A phylogenetic analysis of the tent tortoise *Psammobates tentorius* (Bell, 1828) species complex, using molecular and morphological markers

A thesis submitted for the degree of Doctor of Philosophy
University of the Free State
November 2019

Zhongning Zhao
(Student number: 2008085140)

Department of Zoology and Entomology, Faculty of Natural & Agricultural Sciences

Promoter: Prof Neil Heideman (Department of Zoology and Entomology)

Co-promoter: Prof Paul Grobler (Department of Genetics)

A phylogenetic analysis of the tent tortoise *Psammobates tentorius* (Bell, 1828) species complex, using molecular and morphological markers

ABSTRACT

The rapid development of advanced DNA analytical programs, algorithms and techniques have greatly strengthened molecular systematics in the discovery and clarification of biodiversity, as well as the accuracy of taxonomies and reconstructing the relationships among taxa or operational taxonomic units (OTUs). Furthermore, these sophisticated analytical approaches enhanced traditional systematic biology in terms of temporal and spatial dimensions, thereby bridging the gap between the past, present and future. Phylogeographical analyses allow proper interpretation of how the distribution of organisms correlates with their cladogenesis, while biogeographical analyses enable the determination of how changes in geographic features and systematics are correlated. Lastly, utilizing molecular markers in calibration dating with sophisticated current analytical models enable the tracing of radiation histories, while diversification rate analyses allow the predicting of likely future cladogenic scenarios, which are crucial for conservation management. Modern molecular systematics also makes possible the testing of alternative phylogenetic reconstruction scenarios via computational simulations. All these methodological advances overcome the limitations of traditional phylogenetic analysis in developing evolutionary and cladogenic scenarios, and strengthen our ability to predict and prepare for likely future biodiversity patterns.

Chapter 1 provides the general introduction to the study, which focused on the highly polymorphic and taxonomically confusing tent tortoise (*Psammobates tentorius*) (Testudines: Testudinidae) species complex from southern Africa and had six broad objectives. First, to identify its clades and infer phylogenetic relationships among them, using mitochondrial and nuclear DNA sequence data, as well as microsatellite DNA data. Second, to use the phylogenetic assumptions to delineate the clades into a proper OTUs scheme, which will be crucial for the taxonomic revision of the complex. Third, to elucidate the radiation history of the complex against temporal and spatial parameters, using phylogeographic and biogeographic analyses. And then to use diversification rate analyses to elucidate cladogenic rates and predict future radiation trends, which could benefit conservation management efforts. Fourth, using DNA markers to determine genetic structure and diversity between and within

clades, which should also benefit the conservation of the complex. Fifth, to use computational simulations to determine alternative systematics scenarios, which could potentially explain the radiation history from a more explicit evolutionary viewpoint. Sixth, to use morphometric measurements and morphological character analyses to find diagnostic characters for the different candidate species needed for reviewing the taxonomy of the species complex.

Chapter 2 reports on the investigation of genetic structure, phylogenetic relationships and genetic diversities within the *P. tentorius* complex, which resulted in the discovery of seven clades (C1-C7). The mtDNA and nDNA datasets generated conflicting tree topologies. The mtDNA dataset retrieved seven clades, each associated with a distinct geographic region. The nDNA dataset advocated three lineages generally corresponding with the three currently recognized subspecies, although potential hybridization was found between C1 and C2, and between C2 and C4. The multiple species delimitation analyses and the traditional *p*-distance method generally supported 6 to 7 putative species corresponding to the seven clades, although most of the analyses suggested six putative species. It was found that the currently recognized *Psammobates tentorius verroxii* was not a monophyletic group, but consisted of two candidate species, so its current taxonomic status should be revised. The uniformly brown coloured "*Psammobates bergeri*" mentioned in the literature should not be considered a valid taxon, since it has been found in three clades. The "*Psammobates tentorius trimeni*"- like population found in the Kamiesberg, Hantam Karoo and Roggeveldberge region is genetically the close relative of C1, but distant relative of the real *P. t. trimeni* occurring on the west coast of South Africa.

The findings in Chapter 3 suggest that divergence in the *P. tentorius* complex is deep, with most of the cladogenic events occurring in the Miocene, while radiation within the species is linked to the Pliocene and Pleistocene. The results of the biogeographic, diversification rate and character dependency analyses suggest that the cladogenic radiation of the *P. tentorius* complex was shaped by climatic and topographic changes since the Miocene. The calibration dating results revealed that the cladogenic divergences in the *P. tentorius* complex were comparable to that found in many other genera. This implies that some of the clades may deserve to be elevated to species level.

Chapter 4 reports on the comparison of the phylogenetic topologies generated from the nDNA, mtDNA, mtDNA+nDNA and microsatellite DNA datasets. The results showed that the tree

topologies did not differ greatly between the sequence and microsatellite datasets, despite the incongruence between the mtDNA and nDNA datasets as indicated in Chapter 2. The Bayesian computational simulation results advocated an alternative systematic scenario, different from the best scenario retrieved with the traditional phylogenetic analyses. It was found that this scenario more plausibly explained the cladogenic prosesses of the *P. tentorius complex*. The microsatellite based multivariate analyses and clustering analyses generally suggested 4-6 putative species for the complex, with the difference between C1 and C4 not clear. The findings are generally congruent with the phylogenetic and species delimitation results reported for the sequence datasets in Chapter 1.

Chapter 5 reports on the use of selection analysis, based on codon and dN/dS ratios in the protein coding genes Cyt-b, ND4 and PRLR. The results suggested a unique selection pattern for group C2+C3, which is believed to have been driven by intensification of aridity during the development of the Benguela Current, 8-10 Mya. However, the dN/dS-based assumption should only be regarded as a reasonable assumption or hypothesis, as further long-term ecological experiments or simulations based on historical climate datasets are needed to investigate the real scenario. No new putative species were detected by the phylogenetic inference using codon and amino acid datasets. Nonetheless, codon-based phylogeny, transition and transversion rate-based topologies and phylogenetic reconstruction using amino acid sequences provided additional support for confirming the validity and stability of the seven clades.

Chapter 6 should be considered as preliminary, as sufficient samples of C5, C6 and C7 could not be obtained to allow reliable morphometric and phenotypic analyses. Despite the inadequate and uneven samples, the results still showed clear sexual dimorphism, and also suggested that morphological characters could be used to distinguish among the clades. The discrete (phenotypic) and continuous (morphometric) character datasets both showed clear separation between the two major branches “C1+C4+C5+C7” and “C2+C6”. Four clusters were distinguishable: C2, C3, C6 and (C1+C4+C5+C7). These results showed a good match with the microsatellite DNA results of Chapter 4.

Chapter 7 is the concluding chapter which summarizes the findings of the study along with recommendations.

ACKNOWLEDGEMENTS

The challenging nature of this research project was beyond what any herpetologist could ever have imagined, so I am indebted to many colleagues, individuals and organizations, without whom it would not have been completed by now. My sincere appreciation for the support and assistance from every corner of the network, your support was crucial.

First, I want to thank my family, especially my parents, for giving me the freedom to pursue this project, which has always been my dream, together with the endless support and inspiration which encouraged and spurred me on. I would like to thank my best friend, the late Tu Bo, who enlightened me about the concept of the radiation of the diversity of life, which served as the major driving force enabling me to conquer all the difficulties and achieve this goal.

Special thanks go to my promoter Prof Neil Heideman, who provided adequate resources, and unlimited support and guidance throughout the project. I could not have expected more from him.

I would also like to thank my co-promoter, Prof Paul Grobler, for his guidance and for providing access to a world-class genetics lab, incuding reagents, sequencing and genotyping facilities.

A word of deep gratitude to Prof Margaretha D. Hofmeyr, Professor Emeritus, Department of Biodiversity and Conservation, University of the Western Cape, for providing extensive samples collected over many years spent in the field. This was of great help as collecting tent tortoises is a difficult task, given their crypticity and occurrence at very low densities throughout their distribution range. Her willingness to share her vast expertise on the animal also contributed significantly to the success of the project.

Phillip Bester from the Departmnet of Virology, University of the Free State, is thanked for his outstanding assistance with the use the R packages, as well as assisting with some of the sequencing. Jaco Oosthuizen from the same department is thanked for assisting with most of the microsatellite genotyping work.

Thanks to Adriaan Jordaan, for all the field assistance, long distance driving, organizing field sites with farmers and making world-class maps. This project would not have been completed without your inputs. All our fieldtrip experiences became memorable stories about herping in South Africa, and made the trips more of an adventure than hard work.

I thank Prof Krystal Tolley from the South African National Biodiversity Institute (SANBI) and Werner Conradie from Port Elizabeth Museum for providing additional samples for the DNA analyses. Theo Busschau, David Maguire, Jonathan Palmer, Gary Nicolau, Luke Kemp, Tyrone Ping, Marius Burger, Andries Cilliers and Zandri Van Der Westhuizen are thanked for assisting with tissue sample collections. Sincere gratitude goes to all the farmers who allowed access to their properties for sampling and for being great sources of information regarding the areas on their farm to search. Stephanus Riekert from High-Performance Computing Centre of the University of the Free State is thanked for assisting with the phylogenetic analyses. I also thank Dr. Joaquín Verdú-Ricoy, Postdoctoral Research Fellow from the University of Completense in Madrid, Spain, for assisting with the text editing.

I thank the National Research Foundation (Grant - IFR150216114248), and the Universities of the Free State (UFS Research Grant: A1999/158110) and the Western Cape (SNS Grants) for providing the funding to carry out the study. I also thank the following provincial departments of Nature Conservation for providing collecting permits: Northern Cape Province (Permits No: FAUNA 1061/2/2015, FAUNA 1266/2016, FAUNA 1458/2015, FAUNA 1267/2016, FAUNA 0729/2018, FAUNA 0730/2018), Western Cape Province (Permit No: AAA007-00179-0056) and Eastern Cape Province (Permits No: CRO 171/16CR, CRO 172/16CR), as well as the Ministry of Environment and Tourism, Namibia (Permit No: 1430/2009).

The Animal Research Ethics Committees of the University of the Free State (AREC References No: 180111-005, UFS-AED2015/0013) and University of the Western Cape (ScRiRC2008/39) are thanked for providing ethical clearance for the research.

Finally, I have to thank South Africa, the world's most beautiful country for its mesmerizing biodiversity and diverse cultures. It made herping an absolutely amazing and stupendous experience. Doing fieldwork particular in Namaqualand and the Cape Floral Kingdom, especially during spring, was special, given their highly colourful floral beauty at that time.

Table of contents

List of Figures

List of Supplementary Figures

List of Tables

List of Supplementary Tables

(Supplementary Tables will only be provided independently as supporting documents)

Table S2.5. The uncorrected *p*-distance matrix for the Cyt-b gene. Within-group pairwise *p*-distances are underlined. All pairwise *p*-distances within *Psammobates tentorius* group are in bold.

Table S2.6. The average number of pairwise differences among the seven clades retrieved from Arlequin analyses results. The total number of nucleotide bases used in the analyses was 2437 bp (from mtDNA dataset).

Table S2.7. Summary of the multiple species' delimitation approaches used in the study (the detail given in the text and Table 1). The retrieved number of putative species and speciation events are shown. The criterion used to evaluate the different approaches, the Rtax was calculated following the formula generated from Miralles & Vences 2013.

Table S3.1. List of all samples, their corresponding localities and NCBI GenBank accession numbers across different genes.

Table S3.2. GenBank accession numbers for all outgroups used in the study. Also, the six calibration nodes with their mean and range as used in the study, together with all references.

Table S3.3. Primers used in the study with corresponding oligo sequences, optimized annealing temperatures and source.

Table S3.4. Optimal partition scheme, substitution model, likelihood score (-InL), Gamma shape, proportion of estimated invariance and Homogeneity Test results.

Table S3.5. The BioGeoBEARS model test results tested for six different habitat reconstruction models with consideration of the "founder effect" parameter "J". The selected best model with its criteria is shown in bold.

Table S3.6. The ANOVA based LRT test results retrieved from the likelihood function models of different scenarios investigated under character dependency analyses BiSSE and MuSSE with different objects of Regions, Biomes and two sides of the Orange River.

Table S4.1. List of all samples, their corresponding localities and NCBI GenBank accession numbers across different genes.

Table S4.2. Primers of the sequence datasets used in the study with corresponding oligo sequences, optimized annealing temperatures and sources.

Table S4.3. Allele size ranges, repeat motifs with their NCBI GenBank accession numbers, Multiply-mix reaction grouping schemes, the oligo-nucleotide sequences of primers with dyes, optimal

annealing temperatures and the sources of primers of all microsatellite DNA markers tested in this study.

Table S4.4: Optimal partition scheme, substitution model, likelihood score (-InL), Gamma shape, proportion of estimated invariant and Homogeneity Test results.

Table S4.5. GenBank accession numbers for all outgroups used in the study. Also, the seven calibration nodes with their mean and range as used in the study with literature.

Table S4.6. The genotyping results across all 19 microsatellite DNA loci.

Table S4.7. The genotyping results of 14 microsatellite DNA loci used in this study, together with locality and subpopulation information.

Table S4.8. SAMOVA analysis results for the mtDNA and microsatellite DNA datasets across all scenarios (tests with different cluster schemes from K=2 to K=12). " * " indicates $p < 0.05$, " ** ", $p < 0.01$ and " *** " $p < 0.001$.

Table S5.1. List of all samples used in Chapter 5, their corresponding localities, geographic coordinates and NCBI Genbank accession numbers across different genes.

Table S5.2. GenBank accession numbers for all outgroups used in the study. 222

Table S5.3. The transition rate (Ti), transversion rate (Tv) and selection strength Tv/Ti across all nodes; Neighbour-joining tree used as backbone tree. 223

Table S6.1. The list of all continuous morphometric measurements and ratios used in this study, together with detail information on specimens, clade belonging, location collected, Country, sex and life stage. "F" = female, "M" = male. "PEM R" represents specimens deposited in Port Elizabeth Museum, "ZPT" represents specimens measured from this study, "ZR" represents specimens stored in Cape Town Iziko Museum.

Table S6.2. The list of all discrete phenotypic characters has been used in this study, together with detail information on specimens, clade belonging, location collected, Country, sexes and life stage. The detail on the approaches on coding the characters explained in text in Chapter 6. "F" = female, "M" = male. "A"= Adult, "J" represents juvenile. "PEM R" represents specimens deposited in Port Elizabeth Museum, "ZPT" represents specimens measured from this study, "ZR" represents specimens stored in Cape Town Iziko Museum, "SMN" represents specimens stored in Namibia Windhoek Museum. All other codes in specimens' ID represent collections from local municipality museums.

Table S6.3. A list of all informative morphometric ratios retrieved from ANCOVA analyses results, effects for sex, clade and interactions between were given, "df" = degree of freedom, "F" = F-value; for sex: "F" = females, "M" = males. The significant p-values were bolded.

Table S6.4. A list of all informative morphometric ratios retrieved from Two-way factorial ANOVA analyses. "df" = degree of freedom, "F" = F statistic value, "p (females VS males)" represents p-values for post-hoc test, "F" = females, "M" = males.

Table S6.5. The all informative discrete phenotypic characters retrieved from Two-way factorial ANOVA analyses. "df" = degree of freedom, "F" = F statistic value, in sexes: "F" = females, "M" = males.

Table S6.6 The summary on results retrieved from multivariate analyses approaches PERMANOVA, DFA with continuous and discontinuous datasets with sexes analyzed independently, to determine the most possible clustering scheme.

Chapter 1. General Introduction

1.1 General description

South Africa is the only country that has three officially recognized biodiversity hotspots, the highest number in the world (Mittermeier et al., 1999; Myers et al., 2000). The country not only has exceptional diversity with respect to fauna & flora, but also a variety of climatic zones, biomes, vegetation types and ocean currents (Mucina & Rutherford, 2011).

Southern Africa is recognized world-wide as the biodiversity hotspot for genera and species of the family Testudinidae (Cunningham, 2002; Branch, 2008; Hofmeyr et al., 2014). It is the home of 13 species of tortoise in five different genera. The 13 species make up almost 30% of the currently recognized number of species globally (45 species) and the 5 genera, 33% of the world total (15 genera) (Alexander & Marais 2007; Branch 2008). Among these South African tortoises, 10 are considered endemic or nearly endemic (species endemism = 76.9%) (Alexander & Marais, 2007; Branch, 1998 & 2008; Hofmeyr et al., 2014)

Among this remarkable tortoise diversity, arguably the most attractive genus is *Psammobates* (Fitzinger, 1835), which contains the rarest tortoise species, namely, the geometric tortoise, *Psammobates geometricus* (Linnaeus, 1758), that is listed as "Critically Endangered" in the IUCN list (IUCN, 2018). The tent tortoise, *Psammobates tentorius* (Bell, 1828) has carapace and plastron patterns which vary greatly among populations (Hewitt, 1933 & 1934; Branch, 1998 & 2008). Three subspecies are currently assumed by many herpetologists (Branch, 1998 & 2008; Hofmeyr et al., 2014). The serrated tent tortoise, *Psammobates oculifer* (Kuhl, 1820), which is common and widespread in the Kalahari region shows similar inter-population variation. According to the recent IUCN evaluation conducted by Hofmeyr et al. (2018a), the overall *P. tentorius* species complex was considered as "Near Threatened". *Psammobates tentorius trimeni* was elevated to the "Endangered" category (Hofmeyr et al., 2018b). Several studies of its ecology and population genetics have been conducted for purposes of conservation (Baard, 1995; Cunningham et al., 2002), while studies of its ecology and thermoregulation have also been carried out (Keswick, 2012). By contrast, very little research has been done on the tent tortoise, at either phylogenetic, population or ecological level (Cunningham, 2002; Leuteritz & Hofmeyr, 2007). The only formal study carried out thus far

dealt with the prolonged reproductive period and ecology of *Psammobates tentorius verroxii* (Leuteritz & Hofmeyr, 2007). The highly variable colouration and carapace patterns between different tent tortoise populations ("subspecies") offer a great opportunity to study the influence of South Africa's biome and climate heterogeneity on the evolution of this species.

1.2 Distribution

Tortoises are widespread reptiles occurring in southern Europe, northern Africa, the Middle East, southern Russia, Asia, North America, sub-Saharan Africa and on some islands adjacent to South America, but are absent from Australia and New Zealand (Swingland & Klemens, 1989; Buhlmann et al., 2009). All tortoise species were categorized as CITES II or above, by an international convention that regulates animal trade (Bour et al., 2008).

The tent tortoise, *P. tentorius*, is endemic to South Africa and Namibia (Alexander & Marais, 2007). It has a wide distribution range (Fig. 1.1) that stretches from south-eastern Namibia in the north, to the succulent Karoo region along the West Coast, to Matjiesfontein in the south (Branch, 1998 & 2008), into the Little Karoo, the northern slope of the Langeberg and surrounding valleys, and eastward into Fort Brown and Grahamstown (Hofmeyr et al., 2014). The species is not present on the northern-eastern side of Orange River, but a population is found on the north-western side of the Orange River (Branch, 1998 & 2008; Hofmeyr et al., 2014).

The three subspecies have overlapping ranges and intergrade with each other, their exact distribution range limits having not been clearly verified as yet (Hofmeyr et al., 2014). Among the three subspecies, the typical subspecies is the Karoo tent tortoise, *P. t. tentorius*, which occupies the southern part of the range, and is most distinct in the south-east. Its distribution range extends from Grahamstown and Fort Brown in the Eastern Cape Province to the Langeberg region (Touws River, Matjiesfontein and Montagu) in the Western Cape Province (Branch, 1998 & 2008; Hofmeyr et al., 2014). The northern edge of its range extends into the Kamiesberg area and Victoria West in the Northern Cape Province (Griffin, 2003; Hofmeyr et al., 2014).

The intergrading subspecies, Verreaux's tent tortoise, *P. t. verroxii*, has the widest distribution range of all three subspecies (Fig 1.1). It occurs throughout almost the entire Northern Cape

Province, extending northward into southern Namibia (Branch, 1998 & 2008; Hofmeyr et al., 2014). To the west *P. t. verroxii* extends into the West Coast Namaqualand area, reaching Ceres in the Western Cape Karoo (Griffin, 2003). The Namaqua tent tortoise, *P. t. trimeni*, occurs north of the Orange River at Helmeringhausen in south-western Namibia (Griffin, 2003). South of the Orange River it occurs exclusively in Namaqualand along the West Coast region (Fig. 1.1) of South Africa (Branch, 2008; Griffin, 2003). Incorrectly identifying specimens of the three tent tortoise subspecies is quite common and largely unavoidable due to the exceptional carapace and plastron colour pattern variations.

The ecotone areas between the subspecies are also particularly remarkable, between *P. t. tentorius* and *P. t. verroxii*, and *P. t. verroxii* and *P. t. trimeni*. The main overlapping zone between *P. t. tentorius* and *P. t. verroxii* includes the southern part of the Great Karoo and the border region between the Western Cape and Northern Cape Provinces. In the case of *P. t. trimeni* and *P. t. verroxii,* they have sympatric ranges throughout almost the entire Namaqualand area of the West Coast (Griffin, 2003; Hofmeyr et al., 2014)

The distribution range of the tent tortoise covers the Nama Karoo (most of the Northern Cape Province), Succulent Karoo (almost the entire West Coast area), Fynbos and Albany Thicket (the last two constituting the major biomes in the southern part of its range (Branch, 2008; Mucina & Rutherford, 2011; Hofmeyr et al., 2014). They were considered as occuring at low densities (Branch, 1998; Alexander & Marais, 2007; Branch, 2008; Hofmeyr et al., 2014), and their densities possibly underestimated due to their small size, and incredibly effective camouflage resulting from their carapace patterns (Branch, 1998 & 2008; Cunningham, 2008; Hofmeyr et al., 2014). They are therefore seldomly encountered and relatively little research has been conducted on them, published work being only that of Boycott & Bourquin (2000), Cunningham (2008) and Hofmeyr et al. (2017).

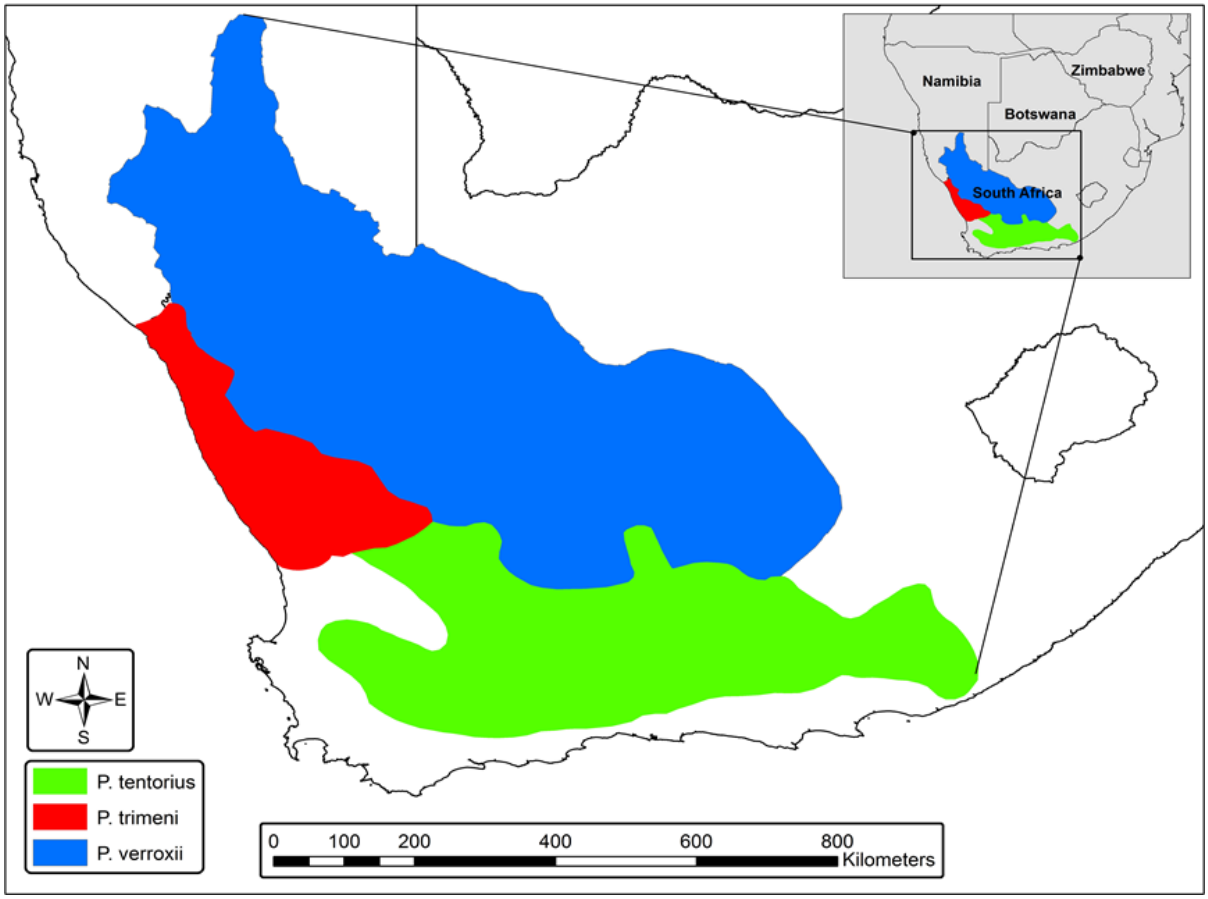


Figure 1.1. Map showing the interpreted distribution ranges of the three currently recognized subspecies of *Psammobates tentorius*, based on data from various sources: *P. t. tentorius* (green area), *P. t. trimeni* (red area) and *P. t. verroxii* (blue area). The data are a combination of own records, Cape Town Iziko South African Museum, Port Elizabeth Bayworld Museum and Pretoria Ditsong Museum of Natural History records, as well as records provided by the Animal Demographic Unit (ADU), University of Cape Town.

1.3 Morphology

Tent tortoises are morphologically highly variable with respect to the colour patterns of their carapace (Fig. 1.2) and plastron (Fig. 1.3), also the shape of the carapace and the smoothness of the carapace scutes, making their taxonomic diagnosis based on these structures very confusing (Branch, 1998 & 2008; Hofmeyr et al., 2014). The *P. t. verroxii* is the most challenging in this regard (Branch, 2008; Hofmeyr et al., 2014).

Current criteria used to distinguish between the three subspecies comprise only a few characters, which unfortunately do not have adequate discriminatory power to reliably distinguish among them. This can be attributed to either the existence of intermediate states (Branch, 2008) or insufficient diagnostic characters discovered so far.

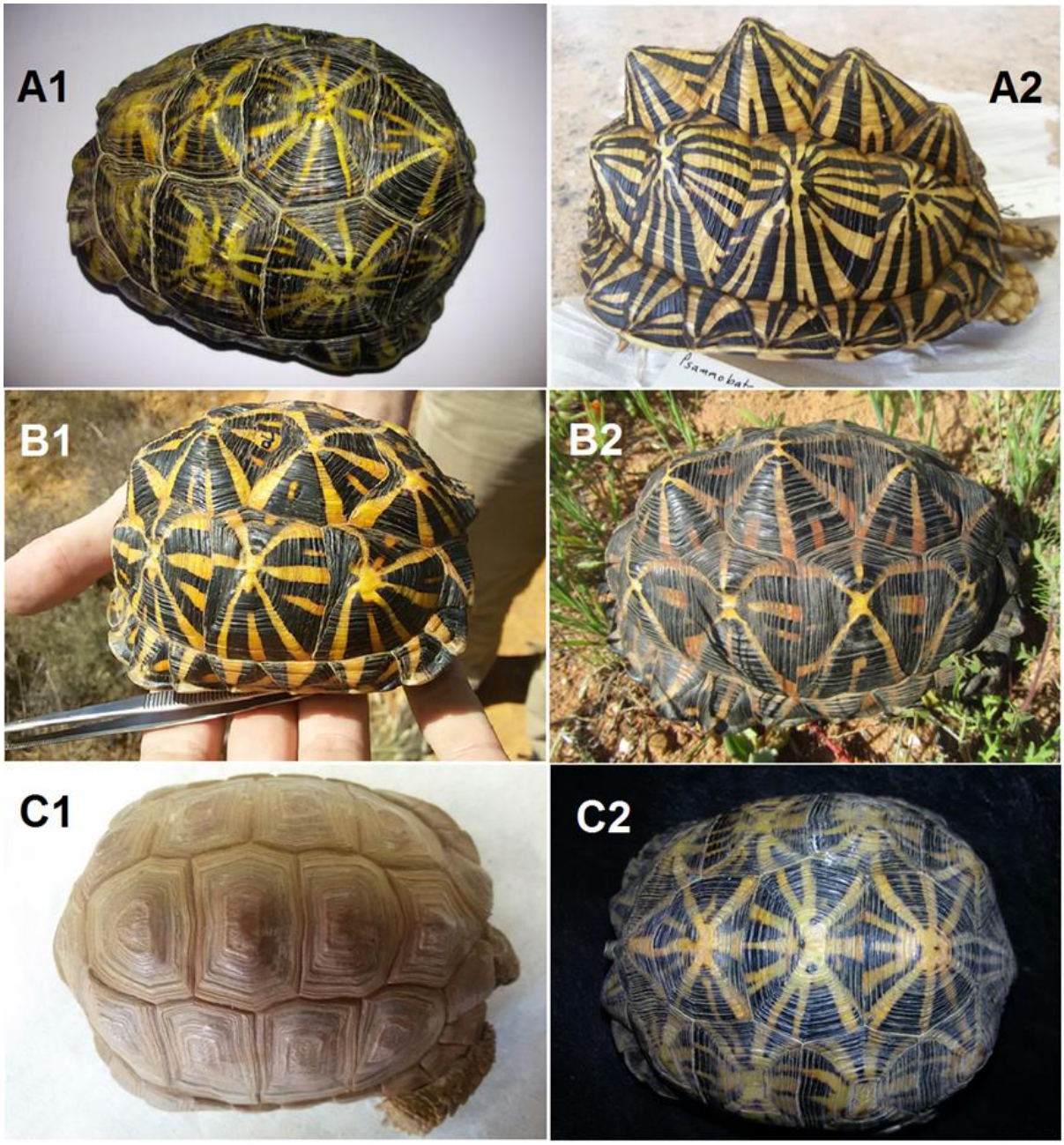


Figure 1.2. Some of the colour patterns and shape variations of the carapace of the tent tortoise complex. A1: *P. t. tentorius* from the central Karoo region, A2: *P. t. tentorius* from the Camdeboo area; B1: *P. t. trimeni* from the Hamtam Karoo region, B2: *P. t. trimeni* from the West Coast region; C1: *P. t. verroxii* from the upper Bushmanland region, C2: *P. t. verroxii* from the upper Great Karoo region.

Based on previous descriptions, *P. t. tentorius* has the largest body size (in terms of carapace length) of all three subspecies, whilst *P. t. trimeni* has the smallest body size (Branch, 1998 & 2008). With respect to the morphological patterns, *P. t. tentorius* has well developed carapace domes, with distinct yellow or orange stripes on each scute on a black background. Its plastron has a large solid central dark brown area. *Psammobates t. verroxii* has a carapace with flat or

slightly domed scutes, dull in colour with stripes on a brown background.

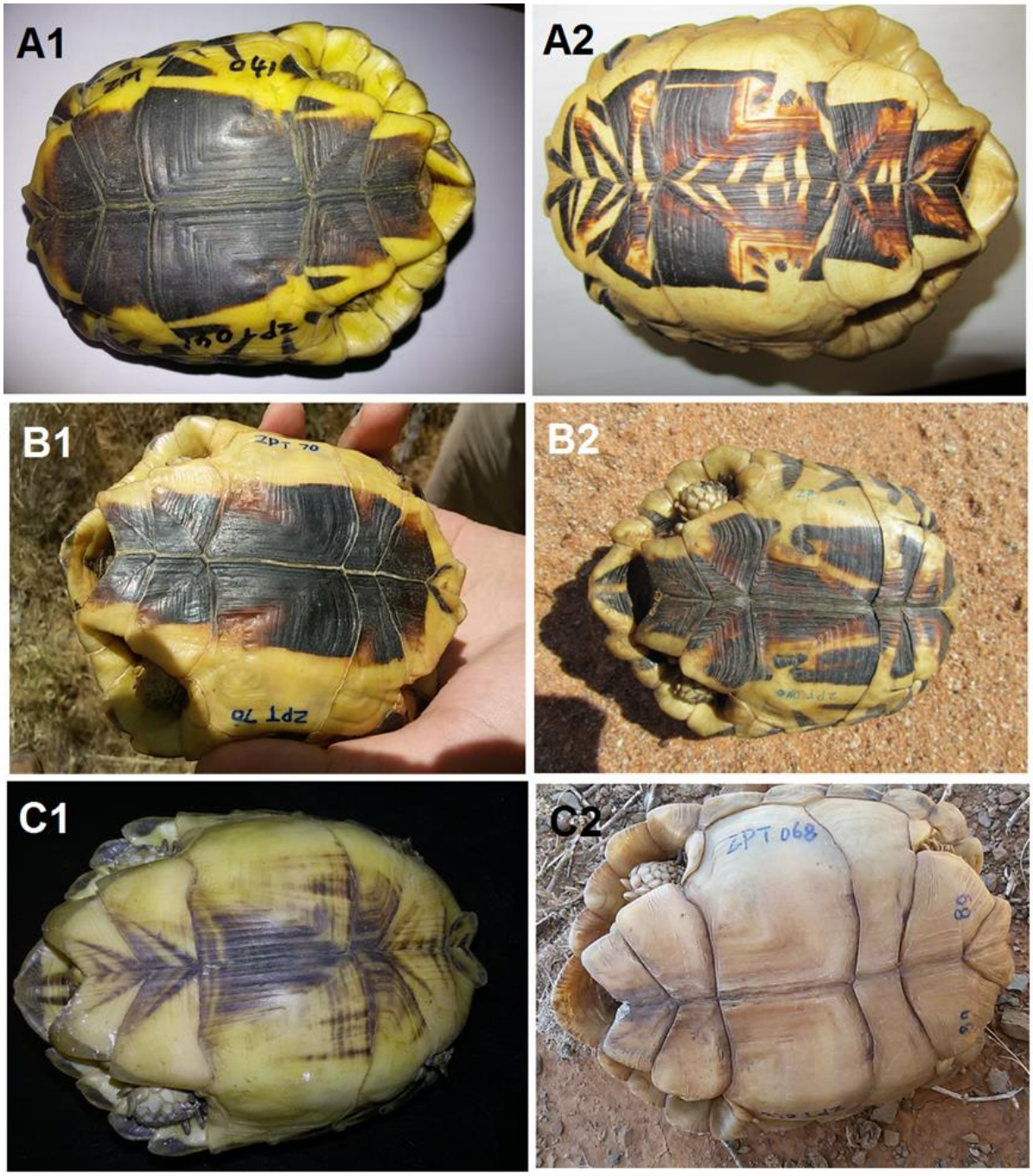


Figure 1.3. Some colour variations of the plastron among the different subspecies. A1: *P. t. tentorius* from the southern Karoo region, A2: *P. t. tentorius* from the Little Karoo region; B1: *P. t. trimeni* from the Hamtam Karoo region, B2: *P. t. trimeni* from the West Coast; C1: *P. t. verroxii* from the upper Great Karoo region, C2: *P. t. verroxii* from the central Karoo region.

Occasionally uniformly brown individuals without stripes may be seen in some localities. The plastron has dull diffuse markings or sparse stripes, and the central mark of the plastron is usually poorly bounded. The *Psammobates t. trimeni* has a fairly raised carapace dome on each

scute with clear orange to yellow stripes which become reddish from the center of each scute to its margin. The plastron has a large dark brown to black central area that may be interspersed with lighter patches or indented stripes (Branch, 1998 & 2008). Exceptions to the colour patterns described above were nevertheless encountered frequently in the field (Zhao, pers. obs.) which may reflect higher taxonomic diversity in the group or simply higher colour polymorphism, which require more comprehensive description.

1.4 Climate, biomes and vegetation

1.4.1 *Psammobates. t. verroxii*

The home range of *P. t. verroxii* falls into the Bushmanland region of the Nama Karoo (usually at an altitude above 900 m), as well as in southern Namibia. These areas receive much lower annual rainfall than the areas in which the other two subspecies are present. Rainfall is extremely unpredictable in these areas and predominantly in summer (Branch, 2008, Mucina & Rutherford, 2011, Hofmeyr et al., 2014). Summer can be extremely hot, especially in the inland areas (Mucina & Rutherford, 2011). A variety of shrub species with some dwarf succulent plant species occur in their habitat (Mucina & Rutherford, 2011, Hofmeyr et al., 2014).

1.4.2 *Psammobates t. trimeni*

This subspecies occurs mainly in Namaqualand along the West Coast where moderate levels of precipitation in winter is experienced. Fog is also not uncommon throughout the year. Summer is usually much cooler than the areas where the other two subspecies occur and this region has an astonishing diversity of succulent plant species, many of them endemic (Milton et al., 1997). The succulent biome is recognized as one of world famous “Biodiversity hotspots” (Mittermeier et al., 1999; Myers et al., 2000). The area is dominated by a huge variety of succulent plant species, some of which are considered as important and vital food resources for the tent tortoise (Branch, 1998; Cunningham & Simang, 2008).

1.4.3 *Psammobates t. tentorius*

The distribution range of *P. t. tentorius* receives rainfall throughout the year, and frequently has fog (Mucina & Rutherford, 2011, Hofmeyr et al., 2014). Temperatures in summer can be very high (South Africa Weather Service), similar in comparison to that experienced by *P. t. verroxii,* but humidity levels in both the soil and air are higher than those of Bushmanland (Mucina &

Rutherford, 2011).

1.5 Modern approaches to systematics analysis

The recent development of a variety of sophisticated R statistical packages and the constant improvement of the capabilities of High-Performance Computers, offer a refined platform for answering deeper molecular evolutionary questions. The methodologies for DNA-based systematics have improved significantly in recent decades to include Bayesian inference (BI) and a variety of advanced phylogenetic R packages (Paradis et al. 2004; Schliep 2011; Revell 2012; Matzke 2013; Zhang et al. 2013; Rabosky et al. 2014; Warnes et al. 2016). These advanced mega-phylogenetic approaches use high powered algorithms, which allow more options for investigating complex and deep evolutionary questions. The main goal of systematics research remains the unraveling of deep evolutionary questions, based on functional genes and principles of molecular ecology. Various species delimitation approaches became popular in the recent time (e.g. Miralles & Vences, 2013), which became widely used in delineating cryptic taxa in species complexes, particularly in cases where species boundaries are not clear.

This study used these modern phylogenetic and phylogeographic analytical procedures, and comprehensively representative DNA and amino acid sequences as well as multiple loci based microsatellite DNA datasets to try and unravel the long-standing systematic puzzle of the *P. tentorius* species complex.

The non-synonymous/synonymous substitution ratios were found to be useful in estimating selection pressure at certain nodes (Poon et al., 2009), and is regarded as useful indicators for evaluating the fitness of organisms in response to environmental changes. This method was therefore also utilized in this study.

Besides using these molecular techniques to infer the phylogeny, phylogeography and population genetics of the *P. tentorius* species complex, it was also important to assess the degree of congruence between these findings and traditional morphological assessments involving morphological character and morphometric measurement analyses. The latter analyses also involved phylogenetic reconstruction and multivariate analyses aimed at retrieving informative morphological characters and morphometric measurements, which will also be important for the taxonomic revision on the species complex and improving the

accuracy of its dichotomous key.

1.6 Study objectives

1. To develop a clear and deep understanding of the phylogenetic relationships among the different evolutionary lineages (clades) in the three currently recognized subspecies of the *P. tentorius* species complex.

2. To identify haplotype groups and to determine whether the current understanding of their taxonomic relationships is correct in terms of operational taxonomic units (OTUs).

3. To establish whether there is congruence between the currently recognized subspecies diagnosed morphologically and the haplotype groups identified using molecular markers.

4. To identify the ecological and geographical factors that possibly drove the evolutionary radiation of the group with special reference to factors such as historical climatic conditions, geographical barriers and habitat changes.

5. To locate potential biological hotspots of the tent tortoise, which should either have high conservation priority as crucial evolutionary centers (important radiation and speciation centroids), or as places where substantially higher population densities of a taxon are found.

6. To deeply understand the historical diversification patterns and predict future radiation trends, aimed at their conservation.

7. To test selection pressure on each group with non-synonymous/synonymous ratio-based selection analyses.

8. To use the molecular and morphological findings for developing the best taxonomic scenario for the *P. tentorius* species complex, and for developing a more accurate dichotomous key.

References

Alexander, G., & Marais, J. (2007). *A Guide to the Reptiles of Southern Africa*. Cape Town: Penguin Random House South Africa.

Baard, E. H. W. (1995). Growth, age at maturity and sexual dimorphism in the geometric tortoise, *Psammobates geometricus*. *The Journal of the Herpetological Association of Africa, 44*(1), 10-15.

Bour, R. (2008). Global diversity of turtles (Chelonii; Reptilia) in freshwater. *Hydrobiologia, 595*(1), 593-598.

Boycott, R. C., & Bourquin, O. (2000). *The Southern African Tortoise Book: A Guide to Southern African Tortoises, Terrapins and Turtles*. O. Bourquin.

Branch, W. R. (1998). *Field Guide to Snakes and Other Reptiles of Southern Africa*. Cape Town: Struik Publishers.

Branch, B. (2008). *Tortoises, Terrapins & Turtles of Africa*. Cape Town: Penguin Random House South Africa.

Buhlmann, K. A., Akre, T. S., Iverson, J. B., Karapatakis, D., Mittermeier, R. A., Georges, A., ... & Gibbons, J. W. (2009). A global analysis of tortoise and freshwater turtle distributions with identification of priority conservation areas. *Chelonian Conservation and Biology*, *8*(2), 116-149.

Cunningham, J. (2002). *A molecular perspective on the family Testudinidae Batsch, 1788* (Doctoral dissertation, University of Cape Town).

Cunningham, P. L., & Simang, A. (2008). Ecology of the Bushmanland tent tortoise (*Psammobates tentorius verroxii*) in southern Namibia. *Chelonian Conservation and Biology*, *7*(1), 119-124.

Griffin, M. (2003). *Annotated checklist and provisional conservation status of Namibian reptiles*. Namibia Scientific Society.

Hofmeyr, M. D., Boycott, R. C., & Baard, E. H. W. (2014). *Psammobates tentorius* (Bell, 1828). In: M. F. Bates, W. R. Branch, A. M. Bauer, M. Burger, J. Marais, G. J. Alexander & M. S. De Villiers (Eds.), *Atlas and Red List of the Reptiles of South Africa, Lesotho and Swaziland* (pp. 70-85). Pretoria: South African National Biodiversity Institute.

Hofmeyr, M. D., Leuteritz, T. & Baard, E. H. W. (2018a). *Psammobates tentorius*. *The IUCN RedList of Threatened Species 2018*: e.T170524A115656793.

Hofmeyr, M.D., Leuteritz, T. & Baard, E.H.W. (2018b). *Psammobates tentorius ssp. trimeni*.*The IUCN Red List of Threatened Species 2018*: e.T121936835A121936853

Hofmeyr, M. D., Vamberger, M., Branch, W., Schleicher, A., & Daniels S. R. (2017). Tortoise (Reptilia, Testudinidae) radiations in Southern Africa from the Eocene to the present. *Zoologica Scripta*, *46*, 389-400.

Hewitt J. (1933). On the Cape species and subspecies of the genus *Chersinella* Gray. Part I. *Annals of the Natal Museum*, *7*, 255-297.

Hewitt, J. (1934). On the Cape species and subspecies of the genus *Chersinella*. Part II. *Annals of the Natal Museum*, *7*, 303-352.

IUCN. (2018). *The IUCN Red List of Threatened Species*. Version 2018-1.

Keswick, T. (2012). *Ecology and morphology of the Kalahari tent tortoise, Psammobates oculifer, in a semi-arid environment* (Doctoral dissertation, University of the Western Cape).

Leuteritz, T. E. J., & Hofmeyr, M. D. (2007). The extended reproductive season of tent tortoises (*Psammobates tentorius tentorius*): a response to an arid and unpredictable environment. *Journal of Arid Environments*, *68*(4), 546-563.

Matzke, N. J. (2014). Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Systematic Biology*, *63*(6), 951-970.

Milton, S. J., Yeaton, R. I., Dean, W. R. J., & Vlok, J. H. J. (1997). Succulent Karoo. *Vegetation of southern Africa*, *649*.

Miralles, A., & Vences, M. (2013). New metrics for comparison of taxonomies reveal striking discrepancies among species delimitation methods in *Madascincus* lizards. *PLoS One*, *8*(7), e68242.

Mittermeier, R. A., Myers, N., Mittermeier, C. G., & Robles, G. (1999). *Hotspots: Earth's biologically richest and most endangered terrestrial ecoregions*. CEMEX, SA, Agrupación Sierra Madre, SC.

Mucina, L., & Rutherford, M. C. (2011). *The vegetation of South Africa, Lesotho and Swaziland*. South African National Biodiversity Institute.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, *403*(6772), 853.

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*(2), 289-290.

Poon, A. F., Frost, S. D., & Pond, S. L. K. (2009). Detecting signatures of selection from DNA sequences using Datamonkey. In *Bioinformatics for DNA Sequence Analysis* (pp. 163-183). Humana Press.

Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, *3*(2), 217-223.

Rabosky, D. L., Donnellan, S. C., Grundler, M., & Lovette, I. J. (2014). Analysis and visualization of complex macroevolutionary dynamics: an example from Australian scincid lizards. *Systematic biology*, *63*(4), 610-627.

Schliep, K. P. (2010). phangorn: phylogenetic analysis in R. *Bioinformatics*, *27*(4), 592-593.

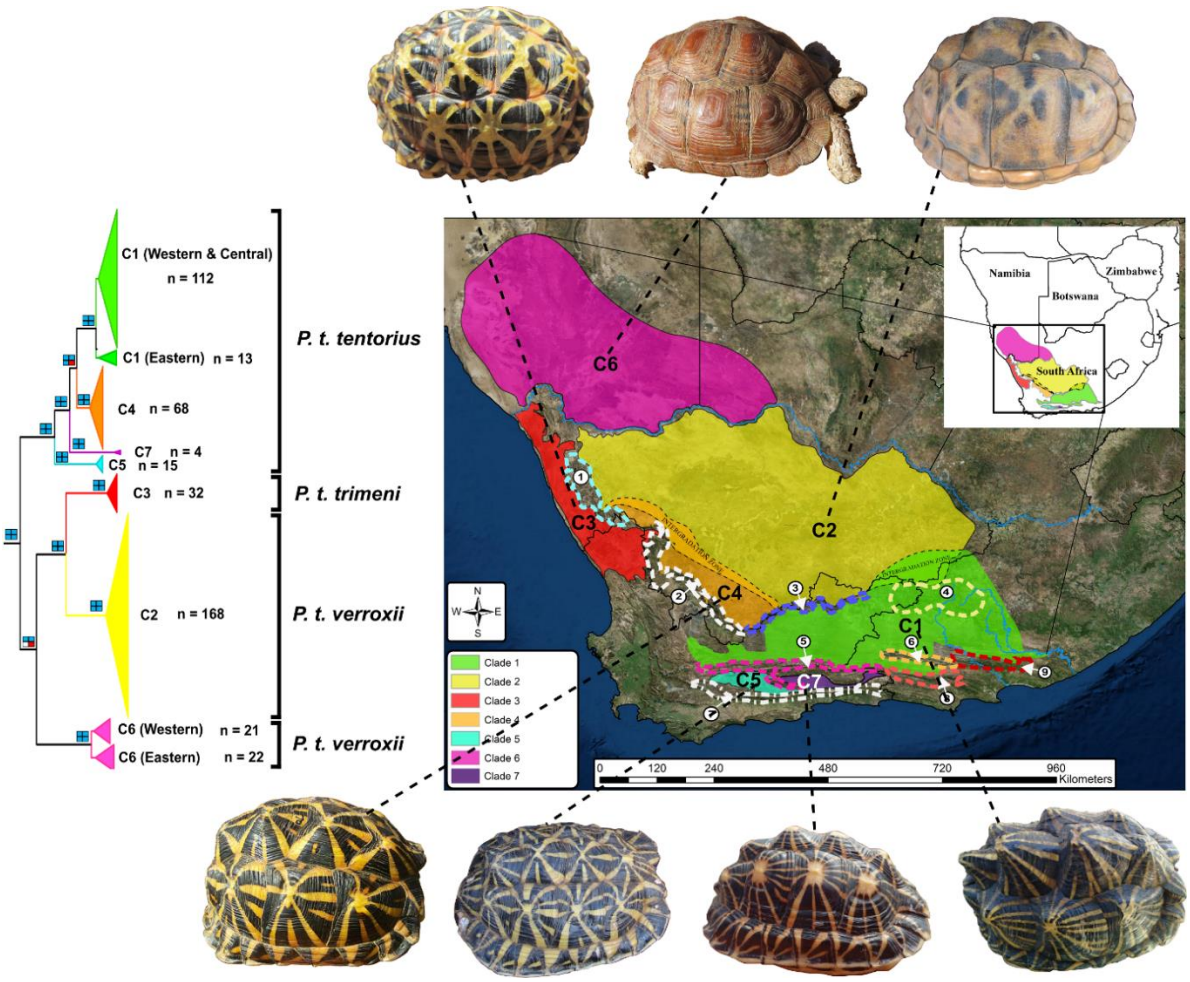
Swingland, I. R., & Klemens, M. W. (Eds.). (1989). *The conservation biology of tortoises* (No. 5). IUCN.

Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., ... & Schwartz, M. (2016). gplots: various R programming tools for plotting data. R package version 3.0. 1. 2016.

Zhang, J., Kapli, P., Pavlidis, P., & Stamatakis, A. (2013). A general species delimitation method with applications to phylogenetic placements. *Bioinformatics*, 29, 2869-2876.

Chapter 2. Unravelling the diversification and systematic puzzle of the highly polymorphic *Psammobates tentorius* species complex through phylogenetic analyses and species delimitation approaches.

Graphical abstract



Abstract

The high level of phenotypic diversity in southern African tent tortoises (*Psammobates tentorius* complex) has for decades prevented systematists from developing a stable taxonomy for the group. Here I used a comprehensive DNA sequence dataset (mtDNA: Cyt-b, ND4, ND4 adjacent tRNA-His and tRNA-Ser, 12S, 16S; and nDNA: PRLR gene) of 455 specimens, and the latest phylogenetic and species delimitation analytical procedures, to unravel the long-standing *P. tentorius* complex systematic puzzle. The results of the mtDNA and nDNA analyses were incongruent, with the poorly supported nDNA phylogeny differentiating the three recognised subspecies and showing potential hybridisation in some regions. In contrast, the concatenated mtDNA phylogeny identified seven operational taxonomic units, with strong support. Clades 1, 4, 5 and 7 corresponded to tortoises identified as *P. t. tentorius*, clade 3 to *P. t. trimeni*, and clades 2 and 6 to *P. t. verroxii*. The analyses showed conflicting topologies for the placement of C6 (*P. t. verroxii* north of the Orange River), with stronger support for it being sister to C2+C3 than to the other clades. Clades 1, 2 and 6 had significantly higher genetic diversity than clades 3, 4, 5 and 7, perhaps because these clades inhabit substantially larger areas. The potential for future cladogenic radiations seems high in C1 and C6, particularly in C6 for which the within-clade diversification level was highest. Further research involving microsatellite DNA, phylogeographic evaluations, and morphological variation among clades, is crucial for understanding the adaptive radiation of the *P. tentorius* species complex and for modifying their taxonomy.

2.1 Introduction

Southern Africa is recognized world-wide as a biodiversity hotspot for genera and species of the family Testudinidae (Cunningham, 2002; Fritz & Bininda-Emonds, 2007; Branch, 2008; Hofmeyr et al., 2014). It is home to 14 species of tortoises in six different genera, with strong indications that the current taxonomy, based on morphology, underestimates diversity (Hofmeyr et al., 2017). Using mitochondrial and nuclear DNA (mtDNA and nDNA), Kindler et al. (2012) distinguished eight sub-Saharan *Kinixys* species in contrast to six previously recognised by morphology. The importance of molecular studies to reveal chelonian diversity is exemplified by the fresh-water turtle *Pelomedusa*, formerly considered monotypic (*P. subrufa*), but recently shown to consist of 10 species with at least five more candidate species (Vargas-Ramírez et al., 2010; Vamberger et al., 2018).

Several species delimitation approaches based on Bayesian algorithms have become available for clarifying species boundaries (Yang & Rannala, 2010; Zhang et al., 2013; Reid, 2014; Jones et al., 2014). These are particularly valuable when species boundaries are diffuse (species complexes), especially if they help to reach agreement on the standardization of criteria for delineating operational taxonomic units (OTUs, Caron et al., 2009) and unifying species concepts, which have been controversial for decades (Scudder, 1974; De Queiroz, 2007; Kunz, 2013; Aldhebiani, 2018). Since OTUs are fundamental units of biodiversity (Hull, 1977; Cracraft, 1983; Balakrishnan, 2005; De Queiroz, 2007; Mysara et al., 2017) and play ecologically important roles as indicators of dynamic changes to their environment (Caron et al., 2009; Preheim et al., 2013), using molecular phylogenies to delineate various aspects of OTUs have become critically important to modern conservation management.

These modern technological and methodological advances have greatly strengthened our ability to answer long-standing phylogenetic and taxonomic questions and seem appropriate to investigate the taxonomic uncertainty of the southern African tent tortoises (*Psammobates tentorius* species complex). The high level of phenotypic diversity shown by this species has prevented systematists for decades from developing a stable taxonomy for the group. Tent tortoises from South Africa and southern Namibia are morphologically highly variable with respect to colour patterns, body shape, and the smoothness of their scutes. This led to the description of many species and subspecies in the past (Hewitt, 1933, 1934). In their taxonomic revision of African tortoises and turtles, Loveridge & Williams (1957) synonymized most *P.*

tentorius taxa to recognize only one species with three subspecies: *P. t. tentorius*, *P. t. trimeni* and *P. t. verroxii* (Supporting information Figure S2.1). This taxonomy was accepted by most subsequent authors (Greig & Burdett, 1976; Branch et al., 1995; Branch, 1998, 2008; Boycott & Bourquin, 2000; Hofmeyr et al., 2014) although some of these authors indicated that there may be unresolved taxonomic issues. A recent phylogenetic evaluation of southern African tortoises identified four lineages for *P. tentorius*, of which two occur within *P. t. verroxii* (Hofmeyr et al., 2017), the other two lineages match the other subspecies boundaries.

This study used molecular systematic techniques to unravel the puzzling diversity of the morphologically variable tent tortoise complex. The main objectives were to (1) verify whether the previous advocated "three subspecies" assumption was valid, (2) resolve the phylogenetic structure and understand the genetic diversity within and among different clades of the *P. tentorius* species complex, (3) address and clarify the OTUs of the complex, and (4) predict the potential for future cladogenesis and adaptive radiation to aid conservation management. To do this, both fast-evolving mtDNA and slow-evolving nDNA markers were used to infer the phylogenetic relationships within the *P. tentorius* species complex.

2.2 Materials and Methods

2.2.1 Sampling strategy

I obtained data of the range of the subspecies of *P. tentorius* from the literature (Hewitt, 1933, 1934; Greig, 1975; Greig & Burdett, 1976; Branch, 1998, 2008; Boycott & Bourquin, 2000; Hofmeyr et al., 2014), the online Virtual Museum records (ADU, available at http://vmus.adu.org.za/), iSpot distribution records (available at https://ispot.org.za/surveys), Bayworld Museum (Port Elizabeth), Iziko South African Museum (Cape Town) and Ditsong Museum of Natural History (Pretoria). I mapped all the records using Google Earth Pro version 7.3. Finally, I selected representative collecting sites to maximize the geographical area covered for each subspecies. A total of 455 specimens of *P. tentorius* were processed from 76 localities throughout its range in South Africa and Namibia, which comprehensively covered nearly the entire distribution range of *P. tentorius* according to the literature (Hewitt, 1933, 1934; Greig, 1975; Greig & Burdett, 1976; Branch, 1998, 2008; Boycott & Bourquin, 2000;

Hofmeyr et al., 2014). Details of the areas sampled are given in Supporting information Table S2.1 and Figure S2.2. Outgroup species selection (Supporting information Table S2.2) to provide structure to the tree topology was based on previous studies (Cunningham, 2002; Le et al., 2006; Hofmeyr et al., 2017).

I collected samples from both live and dead animals. For live animals, I collected fresh tissue from the tail tip (15-25 mg) or blood (0.1 ml) from the subclavian vein. Material from dead animals fell into three categories. I collected (1) connective tissue from inside shells that were in good condition; (2) bone samples from old museum or field shells; and (3) muscle tissue from preserved museum specimens from near the femur or tibia because these tissues had less exposure to preservation chemicals. Fresh tissue was preserved in 96% ethanol and blood was dissolved in 1% of a 10M Sodium-EDTA anticoagulant solution. Connective tissue and bone were kept dry in vials, whereas preserved muscle was stored in 96% ethanol. All samples were stored at -80 ºC in the laboratory until I did DNA extraction.

2.2.2 DNA Extraction, DNA marker selection, Amplification and Sequencing

I used QIAGEN DNeasy Blood and Tissue kits (QIAGEN, Germany) to extract DNA from all sample types. The manufacturer's protocol was followed for fresh tissue and connective tissue from shells, except that for connective tissue the elution buffer volume was reduced from 200 µl to 50 µl in the final elution stage to increase the final DNA concentration.

For heparinised blood samples, I removed the heparin by centrifuge the mixture (blood with heparin) at full speed in centrifuge machine for 10 minutes. I thereafter removed all supernatants and washed the precipitation with 5.0 MM PBS solution via vortex. Finally, I repeated the same centrifuge procedure as above for 10 minutes. The washed precipitation was processed with standard blood QIAGEN extraction protocol.

For preserved museum specimens, I used muscle tissue attached to the surface of the femur and tibia, since samples from these areas often provided better quality DNA. Bone and preserved muscle tissue were pre-treated before extracting DNA and were processed in a room where no *Psammobates* samples had been processed before. To prevent contamination, the work bench was exposed for 13 hours to UV light radiation and air flow between procedures. Bone samples were washed twice with double-distilled water and 96% ethanol to eliminate possible foreign DNA. Washed bones were then dried naturally and ground into powder with

a sterilized mortar and pestle. A total of 25 mg bone powder of each sample was transferred into a micro-centrifuge tube. The preserved muscle tissue was first rehydrated in micro-centrifuge tubes filled with double-distilled water and incubated at 40 ºC for 2-6 days, occasionally up to 2 weeks. This procedure accelerated lysis and digesting processes and helped to dissolve unknown chemicals. Rehydrated samples were then washed two times with 96% ethanol. If the solution still showed colour, I washed it three times with double distilled water using a vortex. I then removed the ethanol and washed the tissue three times with a 10M PBS solution, using the vortex, to eliminate as much formalin as possible as well as other chemical reagents used during preservation.

When extracting DNA from bone powder, incubation time was increased from 24 h to 48 h, whilst 5-10 days were allowed for the extraction from preserved museum samples. To increase the total DNA yield from bone and preserved muscle, I added 30 µl Proteinase K (100µg/ml) on day one and subsequently added 15 µl Proteinase K every day until all tissue was digested. After the digestion stage the standard QIAGEN extraction protocol was followed, except that the volumes of the rest of the reagents were doubled before the elution stage. Furthermore, to optimize the final DNA yield and increase the final DNA concentration and thus the possibility of obtaining longer fragments of genomic DNA, I added only 15-25 µl of QIAGEN elution buffer during the elution stage. The elution buffer was pre-heated to 70 ºC to increase the final yield in terms of DNA concentration. A similar approach was used with museum samples by Daniels et al. (2007) in a study of the phylogeny of the tortoise *Chersina angulata*. In order to check that no foreign DNA was introduced during DNA extraction, a blank control group was set up and treated in the same way as the bone and preserved muscle tissue samples, i.e., using the same reagents but no tissue samples. This blank control was also used for the PCR procedure.

Tortoises are generally slow evolving compared to most other reptiles (Avise et al., 1992); thus, fast evolving mtDNA markers rather than slow evolving nDNA genes should be used to accurately reconstruct their phylogenetic trees, particularly when studying small-scale species complexes (Lourenço et al., 2012). Studies showed that slow evolving nuclear markers are only suitable for studying the phylogeny of organisms with a wide time span range, whilst, faster evolving mtDNA markers become more accurate when focusing on recent events with relatively short time spans (Zheng et al., 2011; Lourenço et al., 2012). This study therefore used six widely used fast-evolving mtDNA loci to study the phylogeny of the *P. tentorius* complex: 12S rRNA (12S, Kocher et al., 1989), 16S rRNA (16S, Palumbi et al., 1991),

Cytochrome b (Cyt-b, Kocher et al., 1989; Pääbo, 1990; Whiting et al., 2003), NADH dehydrogenase subunit 4 (ND4, Stuart & Parham, 2004), and the two ND4 adjacent tRNA genes, tRNA-His and tRNA-Ser (Stuart & Parham, 2004). In addition to these, I used one of the fastest evolving nDNA genes, the Prolactin Receptor Coding gene (PRLR, Townsend et al, 2008).

I used the Polymerase Chain Reaction (PCR) to amplify 12S, 16S, Cyt-b, ND4, tRNA-His & tRNA-Ser, and PRLR genes. Primer details are given in Supporting information Table S2.3; note that I used two primer pairs for Cyt-b to accommodate fresh and ancient DNA. PCR reactions were performed using KAPA2G Robust HotStart, USA. Optimal annealing temperatures for 12S, 16S, Cyt-b, ND4 with tRNA-His & tRNA-Ser, and PRLR were determined through temperature gradient tests. PCR reactions for ancient DNA samples (here defined as bone and preserved muscle tissue) were performed independently from other samples, to minimize the chance of cross-contamination. To ensure that no foreign DNA was introduced during the PCR stage, a blank control group was again set up without adding any template DNA, instead, 2µl from the extraction control group was added. PCR reactions were performed in a BIO-RAD T 100^{TM} Thermal Cycler (Singapore) under the following parameters: an initial 4 min denaturation step at 94 ºC, followed by 37 cycles (43 for ancient DNA) of 30 s denaturation at 94 ºC, 30 s annealing (62 ºC for 12S; 50 ºC for 16S; 51 ºC for Cyt-b; 61 ºC for ND4 with tRNA-His & tRNA-Ser; 58 ºC for PRLR), and 1 min extension at 72 ºC, with a final 10 min extension step at 72 ºC. The PCR products were electrophoresed in 1% agarose gel, visualized under UV light, and purified using a BioFlux PCR Purification Kit (Bioer Technology, China). Purified PCR products were cycle sequenced using BigDye (ABI PRISM® BigDye Terminator v3.1 Cycle Sequencing Kits, USA) and standard methods (with annealing temperatures of 60 ºC for 12S; 50 ºC for 16S; 50 ºC for Cyt-b; 57 ºC for ND4 with tRNA-His & tRNA-Ser; 57 ºC for PRLR). The Big-Dye PCR products were purified by Zymo DNA Sequencing clean-up kit (Epigenetics Company, USA), prior to sequencing in an ABI 3500 genetic analyser.

2.2.3 Sequence Alignment, Modelling, Treatment of Indels and Partitioning

All Sanger sequences were analysed using ABI Prism Sequencing Analysis software v.3.7 (Applied Biosystems), aligned with MUSCLE v.3.2 (Edgar, 2004) and manually checked with MEGA v.7 (Kumar et al., 2016). I employed Partitionfinder v.1.0.1 (Lanfear et al., 2012) under

Python v.2.7 (Python Software Foundation, 2010) to determine the best partition scheme for the concatenated dataset. I also used jModeltest v.2 (Darriba et al., 2012) to determine the best substitution model and parameter settings for each data partition via AIC criterion (Supporting information Table S2.4). Nucleotide base biases across different partitions were determined through the homogeneity test implemented in PAUP v.4.0 (Swofford, 1998). Substitution saturation tests were performed using DAMBE v.6.1.9 (Xia, 2013) at both gene and partition (on protein coding gene) levels, and visually plotted with transition and transversion against GTR-model modified genetic distance diagrams. This was done to investigate potential substitution saturation, particularly at the third codon position in protein coding fast-evolving mtDNA. The DAMBE v.6.1.9 software was used to read sequence frames in order to determine the codon positions. All mtDNA sequences were concatenated through SequenceMatrix (Vaidya et al., 2011) thereafter.

2.2.4 Phylogenetic analyses

2.2.4.1 Phylogenetic inferences

The preliminary maximum likelihood (ML) analysis with RAxML v.8 (Stamatakis, 2014) using RDP4 (Martin, Murrell, Golden, Khoosal, & Muhire, 2015) with 1000 bootstrap replications (Felsenstein, 1985) (results not shown) showed some tree topology conflicts between the mtDNA and nDNA loci, but not among the six mtDNA loci. I therefore ran mtDNA and nDNA separately in all subsequent analyses. I employed four different phylogenetic inference approaches independently for the concatenated mtDNA dataset and the nDNA dataset: ML, maximum parsimony (MP), and two Bayesian inference analyses (BI).

Maximum likelihood analyses were performed using RAxML-VI-HPC (Stamatakis, 2014) implemented in CIPRES Science Gateway (Miller et al., 2010), using non-parametric bootstraps to determine the power of support for each node. The ML analyses results inferred with the mtDNA concatenated dataset and nDNA dataset were partitioned based on results of Partition-Finder, each partition with rate heterogeneity and substitution model GTR-CAI, and 1000 non-parametric bootstrap replications.

Maximum parsimony analyses were conducted using PAUP v.4.0 with default settings, and all gaps were treated as “new states”, since gaps are informative sources that can potentially contribute phylogenetic information. Both mtDNA and nDNA datasets were treated with 1000

bootstrap replications, using a full heuristic search consensus tree and only retaining groups with a frequency > 50%.

The traditional Bayesian inference (BI) analyses (Huelsenbeck & Ronquist, 2001; Huelsenbeck et al., 2001) were conducted with MrBayes 3.2 (Ronquist et al., 2012) on the mtDNA and nDNA datasets independently, using the evolutionary substitution model on each partition specified, based on results from Partition-Finder with the applicable prior parameter settings. Different substitution rates were allowed at different partitions. Tree topologies and tree lengths were linked across partitions, and other parameters treated were unlinked. Indel sites were treated as independent partitions. Analyses consisted of four independent runs (each with 4 chains) of 10 million generations with minimum sampling frequency every 5000 generations, discarding the first 25% of the samples as burn in. I used average standard deviation of split frequencies (ASDSF) to examine the mixing of chains and convergence of each run from the posterior probability distribution and ESS (Effective Sample Size) by Tracer v.1.6 (Rambaut et al., 2014) to determine whether sampling was sufficient. I ran analyses until the ASDSF value was close to 0.02 and the ESS > 200 for most of the parameters.

Multispecies coalescent (MSC) BI analysis using the species tree inference approach (Edward, 2009; Liu, Yu, Pearl, & Edwards, 2009), was performed with the StarBeast package (Heled & Drummond, 2009; Drummond et al., 2012) implemented in BEAST v.2.4.8 (Bouckaert et al., 2014) for the mtDNA and nDNA datasets separately. This approach can evaluate the degree of discordance between gene and species trees by sorting incomplete evolutionary lineages. This method requires inputs of all individuals from different pre-defined phylogenetic groups, which are evaluated and retrieved based on results of the traditional BI phylogenetic analysis with MrBayes. The multispecies coalescent analysis permits the simultaneous estimation of species tree topology, gene tree topology and population size, among others (Heled & Drummond, 2009). I employed the same substitution models and partition scheme previously used for the MrBayes analysis on the mtDNA dataset. Trees from different partitions were linked, and a Yule model selected. For MCMC setting, each analysis comprised four independent runs of 55 million generations, sampling every 5000 generations and discarding the first 10% of samples as burn in. Tracer v.1.6 was used to check if the effective sample size (ESS) reached the threshold value of 200.

In all cases, the consensus trees yielded by the different phylogenetic inference methods were visualized and drawn with FigTree 1.4.3 (Rambaut, 2012) or TreeGraph 2 (Stöver & Müller,

2010). I considered strongly supported nodes as those with bootstrap support (BP) > 70% for ML, and MP analyses (Hillis & Bull, 1993) and with posterior probability (PP) > 0.95 for Bayesian inference (Huelsenbeck & Rannala, 2004).

2.2.4.2 Tree topology and datasets incongruence tests

For the conflicting tree topology from ML result: (((C1+C4)+C7)+C5)))+C6 and C2+C3, and topology retrieved from all other results: ((C1+C4)+C7))+C5 and (C2+C3)+C6, I performed a Bayesian tree topology test under MrBayes, in order to test which assumption had greater likelihood. Different tree topology scenarios were set as constraints to test alternatives. Each hypothesis run had similar parameter settings with Bayesian analysis, allowing rate multipliers among gene partitions, each gene partition incorporating different substitution models and parameters. Each test run was split into two runs with MCMC of 10 million generations, sampling every 5000 generations with the stepping-stone method used to perform the test. The rest of the steps were identical to traditional MrBayes analysis. The marginal likelihood generated from each independent run of each hypothesis was used for comparison, assuming the one with the highest marginal likelihood mean as the best assumption. In terms of threshold, a likelihood difference in the range of 5 log units was considered as strong evidence in favour of the better model (Kass & Raftery, 1995).

To test whether the phylogenetic information retained by the mtDNA dataset was significantly incongruent with that of the nDNA dataset, I used the Incongruence Length Difference (ILD) test (Farris et al., 1994) implemented in PAUP v.4.0. Several studies were, however, skeptical about using the ILD test for determining incongruences between the two types of sequence datasets, pointing out its limited power and shortcomings (Yoder et al., 2001; Darlu and Lecointre, 2002).

2.2.5 Haplotype networks and molecular multivariant (Motif PCA) analysis

I employed TCS v.2.1 (Clement, Posada, & Crandall, 2000) to reconstruct haplotype groups and PopART (Leigh & Bryant, 2015) to draw parsimony networks for the two most informative mtDNA genes, Cyt-b and ND4, respectively. I performed a motif PCA analysis on the mtDNA dataset with DAMBE v.6 to determine the significant motif clusters for the *P. tentorius* complex. The motif PCA analysis is a principal component analysis that can be directly applied

to a sequence matrix. It projects the sequences into vectors and provides a summarized distance matrix. Finally, the PCA scatterplot, using the two principal components which contribute most to the variance, will be used to visualize the distances between each combination of samples (Konish et al., 2019).

2.2.6 Species delimitation analyses

In order to clarify the species boundaries and assign a hierarchical position to each retrieved clade, nine different species delimitation methods were used to determine whether each clade represented a valid putative species. To compare different approaches, summarize species delimitation analyses outcomes from multiple methods, and also to evaluate which species delimitation assumption receive the robustest support, I used the indicator – Relative Taxonomic Resolving Power Index (R_{tax}) advocated by Miralles and Vences (2013) to achieve such a goal. To accurately calculate R_{tax}, using multiple species delimitation approaches is necessary and crucial. In all methods, the congeneric members of genus *Psammobates*, *P. oculifer* and *P. geometricus* were included in the analyses. This purpose of this effort is to ensure each clade identified should be considered as a species.

2.2.6.1 ABGD approach

The ABGD (Automatic Barcode Gap Discovery) species delimitation approach (Puillandre, Lambert, Brouillet, & Achaz, 2012) has been shown to delineate species on the "barcoding gap" between distributions of interspecific pairwise genetic distances and intraspecific pairwise distances. This single locus, distance-based method effectively eliminates the bias of traditional pure p-distance based methods. I applied the ABGD approach on delimited putative species for mtDNA loci 12S, 16S, Cyt-b, ND4+tRNA-His+tRNA-Ser and nDNA PRLR, independently.

2.2.6.2 The rjMCMC BPP

I used the Bayesian reversible-jump MCMC (rjMCMC) algorithm (Rannala & Yang, 2003; Yang, 2015) with an MSC model-based species delimitation approach implemented in the program BPP (Yang, 2015) to delineate seven retrieved clades to putative species. This approach can calculate posterior probabilities in the MSC model on a given species tree by

considering parameters tau (divergence times) and θ (population size). The phylogenetic consensus tree retrieved from BEAST (but with all outgroups removed) was used as guide tree. Two prior parameter settings were used for both tau (shallow divergence: 2, 2000; deep divergence: 1, 10) and θ (small population size: 2, 2000; large population size: 1, 10) parameters, resulting in a total of four possible combinations that were tested by the A00 method implemented in BPP. All four tests were conducted with algorithm 1 and fine-tune parameters alpha = 2 and m = 1. Each test was run for 300000 generations with sampling frequency adjusted to 5% and 10% as burn-in. Each test was repeated three times with different initial seeds to ensure that all analyses generated similar results with different initial seeds. The guide tree used in all four analyses was based on a species tree generated from the BI analysis (mtDNA dataset using BEAST). The mtDNA dataset used in this analysis was partitioned by different gene loci beforehand, to allow rate heterogeneity across different loci.

2.2.6.3 STACEY MSC

I used a BEAST MSC approach with the STACEY package (Jones, 2015) on the mtDNA dataset partitioned by PartitionFinder results. Optimization of the substitution model and parameters was based on the PartitionFinder and jModeltest results. Samples were assigned to clades based on the phylogenetic consensus tree retrieved from BEAST. Analyses involved runs with 55 million generations, sampling every 5000 generations and discarding the first 25% as burn-in. Results were loaded into Tracer to check sample mixing with ESS and convergence diagnostics as in advanced phylogenetic analyses. Both gene and species trees generated from BI analyses were loaded into the TreeAnnotator package implemented in BEAST after the first 10% of trees was discarded as burn-in. The species trees were also loaded into the program DensiTree v.2.0 (Bouckaert, 2010) to visualize the topological overlapping of all trees and to visualize conflict branches and nodes after discarding 10% of the trees as burn-in, to determine the stability of each node and branch.

2.2.6.4 Marginal likelihood based MSC

A Marginal likelihood estimate (MLE) based MSC analysis conducted with the STACEY package implemented in BEAST was used for defining the species boundaries of the clades retrieved with the phylogenetic analysis (C1 to C7, see Results section). Seven different species

allocation scenarios (from one to seven species, see below) were set up by BEAUti under the STACEY template with the same substitution model and parameter settings as used above, employing the stepping-stone method for testing species boundaries based on the Bayesian factor (BF = 2lnBfs = 2 x (best scenario MLE - alternative scenario MLE)) to determine the best putative clustering scheme (Grummer et al. 2013). This could be 1) a one species assumption: C1+C2+C3+C4+C5+C6+C7; 2) a two species assumption: C1+C4+C5+C7 and C2+C3+C6; 3) a three species assumption: C1+C4+C5+C7 and C2+C3 and C6; 4) a four species assumption: C1+C4+C7 and C5 and C2+C3 and C6; 5) a five species assumption: C1+C4+C7 and C5 and C2 and C3 and C6; 6) a six species assumption: C1+C4 and C7 and C5 and C2 and C3 and C6; or 7) a seven species assumption C1 and C4 and C7 and C5 and C2 and C3 and C6. Each scenario was imported into the BEAST package Path sampler to calculate the MLE with 10 million generations, sampling divided into 9 different threads, with the first thread removed as burn-in. Bayes factors were generated with 2lnBf = 0–2 meaning "not worth more than a bare mention", 2lnBf = 2–6 meaning "positive" support, 2lnBf = 6–10 meaning "strong" support, and 2lnBf >10 meaning "decisive" support evidence in favour of the better model (Kass & Raftery, 1995; Grummer et al., 2013).

2.2.6.5 GMYC and bGMYC species delimitation

The GMYC and bGMYC species delimitation analysis was performed with the mtDNA dataset. Since both GMYC and bGMYC species delimitation methods require the input tree to be ultrametric, I removed all redundant sequences (i.e. identical sequences belonging to the same haplotype), thus ensuring that there were no identical sequences in the concatenated mtDNA supermatrix. All non-congeneric outgroups were truncated from the mtDNA supermatrix before the analyses were performed. I thereafter applied the same BI inference and parameter settings as mentioned in the BI MSC above, using BEAST.

The generalized mixed Yule-coalescent (GMYC) model (Pons et al., 2006; Fontaneto et al., 2007; Fujisawa & Barraclough, 2013) implemented in R (R Core Team, 2017) with single threshold GMYC model (Reid and Carstens, 2012) was used to perform GMYC species delimitation analysis by R package Split (Ezard et al., 2009) on the BEAST mtDNA tree file.

I also performed bGMYC Bayesian coalescent model-based analysis to delineate putative species from the BI MSC trees (generated from BEAST) with multiple thresholds. This approach allowed determining the putative species by producing a pairwise probability heat

map among tree leaves to visualize the probability that two nested tree leaves are members of the same putative species. This approach was applied by using the R package GMYC (Reid and Carstens, 2012). I first ran the algorithms on a single random tree from randomly selected 100 BEAST MSC trees, ran MCMC with 100000 generations, cut-off 50000 as burn-in, thinning = 10, leaves prior interval was set between 2 and 200 with start vector value = 25 to confirm sampling is sufficient and well mixed. Then I ran MCMC with 100000 generations, cut-off 50000 as burn-in, using a thinning interval of 100 from the 100 random trees. Besides plotting a probabilities heatmap for nested tips, a log posterior sampling plot, a coalescent ratio plot and a sampling frequency plot was plotted simultaneously to confirm sampling was sufficient and even.

2.2.6.6 PTP and mPTP

A PTP model-based (Zhang et al., 2013) Bayesian species delimitation was performed on bPTP web server (Zhang et al., 2013) with the mtDNA RaxML ML tree, with parameters settings: MCMC generations 500000, 100 of thinning and cut-off 10% as burn in. Finally, I also performed an mPTP model (Kapli et al., 2017), which is based on PTP methods but with multiple rates and faster ML algorithms.

2.2.6.7 TCS network punctuation

I employed TCS v.2.1 (Clement et al., 2000) to determine the number of significantly punctuated clades, with each isolated clade regarded as a valid putative species, since it represented the clade as being substantially isolated from the others.

2.2.6.8 Traditional p-distance method

I used MEGA v.7 to calculate the uncorrected *p*-distances for the Cyt-b gene among clades, within each clade, as well as the *p*-distances between outgroups. I used Cyt-b as the reference gene to calculate *p*-distances, as it has been widely used in testudine studies (Cunningham, 2002; Vargas-Ramírez et al., 2010; Fritz et al., 2012; Kindler et al., 2012; Vamberger et al., 2018).

2.2.7 Genetic diversity and population differentiation

In order to investigate the genetic structure and genetic diversity of each clade, the concatenated mtDNA supermatrix was transformed into a haplotype sequence matrix file, using DnaSP v.5 (Librado & Rozas, 2009), before being imported into Arlequin v. 3.5 (Excoffier & Lischer, 2010). Only 423 samples with full coverage for all mtDNA loci were included in this analysis. Analysis of molecular variance (AMOVA) (Excoffier, Smouse, & Quattro, 1992) was used to test phylogenetic structure (predefined by the seven mtDNA clades), genetic distance (mean pairwise nucleotide differences) and F_{st} variation with corresponding significance levels among, and within each of the groups and populations using 10000 permutations. The Fu's F_s test (Fu, 1997) and Tajima's D test (Tajima, 1989) were used to test whether selection on the different allele sites was even and neutral using a coalescent simulation in Arlequin. A population differentiation analysis method (in Arlequin) was used to determine whether individuals of each assigned population were randomly distributed. Molecular diversity indices θ (k), θ (π), theta (H) and theta (S) were estimated using Arlequin, and these θ values were used to evaluate the molecular diversity among different clades. The diversity indices were estimated separately for all members of each clade, in order to compare the genetic diversities across clades, using multiple indicators. The average pairwise distances and Nei's distances between and among clades were calculated using Arlequin.

2.3. Results

Despite extensive protocol modifications for DNA extraction from historical museum specimens, the success rate was generally poor. For 150 preserved and dry samples from the Port Elizabeth Museum, only 23 samples (15.3%) were successfully amplified for the short sequences of 12S and Cyt-b.

A 407 bp fragment of 12S, 539 bp of 16S, 465 bp (short sequence for recovering ancient DNA from museum preserved specimens) and 693 bp (long sequence) of Cyt-b, 678 bp of ND4, 120 bp of the tRNA-His and tRNA-Ser, and 525 bp of PRLR, were successfully obtained from the

PCR procedure. All the nucleotide sequences generated for this study have been deposited in NCBI GenBank see Supporting Information Table S2.1.

There was no significant substitution saturation has been detected, neither from DAMBE (Iss.c > Iss with $p < 0.0001$ in all cases, for both symmetrical and asymmetrical topology assumptions). The GTR model-based substitution saturation plots (for 1st and 3rd codon positions for Cyt-b and ND4 genes, genes 12S, 16S and tRNA-His & tRNA-Ser) did not reveal any signs of saturation; results not shown).

The IDL test results revealed significantly incongruent phylogenetic information between the nDNA and mtDNA sequence datasets.

2.3.1 Partitions and Substitution Models

Partitionfinder suggested the best partition scheme to be: Subset 1: 12S, Subset 2: 16S, Subset 3: Cyt-b 1st codon position, Subset 4: ND4 2nd codon position, Cyt-b 2nd codon position, Subset 5: Cyt-b 3rd codon position, Subset 6: ND4 1st codon position, Subset 7: ND4 3rd codon position and Subset 8: tRNA-His & tRNA-Ser. Homogeneity tests showed that the third codon positions of Cyt-b and ND4 genes were significantly different from other partitions in their nucleotide composition ($p < 0.0001$). None of the other partitions were significantly different ($p > 0.05$). The optimal substitution models with parameters for different partitions are specified in Supporting information Table S2.4.

2.3.2 Phylogeny of the *P. tentorius* species complex

For the mtDNA dataset (an alignment of 2437bp in total), the MP analysis results recovered a phylogenetic tree with total tree length of 3075 steps, CI (Consistency index) = 0.37, RI (Retention index) = 0.92, and HI (Homoplasy index) = 0.63. Maximum likelihood analysis generated a best tree with ML optimization likelihood = −13866.01. The phylogenetic analyses retrieved seven distinct clades, each with strong support. Apart from ML, the phylogenetic reconstructing approaches generated similar tree topologies ((C1+C4)+C7))+C5 and (C2+C3)+C6 across all major nodes and branches (Fig. 2.1). Each node generally showed high support values (BP > 70 and PP > 0.95) except for ML (Fig. 2.1), where the results revealed a weakly supported (BP < 70) topology (((C1+C4)+C7)+C5)))+C6 and C2+C3. Such topological

conflicts were also visible from BEAST MSC analysis (both gene trees and species trees) by using DensiTree (Supporting information Figure S2.3). Topology tests, however, suggested the most likely topology to be (C2+C3)+C6, since its marginal likelihood value was almost 100 log units higher than that of the alternative topology.

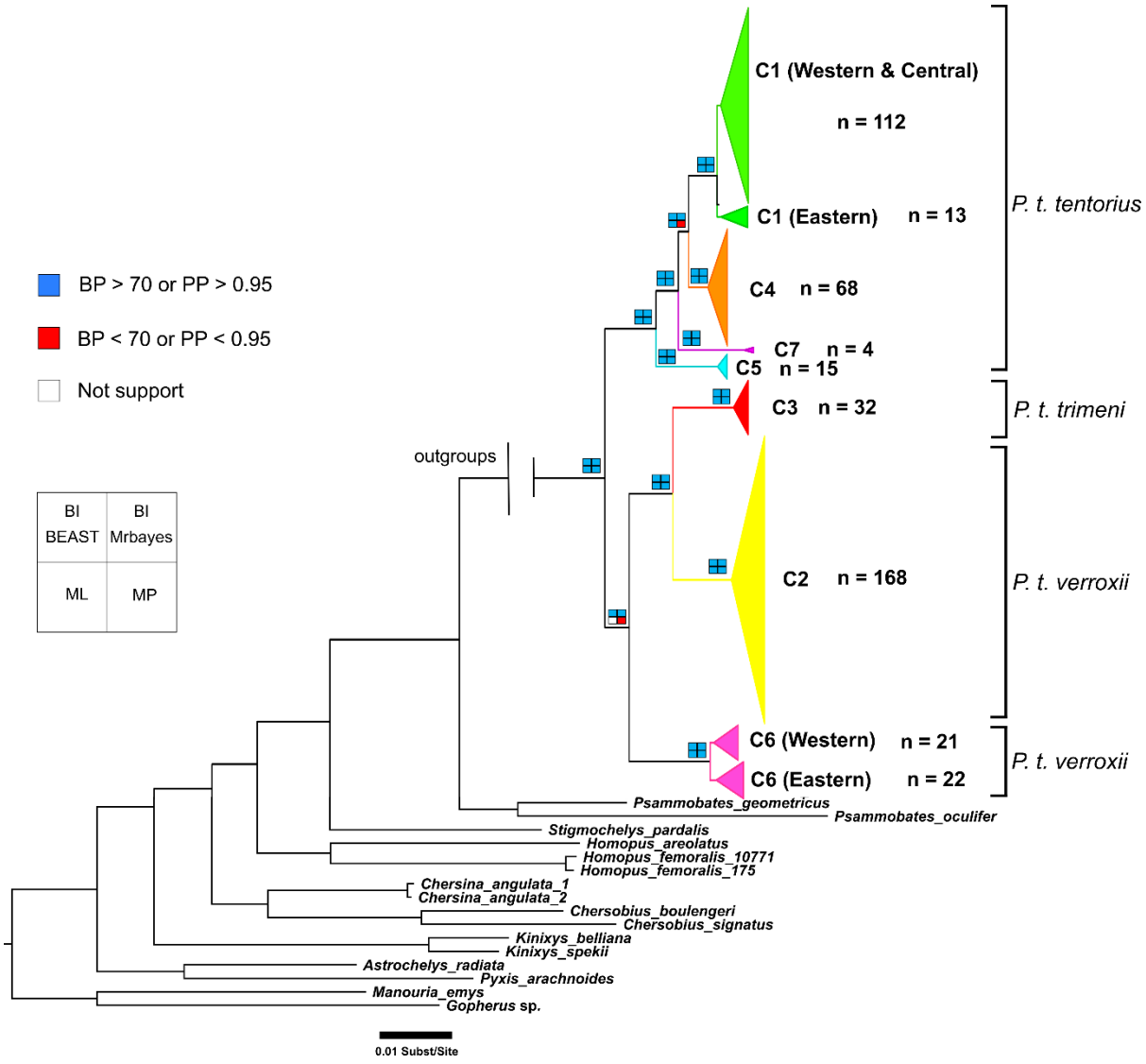


Figure 2.1. Phylogenetic relationships among seven clades of the *P. tentorius* complex inferred from BI (MrBayes) analysis of the combined mtDNA dataset (12S, 16S, Cyt-b, ND4, tRNA-His and tRNA-Ser), detail given in text. The strength of support values from four different approaches were visualized as three different colours as indicated in the figure. Top two are posterior probabilities (PP) for Bayesian inference (BI) from BEAST and MrBayes, whilst, the bootstrap values (BP) of Maximum likelihood (ML) and Maximum Parsimony (MP) are shown at the bottom. Colour scheme: C1 coloured in green, C2 in yellow, C3 in red, C4 in orange, C5 in aqua, C6 in pink and C7 in purple. The same colour scheme will be used throughout this paper.

Clade 1, C7 and C5 included specimens previously recognized as *P. t. tentorius* (Boycott & Bourquin, 2000), whereas C4 comprised individuals previously considered *P. t. trimeni* but recently identified as *P. t. tentorius* from the Kamiesberg, Hamtam Karoo and Roggeveldberge (KHR) region of the western Great Escarpment (GE; Rhodin et al., 2017). Clade 3 included samples of *P. t. trimeni* from the West Coast Succulent Karoo (WCSK) region (Boycott & Bourquin, 2000; Rhodin et al., 2017), and appeared as the sister group of C2, which comprised *P. t. verroxii* (Boycott & Bourquin, 2000; Rhodin et al., 2017) south of the Orange River. Clade 6 also comprised of *P. t. verroxii* populations, occurring north of the Orange River (NOR), being the sister group of C2+C3. These results indicate that *P. t. verroxii* is not monophyletic and thus not a systematically valid taxon.

The nDNA based phylogenetic analyses revealed some topological conflicts compared with the mtDNA results. The PRLR based phylogeny retrieved two clades, each with two subclades, with moderate to low support (Fig. 2.2). Clade 1 consisted of one subclade composed solely of individuals assigned to mtDNA C3 (*P. t. trimeni*), whereas the second subclade consisted of mtDNA C1, C4, C5 and C7, all forming part of *P. t. tentorius*. The second clade included individuals assigned mainly to mtDNA C2 and C6, corresponding to *P. t. verroxii* from south and north of the Orange River. The major subclade of clade 2 also included four samples corresponding to mtDNA C4 (*P. t. tentorius*): three of these were from Sutherland and one from Kliprand, both regions where the distributions of C4 and C2 abut and/or overlap. Two individuals (mtDNA C1) from Victoria West (a region where the distributions of C1 and C2 overlap) formed a distinct subclade of clade 2, possibly indicating intergradation between the two mtDNA clades. The nDNA phylogeny, in general, delineated the three recognised subspecies of *P. tentorius*.

Based on the mtDNA results, C1 occurs throughout the lower and central Karoo (from the Tankwa Karoo on the western side to the Fish River Valley to the east), and is mainly present below the GE, except where it overlaps with C2 at the corridor between the Nuweveldberge and Sneeuberge. Clade 2 occurs throughout the upper Karoo and Bushmanland in the Northern Cape Province and overlaps the range of C4 in the KHR region. Clade 3 is found in the WCSK region, from the Richtersveld in the north to the Knersvlakte in the south. Clade 4 occurs in the KHR region, clade 5 in the south-western Little Karoo, whereas clade 6 is only present NOR. Finally, C7 is found in the Oudtshoorn Basin, a region between Calitzdorp and Oudtshoorn, possibly extending into the Uniondale area (Fig. 2.3).

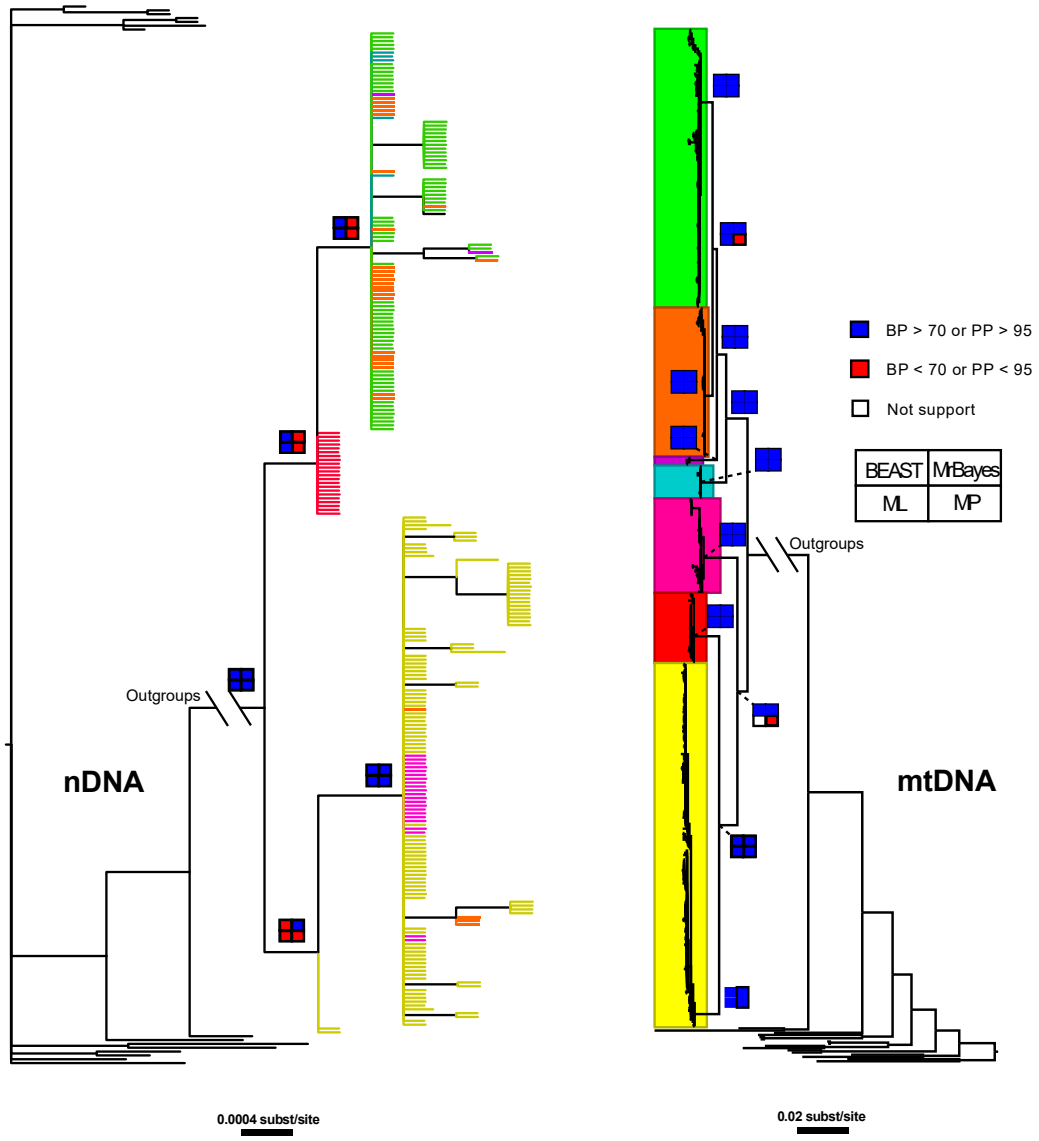


Figure 2.2. Phylogenetic relationships among seven clades of the *P. tentorius* complex inferred by BI (MrBayes) analyses from the nDNA dataset (PRLR) (left) and mtDNA dataset (12S, 16S, Cyt-b, ND4, tRNA-His and tRNA-Ser) (right). Detail given in text. The strength of support values from four different approaches were visualized as three different colours as indicated in the figure. Top two are posterior probabilities (PP) for Bayesian inference (BI) from BEAST and MrBayes, whilst, the bootstrap values (BP) of Maximum likelihood (ML) and Maximum Parsimony (MP) are shown at the bottom.

