (N). Individual models reported in studies were summarised by tissue type, species, and method used, with their associated correlation coefficient (R^2), F-test statistic, t-test statistic, or chi-squared (χ^2) test statistic and P-value. Methylation studies that reported more than one model usually presented models using the same method, but different tissue types or species tested, while a few telomere studies reported multiple models generated to compare different measurement methods on the same samples. Where appropriate statistical measures were not reported, values were calculated from data in the online supplement or data extracted from plots using *WebPlotDigitizer* version 4.6 (Rohatgi, 2022). The data set used in this review is available online (see Section 0. Data availability statement). Deviation between predicted and known age are reported as mean error of prediction (MEP) due to variation in reported measure in studies e.g. median absolute error (MAE) or root mean-square error (MSE). Genome information for species, or their closest relative, was retrieved from the *Animal Genome Size Database* (Gregory, 2023), *Animal Chromosome Count Database* (Román-Palacios *et al.*, 2021), the *Bird 10,000 Genomes (B10K) Project* (Zhang, 2015), and/or the *Bird Chromosome Database* (Degrandi *et al.*, 2020).

(2) Meta-analysis

To conduct a meta-analysis, models reported in studies that were adequately similar in terms of tissue sample and method of detection/assay were identified from the systematic review, including 78 studies for methylation and 109 studies for telomeres. Models reported in studies with a sufficient sample size ($N \geq 15$) were included for consideration in the meta-analysis. For methylation, pan-species models with human samples were excluded from further analyses to focus only on non-human models. An additional two methylation models were also excluded (Ito, Yoshimura & Momoi, 2017; Horvath *et al.*, 2022*d*). The first was excluded because neither the correlation coefficient nor the data plots for the combined model were reported, the latter was excluded due to a discrepancy in the reported sample size for their *Macropus* model (59 in the table but 65 in the figure). The final data set included 60 methylation-based models across the five vertebrate classes: fishes ($N = 3$), amphibians ($N = 2$), reptiles ($N = 3$), birds ($N = 3$), mammals ($N = 49$). For mammals this included aquatic mammals ($N = 6$), bats ($N = 2$), carnivores ($N = 9$), elephants ($N = 2$), marsupials ($N = 3$), primates ($N = 10$), rodents ($N = 4$), shrews ($N = 1$), and ungulates ($N = 12$). For methylation we classified models into three groups. Group 1 ($N = 48$) models were those that collected data for a single species and used those data to create a molecular model for predicting age in the same species. For Group 2 ($N = 8$), the models used a large sample size in a related species to create a model, and subsequently tested the model in another species using fewer samples. Group 3 models ($N = 4$) developed and tested a single model using samples from multiple related species such as cetaceans, bats, cod, or pinnipeds. For the included methylation studies the correlation coefficient and sample size were used to calculate the Fisher's-Z_r transform and the variance (v) as a means of approximating the correlation coefficient effect size (r) for cross-study meta-analysis.

Two telomere studies were excluded from the main analyses, the first was completely excluded due to small sample sizes (Whittemore *et al.*, 2019) while the second was analysed separately, as for methylation, due to the pooling of samples for two species (group 2) of marten (Pauli *et al.*, 2011). Due to the large number of studies on birds, 12 avian models were selected at random for inclusion in the cross-taxa analyses ($N = 60$) while the full set of bird models ($N = 51$) were analysed separately to avoid phylogenetic bias. The first telomere data set, for cross-taxa analyses, included 60 models across the five vertebrate classes: fishes ($N = 6$), amphibians ($N = 3$), reptiles ($N = 6$), birds ($N = 12$), mammals ($N = 33$). For mammals this included aquatic mammals ($N = 2$), bats ($N = 6$), carnivores ($N = 13$), elephants ($N = 1$), primates ($N = 2$), rodents ($N = 4$), and ungulates ($N = 5$). Telomere models for cross-taxa analyses were also grouped, as for methylation, with group 1 studies ($N = 59$) representing studies on single species while group 2 studies ($N = 1$) included models based on pooled samples for multiple species.

The second telomere data set, on studies conducted exclusively in birds, included 51 models and were grouped based on key avian taxonomic groupings. This included specific taxonomic groupings such as: chats and flycatchers ($N = 6$); corvids ($N = 4$); cuckoos ($N = 1$); finches ($N = 3$); geese and waterfowl ($N = 1$); gulls, shorebirds, and waders ($N = 4$); juncos, buntings, and sparrows ($N = 2$), magpies ($N = 1$); manakins ($N = 1$); parrots ($N = 1$); penguins ($N = 3$); seabirds ($N = 11$); swallows ($N = 5$); swifts ($N = 1$); tits ($N = 3$); warblers ($N = 3$); and wrens ($N = 1$). For the included telomere studies statistical methods varied, thus effect size was calculated as before or adapted to the tests used in the study e.g. F- or χ^2-test statistics.

The individual meta-analyses were conducted in RStudio 1.4.1106 (RStudio Team, 2021), running R version 4.0.5 (R Core Team, 2020) with the package *meta 5.5-0* (Schwarzer, Carpenter & Rücker, 2015) and effect sizes calculated by *compute.es 0.2-2* (Del Re, 2013). The relative anticipated levels of author bias resulting from multiple

publications by the same authors was assessed using a custom PYTHON script, *Author Bias Computation* (*ABCal*) version 1.0.1 (see Section 0). Meta-analyses were run with grouping based on class, study group, tissue type, and method. A separate meta-analysis was run between biological clocks, rather than studies, to compare the efficacy of either methylation or telomere length for determining age in animals.

Heterogeneity was measured between studies within groups as well as globally using standard heterogeneity percentages (I^2), tau-squared (τ^2), and the *P*-value for the χ^2-test statistic. Potential causes of heterogeneity were explored with a meta-regression of individual factors such as author, publication year, taxonomic class, tissue type, experimental method and either genome size (methylation) or karyotype (telomere). Type I error was avoided by having adequate sample sizes ($N > 50$), carrying out one-tailed analyses when possible, and reporting significance at standard alpha values along with correlation coefficients (R^2). To avoid potential type I error due to multiple testing, introduced by collinearity, each factor was measured in a separate and independent regression. Publication or small study bias, resulting in asymmetry, was assessed in R. The relationships between statistical power, effect size and sample size (N) for study methods were explored using G*Power version 3.1 (Faul *et al.*, 2009; Kang, 2021) and are graphically depicted in Fig. S3. A separate meta-regression analysis was run between study methods, rather than models, to compare the efficacy of either methylation or telomere length for determining age in animals.

III. BIOLOGICAL CLOCKS AND AGE

(1) DNA methylation

(a) *What is DNA methylation?*

DNA methylation is an epigenetic process that modifies DNA through the addition of a methyl group (CH_3) to the 5′ cytosine of cytosine–guanine pairs (CpGs) (Fig. 2), as part of DNA packaging as well as the regulation of gene expression. Active methylation of unmethylated CpGs is catalysed by two DNA methyltransferase enzymes, DNMT3A and DNMT3B (Chédin, 2011), that add a methyl group to the fifth carbon of cytosines to form 5-methylcytosine. Methylation status of methylated CpGs is maintained between replicative cycles through the enzymatic activity of DNA methyltransferase 1 (DNMT1), an enzyme that is able to recognise hemi-methylated DNA after cell division and add a methyl group to the fifth carbon of cytosines on unmethylated daughter strands (Sen *et al.*, 2010).

DNA is demethylated *via* one of two mechanisms, either active or passive demethylation (Guo *et al.*, 2014). Passive demethylation occurs at cell division, where the newly synthesised daughter strands lack a methyl group at the newly incorporated cytosines and methylation cannot be re-established by DNMT1 prior to subsequent division (Kagiwada *et al.*, 2013). As passive demethylation only occurs during highly active sequential cycles of cell division it is most common during embryogenesis and early development (Andergassen *et al.*, 2021), as well as certain types of malignant growths (Gomez *et al.*, 2007; Zhang *et al.*, 2020).

By contrast, active demethylation is an enzymatic process (Ooi & Bestor, 2008) involving TET methylcytosine dioxygenase enzymes, such as TET2, and thymine DNA glycosylase (TDG). TET enzymes demethylate DNA by repeatedly adding a hydrogen to the methyl group on the fifth carbon of cytosines, first forming 5-hydroxymethylcytosine, followed by 5-formylcytosine, and finally 5-carboxylcytosine (Fig. 2C). This then serves as the substrate for TDG, which decarboxylates 5-carboxycytosine back to an unmethylated cytosine (Dalton & Bellacosa, 2012). This process is typically associated with the uncoiling of DNA to provide access to promotor and enhancer elements for transcription factors such as nuclear factor kappa B (Tilstra *et al.*, 2011; Balistreri *et al.*, 2013), as well as polymerase complexes responsible for transcription and replication (Attwood, Yung & Richardson, 2002).

The methylation of CpGs is variable across the genome and is a heritable trait through parental imprinting (Andergassen *et al.*, 2021) (Fig. 3), as well as trans-generational epigenetic inheritance (Fitz-James & Cavalli, 2022), while still being subject to environmental influence (McEwen *et al.*, 2017), and complements other genetic evolutionary processes (Lind & Spagopoulou, 2018; Parrott & Bertucci, 2019; Cavalli & Heard, 2019). Those located within CpG islands, dense clusters of consecutive CpGs, tend to be hypermethylated and occur in proximity to regulatory elements of genes or transposable elements to maintain genomic stability. Those located outside such islands, within genes or in non-regulatory elements, are most often hypomethylated (Johansson *et al.*, 2013).

(b) *Changes with age*

The changes that occur to methylation during the natural ageing process have formed the subject of multiple extensive reviews of both human and rodent models (Horvath & Raj, 2018) and will therefore only be dealt with briefly. Epigenetic remodelling or reprogramming in the form of altered methylation status is most active during embryogenesis (Andergassen *et al.*, 2021) and gametogenesis (Guo *et al.*, 2014) (Fig. 4), while methylation status is generally more stable in differentiated tissue. Gradual changes in methylation status in differentiated tissue have, however, been observed (Bell *et al.*, 2012) and are sub-classified based on causation into two fundamental processes: epigenetic drift (Jones *et al.*, 2015) and an epigenetic clock (Horvath & Raj, 2018). The tendency for an inconsistent increase in variability in the epigenomes of individuals denotes epigenetic drift, whereas the epigenetic clock comprises site-specific changes that consistently relate to age across individuals (Hannum *et al.*, 2013). Studies have revealed that during ageing a genome-wide trend of hypomethylation occurs while CpG islands tend to become hypermethylated (Hannum *et al.*, 2013; Jones *et al.*, 2015) and that within genes, nearly

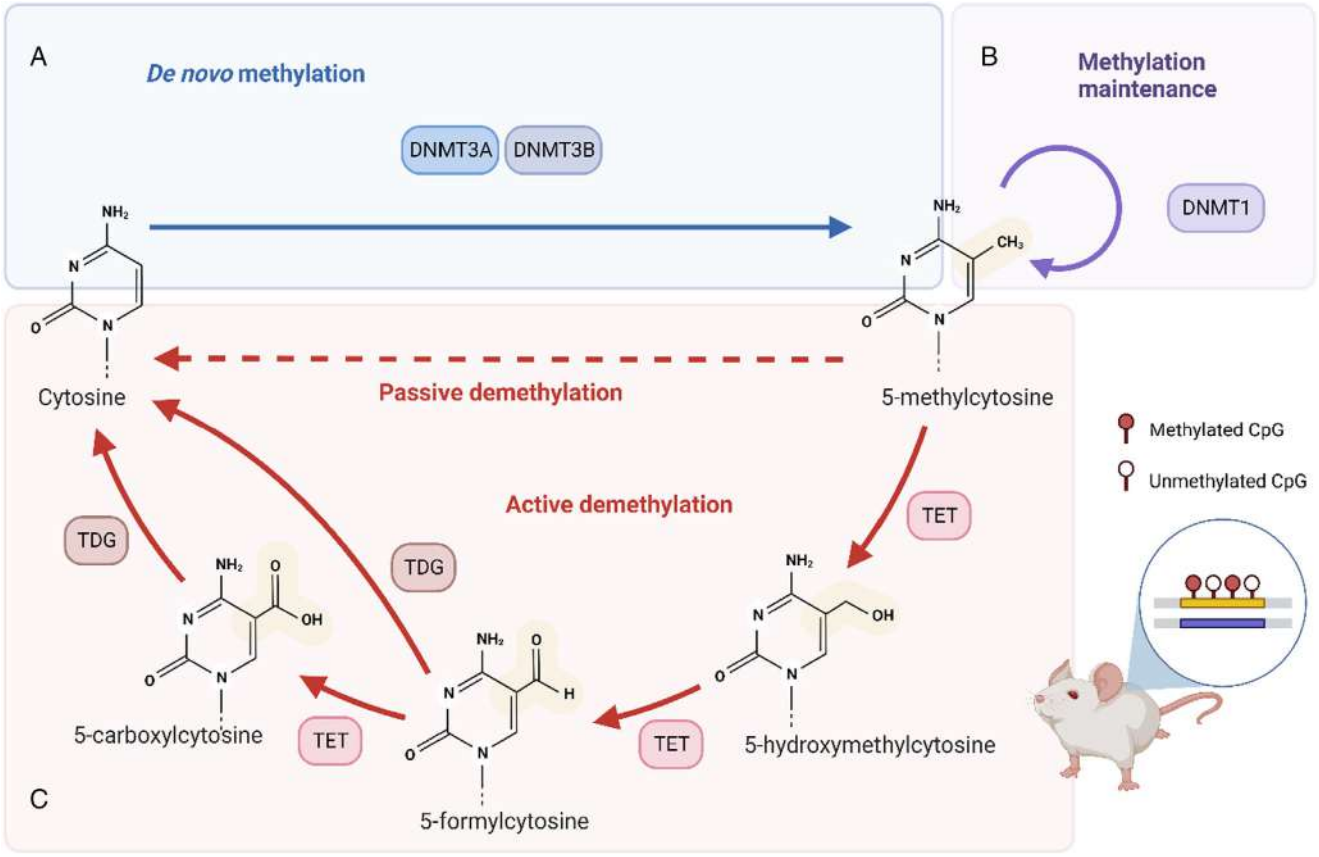


Fig. 2. Diagram of the regulation of methylation. (A) Active *de novo* methylation occurs at unmethylated cytosines through the activity of two DNA methyltransferases 3 (DNMT3A and DNMT3B) that are able to add a methyl group to unmethylated DNA at carbon 5 of cytosines to form 5-methylcytosine. (B) Maintenance of methylation is achieved through the enzymatic activity of DNA methyltransferase 1 (DNMT1) that is able to recognise hemi-methylated DNA after cell division and adds a methyl group to carbon 5 of cytosines on newly synthesised daughter strands. (C) DNA is demethylated through either active or passive demethylation. Passive demethylation happens when cells divide, and the newly synthesised daughter strands do not contain a methyl group at the newly incorporated cytosines. Active demethylation is an enzymatic process involving Tet methylcytosine dioxygenase (TET) enzymes and thymine DNA glycosylase (TDG). TET enzymes sequentially hydroxylate the methyl group on carbon 5 of cytosines to form 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxylcytosine while TDG decarboxylates 5-carboxycytosine back to an unmethylated cytosine. CpG, cytosine–guanine pair. Image created in BioRender.com.

a third of all CpG sites are influenced by age (Johansson *et al.*, 2013).

(*c*) *DNA methylation as a biomarker of animal age*

Given its consistency, the epigenetic clock is a promising avenue for chronological age prediction (Lowe *et al.*, 2018; Hong *et al.*, 2019), particularly within the field of conservation (Rey *et al.*, 2020). Methylation levels are usually quantified by bisulfite sequencing (Ashapkin, Kutueva & Vanyushin, 2020), although several methods exist, and the selection of the most appropriate method depends on the relevant application (Kurdyukov & Bullock, 2016). Bisulfite sequencing methods typically provide data in the form of β-methylation values ranging from 0 to 1, and has been subject to extensive statistical validation to optimise sample sizes (Mayne, Berry & Jarman, 2021*b*) and model methods (Snir, vonHoldt & Pellegrini, 2016), including the development of useful tools to design new assays and analyse CpGs (Statham & Csárdi, 2008; Thompson *et al.*, 2009; Wei *et al.*, 2021).

Thus far, methylation studies have successfully been used to establish age models in several animal species and taxa, including: whales (Polanowski *et al.*, 2014; Tanabe *et al.*, 2020; Bors *et al.*, 2021), fishes (Mayne *et al.*, 2020*a*, 2021*a*), dolphins (Beal *et al.*, 2019), birds (De Paoli-Iseppi *et al.*, 2019), non-human primates (Ito *et al.*, 2018), rodents (Stubbs *et al.*, 2017; Lowe *et al.*, 2020), canids (Thompson *et al.*, 2017), bats (Wilkinson *et al.*, 2021) and equids (Larison *et al.*, 2021*a*,*b*). The results for these studies are summarised in Table 1.

(i) Fishes (classes Agnatha, Chondrichthyes, and Osteichthyes). Methylation studies for age estimation have also been expanded to other aquatic species. The first study explored the utility of this biomarker in zebrafish [*Danio rerio* (Hamilton)], as a model organism (Mayne *et al.*, 2020*a*). This

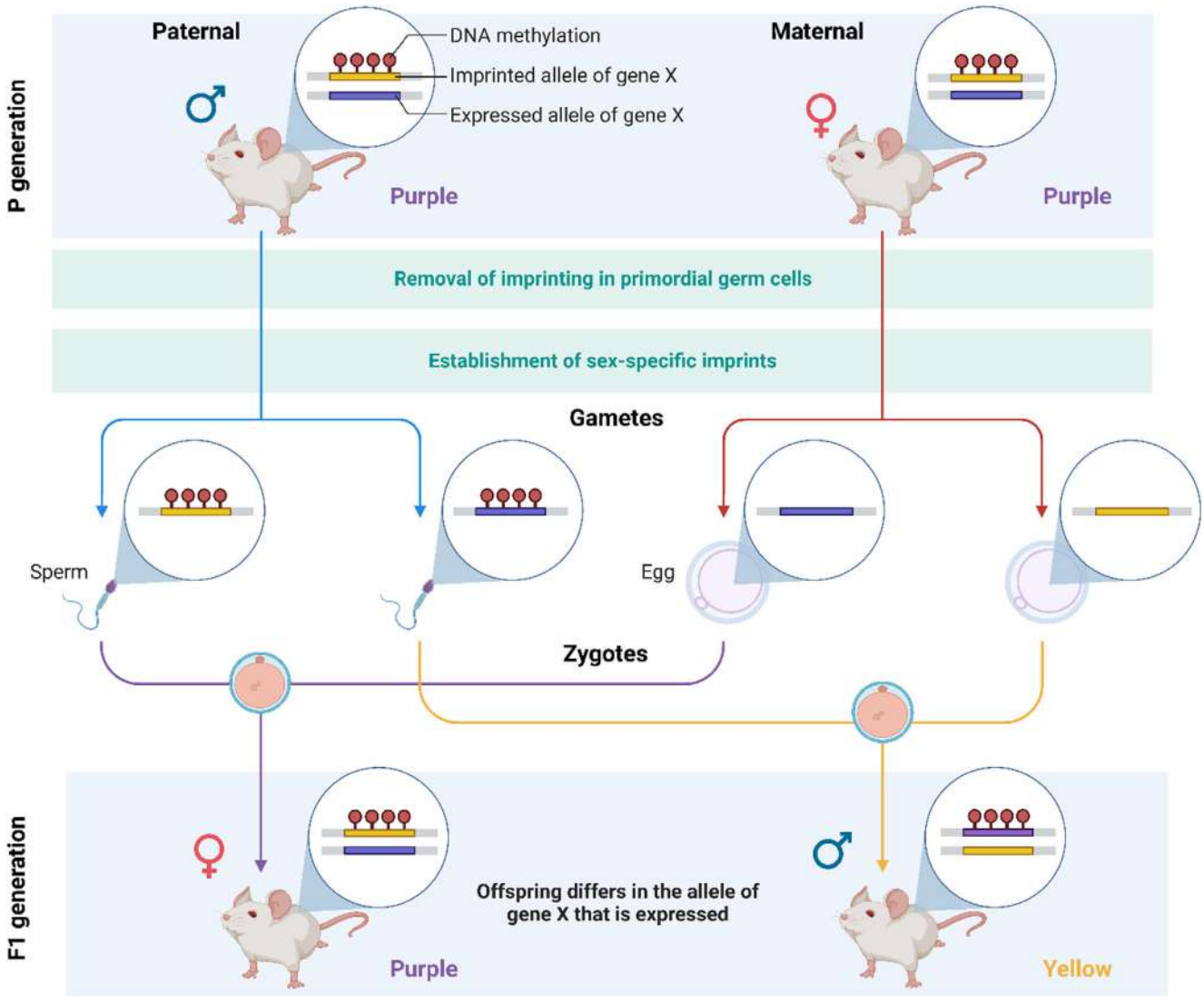


Fig. 3. Diagram of parental imprinting showing the heritability of methylation for a paternally imprinted gene. The top panel shows the parental (P) generation of mice that have identically imprinted (yellow) and expressed (purple) alleles for a hypothetical gene X. Through the removal of inherited imprints in primordial germ cells and the subsequent establishment of sex-specific imprints the alleles inherited paternally are methylated and not expressed, while the maternally inherited alleles are unmethylated and will be expressed. When a zygote is formed through the union of two gametes, the resulting first filial (F1) generation will differ in which allele is expressed, with some expressing the purple allele whilst others express the yellow allele. Image created in BioRender.com.

study used multiplexed methylation-sensitive PCR (MS-PCR) to assay 26 of 29 CpGs shown to follow an age-related decline and was able to establish a predictive model with a standard error of weeks. There was, however, a higher error rate in prediction with a separate test cohort and the authors concluded that MS-PCR may not be sensitive enough to establish age estimation models. The first study on a wild population (Anastasiadi & Piferrer, 2020) assayed methylation in European seabass [*Dicentrarchus labrax* (Linnaeus)]. This study constructed an age clock from 48 CpGs with an error of prediction of 2–3 years, however, the methodology required deep tissue samples and euthanasia similar to otolith age determination. The methods developed in zebrafish were later replicated by the same authors (Mayne *et al.*, 2021*a*) to produce two non-lethal age determination models from skin/fin samples for three threatened fish species: Australian lungfish [*Neoceratodus forsteri* (Krefft)], Murray cod [*Maccullochella peelii* (Mitchell)], and Mary River cod [*Maccullochella mariensis* (Rowland)]. The lungfish clock used 31 CpGs while the cod clock used 26 CpGs, and both had a standard error of prediction of less than a year. The results from a validation test sample showed significantly less error than in zebrafish, which may be related to the different lifespans of these species. Age-related

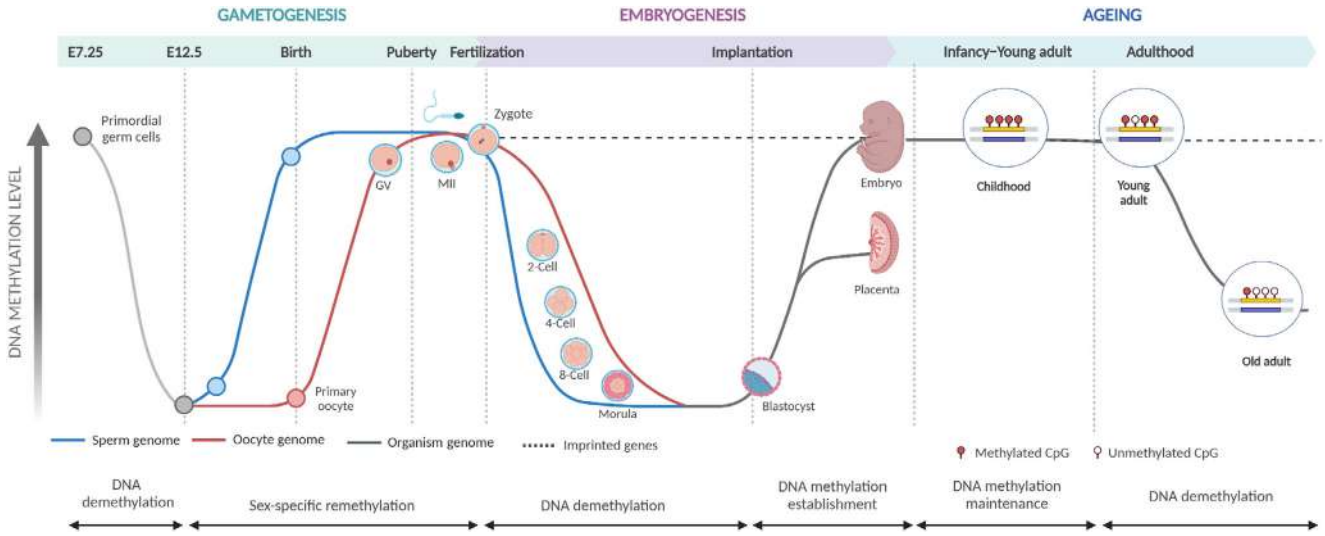


Fig. 4. Diagram of the reprogramming of methylation during key life stages from gametogenesis to embryogenesis and the resulting changes that occur during natural ageing. During gametogenesis the inherited genomic DNA is actively demethylated in primordial germ cells and sex-specific methylation re-established by *de novo* methylation when gametes, either sperm or egg cells, are created. After a zygote is formed from the union of two gametes, passive genomic demethylation takes place due to the rapid replication and division of cells while the methylation of imprinted genes is actively maintained. Once the blastocyst stage/implantation is reached, global methylation is re-established in the embryo. After birth, methylation is maintained throughout childhood from infancy to early adulthood in most tissues with the exception of primordial germ cells forming gametes. During the ageing process methylation is maintained at imprinted genes while global DNA demethylation occurs throughout the remainder of the genome. CpG, cytosine–guanine pair. Image created in BioRender.com.

methylation changes have also been observed in long snout seahorses [*Hippocampus reidi* (Ginsburg)], across different early developmental stages up to adulthood (Suarez-Bregua *et al.*, 2021).

(ii) Amphibians (class Amphibia). The first age clock using methylation for amphibians was recently established in Western clawed frogs [*Xenopus tropicalis* (Gray)] using bisulphite sequencing of the whole genome from web/skin tissue (Morselli *et al.*, 2023). A total of 331 age-correlated CpGs were identified, from which an age clock was established that was able to classify individuals into young, middle, or old age classes with high accuracy; the largest errors occurred in very young or very old individuals. A study published in parallel (Zoller *et al.*, 2022), on both Western clawed frog and African clawed frogs [*Xenopus laevis* (Daudin)], used the same methylation array chip previously developed in mammals, found thousands of conserved CpGs and constructed both individual as well as combined clocks for age. In the Western clawed frog, the mean error of prediction for age was less than a year, while the error for African clawed frog was ∼3 years. The observed difference in accuracy for predictions was likely due to the differences in lifespan as African clawed frogs have a higher average lifespan. A major difference between this and the former study is that the latter models were based on pan-tissue data, including organ and muscle tissue, as the frogs were killed.

(iii) Reptiles (class Reptilia). Early methylation studies in reptiles, done within the context of eco-toxicology (Parrott *et al.*, 2014; Nilsen *et al.*, 2016; Cocci *et al.*, 2018) and sexual differentiation (Venegas *et al.*, 2016), found evidence of an age-related decline in methylation. In the American alligator [*Alligator mississippiensis* (Daudin)], liquid chromatography–tandem mass spectrometry (LC-MS/MS) was used to quantify global methylation patterns in sub-adult and adult groups. The results from both studies revealed a global decline in genomic methylation of ∼8% (Parrott *et al.*, 2014) and 7–9% (Nilsen *et al.*, 2016) respectively. Interestingly both studies also identified a site-directed relationship with the decline in global methylation. This indicates that, while global hypomethylation occurs with ageing, the rate of epigenetic ageing may vary across the genome within a species where specific genes or genomic sites change more slowly or more rapidly than global changes. Subsequently, age clocks based on methylation were investigated in three species of sea turtle, species that have a high infant mortality, are often migratory, and long lived: making long-term tracking for scientific study particularly difficult. The first study exclusively evaluated lifespan prediction in two sea turtles, Australian flatback sea turtle [*Natator depressus* (Garman)] and leatherback sea turtle [*Dermochelys coriacea* (Vandelli)], using CpG promotor density estimates by polymerase chain reaction (PCR) and sequencing (Mayne *et al.*, 2020*b*). The authors were able to predict the lifespan of both turtles as ∼50 and 90 years respectively, in line with previous estimates, and provided the first evidence that methylation studies may be suitable for ageing reptilian species. A subsequent study, using known-age samples from green sea turtle [*Chelonia mydas* (Linnaeus)], developed a methylation-based

Table 1. Summary of studies that assayed global or regional methylation in relation to development or age in animals. Studies that used the data to create a molecular model for predicting animal age, which were included in the meta-analysis, are indicated with an asterisk (*). Species are grouped by class to match the text and listed alphabetically. For each species the genome size in picograms or billions of base-pairs (bbp) is given along with estimates for the average lifespan. For each study, the method of measuring methylation is indicated: a, bisulfite sequencing; b, enzyme-linked immunosorbent assay (ELISA); c, liquid chromatography–tandem mass spectrometry (LC-MS/MS); d, methylation array chip; e, methylation-sensitive polymerase chain reaction (MS-PCR); f, cytosine–guanine (CpG) promotor density (lifespan). Where relevant, citations are also provided for studies on age determination using other biological clocks (tooth annulation/otolith) in the same/similar species. n.a., not available.

Species	Genome (bbp)	Lifespan (years)	Study	Method	Tooth annulation/otolith
(i) Fishes:					
Australian lungfish (*Neoceratodus forsteri*)*	60.73	80–85	Mayne *et al.* (2021*a*)	e	Høie *et al.* (2004)
European seabass (*Dicentrarchus labrax*)*	0.78	15–30	Anastasiadi & Piferrer (2020)	d	Secor *et al.* (1995)
Long snout seahorse (*Hippocampus reidi*)	0.45	1–4	Suarez-Bregua *et al.* (2021)	e	n.a.
Mary River cod (*Maccullochella mariensis*)*	0.83	80–92	Mayne *et al.* (2021*a*)	e	Høie *et al.* (2004)
Murray cod (*Maccullochella peelii*)*	0.83	75–114	Mayne *et al.* (2021*a*)	e	Høie *et al.* (2004)
Zebrafish (*Danio rerio*) – model organism	1.80	1–4	Mayne *et al.* (2020*a*)	e	Higgs *et al.* (2002)
(ii) Amphibians:					
African clawed frog (*Xenopus laevis*)*	3.10	5–15	Zoller *et al.* (2022)	d	n.a.
Western clawed frog (*Xenopus tropicalis*)	3.11	3–12	Morselli *et al.* (2023)	a	n.a.
Western clawed frog (*Xenopus tropicalis*)*	3.11	3–12	Zoller *et al.* (2022)	d	n.a.
(iii) Reptiles:					
American alligator (*Alligator mississippiensis*)*	2.49	30–50	Parrott *et al.* (2014)	c	n.a.
American alligator (*Alligator mississippiensis*)*	2.49	30–50	Nilsen *et al.* (2016)	c	n.a.
Australian flatback sea turtle (*Natator depressus*)	2.64	60–100	Mayne *et al.* (2020*b*)	f	n.a.
Green sea turtle (*Chelonia mydras*)*	2.64	70–80	Mayne *et al.* (2022)	e	n.a.
Leatherback sea turtle (*Dermochelys coriacea*)	2.70	45–50	Mayne *et al.* (2020*b*)	f	n.a.
Loggerhead sea turtle (*Caretta caretta*)	2.70	47–67	Cocci *et al.* (2018)	b	n.a.
(iv) Birds:					
Barn swallow (*Hirundo rustica*)	1.31	1–4	Saino *et al.* (2019)	a	n.a.
Black grouse (*Lyrurus tetrix*)*	1.32	4–5	Soulsbury *et al.* (2018)	a	n.a.
Common tern (*Sterna hirundo*)*	1.44	9–10	Meyer *et al.* (2023)	a	n.a.
Short-tailed shearwater (*Ardenna tenuirostris*)	1.18	19–38	De Paoli-Iseppi *et al.* (2017*b*)	a	n.a.
Short-tailed shearwater (*Ardenna tenuirostris*)*	1.18	19–38	De Paoli-Iseppi *et al.* (2019)	a	n.a.
(v) Mammals:					
Aquatic:					
Antarctic minke whale (*Balaenoptera bonaerensis*)	3.29	100–114	Tanabe *et al.* (2020)	a	Hohn (1980)
Beluga whale (*Delphinapterus leucas*)*	3.29	35–50	Bors *et al.* (2021)	a	Goren *et al.* (1987)
Beluga whale (*Delphinapterus leucas*)	3.29	35–50	Robeck *et al.* (2021*b*)	d	Goren *et al.* (1987)
Bottlenose dolphin (*Tursiops truncatus*)*	3.03	15–16	Beal *et al.* (2019)	a	Kimura (1980)
Bottlenose dolphin (*Tursiops truncatus*)*	3.03	15–16	Robeck *et al.* (2021*a*)	d	Kimura (1980)
Bottlenose dolphin (*Tursiops truncatus*)*	3.03	15–16	Robeck *et al.* (2021*b*)	d	Kimura (1980)
Commerson's dolphin (*Cephalorhynchus commersonii*)*	3.27	10–18	Robeck *et al.* (2021*b*)	d	Myrick *et al.* (1983)
Common dolphin (*Delphinus delphis*)*	3.03	35–40	Robeck *et al.* (2021*b*)	d	Myrick *et al.* (1983)
Harbour porpoise (*Phocoena phocoena*)*	3.46	8–13	Robeck *et al.* (2021*b*)	d	Myrick *et al.* (1983)
Humpback whale (*Megaptera novaeangliae*)*	3.29	45–50	Horvath *et al.* (2022*c*)	d	Ohsumi *et al.* (1965)
Humpback whale (*Megaptera novaeangliae*)*	3.29	45–50	Polanowski *et al.* (2014)	a	Ohsumi *et al.* (1965)
Killer whale (*Orcinus orca*)*	3.18	50–80	Robeck *et al.* (2021*b*)	d	Clark *et al.* (2000)
Pacific white-sided dolphin (*Lagenorhynchus obliquidens*)*	3.46	40–45	Robeck *et al.* (2021*b*)	d	Kasuya & Matsui (1984)
Rough-tooth dolphin (*Steno bredanensis*)*	3.03	32–48	Robeck *et al.* (2021*b*)	d	Myrick *et al.* (1983)
Short-finned pilot whale (*Globicephala macrorhynchus*)*	3.29	45–60	Robeck *et al.* (2021*b*)	d	Kasuya & Matsui (1984)
Sperm whale (*Physeter catodon*)	3.29	50–70	Polanowski *et al.* (2014)	a	Ohsumi *et al.* (1965)
Bats:					
Bechstein's bat (*Myotis bechsteinii*)*	2.72	15–21	Wright *et al.* (2018)	a	Linhart (1973)
Big brown bat (*Eptesicus fuscus*)*	2.25	13–20	Wilkinson *et al.* (2021)	d	Linhart (1973)
Common noctule (*Nyctalus noctule*)*	2.26	5–12	Wilkinson *et al.* (2021)	d	Linhart (1973)
Common vampire bat (*Desmodus rotundus*)*	2.66	12–30	Wilkinson *et al.* (2021)	d	Linhart (1973)

(Continues on next page)

Table 1. *(Cont.)*

Species	Genome (bbp)	Lifespan (years)	Study	Method	Tooth annulation/otolith
Egyptian fruit bat (*Rousettus aegyptiacus*)*	2.11	10–22	Wilkinson *et al.* (2021)	d	Linhart (1973)
Fish-eating bat (*Myotis vivesi*)*	2.72	8–10	Wilkinson *et al.* (2021)	d	Linhart (1973)
Greater horseshoe bat (*Rhinolophus ferrumequinum*)*	2.68	20–30	Wilkinson *et al.* (2021)	d	Linhart (1973)
Greater mouse-eared bat (*Myotis myotis*)*	2.72	6–13	Wilkinson *et al.* (2021)	d	Linhart (1973)
Greater sac-winged bat (*Saccopteryx bilineata*)*	3.16	1–6	Wilkinson *et al.* (2021)	d	Linhart (1973)
Greater spear-nosed bat (*Phyllostomus hastatus*)*	2.62	10–18	Wilkinson *et al.* (2021)	d	Linhart (1973)
Grey-headed flying fox (*Pteropus poliocephalus*)*	2.32	2–5	Wilkinson *et al.* (2021)	d	Linhart (1973)
Indian flying fox (*Pteropus giganteus*)*	2.18	31–44	Wilkinson *et al.* (2021)	d	Linhart (1973)
Jamaican fruit bat (*Artibeus jamaicensis*)*	2.74	9–19	Wilkinson *et al.* (2021)	d	Linhart (1973)
Large flying fox (*Pteropus vampyrus*)*	2.37	15–30	Wilkinson *et al.* (2021)	d	Linhart (1973)
Lesser short-nosed fruit bat (*Cynopterus brachyotis*)*	2.27	20–30	Wilkinson *et al.* (2021)	d	Linhart (1973)
Little brown bat (*Myotis lucifugus*)*	2.72	7–31	Wilkinson *et al.* (2021)	d	Linhart (1973)
Little golden-mantled flying fox (*Pteropus pumilus*)*	2.32	10–17	Wilkinson *et al.* (2021)	d	Linhart (1973)
Mexican free-tailed bat (*Tadarida brasiliensis*)*	2.94	8–12	Wilkinson *et al.* (2021)	d	Linhart (1973)
Pale spear-nosed bat (*Phyllostomus discolor*)*	2.52	6–9	Wilkinson *et al.* (2021)	d	Linhart (1973)
Pallid bat (*Antrozous pallidus*)*	2.67	9–14	Wilkinson *et al.* (2021)	d	Brook *et al.* (2019)
Proboscis bat (*Rhynchonycteris naso*)*	2.72	1–6	Wilkinson *et al.* (2021)	d	Linhart (1973)
Rodrigues flying fox (*Pteropus rodricensis*)*	2.23	28–30	Wilkinson *et al.* (2021)	d	Linhart (1973)
Seba's short-tailed bat (*Carollia perspicillata*)*	2.92	10–17	Wilkinson *et al.* (2021)	d	Linhart (1973)
Small flying fox (*Pteropus hypomelanus*)*	2.33	9–20	Wilkinson *et al.* (2021)	d	Linhart (1973)
Southern long-nosed bat (*Leptonycteris curasoae*)*	2.36	10–30	Wilkinson *et al.* (2021)	d	Linhart (1973)
Straw-coloured fruit bat (*Eidolon helvum*)*	2.03	15–22	Wilkinson *et al.* (2021)	d	Linhart (1973)
Velvety free-tailed bat (*Molossus molossus*)*	2.41	2–5	Wilkinson *et al.* (2021)	d	Linhart (1973)
Carnivores:					
Australian sea lion (*Neophoca cinerea*)*	3.15	17–25	Robeck *et al.* (2023)	d	Rust *et al.* (2019)
Bobcat (*Lynx rufus*)	2.92	7–13	Cantrell *et al.* (2020)	b	Crowe (1972)
California sea lion (*Zalophus californianus*)*	3.15	15–25	Robeck *et al.* (2023)	d	Rust *et al.* (2019)
Cheetah (*Acinonyx jubatus*)	2.56	10–15	Le Clercq *et al.* (2018)	e	Marker & Dickman (2003)
Cheetah (*Acinonyx jubatus*)*	2.56	10–15	Raj *et al.* (2021)	d	Marker & Dickman (2003)
Domestic cat (*Felis catus*)*	2.91	12–18	Qi *et al.* (2021)	e	Grue & Jensen (1979)
Domestic cat (*Felis catus*)*	2.91	12–18	Raj *et al.* (2021)	d	Grue & Jensen (1979)
Domestic dog (*Canis familiaris*)	2.80	10–13	Ito *et al.* (2017)	a	Grue & Jensen (1979)
Domestic dog (*Canis familiaris*)*	2.80	10–13	Thompson *et al.* (2017)	a	Grue & Jensen (1979)
Domestic dog (*Canis familiaris*)*	2.80	10–13	Horvath *et al.* (2022*e*)	d	Grue & Jensen (1979)
Fisher (*Pekania* [*Martes*] *pennanti*)	n.a.	10–14	Lachance *et al.* (2015)	b	Arthur *et al.* (1992)
Gray wolf (*Canis lupus*)*	2.81	14–16	Thompson *et al.* (2017)	a	Walleser *et al.* (2016)
Harbour seal (*Phoca vitulina*)*	2.94	25–35	Robeck *et al.* (2023)	d	Bowen *et al.* (1983)
Harp/Greenland seal (*Pagophilus groenlandicus*)*	2.94	20–30	Robeck *et al.* (2023)	d	Bowen *et al.* (1983)
Lion (*Panthera leo*)*	2.95	10–16	Raj *et al.* (2021)	d	White & Belant (2016)
Pacific walrus (*Odobenus rosmarus divergens*)*	2.40	20–40	Robeck *et al.* (2023)	d	Stewart *et al.* (1996)
Ringed seal (*Phoca hispida*)	2.94	15–28	Sareisian (2014)	b	Stewart *et al.* (1996)
Snow leopard (*Panthera unica*)	3.54	10–13	Qi *et al.* (2021)	e	McCarthy *et al.* (2005)
Steller sea lion (*Eumetopias jubatus*)*	3.15	20–30	Robeck *et al.* (2023)	d	Rust *et al.* (2019)

(Continues on next page)

Table 1. *(Cont.)*

Species	Genome (bbp)	Lifespan (years)	Study	Method	Tooth annulation/otolith
Tiger (*Panthera tigris*)*	2.71	8–10	Raj *et al.* (2021)	d	Sharma *et al.* (2021)
Elephants:					
African elephant (*Loxodonta africana*)*	4.44	60–70	Prado *et al.* (2021)	d	Spinage (1976)
Asian elephant (*Elephas maximus*)*	4.03	48–60	Prado *et al.* (2021)	d	Spinage (1976)
Marsupials:					
Eastern grey kangaroo (*Macropus giganteus*)	4.21	6–8	Horvath *et al.* (2022*d*)	d	Grue & Jensen (1979)
Gray short-tailed opossum (*Monodelphis domestica*)*	4.17	1–3	Horvath *et al.* (2022*d*)	d	Pekelharing (1970)
Red kangaroo (*Macropus rufus*)*	3.13	6–20	Horvath *et al.* (2022*d*)	d	Grue & Jensen (1979)
Red-necked wallaby (*Macropus rufogriseus*)	5.92	7–15	Horvath *et al.* (2022*d*)	d	Grue & Jensen (1979)
Tasmanian devil (*Sarcophilus harrisii*)*	3.63	5–8	Horvath *et al.* (2022*d*)	d	Pekelharing (1970)
Western grey kangaroo (*Macropus fuliginosus*)	3.91	10–29	Horvath *et al.* (2022*d*)	d	Grue & Jensen (1979)
Primates:					
Baboon (*Papio cynocephalus x P. anubis*)*	3.53	20–30	Anderson *et al.* (2021)	a	Wada *et al.* (1978)
Chimpanzee (*Pan troglodytes*)*	3.76	32–39	Guevara *et al.* (2020)	d	Wittwer-Backofen *et al.* (2004)
Chimpanzee (*Pan troglodytes*)*	3.76	32–39	Ito *et al.* (2018)	a	Wittwer-Backofen *et al.* (2004)
Common marmoset (*Callithrix jacchus*)*	3.43	7–16	Horvath *et al.* (2021*b*)	d	Scott *et al.* (1980)
Rhesus macaque (*Macaca mulatta*)*	3.59	25–40	Horvath *et al.* (2021*a*)	d	Newham *et al.* (2021)
Vervet monkey (*Chlorocebus pygerythrus*)*	4.19	15–30	Jasinska *et al.* (2022)	d	Scott *et al.* (1980)
Rodents:					
Mouse (*Mus musculus*)–Model organism	3.25	1–1.5	Spiers *et al.* (2016)	a	Grue & Jensen (1979)
Mouse (*Mus musculus*)–Model organism	3.25	1–1.5	Stubbs *et al.* (2017)	a	Grue & Jensen (1979)
Naked mole-rat (*Heterocephalus glaber*)*	2.90	22–31	Lowe *et al.* (2020)	a	Grue & Jensen (1979)
Naked mole-rat (*Heterocephalus glaber*)*	2.90	22–31	Horvath *et al.* (2022*a*)	d	Grue & Jensen (1979)
Naked mole-rat (*Heterocephalus glaber*) *	2.90	22–31	Kerepesi *et al.* (2022)	a	Grue & Jensen (1979)
Wood mouse (*Apodemus sylvaticus*) *	3.29	1–4	Little *et al.* (2020)	a	Grue & Jensen (1979)
Shrews:					
Masked shrew (*Sorex cinereus*)*	2.91	1–2	Cossette *et al.* (2022)	d	Grue & Jensen (1979)
Ungulates:					
Cattle (*Bos taurus*)*	3.70	10–20	Kordowitzki *et al.* (2021)	d	Grue & Jensen (1979)
Common zebra (*Equus quagga*)*	3.15	20–25	Larison *et al.* (2021*a*)	d	Penzhorn (1982)
Domestic horse (*Equus caballus*)*	3.15	25–30	Larison *et al.* (2021*a*)	d	Penzhorn (1982)
Domestic horse (*Equus caballus*)*	3.15	25–30	Horvath *et al.* (2022*b*)	d	Penzhorn (1982)
Domestic pig (*Sus scrofa domesticus*)*	3.00	15–20	Schachtschneider *et al.* (2021)	d	Grue & Jensen (1979)
Domestic sheep (*Ovis aries*)*	3.30	10–12	Sugrue *et al.* (2021)	d	Saxon & Higham (1968)
Grévy's zebra (*Equus grevyi*)*	3.15	20–25	Larison *et al.* (2021*a*)	d	Penzhorn (1982)
Roe deer (*Capreolus capreolus*)*	3.41	7–10	Lemaître *et al.* (2022)	d	Grue & Jensen (1979)
Somali wild ass (*Equus africanus somaliensis*)*	4.12	27–40	Larison *et al.* (2021*a*)	d	Penzhorn (1982)

clock for age determination that would be transferable across several species of marine turtles (Mayne *et al.*, 2022). This model, based on 17 CpGs, was able to predict age with an error rate of 3–4 years among adult individuals. The method was developed also to include sites that could be assayed in hawksbill sea turtle [*Eretmochelys imbricata* (Linnaeus)], loggerhead sea turtle [*Caretta caretta* (Linnaeus)], and olive Ridley sea turtle [*Lepidochelys olivacea* (Eschscholtz)]; in addition to previously studied flatback and leatherback turtles. Ecotoxicology studies on methylation in loggerhead sea turtles found similar patterns in global methylation as compared to studies conducted in alligators (Cocci *et al.*, 2018).

(iv) Birds (class Aves). Similar to reptiles, many of the early methylation studies in birds had a focus on ecotoxicology (Brandenburg, 2016; Romano *et al.*, 2017) or other aspects of animal ecology. For example, several studies have focussed on the association between methylation levels and migration or breeding phenology (Saino *et al.*, 2017; Liebl *et al.*, 2021), temporal variation on an annual basis in both methylation levels (Viitaniemi *et al.*, 2019; Mäkinen *et al.*, 2019) and the expression of key regulatory elements within the methylation circuitry (Sharma *et al.*, 2018; Trivedi *et al.*, 2019). While these studies were not focussed on age estimation, they did establish the heritability of methylation (Romano *et al.*, 2017) and inter-generational resemblance of methylation patterns (Saino *et al.*, 2019). Studies aimed at exploiting methylation to model age found that 67 CpGs in 13 genes, previously used in mammals, were not well conserved in a long-lived seabird species, the short-tailed shearwater [*Ardenna tenuirostris* (Temminck)], with only 12 CpGs showing a weakly correlated relationship with age (De Paoli-Iseppi *et al.*, 2017*a*). The same authors were, however, able to construct an age clock for this species with seven newly developed CpGs that predicted age with a mean error of 2–3 years (De Paoli-Iseppi *et al.*, 2019). A similar trend of age–dependent changes was later observed in the black grouse [*Lyrurus tetrix* (Linnaeus)] that found four CpGs within one gene to be highly correlated with age (Soulsbury *et al.*, 2018). A recent study in the common tern [*Sterna hirundo* (Linnaeus)], identified several genomic regions that showed shifts in methylation equivalent to epigenetic clocks that changed predictably over time within individuals in a sex-specific manner mostly affecting females (Meyer *et al.*, 2023).

(v) Mammals (class Mammalia)

(*i*.1) *Aquatic mammals (infraorder Cetacea).* The first study to develop an age-estimation model in wildlife based on methylation was conducted on humpback whales and assayed 37 CpGs in eight genes, of which seven CpGs had a significant linear correlation with age (Polanowski *et al.*, 2014). This allowed the creation of a model able to estimate the age of several whale species with a standard error of prediction of ∼3 years. A similar study on humpbacks was reported more recently (Horvath *et al.*, 2022*c*) and used a conserved site mammalian methylation array which was able to reduce the MAE to 2 years. Several similar models have been developed in other whale species as well. Two of the genes assayed by Polanowski *et al.* (2014) were assessed in the Antarctic minke whale [*Balaenoptera bonaerensis* (Burmeister)] in 17 CpGs (Tanabe *et al.*, 2020). The authors, however, only found a significant relationship between methylation and age in one gene and concluded that the methylation–age relationship identified in one species may not be conserved between species. An analysis of 37,491 CpGs using a conserved site mammalian methylation array (Arneson *et al.*, 2022) applied to Beluga whale [*Delphinapterus leucas* (Pallas)] found 23 highly age-correlated CpGs, of which 15 were in genes that enabled the development of an age-estimation model able to determine age with comparable accuracy to the model developed in humpback whales (Bors *et al.*, 2021). Similar methods were later applied to bottlenose dolphins [*Tursiops truncatus* (Montagu)] in a study that replicated previous findings in humpback whales by assaying 17 CpGs in three genes on skin biopsies (Beal *et al.*, 2019). The authors found two of the genes to have comparable methylation levels between species while the third was significantly hypomethylated. Two of the assayed CpGs had significant age-correlated methylation which enabled the construction of a linear ageing model that could accurately predict the age of bottlenose dolphins with a standard error of ∼4 years (Beal *et al.*, 2019). Another study on the same species was able to construct a combined age-estimation model from skin and blood samples with an error in prediction of ∼2–3 years (Robeck *et al.*, 2021*a*).

These studies culminated in the recent validation of a methylation array chip which assayed between 79 and 142 CpGs, depending on tissue type, in nine species of cetaceans to construct a single age-estimation model that is transferable across species (Robeck *et al.*, 2021*b*). This study included four additional species of dolphins, two additional species of whales, and porpoises. The dolphins included the common dolphin [*Delphinus delphis* (Linnaeus)], Commerson's dolphin [*Cephalorhynchus commersonii* (Lacépède)], Pacific white-sided dolphin [*Lagenorhynchus obliquidens* (Gill)], and rough-tooth dolphin [*Steno bredanensis* (Cuvier)], as well as the harbour porpoise [*Phocoena phocoena* (Linnaeus)]. The whales included the killer whale [*Orcinus orca* (Linnaeus)] and short-finned pilot whale [*Globicephala macrorhynchus* (Gray)]. The final model generated from this study was able to predict age accurately in all tested species with a standard error of between 3 and 4 years, depending on tissue type, similar to findings in individual, species-specific studies.

(*i*.2) *Bats (order Chiroptera).* Among the only true flying mammals in the order Chiroptera, which includes bats, megabats, and microbats, two studies have aimed at developing methylation-based clocks for age (Wright *et al.*, 2018; Wilkinson *et al.*, 2021). The first used methylation levels at seven CpGs in three genes to construct an age clock from wing skin samples from Bechstein's bat [*Myotis bechsteinii* (Kuhl)] able to determine age with a MEP of 2–3 years (Wright *et al.*, 2018). A pan-species model was later developed using 162 CpGs from a conserved mammalian methylation array able to predict the age of 26 species within the order Chiroptera with a MEP of less than a year (Wilkinson *et al.*, 2021).

(*i*.3) *Carnivores (order Carnivora).* In carnivores, age clocks were assessed for canids (Ito *et al.*, 2017; Thompson

et al., 2017, Horvath *et al.*, 2022*e*), felines (Cantrell *et al.*, 2020; Raj *et al.*, 2021; Qi *et al.*, 2021), mustelids (Lachance *et al.*, 2015) and pinnipeds (Sareisian, 2014; Robeck *et al.*, 2023). The first model in the domestic dog, *Canis familiaris* (Linnaeus) assayed CpGs in six genes and constructed a clock from the four highly age-correlated CpGs, able to predict age with a MEP of 34–37 months (Ito *et al.*, 2017). The number of breeds included was not stated but likely less than 50 (based on sample size) and the authors postulated that the error in clocks could be improved by screening for age-correlated CpGs to assay prior to modelling age using methylation. The second study (Thompson *et al.*, 2017), which also included both domestic dog and gray wolf [*Canis lupus* (Linnaeus)], was able to construct age clocks for each species from 41 and 67 CpGs respectively with a MEP of ∼10 months. This study also used fewer than 50 breeds but represented a significant improvement in the accuracy of the prediction. Furthermore, a relationship between size (mass) and accelerated epigenetic age was detected in line with the knowledge that smaller breeds live longer while larger breeds have shorter lifespans. A follow-up study in domestic dog by some of the same co-authors (Horvath *et al.*, 2022*e*) was able further to improve the epigenetic clock (MEP of 5–6 months) and predict age-to-death by assaying CpGs in 93 dog breeds.

The first study in felines (Qi *et al.*, 2021) studied methylation in multiple CpGs in two genes, across several breeds of domestic cat [*Felis catus* (Linnaeus)]. Significant correlates with age were detected using high-resolution melting analyses combined with MS-PCR for both genes and a combined ageing model with a MEP of 3.5–4 years was constructed for both healthy and chronically ill individuals. The authors also tested this method in a smaller sample of snow leopard [*Panthera uncia* (Schreber)] and found the trends observed in domestic cat to be sufficiently conserved to determine age in this species with an improved MEP of ∼2 years. The second feline study (Raj *et al.*, 2021), also in domestic cat, used a methylation array to identify 34 CpGs that have a conserved correlation with age and created a model with a MEP of ∼10 months, similar to that observed in canids. The authors further tested the model on smaller sample subsets ($N < 15$) of wild felines including cheetah, lion [*Panthera leo* (Linnaeus)], and tiger [*Panthera tigris* (Linnaeus)]. The cat clock was able to determine age in the related wild species with a MEP of 1–3 years.

Global methylation has also been studied in pinnipeds in relation to age in the ringed seal [*Phoca hispida* (Schreber)], using liver and kidney tissue, however, no clear difference was observed between the sub-adult and adult groups – possibly due to small sample sizes or tissue type (Sareisian, 2014). Similar studies were performed measuring global methylation using enzyme-linked immunosorbent assay (ELISA) in bobcats [*Lynx rufus* (Schreber)], and fishers [*Pekania pennanti* (Erxleben)], using liver samples (Lachance *et al.*, 2015; Cantrell *et al.*, 2020). While both identified the same trend of a decrease in methylation with age, supporting the possibility that these species may be well suited to the development of epigenetic clocks for age determination, neither was able to detect a statistically significant difference between young and mature individuals. This may indicate that site-directed methylation detection methods, such as bisulfite sequencing, are more useful for this purpose than ELISA or MS-PCR (Le Clercq, Dalton & Kotzé, 2018). This was illustrated by a later study in pinnipeds (Robeck *et al.*, 2023) that included six species: three species of sea lion, the Australian sea lion [*Neophoca cinerea* (Péron)], California sea lion [*Zalophus californianus* (Lesson)], and Steller sea lion [*Eumetopias jubatus* (Schreber)]; two seal species, harbour seal [*Phoca vitulina* (Linnaeus)], and harp/Greenland seal [*Pagophilus groenlandicus* (Erxleben)]; as well as one species of walrus, the Pacific walrus [*Odobenus rosmarus divergens* (Linnaeus)]. This study did find several age-correlated changes in methylation in pinnipeds that enabled the construction of an epigenetic clock, from combined skin and blood samples, able to predict age with a MEP of 1–1.5 years.

(*i*.4) *Elephants (order Proboscidea, suborder Elephantiformes)*. Methylation has been used to model age in both extant genera of elephants in the family Elephantidae: the African elephant [*Loxodonta africana* (Blumenbach)] and the Asian elephant [*Elephas maximus* (Linnaeus)]. Prado *et al.* (2021) created three epigenetic models of age, one for each of the genera as well as one general model for all elephants. For the African elephant, 2341 CpGs from a methylation array were found to be significantly correlated with age while the Asian elephant only had approximately a quarter as many. Nevertheless ∼366 of these age-correlated CpGs were shared by both species. The most accurate clock was reconstructed for the African elephant with a MEP of 2–3 years while the clock for Asian elephants had a MEP of 3–4 years. The combined clock for both species performed similarly to the latter, with a MEP of 3–4 years. Overall, this study illustrated that the accuracy of clocks may be improved by a higher number of targets rather than just increased sampling, while also highlighting that substantial differences may exist in age-related CpGs between closely related species.

(*i*.5) *Marsupials (infraorder Marsupialia)*. The first study to assay methylation in relation to age in marsupials (Horvath *et al.*, 2022*d*) was conducted in gray short-tailed opossum [*Monodelphis domestica* (Wagner)], a South American species with a pseudo-pouch, similar to those of male thylacines [*Thylacinus cynocephalus* (Harris)]. This study used a universal mammal array to develop a clock able to predict age with a MEP of 3–4 months. This model was subsequently validated in several other marsupials including the Tasmanian devil [*Sarcophilus harrisii* (Boitard)], eastern grey kangaroo [*Macropus giganteus* (Shaw)], western grey kangaroo [*Macropus fuliginosus* (Desmarest)], red kangaroo [*Macropus rufus* (Desmarest)], and red-necked wallaby [*Macropus rufogriseus* (Desmarest)] from the Australasian region. Despite the taxonomic and phylogenetic distance in the relatedness of these species, the model developed for opossums was well conserved in other marsupials and was able to predict the age of these species with a MEP of 12–13 and 11–13 months for Tasmanian devils and kangaroos/wallabies respectively.

(*i*.6) *Primates (order Primates)*. Epigenetic clocks for age have been studied extensively in humans, *Homo sapiens sapiens* (Linnaeus) (Jarman *et al*., 2015; De Paoli-Iseppi *et al*., 2017*a*). Such studies were expanded into non-human primates including hybrid baboons (*Papio cynocephalus* × *P. anubis*) (Anderson *et al*., 2021), chimpanzees [*Pan troglodytes* (Blumenbach)] (Ito *et al*., 2018; Guevara *et al*., 2020), rhesus macaques [*Macaca mulatta* (Zimmermann)] (Horvath *et al*., 2021*a*), and monkeys (Horvath *et al*., 2021*b*; Jasinska *et al*., 2022). Ito *et al*. (2018) developed the first primate age clock in chimpanzees based on methylation differences in several CpGs analysed in three genes. The most accurate model used two CpGs and was able to predict age with a mean error of prediction of 5–6 years. Improved results were later achieved using a human methylation array (Guevara *et al*., 2020) owing to the high amounts of largely conserved DNA between humans and chimps. This study was able to construct a highly accurate age clock from as little as 10 CpGs with an error rate comparable to that observed in humans (2–3 years). A study conducted on a hybridised baboon population, between congeneric yellow baboon [*Papio cynocephalus* (Linnaeus)], and Anubis baboon [*Papio anubis* (Lesson)], used genome-wide methylation analysis refined to ∼573 CpG sites to create a model with predictive precision of 1–2 years.

In the rhesus macaque, three clocks were constructed using blood, skin, and multi-tissue methylation data (Horvath *et al*., 2021*a*). The most accurate models were derived from blood and multi-tissue data with a MEP of between 12 and 18 months, while the model from skin samples had a MEP of ∼2 years. The pan-tissue models in macaques were based on 71 CpGs from a custom mammalian methylation array. These clocks were also tested on vervet monkeys [*Chlorocebus pygerythrus* (Cuvier)] and were able to predict age with a MEP of 2–3 years using blood or multi-tissue models, and ∼8 years using skin. More detailed analyses of this species were published in a subsequent companion article (Jasinska *et al*., 2022) where clocks were constructed using blood, liver, and brain tissue. Overall performance was improved using species-specific data, with liver and cortex models found to be the most accurate with a MEP of 3–6 months while the blood-derived clock had a MEP of approximately 1 year. A similar study in common marmosets [*Callithrix jacchus* (Linnaeus)] constructed an age clock from blood samples from over 1000 age-related CpGs able to estimate age with a MEP of 8–9 months (Horvath *et al*., 2021*b*). The studies on macaques as well as vervet and marmoset monkeys also compared the accuracy of these models on human data and between species. They found most models to retain a relatively high predictive power across multiple primates and tissue types, despite the ∼12 million years divergence distance between macaques and vervets and 29 million years between humans and vervets (Horvath *et al*., 2021*a*; Jasinska *et al*., 2022).

(*i*.7) *Rodents (order Rodentia)*. Laboratory murine mouse models such as *Mus musculus* (Linnaeus) were used to construct age clocks in multiple tissue types (Stubbs *et al*., 2017). This was later replicated in two wild rodent species, the wood mouse [*Apodemus sylvaticus* (Linnaeus)], and naked mole-rat [*Heterocephalus glaber* (Rüppell)]. In wood mouse (Little *et al*., 2020), ear punches from a laboratory-reared and formerly wild population (Clerc *et al*., 2019) were used to develop an age clock based on 9 CpGs in four genes. The resulting clock provided the first evidence with a high accuracy for the prediction of age in wild mice with a MEP of 59 days. For the naked mole-rat, 23 CpGs that are age correlated were used to construct an initial age clock using liver tissue (Lowe *et al*., 2020). To enable non-lethal age determination of wild individuals this clock was validated using skin samples, with a resulting model able to predict age with a MEP of 3–4 years. The model was found to have slightly lower accuracy between skin and liver samples, indicating that skin may age slower. More recent studies in the same species used reduced-representation bisulfite sequencing (Kerepesi *et al*., 2022) and a universal mammalian methylation array (Horvath *et al*., 2022*a*) to construct more accurate models. The first was able to construct an age-prediction model from blood samples that could predict age with a MEP of 1.1 years (Kerepesi *et al*., 2022) while the second constructed a pan-tissue model able to predict age with a MEP of less than 1 year (Horvath *et al*., 2022*a*).

(*i*.8) *Shrews (order Eulipotyphla)*. To date, one study has explored epigenetic ageing in shrews (Cossette *et al*., 2022), an order closely related to Carnivora, Chiroptera, and Pholidota (pangolins), in the superorder Laurasiatheria. Using a conserved mammalian methylation array, this study explored both the utility of methylation in estimating the age of masked shrew, *Sorex cinereus* (Kerr), as well as the possible divergence in epigenetic ageing between two highly differentiated mainland and insular island populations. The epigenetic clock, constructed from 26 age-correlated CpGs, was able to predict age with a MEP of 0.1 years, however, some evidence was found for the accelerated ageing of island populations. Further epigenome-wide association studies also identified several genes without significant age associations that have differential methylation and possibly expression patterns that may contribute to the morphological differences between mainland and island populations.

(*i*.9) *Ungulates (orders Artiodactyla and Perissodactyla)*. In hooved mammals, distant relatives of whales and other aquatic mammals, age clocks were established in several species of both even-toed (Artiodactyla) and odd-toed (Perissodactyla) ungulates. In odd-toed ungulates, studies included wild and domesticated species of the family Equidae. Epigenetic clocks were modelled in common zebra [*Equus quagga* (Boddaert)] using 70 CpG sites for blood and 99 CpGs for blood and skin pooled samples (Larison *et al*., 2021*a*). The clock developed using blood samples outperformed the skin-based model and was able to predict age with a MEP of 6–7 months, as compared to a MEP of 1–2 years. The blood model was tested further in other species in the genus *Equus*, including domestic horses [*E. caballus* (Linnaeus)], Grévy's zebras [*E. grevyi* (Oustalet)], and Somali wild asses [*E. africanus somaliensis* (Noack)], and achieved results comparable to the skin-based model. A second study on domestic horses created three epigenetic clocks based

on data derived from blood, liver, and pan-tissue sampling (Horvath *et al.*, 2022*b*). The most accurate model was derived from liver tissue with a MEP of 7–8 months while both the blood and pan-tissue models had a MEP of ∼1 year.

Studies conducted on even-toed ungulates included species from the families Bovidae (Kordowitzki *et al.*, 2021; Sugrue *et al.*, 2021), Cervidae (Lemaître *et al.*, 2022), and Suidae (Schachtschneider *et al.*, 2021). A study in roe deer [*Capreolus capreolus* (Linnaeus)], from wild populations in Eastern France, used the same methods as for equines to create sex-specific clocks for age determination with a MEP of 6–8 months (Lemaître *et al.*, 2022). Interestingly, this study also identified acceleration methylation ageing in males and several CpGs that showed differential methylation patterns based on sex. Due to the basal genetic similarities between this species and white-tailed deer [*Odocoileus virginianus* (Zimmermann)], this method may also be transferable to other deer species. Among domesticated species, epigenetic clocks were constructed for cattle [*Bos taurus* (Linnaeus)], domestic sheep [*Ovis aries* (Linnaeus)], and domestic pig [*Sus scrofa domesticus* (Linnaeus)]. Epigenetic clocks constructed for cattle and domestic sheep were able to estimate age with a MEP of 8–9 and 3–5 months, respectively (Kordowitzki *et al.*, 2021; Sugrue *et al.*, 2021), based on between 80 and 140 CpGs with conserved age-related methylation patterns across tissue types. For domestic pig an age-estimation clock was constructed from more than 900 CpGs with a MEP of 1–2 months (Schachtschneider *et al.*, 2021).

(2) Telomeres

(*a*) *What are telomeres?*

Early work in cytogenetics identified knobs of heterochromatin, visualised at the end of linear chromosomes (Fig. 5), which could be considered the natural ends of chromosomes (McClintock, 1941). These ends would later be called telomeres, from the Greek words for 'end' (telos) and 'part' (meros). These studies would pave the way towards the discovery of basic telomere structure, several decades later, and the advancement of science and technology towards the further characterization of their molecular composition and biology (Shay & Wright, 2019).

The telomeres are formed by a tandem repeat of a short nucleotide sequence. In humans, this sequence was identified to be a repeat of 5′-TTAGGG-3′ (Moyzis *et al.*, 1988) which is conserved in most eukaryotes, including all mammals (Gomes *et al.*, 2011). The full length of the telomere sequence varies; from 10 to 15 kilobases (kb) in humans (de Lange *et al.*, 1990) to 20–50 kb in laboratory mice (Kipling & Cooke, 1990). The tandem repeat sequence is partially double stranded (Fig. 6), ending with a free 3′ overhang which protrudes past the 5′ strand by ∼50–500 nucleotides (Wu, Takai & De Lange, 2012). It was initially unclear how the 3′ strand is formed (Palm & de Lange, 2008), as telomerase had already been excluded (Yuan *et al.*, 1999), but further studies elucidated the role of several nucleases (Wu *et al.*, 2012) including Exonuclease I. The 3′ strand, containing the triple G of the repeat is rich in guanosine and thus termed the G-strand, whilst the 5′ strand is termed the C-strand. The 5′-end is well defined and mostly ends with the ATC-5′ codon, in contrast to the variable 3′-end (Sfeir *et al.*, 2005). The telomere sequence folds in upon itself to form a laureate structure (Fig. 6) as a higher order chromatin configuration consisting of two distinct loops, the double-stranded DNA telomere loop (T-loop) and the single-stranded displacement loop (D-loop; Cleal, Norris & Baird, 2018). The D-loop is formed when the 3′ overhang of the free end invades and displaces a 50–300 nucleotide region of the telomeric sequence to close the loop, thereby protecting the complex from recognition by DNA repair

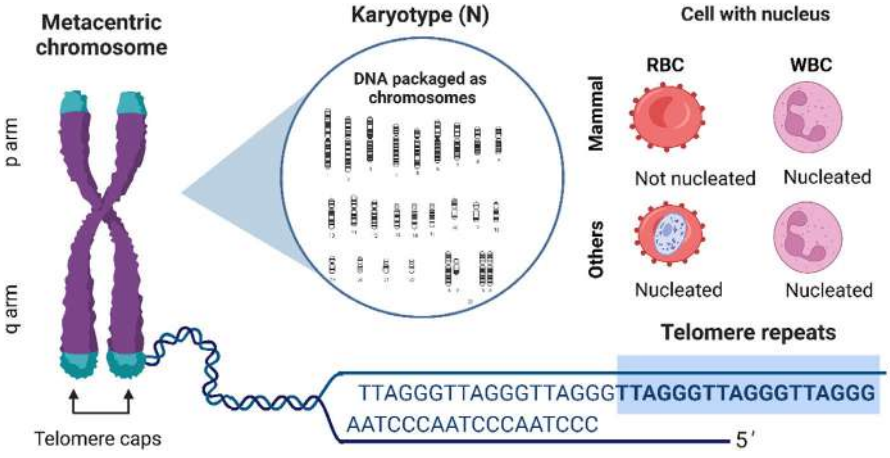


Fig. 5. Picture of metacentric chromosomes showing the telomere caps at either end. On the right the typical mammalian cell is shown with the nucleus highlighted. The nucleus contains the genomic DNA which is packaged in densely coiled structures called chromosomes for which the karyotype (chromosomal organisation) is illustrated in the central circle. The ends of the chromosomes (blue) are capped with telomeres – non-coding repeats of the same 6–8 base pair sequence – to protect the coding genomic DNA (purple) from damage or loss during replications. Cells from the tissue most commonly used in telomere length studies are also indicated showing the nucleated red blood cells (RBCs) found in non-mammalian species as well as the nucleated white blood cells (WBCs).

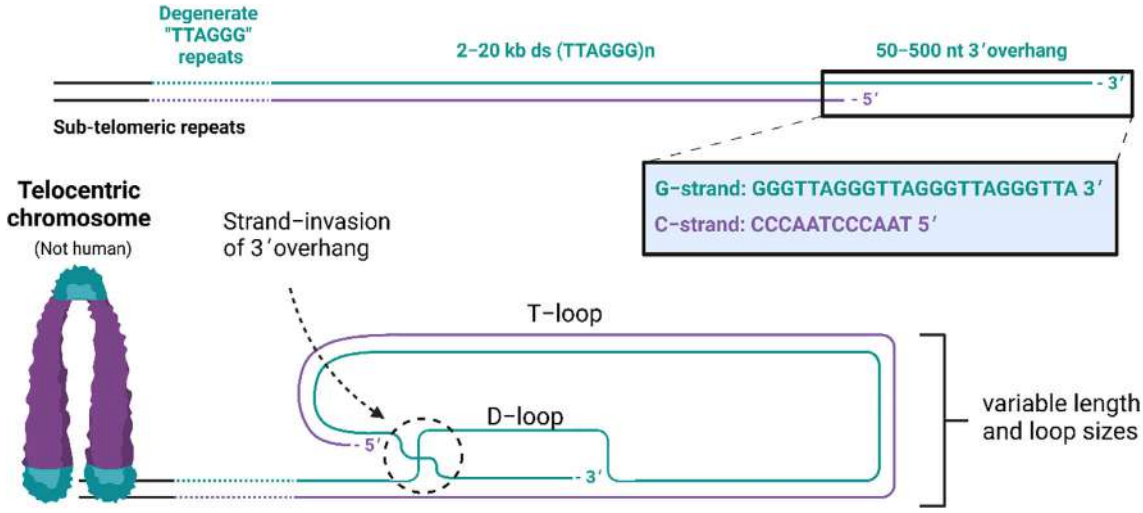


Fig. 6. Segments and higher chromatin structure of telomeric DNA. The DNA strand at the distal end of a chromosome ends is formed by sub-telomeric repeats, followed by degenerate tandem repeats, adjacent to the double stranded (ds) telomeric tandem repeats of 5′-TTAGGG-3′ that span ∼2–20 kilobase pairs (kb). The terminal ends of the telomere DNA comprise the longer, guanine rich, G-strand that forms the 3′ overhang of ∼50–500 nucleotides (nt), as well as the complementary C-strand of the 5′ terminal DNA. The terminal ends of the telomere strand bend back to form a hairpin-like loop, known as the telomere loop (T-loop), where the 3′ overhang inserts itself between a double-stranded segment forming the displacement or D-loop. The structure of telocentric chromosomes found in non-human and non-mammalian species is also illustrated. Image created in BioRender.com.

machinery as single strand breaks (Oganesian & Karlseder, 2011; Kazda *et al.*, 2012; Wu *et al.*, 2012).

Telomere length is maintained through the activity of telomerase (Fig. 7), a reverse transcriptase that elongates telomeres (Greider & Blackburn, 1985), as well as by the alternative lengthening of telomeres (ALT) pathway (Cesare & Reddel, 2010), which relies upon homologous recombination between telomeres. Telomerase is a ribonucleoprotein complex with RNA subunits that elongate chromosomes by adding telomeric repeat sequences to the terminal ends of chromosomes. This complex comprises two catalytic subunits, telomerase reverse transcriptase (TERT) and telomerase RNA component (TERC), which serve as the template to synthesise telomeric repeat sequences, as well as a stabilising protein complex of Dyskerin protein and nucleolar protein family A member proteins (Greider & Blackburn, 1985).

(*b*) *Changes with age*

Similar to methylation, telomeres undergo several phases of alteration throughout gametogenesis, following zygote formation, during embryogenesis, and throughout the lifespan of an organism in different tissues (Fig. 8). For example, telomeres in oocytes are typically shorter than in somatic tissue and undergo activation following zygote formation to elongate throughout the two- and four-cell phases until blastocyst formation (Zhao *et al.*, 2014). This initially happens independent of telomerase, which has been found to be absent from spermatozoa, oocytes and morula stage cells (Wright *et al.*, 1996), followed by an upregulation of telomerase during the morula–blastocyst phase (Schaetzlein *et al.*, 2004). Elongation in the absence of telomerase is likely due to extensive genome-wide demethylation providing a favourable milieu for telomere lengthening via an ALT mechanism (Liu *et al.*, 2007). After birth, telomerase activity is restricted to germline cells and stem cells.

As DNA polymerase would not be able to replicate the segment of DNA to which it binds, telomeres serve as barriers protecting coding sequences located towards the end of chromosomes. In the absence of telomerase, however, telomeres are also subject to the gradual loss of sequence due to the terminal replication problem, and telomere length has since become a well-studied attribute of cellular ageing (Olsson, Wapstra & Friesen, 2018). The terminal replication problem is attributed to DNA polymerase requiring an RNA primer to initiate replication. During replication the RNA primer, which binds to the terminal nucleotides of a replicating strand, dissociates from the newly synthesised strand without being incorporated resulting in a daughter strand that is shorter than the template strand. Consequently, telomere length is typically lost at a rate of 30–50 bp per cell division and ultimately leads to cellular senescence. Furthermore, telomeres have been shown to be highly susceptible to oxidative damage due to reactive oxygen species (ROS), for example because of high stress levels, persistent inflammation, or in response to environmental toxins (Zhao *et al.*, 2014).

(*c*) *Telomeres as a biomarker of animal age*

Following the observations of telomere length reflecting age and cellular senescence in humans there has been increased

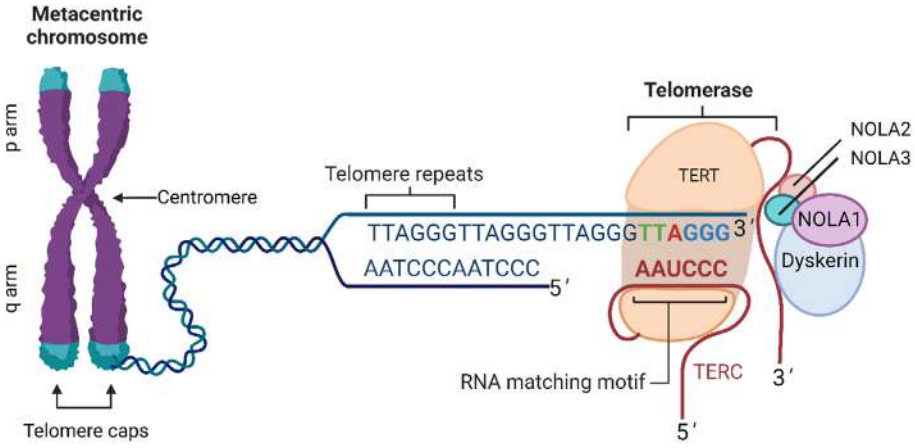


Fig. 7. Diagrammatic representation of the telomerase complex binding to the single–stranded loop of telomeres for elongation. The telomerase complex has a telomerase RNA component (TERC) that contains an RNA matching motif to incorporate nucleotides of the telomeric repeat sequence to the strand being elongated in the 5′ to 3′ direction. The telomerase reverse transcriptase (TERT) element enzymatically catalyses the addition of the newly synthesised telomeric repeat to the existing strand. Other elements of the telomerase complex are also indicated, including small nucleolar ribonucleoproteins Dyskerin, and the three nucleolar protein family A member proteins (NOLA1, NOLA2, and NOLA3). Image created in BioRender.com.

interest in its ability to determine age in animals (Haussmann & Vleck, 2002; Horn, Robertson & Gemmell, 2010) and such correlations have been illustrated in several non-human species including birds (Tricola *et al.*, 2018; Vedder *et al.*, 2022), aquatic mammals (Garde *et al.*, 2010; Izzo *et al.*, 2011), and reptiles (Hatase *et al.*, 2008). Telomere length has been assayed using several methods (Mensà *et al.*, 2019) ranging from terminal restriction fragment (TRF) analyses followed by southern blotting to fluorescence-based methods that include *in situ* hybridization (FISH), flow cytometry coupled FISH (Flow-FISH), and quantitative real time polymerase chain reaction (qPCR). Animal studies that assayed telomere length, included in this review and meta-analysis, are summarised in Table 2. As telomeres and age have been the subject of several reviews (Dunshea *et al.*, 2011; Wilbourn *et al.*, 2018; Remot *et al.*, 2021) this section will highlight a few key studies across

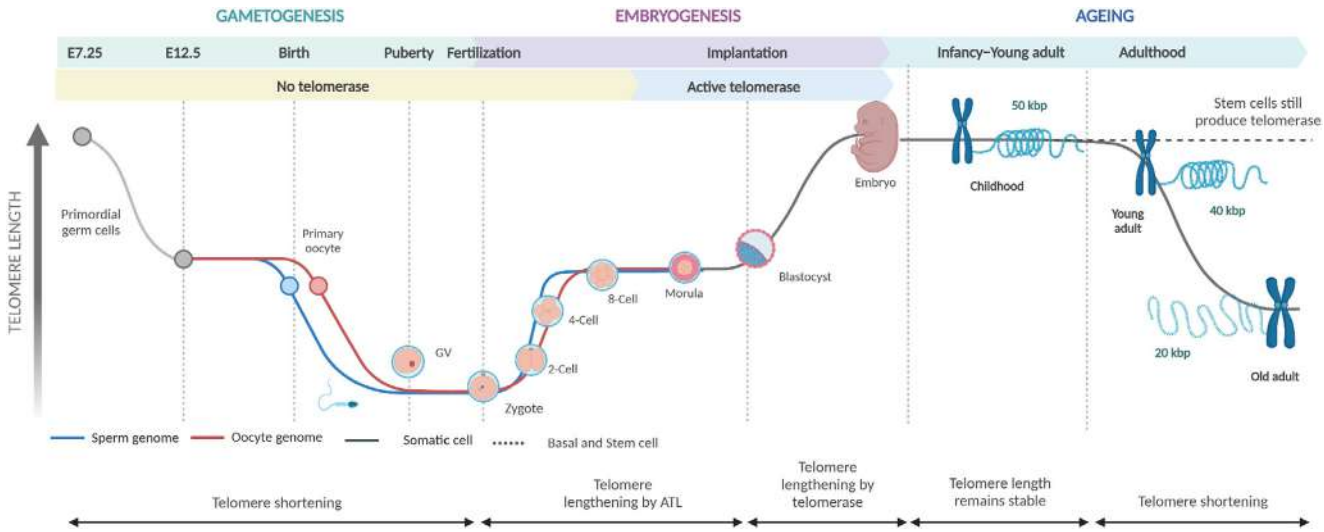


Fig. 8. Diagrammatic illustration of the alterations to telomere length during key life stages from gametogenesis to embryogenesis and the resulting changes that occur during natural ageing in most mammals. During gametogenesis the telomeres shorten in primordial germ cells and gametes, either sperm or egg cells, due to the absence of telomerase. After a zygote is formed from the union of two gametes, passive genomic demethylation happens which aids in telomere lengthening by enabling alternative telomere lengthening (ALT). Between the morula and blastocyst stage, telomerase expression is reactivated, and active telomere lengthening occurs prior to implantation. After birth, telomere length remains stable in most tissues throughout childhood from infancy to early adulthood. During the ageing process telomere length is maintained in stem cells while telomere shortening happens throughout the remainder of somatic cells. Image created in BioRender.com.

Table 2. Summary of studies that assayed animal telomere length in relation to age, the environment, or transgenerational heredity. Studies that used the data to create a molecular model for animal age, which were included in the meta-analysis, are indicated with an asterisk (*) for across-class analyses and a double asterisk (**) for the birds-only analyses. Species are grouped by class and listed alphabetically. For each species the karyotype of diploid chromosomes is given along with estimates for the average lifespan. For each study, the method of measuring telomere length is indicated: a, fluorescent *in situ* hybridization–flow cytometry (Flow-FISH); b, terminal restriction fragment (TRF) length & Southern blot; c, quantitative real-time polymerase chain reaction (qPCR); d, quantitative-fluorescent *in situ* hybridization (Q-FISH). Where relevant, citations are also provided for studies on age determination using other biological clocks (tooth annulation/otolith) in the same/similar species. n.a., not available.

Species	Karyotype	Lifespan (years)	Study	Method	Tooth annulation/otolith
(i) Fishes:					
Common carp (*Cyprinus carpio*)*	100	11–13	Izzo *et al.* (2014)	c	Brown *et al.* (2011)
Common carp (*Cyprinus carpio*)*	100	11–13	Izzo (2010)	c	Brown *et al.* (2011)
European chub (*Squalius cephalus*)*	50	25–30	Molbert *et al.* (2021)	c	Brown *et al.* (2011)
European seabass (*Dicentrarchus labrax*)*	48	15–30	Horn *et al.* (2008)	b	Secor *et al.* (1995)
Rainbow trout (*Oncorhynchus mykiss*)*	60	4–6	Panasiak *et al.* (2020)	d	Payan *et al.* (1998)
Siberian sturgeon (*Acipenser baerii*)*	248	40–60	Simide *et al.* (2016)	c	Campana (2001)
Turquoise killifish (*Nothobranchius furzeri*)	36	4–8 months	Hartmann *et al.* (2009)	b, c	Reichenbacher *et al.* (2007)
Zebrafish (*Danio rerio*) – model organism	48	1–4	Lund *et al.* (2009)	b	Higgs *et al.* (2002)
Zebrafish (*Danio rerio*) – model organism	48	1–4	Anchelin *et al.* (2011)	a, b	Higgs *et al.* (2002)
(ii) Amphibians:					
Australian green tree frog (*Litoria caerulea*)*	26	16–21	Lundsgaard *et al.* (2022)	c	n.a.
Natterjack toad (*Epidalea calamita*)*	22	10–15	Sánchez-Montes *et al.* (2020)	c	n.a.
(iii) Reptiles:					
Algerian psammodromus (*Psammodromus algirus*)*	38	3–5	Burraco *et al.* (2020)	c	n.a.
American alligator (*Alligator mississippiensis*)	32	30–50	Bae *et al.* (2021)	c	n.a.
American alligator (*Alligator mississippiensis*)*	32	30–50	Scott *et al.* (2006)	b	n.a.
Frilled lizard (*Chlamydosaurus kingii*)*	32	10–20	Ujvari *et al.* (2017)	c	n.a.
Loggerhead sea turtle (*Caretta caretta*)*	56	47–67	Hatase *et al.* (2008)	c	n.a.
Leatherback sea turtle (*Dermochelys coriacea*)*	56	45–50	Plot *et al.* (2012)	c	n.a.
Przewalski's toadhead agama (*Phrynocephalus przewalskii*)	46	3–10	Zhang *et al.* (2018)	c	n.a.
Water python (*Liasis fuscus*)*	36	20–30	Ujvari & Madsen (2009)	b	n.a.
(iv) Birds:					
Chats & flycatchers:					
African stonechat (*Saxicola torquatus*)*,**	80	4–5	Apfelbeck *et al.* (2019)	b	n.a.
Bluethroat (*Luscinia svecica*)	82	1–5	Sørheim (2020)	c	n.a.
Collared flycatcher (*Ficedula albicollis*)*,**	80	5–9	Stier *et al.* (2020)	c	n.a.
Collared flycatcher (*Ficedula albicollis*)*,**	80	5–9	Voillemot *et al.* (2012)	c	n.a.
European pied flycatcher (*Ficedula hypoleuca*)**	80	2–15	Kärkkäinen *et al.* (2019)	c	n.a.
European pied flycatcher (*Ficedula hypoleuca*)**	80	2–15	Kärkkäinen *et al.* (2021)	c	n.a.
European stonechat (*Saxicola rubicola*)*,**	80	4–5	Apfelbeck *et al.* (2019)	b	n.a.
Corvids:					
Florida scrub-jay (*Aphelocoma coerulescens*)*,**	80	4–9	Tricola *et al.* (2018)	b	n.a.

(Continues on next page)

Table 2. *(Cont.)*

Species	Karyotype	Lifespan (years)	Study	Method	Tooth annulation/otolith
Jackdaw (*Corvus monedula*)**	80	3–5	Bauch *et al.* (2021)	b	n.a.
Jackdaw (*Corvus monedula*)**	80	3–5	Salomons *et al.* (2009)	b	n.a.
Mexican jay (*Aphelocoma ultramarina*)*,***	80	9–15	Tricola *et al.* (2018)	b	n.a.
Cuckoos:					
Great spotted cuckoo (*Clamator glandarius*)**	78	3–6	Soler *et al.* (2015)	c	n.a.
Eagles:					
White-tailed eagle (*Haliaeetus albicilla*)	66	21–25	Hansen *et al.* (2022)	c	n.a.
Finches:					
Bengalese finch (*Lonchura striata*)*,***	78	4–7	Tricola *et al.* (2018)	b	n.a.
Pine siskin (*Spinus pinus*)	78	5–10	Vernasco & Watts (2022)	c	n.a.
Zebra finch (*Taeniopygia guttata*)*,***	78	2–3	Heidinger *et al.* (2012)	c	n.a.
Zebra finch (*Taeniopygia guttata*)*,***	78	2–3	Tricola *et al.* (2018)	b	n.a.
Geese & waterfowl:					
Barnacle goose (*Branta leucopsis*)*,***	80	14–24	Pauliny *et al.* (2012)	b	n.a.
Gulls, shorebirds & waders:					
Common ruff (*Calidris pugnax*)*,***	86	4–5	Tricola *et al.* (2018)	b	n.a.
Dunlin (*Calidris alpina*)*,***	88	5–9	Pauliny *et al.* (2006)	b	n.a.
Oystercatcher (*Haematopus longirostris*)*,***	72	12–17	Tricola *et al.* (2018)	b	n.a.
Yellow-legged gull (*Larus michahellis*)**	66	18–20	Noguera & Velando (2021)	c	n.a.
Juncos, buntings & sparrows:					
Dark-eyed junco (*Junco hyemalis*)*,***	82	3–11	Bauer *et al.* (2018)	c	n.a.
Dark-eyed junco (*Junco hyemalis*)	82	3–11	Graham (2018)	c	n.a.
House sparrow (*Passer domesticus*)	76	3–5	Meillère *et al.* (2015)	c	n.a.
Savannah sparrow (*Passerculus sandwichensis*)*,***	74	2–6	Tricola *et al.* (2018)	b	n.a.
Magpies:					
Eurasian magpie (*Pica pica*)**	82	3–4	Soler *et al.* (2015)	c	n.a.
Manakins:					
Wire-tailed manakin (*Pipra filicauda*)**	78	10–15	Vernasco *et al.* (2021)	c	n.a.
Ovenbirds:					
Thorn-tailed Rayadito (*Aphrastura spinicauda*)	82	4–9	Quirici *et al.* (2016)	c	n.a.
Parrots:					
African Grey parrot (*Psittacus erithacus erithacus*)*,***	66	60–80	Aydinonat *et al.* (2014)	c	n.a.
Penguins:					
Adelie penguin (*Pygoscelis adeliae*)*,***	96	10–20	Tricola *et al.* (2018)	b	n.a.
King penguin (*Aptenodytes patagonicus*)**	72	26–41	Reichert *et al.* (2015)	c	n.a.
Magellanic penguin (*Spheniscus magellanicus*)*,***	68	20–30	Cerchiara *et al.* (2017)	c	n.a.
Seabirds:					
Black guillemot (*Cepphus grille*)*,***	50	10–11	Tricola *et al.* (2018)	b	n.a.
Common tern (*Sterna hirundo*)*,***	68	9–10	Bauch *et al.* (2014)	b	n.a.
Common tern (*Sterna hirundo*)*,***	68	9–10	Haussmann *et al.* (2003)	b	n.a.
Common tern (*Sterna hirundo*)*,***	68	9–10	Tricola *et al.* (2018)	b	n.a.
Common tern (*Sterna hirundo*)	68	9–10	Vedder *et al.* (2017)	b	n.a.
Common tern (*Sterna hirundo*)	68	9–10	Vedder *et al.* (2022)	b	n.a.
Cory's shearwater (*Calonectris borealis*)*,***	80	20–25+	Bauch *et al.* (2020)	b	n.a.
	80	20–25+	Bauch *et al.* (2022)	b	n.a.

