LINKING GEOGRAPHIC ORIGIN AND MITOGENOMES OF THE GROUND PANGOLIN IN SOUTHERN AFRICA

Zelda du Toit

2021

LINKING GEOGRAPHIC ORIGIN AND MITOGENOMES OF THE GROUND PANGOLIN IN SOUTHERN AFRICA

Zelda du Toit

Thesis submitted in fulfilment of the requirements for the degree of

Philosophiae Doctor

In the Faculty of Natural and Agricultural Science,
Department of Genetics,
University of the Free State

Supervisor:

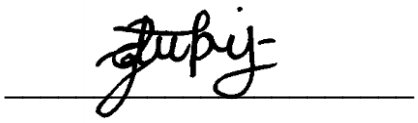
Prof. J.P. Grobler (UFS)

Co-Supervisors:
Prof. A. Kotzé (SANBI/UFS)
Prof. D.L. Dalton (SANBI)
Prof. R Jansen (TUT)

30 November 2021

DECLARATION AND COPYRIGHT

I declare that the thesis hereby submitted by me for the *Philosophiae Doctor* degree at the University of the Free State is my own independent work and has not previously been submitted by me at another University/Facility. I furthermore concede copyright of the thesis in favour of the University of the Free State.



Zelda du Toit

30 November 2021

ACKNOWLEDGEMENTS

Many people and institutions supported this study and allowed me the opportunity to complete the study and make it successful. I would like to express my sincere gratitude too:

Prof. Paul Grobler, Prof. Antoinette Kotzé and Prof. Raymond Jansen, for supporting this study and allowing me the opportunity to learn more in the conservation field. Your expert guidance and advice made me grow as a scientist even with full schedules.

A special expression of gratitude goes out to Prof. Desiré Dalton, Dr. Morné du Plessis and Dr. Monica Mwale. Without your support and guidance this project would not have been a success. Your guidance allowed me to grow and develop as a scientist and awarded me countless opportunities in order to improve in this field. Thank you for always being available and giving advice when I needed motivation as well as for always making time to assist where possible.

The African Pangolin Working Group (APWG) and Darren Pietersen for allowing me to use their data and samples as well as for sharing their knowledge and passion for these wonderful animals.

Agricultural Research Council, Biotechnology Platform (ARC/BTP) and especially Jonathan Featherston for assisting in some of the analysis conducted.

The South African National Biodiversity Institute, National Zoological Garden (SANBI, NZG) and National Research Foundation (NRF) for supporting the project financially under grant no. 86125 – Professional Development Program (PDP) Ph.D.

ACKNOWLEDGEMENTS

The staff of the Research and Scientific Services Department at the South African National Biodiversity Institute, National Zoological Garden (SANBI, NZG) as well as the Department of Genetics at the University of the Free State for your continued support.

A special thank you to my mother Marlene Burbidge, sister Senicia du Toit-Breytenbach and husband Francois Boshoff for all their love and support and allowing me the opportunity to pursue my dream. Without them I would not be where I am today. Thank you for always believing in me and giving me the strength to continue with my passion.

Lastly, to my Heavenly Father for lending me the ability, support and strength to continue and complete this project.

All samples were obtained by qualified individuals or veterinarians to ensure as little harm as possible were done to the animals. No individual pangolins were harmed or adversely affected as a direct result of this study. Ethical approval was obtained from the Animal Research Ethics Committee, University of the Free State, South Africa (UFS-AED2015/0070) and the NZG Research, Ethics and Scientific Committee (NZG/RES/P/001/F/02). Samples transport and possession were sanctioned under a Section 20 permit (12/11/1/1/18) from the Department of Agriculture, Forestry and Fisheries, South Africa with collection also approved under a CPC5 permit (02437) from the Department of Agriculture and Rural Development, South Africa and a Biodiversity permit (FAUNA 714/2012) from the Department of Environment and Nature Conservation, South Africa.

TABLE OF CONTENTS

CHAPTER THREE

MITOCHONDRIAL GENOMES OF AFRICAN PANGOLINS AND INSIGHTS INTO EVOLUTIONARY PATTERNS AND PHYLOGENY OF THE FAMILY MANIDAE

CHAPTER FOUR

PHYLOGEOGRAPHY OF TEMMINCK'S PANGOLIN (*SMUTSIA TEMMINCKII*) POPULATIONS IN SOUTHERN AFRICA USING A GEOREFERENCING APPROACH

LIST OF ABBREVIATIONS

A	Adenine
A	Number of alleles per locus
AIC	Akaike Information Criterion
AMOVA	Analysis of Molecular Variance
ANOVA	Analysis of Varience
APWG	African Pangolin Working Group
ATP	Adenosine Triphosphate
BI	Bayesian Inference
BIC	Bayesian Information Criterion
BLAST	Basic Local Alignment Search Tool
bp	Basepairs
bp	Basis pare
C	Cytocine
c.a.	estimated/approximatedly
CITES	Convention on International Trade in Endangered Species
CoI	Cytochrome *c* oxidase I
Cytb	Cytochrome *b*
DNA	Deoxyribonucleic Acid
DNS	Deoxyribonukleïensuur
DT	Decision Theory Performance-Based Selection
EDTA	Ethylene Diamine Tetra-acetic Acid
F_{CT}	Fixation Index (Among Groups)
Fsc	Among Populations Within Groups
F_{ST}	Fixation Index (Genetic Differentiation)
FGB7	Fibrinogen Beta Chain 7
G	Gamma Parameter
G	Guanine
He	Expected Heterozygosity
H_O	Observed Heterozygosity
H_T	Haplotype Diversity
Hz	Unbiased Heterozygosity

HW	Hardy-Weinberg
I	Invariant Sites
IUCN	International Union for the Conservation of Nature
ML	Maximum Likelihood
MtDNA	Mitochondrial DNA
NA	Not Applicable
NCBI	National Centre for Biotechnology Information
NDNA	Nuclear DNA
NRF	National Research Foundation
NZG	National Zoological Garden
PCA	Principle Component Analysis
PNG	Pangolin
RAD	Restriction Site Associated DNA Markers
SANBI	South African National Biodiversity Institute
SSC	Species Survival Commission
STR	Short Tandem Repeat
SSR	Simple Sequence Repeat
SNP	Single Nucleotide Polymorphism
S*ry*	Sex-determining Region
T	Thymine
TTN	Titin Protein/Antibody

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

GENERAL INTRODUCTION AND LITERATURE REVIEW

1.1. GENERAL INTRODUCTION

Pangolins are a unique order of Mammalia since they are the only mammal species which are covered in keratin scales across their body (Herklots 1937; Heath & Coulson 1998). These animals elicit a spherical defence posture to protect their head and extremities from predators (Grassé 1955; Kingdon 1971; Heath 1992a). Pangolins are myrmechophages and feed on a specialized diet of ant and termite species. This behaviour is considered to be an important ecological factor as some authors claim the pangolins maintain these invertebrates' population numbers (Heath 1992b; Chao 2002; Whitfort 2012; Pietersen *et al.* 2016a).

There are eight extant species of pangolin (order Pholidota), comprising one family, Manidae, and two sub-families (Gaudin *et al*. 2009). Four species are found in Asia, and fall under one genus, *Manis*, within the sub-family Maninae. The remaining four species are situated in Africa under the sub-family Smutsiinae. The African pangolins are further divided into two genera, based on their niche preferences. The genus *Smutsia* was assigned to the two ground-dwelling pangolin species known as Temminck's pangolin (*Smutsia temminckii*) and the giant pangolin (*Smutsia gigantea*). The remaining two arboreal species were assigned to the genus *Phataginus,* which includes the black-bellied pangolin (*Phataginus tetradactyla*) and white-bellied pangolin (*Phataginus tricuspis*) (Gaudin *et al*. 2009).

Pangolins are under numerous threats globally, ranging from the illegal wildlife trade and *Muthi* demand to electrocution on game fences (Bräutigam *et al*. 1994; Swart *et al*. 1999; Soewu & Adekanola 2011; Pietersen *et al*. 2014a, 2019a). It is estimated that more than 10 000 individuals are being illegally traded annually and in 2014 studies indicated that more than a 1 000 000 individuals have been traded within the last decade (Davies 2014; Hance 2014; Sutter 2014a,b; Friedman 2016). Due to the

increased trade of pangolin products, all eight species were up-listed in 2019 on the International Union for the Conservation of Nature (IUCN) Red List of Threatened Species. In 2016 they were up-listed to Appendix I of the Convention on International Trade in Endangered Species of Fauna and Flora (CITES) (CITES 2016; Heinrich *et al.* 2016; IUCN 2016; Pendharkar 2016; Challender *et al.* 2019b,c; Pietersen *et al.* 2019a,b; Schoppe *et al.* 2019; Ingram *et al.* 2019; Mahmood *et al.* 2019; Nixon *et al.* 2019).

Limited research has been performed on the complete order Pholidota with reference to ecology and population genetics. Of the eight species, the four African pangolins are the least studied, especially in Western and Central Africa. However, selected studies have been undertaken on Temminck's pangolin in the southern African distribution range with regards to ecology and conservation genetics (Jacobsen *et al.* 1991; Pietersen 2013; Du Toit 2014; Du Toit *et al.* 2014; Pietersen *et al.* 2014a, 2019a; Hassanin *et al.* 2015; Gaubert *et al.* 2016).

The current study entailed a series of molecular analyses of Temminck's pangolin (*Smutsia temminckii*) populations in southern Africa for potential geo-referencing. Markers employed included mitochondrial loci (mtDNA) and Short Tandem Repeats (STRs). The results were expected to provide more insight regarding population structure and genetic diversity. This study also included a phylogenetic assessment using full mitochondrial genomes of all four African pangolin species, with the main objective to classify African pangolins on a genus level. These data, along with future research and validation of these markers for forensic purposes, will contribute to enhanced detectability of confiscated samples originating from Africa. It will also aid in creating an improved framework for conservation and management of confiscated individuals with regards to husbandry in captive environments and reintroduction into nature (Crozier 1997).

1.2. OVERVIEW OF PANGOLIN TAXONOMY AND BIOLOGY

Pangolin scales consist of keratin and overlap with each other while having sharp edges protruding outwards (Tong *et al*. 1995, 2007; Meyer *et al*. 2013). Apart from rolling into a ball when threatened, they also move their tail from side to side, while being rolled up, which allows the sharp edges of the scales to injure the predator (Grassé 1955; Kingdon 1971; Heath 1992a).

These mammals are predominantly myrmecophagous and feed on selected ant and termite species which they obtain by breaking termite- and ant nests open using their large claws located on the forelimbs (Gebo & Rasmussen 1985; Skinner & Chimimba 2005; Davit-Béal *et al*. 2009). This selective feeding ecology contributes to the extreme difficulty in keeping pangolins in a captive environment since they do not adapt well to an artificial diet (Heath 1992a; Yang *et al*. 2007).

These animals are solitary, predominantly nocturnal with a select few showing diurnal behaviour (Pagés 1975; Richer *et al*. 1997; Pietersen 2013; Ingram *et al*. 2019). Pangolins have small toothless jaws with a long sticky tongue (equivalent to ⅔ to ¾ of the body length) protruding from the abdominal cavity to reach into ant and termite tunnels and used to access their food source (Doran & Allbrook 1973; Dickman & Richer 2001; Davit-Béal *et al*. 2009). They rely mostly on their well developed olfactory senses rather than their eyesight to locate food (Dickman & Richer 2001).

The female pangolin produces one pup per breeding season, with twins in exceptional cases, which are carried on the mother's dorsal side, either on the back or tail (Van Ee 1966, 1978; Jacobsen *et al*. 1991; Soewu & Adekanola 2011). Another notable observation regarding pangolins is their ability to swim; as they are reported to be relatively strong swimmers (Fang 1981; Skinner & Chimimba 2005; Yang *et al*. 2007).

1.2.1. ASIAN PANGOLIN SPECIES

The Asian pangolins are divided into four different species, namely: the Chinese pangolin (*Manis pentadactyla*; Linnaeus 1758), Malayan/Sunda pangolin (*M. javanica*; Desmarest 1822), Philippine pangolin (*M. culionensis*; De Elera 1895) and Indian pangolin (*M. crassicaudata*; Geoffroy 1803). A physical trait that disnguishes Asian pangolins from their African counterparts is the presence of hair bristles between the dorsal scales (Herklots 1937; Dickman & Richer 2001). They also have prominent external ear pinnae; in contrast to the scarce or indistinctive pinnae in the African species (Emry 1970; Kingdon & Hoffmann 2013).

1.2.1.1. CHINESE PANGOLIN (*M. PENTADACTYLA*)

The Chinese pangolin is defined by its helmeted appearance, small scales, large ear pinna as well as a narrowing tail towards the distal end (Figure 1.1*a*) (IUCN-SSC 2020a). They are distributed in the southern parts of China and surrounding countries including Nepal, Bhutan, India, Bangladesh, Myanmar, Lao People's Democratic Republic, Thailand, Vietnam, China, Hong Kong and Taiwan (Figure 1.2). The Chinese pangolin is known to occur at high altitudes such as the hills and mountains in the southern and western parts of it distribution range, however, they are also found at lower altitudes in forest habitats in Hong Kong and the north eastern parts of their distribution range (Challender *et al*. 2019c).

Although the Chinese pangolin has the ability to climb trees, they are predominately a ground-dwelling species with solitary, nocturnal behaviour (Heath & Vanderlip 1988). They are also reported to be occasionally crepuscular. These pangolins maintain large home ranges, containing multiple rotating burrows, and feed primarily on ants and termites by digging long burrows which serves as both access to food and shelter (Allen & Coolidge 1940; Heath & Vanderlip 1988). Their habitat includes primary and secondary tropical forests as well as limestone, bamboo, broad-leaf and coniferous forests. They are also reported in grasslands and agricultural fields (Gurung & Singh 1996). Chinese pangolins are known to produce one pup

(occasionally two) in the spring and also reported to be good swimmers (Allen & Coolidge 1940; Heath & Vanderlip 1988).

1.2.1.2. MALAYAN/SUNDA PANGOLIN (*M. JAVANICA*)

The Malayan pangolin is slightly larger than the Chinese pangolin with fewer rows of scales across its back and a longer tapered prehensile tail (Figure 1.1*b*). They also have shorter forelimb claws to aid in their arboreal lifestyle (IUCN-SSC 2020b). The Malayan pangolin has the largest distribution range in southeast East Asia (Figure 1.2) including Myanmar, Lao People's Democratic Republic, Thailand, Vietnam, Cambodia, Malaysia, Singapore, Brunei Darussalam and Indonesia (Schlitter 2005; Wu *et al*. 2005).

This pangolin species is mostly ground-dwelling, however, they also spend a significant amount of time in trees either foraging or looking for shelter. Malayan pangolins are known to find shelter either in or at the base of trees and occasionally burrow on the ground. They use their prehensile tail to aid in climbing trees to access ant nests, but also feed on termites (Challender *et al*. 2019b). The Malayan pangolin is a solitary nocturnal animal which are known to occupy primary and secondary forest habitats as well as lowland dipterocarp forests. They are also found in cultivated areas such as gardens, oil palm and rubber plantations as well as close to human settlements (Nowak 1999; Azhar *et al*. 2012). These pangolins species are believed to produce one pup after a gestation period of around 130 days.

1.2.1.3. PHILIPPINE PANGOLIN (*M. CULIONENSIS*)

The Philippine pangolin is known as the Asian pangolin with the largest amount of scale rows distributed on the dorsal side (Figure 1.1*c*). The scales are also smaller in size compared to the Malayan pangolin and these animals have a shorter head and torso to tail length ratio (IUCN-SSC 2020c). As suggested by their name, the Philippine pangolin is endemic to the Philippines (Figure 1.2), with sightings reported in Palawan and adjacent islands such as Busuanga-, Coron-, Culion- and Dumaran Island (Heaney *et al*. 1998; Widman *et al*. 2004).

They are found in lowland primary and secondary forests, grasslands and mixed mosaics such as agricultural- and scrublands (Heaney *et al*. 1998; Esselstyn *et al*. 2004). Very little information is available regarding the behaviour, ecology and biology of this species, but they are believed to be myrmecophageous, feeding on ants and termites similar to the other pangolin species (Schoppe *et al*. 2019). Although limited information is available regarding their breeding patterns, the available literature suggest they follow the same trend as the Malayan pangolin with one pup born after an estimated gestation period of 120 days (IUCN-SSC 2020c).

1.2.1.4. INDIAN PANGOLIN (*M. CRASSICAUDATA*)

The Indian pangolin is the largest of the four Asian pangolins with significantly larger scales distributed on the dorsal side of the body compared to the other three species (Figure 1.1*d*). They are fossorial with large forelimb claws used in digging of burrows, similar to the Chinese pangolin (IUCN-SSC 2020d). The Indian pangolin is predominantly distributed in South Asia (Figure 1.2) including Nepal, Pakistan, India and Sri Lanka (Schlitter 2005; Srinivasulu & Srinivasulu 2012).

They are solitary, nocturnal and predominantly ground-dwelling although they are reported to be arboreal in some habitats. They use their large forelimbs for burrowing as well as for climbing trees along with their prehensile tail (Prater 1971; Roberts 1977; Tikader 1983; Heath 1995). They are known to access a variety of habitats such as tropical forests, open land, grasslands and degraded habitat in close proximity to villages (Mahmood *et al*. 2019). The Indian pangolins are also reported to be well adapted in modified habitats as long as their primary food source, ants and termites, are available.

During the mating season, both sexes share the same burrows, which consist of shelter under large rocks or burrows with concealed entrances made from dirt (Prater 1971; Roberts 1977; Tikader 1983). The female pangolin gives birth to one pup (although twins have been reported) after a gestation period of 165 days (Mahmood *et al*. 2019).

1.2.2. AFRICAN PANGOLIN SPECIES

The African pangolin species are divided into two distinct groups, namely ground-dwelling and arboreal, with conspicuous dissimilarities.

The arboreal pangolins include the black-bellied pangolin (*P. tetradactyla*; Linnaeus 1766) and white-bellied pangolin (*P. tricuspis*; Rafinesque 1821), which are characterized by their small body size compared to the ground dwelling pangolins, and significantly longer and slender tail which can be twice their body length. They also have short forelimbs and claws along with a prehensile tail containing tail pads which aid in the climbing of trees. Furthermore, they have hair on the ventral side of their bodies, an irregular arrangement of scales on the dorsal side and large eyes.

The ground-dwelling species are the giant pangolin (*S. gigantea*; Illiger 1815) and Temminck's pangolin (*S. temminckii*; Smuts 1832), which are much larger in size compared to the previous group, with broader and slightly shorter tails. They also lack tail pads and are more adapted for a fossorial lifestyle, with larger forelimbs and claws for foraging. The ground-dwelling species also have an arranged scale pattern on the dorsal side of the body in comparison to the arboreal species (Kingdon & Hoffmann 2013; IUCN-SSC 2020e,f,g,h).