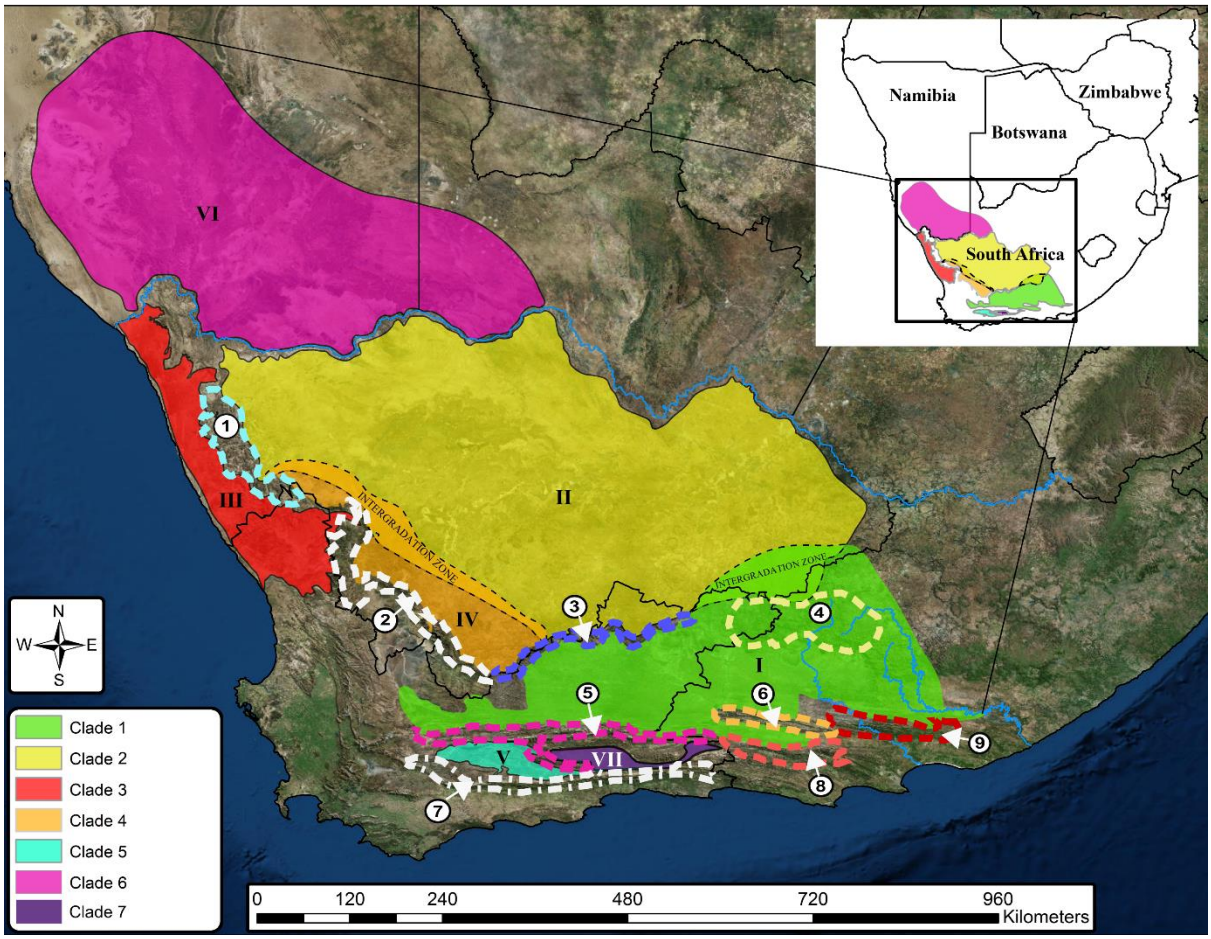


Figure 2.3. The map shows geographic distribution of the seven mtDNA clades throughout their range. The thin light blue line represents the Orange River system between South Africa and Namibia. The major mountain barriers are indicated: 1) Kamiesberg, 2) Hantamberge and Roggeveldberge, 3) Nuweveldberge, 4) Sneeuberge and Kompasberg, 5) Swartberge and Rooiberg, 6) Grootrivierberge, 7) Langeberge, 8) Baviaanskloofberge, 9) Suurberge. The sampling details on each clades are in the Supplementary Figure S2.2.

2.3.3 Networks and motif PCA

The TCS network retrieved six significantly punctuated haplotype groups for Cyt-b (Supporting Information Figure S2.4), and between C1 and C4 there were only ten mutation steps and a zero median vector, whilst, for ND4 (Supporting Information Figure S2.4), only five significantly punctuated haplotype groups were generated, and there were only four mutation steps between C1 and C4. The motif PCA generated a similar clustering pattern and TCS network. The first two principal components explained 74.35% of the total variation and were used to visualize the scatterplot for all seven clades (Supporting Information Figure S2.5).

Clade 1, C4, C7 and C5 were very close to each other, although none of the seven clusters overlapped.

2.3.4 Species delimitation

The ABGD approach in general supported a "seven putative taxa" assumption for Cyt-b and 16S, but not for the ND4, 16S and PRLR genes. The ABGD advocated "six putative species" for ND4 and 12S, but the ABGD of PRLR only retrieved "three putative species", which was congruent with the nDNA phylogenetic results (Table 2.1).

The BPP method generally strongly supported seven clades as valid OTUs for all four scenarios (Table 2.1, Supporting information Figure S2.3). All four possible scenarios: a) small population size (θ: 2, 2000) with shallow divergence (tau: 2, 2000); b) small population size (θ: 2, 2000) with deep divergence (tau: 1, 10); c) large population size (θ: 1, 10) with shallow divergence (tau: 2, 2000) and d) large population size (θ: 1, 10) with deep divergence (tau: 1, 10) all showed strong support (PP > 0.95) at all nodes across topologies. Therefore, it may imply that the depth of divergence and population size do not influence BPP delimitation in the *P. tentorius* complex.

The Bayesian STACEY MSC species tree showed that all nodes within the *P. tentorius* complex were strongly supported (PP > 0.95). Notwithstanding, the DensiTree revealed the uncertainty about the relationship between C6 and C2+C3, and the existence of conflicting placements of C6. Nevertheless, the BF based stepping-stone method strongly supported all seven clades as seven valid OTUs (since the "seven" species scenario generated significantly better MLE and BF scores (see Table 2.2).

The GMYC method strongly suggested seven OTUs (ML of null model = 172.46, ML of GMYC = 219.39, likelihood ratio = 93.86 and LR test: $p < 0.0001$). The bGMYC retrieved seven major nested tip groups as putative OTUs (Supporting Information Figure S2.6), since the pairwise posterior probability differences between all clades were significant ($p < 0.05$ in all cases). Besides, the MCMC outputs from both single-tree runs and multiple-tree runs were well mixed (Supporting information Figure S2.7 and Figure S2.8), implying that the MCMC sampling was sufficient in both cases. The "checkrates" R function revealed that the distribution of ratios of the Coalescence to Yule rates sampled in the analysis were all above zero, indicating that the model may be a good approximation of the data (Supplementary documents Figure S2.9).

Table 2.1. The summary of different species delimitation approaches with different data partitions. The number of putative species is indicated in each case with its putative species cluster components. For the BPP methods: 6a) small population size (θ: 2, 2000) with shallow divergence (tau: 2, 2000), 6b) small population size (θ: 2, 2000) with deep divergence (tau: 1, 10), 6c) large population size (θ: 1, 10) with shallow divergence (tau: 2, 2000), 6d) large population size (θ: 1, 10) with deep divergence (tau: 1, 10).

	1) ABGD (ND4+tRNA)	2) ABGD (PRLR)	3) ABGD (12S)	4) ABGD (16S)	5) ABGD (Cyt-b)	
Putative species no.	6	3	6	7	7	
C1	C1 = C4	C1=C4=C5=C7	C1 = C4	C1	C1	
C2	C2	C2 = C6	C2	C2	C2	
C3	C3	C3	C3	C3	C3	
C4	C1 = C4	C1=C4=C5=C7	C1 = C4	C4	C4	
C5	C5	C1=C4=C5=C7	C5	C5	C5	
C6	C6	C2 = C6	C6	C6	C6	
C7	C7	C1=C4=C5=C7	C7	C7	C7	

	6a) BPP mtDNA	6b) BPP mtDNA	6c) BPP mtDNA	6d) BPP mtDNA	7) STACEY-MSC (mtDNA)	8) STACEY-BF (mtDNA)
Putative species no.	7	7	7	7	7	7
C1	C1	C1	C1	C1	C1	C1
C2	C2	C2	C2	C2	C2	C2
C3	C3	C3	C3	C3	C3	C3
C4	C4	C4	C4	C4	C4	C4
C5	C5	C5	C5	C5	C5	C5
C6	C6	C6	C6	C6	C6	C6
C7	C7	C7	C7	C7	C7	C7

	9) GMYC (mtDNA)	10) bGMYC (mtDNA)	11a) bPTP (mtDNA)	11b) mPTP (mtDNA)	12) TCS (Cyt-b)	13) TCS (ND4)
Putative species no.	7	7	7	7	6	5
C1	C1	C1	C1	C1	C1 = C4	C1 = C4=C7
C2	C2	C2	C2	C2	C2	C2
C3	C3	C3	C3	C3	C3	C3
C4	C4	C4	C4	C4	C1 = C4	C1 = C4=C7
C5	C5	C5	C5	C5	C5	C5
C6	C6	C6	C6	C6	C6	C6
C7	C7	C7	C7	C7	C7	C1 = C4=C7

Both bPTP and mPTP methods retrieved seven significant clusters as OTUs.

The Cyt-b gene based pairwise distance analysis results (see Supporting information Table S2.5) revealed that the greatest among-clades *p*-distance occurred between C6 and C7 (*p*-distance = 0.083), whilst, the smallest among-clades *p*-distance was found between C1 and C4 (*p*-distance = 0.019). At within clade level, C2 and C6 were the clades with the highest within-clade *p*-distance (*p*-distance = 0.007), whilst, C7 had the lowest within-clade *p*-distance (*p*-distance = 0.000).

Table 2.2. The Bayesian Factor (BF) and Marginal Likelihood Estimation (MLE) based BEAST MSC species delimitation results of the seven putative species scenarios. The best scenario with highest MLE and BF is in bold.

Scenario	Putative species	Topology hypothesis	MLE	BF	Rank
S1	1	(C1+C2+C3+C4+C5+C6+C7)	-8515.35	174.82	5
S2	2	(C1+C4+C5+C7) + (C2+C3+C6)	-8561.29	266.7	6
S3	3	(C1+C4+C5+C7) + ((C2+C3)+C6))	-8455.72	55.56	2
S4	4	((C1+C4+C7)+C5)) + ((C2+C3)+C6))	-8499.69	143.5	4
S5	5	((C1+C4+C7)+C5)) + (((C2)+C3)+C6)))	-8596.15	336.42	7
S6	6	(((C1+C4)+C7)+C5))) + (((C2)+C3)+C6)))	-8492.96	130.04	3
S7	**7**	**((((C1)+C4)+C7)+C5) + ((((C2)+C3)+C6))**	**-8427.94**	n/a	**1**

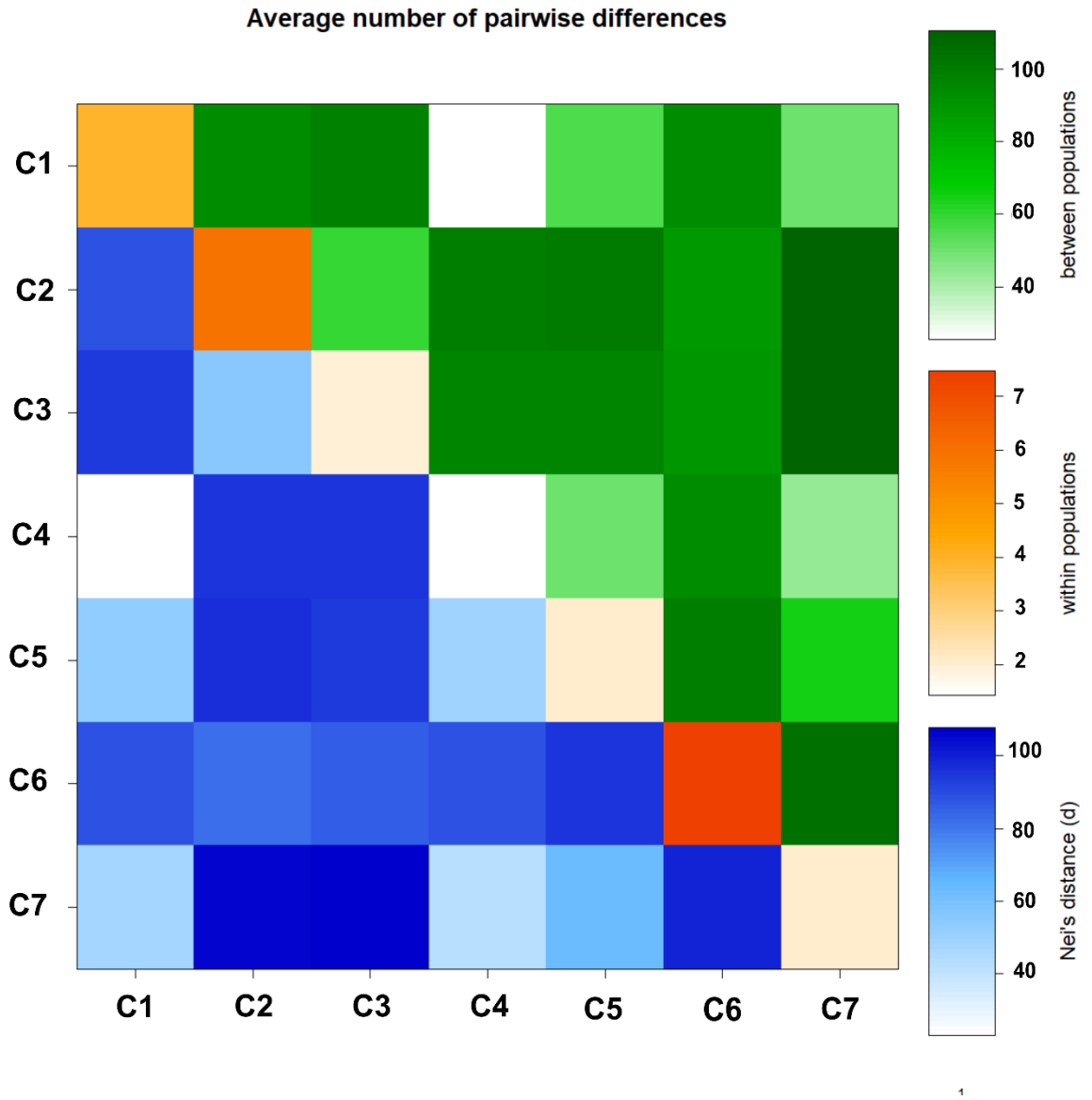


Figure 2.4. The Arlequin genetic diversity heatmap visualizing the average number of pairwise differences between clades and within clades, and also Nei's distance. The matrix used in the analysis was the combined mitogens. The scale unit was the number of pairwise differences.

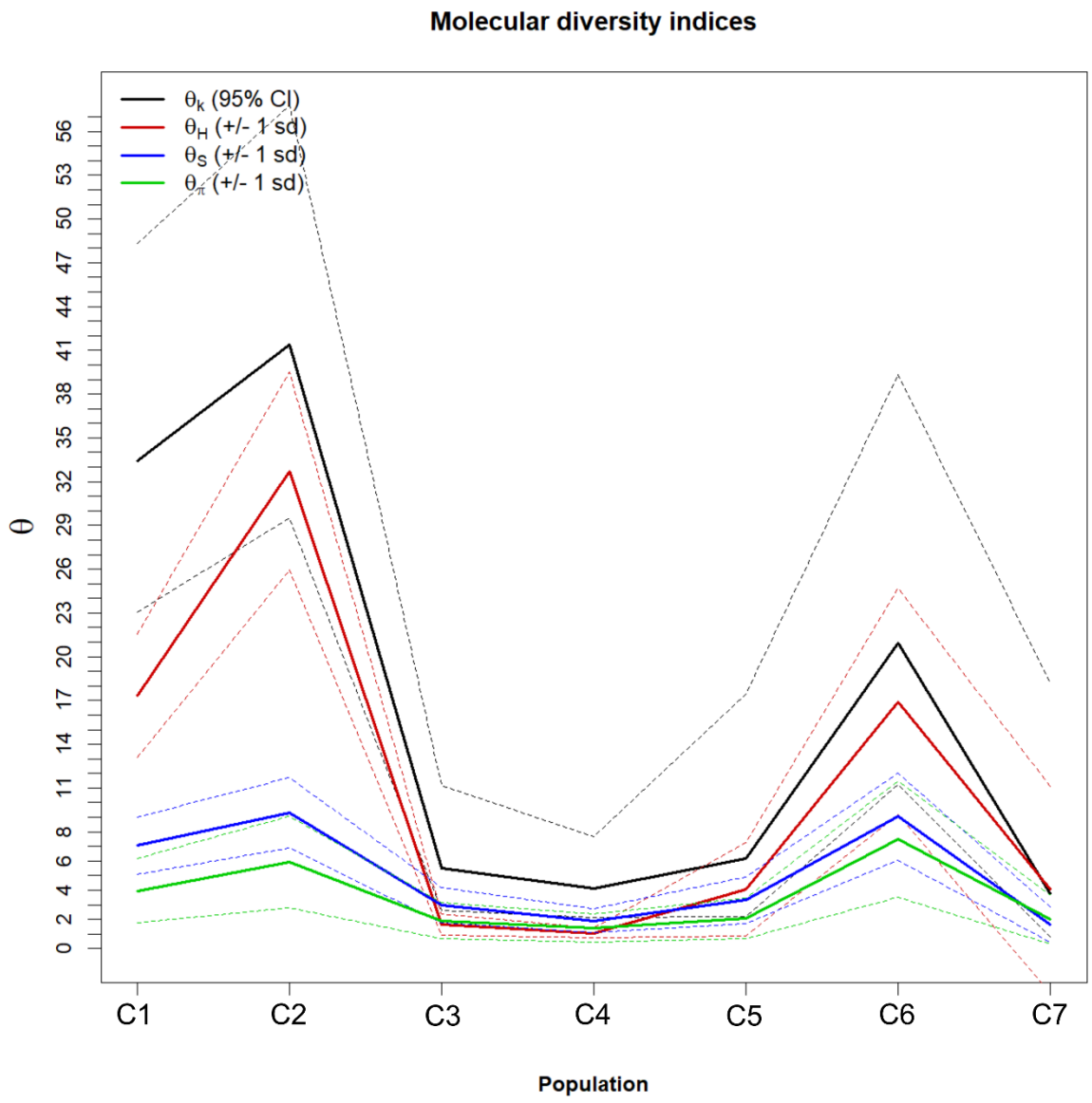


Figure 2.5. The diagram shows molecular indices θ (k), θ (π), theta (H) and theta (S) across all seven mtDNA clades. The indices represent the genetic diversity of each clade.

2.3.5 Genetic diversity and population differentiation

The AMOVA results showed that the demarcations among the seven mtDNA clades were all statistically significant (df = 422, Sum of squares = 13373.04, total variance = 41.37 $p < 0.0001$), 94.71% of the total variation being from among the clades (df = 6, Sum of squares = 12462.64, total variance = 39.18, $p < 0.0001$), and only 5.29% from within the clades (df = 416, Sum of squares = 910.4, total variance = 2.19, $p < 0.0001$). The Tajima's D test suggested

neutral selection on nucleotide substitution in all seven clades and the *P. tentorius* complex overall ($p > 0.05$ in all cases), whilst Fu's F_s test gave incongruent results for C1-C5, for which the neutral selection assumption was rejected ($p < 0.05$ in all cases). The results indicated that C1, C2 and C6 had higher haplotype diversity, molecular diversity (at both Θ π and Θ k) and genetic diversity than the other clades (Table 2.3, Fig. 2.4-2.5 & Supporting Information Table S2.6). The population differentiation test revealed that the seven clades were significantly different from each other (pairwise $p < 0.0001$ in all cases). The global differentiation test also showed significant differences among the seven clades (Exact $p < 0.0001$). The pairwise F_{st} and average pairwise differences (P_{xy}) among the seven clades denoted significant diversification among all pairwise clades (all pairwise p-values of Fst < 0.0001, all pairwise *p*-values of $P_{xy} < 0.0001$).

Table 2.3. The estimation of genetic diversity results showing polymorphic sites, number of haplotypes, Tajima's D and Fu's Fs test results (selective neutral test), and quantification of the genetic diversity at Θ π and Θ k of the seven clades, as well as for the entire *P. tentorius* complex. " * " represents $p < 0.05$, " ** " indicates $p < 0.01$, "***" indicates $p < 0.001$ and "*ns*" indicates not significant.

Clade	N	polymorphic sites	No. Haplotype	Tajima's D test	Fu's Fs test	Θ (π)	Θ (k)	Gene diversity
C1	123	39	52	*ns*	* * *	3.95	33.45	0.95
C2	151	52	64	*ns*	* * *	5.91	41.4	0.97
C3	32	13	11	*ns*	*	1.9	5.5	0.68
C4	63	9	12	*ns*	*	1.4	4.12	0.59
C5	12	10	7	*ns*	*	2.05	6.15	0.83
C6	38	38	22	*ns*	*ns*	7.48	20.95	0.95
C7	4	4	3	*ns*	*ns*	2	3.77	0.83
entire complex	423	288	171	*ns*	*ns*	63.38	106.25	0.98

2.4. Discussion

2.4.1 Phylogeny

Collectively, the phylogenetic analyses retrieved seven mtDNA clades, which implies that the currently advocated "three-subspecies" assumption is not valid, and that the seven clades may deserve to be considered as seven candidate species. The phylogenetic analyses of the *P. tentorius* complex showed some level of discordance between the mtDNA and nDNA results.

The mtDNA data retrieved seven clades, with strong support, with each clade occupying a distinct geographic region. The nDNA data, in contrast, were broadly consistent with the three recognized subspecies, and also indicated the possibility of hybridization in regions where mtDNA clades abut or overlap. Based on morphology, Loveridge & Williams (1957) and Boycott & Bourquin (1988) regarded the south-western Karoo as a three-way contact zone where all three subspecies are present and presumably hybridise. Both the mtDNA and nDNA results, however, showed that only one mtDNA clade, C1, of *P. t. tentorius* occurred there. Nevertheless, additional sampling in the contact zones of mtDNA clades is needed to ascertain which clades hybridise and how extensive hybridization is.

The mtDNA results indicated that *P. t. tentorius* consisted of four clades, each occurring in a different region, generally isolated from each other. Based on morphology, tortoises above the GE in the KHR regions were previously considered to be *P. t. trimeni* (Boycott & Bourquin, 1988; Branch, 2008) but the results confirmed the findings reported in Hofmeyr et al. (2014) and Rhodin et al. (2017) that these tortoises belong to *P. t. tentorius* (clade 4). The sampling showed that *P. t. trimeni* is only present west of the GE in the WCSK region of South Africa. Despite extensive sampling in southwestern Namibia, all individuals from there belonged to C6 (*P. t. verroxii*); consequently, there is reason to doubt that the range of *P. t. trimeni* extends into Namibia as has been reported (Greig & Burdett, 1976; Boycott & Bourquin, 2000; Branch, 2008). The paraphyly of *P. t. verroxii* highlights the need for a taxonomic revision of C2 and C6, a finding also addressed by Hofmeyr et al. (2017).

The fact that morphological characteristics are unreliable to differentiate taxa in a highly polymorphic species such as *P. tentorius* is illustrated by the contention of Hewitt (1933, 1934), Loveridge & Williams (1957) and Branch (2008) that the uniformly brown coloured "*P. bergeri*" morph may deserve recognition as a new taxon. In this study, this uniformly brown morph was found in the Tankwa Karoo (mtDNA assigned to C1), in the Upington and Helmeringhausen regions (mtDNA assigned to C6) and in northern Bushmanland, near Pofadder and Kenhardt (mtDNA assigned to C2). Therefore, "*P. bergeri*" is merely a colour morph present in three separate clades and should not be considered a separate taxon.

Inconsistent phylogenetic patterns of mtDNA and nDNA have been found across a range of organisms (Gonçalves, Martínez-Solano, Ferrand, & García-París, 2007; Shaw, 2002; Rato, Carranza, Perera, Carretero, & Harris, 2010), including turtles (Vargas-Ramirez, Carr, & Fritz, 2013). Such incongruences are usually caused by variation in the depth of coalescent processes

at different gene loci, incomplete lineage sorting resulting in mismatching between gene trees, rate heterogeneity and species trees (Rubinoff & Holland, 2005; Baum & Smith, 2013; Leliaert et al., 2014), as well as stochastic bias and errors from algorithms, various models and programs (Rubinoff and Holland, 2005). Mitochondrial and nuclear DNA carry different evolutionary information and their selection and inheritance patterns may be incongruent with each other, but both nevertheless represent fundamental elements of genomic evolution (Rubinoff & Holland, 2005).

When dealing with a slow evolving group, such as tortoises (Avise et al., 1992; Tollis et al., 2017), the rapid evolving mtDNA markers are possibly more suitable and reliable for elucidating a phylogeny (Rubinoff & Holland, 2005) than the slower evolving nDNA markers. Phylogenetic studies of chelonians (tortoises and turtles) often rely on mtDNA loci because nDNA markers frequently provide incongruent and poorly resolved trees with poor support (e.g., Kindler et al., 2012; Petzold et al., 2014). Caccone et al. (2004) found that nDNA divergence was nearly 30 times slower than that of mtDNA for Galápagos tortoises. Nonetheless, there are pitfalls in using mtDNA to elucidate phylogenetic history. For example, when gene trees bifurcate before speciation, the number of taxa can be overestimated, and in the case of incomplete lineage sorting, gene trees may incorrectly reflect species trees (Shaw, 2002; Rubinoff & Holland, 2005; Baum & Smith, 2013; Leliaert et al., 2014). More research at genomic level is therefore necessary to clarify the phylogeny of the *P. tentorius* complex.

The diversification patterns not only demarcated seven cluster groups within the *P. tentorius* complex, it also showed subclades within C1 and C6 (Fig. 2.1, Fig. 2.2 and Fig. 2.3). Clade 1 tends to show a clear subdivision between western and central populations, and population from the Fish River valley in the east. It is also noteworthy to mention that the eastern population usually has highly domed carapacial scutes compared to those from the central and western region. In terms of C6, populations from the west and east formed separate branches, but, morphological differences between the two are not clear (Zhao et al., unpublished data). These subdivisions may imply that the relatively high level of within group diversifications of C1 and C6 may lead to further cladogenic radiation of the *P. tentorius complex*.

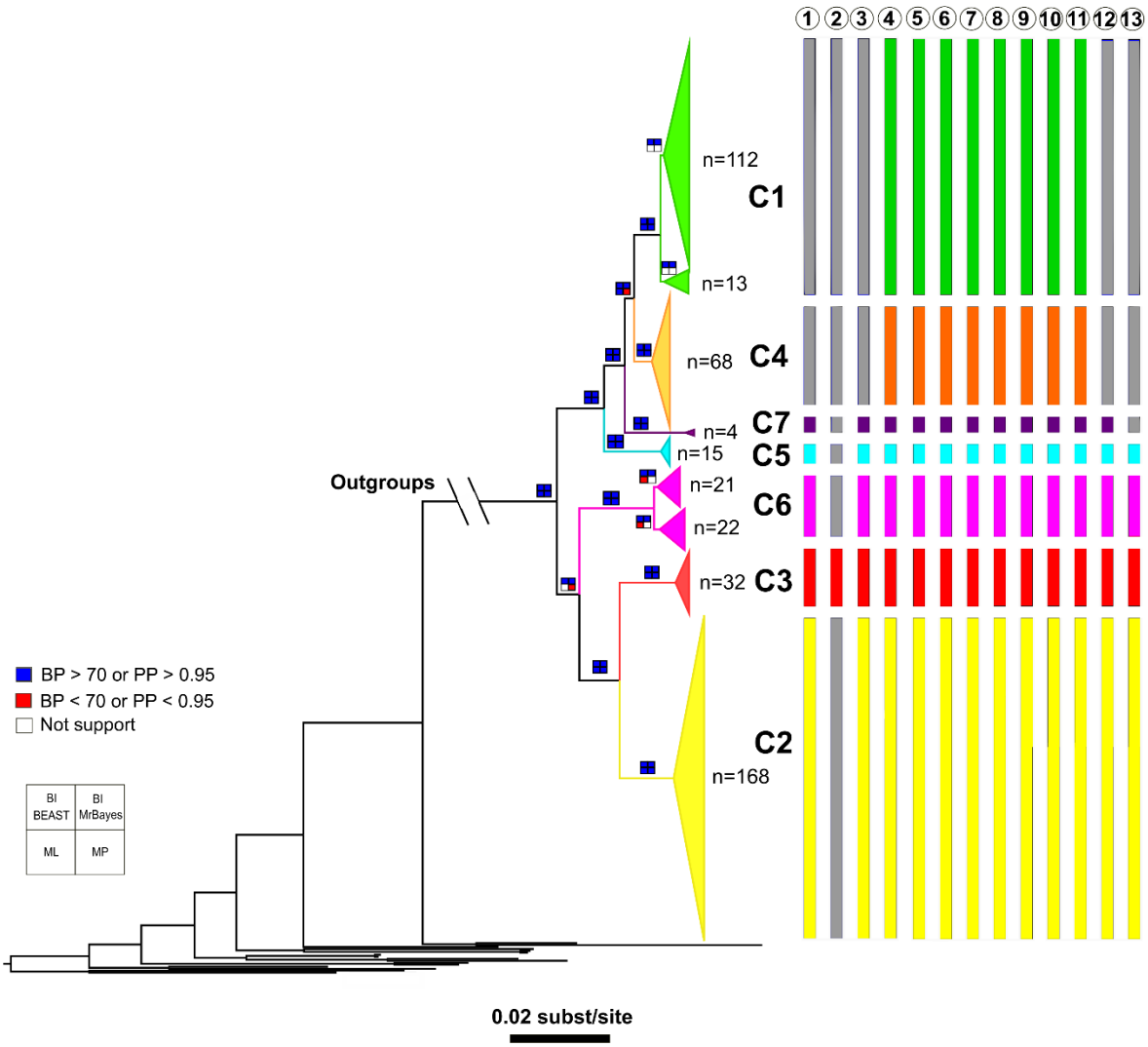


Figure 2.6. The summary graph shows results from multiple species delimitation approaches and the phylogenetic relationships among seven clades of the *P. tentorius* complex inferred from BI (MrBayes) analysis of the combined mtDNA dataset (12S, 16S, Cyt-b, ND4, tRNA-His and tRNA-Ser), detail given in text. The strength of support values from four different approaches were visualized as three different colours as indicated in the figure. Top two are posterior probabilities (PP) for Bayesian inference (BI) from BEAST and MrBayes, whilst, the bootstrap values (BP) of Maximum likelihood (ML) and Maximum Parsimony (MP) are shown at the bottom. The bar plot 1-13 at right side represents multiple species delimitation results, the detail given in Table 1. The grey colour represents clades which were not supported by the corresponding species delimitation approaches.

2.4.2 Species delimitation, Determination of OTUs

The delimitation analyses suggested five to seven putative species (Fig. 2.6 & Table 2.1) for the seven mtDNA clades of the *P. tentorius* complex. The conservative nDNA (PRLR) results distinguished the three subspecies of *P. tentorius*, corresponding to *P. t. tentorius*

(C1+C4+C5+C7), *P. t. trimeni* (C3) and *P. t. verroxii* (C2+C6). It is known that population size and divergence level can impact species delimitation negatively, and that the impact from incomplete lineage sorting (usually ascribed to bias from markers) can also be substantial (Sukumaran & Knowles, 2017; Luo et al., 2018). Six mtDNA markers and one nDNA marker was used, and the power of the markers was possibly not strong enough to perform MSC based coalescent delimitation. Sample sizes for C5 and C7 were also comparatively low. Notwithstanding this, a recent simulation-based study found that increasing the number of loci and sample size caused only limited improvement of multiple species delimitation approaches (Luo et al., 2018), and that even single-locus based delimitations can be useful (Sukumaran & Knowles, 2017).

The MSC model-based species delimitation approach seems useful for delineating putative species, as it outperforms the simple distance threshold-based DNA barcoding methods (Yang & Rannala, 2017). However, a simulation-based study found that the MSC species delimitation approach identifies genetic structure rather than species, resulting in an inflated estimation of the number of species (Sukumaran & Knowles, 2017). A follow-up study based on both simulations and real datasets supported the finding that BPP may detect population splits rather than species divergences, but still advocates the method's usefulness under certain conditions (Leaché, Zhu, Rannala & Yang, 2018). In general, the genomic data-based species delimitation approaches should be considered as hypotheses, which require additional information (e.g., ecological and phenotypic) before delineating species (Sukumaran & Knowles, 2017; Leaché et al., 2018). Moreover, the other Bayesian based methods, and the GMYC approach may also over-split the number of species (Miralles & Vences, 2013).

To set up species delimitation boundaries so as to assign taxa to proper hierarchical positions remain challenging (de Queiroz, 2007; Fujita, Leaché, Burbrink, McGuire, & Moritz, 2012), as the delimitation boundary between species is still controversial in view of different species concepts having completely different criteria (de Queiroz, 1998, 1999, 2005a, b), and none of them can comprehensively explain the delimitation thresholds. Furthermore, none of the existing species concepts can define species boundaries to fully cover species from every aspect of hierarchical systematics (de Queiroz, 1998, 1999, 2005b, 2007). This is especially true when there is no unifying ultimate definition for a species (Mallet, 2013). Even modern likelihood and Bayesian algorithms, based on sophisticated delimitation methods, can only generate results suggesting a particular taxonomy, but are unable to determine absolute species

boundaries (Kass & Raftery, 1995; Leaché & Fujita, 2010; Fujita et al., 2012; Jones, 2015, 2017).

One of the traditional approaches to assigning OTU's to hierarchical taxonomic positions and evaluate species boundaries is to consider uncorrected pairwise genetic distances (*p*-distance), particularly of Cyt-b. The literature reflects wide variation in Cyt-b *p*-distances among congeneric species, ranging from 1.5 - 4.92 % for *Trachemys* taxa (Fritz et al., 2012), 8.6 - 18.3 % for *Pelomedusa* species (Vargas-Ramírez et al., 2010), and 3.76-11.59 % for *Kinixys* species (Kindler et al., 2012). The uncorrected *p*-distances of Cyt-b among five *Testudo* species were 6.9 % to 12.7 % (Fritz & Bininda-Emonds, 2007), and among the three *Indotestudo* species, 3.7 % to 5.9 % (Iverson, Spinks, Shaffer, McCord & Das, 2001). With regard to *p*-distances within species level, the *p*-distances within *Chersina angulata* and *Chersobius signatus* were reported as 2.8 % (Daniels et al., 2007) and 2.5 % (Daniels, Hofmeyr, Henen & Baard, 2010), respectively; within *Aldabrachelys gigantea*, 0.9 % (Austin, Arnold & Bour, 2003); within *Testudo graeca*, 3.4 % (Fritz et al., 2007); within *Testudo horsfieldii*, 1.1 % (Fritz et al., 2009); and within *Chelonoidis chilensis*, 1.21 % (Fritz et al., 2012).

For the *P. tentorius* complex, the Cyt-b average *p*-distances among clades ranged from 1.9 - 8.3 % with only one comparison (C1 vs C4 = 1.9%) being below 3.8 %. When comparing the *p*-distance values within the *P. tentorius* complex to other tortoise species, it is clear that only the *p*-distances between C1 and C4 fall into the range as observed within, but all other pairwise comparison between clades fall into the range as observed between. Based on this index, it is possible that at least six mtDNA clades of *P. tentorius* may qualify for species specific status, but more research is required to investigate the level of hybridization between clades. The traditional absolute *p*-distance based methods are arbitrary and biased, since *p*-distance may vary greatly across different organism groups for the same gene, also, the substitutions may not be evenly distributed in each gene (Nagy et al., 2012). Nonetheless, Malone & Fontenot (2008) found a positive correlation between uncorrected *p*-distance and the degree of postzygotic isolation in toads. This finding may imply that degree of genetic divergence (in terms of *p*-distance) is still a useful indicator reflecting the patterns of reproductive isolation and may still be a useful criterion in species delimitation.

In general, the majority of my species delimitation analysis results supported the “seven putative species” assumption. Collectively, 71 % of the species delimitation approaches used here advocated a “seven putative species” assumption, 18 % suggested “six putative species”,

6 % supported “five putative species” and 6 % supported the “three subspecies” assumption. According to the Relative taxonomic resolving power index (R_{tax}) advocated by Miralles & Vences (2013) (for details see Supporting Information Table S2.7), I found that the 6 - 7 putative species assumption seemed well supported, since their R_{tax} indices were much higher than that of other assumptions. Therefore, it seems likely that some of the mtDNA clades may deserve full species status. Such a decision, however, should obviously be approached with caution. Results of a fine-scale genetic structure study using multiple microsatellite DNA markers (Zhao et al., in prep.) and a morphological marker-based study (Zhao et al., in prep.) are crucial before considering taxonomic revision in this species complex.

2.4.3 Genetic diversity and conservation genetics

The AMOVA results and differentiation analysis also advocated seven distinct clusters in line with the seven clades retrieved from the phylogenetic analyses and species delimitation analyses, since all clades were genetically highly structured. The C1, C2 and C6 showed higher levels of haplotype diversity and genetic diversity than C3, C4, C5 and C7. Although such horizontal comparisons might be biased and arbitrary, since the sample sizes were not even across all clades, C4 showed the lowest genetic diversity level of all the clades (Table 2.3, Fig. 2.4 & 2.5), even though it had the third largest sample size of all. This might imply that the diversification rate of C4 was slow, and that its low genetic diversity may be related to population expansions due to a leading edge effect, a homogeneous environment that promotes gene flow, or even selective sweeps. In contrast, the genetic diversity of C6 is remarkably high (Table 2.3, Fig. 2.4 & 2.5), even though the sample size of C6 only ranked 4th among the clades. This finding corresponds with the phylogenetic results, where a clear subdivision was discovered within C6 (Fig. 2.1, 2.2 and Supporting Information Figure S2.6). It is noteworthy that C6 had the highest within-group average pairwise differences of all clades (Fig. 2.4). This may reasonably lead to the deduction that the diversification rate in C6 is higher than that of the other clades, and further cladogenic radiation might be expected in C6. Furthermore, it is probably related to environmental gradients that limit homogenizing and thus affect gene flow, or could be a refugial area that has persisted over time, resulting in a build-up of diversity. These assumptions obviously need further research for verification. The pairwise F_{st} and P_{xy} showed that C2, C3 and C6 are closer to each other than to the other clades, that is, to C1, C4, C5 and C7 (Fig. 2.4). This finding highlighted the large genetic difference between the two

major branches. The Fu's Fs test results imply that selection was not neutral in C1- C5, therefore further investigation into how selection pressure drives adaptive radiation in the *P. tentorius* species complex would be an interesting topic for future molecular ecological studies.

Since *P. t. tentorius* and *P. t. verroxi* have been evaluated as "Near Threatened", and *P. t. trimeni* (correspond to C3) downgraded to "Endangered" (IUCN, 2018), they all deserve adequate conservation attention. Conservation efforts should focus mainly on conserving areas with cladogenic radiation potential, rather than only on preserving endemic taxa (Allendorf & Luikart, 2009). In this regard, the KHR region should be given adequate conservation attention since the area includes the intergradation zone between C2 and C4. So far, only the Akkerendam Nature Reserve has been established in this region. Furthermore, preserving this intergradation zone will also benefit the conservation of other sympatrically occurring reptiles. The KHR region overlaps with the distribution ranges of the Speckled Dwarf tortoise (*Chersobius signatus*) which is listed as "Endangered" (IUCN, 2018), and the Armadillo girdled lizard (*Ouroborus cataphractus*) categorised as "Vulnerable" (IUCN, 2018). Similar attention is warranted for the biological corridor between the Nuweveldberge and Sneeuberge, where C1 and C2 coexist. It is noteworthy that the sympatrically occurring Karoo Dwarf tortoise (*Chersobius boulengeri*) was also evaluated as "Endangered" (IUCN, 2018). Unfortunately, this region currently has no protected areas. In the NOR region C6 deserves conservation attention, given that it has the highest genetic potential for further radiation. Fortunately, the region already has protected areas like Ai-Ais Transfrontier Park and Tirasberg Conservancy.

2.4.4 Future research

In this study, I used six mtDNA markers and one nDNA marker, and multiple modern species delimitation procedures to clarify the phylogenetic relationships within the *P. tentorius* complex, resulting in the identification of seven mtDNA clades and seven putative species. In order to map the full evolutionary picture of the complex, a further analysis based on multiple microsatellite DNA markers is needed to determine the fine-scale genetic structure among populations and to quantify gene flow among them. This analysis is also crucial for investigating whether the putative species patterns derived from mtDNA, nDNA and microsatellite DNA markers are congruent. Furthermore, the findings of a phylogeographic analysis will also be crucial for understanding the diversification and radiation of the *P.*

tentorius complex under temporal and spatial dimensions. In this regard, the approach will allow tracing back the cladogenic history and aligning historical geographic events with the diversification of the *P. tentorius* complex from an ecological perspective.

In summary, I identified seven operational taxonomic units within the complex, rather than the currently advocated "three-subspecies assumption", though the conservative nDNA marker generated incongruent results. I found that *P. t. verroxii* was not a monophyletic group and should therefore not be considered as taxonomically valid. The genetic diversity analyses revealed that C1, C2 and C6 had significantly higher genetic diversity than C3, C4, C5 and C7. Further cladogenic radiation are expected in C1 and C6, but particularly in C6, since the within-clade genetic diversity level found in the latter clade was the highest. The establishment of protected areas at intergradation zones between C1 and C2, C2 and C4, and in the distribution range of C6, may facilitate conserving the adaptive radiation potential of the *P. tentorius* complex. Future studies at morphological level, at population level using microsatellite DNA markers and at phylogeographic level, will be crucial for understanding the evolutionary radiation of the *P. tentorius* complex, as well as for making appropriate decisions in its taxonomic review.

Supplementary Figures

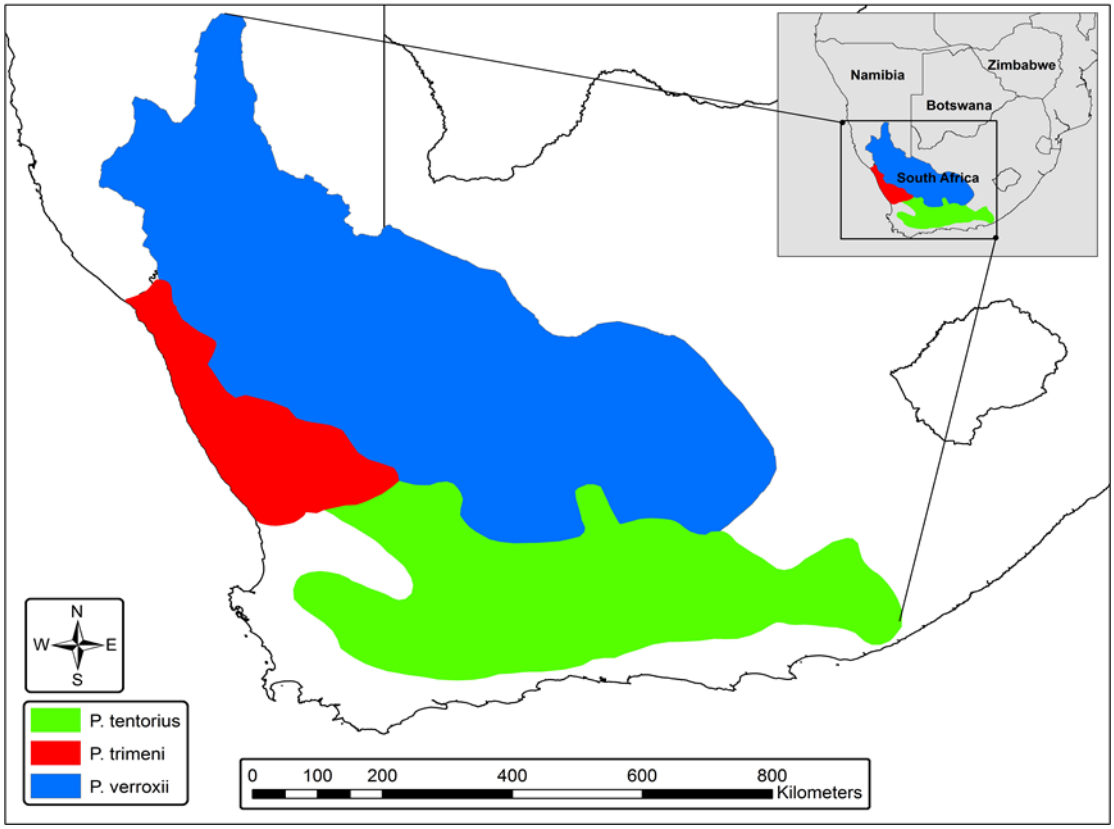


Figure S2.1. Map showing the interpreted distribution ranges of the three currently recognized subspecies of *Psammobates tentorius* was drawn based on data from various sources: *P. t. tentorius* (green area), *P. t. trimeni* (red area) and *P. t. verroxii* (blue area). The data are a combination of own records, Cape Town Iziko South African Museum, Port Elizabeth Bayworld Museum records, Pretoria Ditsong Museum of Natural History records, as well as records provided by the Animal Demographic Unit (ADU), University of Cape Town.

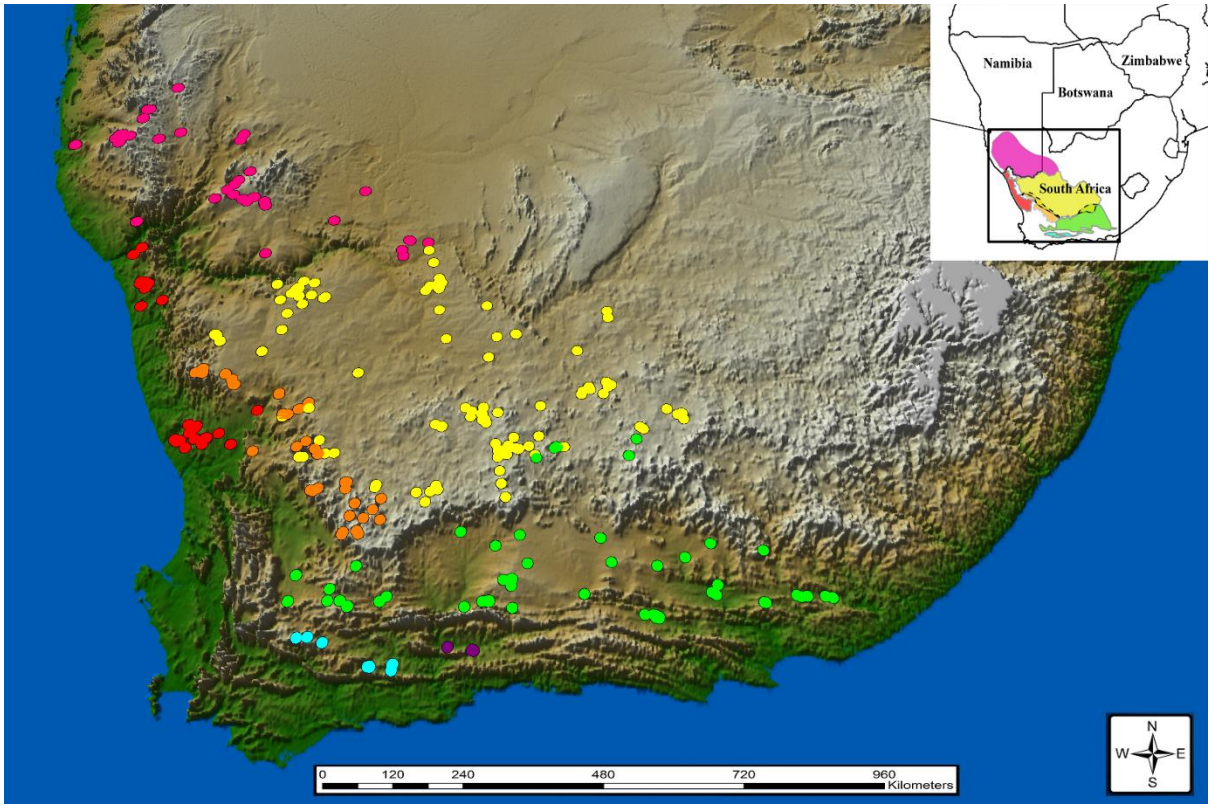


Figure S2.2. The map shows the distribution range of the seven clades retrieved from the phylogenetic analyses, in relation to geographic topology. The samples used in the study are indicated in colours which correspond to their respective clades.

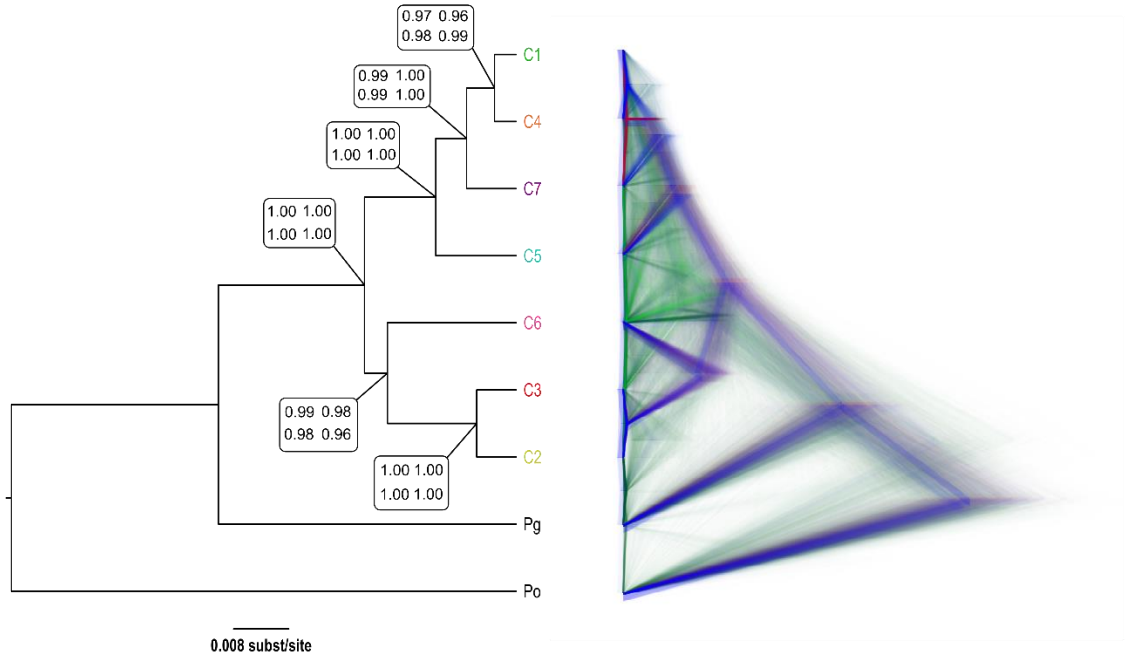


Figure S2.3. Left part: BEAST (STACEY) Bayesian multispecies coalescent species tree (PP > 0.95 at all nodes, not shown), with the four values on the left side of each node indicating the posterior probabilities generated from the BPP species delimitation analysis. 1) Top-left: small population size (theta) with shallow divergence (tau), 2) Top-right: small population size (theta) with deep divergence (tau), 3) Bottom-left: large population size (theta) with shallow divergence (tau), 4) Bottom-right: large population size (theta) with deep divergence (tau). Right: the BEAST (STACEY) Bayesian multispecies coalescent multiple species trees generated from 50 million generations with cut-off of 10% as burn-in, are shown to display node conflicts.

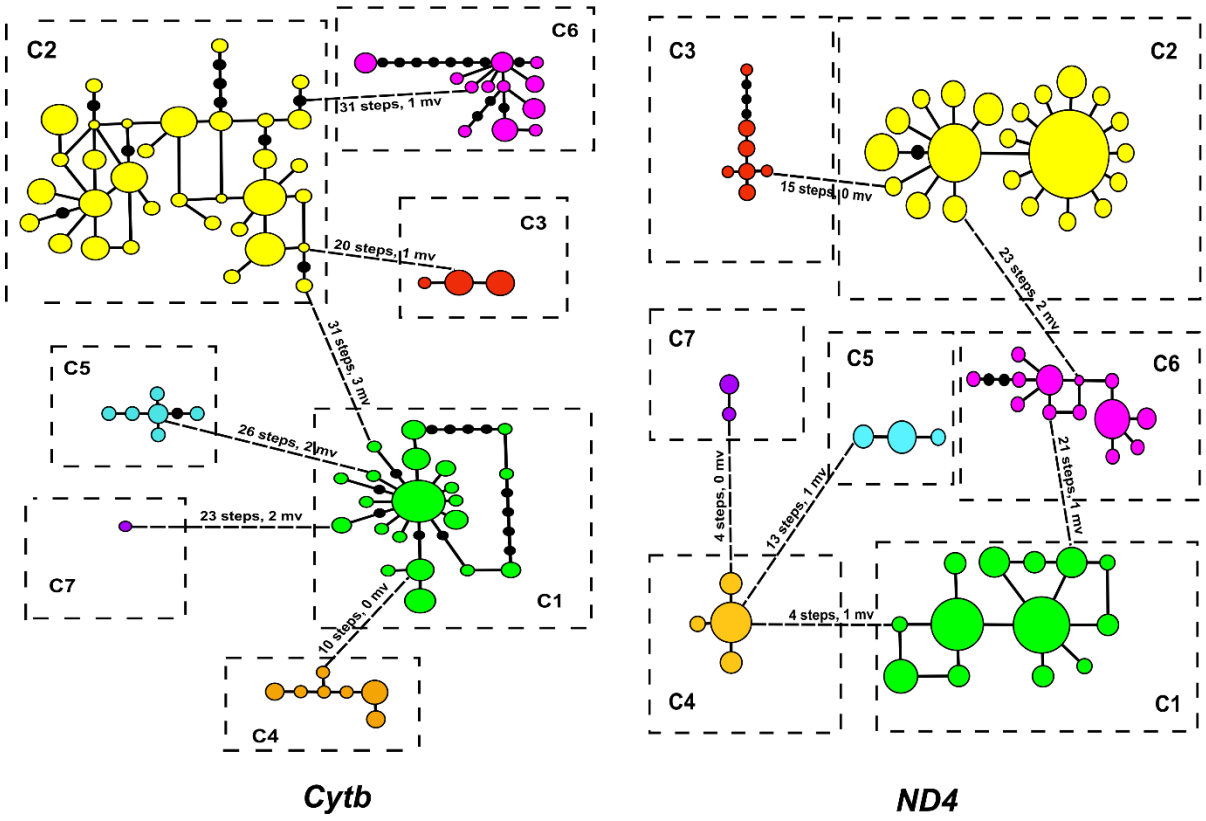


Figure S2.4. TCS networks for the Cyt-b and ND4 genes. The number of mutation steps and hypothetical median vectors (mv) are indicated. The black dots represent mutation steps within each clade. Circle size is proportional to the number of samples for each haplotype.

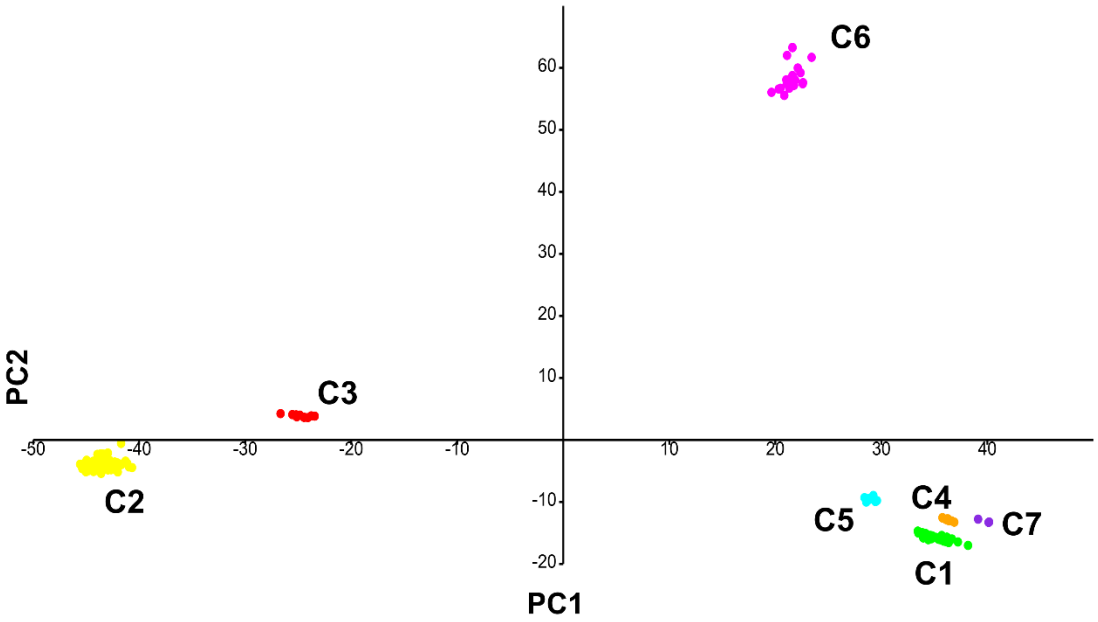


Figure S2.5. Multivariate results of motif PCA analysis of the mtDNA dataset (12S, 16S, Cyt-b, ND4, tRNA-His and tRNA-Ser) derived from DAMBE. All seven retrieved clusters representing C1-C7 are depicted following the same colour schemes applied in the phylogenetic analysis.

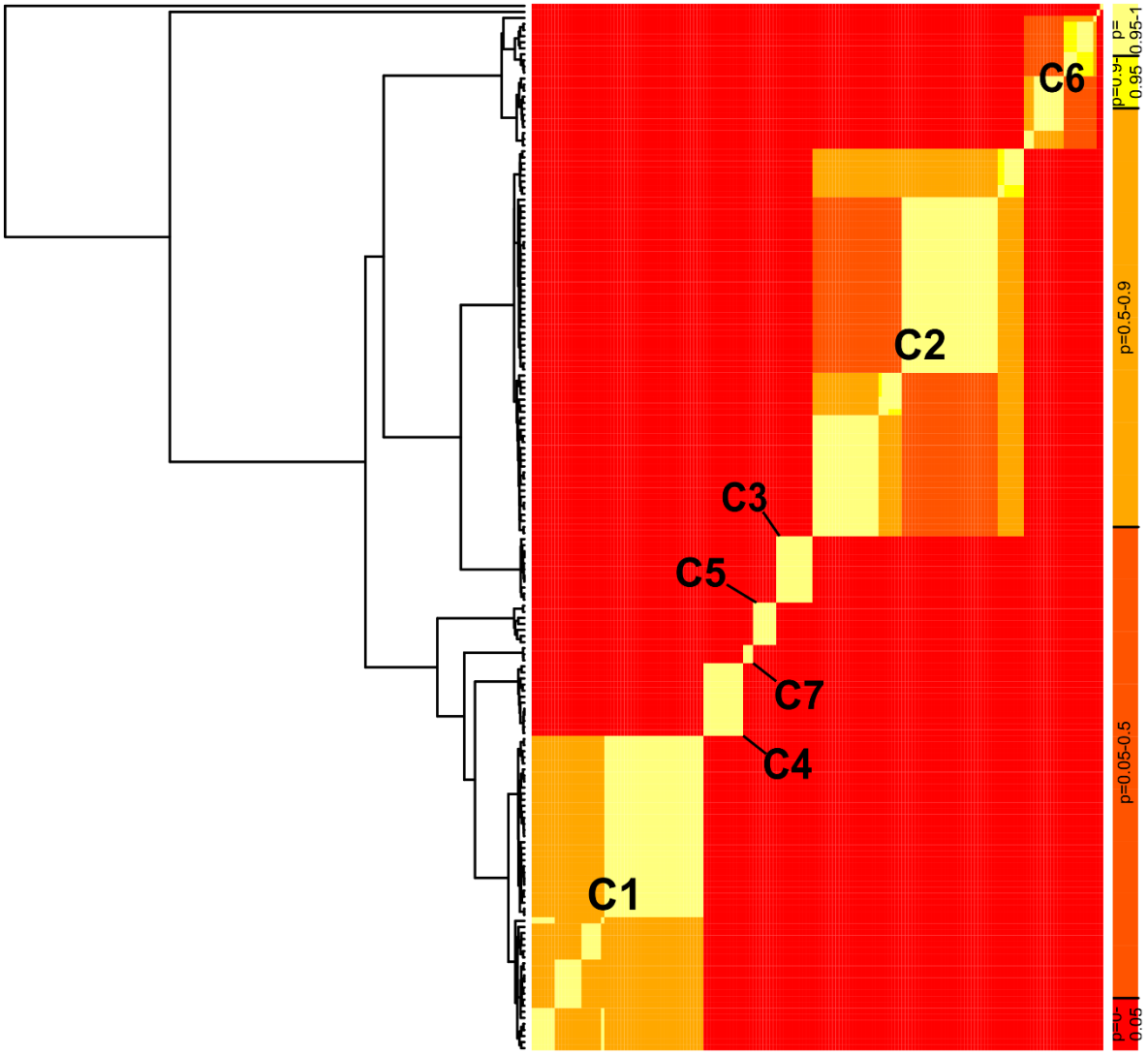


Figure S2.6. The bGMYC pairwise posterior probability matrix visualized as a heat map with the BEAST MSC ultra-metric tree (retrieved from mtDNA). The background colour represents pairwise probabilities with multiple threshold p-values (red colour represents the possibility of the same putative species at its lowest probability, whilst, the white colour represents it at its highest probability). The pairwise posterior probabilities of nested tips from the same clade are indicated in the matrix.

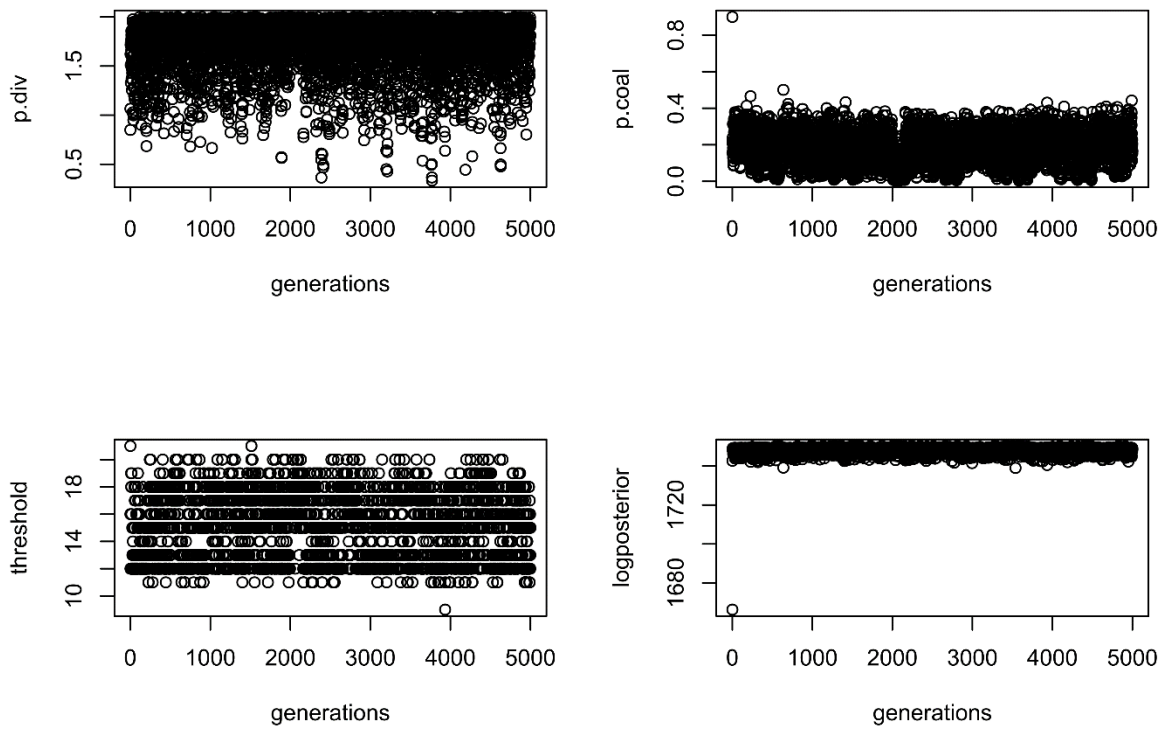


Figure S2.7. The plots are for visualizing three different parameters and log posterior values derived from an MCMC run using the single tree generated from BEAST (STACEY) BI analysis. The four plots facilitate monitoring whether sampling chains were well mixed and sampling sufficient.

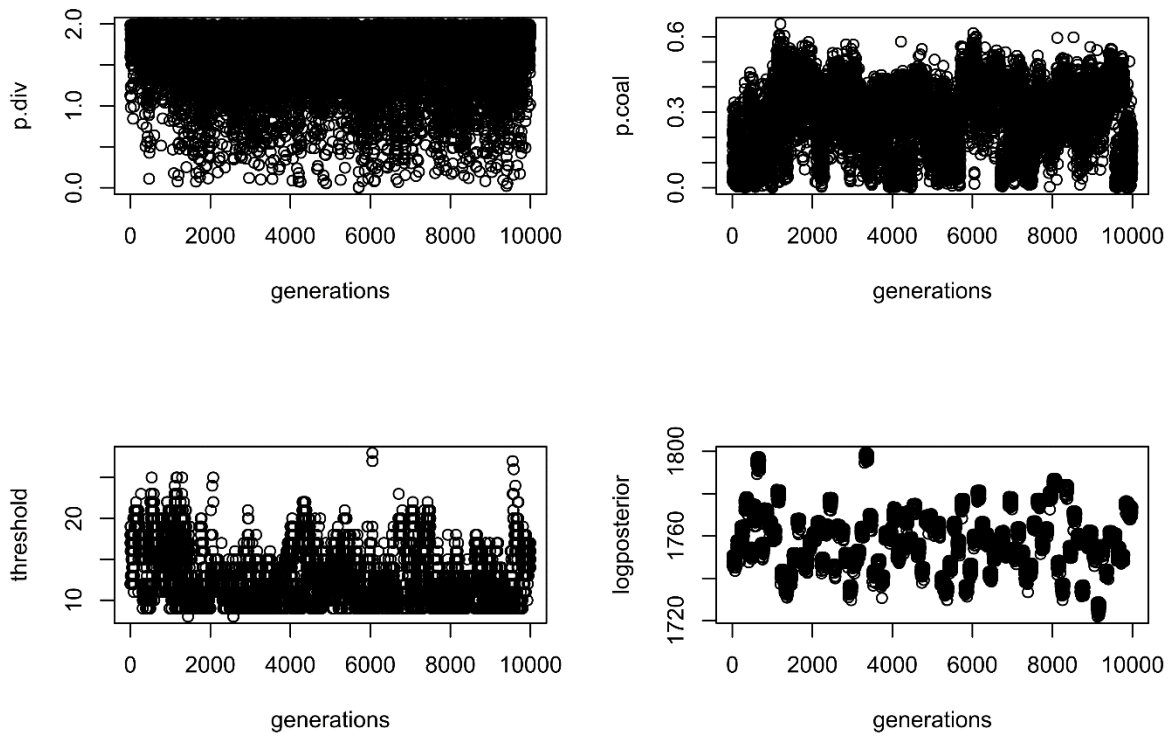


Figure S2.8. The plots are for visualizing three different parameters and log posterior values derived from an MCMC run using the 100 random trees generated from BEAST (STACEY) BI analysis. The four plots facilitate monitoring whether sampling chains were well mixed and sampling sufficient.

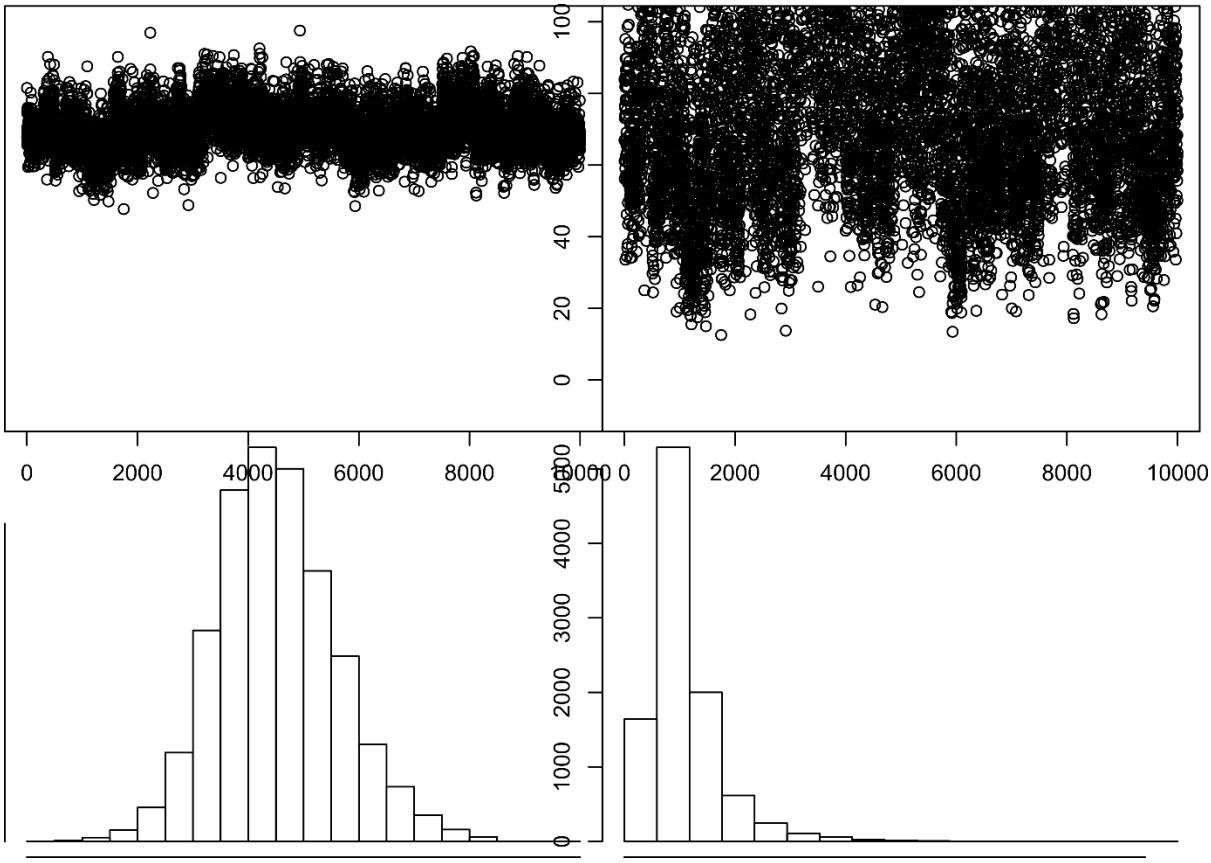


Figure S2.9. The plots for visualizing the rate of branching for the coalescent process across all generations of MCMC runs (top). When the ratios of the Coalescence to Yule rates sampled in the analysis are larger than 0, the model may be considered as a good approximation of the accuracy of the data, and vice versa. The histograms (bottom) indicate whether overall sampling during the MCMC runs was normally distributed, implying that sampling was sufficient. The plots on the left refer to the Coalescent process rates, whilst, the plots on the right refer to the Yule process rates.

Supplementary Tables

All the Supplementary Tables (Table S2.1-S2.7) were not included in the main document of this thesis, since their size are large and unable to fit in the main document. Therefore, all Supplementary Tables will only be provided independently as supporting documents.

References

Aldhebiani, A. Y. (2018). Species concept and speciation. *Saudi Journal of Biological sciences*, *25*(3), 437-440.

Allendorf, F. W., & Luikart, G. (2009). *Conservation and the Genetics of Populations*. Oxford: John Wiley & Sons.

Austin, J. J., Arnold, E. N., & Bour, R. (2003). Was there a second adaptive radiation of giant tortoises in the Indian Ocean? Using mitochondrial DNA to investigate speciation and biogeography of *Aldabrachelys* (Reptilia, Testudinidae). *Molecular Ecology*, *12*(6), 1415-1424.

Avise, J. C., Bowen, B. W., Lamb, T., Meylan, A. B., & Bermingham, E. (1992). Mitochondrial DNA evolution at a turtle's pace: evidence for low genetic variability and reduced microevolutionary rate in the Testudines. *Molecular Biology and Evolution*, *9*(3), 457-473.

Balakrishnan, R. (2005). Species concepts, species boundaries and species identification: a view from the tropics. *Systematic Biology*, *54*(4), 689-693.

Baum, D. A., & Smith, S. D. (2013). *Tree thinking: an introduction to phylogenetic biology*. Greenwood Village: CO: Roberts.

Bouckaert, R. R. (2010). DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics*, *26*(10), 1372-1373.

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.

Boycott, R. C., & Bourquin, O. (1988). The South African Tortoise Book. *A Guide to South African Tortoises, Terrapins and Turtles*. Johannesburg: Southern Book Publisher.

Boycott, R. C., & Bourquin, O. (2000). *The Southern African Tortoise Book: A Guide to Southern African Tortoises, Terrapins and Turtles*. O. Bourquin.

Branch, B. (1998). *Field Guide to Snakes and Other Reptiles of Southern Africa*. Sanibel Island, FL: Ralph Curtis Books. Cape Town.

Branch, B. (2008). *Tortoise, Terrapins and Turtles of Africa*. Cape Town, South Africa: Struik Publisher.

Branch, W. R., Benn, G. A., & Lombard, A. T. (1995). The tortoises (Testudinidae) and terrapins (Pelomedusidae) of southern Africa: their diversity, distribution and conservation. *South African Journal of Zoology*, *30*(3), 91-102.

Caccone, A., Gentile, G., Burns, C. E., Sezzi, E., Bergman, W., Ruelle, M., Saltonstall, K., & Powell, J. R. (2004). Extreme difference in rate of mitochondrial and nuclear DNA evolution in a large ectotherm, Galápagos tortoises. *Molecular Phylogenetics and Evolution*, *31*(2), 794-798.

Caron, D. A., Countway, P. D., Savai, P., Gast, R. J., Schnetzer, A., Moorthi, S. D., Dennett, M. R., Moran, D. M. & Jones, A. C. (2009). Defining DNA-based operational taxonomic units for microbial-eukaryote ecology. *Journal of Applied & Environmental Microbiology*, *75*(18), 5797-5808.

Clement, M., Posada, D. C. K. A., & Crandall, K. A. (2000). TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, *9*(10), 1657-1659.

Core Team, R. C. T. R. (2017). R: A language and environment for statistical computing. *R Foundation for statistical computing, Vienna*. URL http://www.R-project.org/.

Cracraft, J. (1983). Species concepts and speciation analysis. In *Current ornithology* (pp. 159-187). Springer, Boston, MA.

Cunningham, J. (2002). *A molecular perspective on the family Testudinidae Batsch, 1788*, Ph.D. Dissertation, University of Cape Town.

Daniels, S. R., Hofmeyr, M. D., Henen, B. T., & Baard, E. H. W. (2010). Systematics and phylogeography of a threatened tortoise, the speckled padloper. *Animal Conservation*, *13*(3), 237-246.

Daniels, S. R., Hofmeyr, M. D., Henen, B. T., & Crandall, K. A. (2007). Living with the genetic signature of Miocene induced change: evidence from the phylogeographic structure of the endemic angulate tortoise *Chersina angulata*. *Molecular Phylogenetics and Evolution*, *45*(3), 915-926.

Darlu, P., & Lecointre, G. (2002). When does the incongruence length difference test fail?. *Molecular Biology and Evolution*, *19*(4), 432-437.

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, *9*(8), 772.

De Queiroz K. (1998). The general lineage concept of species, species criteria, and the process of speciation: A conceptual unification and terminological recommendations. In D. J. Howard & S. H. Berlocher (Eds.), *Endless forms: Species and speciation* (pp.57-75). New York: Oxford University Press.

De Queiroz K. (1999). The general lineage concept of species and the defining properties of the species category. In R. A. Wilson (Ed.), *Species: New interdisciplinary essays* (pp. 49-89). Cambridge: MIT Press.

De Queiroz, K. (2005a). Different species problems and their resolution. *BioEssays*, *27*(12), 1263-1269.

De Queiroz, K. (2005b). Ernst Mayr and the modern concept of species. *Proceedings of the National Academy of Sciences*, *102*(suppl 1), 6600-6607.

De Queiroz, K. (2007). Species concepts and species delimitation. *Systematic Biology*, *56*(6), 879-886.

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, *29*(8), 1969-1973.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792-1797.

Edwards, S. V. (2009). Is a new and general theory of molecular systematics emerging?. *Evolution: International Journal of Organic Evolution*, *63*(1), 1-19.

Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564-567.

Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, *131*(2), 479-491.

Ezard, T., Fujisawa, T., & Barraclough, T. G. (2009). Splits: species' limits by threshold statistics. *R package version*, *1*(11), r29.

Farris, J. S., Källersjö, M., Kluge, A. G., & Bult, C. (1994). Testing significance of incongruence. *Cladistics*, *10*(3), 315-319.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, *39*(4), 783-791.

Fontaneto, D., Herniou, E. A., Boschetti, C., Caprioli, M., Melone, G., Ricci, C., & Barraclough, T. G. (2007). Independently evolving species in asexual bdelloid rotifers. *PLoS Biology*, *5*(4), e87.

Fritz, U., Auer, M., Chirikova, M. A., Duysebayeva, T. N., Eremchenko, V. K., Kami, H. G., Kashkarov, R. D., Masroor, R., Moodley, Y., Pindrani, A., Široký, P., & Hundsdörfer, A. K. (2009). Mitochondrial diversity of the widespread Central Asian steppe tortoise (*Testudo horsfieldii* Gray, 1844): implications for taxonomy and relocation of confiscated tortoises. *Amphibia-Reptilia*, *30*(2), 245-257.

Fritz, U., & Bininda-Emonds, O. R. (2007). When genes meet nomenclature: tortoise phylogeny and the shifting generic concepts of *Testudo* and *Geochelone*. *Zoology*, *110*(4), 298-307.

Fritz, U., Hundsdörfer, A. K., Široký, P., Auer, M., Kami, H. G., Lehmann, J., Mazanaeva, L. F., Türkozan, O., & Wink, M. (2007). Phenotypic plasticity leads to incongruence between morphology-based taxonomy and genetic differentiation in western Palaearctic tortoises (*Testudo graeca* complex; Testudines, Testudinidae). *Amphibia-Reptilia*, *28*(1), 97-121.

Fritz, U., Stuckas, H., Vargas-Ramírez, M., Hundsdörfer, A. K., Maran, J., & Päckert, M. (2012). Molecular phylogeny of Central and South American slider turtles: implications for biogeography and systematics (Testudines: Emydidae: *Trachemys*). *Journal of Zoological Systematics and Evolutionary Research*, *50*(2), 125-136. https://doi.org/10.1111/j.1439-0469.2011.00647.x

Fu, Y. X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, *147*(2), 915-925.

Fujisawa, T., & Barraclough, T. G. (2013). Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent approach: a revised method and evaluation on simulated data sets. *Systematic Biology*, *62*(5), 707-724.

Fujita, M. K., Leaché, A. D., Burbrink, F. T., McGuire, J. A., & Moritz, C. (2012). Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology & Evolution*, *27*(9), 480-488.

Gonçalves, H., Martínez-Solano, I., Ferrand, N., & García-París, M. (2007). Conflicting phylogenetic signal of nuclear vs mitochondrial DNA markers in midwife toads (Anura, Discoglossidae, Alytes): deep coalescence or ancestral hybridization?. *Molecular Phylogenetics and Evolution*, *44*(1), 494-500.

Greig, J. C. (1975). Patterns In the distribution of Southern African terrestrial tortoises (Chelonia: Cryptodira: Testudinidae). *The Journal of the Herpetological Association of Africa*, *14*(1), 9-9.

Greig, J. C., & Burdett, P. D. (1976). Patterns in the distribution of southern African terrestrial tortoises (Cryptodira: Testudinidae). *African Zoology*, *11*(2), 251-273.

Grummer, J. A., Bryson Jr, R. W., & Reeder, T. W. (2013). Species delimitation using Bayes factors: simulations and application to the *Sceloporus scalaris* species group (Squamata: Phrynosomatidae). *Systematic Biology*, *63*(2), 119-133.

Heled, J., & Drummond, A. J. (2009). Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, *27*(3), 570-580.

Hewitt J. 1933. On the Cape species and subspecies of the genus *Chersinella* Gray. Part I. *Annals of the Natal Museum*, *7*, 255-297.

Hewitt, J. (1934). On the Cape species and subspecies of the genus *Chersinella*. Part II. *Annals of the Natal Museum*, *7*, 303-352.

Hillis, D. M., & Bull, J. J. (1993). An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology*, *42*(2), 182-192.

Hofmeyr, M. D., Boycott, R. C., & Baard, E. H. W. (2014). *Psammobates tentorius* (Bell, 1828). In: M. F. Bates, W. R. Branch, A. M. Bauer, M. Burger, J. Marais, G. J. Alexander & M. S. De Villiers (Eds.), *Atlas and red list of the reptiles of South Africa, Lesotho and Swaziland* (pp. 70-85). Pretoria: South African National Biodiversity Institute.

Hofmeyr, M. D., Vamberger, M., Branch, W., Schleicher, A., & Daniels S. R. (2017). Tortoise (Reptilia, Testudinidae) radiations in Southern Africa from the Eocene to the present. *Zoologica Scripta*, *46*, 389-400.

Huelsenbeck, J. P., & Rannala, B. (2004). Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Systematic Biology*, *53*(6), 904-913.

Huelsenbeck, J. P., & Ronquist, F. (2001). MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics*, *17*(8), 754-755.

Huelsenbeck, J. P., Ronquist, F., Nielsen, R., & Bollback, J. P. (2001). Bayesian inference of phylogeny and its impact on evolutionary biology. *Science*, *294*(5550), 2310-2314.

Hull, D. L. (1977). The ontological status of species as evolutionary units. In *Foundational problems in the special sciences* (pp. 91-102). Springer, Dordrecht.

IUCN. (2018). *The IUCN Red List of Threatened Species*. Version 2018-1.

Iverson, J. B., Spinks, P. Q., Shaffer, H. B., McCord, W. P., & Das, I. (2001). Phylogenetic relationships among the Asian tortoises of the genus *Indotestudo* (Reptilia: Testudines: Testudinidae). *HAMADRYAD-MADRAS-*, *26*, 272-275.

Jones, G. (2015). Species delimitation and phylogeny estimation under the multispecies coalescent. *BioRxiv*, 010199.

Jones, G. (2017). Algorithmic improvements to species delimitation and phylogeny estimation under the multispecies coalescent. *Journal of Mathematical Biology*, *74*(1-2), 447-467.

Jones, G., Aydin, Z., & Oxelman, B. (2014). DISSECT: an assignment-free Bayesian discovery method for species delimitation under the multispecies coalescent. *Bioinformatics*, *31*(7), 991-998.

Kapli, P., Lutteropp, S., Zhang, J., Kobert, K., Pavlidis, P., Stamatakis, A., & Flouri, T. (2017). Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo. *Bioinformatics*, *33*(11), 1630-1638.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773-795.

Kindler, C., Branch, W. R., Hofmeyr, M. D., Maran, J., Široký, P., Vences, M., Harvey, J., Hauswaldt, J. S., Schleicher, A., Stuckas, H., & Fritz, U. (2012). Molecular phylogeny of African hinge-back tortoises (*Kinixys*): implications for phylogeography and taxonomy (Testudines: Testudinidae). *Journal of Zoological Systematics and Evolutionary Research, 50*, 192-201.

Konishi, T., Matsukuma, S., Fuji, H., Nakamura, D., Satou, N., & Okano, K. (2019). Principal Component Analysis applied directly to Sequence Matrix. *Scientific Reports, 9*(1), 1-13.

Kocher, T. D., Thomas, W. K., Meyer, A., Edwards, S. V., Pääbo, S., Villablanca, F. X., & Wilson, A. C. (1989). Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences, 86*(16), 6196-6200.

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution, 33*(7), 1870-1874.

Kunz, W. (2013). *Do species exist?: Principles of taxonomic classification*. New Haven: John Wiley & Sons.

Lanfear, R., Calcott, B., Ho, S. Y., & Guindon, S. (2012). PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution, 29*(6), 1695-1701.

Le, M., Raxworthy, C. J., McCord, W. P., & Mertz, L. (2006). A molecular phylogeny of tortoises (Testudines: Testudinidae) based on mitochondrial and nuclear genes. *Molecular Phylogenetics and Evolution, 40*(2), 517-531.

Leaché, A. D., & Fujita, M. K. (2010). Bayesian species delimitation in West African forest geckos (Hemidactylus fasciatus). *Proceedings of the Royal Society B: Biological Sciences, 277*(1697), 3071-3077.

Leaché, A. D., Zhu, T., Rannala, B., & Yang, Z. (2018). The spectre of too many species. *Systematic Biology, 68*(1), 168-181.

Leigh, J. W., & Bryant, D. (2015). PopART: full-feature software for haplotype network construction. *Methods in Ecology and Evolution, 6*(9), 1110-1116.

Leliaert, F., Verbruggen, H., Vanormelingen, P., Steen, F., López-Bautista, J. M., Zuccarello, G. C., & De Clerck, O. (2014). DNA-based species delimitation in algae. *European Journal of Phycology*, *49*(2), 179-196.

Librado, P., & Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, *25*(11), 1451– 1452.

Liu, L., Yu, L., Pearl, D. K., & Edwards, S. V. (2009). Estimating species phylogenies using coalescence times among sequences. *Systematic Biology*, *58*(5), 468-477.

Lourenço, J. M., Claude, J., Galtier, N., & Chiari, Y. (2012). Dating cryptodiran nodes: origin and diversification of the turtle superfamily Testudinoidea. *Molecular Phylogenetics and Evolution*, *62*(1), 496-507.

Loveridge, A., & Williams, E. E. (1957). Revision of the African tortoises and turtles of the suborder Cryptodira. *Bulletin of the Museum of Comparative Zoology*, *115*(6), 161-557.

Luo, A., Ling, C., Ho, S. Y., & Zhu, C. D. (2018). Comparison of methods for molecular species delimitation across a range of speciation scenarios. *Systematic Biology*, *67*(5), 830-846.

Mallet, J. (2013). Concepts of species. In S. Levin, (Ed), *Encyclopedia of Biodiversity*, volume 6, (pp.679–691). Massachusetts: Academic Press.

Malone, J. H., & Fontenot, B. E. (2008). Patterns of reproductive isolation in toads. *PLoS One*, *3*(12), e3900.

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., & Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution*, *1*(1).

Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Gateway Computing Environments Workshop (GCE)*, 2010, 1–8.

Miralles, A., & Vences, M. (2013). New metrics for comparison of taxonomies reveal striking discrepancies among species delimitation methods in *Madascincus* lizards. *PLoS One*, *8*(7), e68242.

Mysara, M., Vandamme, P., Props, R., Kerckhof, F. M., Leys, N., Boon, N., Raes, J., & Monsieurs, P. (2017). Reconciliation between operational taxonomic units and species boundaries. *FEMS Microbiology Ecology*, *93*(4).

Nagy, Z. T., Sonet, G., Glaw, F., & Vences, M. (2012). First large-scale DNA barcoding assessment of reptiles in the biodiversity hotspot of Madagascar, based on newly designed COI primers. *PLoS One, 7*(3), e34506.

Pääbo S. (1990). Amplifying ancient DNA. In M. A. Innis, D. H. Gelfand, J. J. Sninsky & T. J. White (Eds), *PCR-Protocols and Applications - A Laboratory Manual* (pp. 159– 66). San Diego: Academic Press.

Palumbi, S. R., Martin, A., Romano, S., Mcmillan, W. O., Stice L., & Grabowski G. (1991). *The simple fool's guide to PCR. A collection of PCR protocols*, version 2. Honolulu: University of Hawaii.

Python Software Foundation, (2010). Python Language Reference, version 2.7. Available at http://www.python.org

Petzold, A., Vargas-Ramirez, M., Kehlmaier, C., Vamberger, M., Branch, W. R., Du Preez, L., Hofmeyr, M.D., Meyer, L., Schleicher, A., Široký, P.A.V.E.L., & Fritz, U. (2014). A revision of African helmeted terrapins (Testudines: Pelomedusidae: *Pelomedusa*), with descriptions of six new species. *Zootaxa*, *3795*(5), 523-548.

Pons, J., Barraclough, T. G., Gomez-Zurita, J., Cardoso, A., Duran, D. P., Hazell, S., Kamoun, S., Sumlin, W. D., & Vogler, A. P. (2006). Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, *55*(4), 595-609.

Preheim, S. P., Perrotta, A. R., Martin-Platero, A. M., Gupta, A., & Alm, E. J. (2013). Distribution-based clustering: using ecology to refine the operational taxonomic unit. *Journal of Applied & Environmental Microbiology*, *79*(21), 6593-6603.

Puillandre, N., Lambert, A., Brouillet, S., & Achaz, G. (2012). ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology*, *21*(8), 1864-1877.

Python Software Foundation, (2010). Python Language Reference, version 2.7. Available at http://www.python.org

Rambaut A. http://tree.bio.ed.ac.uk/software/figtree/ (2012).

Rambaut, A., Suchard, M. A., Xie, D., & Drummond, A. J. (2014). Tracer 1.6. *URL: http://beast. bio. ed. ac. uk/tracer*.

Rannala, B., & Yang, Z. (2003). Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics*, *164*(4), 1645-1656.

Rato, C., Carranza, S., Perera, A., Carretero, M. A., & Harris, D. J. (2010). Conflicting patterns of nucleotide diversity between mtDNA and nDNA in the Moorish gecko, *Tarentola mauritanica*. *Molecular Phylogenetics and Evolution*, *56*(3), 962-971.

Reid, N. M. (2014). bGMYC: a Bayesian MCMC implementation of the general mixed Yule-coalescent model for species delimitation. R package version, 1(2).

Reid, N. M., & Carstens, B. C. (2012). Phylogenetic estimation error can decrease the accuracy of species delimitation: a Bayesian implementation of the general mixed Yule-coalescent model. *BMC Evolutionary Biology*, *12*(1), 196.

Rhodin, A. G. J., Iverson, J. B., Bour, R., Fritz U., Georges, A., Shaffer, H. B., & van Dijk P.P. (2017). Turtles of the World: Annotated Checklist and Atlas of Taxonomy, Synonymy, Distribution, and Conservation Status (8th ed.). In: A. G. J. Rhodin, J. B. Iverson, P. P. van Dijk, R. A. Saumure, K. A. Buhlmann, P. C. H Pritchard & R. A. Mittermeier (Eds.), *Conservation Biology of Freshwater Turtles and Tortoises: A Compilation Project of the IUCN/SSC Tortoise and Freshwater Turtle Specialist Group* (pp. 1–292). Lunenburg: Chelonian Research Monographs.

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., & Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, *61*(3), 539-542.

Rubinoff, D., & Holland, B. S. (2005). Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. *Systematic Biology*, *54*(6), 952-961.

Scudder, G. G. E. (1974). Species concepts and speciation. *Canadian Journal of Zoology*, *52*(9), 1121-1134.

Shaw, K. L. (2002). Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proceedings of the National Academy of Sciences*, *99*(25), 16122-16127.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312-1313.

Stöver, B. C., & Müller, K. F. (2010). TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics*, *11*(1), 7.

Stuart, B. L., & Parham, J. F. (2004). Molecular phylogeny of the Critically Endangered Indochinese box turtle (*Cuora galbinifrons*). *Molecular Phylogenetics and Evolution*, *31*(1), 164-177.

Sukumaran, J., & Knowles, L. L. (2017). Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences*, *114*(7), 1607-1612.

Swofford, D. L. (1998). *Paup 4.0 beta version for windows: phylogenetic analysis using parsimony*. Sunderland, Massachusetts: Sinauer Associates.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), 585-595.