(Continues on next page)

Table 2. *(Cont.)*

Species	Karyotype	Lifespan (years)	Study	Method	Tooth annulation/otolith
Cory's shearwater (*Calonectris borealis*)					
Great frigatebird (*Fregata minor*)*,***	78	30–34	Juola *et al.* (2006)	b	n.a.
Great frigatebird (*Fregata minor*)*,***	78	30–34	Tricola *et al.* (2018)	b	n.a.
Leach's storm petrel (*Oceanodroma leucorhoa*)*,***	74	20–36	Tricola *et al.* (2018)	b	n.a.
Northern fulmar (*Fulmarus glacialis*)*,***	78	30–42	Tricola *et al.* (2018)	b	n.a.
Red-footed booby (*Sula sula*)*,***	78	24–40	Tricola *et al.* (2018)	b	n.a.
Thick-billed murre (*Uria lomvia*)*,***	80	20–29	Tricola *et al.* (2018)	b	n.a.
Starlings:					
Spotless starling (*Sturnus unicolor*)	80	8–9	Soler *et al.* (2017)	c	n.a.
Swallows:					
Barn swallow (*Hirundo rustica*)**	80	1–4	Parolini *et al.* (2015)	c	n.a.
Barn swallow (*Hirundo rustica*)**	80	1–4	Tricola *et al.* (2018)	b	n.a.
Mangrove swallow (*Tachycineta albilinea*)*,***	78	4–8	Tricola *et al.* (2018)	b	n.a.
Sand martin (*Riparia riparia*)*,***	80	2–9	Pauliny *et al.* (2006)	b	n.a.
Tree swallow (*Tachycineta bicolor*)*,***	78	4–8	Tricola *et al.* (2018)	b	n.a.
Swifts:					
Alpine swift (*Tachymarptis melba*)*,***	78	10–15	Bize *et al.* (2009)	c	n.a.
Thrushes:					
Eastern bluebird (*Sialia sialis*)	80	6–10	de la Iglesia (2018)	c	n.a.
Eurasian blackbird (*Turdus merula*)	80	3–21	Ibáñez-Álamo *et al.* (2018)	c	n.a.
Tits:					
Blue tit (*Cyanistes caeruleus*)**	80	3–10	Sudyka *et al.* (2014)	c	n.a.
Great tit (*Parus major*)	80	2–3	Grunst *et al.* (2020)	c	n.a.
Great tit (*Parus major*)*,***	80	2–3	Salmón *et al.* (2016)	c	n.a.
Great tit (*Parus major*)*,***	80	2–3	Tricola *et al.* (2018)	b	n.a.
Warblers:					
American redstart (*Setophaga ruticilla*)**	78	5–10	Angelier *et al.* (2013)	c	n.a.
Great reed warbler (*Acrocephalus arundinaceus*)**	80	2–6	Asghar *et al.* (2015)	c	n.a.
Seychelles warbler (*Acrocephalus sechellensis*)*,***	78	5–17	Barrett *et al.* (2013)	c	n.a.
Seychelles warbler (*Acrocephalus sechellensis*)	78	5–17	Spurgin *et al.* (2018)	c	n.a.
Wrens:					
Purple-crowned fairywren (*Malurus coronatus*)	72	3–5	Eastwood *et al.* (2019)	c	n.a.
Purple-crowned fairywren (*Malurus coronatus*)*,***	72	3–5	Roast *et al.* (2022)	c	n.a.
(v) Mammals:					
Aquatic:					
Dugong (*Dugong dugon*)*	48	50–70	Cherdsukjai *et al.* (2020)	c	Marsh, Heinsohn & Marsh (1984)
Humpback whale (*Megaptera novaeangliae*)*	44	45–50	Olsen *et al.* (2014)	c	Ohsumi *et al.* (1965)
Bats:					
Bechstein's bat (*Myotis bechsteinii*)*	44	15–21	Foley *et al.* (2018)	c	Linhart (1973)
Big brown bat (*Eptesicus fuscus*)*	50	6–19	Ineson *et al.* (2020)	c	Linhart (1973)
Common bent-wing bat (*Miniopterus schreibersii*)*	46	13–20	Foley *et al.* (2018)	c	Linhart (1973)
Greater horseshoe bat (*Rhinolophus ferrumequinum*)*	58	20–30	Foley *et al.* (2018)	c	Linhart (1973)
	44	6–13	Foley *et al.* (2018)	c	Linhart (1973)

(Continues on next page)

Table 2. *(Cont.)*

Species	Karyotype	Lifespan (years)	Study	Method	Tooth annulation/otolith
Greater mouse-eared bat (*Myotis myotis*)*					
Greater mouse-eared bat (*Myotis myotis*)	44	6–13	Foley *et al.* (2020)	c	Linhart (1973)
Little brown bat (*Myotis lucifugus*)*	44	7–31	Ineson *et al.* (2020)	c	Linhart (1973)
Carnivores:					
American black bear (*Ursus americanus*)*	74	18–23+	Kirby *et al.* (2017)	c	Marks & Erickson (1966)
American marten (*Martes americana*)*	38	6–15	Pauli *et al.* (2011)	c	Belant *et al.* (2011)
Australian sea lion (*Neophoca cinerea*)*	36	17–25	Izzo *et al.* (2011)	c	Rust *et al.* (2019)
Domestic cat (*Felis catus*)*	38	12–18	Brümmendorf *et al.* (2002)	a	Grue & Jensen (1979)
Domestic cat (*Felis catus*)*	38	12–18	McKevitt *et al.* (2003)	b	Grue & Jensen (1979)
Domestic dog (*Canis familiaris*)	78	10–13	Dutra *et al.* (2020)	c	Grue & Jensen (1979)
Domestic dog (*Canis familiaris*)*	78	10–13	Fick *et al.* (2012)	c	Grue & Jensen (1979)
Domestic dog (*Canis familiaris*)*	78	10–13	McKevitt *et al.* (2002)	b	Grue & Jensen (1979)
European badger (*Meles meles*)*	44	6–16	Beirne *et al.* (2014)	c	Ahnlund (1976)
European badger (*Meles meles*)	44	6–16	van Lieshout *et al.* (2019)	c	Ahnlund (1976)
European badger (*Meles meles*)	44	6–16	van Lieshout *et al.* (2021)	c	Ahnlund (1976)
Grizzly bear (*Ursus arctos horribilis*)*	74	20–25	Wong (2017)	b	Craighead *et al.* (1970)
Harp/Greenland seal (*Pagophilus groenlandicus*)*	32	20–30	Garde *et al.* (2010)	b, c	Bowen *et al.* (1983)
Meerkat (Suricata suricatta)	36	12–14	Cram *et al.* (2018)	c	Rausch & Pearson (1972)
Pacific marten (*Martes caurina*)*	38	6–15	Pauli *et al.* (2011)	c	Belant *et al.* (2011)
Polar bear (*Ursus maritimus*)*	74	25–30	Wong (2017)	c	Christensen-Dalsgaard *et al.* (2010)
Spotted hyena (*Crocuta crocuta*)*	40	12–25	Lewin *et al.* (2015)	b	Grue & Jensen (1979)
Elephants:					
Asian elephant (*Elephas maximus*)*	56	48–60	Buddhachat *et al.* (2017)	c	Spinage (1976)
Marsupials:					
Tasmanian devil (*Sarcophilus harrisii*)	14	5–8	Smith *et al.* (2020)	c	Pekelharing (1970)
Primates:					
Baboon (*Papio hamadryas*)*	42	27–37	Baerlocher *et al.* (2003)	a	Wada *et al.* (1978)
Baboon (*Papio hamadryas*)	42	27–37	Karere *et al.* (2019)	c	Wada *et al.* (1978)
Chimpanzee (*Pan troglodytes*)	48	32–39	Tackney *et al.* (2014)	c	Wittwer-Backofen *et al.* (2004)
Chimpanzee (*Pan troglodytes*)	48	32–39	Eisenberg *et al.* (2017)	c	Wittwer-Backofen *et al.* (2004)
Crab-eating macaque (*Macaca fascicularis*)*	42	15–30	Lee *et al.* (2002)	a, b	Wada *et al.* (1978)
Rodents:					
American red squirrel (*Tamiasciurus hudsonicus*)	46	3–8	Dantzer *et al.* (2020)	b	Adams & Watkins (1967)
Columbian ground squirrel (*Urocitellus columbianus*)*	32	5–10	Viblanc *et al.* (2022)	c	Adams & Watkins (1967)
Damaraland mole-rat (*Fukomys damarensis*)*	74	15–16	Leonida *et al.* (2020)	d	Grue & Jensen (1979)
Eastern chipmunk (*Tamias striatus*)*	38	3–8	Tissier *et al.* (2022)	c	Adams & Watkins (1967)
Hottentot mole-rat (*Cryptomys hottentotus*)*	54	11–15	Leonida *et al.* (2020)	d	Grue & Jensen (1979)
Shrews:					
Iberian shrew (*Sorex granarius*)	42	10–14 months	Zhdanova *et al.* (2010)	d	Grue & Jensen (1979)
Ungulates:					
Domestic horse (*Equus caballus*)*	64	25–30	Denham *et al.* (2019)	b	Penzhorn (1982)
Moose (*Alces alces*)	68	15–25	Fohringer *et al.* (2022)	c	Rolandsen *et al.* (2008)

(Continues on next page)

Table 2. *(Cont.)*

Species	Karyotype	Lifespan (years)	Study	Method	Tooth annulation/otolith
Roe deer (*Capreolus capreolus*)*	70	7–10	Wilbourn *et al.* (2017)	c	Grue & Jensen (1979)
Soay sheep (*Ovis aries*)	54	10–12	Fairlie *et al.* (2016)	c	Saxon & Higham (1968)
Soay sheep (*Ovis aries*)*	54	10–12	Froy *et al.* (2021)	c	Saxon & Higham (1968)
Water buffalo (*Bubalus bubalis*)*	50	25–35+	Seibt *et al.* (2019)	c	Grimsdell (1973)

the same taxonomic range as those done comparing methylation to age.

(i) Fishes (classes Agnatha, Chondrichthyes, and Osteichthyes). In fishes, telomere length has been studied in relation to ecotoxicology and environmental stressors in addition to ageing, and the validity of telomeres as proxies of life history in ectotherms has been the subject of considerable research (Friesen, Wapstra & Olsson, 2021). With regards to age, a relationship has been illustrated in several species including model organisms such as zebrafish (Anchelin *et al.*, 2011) and several wild populations such as rainbow trout [*Oncorhynchus mykiss* (Walbaum)] (Panasiak, Dobosz & Ocalewicz, 2020), European seabass (Horn *et al.*, 2008), common carp [*Cyprinus carpio* (Linnaeus)] (Izzo *et al.*, 2014), turquoise killifish [*Nothobranchius furzeri* (Jubb)] (Hartmann *et al.*, 2009), and Siberian sturgeon [*Acipenser baerii* (Brandt)] (Simide *et al.*, 2016). Although most of these studies used non-lethal sampling techniques to explore telomere length as an alternative to otolith ageing techniques, a few did include lethal sampling and even explored telomere length in the head kidney, an organ unique to teleost fishes (Panasiak *et al.*, 2020, 2022).

(ii) Amphibians (class Amphibia). Two amphibian species have been studied in the context of telomere length changes throughout their lifespan. One was principally focussed on differential changes in the Australian green tree frog [*Litoria caerulea* (White)], as a result of early life exposure to ultraviolet (UV) and different periods of light (Lundsgaard, Cramp & Franklin, 2022). The authors found that differences in early life exposure led to differences in telomere length later in life, as well as a relatedness to several key aspects that influence long-term survival in frogs. The other specifically studied telomere shortening in a wild population of naturally ageing natterjack toads [*Epidalea calamita* (Laurenti)], over a period of 3 years (Sánchez-Montes *et al.*, 2020). Their study found significant telomere shortening in the first 2–3 years of life in most of the studied individuals and a general trend of telomere shortening throughout the lifespan of toads.

(iii) Reptiles (class Reptilia). The observation that telomere length is affected by temperature or light exposure in ectotherms was replicated in both experimental and wild populations of reptiles. Under experimental conditions, telomeres were found to shorten faster and result in shortened lifespans in Przewalski's toadhead agama [*Phrynocephalus przewalskii* (Strauch)] (Zhang *et al.*, 2018). In the wild, Algerian psammodromus [*Psammodromus algirus* (Linnaeus)] lizards showed a cline of variable telomere length across a latitudinal gradient of their habitat (Burraco *et al.*, 2020). In terms of age-related telomere studies, telomere length has been shown to decline with age in American alligators (Scott *et al.*, 2006; Bae *et al.*, 2021), as well as loggerhead sea turtles (Hatase *et al.*, 2008; Plot *et al.*, 2012). Conversely, telomeres were found to increase in length in tropical water python [*Liasis fuscus* (Peters)] (Ujvari & Madsen, 2009) and follow a curvilinear relationship in frilled lizards [*Chlamydosaurus kingii* (Gray)] with both increases and decreases across their full lifespan (Ujvari *et al.*, 2017).

(iv) Birds (class Aves). Birds represent the best studied class of wild vertebrates in terms of telomeres (Criscuolo, Dobson & Schull, 2021) with most studies falling into one of four broad categories: ecotoxicological studies, stress-associated studies, migration-related studies, and age-model studies. As in other organisms, telomere length has been shown to decrease at an increased rate in response to toxins or contaminants in birds, including Arctic-breeding black-legged kittiwakes [*Rissa tridactyla* (Linnaeus)] (Blévin *et al.*, 2016), and Cory's shearwaters [*Calonectris borealis* (Cory)] (Bauch *et al.*, 2022). In stress-associated studies, telomere length dynamics were compared to environmental stressors including noise (Dorado-Correa *et al.*, 2018; Grunst *et al.*, 2020), an urban environment (Salmón *et al.*, 2016; Ibáñez-Álamo *et al.*, 2018), or anthropogenic activity (Caccavo *et al.*, 2021) as well as social factors such as brood size (Voillemot *et al.*, 2012; Quirici *et al.*, 2016; Criscuolo, Zahn & Bize, 2017, Noguera & Velando, 2021), brood parasitism (Soler *et al.*, 2015), proximity to predators (Kärkkäinen *et al.*, 2019), and social isolation (Aydinonat *et al.*, 2014). In most cases, telomere length was negatively impacted by stress and cohort-wide increases in telomere loss were observed in response to measured stressors. It should, however, be noted that many of these studies solely included birds in their early life stages and did not access telomere length changes across the full lifespan of all species.

In terms of the relationship between telomere length and attributes of migration differences have been observed for the differential timing of breeding (Graham, 2018) and migration (Vernasco & Watts, 2022) as well as return rate (Angelier *et al.*, 2013) and reproductive success (Eastwood *et al.*, 2019). Studies have also illustrated that habitat quality

differentially affects telomere shortening within species with different breeding habitats (Angelier *et al.*, 2013) and that generally advanced shortening is a consequence of the metabolic stain that migration creates when comparing migratory and resident populations of the same species (Bauer *et al.*, 2018).

Studies with a focus on cross-sectional changes in telomere length with increased age found varied results depending on the study species (Tricola *et al.*, 2018). For example, small effects were observed for an age correlation in Magellanic penguin [*Spheniscus magellanicus* (Forster)] (Cerchiara *et al.*, 2017), and other penguins (Reichert *et al.*, 2015) while an intermediate effect was observed in several other bird species (Bize *et al.*, 2009; Vedder *et al.*, 2017; Bauer *et al.*, 2018; Roast *et al.*, 2022). The largest effect was observed in three species: common tern (Haussmann, Vleck & Nisbet, 2003), great frigatebird [*Fregata minor* (Gmelin)] (Juola *et al.*, 2006), and sand martin *Riparia riparia* (Linnaeus) (Pauliny *et al.*, 2006). Although telomere length declines were evident in birds from most lineages the rate of decline varies and appears to be correlated with maximum lifespan (Tricola *et al.*, 2018).

(v) Mammals (class Mammalia)

(*i*.1) *Aquatic mammals (infraorder Cetacea and order Sirenia)*. Telomere length was found to decline with age in humpback whales but showed a high level of variation among individuals of the same age, resulting in a model with high error (∼28 years) in the predicted ages. The authors attributed this to the low sample size for calibration as well as the sensitivity of the qPCR assay as even small experimental deviations significantly altered the assay values (Olsen *et al.*, 2014). A recent study that included dolphins found a significant decline in telomere length with age, however, the study was limited by small sample sizes ($N < 10$) with no or few replicates across the same age. One key finding was that telomere length tended to decrease at an increased rate in species with shorter lifespans (Whittemore *et al.*, 2019). In wild dugong [*Dugong dugon* (Müller)], relatives of manatees in the herbivorous aquatic mammal order Sirenia, telomere length was highly correlated with age and showed less variability at the individual level (Cherdsukjai *et al.*, 2020).

(*i*.2) *Bats (order Chiroptera)*. Telomeres have been studied in at least six species of bats including several members of the long lived *Myotis* genus (Foley *et al.*, 2018, 2020; Ineson *et al.*, 2020). In both greater horseshoe bat [*Rhinolophus ferrumequinum* (Schreber)] and common bent-wing bat [*Miniopterus schreibersii* (Kuhl)] telomere length was found to decrease rapidly with age while telomeres in *Myotis* species, including Bechstein's bat, were well maintained throughout their lifespan (Foley *et al.*, 2018). Due to the absence of telomerase in mature blood cells this was likely due to the upregulation of genes involved in the ALT pathway (Foley *et al.*, 2020). A study on two American bat species (Ineson *et al.*, 2020) found telomere length to be correlated with age in both big brown bats [*Eptesicus fuscus* (Beauvois)] and little brown bats [*Myotis lucifugus* (Le Conte)]. The authors did, however, note that the relative predictive power for unknown-age individuals was low and the relationship between telomere attrition and age in *M. lucifugus* was correlated with advanced ageing in response to white nose syndrome, a common fungal infection in some bats.

(*i*.3) *Carnivores (order Carnivora)*. Some of the first studies in non-model animals with regards to telomeres and ageing used domesticated companion animals including the domestic dog and cat. In canids, telomere length was found to shorten at an advanced rate (McKevitt *et al.*, 2002) and to be correlated with the maximum lifespan of breeds (Fick *et al.*, 2012). Similar results were obtained from longitudinal studies in felids (Brümmendorf *et al.*, 2002; McKevitt *et al.*, 2003) including the observation that telomere length varies in sub-populations of leucocytes. Other studies on other members of the suborder Feliformia included the spotted hyena [*Crocuta crocuta* (Erxleben)], and meerkat [*Suricata suricatta* (Schreber)]. For both species, studies focussed on socio-ecological factors affecting variable telomere attrition primarily in relation to social rank and found higher ranking members to have a slower rate of telomere attrition and a corresponding longer life expectancy than low-ranking members (Lewin *et al.*, 2015; Cram *et al.*, 2018).

A study on two species of martens, the American marten [*Martes americana* (Turton)], and Pacific marten [*M. caurina* (Merriam)], specifically used tooth cementum annulation to determine age for cross-validation of a telomere-based PCR method of assigning age (Pauli *et al.*, 2011). The authors were able to establish a model that can assign the age of individuals to one of five functionally relevant classes with ∼80% accuracy, which improved to over 90% when including morphometric data. Several studies in the related European badger [*Meles meles* (Linnaeus)] found diverse factors to contribute to telomere biology in these species including age-related declines (Beirne *et al.*, 2014), disease-related declines (Beirne *et al.*, 2014), transgenerational effects (van Lieshout *et al.*, 2021), and differential fitness (van Lieshout *et al.*, 2019).

Other studied carnivores include several species of bears (Wong, 2017; Kirby, Alldredge & Pauli, 2017) and pinnipeds (Garde *et al.*, 2010; Izzo *et al.*, 2011) for which a small to intermediate effect for age was observed.

(*i*.4) *Elephants (order Proboscidea, suborder Elephantiformes)*. Telomere length has been studied in relation to age in the Asian elephant using 122 samples across the full lifespan of the species (Buddhachat *et al.*, 2017). A significant decline in telomere length was found ($P < 0.05$) independent of any observed effect for sex-linked differences in ageing rate. Furthermore, this study found telomere length to be markedly increased in younger individuals in comparison to shorter lived species, while the general trend and rate of telomere shortening was consistent in both Asian elephants and shorter-lived species. Another study (Whittemore *et al.*, 2019) evaluated differential telomere shortening in species with different lifespans, including Sumatran subspecies of Asian elephants (*Elephas maximus sumatranus*), but was limited in resolution by small sample sizes ($N = 4$). Nevertheless,

this study did detect a decline in telomere length and differential rates of attrition among species of varied lifespans.

(*i*.5) *Marsupials (infraorder Marsupialia)*. Few studies have been conducted on telomere dynamics in marsupials. One study, in Tasmanian devils, focussed on telomere length as a predictive risk factor for facial tumours and found longer telomeres to have a protective role that delays the age of onset for facial tumours in different populations (Smith *et al*., 2020). Studies on sexual dimorphism in telomere length in Tasmanian devils and other marsupials have illustrated that while telomeres shorten during oogenesis they tend to elongate during spermatogenesis and the inherited length therefore shows strong parental effects. In facial tumours, however, all cells showed characteristically short telomeres even though telomerase activity is typically upregulated in cancer cells (Bender *et al*., 2012).

(*i*.6) *Primates (order Primates)*. Beyond the cornucopia of telomere studies conducted in humans, telomeres have been assayed in relation to age in at least three other primates. The first analysed telomere length in an ageing population of laboratory-kept crab-eating macaque [*Macaca fascicularis* (Raffles)] using flow-FISH (Lee *et al*., 2002). This study highlighted two important aspects of ageing in primates. Firstly, telomere length was shown to decline with age in a predictable manner and, secondly, the ratios of different leucocyte sub-types changes with age. A similar study in baboons (Baerlocher *et al*., 2003, 2007) further supported these findings. In chimpanzees, studies have been done on telomere dynamics in relation to sex and age. A study on female chimpanzees found telomere length is increased as compared to humans, however, the rate of telomere attrition is conserved in both species and declines predictably with age (Tackney *et al*., 2014). In male chimpanzees, the heritability of telomere length based on paternal age at conception was explored and an increase in the telomeres of spermatozoa with increased paternal age was observed with consequently high levels of paternal inheritance of telomere length (Eisenberg *et al*., 2017).

(*i*.7) *Rodents (order Rodentia)*. Studies on telomeres in rodents revealed that food availability differentially affects telomere attrition rates in dormouse species (Hoelzl *et al*., 2016*a*) while hibernation tends to increase telomere length (Hoelzl *et al*., 2016*b*). A similar trend was also observed for post-hibernation differences in telomere length among female Columbian ground squirrels [*Urocitellus columbianus* (Ord)] (Viblanc *et al*., 2022). Experimental studies in glucocorticoid-treated mothers with offspring in the American red squirrel [*Tamiasciurus hudsonicus* (Erxleben)] (Dantzer *et al*., 2020) found no evidence of maternal stress as measured by endocrine responses adversely affecting telomere length in offspring either during parturition, lactation, or rearing. This indicates that stress-associated telomere attrition in the young may be independent of a biological maternal effect. Further evidence of telomere elongation in hibernating rodent species was recently established in a longitudinal study of Eastern chipmunks [*Tamias striatus* (Linnaeus)] (Tissier *et al*., 2022). In the non-hibernating mole-rat, two species were, however, found to have shortening telomeres with advanced age in mature bone marrow cells (Leonida *et al*., 2020).

(*i*.8) *Shrews (order Eulipotyphla)*. Studies on telomeres in shrews included species such as the Iberian shrew [*Sorex granaries* (Miller)], however the complex biology of telomeres in these species, including telocentric chromosomes and the inclusion of ribosomal DNA between telomere repeats, makes them unsuitable as a biomarker of age (Zhdanova *et al*., 2010).

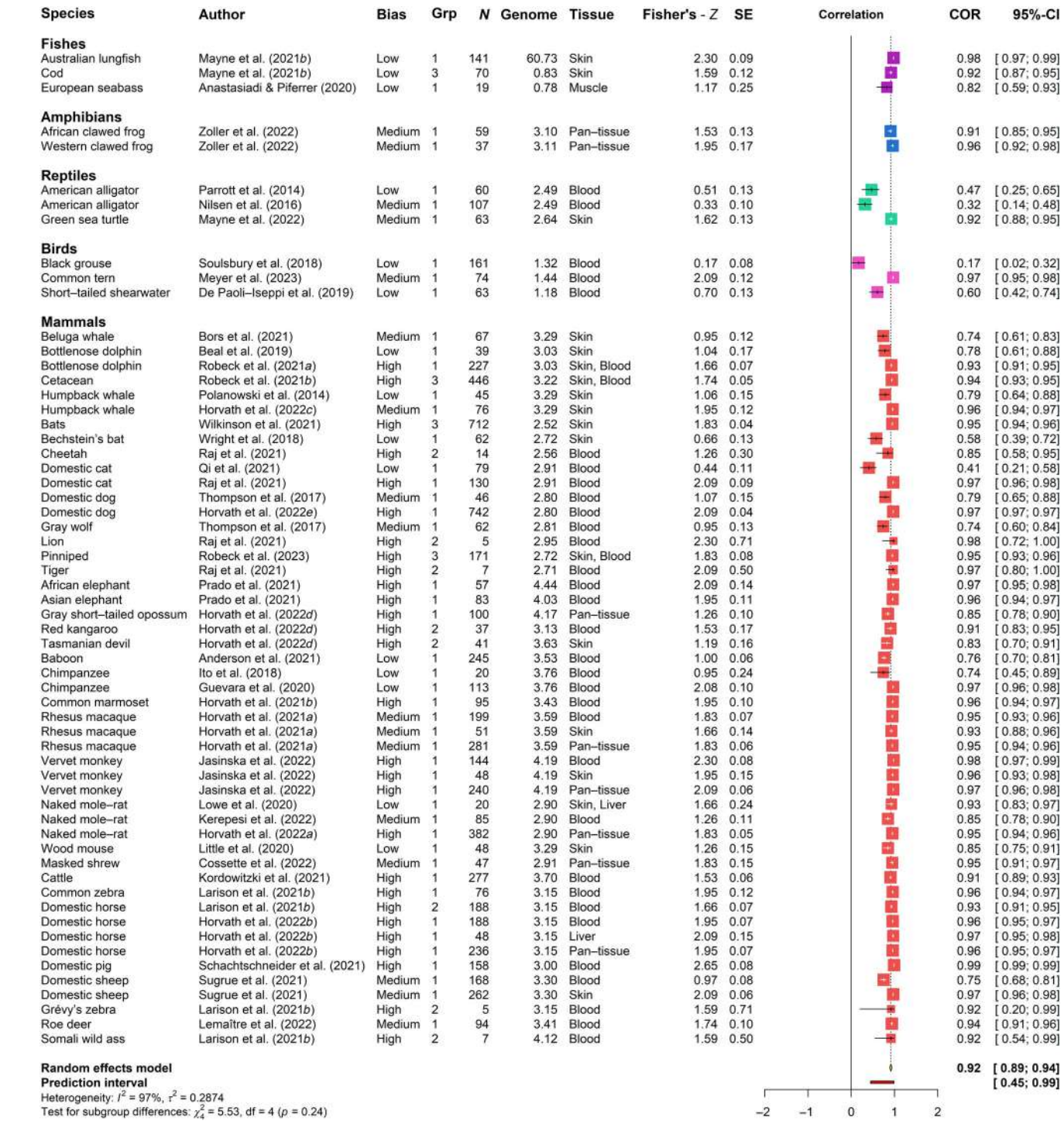
(*i*.9) *Ungulates (orders Artiodactyla and Perissodactyla)*. Several ungulate species for which methylation clocks of age have been established were also studied in telomere–age studies. Among deer species, telomeres of both roe deer and moose [*Alces alces* (Linnaeus)] from cross-sectional studies of populations from different ecoregions found telomere length changes with age to be highly correlated with geographic location, likely due to differential exposure to environmental stressors and food availability (Wilbourn *et al*., 2017; Fohringer *et al*., 2022). In Soay sheep [*Ovis aries* (Linnaeus)], a recently domesticated breed of sheep, longitudinal studies of a largely unmanaged population found evident differences in telomere shortening based on biological sex (Watson *et al*., 2017); while no evidence was found for an effect of parental age at conception (Froy *et al*., 2017) despite the high heredity (Froy *et al*., 2021) previously linked to individual lifespans (Fairlie *et al*., 2016) in this species. Telomere length was also found to decrease with age in domesticated horses (Denham, Stevenson & Denham, 2019) and domesticated water buffalo [*Bubalus bubalis* (Linnaeus)] (Seibt *et al*., 2019).

IV. META-ANALYSIS RESULTS

(1) Methylation-based clocks

Results from the meta-analysis of studies using methylation as a biological clock for age determination are summarised in Fig. 9. The global pooled cross-study effect was significant in favour of methylation being an accurate method of age prediction ($R^2 = 0.92$, $k = 60$, $P < 0.01$). No significant differences were observed in effect size between classes ($Q = 5.53$, df = 4, $P = 0.24$), however, a separate meta-analysis with clustering based on groups found a significant difference in effect sizes between the groups (Fig. S4; $Q = 10.13$, df = 2, $P < 0.01$). Further subgroup differences were also evident for effect size between models based on tissue type (Fig. S5; $Q = 17.52$, df = 6, $P < 0.01$) and study method (Fig. S6; $Q = 232.24$, df = 4, $P < 0.01$).

Between-study heterogeneity was high ($I^2 = 97\%$, $\tau^2 = 0.29$, $P < 0.01$). This heterogeneity was less for within-group analyses for groups 2 and 3 (Fig. S4) indicating that much of the heterogeneity was likely due to the smaller sample size used to validate models in related species not used to train the model. There was, however, still a high level of heterogeneity in group 1. Results from meta-regressions assessing individual factors as possible sources of heterogeneity are summarised in Table 3. As expected, heterogeneity and author were correlated with more than half of the



Species	Author	Bias	Grp	*N*	Genome	Tissue	Fisher's - *Z*	SE	COR	95%-CI
Fishes										
Australian lungfish	Mayne et al. (2021*b*)	Low	1	141	60.73	Skin	2.30	0.09	0.98	[0.97; 0.99]
Cod	Mayne et al. (2021*b*)	Low	3	70	0.83	Skin	1.59	0.12	0.92	[0.87; 0.95]
European seabass	Anastasiadi & Piferrer (2020)	Low	1	19	0.78	Muscle	1.17	0.25	0.82	[0.59; 0.93]
Amphibians										
African clawed frog	Zoller et al. (2022)	Medium	1	59	3.10	Pan–tissue	1.53	0.13	0.91	[0.85; 0.95]
Western clawed frog	Zoller et al. (2022)	Medium	1	37	3.11	Pan–tissue	1.95	0.17	0.96	[0.92; 0.98]
Reptiles										
American alligator	Parrott et al. (2014)	Low	1	60	2.49	Blood	0.51	0.13	0.47	[0.25; 0.65]
American alligator	Nilsen et al. (2016)	Medium	1	107	2.49	Blood	0.33	0.10	0.32	[0.14; 0.48]
Green sea turtle	Mayne et al. (2022)	Medium	1	63	2.64	Skin	1.62	0.13	0.92	[0.88; 0.95]
Birds										
Black grouse	Soulsbury et al. (2018)	Low	1	161	1.32	Blood	0.17	0.08	0.17	[0.02; 0.32]
Common tern	Meyer et al. (2023)	Medium	1	74	1.44	Blood	2.09	0.12	0.97	[0.95; 0.98]
Short–tailed shearwater	De Paoli–Iseppi et al. (2019)	Low	1	63	1.18	Blood	0.70	0.13	0.60	[0.42; 0.74]
Mammals										
Beluga whale	Bors et al. (2021)	Medium	1	67	3.29	Skin	0.95	0.12	0.74	[0.61; 0.83]
Bottlenose dolphin	Beal et al. (2019)	Low	1	39	3.03	Skin	1.04	0.17	0.78	[0.61; 0.88]
Bottlenose dolphin	Robeck et al. (2021*a*)	High	1	227	3.03	Skin, Blood	1.66	0.07	0.93	[0.91; 0.95]
Cetacean	Robeck et al. (2021*b*)	High	3	446	3.22	Skin, Blood	1.74	0.05	0.94	[0.93; 0.95]
Humpback whale	Polanowski et al. (2014)	Low	1	45	3.29	Skin	1.06	0.15	0.79	[0.64; 0.88]
Humpback whale	Horvath et al. (2022*c*)	Medium	1	76	3.29	Skin	1.95	0.12	0.96	[0.94; 0.97]
Bats	Wilkinson et al. (2021)	High	3	712	2.52	Skin	1.83	0.04	0.95	[0.94; 0.96]
Bechstein's bat	Wright et al. (2018)	Low	1	62	2.72	Skin	0.66	0.13	0.58	[0.39; 0.72]
Cheetah	Raj et al. (2021)	High	2	14	2.56	Blood	1.26	0.30	0.85	[0.58; 0.95]
Domestic cat	Qi et al. (2021)	Low	1	79	2.91	Blood	0.44	0.11	0.41	[0.21; 0.58]
Domestic cat	Raj et al. (2021)	High	1	130	2.91	Blood	2.09	0.09	0.97	[0.96; 0.98]
Domestic dog	Thompson et al. (2017)	Medium	1	46	2.80	Blood	1.07	0.15	0.79	[0.65; 0.88]
Domestic dog	Horvath et al. (2022*e*)	High	1	742	2.80	Blood	2.09	0.04	0.97	[0.97; 0.97]
Gray wolf	Thompson et al. (2017)	Medium	1	62	2.81	Blood	0.95	0.13	0.74	[0.60; 0.84]
Lion	Raj et al. (2021)	High	2	5	2.95	Blood	2.30	0.71	0.98	[0.72; 1.00]
Pinniped	Robeck et al. (2023)	High	3	171	2.72	Skin, Blood	1.83	0.08	0.95	[0.93; 0.96]
Tiger	Raj et al. (2021)	High	2	7	2.71	Blood	2.09	0.50	0.97	[0.80; 1.00]
African elephant	Prado et al. (2021)	High	1	57	4.44	Blood	2.09	0.14	0.97	[0.95; 0.98]
Asian elephant	Prado et al. (2021)	High	1	83	4.03	Blood	1.95	0.11	0.96	[0.94; 0.97]
Gray short–tailed opossum	Horvath et al. (2022*d*)	High	1	100	4.17	Pan–tissue	1.26	0.10	0.85	[0.78; 0.90]
Red kangaroo	Horvath et al. (2022*d*)	High	2	37	3.13	Blood	1.53	0.17	0.91	[0.83; 0.95]
Tasmanian devil	Horvath et al. (2022*d*)	High	2	41	3.63	Skin	1.19	0.16	0.83	[0.70; 0.91]
Baboon	Anderson et al. (2021)	Low	1	245	3.53	Blood	1.00	0.06	0.76	[0.70; 0.81]
Chimpanzee	Ito et al. (2018)	Low	1	20	3.76	Blood	0.95	0.24	0.74	[0.45; 0.89]
Chimpanzee	Guevara et al. (2020)	Low	1	113	3.76	Blood	2.08	0.10	0.97	[0.96; 0.98]
Common marmoset	Horvath et al. (2021*b*)	High	1	95	3.43	Blood	1.95	0.10	0.96	[0.94; 0.97]
Rhesus macaque	Horvath et al. (2021*a*)	Medium	1	199	3.59	Blood	1.83	0.07	0.95	[0.93; 0.96]
Rhesus macaque	Horvath et al. (2021*a*)	Medium	1	51	3.59	Skin	1.66	0.14	0.93	[0.88; 0.96]
Rhesus macaque	Horvath et al. (2021*a*)	Medium	1	281	3.59	Pan–tissue	1.83	0.06	0.95	[0.94; 0.96]
Vervet monkey	Jasinska et al. (2022)	High	1	144	4.19	Blood	2.30	0.08	0.98	[0.97; 0.99]
Vervet monkey	Jasinska et al. (2022)	High	1	48	4.19	Skin	1.95	0.15	0.96	[0.93; 0.98]
Vervet monkey	Jasinska et al. (2022)	High	1	240	4.19	Pan–tissue	2.09	0.06	0.97	[0.96; 0.98]
Naked mole–rat	Lowe et al. (2020)	Low	1	20	2.90	Skin, Liver	1.66	0.24	0.93	[0.83; 0.97]
Naked mole–rat	Kerepesi et al. (2022)	Medium	1	85	2.90	Blood	1.26	0.11	0.85	[0.78; 0.90]
Naked mole–rat	Horvath et al. (2022*a*)	High	1	382	2.90	Pan–tissue	1.83	0.05	0.95	[0.94; 0.96]
Wood mouse	Little et al. (2020)	Low	1	48	3.29	Skin	1.26	0.15	0.85	[0.75; 0.91]
Masked shrew	Cossette et al. (2022)	Medium	1	47	2.91	Pan–tissue	1.83	0.15	0.95	[0.91; 0.97]
Cattle	Kordowitzki et al. (2021)	High	1	277	3.70	Blood	1.53	0.06	0.91	[0.89; 0.93]
Common zebra	Larison et al. (2021*b*)	High	1	76	3.15	Blood	1.95	0.12	0.96	[0.94; 0.97]
Domestic horse	Larison et al. (2021*b*)	High	2	188	3.15	Blood	1.66	0.07	0.93	[0.91; 0.95]
Domestic horse	Horvath et al. (2022*b*)	High	1	188	3.15	Blood	1.95	0.07	0.96	[0.95; 0.97]
Domestic horse	Horvath et al. (2022*b*)	High	1	48	3.15	Liver	2.09	0.15	0.97	[0.95; 0.98]
Domestic horse	Horvath et al. (2022*b*)	High	1	236	3.15	Pan–tissue	1.95	0.07	0.96	[0.95; 0.97]
Domestic pig	Schachtschneider et al. (2021)	High	1	158	3.00	Blood	2.65	0.08	0.99	[0.99; 0.99]
Domestic sheep	Sugrue et al. (2021)	Medium	1	168	3.30	Blood	0.97	0.08	0.75	[0.68; 0.81]
Domestic sheep	Sugrue et al. (2021)	Medium	1	262	3.30	Skin	2.09	0.06	0.97	[0.96; 0.98]
Grévy's zebra	Larison et al. (2021*b*)	High	2	5	3.15	Blood	1.59	0.71	0.92	[0.20; 0.99]
Roe deer	Lemaître et al. (2022)	Medium	1	94	3.41	Blood	1.74	0.10	0.94	[0.91; 0.96]
Somali wild ass	Larison et al. (2021*b*)	High	2	7	4.12	Blood	1.59	0.50	0.92	[0.54; 0.99]
Random effects model									**0.92**	**[0.89; 0.94]**
Prediction interval										**[0.45; 0.99]**

Heterogeneity: I^2 = 97%, τ^2 = 0.2874
Test for subgroup differences: χ^2_4 = 5.53, df = 4 (p = 0.24)

Fig. 9. Forest plot and results for the meta-analysis of studies using methylation to estimate animal age grouped by class. The pooled correlation from the random effects model is indicated (yellow diamond) along with measures of between-study heterogeneity and subgroup differences. The prediction interval is indicated with a maroon bar. Key study attributes such as sample size (*N*), genome size (in bbp), and tissue type are indicated along with the computed level of possible bias. CI, confidence interval; COR, correlation; *N*, sample size; SE, standard error.

heterogeneity being accounted for by within-study variances. A relationship with heterogeneity was also illustrated for both year and study method ($P < 0.01$) as well as class ($P < 0.05$) and sample size ($P < 0.10$), however, no relationship was found with maximum lifespan estimates, tissue type, or genome size. There were some significant indicators for

Table 3. Summary of results for multiple meta-regressions for heterogeneity. Lifespan estimates are derived from the lower, average/middle, and upper estimates of average lifespans reported in Tables 1 and 2. Significance: *$P < 0.10$, **$P < 0.05$, and ***$P < 0.01$

Factor	Z-value	P-value	Degrees of freedom (df)	Correlation (R^2)
Methylation:				
Author	3.10	<0.01***	40	0.724
Year	2.65	<0.01***	8	0.467
Bias	20.32	<0.01***	2	0.297
Sample size (N)	1.90	0.057*	1	0.049
Tissue	14.61	0.762	6	<0.01
Method	9.58	<0.01***	4	0.492
Class	5.65	0.022**	4	0.126
Genome size (bbp)	1.63	0.103	1	0.030
Lifespan:				
Lower	11.70	0.290	1	0.054
Middle	10.80	0.390	1	<0.01
Upper	10.32	0.497	1	<0.01
Telomeres (All):				
Author	4.66	<0.01***	46	0.513
Year	3.07	0.060*	17	0.147
Bias	5.63	0.251	2	0.013
Sample size (N)	8.94	0.057*	1	0.043
Tissue	6.83	0.368	7	0.057
Method	2.91	0.225	3	0.020
Class	3.54	0.499	4	<0.01
Karyotype (diploid)	2.26	0.024**	1	0.065
Lifespan:				
Lower	4.48	0.172	1	0.014
Middle	4.08	0.149	1	0.019
Upper	3.87	0.139	1	0.020
Telomeres (Birds):				
Author	3.123	<0.01***	24	0.378
Year	0.008	0.991	1	<0.01
Sample size (N)	6.553	0.507	1	<0.01

asymmetry and publication bias or small study effects (t-statistic = −2.30, df = 58, $P = 0.03$). The quantification of potential author bias using *ABCal* identified several studies (28 models) with a higher risk of bias considering they share a significant number of authors and co-authors, compared to medium- (17 models) and low-risk (15 models) studies. The computed risk of bias also accounted for a significant ($P < 0.01$) amount of the heterogeneity. Notably, eight authors contributed to more than five of the included studies while four contributed to more than 10.

(2) Telomere-based clocks

Results for the meta-analysis of studies using telomeres as a biological clock for age determination across taxa are summarised in Fig. 10 and the separate analysis for birds in Fig. 11. The global pooled cross-study effect was significant in favour of telomere length being an accurate method of age prediction ($R^2 = 0.31$, $k = 60$, $P < 0.01$). A significant difference was observed in effect size between classes ($Q = 46.43$, df = 4, $P < 0.01$), between study groups (Fig. S7; $Q = 28.96$, df = 1, $P < 0.01$), between models based on tissue used (Fig. S8; $Q = 522.36$, df = 7, $P < 0.01$), and study method (Fig. S9; $Q = 44.10$, df = 3, $P < 0.01$). When analysing only studies in birds, the pooled effect was slightly larger ($R^2 = 0.43$, $k = 51$, $P < 0.01$) with a similar prediction interval to cross-taxa studies.

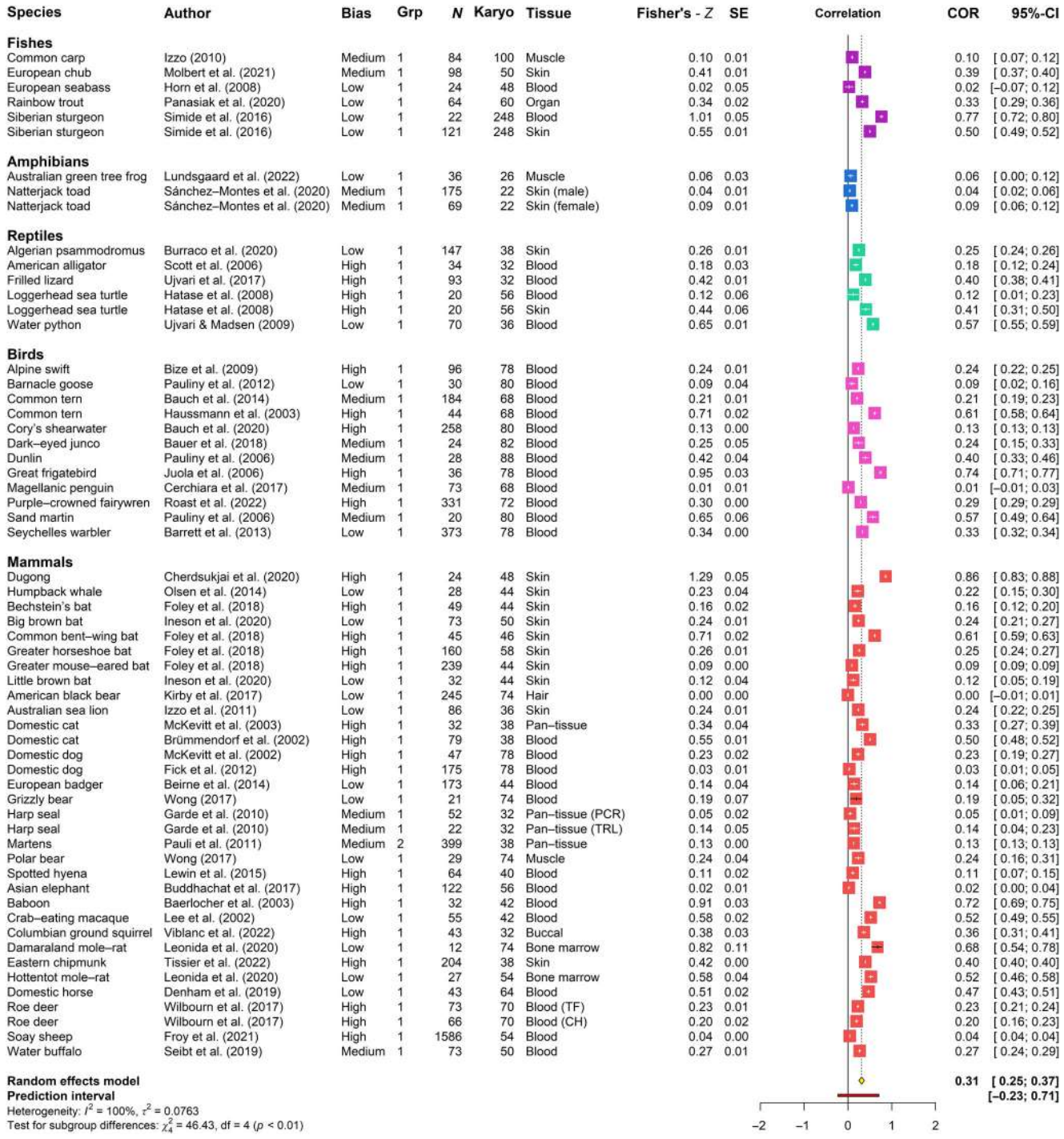
Between-study heterogeneity was high for both cross-taxa studies ($I^2 = 100\%$, $\tau^2 = 0.076$, $P < 0.01$) as well as birds alone. Results from the meta-regression assessing possible contributors to heterogeneity are summarised in Table 3. As for methylation, a correlation was detected for heterogeneity and author (as well as year) accounting for more than half of the heterogeneity, however, no relationship was evident for study method, tissue type, or lifespan. A small effect was observed for sample size and karyotype, however, both only accounted for <10% of the observed heterogeneity. There were some significant indicators for asymmetry and publication bias or small study effects for the cross taxa analyses (t-statistic = 3.33, df = 58, $P < 0.01$), as for methylation, but not for the birds alone (t-statistic = −0.18, df = 49, $P = 0.86$). The quantification of potential author bias using *ABCal* identified a near-equal number of studies (26 models) with a higher risk of bias for the between-study meta-analyses considering they share a significant number of authors and co-authors. This was, however, not found to be a significant contributor to the heterogeneity of studies ($P = 0.25$). This risk of bias was generally lower than for methylation studies as most of the highly represented authors only contributed to 2–4 of the included models.

(3) Methylation *versus* telomeres

Results for the comparison between telomere length and methylation as biomarkers for age determination are indicated in Table 4 and Fig. S10. The overall effect size for methylation-based studies was higher than that observed for telomeres, 0.92 *versus* 0.31, with a narrower prediction interval. The observed difference was statistically significant (Z-value = 21.58, $P < 0.01$) while no significant difference was found between the sample size distributions of the two methods (t-statistic = 0.422, df = 104.5, $P = 0.67$).

V. DISCUSSION

Our systematic review on the use of biological clocks as markers for age determination in wildlife clearly illustrates that several biological clocks are well suited to this purpose. Two methods, tooth cementum annulation and otoliths, have been in use for nearly a century, however, both methods have limitations on their use in studying age in living animals as part of continued conservation efforts. We therefore identified and synthesised the molecular mechanisms and existing



Species	Author	Bias	Grp	*N*	Karyo	Tissue	Fisher's - *Z*	SE	COR	95%-CI
Fishes										
Common carp	Izzo (2010)	Medium	1	84	100	Muscle	0.10	0.01	0.10	[0.07; 0.12]
European chub	Molbert et al. (2021)	Medium	1	98	50	Skin	0.41	0.01	0.39	[0.37; 0.40]
European seabass	Horn et al. (2008)	Low	1	24	48	Blood	0.02	0.05	0.02	[−0.07; 0.12]
Rainbow trout	Panasiak et al. (2020)	Low	1	64	60	Organ	0.34	0.02	0.33	[0.29; 0.36]
Siberian sturgeon	Simide et al. (2016)	Low	1	22	248	Blood	1.01	0.05	0.77	[0.72; 0.80]
Siberian sturgeon	Simide et al. (2016)	Low	1	121	248	Skin	0.55	0.01	0.50	[0.49; 0.52]
Amphibians										
Australian green tree frog	Lundsgaard et al. (2022)	Low	1	36	26	Muscle	0.06	0.03	0.06	[0.00; 0.12]
Natterjack toad	Sánchez–Montes et al. (2020)	Medium	1	175	22	Skin (male)	0.04	0.01	0.04	[0.02; 0.06]
Natterjack toad	Sánchez–Montes et al. (2020)	Medium	1	69	22	Skin (female)	0.09	0.01	0.09	[0.06; 0.12]
Reptiles										
Algerian psammodromus	Burraco et al. (2020)	Low	1	147	38	Skin	0.26	0.01	0.25	[0.24; 0.26]
American alligator	Scott et al. (2006)	High	1	34	32	Blood	0.18	0.03	0.18	[0.12; 0.24]
Frilled lizard	Ujvari et al. (2017)	High	1	93	32	Blood	0.42	0.01	0.40	[0.38; 0.41]
Loggerhead sea turtle	Hatase et al. (2008)	High	1	20	56	Blood	0.12	0.06	0.12	[0.01; 0.23]
Loggerhead sea turtle	Hatase et al. (2008)	High	1	20	56	Skin	0.44	0.06	0.41	[0.31; 0.50]
Water python	Ujvari & Madsen (2009)	Low	1	70	36	Blood	0.65	0.01	0.57	[0.55; 0.59]
Birds										
Alpine swift	Bize et al. (2009)	High	1	96	78	Blood	0.24	0.01	0.24	[0.22; 0.25]
Barnacle goose	Pauliny et al. (2012)	Low	1	30	80	Blood	0.09	0.04	0.09	[0.02; 0.16]
Common tern	Bauch et al. (2014)	Medium	1	184	68	Blood	0.21	0.01	0.21	[0.19; 0.23]
Common tern	Haussmann et al. (2003)	High	1	44	68	Blood	0.71	0.02	0.61	[0.58; 0.64]
Cory's shearwater	Bauch et al. (2020)	High	1	258	80	Blood	0.13	0.00	0.13	[0.13; 0.13]
Dark–eyed junco	Bauer et al. (2018)	Medium	1	24	82	Blood	0.25	0.05	0.24	[0.15; 0.33]
Dunlin	Pauliny et al. (2006)	Medium	1	28	88	Blood	0.42	0.04	0.40	[0.33; 0.46]
Great frigatebird	Juola et al. (2006)	High	1	36	78	Blood	0.95	0.03	0.74	[0.71; 0.77]
Magellanic penguin	Cerchiara et al. (2017)	Medium	1	73	68	Blood	0.01	0.01	0.01	[−0.01; 0.03]
Purple–crowned fairywren	Roast et al. (2022)	High	1	331	72	Blood	0.30	0.00	0.29	[0.29; 0.29]
Sand martin	Pauliny et al. (2006)	Medium	1	20	80	Blood	0.65	0.06	0.57	[0.49; 0.64]
Seychelles warbler	Barrett et al. (2013)	Low	1	373	78	Blood	0.34	0.00	0.33	[0.32; 0.34]
Mammals										
Dugong	Cherdsukjai et al. (2020)	High	1	24	48	Skin	1.29	0.05	0.86	[0.83; 0.88]
Humpback whale	Olsen et al. (2014)	Low	1	28	44	Skin	0.23	0.04	0.22	[0.15; 0.30]
Bechstein's bat	Foley et al. (2018)	High	1	49	44	Skin	0.16	0.02	0.16	[0.12; 0.20]
Big brown bat	Ineson et al. (2020)	Low	1	73	50	Skin	0.24	0.01	0.24	[0.21; 0.27]
Common bent–wing bat	Foley et al. (2018)	High	1	45	46	Skin	0.71	0.02	0.61	[0.59; 0.63]
Greater horseshoe bat	Foley et al. (2018)	High	1	160	58	Skin	0.26	0.01	0.25	[0.24; 0.27]
Greater mouse–eared bat	Foley et al. (2018)	High	1	239	44	Skin	0.09	0.00	0.09	[0.09; 0.09]
Little brown bat	Ineson et al. (2020)	Low	1	32	44	Skin	0.12	0.04	0.12	[0.05; 0.19]
American black bear	Kirby et al. (2017)	Low	1	245	74	Hair	0.00	0.00	0.00	[−0.01; 0.01]
Australian sea lion	Izzo et al. (2011)	Low	1	86	36	Skin	0.24	0.01	0.24	[0.22; 0.25]
Domestic cat	McKevitt et al. (2003)	High	1	32	38	Pan–tissue	0.34	0.04	0.33	[0.27; 0.39]
Domestic cat	Brümmendorf et al. (2002)	High	1	79	38	Blood	0.55	0.01	0.50	[0.48; 0.52]
Domestic dog	McKevitt et al. (2002)	High	1	47	78	Blood	0.23	0.02	0.23	[0.19; 0.27]
Domestic dog	Fick et al. (2012)	High	1	175	78	Blood	0.03	0.01	0.03	[0.01; 0.05]
European badger	Beirne et al. (2014)	Low	1	173	44	Blood	0.14	0.04	0.14	[0.06; 0.21]
Grizzly bear	Wong (2017)	Low	1	21	74	Blood	0.19	0.07	0.19	[0.05; 0.32]
Harp seal	Garde et al. (2010)	Medium	1	52	32	Pan–tissue (PCR)	0.05	0.02	0.05	[0.01; 0.09]
Harp seal	Garde et al. (2010)	Medium	1	22	32	Pan–tissue (TRL)	0.14	0.05	0.14	[0.04; 0.23]
Martens	Pauli et al. (2011)	Medium	2	399	38	Pan–tissue	0.13	0.00	0.13	[0.13; 0.13]
Polar bear	Wong (2017)	Low	1	29	74	Muscle	0.24	0.04	0.24	[0.16; 0.31]
Spotted hyena	Lewin et al. (2015)	High	1	64	40	Blood	0.11	0.02	0.11	[0.07; 0.15]
Asian elephant	Buddhachat et al. (2017)	High	1	122	56	Blood	0.02	0.01	0.02	[0.00; 0.04]
Baboon	Baerlocher et al. (2003)	High	1	32	42	Blood	0.91	0.03	0.72	[0.69; 0.75]
Crab–eating macaque	Lee et al. (2002)	Low	1	55	42	Blood	0.58	0.02	0.52	[0.49; 0.55]
Columbian ground squirrel	Viblanc et al. (2022)	High	1	43	32	Buccal	0.38	0.03	0.36	[0.31; 0.41]
Damaraland mole–rat	Leonida et al. (2020)	Low	1	12	74	Bone marrow	0.82	0.11	0.68	[0.54; 0.78]
Eastern chipmunk	Tissier et al. (2022)	High	1	204	38	Skin	0.42	0.00	0.40	[0.40; 0.40]
Hottentot mole–rat	Leonida et al. (2020)	Low	1	27	54	Bone marrow	0.58	0.04	0.52	[0.46; 0.58]
Domestic horse	Denham et al. (2019)	Low	1	43	64	Blood	0.51	0.02	0.47	[0.43; 0.51]
Roe deer	Wilbourn et al. (2017)	High	1	73	70	Blood (TF)	0.23	0.01	0.23	[0.21; 0.24]
Roe deer	Wilbourn et al. (2017)	High	1	66	70	Blood (CH)	0.20	0.02	0.20	[0.16; 0.23]
Soay sheep	Froy et al. (2021)	High	1	1586	54	Blood	0.04	0.00	0.04	[0.04; 0.04]
Water buffalo	Seibt et al. (2019)	Medium	1	73	50	Blood	0.27	0.01	0.27	[0.24; 0.29]
Random effects model									**0.31**	**[0.25; 0.37]**
Prediction interval										**[−0.23; 0.71]**

Heterogeneity: I^2 = 100%, τ^2 = 0.0763

Test for subgroup differences: χ^2_4 = 46.43, df = 4 (p < 0.01)

Fig. 10. Forest plot and results for the meta-analysis of studies using telomere length to estimate animal age grouped by class. The pooled correlation from the random effects model is indicated (yellow diamond) along with measures of between-study heterogeneity and the prediction interval is indicated with a red bar. Key study attributes such as sample size (*N*), karyotype (Karyo), and tissue type are indicated along with the computed level of possible bias. CH, Chizé region, France; CI, confidence interval; COR, correlation; *N*, sample size; PCR, polymerase chain reaction; SE, standard error; TF, Trois-Fontaines region, France; TRL, terminal restriction length.

primary evidence for the use of two genetic biological clocks, methylation and telomere length, that are suitable alternatives for studying age in living animals.

Methylation has previously been studied in several organisms in the context of eco-toxicology (Sareisian, 2014; Nilsen *et al.*, 2016; Romano *et al.*, 2017), relating changes in methylation to exposure to environmental pollutants, as well as

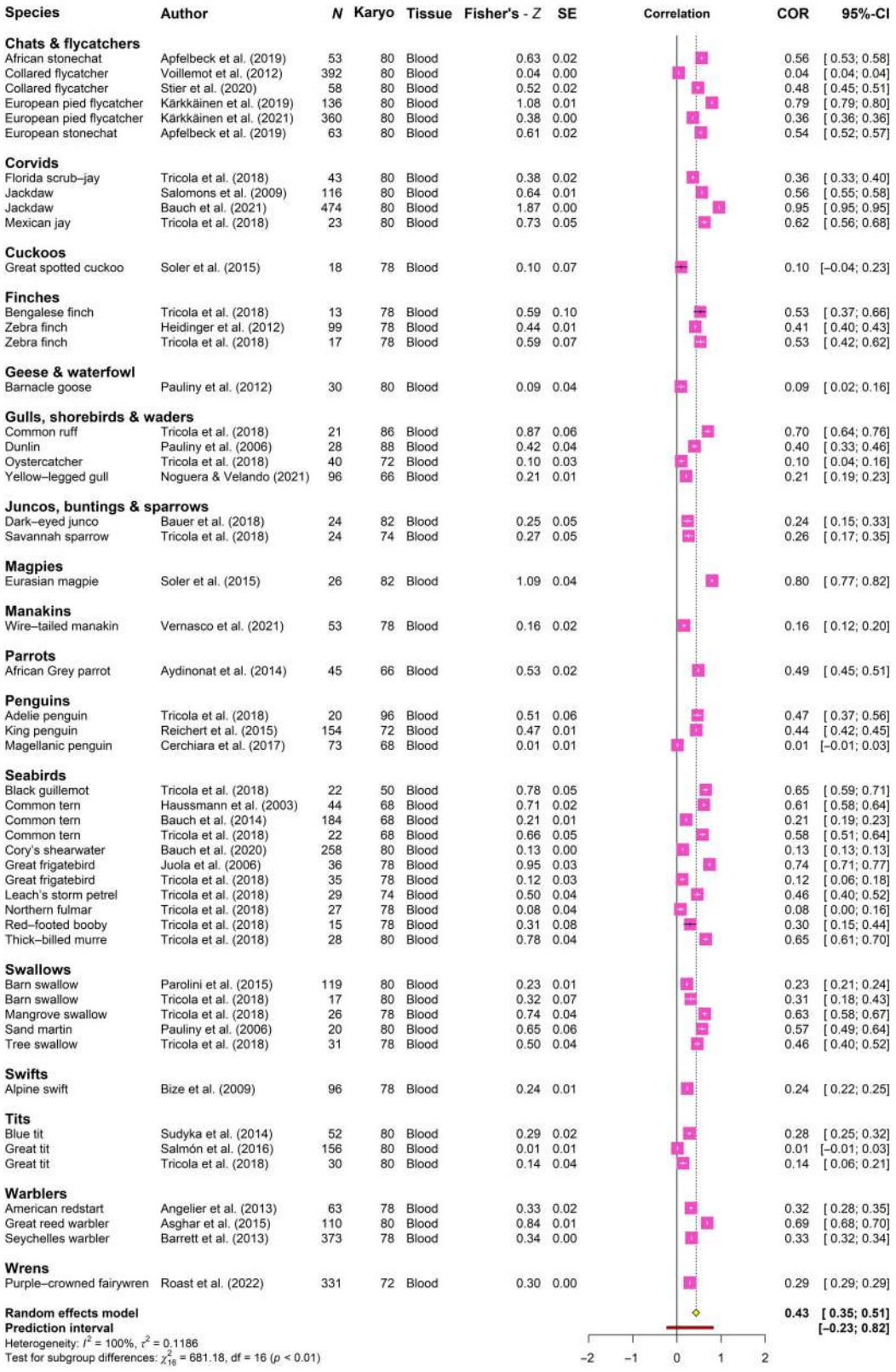


Fig. 11. Forest plot and results for the meta-analysis of studies using telomere length to estimate age in birds. Key study attributes such as sample size (*N*), karyotype (Karyo), and tissue type are indicated. The pooled correlation from the random effects model is indicated (yellow diamond) along with measures of between-study heterogeneity and the prediction interval is indicated with a red bar. CI, confidence interval; COR, correlation; *N*, sample size; SE, standard error.

Table 4. Results from the comparison of meta-analysis results from studies on methylation and telomeres for age determination in animals. A fixed-effect method was used with model ($k = 2$) as a moderator. The overall z-value and P-value indicate that the observed effect sizes and variances between the methylation studies ($N = 60$) and telomere studies ($N = 60$) is statistically significant (***$P < 0.01$) with the methylation studies having a substantially higher estimation efficiency than telomeres

Meta-analysis	N	Estimate	Standard error (σ_M)
Methylation	60	0.92	0.29
Telomeres	60	0.31	0.07
***Z*-value**		**21.58*****	
***P*-value**		**<0.01*****	

***$P < 0.01$.

several developmental or embryological studies (Suarez-Bregua *et al.*, 2021). More recently, however, many studies have shifted in focus towards the potential use of methylation to determine age in wildlife. This followed the great successes in human and murine mouse models and the accumulating evidence that methylation is conserved across most vertebrate (Jarman *et al.*, 2015), and even invertebrate (Macchi, Edsinger & Sadler, 2022), species. Thus far, methylation clocks for age determination have been developed or tested in more than 50 species across six classes in less than 10 years, including reptiles and bird species for which other biological clocks were not suitable.

Variable methods have been applied to several different tissue types to establish these models ranging from studies that used only a small, targeted selection of genes or subsets of CpGs to studies that used methylation arrays to assay thousands of CpGs at once to identify suitable targets empirically. Our review and meta-analysis illustrate that methylation is generally a highly accurate method able to predict age within weeks, months or years in many species and does not suffer significantly from variation due to changes in the lifespan of the organism or the tissue type. By contrast, one study did find a lack of age correlates in liver tissue (Sareisian, 2014) which may be related to the higher regenerative capacity of liver cells causing differential ageing in the liver compared to other tissue. Although our meta-analysis identified substantial study heterogeneity, this was mostly confined either to differences in sample size or to differences in the sensitivity of the specific study method used to assay methylation. Sample size requirements have been explored (Mayne *et al.*, 2021*b*) but remain fundamentally tied to effect size and the number of predictors used. In the case of methylation studies, calibrated with highly age-correlated CpGs, the effect sizes are generally large and sample sizes of 40 of above should be well suited for adequate statistical power, as measured by β-values (Fig. S3).

It should be noted that several authors did suggest that the accuracy of prediction is not always linear across the lifespan of organisms and that age estimates of older individuals are typically under- or overestimated. This does not, however, reduce the utility of this method to classify animals into functionally relevant age classes for conservation purposes (Wright *et al.*, 2018). Several other potential confounding factors have also been identified that may need to be taken into consideration in studies analysing methylation for age determination. For example, some evidence suggests that ageing of the epigenome in terms of methylation may be dependent on biological sex (Beal *et al.*, 2019; Meyer *et al.*, 2023), with females ageing slower, and transgenerational epigenetic inheritance in offspring may be linked to maternal age or brood size in some species (De Paoli-Iseppi *et al.*, 2017*a*). Methylation levels were also found to decrease during moulting and periods of fasting and return to normal later, highlighting the importance of taking life histories into account at the time of sampling (Poganik *et al.*, 2023). Other methods of correction such as accounting for leucocyte composition when blood cells are used have also been suggested (Adalsteinsson *et al.*, 2012), however, while variation in leucocyte ratios with age has been illustrated in some primates it is absent in others (Sato *et al.*, 2005), and the heredity or repeatability of such ratio changes is low across different generations for cross-sectional studies (Webb *et al.*, 2020).

Here we found that telomeres have been studied in relation to age in many of the same species as methylation, however, the total number of publications are nearly double the number of methylation studies with approximately half of the published studies focused on bird species. The abundance of studies on birds is likely attributable to the fact that, as jawless terrestrial vertebrates, neither tooth cementum annulation nor otoliths are suitable for age estimation in these species (Chaney *et al.*, 2003). Telomere length showed a moderate effect size in relation to age and the efficacy in relating telomere changes to age was affected by the average lifespan of some species, consistent with previous findings (Dantzer & Fletcher, 2015; Tricola *et al.*, 2018). Some studies have however found that telomere length can be heritable (Reichert *et al.*, 2015) and that both social (Lewin *et al.*, 2015; Cram *et al.*, 2018) and environmental factors can differentially affect telomere shortening rates within populations (Salmón *et al.*, 2016; Dorado-Correa *et al.*, 2018). While similar effects have been observed on methylation, telomere length tends to be permanently altered, likely due to a lack of telomerase activity in most tissue, while methylation levels recover after periods of stress (Poganik *et al.*, 2023) due to DNMT enzymes being persistently expressed. As a result, methylation outperformed telomere length in its ability to change predictably with age.

Both methylation and telomere length show distinct patterns of inheritance that vary across taxa. For example, methylation is known to follow parental imprinting in subsets of genes where different patterns of methylation are inherited from different parents (Andergassen *et al.*, 2021) while global changes can result in offspring that resemble the maternal methylome (Romano *et al.*, 2017; Saino *et al.*, 2019). Telomere length, however, showed strong paternal inheritance in mammals (Eisenberg *et al.*, 2017) but maternal inheritance in birds (Horn *et al.*, 2011; Reichert *et al.*, 2015), possibly

indicating that the parental inheritance of telomere length is driven by the heterogamous sex as most male mammals are XY while most female birds are ZW. Interestingly, paternal inheritance has also been illustrated in sand lizard species that follow a similar ZZ/ZW pattern of sex chromosome heterogamy (Olsson *et al.*, 2011). More studies are needed to elucidate if similar patterns are absent from vertebrate species that either lack sex chromosomes or are completely homogamous, including most amphibians and mammals with unusual sex-determination systems (Saunders & Veyrunes, 2021).

Using attributes of cellular and molecular ageing as biomarkers for age across many taxa is fairly complicated. It requires not only a high level of consistency but also substantial heritability in the trait to ensure comparability in the measurements from a cross-sectional study accessing levels in age groups for different generations: as opposed to longitudinal studies that track changes in specific individuals across their lifetime. Furthermore, it needs to be correlated with all relevant biological and environmental factors that contribute to variability. Biological factors include biological sex, diet, stress, differences in physiology between ectotherms and endotherms, and differences in growth patterns such as the arrested growth in mammals and birds compared to continuous growth in most fish, reptiles, and amphibians. Environmental factors may include the effects of temperature, latitude, environmental stressors, and possible exposure to toxins or contaminants. When controlling for such complex variables, both methylation and telomere length are promising targets for the development of methods to determine age in animals.

VI. CONCLUSIONS

(1) The first biological clocks used to determine age in animals were tooth cementum/dentin annulation and otoliths but are mostly applicable to toothed vertebrates and fish and remain superior tools for age estimation in the *post-mortem* interval.
(2) Methylation is a conserved molecular ageing attribute able to predict age across most classes, orders and infraorders using as little as two CpGs in calibrated clocks.
(3) Telomere shortening is another biological clock with sufficient age prediction capabilities to measure age in animals but is less conserved.
(4) Methylation may be superior to telomere length in terms of accuracy and cross-taxa portability; however, costs may be higher depending on the method.

VII. ACKNOWLEDGEMENTS

Images were created in BioRender.com. Figure 1 was reproduced with permission from the authors (Jarman *et al.*, 2015).

This work is based on research supported wholly/in part by the National Research Foundation of South Africa (Grant Number: 112062).

VIII. DATA AVAILABILITY STATEMENT

Data and code used in this study are available for download from the Zenodo repository at the following link: https://doi.org/10.5281/zenodo.7893443. The custom PYTHON script, ABCal *version* 1.0.1., used to quantitate possible bias by virtue of authorship contributions is freely available on GitHub (https://github.com/LSLeClercq/ABCal).

IX. REFERENCES

References identified with an asterisk (*) are cited only within the supporting information.

Adalsteinsson, B. T., Gudnason, H., Aspelund, T., Harris, T. B., Launer, L. J., Eiriksdottir, G., Smith, A. V. & Gudnason, V. (2012). Heterogeneity in white blood cells has potential to confound DNA methylation measurements. *PLoS One* **7**, e46705.

*Adams, D. M. & Blanchong, J. A. (2020). Precision of cementum annuli method for aging male white-tailed deer. *PLoS One* **15**, e0233421.

Adams, L. & Watkins, S. G. (1967). Annuli in tooth cementum indicate age in California ground squirrels. *The Journal of Wildlife Management* **31**, 836.

Ahnlund, H. (1976). Age determination in the European badger, *Meles meles* L. *Zeitschrift fur Saugetierkunder* **41**, 119–125.

Anastasiadi, D. & Piferrer, F. (2020). A clockwork fish. Age-prediction using DNA methylation-based biomarkers in the European seabass. *Molecular Ecology Resources* **20**, 387–397.

Anchelin, M., Murcia, L., Alcaraz-Pérez, F., García-Navarro, E. M. & Cayuela, M. L. (2011). Behaviour of telomere and telomerase during aging and regeneration in zebrafish. *PLoS One* **6**, e16955.

Andergassen, D., Smith, Z. D., Kretzmer, H., Rinn, J. L. & Meissner, A. (2021). Diverse epigenetic mechanisms maintain parental imprints within the embryonic and extraembryonic lineages. *Developmental Cell* **56**, 2995–3005.e4.

Anderson, J. A., Johnston, R. A., Lea, A. J., Campos, F. A., Voyles, T. N., Akinyi, M. Y., Alberts, S. C., Archie, E. A. & Tung, J. (2021). High social status males experience accelerated epigenetic aging in wild baboons. *eLife* **10**, e66128.

Angelier, F., Vleck, C. M., Holberton, R. L. & Marra, P. P. (2013). Telomere length, non-breeding habitat and return rate in male American redstarts. *Functional Ecology* **27**, 342–350.

*Ansorge, H. & Suchentrunk, F. (2001). Aging steppe polecats (*Mustela eversmanni*) and polecats (*Mustela putorius*) by canine cementum layers and skull characters. *Wissenschaftliche Mitteilungen des Niederösterreichischen Landesmuseums* **14**, 79–106.

Apfelbeck, B., Haussmann, M. F., Boner, W., Flinks, H., Griffiths, K., Illera, J. C., Mortega, K. G., Sisson, Z., Smiddy, P. & Helm, B. (2019). Divergent patterns of telomere shortening in tropical compared to temperate stonechats. *Ecology and Evolution* **9**, 511–521.

Arneson, A., Haghani, A., Thompson, M. J., Pellegrini, M., Kwon, S. B., Vu, H., Maciejewski, E., Yao, M., Li, C. Z., Lu, A. T., Morselli, M., Rubbi, L., Barnes, B., Hansen, K. D., Zhou, W., *et al.* (2022). A mammalian methylation array for profiling methylation levels at conserved sequences. *Nature Communications* **13**, 1–13.

Arthur, S. M., Cross, R., Paragi, T. F. & Krohn, W. B. (1992). Precision and utility of cementum annuli for estimating ages of fishers. *Wildlife Society Bulletin* **20**, 402–405.

Asghar, M., Bensch, S., Tarka, M., Hansson, B. & Hasselquist, D. (2015). Maternal and genetic factors determine early life telomere length. *Proceedings of the Royal Society B: Biological Sciences* **282**, 20142263.

Ashapkin, V. V., Kutueva, L. I. & Vanyushin, B. F. (2020). Quantitative analysis of DNA methylation by bisulfite sequencing. *Methods in Molecular Biology* **2138**, 297–312.

Attwood, J. T., Yung, R. L. & Richardson, B. C. (2002). DNA methylation and the regulation of gene transcription. *Cellular and Molecular Life Sciences* **59**, 241–257.

Aubert, G. & Lansdorp, P. M. (2008). Telomeres and aging. *Physiological Reviews* **88**, 557–579.

Aydinonat, D., Penn, D. J., Smith, S., Moodley, Y., Hoelzl, F., Knauer, F. & Schwarzenberger, F. (2014). Social isolation shortens telomeres in African Grey parrots (*Psittacus erithacus erithacus*). *PLoS One* **9**, e93839.

*Azorit, C., Analla, M., Hervas, J., Carrasco, R. & Muñoz-Cobo, J. (2002). Growth marks observation: preferential techniques and teeth for ageing of Spanish red deer (*Cervus elaphus hispanicus*). *Anatomia, Histologia, Embryologia* **31**, 303–307.

Bae, J., Bertucci, E. M., Bock, S. L., Hale, M. D., Moore, J., Wilkinson, P. M., Rainwater, T. R., Bowden, J. A., Koal, T., PhamTuan, H. & Parrott, B. B. (2021). Intrinsic and extrinsic factors interact during development to influence telomere length in a long-lived reptile. *Molecular Ecology* **31**, 6114–6127.

Baerlocher, G. M., Mak, J., Röth, A., Rice, K. S. & Lansdorp, P. M. (2003). Telomere shortening in leukocyte subpopulations from baboons. *Journal of Leukocyte Biology* **73**, 289–296.

Baerlocher, G. M., Rice, K., Vulto, I. & Lansdorp, P. M. (2007). Longitudinal data on telomere length in leukocytes from newborn baboons support a marked drop in stem cell turnover around 1 year of age. *Aging Cell* **6**, 121–123.

*Baitchman, E. J. & Kollias, G. V. (2000). Clinical anatomy of the North American river otter (*Lontra canadensis*). *Journal of Zoo and Wildlife Medicine* **31**, 473–483.

Balistreri, C. R., Candore, G., Accardi, G., Colonna-Romano, G. & Lio, D. (2013). NF-κB pathway activators as potential ageing biomarkers: targets for new therapeutic strategies. *Immunity and Ageing* **10**, 1–16.

Barrett, E. L. B., Burke, T. A., Hammers, M., Komdeur, J. & Richardson, D. S. (2013). Telomere length and dynamics predict mortality in a wild longitudinal study. *Molecular Ecology* **22**, 249–259.

Bauch, C., Becker, P. H. & Verhulst, S. (2014). Within the genome, long telomeres are more informative than short telomeres with respect to fitness components in a long-lived seabird. *Molecular Ecology* **23**, 300–310.

Bauch, C., Boonekamp, J. J., Korsten, P., Mulder, E. & Verhulst, S. (2021). High heritability of telomere length and low heritability of telomere shortening in wild birds. *Molecular Ecology* **31**, 6308–6323.

Bauch, C., Gatt, M. C., Granadeiro, J. P., Verhulst, S. & Catry, P. (2020). Sex-specific telomere length and dynamics in relation to age and reproductive success in Cory's shearwaters. *Molecular Ecology* **29**, 1344–1357.

Bauch, C., Gatt, M. C., Verhulst, S., Granadeiro, J. P. & Catry, P. (2022). Higher mercury contamination is associated with shorter telomeres in a long-lived seabird – a direct effect or a consequence of among-individual variation in phenotypic quality? *Science of the Total Environment* **839**, 156359.

Bauer, C. M. C. M., Graham, J. L. J. L., Abolins-Abols, M., Heidinger, B. J. B. J., Ketterson, E. D. E. D. & Greives, T. J. T. J. (2018). Chronological and biological age predict seasonal reproductive timing: an investigation of clutch initiation and telomeres in birds of known age. *The American Naturalist* **191**, 777–782.

Beal, A. P., Kiszka, J. J., Wells, R. S. & Eirin-Lopez, J. M. (2019). The bottlenose dolphin epigenetic aging tool (BEAT): a molecular age estimation tool for small cetaceans. *Frontiers in Marine Science* **6**, 561.

Beirne, C., Delahay, R., Hares, M. & Young, A. (2014). Age-related declines and disease-associated variation in immune cell telomere length in a wild mammal. *PLoS One* **9**, e108964.

Belant, J. L., Etter, D. R., Friedrich, P. D., Cosgrove, M. K., Williams, B. W. & Scribner, K. T. (2011). Comparison of techniques for sex determination of American martens. *Journal of Wildlife Management* **75**, 256–260.

Bell, J. T., Tsai, P. C., Yang, T. P., Pidsley, R., Nisbet, J., Glass, D., Mangino, M., Zhai, G., Zhang, F., Valdes, A., Shin, S. Y., Dempster, E. L., Murray, R. M., Grundberg, E., Hedman, A. K., *et al.* (2012). Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genetics* **8**, e1002629.

Bender, H. S., Murchison, E. P., Pickett, H. A., Deakin, J. E., Strong, M. A., Conlan, C., McMillan, D. A., Neumann, A. A., Greider, C. W., Hannon, G. J., Reddel, R. R. & Graves, J. A. M. (2012). Extreme telomere length dimorphism in the Tasmanian devil and related marsupials suggests parental control of telomere length. *PLoS One* **7**, e46195.

Bird, A. (2007). Perceptions of epigenetics. *Nature* **447**, 396–398.

Bize, P., Criscuolo, F., Metcalfe, N. B., Nasir, L. & Monaghan, P. (2009). Telomere dynamics rather than age predict life expectancy in the wild. *Proceedings of the Royal Society B: Biological Sciences* **276**, 1679–1683.

Blévin, P., Angelier, F., Tartu, S., Ruault, S., Bustamante, P., Herzke, D., Moe, B., Bech, C., Gabrielsen, G. W., Bustnes, J. O. & Chastel, O. (2016). Exposure to oxychlordane is associated with shorter telomeres in Arctic breeding kittiwakes. *Science of the Total Environment* **563**, 125–130.

Bors, E. K., Baker, C. S., Wade, P. R., O'Neill, K. B., Shelden, K. E. W., Thompson, M. J., Fei, Z., Jarman, S. & Horvath, S. (2021). An epigenetic clock to estimate the age of living beluga whales. *Evolutionary Applications* **14**, 1263–1273.

Bowen, W. D., Sergeant, D. E. & Oritsland, T. (1983). Validation of age estimation in the harp seal, *Phoca groenlandica*, using dentinal annuli. *Canadian Journal of Fisheries and Aquatic Sciences* **40**, 1430–1441.

Bradley, R. J. & Safran, R. J. (2014). Conceptual revision and synthesis of proximate factors associated with age-related improvement in reproduction. *Ethology* **120**, 411–426.

Brandenburg, J. (2016). Persistent effects of early life exposure to polycyclic aromatic hydrocarbons on DNA methylation in birds. Masters dissertation: McGill University, Montreal. https://escholarship.mcgill.ca/concern/theses/xd07gw307.

Brelsford, A., Milá, B. & Irwin, D. E. (2011). Hybrid origin of Audubon's warbler. *Molecular Ecology* **20**, 2380–2389.

Brook, C. E., Ranaivoson, H. C., Andriafidison, D., Ralisata, M., Razafimanahaka, J., Héraud, J. M., Dobson, A. P. & Metcalf, C. J. (2019). Population trends for two Malagasy fruit bats. *Biological Conservation* **234**, 165–171.

Brown, P., Green, C., Sivakumaran, K. P., Stoessel, D. & Giles, A. (2011). Validating otolith annuli for annual age determination of common carp. *Transactions of the American Fisheries Society* **133**, 190–196.

Brümmendorf, T. H., Mak, J., Sabo, K. M., Baerlocher, G. M., Dietz, K., Abkowitz, J. L. & Lansdorp, P. M. (2002). Longitudinal studies of telomere length in feline blood cells: implications for hematopoietic stem cell turnover in vivo. *Experimental Hematology* **30**, 1147–1152.

Buddhachat, K., Kriangwanich, W., Kumoun, S., Brown, J. L., Chailangkarn, S., Somgird, C., Thitaram, C., Prasitwattanaseree, S. & Nganvongpanit, K. (2017). Telomeric attrition with increasing age in short- (Chihuahua dog) and long- (Asian elephant) life span animals. *Kafkas Universitesi Veteriner Fakultesi Dergisi* **23**, 643–649.

Burraco, P., Comas, M., Reguera, S., Zamora-Camacho, F. J. & Moreno-Rueda, G. (2020). Telomere length mirrors age structure along a 2200-m altitudinal gradient in a Mediterranean lizard. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* **247**, 110741.

Caccavo, J. A., Raclot, T., Poupart, T., Ropert-Coudert, Y. & Angelier, F. (2021). Anthropogenic activities are associated with shorter telomeres in chicks of Adélie penguin (*Pygoscelis adeliae*). *Polar Biology* **44**, 1391–1399.

Cailliet, G. M. (1990). Elasmobranch age determination and verification: an updated review Elasmobranch age determination and verification: an updated review. In *Elasmobranchs as Living Resources: Advances in the Biology, Ecology, Systematics, and the Status of Fisheries. National Oceanic and Atmospheric Administration Technical Report* **90**, pp. 157–165. NOAA, Seattle, USA.

Campana, S. E. (2001). Accuracy, precision and quality control in age determination, including a review of the use and abuse of age validation methods. *Journal of Fish Biology* **59**, 197–242.

Cantrell, B., Friedman, S., Lachance, H., Bernier, C., Murdoch, B., Frattini, S., Talenti, A., Crepaldi, P. & McKay, S. (2020). A novel understanding of global DNA methylation in bobcat (*Lynx rufus*). *Genome* **63**, 125–130.

Carroll, E. L., Bruford, M. W., DeWoody, J. A., Leroy, G., Strand, A., Waits, L. & Wang, J. (2018). Genetic and genomic monitoring with minimally invasive sampling methods. *Evolutionary Applications* **11**, 1094–1119.

*Casey, G. A. & Webster, W. A. (1975). Age and sex determination of striped skunks (*Mephitis mephitis*) from Ontario, Manitoba, and Quebec. *Canadian Journal of Zoology* **53**, 223–226.

Cavalli, G. & Heard, E. (2019). Advances in epigenetics link genetics to the environment and disease. *Nature* **571**, 489–499.

Cerchiara, J. A., Risques, R. A., Prunkard, D., Smith, J. R., Kane, O. J. & Boersma, P. D. (2017). Magellanic penguin telomeres do not shorten with age with increased reproductive effort, investment, and basal corticosterone. *Ecology and Evolution* **7**, 5682–5691.

Cesare, A. J. & Reddel, R. R. (2010). Alternative lengthening of telomeres: models, mechanisms and implications. *Nature Reviews Genetics* **11**, 319–330.

Chaney, R. C., Blemings, K. P., Bonner, J. & Klandorf, H. (2003). Pentosidine as a measure of chronological age in wild birds. *Auk* **120**, 394–399.

Charpentier, M. J. E., Tung, J., Altmann, J. & Alberts, S. C. (2008). Age at maturity in wild baboons: genetic, environmental and demographic influences. *Molecular Ecology* **17**, 2026–2040.

Chédin, F. (2011). The DNMT3 family of mammalian *de novo* DNA methyltransferases. *Progress in Molecular Biology and Translational Science* **101**, 255–285.

Cherdsukjai, P., Buddhachat, K., Brown, J., Kaewkool, M., Poommouang, A., Kaewmong, P., Kittiwattanawong, K. & Nganvongpanit, K. (2020). Age relationships with telomere length, body weight and body length in wild dugong (*Dugong dugon*). *PeerJ* **8**, e10319.

Christensen-Dalsgaard, S. N., Aars, J., Andersen, M., Lockyer, C. & Yoccoz, N. G. (2010). Accuracy and precision in estimation of age of Norwegian Arctic polar bears (*Ursus maritimus*) using dental cementum layers from known-age individuals. *Polar Biology* **33**, 589–597.

Clark, S. T., Odell, D. K. & Lacinak, C. T. (2000). Aspects of growth in captive killer whales (*Orcinus orca*). *Marine Mammal Science* **16**, 110–123.

Cleal, K., Norris, K. & Baird, D. (2018). Telomere length dynamics and the evolution of cancer genome architecture. *International Journal of Molecular Sciences* **19**, 1–17.

Clerc, M., Fenton, A., Babayan, S. A. & Pedersen, A. B. (2019). Parasitic nematodes simultaneously suppress and benefit from coccidian coinfection in their natural mouse host. *Parasitology* **146**, 1096–1106.