1.2.2.1. TEMMINCK'S PANGOLIN (*S. TEMMINCKII*)

Temminck's pangolin is reported to be the second largest of the extant pangolin species, after its counterpart the giant pangolin, with an average weight of 12 kg and with the largest individual recorded to date weighing 19 kg (Figure 1.1*e*) (APWG 2020a). They are also known to have the widest recorded distribution range in Africa. Their distribution ranges from Chad descending southwards through eastern Africa, into southern Africa (Figure 1.2). The range countries include Chad, Central African Republic, South Sudan, Uganda, Kenya, Rwanda, Tanzania, Zambia, Malawi, Mozambique, Zimbabwe, Botswana, Namibia and South Africa *(Pietersen et al.* 2019a)

Temminck's pangolin is a ground-dwelling fossorial species inhabiting abandoned burrows from other burrowing animals such as the aardvark (Herklots 1937; Heath & Coulson 1997b; Coggins 2003; Skinner & Chimimba 2005; Tong *et al.* 2007; Pietersen *et al.* 2014b).

They are solitary and predominantly nocturnal, with large distribution ranges containing multiple burrows used on a rotational basis (Jacobsen *et al.* 1991; Heath & Coulson 1997a, 1997b). Being the only bipedal species of pangolin, the front forelimbs are highly adapted to digging into hardened termataria and compact soil in search of prey (Steyn *et al.* 2018). Although they are predominantly nocturnal, reports indicate diurnal activity as well, especially in young animals and in adults during the winter months, when they are prone to be more active during daytime (Pietersen *et al.* 2016a).

They inhabit a variety of different habitats including arid, mesic and semi-arid environments such as savannah and woodland as well as floodplain grassland, rocky slopes and sandveld (Coulson 1989; Skinner & Chimimba 2005; Pietersen 2013). Temminck's pangolin are absent from Karroid veld, Highveld grassland and coastal regions as well as tropical and coastal forests. They are sometimes found on livestock and wildlife farms, but are absent from areas where crop farming occurs (Pietersen 2013). Temminck's pangolin is entirely myrmecophagous, with a selective eating habit based on specific ant and termite species. This discriminating practise contributes to the selection of suitable habitat based on the specific food source availability (Jacobsen *et al.* 1991; Swart 1996; Swart *et al.* 1999; Elsner *et al.* 2016; Pietersen *et al.* 2016b). Although they are mostly water independent, they have been observed to utilise free-standing water sources when these are available (Stuart 1980).

Temminck's pangolin reach maturity at two years of age and the males start breeding from the age of six to seven years. The females reach sexual maturity earlier at three to four years and have an estimated gestation period of 105–140 days. They usually produce one pup annually, but have been observed to only reproduce a pup every second year (Pietersen *et al.* 2016a).

1.2.2.2. GIANT PANGOLIN (*S. GIGANTEA*)

The giant pangolin is the largest of the African species, weighing up to 35 kg (Figure 1.1*f*). They are also the rarest species in Africa, with few reported sightings (APWG 2020b). They are mostly distributed in West and central Africa (Figure 1.2) including Senegal, Gambia, Guinea-Bissau, Guinea, Sierra Leone, Liberia, Côte d'Ivoire, Ghana, Cameroon, Central African Republic, Equatorial Guinea, Gabon, Republic of the Congo, Democratic Republic of the Congo, Uganda and Tanzania (Grubb *et al*. 1998; Kingdon & Hoffmann 2013; Nixon *et al*. 2019).

The giant pangolin is a ground-dwelling, solitary and nocturnal species occurring in lowland, tropical and swamp forests as well as forest savannah habitats. They are prone to hiding in burrows or underneath plant debris or uncovered tree roots and feed predominantly on ants and termites (Kingdon & Hoffmann 2013; APWG 2020b). The available literature suggests that they are highly dependent on water and only produce one pup annually (Pietersen 2013). Gathering information and performing population censuses is challenging as they have a scattered distribution range (Bräutigam *et al*. 1994). Therefore, the information regarding the ecology and behaviour of this species is sparse.

1.2.2.3. BLACK-BELLIED PANGOLIN (*P. TETRADACTYLA*)

The black-bellied pangolin is unique amongst the African pangolin species as it has the shortest body of all extant species with the longest tail length, consisting of 46-47 vertebrae. This makes it the record holder for the mammal with the longest tail (Jentink 1882; Kingdon & Hoffmann 2013). They are dark in colour with yellowish and dark brown scales and is covered with dark hair on the ventral side (Figure 1.1*g*) (Rahm 1956; Kingdon & Hoffmann 2013). They are mostly found in western and central Africa (Figure 1.2) including Sierra Leone, Liberia, Côte d'Ivoire, Ghana, Nigeria, Cameroon, Central African Republic, Equatorial Guinea, Gabon, Republic of the Congo and Democratic Republic of the Congo (Ingram *et al*. 2019).

The black-bellied pangolin, unlike other pangolins, is predominantly diurnal with some level of nocturnal activity noted (Challender *et al*. 2019a). Foraging during the day may prevent feeding competition with the white-bellied pangolin (Pappin 2012; Kingdon & Hoffmann 2013; APWG 2020c). Although both black- and white-bellied pangolins are arboreal species, the black-bellied pangolins spends more time in trees compared to the white-bellied pangolins. They inhabit trees of tropical-, moist riverine- and swamp forests where they feed on ants, termites and other small invertebrates (Pappin 2012; APWG 2020c; IUCN-SSC 2020f). These animals occasionally descend from the trees to drink water. Although very little information is available on this species, black-bellied pangolin females are believed to produce one pup after a gestation period of around 140 days (Pagés 1970, 1972; Pappin 2012; APWG 2020c; IUCN-SSC 2020f).

1.2.2.4.WHITE-BELLIED PANGOLIN (*P. TRICUSPIS*)

Of all the African species, the white-bellied pangolin is the most frequently encountered species with high densities in Gabon and Nigeria (Figure 1.1*h*) (Pagés 1975; Angelici *et al*. 1999; Kingdon & Hoffmann 2013). They are found in most central, western and eastern African countries (Figure 1.2) such as Guinea-Bissau, Guinea, Sierra Leone, Liberia, Côte d'Ivoire, Ghana, Togo, Benin, Nigeria, Cameroon, Central African Republic, South Sudan, Equatorial Guinea, Gabon, Republic of the Congo, Democratic Republic of the Congo, Uganda, Kenya, Rwanda, Tanzania, Zambia and Angola (Kingdon & Hoffmann 2013; Pietersen *et al*. 2019b).

The white-bellied pangolin is primarily a nocturnal species, similar to both ground-dwelling species and in contrast to its counterpart, the black-bellied pangolin. Although they are classified as an arboreal species, previous reports suggested they also spend a significant amount of time on the ground (Pagés 1975; Pietersen *et al*. 2019b). They are found in moist tropical lowland forests and secondary growth habitats as well as dense woodlands along waterways (Kingdon 1971; Gaubert 2011; APWG 2020d). They are also sometimes found in primary and secondary rainforests, altered forests and farmlands (Angelici *et al*. 1999). However, according to (Akpona *et al.* 2008), the difference in habitats doesn't seem to affect population

densities, indicating some level of adaptability for different habitat modifications. This species is also myrmecophageous, feeding only on ants and termites. Their breeding pattern is thought to be continuous throughout the year, with one pup born after a gestation period of approximately 150 days (Kingdon & Hoffmann 2013).

1.2.3. THE CONSERVATION STATUS OF PANGOLINS

Following the significant escalation of illegal trade of pangolins over the last few years, all eight extant species have shown a decline in population numbers, with extinctions already recorded within each species' historical distribution range (Challender *et al*. 2019b,c; Ingram *et al*. 2019; Mahmood *et al*. 2019; Nixon *et al*. 2019; Pietersen *et al*. 2019a,b; Schoppe *et al*. 2019).

The first assessment for the IUCN Red List of Threatened Species was performed in 1996, when only seven pangolin species were recognized (Chinese, Malayan, Indian, Temminck's, giant, black-bellied and white-bellied). The three Asian species and Temminck's was listed as *Lower Risk*/*Near Threatened* (LR/NT) while the other three African species were listed as *Lower Risk*/*Least Concern* (LR/LC). The Philippine pangolin was only recognised as a separate species in 1998 (Table 1.1).

Table 1.1. IUCN Red List of Threatened Species indicating the different statuses assigned to the order Pholidota over several evaluations according to specific years.

Species	**IUCN Red List (1996)**	**IUCN Red List (2008)**	**IUCN Red List (2014)**	**IUCN Red List (2019)**
Chinese pangolin, *Manis pentadactyla*	*Near Threatened*	*Endangered*	*Critically Endangered*	NA
Malayan/Sunda pangolin, *M. javanica*	*Near Threatened*	*Endangered*	*Critically Endangered*	NA
Philippine pangolin, *M. culionensis*	NA	*Near Threatened*	*Endangered*	*Critically Endangered*
Indian pangolin, *M. crassicaudata*	*Near Threatened*	*Near Threatened*	*Endangered*	NA
Temminck's pangolin, *Smutsia temminckii*	*Near Threatened*	*Least Concern*	*Vulnerable*	NA
Giant pangolin, *S. gigantea*	*Least Concern*	*Near Threatened*	*Vulnerable*	*Endangered*
Black-bellied pangolin, *Phataginus tetradactyla*	*Least Concern*	*Near Threatened*	*Vulnerable*	NA
White-bellied pangolin, *P. tricuspis*	*Least Concern*	*Least Concern*	*Vulnerable*	*Endangered*

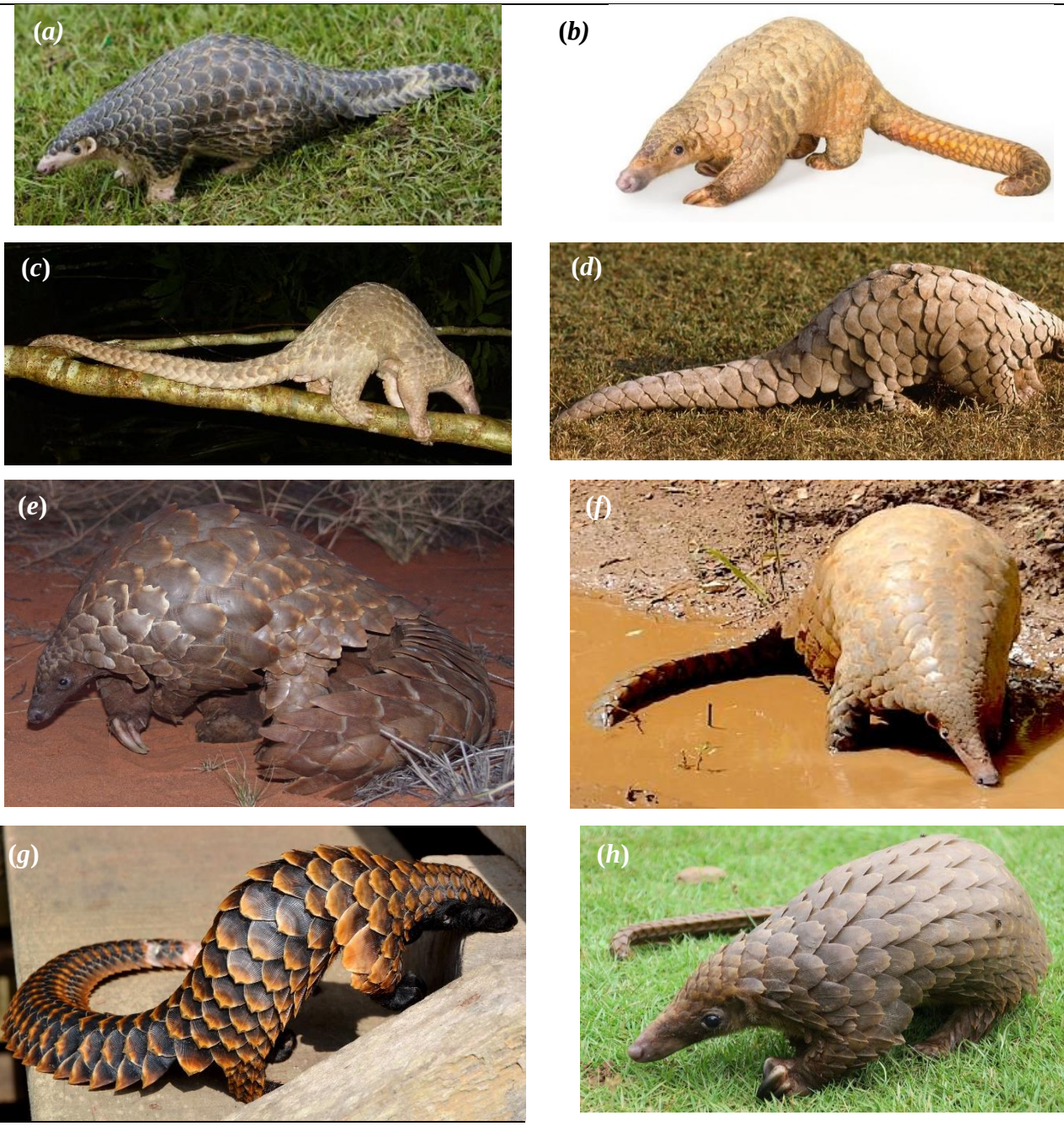


Figure 1.1. The eight extant pangolin species. (*a*) Chinese pangolin, *Manis pentadactyla* (Jason S C Chin, www.pangolinsg.org); (*b*) Malayan/Sunda pangolin, *M. javanica* (Joel Sartore, PhotoArk, www.pangolinsg.org); (*c*) Philippine pangolin, *M. culionensis* (Katala Foundation, www.pangolinsg.org); (*d*) Indian pangolin, *M. crassicaudata* (Zeeshan Merchant, www.pangolinsg.org); (*e*) Temminck's pangolin, *Smutsia temminckii* (Darren Pietersen, www.africanpangolin.org); (*f*) Giant pangolin, *S. gigantea* (Ugandan Wildlife Authority, www.africanpangolin.org); (*g*) Black-bellied pangolin, *Phataginus tetradactyla* (Alexis Kriel, www.africanpangolin.org); (*h*) White-bellied pangolin, *P. tricuspis* (Maxwell Boakye, www.africanpangolin.org).

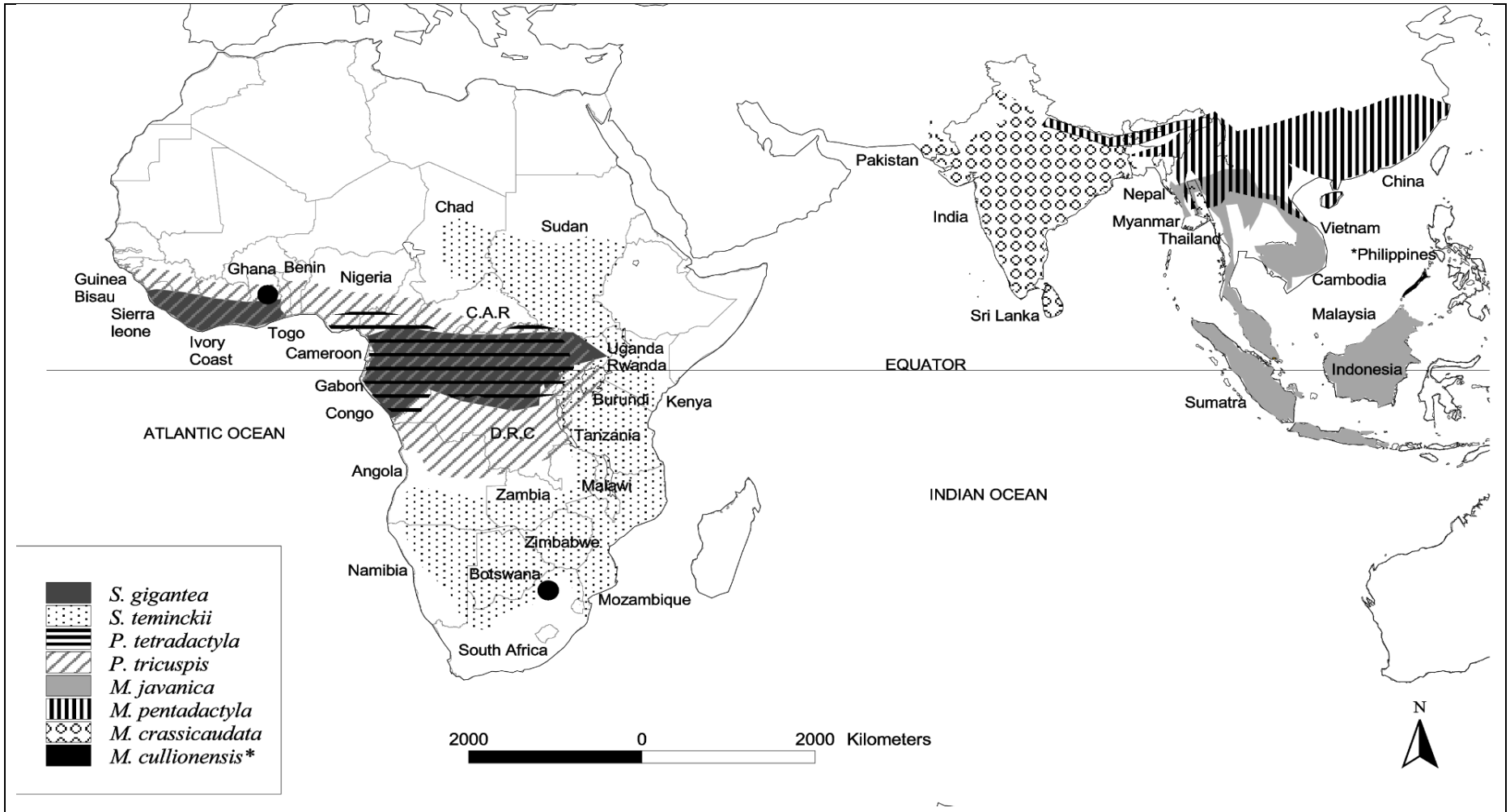


Figure 1.2. Distribution map of the eight extant pangolin species namely Giant pangolin, *S. gigantea* (Nixon *et al*. 2019); Temminck's pangolin, *S. temminckii* (Pietersen *et al*. 2019a); Black-bellied pangolin, *P. tetradactyla* (Ingram *et al*. 2019); White-bellied pangolin, *P. tricuspis* (Pietersen *et al*. 2019b); Malayan/Sunda pangolin, *M. javanica* (Challender *et al*. 2019b); Chinese pangolin, *Manis pentadactyla* (Challender *et al*. 2019c); Indian pangolin, *M. crassicaudata* (Mahmood *et al*. 2019); Philippine pangolin, *M. culionensis* (Schoppe *et al*. 2019).

In 2008, the IUCN re-evaluated the Red List Status of all eight pangolin species (inclusive Philippine pangolin) and up-listed all the species (Table 1.1). The Chinese- and Malayan pangolins were up-listed to *Endangered* (EN) status while the Philippine-, Indian-, giant- and black-bellied pangolins were up-listed as *Near Threatened* (NT). Temminck's pangolin and the white-bellied pangolin were up-listed to *Least Concern* (LC).

Due to an increase in the illegal pangolin trade over the years, from both Africa and Asia, the IUCN once again performed a re-evaluation of the Red List status of all eight pangolin species, and in 2014 a decision was taken to up-list all the species (Table 1.1). The Chinese and Malayan pangolins were up-listed to *Critically Endangered* (CR) while the Philippine and Indian pangolins were up-listed to *Endangered* (EN), as most of the illegal trade was limited to Asia. The African species were all up-listed to the same status of *Vulnerable* (VU).

In 2019, the latest assessment was done and three of the pangolin species were further up-listed. These species include, the Philippine pangolin which were up-listed to *Critically Endangered* (CR) and both the white-bellied and giant pangolin were listed as *Endangered* (EN). The remaining five species statuses remains the same (Challender *et al*. 2019b,c, Ingram *et al*. 2019; Mahmood *et al*. 2019; Nixon *et al*. 2019; Pietersen *et al*. 2019a,b; Schoppe *et al*. 2019).