Tollis, M., DeNardo, D. F., Cornelius, J. A., Dolby, G. A., Edwards, T., Henen, B. T., Karl, A. E., Murphy, R. W., & Kusumi, K. (2017). The Agassiz's desert tortoise genome provides a resource for the conservation of a threatened species. *PloS One*, *12*(5), e0177708.

Townsend, T. M., Alegre, R. E., Kelley, S. T., Wiens, J. J., & Reeder, T. W. (2008). Rapid development of multiple nuclear loci for phylogenetic analysis using genomic resources: an example from squamate reptiles. *Molecular Phylogenetics and Evolution*, *47*(1), 129-142.

Vaidya, G., Lohman, D. J., & Meier, R. (2011). SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics*, *27*(2), 171-180.

Vamberger, M., Hofmeyr, M.D., Ihlow, F. & Fritz, U. (2018). In quest of contact: phylogeography of helmeted terrapins (*Pelomedusa galeata*, *P. subrufa* sensu stricto). *PeerJ*, *6*, e4901.

Vargas-Ramirez, M., Carr, J. L., & Fritz, U. (2013). Complex phylogeography in *Rhinoclemmys melanosterna*: conflicting mitochondrial and nuclear evidence suggests past hybridization (Testudines: Geoemydidae). *Zootaxa*, *3670*(2), 238-254.

Vargas-Ramírez, M., Vences, M., Branch, W. R., Daniels, S. R., Glaw, F., Hofmeyr, M. D., Kuchling, G., Maran, J., Papenfuss, T. J., Široký, P., Vieites, D. R., & Fritz, U. (2010). Deep genealogical lineages in the widely distributed African helmeted terrapin: evidence from mitochondrial and nuclear DNA (Testudines: Pelomedusidae: *Pelomedusa subrufa*). *Molecular Phylogenetics and Evolution*, *56*, 428-440.

Whiting, A. S., Bauer, A. M., & Sites Jr, J. W. (2003). Phylogenetic relationships and limb loss in sub-Saharan African scincine lizards (Squamata: Scincidae). *Molecular Phylogenetics and Evolution*, *29*(3), 582-598.

Xia, X. (2013). DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. *Molecular Biology and Evolution*, *30*(7), 1720-1728.

Yang, Z. (2015). The BPP program for species tree estimation and species delimitation. *Current Zoology*, *61*(5), 854-865.

Yang, Z., & Rannala, B. (2010). Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences*, *107*(20), 9264-9269.

Yang, Z., & Rannala, B. (2017). Bayesian species identification under the multispecies coalescent provides significant improvements to DNA barcoding analyses. *Molecular Ecology*, *26*(11), 3028-3036.

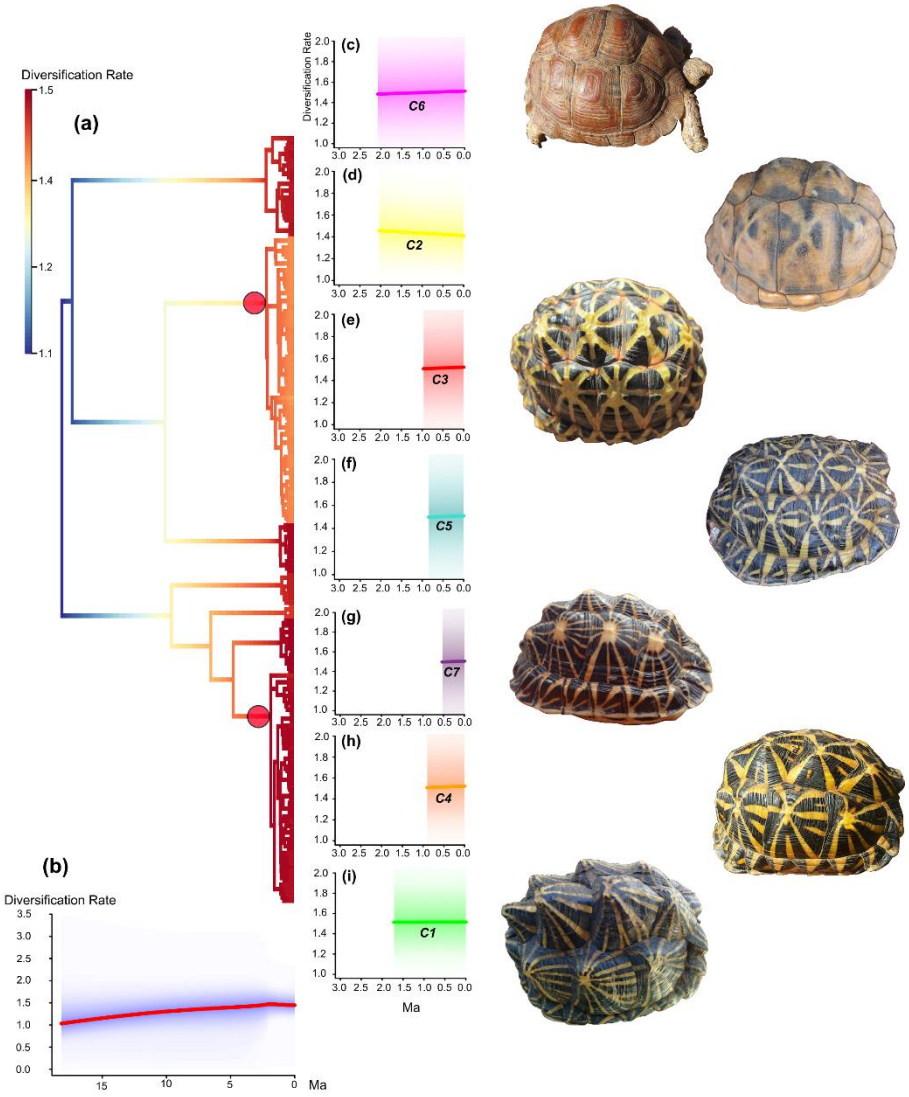
Yoder, A. D., Irwin, J. A., & Payseur, B. A. (2001). Failure of the ILD to determine data combinability for slow loris phylogeny. *Systematic Biology*, *50*(3), 408-424.

Zhang, J., Kapli, P., Pavlidis, P., & Stamatakis, A. (2013). A general species delimitation method with applications to phylogenetic placements. *Bioinformatics*, 29, 2869-2876.

Zheng, Y., Peng, R., Kuro-o, M., & Zeng, X. (2011). Exploring patterns and extent of bias in estimating divergence time from mitochondrial DNA sequence data in a particular lineage: a case study of salamanders (Order Caudata). *Molecular Biology and Evolution*, *28*(9), 2521-2535.

Chapter 3. Climatic and topographic changes since the Miocene may have influenced the radiation and biogeography of tent tortoises (*Psammobates tentorius*) in southern Africa

Graphical abstract



Abstract

Psammobates tentorius ranges over western and central South Africa, and southwestern Namibia. The species consists of seven mtDNA clades showing deep genetic divergences. Based on knowledge of other reptiles, I hypothesised that climatic and topographic changes since the Miocene shaped radiation patterns in *P. tentorius*, and that the species would have higher genetic diversity in the Greater Cape Floristic Region (GCFR) than elsewhere. I used a dated phylogeny to assess if the timing of cladogenic events corresponded to climatic and/or landscape changes and evaluated different evolutionary drivers and diversification rates to understand evolutionary patterns. Cladogenesis in *P. tentorius* started in the early to mid-Miocene when populations dispersed from north to south to form two geographically isolated groups. The mid-Miocene drop in temperature possibly caused populations to retract into refugia, contributing to vicariant evolution. The northern group diverged into a clade north of the Orange River (OR), followed by a split into a western and an interior clade south of the OR. The latter divergence corresponded to the development of the cold Benguela current, which caused western aridification and rainfall seasonality. In the south, tectonic uplift and subsequent exhumation, together with climatic fluctuations, seem to be responsible for radiations among the four southern clades since the late-Miocene. I found that genetic richness was higher in the GCFR than the adjacent Nama Karoo, as also in plants and some animal taxa. I conclude that *P. tentorius* contains cryptic species and that vicariant events allowed allopatric speciation within this group.

3.1 Introduction

Climate fluctuations have often been invoked to explain floral and faunal diversification on a global (deMenocal, 2004; Hewitt, 2000; Jansson & Dynesius, 2002), also at the regional scale, e.g., in southern Africa (Neumann & Bamford, 2015; Tolley et al., 2014). Global cooling after the Early Eocene Climatic Optimum led to the formation of the Antarctic ice-sheet by the early Oligocene, whereafter temperatures increased and remained high from the late Oligocene to the Mid-Miocene Climatic Optimum, 17-15 Ma (Zachos et al., 2001). Subsequently, decreasing temperatures re-established the Antarctic ice-sheet by 10 Ma and established the Arctic ice-sheet by 3.2 Ma (Zachos et al., 2001). Apart from global cooling, southern Africa became progressively more arid since the Oligocene (Burke & Gunnell, 2008), and the development of the Benguela Current 14-10 Ma (Hoffmann et al., 2015; Rommerskirchen et al., 2011) increased western aridity and established seasonal rainfall patterns over southern Africa by the late Miocene (Hoffmann et al., 2015; Neumann & Bamford, 2015).

Besides climate, landscape evolution also contributed to genetic diversification of southern Africa's fauna and flora (Tolley et al., 2014), with the Cape Fold Mountains (CFMs), Great Escarpment (GE), and Orange River (OR) identified as potential geographic barriers to gene exchange in some taxa (Bauer & Lamb, 2005; Scott et al., 2004; Tolley et al., 2009). The CFMs were already formed before rifting of Gondwanaland in the Lower Cretaceous (Grab & Knight, 2015), but differential uplift and exhumation in the last 30 Ma (Green et al., 2017) may have contributed to the current relief, and influenced distribution corridors. The GE runs 50-200 km inland from the coastline and separates the coastal plains from an elevated plateau more than 1000 m above sea level (Clark et al., 2011; Grab & Knight, 2015). Many researchers believe that the GE represents a passive, erosional remnant of the continental margin after the break-up of Gondwanaland, with its current topography dating to the end of the Cretaceous (Clark et al., 2011 and references therein). In contrast, some geomorphological and thermochronological studies advocate that upward flexures by 30 Ma (Burke & Gunnell, 2008) or 20 and 5 Ma (Partridge & Maud, 1987), followed by erosion, established, or altered the topography of the GE. Uplift events changed the course of fluvial systems (Moore et al., 2009), which can also affect genetic exchange among populations (Scott et al., 2004). The current course of the OR represents the confluence of two paleo river systems (de Wit, 1999). The paleo OR (Kalahari River) drained most of Namibia and southern Botswana since the Cretaceous while the Karoo

River (Orange and Vaal Rivers) drained most of South Africa, and had its mouth further south at the current Olifants River (de Wit, 1999; Kounov et al., 2008). The Kalahari River captured the upper courses of the Karoo River by the early Cenozoic (de Wit, 1999).

Floral development in southern Africa has been studied extensively, with strong evidence that late Neogene climatic and geomorphic evolution (due to tectonic events) shaped development of most current biomes (Cowling et al., 2015; Verboom et al., 2014). Reptile diversity is relatively high in southern Africa (Bauer, 1993; Branch, 2006) and genetic diversification in several taxa has been linked to climatic change, particularly during the Pliocene-Pleistocene period, although some radiations date back to the Miocene (Tolley et al., 2014). In addition, landscape features have been postulated as having demarcated phylogeographic clades in some reptile taxa (Bauer & Lamb, 2005; Heinicke et al., 2017; Matthee & Flemming, 2002).

Southern Africa has a relatively high tortoise diversity, harbouring one-third of the world's tortoise genera (Rhodin et al. 2017). Cladogenesis of southern African tortoises at the genus level occurred mainly in the Eocene, with most species diverging between the Oligocene and mid-Miocene (Hofmeyr et al., 2017); thus, divergences in this group is deep. The latter study could not resolve relationships among the three *Psammobates* species, possibly due to rapid speciation between the late Oligocene and early Miocene. The study also showed deep divergences within *Psammobates tentorius*, indicating the possible existence of cryptic taxa. In Chapter 2, I used six mtDNA and one nDNA marker, and multiple species delimitation procedures, to investigate the phylogenetic structure of *P. tentorius*. The results of the mtDNA and nDNA phylogenies were incongruent, as also found in many other studies on slow-evolving chelonian species (Avise et al., 1992; Shaffer et al., 2013). The nDNA phylogeny differentiated the three recognized subspecies, *P. t. tentorius*, *P. t. trimeni*, and *P. t. verroxii*. In contrast, the mtDNA phylogeny distinguished seven monophyletic clades with strong support, with each clade occurring in a distinct geographic region.

The distribution of *P. tentorius* covers two of the 25 biodiversity hotspots, the Succulent Karoo and Cape Floristic Region (CFR); the latter consisting mainly of Fynbos (Myers et al., 2000). In recent years, botanists combined the CFR with the western winter-rainfall region of southern Africa (mostly Succulent Karoo) below and above the western GE, to constitute a region of floral richness and endemism, the Greater Cape Floristic Region (GCFR) (Born et al., 2007). This region also has exceptionally high reptile diversity and endemism (Tolley et al., 2014).

Apart from occurring in the GCFR, the distribution of *P. tentorius* also extends into the Nama Karoo biome of central South Africa and southern Namibia, which has the lowest floral diversity of biomes on the subcontinent (Neumann & Bamford, 2015). The southeasternmost range of *P. tentorius* penetrates the westernmost distribution of Albany Thicket, a tropical biome with low floral diversity in the western region (Cowling et al., 2017).

In view of its extensive genetic diversity and occurrence in regions of high and low floral richness and endemism, *P. tentorius* represents a model species for exploring biogeographic and radiation patterns of reptiles in southern Africa. I hypothesized that climate fluctuations and associated habitat changes over the Neogene influenced cladogenesis in *P. tentorius*, and that physical barriers, in the form of either unsuitable climate or topography, possibly obstructed dispersal routes, leading to isolation and allopatric diversification. I also postulated that diversification would be higher in Fynbos and Succulent Karoo vegetation of the GCFR than in the other biomes, as was found for plants. To test these hypotheses, I calibrated a dated phylogeny to evaluate if the timing of cladogenic events corresponded to specific climatic and/or landscape evolution events. This was followed by a character dependency analysis to assess if barriers such as the GE, CFMs and OR influenced radiation in *P. tentorius*, and a habitat reconstruction analysis to gauge the role of vicariance, dispersal and extinction in the radiation of *P. tentorius*. In addition, I did a diversification rate analysis to better understand evolutionary patterns over time in this group.

3.2 Materials and methods

3.2.1 Taxon sampling

I sampled 455 specimens of *P. tentorius* from 76 localities throughout its distribution range in southern Africa, ensuring comprehensive coverage of its distribution range according to the literature (Hewitt, 1933, 1934; Greig, 1975; Greig & Burdett, 1976; Branch, 2008; Boycott & Bourquin, 2000; Hofmeyr et al., 2014). The list of all samples, their localities and NCBI GenBank accession numbers across different gene loci is provided in Table S3.1, with published outgroup sequence details (plus references) provided in Table S3.2.

3.2.2 DNA markers selection, DNA extraction and amplification

Sequences used in this study were generated from Chapter 2, include 2 rRNA loci, 4 mtDNA loci and 1 nDNA marker. All NCBI GenBank accession numbers of sequence data given in Table S3.1. The optimal annealing temperatures and primers used for six mtDNA markers were given in Table S3.2. I employed Partitionfinder v.1.0.1 (Lanfear *et al*., 2012) under Python v.2.7 (Python Software Foundation, 2010) to determine the best partition scheme for the concatenated dataset. I also used jModeltest v.2 (Darriba *et al*., 2012) to determine the best substitution model and parameter settings for each data partition via AIC criterion (Table S3.3). For details on substitution saturation analysis see Chapter 2.

3.2.3 Calibration dating

In order to generate an ultrametric tree to meet the requirements of habitat reconstruction analysis, I used DnaSP v.5 (Librado & Rozas, 2009) to determine the haplotype sequences, and removed all the excess identical sequences in the concatenated mtDNA dataset. Since no nucleotide saturation has been detected in Chapter 2, I performed the dating analyses with the full mtDNA dataset. I excluded nDNA PRLR from these analyses since it showed conflicting phylogenetic results compared with mtDNA genes, was less informative, and sampling coverage was less extensive (details see Table S3.1). Secondly, in Chapter 2, I discovered a some tree topology conflict between the nDNA and mtDNA based phylogeny. I discarded two haplotype sequences from unknown localities and used 169 haplotype sequences to perform Bayesian calibration dating analysis with the program BEAST v.2.4.8 (Bouckaert et al., 2014). Each partition was applied with different optimized substitution models and parameters (see Table S3.4) obtained from Chapter 2. Trees and clock models were linked, site models unlinked, and dating priors in most recent common ancestor (MRCA) based on published data (Table S3.2, Cunningham, 2002, Hofmeyr *et al*., 2017). All nodes for calibration dating were defined using the literature (Cunningham, 2002, Hofmeyr *et al*., 2017) with a sampling gap between zero and two. The Markov Chain Monte Carlo (MCMC) chain lengths were set as 50 million generations to enable the majority of parameters to be well sampled (with ESS > 200), sampling every 5000 chains. Results from each analysis were inspected with Tracer to check standard deviations and sampling adequacy (ESS >200). In order to compare calibration dating results from different dating models implemented in BEAST 2, I performed the above Bayesian calibration dating analysis on four different dating models: Birth-death (BD), Fossil Birth-death (FBD), Yule, and Coalescent Bayesian Skyline (BS). In all cases, posterior probability (PP) > 0.95 for Bayesian inference (Huelsenbeck & Rannala, 2004) were considered as strongly

supported. The LTT plot (Lineage through time) plot was drawn with the R package 'phytools' (Revell, 2012) implemented in R 3.5.1 (Core Team, 2017) based on BEAST 2 dating results.

3.2.4 Spatial AMOVA to define geographic group structure

In order to define population groups that were geographically homogeneous and maximally differentiated from each other, as well as to identify the potentially crucial genetic barriers that separate these population groups, I used SAMOVA 2.0 (Dupanloup et al., 2002) to delineate groups with temporal and spatial dimensions, K=3 to 7 were run (run n=5) for SAMOVA analyses. Only the run showed sensible matching with my phylogenetic results was regarded best run, and the best results will be in interpreting results.

3.2.5 Habitat reconstruction analysis

I partitioned the distribution range of the *P. tentorius* complex into seven main geographical areas according to the optimal grouping scheme suggested by results of the SAMOVA 2.0 analysis (see Fig. 3.1). Biogeographic reconstructions were first performed with the R package BioGeoBEARS (Matzke, 2013a; Matzke, 2014; Massana et al., 2015) in order to unify some of the traditional methods of biogeographical inference, such as the dispersal-vicariance analysis (DIVA), the Dispersal-Extinction-Cladogenesis (DEC-Lagrange) analysis and the Statistical BayArea analysis. The BioGeoBEARS analysis tests the optimal biogeographic reconstruction model by evaluating the founder effect parameter "J" (Matzke, 2013a, b; Matzke, 2014). Thus, I compared combinations of the three models with and without "J", and selected the optimal model based on AIC corrected (AICc) and AIC weight criteria with the LRT test. I also compared this approach with the Dispersal-Extinction-Cladogenesis (Bayes-Lagrange or S-DEC, Beaulieu et al., 2013) model and Statistical Dispersal-Vicariance Analysis (S-DIVA, Yu et al., 2010) implemented in RASP v.4.0 (Yu et al., 2015). These two models were not available in the package BioGeoBEARS. Finally, I used the Bayesian Binary MCMC (BBM) model in RASP v.4.0 to determine the most likely centre of origin for all possible area scenarios. The phylogenetic tree selected for the biogeographic reconstructions was based on a consensus ultrametric chronogram generated by BEAST 2 from the mtDNA dataset with the FBD model. All outgroups were truncated before initiating analyses to improve the accuracy of the reconstruction results (Yu et al., 2015).

3.2.6 Diversification rate analysis

In order to detect and visualize speciation dynamics, evolution rate heterogeneity and significant rate shifts across different clades on the calibrated BEAST chronogram, I employed the Bayesian statistical framework approach (Rabosky, 2014; Shi and Rabosky, 2015) of the BAMM software (Rabosky, 2014) together with the R packages 'BAMMtools' (Rabosky, Grundler et al., 2014) and Ape (Paradis et al., 2004). By using rjMCMC (reversible jump MCMC) methods, BAMM automatically detects candidate models of radiation and enables sampling rate shift configurations proportional to their posterior probabilities, therefore avoiding problems of single "best" rate shift configuration and making results more reliable. Analyses were conducted with four simultaneous MCMC chains for 50 million generations, using the "speciation-extinction" model, deltaT = 0.1, Swap period = 1000 and default settings for the rest of the parameters. Pairwise probability comparisons obtained were then used to reconstruct a matrix with macro-evolutionary cohort analysis (Rabosky, Donnellan et al., 2014) to look for differences in the macroevolutionary rate regime among the lineages of the *P. tentorius* complex. Distinct macro-evolutionary shift configurations were tabulated by BAMMtools based on BAMM results and posterior probabilities of each configuration. A phylorate plot was constructed through BAMMtools to visualize evolutionary rate variation across the Bayesian phylogenetic tree topology. Rate through time (RTT) plots were constructed using BAMM on all groups to visualize the radiation rate shift against temporal scales in different groups.

In addition, to compute the speciation rate (lambda) and extinction rate (mu) independently for each clade and crucial evolutionary group, R packages 'dplyr' (Wickham, & Wickham, 2019), 'readr' (Wickham et al., 2017) and 'ggplot2' (Wickham et al., 2016) were used to calculate lambda and mu for each target group and also to plot lambda and mu. The net diversification (r = lambda - mu) and turn-over rates (t = mu/lambda) were used as indicators to evaluate the radiation of the *P. tentorius* complex against temporal dimensions (Ricklefs, 2007), as well as predict future trends.

3.2.7 Character dependency analysis

A character dependency analysis was used to verify whether the cladogenesis and radiation of the *P. tentorius* complex was influenced statistically by geographic barriers and biomes. When considering the Orange River (OR) as a geographic barrier, I used the BiSSE approach (Maddison et al., 2007) instead of the GeoSSE approach (Goldberg et al., 2011) because there

is no clear evidence of population overlap across the OR. I performed the BiSSE analysis (6 parameters) with R packages 'ape' (Paradis, et al., 2004), 'diversitree' (FitzJohn, 2012) and 'phytools' (Revell, 2012) to investigate whether the OR significantly influenced the radiation history of the *P. tentorius* complex. To do this, I coded populations north of the Orange River as the "0" state, and populations south of Orange River as the "1" state. I then built likelihood functions into different scenarios: a) full model ((through function 'all.different', lambda1 ≠ lambda2, mu1 ≠ mu2, q01 (transition rate from character "0" to "1") ≠ q10 (transition rate from character "1" to "0")), b) lambda different model (through function 'free.lambda', lambda1 ≠ lambda2, mu1 = mu2 and q01 = q10), c) variable q model (through function 'free.q', lambda1 = lambda2, mu1 = mu2 and q01 ≠ q10). I used the ANOVA based likelihood ratio test (LRT) to determine the "best scenario" under criteria ln likelihood value and AIC criterion. The best-fit model was used to run a Bayesian MCMC analysis. First a preliminary run with 100 steps was done to estimate parameter settings. The formal long run was done with 100000 steps, sampling every 1000 steps.

I used the MuSSE approach (FitzJohn et al., 2009) implemented in R packages 'ape', 'diversitree', 'phytools' and 'geiger' (Harmon et al., 2014) to determine whether barriers like the GE and Swartberg Mountain (SM) significantly influenced the radiation of the *P. tentorius* complex. I partitioned the *P. tentorius* complex's distribution range into three geographic regions as "1" (above the GE), "2" (below the GE, but north of the SM including the western region below the GE) and "3" (south of the SM). In total, 12 parameters were included in the analysis. Different models were built, based on several scenarios as different likelihood functions: a) a "null" model (in function 'Minimal'), which regarded all lambda, mu and q to be equal across the different geographic regions (lambda1 = lambda2 = lambda3, mu1 = mu2 = mu3 and q12 = q21 = q13 = q31 = q23 = q32), b) a "full" model (in function 'all different'), which regarded all lambda, mu and q to be different (lambda1 ≠ lambda2 ≠ lambda3, mu1 ≠ mu2 ≠ mu3 and q12 ≠ q21 ≠ q13 ≠ q31 ≠ q23 ≠ q32), c) considered differences at lambda only (by function 'free.lambda', lambda1 ≠ lambda2 ≠ lambda3, mu1 = mu2 = mu3 and q12 = q21 = q13 = q31 = q23 = q32) d) considered differences at mu only (in function 'free.mu', mu1 ≠ mu2 ≠ mu3, lambda1 = lambda2 = lambda3 and q12 = q21 = q13 = q31 = q23 = q32), e) considered lambda & mu to vary, but that character transitions are orderly (through function 'free.lambda.mu', mu1 ≠ mu2 ≠ mu3, lambda1 ≠ lambda2 ≠ lambda3 and q12 = q21 = q13 = q31 = q23 = q32), f) lambda and mu are constant, but the transition process is fully flexible (through function 'free.q', lambda1 = lambda2 = lambda3, mu1 = mu2 = mu3 and q12 ≠ q21

≠ q13 ≠ q31 ≠ q23 ≠ q32). I then performed an ANOVA based LRT test to determine the most likely scenario. In addition, I used AIC weight criteria to select the best fitting model, which was used to run a Bayesian MCMC. Again, I made a preliminary run to determine prior settings. The formal long run was conducted with 100000 steps, sampling every 1000 steps.

The same method was used to model the interactions between cladogenesis and biome. To simplify biome analysis and interpretation, I excluded Albany Thicket and included only Nama Karoo, Succulent Karoo and Fynbos. Furthermore, I considered only Succulent Karoo as relevant in the Little Karoo because tent tortoises prefer this vegetation type in that region. The biome assessment included 1) Nama Karoo on the northern and southern sides of the GE, 2) Fynbos mixed with Succulent Karoo on the western side of the GE and 3) Succulent Karoo along the west coast and in the Little Karoo, and I encoded them as characters "1", "2" and "3", respectively. The rest of the steps followed the MuSSE analysis of the geographic regions.

3.3. Results

3.3.1 Calibration dating analyses

When comparing the four calibration dating results, the analysis with the FBD model generated the highest posterior likelihood value (Supporting information Figure S3.1). According to the tree topology test results of Zhao et al. (in press), the most likely topology regarding the placing of C6 was (C2+C3)+C6, rather than (C1+C4)+C7)+C5)+C6. I therefore selected the dating model that generated an identical tree topology as the best tree, which was the chronogram generated from the FBD model.

The best chronogram retrieved from the FBD model (Fig. 3.2) revealed that *P. tentorius* branched off from the sister group *P. geometricus* + *P. oculifer* at about 35 Ma (node n2), and that radiation of the *P. tentorius* complex (node n4) started in the early- to mid-Miocene, about 18 Ma, before the mid-Miocene Climatic Optimum (Fig. 3.2). Soon after the start of the radiation, at 17 Ma, C6 split off from C2+C3 (node n5). At about 10 Ma, C2 and C3 diverged (node n6) and more or less simultaneously, C5 segregated from (C1+C4)+C7 about 9.5 Ma

(node n7). The splitting of C7 from (C1+C4) occurred at about 6.4 Ma (node n8) whereas C1 and C4 diverged at about 4.7 Ma (node n9). The rest of the radiations within clades occurred within the Pleistocene epoch.

3.3.2 SAMOVA results and Habitat reconstruction

Since run K=7 gave the most sensible results in comparing the phylogenetic results of Chapter 2. The SAMOVA analysis retrieved seven geographic groups, identical to the distribution of the mtDNA clades, with strong support (fixation indices F_{SC} = 0.11, F_{ST} = 0.82 and F_{CT} = 0.8). In total, 80.04% of the total variance occurred among groups (Va: df = 6, sum of squares = 14555.33, variance = 114.24), 17.73% of the variance came from within populations (Vc: df = 112, sum of squares = 2834.23, variance = 25.31) and only 2.23% of the variance was from within groups (Vb: df = 46, sum of squares = 1583.72, variance = 3.18). Significance tests with 1023 permutations confirm this strong genetic structure against spatial dimensions (Vc and F_{ST}, $p < 0.0001$; Vb and F_{SC}, $p < 0.0001$ and Va and F_{CT}, $p < 0.0001$).

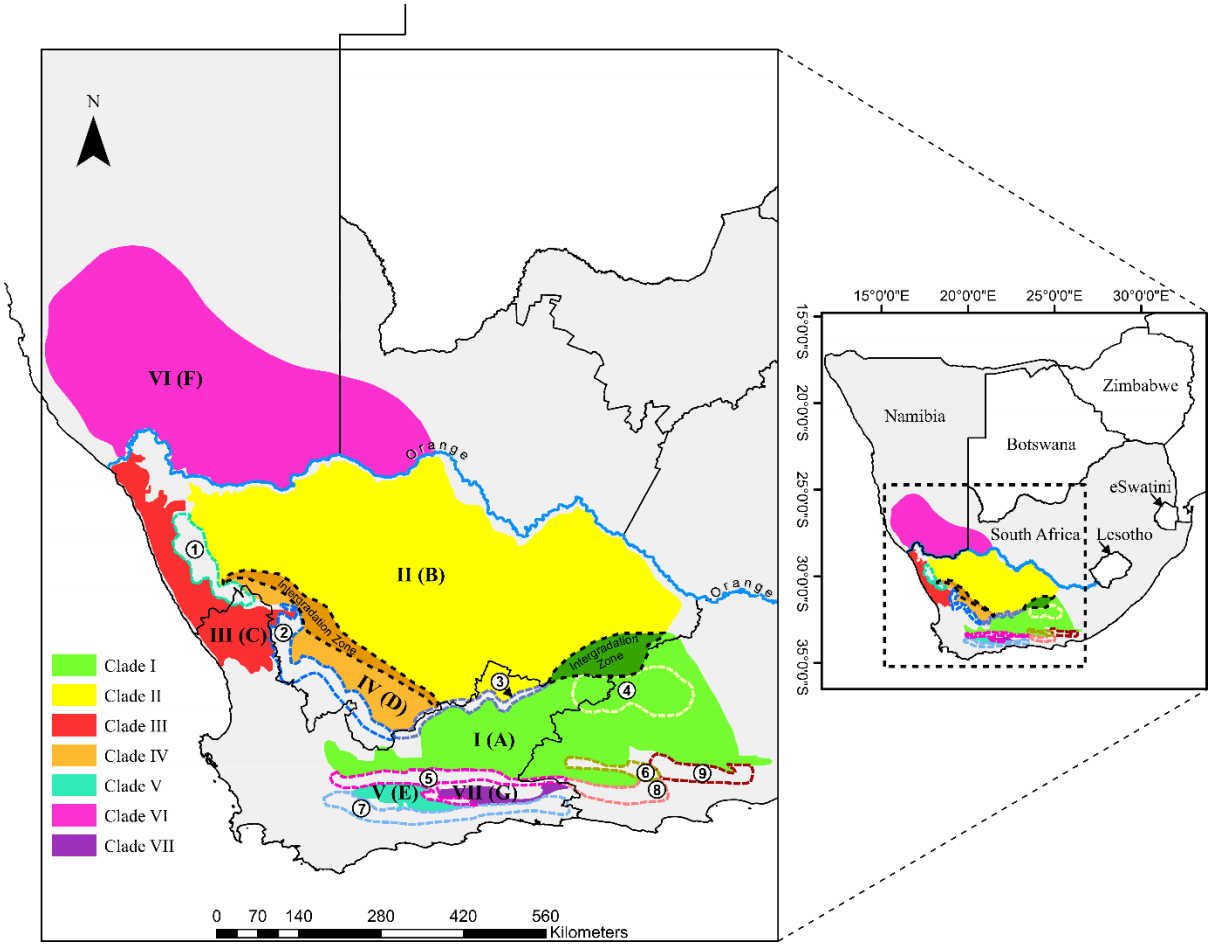


Figure 3.1. Map showing the geographic distribution of the seven mtDNA clades throughout their range. The Orange River between South Africa and Namibia is indicated. Major mountain barriers are indicated: 1) Kamiesberg, 2) Hantamberge and Roggeveldberge, 3) Nuweveldberge, 4) Sneeuberge and Kompasberg, 5) Swartberge and Rooiberg, 6) Grootrivierberge, 7) Langeberge, 8) Baviaanskloofberge, 9) Suurberge. The seven geographic regions retrieved from the SAMOVA analysis are indicated as A-G.

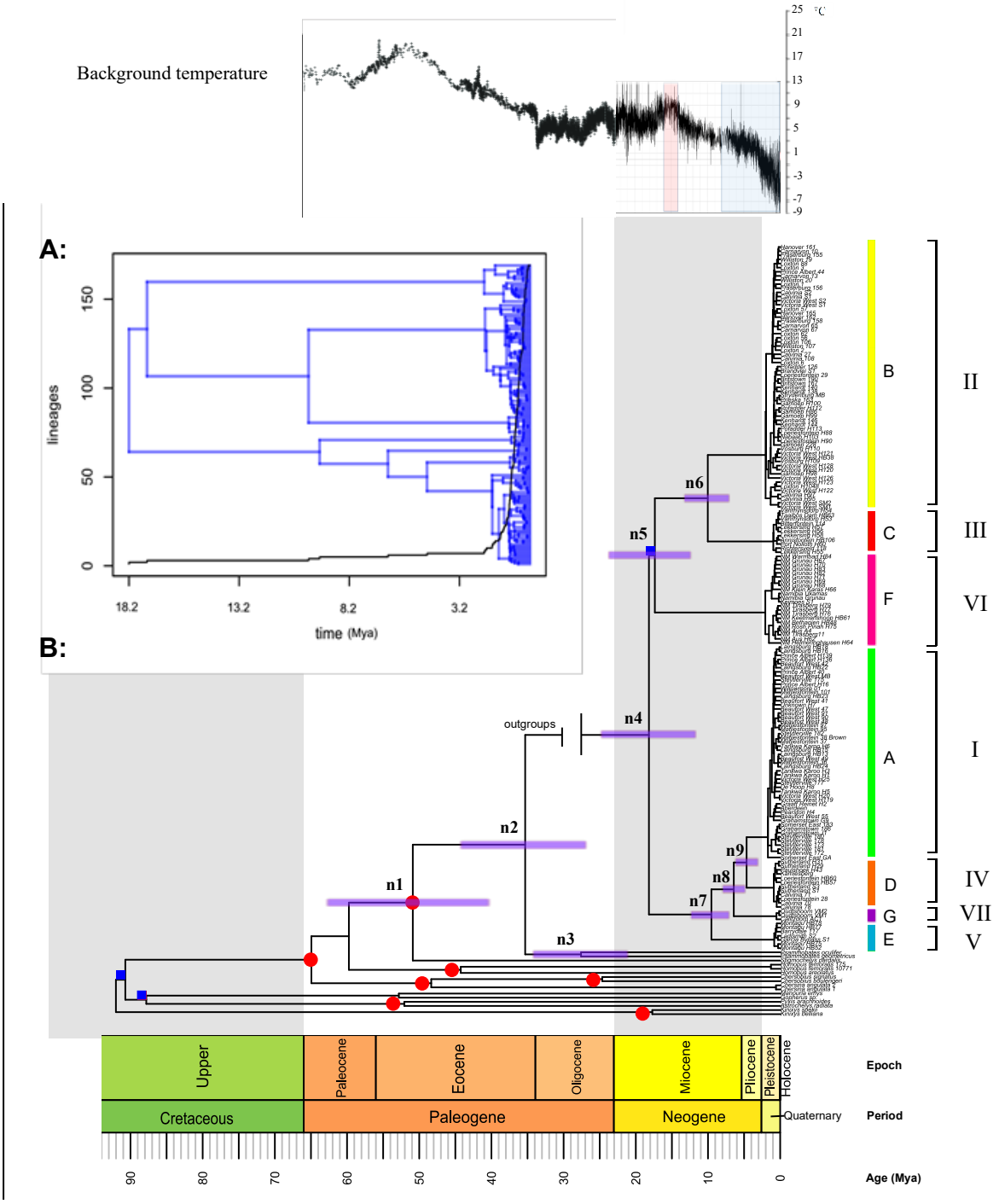


Figure 3.2. (a) The LTT plot displaying the cladogenesis and lineage changes through time in the *Psammobates tentorius* complex. (b) The chronogram generated from BEAST Bayesian calibration dating analysis with a background temperature fluctuation diagram (modified from Zachos et al., 2001 & Zachos et al., 2008). The interval of the mid-Miocene climatic optimum is highlighted in pink colour and the blue colour represents the re-establishment of major ice sheets due to cooling down. The red spots symbolize the seven calibration nodes from the literature used for the calibration dating analysis. The blue blocks indicate nodes with weak support (BP < 70% or PP < 0.95). The seven geographic regions defined by the SAMOVA analysis are indicated as "A" – "G". The distribution of the seven mtDNA clades are labelled as "I" – "VII".

The BioGeoBEARS model test suggested "DIVALIKE + J" as the best model with the highest AICc weight score (see Table S3.5) and a significant p-value in the LRT test ($p < 0.0001$). In general, the geographic region reconstruction results from the DIVALIKE + J model did not show significant incongruence with S-DEC and S-DIVA results from RASP (Table 3.1 & Fig. 3.3), particularly for significant biogeographic events. The majority of the variations between the findings of BioGeoBEARS, S-DEC and S-DIVA were the assumptions at the basal node (the split point between two major branches "(C1+C4)+C7)+C5" and "(C2+C3)+C6"). Nonetheless, the results from the three analyses in general imply a complex cladogenic history at node n4, since extinction, dispersal and vicariance events were all discovered by the DIVALIKE + J model. All the significant biogeographic events retrieved from the three approaches did not show significant incongruence, particularly for the cladogenic events after node n5 (the split between C6 and (C2+C3), see Fig. 3.3 & Table 3.1).

All three models suggested that a vicariant event caused C6 and its habitat to be isolated from that of C2+C3 (n5), although DIVALIKE+J and S-DEC models implied that dispersal, and S-DEC that extinction, also played a role in this cladogenic event. All three models also suggest that the split between C2 and C3 (n6) was caused by vicariance, with the S-DEC model also proposing a dispersal event. For C5 in region E, all three models advocated that the separation between C5 and (C1+C4)+C7) at n7 was caused by a vicariance event, with S-DEC and S-DIVA also suggesting a dispersal event. The separation of C7 and its region G (n8) is likely to have resulted only from vicariance, as suggested by both DIVALIKE+J and S-DIVA, although S-DEC found evidence for dispersal as well. The cause for the isolation between C1 and C4 (n9) were suggested as being vicariance, as all three models advocated a vicariance-only scenario. The BBM analysis supports region F as the ancestral habitat or radiation centre of the *P. tentorius* complex, whereas S-DEC predicted region DF as ancestral region (Table 3.1).

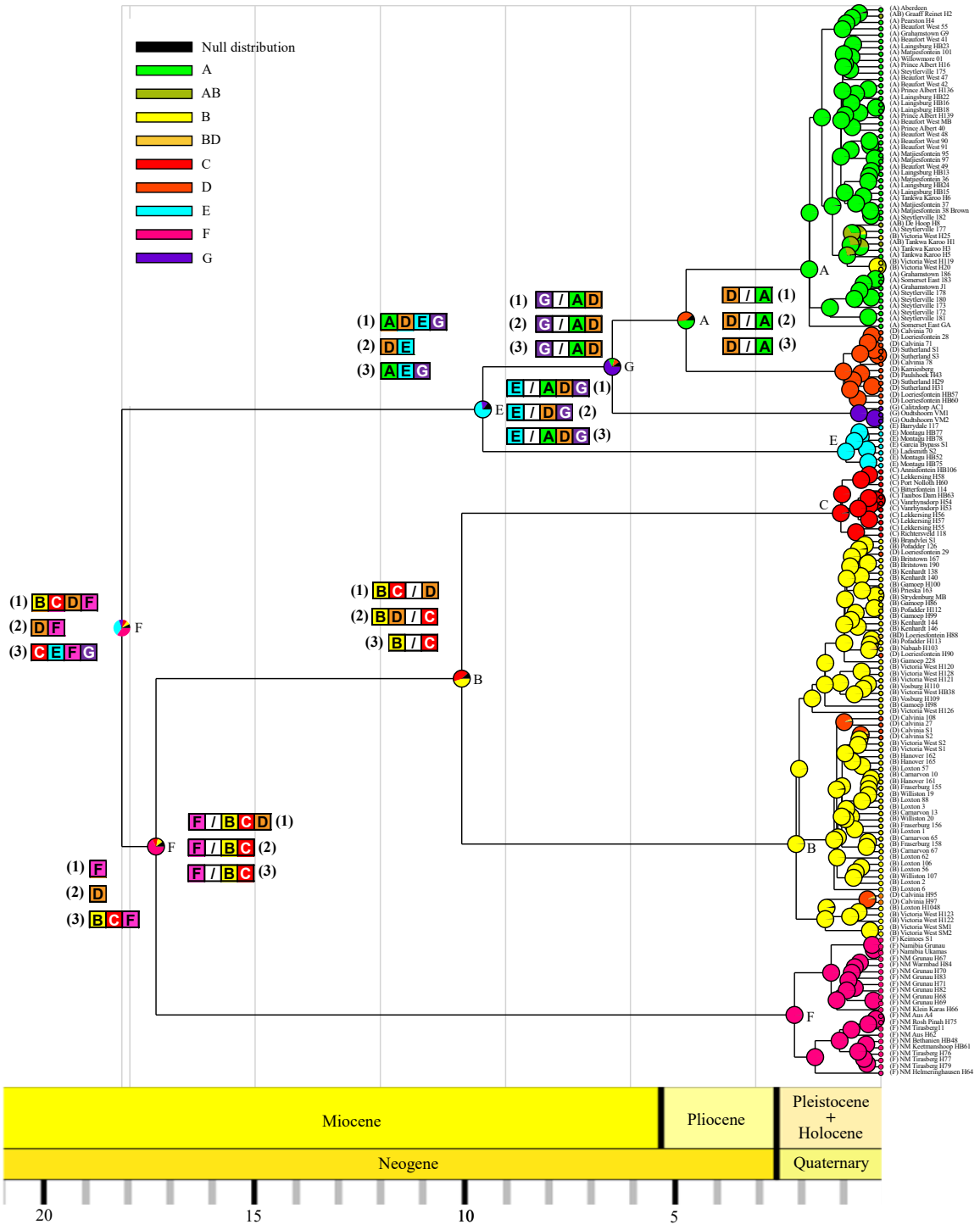


Figure 3.3. Habitat reconstruction analysis results: (1) The results generated from the best model DIVALIKE + J using the BioGeoBEAR package, (2) S-DEC analysis results, (3) D-DIVA habitat reconstruction results. The pie at the centre of each node was retrieved from the BBM analysis. The most likely ancestral habitat is represented by regions A-G. The colour scheme in the tree is the same as the colour scheme used elsewhere. The unit for the timeline was Mya.

Table 3.1. BioGeoBEAR and RASP results showing significant dispersal and vicariance events. RASP habitat reconstruction history is shown, using different phylogeographic analytical methods for the seven geographic regions. It also shows the estimated potential centres of origin using S-DEC and BBM based methods for all three area scenarios.

	BioGeoBEARS (DIVALIKE + J)				S-DEC			
Node	Extinction	Dispersal	Vicariance	RASP route	Extinction	Dispersal	Vicariance	RASP route
n4	Yes	Yes	Yes	BCDF->FD->FADEG->F/ADEG	Yes	Yes	No	DF->D->D^D->DE^D->D/DE
n5	No	Yes	Yes	F->FBCD->F/BCD	Yes	Yes	Yes	D->->FBC->F/BC
n6	No	No	Yes	BCD->BD/C	No	Yes	Yes	BC->BDC->BD/C
n7	No	No	Yes	ADEG->E/ADG	No	Yes	Yes	DE->EDG->E/DG
n8	No	No	Yes	ADG->G/AD	No	Yes	Yes	DG->GAD->G/AD
n9	No	No	Yes	AD->D/A	No	No	Yes	AD->D/A

	S-DIVA				BBM			
Node	Extinction	Dispersal	Vicariance	RASP route	Extinction	Dispersal	Vicariance	RASP route
n4	No	Yes	Yes	CEFG->BCFAEG->BCF/AEG	No	Yes	Yes	F->FE->F/E
n5	No	No	Yes	BCF->F/BC	No	Yes	Yes	F->FB->F/B
n6	No	No	Yes	BC->B/C	No	Yes	Yes	B->BC->B/C
n7	No	Yes	Yes	AEG->EADG->E/ADG	No	Yes	Yes	E->EG->E/G
n8	No	No	Yes	ADG->G/AD	No	Yes	Yes	G->GA->G/A
n9	No	No	Yes	AD->D/A	No	Yes	Yes	A->DA->D/A

Ancestral area prediction	
Model	**Centre**
BBM	F
S-DEC	DF

3.3.3 Diversification rate analysis

From the macro-evolutionary cohort matrices (Fig. 3.4), five comparatively distinct rate regimes seem to have occurred in the *P. tentorius* complex, 1) between C1 and C6, 2) between C1 and C2, 3) between C1 and C3, C5 and C7, 4) between C2 and C4, C7, C5 and C3, 5) between C2 and C6. The highest rate variation was found between C2 and C1+C3+C4+C5+C6. The diversification rate at C2 and C7 was lower than that of the other clades.

The phylorate plot showed two clear evolutionary rate shifts, one at C1 and the other at C2 (Fig. 3.5a). The RTT plot for the entire *P. tentorius* complex showed a trend of increasing diversification rates since its radiation started about 18.2 Ma, levelling off only in recent time (about 2 Ma, Fig. 3.5b). The RTT plots at clade level (Fig. 3.5c-i) revealed a trend of increasing diversification rates at C3, C4, C5 and C6, but especially at C6 (Fig. 3.5c). A clear decreasing

trend was found at C2, and the rate plot for C1 and C7 showed a constant trend. The RTT plot for the "(C2+C3)+C6" branch shows a clear decreasing rate shift at about 2 Ma (Fig. 3.5j), whilst, the RTT plot for branch "(C1+C4)+C7)+C5" revealed an increasing rate shift at the same time (Fig. 3.5k).

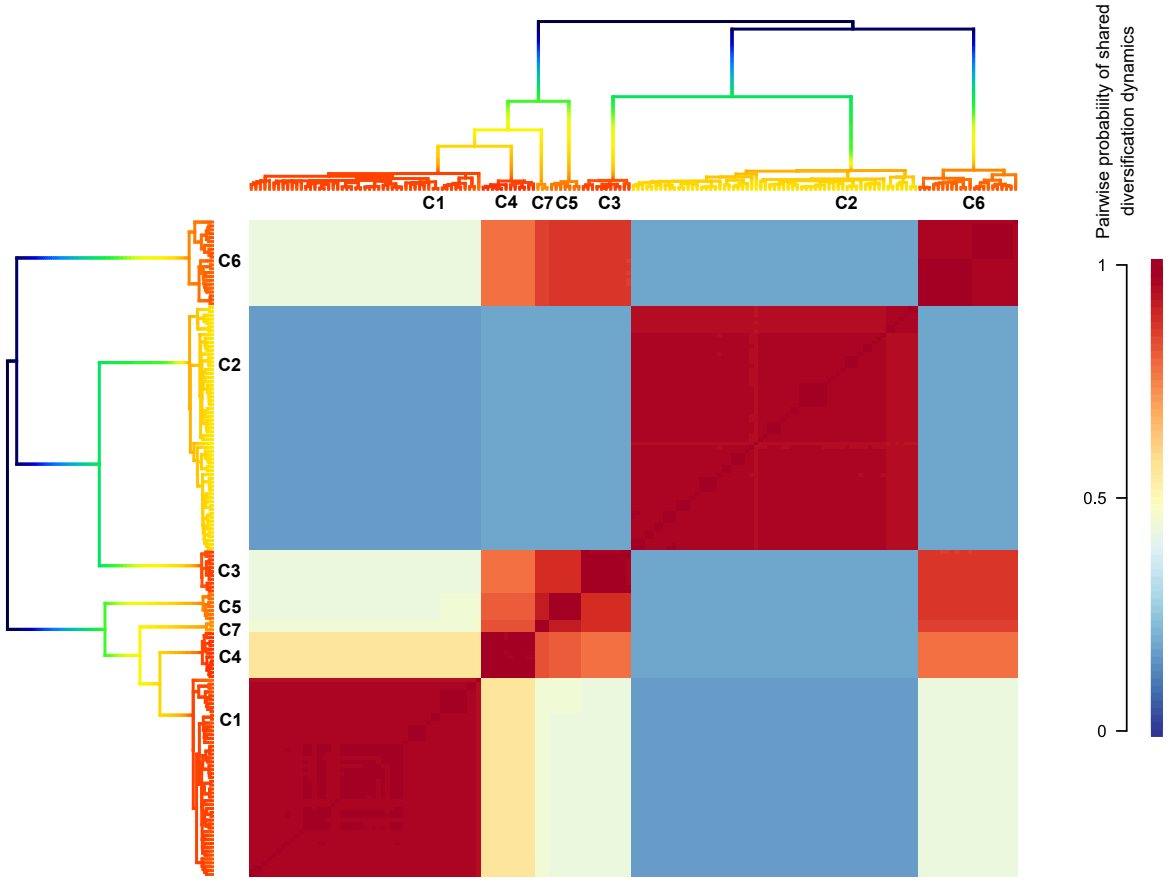


Figure 3.4. Macroevolution cohort matrix for the seven clades of the *P. tentorius* complex. BAMM Bayesian diversification rate analysis based on the mean phylorate plot trees, are shown at the top and on the left side of the cohort matrix, for purposes of comparison. The matrix shows pairwise probabilities of two groups sharing the same evolutionary dynamics. The "warm" colours represent high cohort similarities (highest value "1" refers to 100% similarity), whilst, the "cool" colours represent low cohort similarities (lowest value "0" refers to 0% similarity).

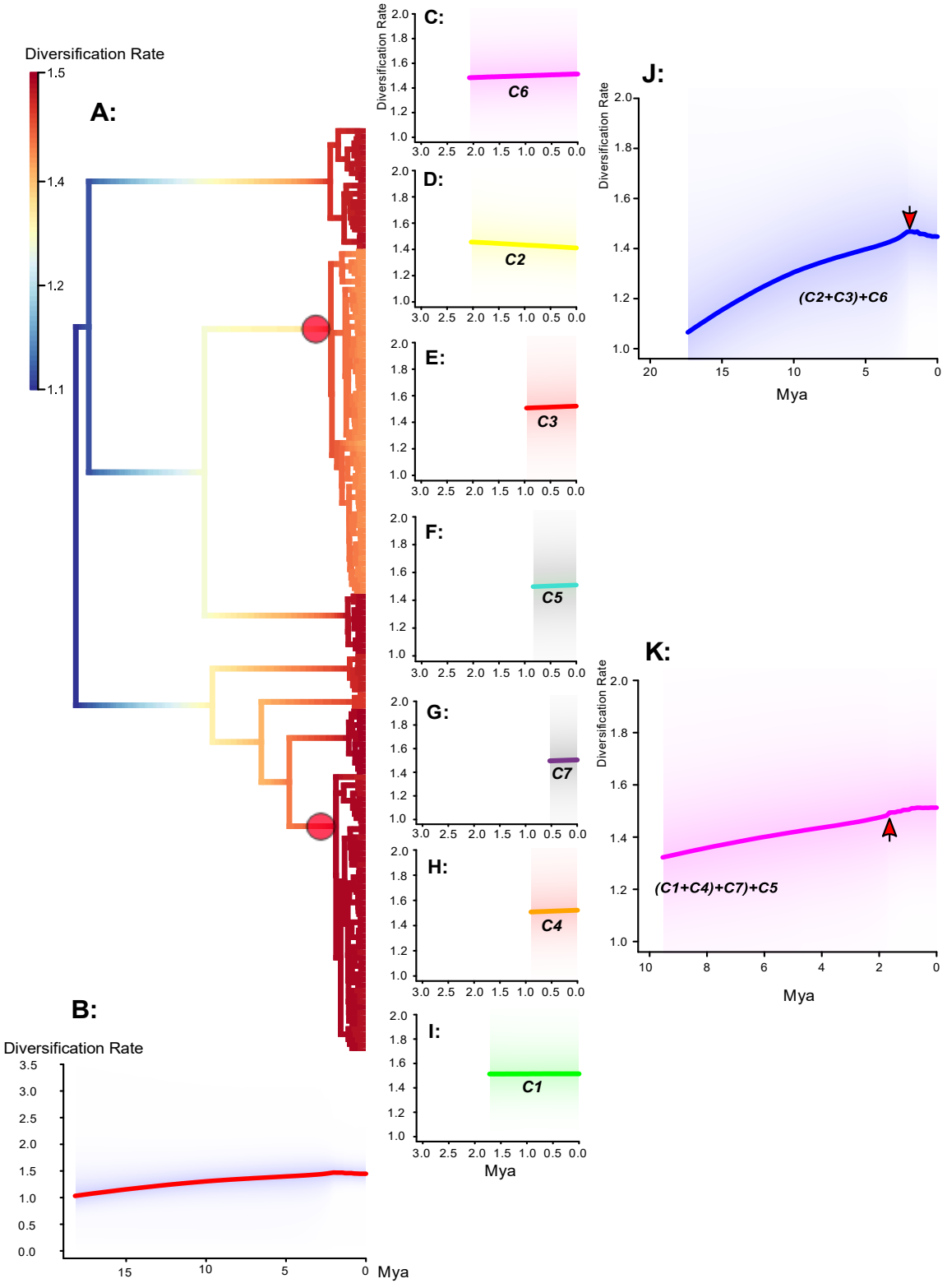


Figure 3.5. (a) The phylorate plot shows the diversification rate across the BEAST chronogram topology. The "warm" colours represent fast rates, whilst "cool" colours represent the branches with slow rates (units per event or lineages per million years). The red spots indicate corresponding nodes with significant rate shifts. (b) The rate though time (RTT) plot for the overall *Psammobates tentorius* complex. (c)-(k) The RTT plots for different groups. The upward arrow indicates an increasing rate shift, and the downward arrow represent a decreasing rate shift.

Macro-evolutionary rate shift configurations at a 95% confidence interval (Figure S3.2) suggested that the most likely scenario was that 24% of the posteriors samples (f = 0.24) can be assigned to two major shift configurations, at C1 and C2. Nine possible rate shift scenarios were detected by BAMM. The first three scenarios occupied 65% of the total posteriors samples (scenario 1: 24% of the samples from posteriors (f = 0.24) supported two rate shift events, at C1 and C2; scenario 2: 23% of the samples from posteriors (f = 0.23) supported only one rate shift event, at C2; Scenario 3: only 18% of samples from posteriors (f = 0.18) suggested no rate shift event across the tree topology). Under these configuration scenarios, the diversification rate within C2 was apparently lower than that of the rest of the clades.

The BAMM results (Table 3.2) revealed C1 and C2 as having substantially higher net diversification rates and lower turn-over rates than the rest of the clades, though C2 exhibited the lowest speciation rate. All clades and groups had higher speciation rates than extinction rates. The overall lineage turn-over rates were relatively high in clades C3, C5-C7 and the two major branches "(C1+C4)+C7)+C5" & "(C2+C3)+C6". The entire complex generally had a low overall net diversification rate and a relatively high lineage turn-over rate The plots of lambda and mu retrieved from the BAMM analysis results (Figures S3.3-3.5) revealed a clear increasing diversification trend at C3 (increasing lambda, relatively constant mu), C4, C5, and C7 (increasing lambda, nearly constant mu), but especially at C6 (increasing lambda, decreasing mu), since the slope of linear-ship between lambda and the time line in its case was significantly steeper than that in the other clades. By contrast, both lambda and mu plots of C1 tended to be flatter, with low levels of fluctuation. Clade 2 showed a remarkable decline in its lambda plot, and a slight increasing trend in its mu, therefore indicating a decreasing trend in net diversification. The lambda and mu plots of the entire complex showed three visible rate shifts, the first one at about 10 Ma, the second one at about 7-8 Ma and the third one at around 5 Ma. Two increasing rate shifts were detected at branch (C1+C4)+C7)+C5, the first at about 5 Ma and the second at around 2 Ma (Figure S3.5). Only one decreasing rate shift was found around 2 Ma at branch (C2+C3)+C6 (Figure S3.5).

Table 3.2. The estimated speciation rate (lambda), extinction rate (mu), net diversification rate (r) and lineage turn-over rate (t) generated from MuSSE analysis (for modelling the effect of geographic regions and biomes), BiSSE (for modelling the effect of the Orange River) and BAMM analysis (for computing the overall diversification pattern for each target group). The unit used in the table was uniform for the cladogenesis events (or lineage events) per Myr.

		Speciation rate	Extinction rate	† Net diversification rate	‡ Lineage turn-over rate
Region (MuSSE)	1. Above GE	2.097	0.164	1.933	0.078
	2. Below GE, north of Swartberg	2.011	1.037	0.974	0.516
	3. South of Swartberg	0.988	0.967	0.021	0.979
Biome (MuSSE)	1. Nama Karoo	2.041	0.2	1.841	0.098
	2. Fynbos Succulent Karoo mix	6.465	6.674	-0.209	1.032
	3. Succulent Karoo	0.852	0.536	0.316	0.629
Orange River (BiSSE)	1. North of Orange River	0.779	0.703	0.076	0.902
	2. South of Orange River	1.323	0.133	1.19	0.100
Clade (BAMM)	C1	1.528	0.678	0.85	0.444
	C2	1.445	0.334	1.111	0.231
	C3	1.526	1.271	0.255	0.833
	C4	1.534	1.162	0.372	0.757
	C5	1.5	1.286	0.214	0.857
	C6	1.51	1.239	0.271	0.820
	C7	1.486	1.276	0.21	0.859
	(C1+C4)+C7)+C5	1.425	1.175	0.25	0.825
	(C2+C3)+C6	1.306	1.177	0.129	0.901
Overall	entire complex	1.308	1.199	0.109	0.917

Note: † Net diversification rate (r) = Speciation rate (lambda) - Extinction rate (mu)
‡ Lineage turn-over rate = mu / lambda

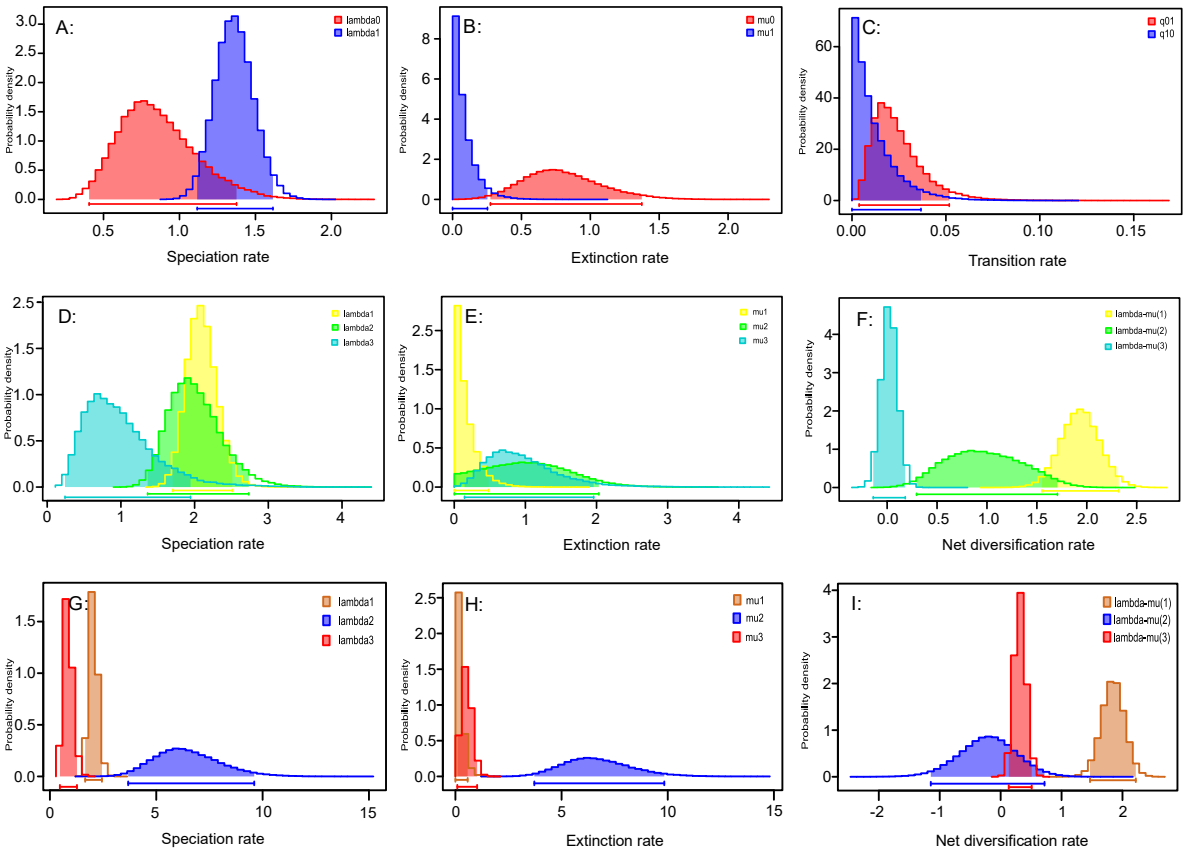


Figure 3.6. The estimated lambda, mu and q (a - c) from the BiSSE analysis: "0"- populations from north of the Orange River, "1" - populations from south of the Orange River; d-f, MuSSE analysis results of three geographic regions: "1" – region above the GE, "2" – region south of the GE, and north of the SM and "3" – region south of the SM. Plots show lambda, mu and net diversification rates (r) across the three different regions; g-i, MuSSE analysis modelling the lambda, mu and r of three biomes: "1" – Nama Karoo, "2" – Fynbos mixed with Succulent Karoo, "3" –Succulent Karoo. Units were "events/million years" in all cases.

3.3.4 Dependent character analysis

The MuSSE analysis suggested that the GE and SM geographic barriers may have influenced the historical radiation of the *P. tentorius* complex substantially, and identified the "full model" (lambda, mu and q all differ across the three regions), with strong statistical support, as the best scenario (model likelihood scores and criteria provided as Table S3.6). The biome based MuSSE analysis statistically confirmed that biomes affected radiation significantly, since the "full model" was favoured according to the LRT test (Table S3.6). In contrast, results of the BiSSE analysis indicated that the OR did not influence radiation in *P. tentorius*, because results for the best model, "free lambda", were not statistically significant (Table S3.6).

The BiSSE analysis results revealed a much lower net diversification rate, and higher lineage turn-over rate, in populations north of the OR than in ones south of the it (Fig. 3.6 & Table 3.2). MuSSE analysis of regions indicated low net diversification rates and high lineage turn-over rates in populations south of the SM, and in populations occupying a mixture of Fynbos and Succulent Karoo vegetation (Fig. 3.6 & Table 3.2).

3.4 Discussion

3.4.1 Radiation patterns and biogeography

Cladogenesis in *P. tentorius* started during the early to mid-Miocene, a period initially characterised by high temperatures (Zachos et al., 2001), which may have facilitated dispersal of ancestral *P. tentorius* over a relatively wide region of southern Namibia and northern South Africa. Ancestral area analyses showed that either region F (north of the OR in Namibia) or region DF (north of the OR and on the western GE of South Africa) were the most likely regions of origin. The presence of a Miocene fossil (19 Ma), identified as either *Psammobates* or *Homopus*, at Arrisdrift on the northern bank of the OR (de Broin, 2003), supports the potential presence of *Psammobates* north of the OR. Nevertheless, it is not possible to decide which of the two proposed areas of origin is the most feasible, particularly since results of RASP routes for different analyses are contradictory for the first few divergences. As DIVALIKE+J received strongest support, the interpretation will focus mostly on results from that analysis.

The absence of fossil records in the biometric analysis may have compromised the accuracy of estimates; consequently, the divergence dates should be treated as ranges rather than means. The timespan when the *P. tentorius* ancestor diverged from other *Psammobates* covered the late Eocene and early Oligocene, when temperatures dropped and the Antarctic ice-sheet formed (Feakins & Demenocal, 2010; Zachos et al., 2001). In addition, the opening of the Drake Passage caused cold water (Benguela Current) to flow northwards along western South Africa (Feakins & Demenocal, 2010) and the southern part of Africa became progressively more arid (Burke & Gunnell, 2008). This cooling and gradual aridification possibly contributed to diversification and the separation of ancestral species into distinct geographical regions. Since the biomes with which these species are associated today did not exist at the time of divergence (Neumann & Bamford, 2015), it is not possible to speculate in which geographic

region each taxon developed. Nevertheless, the fact that *P. geometricus* is restricted to a small region in southwestern South Africa, and *P. tentorius* fills the geographic space between *P. geometricus* in the southwest and *P. oculifer* in the north, suggests that the ancestor of the three species possibly had a wide distribution across southern Africa during the Eocene, and that global cooling in the early Oligocene likely restricted their ranges and isolated populations.

Ecologically, *P. tentorius* only occurs in arid semi-desert areas, never approaching water bodies and cannot swim at all (Branch, 2008; Z. Zhao pers obs). Besides that, *P. tentorius* does not occur on slopes or in mountainous areas, but only in flat areas (Z. Zhao pers obs.), usually with a lose sandy base (Branch, 2008). The reasons for it not occurring in mountainous or sloped areas is possibly due to its morphology. The highly domed carapace (particularly in females) may result in challenges with respect to balance, limiting its ability to climb mountains, or handle slopes. Mountains and rivers are presumably therefore significant barriers for the movement of *P. tentorius*.

Divergence at node 4 divided the ancestral species into two lineages, consisting respectively of the ancestors of the current northernmost (C2+C3+C6) and southernmost (C1+C4+C5+C7) clades. The DIVALIKE+J RASP route indicated that the range of an ancestral group first became restricted to regions F and D, from where a southwards dispersal occurred to region ADEG. The final step was a divergence between F and ADEG, indicating a geographic split between the northern and southern lineages. From this scenario it would appear that ancestral *P. tentorius* crossed the OR despite it representing a potential geographic barrier. The analysis suggested that vicariance, dispersal and extinction played a role in this first divergence. Soon after the *P. tentorius* ancestors divided into two lineages, C6 diverged from C2+C3 at node 5. The DIVALIKE+J RASP route indicated a second crossing of the OR from region F, followed by dispersal into regions BCD on the southern side of the OR in South Africa, and finally a divergence between populations north and south of the OR. This divergence apparently involved both dispersal and vicariance.

The timing of divergence at nodes 4 and 5 does not correspond to any major geological event, although minor uplifts in the early- to mid-Miocene may have influenced the topography of the western GE, and the flow of the OR. Several studies using different methodologies indicate that major uplift events in southern Africa occurred during the Cretaceous and that tectonic uplift during the Cenozoic was at a relatively small scale (Braun et al., 2014; Jacobs, 2005;

Wildman et al., 2015; Wildman et al., 2017). Yet, uplifts of up to 250 m during the early- to mid-Miocene (Dauteuil et al., 2015; Rudge et al., 2015) may have influenced distribution routes of *P. tentorius*. Apart from possible geological changes, rapid cooling after the mid-Miocene Climatic Optimum (15-17 Ma; Zachos et al. 2001) likely contributed to cladogenesis at nodes 4 and 5. I propose that dispersal at node 4 involved movement from the northern into the southern region of South Africa, and that extinction may have been associated with the initial restriction of the range in northern South Africa, and subsequent extinction of populations in central South Africa when temperatures decreased. The character dependency analyses suggested that the OR did not play a role in speciation events, consequently vicariance pertains to climate differences rather than the OR. Equally for node 5, I propose that dispersal relates to *P. tentorius* crossing the OR a second time and dispersed into the northern regions of South Africa, whereas vicariance relates to populations retracting into refugial areas in response to a steep reduction in temperature after the mid-Miocene Climatic Optimum (Zachos et al., 2001). The western GE in Namibia and South Africa may have served as both refugia and corridors for diversification in *P. tentorius*, as it has for other taxa (Clark, 2010). More research is necessary to see if members of the clades still cross the OR on occasion and perhaps hybridize.

There are few reptile taxa with similar distribution patterns as *P. tentorius* in southern Africa, for which divergence dates have been estimated. Nevertheless, Bauer & Lamb (2005) proposed that the deep divergences among major clades of the *Pachydactylus* group date back to the Miocene, and implicated vicariance associated with both geological and climatic events, including uplift of the western GE (see Rudge et al. 2015) and the western north-to-south aridity gradient, which is believed to have had its inception in the early Miocene (Roberts et al. 2013). Of interest is that the three *Chersobius* tortoise species, endemic to southern Africa, diverged within the same time-frame (18-14 Ma) as the first divergences in *P. tentorius*, and show a similar distribution pattern; *Chersobius solus* occurs only north of the OR in Namibia, *C. signatus* is restricted to the western, winter-rainfall region south of the OR, whereas the range of *C. boulengeri* falls mostly in the central Karoo where the species is exposed to summer rains (Hofmeyr et al., 2014; Hofmeyr et al. 2017).

The next two divergences at nodes 6 and 7 occurred more or less at the same time in the late Miocene, with both divergences being ascribed to vicariance, although some methods also implied that dispersal was involved. The divergence at node 6 involved a separation of clade 3 on the western coastal plains (region C) from clade 2 above the GE (region BD). The most

likely explanation for this cladogenic event is increasing western aridity due to intensification of the cold Benguela Upwelling System approximately 10 Ma (Hoetzel et al., 2015; Jung et al., 2014; Rommerskirchen et al., 2011). In parallel with western aridification is the initiation of a winter rainfall regime and the development of Succulent Karoo vegetation in the west during the late Miocene, which came to full development in the Pliocene (Hoffmann et al., 2015; Neumann & Bamford, 2015; Verboom et al., 2009). Vicariance due to climate seems more likely than the western GE developing into a physical barrier between the coastal plains and interior plateau. The GE undoubtedly would have restricted gene exchange, but there are low-lying regions, particularly in the northwest, which could have served as dispersal routes between the plains and plateau, thus between C3 and C2. Nevertheless, there is evidence for a minor uplift in the west in the early Pliocene (Rudge et al., 2015), which possibly strengthened isolation between the two clades.

Psammobates tentorius dispersed to the southern region of South Africa (ADEG) by the early to mid-Miocene, but this southern group only started diverging from the late Miocene to early Pliocene. The first divergence at node 7 isolated clade 5 in region E (the western Little Karoo) from the remaining *P. t. tentorius*. Thereafter, clade 7 in the eastern Little Karoo region G (Oudtshoorn basin) diverged from *P. t. tentorius* clades 1 and 4, in regions A and D. In both instances, vicariance seems to be of primary importance although some models also implicated dispersal. These divergences may be due either to climate change or Cenozoic exhumation. Since climatic and vegetational changes were not restricted to the Little Karoo, it seems unlikely that these changes could have played a major role in vicariance.

Evidence for the effects of geomorphological changes on vicariance of the southern clades seems strong. Thermochronology results indicate that regions west and east of Worcester in the southwestern Cape underwent burial (up to 1.2 km) prior to exhumation (starting 30-20 Ma), which created the current relief with surrounding peaks reaching elevations of 1500 m (Green et al., 2017). It seems likely that over time this exhumation isolated *P. t. tentorius* populations in the Little Karoo from their western and northern counterparts. In addition, thermochronology results indicated differential uplift followed by exhumation (starting 30-20 Ma) in mountain ranges of the Oudtshoorn basin (Green et al., 2017). These events could have isolated clade 7 in region E. Despite being on a more recent time scale, divergence within a Little Karoo endemic plant species, *Berkheya cuneata*, shows two distinct lineages associated respectively with the western and eastern Little Karoo (Potts et al., 2013), similar to clades 5

and 7 of *P. t. tentorius*. Pygmy geckos (*Goggia*) also show a similar differentiation pattern in the southern Cape (Heinicke et al., 2017). On a broader scale, Cowling et al. (2009) proposed that increased topo-edaphic heterogeneity in response to moderate uplift in the early and late Miocene played a major role, together with climatic deterioration, in the rapid diversification of the Cape plant clades.

The final divergence between clades 4 and 1, respectively on the western GE (region D) and below the southern GE (region A), occurred during the Pliocene with vicariance as the main driver, possibly due to both tectonic and climatic events. The late Miocene and early Pliocene (7.7-4.0 Ma) was relatively humid (Diekmann et al., 2003), which may have facilitated movements and gene exchange between populations below and above the southern escarpment. A possible contact zone may have been the relatively low slopes of the GE in the northern Tankwa Karoo, southwest of Calvinia. The Tankwa Karoo is currently one of the most arid sections of the Karoo with <100 mm rain per annum (Rubin, 1998). Aridification may thus have closed this contact zone and isolated C1 and C4. Green et al. 2017 provide evidence for a Cenozoic uplift and subsequent denudation of around 700 m in the Beaufort West GE region, but there is no evidence suggesting that this tectonic event was localised or affected the remaining southern escarpment. Nevertheless, if such an uplift influenced dispersal routes across the GE, it should have contributed to this diversification event.

The phylogeographic analyses detected several Pleistocene radiations within C1, C2, C4 and C6, while the lack of fine structure in C3, C5 and C7 may be due to small sample sizes. Periodic climatic cycles in response to Milankovitch forces caused glaciers to wax and wane over the Pliocene and Pleistocene (Fedorov et al., 2006), which influenced species distribution and diversification worldwide (Hewitt, 2000). Radiations within C2 and C6 were possibly driven by the southward expansion of Kalahari sands in response to Pleistocene glaciation (Partridge & Maud, 1987), as has been shown for the Namaqua rock rat (*Micaelamys namaquensis*) (Russo et al., 2010) and Chacma baboons (*Papio ursinus*) *sensu lato* (Sithaldeen et al., 2015).

3.4.2 Diversification rates and regional patterns

The radiation in *P. tentorius* was not influenced by the OR, but differed among regions and among biomes. The latter results, however, became more complex when considering the association between clades and regions or biomes. Lineage turnover rate was low for the region above the GE, which is paralleled by a low turnover rate for C2, but not for C4 or C6. Lineage

turnover rate was low for the Nama Karoo, as well as for C1 and C2 associated with this biome, but the turnover rate for C6 in Nama Karoo north of the OR was high. The region and biome analyses were not very informative possibly because the division of characters oversimplified the great climatic and topographic heterogeneity of the southern African landscape (Cowling et al., 2017; Partridge et al., 2010). In addition, current biomes developed only recently (late-Miocene to Pliocene; Neumann & Bamford, 2015) over the long cladogenic timespan of *P. tentorius*.

Net diversification rate for the entire *P. tentorius* complex was low but lineage turnover rate was high. Low net diversification rates and high lineage turnover rates applied to all clades, except C1 and C2, occurring in the Nama Karoo in central South Africa. The high net diversification rates of C1 and C2 implied that these clades had accumulated much diversity over a relatively short period of time, which was due to a lower extinction rate rather than a higher speciation rate than in the other clades. Yet, the evolutionary dynamics of C1 and C2 differed, with significant rate shifts in both clades associated with an increase in rate for C1 and a decrease in rate for C2. Species distribution models for *Bitis arietans* indicated that the interior of South Africa became inhospitable for this species during glaciation periods, and that populations continuously retracted and expanded to and from refugia in the northern and southern regions of South Africa (Barlow et al., 2013). I propose that C2 retracted northward during glaciation periods and that the relatively low landscape heterogeneity of their habitat (Bushmanland and Upper Karoo bioregions) reduced rate shifts in it. In contrast, C1 possibly retracted to southern regions near the CFMs in the south, which include several vegetation types such as the Lower Karoo bioregion, Albany Thicket and Fynbos. The higher topographic and vegetation heterogeneity of the southern region may explain the increase in rate shift for C1 in contrast with a decrease in rate shift for C2.

Evolutionary dynamics were similar for C3, C5, and C7 below the GE and for C4 and C6 above the GE. The two biodiversity hotspots of the GCFR cover the full distribution ranges of four *P. tentorius* clades (C3, C4, C5 and C7), and part of the ranges of C1 and C6 (see Fig. 1 in Snijman, 2013). Topographic diversity in the habitats of C4, C5, C6, and C7 is high but it is low for C3 except in the north, in the Richtersveld, which includes the arid western escarpment in South Africa (Clark, 2010). Associated with this topographic diversity is high climatic diversity, with a strong aridity gradient from north (very arid) to south (more mesic) along the western, winter-rainfall region of South Africa (Grab & Knight, 2015). The southern region is

arid south and north of the Swartberg Mountains, but becomes more mesic towards the east, and topographic heterogeneity creates local climates (Partridge et al., 2010) over the range of the southern clades. North of the OR, the habitat of C6 is also diverse not only with regard to topography but also to climate. The southwest is highly arid and receives winter rains, whereas rainfall increases towards the east, which receives summer rains (Irish, 2008). The topographic and climatic heterogeneity of the abovementioned regions probably explain their similarities in evolutionary dynamics, characterised by high speciation rates, countered by high extinction rates, culminating in low net diversification and high lineage turnover rates.

The GCFR is recognised for its great floristic richness and endemism (Born et al., 2007; Cowling et al., 2017), with much of its present diversity ascribed to radiation events during the Pliocene and Pleistocene (Tolley et al., 2014). The Pliocene-Pleistocene glacial cycles involved fluctuations between cool-arid conditions during glacial periods and warm-humid conditions during interglacial phases (deMenocal, 2004; Zachos et al., 2001), causing the contraction and expansion of taxa in and out of refugial areas. Several studies implied the existence of western and eastern refugia in the southern GCFR (Barlow et al., 2013; Vamberger et al., 2018) as well as refugial areas in mountains of the western GCFR, such as the Cederberg and Kamiesberg (Clark, 2010; Daniels et al., 2009). It appears that large patches of Fynbos and Succulent Karoo biomes persisted as refugia over glacial-inter-glacial cycles, which was not the case with the Nama Karoo; recent expansion of the Nama Karoo may thus explain present contact zones of many taxa (Tolley et al., 2014 and references therein). Such contact zones for *P. tentorius* include intergradation zones between C1 and C2, and between C2 and C4, which in both instances may have involved recent expansions of Nama Karoo clades.

3.5 Conclusions

Divergences in *P. tentorius* are deep, with most clades originating in the Miocene. Among southern African reptiles, Miocene divergences usually correspond to cladogenesis within genera, e.g., *Pachydactylus* (Bauer & Lamb, 2005) and *Platysaurus* (Scott et al., 2004), whereas radiation within species is linked to the Pliocene and Pleistocene, e.g., *Agama atra* (Matthee, & Flemming, 2002) and *Bitis arietans* (Barlow et al., 2013). Some clades of *P. tentorius* may therefore need elevation to full species status.

The study provided strong evidence that cladogenesis in *P. tentorius* can be linked to climatic fluctuations and topographic changes in southern Africa since the Miocene, thus supporting the first hypothesis. It appears if climate was of greater importance than purported topographic changes in earlier diversification events, but that uplift events together with climate change played a significant role in later divergences. The climatic and topographic changes linked here to early divergences in *P. tentorius* were also proposed for vicariance in other reptiles at the generic level.

The results also correspond to other studies showing high genetic diversity (species richness in terms of the number of clades occurs there) in the GCFR, not only for plants but also for animals, including reptiles. Consequently, diversification patterns of *P. tentorius* in the late Miocene and Pliocene parallels those of other organisms, supporting the second hypothesis of higher diversity in the GCFR than elsewhere over the distribution range of *P. tentorius*. The strong association of *P. tentorius* clades with particular regions and vegetation types suggests that the clades evolved allopatrically and that contact in restricted areas is recent, following range expansions of some clades. However, although the clades abut, they do not necessarily overlap because vegetation in the regions regarded as possible intergradation zones forms a mosaic, which may still keep clades distinct. Nevertheless, more research is necessary to establish if the clades hybridize in these so-called intergradation zones.

This Chapter in parallel with finding in Chapter 2 provides strong evidence that *P. tentorius* requires taxonomic revision, which would impact the Red List Assessment of the species. As a species, the IUCN currently lists *P. tentorius* as Near Threatened (Hofmeyr et al., 2018a) with *P. t. trimeni* being listed as Endangered (Hofmeyr et al., 2018b).

Supplementary Figures

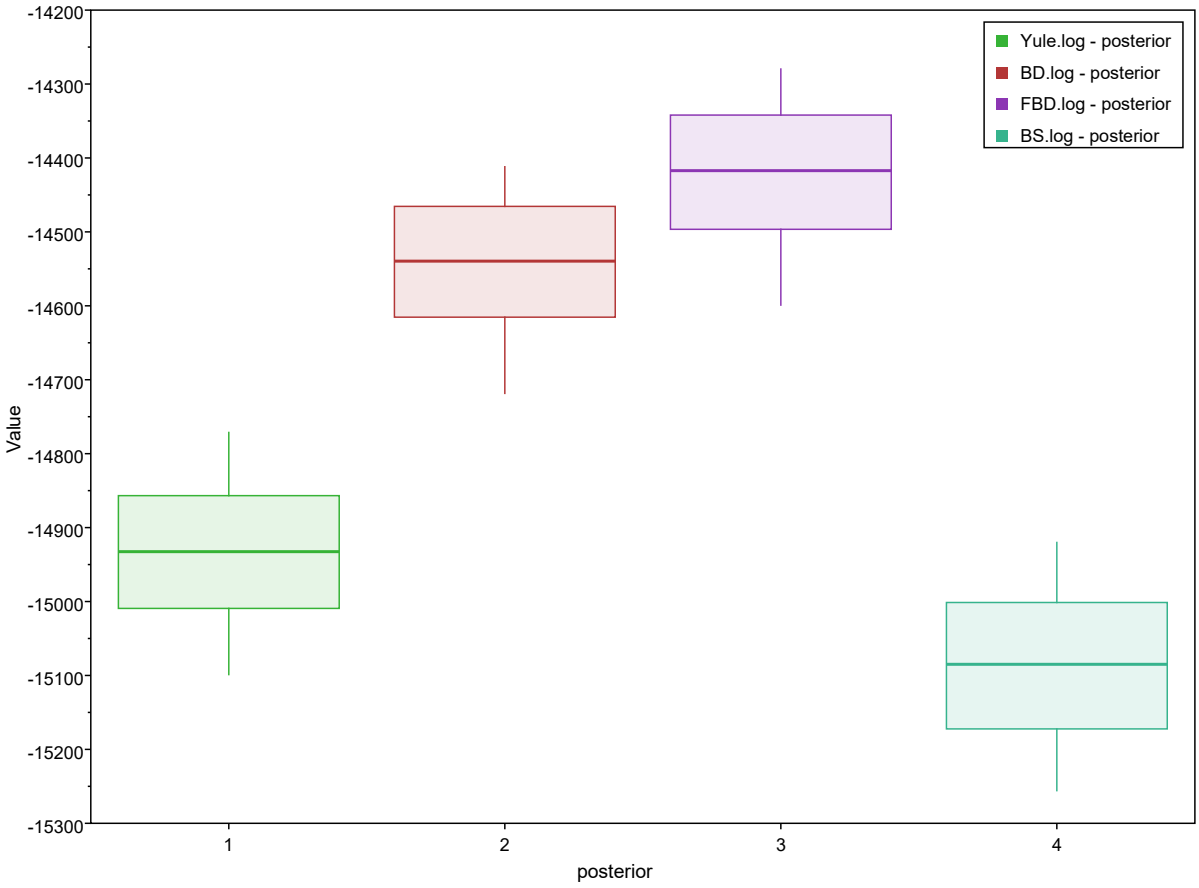


Figure S3.1. The posterior likelihood score (Y-axis) boxplots of the four calibration dating models (X-axis): “Yule”-Yule model, “BD”-Birth Death model, “FBD”-Fossil Birth Death model and “BS”-Bayesian Coalescent Skyline model.

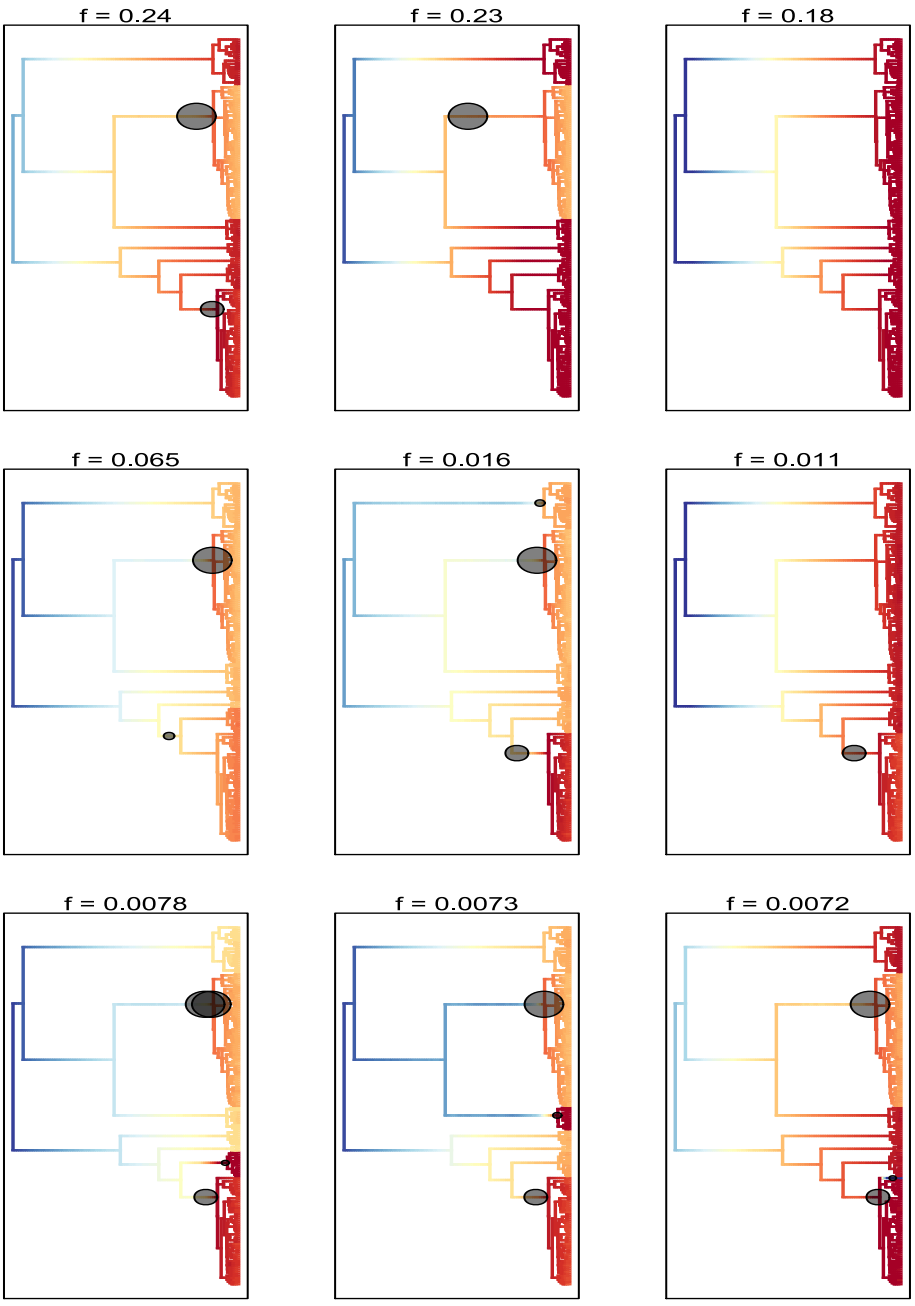


Figure S3.2. BAMM rate shift configurations with their corresponding frequencies at f = 0.24, f = 0.23, f = 0.18, f = 0.065, f = 0.016, f = 0.011, f = 0.0078, f = 0.0073 and f = 0.0072, respectively (critical nodes with significant rate shifts are labelled with a dark grey dot). The frequency at the top of each configuration scenario indicates the percentage of samples in the posterior distribution that can be assigned to each single shift configuration (e.g., a rate shift configuration of f = 0.24 (the best scenario) indicates that 24 % of the samples in the posterior distribution can be assigned to a two-shift configuration at nodes C1 and C2, where a distinct evolutionary rate is found, which differs significantly from other parts of the tree).

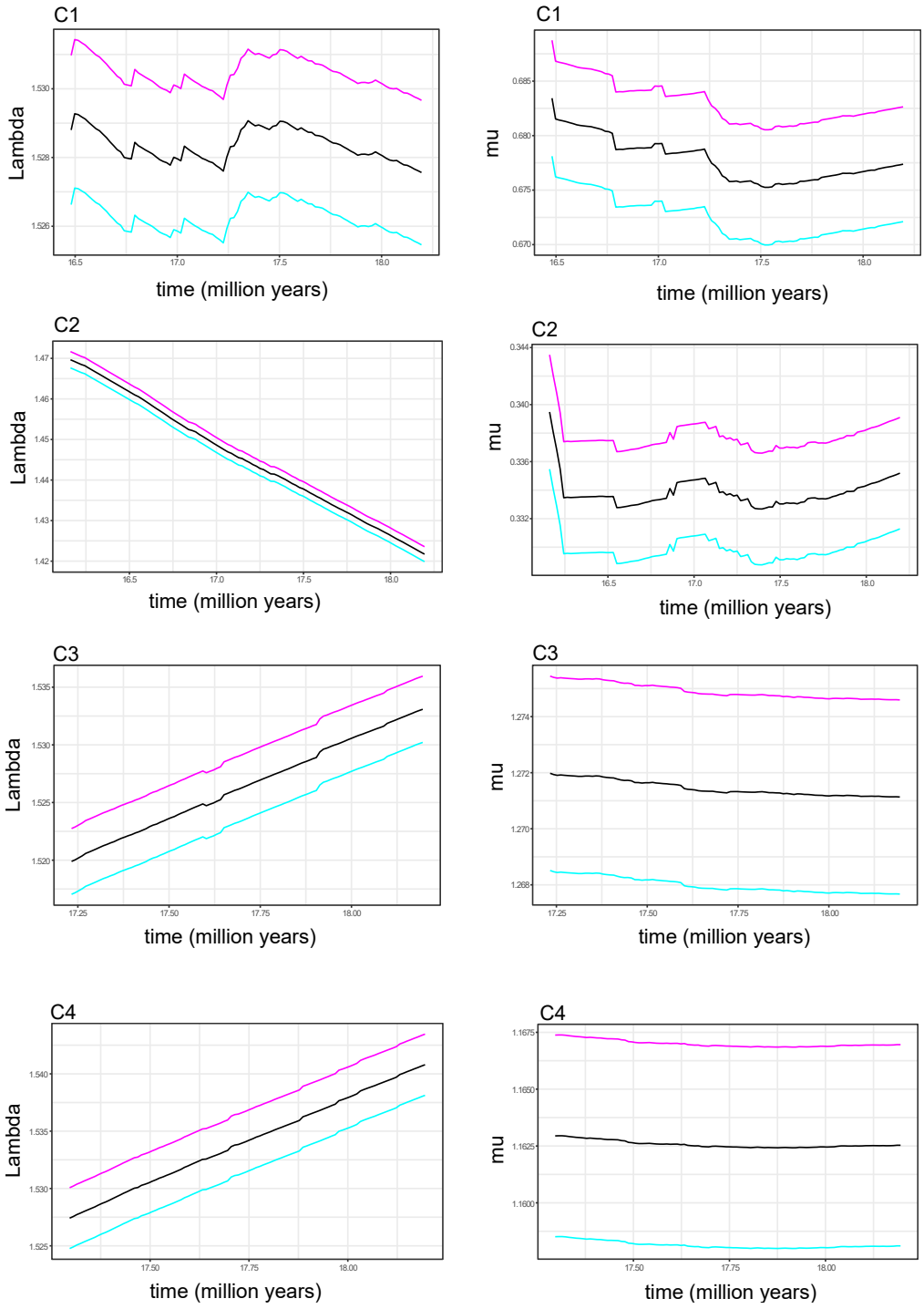


Figure S3.3. The Rate Through Time (RTT) plots for lambda and mu for C1-C4. Pink line at the top is the 95% HPD upper-boundary, the black line in the middle represents the mean value and the blue line at the bottom represents the lower-boundary of 95% HPD. The x-axis represents a recent time slice after the divergence of *Psammobates tentorius* commenced 18 million years ago.