Clutton-Brock, T. & Sheldon, B. C. (2010). Individuals and populations: the role of long-term, individual-based studies of animals in ecology and evolutionary biology. *Trends in Ecology and Evolution* **25**, 562–573.
Cocci, P., Mosconi, G., Bracchetti, L., Nalocca, J. M., Frapiccini, E., Marini, M., Caprioli, G., Sagratini, G. & Palermo, F. A. (2018). Investigating the potential impact of polycyclic aromatic hydrocarbons (PAHs) and polychlorinated biphenyls (PCBs) on gene biomarker expression and global DNA methylation in loggerhead sea turtles (*Caretta caretta*) from the Adriatic Sea. *Science of the Total Environment* **619**, 49–57.
Cook, P. E., Hugo, L. E., Iturbe-Ormaetxe, I., Williams, C. R., Chenoweth, S. F., Ritchie, S. A., Ryan, P. A., Kay, B. H., Blows, M. W. & O'Neill, S. L. (2006). The use of transcriptional profiles to predict adult mosquito age under field conditions. *Proceedings of the National Academy of Sciences* **103**, 18060–18065.
Cossette, M.-L., Stewart, D. T., Haghani, A., Zoller, J. A., Shafer, A. B. A. & Horvath, S. (2022). Epigenetics and Island-mainland divergence in an insectivorous small mammal. *Molecular Ecology* **32**, 152–166.
Craighead, J. J., Craighead, F. C. & McCutchen, H. E. (1970). Age determination of grizzly bears from fourth premolar tooth sections. *The Journal of Wildlife Management* **34**, 353.
Cram, D. L., Monaghan, P., Gillespie, R., Dantzer, B., Duncan, C., Spence-Jones, H. & Clutton-Brock, T. (2018). Rank-related contrasts in longevity arise from extra-group excursions not delayed senescence in a cooperative mammal. *Current Biology* **28**, 2934–2939.e4.
Criscuolo, F., Dobson, F. S. & Schull, Q. (2021). The influence of phylogeny and life history on telomere lengths and telomere rate of change among bird species: a meta-analysis. *Ecology and Evolution* **11**, 12908–12922.
Criscuolo, F., Zahn, S. & Bize, P. (2017). Offspring telomere length in the long lived Alpine swift is negatively related to the age of their biological father and foster mother. *Biology Letters* **13**, 20170188.
Crowe, D. M. (1972). The presence of annuli in bobcat tooth cementum layers. *The Journal of Wildlife Management* **36**, 1330.
*Crowe, D. M. & Strickland, M. D. (1975). Dental annulation in the American badger. *Journal of Mammalogy* **56**, 269–272.
Dalton, S. R. & Bellacosa, A. (2012). DNA demethylation by TDG. *Epigenomics* **4**, 459–467.
Dantzer, B. & Fletcher, Q. E. (2015). Telomeres shorten more slowly in slow-aging wild animals than in fast-aging ones. *Experimental Gerontology* **71**, 38–47.
Dantzer, B., van Kesteren, F., Westrick, S. E., Boutin, S., McAdam, A. G., Lane, J. E., Gillespie, R., Majer, A., Haussmann, M. F. & Monaghan, P. (2020). Maternal glucocorticoids promote offspring growth without inducing oxidative stress or shortening telomeres in wild red squirrels. *Journal of Experimental Biology* **223**, jeb212373.
de la Iglesia, G. (2018). Telomere length as a biomarker for stress in eastern bluebird nestlings (*Sialis sialis*) reared in urban habitats. Masters dissertation: Western Carolina University: Cullowhee. http://libres.uncg.edu/ir/wcu/listing.aspx?id=23338.
de Lange, T., Shiue, L., Myers, R. M., Cox, D. R., Naylor, S. L., Killery, A. M. & Varmus, H. E. (1990). Structure and variability of human chromosome ends. *Molecular and Cellular Biology* **10**, 518–527.
De Paoli-Iseppi, R., Deagle, B. E., McMahon, C. R., Hindell, M. A., Dickinson, J. L. & Jarman, S. N. (2017*b*). Measuring animal age with DNA methylation: from humans to wild animals. *Frontiers in Genetics* **8**, 2010–2017.
De Paoli-Iseppi, R., Deagle, B. E., Polanowski, A. M., McMahon, C. R., Dickinson, J. L., Hindell, M. A. & Jarman, S. N. (2019). Age estimation in a long-lived seabird (*Ardenna tenuirostris*) using DNA methylation-based biomarkers. *Molecular Ecology Resources* **19**, 411–425.
De Paoli-Iseppi, R., Polanowski, A. M., McMahon, C., Deagle, B. E., Dickinson, J. L., Hindell, M. A. & Jarman, S. N. (2017*a*). DNA methylation levels in candidate genes associated with chronological age in mammals are not conserved in a long-lived seabird. *PLoS One* **12**, 1–20.
Degrandi, T. M., Barcellos, S. A., Costa, A. L., Garnero, A. D. V., Hass, I. & Gunski, R. J. (2020). Introducing the bird chromosome database: an overview of cytogenetic studies in birds. *Cytogenetic and Genome Research* **160**, 199–205.
Del Re, A. C. (2013). compute.es: Compute Effect Sizes. *R Package*. Electronic file available at https://cran.r-project.org/package=compute.es Accessed 31.10.2022.
Denham, J., Stevenson, K. & Denham, M. M. (2019). Age-associated telomere shortening in thoroughbred horses. *Experimental Gerontology* **127**, 110718.
*Dietz, R., Heide-Jørgensen, M. P., Härkönen, T., Teilmann, J. & Valentin, N. (1991). Age determination of European harbour seal, *Phoca vitulina* L. *Sarsia* **76**, 17–21.
*Divljan, A., Parray-Jones, K. & Wardle, G. M. (2006). Age determination in the grey-headed flying fox. *Journal of Wildlife Management* **70**, 607–611.
Dorado-Correa, A. M., Zollinger, S. A., Heidinger, B. & Brumm, H. (2018). Timing matters: traffic noise accelerates telomere loss rate differently across developmental stages. *Frontiers in Zoology* **15**, 1–8.
Double, M. C., Dawson, D., Burke, T. & Cockburn, A. (1997). Finding the fathers in the least faithful bird: a microsatellite-based genotyping system for the superb fairy-wren *Malurus cyaneus*. *Molecular Ecology* **6**, 691–693.
Dunshea, G., Duffield, D., Gales, N., Hindell, M., Wells, R. S. & Jarman, S. N. (2011). Telomeres as age markers in vertebrate molecular ecology. *Molecular Ecology Resources* **11**, 225–235.
Dutra, L., Souza, F., Friberg, I., Araújo, M., Vasconcellos, A. & Young, R. (2020). Validating the use of oral swabs for telomere length assessment in dogs. *Journal of Veterinary Behavior* **40**, 16–20.
*Earle, R. D. & Kramm, K. R. (1980). Techniques for age determination in the Canadian porcupine. *The Journal of Wildlife Management* **44**, 413.
Eastwood, J. R., Hall, M. L., Teunissen, N., Kingma, S. A., Hidalgo Aranzamendi, N., Fan, M., Roast, M., Verhulst, S. & Peters, A. (2019). Early-life telomere length predicts lifespan and lifetime reproductive success in a wild bird. *Molecular Ecology* **28**, 1127–1137.
Eisenberg, D. T. A., Tackney, J., Cawthon, R. M., Cloutier, C. T. & Hawkes, K. (2017). Paternal and grandpaternal ages at conception and descendant telomere lengths in chimpanzees and humans. *American Journal of Physical Anthropology* **162**, 201–207.
Erdem, E., Özbaşer, F. T., Gürcan, E. K. & Soysal, M. I. (2021). The morphological and morphometric characteristics of Alabadem pigeons. *Turkish Journal of Veterinary and Animal Sciences* **45**, 372–379.
Fairlie, J., Holland, R., Pilkington, J. G., Pemberton, J. M., Harrington, L. & Nussey, D. H. (2016). Lifelong leukocyte telomere dynamics and survival in a free-living mammal. *Aging Cell* **15**, 140–148.
Faul, F., Erdfelder, E., Buchner, A. & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods* **41**, 1149–1160.
Fick, L. J., Fick, G. H., Li, Z., Cao, E., Bao, B., Heffelfinger, D., Parker, H. G., Ostrander, E. A. & Riabowol, K. (2012). Telomere length correlates with life span of dog breeds. *Cell Reports* **2**, 1530–1536.
Fitz-James, M. H. & Cavalli, G. (2022). Molecular mechanisms of transgenerational epigenetic inheritance. *Nature Reviews Genetics* **23**, 325–341.
Fohringer, C., Hoelzl, F., Allen, A. M., Cayol, C., Ericsson, G., Spong, G., Smith, S. & Singh, N. J. (2022). Large mammal telomere length variation across ecoregions. *BMC Ecology and Evolution* **22**, 1–10.
Foley, N. M., Hughes, G. M., Huang, Z., Clarke, M., Jebb, D., Whelan, C. V., Petit, E. J., Touzalin, F., Farcy, O., Jones, G., Ransome, R. D., Kacprzyk, J., O'Connell, M. J., Kerth, G., Rebelo, H., *et al.* (2018). Growing old, yet staying young: the role of telomeres in bats' exceptional longevity. *Science Advances* **4**, eaao0926.
Foley, N. M., Petit, E. J., Brazier, T., Finarelli, J. A., Hughes, G. M., Touzalin, F., Puechmaille, S. J. & Teeling, E. C. (2020). Drivers of longitudinal telomere dynamics in a long-lived bat species, *Myotis myotis*. *Molecular Ecology* **29**, 2963–2977.
Friesen, C. R., Wapstra, E. & Olsson, M. (2021). Of telomeres and temperature: measuring thermal effects on telomeres in ectothermic animals. *Molecular Ecology* **31**, 6069–6086.
Froy, H., Bird, E. J., Wilbourn, R. V., Fairlie, J., Underwood, S. L., Salvo-Chirnside, E., Pilkington, J. G., Bérénos, C., Pemberton, J. M. & Nussey, D. H. (2017). No evidence for parental age effects on offspring leukocyte telomere length in free-living Soay sheep. *Scientific Reports* **7**, 9991.
Froy, H., Underwood, S. L., Dorrens, J., Seeker, L. A., Watt, K., Wilbourn, R. V., Pilkington, J. G., Harrington, L., Pemberton, J. M. & Nussey, D. H. (2021). Heritable variation in telomere length predicts mortality in Soay sheep. *Proceedings of the National Academy of Sciences of the United States of America* **118**, e2020563118.
Garde, E., Frie, A. K., Dunshea, G., Hansen, S. H., Kovacs, K. M. & Lydersen, C. (2010). Harp seal ageing techniques—teeth, aspartic acid racemization, and telomere sequence analysis. *Journal of Mammalogy* **91**, 1365–1374.
Gomes, N. M. V., Ryder, O. A., Houck, M. L., Charter, S. J., Walker, W., Forsyth, N. R., Austad, S. N., Venditti, C., Pagel, M., Shay, J. W. & Wright, W. E. (2011). Comparative biology of mammalian telomeres: hypotheses on ancestral states and the roles of telomeres in longevity determination. *Aging Cell* **10**, 761–768.
Gomez, A., Karlgren, M., Edler, D., Bernal, M. L., Mkrtchian, S. & Ingelman-Sundberg, M. (2007). Expression of CYP2W1 in colon tumors: regulation by gene methylation. *Pharmacogenomics* **8**, 1315–1325.
Gopalan, S., Carja, O., Fagny, M., Patin, E., Myrick, J. W., McEwen, L. M., Mah, S. M., Kobor, M. S., Froment, A., Feldman, M. W., Quintana-Murci, L. & Henn, B. M. (2017). Trends in DNA methylation with age replicate across diverse human populations. *Genetics* **206**, 1659–1674.
Goren, A. D., Brodie, P. F., Spotte, S., Ray, G. C., Kaufman, H. W., Gwinnett, A. J., Sciubba, J. J. & Buck, J. D. (1987). Growth layer groups (GLGs) in the teeth of an adult beluga whale (*Delphinapterus leucas*) of known age: evidence for two annual layers. *Marine Mammal Science* **3**, 14–21.

*Gosselink, T. E., van Deelen, T. R., Warner, R. E. & Joselyn, M. G. (2003). Temporal habitat partitioning and spatial use of coyotes and red foxes in east-Central Illinois. *The Journal of Wildlife Management* **67**, 90.

*Gourichon, L. & Parmigiani, V. (2016). Preliminary analysis of dental cementum of *Lama guanicoe* for the estimation of age and season at death: studies of modern specimens and further archaeological applications. *Journal of Archaeological Science: Reports* **6**, 856–861.

Graham, J. L. (2018). Reproductive timing in a changing world: understanding mechanisms and costs associated with reproductive timing decisions. Doctoral Thesis: North Dakota State University of Agriculture and Applied Science Fargo. https://hdl.handle.net/10365/27478.

*Grau, G. A., Sanderson, G. C. & Rogers, J. P. (1970). Age determination of raccoons. *The Journal of Wildlife Management* **34**, 364.

Gregory, T. R. (2023). Animal genome size database. Electronic file available at http://www.genomesize.com Accessed 8.5.2023.

Greider, C. W. & Blackburn, E. H. (1985). Identification of a specific telomere terminal transferase activity in tetrahymena extracts. *Cell* **43**, 405–413.

Grimsdell, J. J. R. (1973). Age determination of the African buffalo, *Syncerus caffer* Sparrman. *African Journal of Ecology* **11**, 31–53.

Grue, H. & Jensen, B. (1979). Review of the formation of incremental lines in tooth cementum of terrestrial mammals. *Danish Review of Game Biology* **11**, 1–48.

Grunst, M. L., Grunst, A. S., Pinxten, R. & Eens, M. (2020). Anthropogenic noise is associated with telomere length and carotenoid-based coloration in free-living nestling songbirds. *Environmental Pollution* **260**, 114032.

Guevara, E. E., Lawler, R. R., Staes, N., White, C. M., Sherwood, C. C., Ely, J. J., Hopkins, W. D. & Bradley, B. J. (2020). Age-associated epigenetic change in chimpanzees and humans: chimpanzee epigenetic age. *Philosophical Transactions of the Royal Society B: Biological Sciences* **375**, 20190616.

Guo, F., Li, X., Liang, D., Li, T., Zhu, P., Guo, H., Wu, X., Wen, L., Gu, T. P., Hu, B., Walsh, C. P., Li, J., Tang, F. & Xu, G. L. (2014). Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. *Cell Stem Cell* **15**, 447–459.

Haddaway, N. R., Page, M. J., Pritchard, C. C. & McGuinness, L. A. (2022). PRISMA2020: an R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and open synthesis. *Campbell Systematic Reviews* **18**, e1230.

Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S. V., Klotzle, B., Bibikova, M., Fan, J. B., Gao, Y., Deconde, R., Chen, M., Rajapakse, I., Friend, S., Ideker, T., *et al.* (2013). Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular Cell* **49**, 359–367.

Hansen, E., Skotnes, T., Bustnes, J. O., Helander, B., Eulaers, I., Sun, J., Covaci, A., Bårdsen, B. J., Zahn, S., Criscuolo, F. & Bourgeon, S. (2022). Telomere length in relation to persistent organic pollutant exposure in white-tailed eagle (*Haliaeetus albicilla*) nestlings from Sweden sampled in 1995–2013. *Environmental Research* **208**, 112712.

Hartmann, N., Reichwald, K., Lechel, A., Graf, M., Kirschner, J., Dorn, A., Terzibasi, E., Wellner, J., Platzer, M., Rudolph, K. L., Cellerino, A. & Englert, C. (2009). Telomeres shorten while Tert expression increases during ageing of the short-lived fish *Nothobranchius furzeri*. *Mechanisms of Ageing and Development* **130**, 290–296.

Hatase, H., Sudo, R., Watanabe, K. K., Kasugai, T., Saito, T., Okamoto, H., Uchida, I. & Tsukamoto, K. (2008). Shorter telomere length with age in the loggerhead turtle: a new hope for live sea turtle age estimation. *Genes and Genetic Systems* **83**, 423–426.

Haussmann, M. F. & Vleck, C. M. (2002). Telomere length provides a new technique for aging animals. *Oecologia* **130**, 325–328.

Haussmann, M. F., Vleck, C. M. & Nisbet, I. C. T. (2003). Calibrating the telomere clock in common terns, *Sterna hirundo*. *Experimental Gerontology* **38**, 787–789.

Heidinger, B. J., Blount, J. D., Boner, W., Griffiths, K., Metcalfe, N. B. & Monaghan, P. (2012). Telomere length in early life predicts lifespan. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 1743–1748.

Higgs, D. M., Souza, M. J., Wilkins, H. R., Presson, J. C. & Popper, A. N. (2002). Age- and size-related changes in the inner ear and hearing ability of the adult zebrafish (*Danio rerio*). *Journal of the Association for Research in Otolaryngology* **3**, 174–184.

Hiller, T. L. & Tyre, A. J. (2014). Comparison of two age-estimation techniques for cougars. *Northwestern Naturalist* **95**, 77–82.

Hoelzl, F., Cornils, J. S., Smith, S., Moodley, Y. & Ruf, T. (2016*a*). Telomere dynamics in free-living edible dormice (*Glis glis*): the impact of hibernation and food supply. *Journal of Experimental Biology* **219**, 2469–2474.

Hoelzl, F., Smith, S., Cornils, J. S., Aydinonat, D., Bieber, C. & Ruf, T. (2016*b*). Telomeres are elongated in older individuals in a hibernating rodent, the edible dormouse (*Glis glis*). *Scientific Reports* **6**, 36856.

Hohn, A. A. (1980). Age determination and age related factors in the teeth of Western North Atlantic bottlenose dolphins. *Scientific Reports of the Whales Research Institute* **32**, 39–66.

Høie, H., Andersson, C., Folkvord, A. & Karlsen, Ø. (2004). Precision and accuracy of stable isotope signals in otoliths of pen-reared cod (*Gadus morhua*) when sampled will a high-resolution micromill. *Marine Biology* **144**, 1039–1049.

Hong, S. R., Shin, K. J., Jung, S. E., Lee, E. H. & Lee, H. Y. (2019). Platform-independent models for age prediction using DNA methylation data. *Forensic Science International: Genetics* **38**, 39–47.

Horn, T., Gemmell, N. J., Robertson, B. C. & Bridges, C. R. (2008). Telomere length change in European sea bass (*Dicentrarchus labrax*). *Australian Journal of Zoology* **56**, 207–210.

Horn, T., Robertson, B. C. & Gemmell, N. J. (2010). The use of telomere length in ecology and evolutionary biology. *Heredity* **105**, 497–506.

Horn, T., Robertson, B. C., Will, M., Eason, D. K., Elliott, G. P. & Gemmell, N. J. (2011). Inheritance of telomere length in a bird. *PLoS One* **6**, e17199.

Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biology* **14**, 3156.

Horvath, S., Haghani, A., Macoretta, N., Ablaeva, J., Zoller, J. A., Li, C. Z., Zhang, J., Takasugi, M., Zhao, Y., Rydkina, E., Zhang, Z., Emmrich, S., Raj, K., Seluanov, A., Faulkes, C. G., *et al.* (2022*a*). DNA methylation clocks tick in naked mole rats but queens age more slowly than nonbreeders. *Nature Aging* **2**, 46–59.

Horvath, S., Haghani, A., Peng, S., Hales, E. N., Zoller, J. A., Raj, K., Larison, B., Robeck, T. R., Petersen, J. L., Bellone, R. R. & Finno, C. J. (2022*b*). DNA methylation aging and transcriptomic studies in horses. *Nature Communications* **13**, 1–13.

Horvath, S., Haghani, A., Zoller, J. A., Fei, Z., Bérubé, M. & Robbins, J. (2022*c*). DNA methylation age studies of humpback whales. *bioRxiv* 2022.08.15.503952. https://doi.org/10.1101/2022.08.15.503952.

Horvath, S., Haghani, A., Zoller, J. A., Raj, K., Sinha, I., Robeck, T. R., Black, P., Couzens, A., Lau, C., Manoyan, M., Ruiz, Y. A., Talbott, A., Belov, K., Hogg, C. J. & Sears, K. E. (2022*d*). Epigenetic clock and methylation studies in marsupials: opossums, Tasmanian devils, kangaroos, and wallabies. *GeroScience* **44**, 1825–1845.

Horvath, S., Lu, A. T., Haghani, A., Zoller, J. A., Li, C. Z., Lim, A. R., Brooke, R. T., Raj, K., Serres-Armero, A., Dreger, D. L., Hogan, A. N., Plassais, J. & Ostrander, E. A. (2022*e*). DNA methylation clocks for dogs and humans. *Proceedings of the National Academy of Sciences of the United States of America* **119**, e2120887119.

Horvath, S. & Raj, K. (2018). DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nature Reviews Genetics* **19**, 371–384.

Horvath, S., Zoller, J. A., Haghani, A., Jasinska, A. J., Raj, K., Breeze, C. E., Ernst, J., Vaughan, K. L. & Mattison, J. A. (2021*a*). Epigenetic clock and methylation studies in the rhesus macaque. *GeroScience* **43**, 2441–2453.

Horvath, S., Zoller, J. A., Haghani, A., Lu, A. T., Raj, K., Jasinska, A. J., Mattison, J. A. & Salmon, A. B. (2021*b*). DNA methylation age analysis of rapamycin in common marmosets. *GeroScience* **43**, 2413–2425.

Ibáñez-Álamo, J. D., Pineda-Pampliega, J., Thomson, R. L., Aguirre, J. I., Díez-Fernández, A., Faivre, B., Figuerola, J. & Verhulst, S. (2018). Urban blackbirds have shorter telomeres. *Biology Letters* **14**, 20180083.

Ineson, K. M., O'Shea, T. J., Kilpatrick, C. W., Parise, K. L. & Foster, J. T. (2020). Ambiguities in using telomere length for age determination in two North American bat species. *Journal of Mammalogy* **101**, 958–969.

Isermann, D. A., Wolter, M. H. & Breeggemann, J. J. (2010). Estimating black crappie age: an assessment of dorsal spines and scales as nonlethal alternatives to otoliths. *North American Journal of Fisheries Management* **30**, 1591–1598.

Ito, G., Yoshimura, K. & Momoi, Y. (2017). Analysis of DNA methylation of potential age-related methylation sites in canine peripheral blood leukocytes. *Journal of Veterinary Medical Science* **79**, 745–750.

Ito, H., Udono, T., Hirata, S. & Inoue-Murayama, M. (2018). Estimation of chimpanzee age based on DNA methylation. *Scientific Reports* **8**, 1–5.

Izzo, C. (2010). Patterns of telomere length change with age in aquatic vertebrates and the phylogenetic distribution of the pattern among jawed vertebrates. Doctoral Thesis: University of Adelaide, Adelaide. https://hdl.handle.net/2440/63477.

Izzo, C., Bertozzi, T., Gillanders, B. M. & Donnellan, S. C. (2014). Variation in telomere length of the common carp, *Cyprinus carpio* (Cyprinidae), in relation to body length. *Copeia* **2014**, 87–94.

Izzo, C., Hamer, D. J., Bertozzi, T., Donnellan, S. C. & Gillanders, B. M. (2011). Telomere length and age in pinnipeds: the endangered Australian sea lion as a case study. *Marine Mammal Science* **27**, 841–851.

Jarman, S. N., Polanowski, A. M., Faux, C. E., Robbins, J., De Paoli-Iseppi, R., Bravington, M. & Deagle, B. E. (2015). Molecular biomarkers for chronological age in animal ecology. *Molecular Ecology* **24**, 4826–4847.

Jasinska, A. J., Haghani, A., Zoller, J. A., Li, C. Z., Arneson, A., Ernst, J., Kavanagh, K., Jorgensen, M. J., Mattison, J. A., Wojta, K., Choi, O. W., DeYoung, J., Li, X., Rao, A. W., Coppola, G., *et al.* (2022). Epigenetic clock and methylation studies in vervet monkeys. *GeroScience* **44**, 699–717.

Johansson, Å., Enroth, S. & Gyllensten, U. (2013). Continuous aging of the human DNA methylome throughout the human lifespan. *PLoS One* **8**, 67378.
Jones, M. J., Goodman, S. J. & Kobor, M. S. (2015). DNA methylation and healthy human aging. *Aging Cell* **14**, 924–932.
Jones, O. R., Scheuerlein, A., Salguero-Gómez, R., Camarda, C. G., Schaible, R., Casper, B. B., Dahlgren, J. P., Ehrlén, J., García, M. B., Menges, E. S., Quintana-Ascencio, P. F., Caswell, H., Baudisch, A. & Vaupel, J. W. (2014). Diversity of ageing across the tree of life. *Nature* **505**, 169–173.
Juola, F. A., Haussmann, M. F., Dearborn, D. C. & Vleck, C. M. (2006). Telomere shortening in a long-lived marine bird: cross-sectional analysis and test of an aging tool. *The Auk* **123**, 775–783.
Kagiwada, S., Kurimoto, K., Hirota, T., Yamaji, M. & Saitou, M. (2013). Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. *EMBO Journal* **32**, 340–353.
*Kamler, J. F. & Macdonald, D. W. (2006). Longevity of a wild bat-eared fox. *African Journal of Wildlife Research* **36**, 199–200.
Kang, H. (2021). Sample size determination and power analysis using the G*power software. *Journal of Educational Evaluation for Health Professions* **18**, 1–12.
Karere, G. M., Mahaney, M. C., Newman, D. E., Riojas, A. M., Christensen, C., Birnbaum, S., VandeBerg, J. L. & Cox, L. (2019). Diet-induced leukocyte telomere shortening in a baboon model for early stage atherosclerosis. *Scientific Reports* **9**, 1–9.
Kärkkäinen, T., Laaksonen, T., Burgess, M., Cantarero, A., Martínez-Padilla, J., Potti, J., Moreno, J., Thomson, R. L., Tilgar, V. & Stier, A. (2021). Population differences in the length and early-life dynamics of telomeres among European pied flycatchers. *Molecular Ecology* **31**, 5966–5978.
Kärkkäinen, T., Teerikorpi, P., Panda, B., Helle, S., Stier, A. & Laaksonen, T. (2019). Impact of continuous predator threat on telomere dynamics in parent and nestling pied flycatchers. *Oecologia* **191**, 757–766.
Kasuya, T. & Matsui, S. (1984). Age determination and growth of the short-finned pilot whale off the Pacific coast of Japan. *Scientific Reports of the Whales Research Institute Tokyo* **35**, 57–92.
Kazda, A., Zellinger, B., Rössler, M., Derboven, E., Kusenda, B. & Riha, K. (2012). Chromosome end protection by blunt-ended telomeres. *Genes and Development* **26**, 1703–1713.
*Kenyon, K. W. & Fiscus, C. H. (1963). Age determination in the Hawaiian monk seal. *Journal of Mammalogy* **44**, 280.
Kerepesi, C., Meer, M. V., Ablaeva, J., Amoroso, V. G., Lee, S. G., Zhang, B., Gerashchenko, M. V., Trapp, A., Yim, S. H., Lu, A. T., Levine, M. E., Seluanov, A., Horvath, S., Park, T. J., Gorbunova, V., *et al.* (2022). Epigenetic aging of the demographically non-aging naked mole-rat. *Nature Communications* **13**, 1–10.
Kim, K. S. & Sappington, T. W. (2013). Microsatellite data analysis for population genetics. In *Microsatellites: Methods in Molecular Biology* (Volume **1006**, ed. S. Kantartzi), pp. 271–295. Humana Press, Totowa.
Kimura, M. (1980). Variability in techniques of counting dentinal growth layer groups in a tooth of a known-age dolphin, Tursiops truncatus. *Reports of the International Whaling Commission, Special Issue* **3**, 161–163.
*Kingsmill, E. (1962). An investigation of criteria for estimating age in the marsupials *Trichosurus vulpecula* kerr and *Perameles nasuta* geoffroy. *Australian Journal of Zoology* **10**, 597–616.
Kipling, D. & Cooke, H. J. (1990). Hypervariable ultra-long telomeres in mice. *Nature* **347**, 400–402.
Kirby, R., Alldredge, M. W. & Pauli, J. N. (2017). Environmental, not individual, factors drive markers of biological aging in black bears. *Evolutionary Ecology* **31**, 571–584.
*Klevezal, G. A. & Pucek, Z. (1987). Bisoniana XCIII. Growth layers in tooth cement and dentine of European bison and its hybrids with domestic cattle. *Acta Theriologica* **32**, 115–128.
*Klevezal, G. A. & Stewart, B. S. (1994). Patterns and calibration of layering in tooth cementum of female northern elephant seals, *Mirounga angustirostris*. *Journal of Mammalogy* **75**, 483–487.
Kordowitzki, P., Haghani, A., Zoller, J. A., Li, C. Z., Raj, K., Spangler, M. L. & Horvath, S. (2021). Epigenetic clock and methylation study of oocytes from a bovine model of reproductive aging. *Aging Cell* **20**, e13349.
Kujoth, C. C., Hiona, A., Pugh, T. D., Someya, S., Panzer, K., Wohlgemuth, S. E., Hofer, T., Seo, A. Y., Sullivan, R., Jobling, W. A., Morrow, J. D., Van Remmen, H., Sedivy, J. M., Yamasoba, T., Tanokura, M., *et al.* (2005). Medicine: mitochondrial DNA mutations, oxidative stress, and apoptosis in mammalian aging. *Science* **309**, 481–484.
Kurdyukov, S. & Bullock, M. (2016). DNA methylation analysis: choosing the right method. *Biology* **5**, 3.
Lachance, H., Cantrell, B., Friedman, S., Mckay, S. & Yee, R. (2015). Genome-wide epigenetic survey of Vermont bobcat (*Lynx rufus*) and fisher (*Martes pennanti*). In *74th Annual Conference of the Northeast Association of Fish and Wildlife Agencies*. Northeast Association of Fish and Wildlife Agencies, Vermont.
Larison, B., Pinho, G. M., Hagani, A., Zoller, J. A., Li, C. Z., Finno, C. J., Farrell, C., Kaelin, C. B., Barsh, G. S., Wooding, B., Robeck, T. R., Maddox, D., Pellegrini, M. & Horvath, S. (2021*b*). Epigenetic models predict age and aging in plains zebras and other equids. *bioRxiv* 2021.03.29.437607. https://doi.org/10.1101/2021.03.29.437607.
Larison, B., Pinho, G. M., Haghani, A., Zoller, J. A., Li, C. Z., Finno, C. J., Farrell, C., Kaelin, C. B., Barsh, G. S., Wooding, B., Robeck, T. R., Maddox, D., Pellegrini, M. & Horvath, S. (2021*a*). Epigenetic models developed for plains zebras predict age in domestic horses and endangered equids. *Communications Biology* **4**, 1–9.
*Larter, N. C. & Allaire, D. G. (2016). Longevity and mortality of boreal woodland caribou (*Rangifer tarandus caribou*) of the dehcho region, northwest territories. *Canadian Field-Naturalist* **130**, 222–223.
*Laws, R. M. (1952). A new method of age determination for mammals. *Nature* **169**, 972–973.
*Laws, R. M. (1958). Growth rates and ages of crabeater seals, *Lobodon carcinophagus*. *Proceedings of the Zoological Society of London* **130**, 275–288.
Le Clercq, L. S., Dalton, D. L. & Kotzé, A. (2018). Molecular age estimation based on promotor CpG methylation using methylation sensitive PCR. In *Poster Presented at: 9th Annual Research Symposium of the National Zoological Gardens of South Africa*. SANBI, Pretoria. https://doi.org/10.13140/RG.2.2.11352.21768.
Lee, W. W., Nam, K. H., Terao, K. & Yoshikawa, Y. (2002). Age-related telomere length dynamics in peripheral blood mononuclear cells of healthy cynomolgus monkeys measured by flow FISH. *Immunology* **105**, 458–465.
Lemaître, J. F., Rey, B., Gaillard, J. M., Régis, C., Gilot-Fromont, E., Débias, F., Duhayer, J., Pardonnet, S., Pellerin, M., Haghani, A., Zoller, J. A., Li, C. Z. & Horvath, S. (2022). DNA methylation as a tool to explore ageing in wild roe deer populations. *Molecular Ecology Resources* **22**, 1002–1015.
Leonida, S. R. L., Bennett, N. C., Leitch, A. R. & Faulkes, C. G. (2020). Patterns of telomere length with age in African mole-rats: new insights from quantitative fluorescence *in situ* hybridisation (qFISH). *PeerJ* **8**, e10498.
Lewin, N., Treidel, L. A., Holekamp, K. E., Place, N. J. & Haussmann, M. F. (2015). Socioecological variables predict telomere length in wild spotted hyenas. *Biology Letters* **11**, 20140991.
Liebl, A. L., Wesner, J. S., Russell, A. F. & Schrey, A. W. (2021). Methylation patterns at fledging predict delayed dispersal in a cooperatively breeding bird. *PLoS One* **16**, e0252227.
Linck, E., Freeman, B. G. & Dumbacher, J. P. (2020). Speciation and gene flow across an elevational gradient in New Guinea kingfishers. *Journal of Evolutionary Biology* **33**, 1643–1652.
Lind, M. I. & Spagopoulou, F. (2018). Evolutionary consequences of epigenetic inheritance. *Heredity* **121**, 205–209.
Linhart, S. B. (1973). Age determination and occurrence of incremental growth lines in the dental cementum of the common vampire bat (*Desmodus rotundus*). *Journal of Mammalogy* **54**, 493–496.
Little, T. J., O'Toole, A. N., Rambaut, A., Chandra, T., Marioni, R. & Pedersen, A. B. (2020). Methylation-based age estimation in a wild mouse. *bioRxiv* 2020.07.16.203687. https://doi.org/10.1101/2020.07.16.203687.
Liu, L., Bailey, S. M., Okuka, M., Muñoz, P., Li, C., Zhou, L., Wu, C., Czerwiec, E., Sandler, L., Seyfang, A., Blasco, M. A. & Keefe, D. L. (2007). Telomere lengthening early in development. *Nature Cell Biology* **9**, 1436–1441.
Lockyer, C. H. (1993). A report on patterns of deposition of dentine and cement in teeth of pilot whales, genus *Globicephala*. In *Reports of the International Whaling Commission, Special Issue* (Volume **14**), pp. 137–161. IWC, Impington, UK.
Lockyer, C. H., Hohn, A. A., Doidge, W. D., Heide-Jørgensen, M. P. & Suydam, R. (2007). Age determination in belugas (*Delphinapterus leucas*): a quest for validation of dentinal layering. *Aquatic Mammals* **33**, 293–304.
*Lombaard, L. J. (1971). Age determination and growth curved in the black-backed jackal *Canis mesomelas* Schreber, 1775 (Carnivora: Canidae). *Annals of the Transvaal Museum* **27**, 135–169.
*Low, W. A. (1970). The influence of aridity on reproduction of the collared peccary (*Picotyles tajacu* (Linn)) in Texas. Doctoral Thesis: University of British Columbia, Vancouver. http://hdl.handle.net/2429/34781.
*Low, W. A. & Cowan, I. M. (1963). Age determination of deer by annular structure of dental cementum. *The Journal of Wildlife Management* **27**, 466.
Lowe, R., Barton, C., Jenkins, C. A. C. A., Ernst, C., Forman, O., Fernandez-Twinn, D. S. D. S., Bock, C., Rossiter, S. J. S. J., Faulkes, C. G. C. G., Ozanne, S. E. S. E., Walter, L., Odom, D. T., Mellersh, C. & Rakyan, V. K. V. K. (2018). Ageing-associated DNA methylation dynamics are a molecular readout of lifespan variation among mammalian species. *Genome Biology* **19**, 1–8.
Lowe, R., Danson, A. F., Rakyan, V. K., Yildizoglu, S., Saldmann, F., Viltard, M., Friedlander, G. & Faulkes, C. G. (2020). DNA methylation clocks as a predictor for ageing and age estimation in naked mole-rats, *Heterocephalus glaber*. *Aging* **12**, 4394–4406.

Lund, T. C., Glass, T. J., Tolar, J. & Blazar, B. R. (2009). Expression of telomerase and telomere length are unaffected by either age or limb regeneration in *Danio rerio*. *PLoS One* **4**, e7688.

Lundsgaard, N. U., Cramp, R. L. & Franklin, C. E. (2022). Early exposure to UV radiation causes telomere shortening and poorer condition later in life. *Journal of Experimental Biology* **225**, jeb243924.

Macchi, F., Edsinger, E. & Sadler, K. C. (2022). Epigenetic machinery is functionally conserved in cephalopods. *BMC Biology* **20**, 1–23.

Mäkinen, H., Viitaniemi, H. M., Visser, M. E., Verhagen, I., van Oers, K. & Husby, A. (2019). Temporally replicated DNA methylation patterns in great tit using reduced representation bisulfite sequencing. *Scientific Data* **6**, 1–7.

Marker, L. L. & Dickman, A. J. (2003). Morphology, physical condition, and growth of the cheetah (*Acinonyx jubatus jubatus*). *Journal of Mammalogy* **84**, 840–850.

Marker, L. L., Dickman, A. J., Jeo, R. M., Mills, M. G. L. & Macdonald, D. W. (2003). Demography of the Namibian cheetah, *Acinonyx jubatus jubatus*. *Biological Conservation* **114**, 413–425.

Marks, S. A. & Erickson, A. W. (1966). Age determination in the black bear. *The Journal of Wildlife Management* **30**, 389–410.

Marsh, H., Heinsohn, G. E. & Marsh, L. M. (1984). Breeding cycle, life history and population dynamics of the dugong, *Dugon dugon*. *Australian Journal of Zoology* **32**, 767–788.

Mayne, B., Berry, O. & Jarman, S. (2021*b*). Optimal sample size for calibrating DNA methylation age estimators. *Molecular Ecology Resources* **21**, 2316–2323.

Mayne, B., Espinoza, T., Roberts, D., Butler, G. L., Brooks, S., Korbie, D. & Jarman, S. (2021*a*). Nonlethal age estimation of three threatened fish species using DNA methylation: Australian lungfish, Murray cod and Mary River cod. *Molecular Ecology Resources* **21**, 2324–2332.

Mayne, B., Korbie, D., Kenchington, L., Ezzy, B., Berry, O. & Jarman, S. (2020*a*). A DNA methylation age predictor for zebrafish. *Aging* **12**, 24817–24835.

Mayne, B., Mustin, W., Baboolal, V., Casella, F., Ballorain, K., Barret, M., Vanderklift, M. A., Tucker, A. D., Korbie, D., Jarman, S. & Berry, O. (2022). Age prediction of green turtles with an epigenetic clock. *Molecular Ecology Resources* **22**, 2275–2284.

Mayne, B., Tucker, A. D., Berry, O. & Jarman, S. (2020*b*). Lifespan estimation in marine turtles using genomic promoter CpG density. *PLoS One* **15**, e0236888.

*Mbizah, M. M., Steenkamp, G. & Groom, R. J. (2016). Evaluation of the applicability of different age determination methods for estimating age of the endangered African wild dog (*Lycaon pictus*). *PLoS One* **11**, e0164676.

McCarthy, T. M., Fuller, T. K. & Munkhtsog, B. (2005). Movements and activities of snow leopards in southwestern Mongolia. *Biological Conservation* **124**, 527–537.

McClintock, B. (1941). The stability of broken ends of chromosomes in *Zea mays*. *Genetics* **26**, 234–282.

*McCutchen, H. E. (1969). Age determination of pronghorns by the incisor cementum. *The Journal of Wildlife Management* **33**, 172.

McEwen, L. M., Goodman, S. J., Kobor, M. S. & Jones, M. J. (2017). The DNA methylome: an interface between the environment, immunity, and ageing. In: *The Ageing Immune System and Health* pp. 35–52. Springer, Cham.

McKevitt, T. P., Nasir, L., Devlin, P. & Argyle, D. J. (2002). Telomere lengths in dogs decrease with increasing donor age. *Journal of Nutrition* **132**, 1604S–1606S.

McKevitt, T. P., Nasir, L., Wallis, C. V. & Argyle, D. J. (2003). A cohort study of telomere and telomerase biology in cats. *American Journal of Veterinary Research* **64**, 1496–1499.

*Mclaren, I. (1958). The biology of the ringed seal (*Phoca hispida* Schreber) in the eastern Canadian Arctic. *Bulletin of the Fisheries Research Board of Canada* **118**, 1–97.

Meillère, A., Brischoux, F., Ribout, C. & Angelier, F. (2015). Traffic noise exposure affects telomere length in nestling house sparrows. *Biology Letters* **11**, 20150559.

Mensà, E., Latini, S., Ramini, D., Storci, G., Bonafè, M. & Olivieri, F. (2019). The telomere world and aging: analytical challenges and future perspectives. *Ageing Research Reviews* **50**, 27–42.

Merlin, C. & Liedvogel, M. (2019). The genetics and epigenetics of animal migration and orientation: birds, butterflies and beyond. *The Journal of Experimental Biology* **222**, jeb191890.

Meyer, B. S., Moiron, M., Caswara, C., Chow, W., Fedrigo, O., Formenti, G., Haase, B., Howe, K., Mountcastle, J., Uliano-Silva, M., Wood, J., Jarvis, E. D., Liedvogel, M. & Bouwhuis, S. (2023). Sex-specific changes in autosomal methylation rate in ageing common terns. *Frontiers in Ecology and Evolution* **11**, 20.

*Miller, F. L. (1974). Age determination of caribou by annulations in dental cementum. *The Journal of Wildlife Management* **38**, 47.

*Mitchell, B. (1967). Growth layers in dental cement for determining the age of red deer (*Cervus elaphus* L.). *The Journal of Animal Ecology* **36**, 279.

*Miura, S. (1985). Horn and cementum annulation as age criteria in Japanese serow. *The Journal of Wildlife Management* **49**, 152.

Molbert, N., Angelier, F., Alliot, F., Ribout, C. & Goutte, A. (2021). Fish from urban rivers and with high pollutant levels have shorter telomeres: telomeres and contamination in fish. *Biology Letters* **17**, 20200819.

Morin, P. A., Nestler, A., Rubio-Cisneros, N. T., Robertson, K. M. & Mesnick, S. L. (2005). Interfamilial characterization of a region of the ZFX and ZFY genes facilitates sex determination in cetaceans and other mammals. *Molecular Ecology* **14**, 3275–3286.

Morselli, M., Bennett, R., Shaidani, N.-I., Horb, M., Peshkin, L. & Pellegrini, M. (2023). Age-associated DNA methylation changes in *Xenopus* frogs. *Epigenetics* **18**, 2201517.

Moyzis, R. K., Buckingham, J. M., Crams, L. S., Dani, M., Deavent, L. L., Jones, M. D., Meyne, J., Ratliff, R. L. & Wu, J.-R. (1988). A highly conserved repetitive DNA sequence, $(TTAGGG)_n$, present at the telomeres of human chromosomes. *Proceedings of the National Academy of Sciences* **85**, 6622–6626.

Myrick, A. C. J., Hohn, A. A., Sloan, P. A., Makoto, K. & Stanley, D. D. (1983). Estimating age of spotted and spinner dolphins (*Stenella attenuata* and *Stenella longirostris*) from teeth. NOAA Technical Memorandum NMFS, SWFC-30.

*Nakanishi, N., Ichinose, F., Higa, G. & Izawa, M. (2009). Age determination of the iriomote cat by using cementum annuli. *Journal of Zoology* **279**, 338–348.

Newham, E., Gill, P. G., Robson Brown, K., Gostling, N. J. & Schneider, P. (2021). A robust, semi-automated approach for counting cementum increments imaged 1 with X-ray computed tomography. *PLoS One* **16**, e0249743.

*Nicholson, W. S., Hill, E. P. & Briggs, D. (1985). Denning, pup-rearing, and dispersal in the gray fox in east-Central Alabama. *The Journal of Wildlife Management* **49**, 33.

Nilsen, F. M., Parrott, B. B., Bowden, J. A., Kassim, B. L., Somerville, S. E., Bryan, T. A., Bryan, C. E., Lange, T. R., Delaney, J. P., Brunell, A. M., Long, S. E. & Guillette, L. J. (2016). Global DNA methylation loss associated with mercury contamination and aging in the American alligator (*Alligator mississippiensis*). *Science of the Total Environment* **545**, 389–397.

Noguera, J. C. & Velando, A. (2021). Telomerase activity can mediate the effects of growth on telomeres during post-natal development in a wild bird. *Journal of Experimental Biology* **224**, jeb242465.

Noren, S. R. & Edwards, E. F. (2007). Physiological and behavioral development in delphinid calves: implications for calf separation and mortality due to tuna purse-seine sets. *Marine Mammal Science* **23**, 15–29.

Oganesian, L. & Karlseder, J. (2011). Mammalian 5′ C-rich telomeric overhangs are a mark of recombination-dependent telomere maintenance. *Molecular Cell* **42**, 224–236.

Ohsumi, S., Kasuya, T. & Nishiwaki, M. (1965). Accumulation rate of dentinal growth layers in the maxillary tooth of the sperm whale. *Scientific Reports The Whales Research Institute* **17**, 15–34.

*Olivier, G. C. & Greyling, F. J. (1991). Criteria for age assessment of blesbok. *South African Journal of Wildlife Research* **21**, 1–5.

Olsen, M. T., Robbins, J., Bérubé, M., Rew, M. B. & Palsbøll, P. J. (2014). Utility of telomere length measurements for age determination of humpback whales. *NAMMCO Scientific Publications* **10**, 1–17.

Olsson, M., Pauliny, A., Wapstra, E., Uller, T., Schwartz, T. & Blomqvist, D. (2011). Sex differences in sand lizard telomere inheritance: paternal epigenetic effects increases telomere heritability and offspring survival. *PLoS One* **6**, e17473.

Olsson, M., Wapstra, E. & Friesen, C. R. (2018). Evolutionary ecology of telomeres: a review. *Annals of the New York Academy of Sciences* **1422**, 5–28.

Ooi, S. K. T. & Bestor, T. H. (2008). The colorful history of active DNA demethylation. *Cell* **133**, 1145–1148.

*Oosthuizen, W. H. (1997). Evaluation of an effective method to estimate age of Cape fur seals using ground tooth sections. *Marine Mammal Science* **13**, 683–693.

*Østbye, E., Lauritzen, S.-E. E., Østbye, K. & Wiig, Ø. (2006). Holocene brown bear (*Ursus arctos* L.) from Norwegian caves. *Boreas* **35**, 296–316.

Owen, R. (1840). *Odontography, or, A Treatise on the Comparative Anatomy of the Teeth: Their Physiological Relations, Mode of Development, and Microscopic Structure in the Vertebrate Animals*, Edition (Volume **1**). Balliere, London, UK. https://doi.org/10.5962/bhl.title.16281.

Palm, W. & de Lange, T. (2008). How shelterin protects mammalian telomeres. *Annual Review of Genetics* **42**, 301–334.

Panasiak, L., Dobosz, S. & Ocalewicz, K. (2020). Telomere dynamics in the diploid and triploid rainbow trout (*Oncorhynchus mykiss*) assessed by Q-FISH analysis. *Genes* **11**, 1–12.

Panasiak, L., Szubert, K., Polonis, M. & Ocalewicz, K. (2022). Telomere length variation does not correspond with the growth disturbances in the rainbow trout (*Oncorhynchus mykiss*). *Journal of Applied Genetics* **63**, 133–139.

Parolini, M., Romano, A., Khoriauli, L., Nergadze, S. G., Caprioli, M., Rubolini, D., Santagostino, M., Saino, N. & Giulotto, E. (2015). Early-life telomere dynamics differ between the sexes and predict growth in the barn swallow (*Hirundo rustica*). *PLoS One* **10**, e0142530.

Parrott, B. B. & Bertucci, E. M. (2019). Epigenetic aging clocks in ecology and evolution. *Trends in Ecology and Evolution* **34**, 767–770.
Parrott, B. B., Bowden, J. A., Kohno, S., Cloy-McCoy, J. A., Hale, M. D., Bangma, J. T., Rainwater, T. R., Wilkinson, P. M., Kucklick, J. R. & Guillette, L. J. (2014). Influence of tissue, age, and environmental quality on DNA methylation in *Alligator mississippiensis*. *Reproduction* **147**, 503–513.
Pauli, J. N., Whiteman, J. P., Marcot, B. G., McClean, T. M. & Ben-David, M. (2011). DNA-based approach to aging martens (*Martes americana* and *M. caurina*). *Journal of Mammalogy* **92**, 500–510.
Pauliny, A., Larsson, K. & Blomqvist, D. (2012). Telomere dynamics in a long-lived bird, the barnacle goose. *BMC Evolutionary Biology* **12**, 1–8.
Pauliny, A., Wagner, R. H., Augustin, J., Szép, T. & Blomqvist, D. (2006). Age-independent telomere length predicts fitness in two bird species. *Molecular Ecology* **15**, 1681–1687.
Payan, P., Borelli, G., Boeuf, G. & Mayer-Gostan, N. (1998). Relationship between otolith and somatic growth: consequence of starvation on acid-base balance in plasma and endolymph in the rainbow trout *Oncorhynchus mykiss*. *Fish Physiology and Biochemistry* **19**, 35–41.
Pekelharing, C. J. (1970). Cementum deposition as an age indicator in the brush-tailed possum, *Trichosurus vulpecula* Kerr (Marsupiala). *Australian Journal of Zoology* **18**, 71–76.
Penzhorn, B. L. (1982). Age determination in Cape mountain zebras *Equus zebra zebra* in the Mountain Zebra National Park. *Koedoe* **25**, 89–102.
Pérez-Barberìa, F. J., Duff, E. I., Brewer, M. J. & Guinness, F. E. (2014). Evaluation of methods to age Scottish red deer: the balance between accuracy and practicality. *Journal of Zoology* **294**, 180–189.
*Perez-Barberia, F. J. & Fernandez-Lopez, J. M. (1996). Using cementum annuli to estimate Cantabrian chamois age. *The Journal of Wildlife Management* **60**, 62–67.
Pérez-Pereira, N., López-Cortegano, E. & García-Dorado, A. (2022). Prediction of fitness under different breeding designs in conservation programs. *Animal Conservation* **26**, 86–102.
*Pires, A. E., Caldeira, I. S., Petrucci-Fonseca, F., Viegas, I., Viegas, C., Bastos-Silveira, C. & Requicha, J. F. (2020). Dental pathology of the wild Iberian wolf (*Canis lupus signatus*): the study of a 20th century Portuguese museum collection. *Veterinary and Animal Science* **9**, 100100.
Plot, V., Criscuolo, F., Zahn, S. & Georges, J. Y. (2012). Telomeres, age and reproduction in a long-lived reptile. *PLoS One* **7**, e40855.
Poganik, J. R., Zhang, B., Baht, G. S., Kerepesi, C., Yim, S. H., Lu, A. T., Haghani, A., Gong, T., Hedman, A. M., Andolf, E., Pershagen, G., Almqvist, C., White, J. P., Horvath, S. & Gladyshev, V. N. (2023). Biological age is increased by stress and restored upon recovery. *Cell Metabolism* **35**, 807–820.
Polanowski, A. M., Robbins, J., Chandler, D. & Jarman, S. N. (2014). Epigenetic estimation of age in humpback whales. *Molecular Ecology Resources* **14**, 976–987.
Prado, N. A., Brown, J. L., Zoller, J. A., Haghani, A., Yao, M., Bagryanova, L. R., Campana, M. G., Maldonado, J. E., Raj, K., Schmitt, D., Robeck, T. R. & Horvath, S. (2021). Epigenetic clock and methylation studies in elephants. *Aging Cell* **20**, e13414.
Qi, H., Kinoshita, K., Mori, T., Matsumoto, K., Matsui, Y. & Inoue-Murayama, M. (2021). Age estimation using methylation-sensitive high-resolution melting (MS-HRM) in both healthy felines and those with chronic kidney disease. *Scientific Reports* **11**, 1–10.
Quirici, V., Guerrero, C. J., Krause, J. S., Wingfield, J. C. & Vásquez, R. A. (2016). The relationship of telomere length to baseline corticosterone levels in nestlings of an altricial passerine bird in natural populations. *Frontiers in Zoology* **13**, 1–11.
R Core Team (2020). *R: A Language and Environment for Statisitical Computing*. R Foundation for Statistical Computing, Vienna https://www.r-project.org/.
Raj, K., Szladovits, B., Haghani, A., Zoller, J. A., Li, C. Z., Black, P., Maddox, D., Robeck, T. R. & Horvath, S. (2021). Epigenetic clock and methylation studies in cats. *GeroScience* **43**, 2363–2378.
Rausch, R. A. & Pearson, A. M. (1972). Notes on the wolverine in Alaska and the Yukon territory. *The Journal of Wildlife Management* **36**, 249–268.
*Rautenbach, I. L. (1970). Ageing criteria in the springbok, *Antidorcas marsupialis* (Zimmermann, 1780) (Artiodactyla: Bovidae). *Annals of the Transvaal Museum* **27**, 83–133.
Reichenbacher, B., Sienknecht, U., Küchenhoff, H. & Fenske, N. (2007). Combined otolith morphology and morphometry for assessing taxonomy and diversity in fossil and extant killifish (Aphanius, Prolebias). *Journal of Morphology* **268**, 898–915.
Reichert, S., Rojas, E. R., Zahn, S., Robin, J. P., Criscuolo, F. & Massemin, S. (2015). Maternal telomere length inheritance in the king penguin. *Heredity* **114**, 10–16.
*Reimers, E. & Nordby, O. (1968). Relationship between age and tooth cementum layers in Norwegian reindeer. *The Journal of Wildlife Management* **32**, 957.
Remot, F., Ronget, V., Froy, H., Rey, B., Gaillard, J. M., Nussey, D. H. & Lemaitre, J. F. (2021). Decline in telomere length with increasing age across nonhuman vertebrates: a meta-analysis. *Molecular Ecology* **31**, 5917–5932.
Rey, O., Eizaguirre, C., Angers, B., Baltazar-Soares, M., Sagonas, K., Prunier, J. G. & Blanchet, S. (2020). Linking epigenetics and biological conservation: towards a conservation epigenetics perspective. *Functional Ecology* **34**, 414–427.
Roast, M. J., Eastwood, J. R., Aranzamendi, N. H., Fan, M., Teunissen, N., Verhulst, S. & Peters, A. (2022). Telomere length declines with age, but relates to immune function independent of age in a wild passerine. *Royal Society Open Science* **9**, 212012.
Robeck, T. R., Fei, Z., Haghani, A., Zoller, J. A., Li, C. Z., Steinman, K. J., Dirocco, S., Staggs, L., Schmitt, T., Osborn, S., Montano, G., Rodriguez, M. & Horvath, S. (2021*a*). Multi-tissue methylation clocks for age and sex estimation in the common bottlenose dolphin. *Frontiers in Marine Science* **8**, 1318.
Robeck, T. R., Fei, Z., Lu, A. T., Haghani, A., Jourdain, E., Zoller, J. A., Li, C. Z., Steinman, K. J., DiRocco, S., Schmitt, T., Dold, C. & Horvath, S. (2021*b*). Multi-species and multi-tissue methylation clocks for age estimation in toothed whales and dolphins. *Communications Biology* **4**, 642.
Robeck, T. R., Haghani, A., Fei, Z., Lindemann, D. M., Russell, J., Herrick, K. E. S., Montano, G., Steinman, K. J., Katsumata, E., Zoller, J. A. & Horvath, S. (2023). Multi-tissue DNA methylation aging clocks for sea lions, walruses and seals. *Communications Biology* **6**, 359.
*Robinette, W. L. & Archer, A. L. (1971). Notes on aging criteria and reproduction of Thomson's gazelle. *African Journal of Ecology* **9**, 83–98.
Rohatgi, A. (2022). Webplotdigitizer: Version 4.6. https://automeris.io/WebPlotDigitizer.
Rolandsen, C. M., Solberg, E. J., Heim, M., Holmstrøm, F., Solem, M. I. & Saether, B.-E. (2008). Accuracy and repeatability of moose (*Alces alces*) age as estimated from dental cement layers. *European Journal of Wildlife Research* **54**, 6–14.
Romano, A., De Giorgio, B., Parolini, M., Favero, C., Possenti, C. D., Iodice, S., Caprioli, M., Rubolini, D., Ambrosini, R., Gianfranceschi, L., Saino, N. & Bollati, V. (2017). Methylation of the circadian clock gene in the offspring of a free-living passerine bird increases with maternal and individual exposure to PM10. *Environmental Pollution* **220**, 29–37.
Román-Palacios, C., Medina, C. A., Zhan, S. H. & Barker, M. S. (2021). Animal chromosome counts reveal a similar range of chromosome numbers but with less polyploidy in animals compared to flowering plants. *Journal of Evolutionary Biology* **34**, 1333–1339.
*Rosell, F., Zedrosser, A. & Parker, H. (2010). Correlates of body measurements and age in Eurasian beaver from Norway. *European Journal of Wildlife Research* **56**, 43–48.
RStudio Team (2021). *RStudio: Integrated Development Environment for R*. IDE, Boston http://www.rstudio.com/.
Rust, L. B., Danil, K., Melin, S. R. & Wilkerson, B. (2019). Accuracy and precision of age determination using growth layer groups for California Sea lions (*Zalophus californianus*) with known ages. *Marine Mammal Science* **35**, 1355–1368.
Saino, N., Albetti, B., Ambrosini, R., Caprioli, M., Costanzo, A., Mariani, J., Parolini, M., Romano, A., Rubolini, D., Formenti, G., Gianfranceschi, L. & Bollati, V. (2019). Inter-generational resemblance of methylation levels at circadian genes and associations with phenology in the barn swallow. *Scientific Reports* **9**, 1–16.
Saino, N., Ambrosini, R., Albetti, B., Caprioli, M., De Giorgio, B., Gatti, E., Liechti, F., Parolini, M., Romano, A., Romano, M., Scandolara, C., Gianfranceschi, L., Bollati, V. & Rubolini, D. (2017). Migration phenology and breeding success are predicted by methylation of a photoperiodic gene in the barn swallow. *Scientific Reports* **7**, 45412.
Saino, N., Rubolini, D., Ambrosini, R., Romano, M., Scandolara, C., Fairhurst, G. D., Caprioli, M., Romano, A., Sicurella, B. & Liechti, F. (2015). Light-level geolocators reveal covariation between winter plumage molt and phenology in a trans-Saharan migratory bird. *Oecologia* **178**, 1105–1112.
Salmón, P., Nilsson, J. F., Nord, A., Bensch, S. & Isaksson, C. (2016). Urban environment shortens telomere length in nestling great tits, *Parus major*. *Biology Letters* **12**, 20160155.
Salomons, H. M., Mulder, G. A., Van De Zande, L., Haussmann, M. F., Linskens, M. H. K. & Verhulst, S. (2009). Telomere shortening and survival in free-living corvids. *Proceedings of the Royal Society B: Biological Sciences* **276**, 3157–3165.
Sánchez-Montes, G., Martínez-Solano, Í., Díaz-Paniagua, C., Vilches, A., Ariño, A. H. & Gomez-Mestre, I. (2020). Telomere attrition with age in a wild amphibian population. *Biology Letters* **16**, 20200168.
Sareisian, L. (2014). Global DNA methylation and persistent organic pollutants in ringed seals (*Phoca hispida*) from Svalbard and the Baltic Sea. Masters dissertation: University of Oslo, Oslo. http://urn.nb.no/URN:NBN:no-46760.
Sato, A., Fairbanks, L. A., Lawson, P. T. & Lawson, G. W. (2005). Effects of age and sex on hematologic and serum biochemical values of vervet monkeys (*Chlorocebus aethiops sabaeus*). *Contemporary Topics in Laboratory Animal Science* **44**, 29–34.

Saunders, P. A. & Veyrunes, F. (2021). Unusual mammalian sex determination systems: a cabinet of curiosities. *Genes* **12**, 1170.
Saxon, A. & Higham, C. F. W. (1968). Identification and interpretation of growth rings in the secondary dental cementum of *Ovis aries* L. *Nature* **219**, 634–635.
Schachtschneider, K. M., Schook, L. B., Meudt, J. J., Shanmuganayagam, D., Zoller, J. A., Haghani, A., Li, C. Z., Zhang, J., Yang, A., Raj, K. & Horvath, S. (2021). Epigenetic clock and DNA methylation analysis of porcine models of aging and obesity. *GeroScience* **43**, 2467–2483.
Schaetzlein, S., Lucas-Hahn, A., Lemme, E., Kues, W. A., Dorsch, M., Manns, M. P., Niemann, H. & Rudolph, K. L. (2004). Telomere length is reset during early mammalian embryogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 8034–8038.
*Scheffer, V. B. (1950). Growth layers on the teeth of pinnipedia as an indication of age. *Science* **112**, 309–311.
*Schuler, K. L., Baker, B. B., Mayer, K. A., Perez-Heydrich, C., Holahan, P. M., Thomas, N. J. & White, C. L. (2018). Refining aging criteria for northern sea otters in Washington State. *Journal of Fish and Wildlife Management* **9**, 208–221.
Schwarzer, G., Carpenter, J. R. & Rücker, G. (2015). *Meta-Analysis with R*. Springer, Cham..
Scott, G. G., Sis, R. F., Levy, B. M., Stott, G. G., Sis, R. F. & Levy, B. M. (1980). Cemental annulation as an age criterion in the common marmoset (*Callithrix jacchus*). *Journal of Medical Primatology* **9**, 274–285.
Scott, N. M., Haussmann, M. F., Elsey, R. M., Trosclair, P. L. & Vleck, C. M. (2006). Telomere length shortens with body length in *Alligator mississippiensis*. *Southeastern Naturalist* **5**, 685–692.
Secor, D. H., Trice, T. M. & Hornick, H. T. (1995). Validation of otolith-based ageing and a comparison of otolith and scale-based ageing in mark-recaptured Chesapeake Bay striped bass, *Morone saxatilis*. *Fishery Bulletin* **93**, 186–190.
Seibt, K. D., Häussler, S., Vecchio, D., DeCarlo, E., Ceciliani, F. & Sauerwein, H. (2019). Comparison of telomere lengths in leukocytes and in nasal and vaginal epithelial cells from water buffaloes (*Bubalus bubalis*) of different ages. *Research in Veterinary Science* **124**, 328–333.
Sen, G. L., Reuter, J. A., Webster, D. E., Zhu, L. & Khavari, P. A. (2010). DNMT1 maintains progenitor function in self-renewing somatic tissue. *Nature* **463**, 563–567.
Sfeir, A. J., Chai, W., Shay, J. W. & Wright, W. E. (2005). Telomere-end processing: the terminal nucleotidesof human chromosomes. *Molecular Cell* **18**, 131–138.
Sharma, A., Singh, D., Malik, S., Gupta, N. J., Rani, S. & Kumar, V. (2018). Difference in control between spring and autumn migration in birds: insight from seasonal changes in hypothalamic gene expression in captive buntings. *Proceedings of the Royal Society B: Biological Sciences* **285**, 20181531.
Sharma, V., Sharma, C. P., Sharma, V., Goyal, S. P. & Gupta, S. K. (2021). Fast and cost-effective age estimation in Bengal tiger and Asiatic lion: applicability of cementum analysis method. *bioRxiv* 2021.09.27.461978. https://doi.org/10.1101/2021.09.27.461978.
Shay, J. W. & Wright, W. E. (2019). Telomeres and telomerase: three decades of progress. *Nature Reviews Genetics* **20**, 299–309.
Sherley, R. B., Abadi, F., Ludynia, K., Barham, B. J., Clark, A. E. & Altwegg, R. (2014). Age-specific survival and movement among major African penguin *Spheniscus demersus* colonies. *Ibis* **156**, 716–728.
Short, R., Williams, D. D. & Bowden, D. M. (1987). Cross-sectional evaluation of potential biological markers of aging in pigtailed macaques: effects of age, sex, and diet. *Journals of Gerontology* **42**, 644–654.
Simide, R., Angelier, F., Gaillard, S. & Stier, A. (2016). Age and heat stress as determinants of telomere length in a long-lived fish, the Siberian sturgeon. *Physiological and Biochemical Zoology* **89**, 441–447.
*Simpson, C. D. (1983). *Tooth Replacement, Growth, and Ageing Criteria for the Zambezi Bushbuck, Tragelaphus Scriptus Ornatus Pocock*. National Museums and Monuments of Rhodesia. Harare, Zimbabwe.
*Simpson, C. D. & Elder, W. H. (1969). Tooth cementum as an index of age in greater kudu (*Trageluphus strepsiceros*). *Artioldia* **4**, 1–10.
*Smith, J. R. (1975). Age determination of the woodchuck, *Marmota monax*, by annulations in dental cementum and in the periosteal zone of bones. Doctoral Thesis: Eastern Kentucky University, Richmond, United States of America.
Smith, L. E., Jones, M. E., Hamede, R., Risques, R., Patton, A. H., Carter, P. A. & Storfer, A. (2020). Telomere length is a susceptibility marker for Tasmanian devil facial tumor disease. *EcoHealth* **17**, 280–291.
Smuts, G. L., Anderson, J. L. & Austin, J. C. (1978). Age determination of the African lion (*Panthera leo*). *Journal of Zoology* **185**, 115–146.
Snir, S., vonHoldt, B. M. & Pellegrini, M. (2016). A statistical framework to identify deviation from time linearity in epigenetic aging. *PLoS Computational Biology* **12**, e1005183.
Soler, J. J., Ruiz Castellano, C., Martínez-de la Puente, J., Tomás, G., Ruiz-Rodríguez, M. & Figuerola, J. (2015). Telomere dynamics in parasitic great spotted cuckoos and their magpie hosts. *Journal of Evolutionary Biology* **28**, 1610–1617.
Soler, J. J., Ruiz-Castellano, C., Figuerola, J., Martín-Vivaldi, M., Martínez-de la Puente, J., Ruiz-Rodríguez, M. & Tomás, G. (2017). Telomere length and dynamics of spotless starling nestlings depend on nest-building materials used by parents. *Animal Behaviour* **126**, 89–100.
Sørheim, V. C. (2020). Telomere dynamics during nestling development in bluethroats (*Luscinia s. svecica*). Masters dissertation: University of Oslo, Oslo.
Soulsbury, C. D., Lipponen, A., Wood, K., Mein, C. A., Hoffman, J. I. & Lebigre, C. (2018). Age- and quality-dependent DNA methylation correlate with melanin-based coloration in a wild bird. *Ecology and Evolution* **8**, 6547–6557.
Spiers, H., Hannon, E., Wells, S., Williams, B., Fernandes, C. & Mill, J. (2016). Age-associated changes in DNA methylation across multiple tissues in an inbred mouse model. *Mechanisms of Ageing and Development* **154**, 20–23.
Spinage, C. A. (1976). Incremental cementum lines in the teeth of tropical African mammals. *Journal of Zoology* **178**, 117–131.
Spurgin, L. G., Bebbington, K., Fairfield, E. A., Hammers, M., Komdeur, J., Burke, T., Dugdale, H. L. & Richardson, D. S. (2018). Spatio-temporal variation in lifelong telomere dynamics in a long-term ecological study. *Journal of Animal Ecology* **87**, 187–198.
Starr, M. E. & Saito, H. (2014). Sepsis in old age: review of human and animal studies. *Aging and Disease* **5**, 126–136.
Statham, A. & Csárdi, G. (2008). *RSeqMeth: Package for Analysis of Sequenom EpiTYPER Data*. Sydney. https://github.com/cran/RSeqMeth.
Stewart, R. E. A., Stewart, B. E., Stirling, I. & Street, E. (1996). Counts of growth layer groups in cementum and dentine in ringed seals (*Phoca hispida*). *Marine Mammal Science* **12**, 383–401.
Stier, A., Hsu, B. Y., Marciau, C., Doligez, B., Gustafsson, L., Bize, P. & Ruuskanen, S. (2020). Born to be young? Prenatal thyroid hormones increase early-life telomere length in wild collared flycatchers. *Biology Letters* **16**, 20200364.
*Stirling, I., Archibald, W. R. & DeMaster, D. (1977). Distribution and abundance of seals in the eastern Beaufort Sea. *Journal of the Fisheries Research Board of Canada* **34**, 976–988.
Stubbs, T. M., Bonder, M. J., Stark, A. K., Krueger, F., von Meyenn, F., Stegle, O., Reik, W., Bolland, D., Butcher, G., Chandra, T., Clark, S. J., Corcoran, A., Eckersley-Maslinc, M., Field, L., Frising, U. C., *et al.* (2017). Multi-tissue DNA methylation age predictor in mouse. *Genome Biology* **18**, 1–14.
Suarez-Bregua, P., Rosendo, S., Comesaña, P., Sánchez-Ruiloba, L., Morán, P., Planas, M. & Rotllant, J. (2021). Dynamic changes in DNA methylation during seahorse (*Hippocampus reidi*) postnatal development and settlement. *Frontiers in Zoology* **18**, 1–10.
Sudyka, J., Arct, A., Drobniak, S., Dubiec, A., Gustafsson, L. & Cichoń, M. (2014). Experimentally increased reproductive effort alters telomere length in the blue tit (*Cyanistes caeruleus*). *Journal of Evolutionary Biology* **27**, 2258–2264.
Sugrue, V. J., Zoller, J. A., Narayan, P., Lu, A. T., Ortega-Recalde, O. J., Grant, M. J., Bawden, C. S., Rudiger, S. R., Haghani, A., Bond, D. M., Hore, R. R., Garratt, M., Sears, K. E., Wang, N., Yang, X. W., *et al.* (2021). Castration delays epigenetic aging and feminizes DNA methylation at androgen-regulated loci. *eLife* **10**, e64932.
Tackney, J., Cawthon, R. M., Coxworth, J. E. & Hawkes, K. (2014). Blood cell telomere lengths and shortening rates of chimpanzee and human females. *American Journal of Human Biology* **26**, 452–460.
Tanabe, A., Shimizu, R., Osawa, Y., Suzuki, M., Ito, S., Goto, M., Pastene, L. A. L. A., Fujise, Y. & Sahara, H. (2020). Age estimation by DNA methylation in the Antarctic minke whale. *Fisheries Science* **86**, 35–41.
Thomas, C. & Zapico, S. C. (2017). Role of telomeres in aging. In *Mechanisms Linking Aging, Diseases and Biological Age Estimation*, First Edition (ed. S. C. Zapico), pp. 141–149. CRC Press. Boca Raton, USA.
*Thomas, D. C. (1977). Metachromatic staining of dental cementum for mammalian age determination. *The Journal of Wildlife Management* **41**, 207.
Thompson, M. J., von Holdt, B., Horvath, S. & Pellegrini, M. (2017). An epigenetic aging clock for dogs and wolves. *Aging* **9**, 1055–1068.
Thompson, R. F., Suzuki, M., Lau, K. W. & Greally, J. M. (2009). A pipeline for the quantitative analysis of CG dinucleotide methylation using mass spectrometry. *Bioinformatics* **25**, 2164–2170.
Tilstra, J. S., Clauson, C. L., Niedernhofer, L. J. & Robbins, P. D. (2011). NF-κB in aging and disease. *Aging and Disease* **2**, 449–465.
Tissier, M. L., Bergeron, P., Garant, D., Zahn, S., Criscuolo, F. & Réale, D. (2022). Telomere length positively correlates with pace-of-life in a sex- and cohort-specific way and elongates with age in a wild mammal. *Molecular Ecology* **31**, 3812–3826.
Tricola, G. M., Simons, M. J. P., Atema, E., Boughton, R. K., Brown, J. L., Dearborn, D. C., Divoky, G., Eimes, J. A., Huntington, C. E., Kitaysky, A. S., Juola, F. A., Lank, D. B., Litwa, H. P., Mulder, E. G. A., Nisbet, I. C. T., *et al.* (2018). The rate of telomere loss is related to maximum lifespan in birds. *Philosophical Transactions of the Royal Society B: Biological Sciences* **373**, 20160445.

Trivedi, A. K., Sur, S., Sharma, A., Taufique, S. T., Gupta, N. J. & Kumar, V. (2019). Temperature alters the hypothalamic transcription of photoperiod responsive genes in induction of seasonal response in migratory redheaded buntings. *Molecular and Cellular Endocrinology* **493**, 110454.
Ujvari, B., Biro, P. A., Charters, J. E., Brown, G., Heasman, K., Beckmann, C. & Madsen, T. (2017). Curvilinear telomere length dynamics in a squamate reptile. *Functional Ecology* **31**, 753–759.
Ujvari, B. & Madsen, T. (2009). Short telomeres in hatchling snakes: erythrocyte telomere dynamics and longevity in tropical pythons. *PLoS One* **4**, e7493.
*Van Aarde, R. J. (1985). Age determination of Cape porcupines, *Hystrix africaeaustralis*. *African Zoology* **20**, 232–236.
van Eck, N. J. & Waltman, L. (2017). Citation-based clustering of publications using CitNetExplorer and VOSviewer. *Scientometrics* **111**, 1053–1070.
van Lieshout, S. H. J., Bretman, A., Newman, C., Buesching, C. D., Macdonald, D. W. & Dugdale, H. L. (2019). Individual variation in early-life telomere length and survival in a wild mammal. *Molecular Ecology* **28**, 4152–4165.
Van Lieshout, S. H. J., Sparks, A. M., Bretman, A., Newman, C., Buesching, C. D., Burke, T., Macdonald, D. W. & Dugdale, H. L. (2021). Estimation of environmental, genetic and parental age at conception effects on telomere length in a wild mammal. *Journal of Evolutionary Biology* **34**, 296–308.
*Van Nostrand, F. C. & Stephenson, A. B. (1964). Age determination for beavers by tooth development. *The Journal of Wildlife Management* **28**, 430.
Vedder, O., Moiron, M., Bichet, C., Bauch, C., Verhulst, S., Becker, P. H. & Bouwhuis, S. (2022). Telomere length is heritable and genetically correlated with lifespan in a wild bird. *Molecular Ecology* **31**, 6297–6307.
Vedder, O., Verhulst, S., Bauch, C. & Bouwhuis, S. (2017). Telomere attrition and growth: a life-history framework and case study in common terns. *Journal of Evolutionary Biology* **30**, 1409–1419.
Venegas, D., Marmolejo-Valencia, A., Valdes-Quezada, C., Govenzensky, T., Recillas-Targa, F. & Merchant-Larios, H. (2016). Dimorphic DNA methylation during temperature-dependent sex determination in the sea turtle *Lepidochelys olivacea*. *General and Comparative Endocrinology* **236**, 35–41.
Vernasco, B. J., Dakin, R., Majer, A. D., Haussmann, M. F., Brandt Ryder, T. & Moore, I. T. (2021). Longitudinal dynamics and behavioural correlates of telomeres in male wire-tailed manakins. *Functional Ecology* **35**, 450–462.
Vernasco, B. J. & Watts, H. E. (2022). Telomere length predicts timing and intensity of migratory behaviour in a nomadic songbird. *Biology Letters* **18**, 20220176.
Viblanc, V. A., Criscuolo, F., Sosa, S., Schull, Q., Boonstra, R., Saraux, C., Lejeune, M., Roth, J. D., Uhlrich, P., Zahn, S. & Dobson, F. S. (2022). Telomere dynamics in female Columbian ground squirrels: recovery after emergence and loss after reproduction. *Oecologia* **199**, 301–312.
Viciano, J., López-Lázaro, S. & Tanga, C. (2022). *Post-mortem* dental profile as a powerful tool in animal forensic investigations—a review. *Animals* **12**, 2038.
Viitaniemi, H. M., Verhagen, I., Visser, M. E., Honkela, A., Van Oers, K., Husby, A. & Meyer, M. (2019). Seasonal variation in genome-wide DNA methylation patterns and the onset of seasonal timing of reproduction in great tits. *Genome Biology and Evolution* **11**, 970–983.
Voillemot, M., Hine, K., Zahn, S., Criscuolo, F., Gustafsson, L., Doligez, B. & Bize, P. (2012). Effects of brood size manipulation and common origin on phenotype and telomere length in nestling collared flycatchers. *BMC Ecology* **12**, 1–8.
Wada, K., Ohtaishi, N. & Hachiya, N. (1978). Determination of age in the Japanese monkey from growth layers in the dental cementum. *Primates* **19**, 775–784.
Walleser, L. R., Crimmins, S. M. & Roberts, N. M. (2016). Estimating the age of male gray wolves (*Canis lupus*) using baculum measurements. *Canadian Field-Naturalist* **130**, 212–215.
Watson, R. L., Bird, E. J., Underwood, S., Wilbourn, R. V., Fairlie, J., Watt, K., Salvo-Chirnside, E., Pilkington, J. G., Pemberton, J. M., McNeilly, T. N., Froy, H. & Nussey, D. H. (2017). Sex differences in leucocyte telomere length in a free-living mammal. *Molecular Ecology* **26**, 3230–3240.
Webb, S. J. N., Schapiro, S. J., Sherwood, C. C., Raghanti, M. A. & Hopkins, W. D. (2020). Neutrophil to lymphocyte ratio (NLR) in captive chimpanzees (*Pan troglodytes*): the effects of sex, age, and rearing. *PLoS One* **15**, e0244092.
Wei, T., Nie, J., Larson, N. B., Ye, Z., Eckel-Passow, J. E., Robertson, K. D., Kocher, J. P. A. & Wang, L. (2021). CpGtools: a python package for DNA methylation analysis. *Bioinformatics* **37**, 1598–1599.
White, P. A. & Belant, J. L. (2016). Individual variation in dental characteristics for estimating age of African lions. *Wildlife Biology* **22**, 71–77.
Whittemore, K., Vera, E., Martínez-Nevado, E., Sanpera, C. & Blasco, M. A. (2019). Telomere shortening rate predicts species life span. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 15122–15127.
Wilbourn, R. V., Froy, H., McManus, M. C., Cheynel, L., Gaillard, J. M., Gilot-Fromont, E., Regis, C., Rey, B., Pellerin, M., Lemaître, J. F. & Nussey, D. H. (2017). Age-dependent associations between telomere length and environmental conditions in roe deer. *Biology Letters* **13**, 20170434.
Wilbourn, R. V., Moatt, J. P., Froy, H., Walling, C. A., Nussey, D. H. & Boonekamp, J. J. (2018). The relationship between telomere length and mortality risk in non-model vertebrate systems: a meta-analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences* **373**, 20160447.
Wilkinson, G. S., Adams, D. M., Haghani, A., Lu, A. T., Zoller, J., Breeze, C. E., Arnold, B. D., Ball, H. C., Carter, G. G., Cooper, L. N., Dechmann, D. K. N., Devanna, P., Fasel, N. J., Galazyuk, A. V., Günther, L., *et al.* (2021). DNA methylation predicts age and provides insight into exceptional longevity of bats. *Nature Communications* **12**, 1615.
*Wilson, V. J., Schmidt, J. L. & Hanks, J. (1984). Age determination and body growth of the common duiker *Sylvicapra grimmia* (Mammalia). *Journal of Zoology* **202**, 283–297.
Wittwer-Backofen, U., Gampe, J. & Vaupel, J. W. (2004). Tooth cementum annulation for age estimation: results from a large known-age validation study. *American Journal of Physical Anthropology* **123**, 119–129.
Wong, P. B. Y. (2017). Conservation and biological senescence in polar bears: telomeres and inuit traditional knowledge. Doctoral Thesis: University of Toronto, Toronto. https://hdl.handle.net/1807/79501.
Wright, P. G. R., Mathews, F., Schofield, H., Morris, C., Burrage, J., Smith, A., Dempster, E. L. E. L. & Hamilton, P. B. P. B. (2018). Application of a novel molecular method to age free-living wild Bechstein's bats. *Molecular Ecology Resources* **18**, 1374–1380.
Wright, W. E., Piatyszek, M. A., Rainey, W. E., Byrd, W. & Shay, J. W. (1996). Telomerase activity in human germline and embryonic tissues and cells. *Developmental Genetics* **18**, 173–179.
Wu, P., Takai, H. & De Lange, T. (2012). Telomeric 3′ overhangs derive from resection by Exo1 and apollo and fill-in by POT1b-associated CST. *Cell* **150**, 39–52.
*Yoneda, M. (1982). Growth layers in dental cementum of Saguinus monkeys in South America. *Primates* **23**, 460–464.
Yuan, X., Ishibashi, S., Hatakeyama, S., Saito, M., Nakayama, J. I., Nikaido, R., Haruyama, T., Watanabe, Y., Iwata, H., Iida, M., Sugimura, H., Yamada, N. & Ishikawa, F. (1999). Presence of telomeric G-strand tails in the telomerase catalytic subunit TERT knockout mice. *Genes to Cells* **4**, 563–572.
*Zapata, S. C., Perea, R. G., Beltran, J. F., Ferreras, P. & Delibes, M. (1997). Age determination of Iberian lynx (*Lynx pardinus*) using canine radiograph and cementum annuli enumeration. *Zeitschrift fur Saugetierkunde* **62**, 119–123.
Zapico, S. C. & Ubelaker, D. H. (2016). Relationship between mitochondrial DNA mutations and aging. Estimation of age-at-death. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* **71**, 445–450.
Zhang, G. (2015). Bird sequencing project takes off. *Nature* **522**, 34.
Zhang, Q., Han, X., Hao, X., Ma, L., Li, S., Wang, Y. & Du, W. (2018). A simulated heat wave shortens the telomere length and lifespan of a desert lizard. *Journal of Thermal Biology* **72**, 94–100.
Zhang, W., Klinkebiel, D., Barger, C. J., Pandey, S., Guda, C., Miller, A., Akers, S. N., Odunsi, K. & Karpf, A. R. (2020). Global DNA hypomethylation in epithelial ovarian cancer: passive demethylation and association with genomic instability. *Cancers* **12**, 764.
Zhao, Z., Pan, X., Liu, L. & Liu, N. (2014). Telomere length maintenance, shortening, and lengthening. *Journal of Cellular Physiology* **229**, 1323–1329.
Zhdanova, N. S., Minina, J. M., Karamysheva, T. V., Rubtsov, N. B. & Londono-Vallejo, J. A. (2010). The structure of long telomeres in chromosomes of the Iberian shrew. *Russian Journal of Genetics* **46**, 1084–1086.
*Zhu, X.-M., Xiao, H. & Wu, J.-Y. (1992). Age determination of Goitred gazelle (*Gazella subgutturosa*). *Zoological Research* **13**, 95–99.
Zoller, J. A., Parasyraki, E., Lu, A. T., Haghani, A., Niehrs, C. & Horvath, S. (2022). DNA methylation clocks for clawed frogs reveal evolutionary conservation of epigenetic ageing. *bioRxiv* 2022.08.02.502561. https://doi.org/10.1101/2022.08.02.502561.

X. SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. Summary of all species within the class Mammalia for which tooth cementum annulation (TCA) has been used as biological clock method for age determination.
Fig. S1. PRISMA flow diagram for search, screening, and inclusion of methylation studies for age determination in animals.
Fig. S2. PRISMA flow diagram for search, screening, and inclusion of telomere studies for age determination in animals.

Fig. S3. Relationship between power and sample size for different effect sizes in a multiple linear regression with a minimum of three predictors at a significance of 95% ($\alpha = 0.05$).
Fig. S4. Forest plot and results for the meta-analysis of studies using methylation to estimate animal age with subgrouping based on the defined groups.
Fig. S5. Forest plot and results for the meta-analysis of studies using methylation to estimate animal age with subgrouping based on tissue type.
Fig. S6. Forest plot and results for the meta-analysis of studies using methylation to estimate animal age with subgrouping based on study method.
Fig. S7. Forest plot and results for the meta-analysis of studies using telomere length to estimate animal age with subgrouping based on defined groups.
Fig. S8. Forest plot and results for the meta-analysis of studies using telomere length to estimate animal age with subgrouping based on tissue type.
Fig. S9. Forest plot and results for the meta-analysis of studies using telomere length to estimate animal age with subgrouping based on study method.
Fig. S10. Forest plot indicating and comparing the results obtained from independent meta-analyses of studies using methylation *versus* telomere length to model the relationship with ageing in animals.

(*Received 21 October 2022; revised 4 June 2023; accepted 8 June 2023; published online 25 June 2023*)

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Dataset generated in a systematic review and meta-analysis of biological clocks as age estimation markers in animal ecology

Louis-Stéphane Le Clercq [a,b,*], J. Paul Grobler [b], Antoinette Kotzé [a,b], Desiré Lee Dalton [c]

[a] South African National Biodiversity Institute, Pretoria 0001, South Africa
[b] Department of Genetics, University of the Free State, Bloemfontein 9300, South Africa
[c] School of Health and Life Sciences, Teesside University, Middlesbrough TS1 3BA, United Kingdom

ARTICLE INFO

Article history:
Received 8 September 2023
Revised 18 September 2023
Accepted 19 September 2023
Available online 25 September 2023

Dataset link: Biological clocks (Data) (Reference data)

Keywords:
Animals
Biological clocks
Epigenetics
Methylation
Telomeres
Biomarker
Age determination
Meta-analysis

ABSTRACT

The dataset comprises a comprehensive systematic review and meta-analysis exploring the utility of biological clocks as age estimation markers in the context of animal ecology. The systematic review adhered to PRISMA guidelines and employed optimized Boolean search strings to retrieve relevant studies from Scopus and Dimensions databases. A total of 78 methylation studies and 108 telomere studies were included after rigorous screening. Effect sizes were computed, and statistical transformations were applied when necessary, ensuring compatibility for meta-analysis. Data from these studies were meticulously collected, encompassing statistical measures, study attributes, and additional biological information. The dataset comprises several folders, carefully organized to facilitate access and understanding. It contains raw and processed data used in the systematic review and meta-analysis, including Boolean search strings, database search results, citation network analysis data, PRISMA statements, extracted study data, and input data for meta-analysis. Each folder's contents are described in detail, ensuring clarity and reusability. This dataset aggregates primary research studies

* Corresponding author.
E-mail address: leclercq.l.s@gmail.com (L.-S. Le Clercq).
Social media: @leclercq0 (L.-S. Le Clercq)

https://doi.org/10.1016/j.dib.2023.109615

spanning diverse ecosystems and taxa, providing a valuable resource for researchers, biodiversity managers and policy-makers. This dataset offers a wealth of information and analysis potential for researchers studying age estimation markers in animal ecology, serving as a robust foundation for future investigations and reviews in this evolving field.

Specifications Table

Subject	Biological Science (Aging, Genetics: Epigenetics)
Specific subject area	This dataset pertains to studies and study attributes for two biological clocks, methylation and telomeres, used in animal age determination.
Data format	Raw, Filtered, Analysed
Type of data	Tables, Figures
Data collection	A Boolean search string was used to search two scientific databases. Initial results were narrowed down by automated filtering techniques such as additional search terms as inclusion and exclusion criteria. Results were screened manually by assessing the titles, abstracts, and key words for relevance. Additional studies were identified from ancillary "free term" searches. The final set of preliminary studies for inclusion were sought for full text retrieval. Relationships between included studies were explored by citation network analysis of bibliographic coupling. Data collated from studies included species, sample size, and statistical measures used to calculate effect sizes and variances.
Data source location	Data included in the meta-analysis and review were collated from scientific literature (Scopus and Dimensions databases) and represents a globally distributed dataset from 2001 to 2023, with most studies originating from North America, Europe, and Australia. A smaller number of studies included sampling from South America and Africa. Institution: University of the Free State Address: 205 Nelson Mandela Dr, Park West, Bloemfontein, 9301 Co-ordinates: -29.1068, 26.1922
Data accessibility	Repository name: Zenodo Data identification number: 7091053 Direct URL to data: https://doi.org/10.5281/zenodo.7091053
Related research article	Le Clercq, L.-S., Kotzé, A., Grobler, J.P., and Dalton, D.L. (2023), Biological clocks as age estimation markers in animals: a systematic review and meta-analysis. *Biological Reviews*. https://doi.org/10.1111/brv.12992 [1]

1. Value of the Data

- These datasets represent the most up-to-date collection of studies using two biological clocks based on epigenetics, methylation and telomere length, to determine age in animals and defines the state-of-the-art of the field using a systematic method that enables both transparency and reproducibility.
- This data helps gain novel insights into key attributes of study design such as method used, sample sizes, and tissue type and may guide the design of future studies.
- The side-by-side comparison of the overall utility of either method for age estimation helps provide clarity on the effectiveness of either method or may assist scientists when choosing a method of studying ageing in natural populations.
- While telomere length has been in use for two decades, methylation as a biomarker for age is a relatively new and rapidly evolving field, which may likely require an updated review on the topic in a few years; in which case this dataset serves as a good baseline for subsequent reviews.

- Authorship of included studies for both methylation and telomeres was used to create and benchmark a new measure of bias for meta-analysis, Author Bias, calculated by a custom PYTHON script called ABCal to quantitate the effects of highly represented authors.
- Study attributes, such as year of publication and study location, were also used to benchmark scientometric plotting functionality using the same tool.

2. Data Description

The full dataset consists of several folders and subfolders containing the raw and/or processed data used in the systematic review and meta-analysis. The first folder is named "literature search" and contains three subfolders. The first contains two text files with the Boolean search strings, used to search databases, while the second and third contains the raw results from the search of the Scopus and Dimensions databases as comma separated value files. The column headings for search result files are described in Tables 1 and 2. The second folder is named "Citation Network Analysis" and contains two subfolders, one for raw data and one for results. The data folder contains two files for the methylation studies and two for the telomere studies, of which the comma separated value file contains the studies selected for inclusion after screening, in the same format as Scopus output, while the text file contains the same data transformed to match the input requirements for citation network analysis. The column headings for the transformed data are described in Table 3. The results subfolder contains the images for the citation network analysis. The third folder named "PRISMA Statements" contains the two PRISMA statements for methylation and telomere studies respectively. The fourth folder is named "Extracted Data" and contains two subfolders. The first subfolder is titled "Extracted Data – Studies", and contains a spreadsheet workbook with three tabs, one for methylation studies and two for telomere studies; column headings and associated data is described in Table 4. The second subfolder is titled "Extracted Data – Other" and contains data that was extracted from individual studies where relevant statistics could not be retrieved from the full text. The fifth folder is titled "Meta-Analysis" and contains that final dataset in comma separated value format that was used to perform the meta-analysis. The column headings and as-

Table 1
Field names and data description for search results retrieved from Scopus.