In 1995, all pangolins were listed under Appendix II of the Convention on International Trade in Endangered Species of Fauna and Flora (CITES) at the 9th Conference of Parties (CoP9) (CITES 1995, 2017a). However, due to the increase in illegal trade, all eight pangolin species were transferred to CITES Appendix I at the 17th Conference of Parties (CoP17), which prohibit worldwide trade in any of the pangolin species, applicable from January 2017 (Dixon & Weiskotten 2016; IUCN 2016; Rohr 2016; Swails & McKenzie 2016; CITES 2017b).

1.2.4. THREATS TO PANGOLIN POPULATIONS

Pangolin population numbers across all eight species are rapidly declining due to a number of anthropogenic threats. These threats include illegal trade and over-utilization, electrocution on game fences, habitat loss and deforestation as well as road mortalities (Pietersen *et al.* 2016a).

1.2.4.1.ILLEGAL TRADE AND OVER-UTILIZATION

Pangolins are one of the of most intensely affected illegally traded mammal species globally, with an estimated 10 000 individuals being trafficked annually (Davies 2014; Sutter 2014a; Pendharkar 2016). This poses a significant threat to these species, as all eight pangolins species are traded on an unsustainable level to supply the demand. Due to this unsustainable demand, the Chinese- and Malayan pangolins face extinction with drastic declines in their population numbers and have therefore been up-listed to *Critically Endangered* on the IUCN Red List of Threatened Species as indicated above (Challender *et al*. 2019b,c).
However, this restriction has not reduced the severe demand for pangolin and pangolin related products, as alternative sources were sought to supply the demand. These sources included a shift in focus to the remaining two Asian pangolin species (which were consequently up-listed to *Endangered* status), as well as the four species distributed in Africa. The illicit trade in pangolins and pangolin parts has thus significantly contributed to the decline in population numbers of all eight pangolin species (Challender *et al*. 2019b,c; Ingram *et al*. 2019; Mahmood *et al*. 2019; Nixon *et al*. 2019; Pietersen *et al*. 2019a,b; Schoppe *et al*. 2019).

Asian trade: Pangolins are sought after as a delicacy in certain Asian restaurants, where the meat as well as the blood and foetuses are used in a variety of dishes and sold at an exclusive price. Various other body parts are also utilized for use in traditional medicine, especially the scales, which are considered the most valuable part of the animal. The scales are grounded into a powder and prescribed for a number of ailments (Fang & Wang 1980).

African trade: While being targeted to supply the Asian markets, the African pangolin species also face numerous threats in their local range states where they are utilised for the bush-meat trade and *Muthi* markets (Bräutigam *et al*. 1994; Sodeinde & Adedipe 1994; Alves & Rosa 2005; Chakravorty *et al*. 2011; Boakye *et al*. 2014, 2015, 2016). They are thought to be the most commonly used animals in traditional medicinal practices by traditional healers (Soewu & Ayodele 2009; Boakye *et al*. 2014, 2015). For these purposes, none of the animal's body parts goes to waste with everything, inclusive the scales, tongue, blood, claws, bones, meat and internal organs utilized for various treatments for a variety of ailments (Sodeinde & Adedipe 1994; Soewu 2008; Soewu & Ayodele 2009; Soewu & Adekanola 2011).

The bush-meat trade also show an increase, with pangolins harvested and sold as a popular meat source. Studies have indicated that pangolins are being harvested in Africa at an unsustainable rate and as a result, these species could also face extinction of populations, similar to the Chinese- and Malayan pangolins (Boakye *et al.* 2016; Pietersen *et al.* 2016a).

1.2.4.2.ELECTROCUTION

In addition to the global threat of pangolin over-exploitation for illegal wildlife trade, Temminck's pangolin populations are exposed to a greater threat in southern Africa. Private game- and livestock farms utilize electrified fences to protect their livestock against predators as well as to keep the livestock contained within a particular area (Beck 2008; Pietersen *et al.* 2014a). However, this poses a threat to pangolins that are being electrocuted on the lower wires of the game fences. Electrocution happens as a result of the reflex instinct of pangolins when threatened or hurt to roll into a ball for protection. They curl around the electric wire and receive continuous pulse shocks until death ensues (Beck 2008; Pietersen *et al.* 2014a). Recent studies estimate fence mortalities to be in the region of between 377–1028 individuals electrocuted annually, which contribute to an estimated 2–13% decline in South African pangolin populations every year (Pietersen *et al.* 2016a).

1.2.4.3. ADDITIONAL THREATS

In addition to the threat that pangolins face from illegal trade, there are several other factors that further contribute to the population decline of these mammals, albeit to a lesser extent. In this regard, two notable threats are habitat destruction and deforestation, as well as road mortalities (Pietersen *et al.* 2016a).

Habitat loss and deforestation: The distribution and dispersal of all pangolin species are shaped by their specialized and specific dietary requirements, as they naturally occur in habitats where the food source is abundant (Sodeinde & Adedipe 1994; Yang *et al*. 2007). Pangolin habitats have shown a global decline across all species, resulting from habitat loss as a result of agriculture and livestock farms, urbanization and deforestation (Sodeinde & Adedipe 1994; Yang *et al.* 2007; Whiting *et al.* 2011; Pietersen *et al.* 2016a).

They are known to be important pest control agents as they feed primarily on ant and termites, thereby controlling the population numbers of these insects (Heath 1992b; Whitfort 2012). Reports suggest an individual pangolin can consume more than 70 million, possibly a billion, of these invertebrates annually (Chao 2002; Pietersen *et al.* 2016a). However, pangolins are rarely found in areas where crop agriculture and urbanisation is dominant, as their food source is largely rendered unavailable (Pietersen *et al.* 2016a). This also applies to various forest areas where deforestation occurs and their habitat and food source is demolished (Challender *et al*. 2019b,c; Ingram *et al*. 2019; Mahmood *et al*. 2019; Nixon *et al*. 2019; Pietersen *et al*. 2019a,b; Schoppe *et al*. 2019).

Road mortalities: As pangolins are solitary animals by nature, juvenile pangolins - upon reaching maturity - disperse across vast distances in search of new home ranges (Pietersen *et al.* 2016a). This migration brings them in contact with roads and as a result leads to a significant number of fatalities every year. In South Africa, an estimated 280 individuals fall victim to this threat annually (Pietersen *et al.* 2016a). While less pronounced than fatalities incurred through electric fences and illegal trade, this number may be an under-estimation as few mortalities are reported and

the carcasses are removed from the roads by scavengers or for the bush-meat and illegal trade markets (Pietersen *et al.* 2016a).

1.3. CONSERVATION GENETICS OF PANGOLINS

The order Pholidota is one of the most sparsely studied mammal orders, with little to no information available regarding some of the species in terms of behaviour, ecology and population genetic structure. Through the use of genotypic methods, the population structure of pangolins can be unravelled, providing greater insight into the evolutionary history that underpins their biology and behaviour.

Genetic data will aid in identifying specific populations and population boundaries, or provide information on differences between species, in order to further develop effective conservation strategies. To date, limited molecular studies have been undertaken on the pangolins due to their scarceness, low numbers and nocturnal behaviour. Nevertheless, a variety of molecular markers (nuclear and mitochondrial) have been designed to determine various population genetic parameters of pangolin species (see below). Of the eight extant species studied, the Chinese, Malayan and Temminck's pangolins currently have the most genetic data available. The remaining five pangolin species have some genomic data available in the form of partial gene/loci sequences.

1.3.1. PREVIOUS STUDIES OF EVOLUTION OF THE ORDER PHOLIDOTA

A series of studies have investigated several genes associated with unique characteristics to determine the origin of features unique to the pangolin order from an evolutionary perspective (Meredith *et al*. 2014; Choo *et al*. 2016; Yusoff *et al*. 2016; Ma *et al*. 2017; Opazo *et al*. 2017). These gene groups included the genes associated with hair development and immunity in mammals as well as hormonal influences using Transcriptomes.

In the latter investigation, the authors focussed on these gene groups as the pangolins have a unique body coverage of keratin scales as opposed to hair. These

studies were performed on the Malayan pangolin (Choo *et al*. 2016; Yusoff *et al*. 2016) as well as the seven remaining pangolin species (Choo *et al*. 2016).

Choo and authors (2016) also investigated the genes associated with vision, as pangolins have limited eyesight. These authors used representatives of all eight pangolin species and compared them to various other mammalian species (Choo *et al*. 2016).

Since pangolins are myrmegophagous, they are adapted to a toothless feeding lifestyle, where the sticky tongue becomes a necessity to the diet of these animals. Various studies have investigated the gene group associated with the development and formation of teeth and subsequently, the loss of teeth in various groups of animals. As pangolins are edentate, various authors investigated the loss of teeth genes in the Chinese pangolin (Meredith *et al*. 2014; Choo *et al*. 2016), Malayan pangolin (Choo *et al*. 2016; Opazo *et al*. 2017) and the remaining six pangolin species (Choo *et al*. 2016), which were compared to other toothless species. In 2017, Ma and authors sequenced 22 538 transcripts in the Malayan pangolin associated with myrmegophagy (Ma *et al*. 2017). Of these transcripts, 3 373 new genes were identified and of these 1 459 were annotated. The authors also found 969 genes to be associated with olfactory transduction and 75 genes linked to metabolism (Ma *et al*. 2017).

1.3.2. TAXONOMIC CLASSIFICATION OF PANGOLIN SPECIES

Phenotypic classification uses morphological characteristics to describe organisms and divide those into different species and even sub-species. This was initially the preferred method for identification of specimens; however, the limitations arrived with accessioning cryptic species of which many have the same morphological traits, but were genotypically separate species. Subsequently, this led to genotypic classification becoming a necessity for assisting in the identification of closely related and cryptic species (Knowlton 1993; Jarman & Elliott 2000; Hebert *et al.* 2003; Savolainen *et al.* 2005; Papeş & Gaubert 2007; Kress & Erickson 2008).

1.3.2.1. TAXONOMIC DELINEATION

The taxonomic classification of pangolins has been under debate for a number of years, and is still not fully resolved. This is due to a lack of discriminating morphological and molecular data (Pocock 1924; Emry 1970; Mckenna & Bell 1997; Gaudin & Wible 1999; Nowak 1999; Wilson & Reeder 2005).

The first genus described for the extant pangolins was the genus *Manis* (Linnaeus 1758), which was proposed as the universal genus for all extant pangolin species globally (Linnaeus 1758; Nowak 1999; Wilson & Reeder 2005). However, the genera *Phataginus* (Rafinesque 1821) and *Smutsia* (Gray 1865) were later described, which were assigned to the African pangolins as sub-genera. The name *Phataginus* has been ascribed to the arboreal pangolins and the ground dwelling pangolins has been designated as *Smutsia.*

Later authors suggested two genera for the pangolins, namely *Manis* for the Asian pangolins and *Phataginus* for the African pangolin species (Corbet & Hill 1991), with *Smutsia* as sub-genus for the ground pangolins and *Uromanis* for the black-bellied pangolin. The taxonomy was later changed to four genera, with *Manis* assigned to the Asian pangolins, *Smutsia* for the African ground pangolins, *Phataginus* for the white-bellied pangolin and *Uromanis* for the black-bellied pangolin (Kingdon 1997; Mckenna & Bell 1997).

In 1999, Gaudin and Wible published research suggesting all eight pangolin species be assigned to three genera, namely *Manis* for the Asian pangolins, *Phataginus* for the arboreal African pangolins and *Smutsia* for the ground-dwelling African pangolin species (Gaudin & Wible 1999). This nomenclature was also supported by Koenigswald (1999).

In 2009, Gaudin and co-authors re-assessed this nomenclature of pangolins from examining entire skeletons for osteological characteristics and concluded that all extant pangolin species can be assigned to the three genera mentioned above. These authors also concluded that the two African genera, *Phataginus* and *Smutsia,* form a monophyletic group and should be re-assigned to a separate sub-family,

Smutsiinae, which still fall under the one family Manidae (Gaudin *et al*. 2009). However, some authors still show preference for one genus, *Manis*, for all extant pangolin species (Gaubert & Antunes 2005; Gaudin *et al*. 2006; Botha & Gaudin 2007; Gaubert *et al*. 2016).

1.3.2.2.GENOTYPIC CLASSIFICATION

Currently, various DNA sequencing technologies are available to aid biologists in studying a broad range of applications in a fast, precise, manageable and cost-effective way. Some of these applications include techniques such as cloning and breeding, pathogenic analysis in the identifying of pathogenic genes as well as comparative population and evolutionary studies (Liu *et al*. 2012).

A variety of molecular techniques such as DNA hybridization, restriction enzyme digestion, random PCR amplification, species-specific PCR primer and DNA barcoding were developed as different genotyping approaches (Hayashi *et al*. 1985; Irwin *et al*. 1991; Cano *et al*. 1993; Esposti *et al*. 1993; Collura & Stewart 1995; Brown *et al*. 1996; Birstein & DeSalle 1998; Burgener & Hübner 1998). DNA sequencing can be broadly classified into two main branches using either mitochondrial DNA (mtDNA) or nuclear DNA (nDNA) sequence comparisons.

MITOCHONDRIAL MARKERS

Mitochondrial DNA (mtDNA) is found in the mitochondria of the living cells of Eukaryotic organisms. Each mitochondrion carries two to ten copies of mitochondrial DNA, with an estimated 100 mitochondria found inside a single mammalian cell (Robin & Wong 1988). Mitochondrial DNA is estimated to form one to two percent of all DNA available inside a cell (Yang *et al*. 2014).

Due to the abundance of mitochondria and mitochondrial DNA copies in each cell, it has become an established method in a variety of molecular analysis. These analysis include phylogenetics and evolutionary studies (Gibb *et al*. 2007; Lei *et al*.

2010), species identification for forensic purposes (Seo *et al*. 2015; Mwale *et al*. 2017) as well as kinship analysis of the maternal lineage (Seo *et al*. 2015).

Mitochondrial DNA can be isolated from hair shafts, scales and bones which in turn make identification of species from confiscated scale or bone samples possible (Seo *et al*. 2015; Du Toit *et al*. 2016; Mwale *et al*. 2017). Mitochondrial DNA is also known for a high mutation rate and low recombination, which results in variability that enables the study of animal lineages and subsequent evolutionary analysis through reconstructing population- and species histories (Martin & Palumbi 1993; Rand 1994; Mindell *et al*. 1996; Bi & Sokouri 2015; Seo *et al*. 2015). This variability also aids in maternal lineage traceability, over several generations (Seo *et al*. 2015).

Whole mitochondrial DNA genomes or mitogenomes are more informative, compared to partial gene analysis, in determining phylogenetic and taxonomic characteristics of species. To date, reference mitogenomes for seven of the extant pangolin species are available on Genbank: *S. temminckii*; *S. gigantea; P. tricuspis*; *P. tetradactyla*; *M. pentadactyla*; *M. javanica* and *M. crassicaudata* (Arnason *et al*. 2002; Qin *et al*. 2012; Du Toit *et al*. 2014; Hari & Siew Woh 2015; Hassanin *et al*. 2015; Hari *et al*. 2016; Gaubert *et al*. 2018)*.

Prior to the work by (Du Toit *et al*. 2014), two mitogenomes for black-bellied pangolin (Arnason *et al*. 2002) and the Chinese pangolin (Qin *et al*. 2012) already existed in the database. However, various studies utilising partial mtDNA loci determined that these two mitogenomes were wrongly ascribed to the respective species and these should rather be included with the white-bellied and Malayan pangolin species respectively (Olayemi *et al*. 2011; Hassanin *et al*. 2015; Hari *et al*. 2016).

(*The work from Gaubert *et al.* (2018) was published subsequent to the second research paper, mitogenomes of the four African pangolin species. It is not included here since it forms part of the thesis where it is discussed in detail. See Chapter 3).

The phylogenetic placement of the order Pholidota within the Eutherian tree has been the subject of a number of studies utilizing the comparison of sequenced mitogenomes. From these studies it was evident that the orders Carnivora and Perisodactyla formed a monophyletic group, with Pholidota, contrary to the expected

grouping with Xenathra (Arnason *et al*. 2002; Du Toit *et al*. 2014; Gaubert *et al*. 2018).

Y-CHROMOSOME MARKERS

The Y-chromosome gene (*Sry*) is a locus responsible for sex determination in mammalian species and is located as a single copy on the Y-chromosome of male animals (Pamilo & O'Neill 1997). Analogous to mitochondrial DNA, which is maternally inherited, the Y-borne gene is paternally inherited. This gene is also responsible for the development of bipotential embryonic gonads to form the male reproductive organs, the testes (Yu *et al.* 2011). The use of Y-chromosome markers for phylogenetic purposes is advantageous when determining dispersal and migrating patterns in male individuals of a population.

In 2011, Yu and co-authors sequenced the Y-chromosome (*Sry*) of the Formosan pangolin (*Manis pentadactyla pentadactyla*), which is a sub-species of the Chinese pangolin. The aim of the latter study was to further elucidate the taxanomic placement of Pholidota in the Eutherian tree. The results suggested that pangolins are phylogenetically closer to Carnivora and Perrisodactyla and share a Laurasian origin rather than grouping with the order Xenarthra. The study also concluded that the orders Pholidota and Xenarthra share similar characteristics, which can probably be ascribed to convergent evolution rather than the two orders being related and sharing a common ancestor (Yu *et al*. 2011).

In 2016, Gaubert and co-authors published research using various nuclear and mitochondrial genes and loci, including the Y-chromosome (*Sry*) gene, in order to describe various white-bellied pangolin populations in Africa. However, the results were less reliable compared to the results from mitochondrial loci as inconsistent patterns of population differentiation were observed using the Y-chromosome gene (Gaubert *et al*. 2016). To date, no further studies have been conducted utilising the Y-chromosome for any of the remaining six pangolin species.

WHOLE GENOME DNA ANALYSIS

In general, only one or a few genes or gene segments were traditionally used to determine phylogenetic relations, which can lead to a distorted view of the phylogenetic structure of populations (Brown *et al*. 1979; Moore 1995; Zardoya & Meyer 1996; Springer *et al*. 2001). Whole mtDNA genomes are more reliable since there are more positions (bases) available to reflect variation between species, and this greatly improves the resolving power during analysis (Cao *et al*. 1994, 2000; Cummings *et al*. 1995; Zardoya & Meyer 1996; Ingman *et al*. 2000). The same is true for whole nuclear genome assemblies, as it allows any gene group or chromosome to be studied and aids in comparison studies between different species or individuals on a genomic level.

To date, only two of the extant pangolin species have whole genome data available, at a level where the genome was sequenced and annotated (Choo *et al*. 2016; Tan *et al*. 2016; Zhihai *et al*. 2016; Hu *et al*. 2020). Choo and co-authors published annotated genomes for the Chinese pangolin (*M. pentadactyla*) and the Malayan pangolin (*M. javanica*). Although the genomes are not completely assembled, large scaffolds exist with multiple protein coding gene identities determined. These authors successfully identified 20 298 genes in the Chinese pangolin and 23 446 genes in the Malayan pangolin. Pertaining to the above, 3 465 gene variants were found to be unique to the Chinese pangolin and 4 958 gene variants for the Malayan pangolin, indicating the significant level of divergence between these species. During the annotation of the genomes, the authors also found 1 152 genes which seem to be unique to the order Pholidota, as they were absent in the other species used for comparison (Choo *et al*. 2016).