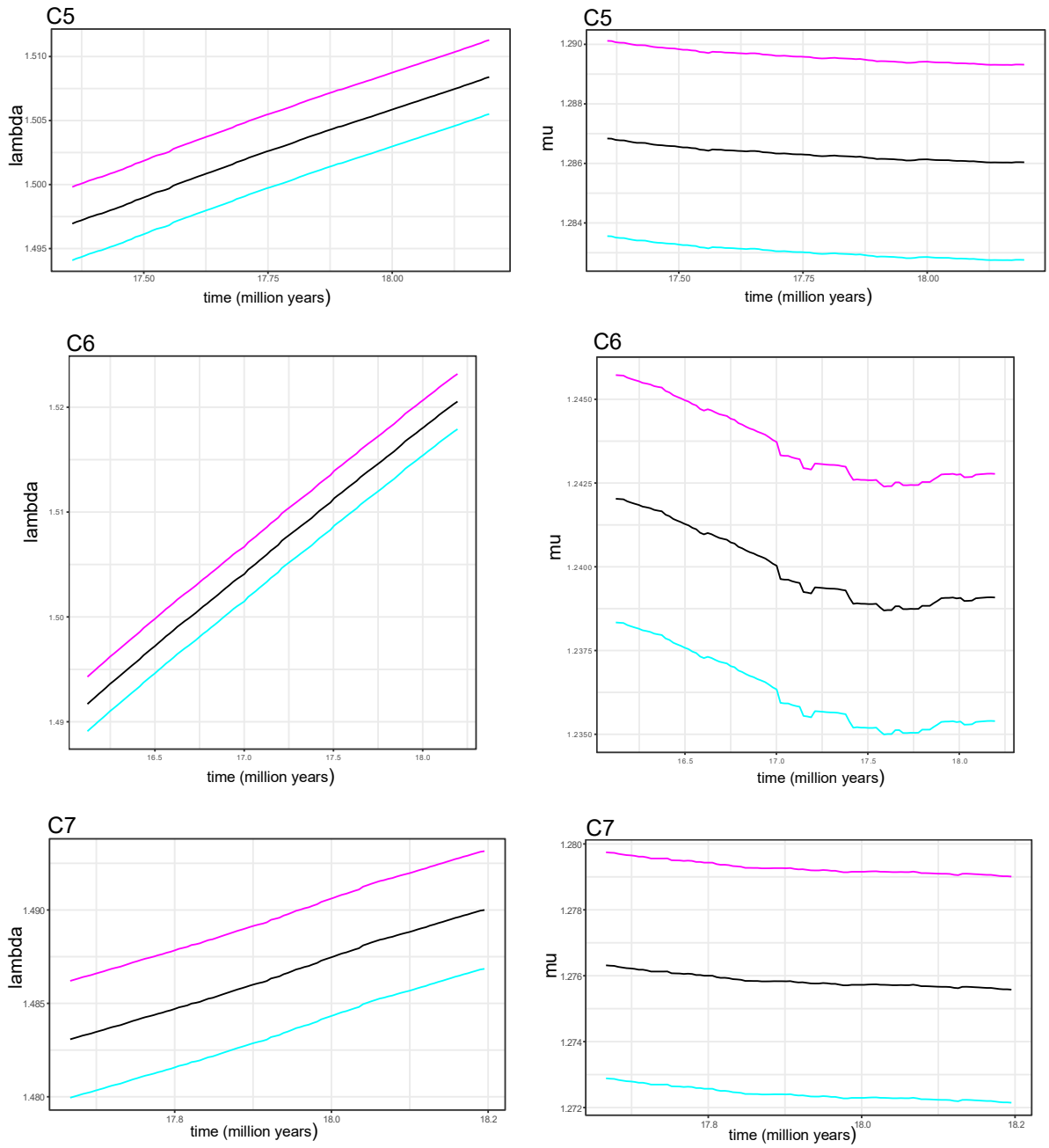


Figure S3.4. The Rate Through Time (RTT) plots for lambda and mu for C5-C7. Pink line at the top is the 95% HPD upper-boundary, black line in the middle represents the mean value and the blue line at the bottom represents lower-boundary of 95% HPD. The x-axis represents a recent time slice after the divergence of *Psammobates tentorius* commenced 18 million years ago.

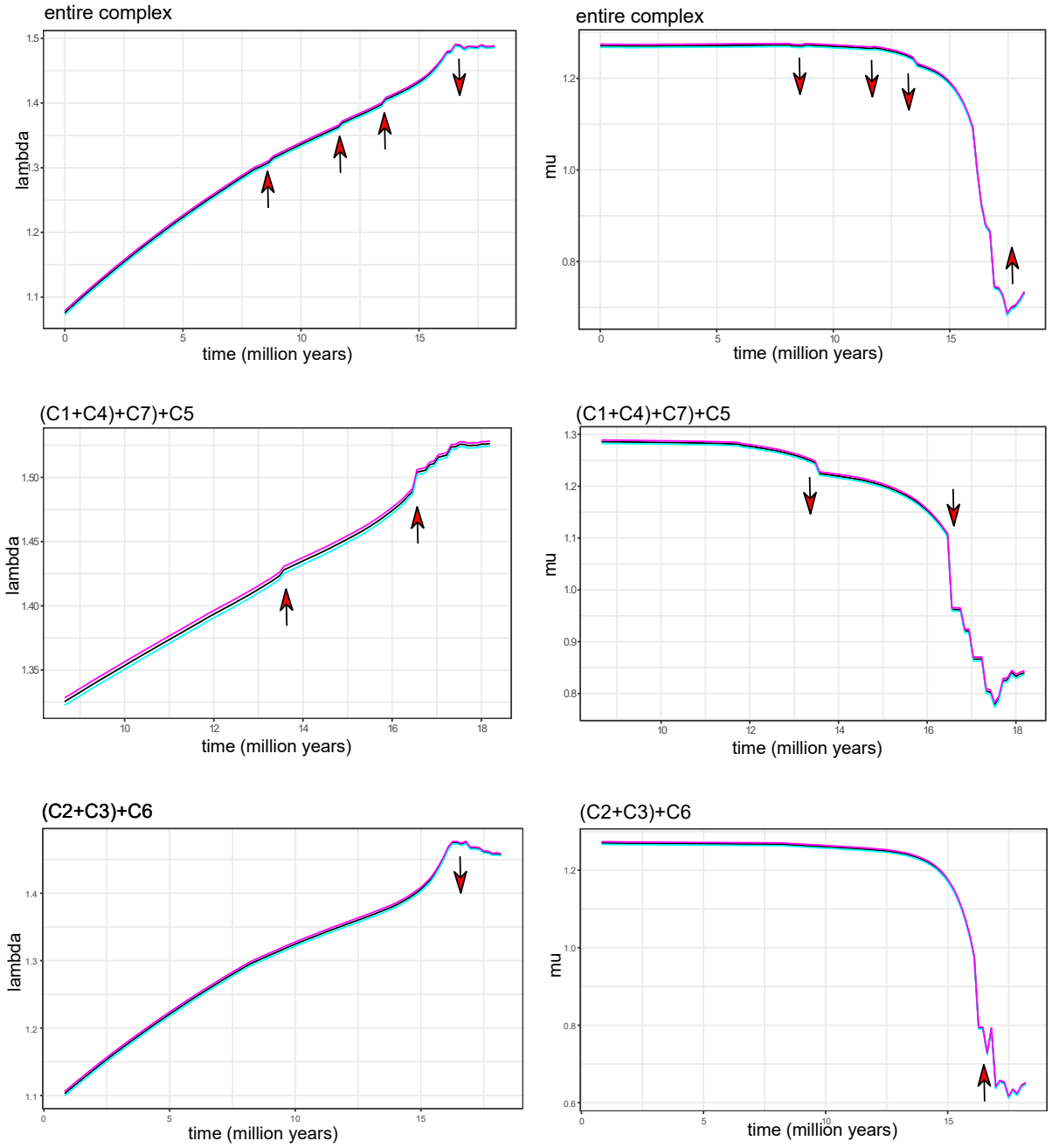


Figure S3.5. The Rate Through Time (RTT) plots for lambda and mu for the entire *Psammobates tentorius* complex, branch "(C1+C4)+C7)+C5" and "(C2+C3)+C6". Pink line at the top is the 95% HPD upper-boundary, black line in the middle represents the mean value and the blue line at the bottom represents the lower-boundary of the 95% HPD. The upward pointing arrows represent rate shifts with increasing trends, whilst, the downward pointing arrows indicate decreasing rate shifts. The x-axis represents a time slice after the divergence of *Psammobates tentorius* commenced 18 million years ago.

Supplementary Tables

All the Supplementary Tables (Table S3.1-S3.6) were not included in the main document of this thesis, as they are too large to fit in the main document. They are therefore provided independently as supporting documents.

References

Barlow, A., Baker, K., Hendry, C. R., Peppin, L., Phelps, T., Tolley, K. A., ... & Wüster, W. (2013). Phylogeography of the widespread African puff adder (*Bitis arietans*) reveals multiple Pleistocene refugia in southern Africa. *Molecular Ecology*, *22*(4), 1134-1157.

Bauer, A. M., & Lamb, T. (2005). Phylogenetic relationships of southern African geckos in the *Pachydactylus* group (Squamata: Gekkonidae). *African Journal of Herpetology*, *54*(2), 105-129.

Beaulieu, J.M., Tank, D.C., Donoghue, M.J., (2013). A Southern Hemisphere origin for campanulid angiosperms, with traces of the break-up of Gondwana. *BMC Evolutionary Biology*. 13(1), 80.

Born, J., Linder, H.P., & Desmet, P. (2007). The Greater Cape Floristic Region. *Journal of Biogeography*, 34, 147-162.

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.

Boycott, R. C., & Bourquin, O. (2000). *The Southern African Tortoise Book: A Guide to Southern African Tortoises, Terrapins and Turtles*. O. Bourquin.

Branch, B. (2008). *Tortoise, Terrapins and Turtles of Africa*. Cape Town: Struik Publisher.

Branch, W. R. (2006). Priorities for systematic studies on southern African reptiles. In W. R. Branch, K. A. Tolley, M. Cunningham, M. Cunningham, A. M. Bauer, G. Alexander, J. A. Harrison, A. A. Turner, & M. F. Bates (Eds.), *A plan for phylogenetic studies of southern African reptiles: proceedings of a workshop held at Kirstenbosch, February 2006* (pp. 2–20). Pretoria: South African National Biodiversity Institute.

Bauer, A. M. (1993). African-South American relationships: a perspective from the Reptilia. *Biological Relationships between Africa and South America*, 244-288.

Braun, J., Guillocheau, F., Robin, C., Baby, G., & Jelsma, H. (2014). Rapid erosion of the Southern African Plateau as it climbs over a mantle superswell. *Journal of Geophysical Research: Solid Earth*, *119*(7), 6093-6112.

Burke, K., & Gunnell, Y. (2008). *The African erosion surface: a continental-scale synthesis of geomorphology, tectonics, and environmental change over the past 180 million years* (Vol. 201). Geological Society of America.

Clark, V. R. (2010). *The phytogeography of the Sneeuberg, Nuweveldberge and Roggeveldberge (Great Escarpment): assessing migration routes and endemism* (unpublished doctoral dissertation). Rhodes University.

Clark, V. R., Barker, N. P., & Mucina, L. (2011). The Great Escarpment of southern Africa: a new frontier for biodiversity exploration. *Biodiversity and Conservation*, *20*(12), 2543.

Cowling, R.M., Bradshaw, P.L., Colville, J.F., Forest, F. (2017). Levyns' Law: explaining the evolution of a remarkable longitudinal gradient in Cape plant diversity. *Transactions of the Royal Society of South Africa*, 72, 184-201.

Cowling, R. M., Procheș, Ș., & Partridge, T. C. (2009). Explaining the uniqueness of the Cape flora: incorporating geomorphic evolution as a factor for explaining its diversification. *Molecular Phylogenetics and Evolution*, *51*(1), 64-74.

Cunningham, J. (2002). A molecular perspective on the family Testudinidae Batsch, 1788, (Unpublished doctoral dissertation). University of Cape Town.

Daniels, S. R., Heideman, N. J., & Hendricks, M. G. (2009). Examination of evolutionary relationships in the Cape fossorial skink species complex (Acontinae: *Acontias meleagris meleagris*) reveals the presence of five cryptic lineages. *Zoologica Scripta*, *38*(5), 449-463.

Dauteuil, O., Bessin, P., & Guillocheau, F. (2015). Topographic growth around the Orange River valley, southern Africa: A Cenozoic record of crustal deformation and climatic change. *Geomorphology*, *233*, 5-19.

de Broin, F. D. L. (2003). The Miocene chelonians from the southern Namibia, B. *Faunas from the southern Namibia, Mem. Geology Survey Namibia, 19*, 67-102.

De Wit, M. C. J. (1999). Post-Gondwana drainage and the development of diamond placers in western South Africa. *Economic Geology*, *94*(5), 721-740.

DeMenocal, P. B. (2004). African climate change and faunal evolution during the Pliocene–Pleistocene. *Earth and Planetary Science Letters*, 220, 3–24.

Diekmann, B., Fälker, M., & Kuhn, G. (2003). Environmental history of the south-eastern South Atlantic since the Middle Miocene: Evidence from the sedimentological records of ODP Sites 1088 and 1092. *Sedimentology*, *50*(3), 511-529.

Dupanloup, I., Schneider, S., Excoffier, L. (2002) A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology*, *11(*12), 2571-81.

Feakins, S. J., & Demenocal, P. B. (2010). Global and African regional climate during the Cenozoic. In L. Werdelin & W. J. Sanders (Eds) *Cenozoic mammals of Africa* (pp. 45-55). Berkeley: University of California Press.

Fedorov, A. V., Dekens, P. S., McCarthy, M., Ravelo, A. C., DeMenocal, P. B., Barreiro, M., ... & Philander, S. G. (2006). The Pliocene paradox (mechanisms for a permanent El Niño). *Science*, *312*(5779), 1485-1489.

FitzJohn, R. G. (2012). Diversitree: comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution*, *3*(6), 1084-1092.

FitzJohn, R. G., Maddison, W. P., & Otto, S. P. (2009). Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Systematic Biology*, *58*(6), 595-611.

Goldberg, E. E., Lancaster, L. T., & Ree, R. H. (2011). Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. *Systematic Biology*, *60*(4), 451-465.

Grab, S., Knight, J. (2015). Landscapes and landforms of South Africa – an overview. In S. Grab & J. Knight (Eds.), *Landscapes and Landforms of South Africa*. (pp. 1-9). Switzerland: Springer International Publishing AG.

Green, P. F., Duddy, I. R., Japsen, P., Bonow, J. M., & Malan, J. A. (2017). Post-breakup burialand exhumation of the southern margin of Africa. *Basin Research*, *29*(1), 96-127.

Greig, J. C. (1975). Patterns In the distribution of Southern African terrestrial tortoises (Chelonia: Cryptodira: Testudinidae). *The Journal of the Herpetological Association of Africa*, *14*(1), 9-9.

Greig, J. C., & Burdett, P. D. (1976). Patterns in the distribution of southern African terrestrial tortoises (Cryptodira: Testudinidae). *African Zoology*, *11*(2), 251-273.

Harmon, L., Weir, J., Brock, C., Glor, R., Challenger, W., Hunt, G., ... & Uyeda, J. (2014). Package 'geiger'. *Bioinformatics*, *24*, 129-131.

Heinicke, M.P., Turk, D., & Bauer, A.M. (2017). Molecular phylogeny reveals strong biogeographic signal and two new species in a Cape Biodiversity Hotspot endemic mini-radiation, the pygmy geckos (Gekkonidae: *Goggia*). *Zootaxa, 4312*, 449–470.

Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature, 405*(6789), 907.

Hewitt, J. (1933). On the Cape species and subspecies of the genus *Chersinella* Gray. Part I. *Annals of the Natal Museum*, *7*, 255-297.

Hewitt, J. (1934). On the Cape species and subspecies of the genus *Chersinella*. Part II. *Annals of the Natal Museum, 7*, 303-352.

Hoetzel, S., Dupont, L. M., Marret, F., Jung, G., and Wefer, G. (2015). Miocene–Pliocene stepwise intensification of the Benguela upwelling over the Walvis Ridge off Namibia, *Climate in the Past Discuss*, 11, 1913-1943.

Hoffmann, V., Verboom, G. A., & Cotterill, F. P. (2015). Dated plant phylogenies resolve neogene climate and landscape evolution in the Cape Floristic Region. *PLoS One*, *10*(9), e0137847.

Hofmeyr, M. D., Boycott, R. C., & Baard, E. H. W. (2014). *Psammobates tentorius* (Bell, 1828). In: M. F. Bates, W. R. Branch, A. M. Bauer, M. Burger, J. Marais, G. J. Alexander & M. S. De Villiers (Eds.), *Atlas and red list of the reptiles of South Africa, Lesotho and Swaziland* (pp. 70-85). Pretoria: South African National Biodiversity Institute.

Hofmeyr, M. D., Vamberger, M., Branch, W., Schleicher, A., & Daniels S. R. (2017). Tortoise (Reptilia, Testudinidae) radiations in Southern Africa from the Eocene to the present. *Zoologica Scripta*, *46*, 389-400.

Hofmeyr, M.D., Leuteritz, T. & Baard, E.H.W. 2018. *Psammobates tentorius ssp. trimeni*. *The IUCN Red List of Threatened Species 2018*: e.T121936835A121936853.

Hofmeyr, M.D., Leuteritz, T. & Baard, E.H.W. 2018a. *Psammobates tentorius*. *The IUCN Red List of Threatened Species 2018*: e.T170524A115656793.

Huelsenbeck, J. P., & Rannala, B. (2004). Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Systematic Biology*, *53*(6), 904-913.

Irish, J. (2008). Biological characterisation of the Orange-Fish River Basin, Namibia. *report for the Ephemeral River Basins in Southern Africa (ERB) project, Desert Research Foundation of Namibia (DRFN), Windhoek.*

Jacob, R. J. (2005). *The erosional and Cainozoic depositional history of the Lower Orange River, southwestern Africa* (unpublished octoral dissertation). University of Glasgow.

Jansson, R., & Dynesius, M. (2002). The fate of clades in a world of recurrent climatic change: Milankovitch oscillations and evolution. *Annual Review of Ecology and Systematics*, 33, 741–77.

Jung, G., Prange, M., Schulz, M. (2014). Uplift of Africa as a potential cause for Neogene intensification of the Benguela upwelling system. *Nature Geoscience*, 7, 741-747.

Kounov, A., Viola, G., De Wit, M. J., & Andreoli, M. (2008). A Mid Cretaceous paleo-Karoo River valley across the Knersvlakte plain (northwestern coast of South Africa): Evidence from apatite fission-track analysis. *South African Journal of Geology*, *111*(4), 409-420.

Librado, P., & Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, *25*(11), 1451– 1452.

Lourenço, J. M., Claude, J., Galtier, N., & Chiari, Y. (2012). Dating cryptodiran nodes: origin and diversification of the turtle superfamily Testudinoidea. *Molecular Phylogenetics and Evolution*, *62*(1), 496-507.

Maddison, W. P., Midford, P. E., & Otto, S. P. (2007). Estimating a binary character's effect on speciation and extinction. *Systematic Biology*, *56*(5), 701-710.

Massana, K. A., Beaulieu, J. M., Matzke, N. J., & O'Meara, B. C. (2015). Non-null effects of the null range in biogeographic models: exploring parameter estimation in the dec model. *bioRxiv*, 026914.

Matthee, C. A., & Flemming, A. F. (2002). Population fragmentation in the southern rock agama, *Agama atra*: more evidence for vicariance in southern Africa. *Molecular Ecology*, 11, 465-471.

Matzke, N. J. (2013a). BioGeoBEARS: BioGeography with Bayesian (and likelihood) evolutionary analysis in R Scripts. *R package, version 0.2, 1*, 2013.

Matzke, N. J. (2013b). Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Frontiers of Biogeography*, *5*(4).

Matzke, N. J. (2014). Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Systematic Biology*, *63*(6), 951-970.

Moore, A., Blenkinsop, T., & Cotterill, F. (2009). Southern African topography and erosion history: plumes or plate tectonics?. *Terra Nova*, *21*(4), 310-315.

Mucina, L., & Rutherford, M. C. (2011). *The vegetation of South Africa, Lesotho and Swaziland*. Pretoria: South African National Biodiversity Institute.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, *403*, 853-858.

Neumann, F. H., & Bamford, M. K. (2015). Shaping of modern southern African biomes: Neogene vegetation and climate changes. *Transactions of the Royal Society of South Africa*, *70*(3), 195-212.

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*(2), 289-290.

Partridge, T. C., & Maud, R. R. (1987). Geomorphic evolution of southern Africa since the Mesozoic. *South African Journal of Geology*, *90*(2), 179-208.

Partridge, T.C., Dollar, E.S.J., Moolman, J., Dollar, L.H. (2010). The geomorphic provinces of South Africa, Lesotho and Swaziland: a physiographic subdivision for earth and environmental scientists. *Transactions of the Royal Society of South Africa*, *65*, 1-47.

Potts, A. J., Hedderson, T. A., Vlok, J. H., & Cowling, R. M. (2013). Pleistocene range dynamics in the eastern Greater Cape Floristic Region: a case study of the Little Karoo endemic *Berkheya cuneata* (Asteraceae). *South African Journal of Botany*, *88*, 401-413.

Rabosky, D. L. (2014). Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PloS One*, *9*(2), e89543.

Rabosky, D. L., Donnellan, S. C., Grundler, M., & Lovette, I. J. (2014). Analysis and visualization of complex macroevolutionary dynamics: an example from Australian scincid lizards. *Systematic Biology*, *63*(4), 610-627.

Rabosky, D. L., Grundler, M., Anderson, C., Title, P., Shi, J. J., Brown, J. W., ... & Larson, J. G. (2014). BAMM tools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods in Ecology and Evolution*, *5*(7), 701-707.

R Core Team. (2017). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna*. URL http://www.R-project.org/

Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, *3*(2), 217-223.

Rhodin, A. G. J., Iverson, J. B., Bour, R., Fritz U., Georges, A., Shaffer, H. B., & van Dijk P.P. (2017). Turtles of the World: Annotated Checklist and Atlas of Taxonomy, Synonymy, Distribution, and Conservation Status (8th ed.). In: A. G. J. Rhodin, J. B. Iverson, P. P. van Dijk, R. A. Saumure, K. A. Buhlmann, P. C. H Pritchard & R. A. Mittermeier (Eds.), *Conservation Biology of Freshwater Turtles and Tortoises: A Compilation Project of the IUCN/SSC Tortoise and Freshwater Turtle Specialist Group* (pp. 1–292). Lunenburg: Chelonian Research Monographs.

Ricklefs, R. E. (2007). Estimating diversification rates from phylogenetic information. *Trends in Ecology & Evolution*, *22*(11), 601-610.

Roberts, D. L., Sciscio, L., Herries, A. I., Scott, L., Bamford, M. K., Musekiwa, C., & Tsikos, H. (2013). Miocene fluvial systems and palynofloras at the southwestern tip of Africa: Implications for regional and global fluctuations in climate and ecosystems. *Earth Science Reviews*, *124*, 184-201.

Rommerskirchen, F., Condon, T., Mollenhauer, G., Dupont, L., & Schefuß, E. (2011). Miocene to Pliocene development of surface and subsurface temperatures in the Benguela Current system. *Paleoceanography and Paleoclimatology*, *26*(3).

Rubin, F. (1998). The physical environment and major plant communities of the Tankwa-Karoo National Park. *Koedoe*, *41*(2), 61-94.

Rudge, J. F., Roberts, G. G., White, N. J., & Richardson, C. N. (2015). Uplift histories of Africa and Australia from linear inverse modelling of drainage inventories. *Journal of Geophysical Research: Earth Surface*, 120, 894-914.

Russo, I. R. M., Chimimba, C. T., & Bloomer, P. (2010). Bioregion heterogeneity correlates with extensive mitochondrial DNA diversity in the Namaqua rock mouse, *Micaelamys namaquensis* (Rodentia: Muridae) from southern Africa-evidence for a species complex. *BMC Evolutionary Biology*, *10*(1), 307.

Scott, I. A., Keogh, J. S., & Whiting, M. J. (2004). Shifting sands and shifty lizards: molecular phylogeny and biogeography of African flat lizards (*Platysaurus*). *Molecular Phylogenetics and Evolution*, *31*(2), 618-629.

Shaffer, H. B., Minx, P., Warren, D. E., Shedlock, A. M., Thomson, R. C., Valenzuela, N., ... & Borchert, G. M. (2013). The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. *Genome Biology*, *14*(3), R28.

Shi, J. J., & Rabosky, D. L. (2015). Speciation dynamics during the global radiation of extant bats. *Evolution*, *69*(6), 1528-1545.

Sithaldeen, R., Ackermann, R. R., & Bishop, J. M. (2015). Pleistocene aridification cycles shaped the contemporary genetic architecture of southern African Baboons. *PloS One*, *10*(5), e0123207.

Snijman, D. A. (2013). *Plants of the Greater Cape Floristic Region. 2: The Extra Cape flora*. Pretoria: South African National Biodiversity Institute.

Tolley, K. A., Bowie, R. C. K., Measey, G. J., Price, B. W., & Forest, F. (2014). The shifting landscape of genes since the Pliocene: terrestrial phylogeography in the Greater Cape Floristic Region. In: N. Allsopp, J. F. Colville, A. Verboom (Eds.), *Fynbos: Ecology, Evolution, and Conservation of a Megadiverse Region* (pp. 142–163). Oxford, UK: Oxford University Press.

Tolley, K. A., Makokha, J. S., Houniet, D. T., Swart, B. L., & Matthee, C. A. (2009). The potential for predicted climate shifts to impact genetic landscapes of lizards in the South African Cape Floristic Region. *Molecular Phylogenetics and Evolution*, *51*(1), 120-130.

Vamberger, M., Hofmeyr, M.D., Ihlow, F., Fritz, U. (2018). In quest of contact: phylogeography of helmeted terrapins (*Pelomedusa galeata*, *P. subrufa* senso stricto). *PeerJ* 6:e4901

Verboom, G. A., Archibald, J. K., Bakker, F. T., Bellstedt, D. U., Conrad, F., Dreyer, L. L., ... Mummenhoff, K. (2009). Origin and diversification of the Greater Cape flora: ancient species repository, hot-bed of recent radiation, or both?. *Molecular Phylogenetics and Evolution*, *51*(1), 44-53.

Wickham, H., & Wickham, M. H. (2019). Package 'plyr'. Obtenido de https://cran. rproject. org/web/packages/dplyr/dplyr. pdf.

Wickham, H., Chang, W., & Wickham, M. H. (2016). Package 'ggplot2'. *Create Elegant Data Visualisations Using the Grammar of Graphics. Version*, *2*(1), 1-189.

Wickham, H., Hester, J., Francois, R., Jylänki, J., Jørgensen, M. (2017). readr: read rectangular text data. R package version 1.1. 1.

Wildman, M., Brown, R., Persano, C., Beucher, R., Stuart, F. M., Mackintosh, V., ... Carter, A. (2017). Contrasting Mesozoic evolution across the boundary between on and off craton regions of the South African plateau inferred from apatite fission track and (U-Th-Sm)/He thermochronology. *Journal of Geophysical Research: Solid Earth*, *122*(2), 1517-1547.

Wildman, M., Brown, R., Watkins, R., Carter, A., Gleadow, A., & Summerfield, M. (2015). Post break-up tectonic inversion across the southwestern cape of South Africa: New insights from apatite and zircon fission track thermochronometry. *Tectonophysics*, *654*, 30-55.

Yu, Y., Harris, A. J., Blair, C., & He, X. (2015). RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. *Molecular Phylogenetics and Evolution*, *87*, 46-49.

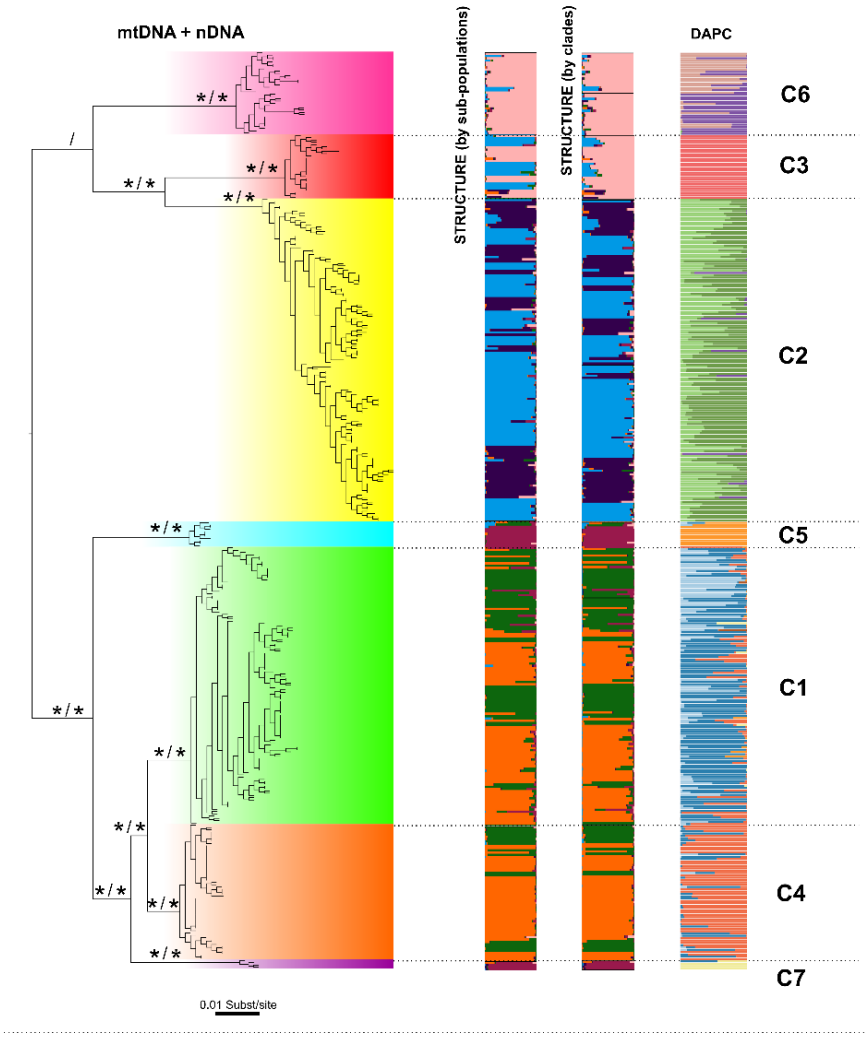
Yu, Y., Harris, A. J., He, X. J. (2010). S-DIVA (statistical dispersal-vicariance analysis): a tool for inferring biogeographic histories. *Molecula Phylogenetics and Evolution*, 56(2):848-850.

Zachos, J., Pagani, M., Sloan, L., Thomas, E., & Billups, K. (2001). Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science*, *292*(5517), 686-693.

Zheng, Y., Peng, R., Kuro-o, M., & Zeng, X. (2011). Exploring patterns and extent of bias in estimating divergence time from mitochondrial DNA sequence data in a particular lineage: a case study of salamanders (Order Caudata). *Molecular Biology and Evolution*, *28*(9), 2521-2535.

Chapter 4. When microsatellite and sequence DNA data square-off under multiple methodologies to test alternative phylogenetic assumptions in the taxonomically confusing tent tortoise species complex (Reptilia, Testudinidae).

Graphical abstract



Abstract

This study compared the different phylogenetic assumptions retrieved from three types of molecular markers (nDNA, mtDNA and microsatellite DNA) in the highly polymorphic and taxonomically confusing tent tortoise (*Psammobates tentorius*) species complex in southern Africa. I also used fine scale population genetic analyses to investigate genetic structure and geneflow using both the mtDNA and microsatellite DNA datasets. The mtDNA and microsatellite DNA datasets yielded similar genetic structure and gene flow patterns among the seven mtDNA clades. The genetic difference between C1 and C4 was, however, not significant for the microsatellite DNA dataset, while the within species level genetic diversities between the two datasets were different. Further evidence was found of possible hybridization between C1 and C2 in their intergradation zone, but no microsatellite based evidence of possible hybridization between C2 and C4. Results of the inbreeding analyses provided strong evidence of inbreeding in the eastern population of C1 in the Fish River valley, which may be indicative of a bottleneck effect. Both the mtDNA and microsatellite DNA datasets revealed a significant genetic subdivision between the eastern and western populations of C6, which suggests that further radiation and therefore speciation within C6 can be expected. The Approximate Bayesian Computation based simulation analyses advocated an alternative phylogenetic assumption for the tent tortoise complex, which was better and more inclusive in explaining its genealogical history and supported the recognition of at least four species.

4.1 Introduction

Southern Africa harbours the world's richest diversity of tortoise species, with a high level of endemism. Taxonomically the highly polymorphic tent tortoises complex of southern Africa is one of the world's most confusing testudine groups (Branch, 2008; Hofmeyr et al., 2014; Rhodin et al., 2017), as manifested, amongst others, in the highly variable shape and colour patterns of its carapace and plastron. Although Hewitt (1933 & 1934) described many species and subspecies based on regional character colour variations, Loveridge and Williams (1957) recognised only three subspecies, *Psammobates tentorius tentorius*, *Psammobates tentorius verroxii* and *Psammobates tentorius trimeni*, also supported by Branch (2008), Hofmeyr et al., (2014) and Rhodin et al. (2017). There are, however, still unresolved taxonomic complexities in some populations, particularly in *P. t. verroxii*, where the levels of polymorphism are the highest, warranting a re-evaluation of some of its synonymised taxa (Hofmeyr et al., 2014; Rhodin et al., 2017). There are substantial distribution range overlaps among the three subspecies, with seemingly hybrid individuals found in the intergradation zones, though many are believed to be misidentifications (Hofmeyr et al., 2014; Rhodin et al., 2017). Conducting research on the *P. tentorius* complex is also challenging as it generally occurs in low densities throughout its distribution range (Branch, 2008; Hofmeyr et al., 2014).

Chapter 2 based on six mtDNA loci and one nDNA marker revealed 7 evolutionary lineages, while multiple species delimitation analyses in general recognised 6 to 7 putative species. The findings placed clades C1, C5 and C7 in the *P. t. tentorius* group, each occurring in a different geographic region generally isolated from each other by significant geographic barriers. The *P. t. veroxii* group was found not to be monophyletic, its populations north of the Orange River (C6) being genetically distinct from the populations south of the Orange River (C2). Furthermore, C2 and C6 were not sister taxa, implying that their designation as *P. t. verroxii* is not valid and should be reviewed. The nDNA analysis, however, did not reveal a significant genetic difference between C2 and C6. These findings along with the uncertainty of the phylogenetic position of C6 (Chapter 2) therefore highlights the need for further investigation to resolve its taxonomic status, using additional markers. With respect to the *P. t. trimeni* group, the study supports the findings of Rhodin et al. (2017), that true *P. t. trimeni* only occurs along the Cape West coast. Its populations occur in the Roggeveldberge, Hantam Karoo and Kamiesberg (RHK) regions, and were also previously considered as *P. t. trimeni* by Boycott &

Bourquin (1988). In Chapter 2, I found the RHK populations, although genetically assigned to C4, are closely related to C1 of the *P. t. tentorius* group. The phylogenetic position of the samples from the Uniondale valley also remains to be determined.

Given their conflicting findings, the mtDNA and nDNA phylogenies failed to satisfactorily resolve the taxonomic dilemma of the *P. tentorius* complex. In this study I attempted to provide further clarity on this matter by using multiple microsatellite DNA loci together with the mtDNA sequence dataset of Chapter 2, in a finer scale genetic structure analysis of the species complex.

The high level of discordance between phylogenies based on mtDNA and nDNA loci has been addressed in a variety of testudine species (Spinks & Shaffer, 2009; Wiens et al., 2010). Despite the emphasis on mtDNA markers to infer phylogenies, given their high mutation rates (Ballard & Rand, 2005; Rubinoff & Holland, 2005), several studies revealed that they can overestimate evolutionary rates and provide misleading phylogenetic inferences (Shaw, 2002; Leaché et al., 2010; Wiens et al., 2010). Although mtDNA and nDNA phylogenies are often incongruent, nDNA generally being less informative because of its conservative evolutionary rate, some studies found that combining mtDNA and nDNA datasets into a supermatrix can improve phylogenetic inference results (Reid and Wiens, 2011). It is, however, crucial to check incongruencies between the two types of data before combining them. The results of Reid and Wiens (2011) further suggest that the combined analysis is not necessarily dominated by the more variable mtDNA dataset. Mitochondrial and nuclear DNA markers are located in different genomes, evolve at different rates and function differently, and thus carry different evolutionary information which are not necessarily congruent, but nonetheless represent fundamental elements of genomic evolution (Rubinoff & Holland, 2005).

Phylogenetic and population genetics studies focus on organisms with varying lifespans, such as species complexes and slow-evolving groups such as tortoises (Avise et al., 1992), often with a relatively shallow depth of coalescence. Gene tree topologies based on a limited number of loci may therefore not always accurately reflect the real phylogenetic information presented in species trees (Zheng et al., 2011; Lourenço et al., 2012). Such incongruent genealogical information by gene and species trees is usually heavily impacted by incomplete lineage sorting and the level of coalescence (Rubinoff & Holland, 2005; Baum & Smith, 2013; Leliaert et al., 2014), and the gap between the two can be particular large in slow evolving groups.

The traditional phylogenetic approach is usually restricted by limited scenario options, and can be misleading in the interpretation of phylogenetic history, particularly when the study uses few markers (Ballard & Rand, 2005; Cornuet et al., 2014). To address this issue and improve phylogenetic and genealogical reconstruction accuracy, I used Approximate Bayesian Computation (ABC) to infer population history by carrying out simulations with the microsatellite DNA and mtDNA datasets under certain constrained scenarios. This was done in order to rank the likelihood of different scenarios and determine the most plausible cladogenic scenarios without being limited to the few scenarios of traditional phylogenetic analytical approaches.

The aims of the the study were, a) to compare phylogenetic topologies between multiple loci sequence (mtDNA and nDNA) and microsatellite DNA datasets to assess their degree of congruence, b) to compare genetic structure patterns between the sequence and microsatellite DNA datasets, c) to carry out simulations with the combined dataset (microsatellite and sequence DNA datasets) under different scenarios obtained from different analyses and datasets, to determine the most plausible cladogenetic scenarios, d) to use the findings to make suggestions about the taxonomic revision of the *P. tentorius* complex, and e) to determine the genetic diversities and inbreeding indices of clades, and potential gene flow among them from a conservation perspective.

4.2 Materials and Methods

4.2.1 Sampling and DNA extraction

A total of 404 specimens of *P. tentorius* collected from 76 localities were used for the sequencing and microsatellite DNA genotyping and analyses in this study, The collection covered its entire distribution range comprehensively (Table S4.1). The details on DNA samples collections, DNA extractions see Chapter 2.

4.2.2 mtDNA and nDNA amplification and sequencing

The methods and materials on DNA amplifications and sequencing were same as in Chapter 2. The optimal annealing temperatures and primers used for the six mtDNA markers are given in Table S4.2. All NCBI GenBank accession numbers of sequence data are given in Table S4.1.

4.2.3 Microsatellite DNA amplification, sequencing and genotyping

Extracted genomic DNA was used as template DNA in the PCR. Twenty two microsatellite DNA loci, which were successful in previous Testudine population genetics studies (Vamberger et al., 2001; Ciofi et al., 2002; Schwartz et al., 2003; Forlani et al., 2005; Paquette et al., 2005; Mandimbihasina et al., 2009; Orozco-terWengel et al., 2013), were tested. They were tested on the target animal using PCR, with all forward primers labelled fluorescently with different dye probes (primer details and optimal annealing temperatures are given in Table S4.3). All PCR reactions were performed using KAPA2G Robust HotStart Readymix, USA, in a BIO-RAD T 100^{TM} Thermal Cycler (Singapore) under the following parameters: an initial 5 min. denaturation step at 94 ºC, followed by 35 cycles of 30 seconds of denaturation at 94 ºC, 30 seconds of annealing (different optimal annealing temperatures were used depending on the loci, details given in Table S4.3), and a 1 min. extension step at 72 ºC, with a final 10 min. extension step at 72 ºC. The PCR annealing temperatures for the different microsatellite DNA loci were optimized using temperature gradient tests. Only loci showing a positive amplicon peak were used for further genotyping.

I applied multiple mixed PCR's to enable the simultaneous amplification of multiple microsatellite DNA markers. I subdivided the microsatellite markers into six multiple mixed PCR groups according to the type of dye probes, fragment lengths and optimal PCR annealing temperatures, to maximize visualization of the genotyping and to minimize the interference among the different loci during genotyping (details given in Table S4.3). The PCR conditions for these six multiply mix reactions were the same as above for the single loci based PCR. The annealing temperature for each multiple mix reaction is given in Table S4.3.

In order to verify the microsatellite repeat motif unit in the *P. tentorius* complex, and to ensure that the amplicons obtained from the dye labelled primer pairs indeed came from the target region, and not from other regions (caused by unspecific primer binding sites), I sequenced some individuals of the seven clades with the same primer pairs but without the dye labels

(with the same PCR conditions as the above genotyping). The PCR products were electrophoresed in 1% agarose gel, visualized under UV light, and purified using a BioFlux PCR Purification Kit (Bioer Technology, China). Purified PCR products were cycle sequenced using BigDye (ABI PRISM® BigDye Terminator v3.1 Cycle Sequencing Kits, USA) and standard methods. The Big-Dye PCR products were purified with Zymo DNA Sequencing clean-up kits (Epigenetics Company, USA), prior to sequencing in an ABI 3500 genetic analyser.

I used PCR amplicon length for microsatellite genotyping and analyses throughout the study. In all cases, CONVERT (Glaubitz, 2004) and PGDSpider (Lischer and Excoffier, 2011) were used to convert datasets into different input file formats for different analyses.

Sanger sequences were analysed using ABI Prism Sequencing Analysis software v.3.7 (Applied Biosystems), then aligned with MUSCLE v.3.2 (Edgar, 2004) and manually checked with MEGA v.7 (Kumar, Stecher, & Tamura, 2016). Microsatellite DNA was genotyped with GeneMarker v.2.4 (Holland & Parson, 2011).

4.2.4 Microsatellite DNA allele range

Among these tested 22 microsatellite loci, 19 of which showed positive results, therefore the 19 loci with positive results were used for carrying further analyses. In order to determine and visualize allele range, and differences among 19 microsatellite DNA markers across the seven clades (see Chapter 2), STATISTICA v.8 (StatSoft) was used to plot allele range. A preliminary Discriminant Function Analysis (DFA) was performed with GenAlex v.6 (Peakall and Smouse, 2006) on the complete microsatellite dataset (see Appendix 4) to check for suspicious samples and thereby eliminating errors in genotyping prior to the formal analyses. Samples were grouped according to the seven mtDNA clades (see Chapter 2) for the preliminary DFA analysis.

4.2.5 Determining of potential null alleles, stuttering and large allele dropout

To ensure that the microsatellite dataset did not contain null alleles, stuttering and large allele dropout (Oosterhout et al., 2004), I checked it using Micro-Checker (Oosterhout et al., 2004)

to avoid misleading genotyping results. For all classes a combined probability of $p < 0.05$ was considered a significant indicator of the presence of null alleles.

4.2.6 Assignment test on unknown individuals

Six individuals came from unknown localities, and in order to assign them to their most likely groups, I performed assignment tests with GeneClass v 2.0 (Piry et al., 2004). The assignment threshold of score was set at 0.05 following Rannala & Mountain's Bayesian methods (Rannala & Mountain, 1997) and the Monte-Carlo resampling approach of Paetkau et al. (2004). I used Arlequin v 3.5 (Excoffier et al., 2009) to perform assignment tests on 2 of the Uniondale samples which had been successfully genotyped, in order to determine their closest group at clade level. The assignment plots were drawn using R v.3.5 with the "Rcmd" function (R Core Team, 2018).

4.2.7 Population genetics analysis

4.2.7.1 Genetic structure

To determine genetic structure and significant clusters in the microsatellite dataset, the Discriminant Analysis of Principal components (DAPC, Jombart et al., 2010) was performed using the R package 'adegenet' (Jombart, 2008) of the program R 3.5. I first plotted a graph of the cumulative variance versus the number of Principal components (PCs), initially retaining 400 PCs to initiate the analyses. I first used the BIC criteria to determine the best k value (the number of clusters that should be retained), as visualized by a screen plot. I then used the a-score to measure the trade-off between discrimination power and over-fitting. The results of the a-score analysis were used to further optimize the number of PCs in the second DAPC analysis. Finally, I conducted a DAPC Cross-Validation test (using function "xvalDapc" to identify the 'goldilocks point' in the trade-off between retaining too few or too many PCs in the model. The optimal group membership cluster scheme was mapped using the "find, clusters" function.

STRUCTURE v 2.3 (Falush et al., 2003) was used to map the best cluster scheme by using Admixture as ancestry model to allow individuals the possibility of having mixed ancestry with

the Bayesian algorithms. In order to test genetic structure at different levels, I partitioned the dataset into two testing schemes. In the first scheme, individuals were divided into seven mtDNA clades with the Uniondale samples treated as a separate group, thus giving a total of eight populations. In the second scheme individuals were divided into 11 populations. Clade 1 was subdivided into a western and an eastern population. Clade 6 was also subdivided into a western and eastern population. Both C1 and C6 showed clear subdivisions in Chapter 2. As for C2, the northern population is morphologically distinct from the southern population, I therefore subdivided it into a northern and southern population (although the mtDNA loci did not show them as clearly separate according to Chapter 2). Analyses were run for 5 million generations using MCMC with the first 50, 000 discarded as burn-in. Time series plots against different parameters were used to establish whether sampling was sufficient. STRUCTURE HARVESTER (Earl, 2012) was used to summarize outputs of STRUCTURE and determine the best K value ((the one with the largest In Pr(X|K)). Program Clumpak (Kopelman et al., 2015) was used to visualize population structure across multiple cluster schemes (clustering under different K values).

SAMOVA v 2.0 (Dupanloup et al., 2002) was used to perform SAMOVA (Spatial Analysis of Molecular Variance) to define populations that were geographically homogeneous and most likely differentiated from each other. It also assisted in identifying significant genetic barriers between these differentiated populations for both the mtDNA and microsatellite datasets. Different scenarios with K values ranging from 2 to 12 were tested independently, and the K scenario with the highest F_{CT} index was considered the best scenario. The SAMOVA incorporates a spatial dimension when considering the best genetic clustering scheme.

4.2.7.2 Genetic structure, diversity and gene flow

For both the mtDNA and microsatellite datasets, Arlequin v 3.5 was used to perform an Analyses of Molecular Variance (AMOVA) to evaluate population genetic structure of the seven mtDNA clades. The molecular diversity Theta (H) was calculated to measure polymorphism and allele diversity across the seven clades for the different microsatellite loci. The pairwise F_{st} matrix, the average number of pairwise differences (π) within and between the seven clades, and Nei's distance matrix (Nei and Li, 1979) were also calculated.

To evaluate the magnitude of gene flow and genetic divergence between the seven clades, I computed Slatkin's linearized F_{st} matrix (Slatkin, 1995) and relative divergence time between clades, based on the F_{st} in Arlequin v 3.5. I also used Migrate v 3.2 (Beerli, 2009) to estimate the potential gene flow rate (M), effective population size (Θ) and magnitude of potential gene flow xNm (number of immigrants per generation, x=4 for nuclear data, x=2 for haploid data).

4.2.8 Phylogenetic and genealogical analysis

POPTREE2 (Takezaki et al., 2009) and Migrate v 3.2 were used to construct a microsatellite data based phylogenetic tree. I selected D_{sw} distance methods (Shriver et al., 1995) with neighbour-joining tree construction methods using 1000 bootstraps to run the analyses.

4.2.9 Hybridization

The R package 'poppr' (Kamvar et al., 2014) and 'adegenet' were used to investigate potential hybridization in potentially admixed individuals. The optimal number of PCs was determined in the same way as in the DAPC analysis. The threshold for detecting admixed individual analyses was set as less than 0.5 in predicted membership probability.

4.2.10 Inbreeding

In order to detect levels of inbreeding across different populations, namely, C1 (western and central), C1 (eastern), C2 (northern), C2 (southern), C3, C4, C5, C6 (eastern), C6 (western) and C7, I performed an inbreeding detection analysis using the R packages 'ape' (Paradis et al., 2014), 'pegas' (Paradis, 2010) and 'adegenet'.

4.2.11 Phylogenetic analysis and calibration dating

To compare the tree topologies generated from the three datasets, I first removed all sequences from alignments having outstanding microsatellite genotyping data to ensure that the mtDNA dataset completely matched the microsatellite DNA dataset. I then combined the trimmed mtDNA dataset with the nDNA dataset. I used BEAST v 2.4 (Bouckaert et al., 2014) and the

STACEY package (Jones, 2017), respectively, to carry out a Bayesian MSC phylogenetic analysis and generate a species tree. As the few new sequences generated in this study did not introduce any new haplotypes (except the three additional museum samples from Uniondale), I used the same methods and parameter settings as Chapter 2 for the Bayesian inference (BI), but added the nDNA dataset (using the same codon partitioning scheme, length of run and substitution model settings as Chapter 2, see Table S4.4). In addition, I also performed a maximum likelihood analysis (ML) using RAxML v.8 (Stamatakis, 2014). To evaluate the strength of support for each node, I used the same partitioning scheme as with the BEAST analysis, with 1000 nonparametric bootstrap replications. The ML algorithm model selected was GTRCAT.

To estimate the cladogenesis of the *P. tentorius* complex in terms of temporal dimensions and the timing of events for further analyses, I performed a Bayesian calibration dating analysis with BEAST v2.4. I sorted the partitions in the same way as in the above mtDNA+nDNA Bayesian inference (BI) by removing all identical sequences after combining the mtDNA and nDNA datasets, using DAMBE V5 (Xia, 2013). All the parameter settings were the same as for the BI (Table S4.4), but with calibration dating. Seven nodes were constrained as monophyletic clades and calibrated with time intervals obtained from published studies (See Table S4.5).

For the three new Uniondale museum samples, I was successful in obtaining 2 amplicons for the 12S gene, but failed to gain any positive result from the other genes. I thus performed independent BI and ML analyses only on the 12S alignment (including all samples from Chapter 2 and a few new sequences generated in this study). For the BI analysis ran BEAST v2.4 for 50 million generations, sampling every 5000 generations and discarding the first 10% as burn-in. Parameter settings followed the partition scheme and model test results (see Table S4.4) of Chapter 2, as I assumed that the addition of the new Uniondale samples was unlikely to influence the general parameter settings for the BI analysis. I used RAxML v.8 to perform the ML analysis with 100 nonparametric bootstrap replications under algorithm model GTRCAI.

4.2.12 Mantel tests

I performed Mantel tests to investigate the correlation between the mtDNA and microsatellite datasets in terms of genetic divergence and gene flow. The tests were performed on the average

number of pairwise distance matrices, Fst matrices and Slatkin's linearized Fst matrices, using GenAlex v.6. In order to test the correlation between genetic divergence and geographic distance, I used GenGIS v.2.5 (Parks et al., 2013) to perform Mantel tests on the correlation between geographic distances and pairwise Fst, and between the number of pairwise distances and geographic distances, on both the mtDNA and microsatellite datasets. For analyses performed under R v3.5 with R-package 'ade4' (Dstripe & Dufour, 2007), the species input tree for the mtDNA dataset generated from BI analysis was modified manually in notepad to remove excess support information before it was readable by GenGIS. Minor modifications were made to the microsatellite based species tree generated by POPTREES2 to enable its reading by GenGIS.

4.2.13 Bayesian simulations using the Approximate Bayesian Computations approach

In order to overcome the weaknesses of traditional likelihood based computational approaches which restrain evolutionary scenarios and mutation models to very narrow ranges, I used the Approximate Bayesian Computations (ABC) approach (Cornuet et al., 2008) implemented in DIYABC v 2.0 (Cornuet et al., 2014). This approach makes it possible to address much more complex and realistic biological situations using simulation methods. Comparing simulation results generated from different predefined scenarios using real datasets, enables detection of the most likely scenario from a wide range of scenarios, which cannot be tested with traditional likelihood based phylogenetic approaches. The ABC approach also allows the tracing of the past history of populations and species from different datasets. In this case, both sequence and microsatellite datasets were used to reconstruct the cladogenic history of the *P. tentorius* complex to improve its accuracy.

To determine the most likely genealogical scenarios, I tested eight scenarios using all the possible tree topologies generated by the phylogenetic analysis, namely, from the mtDNA and nDNA datasets (Chapter 2), the combined mtDNA + nDNA dataset (present study) and the microsatellite dataset (present study).

To define a proper prior for improving run accuracy, the effective population sizes (Ne) of the seven clades were estimated from the microsatellite dataset results generated by Migrate v 3.2. I decided to use the microsatellite dataset to estimate Ne priors, because it had Ne values of higher magnitude than those predicted from the mtDNA dataset. The divergence time priors (tau) were estimated from the combination of three possible topologies, the nDNA dataset tree

topology, BEAST dating results (six mtDNA loci with nDNA-PRLR), and the microsatellite based tree topology for different scenarios (see Fig. 4.1). The prior settings for tau were, tA > t5 > t4 > t3 > t2 > t1 > 0, t4p > t3, t5b < tA and t4q > t3 (see Fig. 4.2).

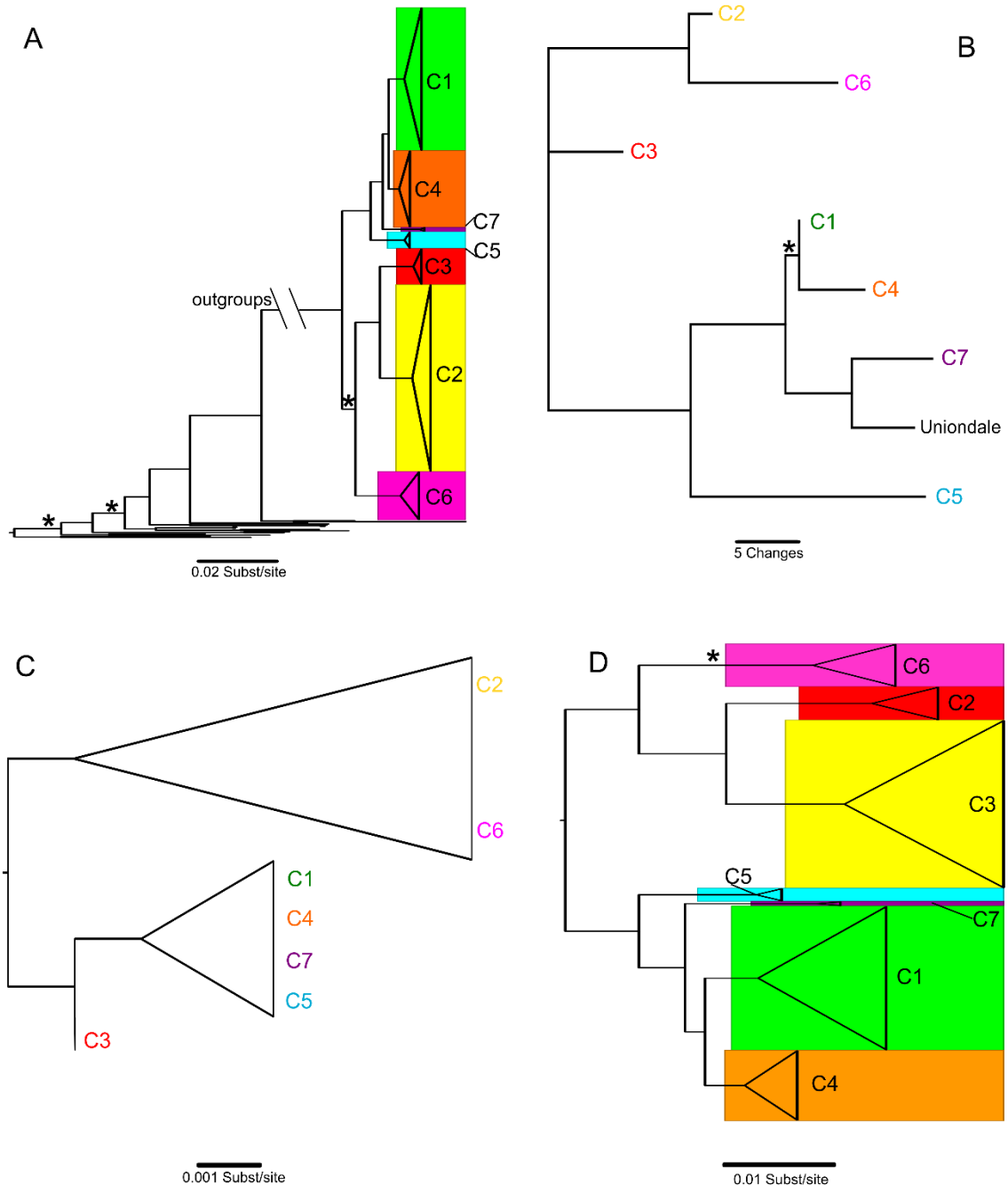


Figure 4.1. Summary of tree topologies generated from different analyses and datasets. A: the tree retrieved from BI and ML analyses with the mtDNA dataset; B: phylogenetic tree retrieved from the Dsw distance methods with a neighbour-joining construction approach using microsatellite DNA; C: tree generated from BI and ML analyses using nDNA-PRLR; D: phylogenetic tree generated from BI and ML analyses using the combined dataset (mtDNA+nDNA). The " * " indicates weak support (BP < 70 or PP < 0.95).

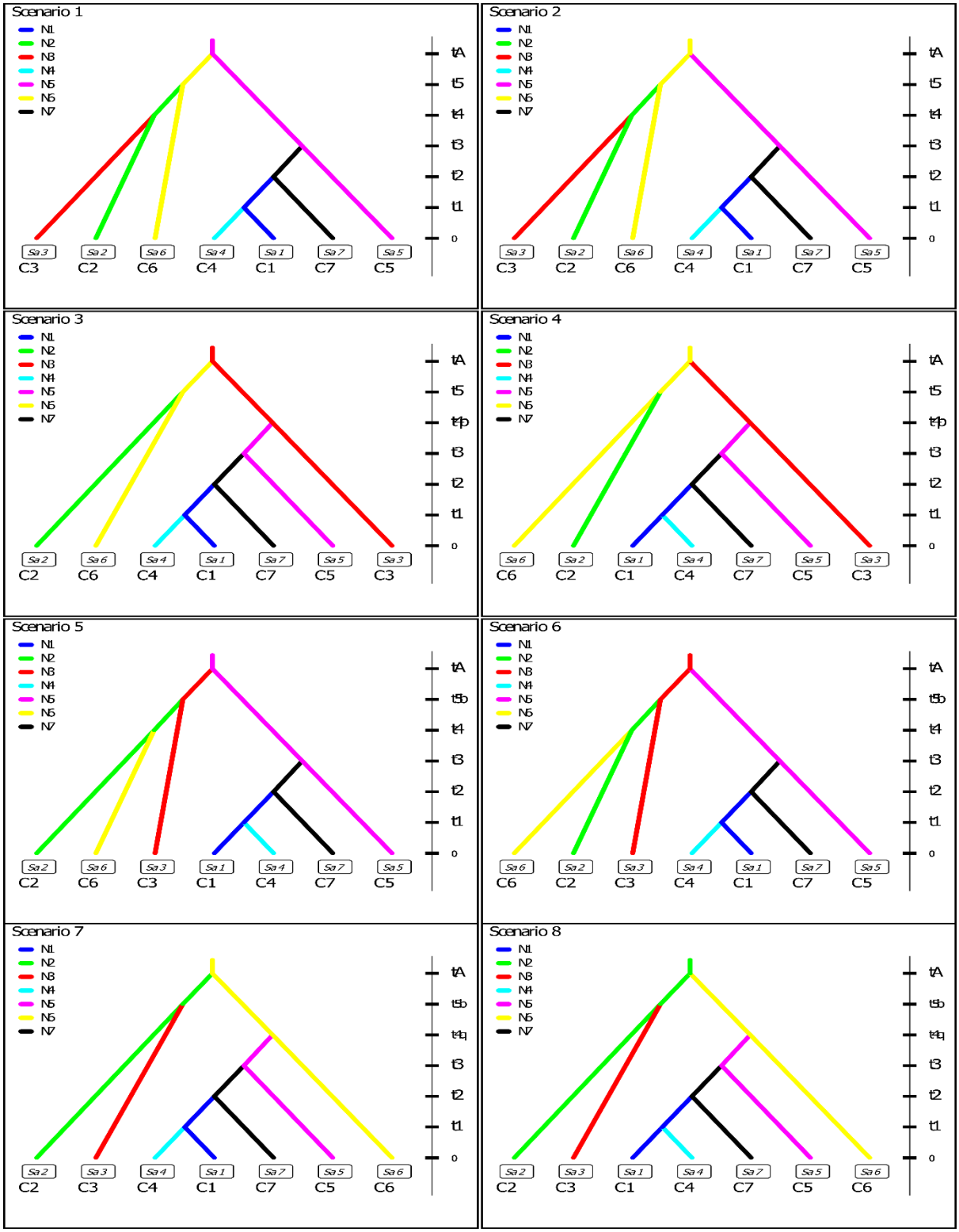


Figure 4.2. The eight possible scenarios used in the Approximate Bayesian Computation (ABC) analyses. Each of these scenarios was constructed based on different possible scenarios retrieved from different phylogenetic analyses with different types of molecular markers (e.g nDna, mtDNA and microsatellite DNA). Each pre-defined genealogical assumption was used in simulations run in ABC analysis. In each pre-defined scenario, the clade names for corresponding branches are given below them.

I divided the dataset into two groups, the microsatellite dataset and the sequence dataset, each with independent mutation models. The priors for the sequence dataset group were specified as, mean mutation = 0.01 per site/per generation/million years, with range 1.0×10^{9}- 1.0×10^{7} based on the evolution rate for the gopher tortoise (*Gopherus agassizii*), which has been estimated as half the general mtDNA evolution rate of vertebrates, this is, 2% per myr (Wilson et al. 1985). The mutation model used was HKY+I+G with invariant sites = 72% and shape of gamma = 0.75, which followed suggestions of the model test results. I followed the suggested default settings for the prior setting scheme in the microsatellite group. In the sequence dataset group, I chose the number of haplotypes, the number of segregating sites, the mean pairwise differences and Tajima's D for the one sample summary statistics, and the mean pairwise differences (W) and F_{st} (Hudson et al., 1992) as objects for the two sample summary statistics. For the microsatellite dataset group, I selected mean number of alleles, mean genetic diversity and mean size of variance for the one sample summary statistics, and mean genetic diversity and F_{st}, classification index, shared allele distance and $(d_{\mu})2$ distance for the two sample summary statistics.

I simulated 9.0×10^{6} datasets in the initial run following the system suggested for the optimal number of runs for obtaining both computationally and statistically robust results, before further processing using the ABC analysis.

For the final summary analysis of the simulation results, all 8 million simulated data sets in the reference table were used. I first performed a pre-evaluative prior-scenario combination analysis to evaluate whether the observed data fell into the simulated prior range of each scenario, using a preliminary Principal Component Analysis (PCA). If the observed dataset fell into the simulated prior scenario range, it meant that the simulated results were reliable and that sampling was adequate and even. I then computed the posterior probability of each scenario in order to determine the most likely scenario, using Linear Discriminant Analysis. The results were plotted to visualize and rank the posterior scores of the different simulated scenarios. To visualize and quantify the posterior distribution of the different simulation parameters, as well as to ensure prior settings were adequate, I estimated the posterior distribution of each parameter and plotted it independently. Finally, I performed a model checking analysis on each simulated scenario to ensure that the observed dataset fell into the range of priors and posteriors of the simulated results of each PCA scenario.

4.3 Results

The 404 tissue samples used in this study of *P. tentorius* complex, which occurs at very low densities throughout its distribution range, was collected over a 20 year period and covers its entire distribution range in southern Africa. The sampling coverage was therefore regarded as adequate for the purpose of the study. All the 404 samples were genotyped using nineteen microsatellite DNA loci as well as six published mtDNA loci and one nDNA locus.

4.3.1 Microsatellite DNA amplification, sequencing and genotyping

The coverage of the mtDNA sequencing was generally good (see Table S4.1), however, I failed to amplify nDNA-PRLR for the majority of carcass samples.

The microsatellite DNA genotyping results showed that 3 of the 21 loci of the *P. tentorius* complex failed to amplify during PCR, namely, GP19, GP30 and GAL75, while the rest of the loci gave positive results (Table S4.6). Despite repeated PCR amplification efforts, the loci RAD932, Maucas22, TEST10, GAL100 and GAL127 showed comparatively large numbers of missing alleles in certain individuals (Table S4.6). The amplifications performed poorly for loci GAL100 and GAL127 in C6, particularly in its eastern population. The microsatellite DNA motif of the 19 loci were determined (details given in Table S4.3). The Micro-Checker analysis results revealed significant null allele signals for loci RAD932, Maucas22 and Test10 ($p < 0.001$ in all cases). The analyses failed for the GAL100 and GAL127 loci, as both had too many missing alleles. No significant signals of potential stuttering peaks and large allele dropout were detected during the analyses. The preliminary DFA analyses indicated that the scatterplot pattern did not change significantly between the 19 loci dataset and 14 loci dataset (see Table S7, with loci RAD932, Maucas22, TEST10, GAL100 and GAL127 removed). The allele range plots (Figure S4.1-4.2) and allele diversity plots (Figure S3) both revealed that these five loci are generally less informative than the rest. I therefore removed them from the dataset before the final analysis commenced. Collectively, fourteen microsatellite markers were ultimately used for carrying out further analyses.

4.3.2 Microsatellite DNA allele range

Among the 14 microsatellite DNA loci used in the analyses, loci 15HDZ20, 15HDZ239, 58HDZ242, GAL45, GP96, Test56 exhibited higher levels of polymorphism and were phylogenetically more informative (Figure S4.1-4.2) when compared to the phylogenetic results of Chapter 2.

4.3.3 Assignment test on unknown individuals

Of the six samples of unknown origin, one, most likely originated from De Hoop Nature Reserve and the mtDNA analysis assigned it to C1 (p = 0.38); one from Van Wyksvlei and one from Louisvale were assigned to C2 (p = 0.25 and 0.31, respectively); one from Nieuwoudtville and two from Calvinia, were assigned to C4 (p = 0.97; p = 0.32 and 0.27, respectively). These assignment test results were applied in the subsequent analyses.

The Arlequin assignment test assigned two samples from the Uniondale area to C1, rather than to C5 or C7 (Figure S4.4).

4.3.4. Population genetic analysis

4.3.4.1. Genetic structure

In the DAPC analysis, the BIC criterion suggested four clusters (K = 4) as optimal clustering scheme (Figure S4.5). The a-score optimisation retrieved fourteen PCs as optimal number for further DAPC analysis (Figure S4.5). The cross-validation test confirmed that the number of PCs retrieved for the analyses was sufficient (Figure S4.5). The DAPC scatterplot (Fig. 4.3A) revealed four significant clusters, cluster 1 consisted of C1 (western and eastern populations), C4, C5 and C7, cluster 2 comprised C3, cluster 3 included the eastern and western populations of C6, while cluster 4 contained the northern and southern populations of C2. Discriminant function 1 showed a clear separation between the two major branches “C1+C4+C5+C7” and “C2+C3+C6” (Fig. 4.3B). The DAPC membership cluster analysis retrieved 6 significant clusters as optimal membership clustering scheme (Fig. 4.4), with only the difference between C1 and C4 not being significant. The Uniondale samples were predictied as a unique cluster.

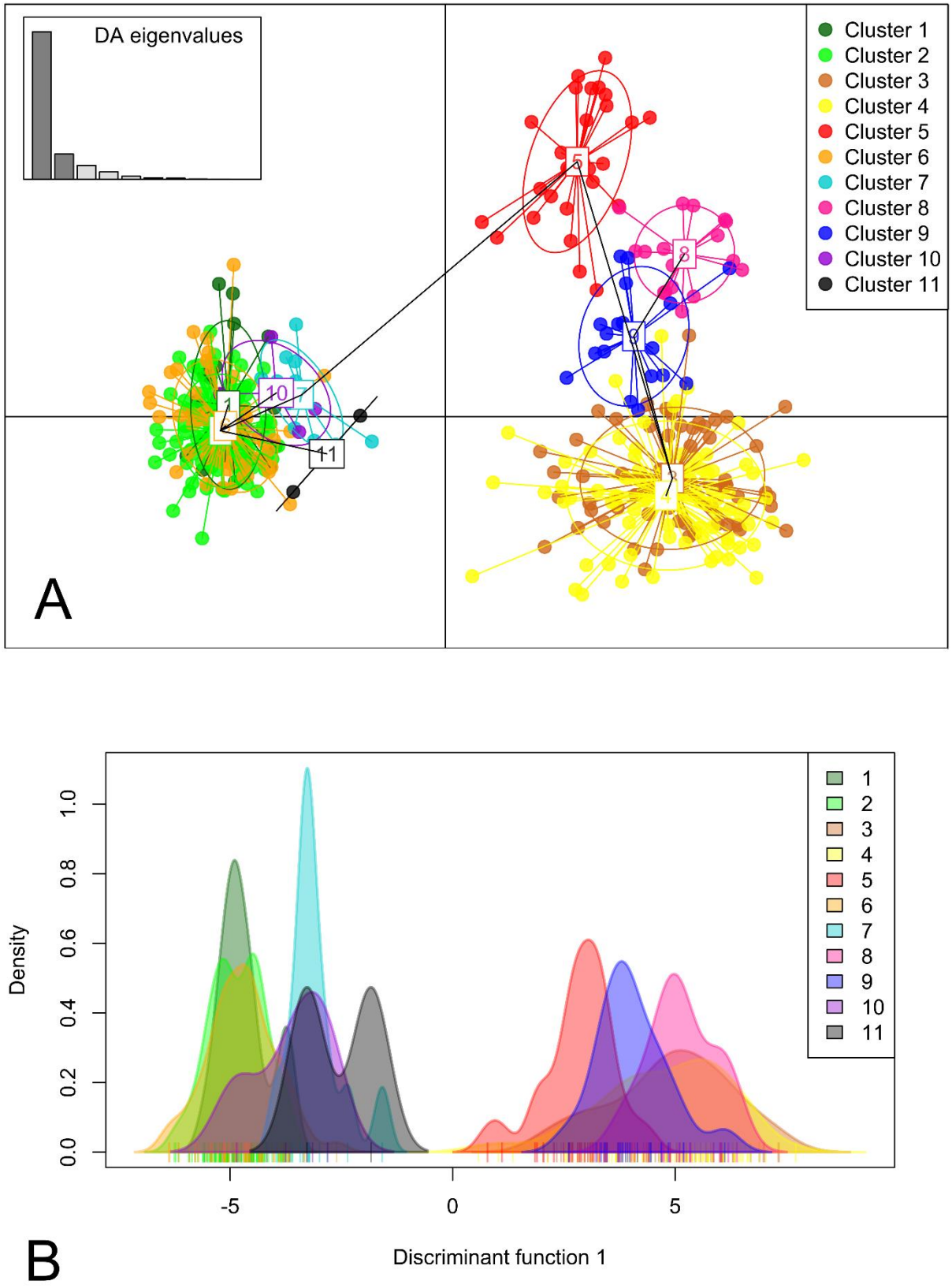


Figure 4.3. A: Scatterplot from DAPC analysis among 11 clusters with 14 microsatellite DNA loci, with the eigenvalues of the first ten discriminant functions also given; B: The distribution of discriminant function scores densities among the 11 clusters in discriminant function one. Cluster 1: C1 (eastern), Cluster 2: C1 (western), Cluster 3: C2 (northern), Cluster 4: C2 (southern), Cluster 5: C3, Cluster 6:

C4, Cluster 7: C5, Cluster 8: C6 (eastern), Cluster 9: C6 (western), Cluster 10: C7 and Cluster 11: Uniondale population.

STRUCTURE results generally supported the species level analysis of Chapter 2, showing 5 clusters that corresponded to the five putative species of the optimal clustering scheme (at K = 6 in all cases) for both datasets partitioned by clades and subclades. The difference between C3 and C6 was, however, not clear (Fig. 4.4). Furthermore, the STRUCTURE results failed to distinguish the differences between C1 and C4, and C5 and C7. The differences between C3 and C6 were nonetheless significant for scenarios K = 7, K = 8, K = 9, K = 10 and K = 11 (results not shown). The two Uniondale samples were generally poorly predicted.

4.3.4.2 Spatial genetic structure

The SAMOVA generally suggested 6 clusters (K=6) as the optimal spatial-scale genetic structure for the microsatellite DNA dataset (Fig. 4.5A), since the over-split event was at its minimum level in each clade, though the FCT value was not the best at K=6 (Table S4.8). In the optimal clustering scheme, the eastern but not the western population of C6 differed significantly from C2, while C1 and C4 did not differ from each other. The spatial clusters C3, C5 and C7 perfectly matched their mtDNA clades.

For the mtDNA dataset, the SAMOVA also retrieved 6 clusters (K=6), but the clustering allocations differed from that of the microsatellite DNA dataset (Fig. 4.5B). No significant difference was found between C1 and C4, while the rest of the clades were significantly different as distinct spatial clusters.

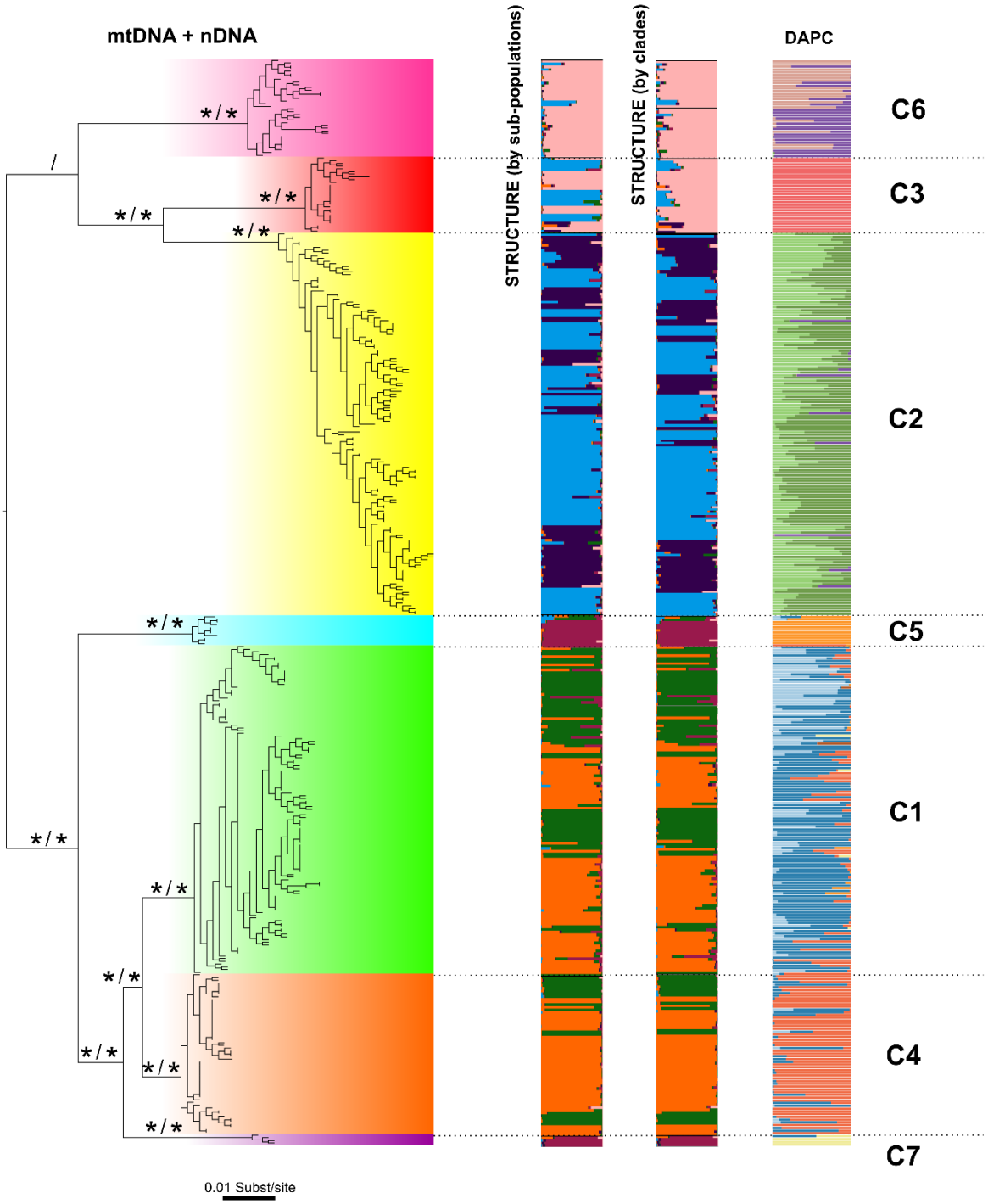


Figure 4.4. Phylogenetic tree retrieved from Bayesian inference (BI) and Maximum likelihood (ML) analyses with the mtDNA+nDNA dataset. The tree topology was generated from ML analysis. Nodes with " * " indicate strong support (BP > 70, PP > 0.95). The clustering results from STRUCTURE analyses (based on sub-populations and clades) and DAPC analysis with 14 microsatellite DNA loci are shown to the right side of the tree.

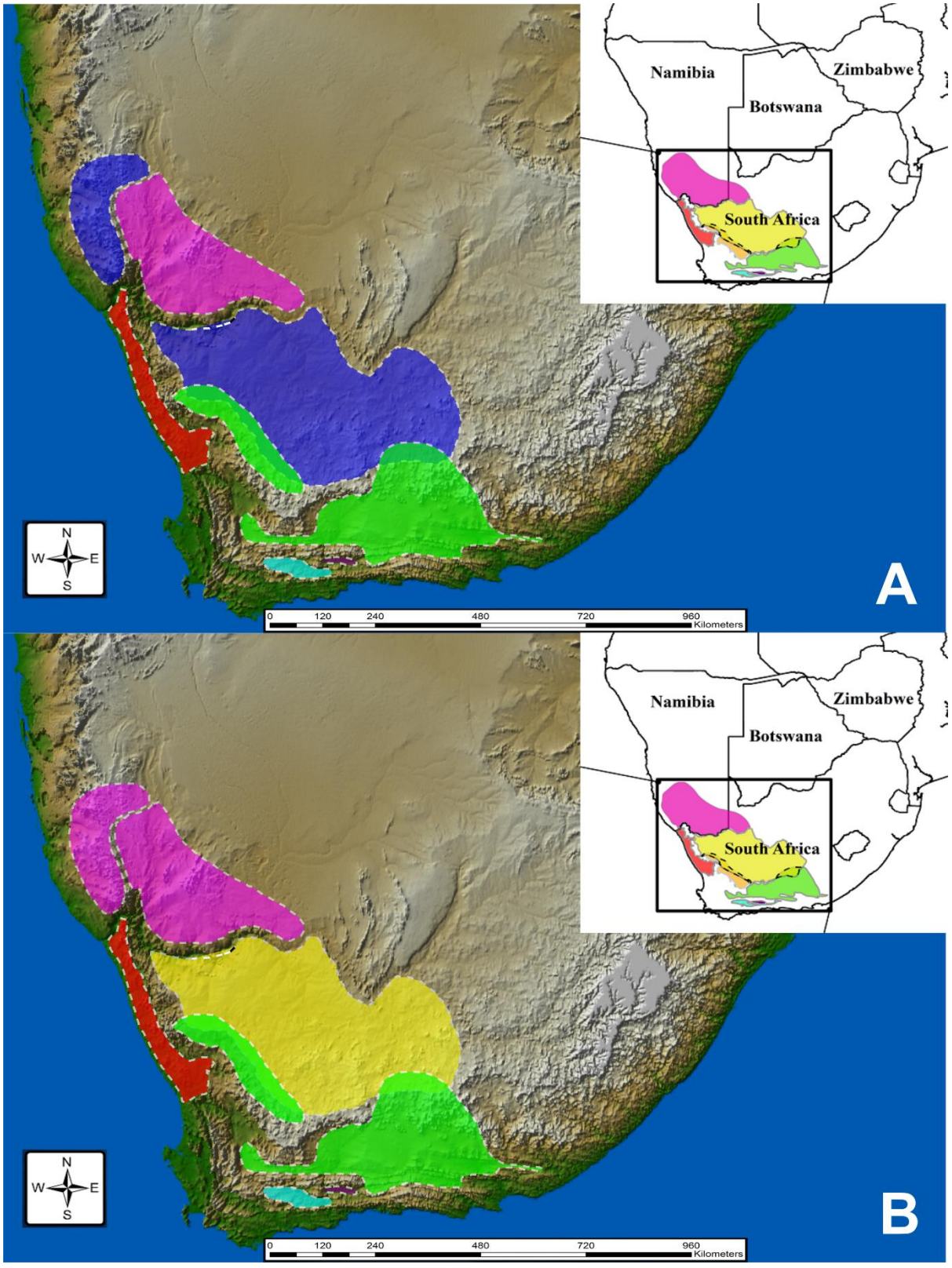


Figure 4.5. A: The optimal clustering scheme advocated by the SAMOVA analyses using the microsatellite DNA dataset; B: the optimal clustering scheme advocated by the SAMOVA analyses using the mtDNA dataset.

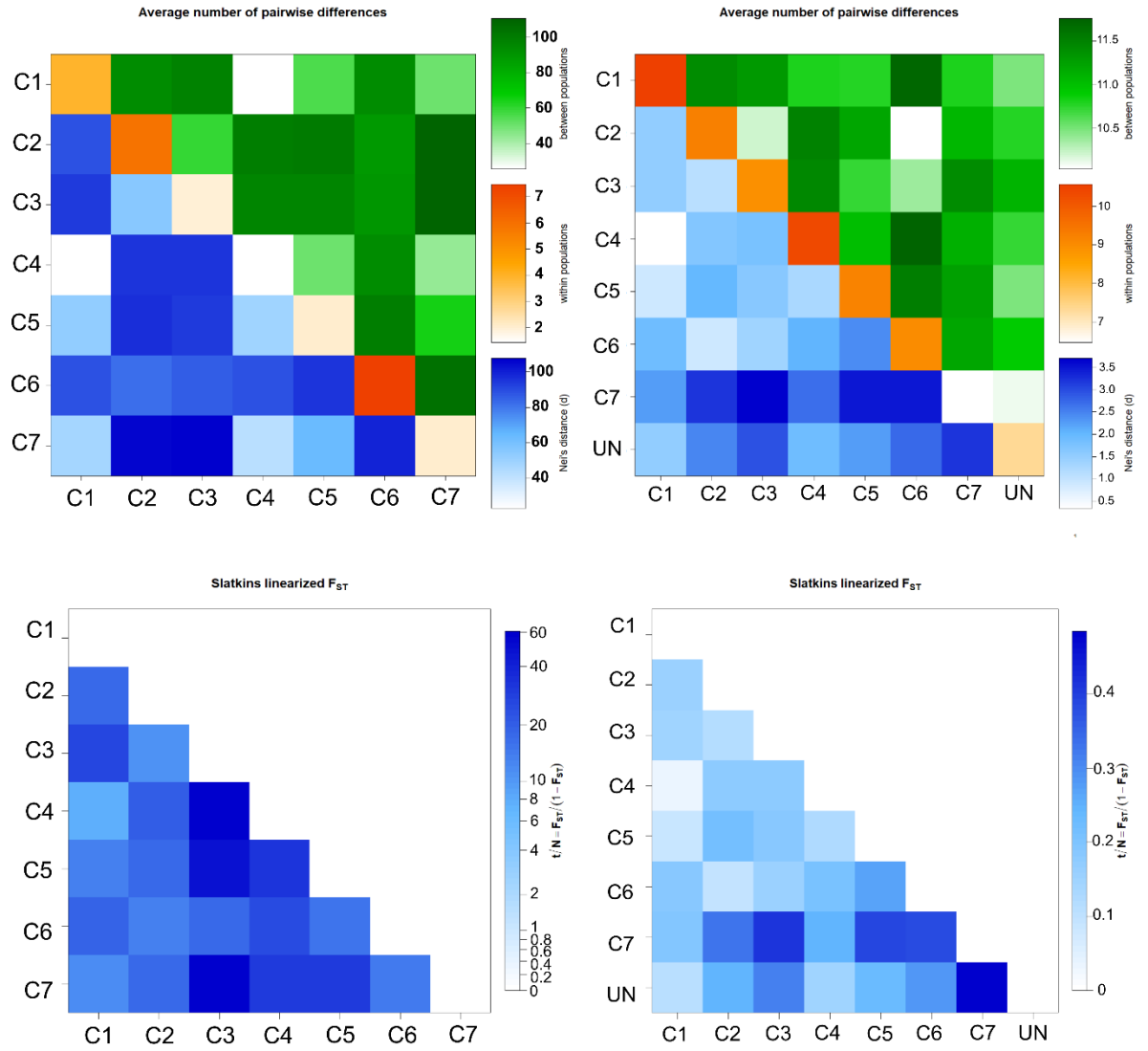


Figure 4.6. Top left: the average number of pairwise differences among and within different clades from the AMOVA-derived analyses using the mtDNA dataset; Top right: the average number of pairwise differences among and within different clades from the AMOVA-derived analyses using the microsatellite DNA dataset; Bottom left: the gene flow indicator-Slatkins linearized F_{st} matrix among different clades from the AMOVA-derived analyses using the mtDNA dataset; Bottom right: the gene flow indicator-Slatkins linearized F_{st} matrix among different clades from the AMOVA-derived analyses using the microsatellite DNA dataset. "UN" refers to the Uniondale population.

4.3.4.3 Genetic diversity and gene flow potential

AMOVA results of the microsatellite DNA dataset revealed that 12.38% of the total genetic variation was from among the seven clades (Va: df=6, sum of squares=453.42 and variance components=0.69), 23.95% of the total variation came from within each clade (Vb: df=395, sum of squares=2458.02, and variance components=1.33) and 63.67% of the variation came

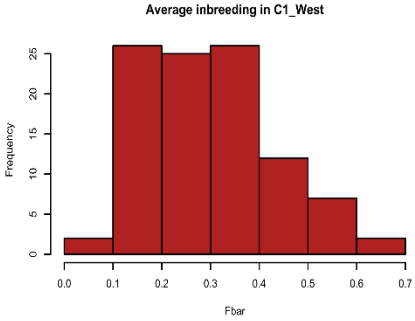
from within individuals (Vc: df=402, sum of squares=1427.5, and variance components=5.58). The fixation indices FIS, FST and FIT were 0.27, 0.13 and 0.36, respectively. The standard genetic diversity of microsatellite DNA was highest in C1 and C4, and lowest in C7 (Table 1). The genetic diversity index, Theta (H), also revealed the highest genetic diversity in C1 and C4, and the lowest in C7 (Table 4.1). The same trend was found in the allele diversity. The matrix of the average number of pairwise differences suggested a major divergence between the two major groups, "C1+C4+C5+C7" and "C2+C3+C6" (Fig. 4.6). In "C1+C4+C5+C7", the lowest pairwise difference was found between C1 and C4, and the highest between C5 and C7, whilst, in "C2+C3+C6", the highest difference was found between C3 and C6, and the lowest between C2 and C6. The highest within-group pairwise distance was found in C1 and C4. The Uniondale population were closest to C1 (Fig. 4.6). The Slatkins linearized FST matrix, tau matrix and FST matrix generally showed identical trends to the pairwise distance matrix. At population level (Table 4.1), the eastern population of C1 exhibited much lower genetic and allele diversity than the western population. The genetic diversity level between the northern and southern populations of C2 did not show a great difference, the southern population having slightly higher genetic diversity (Theta, H) than the northern population. The genetic diversity between western and eastern populations of C6 was almost identical.

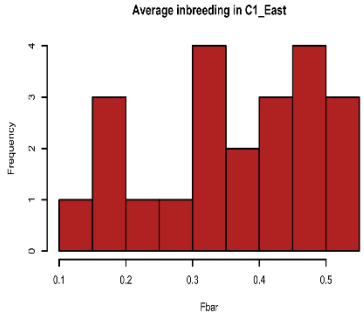
The Arlequin genetic diversity analyses' results of the mtDNA dataset were identical to the findings of Chapter 2. It generally found the highest genetic diversity in C1, C2 and C6, and the lowest in C3 and C4 (Table 4.1). The mtDNA dataset also detected a major divergence between "C1+C4+C5+C7" and "C2+C3+C6" across the different matrices (Fig. 4.6). At population level, the eastern population of C1 showed much lower genetic diversity than the western population. In C2 the genetic diversity was higher in the southern population and in C6 the eastern population exhibited significantly greater genetic diversity than the western population.

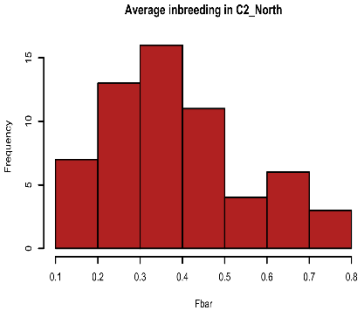
In the microsatellite DNA dataset, the Migrate analysis results suggested that the effective population size (Theta, H) was similar across the seven clades. We therefore did not differentiate between the prior settings for the different Theta values among the clades in the ABC simulation analysis. The posterior distribution plots for the different migration parameters were generally normally distributed, and no significant "double peak" were observed (not shown). In terms of gene flow, the highest rate was found between C1 and C2, and C1 and C4, whilst the lowest was found between C1 and C5, C1 and C7, and C5 and C7. The gene flow

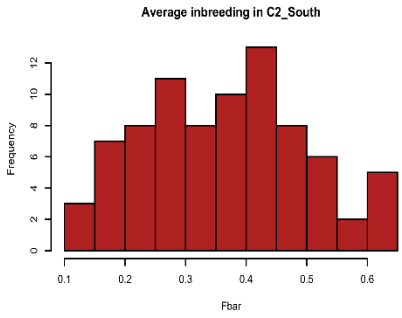
rate between C2 and C4, and between C2 and C6 was comparatively high, but comparatively low between C3 and C6. The rate between C2 and C3 was intermediate.

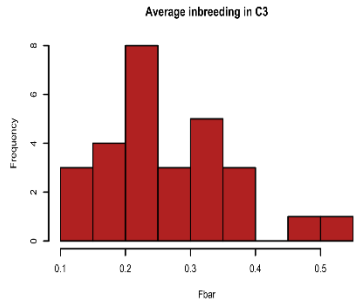
Recent migration events shown by time plots (results too long, not shown) were discovered between C1 and C2, C1 and C4, C3 and C4, C4 and C5, and C4 and C7, but not between C2 and C6, and C3 and C6.

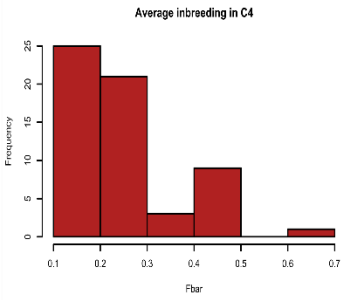


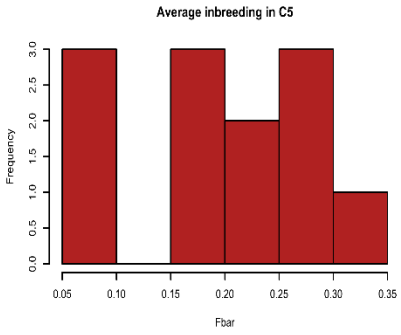


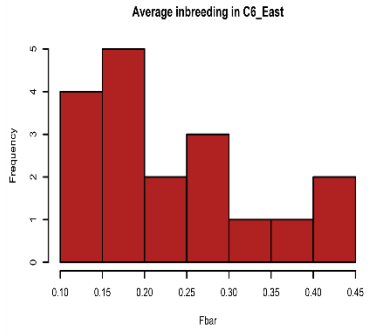


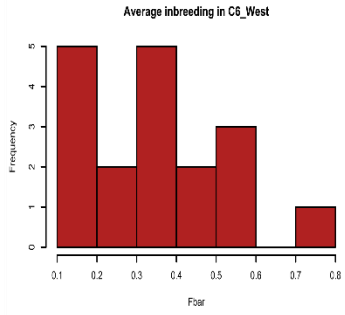












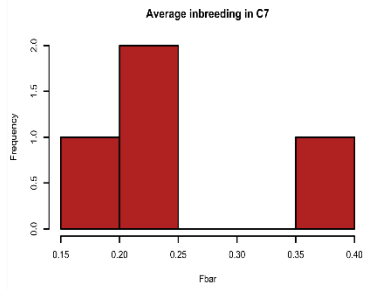


Figure 4.7. Histograms of inbreeding scores among different populations. Frequency bars above 0.4 were considered as groups showing high inbreeding levels.

4.3.4.4 Hybridization (between the putative species)

The DAPC admixture analysis detected 10 suspected hybridized individuals as potential hybridizations between the seven putative species (Figure S4.6), seven of which came from the Victoria West and Richmond area, and suggested as hybridizations between C1 and C2. The remaining three were assigned as crossovers among C4, C5 and C7.

4.3.4.5 Inbreeding

The eastern population of C1 had a significantly high inbreeding signal, with the majority of samples falling into a group with an inbreeding coefficient > 0.4 (Fig. 4.7). A low level of inbreeding was found in the southern population of C2 and the western population of C6.

4.3.5 Phylogeny and calibration dating

Both ML and BI analyses results for the combined dataset (mtDNA+*PRLR*) generated a tree topology similar to the mtDNA based phylogenetic tree of Chapter 2, except for some differences in branch lengths (Fig. 4.1, Fig. 4.4). The different nodes were generally well supported, (BP > 70, PP > 0.95), except the node between (C2+C3) and C6. The mtDNA+nDNA tree topology also revealed low support values for node C6+(C2+C3) (BP < 70, PP < 0.95). These finding were congruent with results from a previous study (Chapter 2).