Field name	Data
Authors	Authors listed by surname and first initials
Author(s) ID	Authors listed by Scopus ID
Title	Main title of the article
Year	Year of publication
Source title	Name of journal in which article is published
Volume	Volume number for article in journal (if available)
Issue	Issue number for article in journal (if available)
Art. No.	Article number in journal, typically used for online only articles
Page start	Page in volume/issue where article starts
Page end	Page in volume/issue where article ends
Page count	Total number of pages
Cited by	Total number of citations
DOI	Digital Object Identifier (DOI) for article
Link	URL to online page for article at publisher
References	Full list of references cited in the article
Document Type	Type of publication e.g., article, review, book etc.
Publication Stage	Status of publication e.g., final or in press
Open Access	Open access status e.g., green or gold
Source	Database used as source to retrieve document details e.g., Scopus
EID	Electronic Identifier (EID), usually the last part of DOI

Table 2
Field names and data description for search results retrieved from dimensions.

Field name	Data
Publication ID	Unique identifier for publication of Dimensions database
DOI	Digital Object Identifier (DOI) for publication
Title	Main title of the publication
Abstract	Abstract for publication
Source title/Anthology title	Name of journal/source of publication
PubYear	Year of publication
Volume	Volume number for publication in journal (if available)
Issue	Issue number for publication in journal (if available)
Pagination	Start and end page numbers for publication
Authors	Authors listed by surname and name or initials
Authors Affiliations - Name of Research organization	List of affiliations (organization) of authors
Authors Affiliations - Country of Research organization	List of affiliations (country) of authors
Dimensions URL	URL to online page for publication on Dimensions
Times cited	Total number of citations
Cited references	Full list of references cited in publication

Table 3
Field names and data description for Scopus results after transformation for citation network analysis.

Field name	Data
AU	Authors listed by surname and initials
TI	Title of publication
PY	Publication year
SO	Name of journal for publication
VL	Volume number for publication in journal
BP	Start/beginning page for publication in volume
DI	Digital Object Identifier (DOI) for publication
CR	List of all cited references

Table 4
Field names and data description for file containing extracted study data from the review.

Field name	Data
Generic name	Common name (English) for species
Scientific name	Latin binomial for species
Author	Study label as first author and publication year e.g., Author et al, 2020
Sample size	Total sample size
R-squared	Correlation coefficient for relationship
p-value	Significance level as per reported probability value
F	Value for F-test statistic (if applicable)
X	Value for Chi-squared test statistic (if applicable)
t	Value for t-test statistic (if applicable)
Outcomes	Tissue type used to measure values
Fishers-Zr	Computed effect size expressed as Fisher's-Z
Var	Computed variance of the effect size

sociated data is described in Table 5. The same folder also contains the related forest plots that were generated as output. The sixth and final folder is named "Code" and contains the R code used to perform the meta-analysis. This file is in the standard R format (.R) and contains relevant labels and annotations to explain which steps were performed by specific lines of code.

Table 5
Field names and data description for file containing extracted data prepared for meta-analysis.

Field names	Data
Generic name	Common name (English) for species
Scientific name	Latin binomial for species
Class	(1) Vertebrate class e.g., Fish, Amphibian, Reptile, Bird, Mammal (All) (2) Grouping of birds e.g., Chats & Flycatchers (Birds)
Class.Num	Numeric encoding for class from 1 to 5
Mammal.Group	Mammal group per systematic review e.g., Aquatic, Carnivore, Primate etc.
Author	Study label as first author and publication year e.g., Author et al, 2020
Group	Grouping based on model type per the review*
N	Sample size used in study
Cor	Correlation coefficient used directly in methylation meta-analysis
Genome.Size	Size of the full genome given in billions of base pairs (methylation)
Karyotype	Total number of chromosomes (telomeres)
Tissue	Tissue type used to measure attributes
Year	Publication year
Method	Quantitative method used to make measurements e.g., PCR
Life Min	Lower end of range for life expectancy
Life Med	Middle value of range for life expectancy
Life Max	Upper end of range for life expectancy
Bias	Level of potential bias from authorship as Low, Medium, or High
Cal.Bias	Raw value for the normalised calculated author bias values

* Group 1: single species for model, validated in same species; Group 2: single species for model, validated in different/related species; Group 3: multiple species for model, validated in multiple species.

3. Experimental Design, Materials and Methods

3.1. Literature search and study screening for systematic review

Literature was searched and screened using systematic review methods (Figs. 1 and 2) per the preferred reporting items for systematic reviews and meta-analysis (PRISMA) statement [2,3], in line with PRISMA Ecology and Evolution guidelines [4] and Cochrane best practices [5]. Literature was searched between September of 2022 and June of 2023 on two databases: Scopus (www.scopus.com) and Dimensions (www.dimensions.ai). Databases were searched using an optimized Boolean search string derived from the PICO terms for the aim and objectives of the review. For methylation studies the search string was: (“Epigenetics” OR “Methylation”) AND (“age” OR “aging”) AND (“determination” OR “model”) AND (“Animals” OR “wild”). For telomere studies the search string was: (“Telomeres”) AND (“age” OR “aging”) AND (“shortening” OR “lengthening”) AND (“Animals” OR “wild”). Initial results were subjected to further automated screening using additional search terms as constraints to reduce the results for specificity and to exclude results from human studies. For the Scopus and Dimensions database searches, the final set of results for screening were exported in the comma separated value (CSV) format. Sources identified were imported (citation and abstract) into Mendeley citation manager (www.mendeley.com) for manual screening. The final set of studies that passed preliminary screening were sought during full text retrieval and added to the imported references if it wasn’t already included. A total of 78 studies were included in the final review for methylation and 108 for telomeres. Further analyses of included studies were done through citation network analyses. For the Scopus database, the results were merged and reformatted with the R package *Scopus2CitNet 0.1.0.0* in RStudio 1.4.1106 [6], running R 4.0.5 [7]. The final included studies from the results of Scopus and Dimensions were subsequently visualized by year in CitNetExplorer 1.0.0. and by group in VOSviewer 1.6.16 [8] keeping only those papers that overlapped in terms of references cited (bibliometric coupling) for the largest connected components (Figs. 3 and 4).

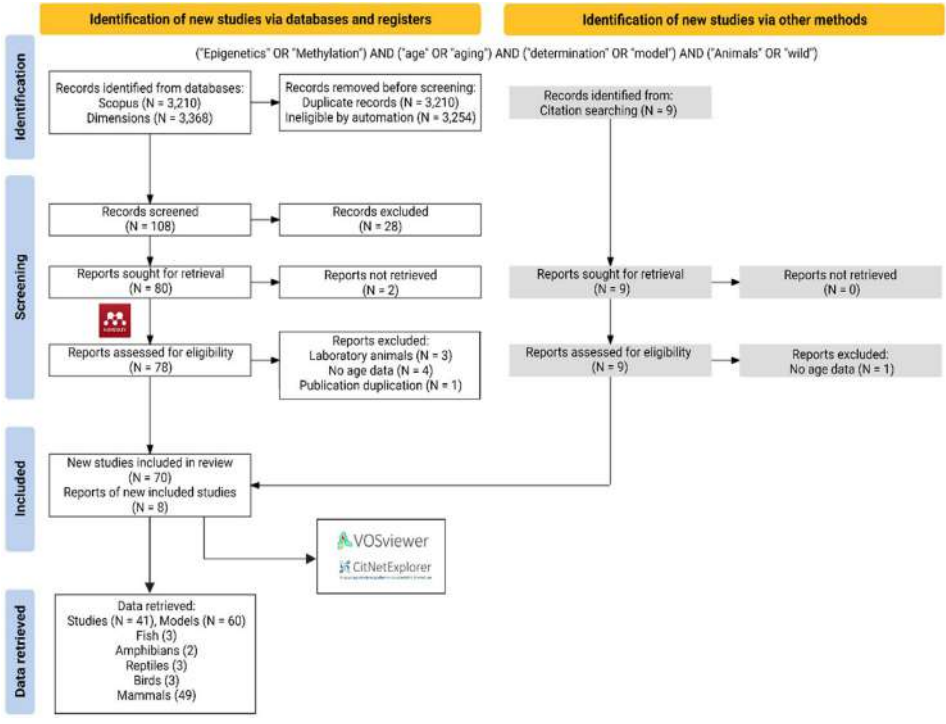


Fig. 1. PRISMA statement for the systematic approach used to identify studies that measured methylation in relation to age to develop methylation as a biomarker for age in animals. Two databases were searched using the indicated Boolean search strings. Initial automated screening removed duplicates and used additional key words to filter the results. Potential studies from the cleaned dataset were sought for retrieval and assessed for eligibility in Mendeley. Additional studies were identified from citation searches. Details are provided for relevant exclusion criteria used at each step. The final set of included studies were analysed by citation network analyses to facilitate the synthesis of the literature. Further details are also provided for the retrieval of model details and summary statistics from individual studies for inclusion in meta-analysis. (image edited in BioRender.com).

3.2. Data collection and processing for meta-analysis

Data was collected from studies that reported models using methylation or telomeres as biomarker to infer the age of animals. Relevant statistics and study attributes were retrieved from tables, figures, or the main text. For methylation studies this included key reported statistics such as the *correlation coefficient* and *p-values* for models. For telomere studies, several different statistical tests were applied beyond linear models. As such, the reported statistics compiled included measures from tests reported as *correlation coefficients* and *p-values* as well as *t-values, F-values*, and *z-values*. Where clear statistical measures were not available, WebPlotDigitizer 4.6 [9] was used to extract data from graphs or plots to repeat the reported tests and derive the relevant statistics. Data was extracted from figures from one methylation study, for snow leopards [10], and two telomere studies, for baboons [11] and chimpanzees [12], for which the correlation coefficients were computed using linear regression. For telomere studies, data was also retrieved from the online supplements to compute relevant statistics for rainbow trout [13] and stonechat species [14], for which the F-statistics were compute using ANOVA. Key study characteristics such as species, sample size, tissue type, and empirical method were also collected. Additional attributes such as lifespan, karyotype, and genome size were retrieved from online databases [15–18].

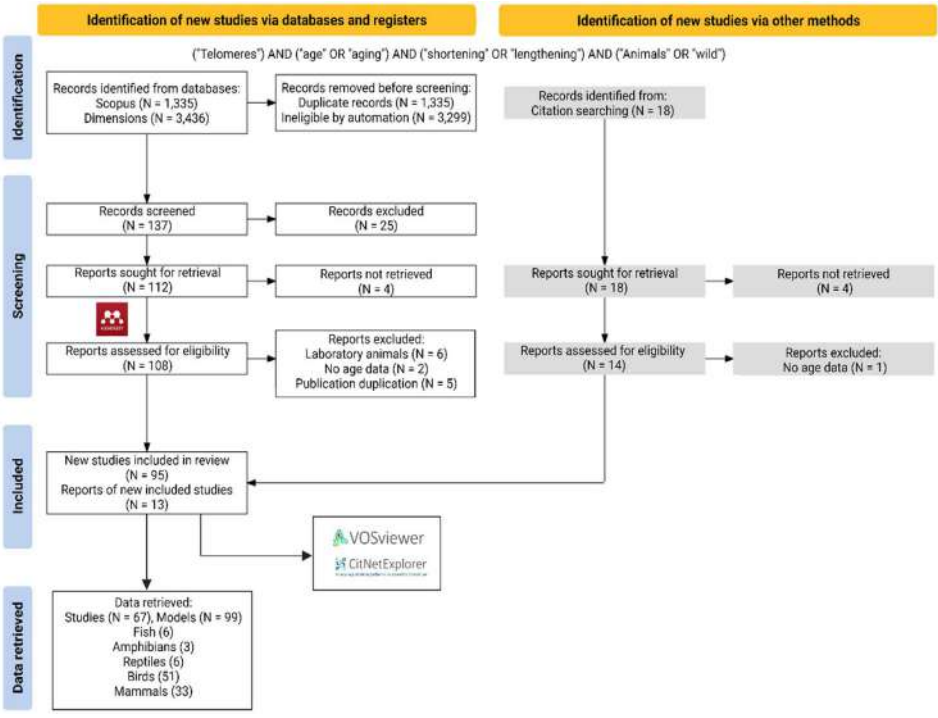


Fig. 2. PRISMA statement for the systematic approach used to identify studies that measured changes in telomere length in relation to age to develop telomere length as a biomarker for age in animals. Two databases were searched using the indicated Boolean search strings. Initial automated screening removed duplicates and used additional key words to filter the results. Potential studies from the cleaned dataset were sought for retrieval and assessed for eligibility in Mendeley. Additional studies were identified from citation searches. Details are provided for relevant exclusion criteria used at each step. The final set of included studies were analysed by citation network analyses to facilitate the synthesis of the literature. Further details are also provided for the retrieval of model details and summary statistics from individual studies for inclusion in meta-analysis. (Image edited in BioRender.com).

The effect sizes of the treatment effect (TE), expressed as *Fisher's-Z*, as well as the variance thereof expressed as standard error (SETE) were computed with the R package *compute effect size 0.2-2* [19] based on equations derived from "The Handbook of Research Synthesis and Meta-Analysis" [20]. Where the *correlation (r)* was available the z-transformed effect sizes, *Fisher's-Z*, were calculated as per equation 1.

$$Fisher's\ Z = 0.5\ \times \log {}^{(1+r)}/_{(1-r)} \tag{1}$$

where the *correlation (r)* was not available, the *Fisher's-Z* was derived from reported statistical measures by first converting between the given measure and the correlation. For *Chi-squared* (χ^2), equation 2 was used, where the correlation is derived from the quotient of the square root of *Chi-squared* and *sample size (n)*.

$$r = \sqrt{\chi 2/\mathrm{n}} \tag{2}$$

For F-test statistics *(f)*, equation 3 was used to first calculate *Cohen's d (d)* by taking the square root of the *f-value* multiplied by the sum of group sizes divided by the product of group sizes. From *d*, the correlation was derived using equation 4, where *d* is divided by the square root of the sum of (i) the square of *d* and (ii) the quotient of the sum of squared group sizes

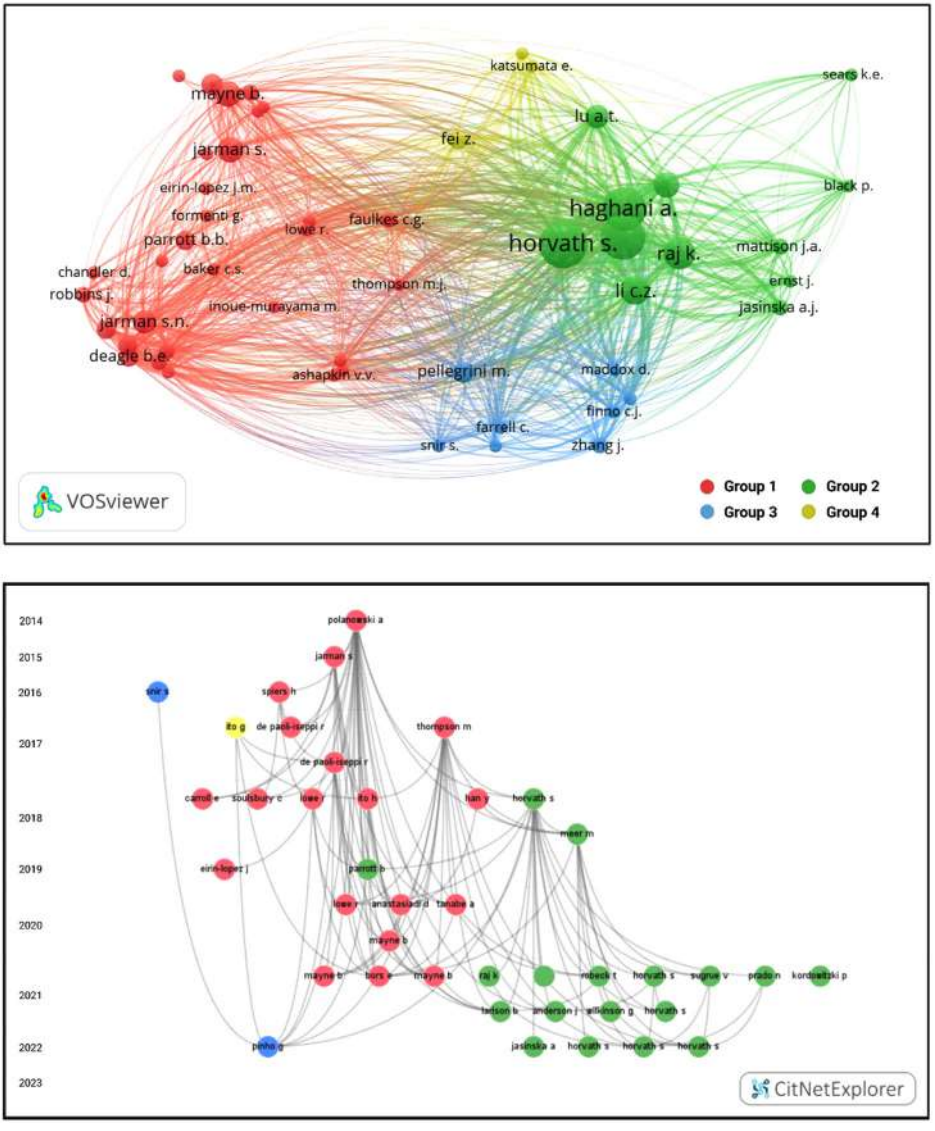


Fig. 3. Visualised citation network for methylation studies identified in database literature searches, visualised in VOSviewer in CitNetExplorer. The top panel indicates clustering analyses performed in VOSviewer, which identified four key groups, labelled 1 through 4. The bubbles indicate key authors labelled by surnames and initials with bubble size corresponding to the number of citation links with other authors. The bottom panel indicates citation network analyses of publications in CitNetExplorer, which are organized by year (2014–2023) with the name and first initial of the first author indicating individual studies. The relationship between studies by virtue of co-citations in the reference lists are indicated by grey lines. Subgroup analyses identified four key clusters, indicated according to the group colours from VOSviewer. (Image edited in BioRender.com).

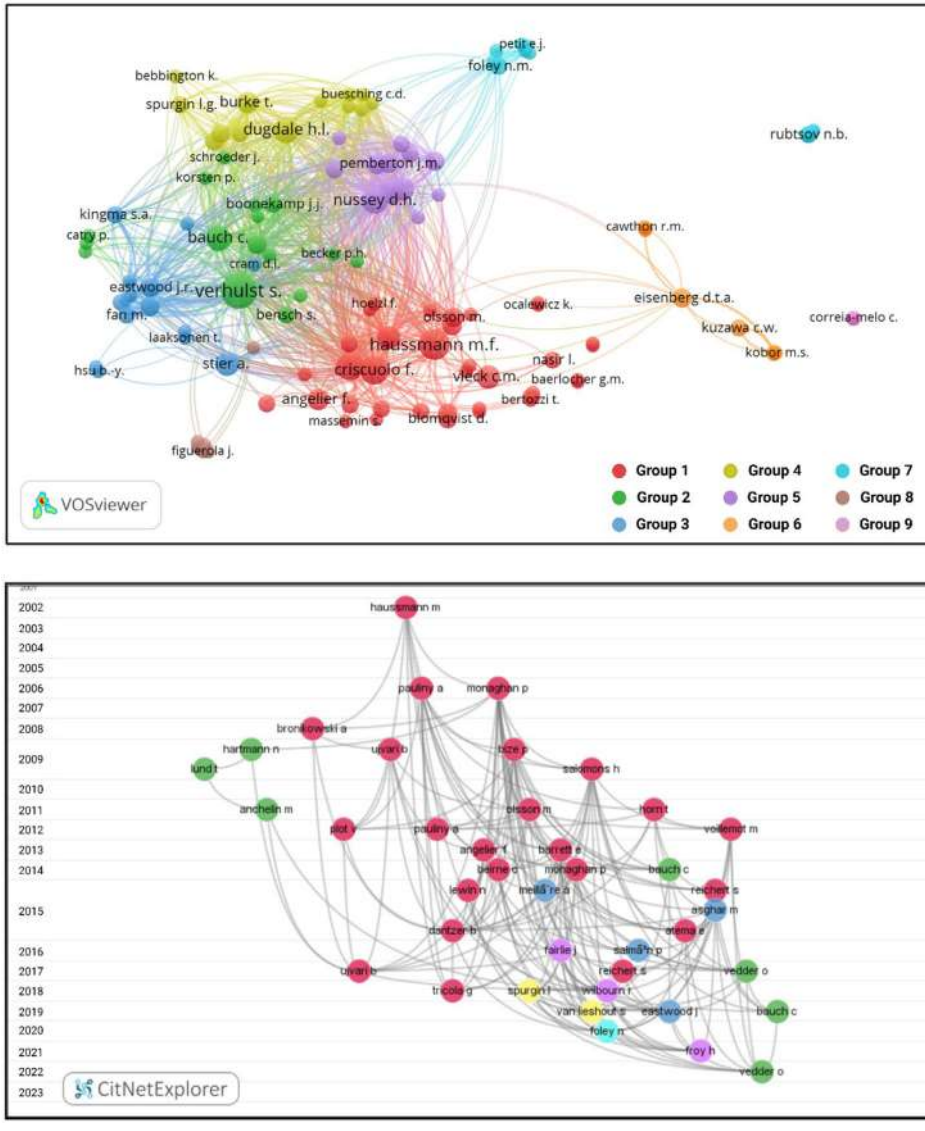


Fig. 4. Visualised citation network for telomere studies identified in database literature searches, visualised in VOSviewer in CitNetExplorer. The top panel indicates clustering analyses performed in VOSviewer, which identified nine key groups, labelled 1 through 9. The bubbles indicate key authors labelled by surnames and initials with bubble size corresponding to the number of citation links with other authors. The bottom panel indicates citation network analyses of publications in CitNetExplorer, which are organized by year (2002–2023) with the name and first initial of the first author indicating individual studies. The relationship between studies by virtue of co-citations in the reference lists are indicated by grey lines. Subgroup analyses identified several key clusters, indicated according to the group colours from VOSviewer. (Image edited in BioRender.com).

and the product of group sizes.

$$d = \sqrt{f \times \left.(n_1 + n_2)\right/(n_1 \times n_2)} \tag{3}$$

$$r = \left.d\right/\sqrt{d^2 + \left.(n_1 + n_2)^2\right/(n_1 \times n_2)} \tag{4}$$

For t-test *(t)* statistics, equation 5 was used to calculate the correlation by taking the square root of the quotient between (i) the square of the *t-value* and (ii) the sum of *t-squared* added to *sample size (n)*, subtracting two.

$$r = \sqrt{\left.t^2\right/\left(t^2 + n - 2\right)} \tag{5}$$

where only the *probability (p-value)* was available, the correlation was derived from first calculating *Cohen's d* using the quantile function of the Student's t-distribution (where q represents the quantile function, α represent the desired significance level from the *p-value*, and *df* represents the degrees of freedom from the sample size) and multiplying by the square root of the quotient for the sum of group sizes and the product of group sizes (equation 6). Hereafter the correlation could be calculated as per equation 4.

$$d = q_{df}^{\alpha} \times \sqrt{\left.(n_1 + n_2)\right/(n_1 \times n_2)} \tag{6}$$

The variance associated with individual correlations, var_r, was calculated using equation 7 by dividing (i) the square of *r-squared* subtracted from one by (ii) the *sample size (n)* minus one.

$$var_r = \left.\left(1 - r^2\right)^2\right/(n - 1) \tag{7}$$

This data was incorporated into a systematic review and meta-analysis as previously published [1].

4. Limitations

Tests for funnel plot asymmetry as a measure of potential publication bias were statistically significant (p-value < 0.05), indicating the potential absence of studies from the primary literature. This is likely due to small study effects, where studies with smaller sample sizes are likely excluded from publication through peer-review. The low levels of studies reported from the global South and Africa may also indicate that similar studies from these regions are missing from the primary literature, possibly due to a lack of priority or suitable resources. The trim-and-fill method was used to infer possibly missing studies and, while this did not substantially alter the overall interpretations of the results, 18 studies were inferred and added as missing from the methylation dataset while 49 studies were inferred and added as missing from the telomere studies. Additionally, it should be noted that the methylation dataset contained an abundance of studies in mammals, with fewer studies from other vertebrate classes, while the telomere studies included an abundance of studies in birds. There was, however, little evidence that vertebrate class had a significant effect on the measured attributes. Lastly, differences in timespan for publications may impact interpretations as reported effect sizes are known to show temporal variation; typically studies that show large effect sizes are published first, establishing the validity of a method or introducing the field, however, later studies—with smaller effect sizes—are often published a decade later. Given that methylation studies have only been published over half of the same period than for telomere studies, future studies may still be published for methylation showing lower effect sizes.

5. Ethics Statement

Data used in our review represent secondary data from published literature and online resources. Included studies complied with the ARRIVE guidelines and were carried out in accordance with the United Kingdom (UK) Animals (Scientific Procedures) Act, 1986 and associated guidelines; European Union Directive 2010/63/EU for animal experiments; the National Institute of Health (UK) guide for the care and use of laboratory animals (NIH Publications No. 8023, revised 1978); or other ethical guidelines as per the country of origin. Ethics approvals for the present study were obtained from the University of the Free State (approval number: UFS-AED2020/0015/1709) as well as the South African National Biodiversity Institute (approval number: SANBI/RES/P2020/30).

Data Availability

Biological clocks (Data) (Reference data) (Zenodo).

CRediT Author Statement

Louis-Stéphane Le Clercq: Conceptualization, Methodology, Data curation, Formal analysis, Writing – original draft, Writing – review & editing, Funding acquisition; **J. Paul Grobler:** Writing – review & editing, Resources, Supervision; **Antoinette Kotzé:** Writing – review & editing, Resources, Supervision; **Desiré Lee Dalton:** Conceptualization, Writing – review & editing, Resources, Supervision.

Acknowledgements

Open access provided by SANBI. Images were created/edited in BioRender.com. This work is based on the research supported wholly/in part by the National Research Foundation of South Africa (Grant Number: 112062).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] L.Le Clercq, A. Kotzé, J.P. Grobler, D.L. Dalton, Biological clocks as age estimation markers in animals: a systematic review and meta-analysis, Biol. Rev. (2023), doi:10.1111/brv.12992.
[2] D. Moher, A. Liberati, J. Tetzlaff, D.G. Altman, Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement, Int. J. Surg. 8 (2010) 336–341, doi:10.1016/J.IJSU.2010.02.007.
[3] N.R. Haddaway, M.J. Page, C.C. Pritchard, L.A. McGuinness, PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis, Campbell Syst. Rev. 18 (2022) e1230, doi:10.1002/cl2.1230.
[4] R.E. O'Dea, M. Lagisz, M.D. Jennions, J. Koricheva, D.W.A. Noble, T.H. Parker, J. Gurevitch, M.J. Page, G. Stewart, D. Moher, S. Nakagawa, Preferred reporting items for systematic reviews and meta-analyses in ecology and evolutionary biology: a PRISMA extension, Biol. Rev. 96 (2021) 1695–1722, doi:10.1111/brv.12721.
[5] J.P. Higgins, S. Green, Cochrane handbook for systematic reviews of interventions: cochrane book series, John Wiley and Sons, 2008, doi:10.1002/9780470712184.
[6] RStudio Team, RStudio: Integrated Development environment for R, (2021). http://www.rstudio.com/.
[7] R Core Team, R: a language and environment for statisitical computing, (2020). https://www.r-project.org/.
[8] N.J. van Eck, L. Waltman, Citation-based clustering of publications using CitNetExplorer and VOSviewer, Scientometrics 111 (2017) 1053–1070, doi:10.1007/s11192-017-2300-7.
[9] A. Rohatgi, WebPlotDigitizer: Version 4.6, (2022). https://automeris.io/WebPlotDigitizer.

[10] H. Qi, K. Kinoshita, T. Mori, K. Matsumoto, Y. Matsui, M. Inoue-Murayama, Age estimation using methylation-sensitive high-resolution melting (MS-HRM) in both healthy felines and those with chronic kidney disease, Sci. Rep. 11 (2021) 1–10, doi:10.1038/s41598-021-99424-4.
[11] G.M. Baerlocher, J. Mak, A. Röth, K.S. Rice, P.M. Lansdorp, Telomere shortening in leukocyte subpopulations from baboons, J. Leukoc. Biol. 73 (2003) 289–296, doi:10.1189/JLB.0702361.
[12] J. Tackney, R.M. Cawthon, J.E. Coxworth, K. Hawkes, Blood cell telomere lengths and shortening rates of chimpanzee and human females, Am. J. Hum. Biol. 26 (2014) 452–460, doi:10.1002/ajhb.22538.
[13] L. Panasiak, S. Dobosz, K. Ocalewicz, Telomere dynamics in the diploid and triploid rainbow trout (*Oncorhynchus mykiss*) assessed by q-fish analysis, Genes 11 (2020) 1–12, doi:10.3390/genes11070786.
[14] B. Apfelbeck, M.F. Haussmann, W. Boner, H. Flinks, K. Griffiths, J.C. Illera, K.G. Mortega, Z. Sisson, P. Smiddy, B. Helm, Divergent patterns of telomere shortening in tropical compared to temperate stonechats, Ecol. Evol. 9 (2019) 511–521, doi:10.1002/ECE3.4769.
[15] C. Román-Palacios, C.A. Medina, S.H. Zhan, M.S. Barker, Animal chromosome counts reveal a similar range of chromosome numbers but with less polyploidy in animals compared to flowering plants, J. Evol. Biol. 34 (2021) 1333–1339, doi:10.1111/jeb.13884.
[16] T.M. Degrandi, S.A. Barcellos, A.L. Costa, A.D.V. Garnero, I. Hass, R.J. Gunski, Introducing the bird chromosome database: an overview of cytogenetic studies in birds, Cytogenet. Genome Res. 160 (2020) 199–205, doi:10.1159/000507768.
[17] T.R. Gregory, Animal genome size database, (2023). http://www.genomesize.com (accessed July 8, 2022).
[18] G. Zhang, Bird sequencing project takes off, Nature 522 (2015) 34–34, doi:10.1038/522034d.
[19] A.C. Del Re, compute.es: compute effect sizes, R Package. (2013). https://cran.r-project.org/package=compute.es.
[20] H.M. Cooper, L.V. Hedges, J.C. Valentine, The handbook of research synthesis and meta-analysis, 2nd ed., Russell Sage Foundation, 2009, doi:10.7758/9781610441384.

Scientometrics (2024) 129:581–600
https://doi.org/10.1007/s11192-023-04880-6

ABCal: a Python package for author bias computation and scientometric plotting for reviews and meta-analyses

Louis-Stéphane Le Clercq[1,2]

Received: 4 August 2023 / Accepted: 10 November 2023 / Published online: 26 November 2023

Abstract
Systematic reviews are critical summaries of the exiting literature on a given subject and, when combined with meta-analysis, provides a quantitative synthesis of evidence to direct and inform future research. Such reviews must, however, account for complex sources of between study heterogeneity and possible sources of bias, such as publication bias. This paper presents the methods and results of a research study using a newly developed software tool called ABCal (version 1.0.2) to compute and assess author bias in the literature, providing a quantitative measure for the possible effect of overrepresented authors introducing bias to the overall interpretation of the literature. ABCal includes a new metric referred to as author bias, which is a measure of potential biases per paper when the frequency or proportions of contributions from specific authors are considered. The metric is able to account for a significant portion of the observed heterogeneity between studies included in meta-analyses. A meta-regression between observed effect measures and author bias values revealed that higher levels of author bias were associated with higher effect measures while lower author bias was evident for studies with lower effect measures. Furthermore, the software's capabilities to analyse authorship contributions and produce scientometric plots was able to reveal distinct patterns in both the temporal and geographic distributions of publications, which may relate to any evident publication bias. Thus, ABCal can aid researchers in gaining a deeper understanding of the research landscape and assist in identifying both key contributors and holistic research trends.

Keywords Reviews · Authorship · Bias detection · Author bias · Scientometric · Plots

Introduction

Scientific studies based on the empirical method contribute dozens of new publications pertaining to a central hypothesis under investigation on an annual basis. Over time, however, it may become apparent that the evidence in support of, or opposing, fundamental views within a discipline may not be unanimous (Le Clercq et al., 2023a, 2023b, 2023c)

✉ Louis-Stéphane Le Clercq
leclercq.l.s@gmail.com

[1] South African National Biodiversity Institute, Pretoria 0001, South Africa

[2] Department of Genetics, University of the Free State, Bloemfontein 9300, South Africa

and can confound the overall interpretation of primary findings. As existing evidence serves as the foundation for informing the direction of prospective studies, however, scientists are often faced with the difficult task of reading and interpreting the literature to derive central tenets for their subject or discipline (Boell & Cecez-Kecmanovic, 2010; Webster & Watson, 2002). One convenient method to synthesise and assess the existing evidence is through systematic review and meta-analysis.

Systematic reviews were first developed as a tool for the synthesis of evidence for causality (Le Clercq et al., 2016) or treatment (Honvo et al., 2019) in medical research and can be defined as: "a review using a systematic method to summarize evidence on questions with a detailed and comprehensive plan of study" (Tawfik et al., 2019). Thus, a systematic review seeks to identify and critically evaluate all those studies (Dickersin et al., 1994) pertaining to a specific research question for the purposes of deriving conclusions from the full body of evidence rather than relying on individual studies alone. Furthermore, systematic reviews attempt to standardise the methods (Moher et al., 2010; O'Dea et al., 2021) used to identify and screen studies in a way that is comprehensive, transparent, and above all reproducible and can serve as independent studies (Kraus et al., 2022). This avoids some of the pitfalls and biases that could influence narrative reviews (Pae, 2015; Tawfik et al., 2019). Another advantage of systematic reviews is the possibility to perform a meta-analysis and scientometric assessment of the included studies (Nakagawa et al., 2023). This is done using primary reported statistics to derive the effect size (Cohen, 1988) or treatment effect (TE), and variance or standard error (SETE), of the measured outcome. This facilitates between-study comparisons and enables the calculation of a pooled effect through a fixed- or random effects model (Borenstein et al., 2010). The pooled effect, therefor, serves as a quantitative measure of the total evidence.

Meta-analysis is not without possible confounders: several factors could contribute to between study differences, called heterogeneity (Higgins & Thompson, 2002), or be a source of bias (Felson, 1992; Sterne et al., 2001). Heterogeneity is expressed by two statistics, the heterogeneity measure (I^2) and the between-study variance or tau-squared (τ^2). The I^2 measure expresses the percentage of total variance in the effect sizes that is explained by between-study variance. The τ^2 approximates between-study variances but is reliant upon the specific effect sizes and needs to be quantitated by a P-value (Higgins, 2008). As heterogeneity could potentially reduce the ability to compare or combine the outcomes from all studies that meet inclusion criteria, it is critical for authors to identify possible sources of heterogeneity and attempt to account for them.

Common factors that contribute to heterogeneity include sample size, quantitative method, and study population. Two approaches can be used to account for these variables: meta-analysis with subgroups and meta-regression of factors. The first, meta-analysis with subgroups, determines if effect sizes and their corresponding variances differ between subgroups (Borenstein & Higgins, 2013). In the case of study populations representing different species, as may be the case in reviews in animal sciences or ecology, several methods have also been developed to account for phylogeny (Chamberlain et al., 2012) and taxonomy by performing a phylogenetic meta-analysis (Adams, 2008; Lajeunesse, 2009). The second approach, meta regression, determines if a significant part of the heterogeneity can be accounted for my individual study attributes, which may be more useful for continuous variables such as sample size (Baker et al., 2009).

The other, and perhaps more difficult, task is to identify and quantify potential sources of bias (Boutron et al., 2019; Felson, 1992; Sterne et al., 2001). The most (Bouyssou & Marchant, 2016; Perianes-Rodriguez et al., 2016; Zhou & Leydesdorff, 2010) established form is publication bias (Lortie et al., 2007; Møller & Jennions, 2001; Thornton & Lee, 2000). This form of bias is detected through funnel-plots or through weighed linear models and focusses on detecting small study effects (Egger et al., 1997; Sterne et al., 2001). This includes the absence of studies that have smaller sample sizes, possibly due to the difficulties associated with the peer-review and publishing of such studies, as well as the inclusion of studies with small sample sizes and very low variance. There is, however, another related source of bias that has received little to no recognition thus far–bias from the over-representation of studies from specific authors (Ausloos, 2013); hereafter, author bias.

Author bias, when not accounted for, has the inherent ability to skew the overview and interpretation of the literature in several ways. In the first, a specific view may be held by a particular group of authors who publish at a much higher frequency than other scientists in their field (Lortie et al., 2007), resulting in many publications in support of a view from a narrow pool of authors. This may create the illusion that opinions reported in their papers represent a majority consensus even when few independent studies support their claims. Secondly, views derived from primary findings based on a novel and 'in-house' method may not be fully reproducible if no independent studies exist where other authors repeated and confirmed the validity of such methods. This can further be confounded by the fact that negative or disconfirming results are often published at a delay (Boutron et al., 2019) or in less prominent journals (Leimu & Koricheva, 2004). Lastly, a majority of studies may have been conducted in a specific country, region (Collyer, 2018), and context (Fohringer et al., 2022) which–in cases where study populations may vary significantly between regions–may result in interpretations and generalisations that aren't universal. This makes the scientometric analysis of studies by author, year, and location, critical in providing an appraisal of the literature.

At present, the most common methods used to access author contributions are the use of fractional citation counts (Bouyssou & Marchant, 2016; Perianes-Rodriguez et al., 2016; Zhou & Leydesdorff, 2010). This method emerged as a practical approach in response to the various complexities of assessing the contribution levels within scholarly works attributed to specific authors. While this method has proven useful in illuminating prominent contributors in differing fields (Bedru et al., 2023; Small & Garfield, 1985) and in citation network analyses (Perianes-Rodriguez et al., 2016), no clear link has been made between scores for individual authors from fractional counting and bias introduced in reviews from contribution levels. Furthermore, many of the methods that have been described still lack available software that implements the method (Bedru et al., 2023), or are available with very limited functionality (Keirstead, 2016; Kozlowski, 2019). To address the current need to quantitate author contribution levels as a measure of bias, and perform scientometric checks on publication year and location, ABCal (version 1.0.2) was created to compute author bias and plot scientometric aspects of studies included in systematic reviews and meta-analyses. In this paper, a full description is given of how the author bias metric is computed along with examples of how ABCal can be used to evaluate potential sources of bias using real data from two datasets from a recent systematic review (Le Clercq et al., 2023a, 2023b, 2023c).

Methods

Author bias metric

To assess relevant attributes of included studies, ABCal includes a new metric referred to as "author bias" to provide a quantitative measure for the possible effect of overrepresented authors introducing bias to the overall interpretation of the literature. This measure is derived in several steps. First, the full list of authors (L_{All}) for all included studies is used to determine the total number of times, n_{Author}, the name of a specific author occurs when iterating through each position from $i = 1$ to $\#L_{\text{All}}$ in the list (Eq. 1).

$$n_{\text{Author}} = \sum_{i=1}^{n=\#L_{\text{All}}} f_i(Author) \tag{1}$$

This value is then divided by the total number of authors in the list, $\#L_{\text{All}}$, for which provides the individual author bias, $AB_{\text{Individual}}$, as the proportion of total authorship contributions that belong to individuals (Eq. 2).

$$AB_{\text{Ind}} = n_{\text{Author}} / \#L_{\text{All}} \tag{2}$$

Next the bias derived from authorship is calculated per study (AB_{Study}) by adding the individual bias values, AB_{Ind}, for each author in the author list for a specific study L_{Paper}, from $i = 1$ to the $\#L_{\text{Paper}}$, per Eq. 3.

$$AB_{\text{Study}} = \sum_{i=1}^{n=\#L_{\text{Paper}}} AB_{\text{Ind}}(i) \tag{3}$$

To facilitate the interpretation of these values, the final steps are calibration (Eq. 4), by dividing total bias per study, AB_{Study}, by the number of authors per paper, $\#L_{\text{Paper}}$.

$$AB_{\text{Calibrated}} = AB_{\text{Study}} / \#L_{\text{Paper}} \tag{4}$$

These are further categorised by assessing the distribution of author bias values by identifying those studies that fall in the bottom, middle, and upper third range, or percentiles (Eq. 5) of thirty-three, to assign bias status as being low, medium, or high based on the calculated quantiles.

$$p = x_{ri} + r_f * \left(x_{ri+1} - x_{ri}\right) \tag{5}$$

The percentile (p) is interpolated by adding the value for x at position ri to the product of the position (r_f) and the difference between the values for x at position $ri + 1$ and ri. ABCal also provides some functionality to assess the normality of the author bias values using three approaches: a Shapiro–Wilk test (Shapiro & Wilk, 1965), a Quantile–Quantile (QQ) plot (Wilk & Gnanadesikan, 1968), and a histogram of distributions. The plotting sub-menu also provides an option for plotting the distribution of z-score transformed bias values which is useful in comparing distributions for different meta-analysis datasets.

The performance of the author bias metric was assessed for both validity and reliability (Cohen et al., 2017; Hammersley, 1987). Validity, in this context, refers to the ability of the metric to accurately measure the intended attribute. This was verified by comparing those studies for which a higher author bias value was computed to whether the authors

listed on the paper ranked within the top 10 contributing authors for the field. Agreement was measured in *R 4.0.6* (R Core Team, 2020) using Cohen's kappa (Cohen, 1960) with the *vcd 1.4–11* package (Meyer et al., 2023). Reliability or repeatability of the measure was assessed by comparing the paper level calibrated author bias levels coded as low (1), medium (2), or high (3), for two different datasets (Le Clercq et al., 2023a, 2023b, 2023c). This was done by assessing their independent validity as well as internal consistency between the distributions using Cronbach's alpha (Cronbach, 1951) with the *ltm 1.2–0* package (Rizopoulos, 2006).

Implementation

ABCal, version 1.0.2 (Le Clercq, 2023), was scripted in the Spyder 5 IDE using the PYTHON 3 language (Python Team, 2021) and should be compatible with all versions upward of version 3.6. The list of packages that form part of the dependencies is provided on the GitHub repository and within the README file, along with detailed instructions for the download and installation. Dependencies include the use of several core PYTHON based libraries such as *NumPy 1.20.1* and *pandas 1.2.4* to handle input files and mould data structures for analyses (Harris et al., 2020; McKinney, 2010). Other dependencies include packages for statistical analyses, such as *SciPy 1.6.2* (Virtanen et al., 2020) and *statsmodels 0.12.2* (Seabold & Perktold, 2010), and packages for graphical plotting, such as *matplotlib 3.3.4* (Barrett et al., 2005) and *GeoPy 2.3.0* (Lopez Gonzalez-Nieto et al., 2020). ABCal further uses the plotting functionality implemented in *folium 0.14.0* with selected functions from the IO tools, for input and output, as well as the PYTHON Image Library (*PIL 10.0.0*). Menu options (detailed in Section "Usage") provide the utilities to calculate author bias, test the distributions of author bias values, and generate several scientometric plots. Scientometric plotting options include the ability to plot publications by the top contributing authors (to identify and visualise the extent to which top authors may skew overall interpretation), by year, and by location.

Input and output file formats

All input files used by ABCal are in the standard comma separated value (CSV) format. For most functions the first column of these files should contain the heading "Paper" with the studies listed by name in the column e.g., "Le Clercq et al. (2023a, b, c)". To calculate the author bias, the CSV file should contain columns for each author, first to last, labelled with appropriate headings such as "Author1" etc. These columns should contain the last name and initials of each author associated with a specific paper e.g., "Le Clercq, L.S". The function to compute author bias moves through several steps to perform each intermediate step to derive the values and provides intermediate output files with relevant measures for later steps or scientometric plotting, detailed under the usage section. These files provide the option to specify unique output file names and are stored as CSV files within the current working directory.

For most of the plotting options, either the CSV files generated from author bias computation or CSV files containing additional study attributes for plotting are used. An example of such a file to plot the distribution of publications by year, includes a file containing two columns with the headers "Paper" and "Year", which should contain the study name as well as the year of publication. Another example is for the plotting of studies by location, where a CSV is required containing two columns with the headers "Paper" and "Location".

For this file, the full name of the country in which the study was conducted is required; if more than one location was included these should be listed on separate lines with the study name and second or third location. The output generated from plotting is saved with a standard name for the type of plot in the portable network graphic (PNG) format, with the exception of the location plot which is also stored in the interactive HTML format.

Usage

To illustrate the usage of ABCal, two datasets (Le Clercq et al., 2023a, 2023b, 2023c) generated as part of a recent systematic review on biomarkers for age in animals (Le Clercq et al. 2023c), comprising age models from included studies on the use of methylation ($N=41$ studies, 60 models) and telomeres ($N=67$ studies, 99 models) respectively, will be used. For each dataset, three input files were generated: one with the paper name and list of authors, a second with the paper name and publication date, and a third with the paper name and study location.

The first file was used to compute the author bias (example A). Once ABCal is initiated, the first option (a) is to calculate the total author bias per paper. Selecting this option initiated the function to perform the needed steps to do the calculation. A prompt appeared to specify the name of the file containing the author lists e.g., "Auth_Meth.csv". The first step generated a list of all authors along with the total counts of times the specific author appeared in an author list. This data was exported as the first file for output from the function and was saved as a CSV containing the number of publications per author. Next, the individual author bias was computed by determining the fraction of total authorship contributions per author. These values were stored in the second output and contained the author names and their associated individual bias. The final step computed the total author bias per paper by adding the individual bias value of each author associated with the author list for a paper. This data was saved as the third output and contained the data used to compute total bias per paper. As an additional step, and to assist in the interpretation of values, the second menu option (b), which takes the final output file from the first option, was used to calibrate the bias value by dividing the total bias by the number of authors per paper. The newly calibrated values were exported as the fourth output file. Furthermore, the third menu option (c) was used for testing the distributions for normality and the fourth menu option (d) was used to get the upper, middle, and lower third quantiles of the author bias distributions.

Hereafter, the author bias values and their respective levels were incorporated into two meta-analyses as part of a review (Le Clercq et al. 2023c). The meta-analysis was done in *RStudio 1.4.1106* (RStudio Team, 2021), running *R 4.0.5* (R Core Team, 2020) with the package *meta 5.5–0* (Harrer et al., 2021; Schwarzer et al., 2015). A meta-regression was done between the random effects model and author bias as a predictor of heterogeneity for a functional test of validity. The results were visualised using a bubble plot implemented in the *metafor 3.8–0* package (Viechtbauer, 2010) with grouping based on the three quantiles. Furthermore, potential publication bias as measured by funnel plot asymmetry was also assessed using the Egger's test (Egger et al., 1997) as implemented in *metafor 3.8–0*, also plotting the relationship between the standardized measured effect and the inverse of the standard error.

Scientometric plotting capabilities of ABCal, accessed via a submenu when selecting option e, were illustrated in example B for methylation studies and example C for telomere studies. For the first option (a) from the submenu, the second file with two columns for

'Paper' and 'Year' was used to plot the total number of publications per year. The first file given as output from the author bias computation steps, containing the authorship counts, was used for the second option (b) to plot the number of contributions from the top ten contributing authors. Lastly, the third file containing two columns for 'Paper' and 'Location' was used to plot a choropleth map of the geographical distribution for study locations using the third option (c) from the submenu.

Results

Example A: author bias computation

Author bias values per paper for methylation studies ranged from 0.0024 to 0.0302 with a mean of 0.0124 (Table 1). The histogram plot of distributions (Fig. 1A) showed many values ($N=19$) were well below the mean, skewing the distribution left, with a moderate number of studies ($N=10$) falling in and around the mean and few studies ($N=12$) having higher values. The position for setting the first (Q1) and third (Q3) quartiles were 0.004 and 0.018 respectively, with 14 studies classified as low, 13 as medium, and 14 as high. The box plot of z-score transformed values (Fig. 1B), to express the bias values in terms of standard deviations from the median, showed that the median was low (approximately − 0.094) with most studies (95%) being evenly distributed around the median. A few studies had higher values; however, they still fell within two standard deviations of the median and no clear outliers were detected. The overall distributions were found to not be normally distributed (Table 1; $P<0.01$). Tests for validity by means of Cohen's kappa showed high

Table 1 Summary of characteristics of calibrated author bias values

	Mean	SE	Min	Max	Q1	Q3	Shapiro–Wilk
Methylation							
Author bias	0.014	0.001	0.002	0.030	0.004	0.018	0.869 ($P<0.01$)
Z-score	-1.08×10^{16}	0.16	− 1.07	1.90	− 1.00	0.75	
(1) Low	14						
(2) Medium	13						
(3) High	14						
Telomeres							
Author bias	0.004	0.0001	0.002	0.012	0.003	0.004	0.780 ($P<0.01$)
Z-score	-2.10×10^{16}	0.12	− 0.79	4.13	− 0.79	0.44	
(1) Low	23						
(2) Medium	21						
(3) High	23						

Values are reported for both the calibrated author bias (AB) values as well as the Z-Score transformed values. The mean, standard error of the mean (SE), minimum (Min), maximum (Max) and positions of the cut-off points for the lower (Q1) and upper (Q3) third percentiles are given. The results for the normality test, tested using the Shapiro–Wilk test, are also given along with the significance. The calibrated author bias for both datasets was not normally distributed ($P<0.01$). The number of studies identified by level as having either low, medium, or high risk of bias are also indicated

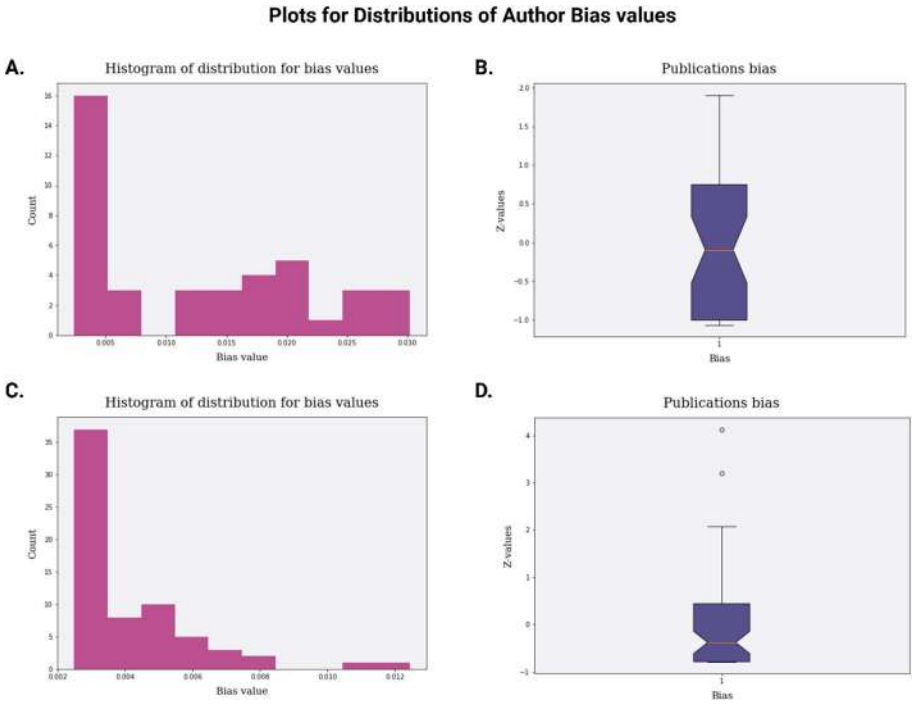


Fig. 1 Plots for the distributions of author bias values. **A** Histogram for the calibrated author bias values for papers in the methylation dataset indicating many studies ($N=16$) with low values, skewing the distribution left, with a moderate number of studies ($N=4$–6) around the median and a similar number of studies ($N=3$–6) with high values. **B** Box plot for the *Z*-score transformed author bias values for papers in the methylation dataset indicating most studies are evenly distributed around the median (orange line) with few studies more than one standard deviation from the median, and a small number of studies in the upper range without appearing as outliers. **C** Histogram for the calibrated author bias values for papers in the telomere dataset indicating many studies ($N=25$) with low values, skewing the distribution left, with a moderate number of studies ($N=5$–10) around the median and two studies with higher values. **D** Box plot for the *Z*-score transformed author bias values for papers in the telomere dataset indicating most studies were slightly above the median (orange line) with most studies within one standard deviation from the median, and a small number of studies in the upper range and two outliers. (image created in BioRender.com)

levels of agreement (Table 2; $\kappa=0.94$, $P<0.01$) between studies ranked as having medium to high risk of bias as compared to the list of top contributing authors.

Author bias values for telomere studies ranged from 0.0024 to 0.0124 with a mean of 0.0040 (Table 1). The position for setting the first (Q1) and third (Q3) quartiles were 0.003 and 0.004 respectively, with 23 studies classified as low, 21 as medium, and 23 as high. The values followed a similar pattern to that observed for methylation studies (Fig. 1C), with a large number of studies ($N=48$) falling below the mean. Most of the remaining studies followed a near bell shape around the mean, with several ($N=9$) having higher values. The box plot (Fig. 1D) showed a slightly higher number of studies fell above the median, while most studies were still within one standard deviation of the median. Several studies had values between one and two standard deviations of the median, with two studies that were more than two standard deviations from the median and thus detected as outliers. Once more, the distributions were found to not be normally distributed (Table 1; $P<0.01$). Tests for validity showed a moderate, yet significant, level of agreement (Table 2; $\kappa=0.52$,

Table 2 Results from tests for validity, by mean of Cohen's kappa (κ), as well as reliability, by means of Cronbach's alpha (α), for the Author Bias metric

	Value	SE	*z*-value	Confidence
Validity				
Methylation	κ: 0.94	0.05	17.36	$P<0.01$
Telomeres	κ: 0.52	0.13	3.89	$P<0.01$
Between	κ: 0.73	0.09		Good
Reliability				
Between	α: 0.84	–	–	CI 0.74–0.91

Validity was calculated for both the methylation and telomere dataset by assessing the agreement between bias rank and authors listed as the top ten contributors. The overall validity was determined by taking the average for individual results. Agreement for ranking between datasets was used to assess between study reliability of the metric. The confidence of the calculated values was assessed by either probability ($P<0.01$) of a z-test or the 98% confidence interval (CI)

$P<0.01$) and an average validity between datasets of 0.73. The reliability tests between datasets also showed a significant (Table 2; $\alpha=0.84$, CI 0.74–0.91) level of reproducibility for levelled classification of studies.

A meta-regression between observed effect measures and author bias values revealed that author bias values were able to account for a significant portion of the observed heterogeneity between studies included in meta-analyses. These relationships are illustrated as bubble plots in Fig. 2. For methylation studies, a strong relationship was observed ($P<0.01$) with author bias values accounting for approximately 23% ($R^2=0.23$) of the study heterogeneity. For telomere studies, a slightly weaker relationship was observed ($P<0.02$) with author bias values accounting for 6% ($R^2=0.06$) of the study heterogeneity. In both instances, higher levels of author bias were associated with higher effect measures while lower author bias was evident for studies with lower effect measures. Tests for publication bias (Fig. S1) detected significant funnel plot asymmetry ($P<0.05$) indicative of possible publication bias. Statistical methods to address publication bias, such as "trim-and-fill" or linear modelling of a fixed-effect model with factorisation, did not significantly alter the overall interpretations from the comparisons (data not shown).

Example B: a meta-analysis of methylation studies

Scientometric assessment of studies included in the methylation dataset was done by plotting three attributes: top contributing authors, publications by year, and publications by location. For publications per author, the number of publications contributed by the top contributing authors (specified as ten) ranged from 3 contributions to a total of 24 contributions (Fig. 3A). Five authors, including Zhang, contributed 3 papers each, respectively. The top three contributing authors, identified as Horvath, Haghani, and Zoller, contributed to approximately half (21–24 out of 41) of the included studies. The bar plot for publication by year (Fig. 4A) revealed the first studies were published in 2014 (Polanowski et al., 2014) with an annual increase leading to 15 publications in 2021 (Bors et al., 2021; Mayne et al., 2021; Robeck et al., 2021; Wilkinson et al., 2021) and several recent studies (Horvath et al., 2022a, 2022b; Horvath et al., 2022a, 2022b; Robeck et al., 2023). The choropleth map (Fig. 4B), showing study locations, showed

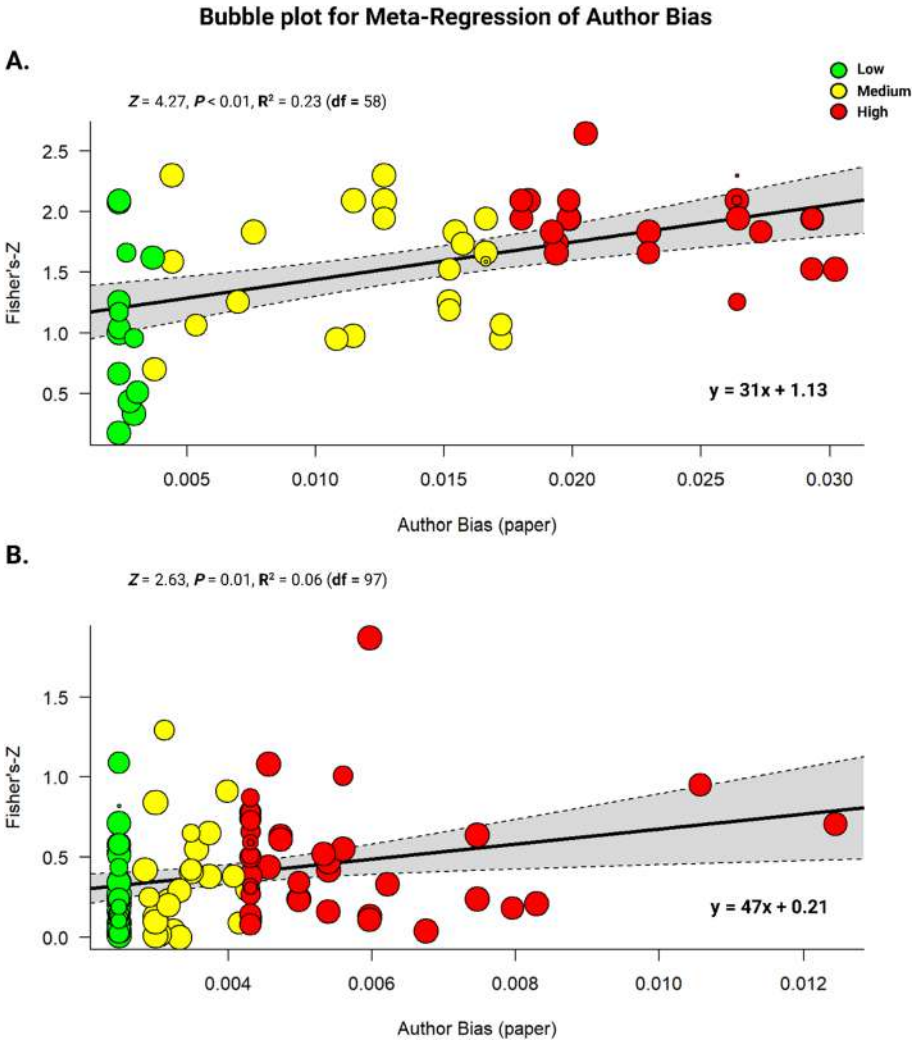


Fig. 2 Bubble plots for the meta-regression of author bias values as a predictor of heterogeneity in the meta-analyses. The Fisher's-Z values (y-axis) were plotted against the calibrated author bias values per paper (x-axis) and studies colour coded according to the quantiles within which they fell and were classified: low (green), medium (yellow), or high (red). The linear equations for the regressions are indicated in the bottom right of each plot. **A** Meta-regression performed on the methylation dataset found a significant association between effect measures and author bias values ($P < 0.01$), accounting for 23 percent of the observed heterogeneity. **B** Meta-regression for the telomere dataset found a moderate association ($P < 0.02$), accounting for only 6 percent of the heterogeneity. (image created in BioRender.com)

the number of studies per country ranged from one study (green) to more than twenty studies (red). Several countries, shown in white, were completely data deficient. The overall distribution showed that most studies emanated from the Northern hemisphere, principally from North America ($N > 20$) and Europe, with Australia ($N > 5$) representing the country with the most publications in the global South.

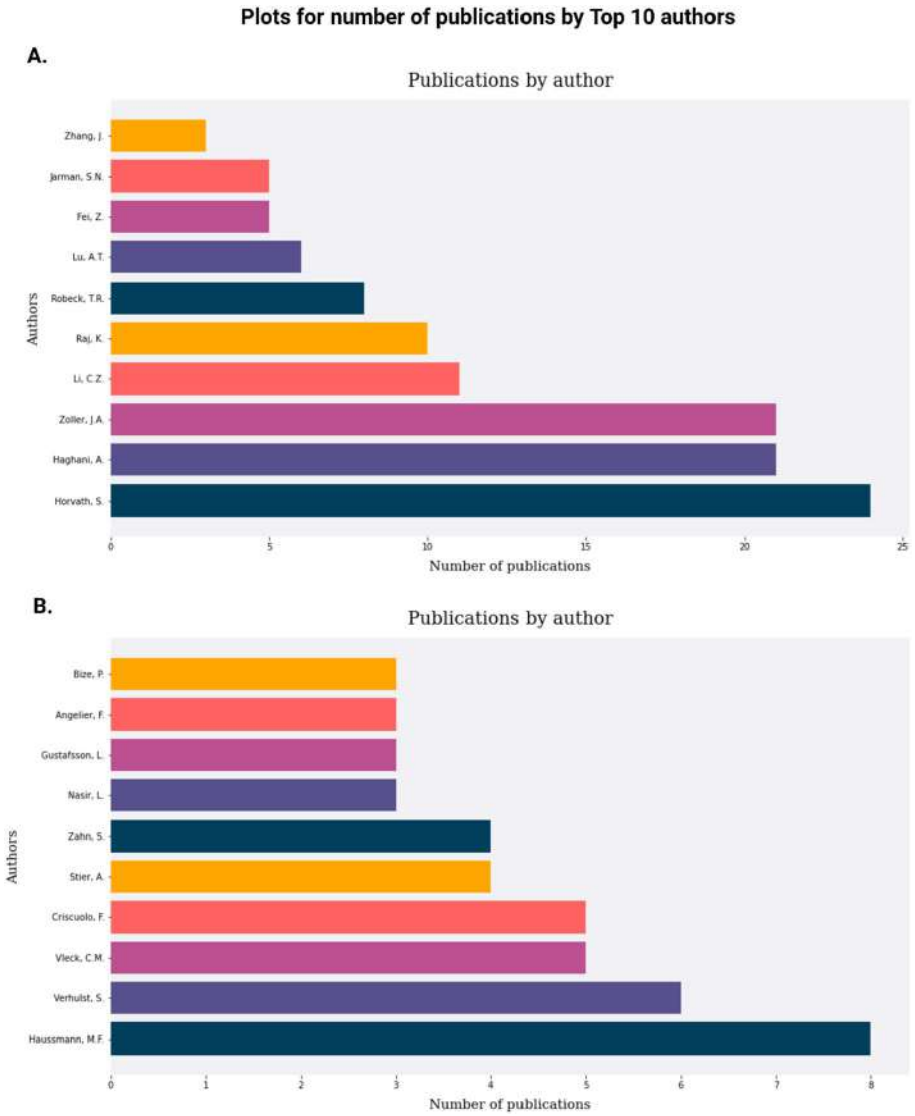


Fig. 3 Scientometric plots for the number of publications attributed to the top 10 contributing authors. **A** For methylation studies, the number of publications attributed to top contributing authors ranged from 4 contributions to as many as 24 contributions: with the top 3 contributing to a half (50%) of the included studies. **B** For telomere studies, the number of publications attributed to the top contributing authors ranged from 3 to 8 contributions: here, the top 3 contributing authors made up approximately a tenth (10%) of the total studies. (image created in BioRender.com)

Example C: a meta-analysis of telomere studies

The same scientometric plots were also generated for the telomere dataset. For publications per author, contributions by the top ten authors ranged from 3 contributions to a total of 8 contributions (Fig. 3B). Several authors contributed 3 papers while the top three authors,

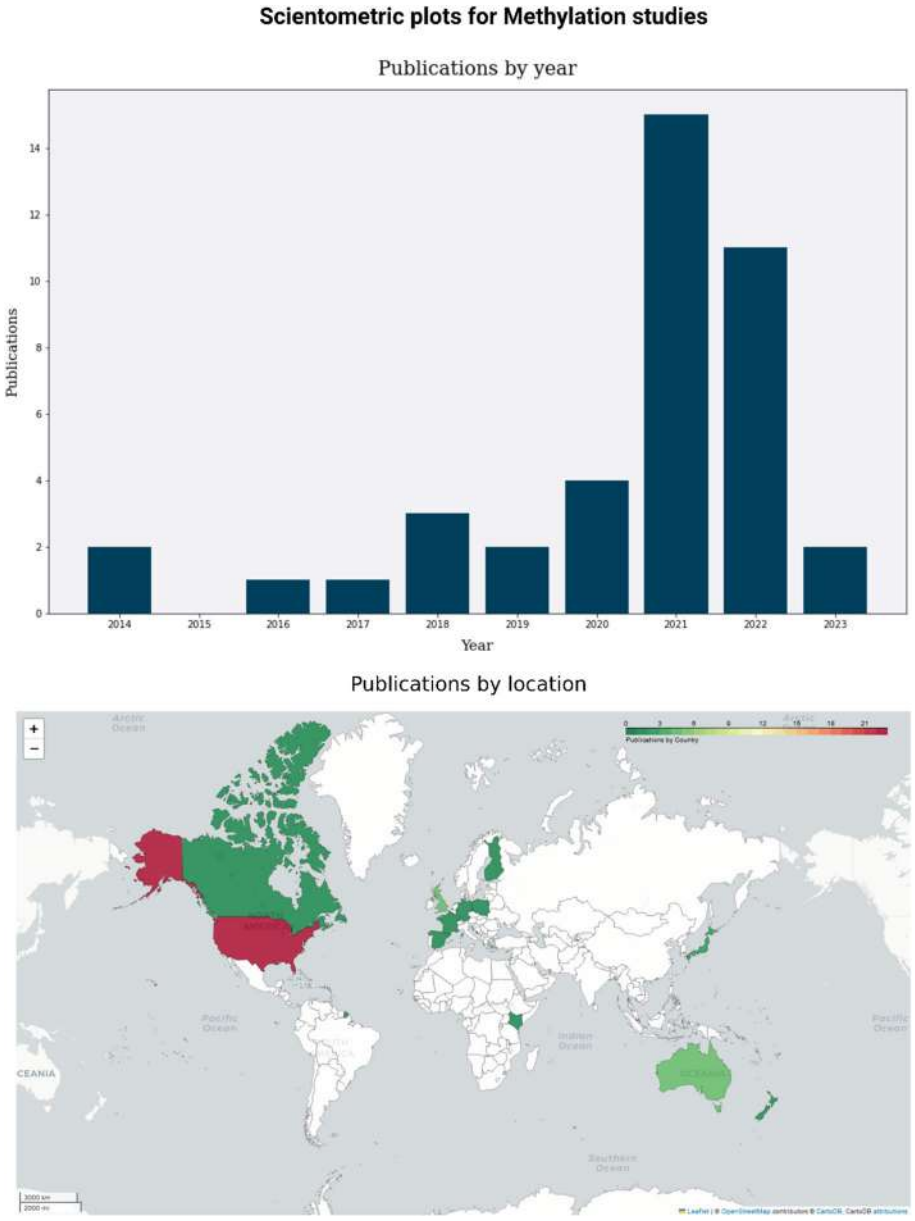


Fig. 4 Scientometric plots for methylation studies, generated with ABCal. **A** Bar plot for publication by year, indicating the first studies published in 2014 with an annual increase leading to 15 publications in 2021. **B** Choropleth map showing study locations. The density gradient plots the number of studies per country ranging from one study (green) to more than twenty studies (red); countries in white are data deficient. The overall distribution shows most studies are from the Northern hemisphere, principally from North America and Europe, as well as Australia. (image created in BioRender.com). (Color figure online)

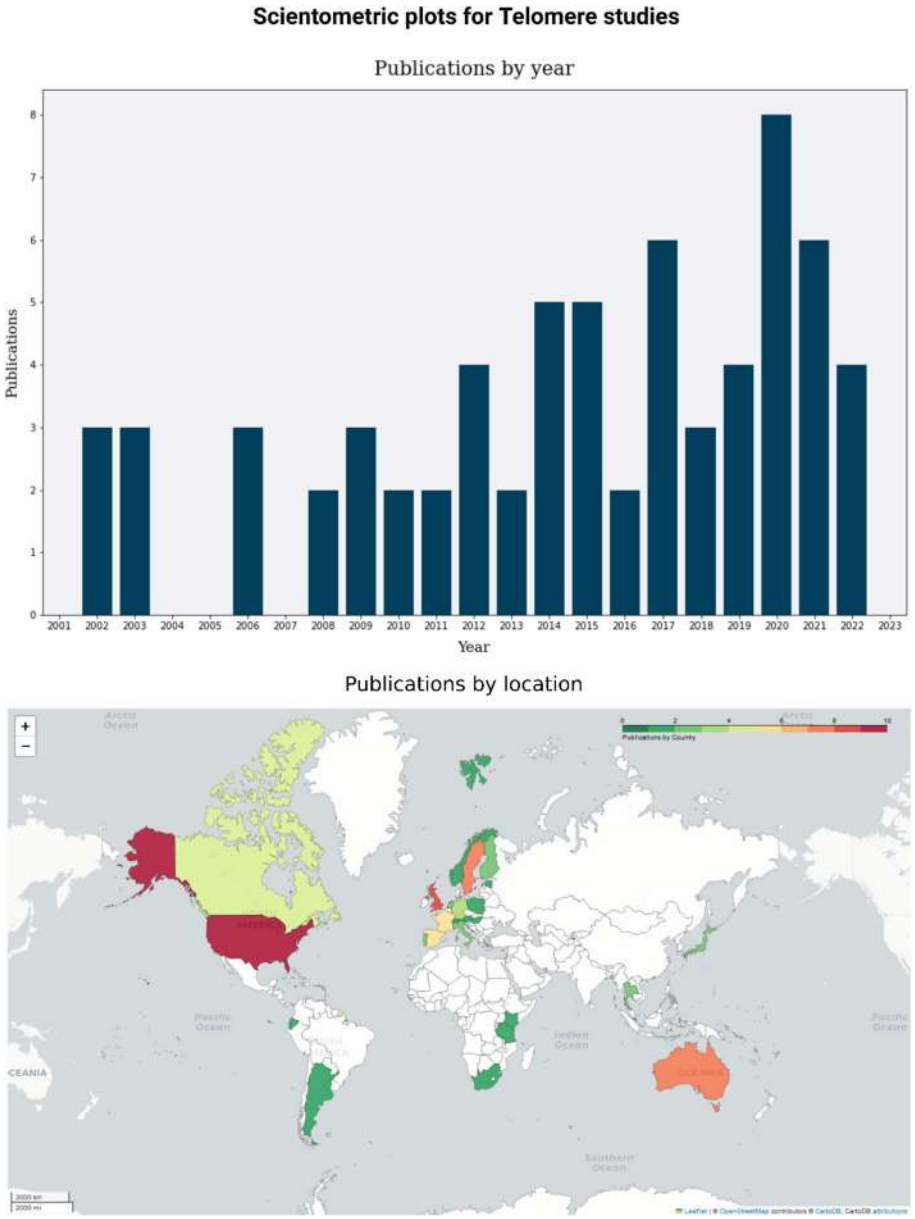


Fig. 5 Scientometric plots for telomere studies, generated with ABCal. **A** Bar plot for publications by year, showing the first included studies were published in 2002 with constant publication of 2–3 studies per year and increasing from 2012 with several spikes in 2017, 2020, and 2021 to between 6–8 publications. **B** Choropleth map showing study locations. The density gradient plots the number of studies per country ranging from one study (green) to ten studies (red); countries in white are data deficient. The distribution shows most studies are from the Northern hemisphere, principally from North America, as well as Australia. (image created in BioRender.com)

identified as Haussmann, Verhulst, Vleck, and Criscuolo, each contributed between 5 and 8 studies. This only accounted for about 10% of the included studies (5–8 out of 68). Publications by year, given as a bar plot (Fig. 5A), indicated the first included studies were published circa 2002 (Brümmendorf et al., 2002; Haussmann & Vleck, 2002) with frequent subsequent publications, around 2–3 studies annually, and an increase seen after 2012 (Fick et al., 2012; Plot et al., 2012) with several spikes in 2017 (Cerchiara et al., 2017; Kirby et al., 2017; Ujvari et al., 2017), 2020 (Bauch et al., 2020; Burraco et al., 2020; Cherdsukjai et al., 2020), and 2021 (Molbert et al., 2021; Vernasco et al., 2021). The maximum number for spikes ranged between 6 to 8 publications. Choropleth mapping of study locations as publications per country (Fig. 5B) ranged from one study (green) to ten studies (red); data deficient countries are indicated in white. The distribution showed most studies originated from the Northern hemisphere, principally from North America ($N > 9$). The most represented country from the Southern hemisphere was Australia ($N > 6$).

Discussion

This original paper presents the methods and results of a research study using a newly developed software tool called ABCal. The tool is implemented in Python and is designed to analyse scientometric data from various studies when conducting systematic reviews and meta-analyses. The primary focus of the study was to compute and assess author bias in the literature, providing a quantitative measure for the possible effect of overrepresented authors introducing bias to the overall interpretation of the literature. The computed author bias values provide a quantitative measure of author influence on the interpretation of the studies. Furthermore, scientometric plots provided valuable insights into the trends and distribution of publications over time and geographic locations.

The distribution of author bias values, as shown in histograms and box plots, showed marginal differences in the raw values but was conserved between the two datasets when using z-score transformed values. This highlighted similar spreads for the distribution with low to medium bias for most studies and a smaller number of studies exhibiting higher bias levels. As such, author bias values were able to identify, in a quantitative manner, the overrepresentation of some authors in both meta-analytic datasets. When combined with scientometric plots of author contributions, it became clear that both datasets contain a significant number of authors who have contributed a larger number of studies than others. This makes the proposed author bias metric useful in addressing the account for such bias when doing reviews (Felson, 1992; Knobloch et al., 2011).

The author bias metric is organically related to fractional citation counting (Zhou & Leydesdorff, 2010), an emerging ‘golden standard’ when comparing author contribution levels. This is due to similarities between calculations used in the initial steps that count the total number of times an author appears in the list of authors, which is divided by the total number of authors in the list (Bouyssou & Marchant, 2016). Considering, however, that this is only done within the context of studies included in a meta-analysis instead of the full reference list—in this instance the new metric represents a special use case of fractional counting. Subsequent steps sum the fractional count for individual authors per included study and divides the total by the number of authors per paper to derive a paper-level value for author bias. Furthermore, rather than relying on raw values, ABCal provides the option to convert between raw values and z-scores as well as three levels of interpretation: low, medium, and high. This facilitates cross-discipline

use of the calibrated author bias metric as several cultural differences may exist between fields in terms of publication and citation behaviour (Bornmann & Daniel, 2008; Zhou & Leydesdorff, 2010). At present ABCal, and the novel author bias metric, also provides enhanced utility in comparison to existing options (Bedru et al., 2023; Keirstead, 2016; Kozlowski, 2019). For example, ABCal uses provided information and does not rely on the indexing of papers on a specific database (Kozlowski, 2019) or the existence of author profiles on a specific platform (Keirstead, 2016). This is particularly important for included studies from publishers that don't index their articles on all databases or when including preprints from e.g., *bioRxiv*. ABCal is also freely available to the community for implementation in other studies while several similar algorithms are not (Bedru et al., 2023).

Any new metric is, however, subject to benchmarking through tests of validity and reliability (Cohen et al., 2017; Hammersley, 1987). Validity, as measured by agreement between medium to high bias studies and the list of top authors using Cohen's kappa (Cohen, 1960), showed moderate to high agreement levels that are generally well suited given the application (Altman, 1990). Reliability, as measured by Cronbach's alpha, found that the author bias metric was able to partition studies into low, medium, and high bias with a high degree of consistency between datasets. Furthermore, the functional validity was assessed by meta-regression for which the results indicated a significant association between author bias and observed effect measures in the meta-analyses. More specifically, higher author bias values were associated with higher effect measures, while lower bias was evident in studies with lower effect measures. This also makes author bias values utile in understanding how author prominence (Cassey et al., 2004), from higher contributions to the field, may interact with reported effect sizes to account for part of the heterogeneity that exists between studies as well as publication bias (Baker et al., 2009).

The presence of publication bias, as revealed by tests of funnel plot asymmetry (Møller & Jennions, 2001), suggests the possibility of selective publication in the literature (Boutron et al., 2019). This is typically attributed to small study effects such as the exclusion of studies with smaller sample sizes, even when sample sizes are sufficient for adequate statistical power of a given test (Cohen, 1988; Faul et al., 2009; Kang, 2021). The concentration of studies in certain regions as seen by geographic mapping of study locations, however, indicates a potential research trend of fewer or missing studies from the global South, as previously suggested (Collyer, 2018), and evidence that research in ecology or animal science may not follow a truly global distribution (Martin et al., 2012). Considering researchers from lower income countries may conduct research on a smaller scale for economic reasons, it is feasible that the existing evidence of publication bias is due to the lack of studies from the Southern hemisphere in the literature.

It's important to acknowledge the limitations of the study. The analysis relies on the accuracy and completeness of the input data (Knobloch et al., 2011; O'Dea et al., 2021), and certain assumptions might have been made during the calculation of author bias. Additionally, the analysis is limited to the specific datasets related to biomarkers for age in animals, and generalization to other research fields might require further investigation. Future work can focus on expanding the application of ABCal to different research areas and datasets to validate its effectiveness and robustness across various domains. Additionally, efforts can be made to address potential limitations and explore enhancements to the tool's functionalities to meet the evolving needs of scientometric analysis in the ecology research community, particularly when conducting systematic reviews and meta-analyses.

Overall, ABCal proves to be a useful tool for scientometric analysis, offering valuable information to researchers in assessing the impact of authors and potential biases in the

literature. The software's capabilities to analyse authorship contributions and produce scientometric plots can aid researchers in gaining a deeper understanding of the research landscape and identifying key contributors and research trends.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11192-023-04880-6.

Acknowledgements All images were edited for publication in BioRender.com with publication licenses granted under the Academic plan. The author would like to thank Dr Desiré Lee Dalton, and Professors Paul Grobler and Antoinette Kotzé, for their support and advice.

Funding Open access funding provided by University of the Free State. Open access provided by the University of the Free State. This work is based on the research supported wholly/in part by the National Research Foundation (Grant Number: 112062), South Africa.

Data availability The custom Python script for ABCal version 1.0.2 is available for download for installation from source code on GitHub (https://github.com/LSLeClercq/ABCal), and includes example files used for testing. Data used in this paper were deposited online (https://doi.org/10.5281/zenodo.7091053) and recently published (Le Clercq et al., 2023a, 2023b, 2023c).

Declarations

Conflict of interest The author has no competing interests to declare.

References

Adams, D. C. (2008). Phylogenetic meta-analysis. *Evolution, 62*(3), 567–572. https://doi.org/10.1111/j.1558-5646.2007.00314.x

Altman, D. (1990). *Practical statistics for medical research* (1st ed.). London: Chapman & Hall.

Ausloos, M. (2013). A scientometrics law about co-authors and their ranking: The co-author core. *Scientometrics, 95*(3), 895–909. https://doi.org/10.1007/s11192-012-0936-x

Baker, W. L., Michael White, C., Cappelleri, J. C., Kluger, J., & Coleman, C. I. (2009). Understanding heterogeneity in meta-analysis: The role of meta-regression. *International Journal of Clinical Practice, 63*(10), 1426–1434. https://doi.org/10.1111/J.1742-1241.2009.02168.X

Barrett, P., Hunter, J., Miller, J. T., Hsu, J.-C., & Greenfield, P. (2005). matplotlib—A portable python plotting package. In *Astronomical data analysis software and systems XIV ASP conference series* (Vol. 347, p. 91). Pasadena

Bauch, C., Gatt, M. C., Granadeiro, J. P., Verhulst, S., & Catry, P. (2020). Sex-specific telomere length and dynamics in relation to age and reproductive success in Cory's shearwaters. *Molecular Ecology, 29*(7), 1344–1357. https://doi.org/10.1111/mec.15399

Bedru, H. D., Zhang, C., Xie, F., Yu, S., & Hussain, I. (2023). CLARA: Citation and similarity-based author ranking. *Scientometrics, 128*(2), 1091–1117. https://doi.org/10.1007/s11192-022-04590-5

Boell, S. K., & Cecez-Kecmanovic, D. (2010). Literature reviews and the hermeneutic circle. *Australian Academic and Research Libraries, 41*(2), 129–144. https://doi.org/10.1080/00048623.2010.10721450

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods, 1*(2), 97–111. https://doi.org/10.1002/JRSM.12

Borenstein, M., & Higgins, J. P. T. (2013). Meta-analysis and subgroups. *Prevention Science, 14*(2), 134–143. https://doi.org/10.1007/S11121-013-0377-7/METRICS

Bornmann, L., & Daniel, H. D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation, 64*(1), 45–80. https://doi.org/10.1108/00220410810844150/FULL/XML

Bors, E. K., Baker, C. S., Wade, P. R., O'Neill, K. B., Shelden, K. E. W., Thompson, M. J., Fei, Z., Jarman, S., & Horvath, S. (2021). An epigenetic clock to estimate the age of living beluga whales. *Evolutionary Applications, 14*(5), 1263–1273. https://doi.org/10.1111/eva.13195

Boutron, I., Page, M., Higgins, J., Altman, D., Lundh, A., & Hróbjartsson, A. (2019). Chapter 7: Considering bias and conflicts of interest among the included studies. In J. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. Page, & V. Welch (Eds.), *Cochrane handbook for systematic reviews of interventions* (2nd ed., pp. 177–204). Chichester: Wiley.

Bouyssou, D., & Marchant, T. (2016). Ranking authors using fractional counting of citations: An axiomatic approach. *Journal of Informetrics, 10*(1), 183–199. https://doi.org/10.1016/J.JOI.2015.12.006

Brümmendorf, T. H., Mak, J., Sabo, K. M., Baerlocher, G. M., Dietz, K., Abkowitz, J. L., & Lansdorp, P. M. (2002). Longitudinal studies of telomere length in feline blood cells: Implications for hematopoietic stem cell turnover in vivo. *Experimental Hematology, 30*(10), 1147–1152. https://doi.org/10.1016/S0301-472X(02)00888-3

Burraco, P., Comas, M., Reguera, S., Zamora-Camacho, F. J., & Moreno-Rueda, G. (2020). Telomere length mirrors age structure along a 2200-m altitudinal gradient in a Mediterranean lizard. *Comparative Biochemistry and Physiology Part a: Molecular & Integrative Physiology, 247*, 110741. https://doi.org/10.1016/j.cbpa.2020.110741

Cassey, P., Ewen, J. G., Blackburn, T. M., & Møller, A. P. (2004). A survey of publication bias within evolutionary ecology. *Proceedings of the Royal Society B Biological Sciences*. https://doi.org/10.1098/rsbl.2004.0218

Cerchiara, J. A., Risques, R. A., Prunkard, D., Smith, J. R., Kane, O. J., & Boersma, P. D. (2017). Magellanic penguin telomeres do not shorten with age with increased reproductive effort, investment, and basal corticosterone. *Ecology and Evolution, 7*(15), 5682–5691. https://doi.org/10.1002/ECE3.3128

Chamberlain, S. A., Hovick, S. M., Dibble, C. J., Rasmussen, N. L., Van Allen, B. G., Maitner, B. S., Ahern, J. R., Bell-Dereske, L. P., Roy, C. L., Meza-Lopez, M., & Carrillo, J. (2012). Does phylogeny matter? Assessing the impact of phylogenetic information in ecological meta-analysis. *Ecology Letters*. https://doi.org/10.1111/j.1461-0248.2012.01776.x

Cherdsukjai, P., Buddhachat, K., Brown, J., Kaewkool, M., Poommouang, A., Kaewmong, P., Kittiwattanawong, K., & Nganvongpaint, K. (2020). Age relationships with telomere length, body weight and body length in wild dugong (Dugong dugon). *PeerJ, 8*, e10319. https://doi.org/10.7717/peerj.10319

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Cohen, L., Manion, L., & Morrison, K. (2017). Validity and reliability. In *Research methods in education* (Vol. 44, pp. 245–284). Eighth edition. New York: Routledge

Collyer, F. M. (2018). Global patterns in the publishing of academic knowledge: Global North, global South. *Current Sociology, 66*(1), 56–73. https://doi.org/10.1177/0011392116680020

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334. https://doi.org/10.1007/BF02310555

Dickersin, K., Scherer, R., & Lefebvre, C. (1994). Systematic reviews: Identifying relevant studies for systematic reviews. *BMJ, 309*(6964), 1286. https://doi.org/10.1136/bmj.309.6964.1286

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315*(7109), 629–634. https://doi.org/10.1136/BMJ.315.7109.629

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Felson, D. T. (1992). Bias in meta-analytic research. *Journal of Clinical Epidemiology, 45*(8), 885–892. https://doi.org/10.1016/0895-4356(92)90072-U

Fick, L. J., Fick, G. H., Li, Z., Cao, E., Bao, B., Heffelfinger, D., Parker, H.G., Ostrander, E.A., & Riabowol, K. (2012). Telomere length correlates with life span of dog breeds. *Cell Reports, 2*(6), 1530–1536. https://doi.org/10.1016/J.CELREP.2012.11.021

Fohringer, C., Hoelzl, F., Allen, A. M., Cayol, C., Ericsson, G., Spong, G., Smith, S., & Singh, N. J. (2022). Large mammal telomere length variation across ecoregions. *BMC Ecology and Evolution, 22*(1), 1–10. https://doi.org/10.1186/S12862-022-02050-5/TABLES/1

Hammersley, M. (1987). Some notes on the terms 'validity' and 'reliability.' *British Educational Research Journal, 13*(1), 73–82. https://doi.org/10.1080/0141192870130107

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2021). *Doing meta-analysis with R: A hands-on guide* (1st ed.). Boca Raton: Chapman & Hall/CRC Press.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., & Kern, R. (2020). Array programming with NumPy. *Nature, 585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Haussmann, M. F., & Vleck, C. M. (2002). Telomere length provides a new technique for aging animals. *Oecologia, 130*(3), 325–328. https://doi.org/10.1007/s00442-001-0827-y

Higgins, J. P. T. (2008). Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified. *International Journal of Epidemiology, 37*(5), 1158–1160. https://doi.org/10.1093/ije/dyn204

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*(11), 1539–1558. https://doi.org/10.1002/sim.1186

Honvo, G., Leclercq, V., Geerinck, A., Thomas, T., Veronese, N., Charles, A., Rabenda, V., Beaudart, C., Cooper, C., Reginster, J. Y., & Bruyère, O. (2019). Safety of topical non-steroidal anti-inflammatory drugs in osteoarthritis: Outcomes of a systematic review and meta-analysis. *Drugs and Aging, 36*(1), 45–64. https://doi.org/10.1007/s40266-019-00661-0

Horvath, S., Haghani, A., Peng, S., Hales, E. N., Zoller, J. A., Raj, K., Larison, B., Robeck, T. R., Petersen, J. L., Bellone, R. R., & Finno, C. J. (2022a). DNA methylation aging and transcriptomic studies in horses. *Nature Communications, 13*(1), 1–13. https://doi.org/10.1038/s41467-021-27754-y

Horvath, S., Haghani, A., Zoller, J. A., Raj, K., Sinha, I., Robeck, T. R., Black, P., Couzens, A., Lau, C., Manoyan, M., & Ruiz, Y. A. (2022b). Epigenetic clock and methylation studies in marsupials: Opossums, Tasmanian devils, kangaroos, and wallabies. *GeroScience, 44*(3), 1825–1845. https://doi.org/10.1007/s11357-022-00569-5

Kang, H. (2021). Sample size determination and power analysis using the G*Power software. *Journal of Educational Evaluation for Health Professions*. https://doi.org/10.3352/JEEHP.2021.18.17. National Health Personnel Licensing Examination Board of the Republic of Korea.