In 2020, Hu and authors published a paper which compared the whole genomes of the Chinese and Malayan pangolin, a more in-depth and robust analysis compared to Choo *et al.* (2016). The authors used the previous assembled scaffolds as guidelines to assemble the whole genomes of the two species (Hu *et al*. 2020). They found 21 785 genes in the Malayan pangolin and 23 865 genes in the Chinese pangolin which were slightly more than reported in the previous study.

1.3.3. GENETIC DIFFERENTIATION AMONG AND WITHIN SPECIES

Species identification has been successfully demonstrated in numerous studies, utilising partial mtDNA regions to determine species identities, especially during confiscation of illegally traded animals such as pangolins.

1.3.3.1 DNA BARCODING APPLIED TO PANGOLIN SCALE TRACEABILITY

Confiscated samples (scales, bush-meat, whole specimens) created a need for a reliable and rapid method to distinguish between the different pangolin species in order to trace origin and to determine which species are targeted frequently.

Currently, five mitochondrial loci exist to aid in reliable identification of confiscated samples across the eight pangolin species. These markers are: the cytochrome c oxidase I (*CoI*) locus which has been successful in identification of *S. temminckii*; *S. gigantea*; *P. tetradactyla*; *P. tricuspis*; *M. javanica* and *M. culionensis* (Du Toit 2014; Gaubert *et al.* 2015; Hassanin *et al.* 2015; Zhang *et al.* 2015; Luczon *et al.* 2016; Mwale *et al.* 2017; Mason *et al.* 2019; Hu *et al.* 2020); the cytochrome b (*Cytb*) locus for all extant pangolin species excluding *M. culionensis* (Hsieh *et al.* 2001; Olayemi *et al.* 2011; Wirdateti 2013; Du Toit 2014; Gaubert *et al.* 2015, 2016; Hassanin *et al.* 2015; Kumar *et al.* 2016b; Tan *et al.* 2016; Mwale *et al.* 2017; Zhang *et al.* 2020; Hu *et al.* 2020); the control region in *S. temminckii*; *S. gigantea*; *P. tetradactyla*; *P. tricuspis*; *M. p. pentadactyla* and *M. javanica* (Hsieh *et al.* 2011; Du Toit 2014; Hassanin *et al.* 2015; Gaubert *et al.* 2016; Mwale *et al.* 2017); *16S rRNA* for *S. temminckii*; *P. tricuspis*; *M. javanica* and *M. crassicaudata* (Gaubert *et al*. 2015; Hassanin *et al*. 2015; Kumar *et al*. 2016b); and *12S rRNA* for *S. temminckii*; *P. tricuspis* and *M. javanica* (Gaubert *et al*. 2015; Hassanin *et al*. 2015).

1.3.4. GENETIC ASSIGNMENT OF INDIVIDUALS TO POPULATIONS

Various techniques have been developed for the assignment of individuals to source populations. These are based on both mitochondrial DNA as well as nuclear DNA research and include sequencing of genes, Simple Tandem Repeats (STRs), Single

Nucleotide Polymorphisms (SNPs) and Restriction Site Associated DNA Markers (RAD).

1.3.4.1.PARTIAL DNA ANALYSIS (SEQUENCING OF SPECIFIC REGIONS)

Several studies have utilised both mitochondrial DNA (mtDNA) and nuclear DNA (nDNA) markers to determine population structure of pangolins, in the context of locating source populations.

MITOCHONDRIAL DNA

The utilization of mitochondrial loci for phylogeographic identification of pangolin populations is useful in determining the geographic origin of individuals as well as revealing population differentiation and diversity. To date, phylogeographic studies, using specimens with known identities and from known localities, have been completed for two African pangolin species utilizing various mtDNA loci.

The first phylogeographic study was performed by Du Toit (2014), who studied three mtDNA regions (*CoI*, *Cytb* and control region) for Temminck's pangolin in order to determine patterns of phylogeographic relationships and genetic differentiation among southern African populations. This author found low levels of genetic diversity among and between populations which indicated a high level of gene flow between these populations. As a result, determining geographic origin of individuals of this species was not possible. Du Toit (2014) recommended that future studies should also include nuclear markers and SNPs.

In 2016, Gaubert and co-authors sampled various white-bellied pangolins in Africa and used phylogenetic data and differences between populations to determine the origin of confiscated bush-meat samples. The results indicated that there were significant levels of genetic differentiation between all the identified populations and this will aid in determining the geographic origin of confiscated samples (Gaubert *et al*. 2016).

NUCLEAR DNA

Nuclear DNA markers are a valuable source of information, since various parts of the genome, or the entire genome, can be studied and compared between different individuals, populations or species. As indicated above, phylogeography can also be performed using a variety of nuclear DNA markers in addition to mitochondrial markers, in order to obtain a more accurate view of population structure.

To date, only one study included nuclear markers to aid in the identification of population structure in a pangolin species, with such data published for the white-bellied pangolin (Gaubert *et al*. 2016). These authors used the genes intron 7 fibrinogen (*FGB7*), titin (*TTN*) along with a series of mtDNA genes and the Y-chromosome gene (*Sry*), as indicated above. The results showed that these nuclear markers, including *Sry*, were less effective in determining the precise number of populations in the white-bellied pangolin when compared to the mtDNA markers.

1.3.4.2.SHORT TANDEM REPEATS (STRS)

Short Tandem Repeats (STRs), also referred to as Simple Sequence Repeats (SSRs) or microsatellites are variable repeat motives in the genome, usually with high heterozygosity levels and multiple alleles. These alleles consist of a variety of motifs, from dinucleotide to tetra-nucleotide repeats (Ellegren 2004). Microsatellites are popular molecular makers since these loci have a relatively high mutation rate. This allows microsatellites to be utilised in a variety of studies such as differentiation between individuals within as well as between various populations (Ellegren 2004; Luo *et al*. 2007). Other analysis also include studies of linkage mapping, linkage-disequilibrium (Ohashi & Tokunaga 2003), relatedness through parentage testing, disease discoveries and forensic application (Ellegren 2004).

In 2007, Luo and co-authors designed a series of 32 species-specific microsatellite markers for the Malayan pangolin (*M. javanica*). These authors tested these markers on the Chinese pangolin (*M. pentadactyla*) as well as the white-bellied pangolin (*P. tricuspis*), with a success rate of 84% (27) and 52% (18) respectively. The study

compared these species in order to determine phylogeography, evolutionary- and population genetic statuses. The results indicated an estimated heterozygosity ranging from 0,321 to 0,708 for the three species and presented the first nuclear data for any of the pangolin species (Luo *et al*. 2007).

In 2013, 14 of the original 32 markers from Luo *et al.* (2007) were tested on Temminck's pangolin (*S. temminckii*) samples with limited success, since these markers were developed and optimized for the Asian pangolins (De Beer 2013). As evident in Luo's study, only 50% of the markers showed a success rate when tested on the white-bellied pangolin and the even lower success rate in Temminck's pangolin meant that results were not robust enough for reliable analysis. In 2020, Sun *et al* developed 10 species specific microsatellite markers for the Chinese pangolin that could be successfully utilized for forensic purposes in species identification as well as between different individuals and siblings (Sun *et al*. 2020).

To date, no microsatellite data exists for the remaining two Asian pangolin species. In 2020, Aguillon and co-authors published an additional 20 species specific STRs for the white-bellied pangolin (Aguillon *et al*. 2020)*.

(*The second chapter of this thesis, first research chapter, involves the development of STR for *S. temminckii* with demonstrated cross amplification capability for the remaining three African pangolin species. Although there is an overlap with Aguillon *et al.* (2020) regarding the STR for white-bellied pangolin, the aim of the research papers were different and the studies done simultaneously where the STR for *S. temminckii* was published prior to the work from Aguillon *et al.* (2020). It is not included here since it forms part of the thesis where it is discussed in detail.)

1.3.4.3.SINGLE NUCLEOTIDE POLYMORPHISMS (SNPS)

Single Nucleotide Polymorphisms (SNPs) are single base variations where at least one nucleotide is exchanged for an alternative nucleotide in different individuals at a specific location in the genome (Shastry 2002). These SNPs are used in genotype calling where the individuals have to display at least one base variation (Nielsen *et al*. 2011). This makes detection of rudimentary polymorphism between different genomes feasible. There is an estimated nine million SNPs found in a single organism, which cover around 90% of their genome (Kwok *et al*. 1996; Collins *et al*. 1999; Rocha *et al*. 2006). These SNPs can occur as frequently as every 250–1000

base pairs (bp) across the genome (Kwok *et al*. 1996; Collins *et al*. 1999; Shastry 2002; Rocha *et al*. 2006). Due to the abundance of SNPs in the genome, it has become one of the most valuable genetic markers for detection of differentiation in populations and species, using a selection of markers across the genome without sequencing the whole genome (Shi *et al*. 2009).

Currently, SNP analysis have only been reported for three of the Asian pangolin species, Chinese (*M. pentadactyla*), Malayan (*M. javanica*) and Indian (*M. crassicaudata)* pangolins (Li *et al.* 2017; Kumar *et al.* 2018; Hu *et al.* 2020). Li and authors (2017) designed 21 SNP markers for the Malayan pangolin which could be used in subsequent studies. In 2020, Hu and authors published an article where >20 000 000 and >21 000 000 SNPs were developed for the Malayan and Chinese pangolin respectively (Hu *et al*. 2020). The results found that there were two distinctive groups present in both the Malayan and Chinese pangolins which could be traced to 300 and 130 thousand years ago respectively (Hu *et al*. 2020).

With regards to the Indian pangolin, mtDNA SNPs were developed for three gene regions: *Cytb*, *16S rRNA* and *12S rRNA* (Kumar *et al.* 2018). The authors used references from online sources to obtain data for six of the extant pangolin species. The only species which didn't have any data available was the Philippine pangolin (*M. culionensis*). The study reported SNPs which were present in all seven species: 25 SNPs for *Cytb*, 15 SNPs for *16S rRNA* and 19 SNPs for *12S rRNA*. Sequence variation in the *Cytb* gene was found to be the highest (Kumar *et al.* 2018).

1.3.4.4. RESTRICTION SITE ASSOCIATED DNA MARKERS (RAD)

Compared to whole genome based studies, restriction enzyme based methods, in the form of Restriction Site Associated DNA (RAD) techniques, are cost effective and simpler since only a subset of the genome of interest is targeted (Miller *et al*. 2007; Baird *et al*. 2008; Van Tassell *et al*. 2008; Elshire *et al*. 2011; Ali *et al*. 2016). These techniques allow for the simultaneous sequencing of a greater sample size by creating a multiplex of all the samples in a single reaction. This in turn reduce the cost per individual for sequencing (Baird *et al*. 2008; Hohenlohe *et al*. 2010; Etter *et al*. 2011).

The RAD techniques utilizes restriction enzymes for targeting a subset of homologous loci which allows for SNP discovery across several individuals simultaneously (Davey *et al*. 2011; Trebbi *et al*. 2011; Peterson *et al*. 2012; Graham *et al*. 2015). Therefore, this technique is beneficial for studies consisting of large population sizes in conservation, ecological, evolutionary and agricultural applications (Poland & Rife 2012; Davey *et al*. 2013; Narum *et al*. 2013). To date, no reported studies have been performed on any of the pangolin species utilising RAD sequencing for SNP identification.

1.4. AIMS AND OBJECTIVES

1.4.1. AIMS OF THE STUDY

The overall aim of this study was to develop molecular markers – both nuclear and mitochondrial – to establish a geo-referencing framework of Temminck's pangolins in southern Africa. From this information, the population structure and relationship with other pangolin species will be unravelled. In addition, an aim was to also have insight into aspects of the genetics of the remaining three African pangolin species. This study is part of an extended project in support of increasing molecular data availability for all four African pangolin species. Understanding the population structure will aid in better and more accurate conservation strategies to protect these animals from the illegal wildlife trade and possible extinction, by enabling positive identification of confiscated samples intercepted from illegal wildlife trafficking.

1.4.2. OBJECTIVES OF THE STUDY

a) Design and optimize species-specific Short Tandem Repeats (STRs) markers for Temminck's pangolin (*S. temminckii*). The hypothesis is to obtain species specific markers with cross-amplication in the remaining African pangolin species.

b) Sequence the whole mitogenomes of all four African pangolin species to perform a comprehensive molecular comparison and to resolve the taxonomy of the family Manidae. The hypothesis is to clearly differentiate between the three pangolin genera using mitogenomic data and provide further support to the recent morphological studies.

c) Geo-referencing of Temminck's pangolin (*S. temminckii*) populations in southern Africa using both Short Tandem Repeats (STRs) and three mitochondrial loci. The hypothesis is to differentiate between different sample localities using samples with known and unknown origin. This will aid forensic cases in determining better release sites for the confiscated animals and in turn provide support for conservation strategies to be properly implemented.

1.5. STRUCTURE OF THE THESIS

A publication format with papers (published or publishable) is used for the research chapters encompassing each study objective. An oversight of the relevant published literature is presented and precedes the rest of the chapters (Chapter 1). The last chapter consists of a synopsis and discussion of the results in a concluding chapter (Chapter 5)

The papers generated and representing the different chapters are as follows:

1. Chapter 2: Du Toit, Z., Dalton, D.L., Du Plessis, M., Jansen, R., Grobler, J.P. and Kotzé, A. 2020. Isolation and characterization of 30 STRs in Temminck's pangolin (*Smutsia temminckii*) and potential for cross amplification in other African species. *Journal of Genetics*, 99, doi.org/10.1007/s12041-020-1176-0.

2. Chapter 3: Du Toit, Z., Du Plessis, M., Dalton, D. L., Jansen, R., Grobler, J. P. and Kotzé, A. 2017. Mitochondrial genomes of African pangolins and insights into evolutionary patterns and phylogeny of the family Manidae. *BMC Genomics*, 18: 746-758.

3. Chapter 4: Phylogeography of Temminck's pangolin (*Smutsia temminckii*) populations in southern Africa using a georeferencing approach. *Unpublished*.

CHAPTER TWO

ISOLATION AND CHARACTERIZATION OF 30 STRS IN TEMMINCK'S PANGOLIN (*SMUTSIA TEMMINCKII*) AND POTENTIAL FOR CROSS AMPLIFICATION IN OTHER AFRICAN SPECIES

2.1. INTRODUCTION

The mounting pressure of pangolin populations due to illicit trade and poaching (Heinrich *et al.* 2016; Pietersen *et al.* 2016a), prompted us to find suitable genetic techniques to aid forensic investigations. Short tandem repeats (STRs) are considered to be the preferred marker in studies such as assessing divergent lineages and evaluating population genetic structure which can subsequently contribute to conservation management (Ellegren 2004; Luo *et al.* 2007). To date, STR markers have been developed for the Malayan pangolin (Asia) *M. javanica* (Luo *et al.* 2007), which displayed cross-species amplification in Chinese pangolin (*M. pentadactyla*) and white-bellied pangolin (Luo *et al.* 2007). However, no STR markers are reported to amplify in all African species. The aim of this study was to isolate and characterise polymorphic STRs for Temminck's pangolin using Next Generation Sequencing (NGS) and to determine cross-amplification potential in the remaining species (*S. gigantea*; *P. tetradactyla* and *P. tricuspis*).

2.2. MATERIALS AND METHODS

Samples from 32 Temminck's pangolin from the Northern Cape Province, South Africa, nine and ten samples respectively from black-bellied and white-bellied pangolins from Ghana and ten giant ground pangolins, from a confiscated batch from Hong Kong, with an unknown origin were used. All samples were stored at the NZG Biobank, National Zoological Garden (NZG). DNA was isolated from tissue or blood samples using the QIAamp Micro Kit (QIAGEN, Novato, CA, U.S.A.) along with the respective manufacturers' protocol. Scale samples were processed according to the protocol of (Du Toit *et al.* 2016) using the PrepFiler^{TM} Forensic DNA Extraction Kit

(Thermo Scientific, Massachusetts, U.S.A.) following the manufacturer's protocol for hair samples. The DNA was quantified on a Qubit v3.0 Fluorometer (Thermo Scientific, Massachusetts, U.S.A.).

One Temminck's pangolin blood sample was used for NGS and a rapid run was performed on the Illumina HiSeq 2500 (Illumina Incorporated, San Diego, CA, U.S.A.) using the TruSeq DNA LT Sample Prep Kit (Illumina Incorporated, San Diego, CA, U.S.A.). The quality of the data was evaluated using the software packages FastQC v0.11.7 (Andrews 2010) and Trimmomatic v0.36 (Bolger *et al.* 2014) was implemented to edit and trim the data by removing poor quality sections and adapters. The program Tandem Repeats Finder v4.09 (Benson 1999) was utilized to identify regions comprising di-, tri- and tetra nucleotide repeats of ≥ eight repeats in length. The unique loci, consisting of di-, tri- and tetra repeats, were identified and primer pair flanking regions was designed using BatchPrimer3 v1.0 (You *et al.* 2008) software.

Polymerase Chain Reaction (PCR) amplification was performed in multiplex reactions with a total reaction volume of 12,5 µl, consisting of *Taq* DNA Polymerase Master Mix Red (Ampliqon A/S, Odense, Denmark), forward and reverse primers (0,25 µM each) from multiple loci and 50 ng genomic DNA template. The conditions for the PCR amplification were as follows: denaturation at 94°C for 5 min, followed by 30 cycles of denaturation at 94°C for 30 sec, annealing at 45–55°C for 30 sec and extension at 72°C for 1 min, followed by final extension at 72°C for 40 min. PCR products were run against a Genescan™ 500 LIZ™ internal size standard on an ABI 3500 Genetic Analyzer (Applied Biosystems, Inc., Foster City, CA, U.S.A.).

Samples were genotyped using GeneMapper v5.0 software (Applied Biosystems, Inc., Foster City, CA, U.S.A.). MICRO-CHECKER v2.2 (Van Oosterhout *et al.* 2004) was used to detect possible genotyping errors, allele dropout and null alleles. The number of alleles per locus (*A*), observed heterozygosity (*Ho*), and unbiased heterozygosity (*Hz*) were calculated with GenAlEx v6.5 (Peakall & Smouse 2012), while Arlequin v3.5 (Excoffier & Lischer 2010) was used to test for deviations from expected Hardy-Weinberg (*HW*) proportions in genotypes and to evaluate loci for

gametic disequilibrium. Associated probability values were corrected for multiple comparisons using Bonferroni adjustment at a significance level of 0,05.

2.3. RESULTS

The NGS rapid run generated more than 10 000 reads containing STRs, of various repeat lengths, from which 40 were identified for further analysis. A total of 30 out of the 40 loci tested, amplified consistently (Table 2.1) in Temminck's pangolin. Annealing temperatures ranged from 45–55°C for all loci. The number of alleles ranged from 2 to 12, with an average of 4,4 alleles per locus. The mean Ho ranged from 0,00–0,968 and He values ranged from 0,11 to 0,75. No evidence of linkage disequilibrium was observed for the analysed loci. However, six markers deviated from Hardy-Weinberg equilibrium following Bonferroni correction, which may suggest population substructure or the presence of null alleles (Table 2.2).

Table 2.1. Primer sequences of 30 STRs isolated for Temminck's pangolin, *S. temminckii*.