The microsatellite DNA based phylogeny (Fig. 4.1B) showed a similar tree topology to the mtDNA and combined (mtDNA+nDNA) tree, except that the relationship between C2, C3 and C6 was different. Clade 2 and C6 formed a sister group, and the relationship between C3 and (C2+C6) was not supported as C3 collapsed into an unique branch resulting in a tree polytomy. The relationships among the three major branches, C2+C6, C3 and (C1+C4)+C7)+C5 were therefore not clear, though C3 was shown as the clade closest to the basal part of the tree. The topology of (C1+C4)+C7)+C5 was nonetheless well supported in the microsatellite based phylogeny, which was identical to the mtDNA and combined (mtDNA+nDNA) tree. Each node

had strong support (BP > 70), except the one between C1 and C4. The Uniondale samples formed a sister clade with C7.

In the single locus 12S based phylogenies both ML and BI approaches suggested C1 as the sister group to the Uniondale samples (Figure S4.7) with strong support (BP > 70, PP > 0.95).

The BI calibration dating results (Fig. 4.8) based on the combined dataset generally suggested that the radiation of the genus *Psammobates* started about 34.63 Ma (late Eocene), while the start of the cladogenic events within the *P. tentorius* complex, dated back to the Mid-Miocene (about 13.82 Ma). The divergence time between C2+C3 and C6 was given as 12.07 Ma, and between C2 and C3, as 8.06 Ma. The branch-off time between C5 and (C1+C4)+C7 was dated as 7.5 Ma, which was similar in age to the node between C2 and C3, but slightly more recent. Clade 7 split from C1+C4 about 5.42 Ma, while C4 diverged from C1 about 3.88 Ma. The dating results of outgroup nodes exhibited high congruence with the calibration dating results of Hofmeyr et al. (2017).

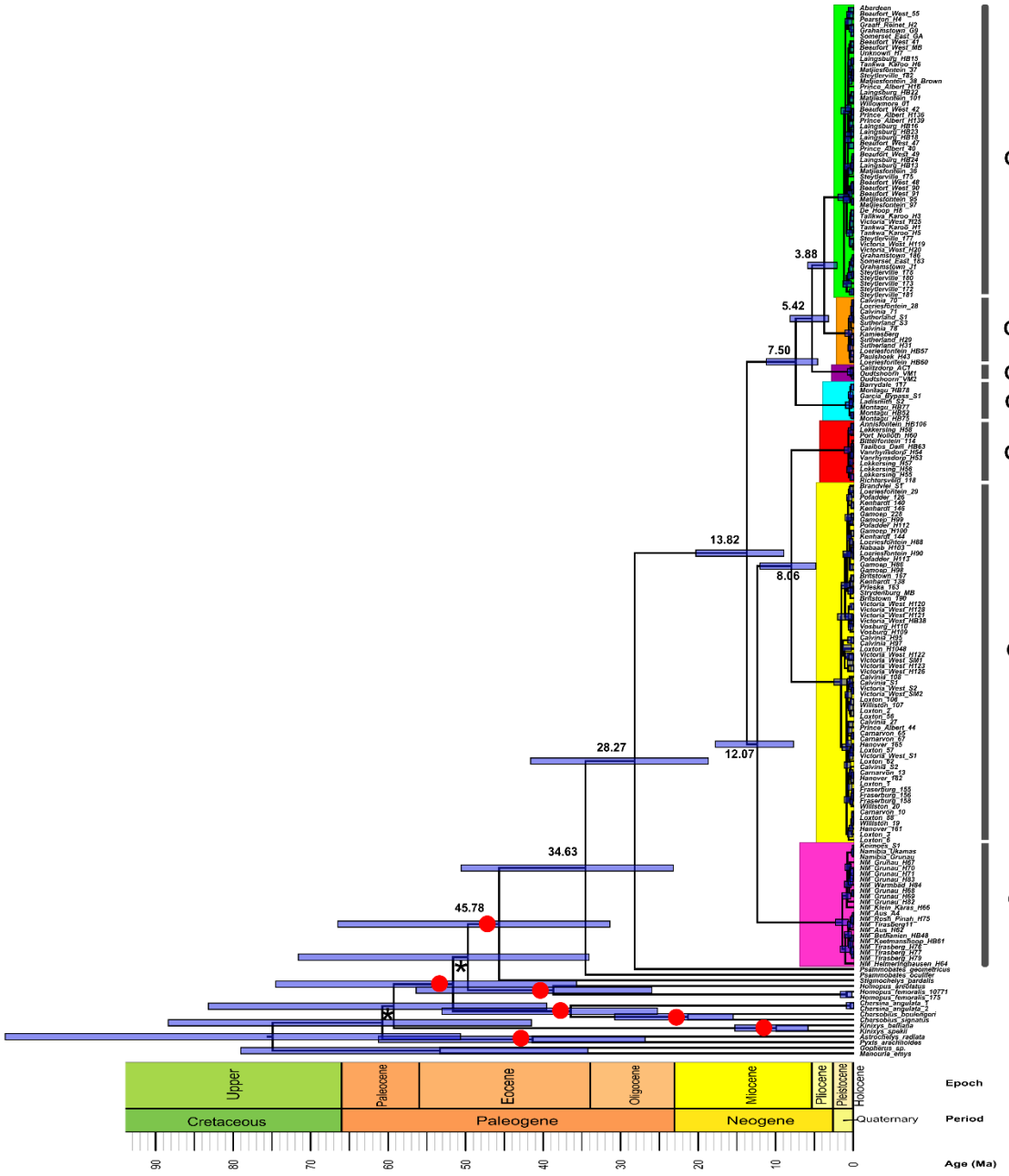


Figure 4.8. Chronogram generated from BEAST BI analyses with the mtDNA+nDNA dataset. The seven clades are indicated in the tree. The red dots in the tree represent the calibration nodes used in priors setting in calibration dating analysis.

Table 4.1. The Arlequin genetic diversity analyses results showing sample size (N), polymorphic sites, number of haplotypes (no. Haplotype), gene diversity, genetic diversity (Theta, H) and number of alleles (no. of alleles) among different populations and clades, using the mtDNA and microsatellite DNA datasets.

mtDNA					
population/clade	N	polymorphic sites	No. Haplotype	gene diversity	Theta (H)
C1	123	39	52	0.95	17
C1_western	104	28	38	0.93	14
C1_eastern	19	11	14	0.71	2
C2	151	52	64	0.97	33
C2_northern	79	25	22	0.91	10
C2_southern	72	30	33	0.95	18
C3	32	13	11	0.68	2
C4	63	9	12	0.59	1
C5	12	10	7	0.83	4
C6	38	38	22	0.95	17
C6_western	19	19	9	0.87	8
C6_eastern	19	19	12	0.9	16
C7	4	4	3	0.83	4
overall	423	288	171	0.98	49

Microsatellite DNA						
population/clade	N	polymorphic loci	No. Haplotype	gene diversity	Theta (H)	No. allele
C1	122	14	244	0.75	3.22	26
C1_western	100	14	200	0.76	3.21	25
C1_eastern	22	13	44	0.58	2.46	10
C2	141	14	282	0.66	2.05	22
C2_northern	60	12	120	0.61	2.02	15
C2_southern	81	14	162	0.67	2	17
C3	28	14	56	0.64	1.77	9
C4	59	13	118	0.75	3.17	17
C5	12	12	24	0.65	1.99	6
C6	36	13	72	0.64	1.8	13
C6_western	18	13	36	0.62	1.65	9
C6_eastern	18	12	36	0.61	1.56	9
C7	4	11	8	0.51	0.92	2
mean	402	14	804	n/a	2.04	12

Table 4.2. The Mantel test results between different matrices with corresponding *p*-values and R-square values; "NS" denotes not significant and " * ", " ** ", " *** " represent $p < 0.05$, $p < 0.01$ and $p < 0.001$, respectively.

Comparison	*p*-value	R-square
Distance vs Pairwise differences (mtDNA)	NS	0.23
Distance vs Fst (mtDNA)	NS	0
Distance vs Pairwise differences (microsatellite DNA)	NS	0.03
Distance vs Fst (microsatellite DNA)	NS	0.05
Pairwise differences (microsatellite DNA vs mtDNA)	*	0.17
Fst (microsatellite DNA vs mtDNA)	*	0.24
Fst vs Pairwise differences (microsatellite DNA)	* *	0.98
Fst vs Pairwise differences (mtDNA)	*	0.21

4.3.6 Mantel tests

The Mantel test results found no significant correlations between genetic divergence among the seven clades and geographic distance. The results between geographic distance and pairwise differences, and between geographic distance and Fst were also not significant for both the mtDNA and microsatellite DNA datasets, $p > 0.05$ in all cases (see Table 4.2). Notwithstanding this, significant correlations were found for pairwise genetic differences between mtDNA and microsatellite DNA ($p < 0.05$, $R^2 = 0.17$), and FST matrices between microsatellite DNA and mtDNA ($p < 0.05$, $R^2 = 0.24$). Significant correlations between FST and pairwise genetic differences were also detected in both the microsatellite DNA ($p < 0.01$, $R^2 = 0.98$) and mtDNA ($p < 0.05$, $R^2 = 0.21$) datasets.

4.3.7 Bayesian simulations of alternative scenarios

The results of the ABC simulation analyses advocated scenario 4 as the most likely scenario for explaining the real genealogical history of the *P. tentorius* complex (Figure S4.8), rather than the expected scenario 1 or 2, which were well supported by the previous mtDNA and mtDNA+nDNA-*PRLR* based phylogeny (Chapter 2). The ABC simulation results were closer to the tree topology retrieved from the nDNA-*PRLR* dataset and microsatellite DNA loci. This suggested C6 as the most likely ancestral lineage, and advocated that C2 and C6 were sister

groups and clustered as a major branch (the *P. t. verroxi* complex). Clade 3 (*P. t. trimeni* complex) clustered with all the *P. t. tentorius* lineages (C1+C4)+C5)+C7, forming a second major branch. The PCA results of priors (Figure S4.9) and posteriors (Figure S4.10) revealed that the real observed dataset fell completely into the range of simulated datasets for all scenarios, which implies adequate sampling with a high level of confidence in its reliability. When comparing the parameter plots between priors and posteriors, I generally found good matches between their distributions (Figure S4.11). It therefore showed that the sampling during simulation was adequate and the simulation results, reliable.

4.4 Discussion

Overall, the analysis results in this section support "four-species" as the most parsimonious clustering scheme. All analyses indicated no significant differences between C1 and C4, and some analyses exhibited diagnosable differences between C5 and (C1+C4+C7). The "four candidate species" are: 1) C1+C4+C7+C5, 2) C3, 3) C2 and 4) C6. The inbreeding analysis results suggested a substantial degree of inbreeding in the eastern population of C1, while a hybridization signal was detected between C1 and C2. This confirmed the findings and conclusions made in Chapter 2, showing that cross-over between putative species C1 and C2 was possible.

4.4.1 Is microsatellite DNA useful for delineating taxa?

The mutation mechanism in microsatellite DNA (changing the length of the variable number of repeat motifs) differs from that of mtDNA and nDNA, and has lower coverage in the genome (Bagshaw, 2017). Microsatellite DNA markers have been shown to be phylogenetically informative and useful in species delimitation analyses (Lumley & Sperling, 2011; Fritz et al., 2012; Vanhaecke et al., 2012; Lagostina et al., 2018), particularly when cryptic species and complex hybridization are involved (Postaire et al., 2016; Pöschel et al., 2018). Insights gained from DNA sequence and microsatellite DNA data may nevertheless not always be fully congruent with each other (Fritz et al., 2012; Feng et al., 2016; Postaire et al., 2016).

Nonetheless, few studies have investigated the utility of variation in allele size across different microsatellite DNA loci in species delimitation. In this study, I found that allele size in many loci was actually very informative and useful evidence for delineating the seven clades, with only a few loci exhibiting no phylogenetic signal (Figure S4.1 – 4.2).

4.4.2 Does the Uniondale population represent a distinct clade?

The single 12S based phylogeny (Figure S4.7) implied that the Uniondale population is a close relative of C1, rather than C5 or C7 from the adjacent localities, which is not completely unexpected. Firstly, the geographic barrier between the northern distribution range of C1 and the Uniondale population is the eastern section of the Swartberge, which is much lower than its western section. The eastern section has kloofs that may allow limited gene flow between the northern distribution range of C1 and the Uniondale population. Nonetheless, it is not clear whether the Uniondale population should only be considered as a subclade within C1, or as a distinct lineage and OTU, given the limited sequence data with a single marker, obtained from museum specimens. Further studies with fresh samples and more genes are therefore required to clarify its phylogenetic position. When comparing the pairwise distances for the 12S gene, the genetic distance between the Uniondale population and C1 was smaller than the distance between C1 and C4. Both this section and Chapter 2 suggest that C1 and C4 should not be regarded as different OTU's or species. The Uniondale population should for now therefore only be considered as a subclade of C1, rather than a distinct lineage or OTU.

The microsatellite DNA analysis, however, did not generate results congruent with the 12S gene, instead, it suggested that the Uniondale population clustered with C7 as sister group. Since the sample size for both C7 and the Uniondale population was very small, the microsatellite results are probably not be reliable, as more samples for both are needed to verify their relationship. Despite the incongruence between the 12S and microsatellite DNA based phylogenetic findings, the microsatellite DNA based assignment test results (Figure S4.4), together with the pairwise differences matrix (Fig. 4.6), Slatkins linearized Fst matrix (Fig. 4.6) and Fst matrix (Figure S4.12) all advocated that C1 was closer to the Uniondale population, than to C5 or C7. In the relative divergence time tau matrix (Figure S4.12), it is clear that the estimated relative divergence time between C1 and the Uniondale population was the smallest of all, which again suggests that the Uniondale population is closest to C1.

4.4.3 DAPC genetic structure and clustering, STRUCTURE clades and subclades, and SAMOVA

The microsatellite DNA based DAPC analyses retrieved 4 significantly distinct cluster groups (Fig. 4.3A), cluster one: C1 (eastern and western population), C4, C5, C7 and the Uniondale population, cluster two: C3, cluster three: C6 (eastern and western population), and cluster four: C2 (northern and southern population). We found no significant differences between C1 (western), C1 (eastern) and C4, but there was detectable but small differences between "C1+C4", C5, C7 and the Uniondale population. The microsatellite DNA results showed a similar clustering pattern to the multivariant analysis results of the mtDNA dataset (see Figure S2.5 in Chapter 2). There was, however, no clear separation between C1 and C4 in the microsatellite DNA analysis. Since the 12S based phylogenetic results of the DAPC analysis did not support a large divergence between C1 and the Uniondale population, their separation by the microsatellite DNA dataset may be due to under-sampling and a lack of genotyping results for some loci of the Uniondale population. The STRUCTURE analyses generated similar results to the DAPC, though C5 and C7 did not show clear separation (Fig. 4.4). The SAMOVA of the microsatellite DNA dataset failed to distinguish C2 from C6, which was unexpected. Clear differences between C6 (western) and C6 (eastern) from the DAPC analysis and the mtDNA based phylogeny (see Chapter 2), suggests that further radiation between C6 (western) and C6 (eastern) is likely, and could lead to further cladogenesis. The phylogenetic results of the mtDNA+nDNA combined dataset generally showed identical results to the mtDNA based phylogeny of Chapter 2, but conflicting tree topologies with the nDNA dataset, respectively (Chapter 2). The phylogenetic congruence between the combined mtDNA+nDNA and the mtDNA dataset may be an artefact resulting from the fast evolving mtDNA markers (faster substitution rate) dominating that of the nDNA markers. A further factor in this regard may be the larger number of mtDNA than nDNA markers used in the analyses. Notwithstanding this, Reid and Wiens (2011) found that the combined mtDNA and nDNA data analyses were not necessarily dominated by the mtDNA dataset because of its more variable sites. In this study, the target organism is slow evolving (Avise et al., 1992; Tolly et al., 2017). Coalescence depth in slow evolving groups is shallow and recent rapid radiation events, particularly in cases where there is a short time span in their rapid radiation history (i.e. in dealing with species complexes and slow evolving organisms like tortoises), in these cases the incomplete lineage sorting (ILS) issue are particularly common. The most common consequence of ILS is incongruence between gene and species trees (Maddison, 1997; Shaw,

2002; Rubinoff & Holland, 2005; Baum & Smith, 2013; Wang et al., 2018). Rubinoff & Holland (2005) suggested that mtDNA markers were more suitable and reliable for inferring phylogeny than slow evolving nDNA loci in studies focusing on slow evolving groups, groups with a short span of period and groups with rapid radiation. Furthermore, nDNA markers were often found to have poor substitution signals due to their slow evolutionary and substitution rate, which results in poorly resolved phylogenetic tree topologies with poorly supported nodes in many studies (Zhao et al., 2019; Busschau et al., 2019), particularly in turtles (Caccone et al., 2004; Kindler et al., 2012; Petzold et al., 2014).

The shortcomings of mtDNA based phylogenetic inferences have also been addressed (Shaw, 2002; Hurst et al., 2005; Rubinoff & Holland, 2005; Wiens et al., 2010; Baum & Smith, 2013; Leliaert et al., 2014). These pitfalls may result in overestimations of species richness, misleading estimations of speciation rates and robustness of branch support, and incorrectly assigning the phylogenetic relationships between taxa. A possible solution to solving the conflict between nDNA and mtDNA analysis results would be by considering nDNA and mtDNA as two independent evolutionary units in the genome, each retaining different evolutionary information, but both being crucial functional units of evolution (Rubinoff & Holland, 2005). Furthermore, introgression (or gene transfer) may also result in tree topologies providing misleading information (Seo, 2008; Bravo et al., 2019).

Anticipated future phylogenomic studies should provide deeper insight into the cladogenesis of the *P. tentorius* complex by overcoming the limitations of traditional small scale datasets. The substantially larger number of informative substitution sites of independently evolving gene loci will lead to greater robustness when inferring phylogenetic relationships among taxa (Rokas et al., 2003). Notwithstanding this, studies found that ILS can have a severe effect on genomic datasets during rapid cladogenesis (Rosenberg, 2013; Roycroft et al., 2019), which may provide misleading phylogenomic relationships among taxa, and the number of clusters recovered (Kubatko & Degnan, 2007; Liu & Edwards, 2009; Giarla & Esselstyn, 2015; Warnow, 2015; Mendes and Hahn, 2018). What this suggests is that in terms of the number of loci used in a study where there is rapid radiation as shown by recent cladogenic events (as in this case) a larger number of gene loci is not necessarily the critical factor, but rather the suitability of the markers utilized.

4.4.4 Genetic diversity and gene flow between sequence and microsatellite DNA datasets

The AMOVA (Fig. 4.6 and Figure S4.12) and Migrate analysis results showed similar genetic diversity and gene flow patterns between the mtDNA and microsatellite DNA datasets. Chapter 2 addressed the uncertainty in the placement of C6 in the tree topology. Despite their tree topology supporting its placement, it clustered with C2+C3 rather than with (C1+C4)+C7)+C5. The AMOVA analyses of both the mtDNA and microsatellite DNA datasets advocated C6 as close relative to C2+C3, rather than the rest of the clades. Nonetheless, it is interesting that the incongruent patterns between the mtDNA and microsatellite DNA datasets were at within-clade level. In the mtDNA dataset the highest within-clade genetic diversity was found in C1, C2 and C6, and the lowest in C3, C4, C5 and C7. In the microsatellite DNA dataset, the highest diversity was in C1 and C4, and the lowest in C3, C5, C6 and C7. This may imply that single marker based genetic diversity analyses underestimates genetic diversity, as seems to be the case in C4 according to Chapter 2.

4.4.5 Hybridization

The microsatellite DNA based DAPC admixed individual analyses revealed possible hybridization between C1 and C2 at Victoria West and in the Richmond area. Similar findings of possible hybridization between C1 and C2 at Victoria West were made by Chapter 2. Victoria West and the Richmond area therefore appear to be intergradation zones for C1 and C2, but this requires verification by further study with extensive sampling. Although Chapter 2 also detected potential hybridization between C2 and C4, the microsatellite DNA based analyses failed to detect any sign of crossing-over between C2 and C4. This may be due to the randomness of crossing-over that affects only part of the nuclear genome and may therefore have missed the microsatellite loci in the hybridized individuals.

4.4.6 Inbreeding

Among the 11 populations, the C1 (eastern) population revealed extensive inbreeding because the majority of the frequency bars fell into the range above 0.4 (Fig. 4.7). According to Jombart (2008) frequency bars above 0.4 are considered a sign of high levels of inbreeding. The C1 (eastern) population occurs in the Fish River Valley and is isolated from other C1 populations. The Fish River Valley is also isolated from the Nama Karoo habitat by Albany Thicket along

the Fish River canyon (Mucina & Rutherford, 2006), and therefore from central and western populations. This isolation may have resulted in fragmentation of its habitat, restraining gene flow between the Fish River valley population and the rest of the C1 populations, resulting in its low genetic diversity; also confirmed by the AMOVA analyses (Table 4.1). This suggests that the Fish River valley population may deserve conservation attention.

4.4.7 Calibration dating based on the combined mtDNA and nDNA dataset

While this study did not focus on the biogeographic and radiation history of the *P. tentorius* complex, for detail on biogeography see Chapter 3. Calibration dating results revealed that *P. tentorius* branched off from the other two congeneric species, *Psammobates oculifer* and *Psammobates geometricus*, at about 28.27 Ma, which suggests that a possible driving force may have been increasing aridity during the Oligocene (Hofmeyr et al., 2017). The nearly identical dating results to that of Hofmeyr et al. (2017) suggest that the radiation of *P. tentorius* was initiated during the Mid-Miocene. Before the Mid-Miocene, temperature and rainfall increased in the Late Oligocene, but then temperature suddenly decreased in the Mid-Miocene (Zachos et al., 2001; Cachel, 2015). This may have facilitated speciation via habitat and shelter shrinking during the cooling down phase. These dating results also coincide with radiation in the genus *Chersobius* (Hofmeyr et al., 2017). Compared to the mtDNA calibration results discussed in Chapter 3, the dating results based on mtDNA+nDNA in this Chapter was more conservative. The latter was most likely due to the much slower nDNA clock rate. As discussed in Chapter 2, the nDNA and mtDNA genomes represent the two fundamental evolutionary units, but they do not necessarily retain congruent evolutionary information.

The results revealed high levels of divergence within the *P. tentorius* complex, that is compatible with that of the genera *Chersobius* and *Kinixys* (Hofmeyr et al., 2017), which implies that some clades in the *P. tentorius* complex deserve to be elevated to species level.

4.4.8 Correlation between geographic distance, mtDNA and microsatellite DNA

Although mtDNA and nDNA are the two fundamental elements of the genome, each contains different evolutionary information (Rubinoff & Holland, 2005) as shown by the conflicting phylogenies they recovered in this study. However, the Mantel tests (Table 4.2) found that gene flow and genetic distance patterns between the mtDNA and microsatellite DNA datasets were

similar. No significant correlation between gene flow and geographic distance was detected by the Mantel tests, or between genetic distance and geographic distance. These results suggest that increasing geographic distance did not necessarily cause deeper genetic divergence in *P. tentorius*, since tortoises are generally highly mobile. The major driving force of cladogenesis was possibly geographic barriers, but this requires further biogeographic investigation.

4.4.9 Alternative cladogenic hypothesis

The ABC analyses did not retrieve results congruent with that of the MSC Bayesian species inference. Instead, it suggested C6 (basal lineage) as ancestral lineage and sister group to C2. In turn, C6+C2 is the sister group to C3. Such results were not completely unexpected, and in fact, are not unreasonable. The mtDNA based phylogenetic results (Chapter 2) gave the following groupings, (C2+C3)+C6 and (C1+C4)+C7)+C5, although there was uncertainty about the phylogenetic position of C6. The microsatellite DNA based phylogeny suggested a sister relationship between C2 and C6, with C3 closer to the basal part of the tree. The rest of the topology was similar to that derived from the mtDNA dataset. The combined mtDNA and nDNA-PRLR phylogeny was generally identical to that of the mtDNA phylogeny. The PRLR phylogeny also suggested that C2 and C6 are sister groups, and that C3 was more closely related to C1+C4+C5+C7. The best scenario simulated from the ABC analyses is indeed compatible with the results of all the datasets. As discussed above, phylogenetic analyses based only on mtDNA may be misleading about the true phylogenetic relationships between taxa and may also overestimate the divergence rate among them (Shaw, 2002; Rubinoff & Holland, 2005; Wiens et al., 2010). The ABC simulated best scenario results of this study also suggest that there should be caution when using solely mtDNA data to infer a phylogeny. Morphologically, C3 is more similar to the *P. t. tentorius* group (C1, C4, C5, C7), rather than the *P. t. verroxii* group (C2, C6) in terms of carapace shape and colour pattern (the detail will be discuss further in Chapter 6). The information retrieved from the morphological features match the ABC analyses and nDNA-PRLR based results. A detailed study based on morphological characters and morphometric will be presented in Chapter 6. Overall, the best scenario retrieved from the ABC simulation analyses is possibly the most likely one for explaining the genealogy of *P. tentorius*, and it highlighted the importance of using multiple types of datasets to infer phylogenetic relationships among taxa.

4.4.10 Determination of OTUs

Species and OTUs are the fundamental entities of biological science (Blaxter et al., 2005; de Queiroz, 2007; Fu et al., 2012; Schmidt et al., 2014), and both are regarded as products and indicators of complex biotic and abiotic ecological interactions. Correctly determining OTUs is crucial in biological conservation and wildlife management (Avise, 1992; Moritz, 1994 & 2002; Thomson et al., 2018). The modern species concept is no longer restricted to morphology, but is rather linked to ecology with its temporal and spatial dimensions. Defining species or OTUs should therefore reflect evolutionary histories and phylogeographic processes (Avise, 1992; Moritz, 1994 & 2002), rather than empirical phenetics and morphology. Moreover, in the conservation of species priority should be given to protecting important historical lineages and maximizing the preservation of speciation potential, rather than only focusing on endangered taxa (Avise, 1992; Moritz, 1994 & 2002). However, to correctly assign OTUs remains challenging (de Queiroz, 2007; Fujita et al., 2012), the major challenges being a) difficulty in defining species boundaries (de Queiroz, 1998, 1999, 2005b, 2007) due to the absence of a unified species concept (Mallet, 2013), and b) unclarity about the criteria used to delineate species (de Queiroz, 1998, 1999, 2005a, b).

Besides the above issues, developing a proper OTU scheme for conservation management can be tough for a species complex when cryptic species are encountered (Heath et al., 2008). Ignoring the importance of correctly identifying cryptic species can lead to ineffective and costly conservation management (Briggs, 2005; Heath et al., 2008). In this study of the *P. tentorius* complex, I discovered several cases of species crypsis with both the microsatellite DNA dataset and sequence datasets, which would not have been possible without the use of molecular genetic markers. This was confirmed by the following findings: 1) most members of the southern population of C2 were previously regarded as *P. t. tentorius* based on morphological characters (Loveridge & Williams, 1957; Hofmeyr et al., 2014; Branch, 1998 & 2008), but were found to actually belong to *P. t. verroxii*; 2) despite being morphologically highly different, the C2 southern and C2 northern populations were found not to differ significantly, genetically; 3) Loveridge & Williams (1957), Hofmeyr et al. (2014), and Branch (1998 & 2008) proposed that the southwestern Karoo (near Matjiesfontein and the Tankwa Karoo) as a possible "three-way contact zone" where the three subspecies coexist but the results confirmed that all individuals found in that region belonged to C1, despite morphologically constituting a highly polymorphic population; 4) Loveridge & Williams (1957) and Branch (2008) suggested that the uniformly brown "*P. bergeri*" may deserve recognition as a valid

taxon but the results showed that these uniformly brown morphs were present in C1, C2 and C6, and should therefore not be considered a valid taxon; 5) Boycott & Bourquin (1988) and Branch (2008) regarded the tortoises found above the Great Escarpment in the Kamiesberg, Hantam Karoo and Roggeveldberge region as *P. t. trimeni*, but the results confirmed the findings of Hofmeyr et al. (2014) and Chapter 2 that these tortoises belong to C4, the closest relative of C1, and not C3 (regarded as *P. t. trimeni*); true *P. t. trimeni* occurs only in the west coast succulent Karoo area; 6) notwithstanding great morphological similarity, C5 from the south-western little Karoo and C7 from Oudtshoorn are significantly different from "C1+C4"; 7) although morphologically similar, C2 and C6 (both previously regarded as *P. t. verroxii*) are genetically significantly different; 8) morphologically C1 and C4 are distinguishable, however, this study together with that of Chapter 2 generally recommend that C1 and C4 not be regarded as different taxa.

Based on all the results, together with the multiple species delimitation results of Chapter 2, I recommend that at least C1, C2, C3 and C6 deserve recognition as full species. And that C1 and C4 should be treated as the same taxon in accordance with the results obtained from the microsatellite DNA dataset, nDNA-PRLR dataset and mtDNA dataset analyses. Clade 5 and C7 were significantly different from C1+C4 according to the microsatellite DNA and mtDNA datasets, although the nDNA-PRLR dataset did not reveal a significant difference between them (see Chapter 2). I suggest that C5 and C7 be recognised as distinct species given their isolation from C1+C4 and their separation by a significant barrier that could promote their further divergence. In terms of *p*-distances, the divergence between C1 vs C5, C1 vs C7 and C5 vs C7, respectively, is higher than the divergence between C2 and C3. Fresh tissue samples are needed to determine the taxonomic status of the Uniondale population, and whether it represents a distinct taxon, or belongs to C1+C4. In accordance with the nomenclature in Hewitt (1933 & 1934) and Loveridge & Williams (1957), I recommend that C1+C4 be assigned as *P. tentorius sp. nov*, C2 as *P. fiski .sp. nov*, C3 as *P. trimeni sp. nov* and C6 as *P. verroxii sp. nov*. Names for the two new taxa, C5 and C7, have still to be decided on. Detailed morphological descriptions still have to be carried out for these proposed new taxa and proper nomenclature developed based on type specimens.

Supplementary Figures

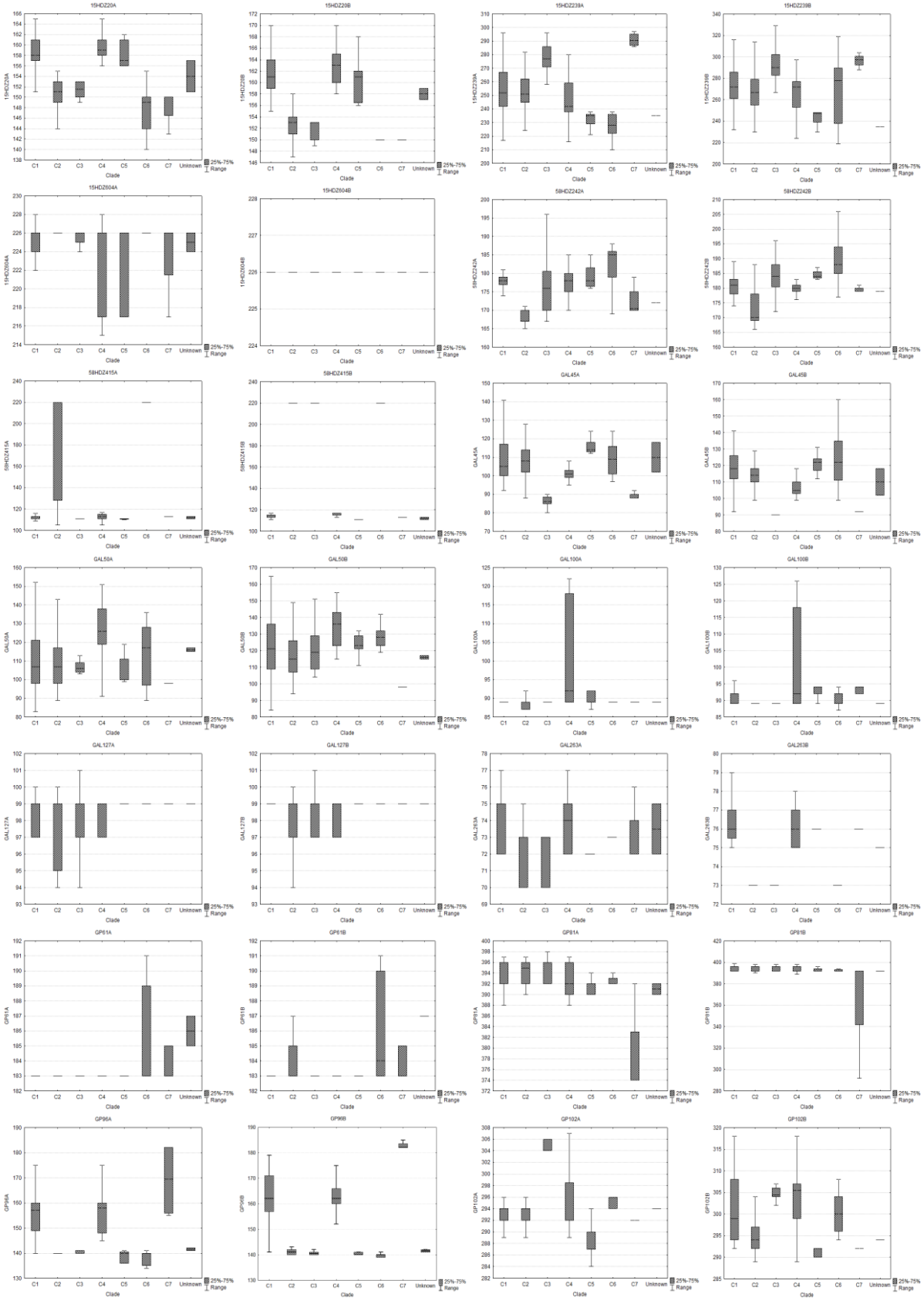


Figure S4.1. Boxplots showing allele size ranges on all diploid microsatellite loci (from locus 15HDZ20 to GP102).

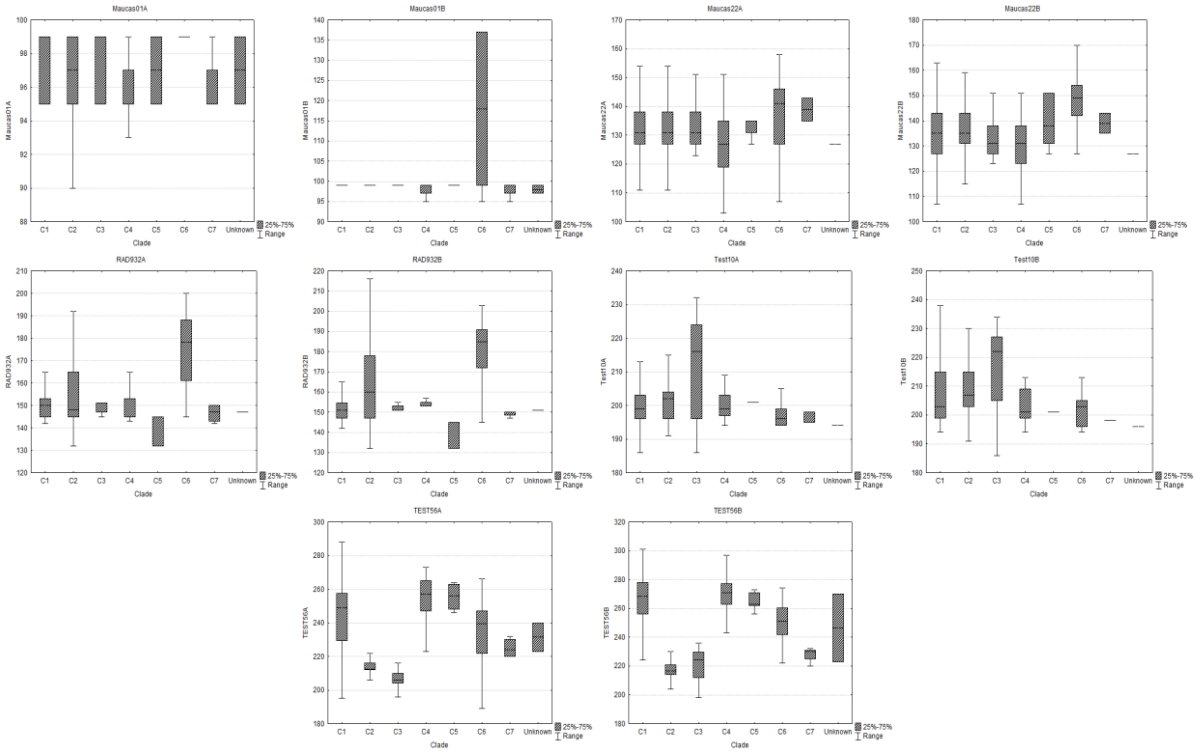


Figure S4.2. The boxplots showing allele size ranges of all diploid microsatellite loci (from locus Maucas01 to TEST56).

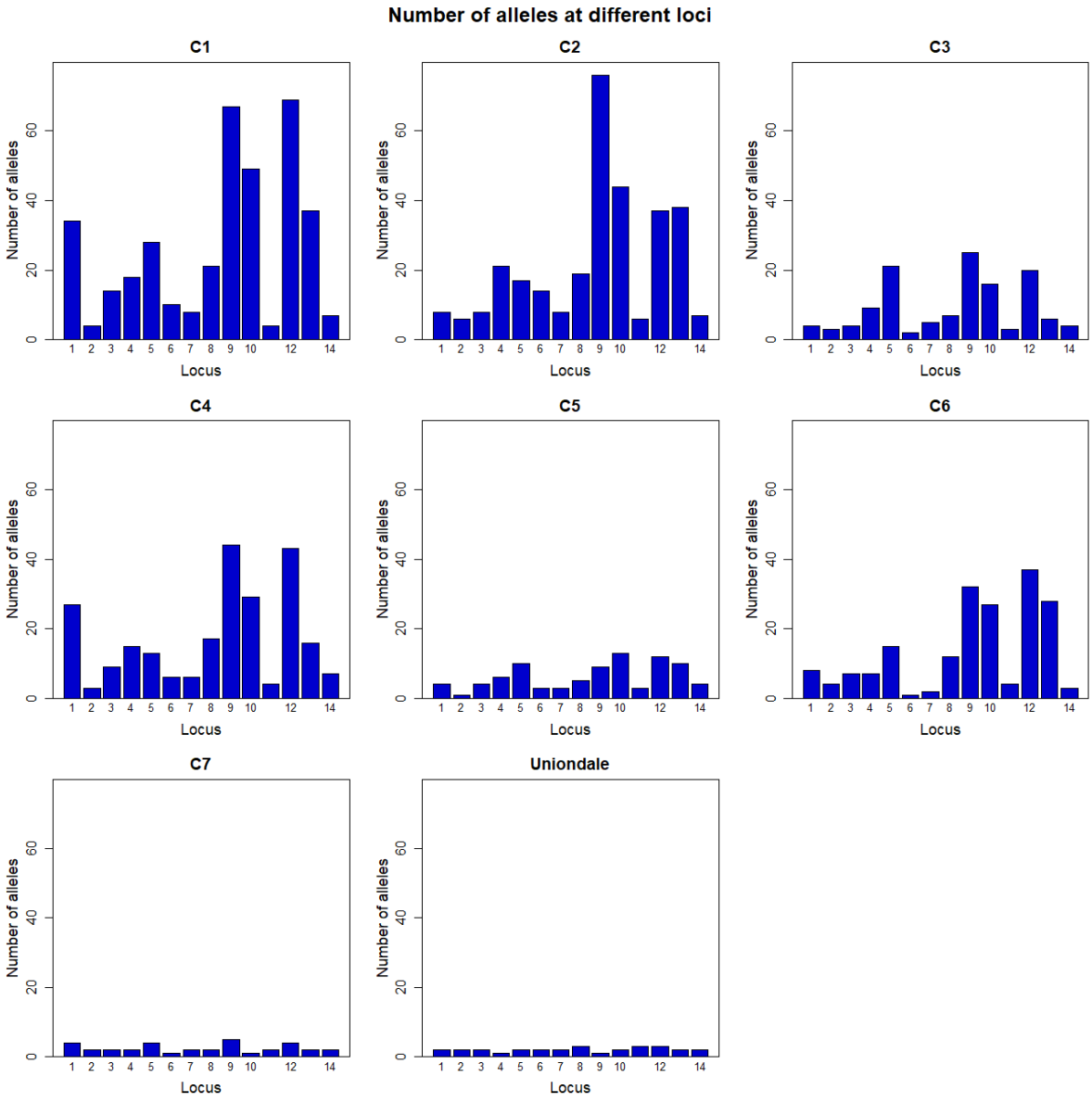


Figure S4.3. Allele diversities among different clades for the 14 different microsatellite DNA loci. The numbers on the x-axis represent different loci.

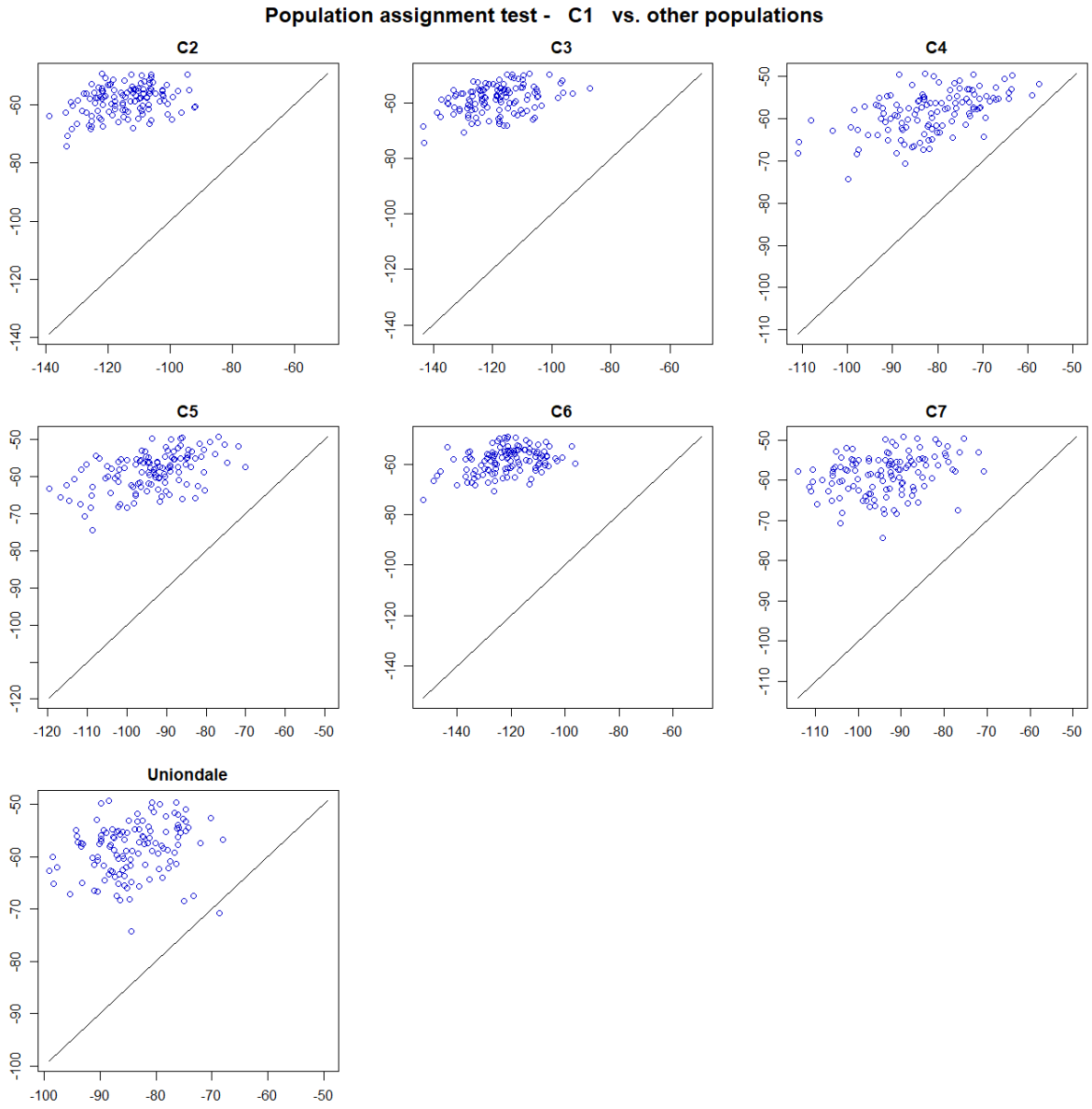


Figure S4.4. Arlequin pairwise assignment test results (using the microsatellite DNA dataset) revealed that the Uniondale samples were closest to C1.

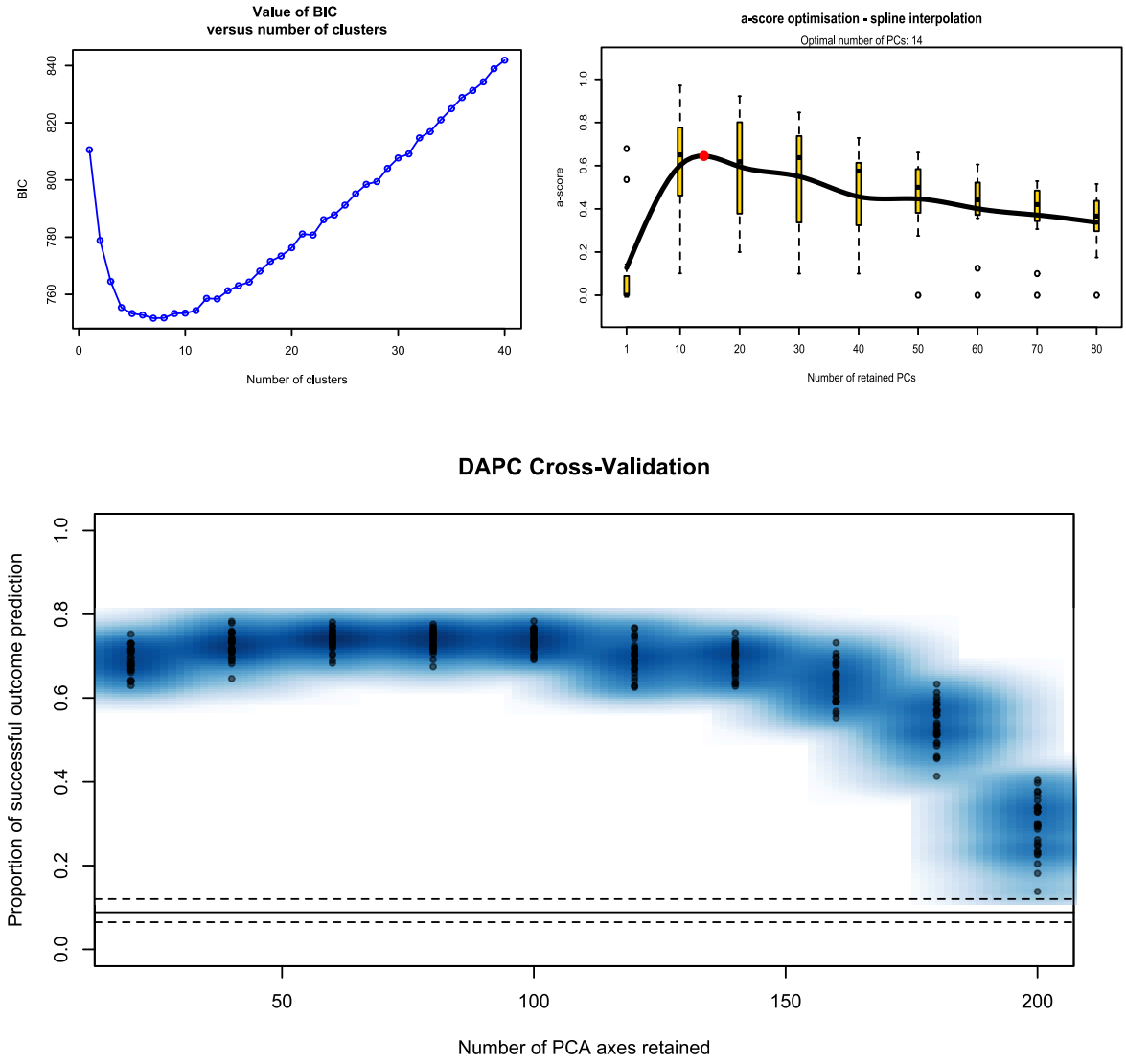


Figure S4.5. Results of BIC value versus the number of clusters to determine the optimal clustering scheme in the DAPC analyses. The a-score optimization determined the optimal number of retained principal components (PCs) and the proportion of successful predictions versus number of PCA axes retained. The Cross-Validation analysis determined the initial number of PC retained.

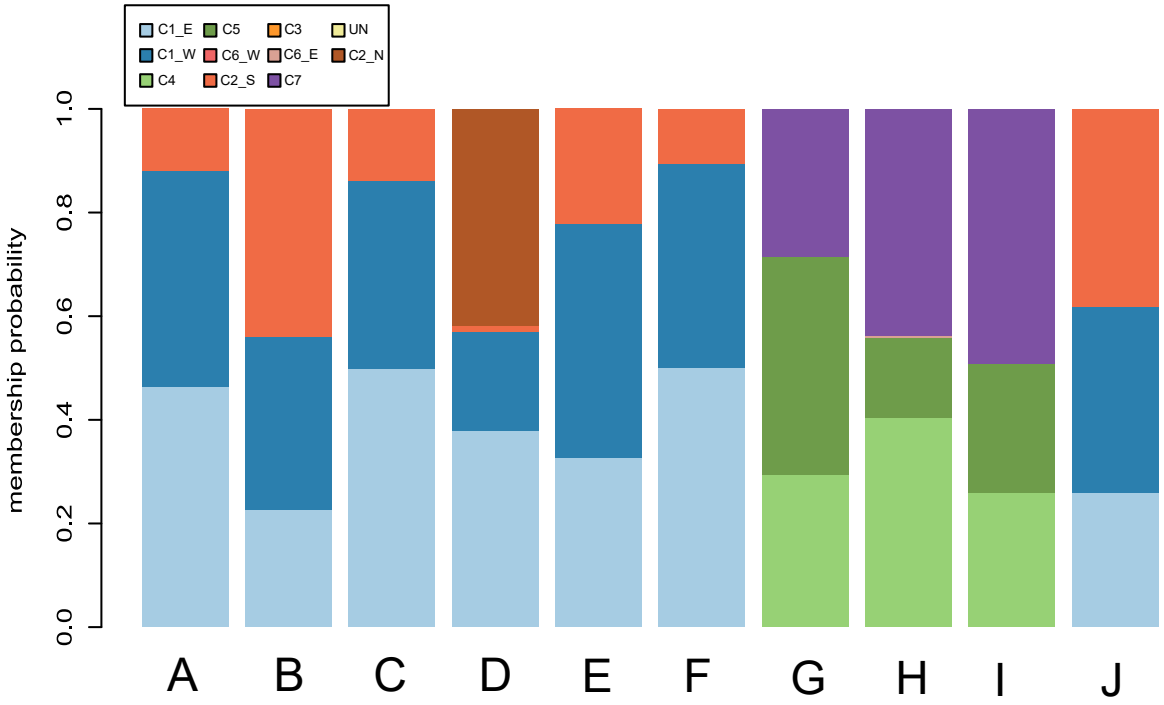


Figure S4.6. Membership prediction plot for ten suspected admixed individuals to determine potential hybridizations. “C1_E”: C1 (eastern), “C1_W”: C1 (western), “C2_N”: C2 (northern), “C2_S”: C2 (southern), “C6_E”: C6 (eastern) and “C6_W”: C6 (western).

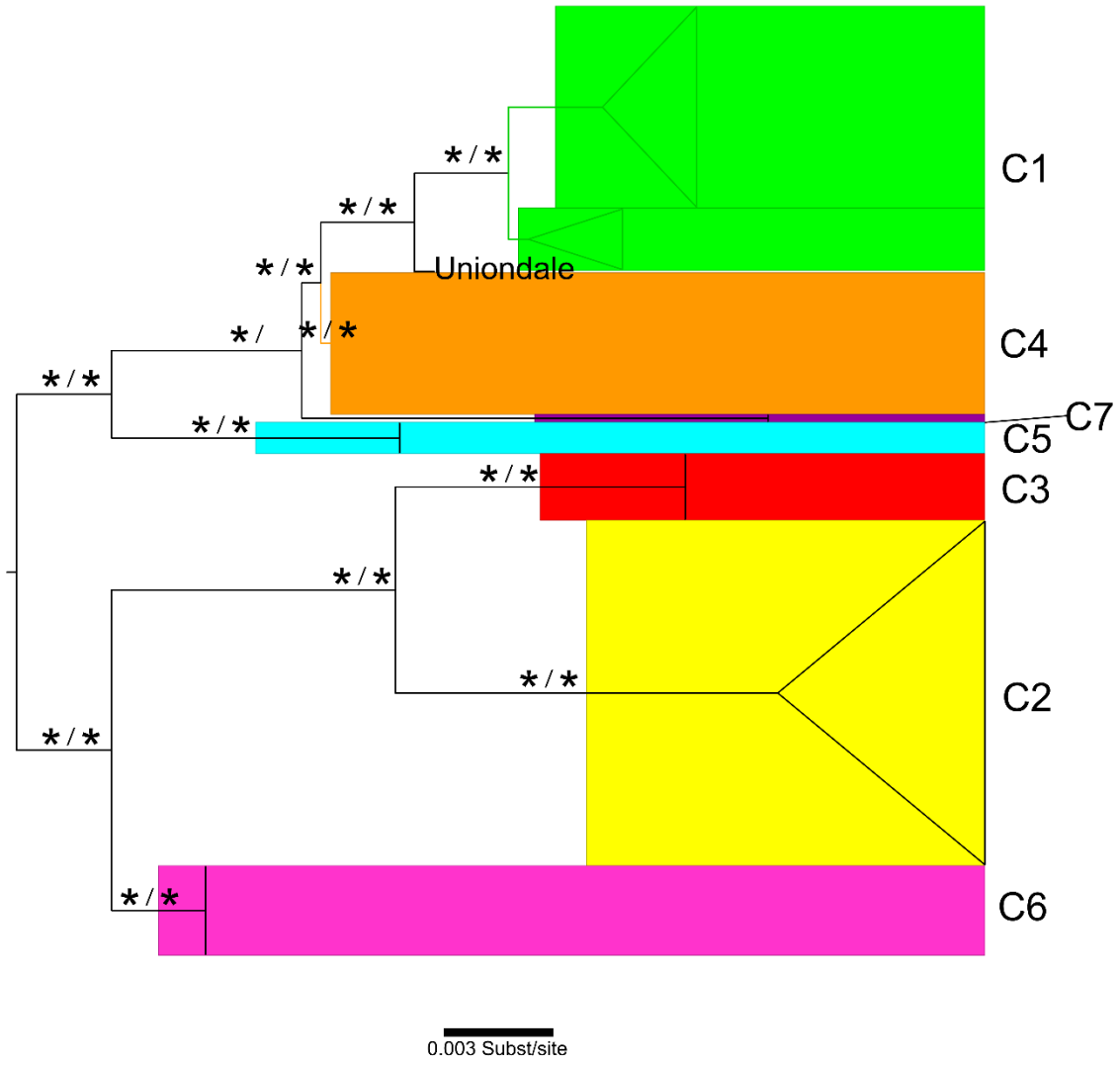


Figure S4.7. The single marker (12S) based tree retrieved from the BI and ML analyses. The tree topology was derived from a ML analysis. Support values are given for BI (left) and ML (right); " * " indicates strong support (BP > 70 or PP > 0.95).

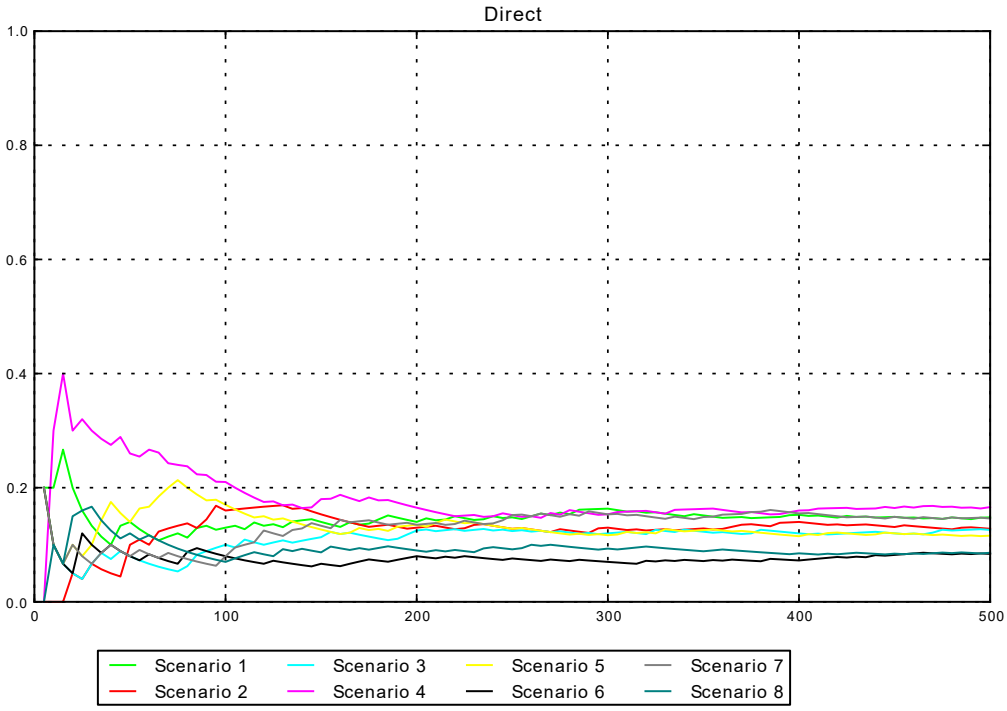


Figure S4.8. Best model selection using model comparison methods implemented in the ABC analyses. The scenario with highest Direct score (occupying the highest proportion) was considered the most plausible scenario.

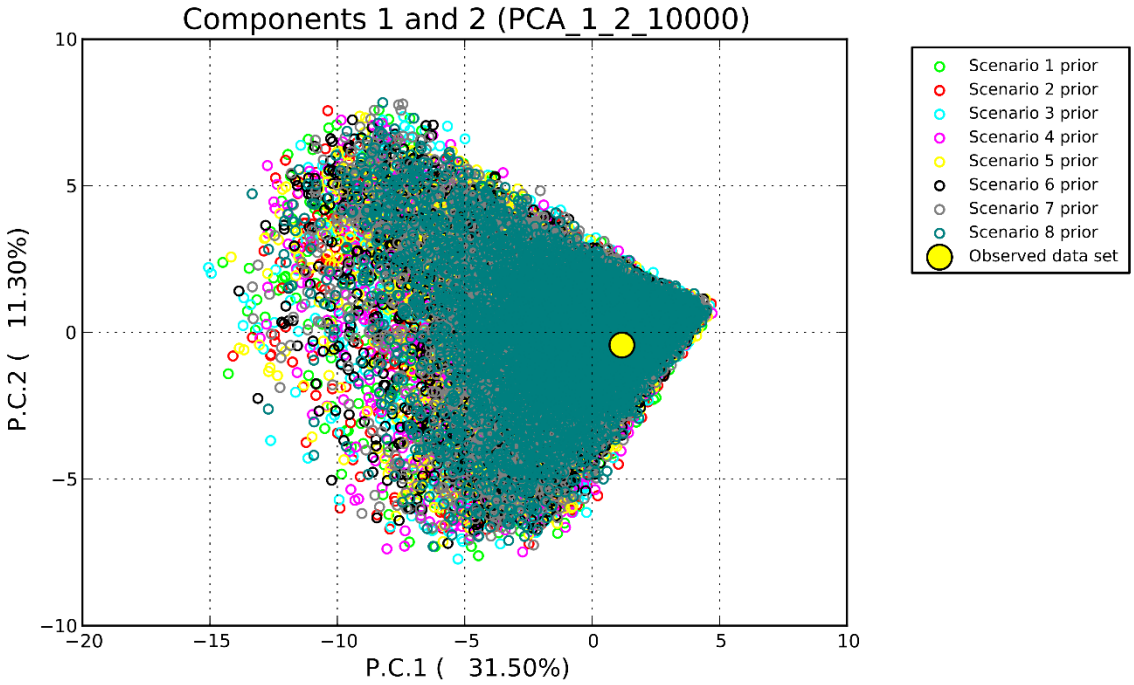


Figure S4.9. Scatterplot showing the PCA results evaluating the scenarios and priors in the ABC analyses. Each small dot represents a simulated dataset from the reference table and the large yellow dot represents the observed dataset. I expected the yellow dot to fall in the range of the small dots, which indicated that sampling was adequate.

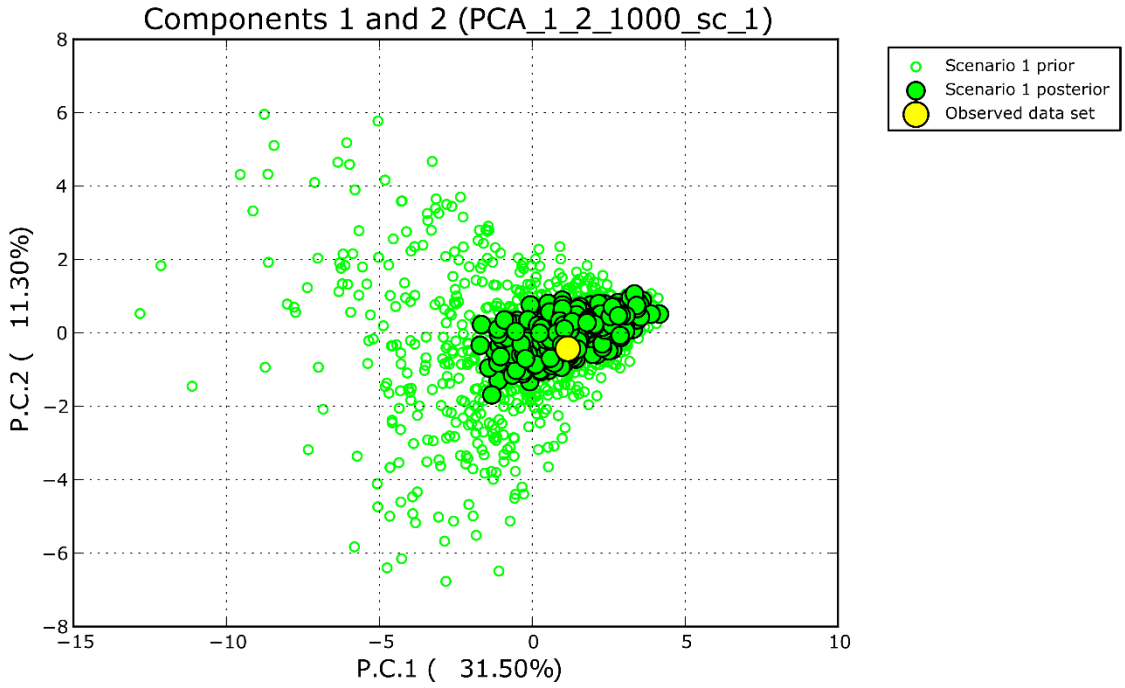


Figure S4.10. Results of the model checking ABC analyses. The observed dataset (shown as the big yellow dot) fell within the range of both the simulated priors (open dots) and posteriors (medium size dots) datasets, showing that sampling was adequate and the simulation analysis results reliable.

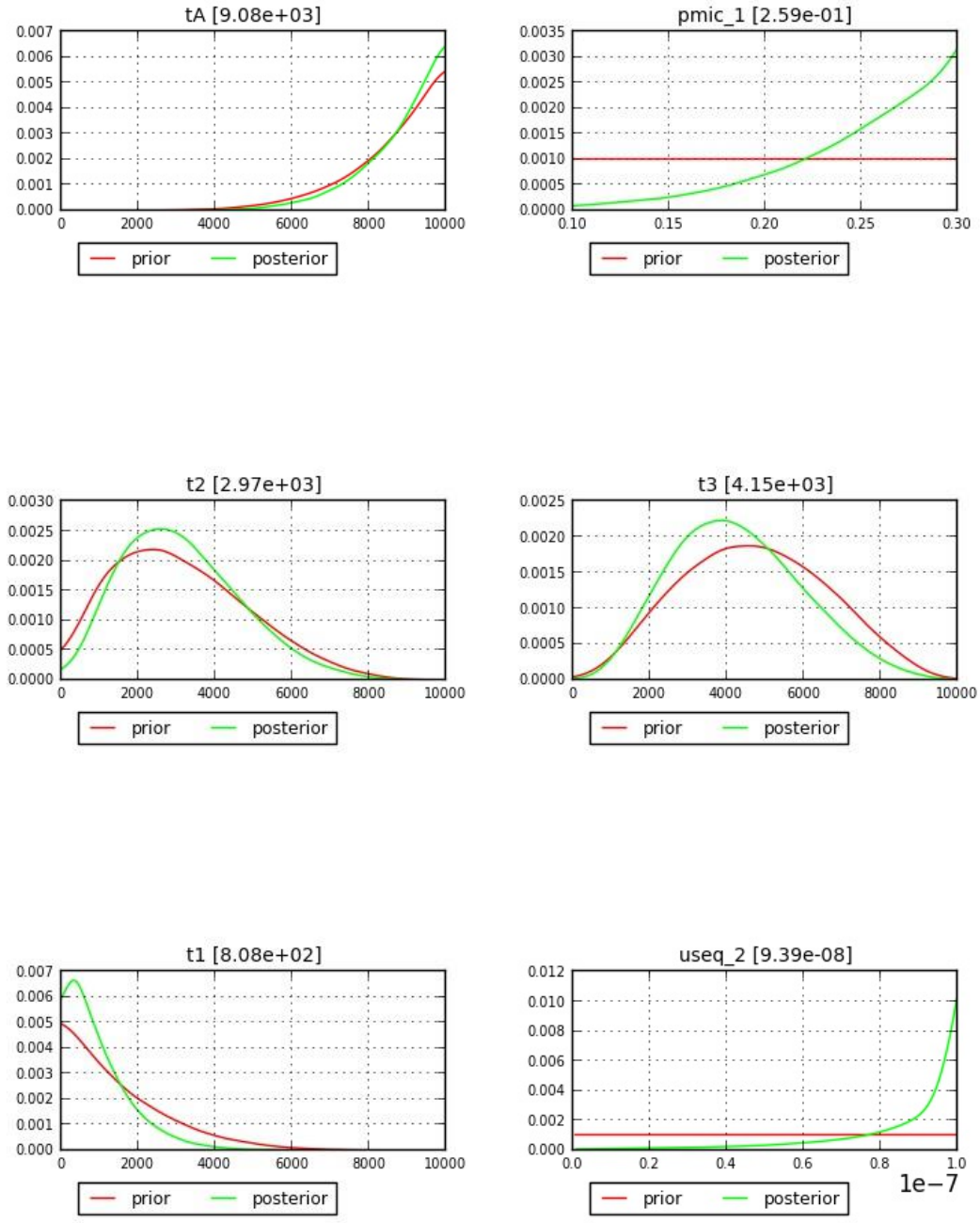


Figure S4.11. The parameters involved in the ABC analyses between priors and posteriors. The majority of the distributions of the priors and posteriors revealed a good match, implying that the simulation analysis results were reliable.

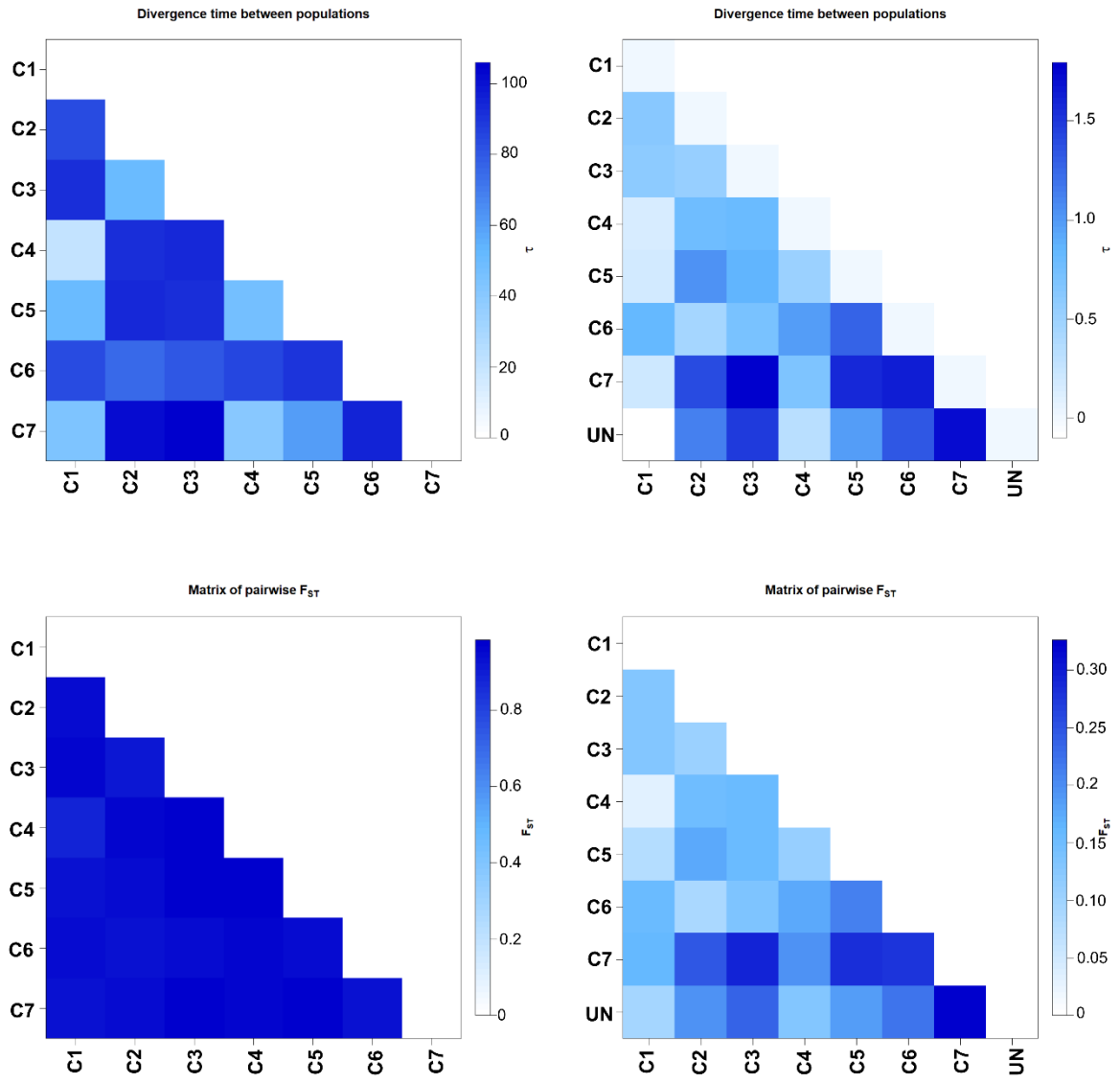


Figure S4.12. The relative divergence time (tau) matrix and the pairwise Fst matrix generated among clades from the AMOVA analyses between mtDNA (shown on the left side) and microsatellite DNA (shown on the right side) datasets. “UN” refers to the Uniondale population.

Supplementary Tables

All the Supplementary Tables (Table S3.1-S3.6) were not included in the main document of this thesis, since their size are large and unable to fit in the main document. Therefore, all supplementary Tables will only be provided independently as supporting documents.

References

Avise, J. C. (1992). Molecular population structure and the biogeographic history of a regional fauna: a case history with lessons for conservation biology. *Oikos*, 62-76.

Avise, J. C., Bowen, B. W., Lamb, T., Meylan, A. B., & Bermingham, E. (1992). Mitochondrial DNA evolution at a turtle's pace: evidence for low genetic variability and reduced microevolutionary rate in the Testudines. *Molecular Biology and Evolution*, *9*(3), 457-473.

Bagshaw, A. T. (2017). Functional mechanisms of microsatellite DNA in eukaryotic genomes. *Genome Biology and Evolution*, *9*(9), 2428-2443.

Ballard, J. W. O., & Rand, D. M. (2005). The population biology of mitochondrial DNA and its phylogenetic implications. *Annual Review of Ecology, Evolution and Systematics*, *36*, 621-642.

Baum, D. A., & Smith, S. D. (2013). *Tree thinking: An introduction to phylogenetic biology*. Greenwood Village: CO: Roberts.

Beaumont, M. A., Zhang, W., & Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics*, *162*(4), 2025-2035.

Beerli, P. (2005). Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics*, *22*(3), 341-345.

Beerli, P. (2009) How to use migrate or why are markov chain monte carlo programs difficult to use? In G. Bertorelle, M. W. Bruford, H. C. Hauffe, A. Rizzoli, and C. Vernesi, editors, Population Genetics for Animal Conservation, volume 17 of Conservation Biology, pages 42–79. Cambridge University Press, Cambridge UK, 2009.

Blaxter, M., Mann, J., Chapman, T., Thomas, F., Whitton, C., Floyd, R., & Abebe, E. (2005). Defining operational taxonomic units using DNA barcode data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1462), 1935-1943.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., ... & Drummond, A. J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, *10*(4), e1003537.

Bravo, G. A., Antonelli, A., Bacon, C. D., Bartoszek, K., Blom, M. P., Huynh, S., ... & Morlon, H. (2019). Embracing heterogeneity: coalescing the Tree of Life and the future of phylogenomics. *PeerJ*, *7*, e6399.

Briggs, J. C. (2005). Coral reefs: conserving the evolutionary sources. *Biological Conservation*, *126*(3), 297-305.

Busschau, T., Conradie, W., & Daniels, S. R. (2019). Evidence for cryptic diversification in a rupicolous forest-dwelling gecko (Gekkonidae: *Afroedura pondolia*) from a biodiversity hotspot. *Molecular Phylogenetics and Evolution*, *139*, 106549.

Cachel, S. (2015). Fossil primates. Cambridge: Cambridge University Press.

Cornuet, J. M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., ... & Estoup, A. (2014). DIYABC v2. 0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics*, *30*(8), 1187-1189.

Cornuet, J. M., Santos, F., Beaumont, M. A., Robert, C. P., Marin, J. M., Balding, D. J., ... & Estoup, A. (2008). Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics*, *24*(23), 2713-2719.

Daïnou, K., Blanc-Jolivet, C., Degen, B., Kimani, P., Ndiade-Bourobou, D., Donkpegan, A. S., ... & Hardy, O. J. (2016). Revealing hidden species diversity in closely related species using nuclear SNPs, SSRs and DNA sequences–a case study in the tree genus *Milicia*. *BMC Evolutionary Biology*, *16*(1), 259.

De Queiroz K. (1998). The general lineage concept of species, species criteria, and the process of speciation: A conceptual unification and terminological recommendations. In D. J. Howard & S. H. Berlocher (Eds.), *Endless forms: Species and speciation* (pp.57-75). New York: Oxford University Press.

De Queiroz K. (1999). The general lineage concept of species and the defining properties of the species category. In R. A. Wilson (Ed.), *Species: New interdisciplinary essays* (pp. 49-89). Cambridge: MIT Press.

De Queiroz, K. (2005a). Different species problems and their resolution. *BioEssays*, *27*(12), 1263-1269.

De Queiroz, K. (2005b). Ernst Mayr and the modern concept of species. *Proceedings of the National Academy of Sciences*, *102*(suppl 1), 6600-6607.

De Queiroz, K. (2007). Species concepts and species delimitation. *Systematic Biology*, *56*(6), 879-886.

Dupanloup, I., Schneider, S., Excoffier, L. (2002) A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology 11*(12), 2571-81.

Earl, D. A. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*(2), 359-361.

Excoffier, L., & Lischer, H. (2009). Arlequin 3.5. *An Integrated Software Package for Population Genetics Data Analysis.* Computational and Molecular Population Genetics Lab (CMPG). Institute of Ecology and Evolution. University of Berne, Switzerland.

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, *164*(4), 1567-1587.

Feng, X., Liu, J., & Gong, X. (2016). Species delimitation of the *Cycas segmentifida* complex (Cycadaceae) resolved by phylogenetic and distance analyses of molecular data. *Frontiers in Plant Science*, *7*, 134.

Fisher-Reid, M. C., & Wiens, J. J. (2011). What are the consequences of combining nuclear and mitochondrial data for phylogenetic analysis? Lessons from *Plethodon* salamanders and 13 other vertebrate clades. *BMC Evolutionary Biology*, *11*(1), 300.

Fritz, U., Alcalde, L., Vargas-Ramírez, M., Goode, E. V., Fabius-Turoblin, D. U., & Praschag, P. (2012). Northern genetic richness and southern purity, but just one species in the *Chelonoidis chilensis* complex. *Zoologica Scripta*, *41*(3), 220-232.

Giarla, T. C., & Esselstyn, J. A. (2015). The challenges of resolving a rapid, recent radiation: empirical and simulated phylogenomics of Philippine shrews. *Systematic Biology*, *64*(5), 727-740.

Glaubitz, J. C. (2004). Convert: a user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. *Molecular Ecology Notes*, *4*(2), 309-310.

Heath, T. A., Hedtke, S. M., & Hillis, D. M. (2008). Taxon sampling and the accuracy of phylogenetic analyses. *Journal of Systematics and Evolution*, *46*(3), 239-257.

Hewitt J. (1933). On the Cape species and subspecies of the genus *Chersinella* Gray. Part I. *Annals of the Natal Museum*, *7*, 255-297.

Hewitt, J. (1934). On the Cape species and subspecies of the genus *Chersinella*. Part II. *Annals of the Natal Museum, 7*, 303-352.

Hofmeyr, M. D., Boycott, R. C., & Baard, E. H. W. (2014). *Psammobates tentorius* (Bell, 1828). In: M. F. Bates, W. R. Branch, A. M. Bauer, M. Burger, J. Marais, G. J. Alexander & M. S. De Villiers (Eds.), *Atlas and red list of the reptiles of South Africa, Lesotho and Swaziland* (pp. 70-85). Pretoria: South African National Biodiversity Institute.

Hurst, G. D., & Jiggins, F. M. (2005). Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1572), 1525-1534.

Ihlow, F., Vamberger, M., Flecks, M., Hartmann, T., Cota, M., Makchai, S., ... & Fritz, U. (2016). Integrative taxonomy of Southeast Asian snail-eating turtles (*Geoemydidae: Malayemys*) reveals a new species and mitochondrial introgression. *PloS One*, *11*(4), e0153108.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403-1405.

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, *11*(1), 94.

Jones, G. (2017). Algorithmic improvements to species delimitation and phylogeny estimation under the multispecies coalescent. *Journal of Mathematical Biology*, *74*(1-2), 447-467.

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*, e281.

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, *15*(5), 1179-1191.

Kubatko, L. S., & Degnan, J. H. (2007). Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology*, *56*(1), 17-24.

Lagostina, E., Dal Grande, F., Andreev, M., & Printzen, C. (2018). The use of microsatellite markers for species delimitation in Antarctic *Usnea* subgenus *Neuropogon*. *Mycologia*, *110(6)*, 1047-1057.

Leaché, A. D. (2010). Species trees for spiny lizards (genus *Sceloporus*): identifying points of concordance and conflict between nuclear and mitochondrial data. *Molecular Phylogenetics and Evolution*, *54*(1), 162-171.

Leliaert, F., Verbruggen, H., Vanormelingen, P., Steen, F., López-Bautista, J. M., Zuccarello, G. C., & De Clerck, O. (2014). DNA-based species delimitation in algae. *European Journal of Phycology*, *49*(2), 179-196.

Lischer, H. E., & Excoffier, L. (2011). PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, *28*(2), 298-299.

Liu, L., & Edwards, S. V. (2009). Phylogenetic analysis in the anomaly zone. *Systematic Biology*, *58*(4), 452-460.

Lourenço, J. M., Claude, J., Galtier, N., & Chiari, Y. (2012). Dating cryptodiran nodes: origin and diversification of the turtle superfamily Testudinoidea. *Molecular Phylogenetics and Evolution*, *62*(1), 496-507.

Loveridge, A., & Williams, E. E. (1957). Revision of the African tortoises and turtles of the suborder Cryptodira. *Bulletin of the Museum of Comparative Zoology*, *115*(6), 161-557.

Lumley, L. M., & Sperling, F. A. (2011). Utility of microsatellites and mitochondrial DNA for species delimitation in the spruce budworm (Choristoneura fumiferana) species complex (Lepidoptera: Tortricidae). *Molecular Phylogenetics and Evolution*, *58*(2), 232-243.

Maddison, W. P. (1997). Gene trees in species trees. *Systematic Biology*, *46*(3), 523-536.

Mallet, J. (2013). Concepts of species. In S. Levin, (Ed), *Encyclopedia of Biodiversity*, volume 6, (pp.679-691). Massachusetts: Academic Press.

Mendes, F. K., & Hahn, M. W. (2017). Why concatenation fails near the anomaly zone. *Systematic Biology*, *67*(1), 158-169.

Moritz, C. (1994). Applications of mitochondrial DNA analysis in conservation: a critical review. *Molecular Ecology*, *3*(4), 401-411.

Moritz, C. (2002). Strategies to protect biological diversity and the evolutionary processes that sustain it. *Systematic Biology*, *51*(2), 238-254.

Mucina, L., & Rutherford, M. C. (2006). *The Vegetation of South Africa, Lesotho and Swaziland*. Pretoria: South African National Biodiversity Institute.

Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, *76*(10), 5269-5273.

Paetkau, D., Slade, R., Burden, M., & Estoup, A. (2004). Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology*, *13*(1), 55-65.

Paradis, E. (2010). pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics*, *26*(3), 419-420.

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*(2), 289-290.

Peakall, R. O. D., & Smouse, P. E. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, *6*(1), 288-295.

Piry, S., Alapetite, A., Cornuet, J. M., Paetkau, D., Baudouin, L., & Estoup, A. (2004). GENECLASS2: a software for genetic assignment and first-generation migrant detection. *Journal of Heredity*, *95*(6), 536-539.

Pöschel, J., Heltai, B., Graciá, E., Quintana, M. F., Velo-Antón, G., Arribas, O., ... & Vamberger, M. (2018). Complex hybridization patterns in European pond turtles (*Emys orbicularis*) in the Pyrenean Region. *Scientific Reports*, *8*(1), 15925.

Postaire, B., Magalon, H., Bourmaud, C. A. F., & Bruggemann, J. H. (2016). Molecular species delimitation methods and population genetics data reveal extensive lineage diversity and cryptic species in Aglaopheniidae (Hydrozoa). *Molecular Phylogenetics and Evolution*, *105*, 36-49.

Raemy, M., Fritz, U., Cheylan, M., & Ursenbacher, S. (2017). Hybridisation between turtle subspecies: a case study with the European pond turtle (*Emys orbicularis*). *Conservation Genetics*, *18*(2), 287-296.

Rannala, B., & Mountain, J. L. (1997). Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences*, *94*(17), 9197-9201.

Rokas, A., Williams, B. L., King, N., & Carroll, S. B. (2003). Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature*, *425*(6960), 798.

Rosenberg, N. A. (2013). Discordance of species trees with their most likely gene trees: a unifying principle. *Molecular Biology and Evolution*, *30*(12), 2709-2713.

Roycroft, E. J., Moussalli, A., & Rowe, K. C. (2019). Phylogenomics Uncovers Confidence and Conflict in the Rapid Radiation of Australo-Papuan Rodents. *Systematic Biology*.

Rubinoff, D., & Holland, B. S. (2005). Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. *Systematic Biology*, *54*(6), 952-961.

Schmidt, T. S., Rodrigues, J. F. M., & Von Mering, C. (2014). Ecological consistency of SSU rRNA-based operational taxonomic units at a global scale. *PLoS Computational Biology, 10(4),* e1003594.

Seo, T. K. (2008). Calculating bootstrap probabilities of phylogeny using multilocus sequence data. *Molecular Biology and Evolution*, *25*(5), 960-971.

Shaw, K. L. (2002). Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proceedings of the National Academy of Sciences*, *99*(25), 16122-16127.

Shriver, M. D., Jin, L., Boerwinkle, E., Deka, R., Ferrell, R. E., & Chakraborty, R. (1995). A novel measure of genetic distance for highly polymorphic tandem repeat loci. *Molecular Biology and Evolution*, *12*(5), 914-920.

Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, *139*(1), 457-462.

Spinks, P. Q., & Shaffer, H. B. (2009). Conflicting mitochondrial and nuclear phylogenies for the widely disjunct *Emys* (Testudines: Emydidae) species complex, and what they tell us about biogeography and hybridization. *Systematic Biology*, *58*(1), 1-20.

Spitzweg, C., Hofmeyr, M. D., Fritz, U., & Vamberger, M. (2019). Leopard tortoises in southern Africa have greater genetic diversity in the north than in the south (Testudinidae). *Zoologica Scripta*, *48*(1), 57-68.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312-1313.

Takezaki, N., Nei, M., & Tamura, K. (2009). POPTREE2: Software for constructing population trees from allele frequency data and computing other population statistics with Windows interface. *Molecular Biology and Evolution*, *27*(4), 747-752.

Team, R. C. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2012. *URL http://www. R-project. org*.

Thomson, S. A., Pyle, R. L., Ahyong, S. T., Alonso-Zarazaga, M., Ammirati, J., Astripea, J. F., ... & Baker, W. J. (2018). Taxonomy based on science is necessary for global conservation. *PLoS Biology*, *16*(3), e2005075.

Tollis, M., DeNardo, D. F., Cornelius, J. A., Dolby, G. A., Edwards, T., Henen, B. T., Karl, A. E., Murphy, R. W., & Kusumi, K. (2017). The Agassiz's desert tortoise genome provides a resource for the conservation of a threatened species. *PloS One*, *12*(5), e0177708.

Vamberger, M., Durkin, L., Kim, C., Handschuh, M., Seng, R., & Fritz, U. (2017). The leaf turtle population of Phnom Kulen National Park (northwestern Cambodia) has genetic and morphological signatures of hybridization. *Journal of Zoological Systematics and Evolutionary Research*, *55*(2), 167-174.

Vamberger, M., Stuckas, H., Sacco, F., D'Angelo, S., Arculeo, M., Cheylan, M., ... & Fritz, U. (2015). Differences in gene flow in a twofold secondary contact zone of pond turtles in southern Italy (Testudines: Emydidae: *Emys orbicularis galloitalica*, *E. o. hellenica*, *E. trinacris*). *Zoologica Scripta*, *44*(3), 233-249.

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P., & Shipley, P. (2004). MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, *4*(3), 535-538.

Vanhaecke, D., De Leaniz, C. G., Gajardo, G., Young, K., Sanzana, J., Orellana, G., ... & Consuegra, S. (2012). DNA barcoding and microsatellites help species delimitation and hybrid identification in endangered galaxiid fishes. *PloS One*, *7*(3), e32939.

Wang, K., Lenstra, J. A., Liu, L., Hu, Q., Ma, T., Qiu, Q., & Liu, J. (2018). Incomplete lineage sorting rather than hybridization explains the inconsistent phylogeny of the wisent. *Communications Biology*, *1*(1), 169.

Warnow, T. (2015). Concatenation analyses in the presence of incomplete lineage sorting. *PLoS Currents*, *7*.

Wiens, J. J., Kuczynski, C. A., & Stephens, P. R. (2010). Discordant mitochondrial and nuclear gene phylogenies in emydid turtles: implications for speciation and conservation. *Biological Journal of the Linnean Society*, *99*(2), 445-461.

Xia, X. (2013). DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. *Molecular Biology and Evolution*, *30*(7), 1720-1728.

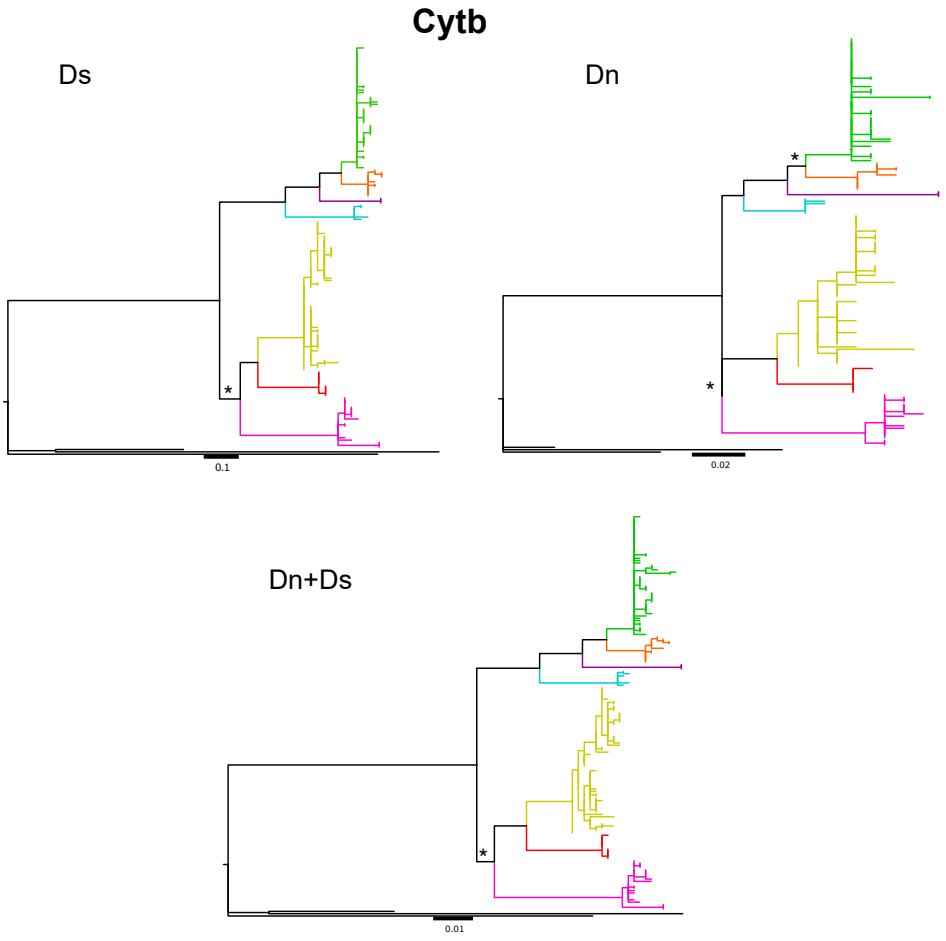
Zachos, J., Pagani, M., Sloan, L., Thomas, E., & Billups, K. (2001). Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science*, *292*(5517), 686-693.

Zhao, Z., Verdú-Ricoy, J., Mohlakoana, S., Jordaan, A., Conradie, W., & Heideman, N. (2019). Unexpected phylogenetic relationships within the world's largest limbless skink species (*Acontias plumbeus*) highlight the need for a review of the taxonomic status of Acontias poecilus. *Journal of Zoological Systematics and Evolutionary Research*, *57*(2), 445-460.

Zheng, Y., Peng, R., Kuro-o, M., & Zeng, X. (2011). Exploring patterns and extent of bias in estimating divergence time from mitochondrial DNA sequence data in a particular lineage: a case study of salamanders (Order Caudata). *Molecular Biology and Evolution*, *28*(9), 2521-2535.

Chapter 5. Selection pressure and genetic structure in the *Psammobates tentorius* species complex based on codon analyses, and phylogeny inferred from amino acid sequences

Graphic abstract



Abstract

This study used codon analysis (dN/dS and Tv/Ti) to investigate selection pressure and genetic structure in the highly polymorphic *Psammobates tentorius* species complex, and amino acid sequences to construct a phylogeny it. The findings based on the dN/dS ratio of the Cyt-b gene showed a strong selection signal at node "C2 + C3", possibly driven by aridity intensification resulting from the development of the Benguela Current, 8-10 Mya. A similar signal was observed at C3, possibly due to the same driving force. These findings suggest that environmental selection pressure favored those groups, and that further cladogenic events were possible. Selection pressure was also found to be high at C1, C4 and C7, which may indicate that they are also favored by the current selection pressure The codon-based phylogeny did not retrieve any potentially new candidate taxa, but nonetheless provided strong additional support for the validity of the seven distinct clades retrieved. The amino acid sequence-based phylogeny generally supported the seven clades as valid putative species, despite the relationships among the clades not perfectly fitting the DNA sequence-based phylogenetic assumptions. Some of the phylogenetic relationships among the clades revealed by the amino acid dataset may, however, be artifacts given the more limited number of informative sites. Investigation at the genomic scale could, however, help to solve the issue. In general, I found the codon, dN, dS, Tv, Ti and amino acid sequences based phylogenetic inferences were useful in species delimitation, and should be considered in species delimitation generally.

5.1 Introduction

5.1.1 Background of the *P. tentorius* species complex

Southern Africa is the only country in the world with four biodiversity hotspots, and home to over 30% of the world's tortoise species (Cunningham, 2002; Branch, 2008; Hofmeyr, Boycott and Baard, 2014). This includes the tent tortoise (*Psammobates tentorius*) species complex with its highly variable and confusing carapace and plastron patterns, which has precluded the development of a stable taxonomy for the group for decades (Hewitt, 1933 and 1934; Loveridge and Williams, 1957; Boycott and Bourquin, 2000; Branch, 2008; Hofmeyr, Boycott and Baard, 2014). Recently Zhao et al. (2020), however, discovered seven distinct lineages in the complex using multiple DNA sequence based phylogenetic analyses and species delimitation approaches. The results implied that some of the lineages may deserve elevation to species level. Furthermore, among the seven gene loci used in the study by Zhao et al. (2020), the Cyt-b and ND4 genes were found to be the most informative loci for phylogenetic inference in the complex.