Keirstead, J. (2016). scholar: analyse citation data from Google Scholar. R package. https://github.com/jkeirstead/scholar

Kirby, R., Alldredge, M. W., & Pauli, J. N. (2017). Environmental, not individual, factors drive markers of biological aging in black bears. *Evolutionary Ecology, 31*(4), 571–584. https://doi.org/10.1007/s10682-017-9885-4

Knobloch, K., Yoon, U., & Vogt, P. M. (2011). Preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement and publication bias. *Journal of Cranio-Maxillofacial Surgery, 39*(2), 91–92. https://doi.org/10.1016/J.JCMS.2010.11.001

Kozlowski, L. P. (2019). fCite: a fractional citation tool to quantify an individual's scientific research output. *bioRxiv*. https://doi.org/10.1101/771485

Kraus, S., Breier, M., Lim, W. M., Dabić, M., Kumar, S., Kanbach, D., & Ferreira, J. J. (2022). Literature reviews as independent studies: Guidelines for academic practice. *Review of Managerial Science, 16*(8), 2577–2595. https://doi.org/10.1007/s11846-022-00588-8

Lajeunesse, M. J. (2009). Meta-analysis and the comparative phylogenetic method. *American Naturalist, 174*(3), 369–381. https://doi.org/10.1086/603628

Le Clercq, L. S. (2023). ABCal: Author bias computation and scientometric plotting. GitHub. https://github.com/LSLeClercq/ABCal

Le Clercq, L., Bazzi, G., Cecere, J. G., Gianfranceschi, L., Grobler, J. P., Kotzé, A., Rubolini, D., Liedvogel, M., & Dalton, D. L. (2023a). Time trees and clock genes: A systematic review and comparative analysis of contemporary avian migration genetics. *Biological Reviews, 98*(4), 1051–1080. https://doi.org/10.1111/brv.12943

Le Clercq, L.-S., Grobler, J. P., Kotzé, A., & Dalton, D. L. (2023b). Dataset generated in a systematic review and meta-analysis of biological clocks as age estimation markers in animal ecology. *Data in Brief, 51*, 109615. https://doi.org/10.1016/j.dib.2023.109615

Le Clercq, L., Kotzé, A., Grobler, J. P., & Dalton, D. L. (2023c). Biological clocks as age estimation markers in animals: A systematic review and meta-analysis. *Biological Reviews, 98*(6), 1972–2011. https://doi.org/10.1111/brv.12992

Le Clercq, C. M. P., van Ingen, G., Ruytjens, L., & van der Schroeff, M. P. (2016). Music-induced hearing loss in children, adolescents, and young adults. *Otology & Neurotology, 37*(9), 1208–1216. https://doi.org/10.1097/MAO.0000000000001163

Leimu, R., & Koricheva, J. (2004). Cumulative meta-analysis: A new tool for detection of temporal trends and publication bias in ecology. *Proceedings of the Royal Society B: Biological Sciences, 271*(1551), 1961–1966. https://doi.org/10.1098/rspb.2004.2828

Lopez Gonzalez-Nieto, P., Gomez Flechoso, M., Arribas Mocoroa, M. E., Muñoz Martin, A., Garcia Lorenzo, M. L., Cabrera Gomez, G., Gomez, J.A., Fraile, A.C., Dagan, J.O., Palomares, R.M., & Lahoz-Beltra, R. (2020). Design and development of a virtual laboratory in PYTHON for the teaching of data analysis and mathematics in geology: GeoPy. In *INTED2020 proceedings* (pp. 2236–2242). Valencia, Spain. https://doi.org/10.21125/inted.2020.0687

Lortie, C. J., Aarssen, L. W., Budden, A. E., Koricheva, J. K., Leimu, R., & Tregenza, T. (2007). Publication bias and merit in ecology. *Oikos*. https://doi.org/10.1111/j.0030-1299.2007.15686.x

Martin, L. J., Blossey, B., & Ellis, E. (2012). Mapping where ecologists work: Biases in the global distribution of terrestrial ecological observations. *Frontiers in Ecology and the Environment*. https://doi.org/10.1890/110154

Mayne, B., Espinoza, T., Roberts, D., Butler, G. L., Brooks, S., Korbie, D., & Jarman, S. (2021). *Non-lethal age estimation of three threatened fish species using DNA methylation: Australian lungfish*. Murray cod and Mary River cod: Molecular Ecology Resources. https://doi.org/10.1111/1755-0998.13440

McKinney, W. (2010). Data structures for statistical computing in Python. In *Proceedings of the 9th Python in science conference* (pp. 56–61). Austin. https://doi.org/10.25080/majora-92bf1922-00a

Meyer, D., Zeileis, A., & Hornik, K. (2023). vcd: Visualizing categorical data. R package. https://cran.r-project.org/package=vcd

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2010). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *International Journal of Surgery, 8*(5), 336–341. https://doi.org/10.1016/J.IJSU.2010.02.007

Molbert, N., Angelier, F., Alliot, F., Ribout, C., & Goutte, A. (2021). Fish from urban rivers and with high pollutant levels have shorter telomeres. *Biology Letters, 17*(1), 20200819. https://doi.org/10.1098/rsbl.2020.0819

Møller, A. P., & Jennions, M. D. (2001). Testing and adjusting for publication bias. *Trends in Ecology & Evolution, 16*(10), 580–586. https://doi.org/10.1016/S0169-5347(01)02235-2

Nakagawa, S., Yang, Y., Macartney, E. L., Spake, R., & Lagisz, M. (2023). Quantitative evidence synthesis: A practical guide on meta-analysis, meta-regression, and publication bias tests for environmental sciences. *Environmental Evidence, 12*(1), 1–19. https://doi.org/10.1186/s13750-023-00301-6

O'Dea, R. E., Lagisz, M., Jennions, M. D., Koricheva, J., Noble, D. W. A., Parker, T. H., Gurevitch, J., Page, M. J., Stewart, G., Moher, D., & Nakagawa, S. (2021). Preferred reporting items for systematic reviews and meta-analyses in ecology and evolutionary biology: A PRISMA extension. *Biological Reviews, 96*(5), 1695–1722. https://doi.org/10.1111/brv.12721

Pae, C. U. (2015). Why systematic review rather than narrative review? *Psychiatry Investigation, 12*(3), 417. https://doi.org/10.4306/PI.2015.12.3.417

Perianes-Rodriguez, A., Waltman, L., & van Eck, N. J. (2016). Constructing bibliometric networks: A comparison between full and fractional counting. *Journal of Informetrics, 10*(4), 1178–1195. https://doi.org/10.1016/J.JOI.2016.10.006

Plot, V., Criscuolo, F., Zahn, S., & Georges, J. Y. (2012). Telomeres, age and reproduction in a long-lived reptile. *PLoS ONE, 7*(7), e40855. https://doi.org/10.1371/JOURNAL.PONE.0040855

Polanowski, A. M., Robbins, J., Chandler, D., & Jarman, S. N. (2014). Epigenetic estimation of age in humpback whales. *Molecular Ecology Resources, 14*(5), 976–987. https://doi.org/10.1111/1755-0998.12247

Python Team. (2021). Python programming language. Wilmington, Delaware, United States. https://www.python.org/

R Core Team. (2020). *R: A language and environment for statisitical computing*. Vienna: R Foundation for Statistical Computing.

Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and item response theory analyses. *Journal of Statistical Software, 17*(5), 1–25.

Robeck, T. R., Fei, Z., Lu, A. T., Haghani, A., Jourdain, E., Zoller, J. A., Li, C. Z., Steinman, K. J., DiRocco, S., Schmitt, T., & Osborn, S. (2021). Multi-species and multi-tissue methylation clocks for age estimation in toothed whales and dolphins. *Communications Biology*. https://doi.org/10.1038/s42003-021-02179-x

Robeck, T. R., Haghani, A., Fei, Z., Lindemann, D. M., Russell, J., Herrick, K. E. S., Montano, G., Steinman, K. J., Katsumata, E., Zoller, J. A., & Horvath, S. (2023). Multi-tissue DNA methylation aging clocks for sea lions, walruses and seals. *Communications Biology, 6*(1), 359. https://doi.org/10.1038/s42003-023-04734-0

RStudio Team. (2021). RStudio: Integrated development environment for R. Boston, MA. http://www.rstudio.com/

Schwarzer, G., Carpenter, J. R., & Rücker, G. (2015). *Meta-analysis with R*. Springer International Publishing. https://doi.org/10.1007/978-3-319-21416-0

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th Python in science conference* (Vol. 57, pp. 10–25080). Austin

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika, 52*(3/4), 591. https://doi.org/10.2307/2333709

Small, H., & Garfield, E. (1985). The geography of science: Disciplinary and national mappings. *Journal of Information Science, 11*(4), 147–159. https://doi.org/10.1177/016555158501100402

Sterne, J. A. C., Egger, M., & Smith, G. D. (2001). Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. *British Medical Journal, 323*(7304), 101–105. https://doi.org/10.1136/bmj.323.7304.101

Tawfik, G. M., Dila, K. A. S., Mohamed, M. Y. F., Tam, D. N. H., Kien, N. D., Ahmed, A. M., & Huy, N. T. (2019). A step by step guide for conducting a systematic review and meta-analysis with simulation data. *Tropical Medicine and Health, 47*(1), 46. https://doi.org/10.1186/s41182-019-0165-6

Thornton, A., & Lee, P. (2000). Publication bias in meta-analysis: Its causes and consequences. *Journal of Clinical Epidemiology, 53*(2), 207–216. https://doi.org/10.1016/S0895-4356(99)00161-4

Ujvari, B., Biro, P. A., Charters, J. E., Brown, G., Heasman, K., Beckmann, C., & Madsen, T. (2017). Curvilinear telomere length dynamics in a squamate reptile. *Functional Ecology, 31*(3), 753–759. https://doi.org/10.1111/1365-2435.12764/SUPPINFO

Vernasco, B. J., Dakin, R., Majer, A. D., Haussmann, M. F., Brandt Ryder, T., & Moore, I. T. (2021). Longitudinal dynamics and behavioural correlates of telomeres in male wire-tailed manakins. *Functional Ecology, 35*(2), 450–462. https://doi.org/10.1111/1365-2435.13715/SUPPINFO

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1–48.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., & Van Der Walt, S. J. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods, 17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly, 26*(2), xiii–xxiii.

Wilk, M. B., & Gnanadesikan, R. (1968). Probability plotting methods for the analysis of data. *Biometrika, 55*(1), 1–17. https://doi.org/10.1093/biomet/55.1.1

Wilkinson, G. S., Adams, D. M., Haghani, A., Lu, A. T., Zoller, J., Breeze, C. E., Arnold, B. D., Ball, H. C., Carter, G. G., Cooper, L. N. and Dechmann, D. K. (2021). DNA methylation predicts age and provides insight into exceptional longevity of bats. *Nature Communications, 12*(1), 1615. https://doi.org/10.1038/s41467-021-21900-2

Zhou, P., & Leydesdorff, L. (2010). Fractional counting of citations in research evaluation: An option for cross- and interdisciplinary assessments. Retrieved October 28, 2023, from http://arxiv.org/abs/1012.0359

CHAPTER 4

Clock gene polymorphism as predictors of phenology in intra-African migratory birds

This chapter represents the methods, results, and discussion for the first biological clock of the study which focuses on the association between polymorphisms within candidate genes of the circadian clock and attributes of migration phenology. The chapter follows the standard article format and was submitted to *Ecology and Evolution* (Le Clercq *et al.*, 2024*a*). Here, the focal species is Diederik cuckoos (*Chrysococcyx caprius*) sampled from three locations in Africa during different parts of the year corresponding to regional breeding seasons. Details are provided for the methods adapted to Diederik Cuckoos for studying polymorphisms in two candidate genes: *Clock* and *Adcyap1.* Furthermore, key results from the experiments—including polymorphism repeat length, population genetics attributes, and correlates between genotypes and phenotypes—are reported. Finally, the results are discussed in terms of the degree of genetic variation in individual genes, the extent of differentiation between individuals from different regions with varied phenology, as well as how these findings relate to previously published studies from the review in Chapter 2. Supplementary materials are included in Appendix E, methods (Le Clercq *et al.*, 2023*g*, 2023*h*) were submitted to the online platform *protocols.io* in support of open methods initiatives.

Articles:

LE CLERCQ, L., KOTZÉ, A., GROBLER, J.P. & DALTON, D.L. (2024*a*) Phenotypic correlates between clock genes and phenology among populations of Diederik cuckoo, *Chrysococcyx caprius*. *Ecology and Evolution* **14**, e70117.

Methods:

LE CLERCQ, L., DALTON, D.L., KOTZÉ, A. & GROBLER, J.P. (2023*g*) PCR Amplification of *Clock* and *Adcyap1* genes with EmeraldAmp® GT PCR Master Mix in Avian species for polymorphism elucidation. *protocols.io*. DOI: https://dx.doi.org/10.17504/protocols.io.6qpvrdwk3gmk/v1.

LE CLERCQ, L., DALTON, D.L., KOTZÉ, A. & GROBLER, J.P. (2023*h*) ABI Sanger Sequencing of Avian Clock genes to elucidate markers for Migration Phenology. *protocols.io*. DOI: https://dx.doi.org/10.17504/protocols.io.3byl4k6zrvo5/v1.

Received: 1 February 2024 | Revised: 12 July 2024 | Accepted: 18 July 2024

DOI: 10.1002/ece3.70117

Ecology and Evolution Open Access WILEY

RESEARCH ARTICLE

Phenotypic correlates between clock genes and phenology among populations of Diederik cuckoo, *Chrysococcyx caprius*

L. S. Le Clercq[1,2] | V. Phetla[1] | S. T. Osinubi[3] | A. Kotzé[1,2] | J. P. Grobler[2] | D. L. Dalton[4]

[1]South African National Biodiversity Institute, Pretoria, South Africa

[2]Department of Genetics, University of the Free State, Bloemfontein, South Africa

[3]FitzPatrick Institute of African Ornithology, University of Cape Town, Cape Town, South Africa

[4]School of Health and Life Sciences, Teesside University, Middlesbrough, UK

Correspondence
L. S. Le Clercq, South African National Biodiversity Institute, P.O. Box 754, Pretoria, 0001, South Africa.
Email: leclercq.l.s@gmail.com

Funding information
National Research Foundation, Grant/Award Number: 112062

Abstract

The Diederik cuckoo, *Chrysococcyx caprius*, is a small Afrotropical bird in the family Cuculidae. It is taxonomically related to 13 other species within the genus *Chrysococcyx* and is migratory in sub-Saharan Africa. It has a unique breeding behaviour of being a brood parasite: Breeding pairs lay their eggs in the nests of a host species and hatchlings expel the eggs of the host species. The aim of the present study was to investigate diversity in two circadian clock genes, *Clock* and *Adcyap1*, to probe for a relationship between genetic polymorphisms and their role in circannual timing and habitat selection (phenology) in intra-African migrants. DNA extracted from blood was used for the PCR amplification and sequencing of clock genes in 30 Diederik cuckoos. Three alleles were detected for *Clock* with similar genotypes between individuals from the Northern and Southern breeding ranges while 10 alleles were detected for *Adcyap1*, having shorter alleles in the North and longer alleles in the South. Population genetic analyses, including allele frequency and zygosity analysis, showed distinctly higher frequencies for the most abundant *Clock* allele, containing 10 polyglutamine repeats, as well as a high degree of homozygosity. In contrast, all individuals were heterozygous for *Adcyap1* and alleles from both regions showed distinct differences in abundance. Comparisons between both clock genes and phenology found several phenotypic correlations. This included evidence of a relationship between the shorter alleles and habitat selection as well as a relationship between longer alleles and timing. In both instances, evidence is provided that these effects may be sex-specific. Given that these genes drive some of the synchronicity between environments and the life cycles of birds, they provide valuable insight into the fitness of species facing global challenges including climate change, urbanisation and expanding agricultural practices.

KEYWORDS
Adcyap1, avian, brood parasitism, *Chrysococcyx caprius*, circannual, *Clock*, cuckoo, intra-African, migration

Samevatting

Die Diederik koekoek, *Chrysococcyx caprius*, is 'n klein Afrotropiese voël in die Cuculidae familie. Dit is taksonomies verwant aan dertien ander spesies in die genus *Chrysococcyx* en is 'n migrant in sub-Sahara Africa met die unieke teling fenologie van broeisparasietisme (broeipare lê nul eiers in die neste van gasheerspesies en kuikens verwyder die eiers van hul gashere). Die doelwit van die huidige studie is om diversiteit in twee gene van die sirkadiese klok, *Clock* en *Adcyap1*, te ondersoek om moontlike verwantskappe tussen polimorfisme en fenologie soos tydsberekening en habitat keuse te ontleed. DNS vanaf bloed monsters was gebruik vir PKR amplifikasie en reeksbepaling in dertig Diederik koekoek. Drie allele was bespeur vir die *Clock* geen, met soortgelyke genotipes onder individue van beide die Noordelike en Suidelike teling streke, terwyl tien allele bespeur is vir die *Adcyap1* geen, met korter allele in die Noord en langer allele in die Suid. Genetiese analise van die populasies, insluitend alleel frekwensie en sigositeit, het kenmerkend hoër frekwensies getoon vir die mees algemene *Clock* geen, met tien poli-glutamien herhalings, sowel as 'n hoë mate van homosigositeit. In teenstelling was alle individue heterosigoties vir die *Adcyap1* geen met afsonderlike verskille in mate. Vergelykings tussen beide gene en fenologie het verskeie korrelasies ontbloot. Onder andere is tekens vir 'n verwantskap tussen korter allele en habitat seleksie sowel as langer allele en tydsberekening gevind. In beide gevalle is verdere bewyse gelewer dat die effek seks-spesifiek mag wees. Aangesien hierdie gene die ritmiek van jaarlikse lewensiklusse reguleer verskaf hul belangrike insig tot die fiksheid van spesies wat globale uitdagings die hoof bied.

1 | INTRODUCTION

Birds in environments that show seasonal changes in temperature (Pancerasa et al., 2018) and food availability (Stephan, 2002) have developed complex annual cycles to carefully time life events including migration and breeding (Cassone, 2014; Cassone & Westneat, 2012). Several heredity studies have also shown high correlations for intergenerational resemblance of attributes related to migration and breeding; with heredity scores (h^2) ranging between 0.38 and 0.67 in several passerine migratory species (Justen & Delmore, 2022). Despite the high heritability, related studies have also indicated that migratory species have been subject to many adaptations to their life history including changes in their departure and arrival times (Jenni & Kéry, 2003; Rubolini et al., 2007; Saino et al., 2009), staging sites (Verhoeven et al., 2018) and ranges (de Vos & Cherry, 2017) in response to environmental change. This has highlighted that while migratory avian species show marked heritability in their annual life events, some degree of adaptability may be required for a species persistence.

Due to the high heritability of these traits, it is thus likely that they are under genetic control and that genetic diversity within key genes confer an element of adaptability to the environment. Genetic studies have linked variability within genes associated with the circadian clock, such as the mean polymorphic repeat number in an exon of the *Clock* gene and 3′-UTR of the *Adcyap1* gene, to several aspects of the annual life events of birds (reviewed in Le Clercq, Bazzi, et al., 2023e; Le Clercq, Bazzi, et al., 2023f). This included aspects such as timing of autumn migration (Justen et al., 2022; Pulido et al., 2001), timing of spring migration (Justen & Delmore, 2022; Krist et al., 2021) and breeding latitude (Johnsen et al., 2007). Thus far, these observations have been most widely illustrated in passerine (order: Passeriformes) and songbird species (Le Clercq, Bazzi, Cecere, et al., 2023) and has largely focused on species from Europe and North America.

Cuckoos present a unique case study because, in addition to being migratory species, cuckoos in the order Cuculiformes follow the unique breeding behaviour of being a brood parasite—adapting not only to their environments but also the species they parasitise. Different species of cuckoos differentially parasitise different host species (Brooker & Brooker, 1990) where cuckoo breeding pairs lay their eggs in the nests of a host species and hatchlings expel the eggs or hatchlings of the host species. For example, the Australian Horsfield's bronze cuckoo, *Chrysococcyx basalis* [Horsfield, 1821], parasitises resident populations of Superb fairy-wren (Langmore et al., 2008), *Malurus cyaneus* [Ellis, 1782], while the Common cuckoo, *Cuculus canorus* [Linnaeus, 1758], parasitises several resident as well as short and long distance migrants (Saino et al., 2009). Due to this 'evolutionary arms race' most cuckoos need to be highly synchronised to their host species and follow a pattern of co-evolution (Rönkä et al., 2022) with their hosts that results in host-specific races; which differ in several phenotypic and behavioural attributes without constituting definitive subspecies but are rather more

similar to ecotypes (Langmore et al., 2008; Reed, 1968; Skead, 1952; Soler & Soler, 2000). These races exhibit phenotypical variation that includes differences in egg characteristics such as coloration (Caves et al., 2015; Reed, 1968) and shell density (Spottiswoode, 2010), behavioural adaptations, such as the mimicking of different host sounds and behaviour (Langmore et al., 2008; Reed, 1968), persistent shared ectoparasite profiles (Lindholm et al., 1998), and high synchronicity in timing of migration and breeding (Douglas et al., 2010).

The Diederik cuckoo, *Chrysococcyx caprius* (Boddaert, 1783), is a small Afrotropical bird in the family *Cuculidae* (order: Cuculiformes) first described by French naturalist Georges-Louis Leclerc, Comte de Buffon, in 1780 as part of his titular series the *Histoire Naturelle des Oiseaux* (Leclerc, 1780). It is taxonomically related to 13 other species within the genus *Chrysococcyx* (Figure 1) that diverged from other cuckoo lineages approximately 23.94 million years ago (MYA) (Jetz et al., 2012; Prum et al., 2015). The genus includes Afro-Asiatic species such as Klaas's cuckoo, *Chrysococcyx klaas* [Stephens, 1815] and Australo-Papuan species such as Horsfield's bronze cuckoo, *Chrysococcyx basalis* [Horsfield, 1821]. Species within the genus diverged approximately 7.71 MYA (diversification rate (r) = 0.24). The Diederik cuckoo is still classified as monotypic and results from detailed molecular study revealed no evidence of an emerging subspecies level divide between different populations based on neutral genomic markers and mitochondrial DNA (Smith et al., 2023). Although this species has been known for more than 300 years, little research has been conducted on them and many aspects of their life history and population structure remain uncharacterised.

The Diederik cuckoo has a wide distribution in sub-Saharan Africa (Figure 2), with possible resident populations of short-winged individuals in West Africa. The remainder are presumed to be short distance, intra-African migrants with annual migrations between the more centrally located populations of the Equator and their Northern and Southern-most breeding ranges, typically moving based on annual rainfall seasons (Payne et al., 2021). In their Northern ranges they typically breed between June and September. Here, they parasitise passerine birds such as the Village weaver (Gee & Heigham, 1977; Honeywell, 1971; Macdonald, 1980; Morel, 1972), *Ploceus cucullatus* [Müller, 1776], in close proximity to other weavers (Macdonald, 1980) such as Vieillot's black weaver, *P. nigerrimus* [Vieillot, 1819], and Little weaver, *P. luteolus*. [Lichtenstein, 1823]. In their Southern ranges they first arrive in September with most rapid arrivals in October before they breed as of November. Thereafter, many birds leave the breeding grounds starting in February, however, evidence from late hatching chicks indicate that some only leave in April (Hockey, 1994; Maclean et al., 1993; Rowan, 1983; Tarboton et al., 1987). In their Southern range they also parasitise the Village weaver, as well as other passerine species: Cape sparrow, *Passer melanurus* [Müller, 1776]; Cape wagtail, *Motacilla capensis* [Linnaeus, 1766]; Cape weaver, *P. capensis* [Linnaeus, 1766]; Southern masked weaver, *P. velatus* [Vieillot, 1819]; and Southern red bishop, *Euplectes orix* [Linnaeus, 1758] (de Vos & Cherry, 2017; Jensen & Jensen, 1969; Rowan, 1983). Unlike the Diederik cuckoo, most host species are year-round resident birds that do not follow a distinct pattern of annual migration but differ in their annual timing for nesting and breeding based on hemisphere and region (Figure S1).

Because of their co-evolution with resident species, Diederik cuckoos are particularly sensitive to changes in their host species in terms of population density (Ferguson, 1994), changing ranges (de Vos & Cherry, 2017) and changes in annual timing of life events (Saino et al., 2009), possibly leading to mismatches

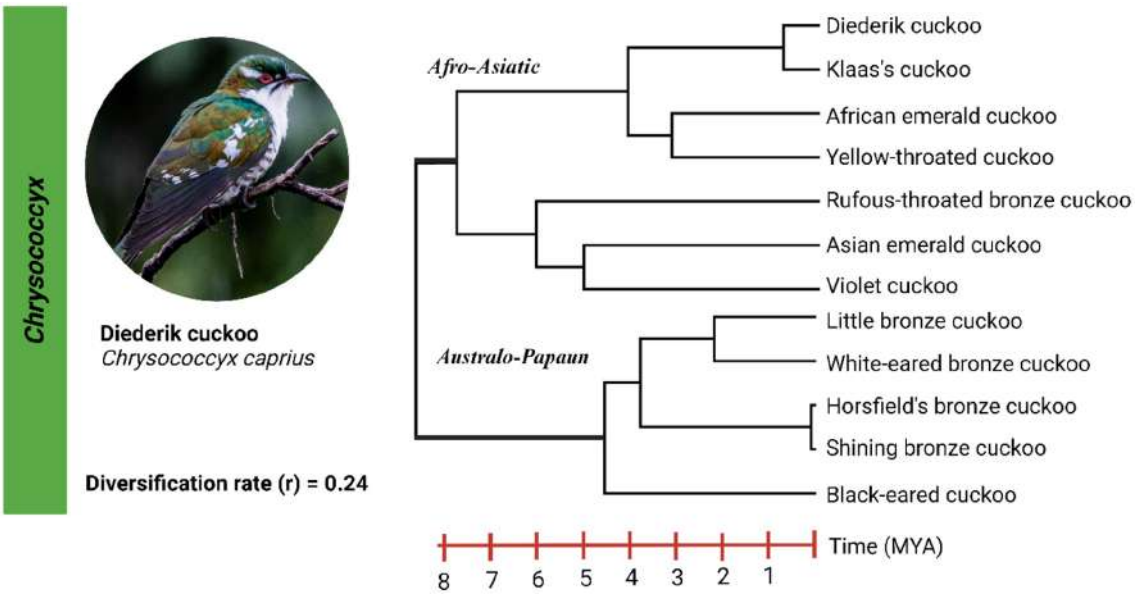


FIGURE 1 Time calibrated phylogenetic tree for species in the genus *Chrysococcyx*. The species tree was downloaded from the Bird Tree website using the Ericson phylogeny for 500 trees (https://birdtree.org/). Image created in BioRender.com. The Diederik cuckoo partitions with Klaas's cuckoo within the upper, Afro-Asiatic, clade of cuckoos including both the African and Asian emerald cuckoos, while the bottom clade contains Australo-Papuan cuckoos such as the Australian Horsefield's bronze cuckoo.

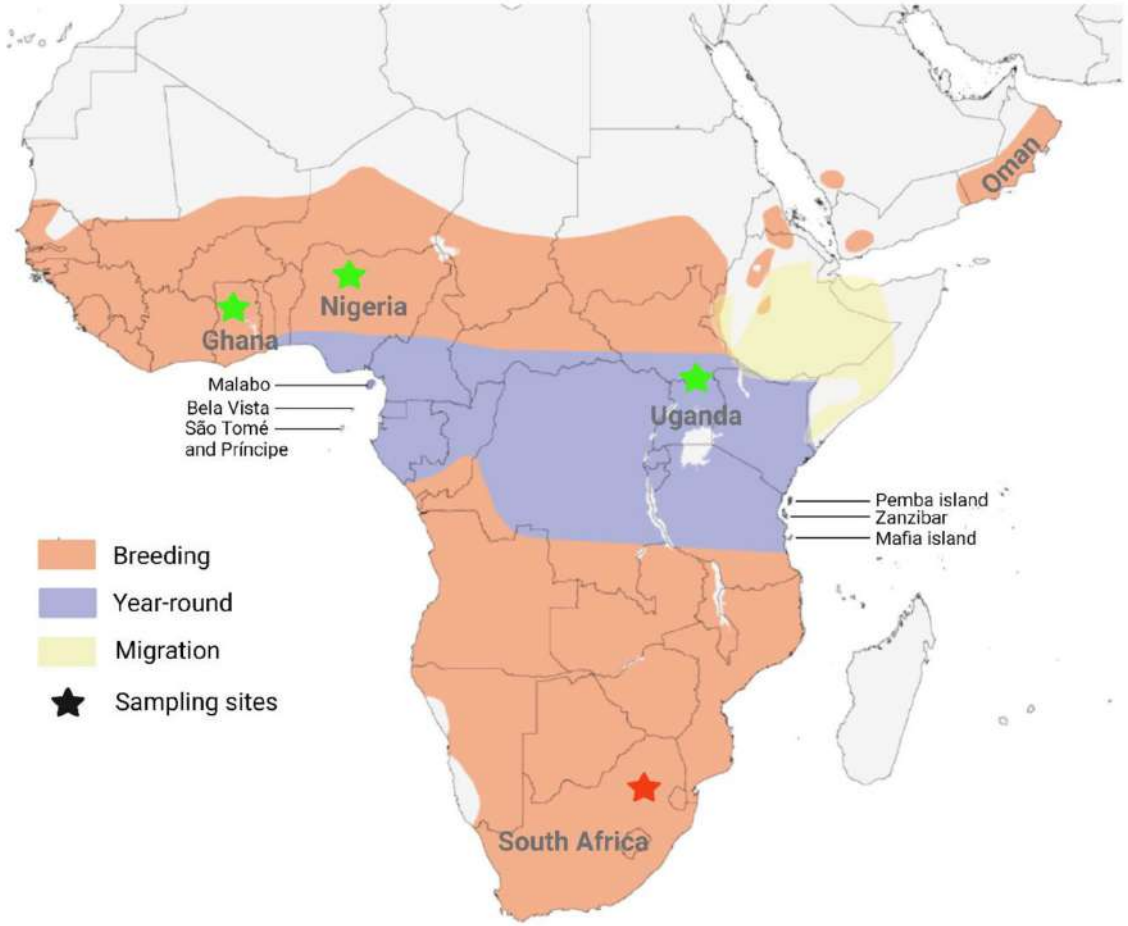


FIGURE 2 Map for the natural range of Diederik cuckoo in sub-Saharan Africa. Regions are colour coded based on the type of habitat use. Ranges indicated in red represent habitat occupied during the local breeding season while regions indicated in purple represent areas with year-round occupation. The small area indicated in yellow represents areas with vagrant sightings during the migration season that serves as staging sites but are typically vacated within a few days. Island populations off the West and East coast are indicated, including São Tomé and Príncipe as well as Zanzibar. The four sampling sites for this study are indicated with stars and included Ghana (green), Nigeria (green), Uganda (green) and South Africa (red). Image adapted from Birds of the World (Payne et al., 2021), created in BioRender.com.

between the host and cuckoo species. This renders cuckoos particularly vulnerable to climate change and the decreased availability of suitable habitat due to anthropogenic activity such as expanding agricultural use and urbanisation in Africa (Greig et al., 2017; Güneralp et al., 2018). Agriculture is less likely to contribute to habitat loss in Southern Africa, as compared to Northern Africa, as widespread land reforms introduced since 1994 have resulted in a more than 10% decrease in actively cultivated land between 1994 and 2014 (Bernstein, 2013; Lidzhegu & Palamuleni, 2012; McCusker, 2004). While this has created novel opportunities for potential rewilding (Hoogendoorn et al., 2019; Navarro & Pereira, 2015), the rewilding of agricultural land after abandonment is a long and arduous process with little land recovering naturally (Carver, 2019; Lidzhegu & Palamuleni, 2012; Navarro & Pereira, 2015). Furthermore, abandoned agricultural land typically experience an initial decline in biodiversity when the habitat becomes unsuitable for insects and birds that are typically found in proximity to farms (Carver, 2019). There is, however, a significant risk in terms of urbanisation—as most counties within the range of Diederik cuckoos are experiencing significant increases in their human populations and a corresponding increase in urbanisation (Güneralp et al., 2018).

Considering migration itself is an evolutionary adaptation in response to environmental changes in temperature (Pancerasa et al., 2018), length of daylight (Robart et al., 2018) and food availability (Stephan, 2002), the ability to adapt migration strategies is crucial for the persistence of a species. Of critical importance is how the aforementioned changes might affect the habitat selection and annual rhythmicity of life events in both resident host species as well as migratory cuckoo species (Carey, 2009). At present, however, no data exist on the diversity and putative adaptive potential, driven by the possibility for selection to shift between genotypes that confer reproductive success, of Diederik cuckoo or their host species within genes that are associated with annual life events including migration. Thus, the aim of the present study is to investigate the existing diversity in two candidate clock genes, *Clock* and *Adcyap1* and probe for a relationship between genetic variance and their potential role in circannual timing in the

Diederik cuckoo. This will be conducted to test several hypothesis including if (i) gene diversity follows a geospatial pattern by latitude and/or longitude, (ii) gene diversity relates to differences in annual timing of breeding, (iii) differences are sex-specific and (iv) if similarity exists between genotypes of cuckoos and their most common hosts.

2 | METHODS

2.1 | Specimens and ethics

Protocol approval for the present study was obtained from the protocol committee of the Department of Genetics, University of the Free State (approval number: Res18/2020). Ethics approvals were obtained from the University of the Free State (approval number: UFS-AED2020/0015/1709) as well as the South African National Biodiversity Institute (approval number: SANBI/RES/P2020/30). The appropriate research permit was also obtained from South African regulatory authorities, specifically the Department of Agriculture, Land Reform, and Rural Development (Section 20 permit: 12/11/1/1/18 (1824JD)). Specimens suited to the present study ($N = 30$) were identified from existing catalogues of samples collected for the Diederik cuckoo (Figure 1), including *C. caprius* lineages from West Africa (Ghana and Nigeria) and East Africa (Uganda) and *C. caprius* from South Africa (Table 1 and Figure 2). Samples were partitioned based on subspecies, sex, migration timing and location. Sex was determined by molecular methods as previously described (Smith et al., 2023) and 21 males and nine females were assayed. Furthermore, samples were partitioned into birds who experienced increased activity early or late during their annual breeding cycles. This was based on time of capture and sampling information from field data, as birds actively breeding or laying eggs tend to remain close to their nesting site while those who have completed these tasks move away from nesting sites and could easily be caught. Furthermore, it is known that some individuals schedule their migration earlier while others migrate later, this likely is reflected in their timing of breeding. Sample details including National Centre for Biotechnology Information (NCBI) BioSample numbers (SAMN31832894–922) are listed in Table S1.

2.2 | DNA extraction, polymerase chain reaction (PCR) and sequencing

DNA was extracted from 100 μL of whole blood with the commercially available Quick-DNA™ Miniprep Plus Kit (Zymo Research, Irvine, California, United States), according to manufacturer's instructions. The final step eluted purified DNA to approximately 25 μL of pure DNA for subsequent analysis. The DNA concentration and purity was determined by spectrophotometric measurement of 1 μL extract applied to the Nanodrop™ One (Thermo Fisher Scientific, Waltham, Massachusetts, USA) instrument. This was determined by measuring the absorbance at wavelengths of A_{260} and A_{280} to calculate the A_{260}/A_{280} ratio to determine purity in addition to concentration in ng/μL.

DNA was amplified by short-range PCR for both the *Clock* (NCBI Gene ID: 9575) and *Adcyap1* (NCBI Gene ID: 116) genes as previously described (Le Clercq et al., 2023d). Briefly, the PCR reaction were set up with 2 μL of DNA and 1.5 μL of each forward or reverse primer respectively (final concentration of 0.2 μM), were added to 15 μL of EmeraldAmp® GT PCR 2× Master Mix (Takara Bio Inc., Kusatsu, Shiga Prefecture, Japan) and adjusted to a final reaction volume of 25 μL with sterile water. Reactions were subjected to thermal cycling on the SimpliAmp™ Thermal Cycler (Thermo Fisher Scientific, Waltham, Massachusetts, USA) as follows: Initial denaturation at 98°C for 10 s, denaturation and annealing at 52°C (*Adcyap1*) or 55°C (*Clock*) for 30 s and elongation for 1 min at 72°C for 40 cycles; followed by a final hold at 4°C. Following PCR, amplicons were resolved by electrophoresis on a 4% TAE-agarose gel, for 1 h 30 at 60 V, along with the Quick-Load® Purple 100 bp DNA ladder (New England Biolabs, Ipswich, Massachusetts, United States) to confirm successful amplification and resolve possible alleles (Bhattacharya & Van Meir, 2019). PCR products were purified using Exonuclease I (Exo)/Shrimp Alkaline Phosphatase (SAP; Thermo Fisher Scientific, Waltham, Massachusetts, USA) following manufacturer's instructions. Standard reaction chain termination sequencing procedures were followed (Le Clercq et al., 2023c) individually for both the *Clock* and *Adcyap1* PCR products using the BigDye™ Terminator v. 3.1 Cycle Sequencing Kit (ABI, Thermo Fisher Scientific, Waltham, Massachusetts, USA). As per manufacturer's recommendations, 10–30 ng of PCR product was used in a 10 μL reaction with the same primers as before (final concentration of 3.2 pmol). Sequence

TABLE 1 Locations, including latitude and longitude, sample sizes (*N*) and breeding seasons for specimens used in the present study.

Location	Latitude	Longitude	*N*	Season
East Africa:				
Uganda	N 9°43'4.79"	E 7°31'55.20"	3	Jun–Sept (North)
West Africa:				
Ghana	N 9°5'16.79"	W 1°48'32.4"	2	Jun–Sept (North)
Nigeria	N 9°52'40.8"	W 8°58'26.4"	8	Jun–Sept (North)
Southern Africa:				
South Africa	S 23°3'21.6"	E 29°8'59.99"	17	Nov–Apr (South)

reactions were cleaned with the BigDye XTerminator™ Purification Kit (ABI, Thermo Fisher Scientific, Waltham, Massachusetts, USA) protocol using 45 μL of SAM™ solution and 5 μL BigDye XTerminator™ bead solution. Sequence reactions were analysed by capillary electrophoresis on the 3500 Genetic Analyser (ABI, Thermo Fisher Scientific, Waltham, Massachusetts, USA) and the Seq Scanner 2 software (ABI, Thermo Fisher Scientific, Waltham, Massachusetts, USA).

2.3 | Supplementary data

Data were supplemented by retrieving additional sequence data from the NCBI sequence read archive (SRA) for host species of the Diederik cuckoo, including specimens for Village weaver (SRR17013387) and Cape wagtail (Table S2). Data from whole genome sequencing experiments were retrieved with the SRA toolkit version 3.0.0 (https://github.com/ncbi/sra-tools) in FASTQ format and aligned in Geneious Prime 2022 (www.geneious.com) to reference sequences for the *Clock* and *Adcyap1* genes, respectively. The aligned reads for each gene were exported in FASTA format and the diploid alleles for the microsatellite repeats determined using the Perl script MEGASAT (Zhan et al., 2017) version 1.0, with Perl implemented in Strawberry Perl (64-bit) version 5.32.1.1 (https://strawberryperl.com/), or through visual inspection. Moult data were compiled from the available literature (Hanmer, 1980; Ramudzuli, 2021). Phylogenetic data for *Chrysococcyx* cuckoos were retrieved from the Bird tree website (www.birdtree.org) using the Ericson phylogeny (Jetz et al., 2012). Phylogenetic trees were summarised using TreeAnnotator 2.6.3, part of BEAST 2.6.3 (Bouckaert et al., 2014), to a 60% consensus tree with a 10% burn in. Time trees and divergence times were retrieved from the Time Tree resource (Kumar et al., 2022) using PAReTT version 1.0.2 (Le Clercq, 2023a; Le Clercq, Kotzé, Grobler, & Dalton, 2023) and were used to calibrate the phylogenetic tree as well as compute diversification rates.

2.4 | Data analysis

Sequence data output was aligned using the MAFFT version 7 multiple and pair-wise aligner (Katoh & Standley, 2013) followed by analysis in BioEdit version 7.2.6.1 (Hall, 2011). The *Clock* gene alleles were transformed to represent only the poly-Q repeat as previously described (Le Clercq, Bazzi, Cecere, et al., 2023). Population genetics analyses were done using POPGENE 1.32 (Yeh et al., 1997) to test for Hardy–Weinberg equilibrium (Hardy, 1908; Weinberg, 1908), based on the equilibrium equation (Equation 1). Here, the equilibrium is expressed as the sum of the squared homozygous genotypes and the product for homozygous genotypes multiplied by two. In the formula, *p* represents the first allele, *q* represents the second allele, and *r* represents the third allele.

$$p^2 + q^2 + r^2 + 2pq + 2pr + 2qr = 1 \quad (1)$$

Both the chi-squared (χ^2) and maximum likelihood tests were used to assess deviation from equilibrium (with significance measured at $\alpha = 0.02$). The same programme was also used to calculate allele frequencies as well as the observed (H_o) and expected (H_e) heterozygosity, based on the standard equation (Equation 2) with slight modification (Levene, 1949). Here, the expected heterozygosity is expressed as the difference between one and the sum of squared allele frequencies (p) from the ith allele to the kth allele. Based on the three-allele system, this corresponds to p, q and r alleles in the equilibrium equation.

$$\text{Expected heterozygosity } (H_e) = 1 - \sum_{i=1}^{n=k} {p_i}^2 \quad (2)$$

Lastly, the number of migrants (N_m) between the two populations was calculated (Equation 3) using the fixation index (F_{ST}). Here, the difference between one and the F_{ST} is multiplied by a quarter and subsequently divided by the F_{ST}.

$$\text{Number of migrants } (N_m) = \frac{0.25(1 - F_{ST})}{F_{ST}} \quad (3)$$

Population structure was accessed using both alleles in STRUCTURE 2.3.4 using the admixture model with populations used as priors (Pritchard et al., 2000). The parameter set had a burn-in of 500 and was run with 10,000 repeats to achieve coalescence. Experiments were run testing a population number (K) between 1 and 10 with a hundred iterations per K. The final results were exported as a zipped archive and the best structural model extracted using the web interface of STRUCTURE HARVESTER (Earl & vonHoldt, 2012). The first allele, second allele and average were compared to timing (sample date) and location (latitude and longitude) in R version 4.0.2 (R Core Team, 2020) by performing individual linear models using a Least Squares model. Models were fitted using the full data sets as well as with subset analyses based on sex to test for potential sex-linked effects.

3 | RESULTS

3.1 | Gene alleles and population genetics

Three alleles were detected in total for the *Clock* gene, corresponding to 8 (Q_8), 10 (Q_{10}), and 11 (Q_{11}) poly glutamine repeats (Table 2). For the Northern population, only two alleles were present with frequencies of 0.115 (Q_8) and 0.885 (Q_{10}) respectively. Tests for deviation from equilibrium, as measured by Hardy–Weinberg, were not significant ($\chi^2 = 0.142$, df = 1, $p = .71$). In the Northern population, the observed homozygosity was 0.769, while the calculated expected homozygosity was 0.788 for the *Clock* gene; the observed heterozygosity (H_o) was 0.231 while the calculated expected heterozygosity (H_e) was 0.212 for the *Clock* gene. For the Southern population, the same two alleles (Q_8 and Q_{10}) were detected, at frequencies of 0.118 (Q_8) and 0.853 (Q_{10}), respectively, while a third

TABLE 2 Allele frequencies and population genetics results for *Clock* and *Adcyap1* genes, including Hardy–Weinberg (HW) and heterozygosity (*H*), for Diederik cuckoo populations.

Pop	*N*	*Clock* gene						*Adcyap1* gene											
		Q_8	Q_{10}	Q_{11}	H_o	H_e	HW	142	144	146	148	150	152	154	156	158	160	H_o	H_e	HW
North	26	0.115	0.885	–	0.231	0.212	0.71	0.039	0.115	0.115	0.039	0.115	0.154	0.115	0.115	0.192	–	1.00	0.91	0.58
South	34	0.118	0.853	0.029	0.177	0.266	0.21	0.031	–	–	0.094	0.188	0.125	0.094	0.188	0.094	0.188	1.00	0.88	0.07

Note: For the *Clock* gene, two alleles (Q_8 and Q_{10}) were detected for specimens from the Northern range while three alleles (an additional Q_{11} allele) were detected for specimens from the Southern range. In the Northern population, the observed heterozygosity (H_o) was higher than the expected heterozygosity (H_e) while the inverse was true for the Southern population. In both populations, the *p*-value for Hardy–Weinberg was not significant and as such there was no evidence for deviation from equilibrium. For the *Adcyap1* gene, nine alleles were detected in the Northern population with 152 and 158 bp being most abundant, while seven alleles were detected for the Southern population with 150, 156 and 160 bp being the most abundant. Both populations were completely heterozygous with the Northern population passing equilibrium tests while the Southern population, still in equilibrium ($p < .05$), was close to deviation.

allele (Q_{11}) was detected in one individual (frequency of 0.029). As for the Northern population, the tests for deviation from equilibrium in the Southern population were not significant ($\chi^2 = 4.513$, df = 3, $p = .21$). In the Southern population, the observed and expected homozygosity were 0.824 and 0.734 respectively, while the observed (H_o) and expected heterozygosity (H_e) was 0.177 and 0.266. In both populations, most of the homozygous individuals were of the Q_{10}/Q_{10} genotype (Figure 3), however, one individual from South Africa was homozygous for the Q_8 allele (Q_8/Q_8). Heterozygous individuals included those with the Q_8/Q_{10} genotype, detected in both Northern and Southern populations, as well as a rare Q_{10}/Q_{11} genotype detected in one individual from South Africa.

Ten alleles were detected for the *Adcyap1* gene (Table 2) ranging from 142 base pairs (bp, 22 bp repeat) to 160 bp (40 bp repeat). In the Northern population, nine alleles were detected of which two represented private alleles of 144 and 146 base pairs, respectively. Five alleles were present at a frequency of 0.115 while the two most common alleles were 152 and 158 bp with a frequency of 0.154 and 0.192 respectively. Tests for deviation from equilibrium, as measured by Hardy–Weinberg, were not significant ($\chi^2 = 33.67$, df = 36, $p = .58$). All individuals were heterozygous while the calculated

FIGURE 3 Genotype distributions for the *Clock* gene in the Northern (green) and Southern (red) populations. In both populations the most common genotype was individuals homozygous for the Q_{10} allele. Furthermore, heterozygous individuals for Q_8 and Q_{10} were detected in both populations. The Northern population had two genotypes (Gt) and no private alleles. The Southern population had slightly higher genetic diversity evidenced by four genotypes, including a heterozygous individual for Q_{10} and the private allele Q_{11}. The grey bell-shaped normal curve was fitted over the distributions and shows evidence of purifying or stabilising selection. Graphs were created in R and the image created in BioRender.com.

expected heterozygosity was 0.91. The Southern population had eight alleles of which one allele (160 bp) was private to this region. Three alleles of 150, 156 and 160 bp represented the most abundant alleles, present at a frequency of 0.188 each. Tests for deviation from equilibrium by Hardy–Weinberg, detected a significant departure from equilibrium ($\chi^2 = 51.17$, df = 28, $p = .01$). As for the Northern population, all individuals were heterozygous with an expected heterozygosity of 0.88. Due to the high number of alleles, a total of 17 genotypes were detected (Figure 4). In the Northern population, nine genotypes were present at a near equal number apart from two genotypes, 144/156 and 150/158 bp, which had additional individuals. The Southern population had nine genotypes of which the two most common genotypes were 150/156 and 152/160 bp respectively with each having four individuals in total. Eight of the genotypes detected in the Northern population were not detected in the Southern population while six genotypes detected in the Southern population were absent from the Northern population; only three genotypes were conserved in both populations.

Based on the *Clock* gene approximately 123 migrants (N_m) were estimated per generation, while calculations based on the *Adcyap1* gene estimated at least 10 migrants (N_m) per generation. Population structure analyses, performed with STRUCTURE, supported a two-population model ($K = 2$) as the best explanatory model for the observed genetic diversity using both the *Clock* and *Adcyap1* genes (Figure 5). The results indicated a high degree of similarity between individuals from the North, indicated in green, and the South, indicated in red. A low level of admixture was observed between for both populations.

3.2 | Relationship between genes and location

Linear models fitted between the first or second alleles for the *Clock* gene, as well as the average or mean repeat length, and measures of locality such as latitude, longitude, and region did not yield any significant results for the total sample set (Table 3). A similar absence of correlation existed when fitting models using the two alleles and their mean length for males. There was, however, a significant correlation with both latitude ($F = 4.41$, df = 6, $R^2 = .42$, $p = .08$) and longitude ($F = 4.49$, df = 6, $R^2 = .43$, $p = .08$) for females in models fitted for the first, shorter allele. Similar correlates were, however, absent for the second, longer allele. This same trend was still present when comparing mean allele size to both latitude ($F = 3.38$, df = 6, $R^2 = .40$, $p = .09$) and longitude ($F = 4.37$, df = 6, $R^2 = .42$, $p = .08$). When comparing alleles to region, as either North or South, a correlation was also detected in relation to the length of the first allele ($F = 4.50$, df = 6, $R^2 = .43$, $p = .08$) as well as the mean allele length ($F = 4.32$, df = 6, $R^2 = .42$, $p = .08$).

Comparison between allele size of *Adcyap1* and latitude, longitude, as well as regions detected a significant relationship for both the shorter, longer and mean allele (Table 3). For the first, shorter, allele a relationship with latitude was detected for both the full data set ($F = 2.97$, df = 27, $R^2 = .10$, $p = .09$) as well as the separate analysis of females only ($F = 4.69$, df = 6, $R^2 = .44$, $p = .07$) while a similar relationship existed for the second, longer, allele for the full data set ($F = 2.84$, df = 27, $R^2 = .10$, $p = .10$). This relationship was still evident when analysing the mean allele size for the full data set ($F = 3.75$, df = 27, $R^2 = .12$, $p = .06$), although sex-specific effects were not detected. Correlations also existed between longitude and allele size for the first ($F = 4.78$, df = 27, $R^2 = .15$, $p = .04$), second ($F = 11.09$, df = 27, $R^2 = .29$, $p = .003$) and mean ($F = 9.17$, df = 27, $R^2 = .25$, $p = .005$) alleles. Sex-specific analyses showed a strong correlation for the first allele ($F = 4.91$, df = 6, $R^2 = .45$, $p = .07$) in females as well as the second ($F = 8.65$, df = 19, $R^2 = .31$, $p = .01$) and mean ($F = 5.28$, df = 19, $R^2 = .23$, $p = .03$) allele in males. When analysing regional differences, a correlation was detected for the first allele in the full dataset ($F = 5.33$, df = 27, $R^2 = .17$, $p = .03$) as well as females only ($F = 5.06$, df = 6, $R^2 = .46$, $p = .07$). A comparable relationship was detected for

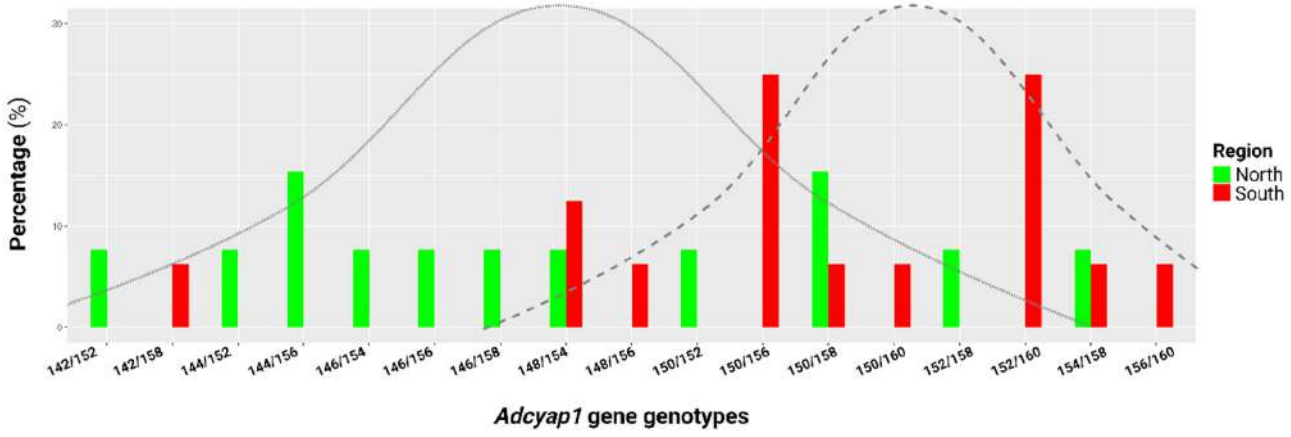


FIGURE 4 Genotype distributions for the *Adcyap1* gene in the Northern (green) and Southern (red) populations. In the Northern populations the most common genotypes were heterozygous 144/156 and 150/158, however, 11 genotypes (Gt) were detected in total with two private alleles (PA). The Southern population had slightly lower genetic diversity evidenced by nine genotypes, with only one private allele. The two most common genotypes were 150/156 and 152/160. The grey bell-shaped normal curve was fitted over the distributions and shows evidence of disruptive selection. Graphs were created in R and the image created in BioRender.com.

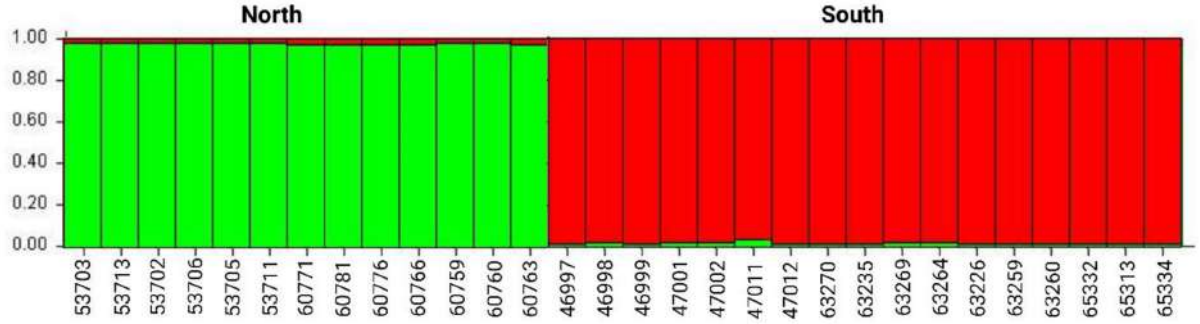


FIGURE 5 Bar plot of results from STRUCTURE analyses based on genotypes of two candidate genes, *Clock* and *Adcyap1*, showing the fraction of population identity on the y-axis and individual samples, grouped by region, on the x-axis. A two-population model was detected as the optimal descriptor of population structure with clear partitioning between the Northern (green) and Southern (red) populations, however, low levels (<1%) of admixture were still detected.

TABLE 3 Results for individual linear models between the *Clock* or *Adcyap1* alleles and phenology, fitted in *R*.

Linear model	*Clock* gene	*Adcyap1* gene
Allele 1:		
Latitude	$F=4.41$, $R^2=.42^*$ (♀)	$F=2.97$, $R^2=.10^*$
		$F=4.69$, $R^2=.44^*$ (♀)
Longitude	$F=4.49$, $R^2=.43^*$ (♀)	$F=4.78$, $R^2=.15^{**}$
		$F=4.91$, $R^2=.45^{**}$ (♀)
Date	–	$F=4.13$, $R^2=0.18^{**}$ (♂)
Region	$F=4.50$, $R^2=.43^*$ (♀)	$F=5.33$, $R^2=.17^{**}$
		$F=5.06$, $R^2=.46^{**}$ (♀)
Allele 2:		
Latitude	–	$F=2.84$, $R^2=.10^*$
Longitude	–	$F=11.09$, $R^2=.29^{**}$
		$F=8.65$, $R^2=.31^{**}$ (♂)
Date	$F=5.08$, $R^2=.15^{**}$	–
	$F=3.95$, $R^2=.17^*$ (♂)	
Region	–	$F=5.74$, $R^2=.18^{**}$
		$F=3.52$, $R^2=.16^*$ (♂)
Mean allele:		
Latitude	$F=3.38$, $R^2=.40^*$ (♀)	$F=3.75$, $R^2=.12^*$
Longitude	$F=4.37$, $R^2=.42^*$ (♀)	$F=9.17$, $R^2=.25^{**}$
		$F=5.28$, $R^2=.23^{**}$ (♂)
Date	–	$F=4.33$, $R^2=.19^{**}$ (♂)
Region	$F=4.32$, $R^2=.42^*$ (♀)	$F=7.23$, $R^2=.21^{**}$
		$F=3.57$, $R^2=.38^*$ (♀)

Note: Models were fitted for the first, shorter allele (allele 1) and second, longer allele (allele 2) as well as the average allele size. Only those models that showed a significant correlation are indicated. Both the *F*-test value (*F*) and the correlation coefficients (R^2) for the models are indicated as well as the group for which significance was achieved as either all, male (♂) or female (♀). Significance: $^*p \le .10$, $^{**}p \le .05$.

the second allele using all samples ($F=5.74$, df = 27, $R^2=.18$, $p=.02$) as well as for the males only ($F=3.52$, df = 19, $R^2=.16$, $p=.08$) analyses. This trend was conserved when assessing the mean allele for the whole population ($F=7.23$, df = 27, $R^2=.21$, $p=.01$) as well as in females ($F=3.57$, df = 6, $R^2=.38$, $p=.10$).

3.3 | Relationship between genes and timing

Models were fitted between both *Clock* alleles and their mean in relation to timing expressed as the date of capture and sampling as well as the date normalised between regions in relation to the Spring Equinox and Summer Solstice in either hemisphere. When comparing alleles for all study samples, no correlation was detected for the first allele, however, a significant correlation was observed for the second, longer allele ($F=5.08$, df = 28, $R^2=.15$, $p=.03$). This correlation was not conserved when analysing the mean allele length nor when normalising dates to either the equinox or solstice. In testing, the same models for either sex, a significant relationship was seen for the second allele in males ($F=5.08$, df = 19, $R^2=.15$, $p=.03$) but not in females ($F=0.50$, df = 6, $R^2=.08$, $p=.50$). Similar to the total population analyses, the relationship with timing was not detected when analysing the mean allele length in either males or females. Models fitted to the *Adcyap1* gene detected no relationship when analysing the total population; however, a relationship was detected when analysing only males for both the first, shorter, allele ($F=4.13$, df = 19, $R^2=.18$, $p=.06$), as well as the mean allele ($F=4.33$, df = 19, $R^2=.19$, $p=.05$).

3.4 | Data for host species

Reference genomes are not currently available for any of the host species which Diederik cuckoos parasitise. SRA sequence data were, however, available (Table S2) for three host species: The Cape wagtail ($N=5$), Red bishop ($N=5$) and Village weaver ($N=1$). BLAST searches of available sequence data from the NCBI SRA did not yield any data for the region of either gene analysed in this study for Red bishops or Cape wagtails. Therefore, data were retrieved for only one Village weaver sampled in Gabon. Based on the data retrieved from SRA, the Village weavers analysed ($N=1$) were homozygous

for the Q_8 allele for *Clock* and heterozygous 139/141 bp (22/24 bp repeat) for *Adcyap1*.

4 | DISCUSSION

Polymorphisms in candidate genes from the circadian clock circuitry have previously been associated with various attributes of migration phenology (Le Clercq, Bazzi, et al., 2023e; Le Clercq, Bazzi, et al., 2023f). Although a large body of evidence already exists (Le Clercq, Bazzi, Cecere, et al., 2023), thus far most studies were focused on passerine species of which only a fraction use habitat on the African continent. In this study, we used a candidate gene approach to assay polymorphisms in clock genes with putative associations to phenology based on the circadian clock, in the intra-African migratory species Diederik cuckoo. Polymorphisms were screened for 30 samples representing three sampling sites in West, East and South Africa. From the individuals screened, three alleles were identified for the *Clock* gene, while 10 alleles were detected for the *Adcyap1* gene. Least squares linear models identified putative associations between the shorter *Clock* alleles and timing and longer alleles were associated with habitat selection as measured by latitude. Similar correlations existed for *Adcyap1* where both alleles were correlated to latitude and longitude while the shorter and mean allele was correlated to timing. Interestingly, while some correlations could be detected at the population level, many of these attributes appear to have sex-specific contributions. This study presents the first evidence for an association between clock gene polymorphisms and phenology in a non-passerine intra-African migrant and provides some insight into extant genetic diversity that natural selection may act upon in response to the changing landscape of Africa.

Population genetic analyses showed distinctly higher frequencies for the most abundant *Clock* allele, Q_{10}, compared with the less abundant Q_8 allele. Due to many individuals carrying the Q_{10}/Q_{10} genotype, a high level of homozygosity and low heterozygosity was detected. In the case of the Northern individuals, the observed values were lower than expected for homozygosity while higher than expected heterozygosity was detected. The inverse trend was observed for the Southern individuals. Nevertheless, the homozygosity and heterozygosity between Northern and Southern populations were close in approximation to each other. The low number of alleles and high homozygosity is indicative of purifying selection (Cvijović et al., 2018; Hughes et al., 2003) which may either relate to the fact that these polymorphisms are contained in exonic sequence or be indicative that the Q_{10} allele provides an advantage in these species (Cassidy, 2021). For the *Adcyap1* gene, more alleles and genotypes were detected for the Northern population in comparison to the Southern population; although both populations showed substantially more diversity in *Adcyap1* as compared to *Clock*. These alleles tended to be shorter in the Northern population while longer alleles were prevalent in the Southern population. This may be indicative of disruptive selection (Gross, 1985; Hendry et al., 2009) due to shorter alleles providing an advantage in Northern individuals while longer alleles are favoured in Southern individuals. Due to the large number of alleles, all individuals in both populations were heterozygous as the probability of heterozygosity increases with alleles numbers while heterozygosity itself can aid in the expansion of allele numbers (Amos et al., 2008). Concurrently, the high heterozygosity may indicate a case of heterozygote advantage (Cassidy, 2021). The expansion of allelic variants for *Adcyap1* may also be related to the polymorphism occurring in the 3′-UTR which, as a non-coding gene segment, is less constrained (Chen et al., 2012; Mazumder et al., 2003; Steri et al., 2018).

With the exception of the *Adcyap1* gene in the Southern population, tests of equilibrium showed no significant deviations which is indicative that allele frequencies may stay the same between generations (Hedrick, 1987) in the Northern ranges with some evidence of differentiation happening in the South; consistent with previous phylogenetic studies in this species using nuclear markers and mitochondrial DNA (Smith et al., 2023). A recent review on clock genes in migrating birds (Le Clercq, Bazzi, et al., 2023e; Le Clercq, Bazzi, et al., 2023f) found similar patterns of low heterozygosity and high homozygosity for populations in equilibrium for the *Clock* gene and higher heterozygosity for *Adcyap1*. This may indicate that the patterns of inheritance and selection in the *Clock* and *Adcyap1* genes are conserved among most avian lineages. It should be noted that heterozygosity has the potential to increase diversity at polymorphic sites (Amos et al., 2008) and, as such, the presence of heterozygous individuals in Diederik cuckoos may confer continued or improved fitness over time by providing more genotypes for natural selection to act upon in adaptive evolution (Cassidy, 2021). This will be critical given the complex landscape of expanding human populations on the African continent with likely increases in agricultural use that could encroach into the habitats of cuckoos and their hosts (Güneralp et al., 2018), possibly requiring shifts in staging sites as well as habitat selection.

Three alleles were detected for the Diederik cuckoo *Clock* gene, which is similar to species such as the Barn swallow, *Hirundo rustica* (Bazzi et al., 2015) and Eurasian hoopoe, *Upupa epops* (Bazzi, Cecere, et al., 2016), but substantially less diverse than alleles detected in warbler species (Le Clercq, Bazzi, et al., 2023e; Le Clercq, Bazzi, et al., 2023f). On the other hand, the *Adcyap1* gene had 10 alleles which is comparable to observations in several warbler species, including the Eurasian reed warbler, *Acrocephalus scirpaceus* (Bazzi, Cecere, et al., 2016), and Willow warbler, *Phylloscopus trochilis* (Bazzi et al., 2017). Although non-passerine species remain largely data deficient and lack data for comparison among cuckoos, a previous study did explore clock gene polymorphisms in the European nightjar, *Caprimulgus europaeus* (Bazzi, Cecere, et al., 2016). Cuckoos, as part of the superorder *Otidimorphae*, and Nightjars, as part of the superorder *Strisores*, both form part of the *Otidae* clade of avian species (Jarvis et al., 2014; Prum et al., 2015). In comparison, the Diederik cuckoo has three alleles ranging from 8 to 11 poly-Q repeats for the *Clock* gene while the European nightjar has two alleles ranging from

eight to nine poly-Q repeats. For the *Adcyap1* gene, Diederik cuckoos had 10 alleles ranging in size from 142 to 160 bp while European nightjars had nine alleles ranging from 146 to 160 bp. It is therefore likely that a range of between 8 and 10 repeats for *Clock*, and 144–160 bp for *Adcyap1*, is common within this lineage. In comparison to the passerine hosts that are parasitised by Diederik cuckoos, the single Village weaver was found to have alleles of approximately eight repeats in *Clock* and *Adcyap1* genes of approximately 139–141 bp (or 22–24 bp repeats). This overlap may indicate either possible co-evolution or specialised selection between cuckoos and their hosts based on clock genes, consistent with previous studies on brood parasites (Soler et al., 2015; Soler & Soler, 2000).

Comparisons between both *Clock* and *Adcyap1* alleles and phenology found several phenotypic correlations. The first was evidence of a relationship between the first, shorter *Clock* alleles and both latitude as well as longitude. The second was a relationship between the first *Clock* allele and region. Interestingly, both observations were made when analysing only female individuals. For *Adcyap1*, phenotypic correlates were found in relation to both latitude and longitude as well as region. In the first, shorter, allele this relationship appeared to be largely present in females, while the correlation found for the second, longer, allele showed stronger evidence for males. This is consistent with previous findings that clock genes show strong spatial patterns in females that are generally more selective of breeding site (Liedvogel et al., 2009). Lastly, evidence of a relationship was detected for the second, longer *Clock* allele and the first, shorter, *Adcyap1* allele with timing, which was observed in males whereas the observed relationship was diminished for females. This is consistent with previous studies that have also found evidence for a sex-specific effect of clock genes on timing (Bazzi et al., 2017). While several of these correlates had moderate statistical support, this does not abrogate their significance and could likely be improved through increased sample sizes to reduce variance.

Our analyses also addressed a pertinent question previously raised regarding the treatment of data from diploid alleles in analyses such as migration genetics (Le Clercq, Bazzi, et al., 2023e; Le Clercq, Bazzi, et al., 2023f). Studies have applied varied methods when analysing diploid polymorphic repeat data including analysing the sum of alleles, the mean of alleles or only the longer alleles. No evidence of parental imprinting, which typically results in the transcription of only one allele (Andergassen et al., 2021), was found in analyses of methylation in the clock genes of birds (Saino et al., 2019). It therefor stands that both genes are transcribed and differentially contribute to phenotypes. There is, however, no knowledge as to whether both genes are co-dominant or if a specific allele contributes more to any given trait. For this reason, each allele was tested individually (in addition to testing mean allele length) and evidence is provided that individual alleles may affect phenotypes more so than could be detected by only analysing either the sum, mean or presumed to be dominant alleles.

Several limitations existed for the present study. Firstly, a high degree of collinearity between measures, such as latitude, longitude and date, precluded the use of multiple linear regression to model several traits in combination with each other. This is because collinearity may inflate statistical measurements such as the correlation coefficients and *p*-values which in turn provides exaggerated effect measures (Belsley et al., 1980; Chen et al., 2024; Mason & Perreault, 1991). Models with multiple testing would also require a statistical correction, which is known to over-correct for small effect size measures, often limiting the detection of biologically relevant observations (Drezner & Drezner, 2016; Tanner et al., 1997). Secondly, the observed differences between sexes might indicate possible sex-specific patterns of inheritance which were outside the scope of the present study (Laurentin Táriba, 2023; Wright et al., 2004). Thirdly, although samples were included from the Northern population that likely contains year-round residents (Uganda), as well as available data for host species (Village weaver) from databases, the sampling number was too low for any meaningful comparison between migratory and resident populations and the comparison to one individual from host species cannot account for the potential existence of greater genetic diversity for clock genes in host species. It should be noted; however, that previous studies have illustrated that neither of the tested genes serve as diagnostic markers to differentiate resident and migratory species (Le Clercq, Bazzi, Cecere, et al., 2023; Lugo Ramos et al., 2017). Lastly, detailed information about the exact timing, direction and distance of migration from, for example, a geolocator (Bazzi et al., 2015; Saino et al., 2015) or isotope study (Bazzi, Galimberti, et al., 2016; Ramudzuli, 2021), was not available and as such a more direct measure for these attributes was not available.

Future studies are needed to expand upon the results of this study. Recommendations include studies on other intra-African migrants from non-passerine orders. For the Diederik cuckoo, future work could include a larger sample size for female individuals as well as include individuals from populations that are more likely to show divergence. This could extend to resident birds on islands off the West and East coast of Africa, such as São Tomé and Príncipe or Zanzibar, and migrants that migrate to Oman in the Middle East to breed. More data are also needed on the clock genes of birds that serve as hosts for cuckoos, such as Village weavers, Cape wagtails and Red bishops, in order to fully explore their co-evolution.

AUTHOR CONTRIBUTIONS

L. S. Le Clercq: Conceptualization (lead); data curation (lead); formal analysis (lead); methodology (lead); validation (lead); visualization (lead); writing – original draft (lead). **V. Phetla:** Data curation (supporting); formal analysis (supporting); writing – review and editing (supporting). **S. T. Osinubi:** Funding acquisition (supporting); investigation (supporting); resources (supporting); writing – review and editing (supporting). **A. Kotzé:** Project administration (supporting); writing – review and editing (supporting). **J. P. Grobler:** Project administration (supporting); supervision (supporting); writing – review and editing (supporting). **D. L. Dalton:** Conceptualization (supporting); funding acquisition (lead); project administration (equal); resources (lead); supervision (lead); writing – review and editing (lead).

ACKNOWLEDGEMENTS

Images were created in BioRender.com. This work is based on the research supported wholly/in part by the National Research Foundation of South Africa (Grant Numbers: 112062) and the African Bird Club (Expedition Award).

CONFLICT OF INTEREST STATEMENT

The authors have no competing interests to declare.

DATA AVAILABILITY STATEMENT

Sample details were deposited in the National Centre for Biotechnology Information (NCBI) BioSample database while sequences generated in this study were deposited to the NBCI Nucleotide collection. Accession numbers with associated links to NCBI BioProject PRJNA737185 are listed in (Table S1) (Genbank accession numbers *Clock*: OR909322–51, *Adcyap1*: PP112918–46). Additional data including individual allele data are available from the Zenodo depository at the following link: https://doi.org/10.5281/zenodo.10211358. Allele data have also been added to the Avian Clocks Data database (Le Clercq, 2023b; Le Clercq, Bazzi, et al., 2023e). Methods are available from Protocols.io (Le Clercq et al., 2023c, Le Clercq et al., 2023d).

ORCID

L. S. Le Clercq https://orcid.org/0000-0002-8713-8920
V. Phetla https://orcid.org/0000-0001-5707-9221
S. T. Osinubi https://orcid.org/0000-0002-5143-6985
A. Kotzé https://orcid.org/0000-0003-2367-1483
J. P. Grobler https://orcid.org/0000-0002-5913-7031
D. L. Dalton https://orcid.org/0000-0001-5975-6425

REFERENCES

Amos, W., Flint, J., & Xu, X. (2008). Heterozygosity increases microsatellite mutation rate, linking it to demographic history. *BMC Genetics*, *9*, 72. https://doi.org/10.1186/1471-2156-9-72

Andergassen, D., Smith, Z. D., Kretzmer, H., Rinn, J. L., & Meissner, A. (2021). Diverse epigenetic mechanisms maintain parental imprints within the embryonic and extraembryonic lineages. *Developmental Cell*, *56*(21), 2995–3005.e4. https://doi.org/10.1016/j.devcel.2021.10.010

Bazzi, G., Ambrosini, R., Caprioli, M., Costanzo, A., Liechti, F., Gatti, E., Gianfranceschi, L., Podofillini, S., Romano, A., Romano, M., Scandolara, C., Saino, N., & Rubolini, D. (2015). Clock gene polymorphism and scheduling of migration: A geolocator study of the barn swallow *Hirundo rustica*. *Scientific Reports*, *5*(July), 12443. https://doi.org/10.1038/srep12443

Bazzi, G., Cecere, J. G., Caprioli, M., Gatti, E., Gianfranceschi, L., Podofillini, S., Possenti, C. D., Ambrosini, R., Saino, N., Spina, F., & Rubolini, D. (2016). Clock gene polymorphism, migratory behaviour and geographic distribution: A comparative study of trans-Saharan migratory birds. *Molecular Ecology*, *25*(24), 6077–6091. https://doi.org/10.1111/mec.13913

Bazzi, G., Galimberti, A., Hays, Q. R., Bruni, I., Cecere, J. G., Gianfranceschi, L., Hobson, K. A., Morbey, Y. E., Saino, N., Guglielmo, C. G., & Rubolini, D. (2016). Adcyap1 polymorphism covaries with breeding latitude in a Nearctic migratory songbird, the Wilson's warbler (*Cardellina pusilla*). *Ecology and Evolution*, *6*(10), 3226–3239. https://doi.org/10.1002/ece3.2053

Bazzi, G., Podofillini, S., Gatti, E., Gianfranceschi, L., Cecere, J. G., Spina, F., Saino, N., & Rubolini, D. (2017). Candidate genes have sex-specific effects on timing of spring migration and moult speed in a long-distance migratory bird. *Current Zoology*, *63*(5), 479–486. https://doi.org/10.1093/cz/zow103

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Identifying influential data and sources of collinearity*. John Wiley & Sons, Inc. https://doi.org/10.1002/0471725153

Bernstein, H. (2013). Commercial agriculture in South Africa since 1994: 'Natural, simply capitalism'. *Journal of Agrarian Change*, *13*(1), 23–46. https://doi.org/10.1111/JOAC.12011

Bhattacharya, D., & Van Meir, E. G. (2019). A simple genotyping method to detect small CRISPR-Cas9 induced indels by agarose gel electrophoresis. *Scientific Reports*, *9*(1), 4437. https://doi.org/10.1038/s41598-019-39950-4

Boddaert, P. (1783). *Table des planches enluminéez d'histoire naturelle de M. D'Aubenton: avec les denominations de M.M. de Buffon, Brisson, Edwards, Linnaeus et Latham, precedé d'une notice des principaux ouvrages zoologiques enluminés*. Utrecht: s. n. https://doi.org/10.5962/BHL.TITLE.39835

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., Suchard, M. A., Rambaut, A., & Drummond, A. J. (2014). BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, *10*(4), e1003537. https://doi.org/10.1371/journal.pcbi.1003537

Brooker, L. C., & Brooker, M. G. (1990). Why are cuckoos host specific? *Oikos*, *57*(3), 301. https://doi.org/10.2307/3565958

Carey, C. (2009). The impacts of climate change on the annual cycles of birds. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *364*(1534), 3321–3330. https://doi.org/10.1098/RSTB.2009.0182

Carver, S. (2019). Rewilding through land abandonment. In N. Pettorelli, S. M. Durant, & J. T. du Toit (Eds.), *Rewilding* (pp. 99–122). Cambridge University Press. https://doi.org/10.1017/9781108560962.006

Cassidy, M. (2021). *Biological evolution: An introduction*. Cambridge University Press. https://worldcat.org/title/1156425756

Cassone, V. M. (2014). Avian circadian organization: A chorus of clocks. *Frontiers in Neuroendocrinology*, *35*(1), 76–88. https://doi.org/10.1016/j.yfrne.2013.10.002

Cassone, V. M., & Westneat, D. F. (2012). The bird of time: Cognition and the avian biological clock. *Frontiers in Molecular Neuroscience*, *5*, 32. https://doi.org/10.3389/fnmol.2012.00032

Caves, E. M., Stevens, M., Iversen, E. S., & Spottiswoode, C. N. (2015). Hosts of avian brood parasites have evolved egg signatures with elevated information content. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1810), 20150598. https://doi.org/10.1098/rspb.2015.0598

Chen, C. Y., Chen, S. T., Juan, H. F., & Huang, H. C. (2012). Lengthening of 3′UTR increases with morphological complexity in animal evolution. *Bioinformatics*, *28*(24), 3178–3181. https://doi.org/10.1093/BIOINFORMATICS/BTS623

Chen, X., Liang, Y., & Feng, X. (2024). Influence of model complexity, training collinearity, collinearity shift, predictor novelty and their interactions on ecological forecasting. *Global Ecology and Biogeography*, *33*(3), 371–384. https://doi.org/10.1111/geb.13793

Cvijović, I., Good, B. H., & Desai, M. M. (2018). The effect of strong purifying selection on genetic diversity. *Genetics*, *209*(4), 1235–1278. https://doi.org/10.1534/GENETICS.118.301058

de Vos, C., & Cherry, M. I. (2017). Declining ranges of southern African breeding cuckoos do not reflect host species' range changes. *Emu - Austral Ornithology*, *117*(2), 188–193. https://doi.org/10.1080/01584197.2017.1292403

Douglas, D. J. T., Newson, S. E., Leech, D. I., Noble, D. G., & Robinson, R. A. (2010). How important are climate-induced changes in host availability for population processes in an obligate brood parasite,

the European cuckoo? *Oikos*, *119*(11), 1834–1840. https://doi.org/10.1111/j.1600-0706.2010.18388.x

Drezner, Z., & Drezner, T. D. (2016). A remedy for the overzealous Bonferroni technique for multiple statistical tests. *The Bulletin of the Ecological Society of America*, *97*(1), 91–98. https://doi.org/10.1002/bes2.1214

Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*(2), 359–361. https://doi.org/10.1007/S12686-011-9548-7/METRICS

Ferguson, J. W. H. (1994). The importance of low host densities for successful parasitism of diederik cuckoos on red bishop birds. *South African Journal of Zoology*, *29*(1), 70–73. https://doi.org/10.1080/02541858.1994.11448329

Gee, J. P., & Heigham, J. B. (1977). Birds of Lagos, Nigeria. *Bulletin of the Nigerian Ornithological Society*, *13*(44), 103–132.

Greig, E. I., Wood, E. M., & Bonter, D. N. (2017). Winter range expansion of a hummingbird is associated with urbanization and supplementary feeding. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1852), 20170256. https://doi.org/10.1098/rspb.2017.0256

Gross, M. R. (1985). Disruptive selection for alternative life histories in salmon. *Nature*, *313*(5997), 47–48. https://doi.org/10.1038/313047a0

Güneralp, B., Lwasa, S., Masundire, H., Parnell, S., & Seto, K. C. (2018). Urbanization in Africa: Challenges and opportunities for conservation. *Environmental Research Letters*, *13*(1), 015002. https://doi.org/10.1088/1748-9326/aa94fe

Hall, T. (2011). BioEdit: An important software for molecular biology. *GERF Bulletin of Biosciences*, *2*(June), 60–61. https://doi.org/10.1002/prot.24632

Hanmer, D. B. (1980). Mensural and moult data of eight species of kingfisher from Mocambique and Malawi. *Ostrich*, *51*(3), 129–150. https://doi.org/10.1080/00306525.1980.9633556

Hardy, G. H. (1908). Mendelian proportions in a mixed population. *Science*, *28*(706), 49–50. https://doi.org/10.1126/SCIENCE.28.706.49

Hedrick, P. W. (1987). Gametic disequilibrium measures: Proceed with caution. *Genetics*, *117*(2), 331–341. https://doi.org/10.1093/genetics/117.2.331

Hendry, A. P., Huber, S. K., De León, L. F., Herrel, A., & Podos, J. (2009). Disruptive selection in a bimodal population of Darwin's finches. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1657), 753–759. https://doi.org/10.1098/rspb.2008.1321

Hockey, P. A. R. (1994). *Birds of southern Africa*. Struik Publishers. https://worldcat.org/title/869997991

Honeywell, R. A. (1971). Some interesting records from Ghana. *Bulletin of the Nigerian Ornithological Society*, *8*(30), 30–32.

Hoogendoorn, G., Meintjes, D., Kelso, C., & Fitchett, J. (2019). Tourism as an incentive for rewilding: The conversion from cattle to game farms in Limpopo province, South Africa. *Journal of Ecotourism*, *18*(4), 309–315. https://doi.org/10.1080/14724049.2018.1502297

Hughes, A. L., Packer, B., Welch, R., Bergen, A. W., Chanock, S. J., & Yeager, M. (2003). Widespread purifying selection at polymorphic sites in human protein-coding loci. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(26), 15754–15757. https://doi.org/10.1073/PNAS.2536718100

Jarvis, E. D., Mirarab, S., Aberer, A. J., Li, B., Houde, P., Li, C., Ho, S. Y., Faircloth, B. C., Nabholz, B., Howard, J. T., Suh, A., Weber, C. C., da Fonseca, R. R., Li, J., Zhang, F., Li, H., Zhou, L., Narula, N., Liu, L., … Zhang, G. (2014). Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, *346*(6215), 1320–1331. https://doi.org/10.1126/science.1253451

Jenni, L., & Kéry, M. (2003). Timing of autumn bird migration under climate change: Advances in long-distance migrants, delays in short-distance migrants. *Proceedings of the Royal Society B: Biological Sciences*, *270*(1523), 1467–1471. https://doi.org/10.1098/rspb.2003.2394

Jensen, R. A. C., & Jensen, M. K. (1969). On the breeding biology of southern African cuckoos. *Ostrich*, *40*(4), 163–181. https://doi.org/10.1080/00306525.1969.9634345

Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., & Mooers, A. O. (2012). The global diversity of birds in space and time. *Nature*, *491*(7424), 444–448. https://doi.org/10.1038/nature11631

Johnsen, A., Fidler, A. E., Kuhn, S., Carter, K. L., Hoffmann, A., Barr, I. R., Biard, C., Charmantier, A., Eens, M., Korsten, P., Siitari, H., Tomiuk, J., & Kempenaers, B. (2007). Avian clock gene polymorphism: Evidence for a latitudinal cline in allele frequencies. *Molecular Ecology*, *16*(22), 4867–4880. https://doi.org/10.1111/j.1365-294X.2007.03552.x

Justen, H., & Delmore, K. E. (2022). The genetics of bird migration. *Current Biology*, *32*(20), R1144–R1149. https://doi.org/10.1016/j.cub.2022.07.008

Justen, H., Hasselmann, T., Illera, J. C., Delmore, K. E., Serrano, D., Flinks, H., & Liedvogel, M. (2022). Population-specific association of clock gene polymorphism with annual cycle timing in stonechats. *Scientific Reports*, *12*(1), 7947. https://doi.org/10.1038/s41598-022-11158-z

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Krist, M., Munclinger, P., Briedis, M., & Adamík, P. (2021). The genetic regulation of avian migration timing: Combining candidate genes and quantitative genetic approaches in a long-distance migrant. *Oecologia*, *196*(2), 373–387. https://doi.org/10.1007/s00442-021-04930-x

Kumar, S., Suleski, M., Craig, J. M., Kasprowicz, A. E., Sanderford, M., Li, M., Stecher, G., & Hedges, S. B. (2022). TimeTree 5: An expanded resource for species divergence times. *Molecular Biology and Evolution*, *39*(8), msac174. https://doi.org/10.1093/molbev/msac174

Langmore, N. E., Maurer, G., Adcock, G. J., & Kilner, R. M. (2008). Socially acquired host-specific mimicry and the evolution of host races in Horsfield's bronze-cuckoo Chalcites basalis. *Evolution*, *62*(7), 1689–1699. https://doi.org/10.1111/j.1558-5646.2008.00405.x

Laurentin Táriba, H. E. (2023). Sex-linked inheritance. *Agricultural Genetics* (pp. 95–101). Springer, Cham. https://doi.org/10.1007/978-3-031-37192-9_7

Le Clercq, L. (2023a). PAReTT: Python automated retrieval of TimeTree data. *GitHub*. https://github.com/LSLeClercq/PAReTT

Le Clercq, L., Bazzi, G., Cecere, J. G., Gianfranceschi, L., Grobler, J. P., Kotzé, A., Rubolini, D., Liedvogel, M., & Dalton, D. L. (2023). Time trees and clock genes: A systematic review and comparative analysis of contemporary avian migration genetics. *Biological Reviews*, *98*(4), 1051–1080. https://doi.org/10.1111/brv.12943

Le Clercq, L., Dalton, D. L., Kotzé, A., & Grobler, J. P. (2023c). ABI sanger sequencing of avian clock genes to elucidate markers for migration phenology. https://doi.org/10.17504/protocols.io.3byl4k6zrvo5/v1

Le Clercq, L., Dalton, D. L., Kotzé, A., & Grobler, J. P. (2023d). PCR amplification of clock and Adcyap1 genes with EmeraldAmp® GT PCR master mix in avian species for polymorphism elucidation. *Protocols*. https://doi.org/10.17504/protocols.io.6qpvrdwk3gmk/v1

Le Clercq, L., Kotzé, A., Grobler, J. P., & Dalton, D. L. (2023). PAReTT: A python package for the automated retrieval and management of divergence time data from the TimeTree resource for downstream analyses. *Journal of Molecular Evolution*, *91*(4), 502–513. https://doi.org/10.1007/s00239-023-10106-3

Le Clercq, L. S. (2023b). AvianClocksData: Dataset of clock and Adcyap1 alleles for birds. *GitHub*. https://github.com/LSLeClercq/AvianClocksData

Le Clercq, L.-S., Bazzi, G., Ferrer Obiol, J., Cecere, J., Gianfranceschi, L., Grobler, J. P., Rubolini, D., Liedvogel, M., & Dalton, D. L. (2023e). Birds of a feather flock together—A compilation of data for two candidate genes for migration genetics. Figshare, Dataset https://doi.org/10.6084/m9.figshare.23623935.v1

Le Clercq, L.-S., Bazzi, G., Ferrer Obiol, J., Cecere, J., Gianfranceschi, L., Grobler, J. P., Rubolini, D., Liedvogel, M., & Dalton, D. L. (2023f). Birds of a feather flock together: A dataset for clock and Adcyap1 genes from migration genetics studies. *Scientific Data*, *10*(1), 787. https://doi.org/10.1038/s41597-023-02717-8

Leclerc, G.-L. (1780). Le coucou vert-doré et blanc. In *Histoire Naturelle des Oiseaux* (Vol. *12*, pp. 42–44). De l'imprimerie Royale. https://doi.org/10.5962/BHL.TITLE.66490

Levene, H. (1949). On a matching problem arising in genetics. *The Annals of Mathematical Statistics*, *20*(1), 91–94. https://doi.org/10.1214/aoms/1177730093

Lidzhegu, Z., & Palamuleni, L. G. (2012). Land use and land cover change as a consequence of the south African land reform program: A remote sensing approach. *Journal of Food, Agriculture and Environment*, *10*(3–4), 1441–1447. https://doi.org/10.1234/4.2012.3708

Liedvogel, M., Szulkin, M., Knowles, S. C. L., Wood, M. J., & Sheldon, B. C. (2009). Phenotypic correlates of clock gene variation in a wild blue tit population: Evidence for a role in seasonal timing of reproduction. *Molecular Ecology*, *18*(11), 2444–2456. https://doi.org/10.1111/j.1365-294X.2009.04204.x

Lindholm, A. K., Venter, G. J., & Ueckermann, E. A. (1998). Persistence of passerine ectoparasites on the diederik cuckoo Chrysococcyx caprius. *Journal of Zoology*, *244*(1), 145–153. https://doi.org/10.1017/S0952836998001162

Lugo Ramos, J. S., Delmore, K. E., & Liedvogel, M. (2017). Candidate genes for migration do not distinguish migratory and non-migratory birds. *Journal of Comparative Physiology A*, *203*(6–7), 383–397. https://doi.org/10.1007/s00359-017-1184-6

Macdonald, M. A. (1980). Observations on the Diederik cuckoo in southern Ghana. *Ostrich*, *51*(2), 75–79. https://doi.org/10.1080/00306525.1980.9633547

Maclean, G. L., Roberts, A., Newman, K., & Lockwood, G. (1993). *Roberts' birds of southern Africa (sixth edit)*. Trustees of the John Voelcker Bird Book Fund Cape Town. https://worldcat.org/title/29301631

Mason, C. H., & Perreault, W. D. (1991). Collinearity, power, and interpretation of multiple regression analysis. *Journal of Marketing Research*, *28*(3), 268–280. https://doi.org/10.1177/002224379102800302

Mazumder, B., Seshadri, V., & Fox, P. L. (2003). Translational control by the 3′-UTR: The ends specify the means. *Trends in Biochemical Sciences*, *28*(2), 91–98. https://doi.org/10.1016/S0968-0004(03)00002-1

McCusker, B. (2004). Land use and cover change as an indicator of transformation on recently redistributed farms in Limpopo Province, South Africa. *Human Ecology*, *32*(1), 49–75. https://doi.org/10.1023/B:HUEC.0000015220.22795.27

Morel, G. (1972). *Liste commentée des oiseaux du Sénégal et de la Gambie, Supplément n 1*. ORSTOM.