Locus	Primer sequence forward (5′–3′)	Primer sequence reverse (5′–3′)	Repeat motif	Annealing temp (°C)
PAN_01	CAGCTGAACACTGAACACAGAA	AGAGCAAGGCAGAGCATTTG	$(GGT)_{12}$	55
PAN_02	TAGAGGCTGTGAGCAGGTTC	GAGGGGATTATGCCAGGAAG	$(TCC)_{9}$	48
PAN_03	ACTTTCCATCCCAAGCCTCT	GGTGGAGAGCGCTAATCAGT	$(GA)_{13}$	48
PAN_05	TTTTCCCCTGTGGAATGTGT	TTTGCGTAAAGTCCCTAGCA	$(GAT)_{11}$	48
PAN_06	CACACAGTAGGGCTCTGCAA	CAACACATCACCTGACAGCA	$(TGAA)_{10}$	48
PAN_07	CTTCTCCCACCTGACCTCCT	CACTTAAGATTTTGGGGCTCA	$(CT)_{15}$	48
PAN_08	CCTCCAGTGTATGTGCGAGT	CCACGGCAAGAACCATTC	$(CA)_{15}$	48
PAN_10	AAAATGGACACAGAGGCATT	ACATTGTGATGTGGACCAGTG	$(TC)_{14}$	48
PAN_11	ACTCGGGTTGTGGGAACAT	CCAGTCCTTGCTGCTCCTAT	$(GT)_{13}$	55
PAN_12	TAGTTCACACACATGGGAAGTT	TGTTATCCTTGCCTCCTGACTT	$(GGA)_{9}$	48
PAN_13	AGCTCCCTGAAGAACCTGCT	CTAGTGCCTGGCCCAACA	$(TCCA)_{9}$	50
PAN_15	GTCCTCAAACTCCTCCGACA	CTGGGGTCCTGTCATCCTG	$(CTC)_{9}$	50
PAN_17	CCCTGTGAAAAATGGAGACC	CTGGTGGCATTATAAGCCTTG	$(GT)_{15}$	45
PAN_18	GCACCTGTGGGTAGTTCCAT	GCATATGGACAAGGCAGGAG	$(CT)_{13}$	55
PAN_19	GAGCATCTATGTTTAGACTGTTCTGC	TGGGCAGAGACCAAGACTGT	$(GT)_{15}$	50
PAN_20	TTTGGGCACAAGTGATCCTA	GAGTCAAGACAGGGCCAGTG	$(AC)_{14}$	50
PAN_21	AGACGGAAGCCCAAGAGG	CCAGAAATGGAAGTTCAGTCC	$(CAC)_{9}$	50
PAN_23	TTCCCATTATGCGACTAGGA	TGCCCCAAATCCAATACTGA	$(GAG)_{9}$	55
PAN_24	AGAAAGGTTCTTGGGGCATT	AGACCATAAGCTCCCTGAAGAA	$(TGGA)_{12}$	55
PAN_26	GGAGATACCAAGGCAAAAAGAA	GGACCATCTGGTCAAATGAGA	$(AG)_{14}$	55
PAN_27	GCTGTTGGCTGGTTACAGGT	CTATGTGTGGGTGTGGGTGT	$(CA)_{13}$	55
PAN_28	CCACCTTTAACCTCAGGGCTA	CACTCCCCTTAACCCATCAA	$(TGAA)_{11}$	50
PAN_29	ACCTTACGCACCTGGAAGAC	AATTCTTGCCTCAGCTCTGC	$(GT)_{17}$	45
PAN_30	TCATATTCATGTGATGTGATGAGG	AAGGCTGCTGCGGTAGAAC	$(GAG)_{8}$	45
PAN_31	CTTCAGTCATCCGTCCGTCT	CACTTGACACAGTGCCTTGC	$(TCCA)_{8}$	45
PAN_32	TCTGAATGATGCCAAGGAGA	TGGCTCTGCAGATTTTCTCA	$(TG)_{14}$	45
PAN_35	GCCTACACTTGCGTGTGAAA	TTCCATCGGTGTGATGATTG	$(CA)_{13}$	55
PAN_36	GGCAGACTTTACCAATGAAGC	AACCATAATGTTGATCATGTAATTG	$(TTG)_{8}$	55
PAN_37	GTCTTCAGCCCACAGTCAGC	TCAAGCTTTCAATAGTCGTCCA	$(TGGA)_{10}$	55
PAN_38	GAAGCCAACAATATCCAATGC	TCGTATTCTCTCAGATAACCCTTTG	$(AC)_{17}$	55

F, forward primer; R, reverse primer.

Cross-species amplification was successful in the three African pangolin species using PCR conditions as described above. However, amplification was not achieved for the following markers: PAN_28 in *S. gigantea*, PAN_29 in *P. tetradactyla* and *P. tricuspis* and PAN_36 in *P. tricuspis*. In addition, 16 (53%) of the markers were identified as monomorphic, including PAN_2, PAN_3, PAN_5, PAN_6, PAN_7, PAN_10, PAN_11, PAN_12, PAN_13, PAN_15, PAN_18, PAN_20, PAN_24, PAN_28, PAN_30, PAN_32 and PAN_38 (Table 2.3).

Table 2.2. Characterization of 30 STRs in Temminck's pangolin, *S. temminckii*.

Locus	GenBank accession number	Size range (bp)	N_a	H_e	H_o	HWE	Null alleles
PAN_01	MN477255	69–118	5	0,813	0,552	***	None
PAN_02	MN477256	85–139	2	0,344	0,285	ns	None
PAN_03	MN477257	89–121	7	0,688	0,713	ns	None
PAN_05	MN477258	98–113	4	0,531	0,546	ns	None
PAN_06	MN477259	97–130	4	0,719	0,655	ns	None
PAN_07	MN477260	91–117	4	0,969	0,558	***	None
PAN_08	MN477261	88–114	5	0,438	0,576	ns	None
PAN_10	MN477262	84–106	3	0,313	0,366	ns	None
PAN_11	MN477263	119–187	5	0,844	0,734	ns	None
PAN_12	MN477264	89–122	5	0,531	0,418	ns	None
PAN_13	MN477265	87–123	5	0,742	0,748	ns	None
PAN_15	MN477266	105–117	3	0,313	0,382	ns	None
PAN_17	MN477267	131–187	12	0,625	0,729	***	None
PAN_18	MN477268	99–121	7	0,844	0,772	***	None
PAN_19	MN477269	105–189	4	0,774	0,674	ns	None
PAN_20	MN477270	96–146	8	0,688	0,688	ns	None
PAN_21	MN477271	97–129	2	0,156	0,144	ns	None
PAN_23	MN477272	111–122	3	0,313	0,389	***	None
PAN_24	MN477273	77–116	5	0,844	0,759	ns	None
PAN_26	MN477274	91–142	4	0,594	0,534	ns	None
PAN_27	MN477275	93–124	9	0,793	0,719	ns	None
PAN_28	MN477276	109–157	3	0,344	0,330	ns	None
PAN_29	MN477277	118–142	5	0,844	0,720	ns	None
PAN_30	MN477278	86–106	2	0,219	0,285	ns	None
PAN_31	MN477279	72–136	2	0,500	0,430	ns	None
PAN_32	MN477280	61–103	2	0,438	0,451	ns	None
PAN_35	MN477281	96–136	7	0,733	0,820	*	None
PAN_36	MN477282	63–111	2	0,000	0,117	***	None
PAN_37	MN477283	86–120	2	0,406	0,359	ns	None
PAN_38	MN477284	80–120	5	0,500	0,599	ns	None

N_a, number of alleles; H_o, observed heterozygosity; H_e, expected heterozygosity; HWE, significant deviation from Hardy–Weinberg proportions; ns, not significant. *P<0.05, *** P<0.001).

In total, 26 (87%) markers successfully amplified and were polymorphic for *S. gigantea* and 18 (60%) markers for *P. tetradactyla* and *P. tricuspis*, respectively. Thus in total, 11 markers were polymorphic across all four African pangolin species. Genetic diversity estimates based on these 11 markers (Table 2.4) varied between the four African pangolin species with *S. temminckii* (Na = 5; Ho = 0,559), *S. gigantea* (Na = 4,909; Ho = 0,514) and *P. tricuspis* (Na = 2,686; Ho = 0,541)

displaying higher diversity in comparison to *P. tetradactyla* (Na = 3; Ho = 0,242). These estimates however should be interpreted with caution due to differences in sample sizes of the populations.

2.4. DISCUSSION

The STRs presented here will contribute to estimates of genetic diversity, population genetic structure and relatedness studies in African pangolin species. Cross-species amplification was achieved for all three African pangolin species. However, the number of markers in which successful polymorphic amplification was achieved was higher in the two ground-dwelling species (*S. gigantea* and *S. temminckii*) in comparison to the arboreal species, which can be attributed to the relative divergence among the two genera (Chapter 3). Genetic diversity estimates were found to be low to moderate in African pangolins (0,168–0,544) in comparison to the Malayan pangolin (*Manis javanica*) (0,708; Luo *et al.* 2007). In conclusion, we developed and optimized 30 STR markers for Temminck's pangolin. We also confirm cross-species amplification in the other African pangolin species with a success ranging from 60–87%. To our knowledge this is the first study to describe STR markers of African pangolin species. This marker set may allow for the investigation of evolutionary history, population structure and genetic differentiation in African pangolins. Furthermore, these markers may inform conservation strategies and aid in the forensic investigation of confiscated scale samples from African pangolins.

Table 2.3. Cross-species amplification of 30 STRs for African pangolins.

Locus	*P. tetradactyla*		*P. tricuspis*		*S. gigantea*	
	Amplification	**HWE**	**Amplification**	**HWE**	**Amplification**	**HWE**
PAN_01	√	ns	√	ns	√	ns
PAN_02	M		M		√	*
PAN_03	M		√	***	√	ns
PAN_05	M		M		M	
PAN_06	√	ns	√	**	M	
PAN_07	√	***	M		√	**
PAN_08	√	**	√	**	√	ns
PAN_10	√	**	M		√	ns
PAN_11	M		√	ns	√	*
PAN_12	M		M		√	ns
PAN_13	M		√	ns	M	
PAN_15	M		√	*	√	ns
PAN_17	√	ns	√	ns	√	ns
PAN_18	√	**	M		√	ns
PAN_19	√	ns	√	*	√	ns
PAN_20	M		M		√	ns
PAN_21	√	ns	√	*	√	ns
PAN_23	√	ns	√	ns	√	ns
PAN_24	M		√	ns	√	***
PAN_26	√	ns	√	ns	√	ns
PAN_27	√	ns	√	ns	√	*
PAN_28	M		√	ns	√	
PAN_29	NA		NA		√	ns
PAN_30	√	ns	M		√	*
PAN_31	√	ns	√	ns	√	**
PAN_32	M		M		√	***
PAN_35	√	***	√	*	√	ns
PAN_36	√	ns	NA		√	ns
PAN_37	√	***	√	*	√	ns
PAN_38	√	ns	M		√	ns

HWE, Hardy–Weinberg equilibrium; ns, not significant; M, positive monomorphic amplification; √, positive polymorphic amplification; NA, no amplification; * P<0.05; ** P<0.01; *** P<0.001.

Table 2.4. Genetic diversity estimates across four African pangolin species based on 11 polymorphic loci.

Population	N_a	N_e	H_o	H_e	H_z
S. temminckii	5,000	2,612	0,559	0,539	0,547
P. tetradactyla	3,000	1,745	0,242	0,349	0,370
P. tricuspis	4,000	2,686	0,541	0,525	0,553
S. gigantea	4,909	3,053	0,514	0,600	0,633
Total	4,227	2,524	0,464	0,503	0,526

N_a, number of alleles; H_o, observed heterozygosity; H_e, expected heterozygosity; H_z, unbiased expected heterozygosity.

CHAPTER THREE

MITOCHONDRIAL GENOMES OF AFRICAN PANGOLINS AND INSIGHTS INTO EVOLUTIONARY PATTERNS AND PHYLOGENY OF THE FAMILY MANIDAE

3.1. INTRODUCTION

To date, full mitochondrial genomes of five pangolin species have been determined including *M. pentadactyla*, *M. javanica*, *S. temminckii*, *P. tetradactyla* and *P. tricuspis* (Arnason *et al*. 2002; Qin *et al*. 2012; Du Toit *et al*. 2014; Hari & Siew Woh 2015; Hassanin *et al*. 2015; Hari *et al*. 2016). However, two mitogenomes include misidentified Genbank records incorrectly accessioned as *M. pentadactyla* and *P. tetradactyla* that were noted in subsequent studies (Gaubert *et al*. 2015; Hassanin *et al*. 2015). Several techniques have been reported to generate whole mitochondrial genomes, however modern techniques such as next generation sequencing (NGS) using 454, Illumina and Ion Torrent technology have simplified and made sequencing mitogenomes from any eukaryotic DNA easier, quicker and more affordable compared to Sanger-based methods (Smith 2012; King *et al*. 2014; Seo *et al*. 2015). The vast suite of Bioinformatics software currently available facilitates the annotation and aids in analyses of large datasets (Smith 2015).

In general, a quarter of the reads generated by RNA/DNA sequence experiments are from mitochondrial genomes (Raz *et al*. 2011; Smith 2012, 2013, 2015) which may be attributed to their high copy numbers as well as their high expression levels. Due to the AT richness of mtDNA, as well as it being polyadenylated it can contribute to an increase in poly-A RNA selection (Raz *et al*. 2011). Assembling mitochondrial genomes are significantly less complex than their nuclear genome counterparts as they are smaller in size, and harbour fewer genes (Smith 2015). The mitogenomes of two Asian pangolin species (*M. pentadactyla* and *M. javanica*) have been assembled using Illumina HiSeq technology, whereby the authors extracted mitochondrial

sequences from nuclear data obtained from NGS techniques (Hari & Siew Woh 2015; Hari *et al*. 2016).

Current phylogenetic assessments of pangolins have been conducted using only two of the four African pangolin species namely; Temminck's pangolin and white-bellied pangolin (Du Toit *et al*. 2014; Hassanin *et al*. 2015). In addition, the current genus-level classification of pangolins is still under debate. Thus, in this study we performed next generation sequencing for all four African pangolins using the Illumina HiSeq 2500 in order to reconstruct complete mitochondrial genomes. Here we present the first whole mitochondrial DNA genomes of two of the African pangolin species; the black-bellied pangolin (*P. tetradactyla*) and the giant pangolin (*S. gigantea*). In addition, we describe the mitochondrial genome features in order to understand the evolutionary forces shaping the mitochondrial genomes of African pangolins. Lastly, we conduct a phylogenetic assessment in order to provide a genus-level classification of African pangolins.

3.2. MATERIALS AND METHODS

3.2.1. SAMPLE COLLECTION AND DNA ISOLATION

This study used six deceased individuals, sampled by the African Pangolin Working Group (APWG) and representing the four African pangolin species. Tissue samples were placed in absolute ethanol and were stored at the National Biobank, National Zoological Gardens of South Africa (NZG), at –80ºC until analysis. The samples were from three Temminck's pangolins (*S. temminckii*; MF536685–MF536687) from South Africa, one white-bellied pangolin (*P. tricuspis*; MF536683) and one black-bellied pangolin (*P. tetradactyla*; MF509825), both from Ghana (Boakye *et al*. 2016). In addition, a giant pangolin (*S. gigantea*; MF536684) scale sample was included from an illegal seizure. The species identity of samples used in this study was confirmed with Sanger sequencing of the *CoI* and *Cytb* loci which were compared to chain-of-custody voucher specimens available from the NZG species reference database (Mwale *et al.* 2017) (see www.barcodewildlife.org). All voucher specimens were verified and identified by an acknowledged authority (Raymond Jansen; African Pangolin Working Group).

DNA was isolated using the QIAamp Micro Kit (QIAGEN, Novato, CA, U.S.A.) and the respective manufacturers' protocol for tissue was followed. DNA was quantified on the Qubit v3.0 Fluorometer (Thermo Scientific, Massachusetts, U.S.A.). Polymerase Chain Reaction (PCR) amplification and sequencing, to verify species identity, were performed as outlined in (Mwale *et al.* 2017).

3.2.2. NEXT-GENERATION SEQUENCING AND ASSEMBLY

The products were run on an Illumina HiSeq 2500 (Illumina Incorporated, San Diego, CA, U.S.A.) using a rapid run and the TruSeq DNA LT Sample Prep Kit (Illumina Incorporated, San Diego, CA, U.S.A.). Data quality was evaluated using FastQC v0.11.2 (Andrews 2010) software, and trimmed and edited through Trimmomatic v0.36 (Bolger *et al*. 2014) to remove the adapters and poor quality sections. Mitogenomes were assembled in CLC Genomics Workbench v6 (https://www.qiagenbioinformatics.com; CLC Bio, Aarhus, Denmark) using *De Novo* alignment, with paired reads. Sequence identity of contigs was validated by performing a BLAST search on the National Centre for Biotechnology Information (NCBI) website (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

3.2.3. MITOGENOME ANNOTATION AND PHYLOGENETIC ANALYSIS

The mitogenomes were annotated with MITOS v806 (Bernt *et al*. 2013) and a circular alignment between the six available pangolin species were drawn in Circos v0.69 (Krzywinski *et al*. 2009). The GC content of the four African pangolin mitogenomes were calculated using GPMiner (Lee *et al*. 2012) with a sliding window of 300 bp. Arlequin v3.5.1 (Excoffier & Lischer 2010) was used to validate the GC scores obtained for the four mitogenomes using ANOVA analysis and the diagrams were plotted in R v3.3.1 (Development Core Team 2008).

The mitogenomes generated in this study comprised six animals from four African pangolin species and were combined and aligned with 11 other genomes using MAFFT v7 (Katoh & Standley 2013) (Table 3.1). The mitogenomes of *Acinonyx jubatus* (Burger *et al*. 2004), *Crocuta crocuta* (Bon *et al*. 2012), *Canis lupus* (Koblmüller *et al*. 2016) *and Arctocephalus pusillus* (Arnason *et al*. 2006) were used as an out-group as the order Pholidota (pangolins) is reported to be evolutionary closer to carnivores (Du Toit *et al*. 2014).

The phylogenetic program jModeltest v2.1.7 (Darriba *et al*. 2012) was used to determine the best fit model of sequence evolution, under the Akaike Information Criterion (AIC) (Akaike 1974), Bayesian Information Criterion (BIC) (Hirose *et al*. 2011) and Decision Theory Performance-Based Selection (DT) (Minin *et al*. 2003). Partition analysis was also implemented using the program PartitionFinder v2 (Lanfear *et al*. 2017) to determine the best fit models of evolution for the different loci in the dataset. The partition was run using linked branch lengths and a greedy search for the models under the AIC.

Phylogenetic analysis was conducted using MrBayes v3.2.6 (Ronquist *et al*. 2012) to infer relationships between the different species using Bayesian Inference (BI). The parameters used for MrBayes were 2 000 000 generations after which 25% of the trees were discarded as burn-in. A Maximum Likelihood (ML) tree was constructed utilizing PhyML v3 (Guindon *et al*. 2010) with the same models used for the Bayesian analysis and was run with 10 000 Bootstrap replications. Individual phylogenetic trees, for each loci, were also created with MrBayes v3.2.6 (Ronquist *et al*. 2012) and PhyML v3 (Guindon *et al*. 2010).