5.1.2 Selection analyses

Evolutionary theory proposes selection pressure as the main driving force shaping genetic variation among organisms (Poon et al., 2009). In all protein-coding genes, among all mutation events, non-synonymous substitutions can directly lead to functional protein changes after gene translation (Poon et al., 2009) if the substitution occurs in the exon region without being spliced out during the post-transcription period (Patthy, 1994). Thus, a non-synonymous substitution (dN) is highly likely to impact the fitness of an organism more than a silent mutation (a synonymous substitution - dS) (Poon et al., 2009; Xia, 2013).

An environmental change may result in one amino acid being replaced by a favorable amino acid under certain selection pressures. Eventually the favorable amino acid can become the fixed form as the new protein becomes part of a new trait, which survives (Poon et al., 2009; Xia, 2013). This process usually takes place over a long period of time; however, sometimes such replacements can take place much faster in response to a sudden environmental selection pressure change (Xia, 2013). Two kinds of selection pressures were used to measure selection direction and strength, namely, negative and positive selection pressure. The relative ratio of

dN/dS is an important indicator for measuring selection pressure (Poon et al., 2009). Under negative selection pressure, dN/dS < 1, certain selection pressures do not favor a mutation, whereas under positive selection pressure, dN/dS > 1, certain selection pressures favor the mutation (Poon et al., 2009). Among the three codon positions, each evolves at a different rate and encodes different functional units (Poon et al., 2009; Xia, 2013).

In the vertebrate mitochondrial genome, second codon positions are the functional sites that evolves the slowest of all three codons, and substitution changes at these positions will eventually lead to non-synonymous substitutions (Xia, 1998). By contrast, third codon positions evolve the fastest of all, and nucleotide substitutions are often synonymous (thus less functional). First codon position evolution nearly parallels amino acid replacement at both transition and transversion levels, is less functional than evolution at second codon positions but more comparable with evolution at third codon positions (Xia, 1998). Therefore, the substitution rate at first, second and third codon positions typically follows the order "third" > "first" > "second" (Poon et al., 2009; Xia, 1998).

Transitions (Ti's) and Transversions (Tv's) together with the Tv/Ti ratio are also believed to be useful indicators for evaluating selection strength and direction. The proportion of Ti's is usually much higher than the proportion of Tv's (Poon et al., 2009; Xia, 1998). In this study, I am therefore also interested in testing whether the Ti and Tv based partitions were useful in delineating putative species.

5.1.3 Amino acid based phylogenetic inferences

During transcription and translation, only the part of the protein coding nucleotide sequence sites falling within the exon region will eventually be translated into amino acid sequences and folded into functional proteins after splicing of the post-transcription stage. In addition, studying phylogenies based on amino acid sequences can also uncover valuable signals for deeply understanding evolution (Margoliash and Fitch, 1971). Furthermore, amino acid-based studies can potentially limit the interference of silent mutations in the reconstruction of phylogenetic trees. This tends to be more in line with the evolution of functional macro-molecules (amino acids in this case), than nucleotide polymorphism based phylogenetic reconstruction alone (Margoliash and Fitch, 1971).

In this study the two most informative protein-coding mtDNA genes in the *P. tentorius* species complex, namely, Cyt-b and ND4 (Zhao et al., 2020), were used for performing codon-based selection analyses, as well as phylogenetic reconstruction based on amino acid datasets (Cyt-b and ND4 separately as well as combined). Since the codon-based selection analyses and amino acid sequence based phylogenetic reconstruction correlate with natural selection, they may provide meaningful insights into the correlation between selection pressure due to environmental change and cladogenesis. Furthermore, this study also aimed to investigate whether the different codons and amino acid sequences generated phylogenetic topologies congruent with those retrieved by Zhao et al. (2020) using traditional phylogenetic analyses. And also, to test whether the codon, dN, dS, Tv, Ti and amino acid based phylogeny were useful in delineating Operational Taxonomic Units (OTUs) in these slow evolving reptiles.

5.2 Materials and Methods

5.2.1 Selection analyses

In order to investigate the evolution of the *P. tentorius* species complex at each codon position and to deeply understand its molecular ecology with respect to selection, DAMBE was first used to determine the reading frame and to locate the correct order of the three codon positions for the protein-coding genes Cyt-b and ND4, respectively. Thereafter, the gene partition was separated into three codon partitions for each gene. The nucleotide polymorphism was calculated for each codon partition per gene, and PAUP* 4.0 beta used to generate a NJ tree to check the tree topology. This additional analysis was used to ensure that the codon position order was correctly assigned in the reading frame. Both the Cyt-b and ND4 genes did not show any stop codon.

Each codon was separated into a codon partition for both Cyt-b and ND4, and the phylogeny of codon partitions analyzed using BEAST V 2.4.6 in a BI analysis. Each partition parameter and substitution model was optimized using the jModeltest results, with 10 million generations and sampling every 2000 chains, discarding the first 10% as burn-in. The rest of the steps followed the same procedure as mentioned in the Bayesian phylogenetic analysis in Chapter 2. Additionally, each codon partition was also analyzed using a simple ML method (program

RDP4, Martin et al., 2015) for comparative purposes, each run with 500 non-parametric bootstrap replications. To investigate the substitution patterns for each protein-coding gene, transitions (Ti) and transversions (Tv) were separated into partitions by MEGA6 (Tamura et al., 2013) for the two genes. Each partition was analyzed by NJ with bootstrap replications of 500 using HYPHY v. 2.2.4.

HYPHY v.2.2.4 (Pond & Muse, 2005) was used to perform selection analyses in terms of both positive and negative selection pressure. The dN and dS at each node and branch of the local substitution model were calculated for the entire tree topology using HYPHY v.2.2.4. The dN and dS were separated into different partitions for each protein coding gene and the NJ method used to generate a tree for each partition using HYPHY v.2.2.4. Bootstrap support values were calculated using RDP4 (Martin et al., 2015). A likelihood ratio test (LRT) was conducted to test whether selection amongst tree topologies was in favor of using the local model (variable rate among sites). A preliminary LRT significantly confirmed the suitability of the local model (variable rate among sites) compared to the global model (uniform rate). Thus, non-synonymous and synonymous substitution rates (dN and dS, respectively) at each branch of the NJ phylogenetic tree were calculated using the local substitution model "MG94X HKY85_3x4" on Cyt-b and ND4. Preliminary results of the nDNA gene, PRLR, showed only a limited number of informative sites, particularly at the 2nd codon position, resulting in it being excluded from the analyses.

Substitution rates are not constant among sites, so heterotachy was proposed as more realistic at different substitution sites of genes (Lopez et al., 2002). In order to test whether each branch evolved differently the DAMBE software was used to perform a relative rate test, although this approach to molecular clock assumptions is controversial (Scherer, 1989). Selection strength and direction in different branches of phylogenetic trees can also be variable (Poon et al., 2009). Thus, batch file "TestBranchDNDS.bf" implemented in HYPHY v.2.2.4 (Pond & Muse, 2005) was used to test nodes for significant dN/dS ratios in comparison with other parts of the tree.

Finally, in order to compare selection patterns between the Cyt-b and ND4 genes of the *P. tentorius* species complex in terms of dN/dS distribution, dN/dS proportions, selection strength of dN/dS and proportion of codons under selection, a "dNdSDistributionComparison.bf" batch file was used, implemented in HYPHY v.2.2.4 to compare selection pattern differences between the protein-coding genes.

5.2.2 Amino acid based phylogenetic inferences

Protein-coding genes were translated into amino acid sequences using DAMBE. MrBayes software was again used to perform amino acid sequence based BI phylogenetic analyses with the Cyt-b and ND4, respectively. Each protein-coding mtDNA gene was translated using the mammalian mitochondrial database, as no reptile-specific amino acid analysis model is available; reptiles being phylogenetically comparatively close relatives of mammals (McNab, 1978; Gauthier et al., 1988; Hedges & Poling, 1999). Each gene alignment was run with a sampling chain length of 12 million generations, and sampled every 5000 generations, with the first 25% discarded as burnin. The methods and thresholds used to monitor diagnostic convergence and sample mixing, were the same as that in the traditional BI phylogenetic analysis mentioned in Chapter 2. MEGA 6 (Tamura et al., 2013) was used to perform ML analysis on the amino acid sequences, with 500 bootstrap replications, using 4 discrete Gamma categories with a full amino acid substitution model, mtREV+I+G. The search method used was the heuristic NNI method. The initial tree was automatically enforced by maximum parsimony for all the protein coding genes. Additionally, ancestral inference analysis was performed using MEGA6 (Tamura et al., 2013) to reconstruct an amino acid based phylogenetic tree (an ML tree generated using MEGA6 as above, using hypothetical ancestral amino acidic traits with an ML approach). Each gene was run with the full mtREV+I+G model. To compare the topological variation between nucleotide sequence and amino acid sequence based phylogenies, MrBayes was used additionally to perform BI analyses on the Cyt-b+ND4 matrix. The dataset was partitioned by the three codon positions. The rest of the prior settings for each run was the same as that of the analysis of the amino acid sequences.

5.3 Results

5.3.1 Selection and codon analyses

The Tv partition-based tree of the Cyt-b gene (Supplementary Figure S5.1) failed to retrieve phylogenetic relationships among the clades in general, despite the monophyly of all clades being strongly supported (BP > 70). The Ti tree was almost identical to the full tree (Ti + Tv).

Notwithstanding this, the Tv tree showed a higher transversion substitution rate and Tv/Ti ratio at the node where C2 and C3 bifurcated.

The Tv partition of the ND4 based NJ tree successfully retrieved six clades: C1-C6, each with strong support (BP > 70) (Supplementary Figure S5.2), but not for C7. The clusters "C1 + C4 + C5 + C6" and "C2 + C3" were strongly supported by the phylogeny based on the full dataset (Ti + Tv: BP > 70). Outgroup members were generally well supported in comparison to the full tree, which may imply that Tv of ND4 is sufficiently informative to separate taxa at species level. In the Ti partition, nodes "(C1 + C4) + C7) + C5" and "(C2 + C3) + C6)" were generally supported which was incongruent with the tree based on the full dataset.

For Cyt-b, the codon 1 tree showed strong conflicts (Fig. 5.1) in comparison with the topology of the Cyt-b gene in the traditional phylogenetic analysis. Clade 5 to C7 was strongly supported by both ML and BI analyses (BP > 70 and PP > 0.95), whilst "(C1 + C4) + C7) + C5" and C6 grouped as a cluster, but with only weak support (PP < 0.95, BP < 70). The monophyly of C1 and C4 was not supported. In terms of evolutionary rate, C6 had the fastest rate of all.

For the codon 2 tree of the Cyt-b gene (Fig. 5.1), five of the expected seven clades were retrieved, and only the monophyly C2 violated, since C2 and C6 seemed to cluster as one group. The relationships among the clades were generally incongruent with that of the full tree of Cyt-b, since most of the nodes were poorly supported. In terms of the evolutionary rate at the codon 2 position, C5, C6 and C7 were shown to evolve at similar rates but faster compared with the rest, whilst two samples of C1 from Matjiesfontein also evolved at a uniquely high rate in comparison with the other clades. Remarkably, in the full tree, *Psammobates oculifer* and C6 evolved much faster than the rest. However, at the second codon position, C5, C6 and C7 still showed relatively fast evolutionary rates, whilst *P. oculifer* seemed to evolve relatively slowly compared with many of the other taxa. Furthermore, although C1 showed a relatively slow evolutionary rate in the full tree, its evolutionary rate was relatively high for the second codon position, which implies higher non-synonymous mutations.

When comparing the trees of codon 1 + 2 and codon 3 for Cyt-b, their tree topologies differed, with C5 grouping with C6, and cluster (C2 + C3) not being supported. Codon 3 generally had a stronger support value at each node, except for the placement of C6, which grouped with (C2 + C3) with weak support (BP < 70) (Fig. 5.1). All seven clades were generally supported for codon 1, 3 and codon 1+2 positions. The support value for each node of the full tree was high and the relationship between nodes strong; only C6 presented a conflict.

For the ND4 gene (Fig. 5.2), the codon 1 partition retrieved seven clades, C1-C7. Clade 2 to C7 each had strong support (PP > 0.95; BP > 70), with only the ML analysis showing weak support for C1 (BP < 70). The node "(C1 + C4) + C7) + C5" was strongly supported (PP > 0.95; BP > 70), while "(C2 + C3) + C6" was not supported (PP > 0.95; BP > 70). The grouping of *Stigmochelys pardalis* with C2 was unexpected, but could be an artifact due to low substitution information at codon 1.

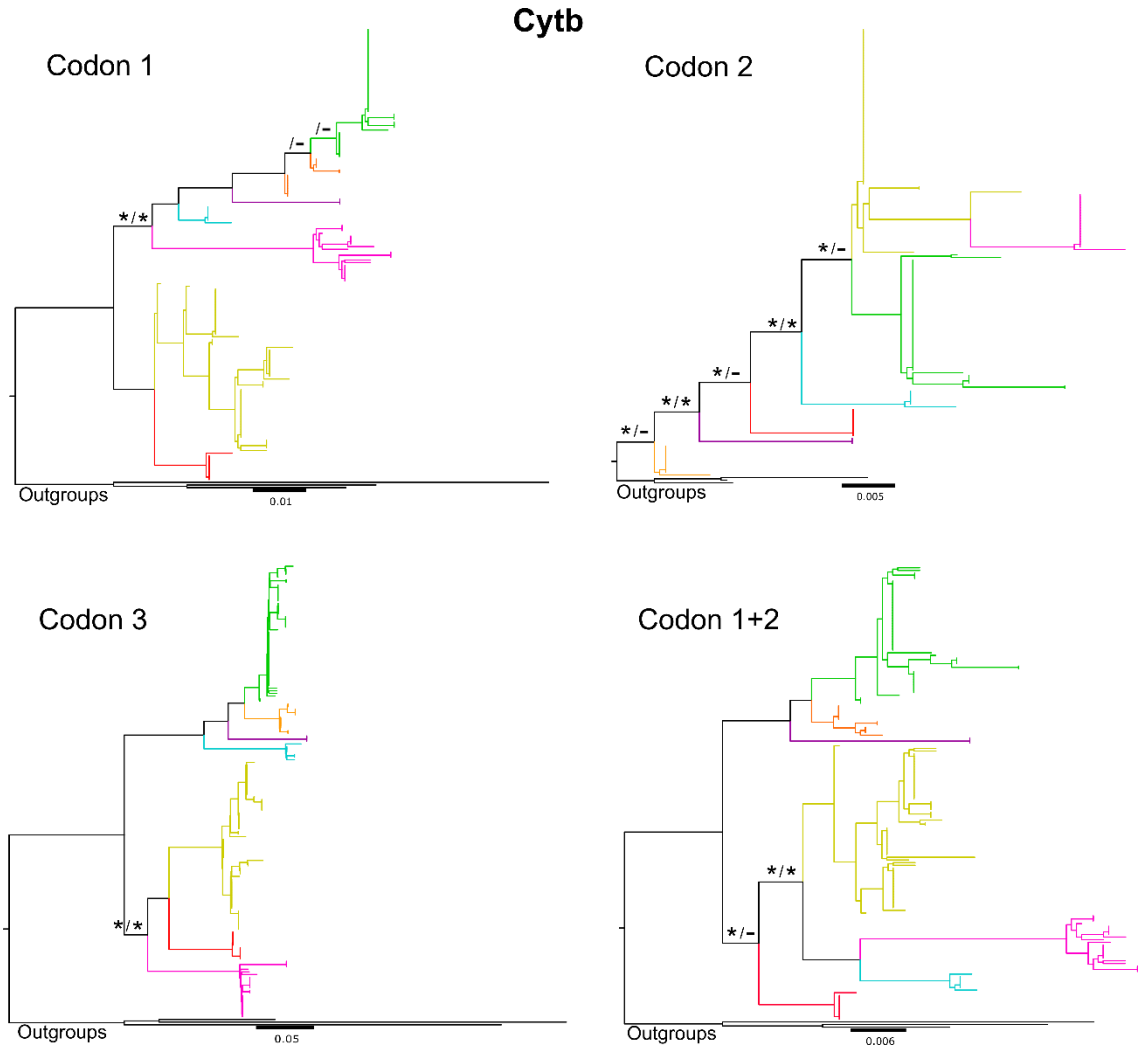


Figure 5.1. ML and BI trees based on different codon partitions for Cyt-b gene. The colour scheme: "green" represents Clade 1 (C1), "orange" represents C4, "yellow" represents C2, "purple" represents C7, "aqua" represents C5, "red" represents C3 and "pink" represents C6. The same colour scheme will be used in other figures throughout this study. Values on the left side refer to bootstrap support (BP) for the ML analysis, and on the right side the BI posterior probability (PP). The " * " represents weak support: PP < 0.95 or BP < 70, " - " indicates node not supported with corresponding approach, while its absence means strong branch support (BP > 70 or PP > 0.95).

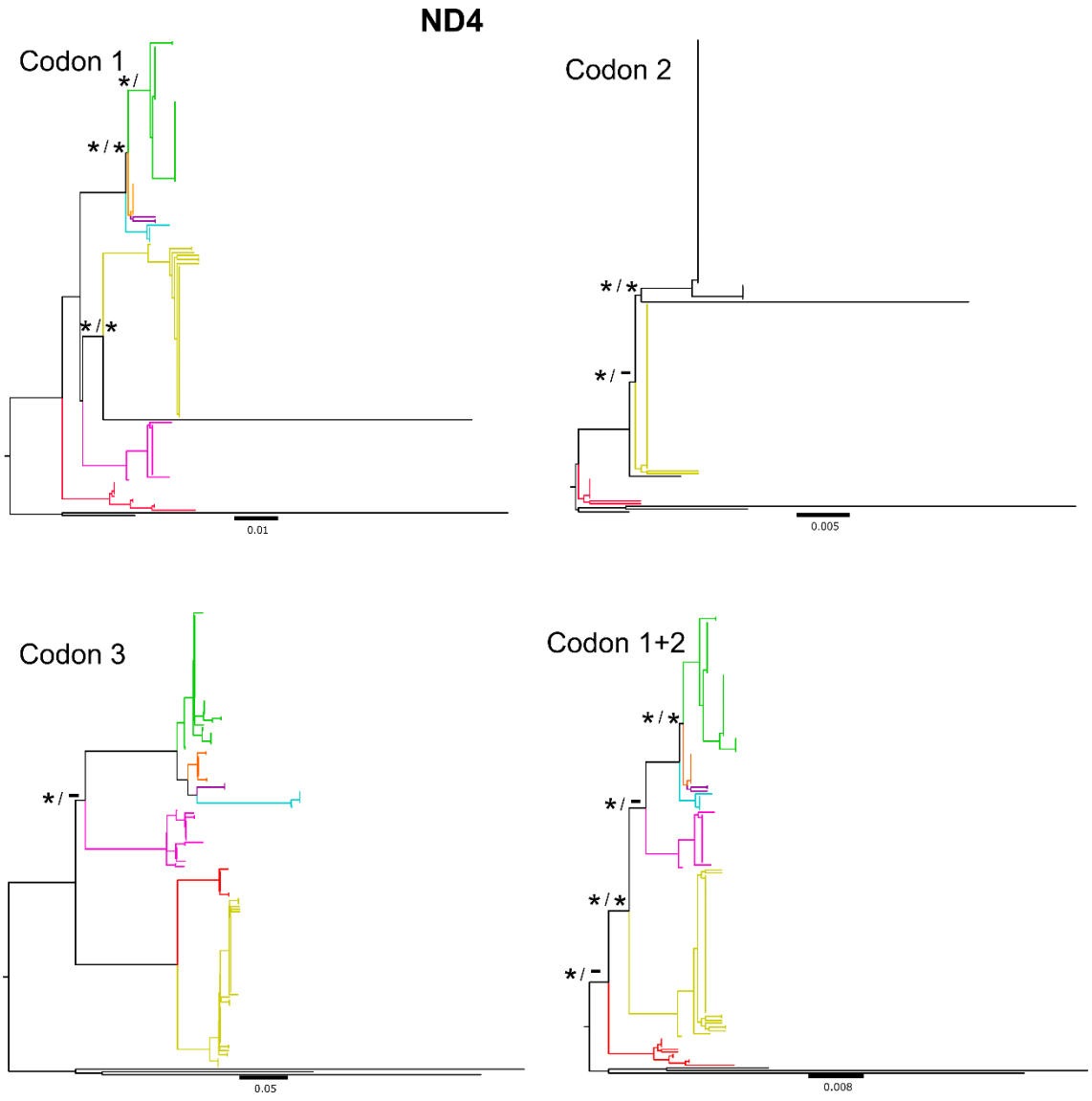


Figure 5.2. ML and BI trees based on different codon partitions for ND4 gene. Values on the left side refer to bootstrap support (BP) for the ML analysis, and on the right side the BI posterior probability (PP). The " * " represents weak support: PP < 0.95 or BP < 70, " - " indicates node not supported with corresponding approach, while its absence means strong branch support (BP > 70 or PP > 0.95).

Codon 2 strongly supported only the monophyly of C2 and C3. The monophyly of *P. tentorius* and *Psammobates* was violated, as *S. pardalis* grouped with "C1 + C4 + C5 + C6" (although the clade support value was low, PP < 0.95, BP < 70). Outgroup member *S. pardalis* showed

higher evolutionary rates at the second codon position. Evolutionary rates within the *P. tentorius* species complex all appeared to be at the same level. Codon 3 strongly supported the seven clades, however, both codon 1 and codon 1 + 2 showed a topological conflict regarding the placement of C3 when compared with the topology of the full tree. Codon 3 showed a topology similar to the full ND4 gene tree with generally strongly supported clusters "C6 + (C5 + (C1 + C4))" and "C2 + C3".

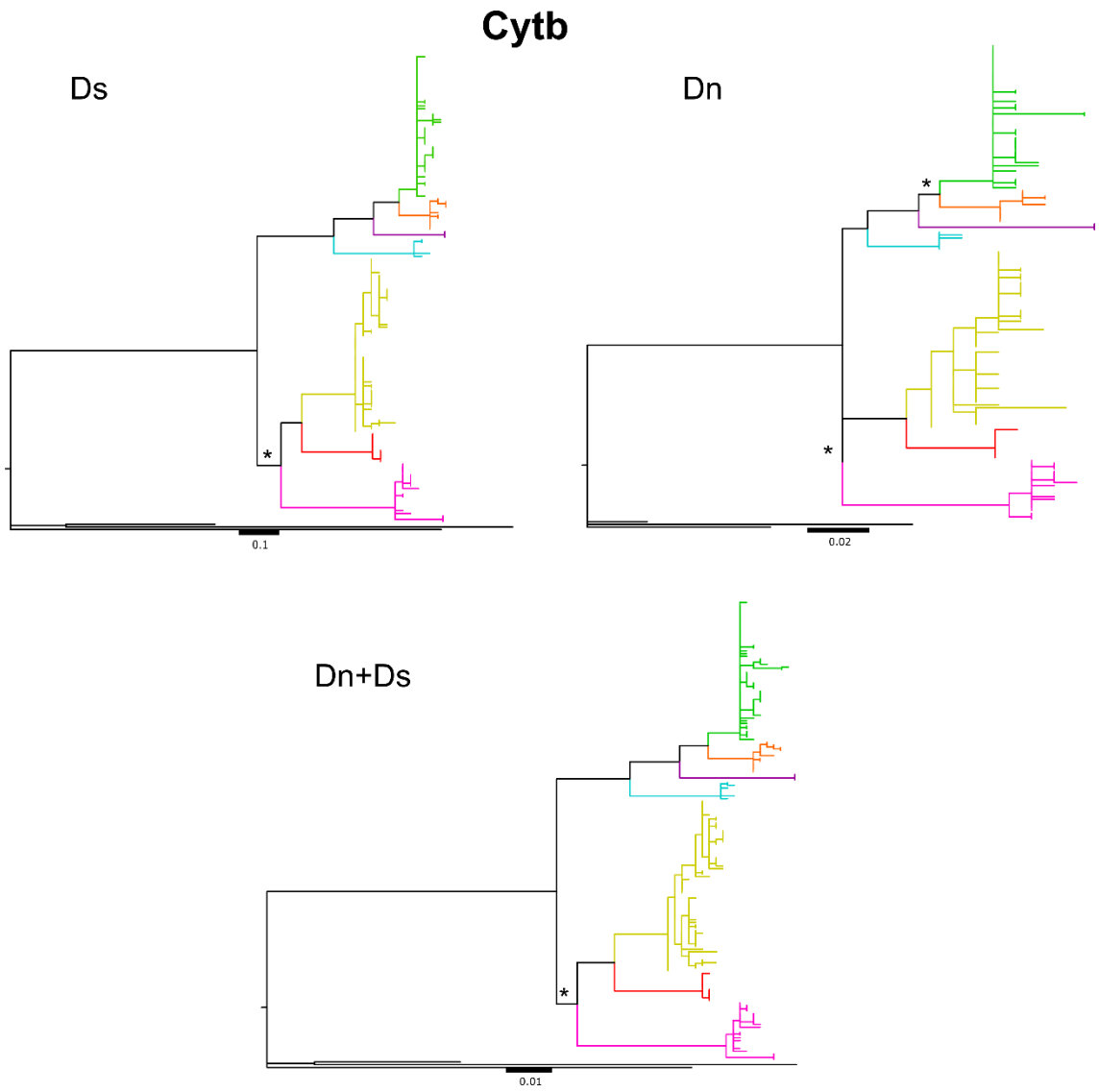


Figure 5.3. NJ trees for the Cyt-b gene for the full partition (dN+dS), dN partition and dS partition, respectively. Bootstrap support is indicated above each branch. The " * " represents weak support: PP < 0.95 or BP < 70, " - " indicates node not supported with corresponding approach, while its absence means strong branch support (BP > 70 or PP > 0.95).

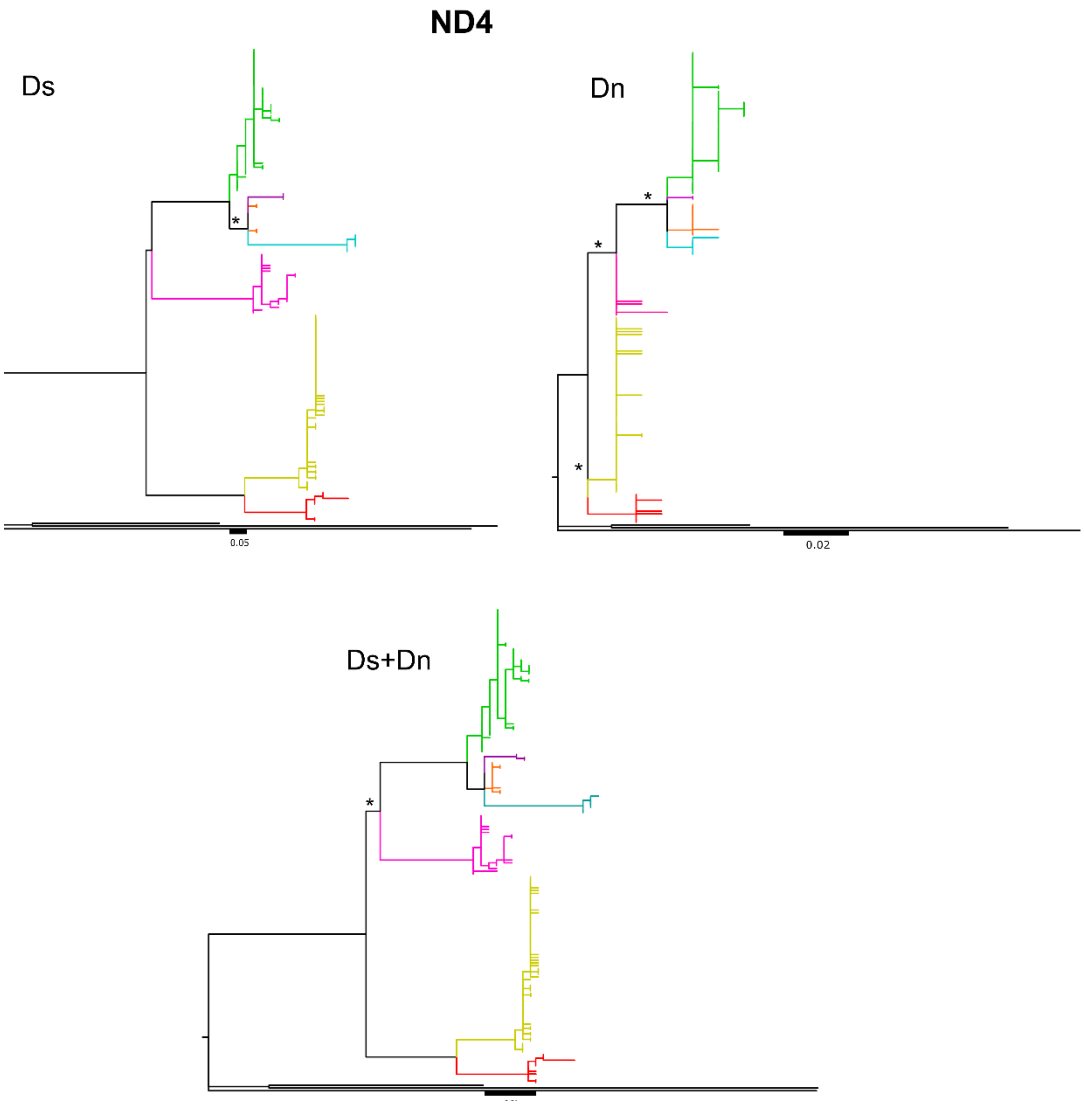


Figure 5.4. NJ trees for the ND4 gene for the full partition (dN+dS), dN partition and dS partition, respectively. Bootstrap support is indicated above each branch. The " * " represents weak support: PP < 0.95 or BP < 70, " - " indicates node not supported with corresponding approach, while its absence means strong branch support (BP > 70 or PP > 0.95).

5.3.2 dN/dS based selection analysis

For the Cyt-b gene (Fig. 5.3), no node in the tree topology was detected with a strong favorable selection signal (Table 5.1). Nevertheless, the favorable selection strength (dN/dS ratio) at C1,

C4, C6, C7 and node (C2 + C3) was found to be significantly greater than that at the rest of the nodes (Table 5.1), particularly that of (C2 + C3). The non-synonymous substitution tree still supported all seven clades, but again with the placement of C6 as uncertain.

Table 5.1. The synonymous rates (dS), non-synonymous rates (dN) and selection strength, dN/dS, across all nodes; neighbour-joining tree used as backbone tree.

Locus	Cyt-b			ND4		
Object	dS	dN	dN/dS	dS	dN	dN/dS
Node						
Pg	0.3654	0.0193	0.053	0.5366	0.0421	0.078
Po	1.0943	0.1044	0.095	1.3339	0.1209	0.091
Pt	0.6044	0.0819	0.136	0.4632	0.0092	0.020
(C2+C3)+C6	0.0589	0	0.000	NS	NS	NS
(C1+C4)+C7+C5	0.1877	0.0081	0.043	0.2228	0.0156	0.070
(C1+C4)+C7)	0.0978	0.0164	0.168	NS	NS	NS
C1+C4	0.0621	0.0068	0.110	NS	NS	NS
C2+C3	0.0503	0.0206	0.410	0.2832	0	0.000
C1	0.0439	0.0171	0.390	0.0216	0.0078	0.361
C2	0.1314	0.0079	0.060	0.1553	0.0086	0.055
C3	0.1737	0.0284	0.164	0.1768	0.0147	0.083
C4	0.0761	0.0194	0.255	0.1342	0	0.000
C5	0.1967	0.0231	0.117	0.2841	0.0078	0.027
C6	0.2793	0.0536	0.192	0.2908	0	0.000
C7	0.1742	0.0565	0.324	0.1009	0	0.000

Note: "Pg" = *Psammobates geometricus*; "Po" = *Psammobates oculifer*

"Pt" = *Psammobates tentorius*

"NS" = unsupported

In the case of the ND4 gene (Fig. 5.4), no positive selection signal (node with dN/dS > 1) was found throughout its tree topology, with the highest selection strength at C1 (Table 5.1). Nevertheless, the additional codon based phylogenetic information is still valuable (in terms of determine the clades). The non-synonymous tree supported the monophyly of all seven clades, but the relationships between them was poorly resolved. Only (C1 + C4) + C7) + C5 was supported as a lineage as well as its sister group (C2 + C3), while the placement of C6 remained unclear (Table 5.1).

When comparing selection patterns between the Cyt-b and ND4 genes, the dN/dS based distribution comparison tests revealed: 1) that there was a significant difference in their dN/dS distribution (LR = 24.184, df = 10, p = 0.007); 2) that the selective regimes (dN/dS and proportions) did not differ significantly between them (LR = 2.699, df = 2, p = 0.259); 3) that the selection strengths (dN/dS) were not different across the two genes (LR = 0.378, df = 1, p = 0.828); and 4) that the proportions of codons under selection did not differ significantly between them (LR = 1.958, df = 1, p = 0.162). These results may imply that the two genes were under similar selection pressure.

5.3.3 Amino acid phylogenetic analysis

The amino acid phylogenetic tree for Cyt-b (Fig. 5.5) showed strong support for clades C3, C4, C5, C6 and C7 (all with PP > 0.95, BP > 70), while C2 was not supported and C1 was only weakly supported (PP < 0.95, BP < 70). The relationship between C1 and C4 + C5 was not supported (PP > 0.95, BP > 70). When comparing the tree topology with the Cyt-b gene tree, topological conflicts occurred at C2, C3 and C4. The amino acid phylogram did not support C2 and C4 as monophyletic groups. Nevertheless, C2 + C6 and C3 still clustered as sister groups with strong support (PP > 0.95, BP > 70). The grouping of C5 with C1, C2, C3, C7 and C6 was unexpected.

The amino acid tree for the ND4 gene (Fig. 5.5) exhibited strong support for clades C3, C4, C5 and C6 (all with PP > 0.95, BP > 70), while C1and C2 were only weakly supported (PP < 0.95, BP < 70). The relationship between C1, C7 and C4 + C5 was also strongly supported (PP > 0.95, BP > 70). When comparing the tree topology with the ND4 gene tree, topological conflicts occur at C2 and C3, while the amino acid phylogram did not support C2 as a monophyletic group. Nevertheless, C2 and C3 still clustered as a sister group with strong support (PP > 0.95, BP > 70).

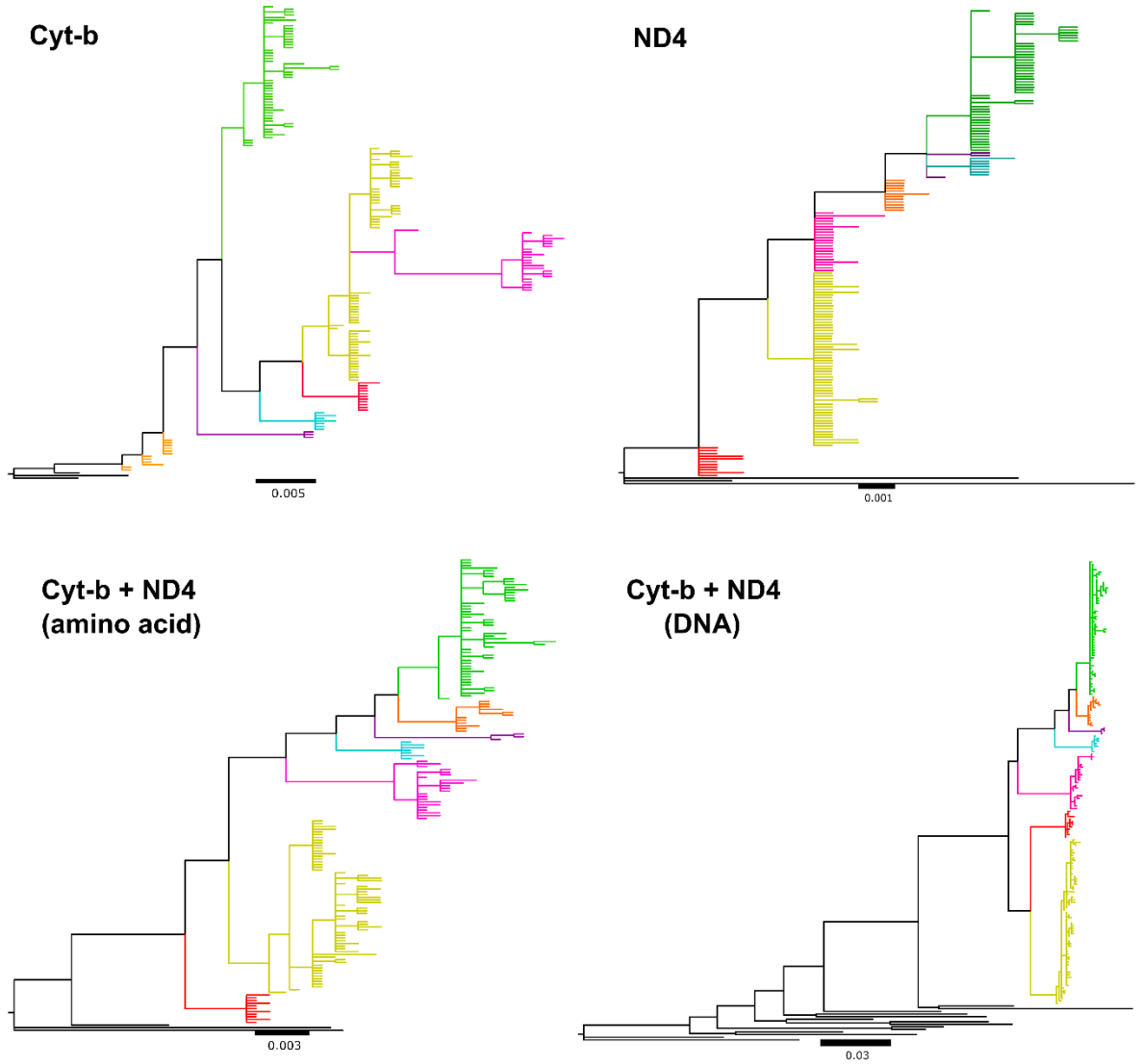


Figure 5.5. Amino acid sequence phylograms for Cyt-b, ND4, Cyt-b+ND4 and the DNA sequence phylogram for Cyt-b+ND4. The same colour scheme as used previously was used to highlight the seven clades. The tree topologies used are the consensus phylograms generated from BI (MrBayes) analysis. " * " indicates weak branch support (BP < 70 or PP < 0.95), while its absence means strong branch support (BP > 70 or PP > 0.95) from the ML and BI analyses.

The ND4 and Cyt-b combined amino acid tree (Fig. 5.5) retrieved all seven clades as the DNA sequence dataset advocated, each node with strong support (PP > 0.95, BP > 70). The relationship among clades, however, was not fully congruent when compared with the DNA sequence based phylogeny. Clade 3 was positioned as a basal clade, rather than clustering with C2 as sister clade. In the Cyt-b + ND4 dataset, the amino acid tree and DNA sequence tree

showed similar topologies, the only difference being the placement of C3. In the amino acid tree, C3 was placed as basal node, whilst, it was the sister clade of C2 in the DNA sequence tree (Fig. 5.5).

The Transversion/Transition ratio (Tv/Ti) was found to be significantly higher in "C2 + C3" and C6 for Cyt-b, whist, the Tv/Ti ratio was higher in group "(C1 + C4 + C7) + C5", "C2 + C3" and C3 for the ND4 gene (see Supplementary Table S5.3).

5.4 Discussion

5.4.1 Selection analyses

The selection analysis results showed that the dN/dS between Cyt-b and ND4 did not differ significantly with respect to selection regimes (dN/dS and its proportions), selection strengths (global dN/dS) and the portions of codons under selection but differed in selection distribution. Thus, it may be concluded that the selection pressure variation on these genes can be disregarded. It may therefore reasonably be deduced that the selection pressures differ on the different phylogenetic groups comes from their respective environments. The amino acid sequences of Cyt-b and ND4 were therefore combined for the phylogenetic inferences.

The transversion partitions of both Cyt-b and ND4 (Supplementary Figures S5.1-5.2) supported the proposition that C2 and C3 were derived from the same common ancestor. This also provided additional support for the sister relationship between C2 and C3. Despite the very conservative transversion rate, the transversion tree for both Cyt-b and ND4 generally retrieved seven clades, which matched the clustering scheme revealed in the mtDNA trees. This could be considered as evidence that the seven clades are indeed distinct from each other. In general, the 3rd codon and 1st + 2nd codon partitions of both Cyt-b (Fig. 5.1) and ND4 (Fig. 5.2) retained sufficient levels of polymorphism to support the monophyly of the seven clades. The second codon position of both genes was much less informative, particularly in the ND4 gene, where it did not even support the monophyly of *Psammobates*. The 2nd codon position in Cyt-b did not support monophyly of C2 + C3, but rather that C2 was mixed with C6 as a single cluster, despite the monophyly of C6 being well supported.

The codon analyses failed to discover more taxa than the traditional phylogenetic analyses. It nevertheless provided additional support for the stability of the seven clades. The unexpected systematic positions of C5 and C7 in the codon-based analyses were perhaps merely artefacts due to insufficient informative sites. Notwithstanding this, it could also be considered as evidence that C5 and C7 are different from C1. As all the substitutions at the 2nd codon position will eventually lead to protein changes (if the substitution takes place in the exon region) (Xia, 1998), the transversion based phylogeny derived from the Cyt-b gene could indicate that C1, C5, C6 and C7 were under similar selection pressure. However, such a coincidence is difficult to interpret given their completely different environments.

The transition and transversion ratio-based analysis supported the distinctiveness at C3 and C6, and also revealed their high degree of genetic isolation, which implies that they may deserve to be treated as separate OTUs to the other clades.

According to the ND4 codon analysis at 2nd codon (Fig. 5.2), C1, C4, C5 and C6 grouped together with *S. pardalis*. Clade 2 and C3 remained as monophyletic clades, which may imply that in the case of the ND4 gene, C6 (north of Orange River) evolved under similar selection pressures as C1, C4 and C5 (despite the monophyly of these groups not being supported). The monophyly of C2 and C3 at the 2nd codon partition may be regarded as additional support for elevating both of them to species level.

In terms of the dN and dS based selection analyses, both genes (Cyt-b and ND4) generally supported the relationships of "C2 + C3" and "C1 + C4 + C5", contrary to their observed placement in relation to C6. Again, these results provide additional evidence supporting the proposed "seven species assumption" in the *P. tentorius* species complex.

For the Cyt-b gene, both dN and dS partitions supported the monophyly of the seven clades, albeit weak in some cases. The dN/dS based analyses suggest significantly higher positive selection pressure on C3, C6, C7 and the congeneric outgroup *P. oculifer*. The relatively high positive selection strength on C3 could be explained as adaptation to the formation of the Succulent Karoo biome under the influence of the cold Benguela current along the west coast region (Partridge, 1997), the increasing severity of summer droughts (Partridge, 1997) and the global cooling trend after the mid Miocene (Zachos et al., 2001). In addition to these the formation of the Kamiesberg as a geographical barrier between the West Coast Succulent Karoo (west of the Kamiesberg) and inland Succulent Karoo and Nama Karoo (east of the Kamiesberg) probably also facilitated the isolation between the two populations.

The rapid evolution of *P. oculifer* was also reported by Cunningham (2002), which could have been caused by the geographic isolation from the rest populations. The Orange River maybe function as a vicariance barrier, as *P. oculifer* only occurs north of the Orange River (Branch, 2008; Hofmeyr, Boycott and Baard, 2014; Hofmeyr et al., 2017), and is generally found in Savanah biomes with relatively high humidity (in terms of rainful, Branch, 2008; Mucina & Rutherford, 2006). The faster rate of evolution in *P. oculifer* may also be correlated to its unique physiological properties in terms of respiration, thermoregulation, or muscle function. Variation in humidity and vegetation type may thus function as additional factors which drive and accelerate cladogenesis in *P. oculifer* (da Silva et al., 2012). These assumptions obviously require verification through further studies. Apart from these, the expansion of the Kalahari Desert (Partridge and Maud, 1987) could also have influenced the diversification of *P. oculifer*. As the phylogenetic results also suggested, there is a sister relationship between *P. oculifer* and *Psammobates geometricus*. The *P. geometricus* occurs exclusively in the Fynbos biome (Branch, 1998 and 2008; Hofmeyr, Boycott and Baard, 2014; Hofmeyr et al., 2017), both habitats having higher moisture levels in comparison to the drier Karoo habitat of *P. tentorius*. This may imply that humidity and precipitation also facilitate rapid cladogenesis in *P. oculifer*. However, these assumptions are merely preliminary hypotheses based on molecular evidence and will require testing using experiments or simulations. The apparently higher dN/dS ratio found in C6 may be adaptative responses to the Kalahari desert expansion and increasing aridification in recent time (Partridge and Maud, 1987). Coincidentally, the ND4 gene showed similar results, except that the selection pressure on C6 was not statistically significant. The high level of within-group divergence and high cladogenic radiation potential has been addressed by previous study (Zhao et al., 2020), which seems to correspond to the findings of this study.

The selection analysis revealed relatively high (Table 5.1) selection strength on node (C2+C3), which may suggest that cladogenesis between C2 and C3 was favored by current environmental selection. The most likely explanation for this cladogenic event would be the increasing western aridity and the development of winter rainfall due to intensification of the Benguela Current (Rommerskirchen et al., 2011; Hoffmann et al., 2015). These conditions favored the development of Succulent Karoo vegetation along western South Africa from about 10-8 Ma, although the Succulent Karoo Biome became fully established only in the Pliocene (Verboom et al., 2009; Neumann and Bamford, 2015). These findings match results obtained from calibration dating and biogeographical analyses (Z. Zhao et al, unpublished data).

5.4.2 Amino acid based phylogenetic analysis

The results of the amino acid sequence based phylogeny obtained from the Cyt-b gene (Fig. 5.5) supported the monophyly of all the clades, except C4 which showed a polytomy. Notwithstanding this, significant genetic distance between members of C1 and C4 is still distinguishable. It was expected that the tree topology of the second codon position would be closer to that of the amino acid tree. However, the results did not show clear congruence between the second codon position tree and amino acid tree, for both genes. A possible interpretation of these findings would be that part of the non-synonymous substitutions came from mutations of the first codon position, though all the substitutions that took place at the second codon position will eventually be translated into changes in protein sequences. Interestingly, the amino acid based phylogeny showed that C5 was sister to "C2 + C3", rather than to "C1 + C4". This may be an indication of the unique evolution of C5 in comparison to "C1 + C4", which may be considered as additional support for C5 to be regarded as a distinct species rather than part of C1.

The amino acid sequences of the ND4 gene (Fig. 5) gave results generally compatible with that of the DNA sequences, except that C3 was presented as basal node in the phylogeny, and C4 was not the sister clade of C1 as expected. This again emphasizes the distinctiveness of C1, and C2, C3 and C4.

The Cyt-b + ND4 combined amino acid dataset confirmed all the clades as being distinct, despite C3 being positioned as basal clade, rather than the sister clade to C2, Again, this phenomenon could be an artefact due to fewer informative sites in comparison to the DNA based phylogeny. Further genome scale DNA and amino acid (translated from genomic sequence data) based phylogenetic analyses would be needed to hopefully finally clarify the relationships among the clades.

5.5 Conclusions

No further subdivisions were detected from the codon and amino acid based phylogeny, compared to the phylogenetic analysis results obtained using the traditional molecular methods by Zhao et al. (2020). Nonetheless, the codon and amino acid-based phylogeny generally provided additional support for the seven distinct evolutionary lineages, each of which should be regarded as a putative species and valid OTU. The selection analysis found that C3 had the highest fitness under the current selection pressures (based on the dN/dS ratios). This implies that the increasing west coast aridity and the development of winter rainfall due to the intensification of the Benguela Current possibly facilitated the divergence between C2 and C3, and may facilitate further radiation within C3. These assumptions should, however, only be regarded as preliminary, since experimental verification is needed for greater certainty. Finally, the high congruence of the dN/dS ratio based patterns between the two mtDNA markers (Cyt-b and ND4) may suggest that selection pressure on the two genes is similar.

Finally, this study revealed that the partitions based on codons, dN, dS, Ti, Tv and amino acid sequences are informative and useful for delineating OTUs, and deserves to be considered as a general approach for species delimitation in all organisms with protein-coding genes.

Supplementary Figures

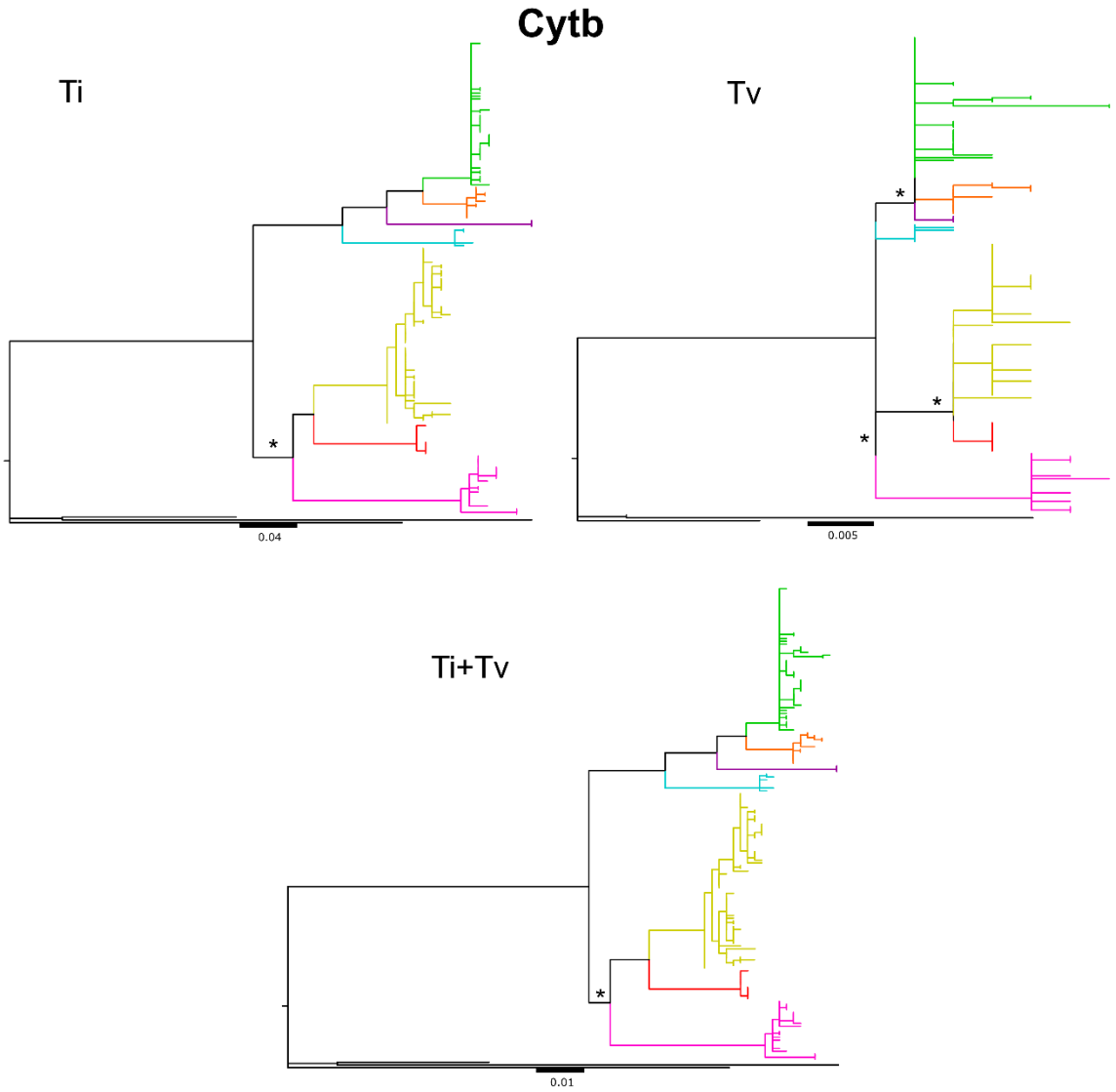


Figure S5.1. The NJ based trees for Cyt-b for Ti, Tv and Ti + Tv, respectively; bootstrap support is indicated above each branch: all nodes without " * " had strong support, whereas nodes with " * " had low support (BP < 70). The colour scheme is the same as the one used in the traditional phylogenetic analysis.

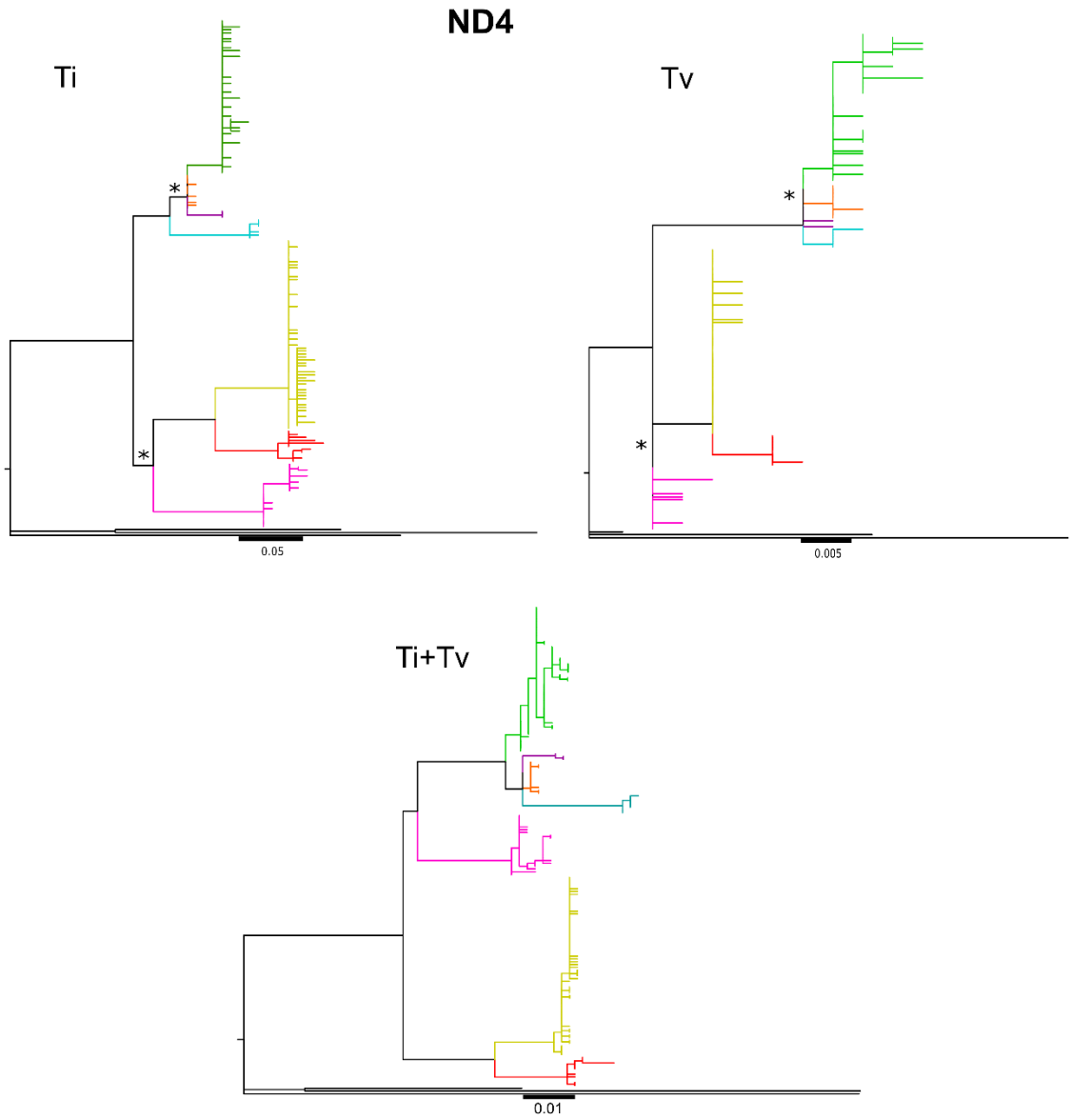


Figure S5.2. The NJ based trees for ND4 for Ti, Tv and Ti + Tv, respectively; bootstrap support is indicated above each branch: all nodes without " * " had strong support, whereas nodes with " * " had low support (BP < 70). The colour scheme is the same as the one used in the traditional phylogenetic analysis.

Supplementary Tables

Supplementary Table S5.1. List of all samples, their corresponding localities and NCBI GenBank accession numbers across different genes. Note: this table is too large in size, given separately in Excel document.

Supplementary Table S5.2. GenBank accession numbers for all outgroups used in the study.

	NCBI Gene Bank Accession Number	
Outgroup sample	***Cyt-b***	***ND4***
Gopherus sp.	AY434562	AY673591
Kinixys belliana	DQ497312	HE662266
Kinixys spekii	AY678398	HE662310
Manouria emys	AY434563	KX355558
Pyxis arachnoides	DQ497319	KX346344
Astrochelys radiata	DQ497304	KX346345
Psammobates geometricus	AY678371	KX346363
Psammobates oculifer	MH429624	MH429618
Stigmochelys pardalis	MH429625	MH429619
Chersina angulata_1	MH429620	MH429614
Chersina angulata_2	MH429621	MH429615
Chersobius boulengeri	DQ497308	KX346358
Chersobius signatus	DQ497309	GU139232
Homopus areolatus	AY678323	KX346350
Homopus femoralis_10771	MH429622	MH429616
Homopus femoralis_175	MH429623	MH429617

Supplementary Table S5.3. The transition rate (Ti), transversion rate (Tv) and selection strength Tv/Ti across all nodes; Neighbour-joining tree used as backbone tree.

Locus	Cyt-b			ND4		
Object	Ti	Tv	Tv/Ti	Ti	Tv	Tv/Ti
Node						
Pg	0.2716	0	0.0000	0.3019	0.0033	0.011
Po	0.3251	0.0307	0.0944	0.1745	0.0281	0.161
Pt	0.1686	0.0225	0.1335	0.0952	0.0063	0.066
(C2+C3)+C6	0.0275	0	0.0000	0.0157	0	0.000
(C1+C4+C7)	0.0308	0.0029	0.0942	0.0135	0	0.000
(C1+C4+C7)+C5	0.0618	0	0.0000	0.0283	0.0149	0.527
C1+C4	0.0249	0	0.0000	NS	NS	NS
C2+C3	0.0145	0.0059	0.4069	0.0478	0.0059	0.123
C1	0.0332	0	0.0000	0.0271	0	0.000
C2	0.0507	0	0.0000	0.0567	0	0.000
C3	0.0711	0.0029	0.0408	0.0484	0.0059	0.122
C4	0.0304	0.0029	0.0954	0.0135	0	0.000
C5	0.0777	0.0029	0.0373	0.0621	0.0029	0.047
C6	0.1163	0.0117	0.1006	0.0851	0	0.000
C7	0.1001	0.0029	0.0290	0.0271	0	0.000

Note: "Pg" = *Psammobates geometricus*; "Po" = *Psammobates oculifer*

"Pt" = *Psammobates tentorius*

"NS" = not supported

References

Branch, B. (1998). *Field guide to snakes and other reptiles of southern Africa*. Sanibel Island (FL): Cape Town: Ralph Curtis Books.

Branch, B. (2008). *Tortoise, Terrapins and Turtles of Africa*. Cape Town: Struik Publisher.

Cunningham J., (2002). *A molecular perspective on the family Testudinidae Batsch, 1788*, Ph.D. Dissertation, University of Cape Town.

da Silva, F. R., Almeida-Neto, M., do Prado, V. H. M., Haddad, C. F. B., & de Cerqueira Rossa-Feres, D. (2012). Humidity levels drive reproductive modes and phylogenetic diversity of amphibians in the Brazilian Atlantic Forest. *Journal of Biogeography*, *39*(9), 1720-1732.

Gauthier, J., Kluge, A. G., & Rowe, T. (1988). Amniote phylogeny and the importance of fossils. *Cladistics*, *4*(2), 105-209.

Hedges, S. B., & Poling, L. L. (1999). A molecular phylogeny of reptiles. *Science*, *283*(5404), 998-1001.

Hoffmann, V., Verboom, G. A., & Cotterill, F. P. (2015). Dated plant phylogenies resolve neogene climate and landscape evolution in the Cape Floristic Region. *PLoS One*, *10*(9), e0137847.

Hofmeyr, M. D., Boycott, R. C., & Baard, E. H. W. (2014). *Psammobates tentorius* (Bell, 1828). In: M. F. Bates, W. R. Branch, A. M. Bauer, M. Burger, J. Marais, G. J. Alexander & M. S. De Villiers (Eds.), *Atlas and red list of the reptiles of South Africa, Lesotho and Swaziland* (pp. 70-85). Pretoria: South African National Biodiversity Institute.

Hofmeyr, M. D., Vamberger, M., Branch, W., Schleicher, A., & Daniels S. R. (2017). Tortoise (Reptilia, Testudinidae) radiations in Southern Africa from the Eocene to the present. *Zoologica Scripta*, *46*, 389-400.

Kolaczkowski, B., & Thornton, J. W. (2004). Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. *Nature*, *431*(7011), 980.

Lopez, P., Casane, D., & Philippe, H. (2002). Heterotachy, an important process of protein evolution. *Molecular Biology and Evolution*, *19*(1), 1-7.

McNab, B. K. (1978). The evolution of endothermy in the phylogeny of mammals. *The American Naturalist*, *112*(983), 1-21.

Margoliash, E., & Fitch, W. M. (1971). Amino Acid Sequences and Phylogeny. *Taxon*, 51-53.

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., & Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution*, *1*(1).

Mucina, L., & Rutherford, M. C. (2006). *The Vegetation of South Africa, Lesotho and Swaziland*. Pretoria: South African National Biodiversity Institute.

Neumann, F. H., & Bamford, M. K. (2015). Shaping of modern southern African biomes: Neogene vegetation and climate changes. *Transactions of the Royal Society of South Africa*, *70*(3), 195-212.

Partridge, T.C., 1997. Evolution of landscapes. In: RM Cowling, DM Richardson, SM Pierce. Vegetation in southern Africa. Cambridge University Press, Cambridge, pp 5-20.

Patthy, L. (1994). Introns and exons. *Current Opinion in Structural Biology*, *4*(3), 383-392.

Pond, S. L. K., & Muse, S. V. (2005). HyPhy: hypothesis testing using phylogenies. In *Statistical methods in molecular evolution* (pp. 125-181). Springer, New York, NY.

Rommerskirchen, F., Condon, T., Mollenhauer, G., Dupont, L., & Schefuß, E. (2011). Miocene to Pliocene development of surface and subsurface temperatures in the Benguela Current system. *Paleoceanography*, *26*(3).

Scherer, S. (1989). The relative-rate test of the molecular clock hypothesis: a note of caution. *Molecular Biology and Evolution*, *6*(4), 436-441.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, *30*(12), 2725-2729.

Verboom, G. A., Archibald, J. K., Bakker, F. T., Bellstedt, D. U., Conrad, F., Dreyer, L. L., ... & Mummenhoff, K. (2009). Origin and diversification of the Greater Cape flora: ancient species repository, hot-bed of recent radiation, or both?. *Molecular Phylogenetics and Evolution*, *51*(1), 44-53.

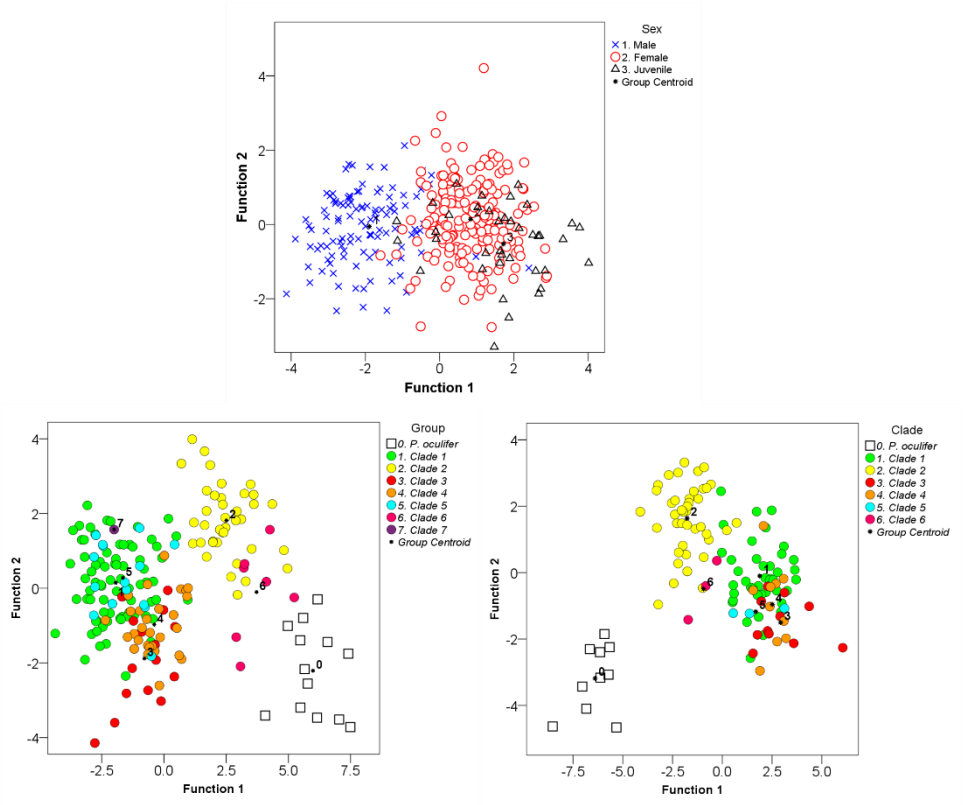
Xia, X. 1998. How optimized is the translational machinery in *Escherichia coli*, *Salmonella typhimurium* and *Saccharomyces cerevisiae*? *Genetics*, *149*, 37-44.

Xia, X. (2013). DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. *Molecular Biology and Evolution*, *30*(7), 1720-1728.

Zachos, J., Pagani, M., Sloan, L., Thomas, E., & Billups, K. (2001). Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science*, *292*(5517), 686-693.

Chapter 6. Does body size, shape and phenotypic characters of the seven clades of the highly polymorphic tent tortoise (*Psammobates tentorius*) species complex show congruence with the molecular findings with respect to its taxonomy?

Graphic abstract



Abstract

The aim of this study was to compare phenotypic and morphometric characters between the seven clades identified by DNA sequence analyses, to establish whether the morphological and genetic findings were congruent in terms of the taxonomic diversity in the *Psammobates tentorius* species complex. The phenotypic characters (discrete variables) were coded and analysed using Permutation based Multivariate Analysis of Variance (PERMANOVA). The morphometric datasets (continuous variables) were tested for normal distribution using the Shapiro-Wilks test, and log-transformed to achieve this where necessary. They were then tested for significant sexual size differences in each clade, using Analysis of Covariance (ANCOVA) which removes the effect of size. In those cases where significant sexual size dimorphism existed, male and female datasets were analysed separately. Multiple ratios were used to test for differences in body shape between sexes and clades, using Analysis of Variance (ANOVA). Discriminant Function Analysis (DFA) of the continuous datasets was used to determine how clades clustered in morphometric space and the linear distances between them. Females generally had highly domed, round shaped carapaces with well developed knobs, a flat plastron, smaller shell openings and larger body size in comparison to male. Males exhibited a narrow, elongated body shape, flattened carapace with poorly developed domes and carapace knobs, concave plastron cavities, larger apical rings, tails and shell openings. Overall, the morphological analyses distinguished four distinct clusters, namely, "C1+C4+C5+C7", "C3", "C2" and "C6", which matched the DNA molecular findings. The four species scheme thus appears to be the best taxonomic solution for the systematic puzzle of the *P. tentorius* species complex, but should serve as working hypothesis for further investigation.

6.1 Introduction

6.1.1 The story of the *P. tentorius* species complex

The southern African tent tortoise (*Psammobates tentorius*) species complex is taxonomically one of the most confusing reptile groups, both phenotypically (Loveridge and Williams, 1957; Greig and Burdett, 1976; Branch 1998 and 2008, Hofmeyr et al. 2014) and genetically (Cunningham, 2002; Hofmeyr et al., 2014; Zhao et al., 2020). Phenotypically, this stems especially from their carapace and plastron colour patterns, and the shape of their shell, which are extremely variable (see Supplementary Figures S6.1-6.13). At some stage, the complex was subdivided into 20-30 species and subspecies, based on descriptions of only a few characters (Hewitt, 1933 and 1934). However, many of the taxa were considered as simply different morphs, rather than valid taxa (Loveridge and Williams, 1957; Greig and Burdett, 1976; Branch, 1998 and 2008, Hofmeyr et al. 2014).

A recently published phylogenetic study based on mtDNA sequence data, by Zhao et al. (2020,), revealed that there were seven clades (C1-C7) within the *P. tentorius* species complex. The study showed that the currently advocated "three-subspecies" assumption was not valid, and that *P. t. verroxii* was not monophyletic, and therefore not a valid taxon (see Figs. 6.1-6.2). Morphological variations are especially remarkable in C1 (Supplementary Figures S6.1-6.3) and C2 (Supplementary Figures S6.4-6.6).

Several researchers highlighted the need for a review of the taxonomy of the tent tortoise species complex, suggesting that some of the taxa may merit elevation to the rank of species (Cunningham, 2002; Branch, 1998 and 2008, Hofmeyr et al., 2014).

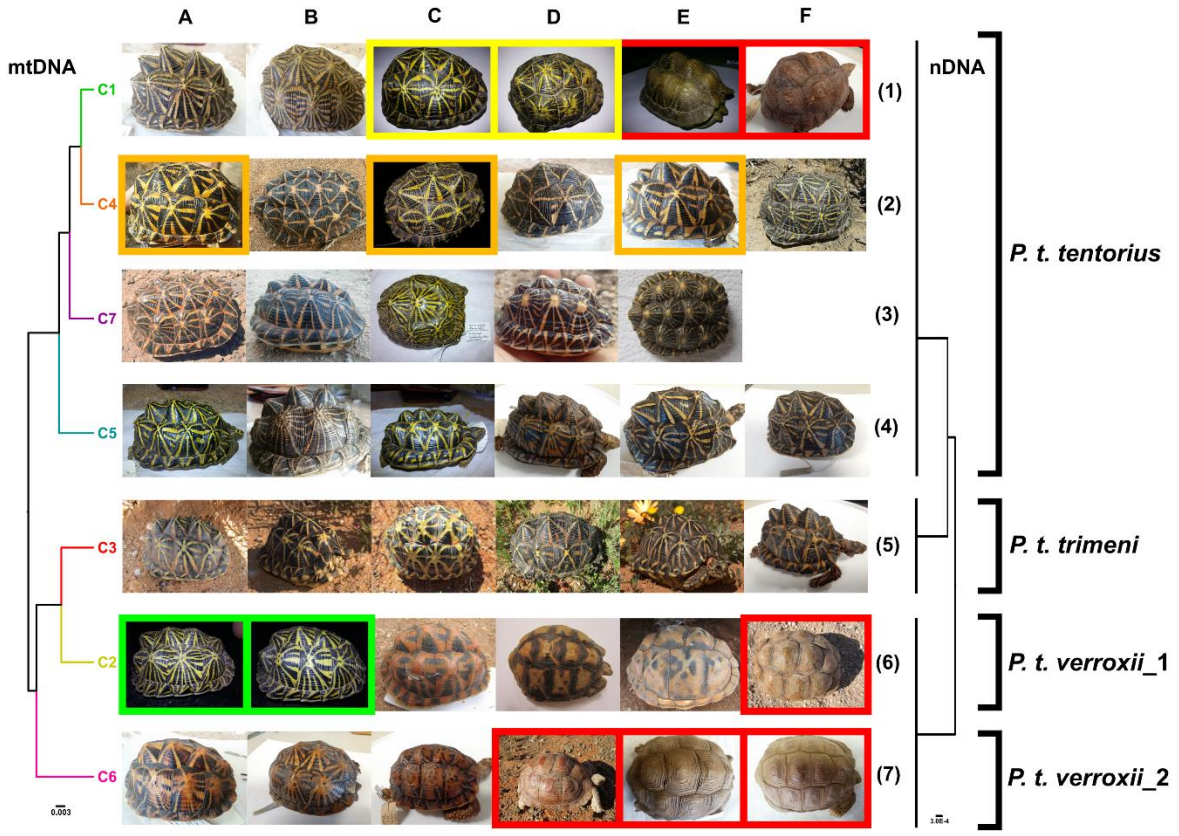


Figure 6.1. Morphs in the *Psammobates tentorius* species complex across the seven mtDNA clades, showing the high level of carapace phenotypic variation. The phylogenetic trees are modified versions of the ones published by Zhao et al. (2020). Left side: the phylogenetic relationships across the seven mtDNA clades; right side: the phylogenetic relationships among the seven clades based on nDNA data. The uniformly brown "*Psammobates bergeri*" morph was placed in red boxes; individuals which look like "*P. t. trimeni*" but belong to Clade 4 (a *P. t. tentorius* clade) were placed in orange boxes; individuals from Clade 2 (*P. t. verroxii*, south of the Orange River, Upper Karoo region) which look like "*P. t. tentorius*" were placed in green boxes; and individuals assigned to Clade 1, but which look like "*P. t. verroxii*, were placed in yellow boxes. The configuration of the phylogenetic tree showing the "three-subspecies" assumption is indicated on the right side.

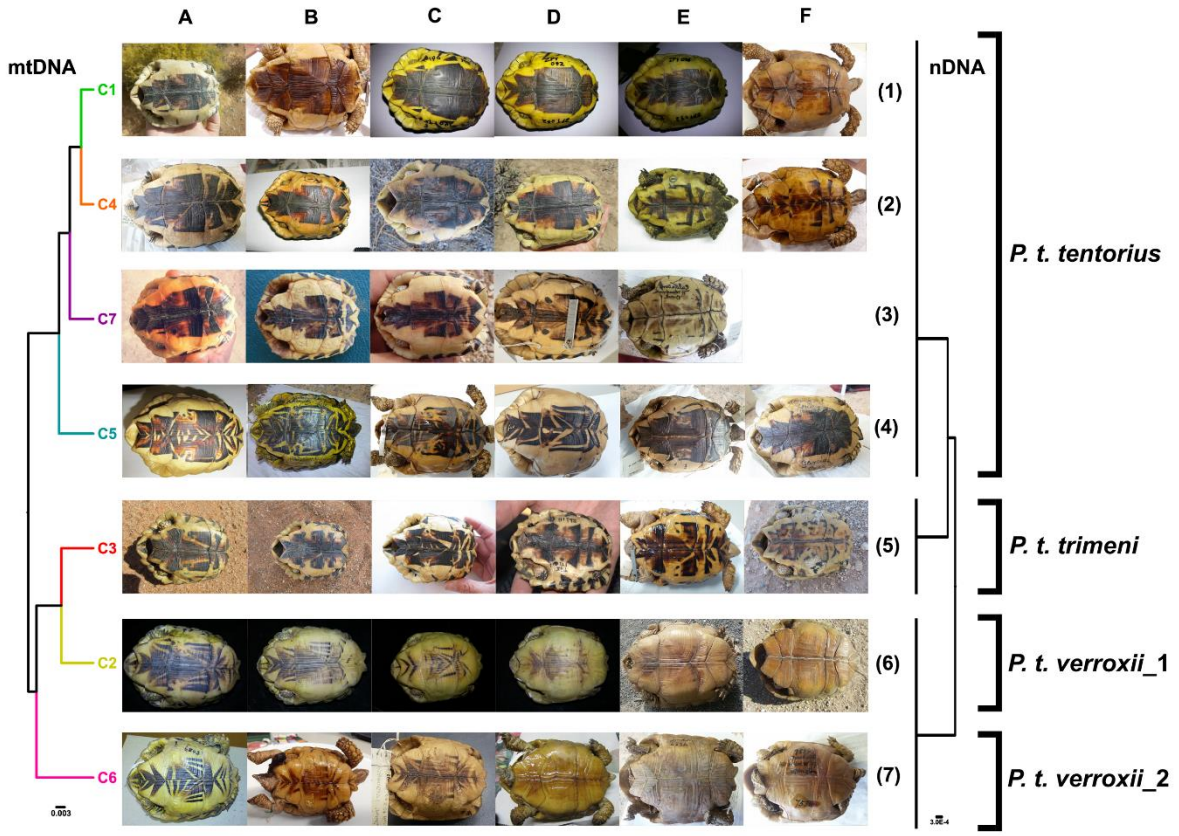


Figure 6.2. Phenotypic variation in plastron patterns in the seven clades of the *Psammobates tentorius* species complex. The mtDNA clades (left side) and nDNA clades showing the currently recognised "three-subspecies" assumption (right side) are modified versions of the phylogenetic tree in Zhao et al. (2020).

When mapping the distribution of the seven clades reported by Zhao et al. (2020) on to the biomes and vegetation types of southern Africa as described by Mucina and Rutherford (2011), they seem to align well with the latter, suggesting a correlation between clade distribution and habitat type. This was also confirmed by the findings in Chapter 3. The influence of habitat on the morphology of organism has also been reported in other studies ((e.g., coastal fish by Farré et al. (2015); the snail, *Trochulus hispidus*, by Proćków et al. (2018)). Southern Africa is a region characterized by high levels of climatic, biome and vegetation heterogeneity (Mucina and Rutherford, 2011), which can be expected to drive the evolution of high levels of phenotypic plasticity in its fauna (Reed et al., 2011; Bonamour et al., 2019), as seems to have been the case in the *P. tentorius* species complex.

Mismatches between the genotype and phenotype of species due to phenotypic plasticity under varying environmental conditions, are fairly common (Via and Lande, 1985). This is especially true in cryptic species complexes. Species richness can therefore be over- or under-estimated in studies based purely on morphological characters. A good example among tortoises would be the genus *Kinixys*. Six morpho-species previous recognised were shown to contain cryptic taxa, and that the actual number of *Kinixys* species was eight (Kindler et al., 2012). *Chersobius signatus* was initially believed to comprise two subspecies, *C. s. signatus* and *C. s. cafer*, based on morphology, but mtDNA and nDNA sequence data showed that the two "subspecies" were phylogenetically invalid (Daniels et al., 2010).

6.1.2 Sexual size and shape dimorphism

The literature abounds in studies documenting sexual size dimorphism in testudine species (Berry and Shine, 1980; McRae et al., 1981; Grover and DeFalco, 1995; Lagarde et al., 2001; Kabigumila, 2002; Willemsen and Hailey, 2003; Paquette and Lapointe, 2007; Chiari et al., 2009; Macale et al., 2011; Rezazadeh et al., 2014), including *P. tentorius*'s congeneric relative, *P. geometricus* (Baard, 1995).

Sexual size dimorphism is a complex phenomenon shaped by selection for survival and reproductive success (Berry and Shine, 1980; Paquette and Lapointe, 2007). The direction of sexual size dimorphism varies across species and can be female biased or male biased (Berry and Shine, 1980; Fairbairn et al., 2007; Ceballos et al., 2013). Collectively, fitness for survival involves multiple forces acting on body size variation through different pathways which determined the direction of sexual size dimorphism (Fairbairn et al., 2007; Ceballos et al., 2013). For example, in the angulate tortoise, sexual selection favored large males because of male-male combat during courtship (Branch, 2008; Berry and Shine, 1980). On the other hand, natural selection may favor bigger females, since bigger size would result in a larger body cavity for accommodating eggs in fecundity selection (Valenzuela, 2001; Stephens and Wiens, 2009).

Before embarking on character based morphometric approaches to infer the degree of distinctiveness among taxonomic groups, it is crucial to first establish whether any significant sexual dimorphism exist in each group and to analyse the sexes separately where that is the case.

Apart from sexual size dimorphism, colour pattern variation between the sexes also exists, for example, in the common padloper tortoise (*Homopus signatus*) (Branch, 2008) and Kalahari serrated tent tortoise (*Psammobates oculifer*) (Keswick and Hofmeyr, 2015). Sexual dimorphism can thus further complicate inter- and intraspecific variation (Chiari et al., 2009). It was therefore crucial to check whether the sexual dimorphism was present in the *P. tentorius* species complex before using character based morphometric approaches to infer the degree of distinctiveness among its different groups.

Apart from sexual size dimorphism, males and females may also differ with respect to the shape of corresponding body parts, something which is common in squamates (Vincent and Herrel, 2007). Sexual shape dimorphism is driven mainly by sexual selection and fecundity selection (Olsson et al., 2002; Herrel et al., 2009). Natural selection can therefore also be responsible for the evolution of sexual shape dimorphism (Schoener, 1967; Slatkin, 1984). Sexual shape dimorphism is common in many chelonian species (Bonnet et al., 2010).

Using a range of statistical procedures, this study investigated whether there was congruence between the findings based on DNA sequence data, which recognized seven clades in the *P. tentorius* species complex, and morphological findings based on variation in body size and shape, and carapace and plastron colour patterns. It also tested the utility of body measurement ratios as a potentially quick and reliable method for identifying clades in the field.

6.2 Materials and Methods

When comparing relative body size among sexes or between groups of organisms, the use of ratios to remove the effect of body size variation, is permissible only if the datasets used are isometric (Bartels et al., 2011; Baur and Leuenberger, 2011). However, when the variables of interest are allometric, using ratios cannot remove the effect of body size variation, and becomes an inappropriate approach (Lieonart et al., 2000; Bartels et al., 2011; Baur and Leuenberger, 2011). Despite these limitations, ratios are important for evaluating and comparing shape between sexes and among clades or taxa. Once such morphological

differences have been detected, a table of ratios can be constructed for easily measurable characters, which could be used for quick identifications in the field and for establishing the sex of individuals.

A commonly used approach for removing the effect of body size when studying sexual size dimorphism is Analysis of Covariance (ANCOVA) with fixed total body length as covariate. This method removes the effect of body size for both allometric and isometric variables.

The Discriminant Function Analysis (DFA) is a parametric multivariate analysis and powerful tool for classifying groups using multiple variables. It does so by separating total variation into different functions, using the functions which contribute most to the total variation. It can also estimate the morphometric distance between different groups in three dimensional spaces as a Mahalanobis distance and gives a *p*-value for pairwise groups indicating the extent to which they are morphometrically different (McHenry and Giles, 1971). This approach is suitable in multivariate analyses of continuous datasets (measurement data).

Permutation based Multivariate Analysis of Variance (PERMANOVA) is a non-parametric multivariate procedure with permutations. It can detect significant variation among all categories and groups. Should there be significant variation among the groups, a Post-hoc test is used to detect the significant pairwise differences (French et al., 2008; Anderson and Walsh, 2013). An extension of MANOVA which takes into consideration the effect of a covariate, namely, Multivariate Analysis of Covariance (MANCOVA) has been widely used in multivariate analyses to eliminate the effect of body size. All these procedures can be applied to both continuous and discrete datasets (coded phenotypic character data).

Unlike in the case of relative size comparisons, ratios can be used to compare the shapes of organisms without the restrictions of data isometry or allometry (Bartels et al., 2011; Baur and Leuenberger, 2011)

6.2.1 Specimen and data collection

In order to obtain adequate datasets for the phenotypic characters (discrete data) and morphometric measurements (continuous data), specimens were collected in the field as well as obtained from the following museums, Bayworld Museum (Port Elizabeth), Iziko Museum (Cape Town), the Ditsong Museum (Pretoria) and the National Museum of Namibia (Windhoek). Specimens were carefully selected, and individuals with visibly abnormal

character states were excluded from the analyses. All specimens were labeled and photographed dorsally, ventrally, anteriorly, posteriorly and laterally. In terms of wild caught specimens, a permanent marker was used to mark specimens prior to their release at the site of capture, in order to avoid repeat sampling. As for the continuous data, measurements were taken of the different linear aspects of the carapace and plastron using an electronic Vernier caliper and a tape measure. Each variable of interest was measured twice, and the average used in order to increase accuracy. In terms of the morphometric analysis, a closely related congeneric species, *Psammobates oculifer* (Cunningham, 2002), was included in the multivariate analyses to assess how the relative morphometric distances among the seven clades recognized with the DNA sequence data compared to their relative morphometric distances to *P. oculifer*.

6.2.2 Grouping variables

The seven clades retrieved by Zhao et al. (2020) were used as grouping variables for the univariate and multivariate analyses of both the discrete and continuous datasets.

6.2.3 Data partitions

The morphological data were partitioned into phenotypic and morphometric datasets. Males were regarded as adult if their straight-line plastron length (PLS) exceeded 55.0 mm and straight-line carapace length (CLS) exceeded 65.0 mm. Females were regarded as adult if their PLS exceeded 75.0 mm and CLS exceeded 85.0 mm. Individuals in which PLS was less than 55.0 mm and CLS less than 65.0 mm were regarded as juveniles.

6.2.4 The morphometric datasets

A preliminary univariate factorial ANCOVA was carried out to determine whether variables were informative and stable (i.e., consistently showed significant results for all the clades) within groups, using IBM®SPSS® version 20, before they were included in the dataset. The CLS (carapace straight-line length) was used as covariate.

The following carapace and plastron measurements were taken, see Supplementary Figure S6.15, CLS, UW (upper width of carapace), CW (central width of carapace), LW (lower width

of carapace), CLE (exact midline carapace surface length), HL (carapace height length), PLE (exact plastron length), AILI (inner axillary inguinal length) and AILO (outer axillary inguinal length); for detail about measurements, see Supplementary Table S6.1.

In Supplementary Figure S6.16 the following is shown, (1) PLS (straight-line plastron length); (2) AWD (lower anal width); (3) AWT (upper anal width); (4) FWT (upper femoral width); (5) FALS (lateral femoral to anal length); (6) FALC (midline femoral to anal length); (7) FWD (posterior femoral width); (8) GPLC (midline gular to pectoral length); (9) GPLS (lateral gular to pectoral length); (10) GWT (anterior gular width); (11) GWD (posterior gular width); (12) HWT (upper humeral width); (13) HWD (humeral width at bottom); (14) OHW (humeral width outward); (15) OPW (pectoral width outward) and (16) OAW (Anal side width outward); for detail, see Supplementary Table S6.1.

Several ratios were calculated using the absolute values of the above variables, in order to evaluate body shape with espect to each character of interest (McLuckie et al., 1999; Macale et al., 2011; Chiari and Claude, 2012). The selected covariate for these ratios was CLS; for detail, see Table 6.1 and Supplementary Table S6.1.

Table 6.1. Some of the descriptive ratios used in the comparative uni- and multivariate shape analyses of the body, carapace and plastron dimensions in the *P. tentorius* species complex.

Ratio	Purpose
UW/CW	to determine the shape of the anterior part of the carapace; the smaller the value, the more elongated or rectangular its shape.
LW/CW	to determine the shape of the posterior part of the carapace; the smaller the value, the more elongated or rectangular the carapace.
UW/LW	to determine the shape of the carapace relative to the rest of the body; the larger the value the more elongated the body; the body tends to be rectangular in shape if UW < CW < LW.
FALS/FALC	to determine how concave the anal scute fork (opening) is; the smaller the value, the less concave the fork (opening)
GWT/GWD	to determine the sharpness of the anterior gular projection; the smaller the value the sharper the projection.
CLE/PLS	to determine the relative height of a carapace knob; the larger the value, the higher the knob.
CLS/PLS	to determine the length of the carapace relative to that of the plastron; the larger the value the greater the relative carapace length.
PLE/PLS	to determine how concave the plastron is; the larger the value the more concave the plastron.
CW/CLS	to determine the relative width of the central part of the carapace; the larger the ratio, the wider the central carapace.
LW/CLS	to determine the relative width of the posterior part of the carapace.
HL/CW and HL/CLS	to determine the relative height of the carapace as well as its shape in two body orientations (frontal and lateral orientations); the smaller these values the flatter the carapace, and vice versa.
AILI/CLS	to determine the relative width of the plastron bridge. The larger the ratio, the wider the plastron bridge.
AWD/CLS	to determine the relative width of the anal scute of the plastron; the larger the value, the bigger the anal opening.
FALS/CLS	to determine the relative length of the posterior half of the plastron; the larger the value; the more elongated posterior part of the plastron.
GWT/CLS	to determine the relative width of the anterior side of the gular; the larger the ratio, the wider for the gular.
GPLS/PLS	to determine the area of the anterior half of the plastron relative to its posterior half; the larger the value, the greater the anterior half relative to the posterior half.

6.2.5 The phenotypic datasets

The same criteria used for selecting the continuous variables were also applied in the case of the discrete variables. The majority of the assumed informative variables selected are shown in Supplementary Figures S6.17-6.24.

Characters were encoded for the multivariate PERMANOVA (Tatsuoka, 1988). Two categories of coding were applied: 1) if a character only had two states, namely, present or absent, it was encoded as "1", for "present" and "2" for "absent". Character states encoded as 1 included: radiated sparse PCP, unmarked plastron, cone stripe, apical ring X, apical ring knob, orange head skin, orange buttock, white head block, black head apical speckle, gular triangular mark and triple mark. 2) in those cases where characters had multiple states, the following coding was used: "absent", encoded as "0", "present at low level", encoded as "1", "present at low to moderate level", encoded as "2", "present at moderate level", encoded as "3", "present at moderate to high level", encoded as "4", "present at high level", encoded as "5", present at very high level", encoded as "6" (for character coding details, see Supplementary Figures S6.17-6.24).

Characters encoded according to the category 2 system included: striped fork, dull striped, solid central plastron pattern (PCP), radiated dense PCP, plastron side stripe, thick stripe richness, thin stripe richness, increasing stripe richness, equal stripe richness, stripe general thickness, discontinued stripe level, head smoothness, apical ring size, apical ring spot, drab stripe, orange plastron, orange striped, red striped, red marginal, light phase, dark phase, yellowish head block, buttock tubercle, carapace smoothness, knob scute status, domed carapace, forelimb distinct border, shell thickness, bent stripes and stripe density. The summarized character dataset is given in Supplementary Table S6.2.

6.2.6 Statistical analyses

6.2.6.1 Determining sexual size and shape dimorphism

Prior to carrying out any of the analyses, the datasets were first tested for normality using the Shapiro-Wilks test in the program SPSS version 20. If the requirements for normality were not met, the variables were log-transformed and again checked for normality.

1) Sexual size dimorphism was determined using two-way ANCOVA with the log-transformed CLS as covariate and sex as factor. In addition to this, a multivariate two-way PERMANOVA was used to determine sexual size dimorphism in multivariate view with PAST (Paleontological Statistics) Version 3.08 (Hammer et al. 2001) with 10000 permutations on the discrete and continuous dataset, respectively. A DFA was performed on the dataset to visualize the degree of differences between sexes. The dataset was then linear regressed with log (CLS) as covariate using a general linear model. The studentized linearized residuals were retrieved and used as input data in the DFA analysis. The use of linearized residuals was to eliminate the effect of body size.

2) Sexual shape dimorphism was determined using two-way ANOVA on the ratios in STATISTICA version 8, with sex, clade and the interaction between them as effects. The Bonferroni corrected post-hoc test was used to evaluate pairwise significance between the sexes and among the clades. The post-hoc test results were plotted in order to visualize character differences between the clades. Ratios which differ significantly between clades can provide a quick method for field identifications.

3) Phenotypic character dimorphism among the sexes was determined, also using a two-way ANOVA followed by the Bonferroni post-hoc test. Again, the effect of sex, clades and the interaction between them was tested.

6.2.6.2 Distinguishing among the clades

To determine whether each of the variables differed significantly between sex and across clades, two-way ANCOVAs and Bonferroni post-hoc tests in STATISTICA version 8 were used.

To investigate how the clades clustered relative to each other, DFA analyses were performed on the linearized residuals in STATISTICA version 8. The scores of the first two discriminant functions were exported to SPSS version 20 to generate better quality scatterplots for visualization. Since significant sexual size and shape dimorphism was found, the continuous datasets were partitioned by sexes prior to the DFA, and separate analyses performed.

For the discrete dataset, two-way ANOVAs and Bonferroni post-hoc tests were used to test for characters that differed significantly between sexes and clades in STATISTICA version 8. The post-hoc test results of those variables showing significant differences were plotted.

Since two-way PERMANOVA does not generate post-hoc pairwise comparisons with corresponding *p*-values to indicate significant differences among clades, one-way PERMANOVAs (with "clades" as fixed factor) were performed on both the discrete (using the coded dataset) and continuous datasets (using the linearized residuals).

The one-way PERMANOVA analysis with 10000 permutations in PAST was used to determine whether the seven clades were significantly different from each other in terms of their morphology (morphometric and phenotypic data, separately). Because of significant sexual dimorphism, separate analyses were carried out for each sex.

6.3. Results

6.3.1 Specimen and data collection

For continuous dataset, A total of 389 specimens were used for the morphometric analyses, comprising 221 females, 128 males and 39 juveniles of unknown (see Supplementary Table S6.1). Twenty *P. oculifer* specimens were included in the analyses for purposes of comparison. For the phenotypic analyses 359 specimens were used, which included 224 females, 89 males and 46 juveniles of unknown sex (see Supplementary Table S6.2).

6.3.2 Sexual size dimorphism

The Shapiro-Wilks test results and histograms indicated that log-transforming the datasets had improved their distribution but that they still did not fully conform to the requirements of a normal distribution.

The ANCOVA results (details given in Supplementary Table S6.3) revealed significant sexual size dimorphism with respect to the following variables: HWD, FALC, FWT, AWD, AILO, AILI, UW, LW, CW, HL, PLS, PLE, CLE, OHW, OPW and OAW. Females were larger than males in all these variables, except AWD in which males were larger. When comparing body size (with log (CLS) as covariate), females were significantly larger than males ($p < 0.0001$).