Navarro, L. M., & Pereira, H. M. (2015). Rewilding abandoned landscapes in Europe. In *Rewilding European landscapes* (pp. 3–23). Springer Nature. https://doi.org/10.1007/978-3-319-12039-3_1

Pancerasa, M., Ambrosini, R., Saino, N., & Casagrandi, R. (2018). Barn swallows long-distance migration occurs between significantly temperature-correlated areas. *Scientific Reports*, *8*(1), 12359. https://doi.org/10.1038/s41598-018-30849-0

Payne, R. B., Bonan, A., & Boesman, P. F. D. (2021). Dideric cuckoo (*Chrysococcyx caprius*), version 1.1. In J. del Hoyo, A. Elliott, J. Sargatal, D. Christie, & E. de Juana (Eds.), *Birds of the world*. Cornell Lab of Ornithology. https://doi.org/10.2173/bow.didcuc1.01.1

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959. https://doi.org/10.1093/genetics/155.2.945

Prum, R. O., Berv, J. S., Dornburg, A., Field, D. J., Townsend, J. P., Lemmon, E. M., & Lemmon, A. R. (2015). A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*, *526*(7574), 569–573. https://doi.org/10.1038/nature15697

Pulido, F., Berthold, P., Mohr, G., & Querner, U. (2001). Heritability of the timing of autumn migration in a natural bird population. *Proceedings of the Royal Society B: Biological Sciences*, *268*(1470), 953–959. https://doi.org/10.1098/rspb.2001.1602

R Core Team. (2020). *R: A language and environment for Statisitical computing*. R Foundation for Statistical Computing. https://www.r-project.org/

Ramudzuli, A. (2021). *Flight feather moult patterns and stable isotope analysis in the woodland Kingfisher (Halcyon senegalensis)*. University of Cape Town.

Reed, R. A. (1968). Studies of the Diederik cuckoo *Chrysococcyx caprius* in the Transvaal. *Ibis*, *110*(3), 321–331. https://doi.org/10.1111/j.1474-919X.1968.tb00040.x

Robart, A. R., McGuire, M. M. K., & Watts, H. E. (2018). Increasing photoperiod stimulates the initiation of spring migratory behaviour and physiology in a facultative migrant, the pine siskin. *Royal Society Open Science*, *5*(8), 180876. https://doi.org/10.1098/rsos.180876

Rönkä, K. H., Eroukhmanoff, F., Kulmuni, J., Nouhaud, P., & Thorogood, R. (2022). Beyond genes-for-behaviour: The potential for genomics to resolve questions in avian brood parasitism. https://doi.org/10.32942/X21592

Rowan, M. K. (1983). *The doves, parrots, louries, and cuckoos of southern Africa*. D. Philip.

Rubolini, D., Møller, A. P., Rainio, K., & Lehikoinen, E. (2007). Intraspecific consistency and geographic variability in temporal trends of spring migration phenology among European bird species. *Climate Research*, *35*(1–2), 135–146. https://doi.org/10.3354/cr00720

Saino, N., Albetti, B., Ambrosini, R., Caprioli, M., Costanzo, A., Mariani, J., & Bollati, V. (2019). Inter-generational resemblance of methylation levels at circadian genes and associations with phenology in the barn swallow. *Scientific Reports*, *9*(1), 6505. https://doi.org/10.1038/s41598-019-42798-3

Saino, N., Rubolini, D., Ambrosini, R., Romano, M., Scandolara, C., Fairhurst, G. D., Caprioli, M., Romano, A., Sicurella, B., & Liechti, F. (2015). Light-level geolocators reveal covariation between winter plumage molt and phenology in a trans-Saharan migratory bird. *Oecologia*, *178*(4), 1105–1112. https://doi.org/10.1007/S00442-015-3299-1

Saino, N., Rubolini, D., Lehikoinen, E., Sokolov, L. V., Bonisoli-Alquati, A., Ambrosini, R., & Møller, A. P. (2009). Climate change effects on migration phenology may mismatch brood parasitic cuckoos and their hosts. *Biology Letters*, *5*(4), 539–541. https://doi.org/10.1098/RSBL.2009.0312

Skead, C. J. (1952). Cuckoo studies on a south African farm (part II). *Ostrich*, *23*(1), 2–15. https://doi.org/10.1080/00306525.1952.9632597

Smith, R. M., Dalton, D. L., Mwale, M., Nupen, L. J., Pretorius, C., Bojko, J., & Osinubi, S. T. (2023). Assessment of genetic and morphological differentiation among populations of the Diederik cuckoo Chrysococcyx caprius. *Ostrich*, *94*(2), 86–99. https://doi.org/10.2989/00306525.2023.2222230

Soler, J. J., Ruiz Castellano, C., Martínez-de la Puente, J., Tomás, G., Ruiz-Rodríguez, M., & Figuerola, J. (2015). Telomere dynamics in parasitic great spotted cuckoos and their magpie hosts. *Journal of Evolutionary Biology*, *28*(9), 1610–1617. https://doi.org/10.1111/JEB.12680

Soler, J., & Soler, M. (2000). Brood-parasite interactions between great spotted cuckoos and magpies: A model system for studying coevolutionary relationships. *Oecologia*, *125*(3), 309–320. https://doi.org/10.1007/s004420000487

Spottiswoode, C. N. (2010). The evolution of host-specific variation in cuckoo eggshell strength. *Journal of Evolutionary Biology*, *23*(8), 1792–1799. https://doi.org/10.1111/j.1420-9101.2010.02010.x

Stephan, F. K. (2002). The "other" circadian system: Food as a zeitgeber. *Journal of Biological Rhythms*, *17*(4), 284–292. https://doi.org/10.1177/074873002129002591

Steri, M., Idda, M. L., Whalen, M. B., & Orrù, V. (2018). Genetic variants in mRNA untranslated regions. *Wiley Interdisciplinary Reviews: RNA*, *9*(4), e1474. https://doi.org/10.1002/WRNA.1474

Tanner, M. S., Sharrard, M. J., & Rigby, A. S. (1997). Gene polymorphisms and the use of the Bonferroni correction factor: When and when not to apply? [4]. *Archives of Disease in Childhood*, *76*(4), 386. https://doi.org/10.1136/adc.76.4.385c

Tarboton, W. R., Kemp, M., & Kemp, A. C. (1987). *Birds of the Transvaal*. Transvaal Museum Pretoria. https://worldcat.org/title/18808605

Verhoeven, M. A., Loonstra, A. H. J., Hooijmeijer, J. C. E. W., Masero, J. A., Piersma, T., & Senner, N. R. (2018). Generational shift in spring staging site use by a long-distance migratory bird. *Biology Letters*, *14*(2), 20170663. https://doi.org/10.1098/RSBL.2017.0663

Weinberg, W. (1908). On the demonstration of heredity in man. *(1963) papers on human genetics*.

Wright, T. F., Brittan-Powell, E. F., Dooling, R. J., & Mundinger, P. C. (2004). Sex-linked inheritance of hearing and song in the Belgian Waterslager canary. *Proceedings of the Royal Society B: Biological Sciences*, *271*, S409–S412. https://doi.org/10.1098/rsbl.2004.0204

Yeh, F. C., Yang, R. C., Boyle, T. B. J., Ye, Z. H., & Mao, J. X. (1997). POPGENE, the user-friendly shareware for population genetic analysis. *Molecular Biology and Biotechnology Centre, University of Alberta, Canada*, *10*, 295–301.

Zhan, L., Paterson, I. G., Fraser, B. A., Watson, B., Bradbury, I. R., Nadukkalam Ravindran, P., & Bentzen, P. (2017). MEGASAT: Automated inference of microsatellite genotypes from sequence data. *Molecular Ecology Resources*, *17*(2), 247–256. https://doi.org/10.1111/1755-0998.12561

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Le Clercq, L. S., Phetla, V., Osinubi, S. T., Kotzé, A., Grobler, J. P., & Dalton, D. L. (2024). Phenotypic correlates between clock genes and phenology among populations of Diederik cuckoo, *Chrysococcyx caprius*. *Ecology and Evolution*, *14*, e70117. https://doi.org/10.1002/ece3.70117

CHAPTER 5

Differential CpG methylation as predictors of age in African Cheetah

This chapter represents the methods, results, and discussion for the second biological clock of the study which focuses on the use of differential methylation in candidate genes as an epigenetic clock for age estimation. The chapter follows the standard article format and was submitted to *Molecular Ecology Resources* (Le Clercq *et al.*, 2024*b*). Details are provided for the methods used to design an experiment to assay orthologous CpG's in genes as potential biomarkers for age for cheetah (*Acinonyx jubatus*) using a candidate gene approach for six genes: *EDARADD*, *ELOVL2*, *FHL2*, *GRIA2*, *ITGA2B*, and *PENK*. Furthermore, the primary results from the experiments—including CpG's assayed, significant correlations detected, and the final constructed model—are reported. Finally, the results are discussed in terms of the degree of a methylation–age correlation for individual genes and CpG's, the performance of the best model for age prediction, and how these findings relate to previously published studies from the review in Chapter 3. Supplementary materials and methods are included in Appendix F, methods (Le Clercq *et al.*, 2023*h*, 2023*i*) were submitted to the online platform *protocols.io* in support of open methods initiatives.

Articles:

LE CLERCQ, L., KOTZÉ, A., GROBLER, J.P. & DALTON, D.L. (2024*b*) Methylation-based markers for the estimation of age in African Cheetah, *Acinonyx jubatus*. *Molecular Ecology Resources* **24**, e13940.

Protocols:

LE CLERCQ, L., DALTON, D.L., KOTZÉ, A. & GROBLER, J.P. (2023*h*) Designing an EpiTYPER bisulfite sequencing assay for age estimation in *Acinonyx jubatus* based on human orthologues. *protocols.io*. DOI: https://dx.doi.org/10.17504/protocols.io.j8nlk4yk1g5r/v1.

LE CLERCQ, L., DALTON, D.L., KOTZÉ, A. & GROBLER, J.P. (2023*i*) DNA extraction protocol for animal blood samples using the E.Z.N.A blood mini kit. *protocols.io*. DOI: https://dx.doi.org/10.17504/protocols.io.ewov141xpvr2/v1.

Received: 20 November 2023 | Revised: 5 February 2024 | Accepted: 9 February 2024

DOI: 10.1111/1755-0998.13940

MOLECULAR ECOLOGY RESOURCES WILEY

RESOURCE ARTICLE

Methylation-based markers for the estimation of age in African cheetah, *Acinonyx jubatus*

Louis-Stéphane Le Clercq[1,2] | Antoinette Kotzé[1,2] | J. Paul Grobler[2] | Desiré L. Dalton[3]

[1]South African National Biodiversity Institute, Pretoria, South Africa

[2]Department of Genetics, University of the Free State, Bloemfontein, South Africa

[3]School of Health and Life Sciences, Teesside University, Middlesbrough, UK

Correspondence
Louis-Stéphane Le Clercq, South African National Biodiversity Institute, Pretoria 0001, South Africa.
Email: leclercq.l.s@gmail.com

Funding information
National Research Foundation, Grant/Award Number: 112062

Handling Editor: Alana Alexander

Abstract

Age is a key demographic in conservation where age classes show differences in important population metrics such as morbidity and mortality. Several traits, including reproductive potential, also show senescence with ageing. Thus, the ability to estimate age of individuals in a population is critical in understanding the current structure as well as their future fitness. Many methods exist to determine age in wildlife, with most using morphological features that show inherent variability with age. These methods require significant expertise and become less accurate in adult age classes, often the most critical groups to model. Molecular methods have been applied to measuring key population attributes, and more recently epigenetic attributes such as methylation have been explored as biomarkers for age. There are, however, several factors such as permits, sample sovereignty, and costs that may preclude the use of extant methods in a conservation context. This study explored the utility of measuring age-related changes in methylation in candidate genes using mass array technology. Novel methods are described for using gene orthologues to identify and assay regions for differential methylation. To illustrate the potential application, African cheetah was used as a case study. Correlation analyses identified six methylation sites with an age relationship, used to develop a model with sufficient predictive power for most conservation contexts. This model was more accurate than previous attempts using PCR and performed similarly to candidate gene studies in other mammal species. Mass array presents an accurate and cost-effective method for age estimation in wildlife of conservation concern.

KEYWORDS

Acinonyx jubatus, African, age, age estimation, biological clocks, biomarker, cheetah, genes, mass array, methylation

1 | INTRODUCTION

Being able to accurately identify the age of an animal is often helpful to wildlife conservation. In understanding population dynamics—including effective population size—the number of individuals that are of young, reproductive, or old age serve as indicators for the present and future fitness of that species (Cooper et al., 2021). Thus, knowledge of these classes are crucial in monitoring populations and measuring the efficacy of current conservation efforts, as most population models and algorithms assume a fixed population distribution of known ages (Durant et al., 2017). In conservation genetics, many methods have been developed to use DNA to study pertinent individual as well as population characteristics, including biological sex (Morin et al., 2005), parentage (Double et al., 1997), population assignment (Kim & Sappington, 2013), hybridization (Brelsford et al., 2011), speciation (Linck et al., 2019), and migration (Merlin & Liedvogel, 2019). As ageing is a biological process under genetic control (Horvath, 2013) or the result of unrepaired damage to DNA (Kujoth et al., 2005), the cellular and genetic hallmarks of ageing can be used to determine age. One genetic method for age estimation is the epigenetic regulatory elements including methylation (Johansson et al., 2013) which has become a well-established method in humans (Gopalan et al., 2017). Recently, these methods have been extended to many species (Le Clercq, Kotzé, et al., 2023), including many of the more than 280 carnivore species (order: Carnivora) such as canids (Horvath et al., 2022; Ito et al., 2017; Thompson et al., 2017), pinnipeds (Robeck et al., 2023; Sareisian, 2014), felines (Cantrell et al., 2020; Qi et al., 2021; Raj et al., 2021), and bears (Nakamura et al., 2023).

The cheetah, *Acinonyx jubatus* [von Schreber, 1775], is a carnivore and the last remaining species in the genus *Acinonyx* (O'Brien et al., 1985). There are currently four recognized subspecies based on geographic location and morphology. These include the most abundant *A. j. jubatus* [von Schreber, 1775] in Southern Africa, *A. j. venaticus* [Griffith, 1821] in Iran, *A. j. hecki* [Hilzheimer, 1913] in Northern Africa, and *A. j. soemmeringii* [Fitzinger, 1855] in Central Africa (Krausman & Morales, 2005). *Felidae* species diverged from other carnivores approximately 100 million years ago (Nyakatura & Bininda-Emonds, 2012). The *Acinonyx* genus diverged from other *Felidae* species approximately 4.02 to 17.30 million years ago, while a divergence time between 32 and 67 thousand years ago is estimated between Asiatic and African cheetahs and African subspecies, *A. j. soemmeringii* and *A. j. jubatus*, diverged in parallel, approximately 16 to 72 thousand years ago (Charruau et al., 2011).

The International Union for Conservation of Nature (IUCN) uses a particular 'Red List' algorithm which considers several key ecological parameters as criteria to categorize individual species for their extinction risk to facilitate the prioritization of conservation efforts to those with the highest risk (Mace et al., 2008). Cheetah, along with many other *Felidae* species including the fishing cat (*Prionailurus viverrinus*), clouded leopard (*Neofelis nebulosa*), snow leopard (*Panthera uncia*), leopard (*Panthera pardus*), lion (*Panthera leo*), and tiger (*Panthera tigris*), are listed as either vulnerable (Durant et al., 2015) or endangered by the IUCN. Serious concerns exist about the potential extinction of these species considering the American lion, *Panthera atrox*, three species of American cheetah, genus *Miracinonyx*, and three species of Sabre-tooth tiger, genus *Smilodon*, went extinct over the past 15 thousand years, while several subspecies of lion (Black et al., 2013), cougar (Cardoza & Langlois, 2002), and tiger (Yamaguchi et al., 2013) were declared extinct between 1960 and 2018.

Cheetah once ranged across large stretches of Asia, India, the Middle East, and Africa; however, the population trend is estimated to be decreasing (Durant et al., 2015). As of 2016, approximately 7100 cheetahs remain in the wild; dispersed into 33 populations with fluctuating effective population sizes of around one hundred (Durant et al., 2015). At present, they only occupy about 6%–9% or less of their historical range, mostly in Eastern and Southern Africa, with an estimated 1326 in South Africa; and the majority (77%) of that range falls outside of protected areas (Durant et al., 2017). The cheetah has a maximum lifespan of up to 15 years in captivity; however, in the wild there is a less than 5% survival rate for cubs and most individuals only live between 6 and 8 years, with high mortality rates around the age of 2 years old when they first mate (Caro, 1994).

In the cheetah, individuals are classified into eight age groups, ranging from young cubs (0–6 months) to very old adults (>144 months) (Table 1). Current methods to age wild cheetah largely consist of physical examination of phenotypical traits such as teeth, fur, and body condition (Marker et al., 2003), which become progressively less accurate with older age groups (summarized in Table 1). The morphological method of age determination in cheetah was developed through observations made from pet cheetah (Adamson, 1970; Burney, 1980) along with published data (Broom, 1949) as well as fieldwork from the Serengeti Cheetah Project (Caro, 1994; Kelly et al., 1998) in Tanzania and Cheetah Conservation Fund (Marker et al., 2003) in Namibia. The accuracy of physical features was cross-validated using the more established method of tooth cementum annulation (TCA) on teeth extracted from dead individuals (Marker & Dickman, 2003). The TCA method uses a biological clock that follows predictable annual changes, the deposition of a dark and light layer each year following the eruption of the tooth, that enable histological age estimation (Naji et al., 2016; Sharma et al., 2021). The accuracy of TCA itself for age estimation in cheetah, however, remains dubious as they are known to suffer from frontal palatine erosion, particularly in captive individuals, due to mandible malocclusion (Fitch & Fagan, 1982) which affects tooth wear patterns and could obscure annuli or lines of annulated cementum (Nakanishi et al., 2009). Furthermore, the utility of morphometric observations for age determination is reliant upon specialized training and considerable experience, while an exploratory or validating study remains to be published in the main scientific literature. This has already been done in, for example, many

TABLE 1 Summary of morphological features used to determine age in cheetah (Marker & Dickman, 2003).

Class	Teeth	Coat	Body
1. Young cubs 0–6 months	Deciduous canines and incisors erupt at 28–30 days; molars erupt at 45–50 days	Spots on legs and yellow hair colouring develop at 6–7 weeks, mantle present at 4 weeks and lost at 3–4 months	Eyes open at 7–10 days, cubs emerge from den at about 6 weeks
2. Large cubs 6–12 months	Lower incisors fall at about 7 months, adult teeth erupt at about 8 months	Long hair on back of neck remains, although it is no longer a defined mantle	Lanky appearance until about 9 months, then body begins to fill out; body mass about two-thirds by 12 months
3. Adolescents 12–18 months	No tartar or yellowing of teeth	Some long fur on back of neck; fur on face and body fuzzy and scruffy rather than smooth	Attain full height but not adult weight, leggy, with dam
4. Independent 18–30 months	No tartar or yellowing of teeth	Some long fur on back of neck, smooth, sleek coat	Develop muscle tone, usually not with dam but may be with littermates
5. Young adults 30–48 months	Slight tartar and yellowing of teeth	Slight mane still	Fully grown but not fully muscled, in prime physical condition. Males have scars, females usually pregnant or with cubs
6. Prime adults 48–96 months	Tartar and yellowing of teeth, slight gum recession, some gingivitis	Mane on back of neck is gone	Fully muscled, prime physical condition but starting to show signs of ageing
7. Old adults 96–144 months	Tartar and yellowing of teeth, gum recession, gingivitis, canines tipped, loss of teeth, especially incisors	Coat beginning to look ragged, poorly groomed, scarred	Pads becoming smooth and elongated, sunken face, thinner, loss of muscle tone
8. Very old adults >144 months	Tartar and yellowing of teeth, gum recession, gingivitis, canines tipped, loss of teeth, especially incisors and canines, broken teeth	Ragged, poorly groomed, scarred coat	Pads quite smooth and elongated, sunken face, body delicate, and frail

other species in the family *Felidae* such as leopards (Stander, 1997), tigers (Fàbregas & Garcés-Narro, 2014), and African lions (White & Belant, 2016). Consequently, this makes the morphological method difficult to apply to cheetahs across their full range with serious questions remaining regarding the reliability of the traits used to assess age.

While the utility of methylation for age determination has been previously illustrated in the cheetah (Raj et al., 2021), there are difficulties with applying the Horvath Mammalian Array to cheetah, due to low-throughput sample processing that increases the per-sample cost when applied in a conservation context. Although the Clock Foundation does facilitate research collaborations that could reduce the cost, this requires submitting samples internationally, due to CITES regulations for species of conservation concern and potential concerns over sample/data sovereignty (Knight et al., 2022; Zainol et al., 2011) mean this is not currently feasible. Therefore, the aim of the present study is to investigate the potential of age-related changes in methylation, using a candidate gene approach for genes under epigenetic clock control assayed in previous studies, to be used to establish an accurate age estimation model using *Acinonyx jubatus* as a case study. Beyond cheetahs, the process we outline here for developing a candidate gene approach should be broadly useful to other species, where resource limitation and/or difficulty in internationally shipping samples precludes the use of existing array technology.

2 | METHODS

2.1 | Samples and ethics

Approval for the present study was obtained from the protocol committee of the Department of Genetics, University of the Free State (approval number: Res18/2020). Ethics approvals were obtained from the University of the Free State (approval number: UFS-AED2020/0015/1709) as well as the South African National Biodiversity Institute (approval number: SANBI/RES/P2020/30). Appropriate research permits were also obtained from South African regulatory authorities including the Department of Agriculture, Land Reform, and Rural Development (Section 20 permit: 12/11/1/1/18(1824JD)) and the Department of Environmental Affairs (Threatened Or Protected Species (TOPS) permit: O-52903). An appropriate minimum sample size required was calculated using G*Power 3.1 (Faul et al., 2009) given an anticipated effect size large enough to provide a correlation coefficient of 90% and power of at least 0.8; close to previous approximations with similar data (Mayne et al., 2021). Fifty cheetah samples (25 male, 25 female) were subsequently received from the SANBI Biobank, where they were stored at −80°C. Blood samples were previously collected during routine screening and health checks of captive cheetah from several localities in South Africa (Figure 1). Venous blood was collected from either the leg or tail

in an EDTA collection tube using an 18-gage needle. The samples were selected in such a way that each year of the expected life span of 15 years old was represented. Samples included young cubs ($N = 12$), adolescents ($N = 2$), independents ($N = 8$), young adults ($N = 7$), prime adults ($N = 6$), old adults ($N = 12$), and very old adults ($N = 3$). This project was registered in the National Centre for Biotechnology Information (NCBI) BioProject database as PRJNA737185 and sample details along with NCBI BioSample numbers are listed in Table S1. Mapping was done in QGIS 3.22.6 based on distribution maps for terrestrial mammals from the IUCN (International Union for Conservation of Nature, 2022) used in the assessment of cheetah (Durant et al., 2015).

2.2 | DNA extraction and quantitation

DNA was extracted from 100 μL of whole blood with the commercially available E.Z.N.A.® blood DNA mini kit (Omega Bio-Tek Inc., Norcross, Georgia, USA), according to manufacturer's instructions (Le Clercq, Dalton, et al., 2023b). This kit makes use of the basic digest-bind-wash-elute spin-column technology where the blood sample was digested with a protease and applied to a column with a binding buffer. These columns were centrifuged to remove debris using a wash buffer. The final step was eluting the bound and purified DNA in 50 μL elution buffer. The DNA concentration and purity were determined by spectrophotometric measurement of 1 μL extract applied to the Nanodrop™ 1000 (Thermo Fisher Scientific, Waltham, Massachusetts, USA) instrument. This is achieved by measuring the absorbance at wavelengths A_{260} and A_{260}/A_{280} to report concentration in ng/μl as well as indicate the purity. Additional screening was done by determining concentration and purity using the Qubit® 4 fluorometer (Invitrogen, Thermo Fisher Scientific, Waltham, Massachusetts, USA) with the Qubit® dsDNA Broad Range Assay Kits.

2.3 | Study design and target genes

Methylation status was assayed using methods adapted from human studies (Le Clercq, Dalton, et al., 2023a) for CpG sites within six promising gene regions (Table 2). Genes were selected based on several criteria, including (1) multiple studies (>2) providing evidentiary support of an epigenetic clock, (2) high correlation with age (>70%) in previous studies, (3) presence of multiple, potentially assayable, CpG's within the target region, and (4) availability of coordinate data in the form of Illumina 'cg' numbers (e.g. cg09809672) to facilitate sequence retrieval. Cheetah orthologues were retrieved using human sequences from the University of California Santa Cruz (UCSC) Genome Browser (Haeussler et al., 2019) in basic local alignment search tool (BLAST) searches (Agarwala et al., 2018) against the cheetah reference genome (NCBI RefSeq: GCF027475565.1). The final set of genes were: *EDARADD* (NCBI Gene ID: 106982554), *ELOVL2* (NCBI Gene ID: 106972032), *FHL2* (NCBI Gene ID: 106977955), *GRIA2* (NCBI Gene ID: 106986586), *ITGA2B* (NCBI Gene ID: 106974321), and *PENK* (NCBI Gene ID: 106982523). The degree to which the target regions are conserved among species

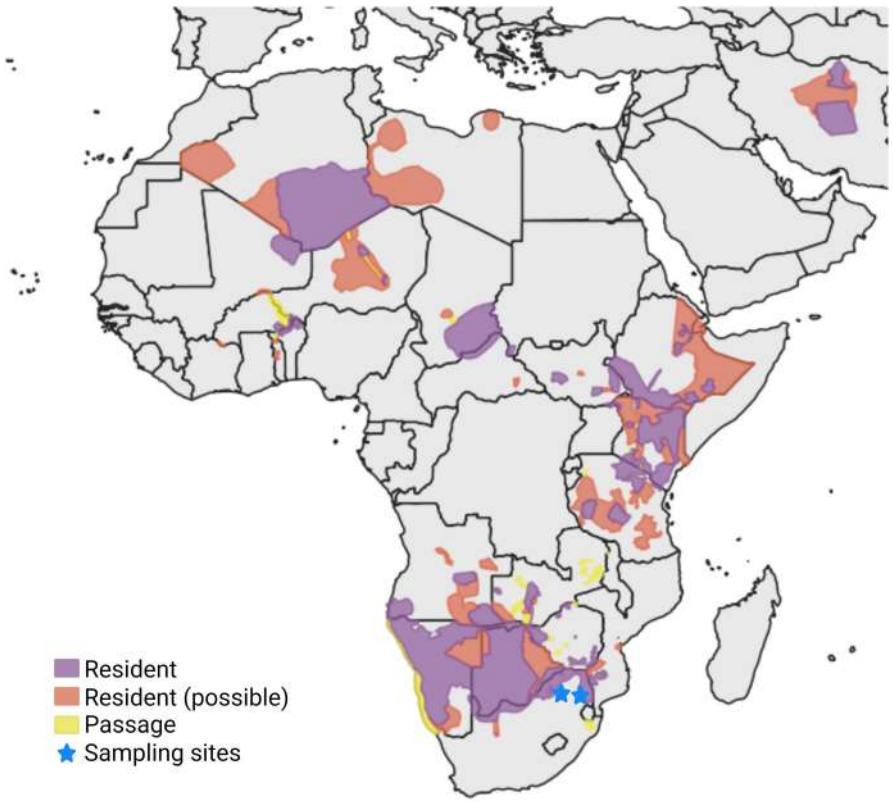


FIGURE 1 Map of distribution for *Acinonyx jubatus* species. Areas indicated in purple represent known resident ranges of cheetah, while pink areas indicate possible additional resident ranges. Areas indicated in yellow represent vagrant ranges where cheetah may occur sporadically as part of passage between established ranges. Sampling sites are indicated by blue stars. Image created in BioRender.com.

TABLE 2 Results from individual linear regressions for CpGs screened in six cheetah genes found to have a near-significant or significant correlation.

Gene	Evidence	CpG	Correlation (R^2)	*p*-Value
EDARADD	(Bocklandt et al., 2011)	**CpG 3**	**.16**	**.007*****
CpGs: 10		CpG 6	.02	.18
		CpG 8	.01	.28
ELOVL2	(Bekaert et al., 2015)	CpG 17	.06	.12
CpGs: 39		**CpG 18/19/20**	**.13**	**.03****
		CpG 21	.01	.26
FHL2	(Giuliani et al., 2016)	CpG 1	.02	.16
CpGs: 16		**CpG 9**	**.12**	**.08***
		CpG 14	.01	.24
GRIA2	(Koch & Wagner, 2011)	CpG 2	.02	.39
CpGs: 6		CpG 3	.02	.38
		CpG 4	**.09**	**.04****
ITGA2B	(Weidner et al., 2014)	CpG 4	.03	.15
CpGs: 6				
PENK	(Giuliani et al., 2016)	**CpG 2**	**.12**	**.05****
CpGs: 23		**CpG 3/4**	**.13**	**.05****
		CpG 6	.05	.11
		CpG 11/12	.02	.22
		CpG 16/17/18	.01	.24
		CpG 19/20	.02	.22

Note: Gene abbreviations are indicated along with the relevant reference study that reported on the use of these genes for age determination. Individual CpG's are listed according to their numbering on the EpiGrams generated with EpiTYPER® with significant CpG's indicated in bold.
*$p < .10$; **$p \leq .05$; ***$p < .01$.

within the family *Felidae* was determined by comparing amplicon sequences for *A. jubatus* to reference genomes using BLAST. Primers were designed using the EpiDesigner website (Agena Bioscience Inc., 2017) and are listed in Table S2. Success of the designed assay was determined in R 4.0.2 (R Core Team, 2020) with the RSeqMeth 1.0.2 (Statham & Csárdi, 2008) and MassArray 1.46.0 (Thompson & Greally, 2020) packages that analyses the predicted amplicon for fragmentation pattern in terms of size and resolution. This gave a direct indication of the number of assayable CpG's per gene. This platform was registered to the NCBI Genome Expression Omnibus (GEO) database as GPL32927.

2.4 | Bisulphite conversion and EpiTYPER® MASS array sequencing

Genomic DNA from all samples was treated with EpiMark® Bisulfite Conversion Kit (New England Biolabs, Ipswich, Massachusetts, USA). Briefly, 10 μL of DNA was mixed with prepared bisulphite mix (sodium metabisulphite) followed by brief vortexing and thermal cycled using alternating denaturation (95°C, 5 min) and incubation (65°C, 30–60–90 s) phases for three cycles. This treatment converted any non-methylated cytosine residues into uracil, while methylated cytosine residues remained unaffected. Samples were purified with spin columns, as for extractions, using desulphonation and wash buffers. Finally, the converted DNA was eluted and quantitated. Complete bisulphite conversion of DNA was monitored with 'TpC' internal conversion controls (Suchiman et al., 2015) that have been identified within each gene amplicon. At a 'TpC', the 'C' is unmethylated and is expected to produce 100% 'TpA' sequence reads in the final data for those sites when conversion was complete.

The EpiTYPER® Assay methylation analysis was run in three phases, starting with (1) PCR using T7-promoter tagged reverse primers to amplify the target regions while preserving the bisulphite-induced sequence changes and treatment of PCR product with Shrimp Alkaline Phosphatase, (2) in vitro transcription was performed, and (3) the resulting RNA transcripts were specifically cleaved at uracil residues. The resulting fragments differed in size and mass, depending on the sequence changes generated through bisulphite treatment, and through enzymatic cleavage of uracils. The resulting fragments were loaded onto a SpectroCHIP® Array (Chip) that was placed in the MASSarray (MALDI-TOF) mass spectrometer for data acquisition. Results showed individual peaks for the generated fragments. Fragmentation pattern was analysed for mass for each fragment (based on size) and the peak height, which corresponds to the percentage of methylation compared to an internal control. The output was automatically loaded into a database for data analysis with EpiTYPER® software.

2.5 | Statistical analysis

The percentage of methylation per gene site was compared to known age of each sample. Methylation data were first analysed for normality and 'goodness of fit' to a Gaussian distribution using the Anderson-Darling test for normality to eliminate possible errors due to skewedness or kurtosis. Hereafter, a standard linear regression was performed individually for each gene and CpG to identify those with the strongest correlation in RStudio 1.4.1106 (Rstudio Team, 2021), running R version 4.0.5 (R Core Team, 2020). Subset analyses to identify the best targets for regression were explored with *leaps* version 3.1 (Lumley, 2020). Those sites with a significant correlation coefficient and probability (at least $p < .10$), explaining the larger part of the Sum of Standard Error (SSE), were used to construct and test multiple-regression models.

2.6 | Model training and validation

Models were trained using machine learning approaches implemented through elastic net regression with the *glmnet* version 4.1–7 (Friedman et al., 2010) and *caret* version 6.0–94 (Kuhn, 2008) packages in R. Models were optimized based on the alpha (α) and lambda (λ) parameters. Furthermore, model performance was accessed based on the values obtained for the Root Mean Standard Error (RMSE) of the model, as per Equation 1, as well as the correlation coefficient (R^2) for the model. Here, the RMSE is derived from the square root of the sum for the quotient of the squared difference between the sample mean ($\hat{y}_i$) and the inferred value (y_i) and the total sample size (n).

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{n} \frac{\left(\hat{y}_i - y_i\right)^2}{n}} \quad (1)$$

These models were validated using Leave-Group-Out Cross-Validation (LGOCV), also known as Monte Carlo Cross-Validation (MCCV), best suited to hierarchical data such as age groups and smaller sample sizes (Shan, 2022; Shao, 1993). LGOCV has the added benefit of avoiding over-fitting while being able to subsample test sets of variable size (between 50% and 75% of the data). This method was applied using fivefold repeats ($k = 5$) with subsample sets of 39 samples (75%) from the test data ($N = 50$). The accuracy of the model predictions was expressed as the Mean Absolute Error (MAE) in months, according to Equation 2. Here, the MAE is derived from the sum of the absolute difference between the inferred (y_1) and actual (x_1) values, divided by the sample size (n).

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} | y_1 - x_1 |}{n} \quad (2)$$

The final model predictions were assessed for both validity and reliability (Cohen et al., 2017; Hammersley, 1987) on a subset of 12 samples, in lieu of samples of unknown ages, that had an equal distribution of males and females and included at least one sample per cheetah age group. Validity, in this context, refers to the ability of the metric to accurately measure the intended attribute. This was verified by comparing the known chronological age of individuals to their predicted age from the model. Reliability of the measure was assessed by comparing the age class assignment correctness for the predicted ages. Although the ultimate purpose of designing an age estimation model such as that presented in this paper is to apply the panel to wild animals of unknown age, due to travel restrictions related to reducing human-to-human spread of COVID and concern about potential transmission of COVID to cheetahs (Heeney et al., 1990; Kennedy et al., 2002; Tewari et al., 2023), the application to wild animals was beyond the scope of this study at the time the study was carried out.

3 | RESULTS

3.1 | CpG screening and selection

Orthologous gene regions were identified for cheetah in all six of the selected target genes. Initial estimates based on sequence revealed a total of 100 CpG's (Table 2) within the target regions. Fragmentation patterns for the selected amplicons in R showed that once fragmentation is considered a total of 77 CpG clusters that could potentially be screened using EpiTYPER® mass assay analyses. EpiTYPER® analyses (see Figure S1) were run on 50 samples. Good amplification was observed; however, eight of the screened CpG fragments fell outside the emission spectra and were not quantifiable: two for *EDARADD*, four for *ELOVL2*, one for *FHL2*, and one for *ITGA2B*. Linear models constructed for the relationship between age and differential methylation detected some correspondence in 19 of the remaining CpG's, of which six showed a statistically significant relationship (Table 2, Figure 2). These CpG's were present in five of the six candidate genes, excluding *ITGA2B*, with four CpG's showing hypomethylation with age and two showing hypermethylation. For *EDARADD*, CpG 3 showed progressive hypomethylation with methylation levels ranging from $\beta = .01$ to $\beta = .29$ and a significant ($F = 8.11$, $R^2 = .16$, $p < .01$) relationship with age. *FHL2* showed similar hypomethylation for CpG 9, with values ranging between $\beta = .01$ and $\beta = .28$ and a significant ($F = 3.28$, $N = 25$, $R^2 = .12$, $p < .10$) age correspondence. *PENK* was the only gene with several significant CpG's, both CpG 2 ($F = 4.1$, $R^2 = .12$, $p \leq .05$) and the cluster for CpG's 3 and 4 ($F = 4.32$, $R^2 = .13$, $p \leq .05$), with beta values of 0.21–0.56 and 0.50–0.84, respectively. *ELOVL2* (CpG cluster of 18, 19, and 20) and *GRIA2* (CpG 4) both showed hypermethylation with age. For *ELOVL2*, methylation levels ranged from $\beta = .01$ to $\beta = .14$ and reached statistical significance ($F = 3.99$, $R^2 = .14$, $p < .05$). Similar to *EDARADD* and *FHL2*, *GRIA2* showed methylation levels ranging between $\beta = .04$ and $\beta = .29$ but was significantly hypermethylated ($F = 4.35$, $R^2 = .09$, $p < .05$) with age. Average methylation levels for all CpG's per gene were not significant with only *PENK* showing a

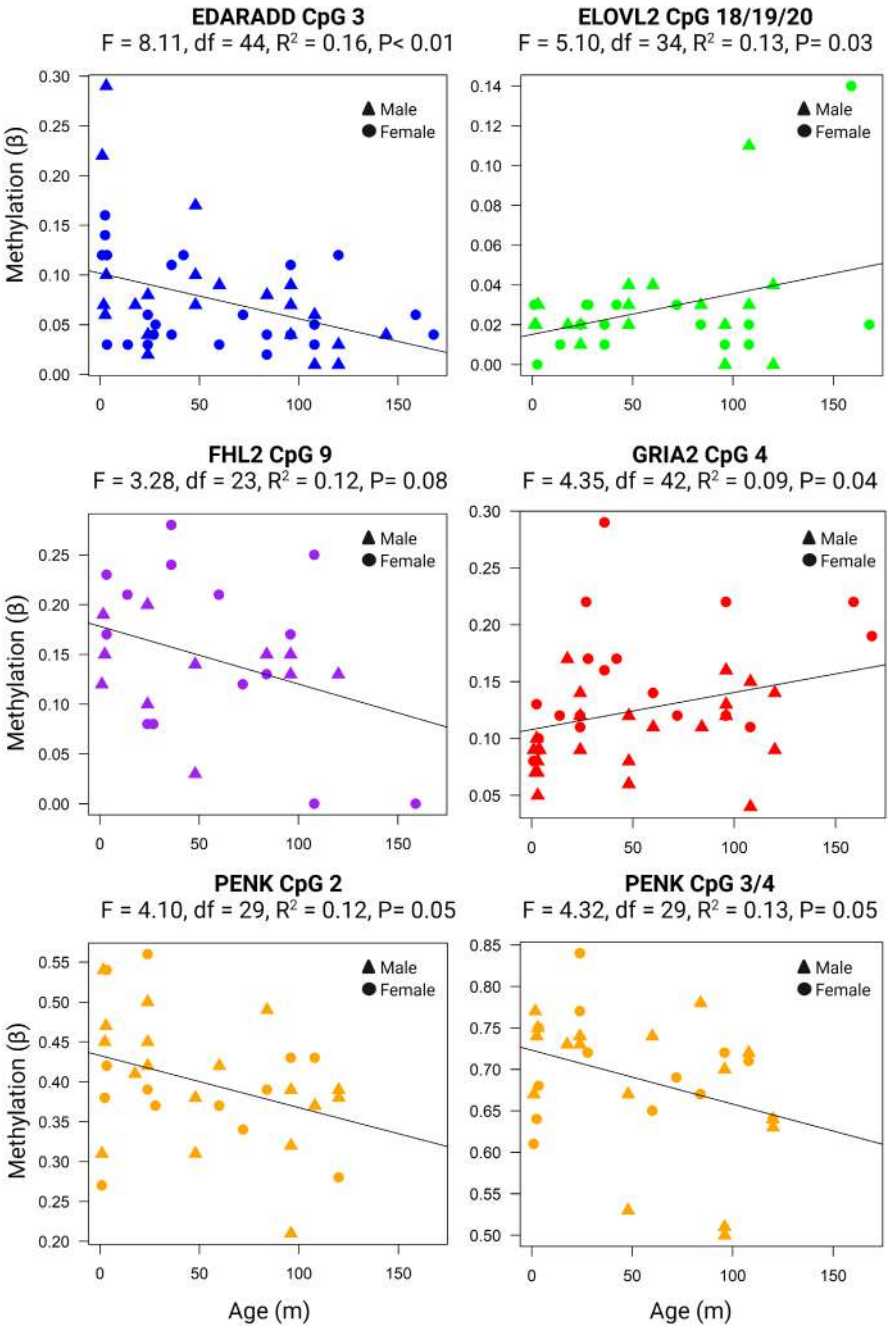


FIGURE 2 Scatter plots for the six CpGs identified as having a significant relationship with age. The y-axis indicates the relative level of methylation as expressed by the β-values, while the *x*-axis indicates the known ages expressed as months. Four CpG's, in the EDARADD (blue), FHL2 (purple), and PENK (orange) genes, showed a linear trend on hypomethylation with age and another two CpG's, in the ELOVL2 (green) and GRIA2 (red) genes, showed a linear trend of hypermethylation with age. Males are indicated by triangles, while females are indicated by circles. No statistically significant trend for methylation was detected for sex. Image created in BioRender.com.

near-significant correlation ($F = 2.64$, $R^2 = .08$, $p = .11$). A CpG subset analysis prior to model building identified four CpG's in *FHL2*, *GRIA2*, and *PENK*, as the most promising targets.

3.2 | Creating an age estimation model

The six significant CpG's were used to train and test several elastic net regression models using between two and six CpG's (Table 3). Models generated with fewer CpG's tended to have higher RMSE and MAE values and explained a lower percentage of the variance based on the correlation coefficient. For example, both the model based on two CpG's and the model based on three CpG's had an RMSE in the forties and explained less than 50% of the variance. The two best models were based on six CpG's (RMSE = 31.30, MAE = 25, $R^2 = .70$) and four CpG's (RMSE = 30.29, MAE = 25, $R^2 = .59$). While both models performed similarly based on RMSE and MAE, the model based on six CpG's performed better than the four CpGs by explaining a higher percentage of the variance as assessed by the correlation coefficient.

3.3 | Characteristics of the age estimation model

The best model was based on six CpG's and had an alpha value of .05 and lambda of 0.01. Subsets sampled from the dataset used to

TABLE 3 Results from machine learning experiments to create a model for age prediction in cheetah.

Model type	RMSE	MAE	R^2
6 CpG's ($\alpha = .05$, $\lambda = 0.01$)	31.30	25	.70
EDARADD CpG 3 + ELOVL2 CpG 18/19/20 + FHL2 CpG 9 + GRIA2 CpG 4 + PENK CpG 2 + PENK CpG 3/4			
5 CpG's ($\alpha = .97$, $\lambda = 5.17$)	35.31	29	.54
EDARADD CpG 3 + ELOVL2 CpG 18/19/20 + GRIA2 CpG 4 + PENK CpG 2 + PENK CpG 3/4			
4 CpG's ($\alpha = .05$, $\lambda = 0.001$)	30.29	25	.59
EDARADD CpG 3 + FHL2 CpG 9 + GRIA2 CpG 4 + PENK CpG 3/4			
3 CpG's ($\alpha = .07$, $\lambda = 5.90$)	40.09	35	.36
EDARADD CpG 3 + ELOVL2 CpG 18/19/20 + GRIA2 CpG 4			
2 CpG's ($\alpha = .73$, $\lambda = 6.28$)	48.56	43	.14
EDARADD CpG 3 + ELOVL2 CpG 18/19/20			

Note: Five models were tested using between two and six CpG's, respectively. In each step, the CpG with the lowest significance was omitted with the exception of the four CpG models which was based on the best targets from the subset analyses. Models were compared in terms of Root Mean Standard Error (RMSE), Mean Absolute Error (MAE), and correlation (R^2). The model with the lowest RMSE and MAE and highest correlation was based on six CpG's and was selected for validation.

train the model for validation testing using LGOCV performed well (Figure 3, $R^2 = .70$), with most samples having a near-perfect correlation between predicted and chronological ages with an accuracy of approximately 25 months or 2 years and 1 month of their actual age. When comparing predicted age versus chronological age, using only 12 samples, a high degree of correlation was observed ($R^2 = .65$) with most samples falling within the 95% confidence interval. There was, however, some evidence that the accuracy of the model is dependent on the age class as estimates for young individuals and very old adults showed the greatest deviance from their chronological age. Using predicted age for age class assignments, more than 80% of samples were assigned to the correct age class based on the predicted age from the model, showing a reasonably high level of correspondence.

3.4 | Degree of gene conservation within the family *Felidae*

Species within the family *Felidae* diverged from approximately 12 million years ago (MYA). Using gene sequences from *Acinonyx jubatus*, BLAST matches were retrieved for 14 additional species from seven genera within the family *Felidae* including the most closely related *Puma*, *Lynx*, and *Felis* species as well as the more distant *Panthera* species (Figure 4). BLAST matches ranged from 75% to 97% for gene similarity depending upon the gene and species with *EDARADD* and *GRIA2* returning the highest number of matches. As expected, species most closely related to cheetah, including the *Puma*, *Lynx*, *Felis*, and *Panthera* genera, showed the highest degree of conservation. Most species within this family also have similar lifespans of between 10 and 20 years in the wild and captivity, respectively (Le Clercq, Kotzé, et al., 2023), and several species from the matches have previously been used as model species in age and methylation studies. This includes studies in the bobcat (Cantrell et al., 2020; Lachance et al., 2015), *Lynx rufus*, domestic cat (Qi et al., 2021), *Felis catus*, lion (Raj et al., 2021), snow leopard (Qi et al., 2021), and tiger (Raj et al., 2021).

4 | DISCUSSION

This study investigated the potential use of age-related changes in methylation within candidate genes, previously identified in human studies, to establish an age estimation model in wildlife using mass array technology. For this purpose, orthologous gene regions from numerous human studies were retrieved for six target genes with a hundred CpG's within the gene regions of interest. Mass array technology was chosen as a sensitive but cost-effective alternative to existing methods, to assay differential methylation. The predicted fragmentation patterns resulted in a total of 77 CpG clusters that could be analysed for the study species. Screening, performed on 50 samples, provided data for 69 of the fragments: the remainder falling outside the weight spectra for the analyses. Correlation analyses between CpG methylation and known chronological age identified six CpG's with a significant relationship of which four showed progressive hypomethylation with age while two showed hypermethylation. Regression models were fitted using different combinations of CpG's and the optimal model, based on performance metrics and validation experiments, indicated that age models using either four or six CpG's were most accurate. The model based on six CpG's did, however, outperform the four CpG models in terms of correlation and predictive power.

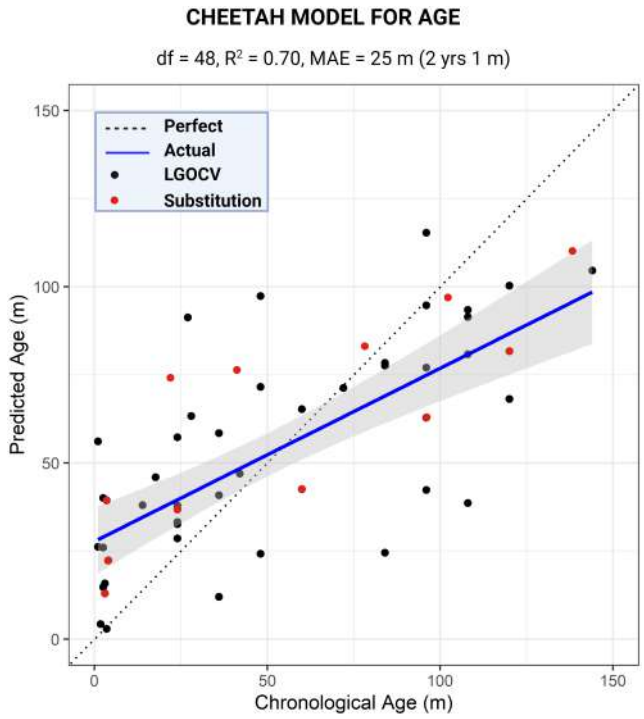


FIGURE 3 Scatter plot for the validation testing of the cheetah model for predicting age based on methylation at six CpG's. The model was validated using Leave-Group-Out Cross-Validation (LGOCV) on a test set of randomly selected samples from the original set used to train the model (black) as well as through the resubstitution of 12 samples (red). The age, as predicted by the methylation model, is plotted against the known chronological age of the individual in months. The dotted line indicated a theoretically perfect correspondence. The blue line shows the regression for predicted versus chronological age and the 95% confidence interval (grey). The model showed a high degree of accuracy (MAE = 25 months or 2 years and 1 month, $R^2 = .70$, $p < .05$) with many samples having a linear correlation between their predicted and actual ages and >55% of samples falling within or close to the 95% confidence interval. Image created in BioRender.com.

Previous studies aimed at creating age estimation models based on differential CpG methylation in candidate genes have used as little as one to four (Nakamura et al., 2023; Weidner et al., 2014) CpG's to subsets of seven (Freire-Aradas et al., 2016; Polanowski et al., 2014; Wright et al., 2018) or eight (Fleckhaus & Schneider, 2020). While most of these studies also used similar methods, such as pyromark bisulphite sequencing, methods such as methylation-sensitive PCR (Mawlood et al., 2016) and mass array technology (Freire-Aradas et al., 2016; Giuliani et al., 2016) have also proven useful. Of the six genes assayed in this study, six CpG's in five genes showed a correlation with age for cheetah. For most genes, apart from several CpG's in PENK, methylation levels were below 50% and differed between age classes by less than 10%. This was consistent with previous findings for *EDARADD* (Bekaert et al., 2015; De Paoli-Iseppi et al., 2017; Fleckhaus & Schneider, 2020; Pan et al., 2020), *ELOVL2* (Giuliani et al., 2016; Márquez-Ruiz et al., 2020; Spiers et al., 2016), *FHL2* (Dias et al., 2020; Fleckhaus & Schneider, 2020; Giuliani et al., 2016), *GRIA2* (Koch & Wagner, 2011; Mawlood et al., 2016; Tanabe et al., 2020), and *PENK* (Garagnani et al., 2012; Giuliani et al., 2016). Four CpG's showed hypomethylation which is characteristic of global hypomethylation associated with age (Le Clercq, Kotzé, et al., 2023) or epigenetic drift (Zampieri et al., 2015); however, the strong correlation with age may indicate that methylations at these sites are not random. Furthermore, two CpG's showed evidence of hypermethylation. As such, these CpG's likely represent highly regulated and site-directed epigenetic clocks (Horvath & Raj, 2018).

The best model for cheetah age was generated using six CpG's and was able to accurately predict age in validation experiments, with a correlation of 0.70 and mean absolute error of 2 years and 1 month, accounting for a large percent of the observed variance. A systematic review and meta-analysis comparing methylation-based age estimation models across several vertebrate classes found that the average effect for included studies was approximately 0.92 with a confidence interval ranging between 0.89 and 0.94 (Le Clercq, Kotzé, et al., 2023). While the model of this study falls below this range, it is comparable to models from several previous studies including 0.74 for beluga whales (Bors et al., 2021), chimpanzees (Ito et al., 2018), grey wolves (Thompson et al., 2017), and sheep (Sugrue et al., 2021); and between 0.76 and 0.79 for baboons (Anderson et al., 2021), bottlenose dolphins (Beal et al., 2019), domestic dog (Thompson et al., 2017), and humpback whales (Polanowski et al., 2014). In comparison to previous model attempts in cheetah, mass array proved more useful than methylation-sensitive PCR (Le Clercq et al., 2018) but was slightly less accurate compared to cheetah samples tested using a feline-specific model from probe-based array bisulphite sequencing (Raj et al., 2021). Furthermore, while the error of the estimate was approximately 2 years, the functional age classes of cheetah typically span more than 2 years of age for adults (Marker & Dickman, 2003), meaning the model could still classify individuals to a close enough approximation of chronological age.

In comparison to previous studies for *Felidae* species, mass array provided more informative data than studies using global methylation detection methods such as ELISA as applied to domestic cats (Tamazian et al., 2014), snow leopards (Jabbari et al., 1997), and bobcats (Cantrell et al., 2020). The study model also performed better than previous models for the domestic cat ($R^2 = .41$, MAE = 2 years and 11 months) and snow leopard ($R^2 = .63$, MAE = 2 years and 10 months), which used methylation-sensitive PCR and included 13 CpG's in the *ELOVL2* gene (Qi et al., 2021). It was, however, less accurate than models tested in lions and tigers (Raj et al., 2021). BLAST searches revealed that many of the genes used to design the assay for cheetah are conserved in other *Felidae*, and possibly carnivore, species with similar lifespans. As such future research could extend the method described here to other species either directly, for those with a near-perfect sequence match, or through only minor modifications to the primers.

It should be noted that there are several limitations that were identified. First, the model predictions were most accurate in adult age classes and performed poorly on young individuals such

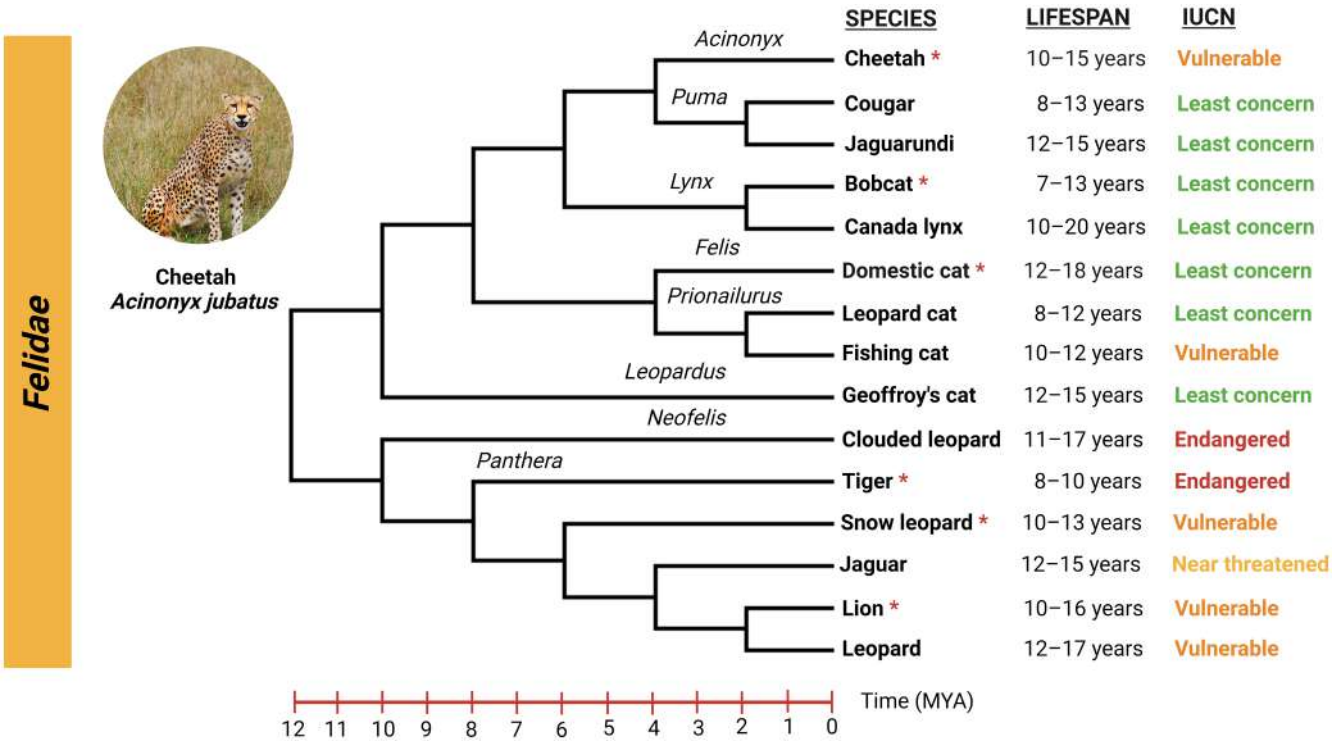


FIGURE 4 Time tree for *Felidae* species for which BLAST matches were identified from study genes. Species within this family diverged over the past 12 million years. Matches were retrieved for 14 additional species from 7 genera within the family *Felidae* including 5 species used in previous methylation studies (red asterisks). BLAST matches ranged between 75% and 97% gene conservation depending upon the gene, with the *Puma*, *Lynx*, and *Felis* genera showing the highest degree of gene similarity compared to *Acinonyx*. Most species within this family also have similar lifespans of between 10 and 20 years in the wild and captivity, respectively. In terms of species vulnerability as assessed by the IUCN, six species are of least concern while one species is classified as near threatened, five species are classified as vulnerable, and two species are classified as endangered. Image created in BioRender.com.

as cubs and adolescents. This does not decrease the overall utility of the method, considering other age estimation methods are highly accurate in the younger classes and lose resolution in adult classes (Marker & Dickman, 2003) but should be considered when using methylation-based models. More specifically, future studies should consider if the inclusion of equal numbers for each functional age class is desirable or if a focus should be placed on adult age classes. Second, a large amount of variance in methylation levels was observed between individuals of the same age, which increased the mean absolute error of the predictions. Variance can often be reduced through either including larger sample sizes for a given measure or using a method with greater sensitivity (James, 1985). While sample size was considered in the study design and based on previously recommended sample sizes (Mayne et al., 2021), some samples did fail for difficult-to-amplify targets such as *FHL2* and *PENK*, resulting in fewer data being available for these regions than is ideal with many of the available data being for younger individuals such as cubs that were less informative. Lastly, the model presented here is population-specific and was calibrated from samples of individuals in captivity and, considering the effects of stress (Poganik et al., 2022), environment (Viñuela et al., 2018), and damage (Kujoth et al., 2005) on epigenetics, may not be suitable to age determination in wild cheetah without further testing.

In conclusion, the present study was able to use age-correlated CpG's from human studies to identify orthologous regions in an unrelated mammal for which the correlations were conserved between species. This enabled the development of an age estimation model that performed similarly to previous studies using a candidate gene approach and was able to assign adult individuals to their relevant functional age classes. Therefore, methylation measured by mass array is a useful molecular tool for the development of age estimation models and could provide useful age information for individuals and populations. Furthermore, the flexibility of the assay design procedure described here could assist in the development of methylation-based age determination methods in most vertebrates. Thus, mass array is a cost-effective alternative to whole-genome bisulphite sequencing approaches and may find widespread applications in conservation—particularly for threatened, vulnerable, and endangered species subject to strict local and global regulations for sampling.

ACKNOWLEDGEMENTS

Images were created in BioRender.com. We would like to thank Professor David Burney for providing a scanned copy of his M.Sc. dissertation as the original was lost during student unrest and political activity on the University of Nairobi campus. We would also like to thank our Post Graduate Research Librarian (UFS), Carmen Nel,

for assisting in the retrieval of several difficult-to-find resources. Parts of this research were presented at the Student Symposium of the South African Society for Genetics. This work is based on the research supported wholly/in part by the National Research Foundation of South Africa (Grant Number: 112062).

CONFLICT OF INTEREST STATEMENT

The authors have no competing interests to declare.

DATA AVAILABILITY STATEMENT

Sample details were deposited in the National Centre for Biotechnology Information (NCBI) BioSample database (SAMN31850912-60), while CpG methylation data generated in this study were deposited to the NBCI Genome Expression Omnibus (GEO) database as dataset GSE252541, associated with GEO platform GPL32927. Accession numbers linked to NCBI BioProject PRJNA737185 are listed in Table S1. The method used to design the assay is available on Protocols.io (Le Clercq, Dalton, et al., 2023a). Additional data including R code used for data analyses are available from the Zenodo depository at the following link: https://doi.org/10.5281/zenodo.10159400. The PYTHON scripted algorithm based on the study model, useful to calculate the age estimates of unknown specimens, is under development and will be available on GitHub at the following link: https://github.com/LSLeClercq/ChEATAE.

BENEFIT-SHARING STATEMENT

Benefits from this research accrue from the sharing of our data and results on public databases as described above. This is particularly important considering that cheetah, used as a case study, are classified as vulnerable and access to sufficient samples may be restricted to specific parts of Africa. As we are aware of historical inequality and underrepresentation of specific demographics within the scientific community and publication, efforts were made to ensure a diverse and balanced author team. This paper includes an LGBT scientist, as well as two authors who are women. The authors are from four institutions and two countries including the UK and South Africa.

ORCID

Louis-Stéphane Le Clercq https://orcid.org/0000-0002-8713-8920
Antoinette Kotzé https://orcid.org/0000-0003-2367-1483
J. Paul Grobler https://orcid.org/0000-0002-5913-7031
Desiré L. Dalton https://orcid.org/0000-0001-5975-6425

REFERENCES

Adamson, J. (1970). *Pippa, the cheetah, and her cubs*. Harcourt.

Agarwala, R., Barrett, T., Beck, J., Benson, D. A., Bollin, C., Bolton, E., Bourexis, D., Brister, J. R., Bryant, S. H., Canese, K., Cavanaugh, M., Charowhas, C., Clark, K., Dondoshansky, I., Feolo, M., Fitzpatrick, L., Funk, K., Geer, L. Y., Gorelenkov, V., … Zbicz, K. (2018). Database resources of the National Center for biotechnology information. *Nucleic Acids Research*, *46*(D1), D8–D13. https://doi.org/10.1093/nar/gkx1095

Agena Bioscience Inc. (2017). *EpiDesigner*. http://www.epidesigner.com/

Anderson, J. A., Johnston, R. A., Lea, A. J., Campos, F. A., Voyles, T. N., Akinyi, M. Y., Alberts, S. C., Archie, E. A., & Tung, J. (2021). High social status males experience accelerated epigenetic aging in wild baboons. *eLife*, *10*, e66128. https://doi.org/10.7554/ELIFE.66128

Beal, A. P., Kiszka, J. J., Wells, R. S., & Eirin-Lopez, J. M. (2019). The bottlenose dolphin epigenetic aging tool (BEAT): A molecular age estimation tool for small cetaceans. *Frontiers in Marine Science*, *6*, 561. https://doi.org/10.3389/fmars.2019.00561

Bekaert, B., Kamalandua, A., Zapico, S. C., Van De Voorde, W., & Decorte, R. (2015). Improved age determination of blood and teeth samples using a selected set of DNA methylation markers. *Epigenetics*, *10*(10), 922–930. https://doi.org/10.1080/15592294.2015.1080413

Black, S. A., Fellous, A., Yamaguchi, N., & Roberts, D. L. (2013). Examining the extinction of the barbary lion and its implications for felid conservation. *PLoS One*, *8*(4), e60174. https://doi.org/10.1371/journal.pone.0060174

Bocklandt, S., Lin, W., Sehl, M. E., Sánchez, F. J., Sinsheimer, J. S., Horvath, S., & Vilain, E. (2011). Epigenetic predictor of age. *PLoS One*, *6*(6), e14821. https://doi.org/10.1371/JOURNAL.PONE.0014821

Bors, E. K., Baker, C. S., Wade, P. R., O'Neill, K. B., Shelden, K. E. W., Thompson, M. J., Fei, Z., Jarman, S., & Horvath, S. (2021). An epigenetic clock to estimate the age of living beluga whales. *Evolutionary Applications*, *14*(5), 1263–1273. https://doi.org/10.1111/eva.13195

Brelsford, A., Milá, B., & Irwin, D. E. (2011). Hybrid origin of Audubon's warbler. *Molecular Ecology*, *20*(11), 2380–2389. https://doi.org/10.1111/j.1365-294X.2011.05055.x

Broom, R. (1949). Notes on the milk dentition of the lion, leopard and cheetah. *Annals of the Transvaal Museum*, *21*(1), 183–185.

Burney, D. A. (1980). *The effects of human activities on cheetah (Acinonyx jubatus) in the Mara region of Kenya*. Masters dissertation. University of Nairobi, Kenya.

Cantrell, B., Friedman, S., Lachance, H., Bernier, C., Murdoch, B., Frattini, S., Talenti, A., Crepaldi, P., & McKay, S. (2020). A novel understanding of global DNA methylation in bobcat (*Lynx rufus*). *Genome*, *63*, 125–130.

Cardoza, J. E., & Langlois, S. A. (2002). The eastern cougar: A management failure? *Wildlife Society Bulletin*, *30*(1), 265–273.

Caro, T. (1994). In G. B. Schaller (Ed.), *Cheetahs of the Serengeti Plains: Group living in an asocial species* (1st ed.). University of Chicago Press.

Charruau, P., Fernandes, C., Orozco-Terwengel, P., Peters, J., Hunter, L., Ziaie, H., Jourabchian, A., Jowkar, H., Schaller, G., Ostrowski, S., Vercammen, P., Grange, T., Schlötterer, C., Kotze, A., Geigl, E. M., Walzer, C., & Burger, P. A. (2011). Phylogeography, genetic structure and population divergence time of cheetahs in Africa and Asia: Evidence for long-term geographic isolates. *Molecular Ecology*, *20*(4), 706–724. https://doi.org/10.1111/j.1365-294X.2010.04986.x

Cohen, L., Manion, L., & Morrison, K. (2017). Validity and reliability. In *Research methods in education* (Vol. *44*, 8th ed., pp. 245–284). Routledge. https://doi.org/10.4324/9781315456539-14

Cooper, E. B., Bonnet, T., Osmond, H. L., Cockburn, A., & Kruuk, L. E. B. (2021). Aging and senescence across reproductive traits and survival in superb fairy-wrens (*Malurus cyaneus*). *American Naturalist*, *197*(1), 111–127. https://doi.org/10.1086/711755

De Paoli-Iseppi, R., Polanowski, A. M. A. M., McMahon, C., Deagle, B. E. B. E., Dickinson, J. L. J. L., Hindell, M. A. M. A., & Jarman, S. N. S. N. (2017). DNA methylation levels in candidate genes associated with chronological age in mammals are not conserved in a long-lived seabird. *PLoS One*, *12*(12), 1–20. https://doi.org/10.1371/journal.pone.0189181

Dias, H. C., Cordeiro, C., Pereira, J., Pinto, C., Real, F. C., Cunha, E., & Manco, L. (2020). DNA methylation age estimation in blood samples

of living and deceased individuals using a multiplex SNaPshot assay. *Forensic Science International*, *311*, 110267. https://doi.org/10.1016/j.forsciint.2020.110267

Double, M. C., Dawson, D., Burke, T., & Cockburn, A. (1997). Finding the fathers in the least faithful bird: A microsatellite-based genotyping system for the superb fairy-wren *Malurus cyaneus*. *Molecular Ecology*, *6*(7), 691–693. https://doi.org/10.1046/J.1365-294X.1997.00228.X

Durant, S. M., Mitchell, N., Groom, R., Pettorelli, N., Ipavec, A., Jacobson, A. P., Woodroffe, R., Böhm, M., Hunter, L. T. B., Becker, M. S., Broekhuis, F., Bashir, S., Andresen, L., Aschenborn, O., Beddiaf, M., Belbachir, F., Belbachir-Bazi, A., Berbash, A., de Matos Machado, I. B., … Young-Overton, K. (2017). The global decline of cheetah *Acinonyx jubatus* and what it means for conservation. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(3), 528–533. https://doi.org/10.1073/pnas.1611122114

Durant, S. M., Mitchell, N., Ipavec, A., & Groom, R. (2015). *Acinonyx jubatus*, cheetah. The IUCN Red List of Threathened Species, 8235.

Fàbregas, M. C., & Garcés-Narro, C. (2014). Validation of gum-line recession as a reliable technique to age tigers. *European Journal of Wildlife Research*, *60*(6), 947–950. https://doi.org/10.1007/s10344-014-0869-1

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fitch, H. M., & Fagan, D. A. (1982). Focal palatine erosion associated with dental malocclusion in captive cheetahs. *Zoo Biology*, *1*(4), 295–310. https://doi.org/10.1002/zoo.1430010403

Fleckhaus, J., & Schneider, P. M. (2020). Novel multiplex strategy for DNA methylation-based age prediction from small amounts of DNA via pyrosequencing. *Forensic Science International: Genetics*, *44*, 102189. https://doi.org/10.1016/j.fsigen.2019.102189

Freire-Aradas, A., Phillips, C., Mosquera-Miguel, A., Girón-Santamaría, L., Gómez-Tato, A., de Casares Cal, M., Álvarez-Dios, J., Ansede-Bermejo, J., Torres-Español, M., Schneider, P. M., Pośpiech, E., Branicki, W., Carracedo, Á., & Lareu, M. V. (2016). Development of a methylation marker set for forensic age estimation using analysis of public methylation data and the Agena Bioscience EpiTYPER system. *Forensic Science International: Genetics*, *24*, 65–74. https://doi.org/10.1016/j.fsigen.2016.06.005

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1), 1–22. https://doi.org/10.18637/jss.v033.i01

Garagnani, P., Bacalini, M. G., Pirazzini, C., Gori, D., Giuliani, C., Mari, D., Di Blasio, A. M., Gentilini, D., Vitale, G., Collino, S., Rezzi, S., Castellani, G., Capri, M., Salvioli, S., & Franceschi, C. (2012). Methylation of ELOVL2 gene as a new epigenetic marker of age. *Aging Cell*, *11*(6), 1132–1134. https://doi.org/10.1111/acel.12005

Giuliani, C., Cilli, E., Bacalini, M. G., Pirazzini, C., Sazzini, M., Gruppioni, G., Franceschi, C., Garagnani, P., & Luiselli, D. (2016). Inferring chronological age from DNA methylation patterns of human teeth. *American Journal of Physical Anthropology*, *159*(4), 585–595. https://doi.org/10.1002/ajpa.22921

Gopalan, S., Carja, O., Fagny, M., Patin, E., Myrick, J. W., McEwen, L. M., Mah, S. M., Kobor, M. S., Froment, A., Feldman, M. W., Quintana-Murci, L., & Henn, B. M. (2017). Trends in DNA methylation with age replicate across diverse human populations. *Genetics*, *206*(3), 1659–1674. https://doi.org/10.1534/genetics.116.195594

Haeussler, M., Zweig, A. S., Tyner, C., Speir, M. L., Rosenbloom, K. R., Raney, B. J., Lee, C. M., Lee, B. T., Hinrichs, A. S., Gonzalez, J. N., Gibson, D., Diekhans, M., Clawson, H., Casper, J., Barber, G. P., Haussler, D., Kuhn, R. M., & Kent, W. J. (2019). The UCSC Genome Browser database: 2019 update. *Nucleic Acids Research*, *47*(D1), D853–D858. https://doi.org/10.1093/nar/gky1095

Hammersley, M. (1987). Some notes on the terms 'Validity' and 'Reliability'. *British Educational Research Journal*, *13*(1), 73–82. https://doi.org/10.1080/0141192870130107

Heeney, J. L., Evermann, J. F., McKeirnan, A. J., Marker-Kraus, L., Roelke, M. E., Bush, M., Wildt, D. E., Meltzer, D. G., Colly, L., & Lukas, J. (1990). Prevalence and implications of feline coronavirus infections of captive and free-ranging cheetahs (*Acinonyx jubatus*). *Journal of Virology*, *64*(5), 1964–1972. https://doi.org/10.1128/JVI.64.5.1964-1972.1990

Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biology*, *14*(10), 3156. https://doi.org/10.1186/gb-2013-14-10-r115

Horvath, S., Lu, A. T., Haghani, A., Zoller, J. A., Li, C. Z., Lim, A. R., Brooke, R. T., Raj, K., Serres-Armero, A., Dreger, D. L., Hogan, A. N., Plassais, J., & Ostrander, E. A. (2022). DNA methylation clocks for dogs and humans. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(21), e2120887119. https://doi.org/10.1073/pnas.2120887119

Horvath, S., & Raj, K. (2018). DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nature Reviews Genetics*, *19*(6), 371–384. https://doi.org/10.1038/s41576-018-0004-3

International Union for Conservation of Nature (IUCN). (2022). *Digital Distribution Maps on The IUCN Red List of Threatened Species* (p. Version 6.2). p. Version 6.2. https://www.iucnredlist.org/resources/spatial-data-download

Ito, G., Yoshimura, K., & Momoi, Y. (2017). Analysis of DNA methylation of potential age-related methylation sites in canine peripheral blood leukocytes. *Journal of Veterinary Medical Science*, *79*(4), 745–750. https://doi.org/10.1292/jvms.16-0341

Ito, H., Udono, T., Hirata, S., & Inoue-Murayama, M. (2018). Estimation of chimpanzee age based on DNA methylation. *Scientific Reports*, *8*(1), 1–5. https://doi.org/10.1038/s41598-018-28318-9

Jabbari, K., Cacciò, S., Païs De Barros, J. P., Desgrès, J., & Bernardi, G. (1997). Evolutionary changes in CpG and methylation levels in the genome of vertebrates. *Gene*, *205*(1–2), 109–118. https://doi.org/10.1016/S0378-1119(97)00475-7

James, B. A. P. (1985). Variance reduction techniques. *Journal of the Operational Research Society*, *36*(6), 525–530. https://doi.org/10.1057/JORS.1985.88/METRICS

Johansson, Å., Enroth, S., & Gyllensten, U. (2013). Continuous aging of the human DNA methylome throughout the human lifespan. *PLoS One*, *8*(6), e67378. https://doi.org/10.1371/journal.pone.0067378

Kelly, M. J., Laurenson, M. K., FitzGibbon, C. D., Collins, D. A., Durant, S. M., Frame, G. W., Bertram, B. C., & Caro, T. M. (1998). Demography of the Serengeti cheetah (*Acinonyx jubatus*) population: The first 25 years. *Journal of Zoology*, *244*(4), 473–488. https://doi.org/10.1111/J.1469-7998.1998.TB00053.X

Kennedy, M., Citino, S., Hillis McNabb, A., Serino Moffatt, A., Gertz, K., & Kania, S. (2002). Detection of feline coronavirus in captive Felidae in the USA. *Journal of Veterinary Diagnostic Investigation*, *14*(6), 520–522. https://doi.org/10.1177/104063870201400615

Kim, K. S., & Sappington, T. W. (2013). Microsatellite data analysis for population genetics. *Methods in Molecular Biology*, *1006*, 271–295. https://doi.org/10.1007/978-1-62703-389-3_19

Knight, J., Flack-Davison, E., Engelbrecht, S., Visagie, R. G., Beukes, W., Coetzee, T., Mwale, M., & Ralefala, D. (2022). A literature review analysis of engagement with the Nagoya Protocol, with specific application to Africa. *South African Journal of Bioethics and Law*, *15*(2), 69–74. https://doi.org/10.7196/SAJBL.2022.v15i2.811

Koch, C. M., & Wagner, W. (2011). Epigenetic-aging-signature to determine age in different tissues. *Aging*, *3*(10), 1018–1027.

Krausman, P. R., & Morales, S. M. (2005). Acinonyx jubatus. *Mammalian Species*, *771*(1), 1. https://doi.org/10.1644/1545-1410(2005)771[0001:aj]2.0.co;2

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, *28*(5), 1–26. https://doi.org/10.18637/jss.v028.i05

Kujoth, C. C., Hiona, A., Pugh, T. D., Someya, S., Panzer, K., Wohlgemuth, S. E., Hofer, T., Seo, A. Y., Sullivan, R., Jobling, W. A., Morrow, J. D., Van Remmen, H., Sedivy, J. M., Yamasoba, T., Tanokura, M., Weindruch, R., Leeuwenburgh, C., & Prolla, T. A. (2005). Medicine: Mitochondrial DNA mutations, oxidative stress, and apoptosis in mammalian aging. *Science*, *309*(5733), 481–484. https://doi.org/10.1126/science.1112125

Lachance, H., Cantrell, B., Friedman, S., Mckay, S., & Yee, R. (2015). Genome-wide epigenetic survey of Vermont bobcat (*Lynx rufus*) and fisher (*Martes pennanti*). In: 74th Annual Conference of the Northeast Association of Fish and Wildlife Agencies.

Le Clercq, L., Dalton, D. L., Kotzé, A., & Grobler, J. P. (2023a). Designing an EpiTYPER bisulfite sequencing assay for age estimation in *Acinonyx jubatus* based on human orthologues. *Protocols.Io*. https://doi.org/10.17504/protocols.io.j8nlk4yk1g5r/v1

Le Clercq, L., Dalton, D. L., Kotzé, A., & Grobler, J. P. (2023b). DNA extraction protocol for animal blood samples using the E.Z.N.A blood mini kit. *Protocols.Io*. https://doi.org/10.17504/protocols.io.ewov141xpvr2/v1

Le Clercq, L., Kotzé, A., Grobler, J. P., & Dalton, D. L. (2023). Biological clocks as age estimation markers in animals: A systematic review and meta-analysis. *Biological Reviews*, *98*(6), 1972–2011. https://doi.org/10.1111/brv.12992

Le Clercq, L. S. I., Dalton, D. L., & Kotze, A. (2018). Molecular age estimation based on promotor CpG methylation using methylation sensitive PCR. 9th Annual Research Symposium of the National Zoological Gardens. Pretoria: ScienceOpen https://doi.org/10.14293/S2199-1006.1.SOR-.PPROAEF.V1

Linck, E., Freeman, B. G., & Dumbacher, J. P. (2019). Speciation and gene flow across an elevational gradient in New Guinea kingfishers. *BioRxiv*. https://doi.org/10.1101/589044

Lumley, T. (2020). *leaps: Regression Subset Selection. R package*. https://cran.r-project.org/package=leaps

Mace, G. M., Collar, N. J., Gaston, K. J., Hilton-Taylor, C., Akçakaya, H. R., Leader-Williams, N., Milner-Gulland, E. J., & Stuart, S. N. (2008). Quantification of extinction risk: IUCN's system for classifying threatened species. *Conservation Biology*, *22*(6), 1424–1442. https://doi.org/10.1111/j.1523-1739.2008.01044.x

Marker, L. L., & Dickman, A. J. (2003). Morphology, physical condition, and growth of the cheetah (*Acinonyx jubatus jubatus*). *Journal of Mammalogy*, *84*(3), 840–850. https://doi.org/10.1644/BRB-036

Marker, L. L., Dickman, A. J., Jeo, R. M., Mills, M. G. L., & Macdonald, D. W. (2003). Demography of the Namibian cheetah, *Acinonyx jubatus jubatus*. *Biological Conservation*, *114*(3), 413–425. https://doi.org/10.1016/S0006-3207(03)00069-7

Márquez-Ruiz, A. B., González-Herrera, L., de Luna, J. D., & Valenzuela, A. (2020). DNA methylation levels and telomere length in human teeth: Usefulness for age estimation. *International Journal of Legal Medicine*, *134*(2), 451–459. https://doi.org/10.1007/s00414-019-02242-7

Mawlood, S. K., Dennany, L., Watson, N., & Pickard, B. S. (2016). The EpiTect methyl qPCR assay as novel age estimation method in forensic biology. *Forensic Science International*, *264*(April), 132–138. https://doi.org/10.1016/j.forsciint.2016.03.047

Mayne, B., Berry, O., & Jarman, S. (2021). Optimal sample size for calibrating DNA methylation age estimators. *Molecular Ecology Resources*, *21*(7), 2316–2323. https://doi.org/10.1111/1755-0998.13437

Merlin, C., & Liedvogel, M. (2019). The genetics and epigenetics of animal migration and orientation: Birds, butterflies and beyond. *The Journal of Experimental Biology*, *222*(Pt Suppl 1), jeb191890. https://doi.org/10.1242/jeb.191890

Morin, P. A., Nestler, A., Rubio-Cisneros, N. T., Robertson, K. M., & Mesnick, S. L. (2005). Interfamilial characterization of a region of the ZFX and ZFY genes facilitates sex determination in cetaceans and other mammals. *Molecular Ecology*, *14*(10), 3275–3286. https://doi.org/10.1111/j.1365-294X.2005.02651.x

Naji, S., Colard, T., Blondiaux, J., Bertrand, B., D'Incau, E., & Bocquet-Appel, J. P. (2016). Cementochronology, to cut or not to cut? *International Journal of Paleopathology*, *15*, 113–119. https://doi.org/10.1016/j.ijpp.2014.05.003

Nakamura, S., Yamazaki, J., Matsumoto, N., Inoue-Murayama, M., Qi, H., Yamanaka, M., Nakanishi, M., Yanagawa, Y., Sashika, M., Tsubota, T., Ito, H., & Shimozuru, M. (2023). Age estimation based on blood DNA methylation levels in brown bears. *Molecular Ecology Resources*, *23*(6), 1211–1225. https://doi.org/10.1111/1755-0998.13788

Nakanishi, N., Ichinose, F., Higa, G., & Izawa, M. (2009). Age determination of the iriomote cat by using cementum annuli. *Journal of Zoology*, *279*(4), 338–348. https://doi.org/10.1111/j.1469-7998.2009.00624.x

Nyakatura, K., & Bininda-Emonds, O. R. P. (2012). Updating the evolutionary history of Carnivora (Mammalia): A new species-level supertree complete with divergence time estimates. *BMC Biology*, *10*, 1–31. https://doi.org/10.1186/1741-7007-10-12

O'Brien, S. J., Roelke, M. E., Marker, L., Newman, A., Winkler, C. A., Meltzer, D., Colly, L., Evermann, J. F., & Wildt, D. E. (1985). Genetic basis for species vulnerability in the cheetah. *Science*, *227*, 1428–1434. https://doi.org/10.1126/science.2983425

Pan, C., Yi, S., Xiao, C., Huang, Y., Chen, X., & Huang, D. (2020). The evaluation of seven age-related CpGs for forensic purpose in blood from Chinese Han population. *Forensic Science International: Genetics*, *46*, 102251. https://doi.org/10.1016/j.fsigen.2020.102251

Poganik, J. R., Zhang, B., Baht, G. S., Kerepesi, C., Yim, S. H., Lu, A. T., Haghani, A., Gong, T., Hedman, A. M., Andolf, E., Pershagen, G., Almqvist, C., Clish, C. B., Horvath, S., White, J. P., Horvath, S., & Gladyshev, V. N. (2022). Biological age is increased by stress and restored upon recovery. *BioRxiv*. https://doi.org/10.1101/2022.05.04.490686

Polanowski, A. M., Robbins, J., Chandler, D., & Jarman, S. N. (2014). Epigenetic estimation of age in humpback whales. *Molecular Ecology Resources*, *14*(5), 976–987. https://doi.org/10.1111/1755-0998.12247

Qi, H., Kinoshita, K., Mori, T., Matsumoto, K., Matsui, Y., & Inoue-Murayama, M. (2021). Age estimation using methylation-sensitive high-resolution melting (MS-HRM) in both healthy felines and those with chronic kidney disease. *Scientific Reports*, *11*(1), 1–10. https://doi.org/10.1038/s41598-021-99424-4

R Core Team. (2020). *R: A language and environment for statisitical computing*. R Foundation for Statistical Computing. https://www.r-project.org/

Raj, K., Szladovits, B., Haghani, A., Zoller, J. A., Li, C. Z., Black, P., Maddox, D., Robeck, T. R., & Horvath, S. (2021). Epigenetic clock and methylation studies in cats. *GeroScience*, *43*(5), 2363–2378. https://doi.org/10.1007/S11357-021-00445-8/FIGURES/5

Robeck, T. R., Haghani, A., Fei, Z., Lindemann, D. M., Russell, J., Herrick, K. E. S., Montano, G., Steinman, K. J., Katsumata, E., Zoller, J. A., & Horvath, S. (2023). Multi-tissue DNA methylation aging clocks for sea lions, walruses and seals. *Communications Biology*, *6*(1), 359. https://doi.org/10.1038/s42003-023-04734-0

RStudio Team. (2021). RStudio: Integrated Development for R. *RStudio, PBC, Boston*, MA URL http://www.rstudio.com/.

Sareisian, L. (2014). *Global DNA methylation and persistent organic pollutants in ringed seals (Phoca hispida) from Svalbard and the Baltic Sea*. University of Oslo.

Shan, G. (2022). Monte Carlo cross-validation for a study with binary outcome and limited sample size. *BMC Medical Informatics and*

Decision Making, *22*(1), 1–15. https://doi.org/10.1186/s12911-022-02016-z

Shao, J. (1993). Linear model selection by cross-validation. *Journal of the American Statistical Association*, *88*(422), 486–494. https://doi.org/10.1080/01621459.1993.10476299

Sharma, V., Sharma, C. P., Sharma, V., Goyal, S. P., & Gupta, S. K. (2021). Fast and cost-effective age estimation in Bengal tiger and Asiatic lion: Applicability of cementum analysis method. *BioRxiv*. https://doi.org/10.1101/2021.09.27.461978

Spiers, H., Hannon, E., Wells, S., Williams, B., Fernandes, C., & Mill, J. (2016). Age-associated changes in DNA methylation across multiple tissues in an inbred mouse model. *Mechanisms of Ageing and Development*, *154*(21), 20–23. https://doi.org/10.1016/j.mad.2016.02.001

Stander, P. E. (1997). Field age determination of leopards by tooth wear. *African Journal of Ecology*, *35*(2), 156–161. https://doi.org/10.1111/j.1365-2028.1997.068-89068.x

Statham, A., & Csárdi, G. (2008). *RSeqMeth: Package for analysis of Sequenom EpiTYPER data*. Sydney. https://github.com/cran/RSeqMeth

Suchiman, H. E. D., Slieker, R. C., Kremer, D., Slagboom, P. E., Heijmans, B. T., & Tobi, E. W. (2015). Design, measurement and processing of region-specific DNA methylation assays: The mass spectrometry-based method EpiTYPER. *Frontiers in Genetics*, *6*, 287. https://doi.org/10.3389/fgene.2015.00287

Sugrue, V. J., Zoller, J. A., Narayan, P., Lu, A. T., Ortega-Recalde, O. J., Grant, M. J., Bawden, C. S., Rudiger, S. R., Haghani, A., Bond, D. M., Hore, R. R., Garratt, M., Sears, K. E., Wang, N., Yang, X. W., Snell, R. G., Hore, T. A., & Horvath, S. (2021). Castration delays epigenetic aging and feminizes dna methylation at androgen-regulated loci. *eLife*, *10*, e64932. https://doi.org/10.7554/eLife.64932

Tamazian, G., Simonov, S., Dobrynin, P., Makunin, A., Logachev, A., Komissarov, A., Shevchenko, A., Brukhin, V., Cherkasov, N., Svitin, A., Koepfli, K. P., Pontius, J., Driscoll, C. A., Blackistone, K., Barr, C., Goldman, D., Antunes, A., Quilez, J., Lorente-Galdos, B., … O'Brien, S. J. (2014). Annotated features of domestic cat – Felis catus genome. *GigaScience*, *3*(1), 1–3. https://doi.org/10.1186/2047-217X-3-13/TABLES/1

Tanabe, A., Shimizu, R., Osawa, Y., Suzuki, M., Ito, S., Goto, M., Pastene, L. A., Fujise, Y., & Sahara, H. (2020). Age estimation by DNA methylation in the Antarctic minke whale. *Fisheries Science*, *86*(1), 35–41. https://doi.org/10.1007/s12562-019-01371-7

Tewari, D., Miller, R., Livengood, J., Wang, L., Killian, M. L., Bustamante, F., Kessler, C., Thirumalapura, N., Terio, K., Torchetti, M., Lantz, K., & Rosenberg, J. (2023). SARS-CoV-2 infection dynamics in the Pittsburgh zoo wild felids with two viral variants (Delta and alpha) during the 2021–2022 pandemic in the United States. *Animals*, *13*(19), 3094. https://doi.org/10.3390/ani13193094

Thompson, M. J., VonHoldt, B., Horvath, S., & Pellegrini, M. (2017). An epigenetic aging clock for dogs and wolves. *Aging*, *9*(3), 1055–1068. https://doi.org/10.18632/aging.101211

Thompson, R. F., & Greally, J. M. (2020). *MassArray: Analytical Tools for MassArray Data*. https://doi.org/10.18129/B9.bioc.MassArray

Viñuela, A., Brown, A. A., Buil, A., Tsai, P. C., Davies, M. N., Bell, J. T., Dermitzakis, E. T., Spector, T. D., & Small, K. S. (2018). Age-dependent changes in mean and variance of gene expression across tissues in a twin cohort. *Human Molecular Genetics*, *27*(4), 732–741. https://doi.org/10.1093/hmg/ddx424

von Schreber, J. C. D. (1775). Der Gepard. In *Die Säugthiere in Abbildungen nach der Natur mit Beschreibungen*. Walther'sche Kunst- und Buchhandlung. https://doi.org/10.11588/diglit.3543

Weidner, C. I., Lin, Q., Koch, C. M., Eisele, L., Beier, F., Ziegler, P., Bauerschlag, D. O., Jöckel, K. H., Erbel, R., Mühleisen, T. W., Zenke, M., Brümmendorf, T. H., & Wagner, W. (2014). Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biology*, *15*(2), 1–12. https://doi.org/10.1186/gb-2014-15-2-r24

White, P. A., & Belant, J. L. (2016). Individual variation in dental characteristics for estimating age of African lions. *Wildlife Biology*, *22*(3), 71–77. https://doi.org/10.2981/wlb.00180

Wright, P. G. R., Mathews, F., Schofield, H., Morris, C., Burrage, J., Smith, A., Dempster, E. L., & Hamilton, P. B. P. B. (2018). Application of a novel molecular method to age free-living wild Bechstein's bats. *Molecular Ecology Resources*, *18*(6), 1374–1380. https://doi.org/10.1111/1755-0998.12925

Yamaguchi, N., Driscoll, C. A., Werdelin, L., Abramov, A. V., Csorba, G., Cuisin, J., Fernholm, B., Hiermeier, M., Hills, D., Hunter, L., Itakura, H., Johansson, U. S., Kascheev, V., Krohmann, K., Martin, T., Nowak-Kemp, M., Pavlinov, I. Y., Renoud, F., Tomsett, L., … Macdonald, D. W. (2013). Locating specimens of Extinct Tiger (*Panthera tigris*) subspecies: Javan Tiger (P. T. sondaica), Balinese Tiger (P. T. balica), and Caspian Tiger (P. T. virgata), including previously unpublished specimens. *Mammal Study*, *38*(3), 187–198. https://doi.org/10.3106/041.038.0307

Zainol, Z. A., Amin, L., Akpoviri, F., & Ramli, R. (2011). Biopiracy and states' sovereignty over their biological resources. *African Journal of Biotechnology*, *10*(58), 12395–12408.