3.2.4. CODON USAGE ANALYSIS FOR AFRICAN AND ASIAN SPECIES

The Relative Synonymous Codon usage (RSCU) values for mitochondrial genes were established using the Mega v7 (Kumar *et al*. 2016a) software. This was performed on the four African pangolin species evaluated in this study as well as for the previously published *M. pentadactyla* (KT445978.1) and *M. Javanica* (KT445979.1). The Principle Component Analysis (PCAs) generated from this data was performed using the FactoMineR package in R (Lê *et al*. 2008). Codon usage

bias (AT3 and GC3 content) was calculated using Mega v7, for each of the protein coding genes, where the A and T values at the third base were summed for the AT3 value. The same was performed with G and C for the GC3 content. The ratios were reported as percentages.

Table 3.1. List of 17 mitogenomes used in the study presented here. The common name, scientific name, Genbank accession number and reference were noted for each individual. [1] = Misidentified Chinese pangolin genome; [2] = Misidentified Black-bellied pangolin genome.

Common Name	Scientific Name	Genbank Accession Number	Reference
Cheetah	*Acinonyx jubatus*	AY463959.1	(Burger *et al*. 2004)
Spotted Hyena	*Crocuta crocuta*	JF894378.1	(Bon *et al*. 2012)
Grey Wolf	*Canis lupus*	KU696410.1	(Koblmüller *et al*. 2016)
Brown Fur Seal	*Arctocephalus pusillus*	NC_008417.1	(Arnason *et al*. 2006)
Chinese Pangolin[1]	*Manis pentadactyla*	JN411577.1	(Qin *et al*. 2012)
Chinese Pangolin	*Manis pentadactyla*	KT445978.1	(Hari *et al*. 2016)
Malayan Pangolin	*Manis javanica*	KP306515.1	(Hassanin *et al*. 2015)
Malayan Pangolin	*Manis javanica*	KT445979.1	(Hari & Siew Woh 2015)
Black-Bellied Pangolin[2]	*Phataginus tetradactyla*	AJ421454.1	(Arnason *et al*. 2002)
White-Bellied Pangolin	*Phataginus tricuspis*	KP306514.1	(Hassanin *et al*. 2015)
Temminck's pangolin	*Smutsia temminckii*	KP125951.1	(Du Toit *et al*. 2014)
Temminck's pangolin	*Smutsia temminckii*	KP306516.1	(Hassanin *et al*. 2015)
Black-Bellied Pangolin	*Phataginus tetradactyla*	MF509825	Current Study
White-Bellied Pangolin	*Phataginus tricuspis*	MF536683	Current Study
Giant Pangolin	*Smutsia gigantea*	MF536684	Current Study
Temminck's pangolin	*Smutsia temminckii*	MF536685	Current Study
Temminck's pangolin	*Smutsia temminckii*	MF536686	Current Study
Temminck's pangolin	*Smutsia temminckii*	MF536687	Current Study

3.2.5. CONFIRMATION OF INSERTIONS OBSERVED IN THE CONTROL REGION OF THE PANGOLIN MITOGENOME

Sanger sequencing of the control region of the mitochondrial genome was performed using five additional samples from each of the African pangolin species. The white-bellied and black-bellied pangolin samples were from Ghana and the Temminck's pangolin samples from South Africa and Tanzania (Hassanin *et al*. 2015). The giant ground pangolin samples were obtained from the collection of the Zoological

Museum, University of Copenhagen. The protocol and cycle conditions outlined in (Du Toit *et al.* 2014) were used for all the samples. Sequencing was conducted in order to verify the presence of insertions in the D-loop observed in the mitogenomes obtained from next-generation sequencing during this study. A sequence fragment of around 500 bp was targeted using the primer pair:
PNG_Dloop forward 5'-CGTTCCTCTTAAATAAGACATCTCG-3' and reverse 5'-TCTTGCTTTTGGGGTTTGAC-3' for verification.

3.3. RESULTS AND DISCUSSION

3.3.1. NEXT-GENERATION SEQUENCING

The HiSeq rapid run resulted in approximately 22 000 000 reads per sample, with an average read length of 250 nucleotides. These reads were used for a *De Novo* assembly of each sample (CLC Bio version 6.0). This resulted in 10 207 contigs for *P. tricuspis* with the largest contig being 16 565 nucleotides, consisting of 78 099 reads at an average coverage of 986x. For *P. tetradactyla*, there were 2 801 contigs with the largest contig, 16 649 nucleotides consisting of 47 686 reads at an average coverage of 369x. For *S. gigantea*, there were 11 346 contigs with the largest contig, 16 540 nucleotides consisting of 13 076 reads and an average coverage of 98x. For the three *S. temminckii* samples (MF536685–MF536687), contigs ranged from 2 742–5 560. The largest contig in all three samples was 16 558 nucleotides consisting of 63 759; 29 702 and 6 820 reads and an average coverage of 494x; 248x and 53x respectively.

The contigs were identified as the mitogenomes of the pangolin species based on the (i) estimated length (≈16,5 kb); (ii) the occurrence of the proteins *Cox*, *Cytb*, NADH dehydrogenase V (*NadV*) and NADH dehydrogenase VI (*NadVI*) (Table 3.2) and (iii) correspondence with mitochondrial sequences from other Pholidota based on NCBI BLAST searches.

Table 3.2. List of mitochondrial genes and loci, indicating size in base pairs from four African pangolin species, *Smutsia gigantea*, *S. temminckii*, *Phataginus tricuspis* and *P. tetradactyla*.

Gene Regions	***S. gigantea*** **(Giant ground pangolin)**	***S. temminckii*** **(Temminck's pangolin)**	***P. Tricuspis*** **(White-bellied pangolin)**	***P. tetradactyla*** **(Black-bellied pangolin)**
Mitogenome (bp)	16,540	16,558	16,565	16,649
12S Ribosomal RNA (rRNA)	960	959	958	958
16S Ribosomal RNA (rRNA)	1,560	1,556	1,555	1,561
NADH dehydrogenase I (*NadI*)	951	945	945	945
NADH dehydrogenase II (*NadII*)	1,038	1,038	1,038	1,038
Cytochrome c oxidase I (*CoxI*)	1,536	1,533	1,536	1,515
Cytochrome c oxidase II (*CoxII*)	681	681	681	681
ATP synthase VIII (*AtpVIII*)	195	195	198	192
ATP synthase VI (*AtpVI*)	675	675	675	675
Cytochrome c oxidase III (*CoxIII*)	783	783	783	783
NADH dehydrogenase III (*NadIII*)	345	345	345	345
NADH dehydrogenase IV-L (*NadIV-L*)	294	294	294	294
NADH dehydrogenase IV (*NadIV*)	1,371	1,368	1,371	1,368
NADH dehydrogenase V (*NadV*)	1,791	1,788	1,794	1,803
NADH dehydrogenase VI (*NadVI*)	519	516	519	519
Cytochrome b (*Cob*)	1,134	1,134	1,134	1,134
Control region (D-loop)	1,135	1,155	1,167	1,265

3.3.2. GENOMIC ORGANISATION

The mitogenome of the *S. temminckii* samples consisted of 16 558bp while *S. gigantea* was 16 540 bp; *P. tetradactyla* was 16 649 bp and *P. tricuspis* was 16,565bp in length (Table 3.2, Figure 3.1). The light and heavy strands each contain their own arrangement of genes, proteins or loci respectfully located on each strand (Figure 3.1).

The heavy strand, or plus strand, comprises of the following loci: two Ribosomal RNAs (12S rRNA, 16S rRNA); 12 Protein-coding genes (*NadI, NadII*, *CoxI*, *CoxII*, *AtpVIII*, *AtpVI*, *CoxIII*, *NdIII, NadIV-L*, *NadIV*, *NadV*, *Cob*) and 14 Transfer RNAs (trnF, trnV, trnL2, trnI, trnM, trnW, trnD, trnK, trnG, trnR, trnH, trnS1; trnL1, trnT). The light or minus strand comprises one Protein-coding gene (*NadVI*) and eight Transfer RNAs (trnQ, trnA, trnN, trnC, trnY, trnS2, trnE, trnP). As indicated in Table 3.2, the

mitogenome of the four African pangolins varied in terms of gene region size at several loci.

Six regions [cytochrome oxidase II (*CoxII*), ATP synthase VI (*AtpVI*), cytochrome oxidase III (*CoxIII*), NADH dehydrogenase III (*NadIII*), NADH dehydrogenase IV-L (*NadIV-L*), cytochrome b (*Cob*)] were found to be the same length in all four species and three loci [16S Ribosomal RNA, NADH dehydrogenase V (*NadV*), D-loop] each have different lengths in each of the four species. The remaining loci [12S Ribosomal RNA, NADH dehydrogenase I (*NadI*), NADH dehydrogenase II (*NadII*), cytochrome oxidase I (*CoxI*), ATP synthase VIII (*AtpVIII*), NADH dehydrogenase IV (*NadIV*), NADH dehydrogenase VI (*NadVI*)] had lengths that were generally consistent, with some pangolin species showing variation in length in comparison to the other species (Table 3.2).

3.3.3. PHYLOGENETIC ANALYSIS OF AFRICAN PANGOLINS

The best fit model of sequence evolution for the dataset under the AIC was the General Time Reversal model (GTR+I+G) with invariable site and gamma distribution of 0,822 (Lanave *et al*. 1984; Rodríguez *et al*. 1990). The best fit model under the BIC and DT was the Transition model two (TIM2+I+G) with invariable site and gamma distribution values of 0,998 and 0,990 respectively (Posada 2003).

From these two models three Bayesian phylogenetic trees along with three maximum likelihood trees were generated for the datasets. The different trees all showed consistent branching patterns, posterior probability values for the BI trees and bootstrap values for the ML trees. The trees were subsequently concatenated into a single consensus tree with the different support values indicated on the respective branches (Figure 3.2). Partition analysis indicated a variety of models for the individual loci and was subjected to individual BI and ML analysis to confirm the results of the whole mitochondrial data.

In the phylogenetic tree, using all available pangolin mitogenomes, it is evident that all the African pangolins group according to species, with the exception of the black-bellied pangolin genome (AJ421454.1) (Arnason *et al*. 2002). The latter mitogenome

has previously been reported to be misclassified based on partial *Cob* analysis (Gaubert *et al*. 2015; Hassanin *et al*. 2015) and was confirmed in this analysis as a white-bellied pangolin genome (Supp. Figure 3.1). In addition, the misclassified *M. pentadactyla* (JN411577.1) (Qin *et al*. 2012) grouped with *M. javanica* (Hassanin *et al*. 2015; Hari *et al*. 2016), confirming an error also reported in previous studies (Add. Figure 3.1).

The phylogenetic tree (Figure 3.2) which excluded the misclassified samples provided support for the Asian and African pangolin species separation into two distinct monophyletic clades with the latter consisting of all African pangolins species, *P. tricuspis, P. tetradactyla, S. temminckii* and *S. gigantea*. Within the African clade the giant and Temminck's pangolin clustered separately from the white-bellied and black-bellied pangolins with significant Bayesian and ML support (Posterior Probabilities of 1). For the African pangolin species, the observed branching pattern thus provides support for the classification of the ground-dwelling and arboreal species into two separate genera; *Phataginus* and *Smutsia*. In addition, results from this analysis suggests the overall classification of pangolin into three genera; *Manis* (Asian pangolins), *Phataginus* (African tree pangolins) and *Smutsia* (African ground pangolins). However, further analysis should be undertaken for Asian pangolins to include the full mitochondrial genomes of the Philippine (*M. culionensis*) and Indian pangolin (*M. crassicaudata*).

The above branching patterns were also confirmed with the individual loci from both BI and ML trees consisting of a variety of models. The control region, rRNAs, light strand proteins, light strand tRNAs and heavy strand proteins (exclusive of CoxII) were all concurrent with the whole mtDNA tree for both BI and ML results with high support values. The heavy strand tRNAs showed differentiation in the tree pangolins and again in the ground pangolins of Africa for both BI and ML trees. Although the internal grouping differs, they still conform to the same three genera identified in the trees above namely Smutsia, Phataginus and Manis. The ML analysis for the heavy strand protein CoxII was in accordance with the results above. However, the BI analysis indicated that the black-bellied pangolin (MF509825) branched separately from the African pangolins, but still formed a monophyletic group with the African pangolin species adjacent to the Asian pangolins. Overall, the majority of the

individual loci subject to different evolutionary rates and models along with a variety of phylogenetic analysis concurred with the results obtained from the whole mtDNA data.

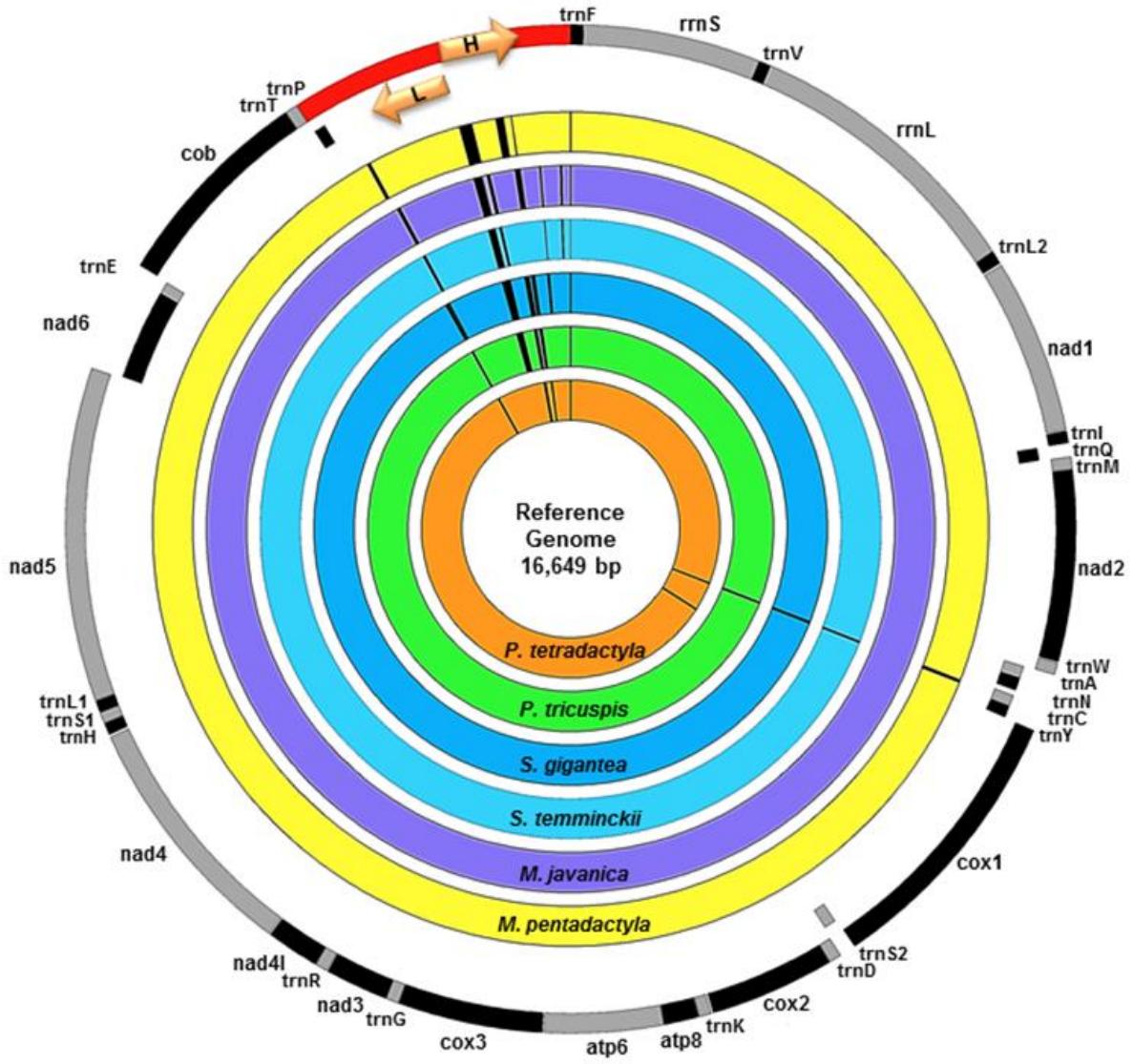


Figure 3.1. Circular diagram of six pangolin mitogenomes. The six coloured circles represent the six different pangolin species (four African and two Asian) aligned to each other to indicate the differences between individuals with the reference genome represented by *P. tetradactyla*. The outer ring represents the annotated loci located on the plus/heavy strand. The second ring represents the loci located on the minus/light strand of the mitogenome. Arrows are representative of the direction of the light and heavy strands; the heavy strand is located clockwise and the light strand anti-clockwise.

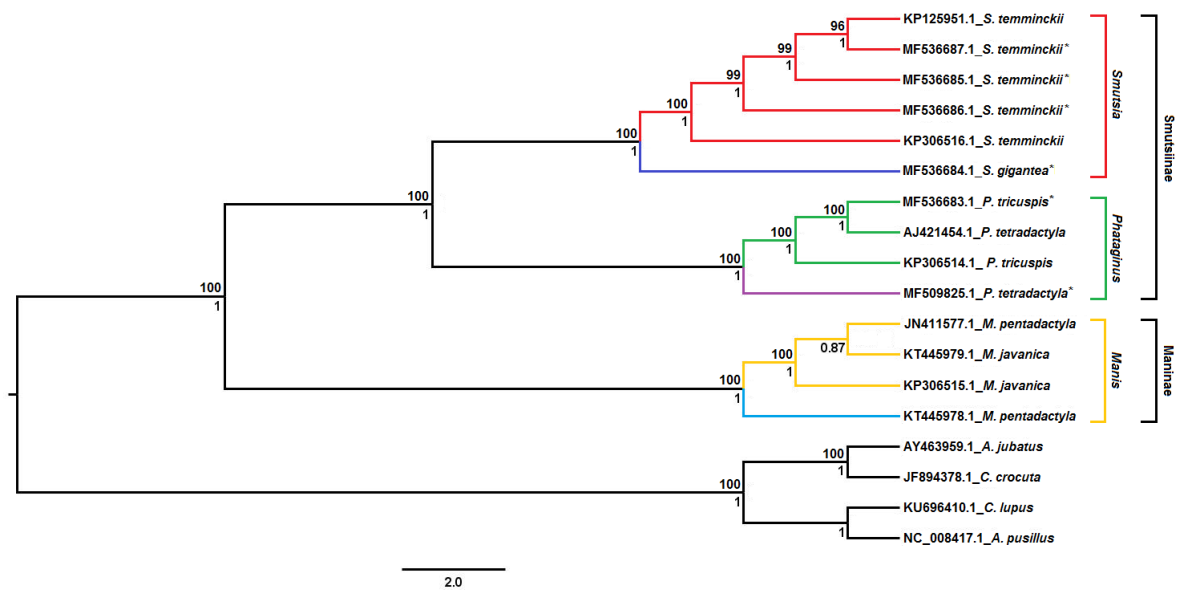


Figure 3.2. Combined Bayesian Inference (BI) and Maximum Likelihood (ML) tree of pangolin species. Bayesian Posterior Probabilities are indicated on the bottom of each node whereas the Maximum Likelihood Bootstrap values are indicated on top of the node. Only Bootstrap values equal or greater than 70% (≥70%) were noted on the tree. The mitogenome of *Acinonyx. jubatus*, *Crocuta crocuta*, *Canis lupus* and *Arctocephalus pusillus* was selected as outgroups. * indicates the six mitogenomes sequenced during this study.