A significant interaction was found between clades and sexes for relative HWD, AILO and CW ($p < 0.01$ in all cases). These results imply that the direction of sexual dimorphism was not consistent among clades.

Two-way PERMANOVA results revealed significant morphometric differences among clades ($F = 10.54$, $p < 0.0001$) and size dimorphism between the sexes ($F = 21.426$, $p < 0.0001$). It also showed a significant interaction between sex and clades ($F = 12.651$, $p < 0.005$). This result also implies that the direction of sexual dimorphism was not consistent across clades.

The DFA results (Fig. 6.3) likewise showed significant size dimorphism (based on the Squared Mahalanobis Distances) between males and females ($F = 21.63$, $p < 0.0001$). In total, 85.34% of the originally grouped cases were correctly classified, i.e., 94.87% of females, 84.87% of males and 32.35% of juveniles. The first two discriminant functions with highest Eigenvalues contributed 96.74% of the total variance and successfully differentiated between the sexes (Wilk's Lambda = 0.32, $p < 0.0001$). The scatterplot (Fig. 6.3) showed minimal overlap between sexes.

With regard to sexual shape dimorphism, the two-way ANOVA results showed significant sexual shape dimorphism in many of the ratios (see Supplementary Table S6.4 and Supplementary Figures S6.25-6.28). Significant interactions between clades and sex were found in the following ratios: GPLS/CLS, UW/CLS, CW/CLS, FALS/FALC, AWT/FALC, AILO/CLS, AILI/CLS, OPW/CLS ($p < 0.05$ in all cases). These interactions indicated that the direction of sexual shape dimorphism among the clades was not consistent.

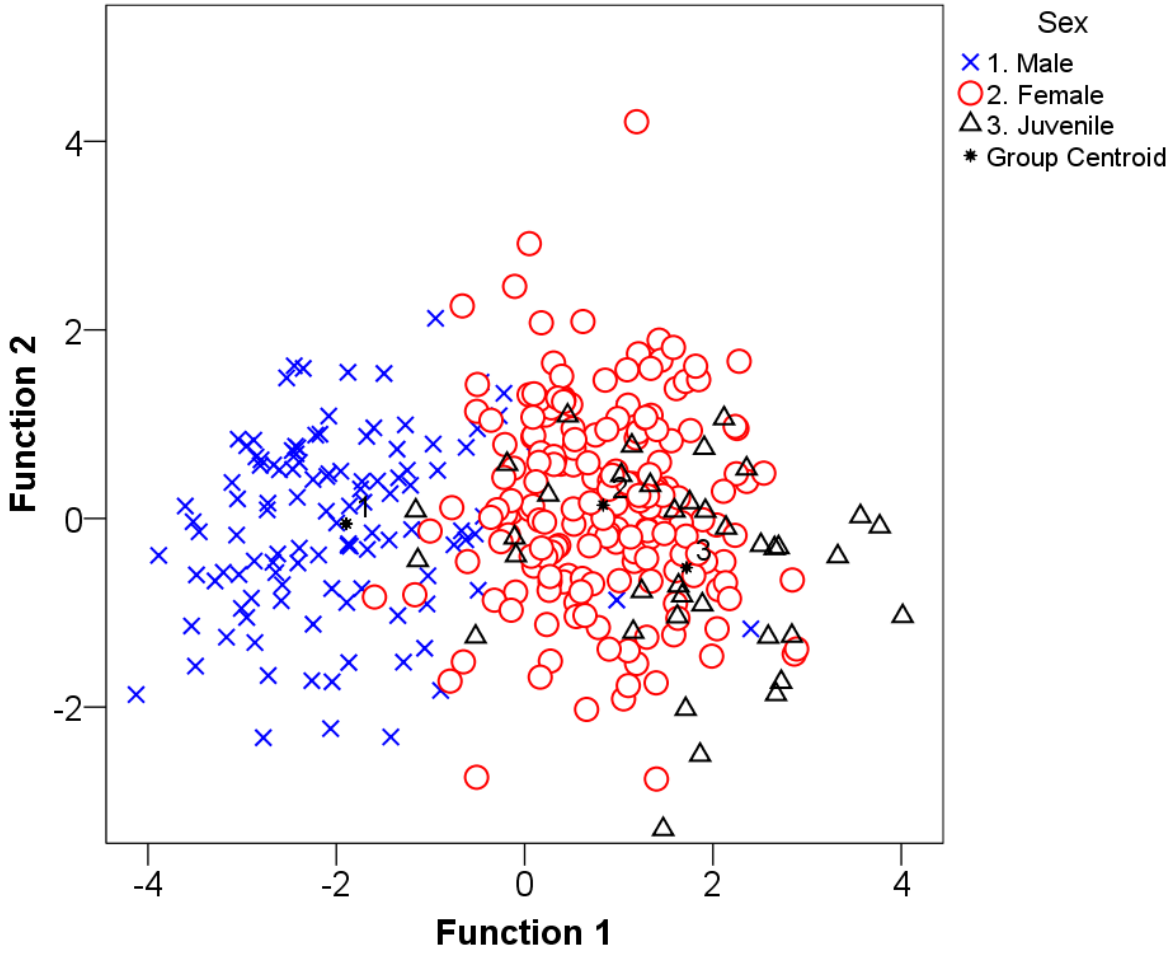


Figure 6.3. DFA scatter plots using the first two discriminant functions to visualize sexual size dimorphism in the *P. tentorius* species complex; the key to the symbols is shown to the right of the scatterplot.

For the phenotypic sexual dimorphism, the coded character-based ANOVA results also revealed significant dimorphisms (details see Supplementary Table S6.5 and Supplementary Figures S6.29-6.33). The Bonferroni post-hoc test results indicated significant sexual dimorphism in the following characters: stripe fork, radiated sparse PCB, distinct horizontal plastron stripe, equal stripe, discontinued stripe, apical ring size, buttock tubercle, knob scute status, domed carapace and forelimb BW border ($p < 0.05$ in all cases). Among these discrete characters, significant interactions between clades and sex were found for distinct horizontal plastron stripe, increasing stripe, equal stripe, X-stripe mix, red marginal, black head tip

speckle, carapace smoothness, knob scute status, domed carapace and forelimb BW border ($p < 0.05$ in all cases). These interactions show that the direction of sexual character dimorphisms were also not consistent among the clades.

Two-way PERMANOVA results revealed significant character differences among clades ($F = 7.81$, $p < 0.0001$) and between sexes ($F = 2.86$, $p < 0.0001$). No significant interactions between sex and clades were, however, observed ($F = 14.35$, $p = 0.14$).

6.3.3 Distinguishing among clades

For the morphometric dataset, the ANCOVA based on the log-transformed variables (see Supplementary Table S6.3) revealed that all variables differed significantly among the clades ($p < 0.005$ in all cases). The ratio-based ANOVA (see results Supplementary Table S6.4) results exhibited significant shape variations for the majority of ratios among the clades. The one-way PERMANOVA results revealed significant overall differences among clades for both females ($F = 8.304$, $p < 0.0005$) and males ($F = 8.264$, $p < 0.0005$). In females the Bonferroni corrected post-hoc test results (see Supplementary Table S6.6) showed significant pairwise differences among all clades (with $p < 0.01$ in all cases), whilst, the results in males only failed to distinguish C1 from C5 ($p > 0.05$), and C3 from C4 ($p > 0.05$).

In the DFA analyses in females (details given in Supplementary Table S6.6), the first two discriminant functions with the highest Eigen values contributed 81.83% of the total variance and significantly differentiated the seven clades from each other (Wilks' lambda = 0.014, $p < 0.0001$). Overall, 87.18% of the cases were correctly assigned (C1: 89.61%, C2: 94.74%, C3: 86.67%, C4: 79.31%, C5: 53.33%, C6: 100%, C7: 100% and *P. oculifer*: 100%,). The pairwise Mahalanobis Distances revealed significant pairwise differences among the clades ($p < 0.05$ in all cases), except the pairwise difference between C1 and C7 which was not significant ($F = 1.12$, $p > 0.05$). The DFA scatterplots (Fig. 6.4) generally exhibited four clusters: 1) C1+C4+C5+C7, 2) C3, 3) C2 and 4) C6. When comparing the pairwise Mahalanobis distances between the clades to their individual distances from *P. oculifer*, the interclade differences were smaller, showing their closer relatedness to each other than to *P. oculifer*. However, C1, C4, C5 and C7 did not show clear separation from each other.

As for the DFA in males, only six clades were included as no C7 samples were available. The first two discriminant functions with the highest Eigen values contributed 82.33% of the total

variance and significantly differentiated the clades from each other (Wilks' lambda = 0.007, $p < 0.0001$). In total 93.28% of the cases were correctly assigned (C1: 90.24%, C2: 100%, C3: 81.82%, C4: 80%, C5: 100%, C6: 100% and *P. oculifer*: 100%). The pairwise Mahalanobis Distances revealed significant pairwise differences among the clades ($p < 0.05$ at all cases), except for the pairwise difference between C3 and C4, which was not significant (F = 0.882, $p > 0.05$). As in the females, the DFA scatterplot of the males also advocated four clusters. Within cluster 1, C1, C4 and C5 were not visually diagnosable.

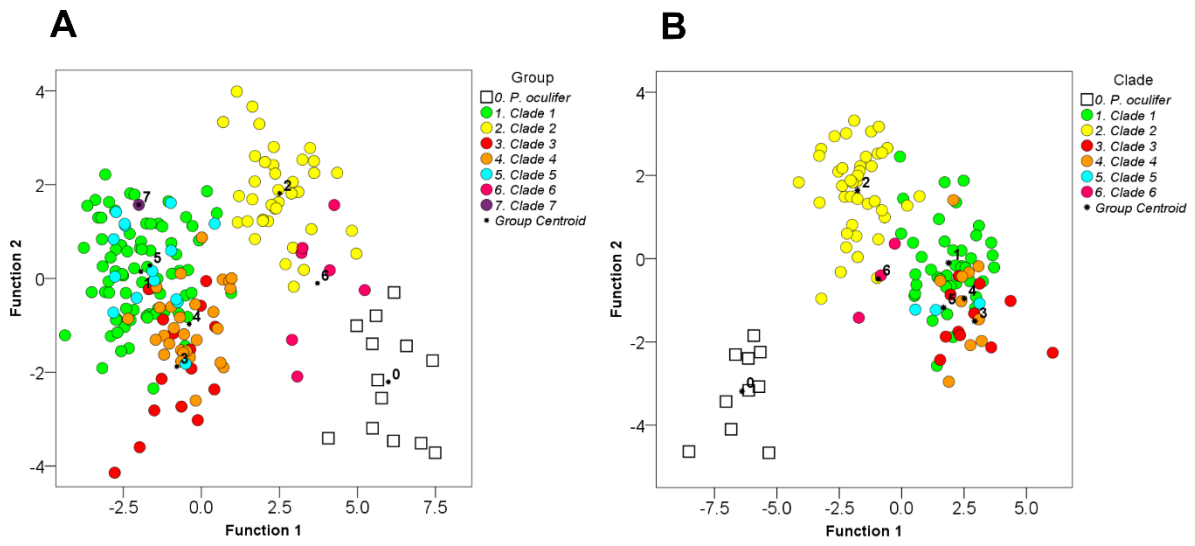


Figure 6.4. DFA scatterplots for the seven clades of the *P. tentorius* species complex, using the first two discriminant functions. A: DFA scatterplot of females, B: DFA scatterplot of males. The congeneric *P. oculifer* clearly stands separately from the clades of this study.

The ANOVA results (see Supplementary Table S6.5) of the phenotypic data revealed that all 32 coded characters were informative in distinguishing among the clades ($p < 0.05$ in all cases). The one-way PERMANOVA results of the female phenotypic data (see Supplementary Table S6.6), showed that the differences among the clades were significant (F = 21.172, $p < 0.0001$). The Bonferroni corrected post-hoc test results confirmed that all seven clades were significantly different from each other ($p < 0.005$ in all cases). In males the one-way

PERMANOVA results (see Supplementary Table S6.6) also revealed overall significant differences among the clades ($F = 12.280$, $p < 0.0001$). The Bonferroni post-hoc test indicated that the difference between C1 and C5 was not significant ($p > 0.05$), whilst, the differences among the rest of the clades were significant ($p < 0.01$ in all cases).

6.4 Discussion

6.4.1 Sexual dimorphism

Understanding the causes of complex sexually dimorphic patterns has been a pervasive question in evolutionary biology, and such thinking drove Darwin to unmask the puzzle of the principles of sexual selection (Darwin, 1871; Ceballos et al., 2013). Generally, body size is highly variable, and strongly correlated with individual fitness through its effect on survival and reproduction, as driven by various evolutionary forces in a sex-specific manner (Ceballos et al., 2013).

The analyses in all respects (univariate and multivariate) found substantial sexual dimorphism in the *P. tentorius species* complex. The sexual size, shape and phenotypic dimorphisms were found to be significant in the vast majority of cases, with non-significant differences in very few cases.

The large samples sizes used in this study confirmed the findings of Loveridge and Williams (1957) that females were larger than males in the *P. tentorius* species complex. All the characters measured had relatively larger sizes in females than in males, except AWD, which was relatively larger in males. The latter finding may be the result of a male-biased growth trajectory in the posterior shell opening, possibly to allow greater ease of locomotion and freedom of movement by the tail. Ease of locomotion may increase the chances of finding a female, while freedom of tail movement may facilitate copulation. These suggestions obviously require verification through further investigation. Female-biased sexual size dimorphism was considered as the basal state in chelonians by Ceballos et al. (2013). The majority of tortoise species have, however, not retained this pattern, as most of them show male-biased sexual size dimorphism (Berry and Shine, 1980; Ceballos et al., 2013). The female-biased sexual

dimorphism of the *P. tentorius* species complex is consistent with that of its close relatives *P. oculifer* and *P. geometricus*, which confirms that all *Psammobates* species have a female-biased sexual dimorphism pattern (Keswick and Hofmeyr, 2015; Baard, 1995, Branch, 2008). This pattern may represent a phylogenetic constraint.

Physiologically, a major advantage of a small male body size would be that it requires less energy for maintenance than a large one, allowing higher energy expenditure on locomotion to find females for mating (Gibbons and Lovich, 1990). Besides this, the small body size should allow it to access smaller shelters which may mean better protection if it minimizes detection by predators or makes it difficult for the latter to get hold of them. Nonetheless, small body size may also have disadvantages, for example, it may make them easier to handle by predators (Janzen et al., 2000), such as crows in the case of *P. tentorius* (Zhao Z, unpublished data). Small body size also means a larger surface area to volume ratios, which should increase vulnerability to thermal stress and desiccation (Cloudsley-Thompson, 1999). The large relative and absolute sizes of characters in females would have advantages which should maximize reproductive output. A large abdominal cavity can accommodate a larger number of eggs, thus increasing fecundity and also enables the storage of larger energy reserves, essential for vitellogenesis and egg production (Bonnet et al., 2001; Loehr et al., 2006).

In the case of sexual shape dimorphism, many of the ratios were informative in terms of describing divergent growth trajectories between the sexes. In summary (details given in Supplementary Table S6.4 and Supplementary Figures S6.25-6.28), males generally had a narrower, more elongated and flattened carapace, which was less domed with carapace knobs not as well developed as in females. Males also had a “hook-shaped” anal shield, and larger anterior and posterior shell openings. They generally have a concave plastron. By contrast, females generally had a rounder, higher and more domed carapace with well-developed carapace knobs. They lacked a “hook-shaped” anal shield, and had smaller shell openings at both the anterior and posterior sides, and their plastron was flat.

A possible interpretation of the adaptative value of a higher and domed carapace with well-developed knobs in females, could be that it decreases the surface area to volume ratio of the body, which would reduce heat uptake and the risk of thermal stress and desiccation (Cloudsley-Thompson, 1999). The function of the concave plastron and “hook-shaped” anal shield in males is presumably to facilitate copulation. The overall flattened, elongated and narrow carapace of males presumably enables them to better access small shelters and crevices

to escape from heat and predators (Janzen et al., 2000; Keswick and Hofmeyr, 2014). The possible adaptive advantage of larger anterior and posterior shell openings in males has already been discussed earlier under sexual size dimorphism. Furthermore, the bigger cranial space could provide extra room for head bobbing during copulation (Keswick and Hofmeyr, 2015).

With regard to the phenotypic sexual dimorphism analyses, significant dimorphism was found with respect to the colour patterns (details given as Supplementary Table S6.5 and Supplementary Figures S6.29-6.33). The significantly larger apical rings on the carapace knobs and more distinctive colour patterns on the front limbs of males, could possibly be adaptations for attracting females during courtship, but this needs to be verified by further investigation.

6.4.2 On distinguishing of clades

Efforts to develop an accurate and stable dichotomous key for the seven clades of the *P. tentorius* species complex is jeopardized by too much variations in potentially useful morphological diagnostic characters. A considerable amount of informative morphometric and phenotypic characters nevertheless exists, which enable differentiation among the clades (details given in Supplementary Tables S6.3-6.6 and Supplementary Figures S6.25-6.33). Three major clusters were diagnosable in this way, namely, 1) C1+C4+C5+C7, 2) C3 and 3) C2+C6.

The multivariate analyses generally advocated four highly parsimonious clusters (see Fig. 6.4): cluster 1) C1+C4+C5+C7, cluster 2) C3, cluster 3) C2 and cluster 4) C6, though some of the analyses advocated further subdivisions within cluster 1. When comparing this morphological clustering scheme with results retrieved from the molecular analyses, both the DNA sequence data (nDNA and mtDNA, Chapter 2) and microsatellite DNA data (Chapter 4) advocated a similar clustering pattern. Collectively, it therefore seems as though a four-species scheme would be the best taxonomic solution for the *P. tentorius* species complex. When comparing the pairwise Mahalanobis distances between the seven clades, and between the seven clades and *P. oculifer*, the distances between the four morphological clusters were all substantially diagnostic. The pairwise Mahalanobis distances between these clusters were in some cases smaller than their distance from *P. oculifer*, but in other cases larger, specifically between clusters 1, 3 and 4. These morphological distances provided clear evidence that the clades were different at the interspecific rather than intraspecific level. However, given the relatively small

sample sizes in the case of C5 and C7, but particularly in C7, verification of these findings with adequate samples is needed in their case.

The major morphological confusion in terms of both the carapace and plastron comes from C1 (see Figs. 6.1-6.2, Supplementary Figures S6.1-6.3), C2 (see Figs. 6.1-6.2, Supplementary Figures S6.4-6.6), C6 (see Figs 6.1-6.2, Supplementary Figures S6.12-6.13) and between C3 (see Figs. 6.1-6.2, Supplementary Figures S6.7-6.8) and C4 (see Figs. 6.1-6.2, Supplementary Figure S6.9). Clade 5 (Figs. 6.1-6.2, Supplementary Figures S6.10-6.11) and C7 (Figs. 6.1-6.2, Supplementary Figure S14) had far less within-group morphological variations than the rest of the clades, but this may be the result of their low sample sizes. Within C1, some individuals encountered in the Tankwa Karoo (south-eastern population) exhibited a uniformly brown carapace, which is also found in C2 (northern populations) and C6 (see Fig. 6.1). This was confirmed by both the DNA sequence data (see Chapter 2) and microsatellite DNA data (see Chapter 4). The occurrence of the brown coloured morph in the clades mentioned, could be the result of microhabitat adaptation and convergent evolution, since they are only found in habitats with sparse vegetation and barren sandy substrates (Mucina and Rutherford, 2011; Z. Zhao, pers. obs). Furthermore, some individuals from the central and south-western populations of C1 resembled similar phenotypes within C2 ((e.g. highly dull carapace pattern, faintly developed carapace dome and knobs – considered as typical characters of *P. t. verroxii* (Branch, 2008)). The molecular results (Chapter 2 and 4) confirmed that those individuals belonged to *P. t. tentorius* (C1), and not to *P. t. verroxii* (C2), which certainly helps to entangle its taxonomy. The microsatellite results (Chapter 4) revealed none of these strange "*P. t. verroxii* - like" individuals, which may be a consequence of hybridization between putative taxa.

In the equally morphologically confusing group, C2, the northern and southern populations differed substantially (see Fig. 6.1). The molecular results (all types of markers), however, did not reveal any significant differences between the two populations. These results suggest that the phenotypic variations within C2 may have been shaped by short-term environmental adaptations, rather than long-term phylogenetic adaptive radiation. It is also noteworthy that many individuals encountered in the southern population were morphologically substantially similar to the typical phenotypes of C1 ((e.g. highly domed carapace, distinctive carapace stripes – considered as typical characters of *P. t. tentorius* (Branch, 2008)). These confusing patterns would have resulted in substantial misidentifications between *P. t. verroxii* and *P. t. tentorius* in the past ((see records of Hofmeyr et al., (2014)), since many of the questionable

individuals were found in the range of the southern population of C2. The microsatellite data (Chapter 4) failed to find any evidence to support any of these "*P. t. tentorius* – like cases". These may have been the result of hybridization between taxa of putative C1 and C2, though both Chapter 2 and 4 advocated the possible crossing-over between C1 and C2.

As briefly mentioned in Chapter 2, the last part of the major confusion in the *P. tentorius* species complex is cause by the great morphological similarity between C3 and C4 ((both previously regarded as "*P. t. trimeni*" by Boycott & Bourquin (1988) and Branch (2008)), despite all the molecular evidence showing them to be distant relatives. The similarity could therefore be ascribed to homoplasy due to convergent evolution, as both occur in Succulent Karoo biomes (Branch, 1998 and 2008; Mucina and Rutherford, 2011; Hofmeyr et al., 2014), thus potentially facing similar selection pressures.

Colour pattern variation is not only linked to genetic variation but can also be the outcome of the interaction between the genotype and the environment, Hoekstra (2006), and Gray and McKinnon (2007) stating that the environment and genetic make-up co-influence the plasticity of a phenotype. The high phenotypic plasticity and morphological variation in the *P. tentorius* species complex, may make it impossible to develop an accurate and stable dichotomous key for it. Nevertheless, the multivariate analyses of multiple characters and ratios were still able to delineate the seven clades into four clusters, C2, C3, C6 and (C1+C4+C5+C7). A future study based on their recovered genomes from the museum type specimens, using ancient DNA techniques, is recommended for achieving final taxonomic resolution.

6.5 Conclusions

From the analyses carried out it was clear that consideration has to be given to morphological characters exhibiting sexual dimorphism (size, shape and phenotype) when performing multivariate clustering analyses in the *P. tentorius* tent tortoise species complex.

The multivariate findings show agreement with the molecular findings (Chapter 2 and 4), both suggesting a four cluster scheme (C2, C3, C6 and "C1+C4+C5+C7") for the seven clades retrieved by the molecular analyses. This may be the best configuration when revising the taxonomy of this highly confusing species complex. This also suggests that species delimitation approach, particularly in cases of highly polymorphic or cryptic species complexes, should consider multiple types of markers (i.e. morphometric, phenotypic, nDNA and mtDNA sequence data, and microsatellite DNA data) as well as other types of evidence (i.e. ecology, geography). To simply erect a "new species" based on a monophyletic clade and *p*-distances (particularly for those ones based on single mtDNA loci) may over-split species, which should be avoided as far as possible.

Despite considerable homoplasy among the morphometric measurements and phenotypic characters, which will be a challenge in the development of a dichotomous key for the species complex, there were nevertheless key characters which could be useful in that regard.

References

Anderson, M. J., and Walsh, D. C. (2013). PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: what null hypothesis are you testing?. *Ecological Monographs, 83*, 557-574.

Baard, E. H. W. (1995). Growth, age at maturity and sexual dimorphism in the geometric tortoise, *Psammobates geometricus*. *The Journal of the Herpetological Association of Africa, 44*, 10-15.

Bartels, P. J., Nelson, D. R., & Exline, R. P. (2011). Allometry and the removal of body size effects in the morphometric analysis of tardigrades. *Journal of Zoological Systematics and Evolutionary Research*, *49*, 17-25.

Baur, H., & Leuenberger, C. (2011). Analysis of ratios in multivariate morphometry. *Systematic Biology*, *60*(6), 813-825.

Berry, J. F., and Shine, R. (1980). Sexual size dimorphism and sexual selection in turtles (Order Testudines). *Oecologia*, *44*, 185-191.

Bonamour, S., Chevin, L. M., Charmantier, A., & Teplitsky, C. (2019). Phenotypic plasticity in response to climate change: the importance of cue variation. *Philosophical Transactions of the Royal Society B*, *374*(1768), 20180178.

Bonnet, X., Delmas, V., El-Mouden, H., Slimani, T., Sterijovski, B., & Kuchling, G. (2010). Is sexual body shape dimorphism consistent in aquatic and terrestrial chelonians?. *Zoology*, *113*(4), 213-220.

Bonnet, X., Lagarde, F., Henen, B. T., Corbin, J., Nagy, K. A., Naulleau, G., ... & Cambag, R. (2001). Sexual dimorphism in steppe tortoises (*Testudo horsfieldii*): Influence of the environment and sexual selection on body shape and mobility. *Biological Journal of the Linnean Society*, *72*(3), 357-372.

Boycott, R. C., and Bourquin, O. (1988). *The South African Tortoise Book. A Guide to South African Tortoises, Terrapins and Turtles*. Southern Book Publisher. Johannesburg.

Branch, B. (1998). *Field guide to snakes and other reptiles of southern Africa*. Sanibel Island (FL): Ralph Curtis Books. Cape Town.

Branch, B. (2008). *Tortoise, Terrapins and Turtles of Africa*. Struik Publisher, Cape Town.

Ceballos, C. P., Adams, D. C., Iverson, J. B., & Valenzuela, N. (2013). Phylogenetic patterns of sexual size dimorphism in turtles and their implications for Rensch's rule. *Evolutionary Biology*, *40*(2), 194-208.

Chiari, Y., and Claude, J. (2012). Morphometric identification of individuals when there are more shape variables than reference specimens: a case study in Galápagos tortoises. *Comptes Rendus Biologies, 335*, 62-68.

Chiari, Y., Hyseni, C., Fritts, T. H., Glaberman, S., Marquez, C., Gibbs, J. P., ... and Caccone, A. (2009). Morphometrics parallel genetics in a newly discovered and endangered taxon of Galápagos tortoise. *PLoS One, 4*, e6272.

Cloudsley-Thompson, J. L. (1999). *The Diversity of Amphibians and Reptiles: An Introduction*. Berlin: Springer-Verlag.

Cunningham J., (2002). *A molecular perspective on the family Testudinidae Batsch, 1788*, (unpublished Ph.D thesis), University of Cape Town.

Daniels, S. R., Hofmeyr, M. D., Henen, B. T., & Baard, E. H. W. (2010). Systematics and phylogeography of a threatened tortoise, the speckled padloper. *Animal Conservation*, *13*(3), 237-246.

Darwin, C. (1871). *The descent of man, and selection in relation to sex*. London: John Murray.

Fairbairn, D. J., Blanckenhorn, W. U., & Szekely, T. (2007). *Sex, size and gender roles. Evolutionary studies of sexual size dimorphism*. New York: Oxford University Press.

Farré, M., Lombarte, A., Recasens, L., Maynou, F., & Tuset, V. M. (2015). Habitat influence in the morphological diversity of coastal fish assemblages. *Journal of Sea Research*, *99*, 107-117.

French, A., Macedo, M., Poulsen, J., Waterson, T., and Yu, A. (2008). *Multivariate analysis of variance (MANOVA)*. San Francisco State University.

Gibbon, J. W., & Lovich, J. E. (1990). On the slider turtle (*Trachemys scripta*). *Herpetological Monographs*, *4*, 1-29.

Gray, S. M., and McKinnon, J. S. (2007). Linking colour polymorphism maintenance and speciation. *Trends in Ecology and Evolution, 22*, 71-79.

Greig, J. C., and Burdett, P. D. (1976). Patterns in the distribution of southern African terrestrial tortoises (Cryptodira: Testudinidae). *African Zoology, 11*, 251-273.

Grover, M. C., and DeFalco, L. A. (1995). Desert tortoise (*Gopherus agassizii*): status-of-knowledge outline with references. General Technical Report, INT-GTR-316. Ogden, UT: US Department of Agriculture, Intermountain Research Station. 134 p., 316.

Hammer, Ø., Harper, D. A., and Ryan, P. D. (2001). PAST: paleontological statistics software package for education and data analysis. *Palaeontologia Electronica, 4*, 9.

Herrel, A., Schaerlaeken, V., Moravec, J., & Ross, C. F. (2009). Sexual shape dimorphism in Tuatara. *Copeia, 2009*(4), 727-731.

Hewitt J. 1933. On the Cape species and subspecies of the genus *Chersinella* Gray. Part I. *Annals of the Natal Museum, 7*, 255-297.

Hewitt, J. (1934). On the Cape species and subspecies of the genus *Chersinella*. Part II. *Annals of the Natal Museum, 7*, 303-352.

Hoekstra, H. E. (2006). Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity, 97*, 222.

Hofmeyr, M. D., Boycott, R. C., and Baard, E. H. W. (2014). *Psammobates tentorius* (Bell, 1828). In: M. F. Bates, W. R. Branch, A. M. Bauer, M. Burger, J. Marais, G. J. Alexander and M. S. De Villiers (Eds.), Atlas and red list of the reptiles of South Africa, Lesotho and Swaziland (pp. 70-85). Pretoria: South African National Biodiversity Institute.

Janzen, F. J., Tucker, J. K., & Paukstis, G. L. (2000): An experimental analysis of an early life history stage: selection on size of hatchling turtles. *Ecology, 81*, 2290-2304.

Kabigumila, J. (2002). Morphometrics of the pancake tortoise (*Malacochersus tornieri*) in Tanzania. *Tanzania Journal of Science, 28*, 33-46.

Keswick, T., & Hofmeyr, M. (2014). Refuge characteristics and preferences of *Psammobates oculifer* in semi-arid savanna. *Amphibia-Reptilia, 35*(*1*), 41-51.

Keswick, T., & Hofmeyr, M. D. (2015). Sexual dimorphism and geographic variation in the morphology of a small southern African tortoise *Psammobates oculifer*. *Amphibia-Reptilia, 36*(*1*), 55-64.

Kindler, C., Branch, W. R., Hofmeyr, M. D., Maran, J., Široký, P., Vences, M., Harvey, J., Hauswaldt, J. S., Schleicher, A., Stuckas, H., & Fritz, U. (2012). Molecular phylogeny of African hinge-back tortoises (*Kinixys*): implications for phylogeography and taxonomy (Testudines: Testudinidae). *Journal of Zoological Systematics and Evolutionary Research*, *50*, 192-201.

Lagarde, F., Bonnet, X., Henen, B. T., Corbin, J., Nagy, K. A., and Naulleau, G. (2001). Sexual size dimorphism in steppe tortoises (*Testudo horsfieldi*): growth, maturity, and individual variation. *Canadian Journal of Zoology, 79*, 1433-1441.

Lieonart, J., Salat, J., & Torres, G. J. (2000). Removing allometric effects of body size in morphological analysis. *Journal of Theoretical Biology*, *205*, 85-93.

Loehr, V. J., Henen, B. T., & Hofmeyr, M. D. (2006). Shell characteristics and sexual dimorphism in the Namaqualand speckled padloper, *Homopus signatus signatus*. *African Journal of Herpetology*, *55*(*1*), 1-11.

Loveridge, A., & Williams, E. E. (1957). Revision of the African tortoises and turtles of the suborder Cryptodira. *Bulletin of the Museum of Comparative Zoology, 115*, 161-557.

Macale, D., Venchi, A., and Scalici, M. (2011). Shell shape and size variation in the Egyptian tortoise *Testudo kleinmanni* (Testudinidae, Testudines). *Folia Zoologica, 60*, 167-176.

McHenry, H., and Giles, E. (1971). Morphological variation and heritability in three Melanesian populations: a multivariate approach. *American Journal of Physical Anthropology, 35*, 241-253.

McLuckie, A. M., Lamb, T., Schwalbe, C. R., and McCord, R. D. (1999). Genetic and morphometric assessment of an unusual tortoise (*Gopherus agassizii*) population in the Black Mountains of Arizona. *Journal of Herpetology, 33*, 36-44.

McRae, W. A., Landers, J. L., and Cleveland, G. D. (1981). Sexual dimorphism in the gopher tortoise (*Gopherus polyphemus*). *Herpetologica, 37*, 46-52.

Mucina, L., and Rutherford, M. C. (2011). *The vegetation of South Africa, Lesotho and Swaziland*. Pretoria: South African National Biodiversity Institute.

Olsson, M., Shine, R., Wapstra, E., Ujvari, B., & Madsen, T. (2002). Sexual dimorphism in lizard body shape: the roles of sexual selection and fecundity selection. *Evolution*, *56*(7), 1538-1542.

Paquette, S. R., and Lapointe, F. J. (2007). The use of shell morphometrics for the management of the endangered malagasy radiated tortoise (*Geochelone radiata*). *Biological Conservation, 134*, 31-39.

Proćków, M., Proćków, J., Błażej, P., & Mackiewicz, P. (2018). The influence of habitat preferences on shell morphology in ecophenotypes of *Trochulus hispidus* complex. *Science of the Total Environment, 630*, 1036-1043.

Reed, T. E., Schindler, D. E., & Waples, R. S. (2011). Interacting effects of phenotypic plasticity and evolution on population persistence in a changing climate. *Conservation Biology*, *25*(*1*), 56-63.

Rezazadeh, E., Alucheh, R. M., and Kami, H. G. (2014). A preliminary study on the Mediterranean spur-thighed tortoise *Testudo graeca* Linnaeus, 1758 from northwestern Iran. *Herpetology Notes, 7*, 127-133.

Schoener, T. W. (1967). The ecological significance of sexual dimorphism in size in the lizard *Anolis conspersus*. *Science, 155*(3761), 474-477.

Slatkin, M. (1984). Ecological causes of sexual dimorphism. *Evolution, 38*(3), 622-630.

Stephens, P. R., & Wiens, J. J. (2009). Evolution of sexual size dimorphisms in emydid turtles: Ecological dimorphism, Rensch's rule, and sympatric divergence. *Evolution, 63*, 910-925.

Tatsuoka, M. M. (1988). Multivariate Analysis of Variance. In: J. R. Nesselroade and R. B. Cattell (Eds), Handbook of Multivariate Experimental Psychology. Perspectives on Individual Differences (pp. 399-419). Boston: Springer.

Valenzuela, N. (2001). Maternal effects on life-history traits in the Amazonian giant river turtle *Podocnemis expansa*. *Journal of Herpetology, 35*, 368–378.

Via, S., & Lande, R. (1985). Genotype-environment interaction and the evolution of phenotypic plasticity. *Evolution*, *39*(*3*), 505-522.

Vincent, S. E., & Herrel, A. (2007). Functional and ecological correlates of ecologically-based dimorphisms in squamate reptiles. *Integrative and Comparative Biology*, *47*(2), 172-188.

Willemsen, R. E., and Hailey, A. (2003). Sexual dimorphism of body size and shell shape in European tortoises. *Journal of Zoology, 260*, 353-365. Zhao, Z., Heideman, N., Grobler, P., Jordaan, A., Bester, P., and Hofmeyr, M. D. (2020). Unraveling the diversification and systematic puzzle of the highly polymorphic *Psammobates tentorius* (Bell, 1828) complex (Reptilia: Testudinidae) through phylogenetic analyses and species delimitation approaches. *Journal of Zoological Systematics and Evolutionary Research, 58*, 308-326.

Supplementary Tables

All the Supplementary Tables (Table S6.1-S6.6) were not included in the main document of this thesis, since their size are large and unable to fit in the main document. Therefore, all Supplementary Tables will only be provided independently as supporting documents.

Supplementary Figures

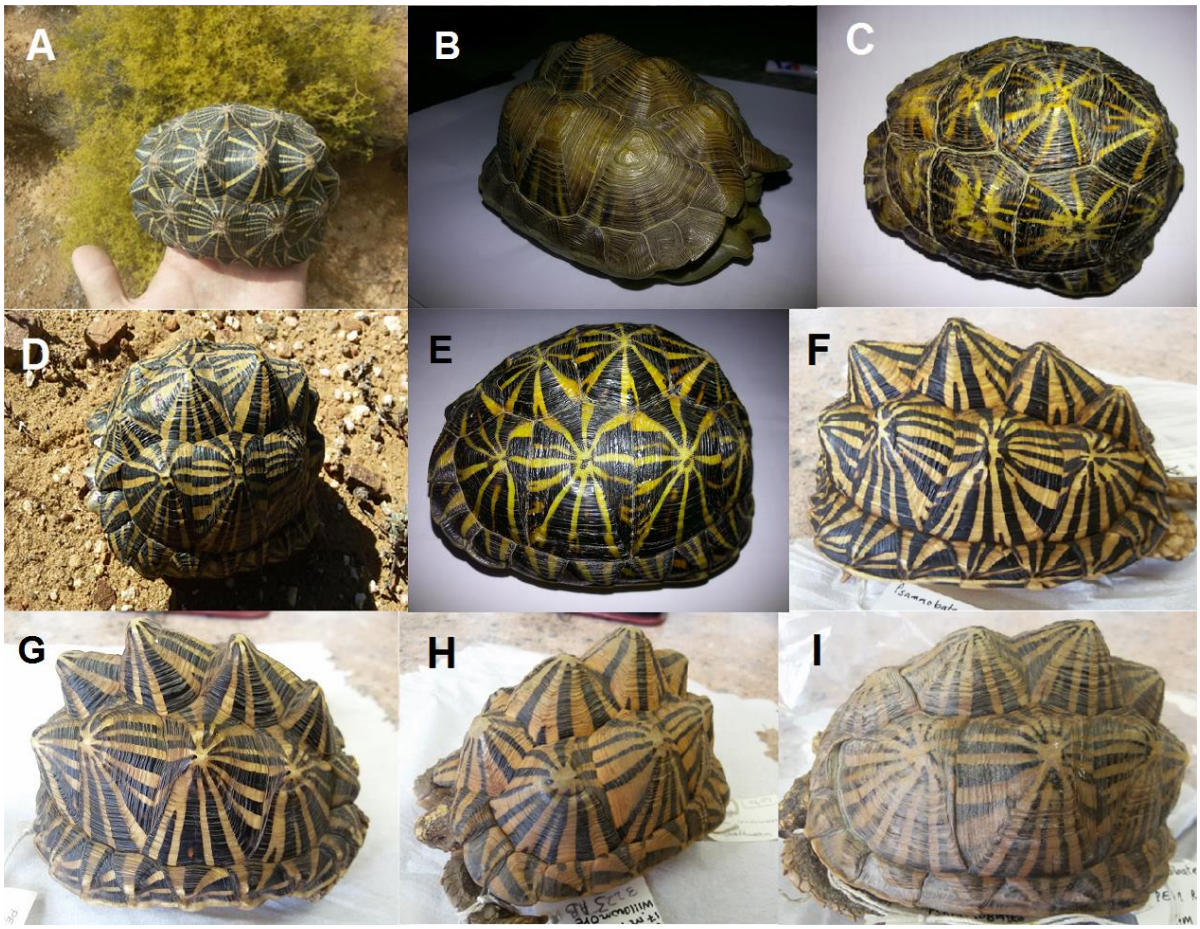


Figure S6.1. Morphological variations in carapace structure and patterns in females of C1 of the *Psammobates tentorius* species complex. The greatest variations within the group are clearly with respect to stripe thickness, stripe density, stripe dullness and degree to which scute knobs are domed. A. Specimen from Matjiesfontein, with very dark phase, highly domed carapace, high carapace scute knobs, thin stripes of equal width, and the absence of dull stripes. B. Specimen from Matjiesfontein, showing a nearly uniformly light brown phase, highly domed carapace, well-developed carapace scute knobs, and extremely dull stripes. C. Specimen from Beaufort West, with medium dark phase, highly domed carapace but only slightly developed carapace knobs, generally thin carapace stripes, highly dull stripes and moderately drab stripes. D. Specimen from Matjiesfontein, of moderately light phase, with increasingly thicker stripes from each knob apex to the margin of the scute, high domed carapace and well-developed knobs, with no dull or drab stripes. E. Specimen from Beaufort West, showing a low level light phase, moderately domed carapace, poorly developed carapace knobs, uniformly thin stripes mixed with increasingly thicker stripes, with only a few dull stripes and no drab stripes. F. Specimen from Steytlerville (PE Bayworld Museum, Voucher no: PEM R10768), showing a moderately high light phase, a very high domed carapace and highly developed carapace knobs, medium to increasingly thicker stripes, and no dull or drab stripes. G. Specimen from Somerset East (PE Bayworld Museum, Voucher no: PEM R15216), of a low level light phase, highly domed carapace and moderately developed carapace knobs, medium to increasingly thicker stripes, and no dull or drab stripes. H. Specimen from Willowmore (PE Bayworld Museum, Voucher no: PEM R15223), showing a very high level light phase, highly domed carapace, well developed carapace knobs, increasingly thicker stripes and no dull or drab stripes. I. Specimen from Jansenville (PE Bayworld Museum, Voucher no: PEM R10767), showing a high level light phase, highly domed carapace, high carapace knobs, increasingly thicker stripes mixed with thin uniform width stripes, either no dull stripes or drab stripes.

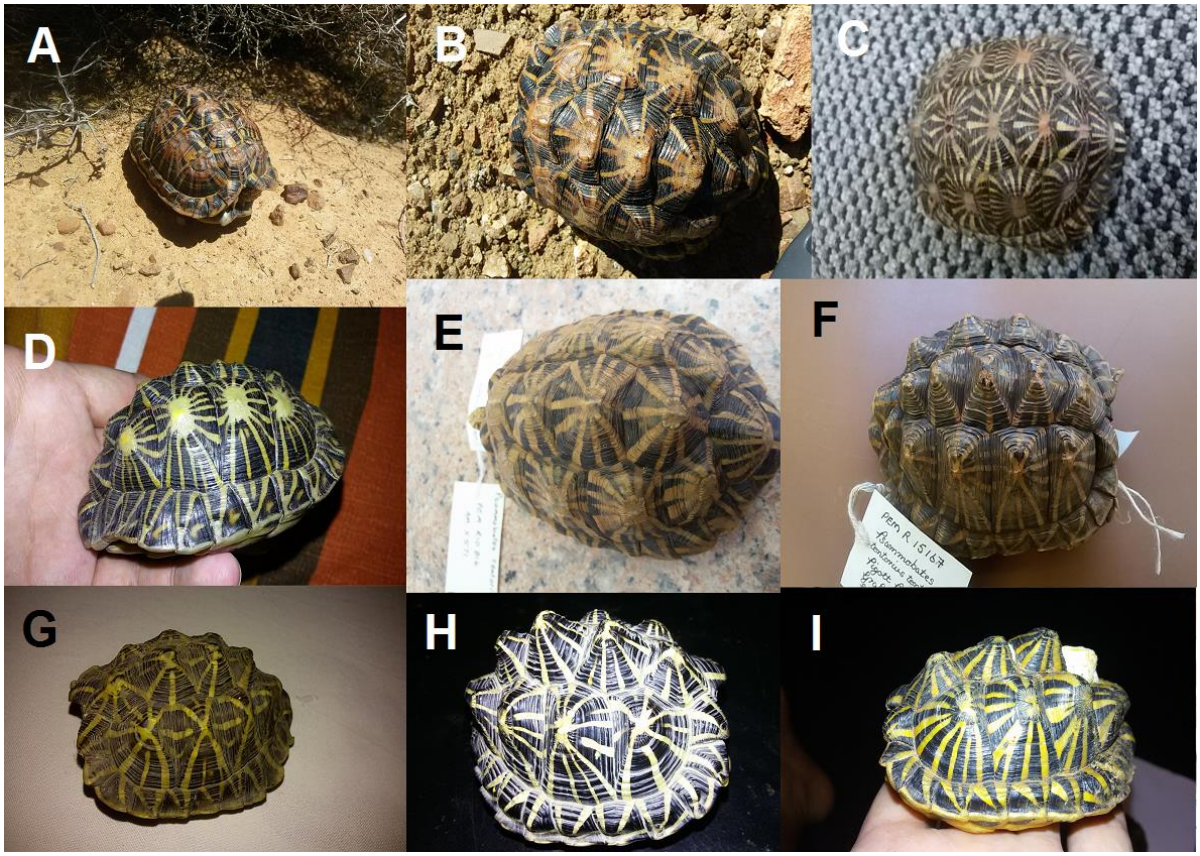


Figure S6.2. Morphological variations of the carapace in males of C1 of the *Psammobates tentorius* species complex. Most males showed a significantly enlarged apical ring in the center of each carapace scute. Morphological variation within males of the *P. t. tentorius* appeared to be less than that of females. A. Specimen from Matjiesfontein, with uniformly wide thin stripes mixed with increasingly thicker stripes, carapace knobs very poorly developed and of low to moderate height, with dull and drab stripes. B. Specimen from Matjiesfontein, with thin stripes of uniform width, low poorly developed carapace knobs, with few dull and drab stripes. C. Specimen from Matjiesfontein, with a relative high density of uniformly thin stripes, a flattened carapace dome, carapace knobs undeveloped, without dull or drab stripes, high stripe density, and large apical rings. D. Specimen from Beaufort West, with a relatively flat, smooth carapace, apical ring significantly enlarged, stripes are thin and of equal width from the apical ring to the margin of scutes, knob undeveloped, without dull or drab stripes. E. Specimen from Willowmore (PE Bayworld Museum, Voucher no: PEM R10816), with low to moderately domed carapace and moderately developed carapace knobs, mixed thick and increasingly thin stripes, no dull or drab stripes. F. Specimen from Grahamstown (PE Bayworld Museum, Voucher no: PEM R15167), with flat carapace dome, with poorly developed knobs, small apical ring, uniformly thin stripes, without dull and drab stripes. G. Specimen from Richmond, with flat carapace but moderate to high carapace knobs, a relatively large number of striped forks. H. Specimen from Grahamstown, with highly domed carapace together with distinctly high carapace knobs, with equally thin stripes. I. Specimen from Steytlerville, with a low domed carapace but moderate to high carapace knobs, carapace stripes moderate to thick.

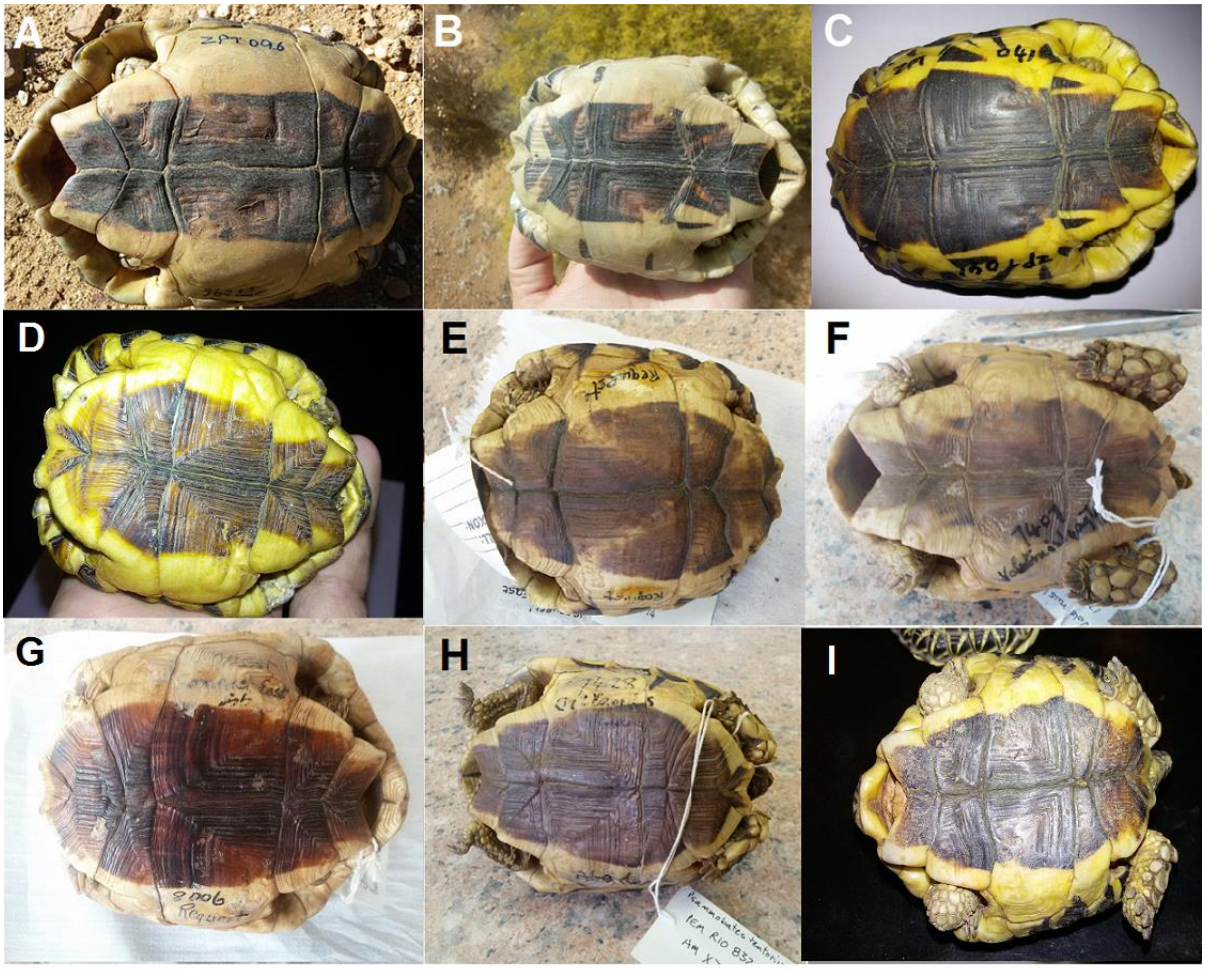


Figure S6.3. Plastron pattern variations in C1 of the *P. tentorius* species complex. A. Specimen from Matjiesfontein, with centrally situated, clearly delimited mahogany coloured plastron block (solid PCB). B. Specimen from Matjiesfontein, with a centrally situated clearly delimited mahogany coloured plastron block with slightly indented marks and stripes. C. Specimen from Beaufort West, with black clearly delimited plastron mark with side stripes. D. Specimen from Steytlerville, with clearly delimited mahogany coloured block. E. Specimen from Somerset East (PE Bayworld Museum Voucher No: PEM R10781), with plastron nearly fully covered by an oval shaped mahogany block, with few indented stripes. F. Sample from Willowmore (PE Bayworld Museum Voucher No: PEM R15223), with plastron centered, lightly coloured diffuse partially bounded block. G. Specimen from Somerset East (PE Bayworld Museum Voucher No: PEM R15216), with plastron having a solid, well bounded oval mahogany block. H. Specimen from Aberdeen (PE Bayworld Museum Voucher No: PEM R10832), with plastron almost fully covered by an oval shaped light mahogany coloured block. I. Specimen from Grahamstown, with plastron with a solid, well bounded near rectangular shaped, near black mahogany coloured block.

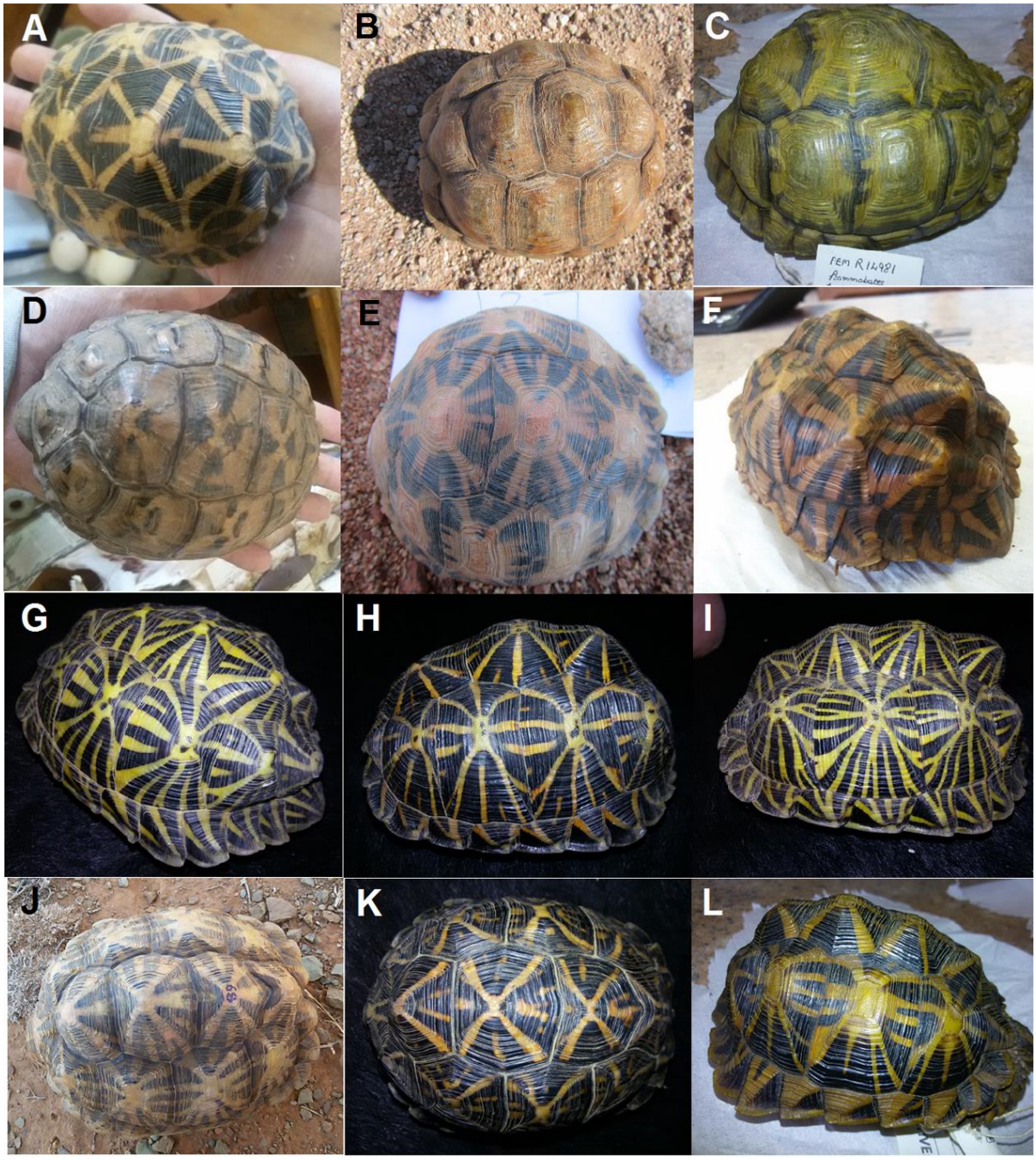


Figure S6.4. Carapace shape and pattern variation among females in C2 of the *P. tentorius* species complex. A. Specimen from Loeriesfontein, with sparse uniformly thick stripes on a high domed carapace, carapace knobs poorly developed, few dull stripes, no drab stripes, very dark phase. B. Specimen from Kenhardt, with an almost uniformly brown carapace, carapace moderate to highly domed, carapace knobs undeveloped, very light phase. C. Specimen from Pofadder (PE Bayworld Museum Voucher No: PEM R14981), nearly uniformly brown carapace with only few very dull and drab stripes, carapace highly domed with faintly developed knobs, very light phase. D. Specimen from Loeriesfontein, carapace dome high with thick very dull and drab stripes, knobs only poorlyly developed, stripes all increasing in thickness, very light phase. E. Specimen from Kenhardt, with very dull stripes, carapace dome moderate to high, knobs undeveloped, light phase. F. Specimen from Strydenburg (PE Bayworld Museum Voucher No: PEM R10834), moderate to highly domed carapace with increasing thick and well-marked stripes, no dull or drab stripes, knobs moderately to highly

developed, moderately light phase. G. Specimen from Carnarvon, with increasingly thick stripes, no dull or drab stripes, carapace dome moderate, knobs moderately to highly developed, very easily confused with *P. t. tentorius*, very light phase. H. Specimen from Loxton, with equal width thin stripes on its moderate to highly domed carapace, knobs slightly developed, without dull or drab stripes, very dark phase. I. Specimen from Carnarvon, with thin to increasingly medium width stripes, no dull or drab stripes, carapace dome high, knobs well developed, very easily confused with *P. t. tentorius*, very light phase. J. Specimen from Carnarvon, increasingly thick stripes mixed with equal width thick stripes on a highly domed carapace, knobs moderately developed, with few dull stripes but no drab stripes, moderately light phase. K. Sample from Carnarvon, with increasingly diffuse stripes, which are medium to thick, with dull stripes but no drab stripes, moderately domed carapace, knobs faintly developed, moderate to very dark phase. L. Sample from Hopetown (PE Bayworld Museum Voucher No: PEM R10770), highly domed carapace with sparse increasingly thick stripes mix with uniformly thin stripes, no dull or drab stripes, knobs moderately developed, slightly dark phase.

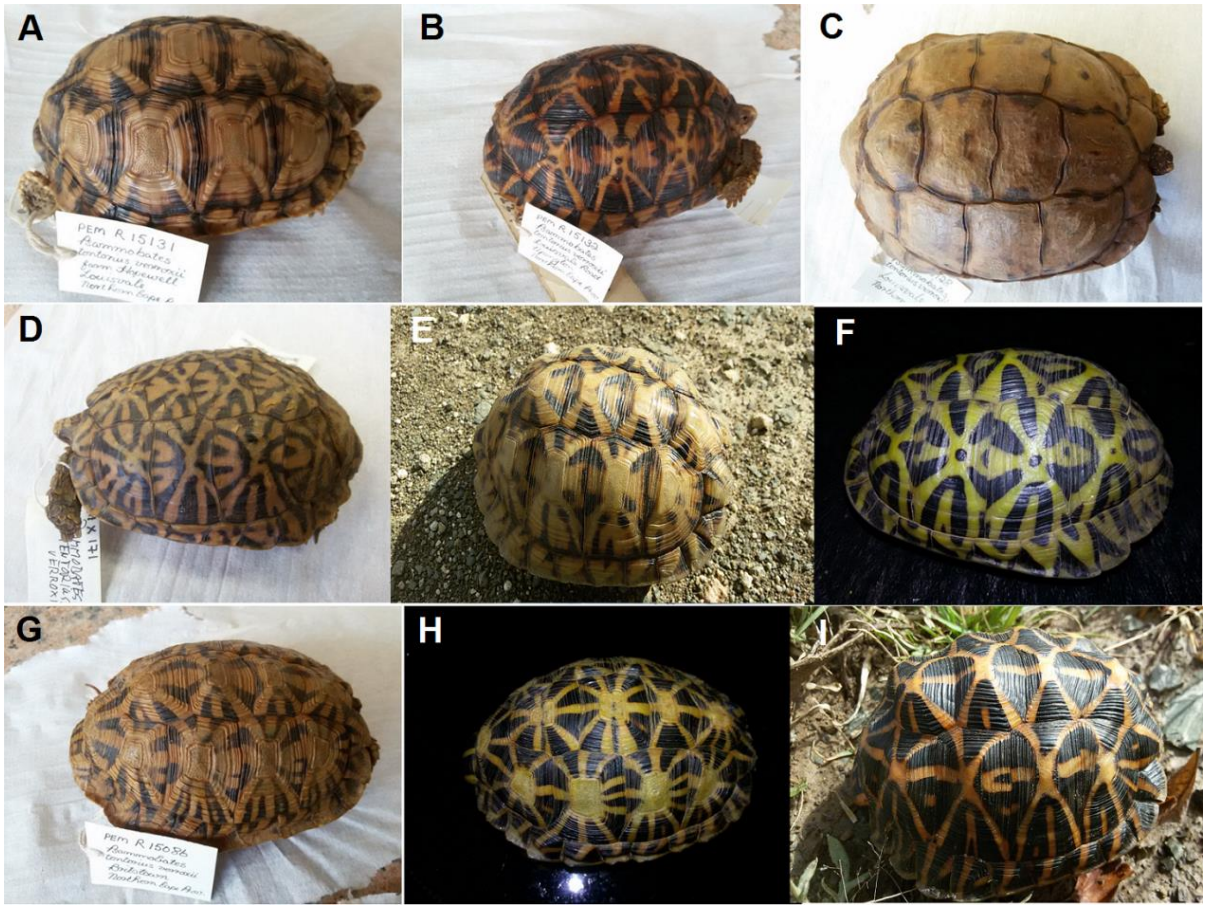


Figure S6.5. Carapace shape and pattern variation among males in C2 of the *P tentorius* species complex. In general males of C2 have a flattened carapace with undeveloped carapace knobs, except for a sample of the Upper Karoo area. A. Specimen from Louisvale, near Upington (PE Bayworld Museum Voucher No: PEM R15131), shows many dull and drab stripes, stripes vary from thick to thin, very light phase. B. Specimen from Louisvale Road, near Upington (PE Bayworld Museum Voucher No: PEM R15132), showing few drab stripes but many dull stripes, thin stripes mixed with thick ones, low to moderate dark phase. C. Specimen from Louisvale (PE Bayworld Museum Voucher No: PEM R15128), showing an almost uniformly brown colour, very feint dull and drab stripes, very light phase. D. Specimen from the Upington area (PE Bayworld Museum Voucher No: PEM R15102), with increasingly thick stripes, but no dull or drab stripes, very light phase. E. Specimen from Prieska, with increasingly very thick stripes, many dull or drab stripes, very light phase. F. Sample from Carnarvon, with thick and increasingly well-marked stripes, no dull or drab stripes, and medium light phase. G. Sample from the Britstown (PE Bayworld Museum Voucher No: PEM R15086), with uniformly thick stripes, only few dull stripes but no drab stripes, very light phase. H. Sample from Williston, with uniformly thin stripes and large apical rings, no dull or drab stripes. I. Sample from Hanover, with equal width distinctly thin stripes, no drab or dull stripes, moderately domed carapace, and carapace knobs moderately high.

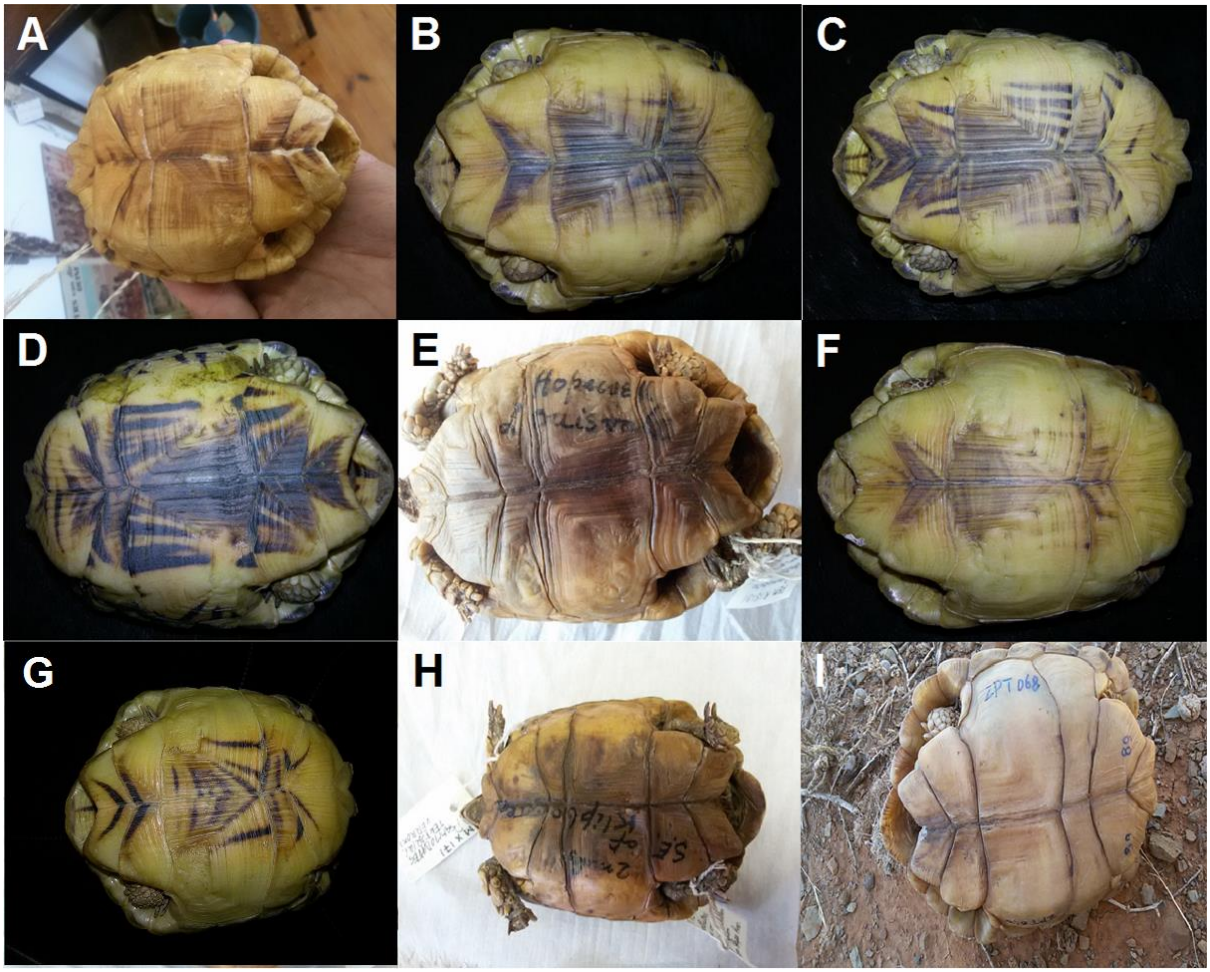


Figure S6.6. Plastron pattern variations in C2 of the *P. tentorius* species complex are also considerably high. From no pattern (See E, H and I) to diffusely marked with a radiated sparse central plastron block (PCB), (see A, B, G and F) or distinctly bounded but indented or with broken stripes (see C, D). A. Loeriesfontein. B. Loxton. C. Carnarvon. D. Loxton. E. Louisvale (PEM R15131). F. Carnarvon. G. Van Rooyensvlei, Gordonia (PEM R15112). H. Upington (PEM R15102). I. Carnarvon.

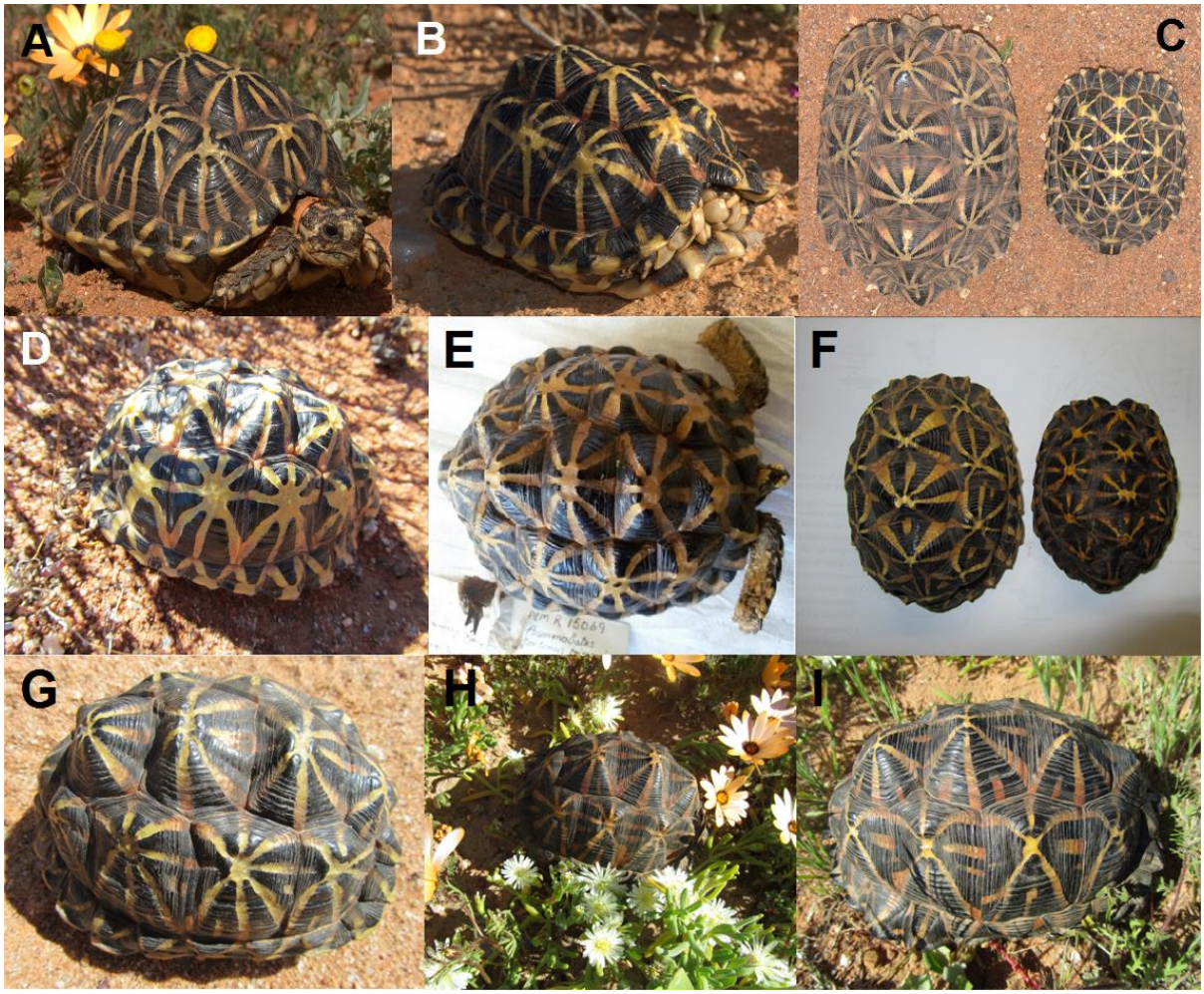


Figure S6.7. Carapace variation in C3. All individuals of C3 have well-marked stripes, no dull or drab stripes, orange to reddish stripes, each scute usually has a reddish margin, stripe density is generally low. A. Specimen from Bitterfontein (female), with equal width well-marked stripes, stripes generally thin to medium width, carapace dome high, with highly developed knobs, lower level of dark phase. B. Specimen from Nuwerus (male), with thin equal width stripes, carapace dome low, with low knobs, moderate to highly dark phase. C. Specimens from Nuwerus (left: female, right: male), female with thick stripes of increasing width, high dome, moderately developed knobs, very light phase; male with thin equal width stripes, moderate to highly domed carapace, moderately developed knobs, dark phase. D. Specimen from the Richtersveld (male), with uniformly thin to medium width stripes, low carapace dome, knobs moderately developed, lower light phase. E. Specimen from southern Namaqualand (PEM R15069), with thick equally well-marked stripes, carapace dome is moderately high, with moderate to high knobs, low light phase. F. Bitterfontein specimens (left: female, right: male), body shape appears not to differ as significantly as in the other clades, all with moderate to increasingly thick well-marked stripes, carapace highly domed, with well-developed knobs, low dark phase. G. Nuwerus specimen, with uniformly well-marked medium thickness stripes, carapace highly domed, knobs highly developed, light dark phase. H. and I. Bitterfontein specimens, all with well-marked increasingly thick stripes, carapaces highly domed, knobs highly developed, all moderately dark phase.

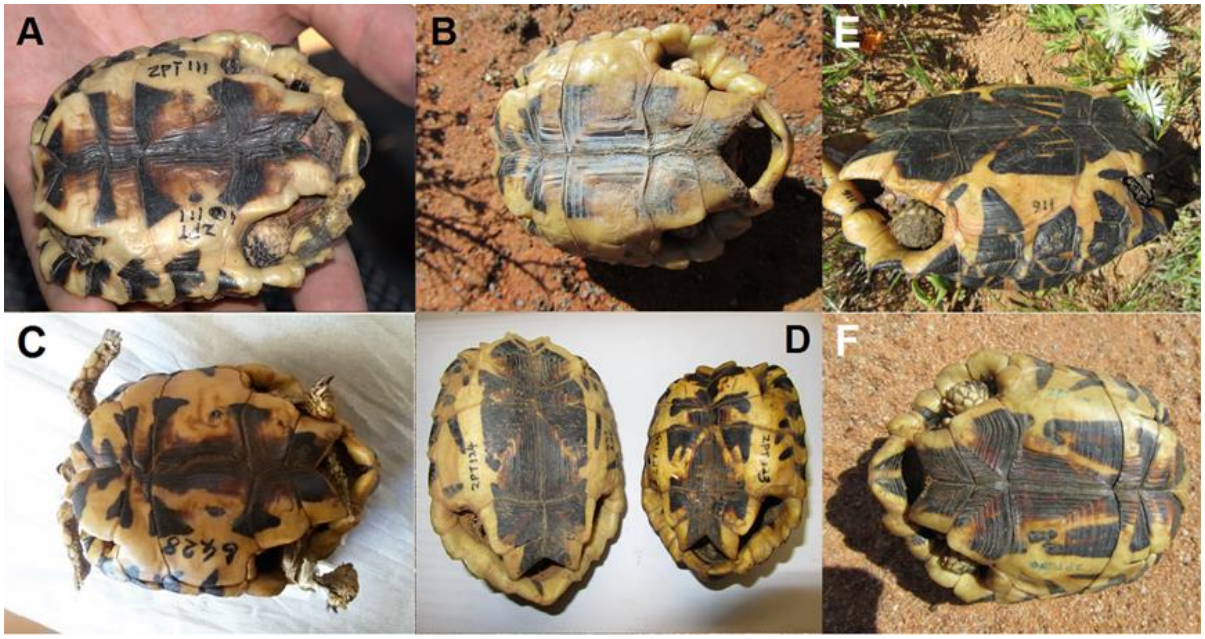


Figure S6.8. Plastron pattern variations within C3. A and F: specimens from Nuwerus; D and E: specimens from Bitterfontein; B: specimen from the Richtersveld; C: specimen from Van Rhynsdorp.

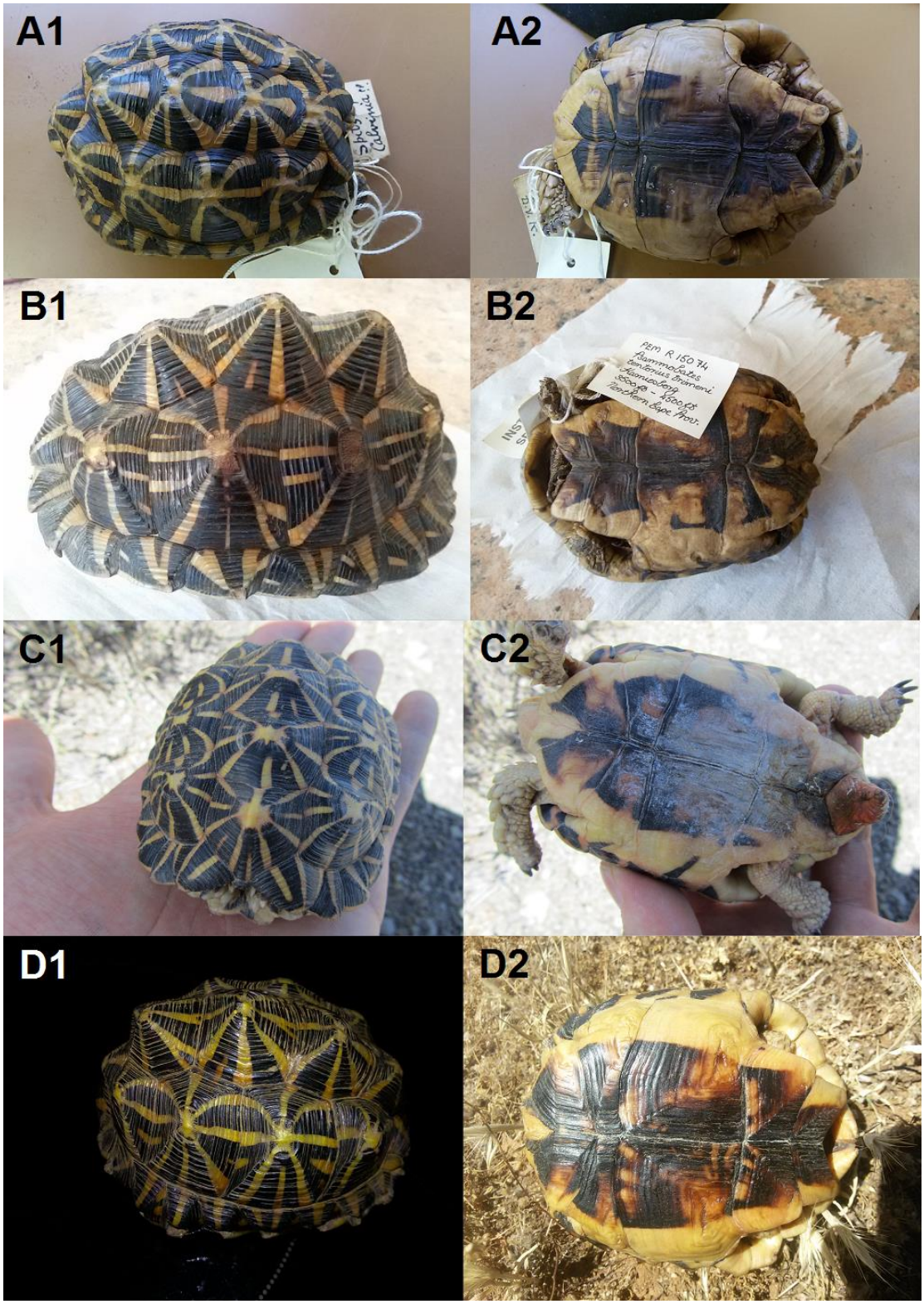


Figure S6.9. Carapace and plastron pattern variations within C4 are relatively small. Clade 4 normally has well-marked, light-yellow carapace stripes. Generally, C4 has well-bounded black or brown mahogany plastron markings (solid PCB) with indentations or broken stripes (but usually not as many as in C3). Thus, C4 is morphologically similar to C3 in general. A. A male specimen from (PE museum voucher: PEM R10835). B. A female specimen from the Kamiesberg region (PE museum voucher: PEM R15074). C. A male specimen from Sutherland. D. A female specimen from Loeriesfontein.

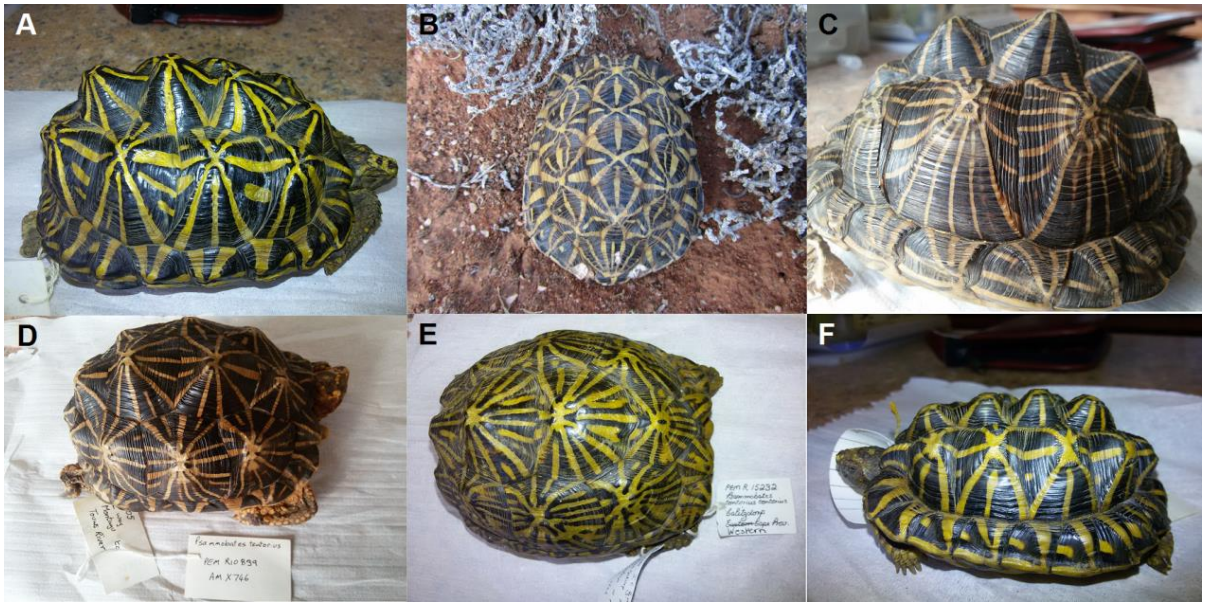


Figure S6.10. Carapace pattern variations within C5 are relatively small. The key characters distinguishing C5 from the other clades are the relatively dense and bent carapace stripes. A. A female specimen from Ladismith (PE Museum: PEM R10782). B. A male specimen from Barrydale. C. A female specimen from Montagu (PE Museum: PEM R15210). D. A female specimen from Touws Rivier (PE Museum: PEM R10839). E. A female specimen from the Montagu area (PE Museum: PEM R15232). F. A male specimen from Garcia Pass (PE Museum: PEM 15213)

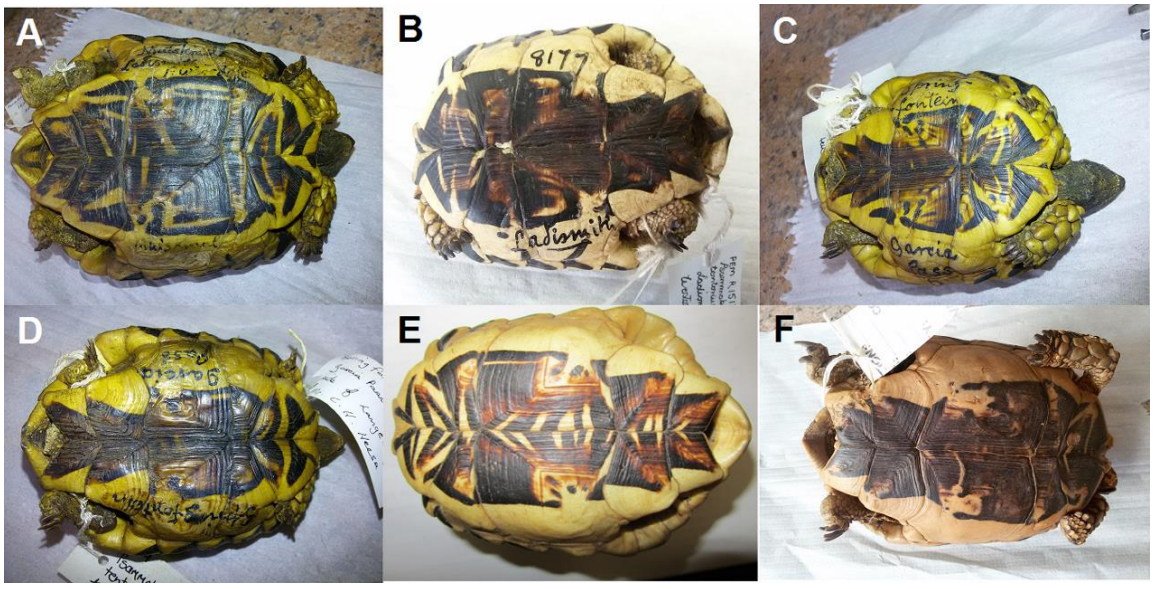


Figure S6.11. Plastron pattern variations within C5 are also relatively small. The key character distinguishing C5 from the rest of the clades are the distinct stripes on the plastron. A. A female specimen from Ladismith (PE Museum: PEM R10782). B. A male specimen from Ladismith (PE Museum: PEM 15179). C. A male specimens from Garcia Pass (PE Museum: PEM R15175 and PEM R15213). E. A female specimen from Barrydale. F. A female specimen from Montagu (PE Museum: PEM 10839).

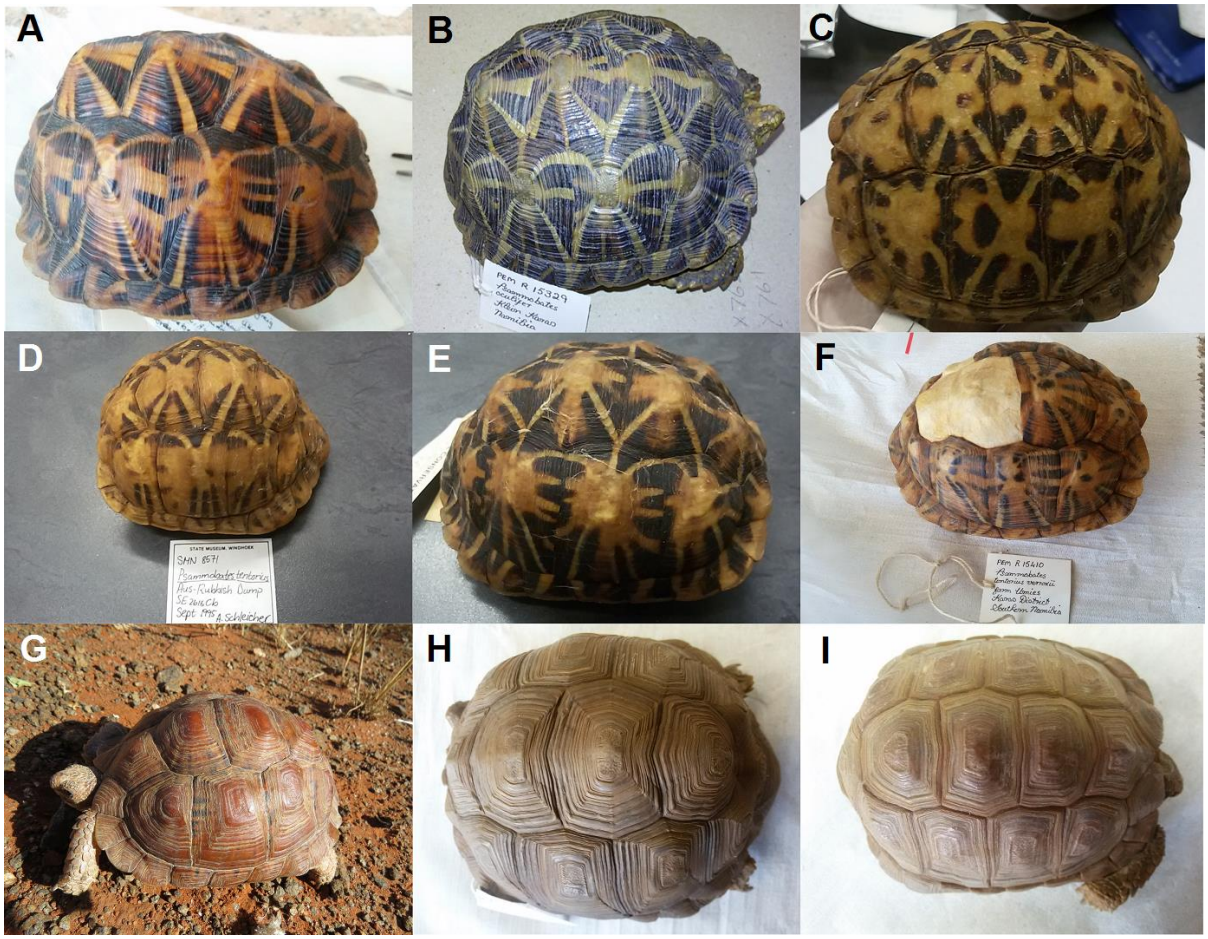


Figure S6.12. Carapace pattern variations within C6 are relatively large. It includes the Namibia populations (A – F) and the "*bergeri*" morph (G – I). The key character distinguishing C6 from the other clades is the unique "X" mark on the carapace or the uniformly brown coloured carapace without any stripes (G – I). A. A female specimen from Ukamas - Namibia (PE Museum: PEM R15111). B. A female specimen from Klein Karas - Namibia (PE Museum: PEM R15329). C. A male specimen from Lüderitz - Namibia (State Museum of Windhoek, Namibia: SMR 8168). D. A female specimen from Aus - Namibia (State Museum of Windhoek: SMR 8571). E. A female specimen from Aus - Namibia (State Museum of Windhoek: SMR 8671). F. A female specimen from southern Namibia (PE Museum: PEM R15410). G. Female "*bergeri*" specimens from Noenieput and the Gordonia region (PE Museum: PEM R15112). I: A male "*bergeri*" specimen from the Upington region (north of Orange River, PE Museum: PEM R15331).

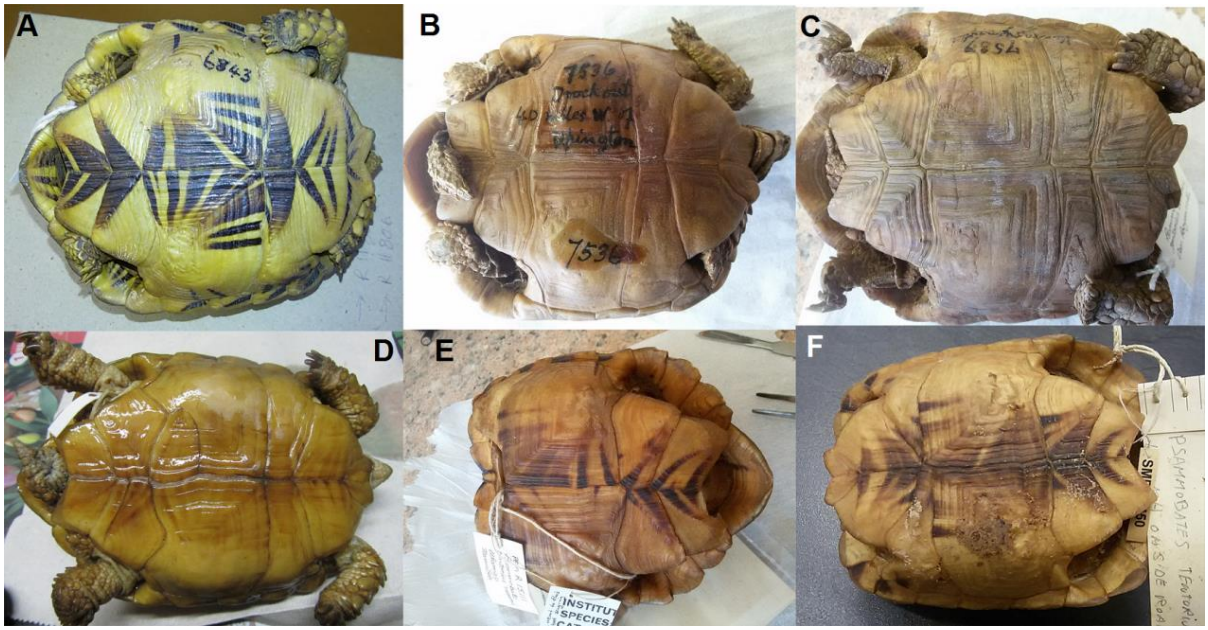


Figure S6.13. Plastron differences found between the "*bergeri*" morph (B, C and D) and the rest of the morphs (A, E and F) in C6. The plastron patterns of non - "*bergeri*" (A, E and F) populations are similar, while all "*bergeri*" morphs (B, C and D) have no plastron markings.

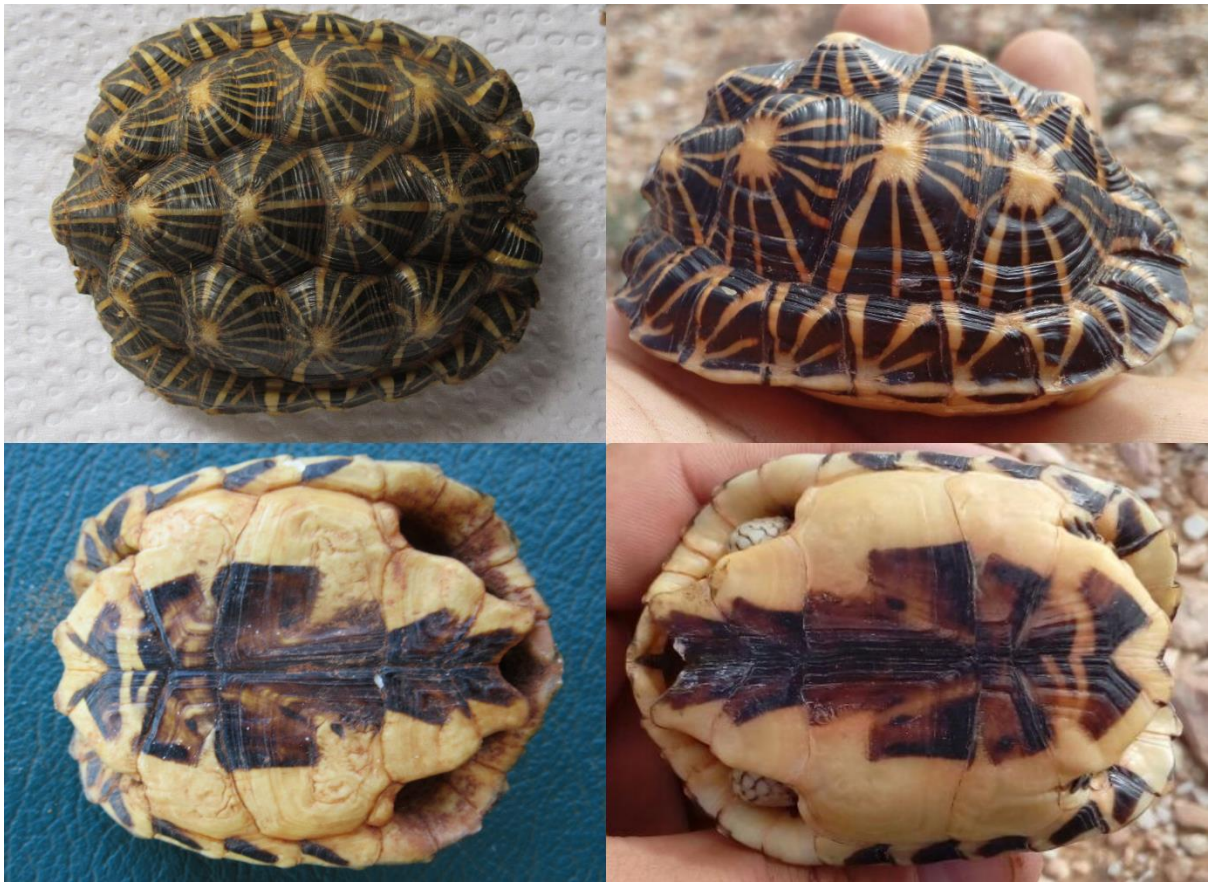


Figure S6.14. The carapace and plastron colour patterns in C7 are relative stable (less variable) compared to the other clades in the *P. tentorius* species complex. It usually has a plastron with few broken stripes. Carapace stripes are often thin and orange in colour, and the apical ring distinct compared to the other clades. The individual on the left was found in Oudtshoorn, and the one on the right was found near Calitzdorp.