Zampieri, M., Ciccarone, F., Calabrese, R., Franceschi, C., Bürkle, A., & Caiafa, P. (2015). Reconfiguration of DNA methylation in aging. *Mechanisms of Ageing and Development*, *151*, 60–70. https://doi.org/10.1016/j.mad.2015.02.002

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Le Clercq, L.-S., Kotzé, A., Grobler, J. P., & Dalton, D. L. (2024). Methylation-based markers for the estimation of age in African cheetah, *Acinonyx jubatus*. *Molecular Ecology Resources*, *24*, e13940. https://doi.org/10.1111/1755-0998.13940

CHAPTER 6

Conclusion and Recommendations

6.1 CONCLUDING REMARKS

The study of biological traits in animals, at the biochemical and molecular level, has created many specialised tools that directly benefit ecological studies and conservation efforts (Carroll *et al.*, 2018; Riekkola *et al.*, 2018). While many scientists focus on the measurement of statistic traits such as determining sex (Rovatsos & Kratochvíl, 2017) or parentage (Olson *et al.*, 2008), others have also been exploring the use of molecular methods to measure traits that vary temporally within individuals or within populations. Such studies have typically exploited the molecular clock (Thorpe, 1982; Li, Tanimura, & Sharp, 1987; Robinson & Robinson, 2001) or other biological clocks (Horvath & Raj, 2018; Lincoln, 2019). The aim of the present study was to explore the use of two biological clocks, the circadian clock and the epigenetic clock. The first within the context of differential migration patterns of birds and the potential for polymorphisms within clock genes to explain behavioural difference, and the second with regards to the ability to use epigenetic clocks to model and predict age in animals.

Various studies have compared annual life events including migratory phenology to putative variation within clock genes (Caprioli *et al.*, 2012; Saino *et al.*, 2015; Delmore *et al.*, 2016). While multiple studies found significant correlations between variability and annual synchronicity (Liedvogel *et al.*, 2009; Caprioli *et al.*, 2012; Bazzi *et al.*, 2015, 2017), a number of other studies failed to detect a significant correlation (Peterson *et al.*, 2013; Contina *et al.*, 2018; Parody-Merino *et al.*, 2019). Most previous studies were, however, focussed on relationships between clock gene diversity and migration patterns of groups within a singular species with little evidence as to whether such patterns may be lineage specific or generalisable to most avian species (Bazzi *et al.*, 2016; Lugo Ramos, Delmore, & Liedvogel, 2017). Thus, to clarify the existing evidence for an association between clock gene polymorphisms and annual timing, the first objective of this study was to perform a systematic review and standardised comparative analysis of clock genes in the context of annual life events. The first objective was achieved in article one (Le Clercq *et al.*, 2023c) of Chapter 2, published in the journal *Biological Reviews*. As

part of this review, a large dataset (Le Clercq, 2023b) comprising 8000 individuals from nearly 50 species was generated which was published in *Scientific Data* as the second article (Le Clercq *et al.*, 2023a) of Chapter 2. The second objective, forming part of the systematic review, was to explore or design methods of incorporating phylogenetic, taxonomic, and the molecular clock (e.g., divergence time data) into ecological and evolutionary reviews. For this purpose, a novel PYTHON tool called PAReTT (Le Clercq, 2023a) was developed and published in the *Journal of Molecular Evolution* as the third article (Le Clercq *et al.*, 2023f) of Chapter 2. PAReTT was used to facilitate the retrieval of time data such as divergence time as well as time-calibrated phylogenies from available online resources (Kumar *et al.*, 2022). PAReTT was further used to illustrate the importance of considering differences in the rates for molecular clocks between lineages to distinguish between species that remain monotypic for extended periods of time (Fry & Boesman, 2020), versus “great speciators” (O’Connell *et al.*, 2019) that produce many subspecies within the same epoch (Le Clercq *et al.*, 2023f).

The systematic review identified several candidate genes within which length polymorphisms have been hypothesised to play a putative role in timekeeping over the past decade, including: *Clock* (Steinmeyer, Mueller, & Kempenaers, 2009), *Adcyap1* (Mueller, Pulido, & Kempenaers, 2011), *CREB1*, *NPAS2* (Steinmeyer *et al.*, 2009), *DRD4* (Korsten *et al.*, 2010; Edwards *et al.*, 2015), and *SERT* (Mueller *et al.*, 2013). Primary results from such studies were contextualised within the existing published literature relating polymorphisms in clock genes to seasonality in a phylogenetically and taxonomically informed manner. While the review clearly illustrated that primary included studies yielded conflicted results, several limitations were identified among studies reporting negative results, and a combined re-analysis provided evidence of a putative associations between *Clock* gene variation and autumn migration as well as *Adcyap1* gene variation and spring migration in migratory species. The results did, however, show that genotypes in neither of the candidate genes could clearly distinguish migratory from sedentary birds, consistent with previous findings (Lugo Ramos *et al.*, 2017). When compared to divergence time data, correlations were evident in both genes and this may indicate that the genotypes contained in modern populations still reflect ancestral states rather than contemporary changes driven by selection (Kuhn *et al.*, 2013; Le Clercq *et al.*, 2023c). This highlights both the importance of including time calibrated phylogenies

in cross taxa analyses as well as possible evolutionary constraints within genes that affect internal time keeping and migration behaviour.

Scientometric considerations revealed that most studies emanated from North America and Europe, with a high number of studies focused on passerine species (Le Clercq *et al.*, 2023a, 2023c). There are, however, many examples of African species, including many non-passerine species, that follow patterns of differential migration on the continent. This includes previously mentioned species such as the European Bee-eater (Fry & Boesman, 2020), Woodlands Kingfisher (Woodall, 2020), and Lilac-Breasted Roller (Fry, Kirwan, & Boesman, 2021). Thus, the subsequent objectives were to sample different populations of the non-passerine and African species, Diederik cuckoo, which shows divergent migratory behaviour. In line with the study objectives, sampling was done in such a manner that comparisons could be made between resident versus migrant individuals, migrants of two different sites, as well as early migrating versus late migrating birds. Subsequently, polymorphism in two candidate genes linked to the circadian clock, the *Clock* and *Adcyap1* genes, were amplified and sequenced for comparison to various attributes of Diederik cuckoo migration phenology (migratory distance, time of departure/arrival, and breeding latitude).

The primary results, reported in the article published in *Ecology and Evolution* (Le Clercq *et al.*, 2024b), report on the three alleles that were detected for *Clock*, with comparable genotypes between cuckoos sampled from the North- and South African breeding ranges. Ten alleles were detected for *Adcyap1*, having shorter alleles in the North and longer alleles in the South. Population genetic analyses showed a substantially higher frequency for the Q_{10} allele, with many individuals being homozygotic while the *Adcayp1* gene showed substantially more diversity and complete heterozygosity. Upon comparing *Clock* and *Adcyap1* alleles to phenology attributes, several phenotypic correlations were detected. This included evidence of a relationship between the shorter alleles and habitat selection (Liedvogel *et al.*, 2009), as measured by latitude and longitude, as well as a relationship between longer alleles and timing (Saino *et al.*, 2015). In both instances, evidence is provided that these effects may be sex specific (Bazzi *et al.*, 2017). This provides clear evidence that complex associations exist for these candidate genes in African birds.

Age is the second biological trait investigated that changes over time and is critical to understanding population dynamics in animals (Tarnita *et al.*, 2009; Durant

et al., 2017; Stalmans *et al.*, 2019). Two biological clocks have previously been employed in the estimation of age in wildlife: tooth cementum annulation (Takken Beijersbergen, 2019) in toothed vertebrates and otolith annulation in fish (Campana, 2001). Both of these clocks are highly accurate; however, their invasiveness and lethal sampling needed makes these methods less useful in population monitoring for vulnerable species. Aging is a programmed process under genetic control (Horvath and Raj, 2018), making molecular methods to determine age in wildlife a promising avenue of research. Several molecular methods have been suggested including differential methylation (Jarman *et al.*, 2015), telomere length (Remot *et al.*, 2021), and mitochondrial mutations (Zapico & Ubelaker, 2016). The second objective of this study was therefore to perform a systematic review and meta-analysis to summarise and evaluate the existing literature exploring the use of epigenetic clocks as biomarkers to model age in animals.

The second objective was accomplished in the first and second articles of Chapter 3, published in the journal *Biological Reviews* (Le Clercq *et al.*, 2023e) and *Data in Brief* (Le Clercq *et al.*, 2023b). The systematic review summarised the two most frequently used methods, methylation and telomere length, by describing what they are, how they change with age, and providing a detailed summary of studies that have used them to develop age-estimation models. The results showed that both methods have been studied across multiple vertebrate classes over the past two decades, however, telomere studies were used prior to methylation studies and has been modelled in nearly twice as many studies. Telomere length studies often related changes to stress responses and illustrated that telomere length is sensitive to environmental and social stressors that, in the absence of repair mechanisms, does not have the ability to recover. In contrast, methylation studies also detected sensitivity to stressors, however, methylation levels were able to recover after a period of stress-induced accelerated ageing.

The meta-analysis provided a quantitative evaluation of the strength of association and magnitude of effect for both methods across five vertebrate classes. The results showed that methylation outperforms telomere length in terms of predictive power based on higher effect sizes and smaller prediction intervals. The review and meta-analysis further attempted to account for possible factors that could be attributed to any error or variation in the reliability of either method. This included meta-regressions comparing lifespan of the species, genome size, karyotype, or the

tissue type, to the differences in effect sizes. Both methods seemed to perform consistently across studies, independent of such factors. Rather, other factors such as quantitative method, patterns of inheritance, and environmental factors, played a more significant role. Thus, provided that these factors can be accounted for, both methylation and telomere length are promising targets to develop as biomarkers for age determination in animals, with methylation yielding more consistent results.

An ancillary objective to this review was to explore potential sources of heterogeneity and bias within the included studies, as well as possible methods to account for bias in systematic reviews in evolution and ecology. This was achieved through meta-regression included in the review (Le Clercq *et al.*, 2023e) as well as a novel PYTHON program called ABCal (Le Clercq, 2023c); the novel metrics designed as part of ABCal and scientometric capabilities was published in *Scientometrics* as the third article (Le Clercq, 2024) of Chapter 3. The development of ABCal was necessitated by the clear abundance of publications from the same set of authors, introducing possibly publication bias, which may be an inherent problem in evolutionary and ecological reviews (Thornton & Lee, 2000; O'Dea *et al.*, 2021). Scientometric evaluation of studies included in the second review (Le Clercq *et al.*, 2023b) revealed a similar pattern to the first review, with most studies emanating from North America and Europe and largely excluding African species. However, Africa has a large amount of biodiversity of which many are classified as threatened or vulnerable and need population modelling (Tarnita *et al.*, 2009) and monitoring (Stalmans *et al.*, 2019) to access recovery.

As such, methylation, the top performing epigenetic clock from the review, was selected as the biological clock for developing an age-estimation model for African cheetah. Specific objectives were to select a cohort of samples for cheetah that includes most known age classes and design a method to assay differential methylation in candidate genes identified in human studies using mass array technology. Those genes with clear evidence of age-related differential methylation would then be used to create a model to predict age in cheetah.

The primary results, reported in Chapter 5, were published in *Molecular Ecology Resources* (Le Clercq *et al.*, 2024a). A cohort of fifty samples were selected based on sample size recommendations (Mayne, Berry, & Jarman, 2021) with an equal distribution of sexes and representing seven of the eight functional age classes for cheetah (Burney, 1980; Caro, 1994; Marker & Dickman, 2003). This

study identified six candidate genes, *EDARADD* (Bocklandt *et al.*, 2011), *ELOVL2* (Bekaert *et al.*, 2015), *FHL2* (Giuliani *et al.*, 2016), *GRIA2* (Koch & Wagner, 2011), *ITGA2B* (Weidner *et al.*, 2014), and *PENK* (Giuliani *et al.*, 2016), with gene orthologues between humans and cheetah. This enabled the design of an assay (Thompson *et al.*, 2009; Suchiman *et al.*, 2015; Le Clercq *et al.*, 2023d) that could assay target regions containing a hundred CpG's. Linear regressions between CpG methylation and known age identified six CpG's with an age relationship. Four of these CpG's were hypomethylated, characteristic of expected methylation trends with aging (Johansson, Enroth, & Gyllensten, 2013). In contract, two CpG's were hypermethylated which is indicative of a strong epigenetic clock (Jarman *et al.*, 2015; Horvath & Raj, 2018), and is likely associated with increased transcriptional regulation with age (Akintola *et al.*, 2008). Machine learning, as implemented in elastic net regression models, was used to fit different combinations of CpG's to select an optimal age estimation model. The results showed that cheetah age models using four or six CpG's were most accurate, with a six CpG model having a comparatively higher correlation and predictive accuracy. This model was more accurate than previous attempts using methylation sensitive Polymerase Chain Reaction (Le Clercq, Dalton, & Kotze, 2018), provides the first calibrated clock designed and tested in cheetah specifically (Raj *et al.*, 2021), and performed similarly to models created using a candidate gene approach in several other mammal species; included in the review (Le Clercq *et al.*, 2023e).

In summary, this study illustrates that biological clocks are useful in studying traits that change temporally, from the annual cycle of life events in birds to lifelong aging in mammals. The circadian clock and clock genes contain several polymorphisms that are correlated with migratory phenotypes and given the high dependence on the optimal use of seasonal environments for survival in migratory species, may serve as important indicators for their fitness. This makes the study of such polymorphism particularly useful in the face of persistent threats to habitats as well as the changing landscapes in response to climate change. The epigenetic clocks provide the unique opportunity to develop less invasive age estimation methods which are critical for populations that require monitoring. Furthermore, this study generated large datasets and several new tools that can be used in various aspects of research from accounting for author bias in ecological systematic reviews to incorporating molecular clock data into both reviews and individual studies. The

multiple publications emanating from this study have broadened and enhanced the current understanding of each biological clock and will hopefully serve as a directive for future research.

6.2 FUTURE PERSPECTIVES

Biological clocks, given their clear utility in answering pertinent questions in ecology and conservation, warrant further study. This is particularly needed for those geographic regions identified as being data deficient or for species that are classified as vulnerable to threatened and in need of improved long-term monitoring. It is, however, important that factors affecting the success of such studies are considered. Throughout this study, several key factors associated with the performance of either biological clock were identified.

For studies focused on relating polymorphisms in candidate genes to annual life cycles, the most pertinent considerations for optimal study resolution are the taxonomic level at which genetic diversity is studied, the functional grouping of subpopulations in the analyses, and study design attributes common to many conservation genetics studies. In terms of taxonomic level, studies were most successful when assaying candidate genes between species of the same genus such as *Saxicola* stonechats (Justen *et al.*, 2022), *Tachycineta* swallows (Dor *et al.*, 2012), or *Catharus* thrushes (Voelker, Bowie, & Klicka, 2013). Studies were less successful at detecting associations within populations of a singular subspecies (Parody-Merino *et al.*, 2019), although between subspecies analyses did yield some evidence for differentiation (Peterson *et al.*, 2013; Contina *et al.*, 2018); at least in terms of genotypes present and allele frequencies. These studies did, however, illustrate that subgrouping members of the same subspecies as individual population based on geography could obscure the results. Lastly, pertinent factors to consider in study design are sample sizes (with at least twenty to thirty samples required in most cases), the direct measurement of the polymorphic repeat itself (barring any secondary source of length variation), and performing group analyses within the taxonomic context rather that using arbitrary population clustering (e.g., based on sample site).

For studies exploring the use of epigenetic clocks such as methylation to develop age estimation models in animals, the most pertinent considerations were

quantitative method, the age demographics for included individuals, and similar study design attributes to those mentioned for clock genes. In terms of quantitative method, a site directed approach using bisulfite sequencing methods outperformed other methods such as methylation sensitive PCR (Le Clercq *et al.*, 2018) or ELISA (Sareisian, 2014). For age demographics the younger individuals tended to provide less informative data (Le Clercq *et al.*, 2024a) and, given that methylation-based methods are more likely to be of benefit in age determination in adult classes, studies aimed at developing molecular age determination methods may benefit from including more individuals from adult classes. Lastly, study design attributes such as sample sizes also factor into studies of epigenetic clocks, with sample sizes above thirty but preferably close to sixty being optimal (Mayne *et al.*, 2021). It should also be noted that the number of sites assayed can greatly affect the sample size requirements and the inclusion of sites with large effect sizes may offset these recommendations.

Given these factors can be accounted for, future research on biological clocks will continue to advance our understanding of traits that vary temporally and provide useful information to be incorporated into continued conservation efforts.

REFERENCES

AKINTOLA, A.D., CRISLIP, Z.L., CATANIA, J.M., CHEN, G., ZIMMER, W.E., BURGHARDT, R.C. & PARRISH, A.R. (2008) Promoter methylation is associated with the age-dependent loss of N-cadherin in the rat kidney. *American Journal of Physiology-Renal Physiology* **294**, F170–F176.

BAZZI, G., AMBROSINI, R., CAPRIOLI, M., COSTANZO, A., LIECHTI, F., GATTI, E., GIANFRANCESCHI, L., PODOFILLINI, S., ROMANO, A., ROMANO, M., SCANDOLARA, C., SAINO, N. & RUBOLINI, D. (2015) Clock gene polymorphism and scheduling of migration: A geolocator study of the barn swallow *Hirundo rustica*. *Scientific Reports* **5**, 12443.

BAZZI, G., CECERE, J.G., CAPRIOLI, M., GATTI, E., GIANFRANCESCHI, L., PODOFILLINI, S., POSSENTI, C.D., AMBROSINI, R., SAINO, N., SPINA, F. & RUBOLINI, D. (2016) Clock gene polymorphism, migratory behaviour and geographic distribution: a comparative study of trans-Saharan migratory birds. *Molecular Ecology* **25**, 6077–6091.

BAZZI, G., PODOFILLINI, S., GATTI, E., GIANFRANCESCHI, L., CECERE, J.G., SPINA, F., SAINO, N. & RUBOLINI, D. (2017) Candidate genes have sex-specific effects on timing of spring migration and moult speed in a long-distance migratory bird. *Current Zoology* **63**, 479–486.

BEKAERT, B., KAMALANDUA, A., ZAPICO, S.C., VAN DE VOORDE, W. & DECORTE, R. (2015) Improved age determination of blood and teeth samples using a selected set of DNA methylation markers. *Epigenetics* **10**, 922–930.

BOCKLANDT, S., LIN, W., SEHL, M.E., SÁNCHEZ, F.J., SINSHEIMER, J.S., HORVATH, S. & VILAIN, E. (2011) Epigenetic Predictor of Age. *PLOS ONE* **6**, e14821.

BURNEY, D.A. (1980) The effects of human activities on cheetah (*Acinonyx jubatus*) in the Mara region of Kenya. Masters dissertation. University of Nairobi, Kenya.

CAMPANA, S.E. (2001) Accuracy, precision and quality control in age determination, including a review of the use and abuse of age validation methods. *Journal of Fish Biology* **59**, 197–242.

CAPRIOLI, M., AMBROSINI, R., BONCORAGLIO, G., GATTI, E., ROMANO, A., ROMANO, M., RUBOLINI, D., GIANFRANCESCHI, L. & SAINO, N. (2012) Clock gene variation is associated with breeding phenology and maybe under directional selection in the migratory barn swallow. *PLoS ONE* **7**, e35140.

CARO, T. (1994) *Cheetahs of the Serengeti Plains: group living in an asocial species*, 1st edition. University of Chicago Press, Chicago.

CARROLL, E.L., BRUFORD, M.W., DEWOODY, J.A., LEROY, G., STRAND, A., WAITS, L. & WANG, J. (2018) Genetic and genomic monitoring with minimally invasive sampling methods. *Evolutionary Applications* **11**, 1094–1119.

CONTINA, A., BRIDGE, E.S., ROSS, J.D., SHIPLEY, J.R. & KELLY, J.F. (2018) Examination of *Clock* and *Adcyap1* gene variation in a neotropical migratory passerine. *Plos One* **13**, e0190859.

DELMORE, K.E., TOEWS, D.P.L., GERMAIN, R.R., OWENS, G.L. & IRWIN, D.E. (2016) The Genetics of Seasonal Migration and Plumage Color. *Current Biology* **26**, 2167–2173.

DOR, R., COOPER, C.B., LOVETTE, I.J., MASSONI, V., BULIT, F., LILJESTHROM, M. & WINKLER, D.W. (2012) Clock gene variation in *tachycineta* swallows. *Ecology and Evolution* **2**, 95–105.

DURANT, S.M., MITCHELL, N., GROOM, R., PETTORELLI, N., IPAVEC, A., JACOBSON,

A.P., WOODROFFE, R., BÖHM, M., HUNTER, L.T.B., BECKER, M.S., BROEKHUIS, F., BASHIR, S., ANDRESEN, L., ASCHENBORN, O., BEDDIAF, M., ET AL. (2017) The global decline of cheetah *Acinonyx jubatus* and what it means for conservation. *Proceedings of the National Academy of Sciences* **114**, 528–533.

EDWARDS, H.A., HAJDUK, G.K., DURIEUX, G., BURKE, T. & DUGDALE, H.L. (2015) No association between personality and candidate gene polymorphisms in a wild bird population. *PLoS ONE* **10**, e0138439.

FRY, H. & BOESMAN, P.F.D. (2020) European Bee-eater (*Merops apiaster*), version 1.0. In *Birds of the World* (eds J. DEL HOYO, A. ELLIOTT, J. SARGATAL, D.A. CHRISTIE & E. DE JUANA), Cornell Lab of Ornithology, Ithaca, NY, USA.

FRY, H., KIRWAN, G.M. & BOESMAN, P.F.D. (2021) Lilac-breasted Roller (*Coracias caudatus*). In *Birds of the World* (eds J. DEL HOYO, A. ELLIOTT, J. SARGATAL, D. CHRISTIE & E. DE JUANA), Cornell Lab of Ornithology, Ithaca, NY, USA.

GIULIANI, C., CILLI, E., BACALINI, M.G., PIRAZZINI, C., SAZZINI, M., GRUPPIONI, G., FRANCESCHI, C., GARAGNANI, P. & LUISELLI, D. (2016) Inferring chronological age from DNA methylation patterns of human teeth. *American Journal of Physical Anthropology* **159**, 585–595.

HORVATH, S. & RAJ, K. (2018) DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nature Reviews Genetics* **19**, 371–384.

JARMAN, S.N., POLANOWSKI, A.M., FAUX, C.E., ROBBINS, J., DE PAOLI-ISEPPI, R., BRAVINGTON, M. & DEAGLE, B.E. (2015) Molecular biomarkers for chronological age in animal ecology. *Molecular Ecology* **24**, 4826–4847.

JOHANSSON, Å., ENROTH, S. & GYLLENSTEN, U. (2013) Continuous Aging of the Human DNA Methylome Throughout the Human Lifespan. *PLoS ONE* **8**, e67378.

JUSTEN, H., HASSELMANN, T., ILLERA, J.C., DELMORE, K.E., SERRANO, D., FLINKS, H., SENZAKI, M., KAWAMURA, K., HELM, B. & LIEDVOGEL, M. (2022) Population-specific association of *Clock* gene polymorphism with annual cycle timing in stonechats. *Scientific Reports* **12**, 1–13.

KOCH, C.M. & WAGNER, W. (2011) Epigenetic-aging-signature to determine age in different tissues. *Aging* **3**, 1018–1027.

KORSTEN, P., MUELLER, J.C., HERMANNSTÄDTER, C., BOUWMAN, K.M., DINGEMANSE, N.J., DRENT, P.J., LIEDVOGEL, M., MATTHYSEN, E., VAN OERS, K., VAN

OVERVELD, T., PATRICK, S.C., QUINN, J.L., SHELDON, B.C., TINBERGEN, J.M. & KEMPENAERS, B. (2010) Association between *DRD4* gene polymorphism and personality variation in great tits: A test across four wild populations. *Molecular Ecology* **19**, 832–843.

KUHN, K., SCHWENK, K., BOTH, C., CANAL, D., JOHANSSON, U.S., VAN DER MIJE, S., TÖPFER, T. & PÄCKERT, M. (2013) Differentiation in neutral genes and a candidate gene in the pied flycatcher: Using biological archives to track global climate change. *Ecology and Evolution* **3**, 4799–4814.

KUMAR, S., SULESKI, M., CRAIG, J.M., KASPROWICZ, A.E., SANDERFORD, M., LI, M., STECHER, G. & HEDGES, S.B. (2022) TimeTree 5: An Expanded Resource for Species Divergence Times. *Molecular Biology and Evolution* **39**, msac174.

LE CLERCQ, L.S. (2023a) PAReTT: Python Automated Retrieval of TimeTree data. GitHub. https://github.com/LSLeClercq/PAReTT.

LE CLERCQ, L.S. (2023*b*) AvianClocksData: Dataset of Clock and Adcyap1 alleles for Birds. GitHub. https://github.com/LSLeClercq/AvianClocksData.

LE CLERCQ, L.S. (2023*c*) ABCal: Author Bias Computation and Scientometric Plotting. GitHub. https://github.com/LSLeClercq/ABCal.

LE CLERCQ, L.S. (2024) ABCal: a Python package for Author Bias Computation and Scientometric Plotting for Reviews and Meta-Analyses. *Scientometrics* **129**, 581–600.

LE CLERCQ, L.S., DALTON, D.L. & KOTZE, A. (2018) Molecular age estimation based on promotor CpG methylation using Methylation Sensitive PCR. In *9th Annual Research Symposium of the National Zoological Gardens*. Pretoria.

LE CLERCQ, L.-S., BAZZI, G., FERRER OBIOL, J., CECERE, J., GIANFRANCESCHI, L., GROBLER, J.P., KOTZE, A., RIUTORT, M., GONZÁLEZ-SOLÍS, J., RUBOLINI, D., LIEDVOGEL, M. & DALTON, D.L. (2023*a*) Birds of a feather flock together: a dataset for Clock and Adcyap1 genes from migration genetics studies. *Scientific Data* **10**, 787.

LE CLERCQ, L.-S., GROBLER, J.P., KOTZÉ, A. & DALTON, D.L. (2023*b*) Dataset generated in a systematic review and meta-analysis of biological clocks as age estimation markers in animal ecology. *Data in Brief* **51**, 109615.

LE CLERCQ, L.S., BAZZI, G., CECERE, J.G., GIANFRANCESCHI, L., GROBLER, J.P., KOTZÉ, A., RUBOLINI, D., LIEDVOGEL, M. & DALTON, D.L. (2023*c*) Time trees and clock genes: a systematic review and comparative analysis of contemporary

avian migration genetics. *Biological Reviews* **98**, 1051–1080.

LE CLERCQ, L.S., DALTON, D.L., KOTZÉ, A. & GROBLER, J.P. (2023*d*) Designing an EpiTYPER bisulfite sequencing assay for age estimation in *Acinonyx jubatus* based on human orthologues. *protocols.io*.

LE CLERCQ, L.S, KOTZÉ, A., GROBLER, J.P. & DALTON, D.L. (2023*e*) Biological clocks as age estimation markers in animals: a systematic review and meta-analysis. *Biological Reviews* **98**, 1972–2011.

LE CLERCQ, L.S, KOTZÉ, A., GROBLER, J.P. & DALTON, D.L. (2023*f*) PAReTT: A Python Package for the Automated Retrieval and Management of Divergence Time Data from the TimeTree Resource for Downstream Analyses. *Journal of molecular evolution* **91**, 502–513.

LE CLERCQ, L.S., KOTZÉ, A., GROBLER, J.P. & DALTON, D.L. (2024*a*) Methylation-based markers for the estimation of age in African cheetah, *Acinonyx jubatus*. *Molecular Ecology Resources* **24**, e13940.

LE CLERCQ, L.S., PHETLA, V., OSINUBI, S.T., KOTZÉ, A., GROBLER, J.P. & DALTON, D.L. (2024*b*) Phenotypic correlates between clock genes and phenology among populations of Diederik cuckoo, *Chrysococcyx caprius*. *Ecology and Evolution* **14**, e70117.

LI, W.H., TANIMURA, M. & SHARP, P.M. (1987) An evaluation of the molecular clock hypothesis using mammalian DNA sequences. *Journal of Molecular Evolution* **25**, 330–342.

LIEDVOGEL, M., SZULKIN, M., KNOWLES, S.C.L., WOOD, M.J. & SHELDON, B.C. (2009) Phenotypic correlates of clock gene variation in a wild blue tit population: Evidence for a role in seasonal timing of reproduction. *Molecular Ecology* **18**, 2444–2456.

LINCOLN, G. (2019) A Brief History of Circannual Time. *Journal of Neuroendocrinology* **31**, e12694.

LUGO RAMOS, J.S., DELMORE, K.E. & LIEDVOGEL, M. (2017) Candidate genes for migration do not distinguish migratory and non-migratory birds. *Journal of Comparative Physiology A* **203**, 383–397.

MARKER, L.L. & DICKMAN, A.J. (2003) Morphology, Physical condition, and Growth of the Cheetah (*Acinonyx jubatus jubatus*). *Journal of Mammalogy* **84**, 840–850.

MAYNE, B., BERRY, O. & JARMAN, S. (2021) Optimal sample size for calibrating DNA

methylation age estimators. *Molecular Ecology Resources* **21**, 2316–2323.

MUELLER, J.C., PARTECKE, J., HATCHWELL, B.J., GASTON, K.J. & EVANS, K.L. (2013) Candidate gene polymorphisms for behavioural adaptations during urbanization in blackbirds. *Molecular Ecology* **22**, 3629–3637.

MUELLER, J.C., PULIDO, F. & KEMPENAERS, B. (2011) Identification of a gene associated with avian migratory behaviour. *Proceedings of the Royal Society B: Biological Sciences* **278**, 2848–2856.

O'CONNELL, D.P., KELLY, D.J., LAWLESS, N., KARYA, A., ANALUDDIN, K. & MARPLES, N.M. (2019) Diversification of a 'great speciator' in the Wallacea region: differing responses of closely related resident and migratory kingfisher species (Aves: *Alcedinidae*: *Todiramphus*). *Ibis* **161**, 806–823.

O'DEA, R.E., LAGISZ, M., JENNIONS, M.D., KORICHEVA, J., NOBLE, D.W.A., PARKER, T.H., GUREVITCH, J., PAGE, M.J., STEWART, G., MOHER, D. & NAKAGAWA, S. (2021) Preferred reporting items for systematic reviews and meta-analyses in ecology and evolutionary biology: a PRISMA extension. *Biological Reviews* **96**, 1695–1722.

OLSON, D.E., FITZGERALD, G., HAND, D.M., ARDREN, W.R., BAUMSTEIGER, J. & SPATEHOLTS, R. (2008) Use of Parentage Analysis to Determine Reproductive Success of Hatchery-Origin Spring Chinook Salmon Outplanted into Shitike Creek, Oregon. *North American Journal of Fisheries Management* **28**, 1472–1485.

PARODY-MERINO, Á.M., BATTLEY, P.F., CONKLIN, J.R. & FIDLER, A.E. (2019) No evidence for an association between *Clock* gene allelic variation and migration timing in a long-distance migratory shorebird (*Limosa lapponica baueri*). *Oecologia* **191**, 843–859.

PETERSON, M.P., ABOLINS-ABOLS, M., ATWELL, J.W., RICE, R.J., MILÁ, B. & KETTERSON, E.D. (2013) Variation in candidate genes CLOCK and ADCYAP1 does not consistently predict differences in migratory behavior in the songbird genus Junco. *F1000Research* **2**, 115.

RAJ, K., SZLADOVITS, B., HAGHANI, A., ZOLLER, J.A., LI, C.Z., BLACK, P., MADDOX, D., ROBECK, T.R. & HORVATH, S. (2021) Epigenetic clock and methylation studies in cats. *GeroScience* **43**, 2363–2378.

REMOT, F., RONGET, V., FROY, H., REY, B., GAILLARD, J.M., NUSSEY, D.H. & LEMAITRE, J.F. (2021) Decline in telomere length with increasing age across

nonhuman vertebrates: A meta-analysis. *Molecular Ecology* **31**, 5917–5932.

RIEKKOLA, L., ZERBINI, A.N., ANDREWS, O., ANDREWS-GOFF, V., BAKER, C.S., CHANDLER, D., CHILDERHOUSE, S., CLAPHAM, P., DODÉMONT, R., DONNELLY, D., FRIEDLAENDER, A., GALLEGO, R., GARRIGUE, C., IVASHCHENKO, Y., JARMAN, S., ET AL. (2018) Application of a multi-disciplinary approach to reveal population structure and Southern Ocean feeding grounds of humpback whales. *Ecological Indicators* **89**, 455–465.

ROBINSON, N.E. & ROBINSON, A.B. (2001) Molecular clocks. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 944–949.

ROVATSOS, M. & KRATOCHVÍL, L. (2017) Molecular sexing applicable in 4000 species of lizards and snakes? From dream to real possibility. *Methods in Ecology and Evolution* 8, 902–906.

SAINO, N., BAZZI, G., GATTI, E., CAPRIOLI, M., CECERE, J.G., POSSENTI, C.D., GALIMBERTI, A., ORIOLI, V., BANI, L., RUBOLINI, D., GIANFRANCESCHI, L. & SPINA, F. (2015) Polymorphism at the Clock gene predicts phenology of long-distance migration in birds. *Molecular Ecology* **24**, 1758–1773.

SAREISIAN, L. (2014) Global DNA methylation and persistent organic pollutants in ringed seals (*Phoca hispida*) from Svalbard and the Baltic Sea. University of Oslo.

STALMANS, M.E., MASSAD, T.J., PEEL, M.J.S., TARNITA, C.E. & PRINGLE, R.M. (2019) War-induced collapse and asymmetric recovery of large-mammal populations in Gorongosa National Park, Mozambique. *PLoS ONE* **14**, e0212864.

STEINMEYER, C., MUELLER, J.C. & KEMPENAERS, B. (2009) Search for informative polymorphisms in candidate genes: Clock genes and circadian behaviour in blue tits. *Genetica* **136**, 109–117.

SUCHIMAN, H.E.D., SLIEKER, R.C., KREMER, D., SLAGBOOM, P.E., HEIJMANS, B.T. & TOBI, E.W. (2015) Design, measurement and processing of region-specific DNA methylation assays: The mass spectrometry-based method EpiTYPER. *Frontiers in Genetics* **6**, 287.

TAKKEN BEIJERSBERGEN, L.M. (2019) Determining Age and Season of Death by Use of Incremental Lines in Norwegian Reindeer Tooth Cementum. *Environmental Archaeology* **24**, 49–60.

TARNITA, C.E., ANTAL, T., OHTSUKI, H. & NOWAK, M.A. (2009) Evolutionary dynamics in set structured populations. *Proceedings of the National Academy*

of Sciences of the United States of America **106**, 8601–8604.

THOMPSON, R.F., SUZUKI, M., LAU, K.W. & GREALLY, J.M. (2009) A pipeline for the quantitative analysis of CG dinucleotide methylation using mass spectrometry. *Bioinformatics* **25**, 2164–2170.

THORNTON, A. & LEE, P. (2000) Publication bias in meta-analysis: Its causes and consequences. *Journal of Clinical Epidemiology* **53**, 207–216.

THORPE, J.P. (1982) The molecular clock hypothesis: biochemical evolution, genetic differentiation and systematics. *Annual review of ecology and systematics. Volume 13*, 139–168.

VOELKER, G., BOWIE, R.C.K. & KLICKA, J. (2013) Gene trees, species trees and Earth history combine to shed light on the evolution of migration in a model avian system. *Molecular Ecology* **22**, 3333–3344.

WEIDNER, C.I., LIN, Q., KOCH, C.M., EISELE, L., BEIER, F., ZIEGLER, P., BAUERSCHLAG, D.O., JÖCKEL, K.-H.H., ERBEL, R., MÜHLEISEN, T.W., ZENKE, M., BRÜMMENDORF, T.H. & WAGNER, W. (2014) Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biology* **15**, 1–12.

WOODALL, P.F. (2020) Woodland Kingfisher (*Halcyon senegalensis*), version 1.0. In *Birds of the World* (eds J. DEL HOYO, A. ELLIOTT, J. SARGATAL, D. CHRISTIE & E. DE JUANA), Cornell Lab of Ornithology, Ithaca, NY, USA.

ZAPICO, S.C. & UBELAKER, D.H. (2016) Relationship Between Mitochondrial DNA Mutations and Aging. Estimation of Age-at-death. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* **71**, 445–450.

APPENDIX A: Supplementary information for Article 1

LE CLERCQ, L.-S., BAZZI, G., CECERE, J.G., GIANFRANCESCHI, L., GROBLER, J.P., KOTZÉ, A., RUBOLINI, D., LIEDVOGEL, M. & DALTON, D.L. (2023*a*) Time trees and clock genes: a systematic review and comparative analysis of contemporary avian migration genetics. *Biological Reviews* **98**, 1051–1080.

Table S1. Heterozygosity and Hardy–Weinberg equilibrium test results on *Clock* and *Adcyap1* alleles for 40 avian species. The sample size (*N*) is given along with the total number of alleles detected and the most abundant (Freq) allele for *Clock* and *Adcyap1*. Both the observed (H_o) and expected (H_e) heterozygosity are shown, with the relative relationship between them indicated by arrows. Results of a chi-squared test for departure from Hardy–Weinberg (HW) equilibrium is also shown together with its significance level. *, $P < 0.10$; **, $P < 0.05$; ***, $P < 0.02$.

Species	**Study**	***Clock***						***Adcyap1***				
		N	**Alleles**	**Freq**	H_o	H_e	**HW**	**Alleles**	**Freq**	H_o	H_e	**HW**
1(a) Tits:												
Blue tit	Johnsen *et al.* (2007)	950	6	Q_{12}	↑0.559	↓0.550	0.812	–	–	–	–	–
1(b) Warblers:												
American redstart	Authors (unpublished)	26	2	Q_9	↓0.154	↑0.208	0.149	7	161	↑0.778	↓0.759	0.545
Blackpoll warbler	Ralston *et al.* (2019)	72	4	Q_6	↑0.722	↓0.714	0.968	13	189	↓0.722	↑0.786	0.044**
Common chiffchaff	Authors (unpublished)	55	4	Q_{11}	↓0.291	↑0.426	0.007***	–	–	–	–	–
Common whitethroat	Bazzi *et al.* (2016*a*)	25	5	Q_{10}	↓0.520	↑0.680	<0.001***	13	172	↓0.769	↑0.879	<0.001***
Common yellowthroat	Authors (unpublished)	31	4	Q_{10}	↑0.484	↓0.445	0.870	9	159	↓0.742	↑0.768	0.940
Eastern subalpine warbler	Bazzi *et al.* (2016*a*)	31	6	Q_9	↑0.774	↓0.758	0.490	7	168	↑0.773	↓0.664	0.987
Eurasian blackcap	Mueller *et al.* (2011)	936	–	–	–	–	–	11	161	↑0.700	↓0.699	0.336
Eurasian reed warbler	Bazzi *et al.* (2016*a*)	24	1	Q_{11}	–	–	–	10	169	↑0.875	↓0.813	0.996
Garden warbler	Bazzi *et al.* (2016*a*)	31	6	Q_{12}	↑0.839	↓0.718	0.904	6	169	↓0.710	↑0.721	0.189
Great reed warbler	Bazzi *et al.* (2016*a*)	20	2	Q_{12}	0.050	0.050	1.000	2	163	↑0.191	↓0.177	0.679
Icterine warbler	Bazzi *et al.* (2016*a*)	29	2	Q_8	↑0.069	↓0.068	0.893	7	169	↓0.517	↑0.662	0.497
Magnolia warbler	Authors (unpublished)	33	5	Q_9	↓0.273	↑0.326	0.001***	8	159	↑0.914	↓0.835	0.751
Sedge warbler	Bazzi *et al.* (2016*a*)	30	1	Q_{11}	–	–	–	4	163	↑0.433	↓0.397	0.991
Willow warbler	Bazzi *et al.* (2016*a*)	495	5	Q_9	↓0.366	↑0.397	0.025**	10	174	↑0.839	↓0.825	<0.001***
Wilson's warbler	Bazzi *et al.* (2016*b*)	102	2	Q_9	0.020	0.020	0.944	9	158	↓0.775	↑0.814	0.002***

Wood warbler	Bazzi *et al.* (2016*a*)	30	5	Q_{11}	↑0.633	↓0.628	0.970	5	162	↑0.552	↓0.531	0.734
1(c) Swallows:												
Barn swallow	Bazzi *et al.* (2015)	64	3	Q_{7}	↑0.094	↓0.091	0.988	–	–	–	–	–
Tree swallow	Bourret & Garant (2015)	921	4	Q_{8}	↓0.503	↑0.507	0.680	25	173	↓0.902	↑0.907	0.954
1(e) Sparrow, juncos, and buntings												
Dark-eyed junco	Peterson *et al.* (2013)	36	4	Q_{11}	↑0.417	↓0.349	0.890	7	161	↑0.800	↓0.797	0.210
White-throated sparrow	Authors (unpublished)	32	1	Q_{10}	–	–	–	10	161	↑0.781	↓0.775	0.853
1(f) Cardinals:												
Painted bunting	Contina *et al.* (2018)	60	6	Q_{11}	↓0.583	↑0.601	0.986	7	169	↑0.817	↓0.709	0.801
1(g) Flycatchers and chats:												
Collared flycatcher	Krist *et al.* (2021)	406	4	Q_{12}	↑0.328	↓0.325	0.306	6	182	↑0.595	↓0.581	0.558
Common nightingale	Bazzi *et al.* (2016*a*)	150	5	Q_{12}	↓0.527	↑0.564	0.006***	7	163	↓0.440	↑0.474	0.637
Common redstart	Bazzi *et al*. (2016*a*)	43	6	Q_{14}	↓0.512	↑0.524	0.907	13	169	↓0.771	↑0.804	<0.001***
European pied flycatcher	Bazzi *et al.* (2016*a*)	226	5	Q_{12}	↑0.478	↓0.464	0.716	11	180	↓0.695	↑0.709	1.000
Northern wheatear	Bazzi *et al.* (2016*a*)	30	4	Q_{12}	↓0.500	↑0.540	0.982	6	167	↓0.567	↑0.577	0.863
Spotted flycatcher	Bazzi *et al.* (2016*a*)	29	2	Q_{9}	↑0.103	↓0.100	0.812	5	162	↑0.690	↓0.683	0.660
Whinchat	Bazzi *et al.* (2016*a*)	208	7	Q_{14}	↑0.125	↑0.120	1.000	13	169	↑0.787	↓0.769	0.947
1(h) Pipits:												
Tree pipit	Bazzi *et al.* (2016*a*)	153	5	Q_{9}	↓0.242	↑0.256	0.101	12	170	↓0.763	↑0.802	0.137
1(i) Thrushes:												
Eurasian blackbird	Mueller *et al.* (2013*b*)	792	2	Q_{7}	↓0.028	↑0.030	0.045	25	165	↓0.782	↑0.820	0.846
Hermit thrush	Authors (unpublished)	30	3	Q_{10}	↓0.233	↑0.244	0.455	7	165	↑0.833	↓0.794	0.871
Swainson's thrush	Authors (unpublished)	29	2	Q_{10}	↑0.276	↓0.242	0.424	11	171	↓0.828	↑0.846	0.324
1(j) Shrikes and orioles:												
Eurasian golden oriole	Bazzi *et al.* (2016*a*)	30	2	Q_{6}	↑0.067	↓0.066	0.895	7	163	↓0.633	↑0.636	0.815
Woodchat shrike	Bazzi *et al.* (2016*a*)	20	4	Q_{7}	↑0.550	↓0.537	0.949	8	176	↑0.722	↓0.649	0.998

2(b) Hoopoes:												
Eurasian hoopoe	Bazzi *et al.* (2016*a*)	25	3	Q_8	↑0.560	↓0.496	0.687	3	157	↓0.500	↑0.559	0.771
2(c) Wrynecks:												
Eurasian wryneck	Bazzi *et al.* (2016*a*)	30	4	Q_8	↓0.433	↑0.438	0.976	5	135	↑0.710	↓0.684	0.856
2(d) Bee-eaters:												
European bee-eater	Bazzi *et al.* (2016*a*)	35	1	Q_4	–	–	–	6	163	↑0.194	↓0.185	1.000
2(e) Nightjars:												
European nightjar	Bazzi *et al.* (2016*a*)	39	2	Q_8	↓0.333	↑0.345	0.824	9	154	↓0.744	↑0.806	<0.001***
2(f) Doves:												
European turtle dove	Bazzi *et al.* (2016*a*)	29	2	Q_7	0.035	0.035	1.000	5	150	↓0.241	↑0.256	0.577

Table S2. Fixation index and Ewens–Watterson test results on *Clock* and *Adcyap1* alleles for 40 avian species. For each species the fixation index F_{IS}) is reported together with *F*-values from the Ewens–Watterson (EW) test for neutrality or selection for *Clock* and *Adcyap1*. Both the observed *F*-value and upper and lower bounds of its 95% confidence interval CI), estimated from 10,000 random samplings, are shown. *P*-values are from Slatkin's implementation of the EW test and show departure from neutrality. *, $P < 0.10$; **, $P < 0.05$; ***, $P < 0.02$.

Species	**Study**	***Clock***					***Adcyap1***				
		F_{IS}	***F*-value**	**Lower 95% CI**	**Upper 95% CI**	***P***	**F_{IS}**	***F*-value**	**Lower 95% CI**	**Upper 95% CI**	***P***
1(a) Tits:											
Blue tit	Johnsen *et al.* (2007)	–0.017	0.450	0.281	0.955	0.307	–	–	–	–	–
1(b) Warblers:											
American redstart	Authors (unpublished)	0.246	0.796	0.501	0.962	0.426	-0.044	0.255	0.191	0.596	0.264
Blackpoll warbler	Ralston *et al.* (2019)	–0.018	0.291	0.319	0.932	0.009***	0.075	0.219	0.128	0.443	0.562
Common chiffchaff	Authors (unpublished)	0.311	0.578	0.310	0.912	0.561	–	–	–	–	–
Common whitethroat	Bazzi *et al.* (2016*a*)	0.220	0.334	0.252	0.778	0.341	0.108	0.138	0.109	0.297	0.291
Common yellowthroat	Authors (unpublished)	–0.106	0.562	0.302	0.877	0.614	0.019	0.244	0.156	0.499	0.479
Eastern subalpine warbler	Bazzi *et al.* (2016*a*)	–0.038	0.254	0.222	0.710	0.093*	–0.191	0.351	0.188	0.576	0.712
Eurasian blackcap	Mueller *et al.* (2011)	–	–	–	–	–	–0.001	0.301	0.188	0.772	0.320
Eurasian reed warbler	Bazzi *et al.* (2016*a*)	–	1.000	–	–	–	–0.099	0.204	0.138	0.409	0.478
Garden warbler	Bazzi *et al.* (2016*a*)	–0.187	0.293	0.221	0.711	0.236	–0.001	0.291	0.222	0.710	0.225
Great reed warbler	Bazzi *et al.* (2016*a*)	–0.026	0.951	0.501	0.951	1.000***	–0.105	0.828	0.501	0.954	0.490
Icterine warbler	Bazzi *et al.* (2016*a*)	–0.036	0.933	0.502	0.966	0.783	0.206	0.349	0.194	0.615	0.629
Magnolia warbler	Authors (unpublished)	0.150	0.679	0.257	0.803	0.887	–0.111	0.177	0.177	0.585	0.023**
Sedge warbler	Bazzi *et al.* (2016*a*)	–	1.000	–	–	–	–0.111	0.610	0.301	0.873	0.684
Willow warbler	Bazzi *et al.* (2016*a*)	0.078	0.604	0.302	0.964	0.543	–0.021	0.178	0.168	0.626	0.045**
Wilson's warbler	Bazzi *et al.* (2016*b*)	–0.010	0.981	0.503	0.990	0.831	0.044	0.190	0.182	0.666	0.040**
Wood warbler	Bazzi *et al.* (2016*a*)	–0.026	0.383	0.254	0.786	0.372	–0.058	0.479	0.253	0.778	0.657
1(c) Swallows:											
Barn swallow	Bazzi *et al.* (2015)	-0.042	0.910	0.373	0.969	0.855	–	–	–	–	–
Tree swallow	Bourret & Garant (2015)	0.008	0.493	0.351	0.989	0.186	0.005	0.094	0.103	0.390	0.008***

1(e) Sparrows, Juncos, and Buntings:											
Dark–eyed junco	Peterson *et al.* (2013)	–0.209	0.656	0.307	0.893	0.729	–0.018	0.214	0.198	0.650	0.060*
White–throated sparrow	Authors (unpublished)	–	1.000	–	–	–	–0.024	0.237	0.144	0.448	0.573
1(f) Cardinals:											
Painted bunting	Contina *et al.* (2018)	0.021	0.404	0.233	0.798	0.490	–0.162	0.297	0.209	0.728	0.285
1(g) Flycatchers and Chats:											
Collared flycatcher	Krist *et al.* (2021)	–0.010	0.676	0.338	0.980	0.518	–0.025	0.420	0.269	0.926	0.301
Common nightingale	Bazzi *et al.* (2016*a*)	0.063	0.438	0.285	0.922	0.307	0.068	0.528	0.219	0.770	0.791
Common redstart	Bazzi *et al.* (2016*a*)	0.013	0.482	0.228	0.748	0.729	0.026	0.208	0.114	0.338	0.737
European pied flycatcher	Bazzi *et al.* (2016*a*)	–0.033	0.537	0.290	0.939	0.502	0.015	0.295	0.157	0.560	0.637
Northern wheatear	Bazzi *et al.* (2016*a*)	0.058	0.469	0.299	0.873	0.420	0.002	0.432	0.221	0.703	0.701
Spotted flycatcher	Bazzi *et al.* (2016*a*)	–0.055	0.902	0.502	0.966	0.591	–0.028	0.329	0.251	0.778	0.204
Whinchat	Bazzi *et al.* (2016*a*)	–0.041	0.880	0.230	0.846	0.985***	–0.026	0.233	0.144	0.544	0.428
1(h) Pipits:											
Tree pipit	Bazzi *et al.* (2016*a*)	0.051	0.745	0.287	0.924	0.809	0.044	0.202	0.143	0.526	0.294
1(i) Thrushes:											
Eurasian blackbird	Mueller *et al.* (2013*b*)	0.069	0.970	0.505	0.999	0.525	0.046	0.181	0.102	0.396	0.544
Hermit thrush	Authors (unpublished)	0.026	0.761	0.361	0.935	0.685	–0.067	0.219	0.195	0.633	0.087*
Swainson's thrush	Authors (unpublished)	–0.160	0.762	0.502	0.966	0.375	0.005	0.168	0.129	0.386	0.273
1(j) Shrikes and Orioles:											
Eurasian golden oriole	Bazzi *et al.* (2016*a*)	–0.035	0.936	0.502	0.967	0.784	–0.013	0.375	0.194	0.629	0.695
Woodchat shrike	Bazzi *et al.* (2016*a*)	–0.050	0.476	0.295	0.814	0.521	–0.144	0.369	0.162	0.463	0.902*
2(b) Hoopoes:											
Eurasian hoopoe	Bazzi *et al.* (2016*a*)	–0.151	0.514	0.360	0.922	0.319	0.088	0.452	0.359	0.925	0.153
2(c) Wrynecks:											
Eurasian wryneck	Bazzi *et al.* (2016*a*)	–0.007	0.569	0.299	0.873	0.627	–0.055	0.327	0.255	0.791	0.190
2(d) Bee-eaters:											
European bee–eater	Bazzi *et al.* (2016*a*)	–	1.000	–	–	–	–0.063	0.817	0.225	0.726	0.996***
2(e) Nightjars:											
European nightjar	Bazzi *et al.* (2016*a*)	0.022	0.659	0.501	0.975	0.252	0.066	0.204	0.163	0.528	0.191

2(f) Doves:											
European turtle dove	Bazzi *et al.* (2016*a*)	–0.018	0.966	0.502	0.966	0.783	0.040	0.749	0.252	0.779	0.946*

Table S3. Linkage disequilibrium test results comparing *Clock versus Adcyap1* alleles for – 36 avian species. Two measures of linkage disequilibrium are reported, *D'* and W_n, along with the calculated *S*–statistic and significance. *, $P < 0.10$; **, $P < 0.05$; ***, $P < 0.02$.

Species	Study	*D'*	W_n	*S*	*P*
1(b) Warblers:					
American redstart	Authors (unpublished)	0.288	0.223	2.320	0.935
Blackpoll warbler	Ralston *et al*. (2019)	0.221	0.316	26.350	0.927
Common whitethroat	Bazzi *et al*. (2016*a*)	0.617	0.741	47.060	0.026**
Common yellowthroat	Authors (unpublished)	0.394	0.391	18.990	0.487
Eastern subalpine warbler	Bazzi *et al*. (2016*a*)	0.616	0.453	18.490	0.722
Eurasian hoopoe	Bazzi *et al*. (2016*a*)	0.560	0.290	6.130	0.120
Eurasian reed warbler	Bazzi *et al*. (2016*a*)	–	–	–	–
Garden warbler	Bazzi *et al*. (2016*a*)	0.381	0.455	23.430	0.497
Great reed warbler	Bazzi *et al*. (2016*a*)	1.000	0.053	0.210	0.192
Icterine warbler	Bazzi *et al*. (2016*a*)	0.945	0.556	6.990	0.153
Magnolia warbler	Authors (unpublished)	0.436	0.479	24.060	0.269
Sedge warbler	Bazzi *et al*. (2016*a*)	–	–	–	–
Willow warbler	Bazzi *et al*. (2016*a*)	0.290	0.319	37.790	0.308
Wilson's warbler	Bazzi *et al*. (2016*b*)	1.000	0.176	4.260	0.370
Wood warbler	Bazzi *et al*. (2016*a*)	0.403	0.280	12.960	0.514
1(c) Swallows:					
Tree swallow	Bourret & Garant (2015)	0.116	0.159	64.420	0.580
1(e) Sparrows, juncos, and buntings:					
Dark–eyed junco	Peterson *et al*. (2013)	0.476	0.501	16.500	0.279
White–throated sparrow	Authors (unpublished)	–	–	–	–
1(f) Cardinals:					
Painted bunting	Contina *et al*. (2018)	0.345	0.281	24.550	0.364
1(g) Flycatchers and chats:					
Collared flycatcher	Authors (unpublished)	0.143	0.100	12.780	0.517
Common nightingale	Bazzi *et al*. (2016*a*)	0.259	0.218	17.560	0.403
Common redstart	Bazzi *et al*. 2016*a*)	0.458	0.417	34.740	0.059*
European pied flycatcher	Bazzi *et al*. (2016*a*)	0.309	0.258	27.040	0.444
Northern wheatear	Bazzi *et al*. (2016*a*)	0.258	0.381	11.510	0.506
Spotted flycatcher	Bazzi *et al*. (2016*a*)	1.000	0.242	3.160	0.334
Whinchat	Bazzi *et al*. (2016*a*)	0.391	0.228	36.950	0.538
1(h) Pipits:					
Tree pipit	Bazzi *et al*. (2016*a*)	0.384	0.280	20.410	0.691
1(i) Thrushes:					
Eurasian blackbird	Mueller *et al*. (2013*b*)	0.291	0.179	12.870	0.625
Hermit thrush	Authors (unpublished)	0.550	0.543	16.240	0.074*
Swainson's thrush	Authors (unpublished)	0.555	0.415	5.040	0.969
1(j) Shrikes and orioles:					
Eurasian golden oriole	Bazzi *et al*. (2016*a*)	0.964	0.782	11.650	0.004***
Woodchat shrike	Bazzi *et al*. (2016*a*)	0.466	0.556	16.480	0.686
2(c) Wrynecks:					
Eurasian wryneck	Bazzi *et al*. (2016*a*)	0.437	0.384	19.020	0.025**

2(d) Bee-eaters:					
European bee–eater	Bazzi *et al*. (2016*a*)	–	–	–	–
2(e) Nightjars:					
European nightjar	Bazzi *et al*. (2016*a*)	0.427	0.486	9.600	0.396
2(f) Doves:					
European turtle dove	Bazzi *et al*. (2016*a*)	1.000	0.053	0.300	0.237

LE CLERCQ, L., KOTZÉ, A., GROBLER, J.P. & DALTON, D.L. (2023*c*) PAReTT: A Python Package for the Automated Retrieval and Management of Divergence Time Data from the TimeTree Resource for Downstream Analyses. *Journal of molecular evolution* **91**, 502–513.

Table 1: Table of memory consumption and speed profiling of individual functions as benchmarking.

```
Line #    Mem usage    Increment  Occurrences   Line Contents
=============================================================
    22   99.684 MiB  295.957 MiB           3   @profile
    23                                         def menu_choice():
    24                                             """ Find out what the user wants to do next. """
    25   99.684 MiB    0.000 MiB           3       print("MAIN MENU:")
    26   99.684 MiB    0.000 MiB           3       print("---------------------------------------")
    27   99.684 MiB    0.000 MiB           3       print("Choose one of the following options?")
    28   99.684 MiB    0.000 MiB           3       print("   *) Check data availability")
    29   99.684 MiB    0.000 MiB           3       print("   a) Get Divergence Times (pair)")
    30   99.684 MiB    0.000 MiB           3       print("   b) Get Divergence Times (batch)")
    31   99.684 MiB    0.000 MiB           3       print("   c) Get Evolutionary Timeline")
    32   99.684 MiB    0.000 MiB           3       print("   d) Build a Time Tree")
    33   99.684 MiB    0.000 MiB           3       print("   e) Print citation")
    34   99.684 MiB    0.000 MiB           3       print("   f) Validate datafile")
    35   99.684 MiB    0.000 MiB           3       print("   q) Quit")
    36   99.684 MiB    0.000 MiB           3       print("---------------------------------------")
    37   99.684 MiB   -0.008 MiB           3       choice = input("Choice: ")
    38   99.684 MiB    0.000 MiB           3       print()
    39   99.684 MiB    0.000 MiB           3       if choice.lower() in ['*','a','b','c','d','e','f','q']:
    40   99.684 MiB    0.000 MiB           3           return choice.lower()
    41                                             else:
    42                                                 print(choice +"?")
    43                                                 print("Invalid option")
    44                                                 print()
    45                                                 Time.sleep(3)
    46                                                 return None
    47   98.680 MiB  196.270 MiB           2   @profile
    48                                         def check_avail_menu():
    49                                             """Menu options for checking data availability function"""
    50   98.680 MiB    0.000 MiB           2       print('AVAILABILITY MENU:')
    51   98.680 MiB    0.000 MiB           2       print("---------------------------------------")
    52   98.680 MiB    0.000 MiB           2       print("Choose one of the following options?")
    53   98.680 MiB    0.000 MiB           2       print("     i) Individual")
    54   98.680 MiB    0.000 MiB           2       print("     l) List")
    55   98.680 MiB    0.000 MiB           2       print("     m) Main menu")
```

```
56   98.680 MiB    0.000 MiB           2       print("----------------------------------------")
57   98.680 MiB    0.000 MiB           2       choice = input("Choice: ")
58   98.680 MiB    0.000 MiB           2       print()
59   98.680 MiB    0.000 MiB           2       if choice.lower() in ['i','l','m']:
60   98.680 MiB    0.000 MiB           2           return choice.lower()
61                                             else:
62                                                 print(choice +"?")
63                                                 print("Invalid option")
64                                                 Time.sleep(3)
65                                                 return None
66   98.680 MiB  196.270 MiB           2   @profile
67                                             def check_available():
68                                                 """Checks the time tree website to confirm if data is available for a species"""
69   98.680 MiB  196.270 MiB           2       choice = check_avail_menu()
70   98.680 MiB    0.000 MiB           2       if choice == "i":
71   97.590 MiB    0.000 MiB           1           name = input("Name: ")
72   98.574 MiB    0.984 MiB           1           with Browser('firefox', headless=True) as browser:
73   98.629 MiB    0.055 MiB           1               browser.visit('http://timetree.org')
74   98.723 MiB    0.094 MiB           1               browser.fill('timeline_taxon', name)
75   98.723 MiB    0.000 MiB           1               browser.find_by_id('timeline-search-button1').click()
76   98.723 MiB    0.000 MiB           1               check = browser.is_text_present('Evolutionary Timeline for',
                                                         wait_time=15)
77   98.723 MiB    0.000 MiB           1               if check is True:
78   98.723 MiB    0.000 MiB           1                   value = 'Available'
79   98.723 MiB    0.000 MiB           1                   time = tm.datetime.now()
80   98.723 MiB    0.000 MiB           1                   print('{:<25}'.format(name), end="")
81   98.723 MiB    0.000 MiB           1                   print('{:<25}'.format(value), end="")
82   98.723 MiB    0.000 MiB           1                   print(time)
83                                                     elif check is False:
84                                                         value = 'Not Available'
85                                                         time = tm.datetime.now()
86                                                         print('{:<25}'.format(name), end="")
87                                                         print('{:<25}'.format(value), end="")
88                                                         print(time)
89   98.723 MiB    0.000 MiB           1               print("Done!")
90   98.738 MiB    0.016 MiB           1               print()
91   98.680 MiB    0.000 MiB           1       elif choice == 'l':
92   99.031 MiB    0.352 MiB           1           data = pd.DataFrame(columns = ['Species', 'TimeTree.Data'])
```

```
 93   99.031 MiB    0.000 MiB           1           infile = input("Input file in .txt format: ")
 94   99.031 MiB    0.000 MiB           1           if infile.endswith('.txt') is True:
 95   99.031 MiB    0.000 MiB           1               pass
 96                                                 elif infile.endswith('.txt') is False:
 97                                                     infile = infile + '.txt'
 98   99.031 MiB    0.000 MiB           1           try:
 99   99.031 MiB    0.000 MiB           1               my_file = open(infile, "r")
100                                                 except FileNotFoundError:
101                                                     print("File not found")
102                                                     return
103                                                 else:
104   99.039 MiB    0.008 MiB           1               my_file = open(infile, "r")
105   99.039 MiB    0.000 MiB           1           content = my_file.read()
106   99.039 MiB    0.000 MiB           1           content_list = content.split("\n")
107   99.039 MiB    0.000 MiB           1           my_file.close()
108   99.039 MiB    0.000 MiB           1           converted_list =[]
109   99.039 MiB    0.000 MiB           4           for element in content_list:
110   99.039 MiB    0.000 MiB           3               converted_list.append(element.strip())
111   99.039 MiB    0.000 MiB           1           species = tuple(converted_list)
112   99.582 MiB    0.000 MiB           4           for i in range(0,len(species)):
113   99.582 MiB    0.000 MiB           3               name = species[i]
114   99.605 MiB    0.293 MiB           3               with Browser('firefox', headless=True) as browser:
115   99.582 MiB   -0.047 MiB           3                   browser.visit('http://timetree.org')
116   99.582 MiB    0.012 MiB           3                   browser.fill('timeline_taxon', name)
117   99.582 MiB    0.000 MiB           3                   browser.find_by_id('timeline-search-button1').click()
118   99.582 MiB    0.000 MiB           3                   browser.is_text_present('Evolutionary Timeline for', wait_time=120)
119   99.582 MiB    0.000 MiB           3                   check = browser.is_text_present('Evolutionary Timeline for',
                                                                 wait_time=15)
120   99.582 MiB    0.000 MiB           3                   if check is True:
121   99.582 MiB    0.000 MiB           3                       value = 'Available'
122   99.582 MiB    0.000 MiB           3                       time = tm.datetime.now()
123   99.582 MiB    0.000 MiB           3                       print('{:<25}'.format(name), end="")
124   99.582 MiB    0.000 MiB           3                       print('{:<25}'.format(value), end="")
125   99.582 MiB    0.000 MiB           3                       print(time)
126   99.582 MiB    0.262 MiB           6                       data = data.append({'Species': name, 'TimeTree.Data':
                                                                                      'Available'},
127   99.582 MiB    0.000 MiB           3                                          ignore_index=True)
128                                                         elif check is False:
```

```
129                                                     value = 'Not Available'
130                                                     time = tm.datetime.now()
131                                                     print('{:<25}'.format(name), end="")
132                                                     print('{:<25}'.format(value), end="")
133                                                     print(time)
134                                                     print("Done!")
135                                                     data = data.append({'Species': name, 'TimeTree.Data': 'Not
                                                                        Available'},
136                                                                         ignore_index=True)
137   99.582 MiB    0.000 MiB           1           output_name = input("File name (.csv): ")
138   99.582 MiB    0.000 MiB           1           if output_name.endswith('.csv') is True:
139   99.582 MiB    0.000 MiB           1               pass
140                                                   elif output_name.endswith('.csv') is False:
141                                                       output_name = output_name + '.csv'
142   99.684 MiB    0.102 MiB           1           data.to_csv(output_name, index=False)
143   99.684 MiB    0.000 MiB           1           print("Done!")
144   99.684 MiB    0.000 MiB           1           print()
145   97.457 MiB   97.457 MiB           1   @profile
146                                              def div_times_sing():
147                                                  """Get the times for a single pair specified as input"""
148   97.457 MiB    0.000 MiB           1       taxon_a = input("Taxon a: ")
149   97.484 MiB    0.027 MiB           1       taxon_b = input("Taxon b: ")
150   98.711 MiB    1.227 MiB           1       with Browser('firefox', headless=True) as browser:
151   98.715 MiB    0.004 MiB           1           browser.visit('http://timetree.org')
152   98.832 MiB    0.117 MiB           1           browser.fill('taxon_a', taxon_a)
153   98.832 MiB    0.000 MiB           1           browser.fill('taxon_b', taxon_b)
154   98.832 MiB    0.000 MiB           1           browser.find_by_id('pairwise-search-button1').click()
155   98.793 MiB   -0.039 MiB           1           browser.is_text_present('Median Time', wait_time=320)
156   98.793 MiB    0.000 MiB           1           var_y = '#pairwise-results > text:nth-child(11)'
157   98.793 MiB    0.000 MiB           1           divtime = browser.find_by_css(var_y).first.value
158   98.793 MiB    0.000 MiB           1           divtime_2 = divtime.replace(' MYA', '')
159   98.793 MiB    0.000 MiB           1           time = tm.datetime.now()
160   98.793 MiB    0.000 MiB           1           print('{:<25}'.format(taxon_a), end="")
161   98.793 MiB    0.000 MiB           1           print('{:<25}'.format(taxon_b), end="")
162   98.793 MiB    0.000 MiB           1           print('{:<5}'.format(divtime_2), end="")
163   98.793 MiB    0.000 MiB           1           print(time)
164   98.809 MiB    0.016 MiB           1           print()
165   98.809 MiB   98.809 MiB           1   @profile
```

```
166                                           def div_times_batch():
167                                               """Get the divergence times for a list of species and stores it as a data-
168                                               frame which gets exported as a .csv with 3 columns"""
169   99.164 MiB    0.355 MiB           1       data = pd.DataFrame(columns = ['Taxa1', 'Taxa2', 'Div.Time'])
170   99.164 MiB    0.000 MiB           1       infile = input("Input file in .txt format: ")
171   99.164 MiB    0.000 MiB           1       if infile.endswith('.txt') is True:
172   99.164 MiB    0.000 MiB           1           pass
173                                               elif infile.endswith('.txt') is False:
174                                                   infile = infile + '.txt'
175   99.164 MiB    0.000 MiB           1       try:
176   99.164 MiB    0.000 MiB           1           my_file = open(infile, "r")
177                                               except FileNotFoundError:
178                                                   print("File not found")
179                                                   return
180                                               else:
181   99.172 MiB    0.008 MiB           1           my_file = open(infile, "r")
182   99.172 MiB    0.000 MiB           1       content = my_file.read()
183   99.172 MiB    0.000 MiB           1       content_list = content.split("\n")
184   99.172 MiB    0.000 MiB           1       my_file.close()
185   99.172 MiB    0.000 MiB           1       converted_list =[]
186   99.172 MiB    0.000 MiB           4       for element in content_list:
187   99.172 MiB    0.000 MiB           3           converted_list.append(element.strip())
188   99.172 MiB    0.000 MiB           1       species = tuple(converted_list)
189   99.172 MiB    0.000 MiB           1       length = len(species)
190   99.625 MiB   -0.004 MiB           4       for variable_x in range(0,length):
191   99.625 MiB   -0.008 MiB          12           for i in range(0,length):
192   99.625 MiB   -0.020 MiB           9               taxon_a = species[variable_x]
193   99.625 MiB   -0.020 MiB           9               taxon_b = species[i]
194   99.625 MiB   -0.020 MiB           9               if taxon_b == species[-1]:
195   99.625 MiB   -0.008 MiB           3                   if taxon_a == taxon_b:
196   99.625 MiB    0.000 MiB           1                       time = tm.datetime.now()
197   99.625 MiB    0.000 MiB           1                       print('{:<25}'.format(taxon_a), end="")
198   99.625 MiB    0.000 MiB           1                       print('{:<25}'.format(taxon_b), end="")
199   99.625 MiB    0.000 MiB           1                       print('{:<5}'.format('0'), end="")
200   99.625 MiB    0.000 MiB           1                       print(time)
201   99.625 MiB    0.000 MiB           2                       data = data.append({'Taxa1' : taxon_a, 'Taxa2' : taxon_b,
                                                                  'Div.Time' : '0'},
202   99.625 MiB    0.000 MiB           1                       ignore_index = True)
```

```
203                                                                         else:
204   99.574 MiB    0.059 MiB           2                                     with Browser('firefox', headless=True) as browser:
205   99.574 MiB    0.000 MiB           2                                         try:
206   99.578 MiB    0.004 MiB           2                                             browser.visit('http://timetree.org')
207   99.582 MiB    0.004 MiB           2                                             browser.fill('taxon_a', taxon_a)
208   99.582 MiB    0.000 MiB           2                                             browser.fill('taxon_b', taxon_b)
209   99.582 MiB    0.000 MiB           2                                             browser.find_by_id('pairwise-search-button1').click()
210                                                                             except:
211                                                                                 Time.sleep(600)
212                                                                                 browser.visit('http://timetree.org')
213                                                                                 browser.fill('taxon_a', taxon_a)
214                                                                                 browser.fill('taxon_b', taxon_b)
215                                                                                 browser.find_by_id('pairwise-search-button1').click()
216   99.574 MiB   -0.008 MiB           2                                         browser.is_text_present('Median Time', wait_time=320)
217   99.574 MiB    0.000 MiB           2                                         var_y = '#pairwise-results > text:nth-child(11)'
218   99.574 MiB    0.000 MiB           2                                         try:
219   99.578 MiB    0.008 MiB           2                                             divtime = browser.find_by_css(var_y).first.value
220                                                                             except:
221                                                                                 divtime = 'NA MYA'
222                                                                             else:
223   99.578 MiB   -0.004 MiB           2                                             divtime = browser.find_by_css(var_y).first.value
224   99.578 MiB   -0.004 MiB           2                                         divtime_2 = divtime.replace(' MYA', '')
225   99.578 MiB   -0.004 MiB           2                                         time = tm.datetime.now()
226   99.578 MiB   -0.004 MiB           2                                         print('{:<25}'.format(taxon_a), end="")
227   99.578 MiB   -0.004 MiB           2                                         print('{:<25}'.format(taxon_b), end="")
228   99.578 MiB   -0.004 MiB           2                                         print('{:<5}'.format(divtime_2), end="")
229   99.578 MiB   -0.004 MiB           2                                         print(time)
230   99.582 MiB   -0.004 MiB           6                                         data = data.append({'Taxa1' : taxon_a, 'Taxa2' : taxon_b,
231   99.578 MiB   -0.016 MiB           4                                                             'Div.Time' : divtime_2}, ignore_index = True)
232   99.625 MiB   -0.004 MiB           3                                     print(taxon_a,'DONE!')
233                                                                         else:
234   99.621 MiB   -0.012 MiB           6                                     if taxon_a == taxon_b:
235   99.574 MiB   -0.008 MiB           2                                         time = tm.datetime.now()
236   99.574 MiB    0.000 MiB           2                                         print('{:<25}'.format(taxon_a), end="")
237   99.574 MiB    0.000 MiB           2                                         print('{:<25}'.format(taxon_b), end="")
238   99.574 MiB    0.000 MiB           2                                         print('{:<5}'.format('0'), end="")
239   99.574 MiB    0.000 MiB           2                                         print(time)
240   99.574 MiB    0.180 MiB           4                                         data = data.append({'Taxa1' : taxon_a, 'Taxa2' : taxon_b,
```

```
                                                                                    'Div.Time' : '0'},
241   99.574 MiB    0.000 MiB           2                                       ignore_index = True)
242                                                                         else:
243   99.621 MiB    0.082 MiB           4                                       with Browser('firefox', headless=True) as browser:
244   99.621 MiB    0.000 MiB           4                                           try:
245   99.621 MiB    0.000 MiB           4                                               browser.visit('http://timetree.org')
246   99.621 MiB    0.098 MiB           4                                               browser.fill('taxon_a', taxon_a)
247   99.621 MiB    0.004 MiB           4                                               browser.fill('taxon_b', taxon_b)
248   99.621 MiB    0.004 MiB           4                                               browser.find_by_id('pairwise-search-button1').click()
249                                                                                   except:
250                                                                                       Time.sleep(600)
251                                                                                       browser.visit('http://timetree.org')
252                                                                                       browser.fill('taxon_a', taxon_a)
253                                                                                       browser.fill('taxon_b', taxon_b)
254                                                                                       browser.find_by_id('pairwise-search-button1').click()
255   99.621 MiB    0.004 MiB           4                                           browser.is_text_present('Median Time', wait_time=320)
256   99.621 MiB    0.000 MiB           4                                           var_y = '#pairwise-results > text:nth-child(11)'
257   99.621 MiB    0.000 MiB           4                                           try:
258   99.621 MiB    0.000 MiB           4                                               divtime = browser.find_by_css(var_y).first.value
259                                                                                   except:
260                                                                                       divtime = 'NA MYA'
261                                                                                   else:
262   99.621 MiB    0.000 MiB           4                                               divtime = browser.find_by_css(var_y).first.value
263   99.621 MiB    0.000 MiB           4                                           divtime_2 = divtime.replace(' MYA', '')
264   99.621 MiB    0.000 MiB           4                                           time = tm.datetime.now()
265   99.621 MiB    0.000 MiB           4                                           print('{:<25}'.format(taxon_a), end="")
266   99.621 MiB    0.000 MiB           4                                           print('{:<25}'.format(taxon_b), end="")
267   99.621 MiB    0.000 MiB           4                                           print('{:<5}'.format(divtime_2), end="")
268   99.621 MiB    0.000 MiB           4                                           print(time)
269   99.625 MiB    0.004 MiB          12                                           data = data.append({'Taxa1' : taxon_a, 'Taxa2' : taxon_b,
270   99.621 MiB    0.000 MiB           8                                                               'Div.Time' : divtime_2}, ignore_index = True)
271   99.625 MiB    0.000 MiB           1       if i == len(species):
272                                                   print(taxon_a," DONE!")
273   99.645 MiB    0.020 MiB           1       data.replace(to_replace=[r"\\t|\\n|\\r", "\t|\n|\r"], value=["",""], regex=True,
                                                inplace=True)
274   99.645 MiB    0.000 MiB           1       print()
275   99.832 MiB    0.188 MiB           1       print(data)
276   99.832 MiB    0.000 MiB           1       print()
```

```
277   99.832 MiB    0.000 MiB           1       output_name = input("File name (.csv): ")
278   99.832 MiB    0.000 MiB           1       if output_name.endswith('.csv') is True:
279   99.832 MiB    0.000 MiB           1           pass
280                                               elif output_name.endswith('.csv') is False:
281                                                   output_name = output_name + '.csv'
282   99.961 MiB    0.129 MiB           1       data.to_csv(output_name, index=False)
283   99.961 MiB    0.000 MiB           1       print("Done!")
284   99.961 MiB    0.000 MiB           1       print()
285   98.828 MiB  196.559 MiB           2   @profile
286                                           def timeline_menu():
287                                               """Menu options for timeline function"""
288   98.828 MiB    0.000 MiB           2       print('TIMELINE MENU:')
289   98.828 MiB    0.000 MiB           2       print("----------------------------------------")
290   98.828 MiB    0.000 MiB           2       print("Choose one of the following options?")
291   98.828 MiB    0.000 MiB           2       print("     i) Individual")
292   98.828 MiB    0.000 MiB           2       print("     l) List")
293   98.828 MiB    0.000 MiB           2       print("     m) Main menu")
294   98.828 MiB    0.000 MiB           2       print("----------------------------------------")
295   98.828 MiB    0.000 MiB           2       choice = input("Choice: ")
296   98.828 MiB    0.000 MiB           2       print()
297   98.828 MiB    0.000 MiB           2       if choice.lower() in ['i','l','m']:
298   98.828 MiB    0.000 MiB           2           return choice.lower()
299                                               else:
300                                                   print(choice +"?")
301                                                   print("Invalid option")
302                                                   Time.sleep(3)
303                                                   return None
304   98.828 MiB  196.559 MiB           2   @profile
305                                           def evol_timeline():
306                                               """Gets the evolutionary timeline for a species or taxa given a single
307                                               name or a list"""
308   98.828 MiB  196.559 MiB           2       choice = timeline_menu()
309   98.828 MiB    0.000 MiB           2       if choice == "i":
310   97.730 MiB    0.000 MiB           1           name = input("Name: ")
311   98.711 MiB    0.980 MiB           1           with Browser('firefox', headless=True) as browser:
312   98.762 MiB    0.051 MiB           1               browser.visit('http://timetree.org')
313   98.859 MiB    0.098 MiB           1               browser.fill('timeline_taxon', name)
314   98.859 MiB    0.000 MiB           1               browser.find_by_id('timeline-search-button1').click()
```

```
315   98.863 MiB    0.004 MiB           1                browser.is_text_present('Evolutionary Timeline for', wait_time=15)
316   98.863 MiB    0.000 MiB           1                browser.find_by_id('timeline-timeline-svg-export-btn').click()
317   98.863 MiB   98.863 MiB           1                downloads = get_download_path()
318   98.863 MiB    0.000 MiB           1                current = os.getcwd()
319   98.863 MiB    0.000 MiB           1                file = name + '_' + 'timeline.jpg'
320   98.863 MiB    0.000 MiB           1                indir = str(downloads) + '\\' + file
321   98.863 MiB    0.000 MiB           1                outdir = str(current) + '\\' + file
322   98.871 MiB    0.008 MiB           1                shutil.move(indir, outdir)
323   98.871 MiB    0.000 MiB           1                print("Done!")
324   98.887 MiB    0.016 MiB           1                print()
325   98.828 MiB    0.000 MiB           1            elif choice == 'l':
326   98.828 MiB    0.000 MiB           1                infile = input("Input file in .txt format: ")
327   98.828 MiB    0.000 MiB           1                if infile.endswith('.txt') is True:
328   98.828 MiB    0.000 MiB           1                    pass
329                                                     elif infile.endswith('.txt') is False:
330                                                         infile = infile + '.txt'
331   98.828 MiB    0.000 MiB           1                try:
332   98.828 MiB    0.000 MiB           1                    my_file = open(infile, "r")
333                                                     except FileNotFoundError:
334                                                         print("File not found")
335                                                         return
336                                                     else:
337   98.836 MiB    0.008 MiB           1                    my_file = open(infile, "r")
338   98.836 MiB    0.000 MiB           1                content = my_file.read()
339   98.836 MiB    0.000 MiB           1                content_list = content.split("\n")
340   98.836 MiB    0.000 MiB           1                my_file.close()
341   98.836 MiB    0.000 MiB           1                converted_list =[]
342   98.836 MiB    0.000 MiB           4                for element in content_list:
343   98.836 MiB    0.000 MiB           3                    converted_list.append(element.strip())
344   98.836 MiB    0.000 MiB           1                species = tuple(converted_list)
345   99.297 MiB    0.000 MiB           4                for i in range(0,len(species)):
346   99.234 MiB    0.000 MiB           3                    name = species[i]
347   99.297 MiB    0.305 MiB           3                    with Browser('firefox', headless=True) as browser:
348   99.297 MiB    0.008 MiB           3                        browser.visit('http://timetree.org')
349   99.297 MiB    0.105 MiB           3                        browser.fill('timeline_taxon', name)
350   99.297 MiB    0.062 MiB           3                        browser.find_by_id('timeline-search-button1').click()
351   99.297 MiB    0.004 MiB           3                        browser.is_text_present('Evolutionary Timeline for', wait_time=120)
352   99.297 MiB   -0.023 MiB           3                        browser.find_by_id('timeline-timeline-svg-export-btn').click()
```

```
353   99.297 MiB    0.000 MiB           3                   time = tm.datetime.now()
354   99.297 MiB    0.000 MiB           3                   print('{:<25}'.format(name), end="")
355   99.297 MiB    0.000 MiB           3                   print(time)
356   99.297 MiB  297.707 MiB           3                   downloads = get_download_path()
357   99.297 MiB    0.000 MiB           3                   current = os.getcwd()
358   99.297 MiB    0.000 MiB           3                   file = name + '_' + 'timeline.jpg'
359   99.297 MiB    0.000 MiB           3                   indir = str(downloads) + '\\' + file
360   99.297 MiB    0.000 MiB           3                   outdir = str(current) + '\\' + file
361   99.297 MiB    0.000 MiB           3                   shutil.move(indir, outdir)
362   99.297 MiB    0.000 MiB           1               print("Done!")
363   99.297 MiB    0.000 MiB           1               print()
364                                                   elif choice == 'm':
365                                                       main_loop()
366   98.832 MiB  196.523 MiB           2   @profile
367                                                   def time_tree_menu():
368                                                       """Menu options for time tree function"""
369   98.832 MiB    0.000 MiB           2       print('TIME TREE MENU:')
370   98.832 MiB    0.000 MiB           2       print("----------------------------------------")
371   98.832 MiB    0.000 MiB           2       print("Choose one of the following options?")
372   98.832 MiB    0.000 MiB           2       print("     t) Taxon")
373   98.832 MiB    0.000 MiB           2       print("     s) Species list")
374   98.832 MiB    0.000 MiB           2       print("     m) Main menu")
375   98.832 MiB    0.000 MiB           2       print("----------------------------------------")
376   98.832 MiB    0.000 MiB           2       choice = input("Choice: ")
377   98.832 MiB    0.000 MiB           2       if choice.lower() in ['t','s','m']:
378   98.832 MiB    0.000 MiB           2           return choice.lower()
379                                               else:
380                                                   print(choice +"?")
381                                                   print("Invalid option")
382                                                   Time.sleep(3)
383                                                   return None
384   98.832 MiB  196.523 MiB           2   @profile
385                                                   def time_tree():
386                                                       """Takes a species list as input and returns a time tree in one of the specified
387                                                       formats"""
388   98.832 MiB  196.523 MiB           2       choice = time_tree_menu()
389   98.832 MiB    0.000 MiB           2       if choice == 't':
390   97.691 MiB    0.000 MiB           1           name = input("Name: ")
```

```
391   98.734 MiB    1.043 MiB           1           with Browser('firefox', headless=True) as browser:
392   98.805 MiB    0.070 MiB           1               browser.visit('http://timetree.org')
393   98.852 MiB    0.047 MiB           1               browser.fill('timetree_taxon', name)
394   98.855 MiB    0.004 MiB           1               browser.find_by_id('timetree-search-button1').click()
395   98.859 MiB    0.004 MiB           1               browser.find_by_id('generate-download-button').click()
396   98.859 MiB   98.859 MiB           1               downloads = get_download_path()
397   98.859 MiB    0.000 MiB           1               current = os.getcwd()
398   98.859 MiB    0.000 MiB           1               file = name + '_' + 'species.nwk'
399   98.859 MiB    0.000 MiB           1               indir = str(downloads) + '\\' + file
400   98.859 MiB    0.000 MiB           1               outdir = str(current) + '\\' + file
401   98.867 MiB    0.008 MiB           1               shutil.move(indir, outdir)
402   98.883 MiB    0.016 MiB           1               print("Done!")
403   98.883 MiB    0.000 MiB           1           print()
404   98.832 MiB    0.000 MiB           1       elif choice == 's':
405   98.832 MiB    0.000 MiB           1           infile_name = input("Input file in .txt format: ")
406   98.832 MiB    0.000 MiB           1           if infile_name.endswith('.txt') is True:
407   98.832 MiB    0.000 MiB           1               pass
408                                                   elif infile_name.endswith('.txt') is False:
409                                                       infile_name = infile_name + '.txt'
410   98.832 MiB    0.000 MiB           1           try:
411   98.840 MiB    0.008 MiB           1               open(infile_name, "r")
412                                                   except FileNotFoundError:
413                                                       print("File not found")
414                                                       return
415   98.840 MiB    0.000 MiB           1           if os.path.isfile(infile_name) is True:
416   98.840 MiB    0.000 MiB           1               pass
417                                                   elif os.path.isfile(infile_name) is False:
418                                                       print("No such file!")
419                                                       return
420   98.840 MiB    0.000 MiB           1           path = os.getcwd()
421   98.840 MiB    0.000 MiB           1           infile_path = path + "\\" + infile_name
422   98.840 MiB    0.000 MiB           1           print(infile_path)
423   98.848 MiB    0.008 MiB           1           with Browser('firefox', headless=True) as browser:
424   98.840 MiB   -0.008 MiB           1               browser.visit('http://timetree.org')
425   98.883 MiB    0.043 MiB           1               browser.attach_file("prunetree_upload", infile_path)
426   98.883 MiB    0.000 MiB           1               browser.find_by_id('prunetree-upload-button1').click()
427   98.895 MiB    0.012 MiB           1               browser.is_text_present('Geologic Timescale', wait_time=320)
428   98.895 MiB    0.000 MiB           1               browser.find_by_id('prunetree-msg-btn').click()
```

```
429   98.895 MiB    0.000 MiB           1                try:
430   98.902 MiB    0.008 MiB           1                    replaced = browser.find_by_id('unresolved-names').value
431                                                       except BaseException:
432                                                           pass
433                                                       else:
434   98.902 MiB    0.000 MiB           1                    replaced = browser.find_by_id('unresolved-names').value
435   98.902 MiB    0.000 MiB           1                    file = open('Unresolved names.txt','w')
436   98.902 MiB    0.000 MiB           1                    file.write(replaced)
437   98.902 MiB    0.000 MiB           1                    file.close()
438   98.902 MiB    0.000 MiB           1                browser.find_by_id('prunetree-newick-export-btn').click()
439   98.902 MiB   98.902 MiB           1                downloads = get_download_path()
440   98.902 MiB    0.000 MiB           1                current = os.getcwd()
441   98.902 MiB    0.000 MiB           1                file = infile_name.replace('.txt', '.nwk')
442   98.902 MiB    0.000 MiB           1                indir = str(downloads) + '\\' + file
443   98.902 MiB    0.000 MiB           1                outdir = str(current) + '\\' + file
444   98.926 MiB    0.023 MiB           1                shutil.move(indir, outdir)
445   98.926 MiB    0.000 MiB           1                print("Done!")
446   98.926 MiB    0.000 MiB           1                print()
447                                                   elif choice == 'm':
448                                                       main_loop()
449   97.516 MiB   97.516 MiB           1   @profile
450                                           def citation():
451                                               """Prints the citation for the timetree resource"""
452   97.516 MiB    0.000 MiB           1       print()
453   97.516 MiB    0.000 MiB           1       print("CITE THE TIME TREE RESOURCE AS:")
454   97.516 MiB    0.000 MiB           2       print("S. Kumar, G. Stecher, M. Suleski, and S.B. Hedges, 2017.",
455   97.516 MiB    0.000 MiB           1             " TimeTree: a resource for timelines, timetrees, and divergence times.",
456   97.516 MiB    0.000 MiB           1             " Molecular Biology and Evolution 34: 1812-1819,",
457   97.516 MiB    0.000 MiB           1             " DOI: 10.1093/molbev/msx116")
458   97.516 MiB    0.000 MiB           1       print()
459  100.078 MiB  297.543 MiB           3   @profile
460                                           def validate_menu():
461                                               """Menu options for validate function"""
462  100.078 MiB    0.000 MiB           3       print("VALIDATE MENU:")
463  100.078 MiB    0.000 MiB           3       print("---------------------------------------")
464  100.078 MiB    0.000 MiB           3       print("      a) Check missing")
465  100.078 MiB    0.000 MiB           3       print("      b) Replace missing")
466  100.078 MiB    0.000 MiB           3       print("      c) View tree")
```

```
467  100.078 MiB    0.000 MiB           3       print("      m) Main menu")
468  100.078 MiB    0.000 MiB           3       print("----------------------------------------")
469  100.078 MiB   -0.031 MiB           3       choice = input("Choice: ")
470  100.078 MiB    0.000 MiB           3       if choice.lower() in ['a','b','c','m']:
471  100.078 MiB    0.000 MiB           3           return choice.lower()
472                                             else:
473                                                 print(choice +"?")
474                                                 print("Invalid option")
475                                                 Time.sleep(3)
476                                                 return None
477  100.078 MiB  297.543 MiB           3   @profile
478                                         def validate():
479                                             """Checks output file for missing values"""
480  100.078 MiB  297.512 MiB           3       choice = validate_menu()
481  100.078 MiB    0.000 MiB           3       if choice == 'a':
482   97.488 MiB    0.000 MiB           1           infile = input("File to check (.csv): ")
483   97.488 MiB    0.000 MiB           1           if infile.endswith('.csv') is True:
484   97.488 MiB    0.000 MiB           1               pass
485                                               elif infile.endswith('.csv') is False:
486                                                   infile = infile + '.csv'
487   97.922 MiB    0.434 MiB           1           data = pd.read_csv(infile)
488   97.922 MiB    0.000 MiB           1           length = len(data.index)
489   98.160 MiB    0.238 MiB           1           indices = data.loc[pd.isna(data).any(1), :].index
490   98.160 MiB    0.000 MiB           1           check = len(indices)
491   98.160 MiB    0.000 MiB           1           if check >= 1:
492   98.242 MiB    0.082 MiB           1               missing = pd.DataFrame(columns = ['Taxa1', 'Taxa2', 'Div.Time'])
493   99.910 MiB    0.000 MiB          10               for i in range (0, length):
494   99.910 MiB    0.066 MiB           9                   if i in indices:
495   99.906 MiB    0.027 MiB           2                       taxon_a = data.iloc[i]['Taxa1']
496   99.906 MiB    0.000 MiB           2                       taxon_b = data.iloc[i]['Taxa2']
497   99.914 MiB    1.238 MiB           2                       with Browser('firefox', headless=True) as browser:
498   99.914 MiB    0.000 MiB           2                           try:
499   99.914 MiB    0.066 MiB           2                               browser.visit('http://timetree.org')
500   99.914 MiB    0.047 MiB           2                               browser.fill('taxon_a', taxon_a)
501   99.914 MiB    0.000 MiB           2                               browser.fill('taxon_b', taxon_b)
502   99.914 MiB    0.000 MiB           2                               browser.find_by_id('pairwise-search-button1').click()
503                                                               except KeyboardInterrupt as error:
504                                                                   print('An exception occurred: {}'.format(error))
```