3.3.4. ANALYSIS OF GC CONTENT AND CODON USAGE

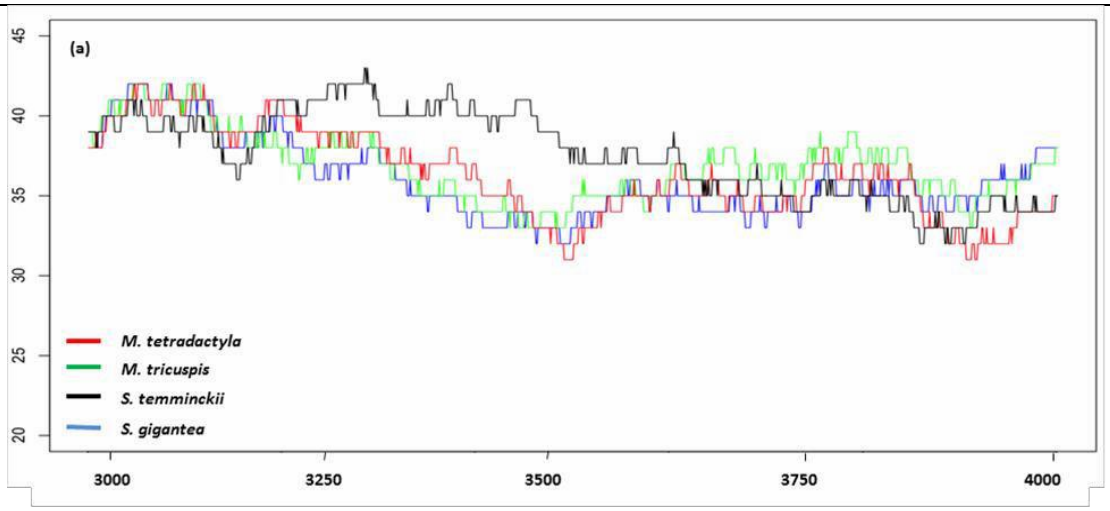
Total GC content of the African pangolin species was observed to be similar between species (*P. tetradactyla* = 36,5%, *S. gigantea* = 36,9%, *S. temminckii* = 37,3% and *P. Tricuspis* = 36,7%). These results confirm an AT-bias that has been reported in several other mammal species (Dou *et al*. 2016). Analysis of codon usage and pattern of mitochondrial genes; *AtpVI*, *AtpVIII*, *Cob*, *CoxI*, *CoxII*, *NadI*, *NadII*, *NadIII*, *NadIV*, *NadIV-L*, *NadV* and *NadVI* provided evidence of bias in terms of the use of codons with A and C occurring most frequently at the third codon (Supp. Table 3.1). Variation in base compositions within and among species has been suggested to occur as a result of two evolutionary processes namely biases in the process of mutation and/or natural selection (Mooers & Holmes 2000). Selective nucleotide compositional biases have been reported in chiropteran mitochondrial genomes (Meganathan *et al*. 2012). Uddin and Chakraborty (Uddin & Chakraborty 2014) similarly observed A or C as the most frequent codon at the 3rd position in a study of mitochondrial *AtpVI* in a variety of mammalian species. The authors

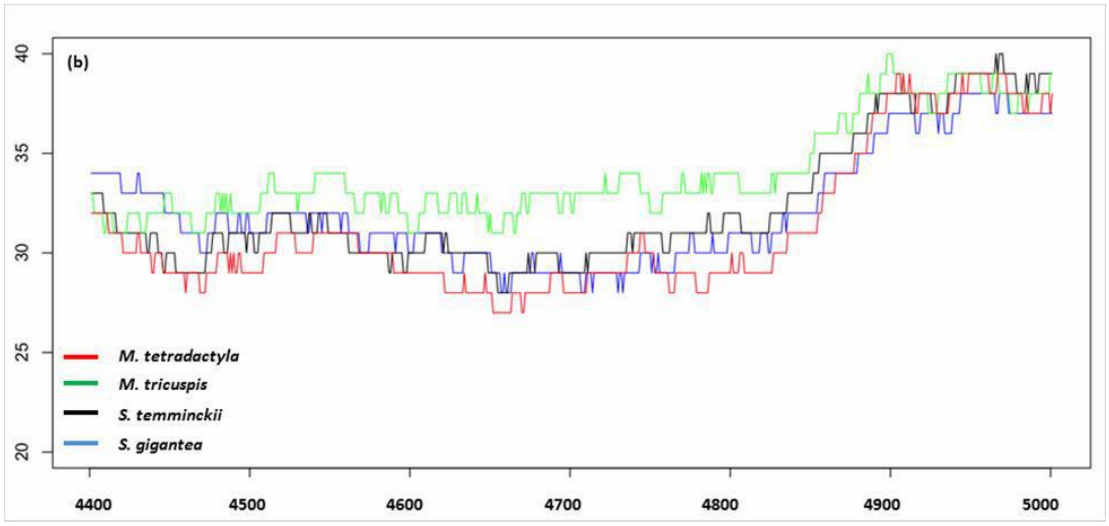
attributed this bias to mutational pressure that can influence codon usage bias in mitochondria.

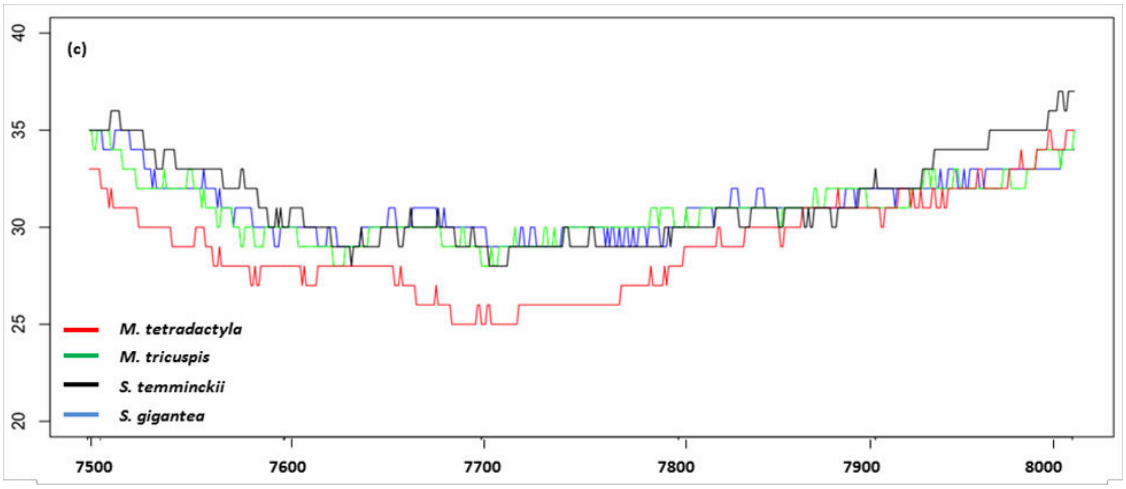
3.3.5. GC CONTENT AND CODON USAGE VARIATION BETWEEN PANGOLIN SPECIES

The percentage of GC content and codon frequencies calculated for five regions (3 100–3 700 bp, 4 500–4 900 bp, 7 500–7 900 bp, 9 700–9 900 bp and 15 000–16 000 bp) of the mitogenome was significantly different between the four African pangolin species based on ANOVA analysis (Figure 3.3 a-e). The respective genes that correspond to these regions include *NadI*, *NadII*, *CoxII*, *NadIII* and *Cob*. Analysis of codon usage of three mitochondrial genes; *CoxI*, *NadI* and *NadIII* reveals a clear distinction between the African and Asian pangolins, as well as within the African clade (Supp. Figure 3.2 a-c).

The PCA plots for these three genes therefore identified a connection between codon usage and phylogeny and provide further support for the phylogenetic analysis based on the whole mitochondrial genome. The PCA plots of the remaining protein coding genes (*AtpVIII*, *NadIV* and *NadVI*), whilst not achieving the same resolution within the African clade, show a clear separation of the Asian and African pangolins (Supp. Figure 3.2 d-f). Combined codon usage patterns of mitochondrial genes (*AtpVI*, *AtpVIII*, *Cob*, *CoxI*, *CoxII*, *NadI*, *NadII*, *NadIII*, *NadIV*, *NadIV-L*, *NadV* and *NadVI*) were plotted in order to perform a hierarchical clustering of each species to investigate the role of codon bias in the evolution of African pangolins. The resulting dendrogram is presented in Figure 3.4. African pangolin species varied in the percentage of codon bias with Asian pangolins displaying a lower degree of AT3 bias. Variations in GC content and codon frequencies between pangolin species may indicate that two selective forces; mutational pressure and/or natural selection may play important roles in the molecular evolution of pangolins with different evolutionary forces acting to shape the mitochondrial genomes of the African and Asian pangolin species.


(a)
M. tetradactyla
M. tricuspis
S. temminckii
S. gigantea
20
25
30
35
40
45
3000
3250
3500
3750
4000


(b)
M. tetradactyla
M. tricuspis
S. temminckii
S. gigantea
20
25
30
35
40
4400
4500
4600
4700
4800
4900
5000


(c)
M. tetradactyla
M. tricuspis
S. temminckii
S. gigantea
20
25
30
35
40
7500
7600
7700
7800
7900
8000

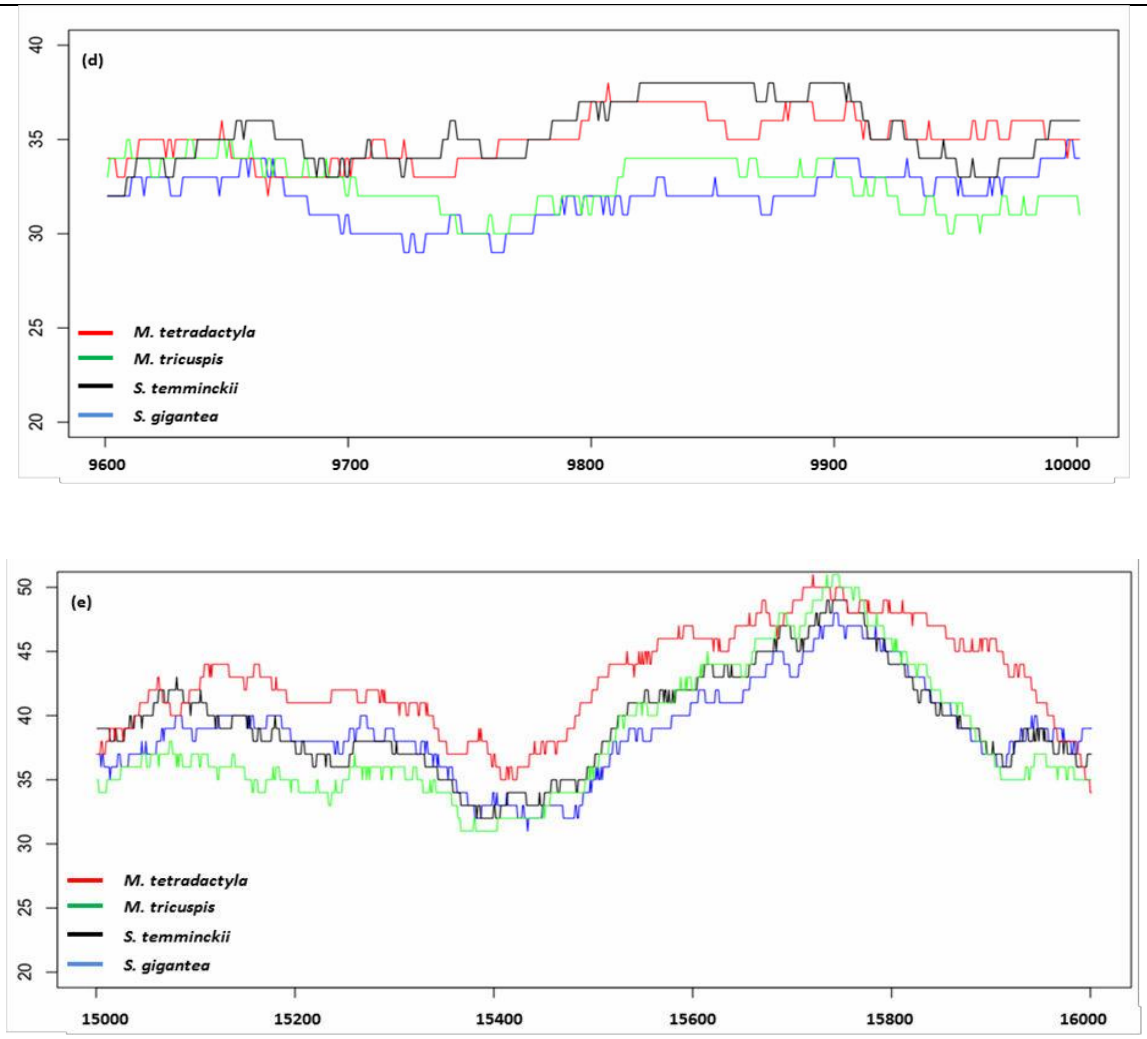


Figure 3.3. Representation of regions which display significant differences in terms of GC content among African pangolin species (*a*-*e*). Image (*a*) showing increased GC content (~3 100–3 700 bp) in *S. temminckii*; (*b*) showing an increased GC content (~4 500–4 900 bp) in *M. tricuspis*; (*c*) showing a decreased GC content (~7 500–7 900 bp) in *M. tetradactyla*; (*d*) showing an increased GC content (~9 700–9 900 bp) in *M. tetradactyla* and *S. temminckii*; (*e*) showing an increased GC content (~15 000–16 000 bp) in *M. tetradactyla.*

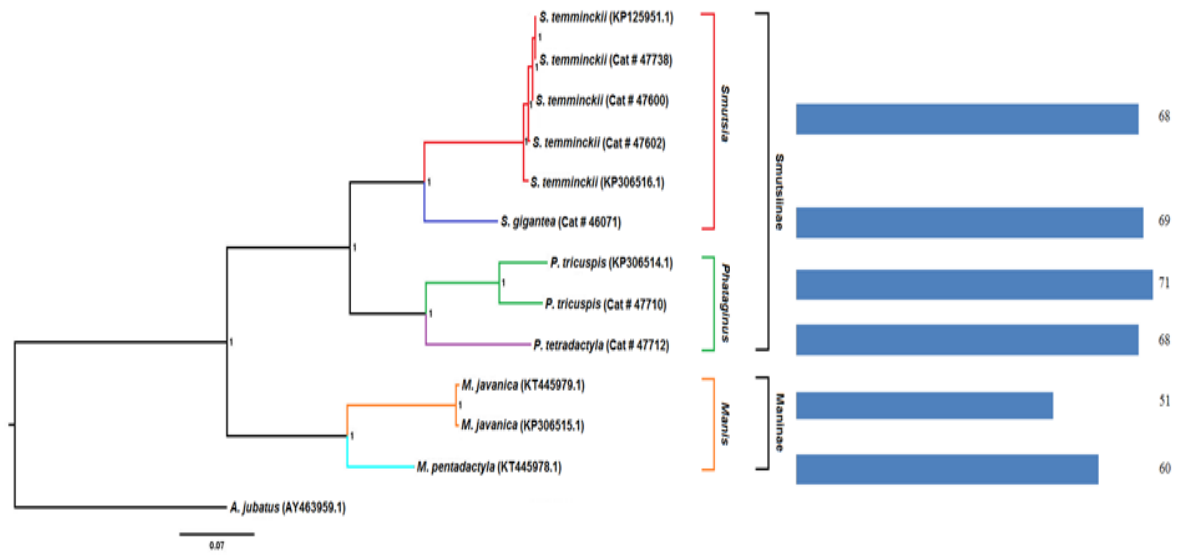


Figure 3.4. Clustering of pangolin species according to the variation of codon usage and phylogeny. The dendrogram was inferred by hierarchical clustering with the AT3 frequency metric for each species. The value of the AT3 is indicated as a percentage.

3.3.6. MITOGENOME COMPARISON BETWEEN *SMUTSIA* AND *PHATAGINUS*

Two insertions (80 bp and 28 bp in length) in the D-loop region of the black-bellied pangolin were observed (Figure 3.5) which was absent in the other African pangolin species. This insertion was validated with Sanger sequencing that included additional African pangolin species. When compared to the Asian pangolin species the first insertion was slightly shorter in length (62 bp).

Length variation in the D-loop has been reported in various species including bats (Wilkinson *et al*. 1997), rodents (Larizza *et al*. 2002) and primates (Saitou & Ueda 1994). In addition, substantial nucleotide sequence differences within and between species have been identified in the D-loop (Aquadro & Greenberg 1983; Foran *et al*. 1988). Lastly, heteroplasmy where individuals had more than one mtDNA form due to variation in numbers of tandem repeats in D-loop has been reported in shad (Bentzen *et al*. 1988), sturgeon (Buroker *et al*. 1990), whiptail lizards (Moritz & Brown 1987) and rabbits (Mignotte *et al*. 1990).

Length variation has been proposed to occur via four different mechanisms including illegitimate elongation (Buroker *et al*. 1990), intra- and intermolecular recombination

(Rand & Harrison 1989), transposition (Rand & Harrison 1989) and slipped miss pairing (Streisinger *et al*. 1966). The identification of the two insertions in the D-loop region may indicate that this region is under selection in black-bellied pangolins, which may demonstrate drift following the initial mutation event. However, further analysis among closely related species should be conducted in order to determine how selection impacts on the length and sequence variation within this region.