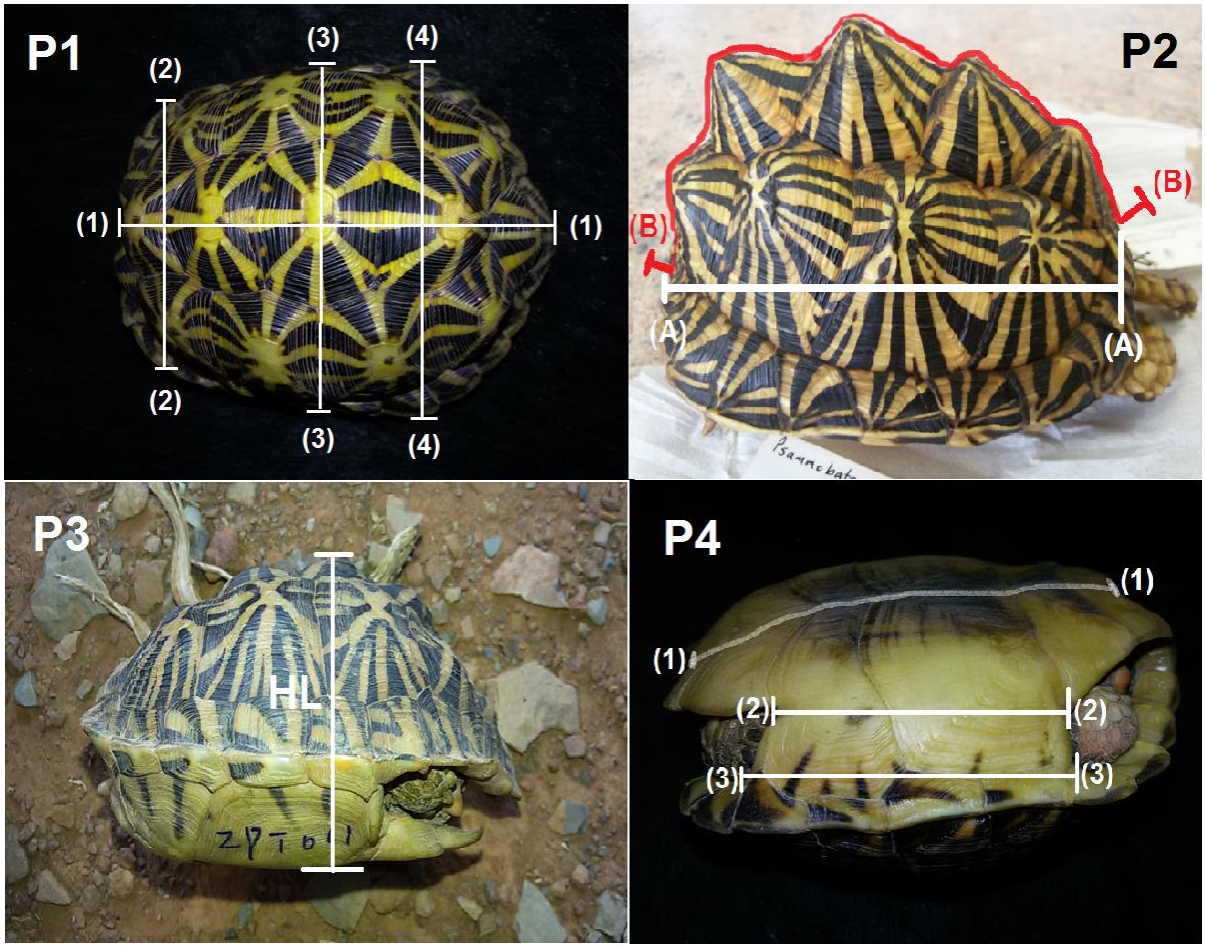


Figure S6.15. Measurements of the plastron and carapace. **P1**: (1) CLS, (2) UW, (3) CW and (4) LW. **P2**: (1) CLS and (2) CLE. **P3**: HL. **P4**: (1) PLE, (2) AILI and (3) AILO.

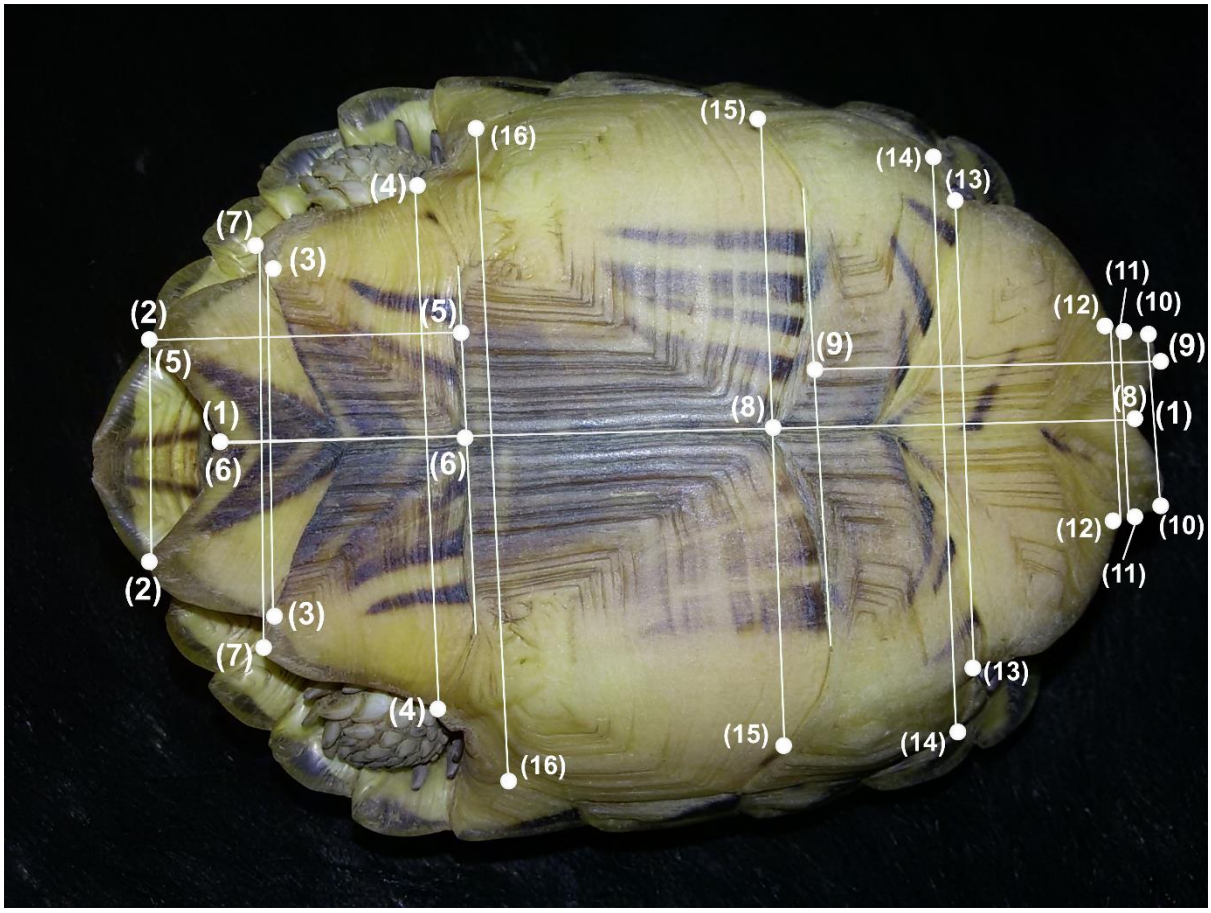


Figure S6.16. Measurements of the plastron. (1) PLS, (2) AWD, (3) AWT, (4) FWT, (5) FALS, (6) FALC, (7) FWD, (8) GPLC, (9) GPLS, (10) GWT, (11) GWD, (12) HWT, (13) HWD, (14) OHW. (15) OPW and (16) OAW.

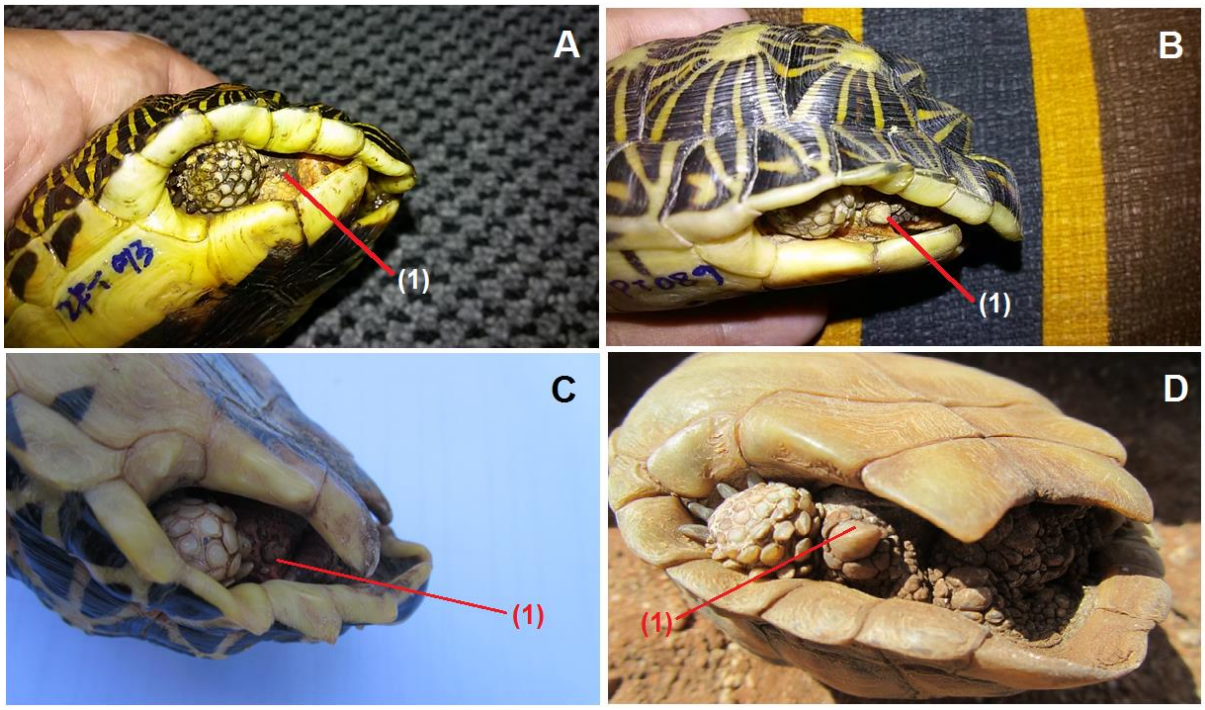


Figure S6.17. Buttock tubercle morphological variations among the three subspecies. "(1)" indicates the buttock tubercle position. A. The highly vestigial buttock tubercle C1 of *P. t. tentorius* (Matjiesfontein). B. The poor to moderately developed buttock tubercle of C1 of *P. t. tentorius* (Beaufort West). C. The poorly developed buttock tubercle of C3 of *P. t. trimeni* (Richtersveld). D. The highly developed buttock tubercle of C2 of *P. t. verroxii* (Kenhardt).

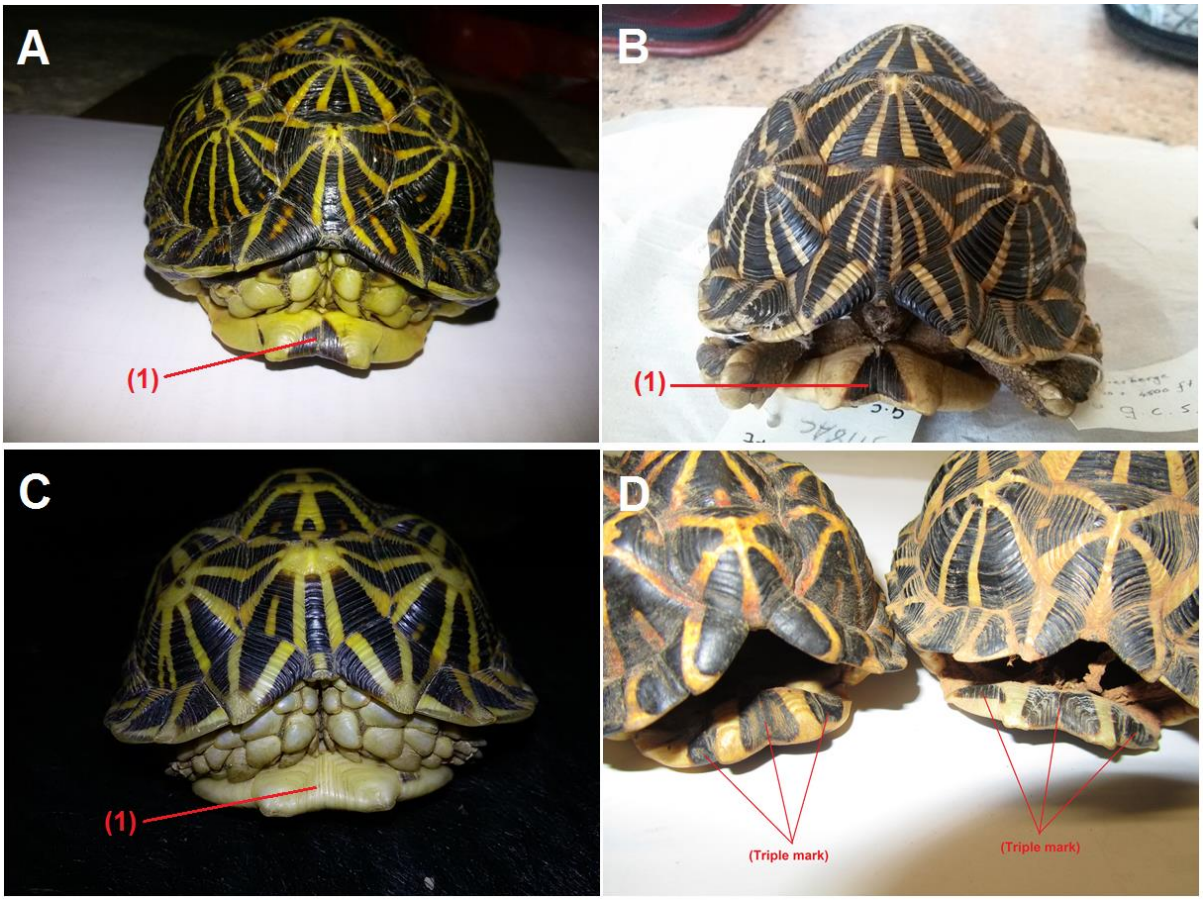


Figure S6.18. The number "(1)" is used to indicate where the triangular gular mark is supposed to be located. A. C1 with its triangular gular mark (occasionally absent). B. C4 with its triangular gular mark (though sometimes triple mark can also be found). C. C2 without a triangular gular mark. D. C3 with its uniquely triple gular marks.

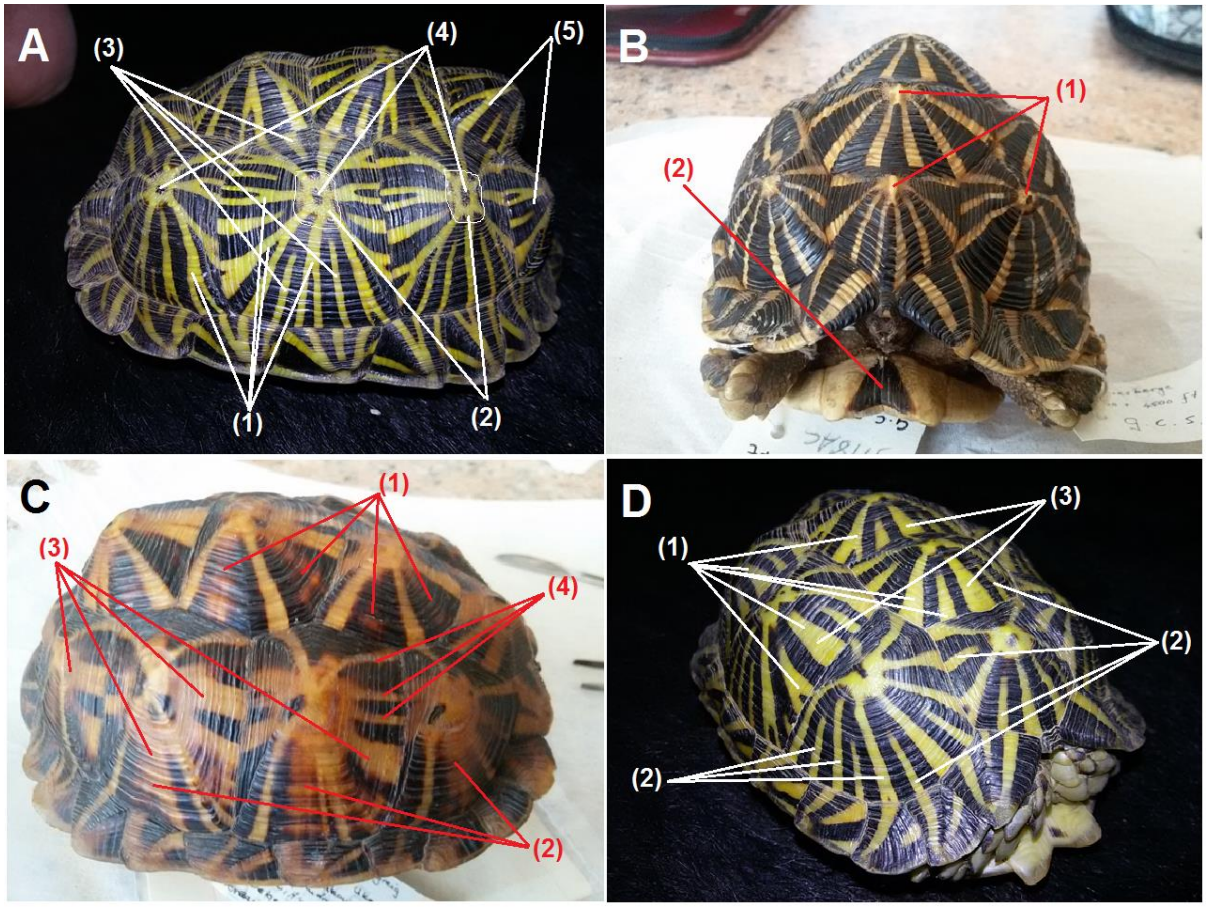


Figure S6.19. Morphological characters used in the morphometric and phenotypic analyses. **A**: (1) Stripe forks. (2) Apical ring. (3) Discontinuous stripe. (4) Apical ring spot (5) Carapace knob. **B**: (1) Apical ring knob. (2) Triangular mark. **C**: (1) Dull stripe. (2) Drab stripe. (3) Thick stripe. (4) Thin stripe. **D**: (1) Increasing width stripe. (2) Uniformly thick stripe or uniformly thin stripe. (3) Cone stripe; "dull stripe" refers to a carapace stripe which is not well bounded and shallow (not clearly defined) (see Fig. 6.19 C-1), and"drab stripe" is a thick carapace stripe with thin rays inside (see Fig. 6.19 C-2).

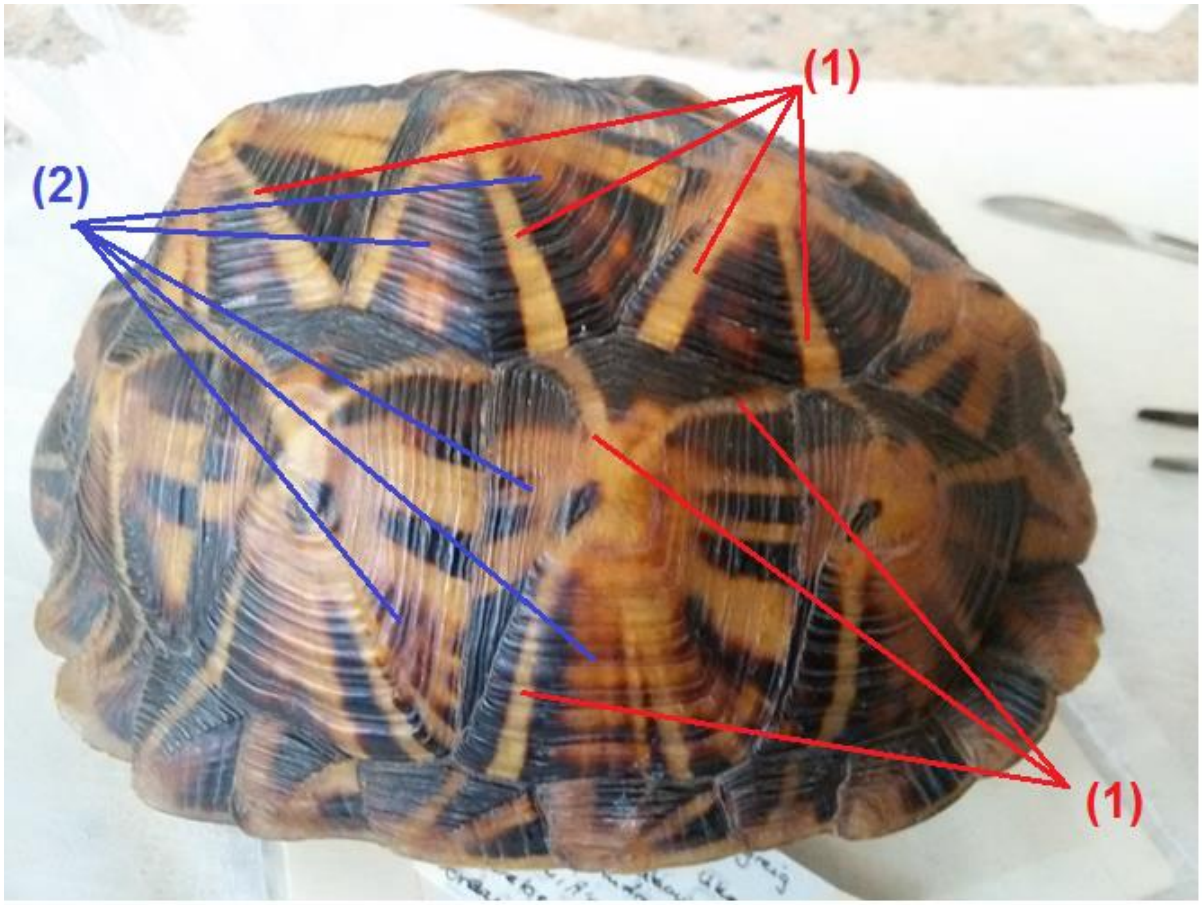


Figure S6.20. Carapace pattern configuration of C6 (Namibia) showing “X” shaped stripes (1) linking the four corners of each carapace scute. (2) Highly drab horizontal or vertical stripes on each carapace scute.

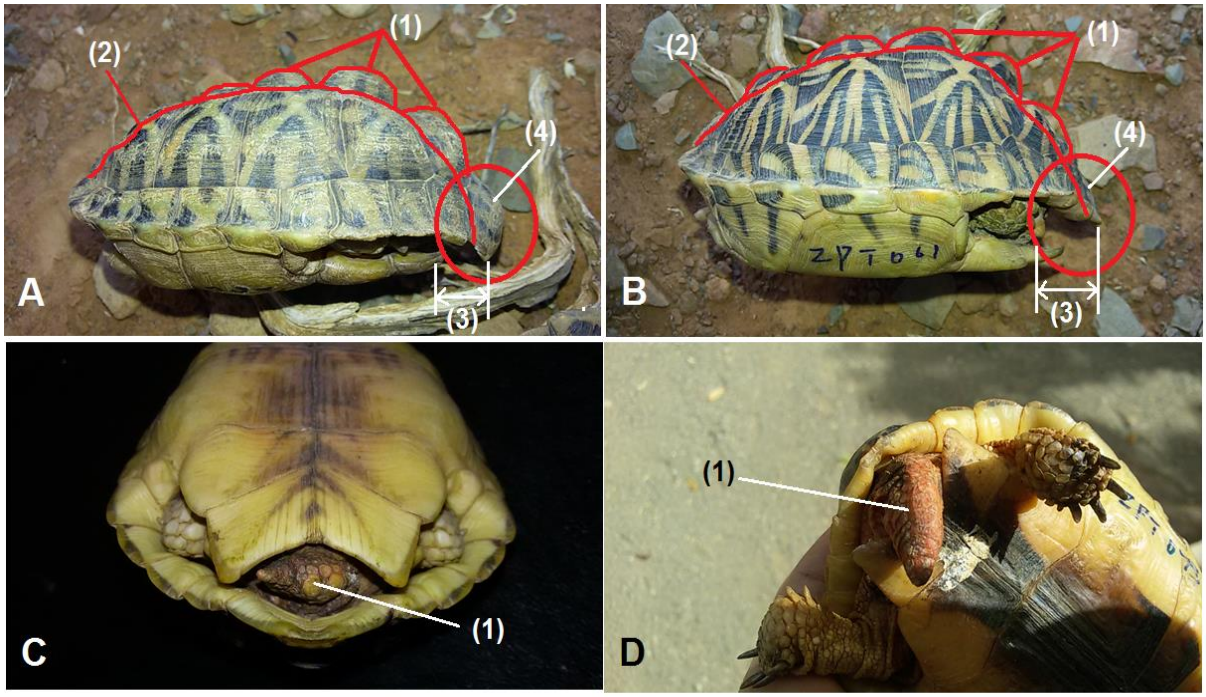


Figure S6.21. Sexually dimorphic features in *P. tentorius*. **A**: Male, (1) Carapace knob. (2) Carapace dome. (3) Distance between tip of caudal scute and central anal scute. (4) Caudal scute. **B**: Female, (1) Carapace knob. (2) Carapace dome. (3) Distance between tip of caudal scute and central anal scute. (4) Caudal scute. **C**. Short tail of female. **D**. Long tail of male.

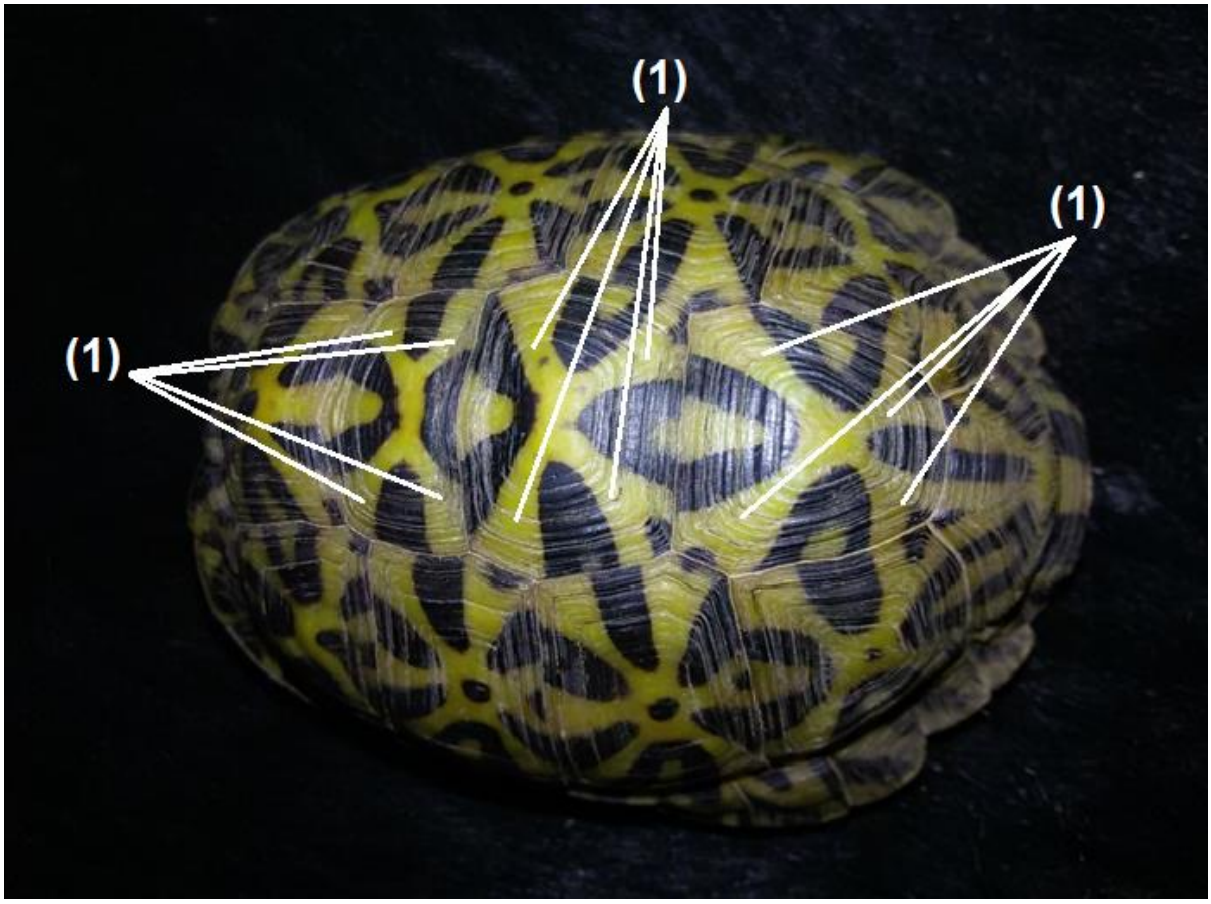


Figure S6.22. (1). Apical ring with “X”-shaped stripe on carapace.

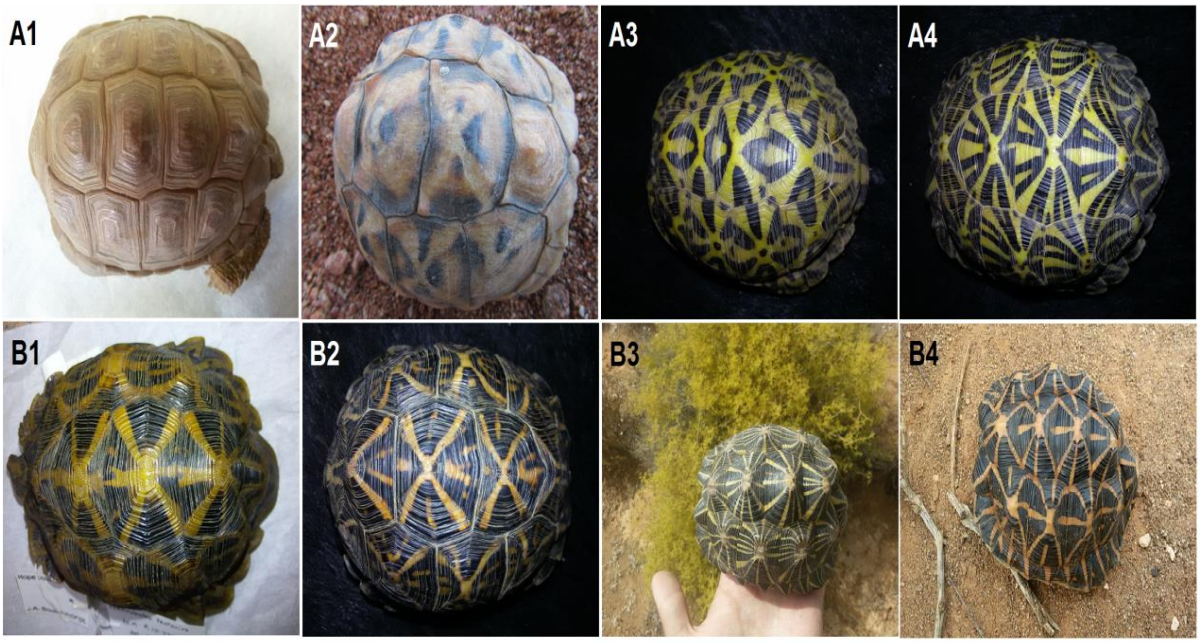


Figure S6.23. Light to dark phase colour morph gradients and scores of weighting. **A1**. Very light phase (weight score 3). **A2**. Light phase (score 2). **A3**. Moderately light phase (score 1). **A4**. Light phase (score 0). **B1**. Slightly dark phase (score 0). **B2**. Moderately dark phase (score 1). **B3**. Dark phase (score 2). **B4**. Very dark phase (score 3).

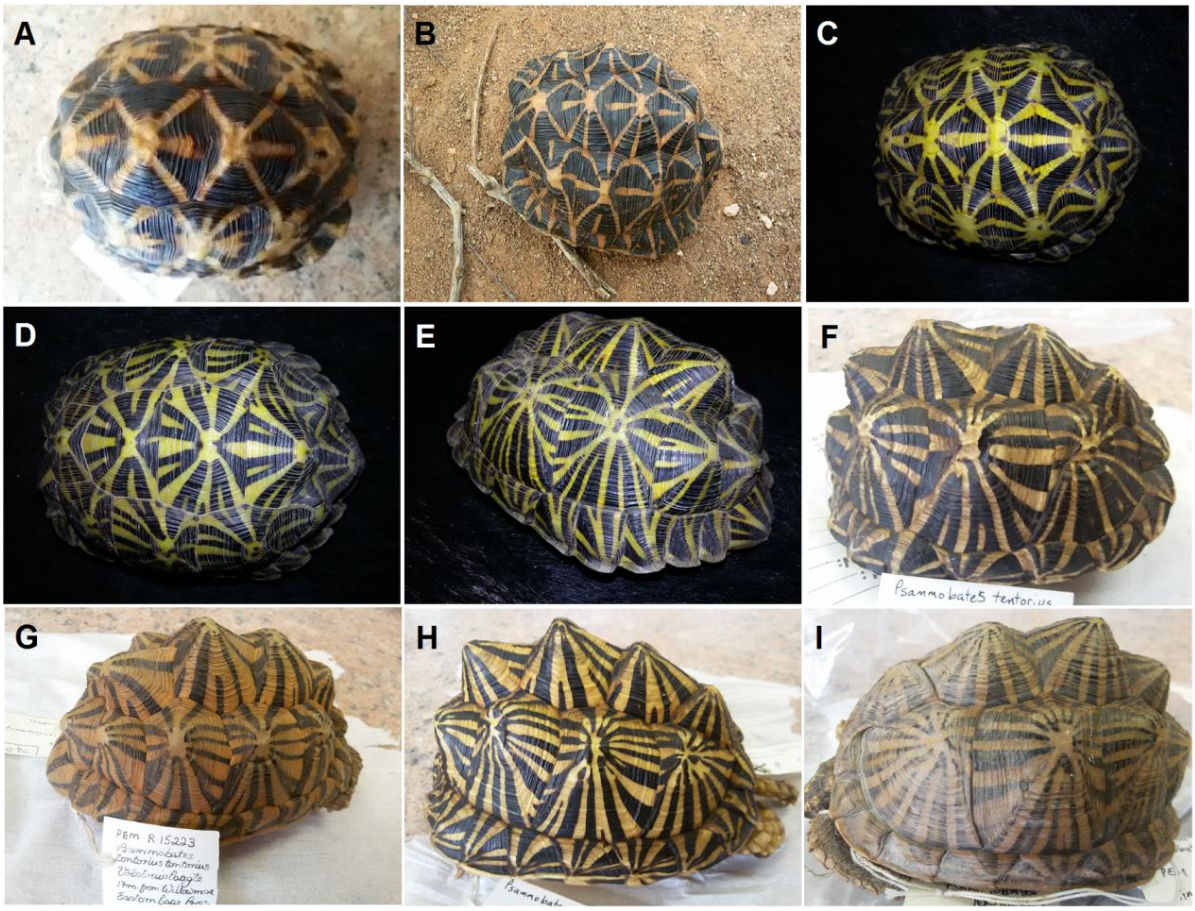


Figure S6.24. Carapace stripe density with weighting scores: A-B: low (score 0), C-D: Moderate (score 1), E-F: ordinary (score 2), G-I: high (score 3).

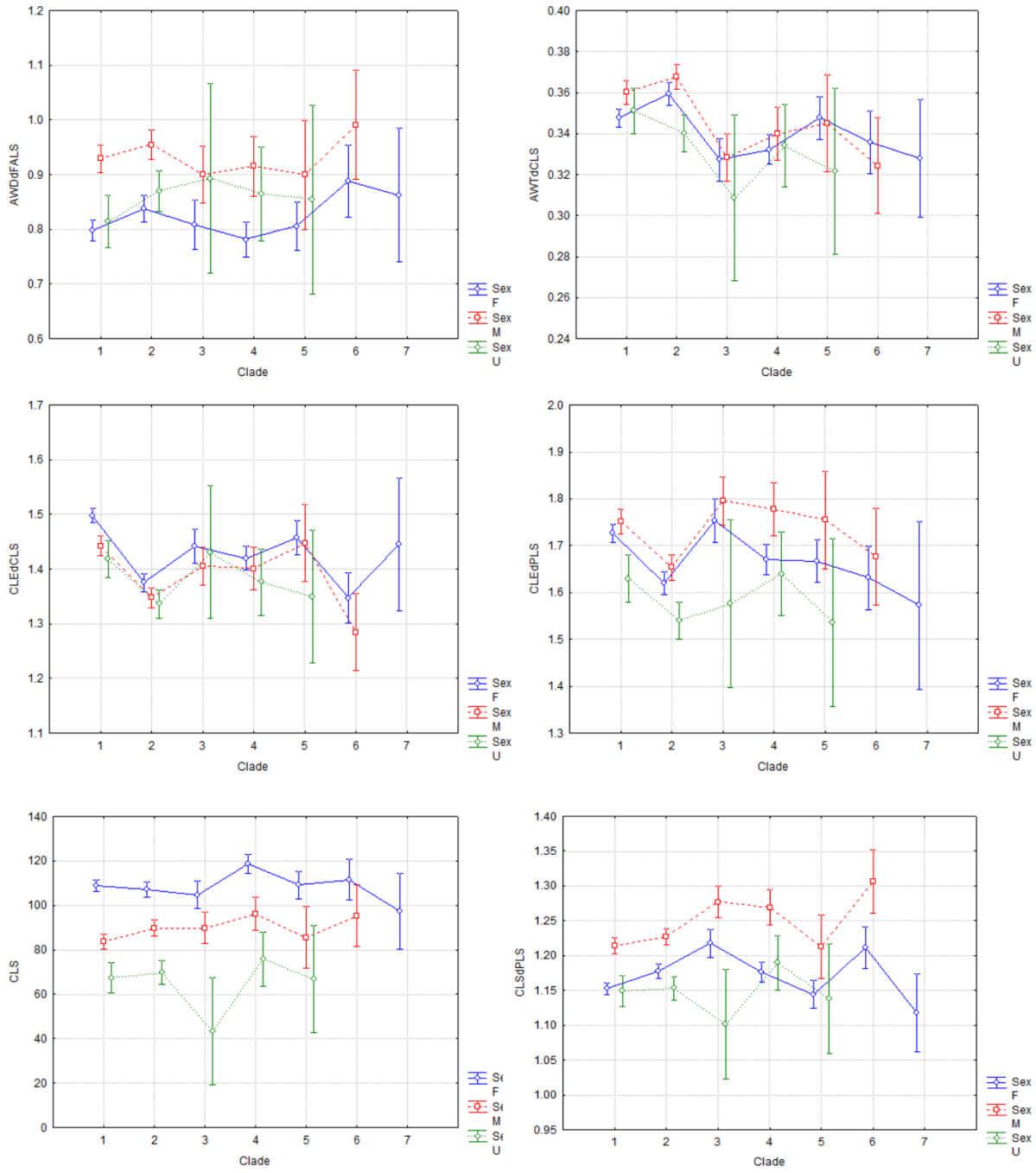


Figure S6.25. CLS ratios based on absolute values (in mm): AWD/FALS, AWT/CLS, CLE/CLS, CLE/PLS and CLS/PLS among the seven clades for the sex groups retrieved from LSD Post-hoc analyses in factorial Two-way ANOVA. F: Female, M: Male, U: Juvenile (similar for all figures in this study).

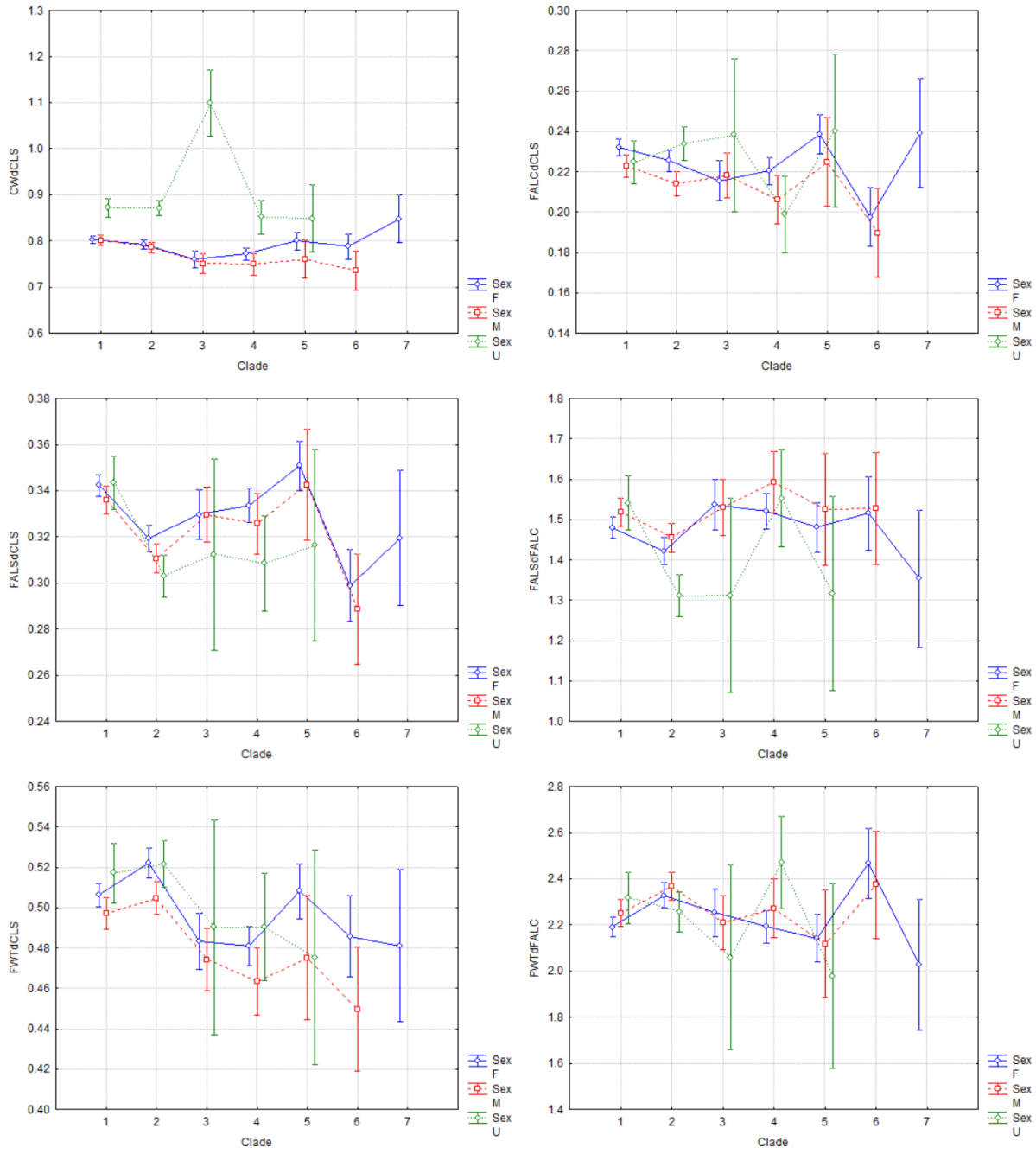


Figure S6.26. The ratios: CW/CLS, FALC/CLS, FALS/CLS, FALS/FALC, FWT/CLS and FWT/FALC among the seven clades for the sex groups retrieved from LSD Post-hoc analyses in factorial Two-way ANOVA.

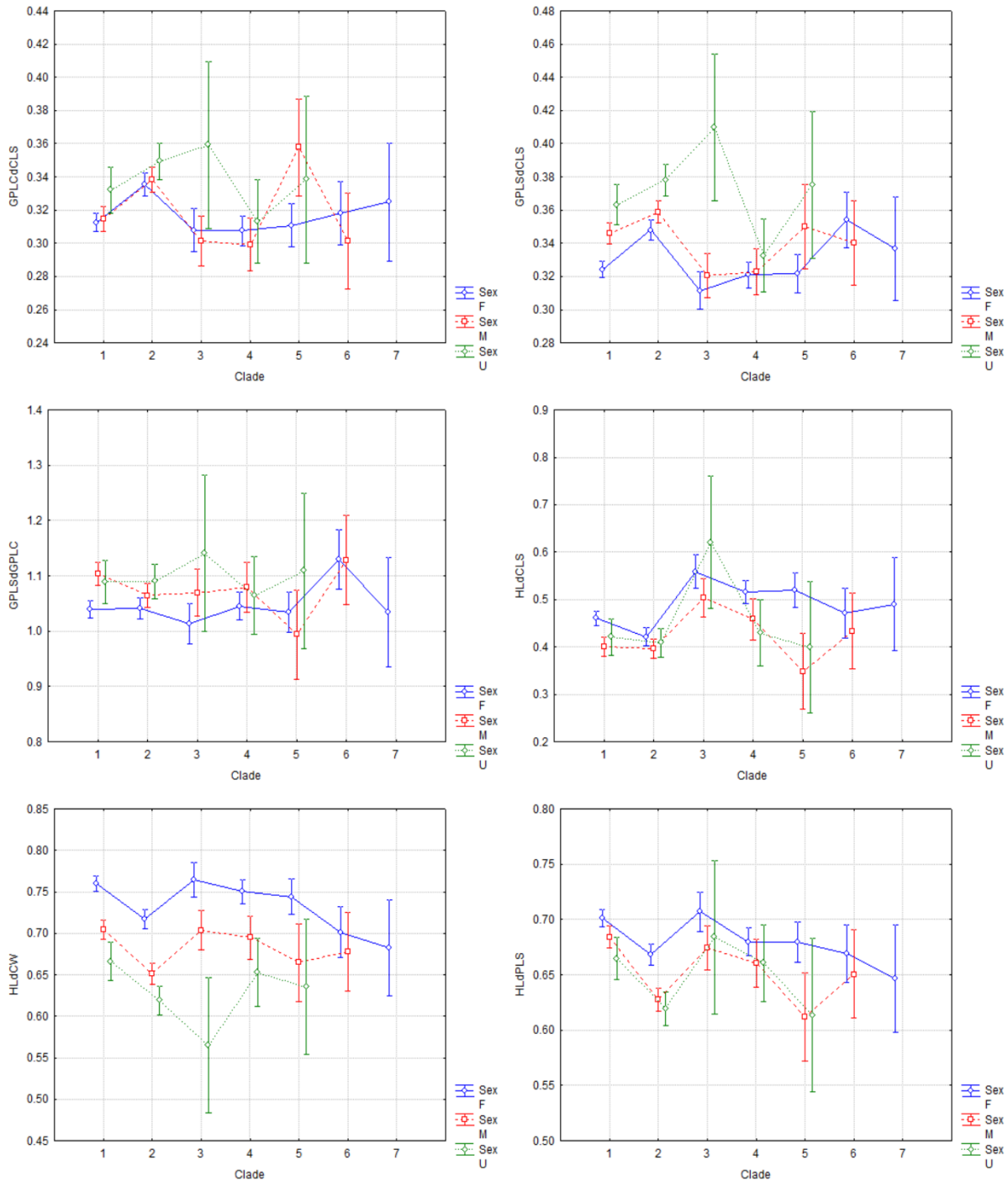


Figure S6.27. The ratios: GPLC/CLS, GPLS/CLS, GPLS/GPLC, HL/CLS, HL/CW and HL/PLS among the seven clades for the sex groups retrieved from LSD Post-hoc analyses in factorial Two-way ANOVA.

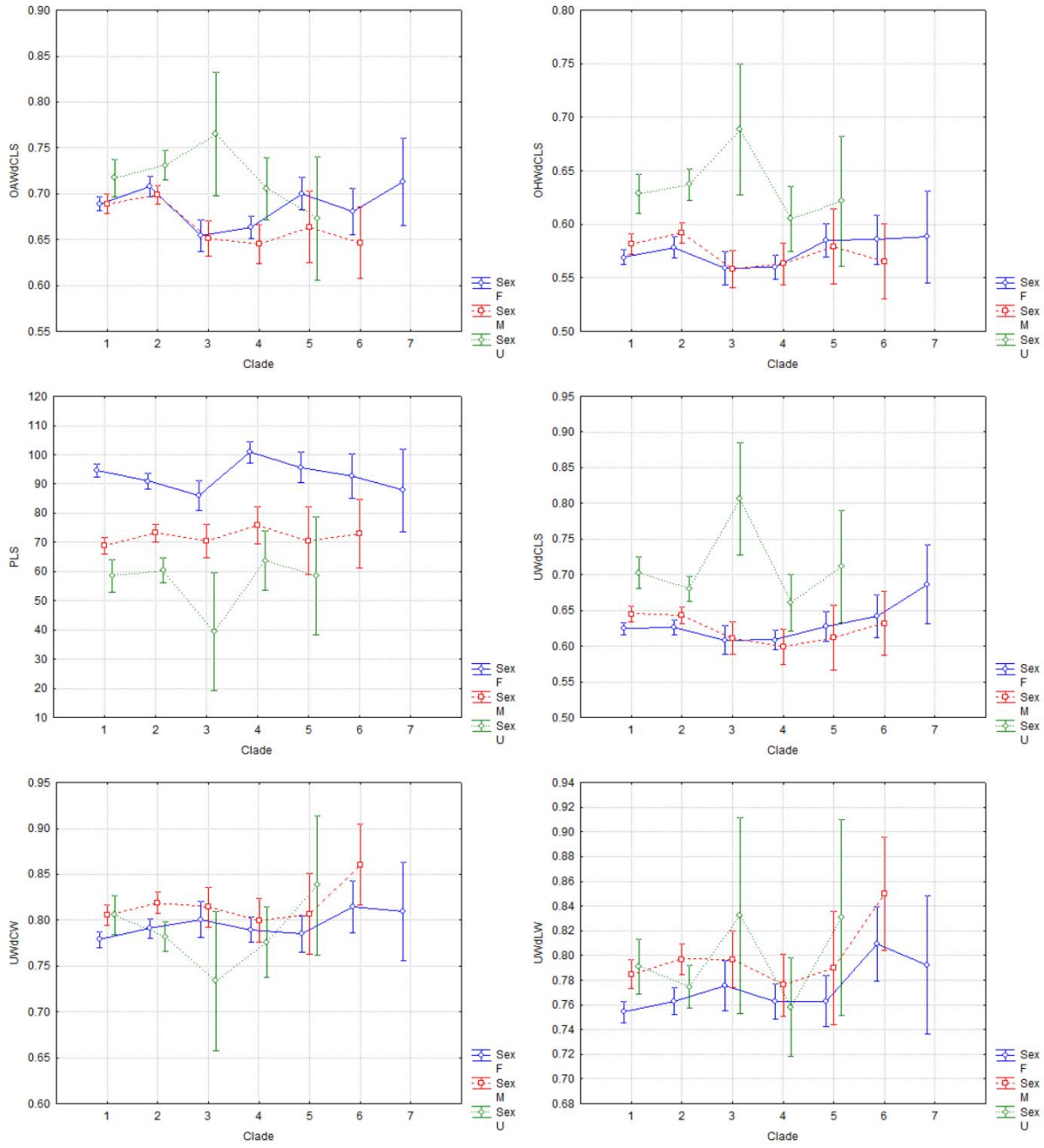


Figure S6.28. The absolute value of PLS used to evaluate body size variation across the seven clades and sexual size dimorphism; The ratios: OAW/CLS, OHW/CLS, UW/CLS, UW/CW and UW/LW among the seven clades for the sex groups retrieved from LSD Post-hoc analyses in factorial Two-way ANOVA.

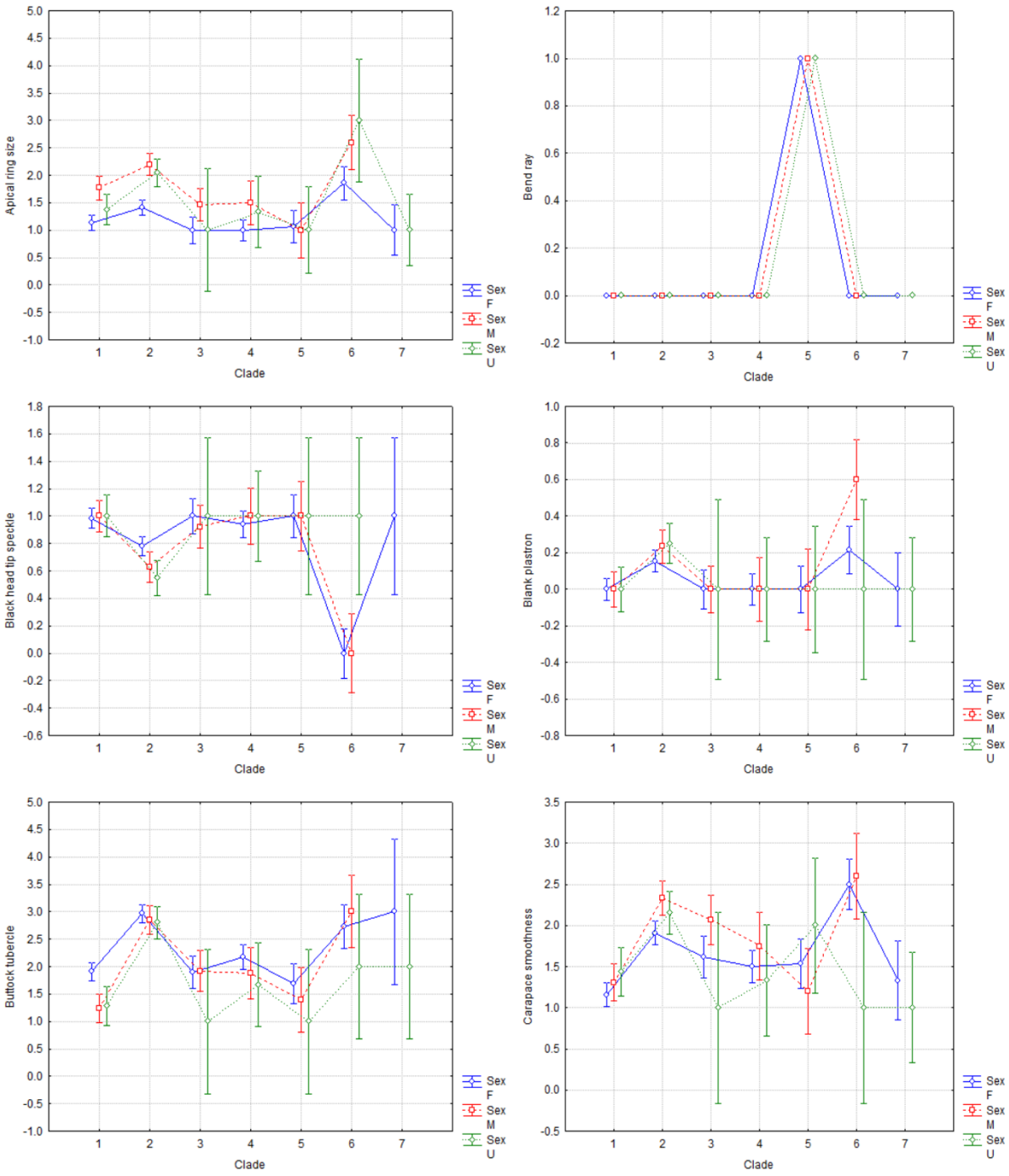


Figure S6.29. The characters: Apical ring size, Bent stripe, Black head tip speckle, Blank plastron, Buttock tubercle and Carapace smoothness level among the seven clades for the sex groups retrieved from LSD Post-hoc analyses in factorial Two-way ANOVA of the morphological character dataset.

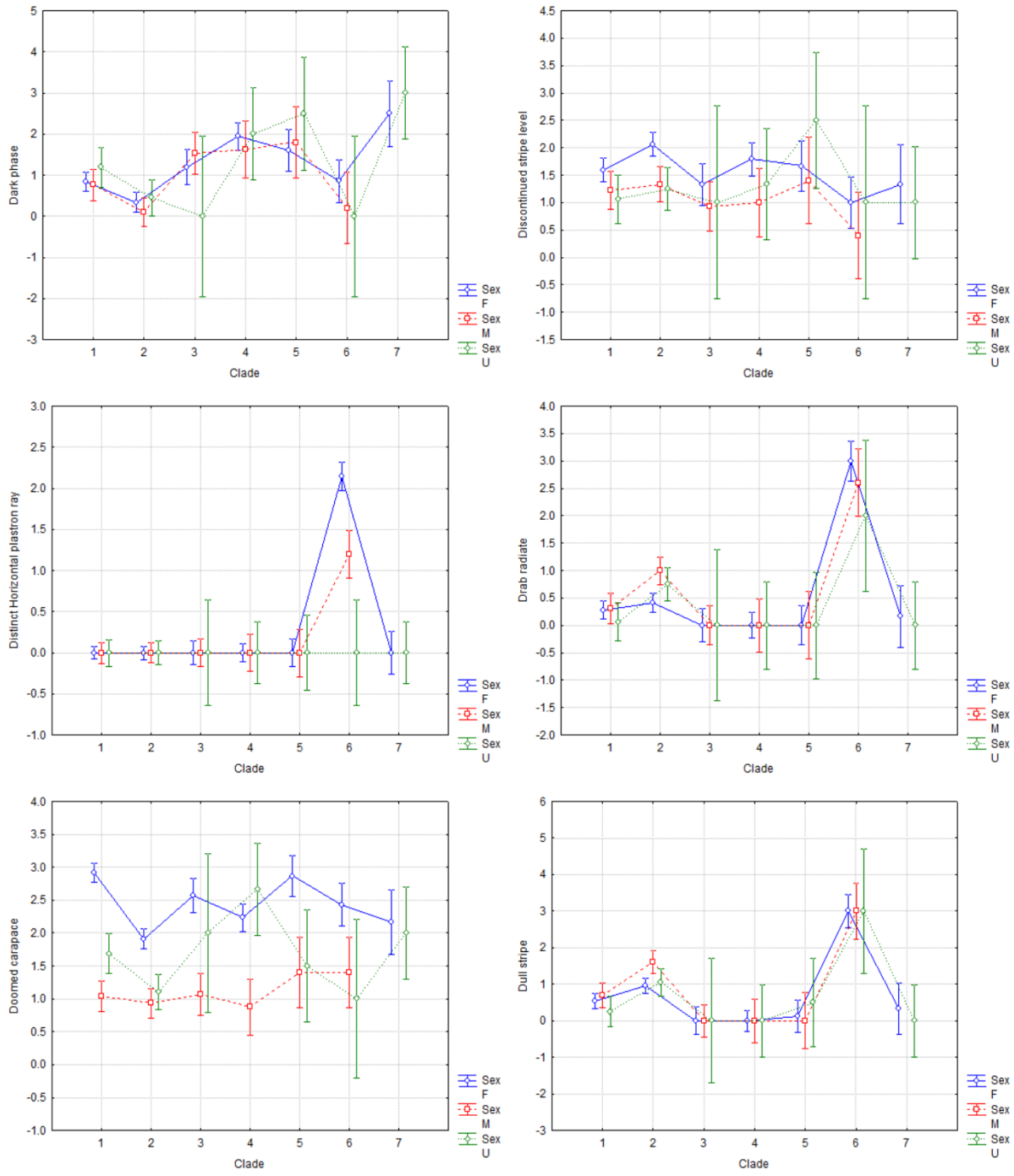


Figure S6.30. The characters: Dark phase, Discontinued stripe level, Distinct Horizontal plastron stripe, Drab radiate, Doomed carapace and Dull stripe among the seven clades for the sex groups retrieved from LSD Post-hoc analyses in factorial Two-way ANOVA of the morphological character dataset.

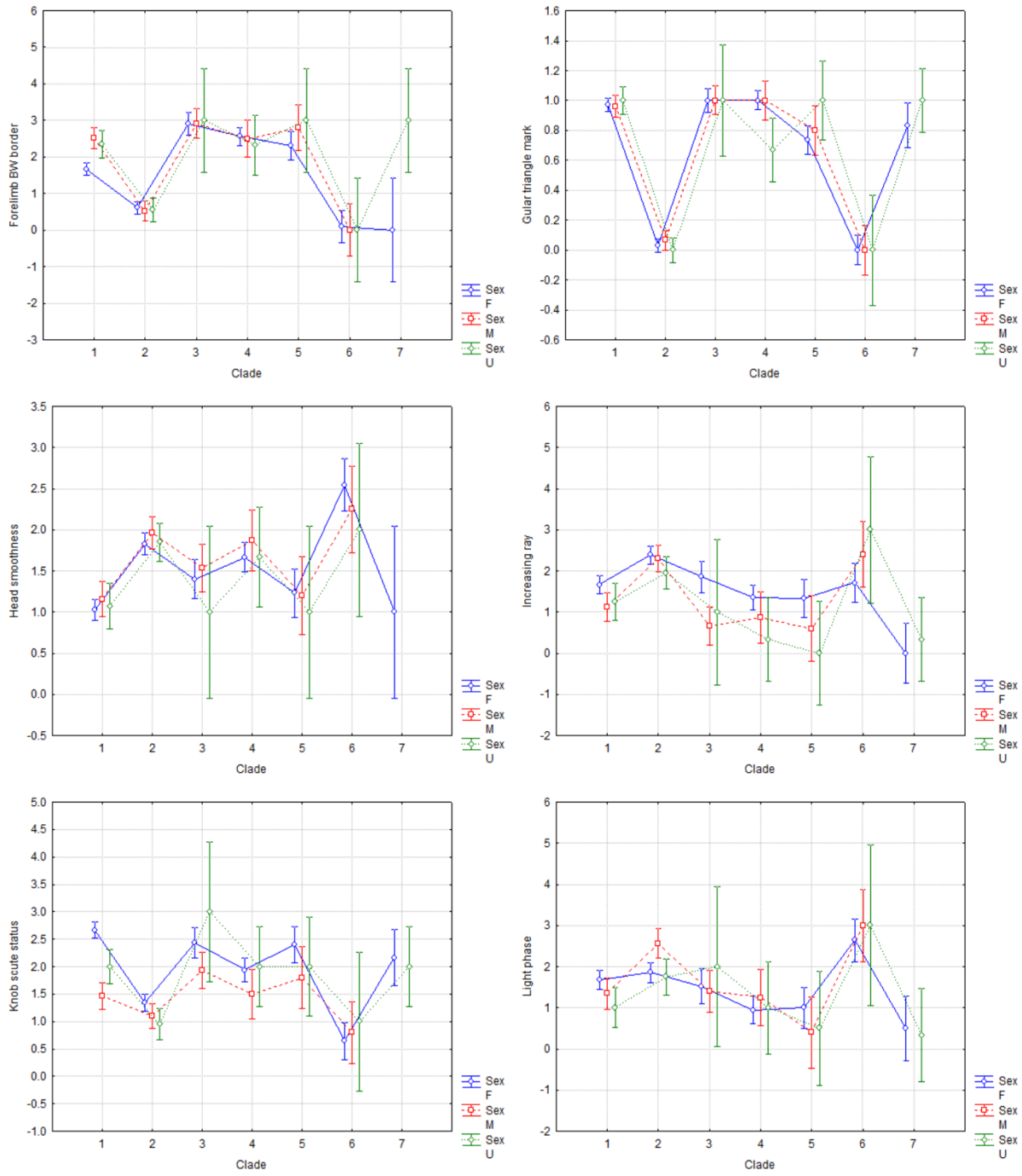


Figure S6.31. The characters: forelimb BW border, gular triangle mark, head smoothness, increasing stripe, knob scute status and light phase among the seven clades for the sex groups retrieved from LSD Post-hoc analyses in factorial Two-way ANOVA of the morphological character dataset.

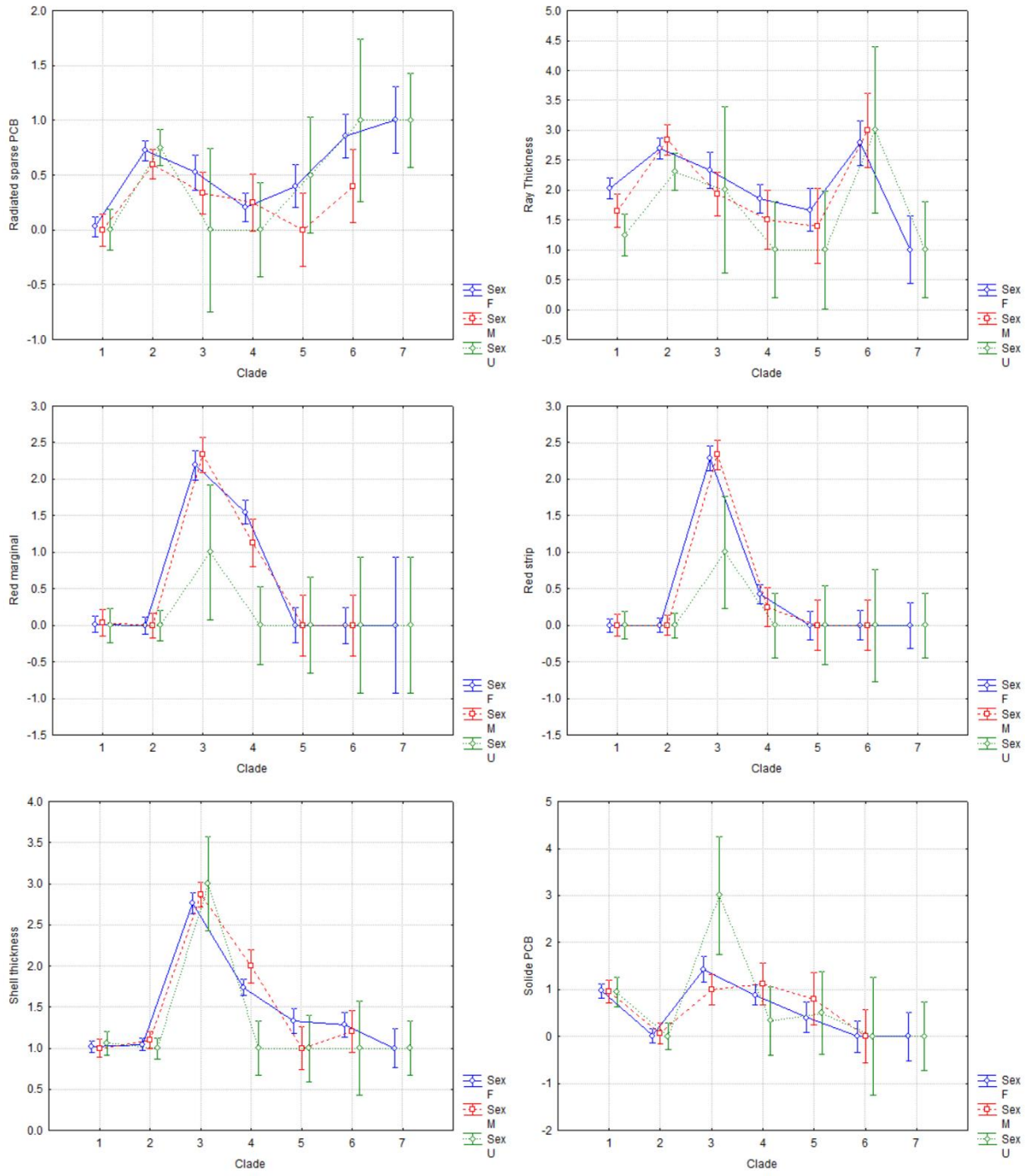


Figure S6.32. The characters: radiated sparse PCB, stripe thickness, red marginal, red stripe, shell thickness and solid PCB among the seven clades for the sex groups retrieved from LSD Post-hoc analyses in factorial Two-way ANOVA of the morphological character dataset.

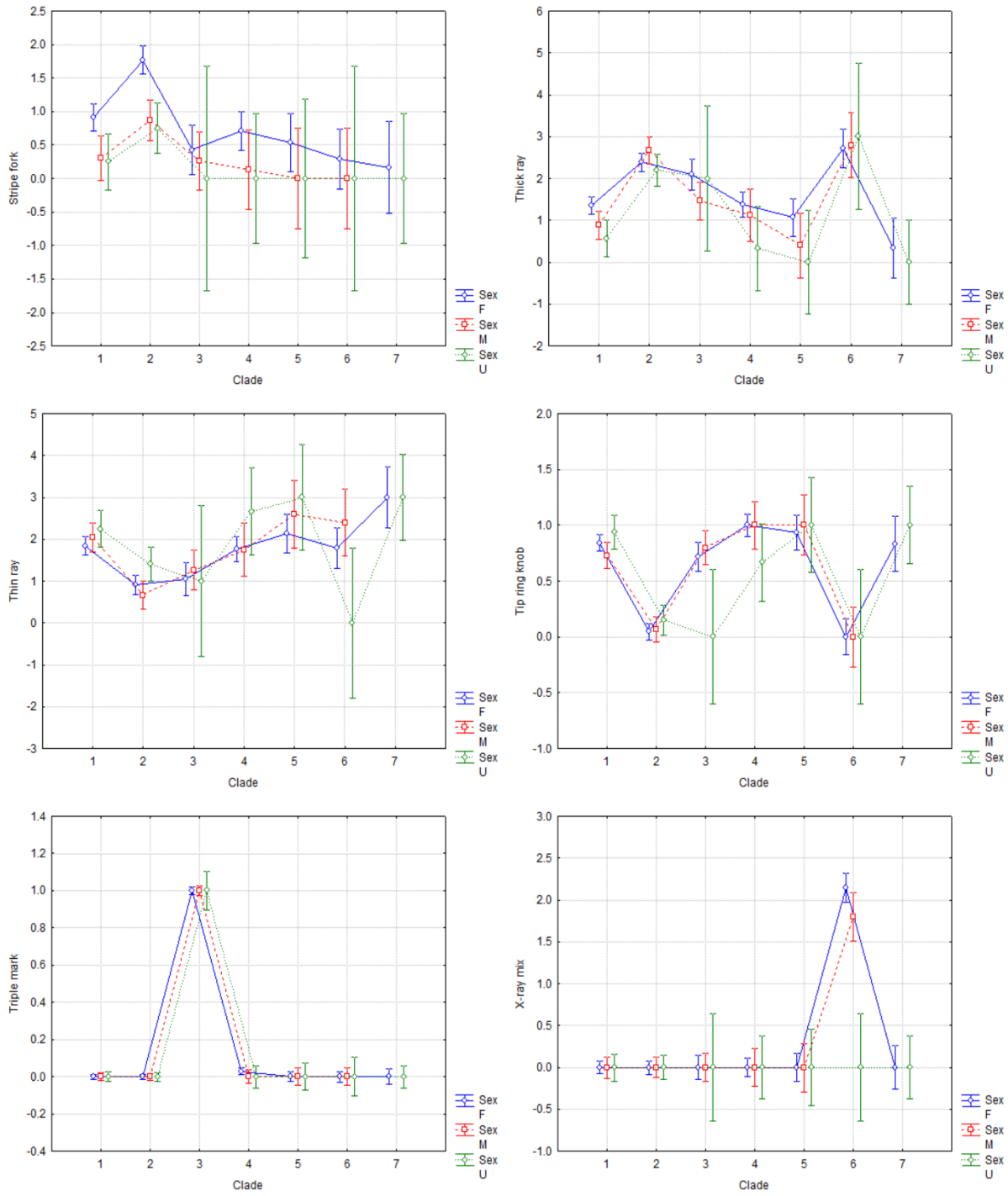


Figure S6.33. The characters: stripe fork, thick stripe, thin stripe, tip ring knob, triple mark and X-stripe mix among the seven clades for the sex groups retrieved from LSD Post-hoc analyses in factorial Two-way ANOVA of the morphological character dataset.

Chapter 7. Concluding Chapter

In Chapter 2, seven operational taxonomic units were identified within the *Psammobates tentorius* species complex, rather than the currently advocated "three-subspecies assumption", though the conservative nDNA marker generated incongruent results. It was found that *P. t. verroxii* was not a monophyletic group and should therefore undergo a taxonomic revision. The genetic diversity analyses revealed that C1, C2 and C6 had significantly higher genetic diversity than C3, C4, C5 and C7. Further cladogenic radiation is therefore expected in C1 and C6, but particularly in C6, since the within-clade genetic diversity level found in it was the highest. The establishment of protected areas at intergradation zones between C1 and C2, C2 and C4, and in the distribution range of C6, may facilitate conserving the adaptive radiation potential of the *P. tentorius* species complex.

The uniformly brown "*Psammobates bergeri*" should not be considered as a valid taxon, since that phenotype was found across three clades. The southwest Nama Karoo was previously also erroneously regarded as a "three-way" contact zone because of the high morphological variation found. This study, however, identified only one clade for the area from both the mtDNA and nDNA datasets, with high within clade genetic diversity.

The results of Chapter 3 revealed the divergences in the *P. tentorius* species complex are deep, with most clades originating in the Miocene. Among southern African reptiles, Miocene divergences usually correspond to cladogenesis within genera, e.g., *Pachydactylus* and *Platysaurus*, whereas radiation within species is linked to the Pliocene and Pleistocene, e.g., *Agama atra* and *Bitis arietans*. Some clades of the *P. tentorius* species complex may therefore justify elevation to full species status.

The study provided strong evidence that cladogenesis in the *P. tentorius* species complex can be linked to climatic fluctuations and topographical changes in southern Africa since the Miocene, thus supporting the first hypothesis (the climate fluctuations and associated habitat shifts since Neogene influenced the cladogenic pattern in *P. tentorius* species complex, and the physical barriers leading to isolation and allopatric diversification). It appears as if climate was of greater importance than purported topographical changes in earlier diversification events, but that uplift events together with climate change played a significant role in later divergences.

The climatic and topographic changes linked here to early divergences in the *P. tentorius* species complex were also proposed for vicariance in other reptiles at the generic level.

The results also correspond to other studies showing high genetic diversity (species richness) in the GCFR, not only for plants but also for animals, including reptiles. Consequently, diversification patterns of the *P. tentorius* species complex in the late Miocene and Pliocene parallels those of other organisms, supporting the second hypothesis of higher diversity in the GCFR than elsewhere over the distribution range of *P. tentorius*. The strong association of *P. tentorius* clades with particular regions and vegetation types suggests that the clades evolved allopatrically and that contact in restricted areas is recent, following range expansions of some clades. However, although the clades abut, they do not necessarily overlap because vegetation in the regions regarded as possible intergradation zones forms a mosaic, which may still keep clades distinct. Nevertheless, more research is necessary to establish if the clades hybridize in these so-called intergradation zones.

The results in Chapter 3 align with the findings in Chapter 2, and provide strong support for *P. tentorius* to undergo a taxonomic revision.

In Chapter 4, the results show a good match between the mtDNA and microsatellite DNA datasets, except for the genetic difference between C1 and C4, which the microsatellite DNA dataset was not significant. The ABC simulation analyses of the combined microsatellite and mtDNA sequence dataset advocated an alternative systematic scenario, which better explained the radiation history of the *P. tentorius* species complex. The latter scenario was supported by all three datasets, mtDNA, nDNA and microsatellite DNA, and should thus be regarded as the most likely scenario. Under this scenario the most likely ancestral lineage is C6, from which C3 had branched off. The assumption of C6 as basal lineage is also supported by some models in the biogeographic analyses (Chapter 3). Moreover, the results of Chapter 4 also imply that phylogenetic assumptions based purely on a mtDNA dataset can sometimes be misleading, emphasizing the importance of using multiple types of markers which making phylogenetic inferences.

The findings also point to possible hybridization between C1 and C2, and possibly also between C2 and C4, although evidence of crossing-over between them was lacking in the microsatellite DNA dataset.

When comparing the results of Chapter 2 and 3, it is evident that microsatellite DNA can be phylogenetically informative and useful in species delimitation. Even the allele range was found to be informative in terms of detecting phylogenetic signal.

A significant signal of inbreeding was detected in the eastern populations of C1, as genetic diversity in those populations were low for both the mtDNA sequence and microsatellite DNA datasets. This suggests the possible occurrence of a bottle-neck event that may have influenced the population dynamics of eastern population of C1 (population occur near Fish River valley in Eastern Cape Province).

In the selection analysis chapter (Chapter 5), additional evidence was provided that support the validity of the seven distinct evolutionary lineages. The amino acid based phylogeny supported all seven clades, despite the relationships among them being incongruent with the phylogeny retrieved from the DNA datasets. This incongruence may possibly be an artefact due to the limited informative sites in the amino acid sequences, rather than a real conflict between the DNA and amino acid datasets. Future studies using genome scale markers should assist in clarifyning the full picture.

When turning to morphology, Chapter 6, the discrete and continuous morphometric characters generally revealed the major difference between the two major groups “C1+C4+C7+C5” and “C2+C6” was clear, C3 clustered morphometrically with “C1+C4+C7+C5”, rather than with its genetically close relative group “C2+C6”, which was unexpected. Notwithstanding this, an interesting finding was the high congruence between nDNA phylogeny and morphological character analysis (both revealed that C3 was closer to “C1+C4+C7+C5”, rather than “C2+C6).

Nonetheless, the multivariate analyses (both discrete and continuous characters datasets) results seemed promising in distinguishing the seven clades, except the differences between C1, C5 and C7 were not significant from some analyses, though in DFA analysis (at both sexes), it advocated all seven clades should be considered as distinctive. I detected a significant sexual dimorphism (both sexual character dimorphism and sexual size dimorphism) in the *P. tentorius* complex. Nevertheless, it is strongly recommended that the interpretation on results at C5, C6 and C7 should be cautious to avoid premature conclusion, since the relatively small sample size (due to the scarcity in museum collection and field sampling, I was unable to gain

more samples for morphological analyses), therefore, further study is needed to variefy the full picture.

When comparing the morphological, mtDNA sequence and microsatellite DNA results, it is recommended that at least four species be recognized in the *P. tentorius* species complex, namely, 1) C1+C4+C7+C5, 2) C2, 3) C3, and 4) C6.

Future studies of this complex should ideally be based on genome wide DNA data in order to unmask its full evolutionary history, and also to investigate the relationship between the genome and the extraordinary colour pattern variations by looking at gene expression and epigenetics.

Publication and conference contributions

Publications

Chapter 2:

Zhao, Z., Heideman, N., Grobler, P., Jordaan, A., Bester, P., & Hofmeyr, M. D. (2020). Unraveling the diversification and systematic puzzle of the highly polymorphic *Psammobates tentorius* (Bell, 1828) complex (Reptilia: Testudinidae) through phylogenetic analyses and species delimitation approaches. *Journal of Zoological Systematics and Evolutionary Research*, *58*(1), 308-326.

Chapters 3 and **4**:

Zhao, Z., Heideman, N., Bester, P., Jordaan, A., & Hofmeyr, M. D. (2020). Climatic and topographic changes since the Miocene influenced the diversification and biogeography of the tent tortoise (*Psammobates tentorius*) species complex in Southern Africa. *BMC Evolutionary Biology*, *20*(1), 1-33.

Chapter 5:

Zhao, Z., Heideman, N., & Hofmeyr, M. D. (2021). Codon-based analysis of selection pressure and genetic structure in the *Psammobates tentorius* (Bell, 1828) species complex, and phylogeny inferred from both codons and amino acid sequences. *African Journal of Ecology*, *59*(2), 497-509.

Chapter 6:

Zhao, Z., Oosthuizen, J., & Heideman, N. (Accepted and in press). How many species does the *Psammobates tentorius* (tent tortoise) species complex (Reptilia, Testudinidae) comprise? A taxonomic solution potentially applicable to species complexes. *Journal of Zoological Systematics and Evolutionary Research*

Conference presentation

Zhao, Z., Heideman, N., Grobler, P., Jordaan, A., Bester, P. and Hofmeyr. M. 2019. *Unravelling the diversification and systematic puzzle of the highly polymorphic Psammobates tentorius complex.* Paper presented at the 14th Conference of the Herpetological Association of Africa, Cape St. Frances Resort, South Africa. 9 -13 September.

Chapter Two

Received: 12 April 2019 | Revised: 29 July 2019 | Accepted: 6 August 2019

DOI: 10.1111/jzs.12338

ORIGINAL ARTICLE

JOURNAL OF ZOOLOGICAL SYSTEMATICS AND EVOLUTIONARY RESEARCH

WILEY

Unraveling the diversification and systematic puzzle of the highly polymorphic *Psammobates tentorius* (Bell, 1828) complex (Reptilia: Testudinidae) through phylogenetic analyses and species delimitation approaches

Zhongning Zhao[1] | Neil Heideman[1] | Paul Grobler[2] | Adriaan Jordaan[1] | Phillip Bester[3] | Margaretha D. Hofmeyr[4]

[1]Department of Zoology and Entomology, University of the Free State, Bloemfontein, South Africa

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

[3]Department of Virology, University of the Free State and National Health Laboratory Service (NHLS), Bloemfontein, South Africa

[4]Chelonian Biodiversity and Conservation, Department of Biodiversity and Conservation Biology, University of the Western Cape, Bellville, South Africa

Correspondence
Zhongning Zhao, Department of Zoology and Entomology, University of the Free State, Biology building B19, 205 Nelson Mandela Dr, Park West, Bloemfontein, South Africa.
Email: orochi19851020@yahoo.com

Funding information
University of the Western Cape; National Research Foundation, Grant/Award Number: IFR150216114248; Universiteit van die Vrystaat, Grant/Award Number: A1999/158110

Abstract

The high level of phenotypic diversity in southern African tent tortoises (*Psammobates tentorius* complex) has for decades prevented systematists from developing a stable taxonomy for the group. Here, we used a comprehensive DNA sequence dataset (mtDNA: *Cytb*, *ND4*, *ND4* adjacent *tRNA-His*, and *tRNA-Ser*, *12S*, *16S*; and nDNA: *PRLR* gene) of 455 specimens, and the latest phylogenetic and species delimitation analytical procedures, to unravel the long-standing *P. tentorius* complex systematic puzzle. Our results for mtDNA and nDNA were incongruent, with the poorly supported nDNA phylogeny differentiating the three recognized subspecies, and showing potential hybridization in some regions. In contrast, the concatenated mtDNA phylogeny identified seven operational taxonomic units, with strong support. Clades 1, 4, 5, and 7 corresponded to tortoises identified as *P. t. tentorius*, clade 3 to *P. t. trimeni*, and clades 2 and 6 to *P. t. verroxii*. Our analyses showed conflicting topologies for the placement of C6 (*P. t. verroxii* north of the Orange River), with stronger support for it being sister to C2 + C3 than to the other clades. Clades 1, 2, and 6 had significantly higher genetic diversity than clades 3, 4, 5, and 7, perhaps because these clades inhabit substantially larger areas. The potential for future cladogenic radiations seems high in C1 and C6, particularly in C6 for which the within-clade diversification level was highest. Further research involving microsatellite DNA, phylogeographic evaluations, and morphological variation among clades is crucial for understanding the adaptive radiation of the *P. tentorius* complex and for modifying their taxonomy.

KEYWORDS
mtDNA, phylogeny, reptile, southern Africa, tent tortoise

Contributing authors: Neil Heideman (heidemannj@ufs.ac.za), Paul Grobler (GroblerJP@ufs.ac.za), Adriaan Jordaan (JordaanA2@ufs.ac.za), Phillip Bester (BesterPA@ufs.ac.za) and Margaretha D. Hofmeyr (mdhofmeyr@gmail.com)

Chapters Three and Four

Zhao et al. BMC Evol Biol (2020) 20:153
https://doi.org/10.1186/s12862-020-01717-1

BMC Evolutionary Biology

RESEARCH ARTICLE Open Access

Climatic and topographic changes since the Miocene influenced the diversification and biogeography of the tent tortoise (*Psammobates tentorius*) species complex in Southern Africa

Zhongning Zhao[1*], Neil Heideman[1], Phillip Bester[2], Adriaan Jordaan[1] and Margaretha D. Hofmeyr[3]

Abstract

Background: Climatic and topographic changes function as key drivers in shaping genetic structure and cladogenic radiation in many organisms. Southern Africa has an exceptionally diverse tortoise fauna, harbouring one-third of the world's tortoise genera. The distribution of *Psammobates tentorius* (Kuhl, 1820) covers two of the 25 biodiversity hotspots in the world, the Succulent Karoo and Cape Floristic Region. The highly diverged *P. tentorius* represents an excellent model species for exploring biogeographic and radiation patterns of reptiles in Southern Africa.

Results: We investigated genetic structure and radiation patterns against temporal and spatial dimensions since the Miocene in the *Psammobates tentorius* species complex, using multiple types of DNA markers and niche modelling analyses. Cladogenesis in *P. tentorius* started in the late Miocene (11.63–5.33 Ma) when populations dispersed from north to south to form two geographically isolated groups. The northern group diverged into a clade north of the Orange River (OR), followed by the splitting of the group south of the OR into a western and an interior clade. The latter divergence corresponded to the intensification of the cold Benguela current, which caused western aridification and rainfall seasonality. In the south, tectonic uplift and subsequent exhumation, together with climatic fluctuations seemed responsible for radiations among the four southern clades since the late Miocene. We found that each clade occurred in a habitat shaped by different climatic parameters, and that the niches differed substantially among the clades of the northern group but were similar among clades of the southern group.

Conclusion: Climatic shifts, and biome and geographic changes were possibly the three major driving forces shaping cladogenesis and genetic structure in Southern African tortoise species. Our results revealed that the cladogenesis of the *P. tentorius* species complex was probably shaped by environmental cooling, biome shifts and topographic uplift in Southern Africa since the late Miocene. The Last Glacial Maximum (LGM) may have impacted the distribution of *P. tentorius* substantially. We found the taxonomic diversify of the *P. tentorius* species complex to be highest in the Greater Cape Floristic Region. All seven clades discovered warrant conservation attention, particularly Ptt-B–Ptr, Ptt-A and Pv-A.

*Correspondence: orochi19851020@yahoo.com
[1] Department of Zoology and Entomology, University of the Free State, Biology Building B19, 205 Nelson Mandela Dr, Park West, Bloemfontein, South Africa
Full list of author information is available at the end of the article

Chapter Five

Received: 11 April 2020 | Revised: 4 October 2020 | Accepted: 21 November 2020

DOI: 10.1111/aje.12840

ARTICLE

African Journal of Ecology WILEY

Codon-based analysis of selection pressure and genetic structure in the *Psammobates tentorius* (Bell, 1828) species complex, and phylogeny inferred from both codons and amino acid sequences

Zhongning Zhao[1] | Neil Heideman[1] | Margaretha D. Hofmeyr[2]

[1]Department of Zoology and Entomology, University of the Free State, Bloemfontein, South Africa

[2]Chelonian Biodiversity and Conservation, Department of Biodiversity and Conservation Biology, University of the Western Cape, Bellville, South Africa

Correspondence
Zhongning Zhao, Department of Zoology and Entomology, University of the Free State, Bloemfontein, South Africa.
Email: orochi19851020@yahoo.com

Funding information
Universiteit van die Vrystaat, Grant/Award Number: A1999/158110; National Research Foundation, Grant/Award Number: IFR150216114248; University of the Western Cape, Grant/Award Number: SNS Grants

Abstract

This study used codon analysis (dN/dS and Tv/Ti) to investigate selection pressure and genetic structure in the highly polymorphic ***Psammobates tentorius*** species complex, and amino acid sequences to construct a phylogeny tree for it. Our results revealed a strong selection signal at node 'C2 + C3', possibly driven by aridity intensification resulting from the development of the Benguela Current. A similar signal was noticed at C3, possibly due to the same driving force. These findings suggest that environmental selection pressure favoured those groups and that further cladogenic events were possible. Selection pressure was also found to be high at C1, C4 and C7, which may indicate that they are also favoured by the current selection pressure. The codon-based phylogeny did not retrieve any potentially undescribed species, but nonetheless provided support for the validity of the seven distinct clades retrieved with the DNA sequence data. The amino acid sequence-based phylogeny generally supported the seven lineages as valid putative species. Investigation at the genomic scale could, however, help to solve the issue. In general, we found the codon, dN, dS, Tv, Ti and amino acid sequence-based phylogenetic inferences useful in species delimitation and recommend their use in species delimitation studies.

KEYWORDS

amino acid sequences, nonsynonymous and synonymous substitution, southern Africa, transition, transversion

Résumé

Cette étude s'est appuyée sur une analyse des codons (dN/ dS et Tv/ Ti) afin d'étudier la pression sélective et la structure génétique au sein du complexe d'espèces *Psammobates tentorius* hautement polymorphe, ainsi que sur les séquences d'acides aminés pour élaborer un arbre phylogénique à ces fins. Nos résultats ont révélé un important signal de sélection au nœud « C2 + C3 », probablement induit par l'intensification de l'aridité résultant du développement du courant de Benguela. Un signal similaire a été remarqué au C3, probablement en raison de la même force motrice. Ces résultats suggèrent que la pression sélective environnementale a favorisé

Chapter Six

Received: 15 March 2021 | Revised: 14 July 2021 | Accepted: 18 July 2021

DOI: 10.1111/jzs.12525

ORIGINAL ARTICLE

JOURNAL OF ZOOLOGICAL SYSTEMATICS AND EVOLUTIONARY RESEARCH

WILEY

How many species does the *Psammobates tentorius* (tent tortoise) species complex (Reptilia, Testudinidae) comprise? A taxonomic solution potentially applicable to species complexes

Zhongning Zhao[1] | Jaco Oosthuizen[2] | Neil Heideman[1]

[1]Department of Zoology and Entomology, University of the Free State, Bloemfontein, South Africa

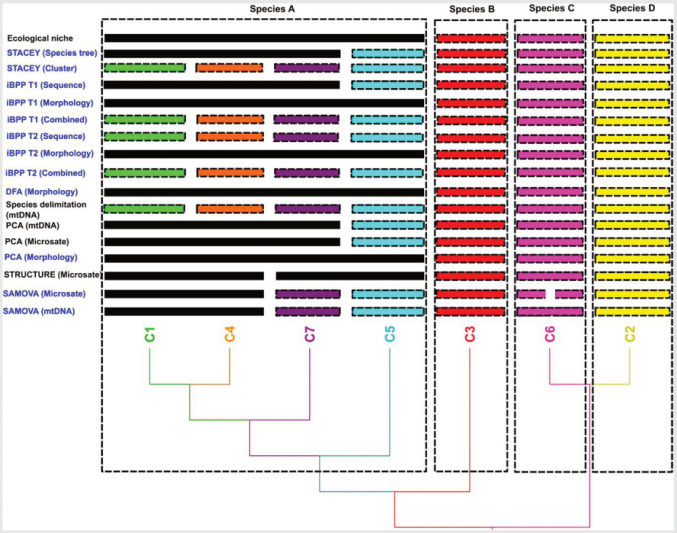
[2]School of Pathology, University of the Free, Bloemfontein, South Africa

Correspondence
Zhongning Zhao, Department of Zoology and Entomology, University of the Free State, Biology building B19, 205 Nelson Mandela Dr, Park West, Bloemfontein, South Africa.
Email: orochi19851020@yahoo.com

Funding information
National Research Foundation, Grant/Award Number: IFR150216114248; Universiteit van die Vrystaat, Grant/Award Number: A1999/158110

Graphical Abstract

The contents of this page will be used as part of the graphical abstract of html only. It will not be published as part of main article.



After comparing species delineation, results obtained from multiple evidences, ecological niche, multi-locus DNA sequences, multiple microsatellite DNA markers, morphology. The four species scheme appears as the best taxonomic solution for the *Psammobates tentorius* species complex. The alternative species tree topology suggested by ABC simulation analysis is better and more inclusive in explaining the genealogical history in the species complex.

Contributing authors: Jaco Oosthuizen (OosthuizenJ@ufs.ac.za) and Neil Heideman (heidemannj@ufs.ac.za)