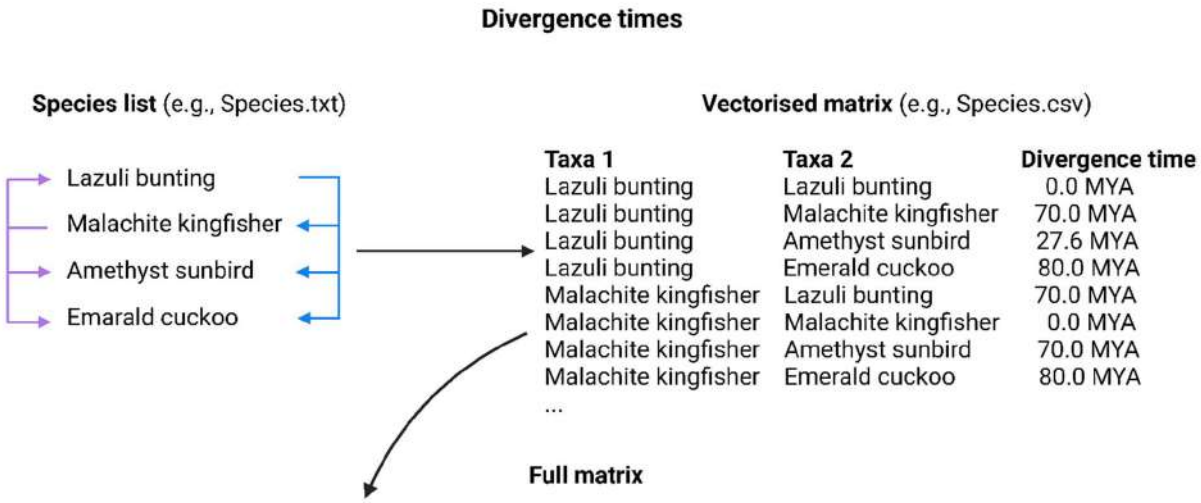
```
505                                                                                     return
506                                                                                 else:
507   99.922 MiB    0.008 MiB           2                                             browser.visit('http://timetree.org')
508   99.922 MiB    0.000 MiB           2                                             browser.fill('taxon_a', taxon_a)
509   99.922 MiB    0.000 MiB           2                                             browser.fill('taxon_b', taxon_b)
510   99.922 MiB    0.000 MiB           2                                             browser.find_by_id('pairwise-search-button1').click()
511   99.910 MiB   -0.008 MiB           2                                         browser.is_text_present('Median Time', wait_time=320)
512   99.910 MiB    0.000 MiB           2                                         var_y = '#pairwise-results > text:nth-child(11)'
513   99.910 MiB    0.090 MiB           2                                         divtime = browser.find_by_css(var_y).first.value
514   99.910 MiB    0.000 MiB           2                                         divtime_2 = divtime.replace(' MYA', '')
515   99.910 MiB    0.133 MiB           6                                         missing = missing.append({'Taxa1' : taxon_a, 'Taxa2' :
                                                                                                            taxon_b,
516   99.910 MiB    0.000 MiB           4                                                                       'Div.Time' : divtime_2}, ignore_index
                                                                                                              = True)
517   99.910 MiB    0.000 MiB           2                                     time = tm.datetime.now()
518   99.910 MiB    0.000 MiB           2                                     print('{:<25}'.format(taxon_a), end="")
519   99.910 MiB    0.000 MiB           2                                     print('{:<25}'.format(taxon_b), end="")
520   99.910 MiB    0.000 MiB           2                                     print('{:<5}'.format(divtime_2), end="")
521   99.910 MiB    0.000 MiB           2                                     print(time)
522   99.910 MiB    0.000 MiB           1                                 output_name = input("File name (.csv) for missing values: ")
523   99.910 MiB    0.000 MiB           1                                 if output_name.endswith(".csv") is True:
524   99.910 MiB    0.000 MiB           1                                     pass
525                                                                         elif output_name.endswith('.csv') is False:
526                                                                             output_name = output_name + '.csv'
527   99.945 MiB    0.035 MiB           1                                 missing.to_csv(output_name, index=False)
528                                                                     else:
529                                                                         print("No missing values detected!")
530  100.078 MiB    0.000 MiB           2                           elif choice == 'b':
531   99.945 MiB    0.000 MiB           1                               infile_1 = input("File to check (.csv): ")
532   99.945 MiB    0.000 MiB           1                               if infile_1.endswith('.csv') is True:
533   99.945 MiB    0.000 MiB           1                                   pass
534                                                                     elif infile_1.endswith('.csv') is False:
535                                                                         infile_1 = infile_1 + '.csv'
536   99.957 MiB    0.012 MiB           1                               data = pd.read_csv(infile_1)
537   99.941 MiB   -0.016 MiB           1                               infile_2 = input("Missing value file (.csv): ")
538   99.941 MiB    0.000 MiB           1                               if infile_2.endswith('.csv') is True:
539   99.941 MiB    0.000 MiB           1                                   pass
540                                                                     elif infile_2.endswith('.csv') is False:
```

```
541                                                             infile_2 = infile_2 + '.csv'
542   99.953 MiB    0.012 MiB           1           missing = pd.read_csv(infile_2)
543   99.953 MiB    0.000 MiB           1           length = len(missing.index)
544  100.008 MiB    0.000 MiB           3           for i in range(0, length):
545  100.008 MiB    0.000 MiB           2               new_value = missing.iloc[i]['Div.Time']
546  100.008 MiB    0.055 MiB           2               data['Div.Time'].fillna(new_value,inplace=True, limit=1)
547  100.008 MiB    0.000 MiB           1           output_name = input("File name (.csv) for replaced values: ")
548  100.008 MiB    0.000 MiB           1           if output_name.endswith('.csv') is True:
549  100.008 MiB    0.000 MiB           1               pass
550                                                 elif output_name.endswith('.csv') is False:
551                                                     output_name = output_name + '.csv'
552  100.078 MiB    0.070 MiB           1           data.to_csv(output_name, index=False)
553  100.078 MiB    0.000 MiB           1           print('Done!')
554  100.078 MiB    0.000 MiB           1           print()
555  100.078 MiB    0.000 MiB           1       elif choice == 'c':
556  100.078 MiB    0.000 MiB           1           infile_1 = input("Input tree file in .nwk format: ")
557  100.078 MiB    0.000 MiB           1           tree = Phylo.read(infile_1, "newick")
558  100.082 MiB    0.004 MiB           1           Phylo.draw_ascii(tree)
559                                               elif choice == 'm':
560                                                   main_loop()
561   99.297 MiB  396.570 MiB           4   @profile
562                                           def get_download_path():
563                                               """Returns the default downloads path for linux or windows"""
564   99.297 MiB    0.000 MiB           4       if os.name == 'nt':
565   99.297 MiB    0.000 MiB           4           sub_key = r'SOFTWARE\Microsoft\Windows\CurrentVersion\Explorer\Shell Folders'
566   99.297 MiB    0.000 MiB           4           downloads_guid = '{374DE290-123F-4565-9164-39C4925E467B}'
567   99.297 MiB    0.000 MiB           4           with winreg.OpenKey(winreg.HKEY_CURRENT_USER, sub_key) as key:
568   99.297 MiB    0.000 MiB           4               location = winreg.QueryValueEx(key, downloads_guid)[0]
569   99.297 MiB    0.000 MiB           4           return location
570                                               else:
571                                                   return os.path.join(os.path.expanduser('~'), 'downloads')
572   97.535 MiB   97.535 MiB           1   @profile
573                                           def main_loop():
574                                               """The main loop of the script"""
575                                               while True:
576   99.684 MiB  295.953 MiB           3           choice = menu_choice()
577   99.684 MiB    0.000 MiB           3           if choice is None:
578                                                   continue
```

```
579   99.684 MiB    0.000 MiB           3           elif choice == 'q':
580   99.684 MiB    0.000 MiB           1               print( "Exiting...")
581   99.684 MiB    0.000 MiB           1               break     # jump out of while loop
582   98.680 MiB    0.000 MiB           2           elif choice == '*':
583   99.684 MiB  198.422 MiB           2               check_available()
584                                                   elif choice == 'a':
585                                                       div_times_sing()
586                                                   elif choice == 'b':
587                                                       div_times_batch()
588                                                   elif choice == 'c':
589                                                       evol_timeline()
590                                                   elif choice == 'd':
591                                                       time_tree()
592                                                   elif choice == 'e':
593                                                       citation()
594                                                   elif choice == 'f':
595                                                       validate()
596                                                   else:
597                                                       print("Invalid choice.")
```

SUPPLEMENTARY FIGURES

Sup. Figure 1: Diagrammatic representation of the batch retrieval of divergence times using a list of species as input. The list is iterated through to compile the divergence times of every possible combination and output is given as a three-column vectorized matrix. The output can then be converted to a full matrix for statistical analyses of distance matrices. (Image created in BioRender.com)



Species	Lazuli bunting	Malachite kingfisher	Amethyst sunbird	Emerald cuckoo
Lazuli bunting	0.0	70	27.6	80
Malachite kingfisher	70.0	0	70.0	80
Amethyst sunbird	27.6	70	0.0	80
Emerald cuckoo	80.0	80	80.0	0

LE CLERCQ, L., KOTZÉ, A., GROBLER, J.P. & DALTON, D.L. (2023*d*) Biological clocks as age estimation markers in animals: a systematic review and meta-analysis. *Biological Reviews* **98**, 1972–2011.

Biological clocks as age estimation markers in animals: a systematic review and meta-analysis

SUPPORTING INFORMATION

Louis-Stéphane Le Clercq [1,2], Antoinette Kotzé [1,2], J. Paul Grobler [2] and Desiré Lee Dalton [3]

[1] *South African National Biodiversity Institute, P.O. Box 754, Pretoria, 0001, South Africa.*

[2] *Department of Genetics, University of the Free State, P.O. Box 339, Bloemfontein, 9300, South Africa.*

[3] *School of Health and Life Sciences, Teesside University, Middlesbrough, TS1 3BA, United Kingdom.*

Table S1. Summary of all species within the class Mammalia for which tooth cementum annulation (TCA) has been used as biological clock method for age determination. Species are grouped first by order and then by family and listed alphabetically by scientific name.

Common name	Scientific name	Study/source
Ungulates (order Artiodactyla)		
Family Antilocapridae		
Pronghorn	*Antilocapra americana*	McCutchen (1969)
Family Bovidae		
Springbok	*Antidorcas marsupialis*	Rautenbach (1970)
European bison	*Bison bonasus*	Klevezal & Pucek (1987)
Cattle	*Bos taurus*	Grue & Jensen (1979)
Water buffalo	*Bubalus bubalis*	Grue & Jensen (1979)
East Caucasian tur	*Capra cylindricornis*	Grue & Jensen (1979)
Japanese serow	*Capricornis crispus*	Miura (1985)
Blesbok	*Damaliscus pygargus phillipsi*	Olivier & Greyling (1991)
Thomson's gazelle	*Eudorcas thomsonii*	Robinette & Archer (1971)
Goitered gazelle	*Gazella subgutturosa*	Zhu *et al.* (1992)
Waterbuck	*Kobus ellipsiprymnus*	Spinage (1976)
Grant's gazelle	*Nanger granti*	Spinage (1976)
Mountain goat	*Oreamnos americanus*	Thomas (1977)
Muskox	*Ovibos moschatus*	Grue & Jensen (1979)
Sheep	*Ovis aries*	Saxon & Higham (1968)
Mouflon	*Ovis gmelini*	Grue & Jensen (1979)
Cantabrian chamois	*Rupicapra pyrenaica parva*	Perez-Barberia & Fernandez-Lopez (1996)
Chamois	*Rupicapra rupicapra*	Grue & Jensen (1979)
Common duiker	*Sylvicapra grimmia*	Wilson *et al.* (1984)
African buffalo	*Syncerus caffer*	Grimsdell (1973)
Common bushbuck	*Tragelaphus scriptus*	Simpson (1983)
Greater kudu	*Tragelaphus strepsiceros*	Simpson & Elder (1969)
Family Camelidae		
Guanaco	*Lama guanicoe*	Gourichon & Parmigiani (2016)
Family Cervidae		
Moose/Elk	*Alces alces*	Rolandsen *et al.* (2008)
Roe deer	*Capreolus capreolus*	Grue & Jensen (1979)
Red deer	*Cervus elaphus*	Mitchell (1967)
Spanish red deer	*Cervus elaphus hispanicus*	Azorit *et al.* (2002)
Sika deer	*Cervus nippon*	Grue & Jensen (1979)
Fallow deer	*Dama dama*	Grue & Jensen (1979)
Mule deer	*Odocoileus hemionus*	Low & Cowan (1963)
Black-tailed deer	*Odocoileus hemionus columbianus*	Low & Cowan (1963)
White-tailed deer	*Odocoileus virginianus*	Adams & Blanchong (2020)
Boreal woodland caribou	*Rangifer tarandus caribou*	Larter & Allaire (2016)
Barren-ground caribou	*Rangifer tarandus groenlandicus*	Miller (1974)
Norwegian reindeer	*Rangifer tarandus tarandus*	Reimers & Nordby (1968)
Family Giraffidae		
Giraffe	*Giraffa camelopardalis*	Spinage (1976)
Family Suidae		
Wild boar	*Sus scrofa*	Grue & Jensen (1979)

Family Tayassuidae		
Collared peccary	*Pecari tajacu*	Low (1970)
	Carnivores (order Carnivora)	
Family Canidae		
Domestic dog	*Canis familiaris*	Grue & Jensen (1979)
Coyote	*Canis latrans*	Gosselink *et al.* (2003)
Grey wolf	*Canis lupus*	Walleser *et al.* (2016)
Dingo	*Canis lupus dingo*	Grue & Jensen (1979)
Black-backed jackal	*Lupulella mesomelas*	Lombaard (1971)
Raccoon dog	*Nyctereutes procyonoides*	Grue & Jensen (1979)
Grey fox	*Urocyon cinereoargenteus*	Nicholson *et al.* (1985)
Arctic fox	*Vulpes lagopus*	Grue & Jensen (1979)
Red fox	*Vulpes vulpes*	Gosselink *et al.* (2003)
Bat-eared fox	*Otocyon megalotis*	Kamler & Macdonald (2006)
African wild dog	*Lycaon pictus*	Mbizah *et al.* (2016)
Iberian wolf	*Canis lupus signatus*	Pires *et al.* (2020)
Family Felidae		
Cheetah	*Acinonyx jubatus*	Marker & Dickman (2003)
Cat	*Felis catus*	Grue & Jensen (1979)
Wildcat	*Felis silvestris*	Grue & Jensen (1979)
Eurasian lynx	*Lynx lynx*	Grue & Jensen (1979)
Iberian lynx	*Lynx pardinus*	Zapata *et al.* (1997)
Bobcat	*Lynx rufus*	Crowe (1972)
Lion	*Panthera leo*	White & Belant (2016)
Iriomote cat	*Prionailurus bengalensis iriomotensis*	Nakanishi *et al.* (2009)
Cougar	*Puma concolor*	Hiller & Tyre (2014)
Snow leopard	*Panthera uncia*	McCarthy *et al.* (2005)
Family Hyaenidae		
Spotted hyena	*Crocuta crocuta*	Grue & Jensen (1979)
Family Mephitidae		
Striped skunk	*Mephitis mephitis*	Casey & Webster (1975)
Family Mustelidae		
Sea otter	*Enhydra lutris*	Schuler *et al.* (2018)
Wolverine	*Gulo gulo*	Rausch & Pearson (1972)
North American river otter	*Lontra canadensis*	Baitchman & Kollias (2000)
Eurasian otter	*Lutra lutra*	Grue & Jensen (1979)
American marten	*Martes americana*	Belant *et al.* (2011)
Beech marten	*Martes foina*	Grue & Jensen (1979)
European pine marten	*Martes martes*	Grue & Jensen (1979)
European badger	*Meles meles*	Ahnlund (1976)
Short-tailed weasel	*Mustela erminea*	Grue & Jensen (1979)
Steppe polecat	*Mustela eversmanii*	Ansorge & Suchentrunk (2001)
Least weasel	*Mustela nivalis*	Grue & Jensen (1979)
European polecat	*Mustela putorius*	Ansorge & Suchentrunk (2001)
American mink	*Neovison vison*	Grue & Jensen (1979)
Fisher	*Pekania pennanti*	Arthur *et al.* (1992)
American badger	*Taxidea taxus*	Crowe & Strickland (1975)
Family Otariidae		
Cape fur seal	*Arctocephalus pusillus*	Oosthuizen (1997)
Northern fur seal	*Callorhinus ursinus*	Scheffer (1950)
California sea lion	*Zalophus californianus*	Rust *et al.* (2019)
Family Phocidae		

Bearded seal	*Erignathus barbatus*	Stirling *et al.* (1977)
Crabeater seal	*Lobodon carcinophaga*	Laws (1958)
Northern elephant seal	*Mirounga angustirostris*	Klevezal & Stewart (1994)
Southern elephant seal	*Mirounga leonina*	Laws (1952)
Hawaiian monk seal	*Neomonachus schauinslandi*	Kenyon & Fiscus (1963)
Harp seal	*Pagophilus groenlandicus*	Bowen *et al.* (1983)
European harbor seal	*Phoca vitulina*	Dietz *et al.* (1991)
Ringed seal	*Pusa hispida*	Mclaren (1958)
Family Procyonidae		
Raccoon	*Procyon lotor*	Grau *et al.* (1970)
Family Ursidae		
American black bear	*Ursus americanus*	Marks & Erickson (1966)
Brown bear	*Ursus arctos*	Østbye *et al.* (2006)
Grizzly bear	*Ursus arctos horribilis*	Craighead *et al.* (1970)
Polar bears	*Ursus maritimus*	Christensen-Dalsgaard *et al.* (2010)
Bats (order Chiroptera)		
Family Phyllostomidae		
Vampire bat	*Desmodus rotundus*	Linhart (1973)
Family Pteropodidae		
Madagascan fruit bat	*Eidolon dupreanum*	Brook *et al.* (2019)
Black flying fox	*Pteropus alecto*	Grue & Jensen (1979)
Grey-headed flying fox	*Pteropus poliocephalus*	Divljan *et al.* (2006)
Madagascan flying fox	*Pteropus rufus*	Brook *et al.* (2019)
Family Vespertilionidae		
Common noctule	*Nyctalus noctula*	Grue & Jensen (1979)
Common pipistrelle	*Pipistrellus pipistrellus*	Grue & Jensen (1979)
Marsupials (order Diprotodontia)		
Family Macropodidae		
Agile wallaby	*Macropus agilis*	Grue & Jensen (1979)
Family Phalangeridae		
Common brush-tail possum	*Trichosurus vulpecula*	Pekelharing (1970)
Shrews (order Eulipotyphla)		
Family Erinaceidae		
European hedgehog	*Erinaceus europaeus*	Grue & Jensen (1979)
Family Talpidae		
European mole	*Talpa europaea*	Grue & Jensen (1979)
Bandicoots (order Peramelemorphia)		
Family Peramelidae		
Long-nosed bandicoot	*Perameles nasuta*	Kingsmill (1962)
Ungulates (order: Perissodactyla)		
Family Equidae		
Cape mountain zebra	*Equus zebra zebra*	Penzhorn (1982)
Family Rhinocerotidae		
Black rhinoceros	*Diceros bicornis*	Spinage (1976)
Primates (order: Primates)		
Family Callitrichidae		

Common marmoset	*Callithrix jacchus*	Scott *et al.* (1980)
Brown-mantled tamarin	*Leontocebus fuscicollis*	Yoneda (1982)
White-lipped tamarin	*Saguinus labiatus*	Yoneda (1982)
Mustached tamarin	*Saguinus mystax*	Yoneda (1982)
Family Cercopithecidae		
Japanese macaque	*Macaca fuscata*	Wada *et al.* (1978)
Rhesus macaques	*Macaca mulatta*	Newham *et al.* (2021)
Family Hominidae		
Humans	*Homo sapiens*	Wittwer-Backofen *et al.* (2004)
Loxodonts (order Proboscidea)		
Family Elephantidae		
African elephant	*Loxodonta africana*	Spinage (1976)
Rodents (order Rodentia)		
Family Castoridae		
North American beaver	*Castor canadensis*	Nostrand & Stephenson (1964)
Eurasian beaver	*Castor fiber*	Rosell *et al.* (2010)
Family Cricetidae		
European water vole	*Arvicola amphibius*	Grue & Jensen (1979)
Family Dipodidae		
Northern birch mouse	*Sicista betulina*	Grue & Jensen (1979)
Family Erethizontidae		
Canadian porcupine	*Erethizon dorsatum*	Earle & Kramm (1980)
Family Gliridae		
Hazel dormouse	*Muscardinus avellanarius*	Grue & Jensen (1979)
Family Hystricidae		
Cape porcupine	*Hystrix africaeaustralis*	Van Aarde (1985)
Family Muridae		
Brown rat	*Rattus norvegicus*	Grue & Jensen (1979)
Family Sciuridae		
Groundhog	*Marmota monax*	Smith (1975)
California ground squirrel	*Otospermophilus beecheyi*	Adams & Watkins (1967)
Eurasian red squirrel	*Sciurus vulgaris*	Grue & Jensen (1979)

Fig. S1. PRISMA flow diagram for search, screening, and inclusion of methylation studies for age determination in animals. Free term searches to retrieve background information through *Google Scholar* are excluded.

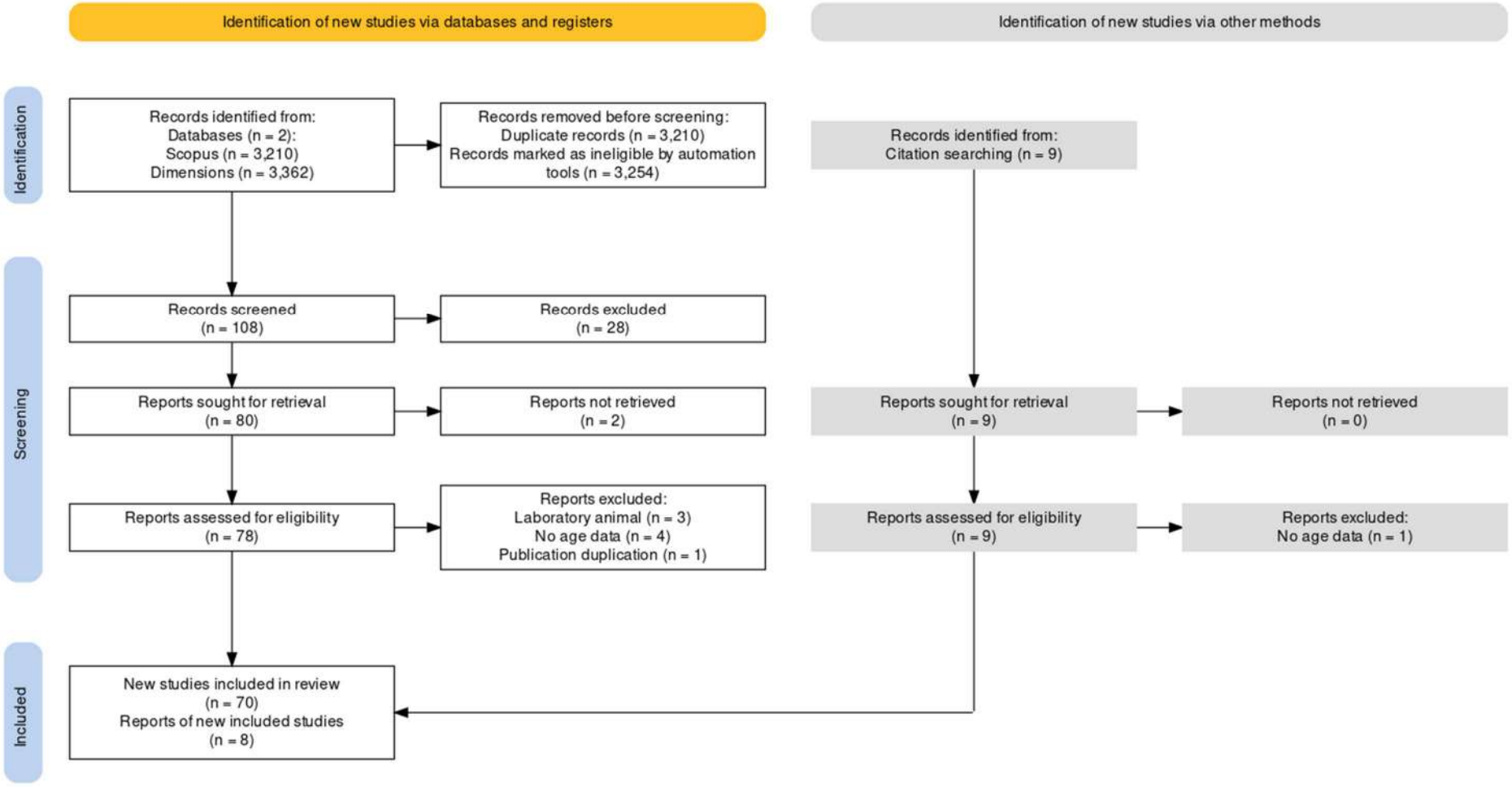


Fig. S2. PRISMA flow diagram for search, screening, and inclusion of telomere studies for age determination in animals. Free term searches to retrieve background information through *Google Scholar* are excluded.

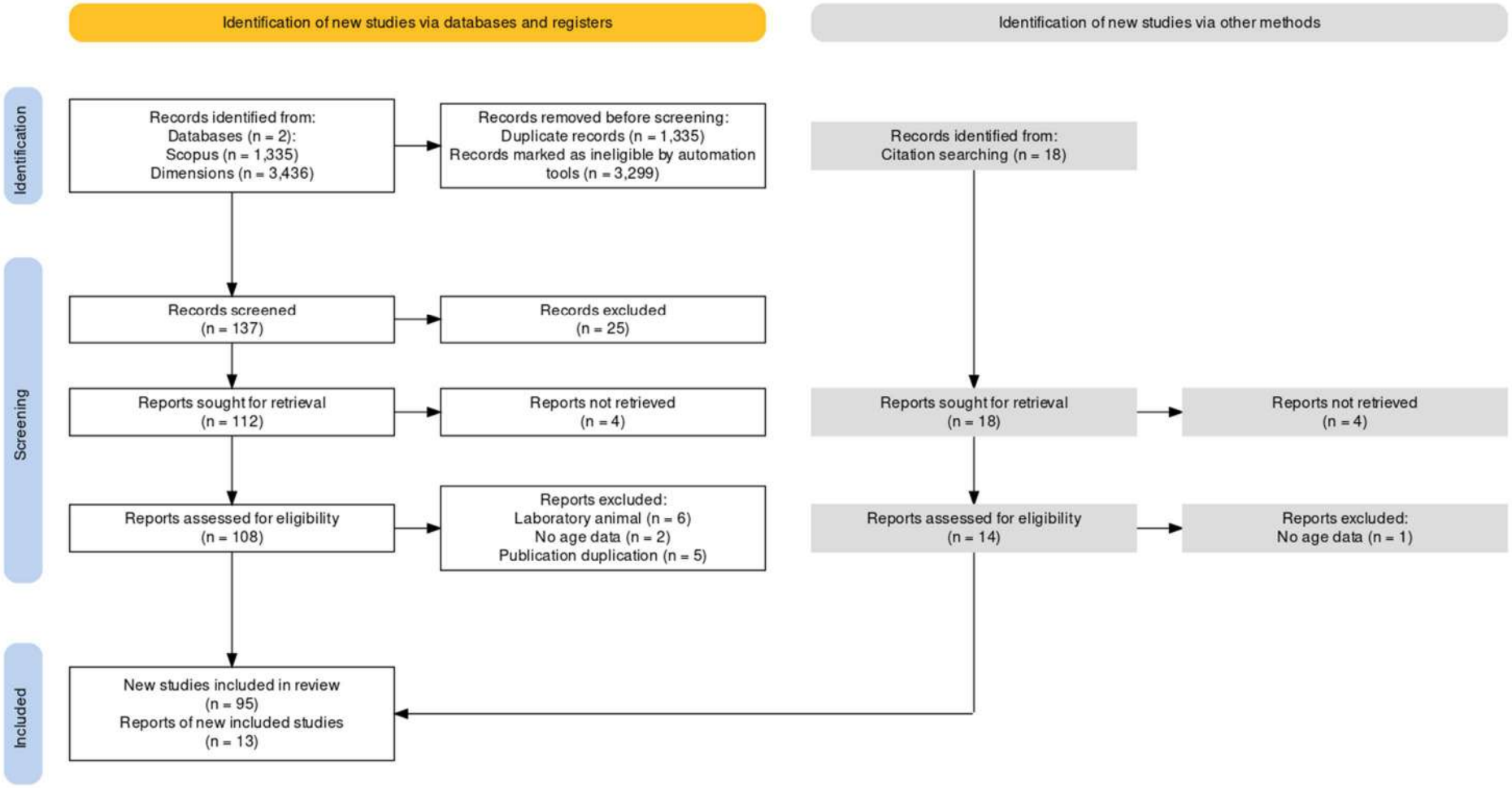


Fig. S3. Relationship between power and sample size for different effect sizes in a multiple linear regression with a minimum of three predictors at a significance of 95% ($\alpha = 0.05$). The range of effect sizes (f^2) are plotted from small effects of approximately 0.1 (red), and medium effects at 0.2 (green), to large effects of 0.3 (pink). For most biological studies a minimum power of at least 0.8 is required. Based on the graph, methylation studies that have a high effect size will achieve a power of 0.8 at approximately 40 samples while studies with smaller effect sizes can achieve comparable power with 60–70 samples. Studies with very small effect sizes can generally achieve a power of at least 0.7 with sample sizes of 100 or more.

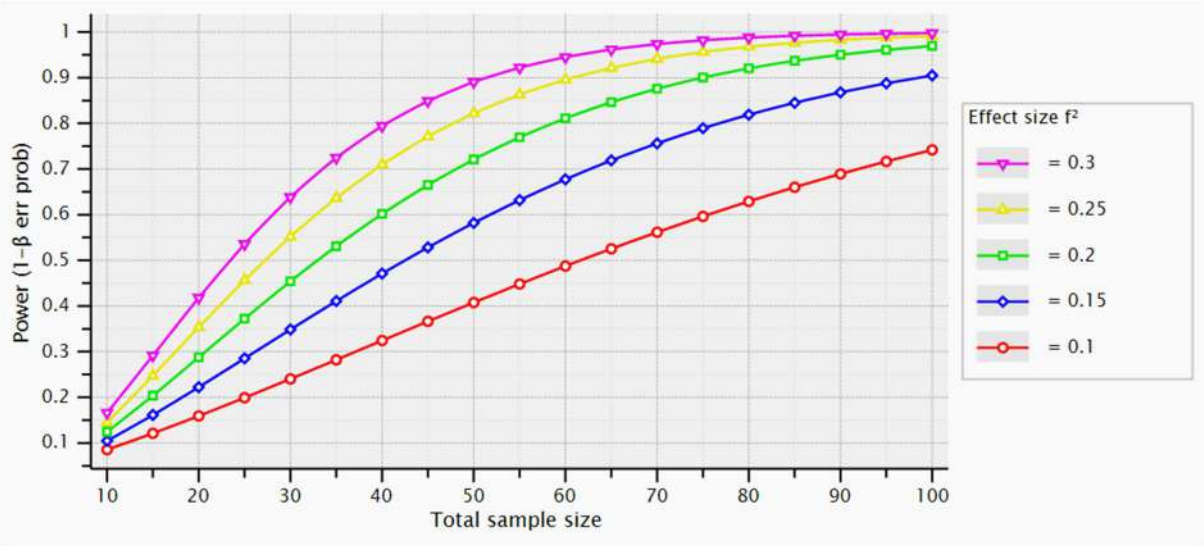


Fig. S4. Forest plot and results for the meta-analysis of studies using methylation to estimate animal age with subgrouping based on the defined groups. Key study attributes such as sample size (*N*), genome size (in billions of base pairs, bbp), and tissue type are indicated along with the computed level of possible bias. CI, confidence interval; COR, correlation; *N*, sample size; SE, standard error.

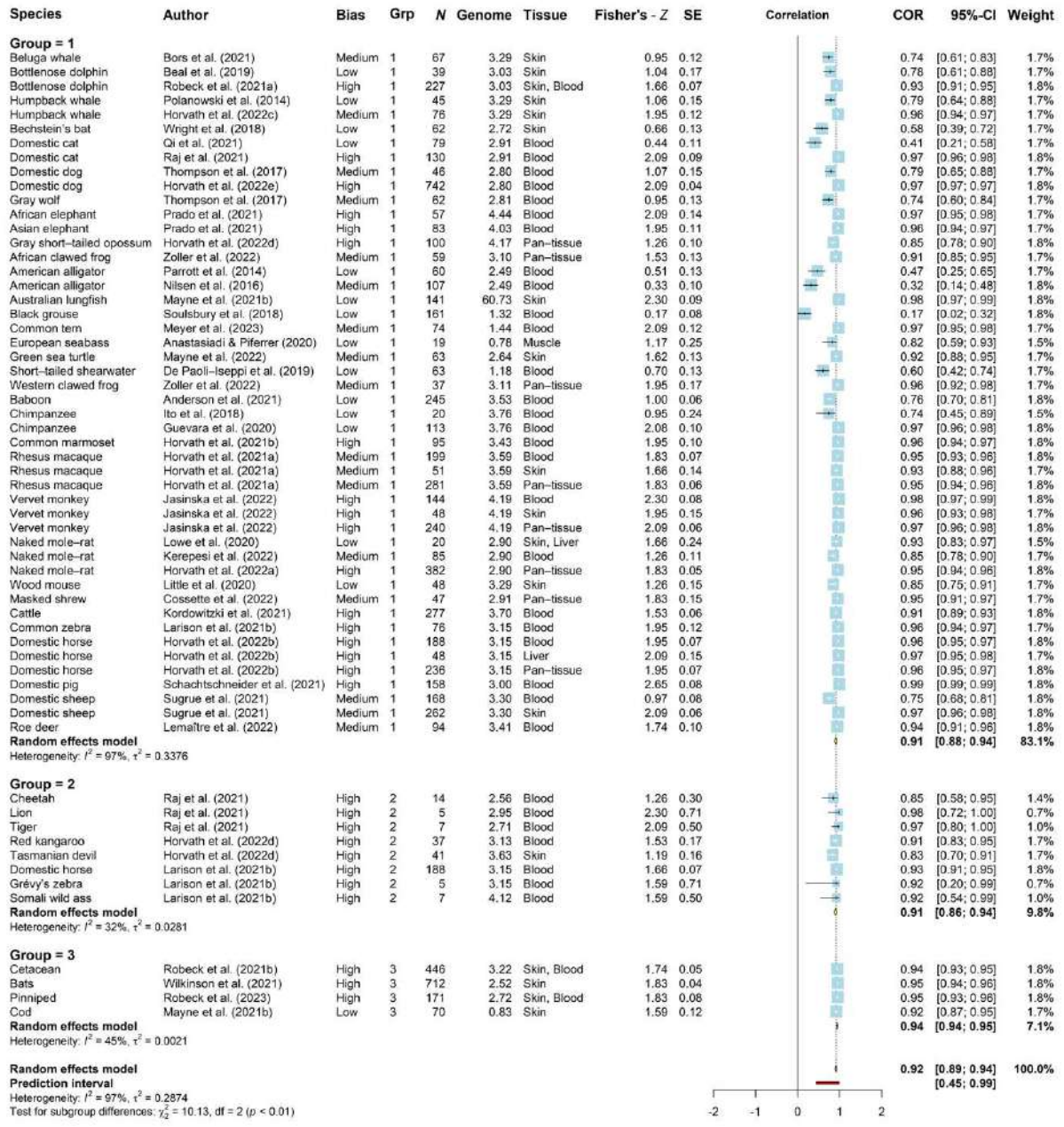


Fig. S5. Forest plot and results for the meta-analysis of studies using methylation to estimate animal age with subgrouping based on tissue type. Key study attributes such as sample size (*N*), genome size (in bbp), and tissue type are indicated along with the computed level of possible bias.CI, confidence interval; COR, correlation; *N*, sample size; SE, standard error.

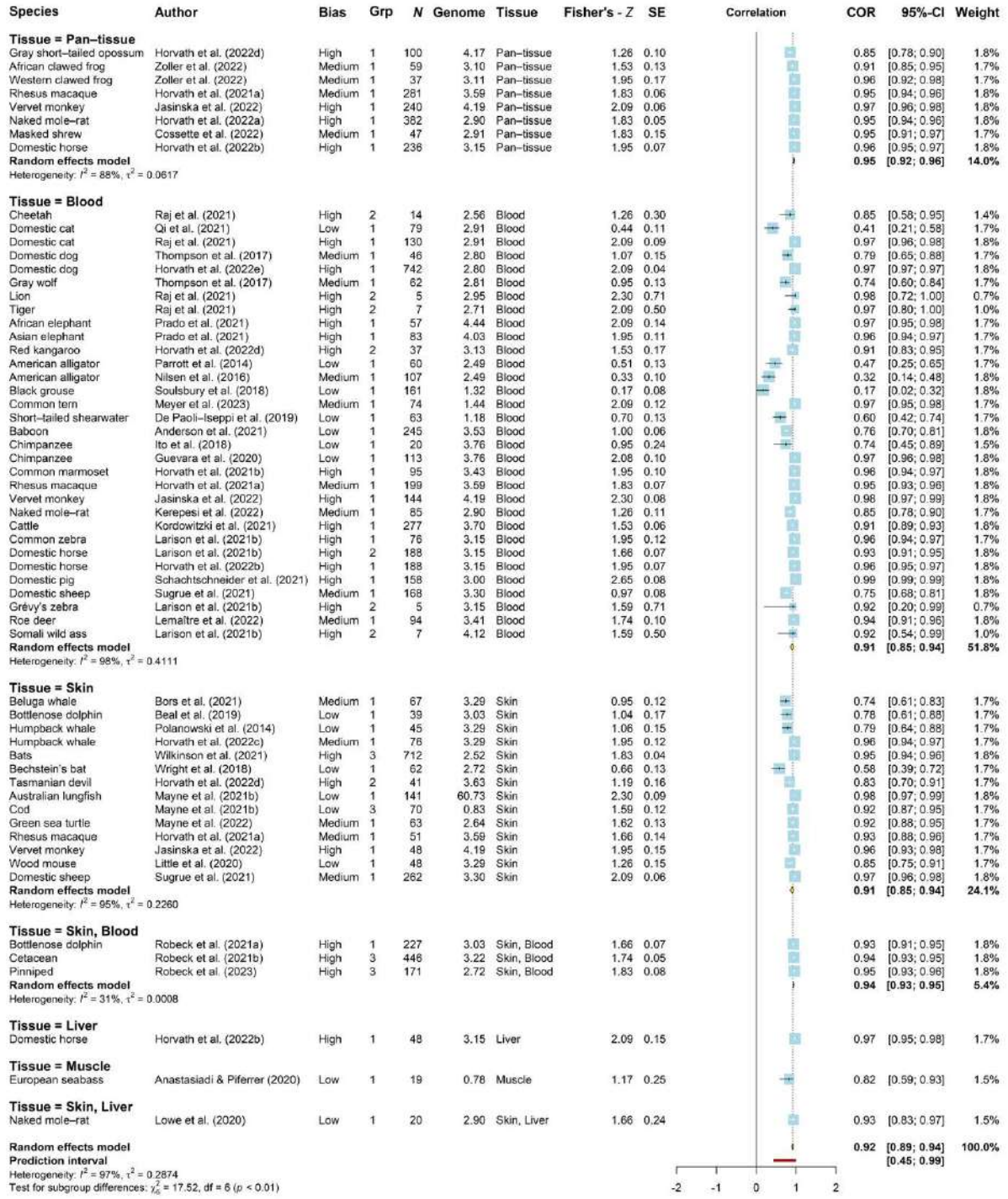


Fig. S6. Forest plot and results for the meta-analysis of studies using methylation to estimate animal age with subgrouping based on study method. Key study attributes such as sample size (*N*), genome size (in bbp), and tissue type are indicated along with the computed level of possible bias. CI, confidence interval; COR, correlation; ELISA, enzyme-linked immunosorbent assay; LC-MS/MS, liquid chromatography-mass spectrometry; MS-PCR, methylation specific/sensitive polymerase chain reaction; N, sample size; SE, standard error.

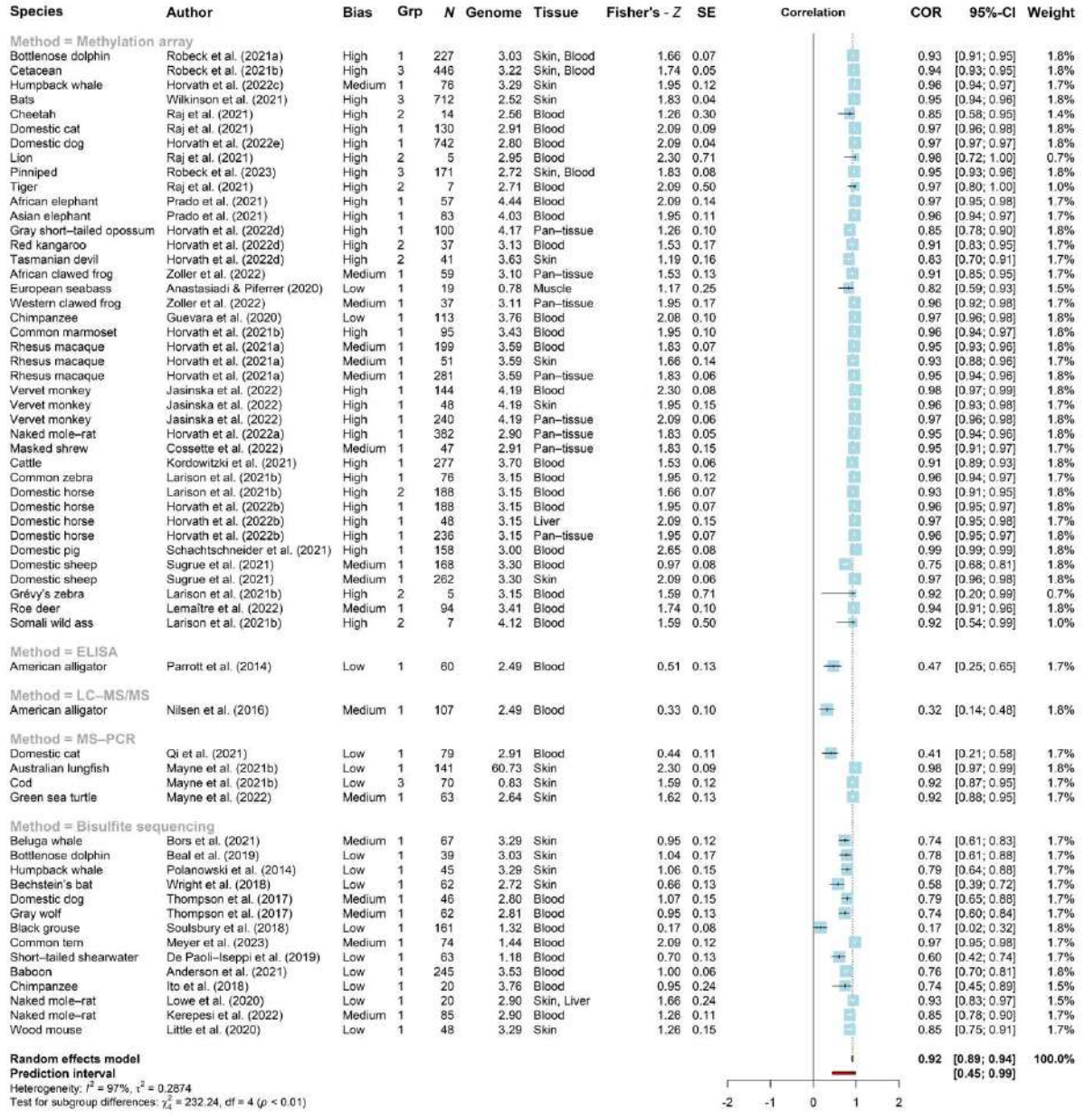


Fig. S7. Forest plot and results for the meta-analysis of studies using telomere length to estimate animal age with subgrouping based on defined groups. Key study attributes such as sample size (*N*), karyotype (Karyo), and tissue type are indicated along with the computed level of possible bias. CH, Chizé region, France; CI, confidence interval; COR, correlation; N, sample size; PCR, polymerase chain reaction; SE, standard error; TF, Trois-Fontaines region, France; TRL, terminal restriction length.

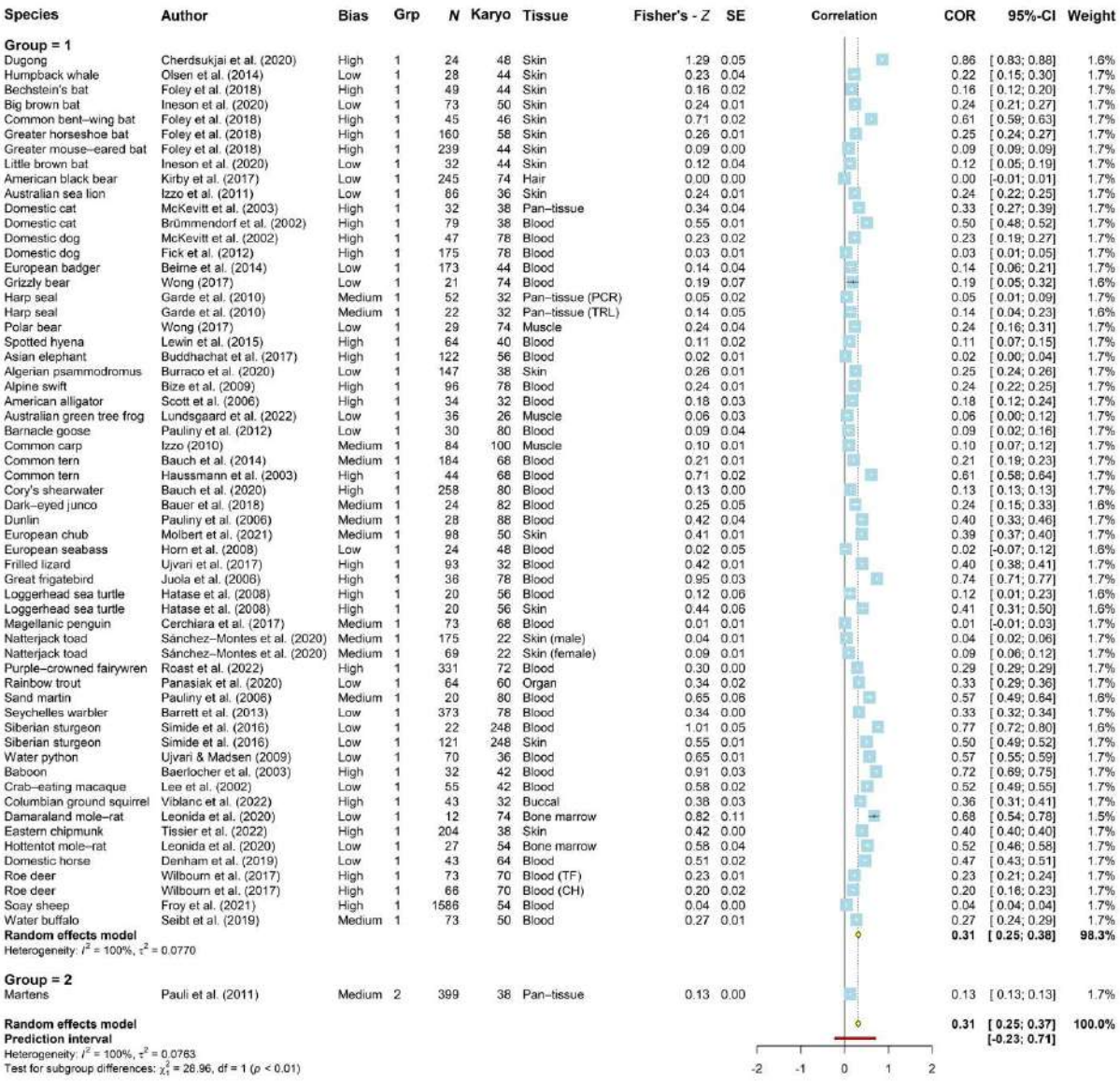


Fig. S8. Forest plot and results for the meta-analysis of studies using telomere length to estimate animal age with subgrouping based on tissue type. Key study attributes such as sample size (*N*), karyotype (Karyo), and tissue type are indicated along with the computed level of possible bias. CH, Chizé region, France; CI, confidence interval; COR, correlation; N, sample size; PCR, polymerase chain reaction; SE, standard error; TF, Trois-Fontaines region, France; TRL, terminal restriction length.

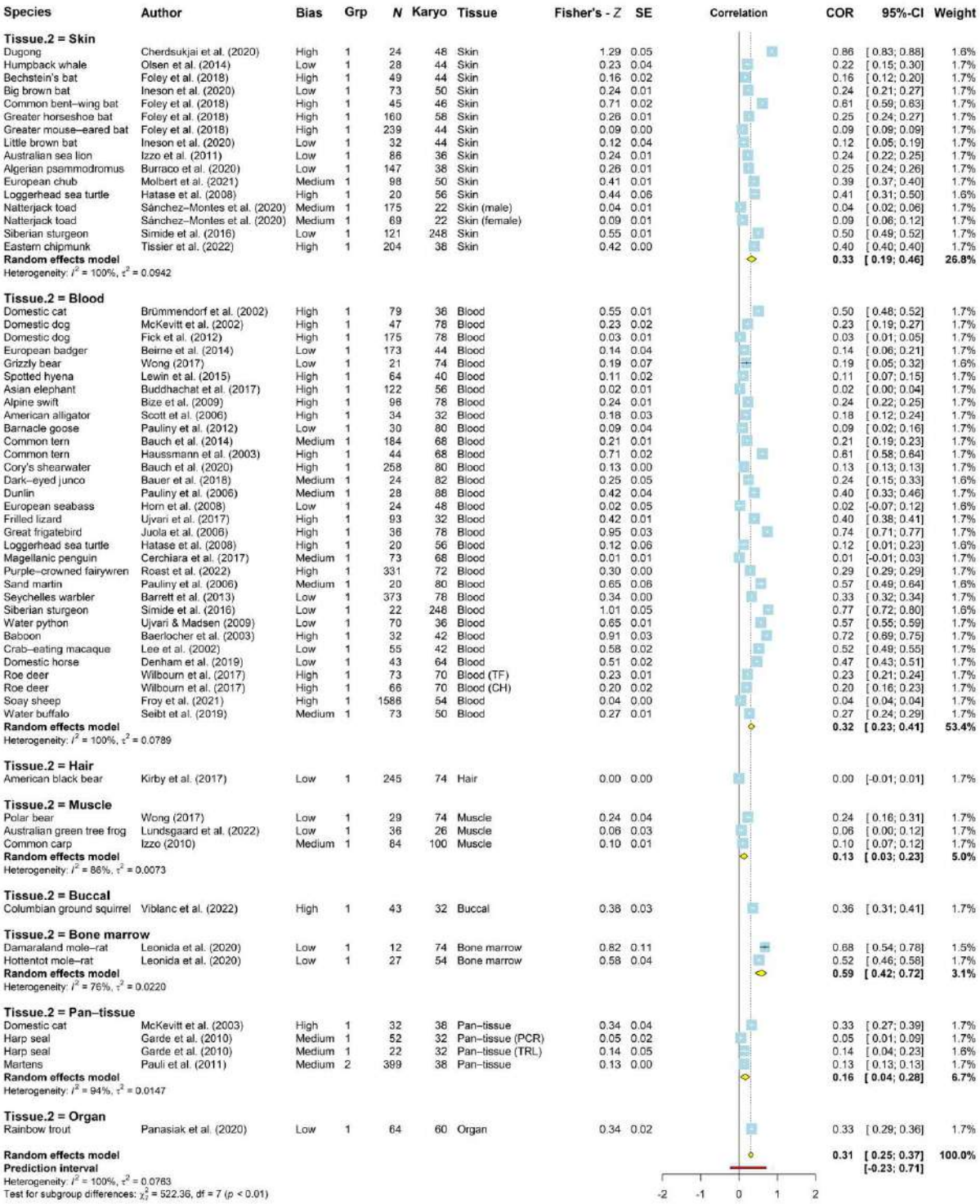


Fig. S9. Forest plot and results for the meta-analysis of studies using telomere length to estimate animal age with subgrouping based on study method. Key study attributes such as sample size (*N*), karyotype (Karyo), and tissue type are indicated along with the computed level of possible bias. Methods are labelled as follows: fluorescent *in situ* hybridization–flow cytometry (Flow-FISH); terminal restriction fragment (TRF) length & Southern blot; quantitative real-time polymerase chain reaction (qPCR); quantitative-fluorescent *in situ* hybridization (Q-FISH). CH, Chizé region, France; CI, confidence interval; COR, correlation; N, sample size; PCR, polymerase chain reaction; SE, standard error; TF, Trois-Fontaines region, France; TRL, terminal restriction length.

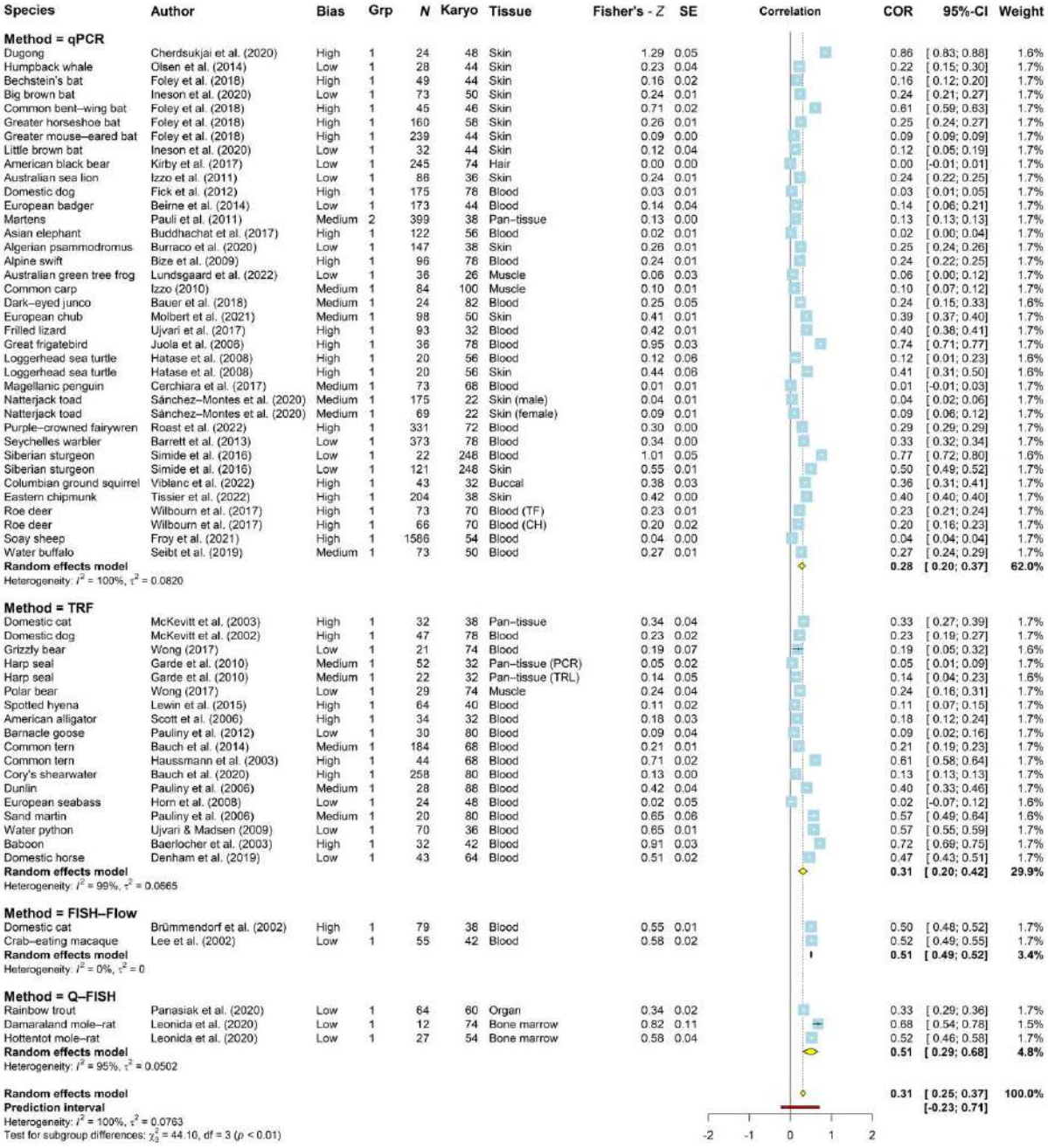
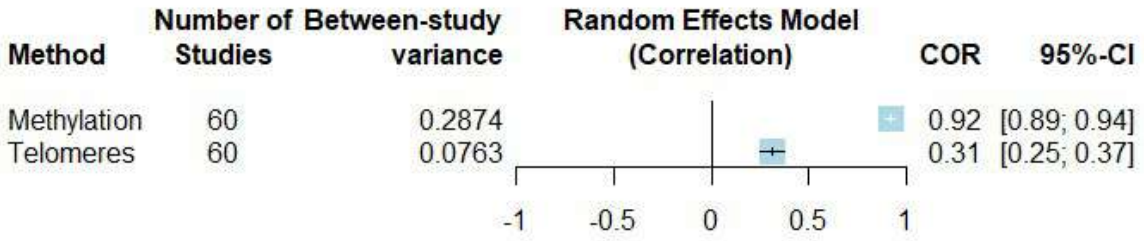


Fig. S10. Forest plot indicating and comparing the results obtained from independent meta-analyses of studies using methylation *versus* telomere length to model the relationship with ageing in animals. CI, confidence interval; COR, correlation.



LE CLERCQ, L. (2024) ABCal: a Python package for Author Bias Computation and Scientometric Plotting for Reviews and Meta-Analyses. *Scientometrics* **129**, 581–600.

SUPPLEMENTARY FIGURES

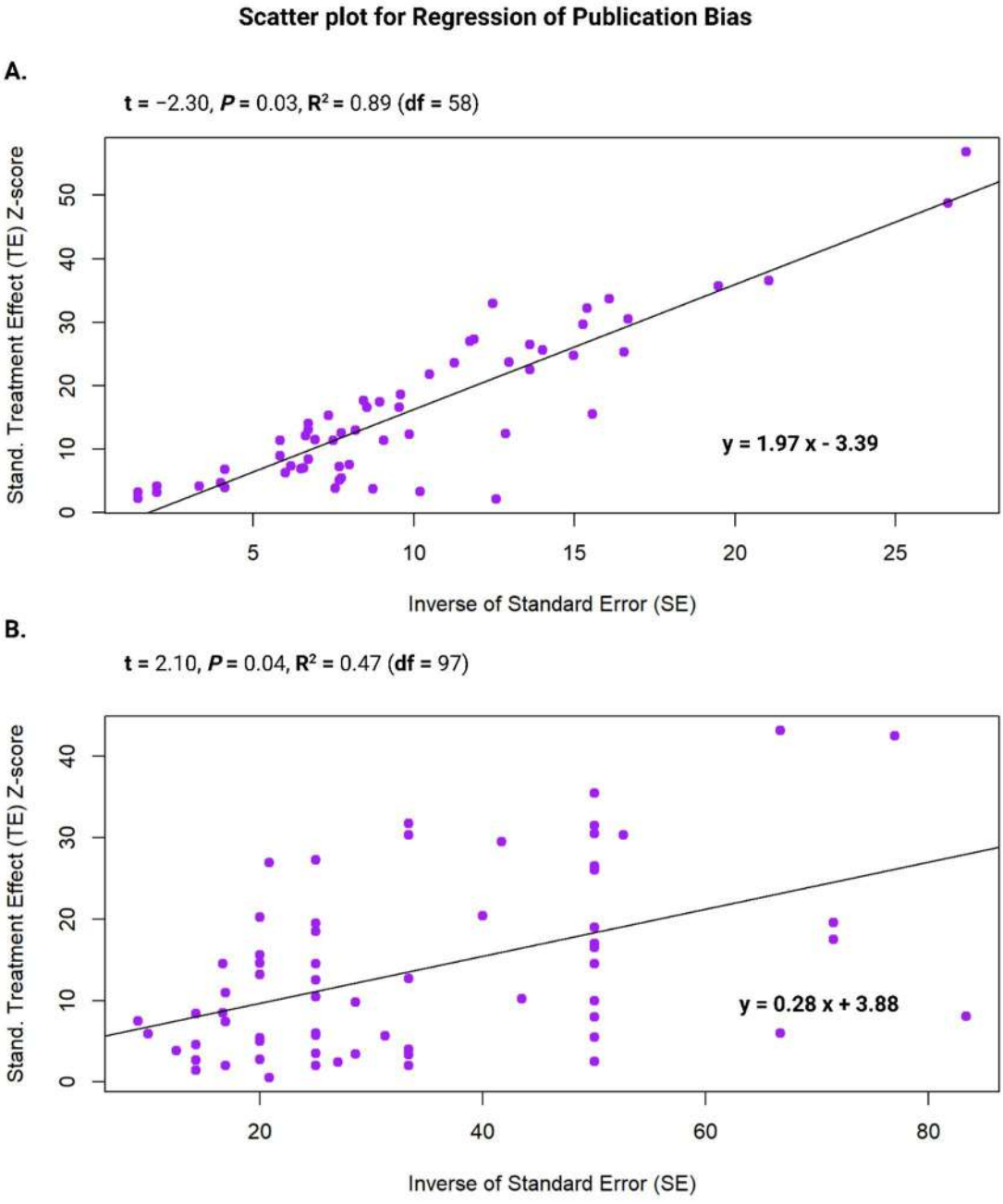


Figure S1: Scatter plots for the regression done to detect publication bias through funnel-plot asymmetry. The treatment effect (TE), Fisher's-Z, as standardised to Z-scores (y-axis) were plotted against the inverse of the standard error (x-axis) for individual studies (purple). The fitted regression line from the Egger's test is shown, along with the linear equation in the bottom right. **A.** For methylation studies, a strong correlation (R^2=0.89) was detected along with significant indicators of potential publication bias ($P < 0.05$). **B.** For telomere studies, a slightly weaker correlation was detected ($R^2 = 0.43$), however, there were still significant indicators ($P < 0.05$) for potential publication bias. (image created in BioRender.com)

LE CLERCQ, L., KOTZÉ, A., GROBLER, J.P. & DALTON, D.L. (2024*a*) Phenotypic correlates between clock genes and phenology among populations of Diederik cuckoo, *Chrysococcyx caprius*. *Ecology and Evolution* **14**, e70117.

SUPPLEMENTARY MATERIAL

Table S1: Details of samples used in the present study along with BioSample and GenBank accession numbers.

Catalogue nr	BioSample nr	Sex	Location	*Clock*	*Adcyap1*
60759CC	SAMN31832894	male	Uganda	OR909323	PP112922
60760CC	SAMN31832895	male	Uganda	OR909324	PP112942
60763CC	SAMN31832896	male	Uganda	OR909322	PP112934
46997CC	SAMN31832897	male	South Africa	OR909340	PP112940
46998CC	SAMN31832898	male	South Africa	OR909346	PP112927
46999CC	SAMN31832899	male	South Africa	OR909338	PP112941
47001CC	SAMN31832900	male	South Africa	OR909336	PP112933
47002CC	SAMN20928759	female	South Africa	OR909337	NA
47011CC	SAMN31832901	male	South Africa	OR909325	PP112918
47012CC	SAMN31832902	male	South Africa	OR909347	PP112929
63270CC	SAMN31832903	male	South Africa	OR909342	PP112943
63235CC	SAMN31832904	male	South Africa	OR909331	PP112935
63269CC	SAMN31832905	male	South Africa	OR909341	PP112926
63264CC	SAMN31832906	female	South Africa	OR909339	PP112945
63226CC	SAMN31832907	female	South Africa	OR909330	PP112931
63259CC	SAMN31832908	female	South Africa	OR909332	PP112932
63260CC	SAMN31832909	female	South Africa	OR909335	PP112939
65332CC	SAMN31832910	female	South Africa	OR909343	PP112938
65313CC	SAMN31832911	female	South Africa	OR909345	PP112946
65334CC	SAMN31832912	male	South Africa	OR909344	PP112937
53703CC	SAMN31832913	male	Nigeria	OR909333	PP112924
53713CC	SAMN31832914	male	Nigeria	OR909350	PP112925
53702CC	SAMN31832915	male	Nigeria	OR909351	PP112923
53706CC	SAMN31832916	male	Nigeria	OR909348	PP112921
53705CC	SAMN31832917	male	Nigeria	OR909334	PP112944
53711CC	SAMN31832918	male	Ghana	OR909349	PP112920
60771CC	SAMN31832919	female	Nigeria	OR909327	PP112936
60781CC	SAMN31832920	female	Nigeria	OR909329	PP112919
60776CC	SAMN31832921	male	Nigeria	OR909328	PP112928
60766CC	SAMN31832922	male	Ghana	OR909326	PP112930

Table S2: Details of sequence data retrieved for host species.

Host	Latin name	SRA accession	*Clock*	*Adcyap1*
Village weaver	*Ploceus cucullatus*	SRR17013387	Q_8/Q_8	139/141
Cape wagtail	*Motacilla capensis*	SRR5107518	–	–
		SRR5107510	–	–
		SRR5107439	–	–
		SRR5107432	–	–
		SRR5107354	–	–
Southern red bishop	*Euplectes orix*	SRX14011259	–	–
		SRX14011258	–	–
		SRX14010987	–	–
		SRX14010986	–	–
		SRX14010985	–	–

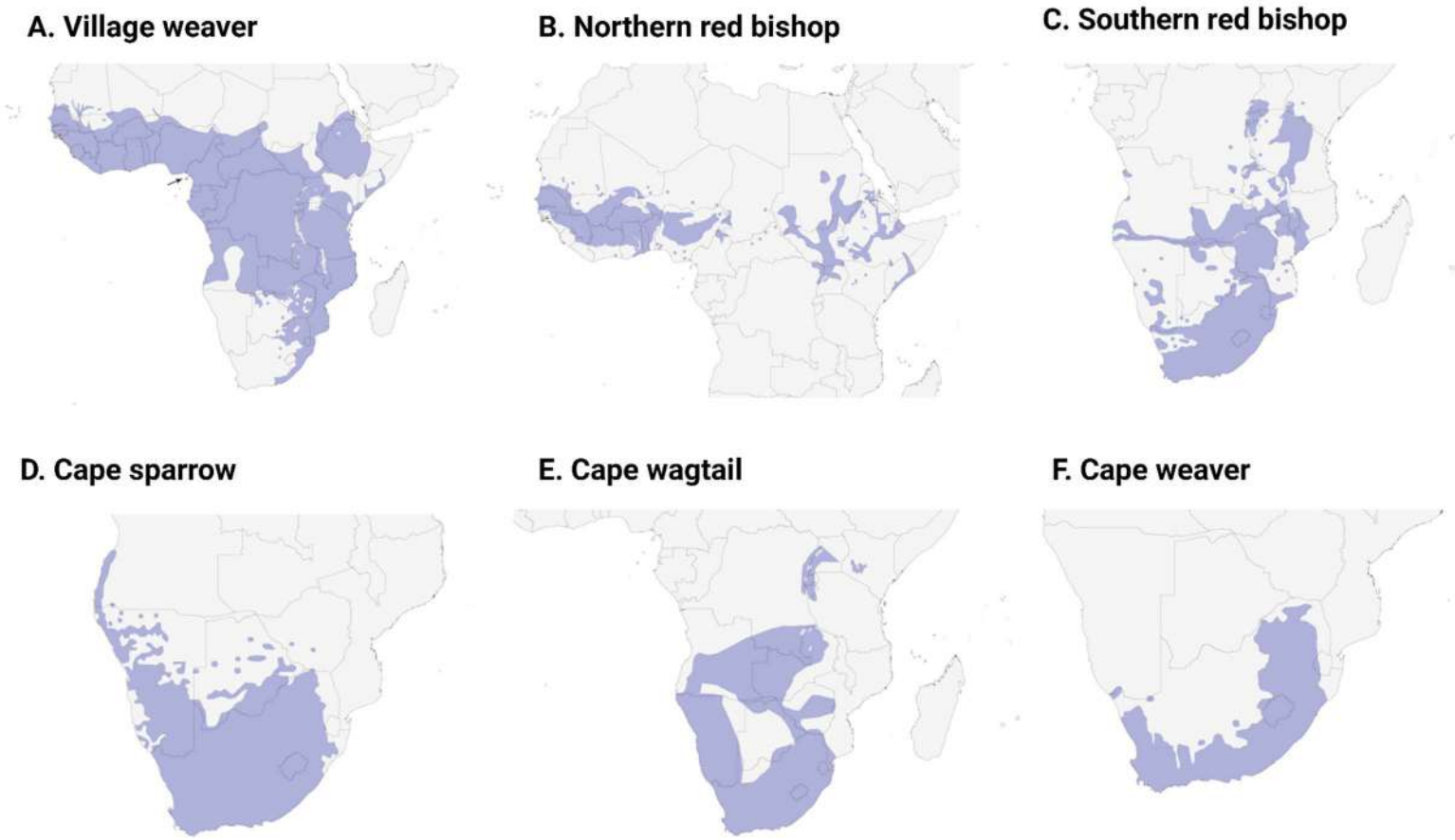


Figure S1: Host ranges for six bird species which are most commonly parasitised by Diederik cuckoo. In their Northern range they are known to parasitise the Village weaver (**A**) as well as the Northern red bishop (**B**), while in their Southern range they parasitise at least four species. In South-West Africa, they most commonly parasitise the Cape sparrow (**D**) and Cape wagtail (**E**), while in South-East Africa they additionally parasitise the Southern red bishop (**C**) and Cape weaver (**F**).

LE CLERCQ, L., KOTZÉ, A., GROBLER, J.P. & DALTON, D.L. (2024*b*) Methylation-based markers for the estimation of age in African Cheetah, *Acinonyx jubatus*. *Molecular Ecology Resources* **24**, e13940.

MOLECULAR ECOLOGY RESOURCES

Supplemental Information for:

Methylation-based markers for the estimation of age in African Cheetah, *Acinonyx jubatus*

Louis-Stéphane Le Clercq[1,2,*], Antoinette Kotzé[1,2], J. Paul Grobler[2], Desiré L. Dalton[3]

[1]*South African National Biodiversity Institute, Pretoria, 0001, South Africa.*
[2]*Department of Genetics, University of the Free State, Bloemfontein, 9300, South Africa.*
[3]*School of Health and Life Sciences, Teesside University, Middlesbrough, TS1 3BA, United Kingdom.*

*Corresponding Author: Louis-Stéphane Le Clercq (leclercq.l.s@gmail.com)

Table of Contents:

Table S1: Details for samples used in the present study.

Catalogue nr	NCBI BioSample nr	Sex	Age (months)	Age Class
49153	SAMN31850912	female	96	Old Adult
49922	SAMN31850913	female	24	Independent
49930	SAMN31850914	male	24	Independent
50461	SAMN31850915	female	60	Prime Adult
51137	SAMN31850916	male	4	Young Cub
51146	SAMN31850917	male	96	Old Adult
52316	SAMN31850918	male	84	Prime Adult
52317	SAMN31850919	female	24	Independent
52353	SAMN31850920	male	48	Young Adult
52387	SAMN31850921	male	96	Old Adult
52608	SAMN31850928	female	14	Adolescent
52616	SAMN31850930	female	27	Independent
52986	SAMN31850931	male	108	Old Adult
53387	SAMN31850932	male	36	Young Adult
53388	SAMN31850933	female	36	Young Adult
53389	SAMN31850934	female	36	Young Adult
53393	SAMN31850935	male	144	Very Old Adult
53933	SAMN31850936	female	4	Young Cub
55031	SAMN31850937	female	108	Old Adult
55040	SAMN31850938	female	24	Independent
55044	SAMN31850939	female	120	Old Adult
55061	SAMN31850940	female	72	Prime Adult
55235	SAMN31850941	female	3	Young Cub
55237	SAMN31850942	male	3	Young Cub
55238	SAMN31850943	female	3	Young Cub
59394	SAMN31850944	male	24	Independent
59397	SAMN31850945	male	24	Independent
60702	SAMN31850946	female	84	Prime Adult
61437	SAMN31850947	female	168	Very Old Adult
61863	SAMN31850948	male	120	Old Adult
61886	SAMN31850949	male	18	Adolescent
61899	SAMN31850950	male	3	Young Cub
61900	SAMN31850951	female	4	Young Cub
61901	SAMN31850952	male	3	Young Cub
62532	SAMN31850953	female	159	Very Old Adult
63351	SAMN31850954	male	3	Young Cub
64684	SAMN31850955	female	42	Young Adult
65180	SAMN31850956	male	60	Prime Adult
65985	SAMN31850957	male	2	Young Cub
66339	SAMN31850958	female	1	Young Cub
66340	SAMN31850959	male	1	Young Cub
66735	SAMN31850960	female	84	Prime Adult
51135	SAMN31850961	female	96	Old Adult
52540	SAMN31850922	female	108	Old Adult
52568	SAMN31850923	male	48	Young Adult
52569	SAMN31850924	male	108	Old Adult
52571	SAMN31850925	male	96	Old Adult
52572	SAMN31850926	male	48	Young Adult
52588	SAMN31850927	male	120	Old Adult
52613	SAMN31850929	female	28	Independent

2

Table S2: Genes and primers used in the present study.

Gene	Name	Forward primer	Reverse primer
EDARADD	*Extodysplasin A receptor associated death domain* (NCBI Gene ID: 106982554)	5'-TTTGAAAGGTTAATTAGTTTTATGGT-3' **T_m:** 59.2°C	5'-CAACCCAAAACTCAACTTTAAAAAA-3' **T_m:** 66.1°C
ELOVL2	*Elongation of very long fatty acids 2* (NCBI Gene ID: 106972032)	5'-TTTAGAGTTAAGTATTGGTTTGGGAAA-3' **T_m:** 61.2°C	5'-TCCAACCCCCTATTACAAACAC-3' **T_m:** 67.3°C
FHL2	*Four and half LIM domains protein 2* (NCBI Gene ID: 106977955)	5'-GGGGAGTTTATAGTTTGGGG-3' **T_m:** 63.6°C	5'-ACTTCCCTCTCCAAAAATCTC-3' **T_m:** 66.4°C
GRIA2	*Glutamate ionotrophic receptor AMPA type subunit 2* (NCBI Gene ID: 106986586)	5'-TATGGGAGGGTGTTGAATATTTTAAGGT-3' **T_m:** 63.3°C	5'-CATTTTAAAACAACAATTTAACACCAA-3' **T_m:** 65.4°C
ITGA2B	*Integrin subunit alpha 2b* (NCBI Gene ID: 106974321)	5'-TGTGTGGTTTTATGTTTGTATTTTTTT-3' **T_m:** 60.1°C	5'-TCCATCTCAAACCACATCTTAATATAAC-3' **T_m:** 65.7°C
PENK	*Proenkephalin* (NCBI Gene ID: 106982523)	5'-TTAAGATTTGGGAAATTTGTAAGGA-3' **T_m:** 60.3°C	5'-CCAATTAAAAACTACCCTTCAAAAC-3' **T_m:** 65.9°C

4

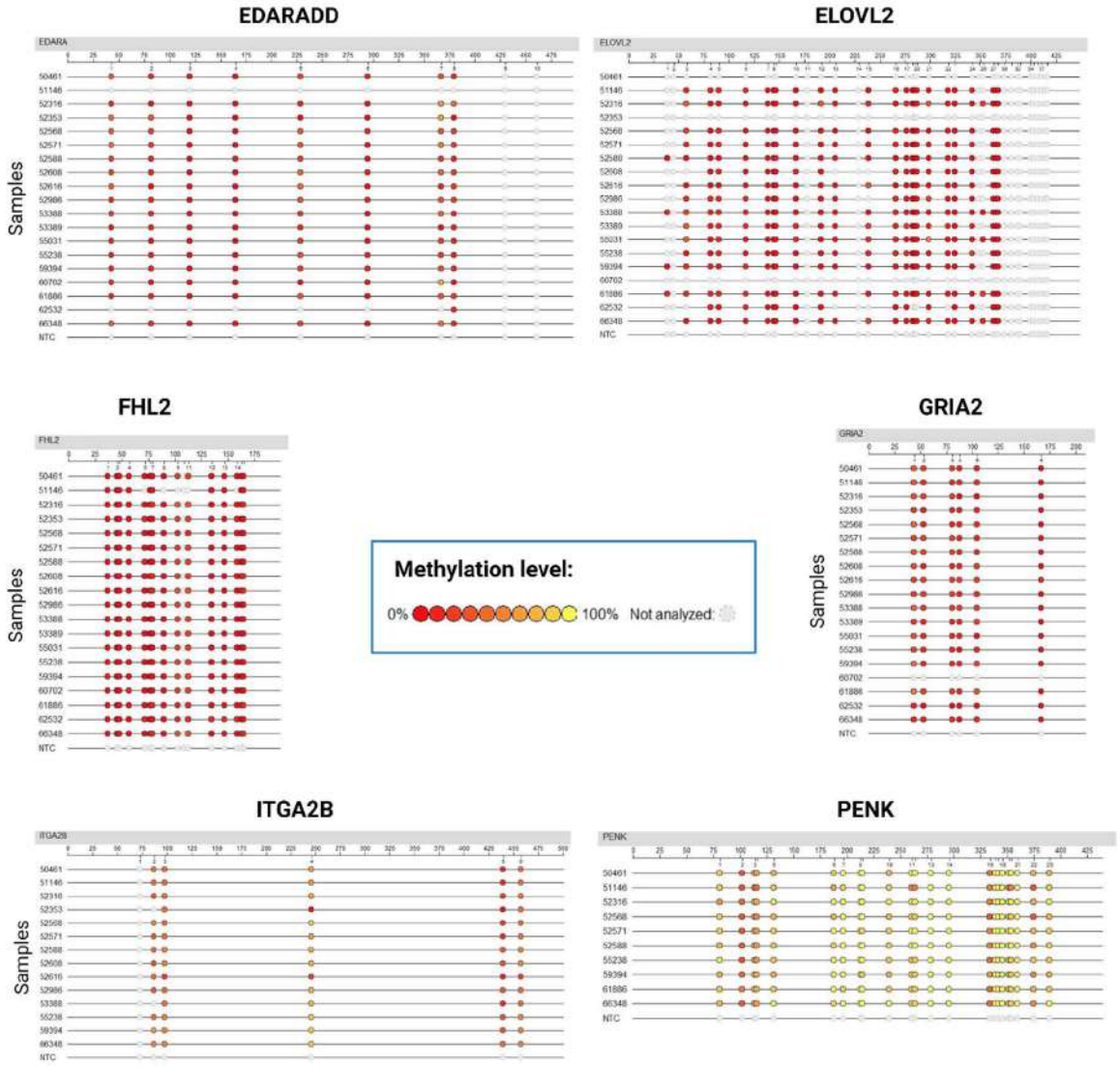
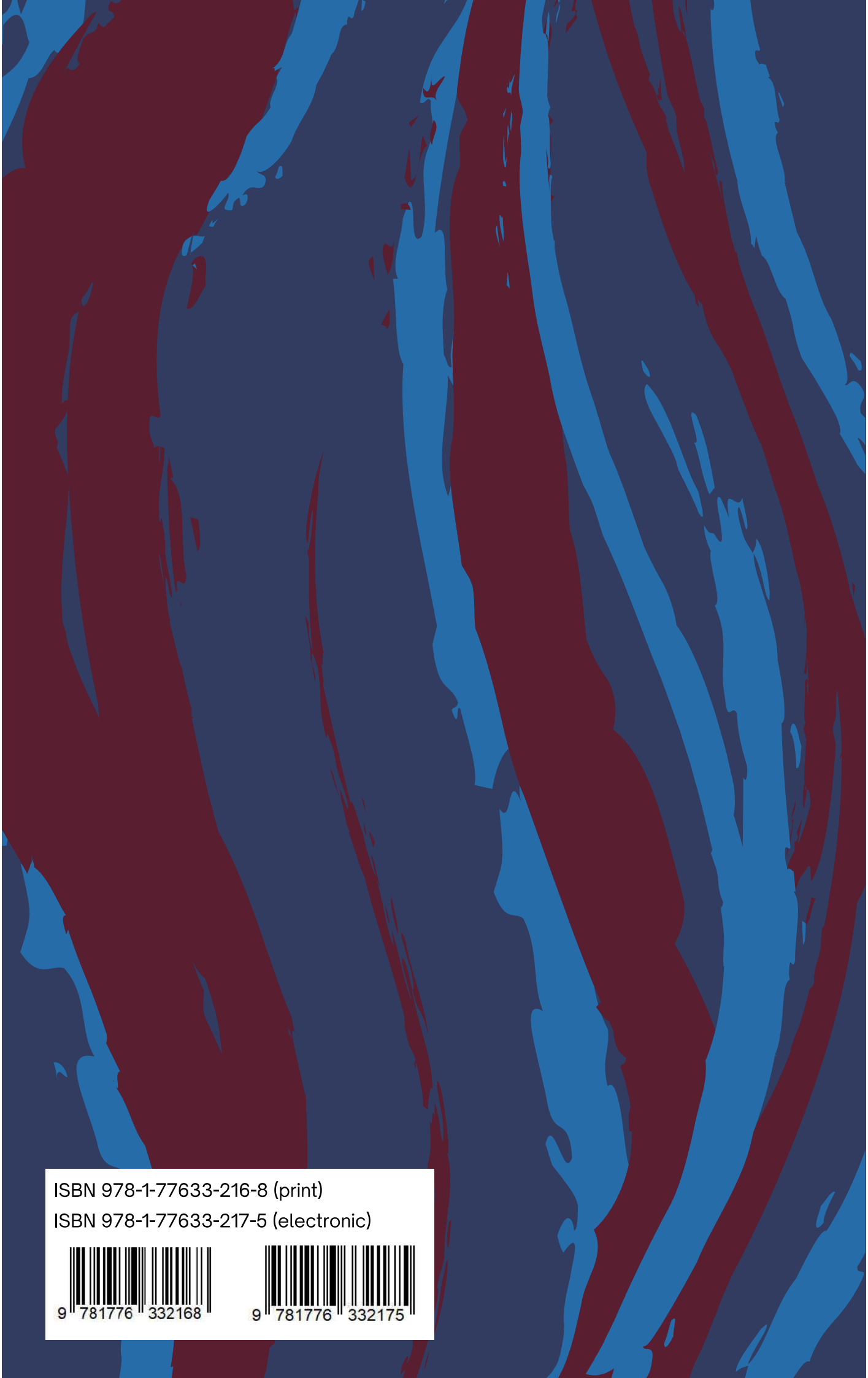


Figure S1: Examples of EpiGrams generated for target genes. Each line indicates an individual sample. The bar under the gene name indicates the total length of the amplified fragment as well as the numbering of CpG's on the fragment. Circles on a line indicate the position of a specific CpG and is colour coded from red (0%) to yellow (100%) to indicate the level of methylation that was detected. Grey circles represent CpG's that could not be assayed. For each gene the last sample, indicated as NTC, represents the unmethylated negative control included with the run.



ISBN 978-1-77633-216-8 (print)
ISBN 978-1-77633-217-5 (electronic)