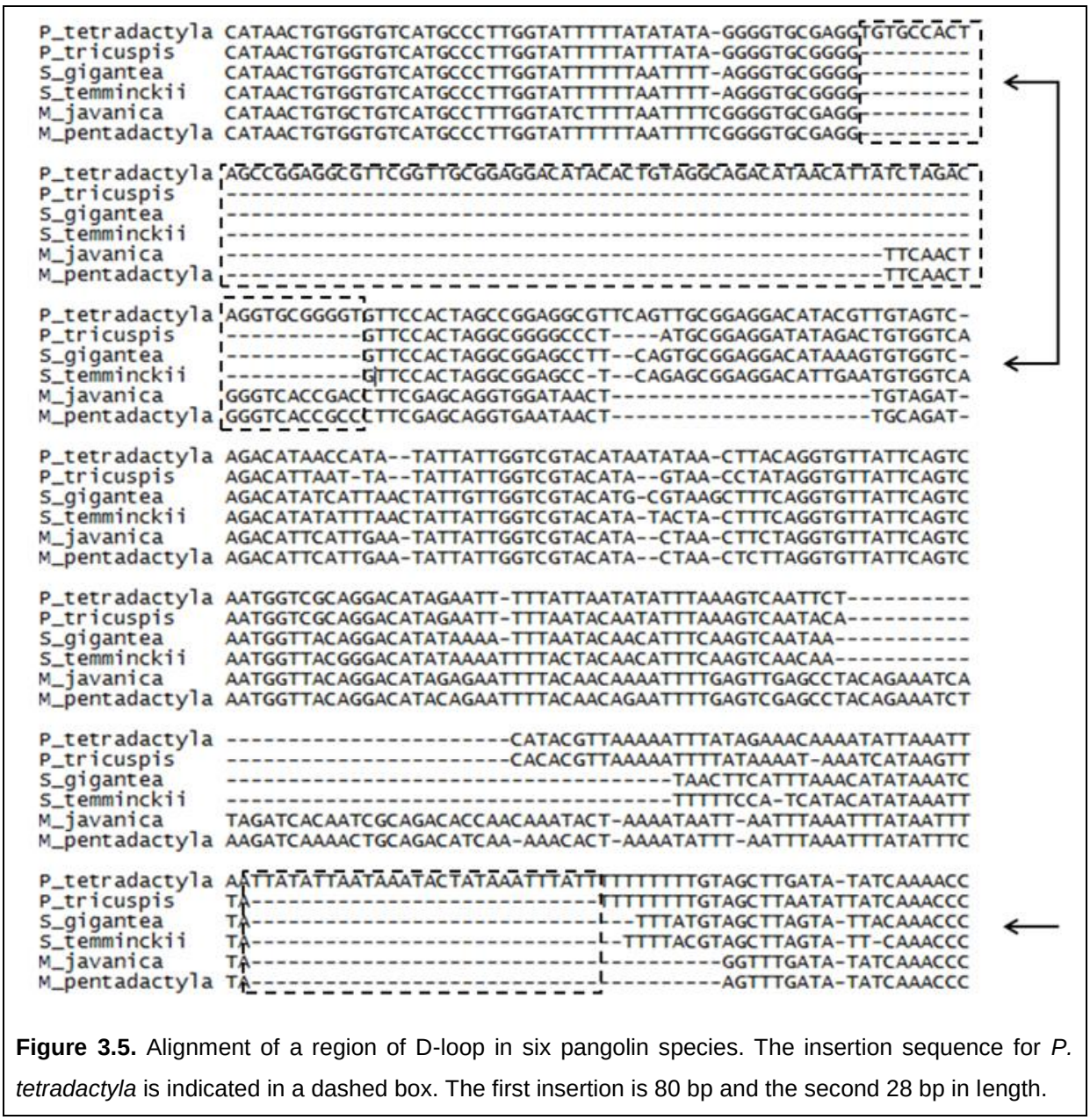
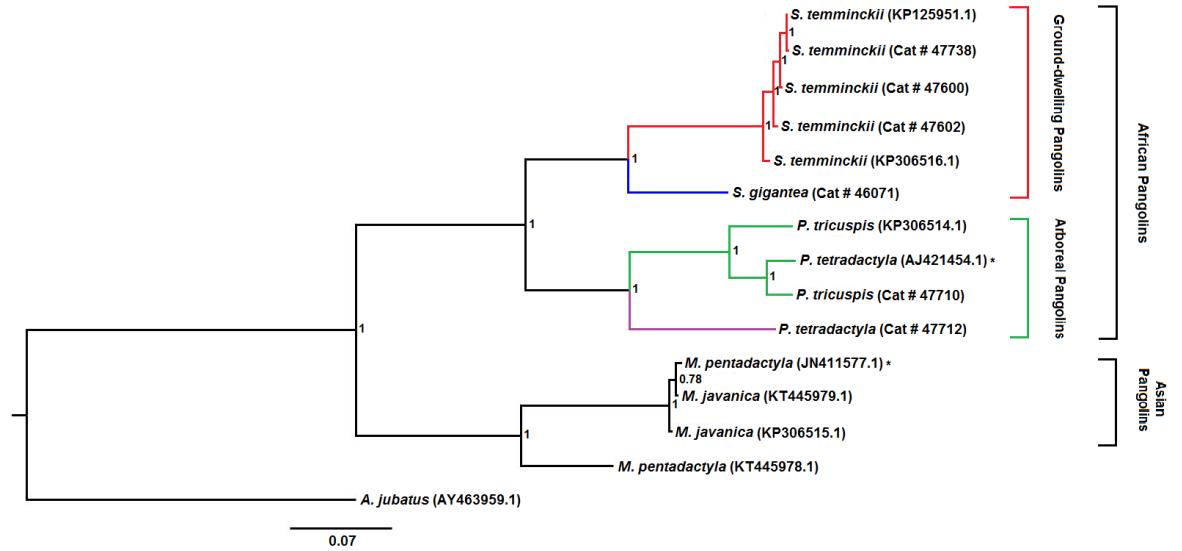


Figure 3.5. Alignment of a region of D-loop in six pangolin species. The insertion sequence for *P. tetradactyla* is indicated in a dashed box. The first insertion is 80 bp and the second 28 bp in length.

3.4. SUPPLEMENTARY FIGURES AND TABLE

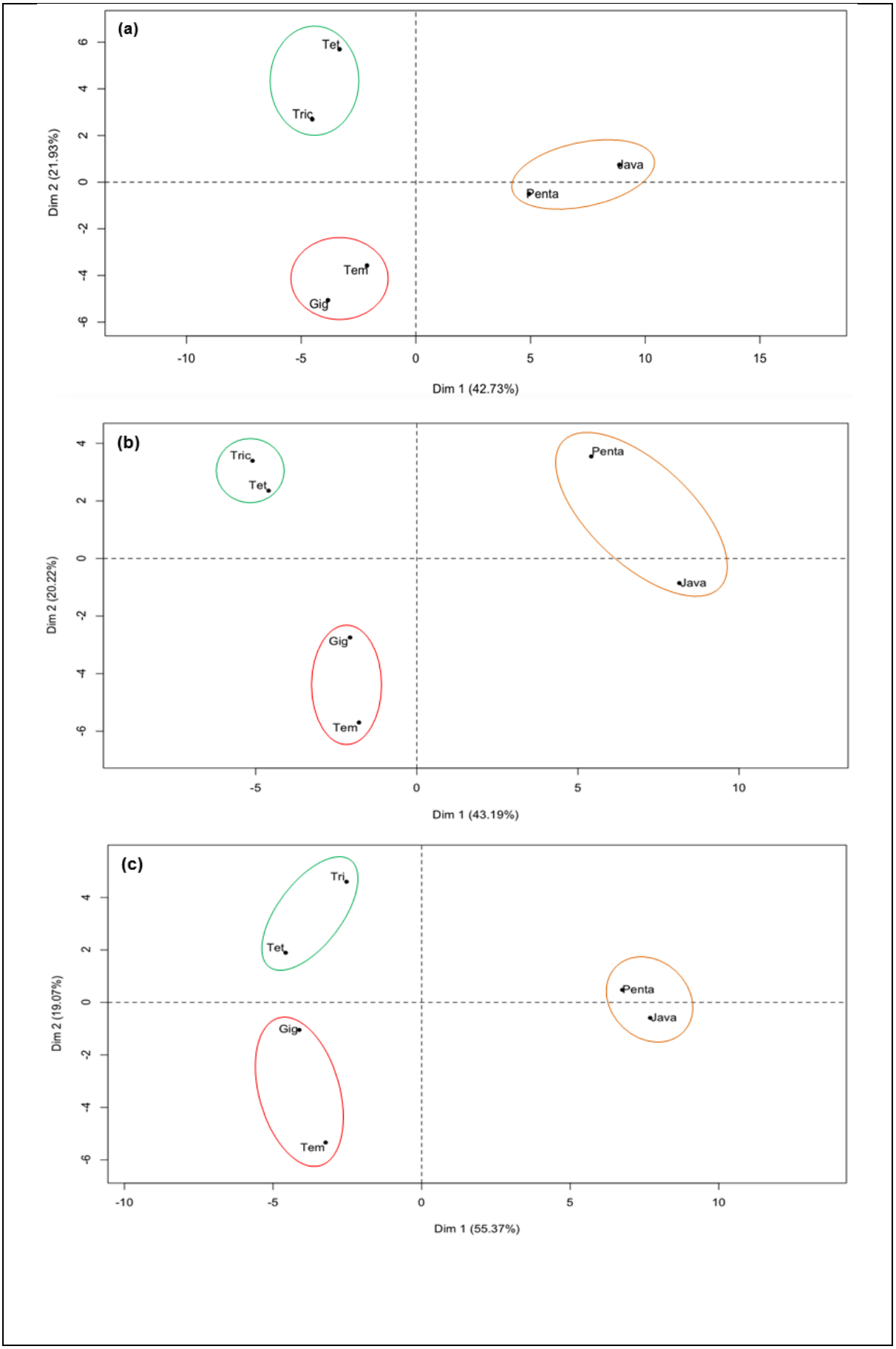


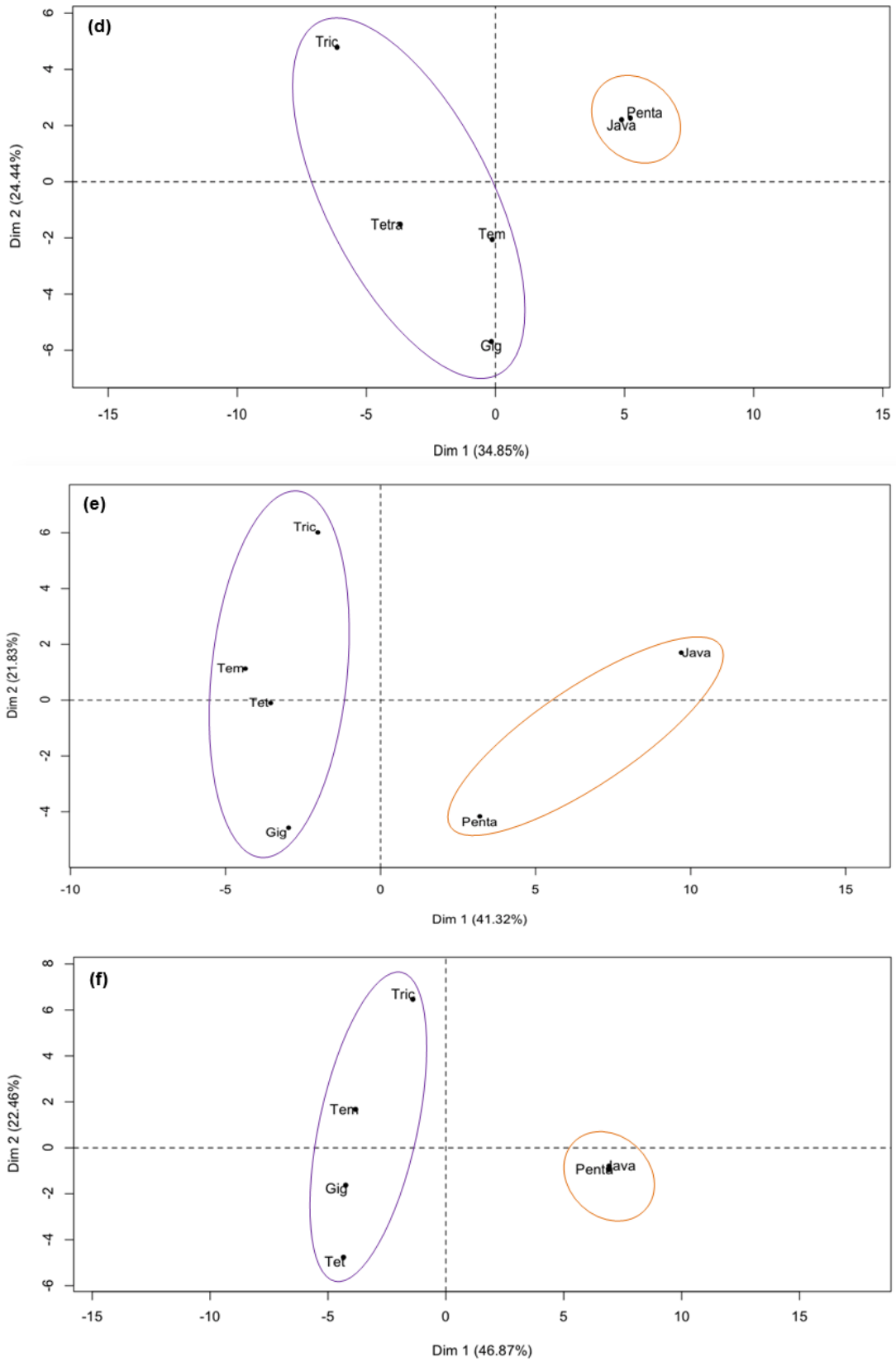
Supplementary Figure 3.1. Bayesian mitogenome phylogenetic tree of all available pangolin mitogenomes. Posterior probabilities are indicated on the respective branches. *A. jubatus* was selected as an outgroup as pangolins are more closely related to the order Carnivora. Asterisks indicate the misidentified mitogenomes.

Supplementary Table 3.1: List of nucleotide percentages and its 3rd codon position percentage (%).

Species	% of A	A3%	% of T	T3%	% of G	G3%	% of C	C3%
				AtpVI				
M. pentadactyla	34.4	46.7	26.4	20	9.8	2.6	29.4	30.8
M. javanica	31.6	38.3	23.6	12	11.9	8.8	32.9	41.0
P. tricuspis	32.4	40.4	31.7	32	10.7	6.2	25.2	20.9
P. tetradactyla	33.9	44.0	31.1	30	9.5	3.6	25.5	22.7
S. gigantea	33.3	44.0	29.3	25	9.5	3.1	27.9	27.6
S. temminckii	32.0	40.9	29.5	26	10.8	5.8	27.7	27.1
				AtpVIII				
M. pentadactyla	38.8	43.3	23.9	16	7.5	9.0	29.9	31.3
M. javanica	38.7	45.6	22.5	13	7.8	8.8	30.9	32.4
P. tricuspis	35.9	39.4	33.3	41	6.1	6.1	24.7	13.6
P. tetradactyla	39.1	43.8	34.4	39	5.7	4.7	20.8	12.5
S. gigantea	37.4	41.5	31.8	29	5.6	4.6	25.1	24.6
S. temminckii	39.5	47.7	30.3	28	5.6	4.6	24.6	20.0
				Cob				
M. pentadactyla	30.9	42.9	25.6	14	13.4	3.4	30.1	40.0
M. javanica	30.0	42.3	24.4	10	14.2	5.6	31.4	42.1
P. tricuspis	30.9	41.5	31.1	26	13.1	3.2	25.0	29.6
P. tetradactyla	31.2	42.1	29.5	23	12.9	3.4	26.4	31.5
S. gigantea	32.5	44.7	28.0	18	12.2	2.4	27.3	34.9
S. temminckii	31.7	43.4	28.3	19	12.4	2.9	27.5	34.7
				CoxI				
M. pentadactyla	28.4	39.3	27.9	21	16.6	5.8	27.1	33.7
M. javanica	26.1	32.8	26.6	18	18.6	11.4	28.6	38.0
P. tricuspis	27.0	26.8	34.1	26	16.8	29.2	22.0	18.4
P. tetradactyla	27.9	38.0	33.4	33	15.7	4.4	23.0	25.0
S. gigantea	29.0	41.8	32.1	31	15.8	3.3	23.0	24.2
S. temminckii	27.7	39.3	33.3	35	16.8	4.7	22.2	21.3

				CoxII				
M. pentadactyla	32.6	41.7	24.9	17	13.5	6.1	29.1	35.5
M. javanica	31.3	37.3	21.9	11	14.8	10.1	32.0	42.1
P. tricuspis	32.6	42.7	30.8	31	13.4	5.7	23.2	20.7
P. tetradactyla	33.0	44.1	30.2	28	13.7	6.6	23.1	21.1
S. gigantea	33.9	46.7	29.2	28	12.3	2.6	24.5	22.5
S. temminckii	33.2	42.7	29.2	27	12.9	5.7	24.7	24.2
				CoxIII				
M. pentadactyla	29.3	42.5	26.8	17	14.5	3.4	29.3	36.8
M. javanica	27.8	39.1	24.4	11	15.8	6.9	32.0	42.9
P. tricuspis	26.6	35.6	34.1	33	15.6	5.7	23.8	25.3
P. tetradactyla	28.2	39.1	33.2	30	13.8	1.9	24.8	29.1
S. gigantea	29.6	44.1	29.8	24	13.9	1.5	26.7	30.3
S. temminckii	28.1	37.5	31.7	27	14.9	6.5	25.3	29.1
				NadI				
M. pentadactyla	32.8	52.8	24.8	11	12.3	3.1	30.0	33.0
M. javanica	30.3	45.9	22.6	5	14.9	10.1	32.2	38.7
P. tricuspis	31.0	46.3	30.9	26	13.1	5.4	25.0	22.2
P. tetradactyla	31.4	48.6	29.0	23	13.4	6.0	26.1	22.9
S. gigantea	33.0	51.4	29.0	21	11.7	2.2	26.3	24.9
S. temminckii	31.9	47.6	28.0	20	13.4	7.3	26.7	25.1
				NadII				
M. pentadactyla	37.3	51.9	24.3	13	8.9	2.9	29.5	32.5
M. javanica	37.1	52.3	21.2	5	10.5	6.4	31.3	36.1
P. tricuspis	35.9	50.0	28.6	18	10.3	6.4	25.1	25.4
P. tetradactyla	38.2	54.9	29.5	23	9.1	2.9	23.3	19.7
S. gigantea	38.9	55.8	27.2	15	7.9	1.7	26.0	27.5
S. temminckii	38.3	54.3	28.3	18	8.0	2.6	25.3	24.6
				NadIII				
M. pentadactyla	34.6	53.9	25.1	9	11.0	3.5	29.4	33.9
M. javanica	31.5	45.2	22.0	3	13.9	10.4	32.7	41.7
P. tricuspis	31.9	51.3	31.6	20	11.3	3.5	25.2	25.2
P. tetradactyla	32.8	52.2	32.5	24	9.9	.9	24.9	22.6
S. gigantea	34.1	54.4	30.9	20	9.6	.9	25.4	24.6
S. temminckii	31.9	48.7	33.0	24	11.9	6.1	23.2	20.9
				NadIV				
M. pentadactyla	34.4	47.7	25.5	15	10.6	2.8	29.5	34.4
M. javanica	32.4	44.0	21.9	6	12.6	6.8	33.0	43.1
P. tricuspis	33.0	44.2	30.1	24	11.6	5.7	25.3	26.0
P. tetradactyla	33.7	45.6	30.5	28	10.3	3.3	25.5	23.5
S. gigantea	34.1	47.7	28.8	22	10.1	2.2	27.1	28.2
S. temminckii	34.0	45.8	29.2	26	10.7	4.6	26.0	23.9
				NadIV-L				
M. pentadactyla	32.3	47.5	26.9	8	11.4	6.1	29.3	32.3
M. javanica	30.6	41.4	26.3	10	12.5	10.1	30.6	30.6
P. tricuspis	29.3	37.8	36.1	33	11.2	5.1	23.5	29.3
P. tetradactyla	27.2	34.7	35.0	30	12.6	6.1	25.2	27.2
S. gigantea	30.6	44.9	36.4	31	10.2	1.0	22.8	30.6
S. temminckii	28.9	40.8	32.0	26	12.6	5.1	26.5	28.9
				NadV				
M. pentadactyla	34.9	44.6	24.8	13	9.8	1.3	30.4	40.9
M. javanica	33.6	42.2	20.9	6	11.5	4.8	34.0	47.4
P. tricuspis	34.1	44.8	29.9	24	11.4	4.3	24.5	26.8
P. tetradactyla	35.1	46.1	30.1	24	10.5	2.7	24.3	26.8
S. gigantea	35.0	45.7	27.3	19	10.1	1.7	27.7	34.0
S. temminckii	33.8	43.8	28.4	21	10.9	3.2	26.9	32.0
				NadVI				
M. pentadactyla	45.3	30.3	17.5	21	6.5	4.0	30.7	44.6
M. javanica	43.8	27.4	13.7	19	9.0	6.9	33.5	46.3
P. tricuspis	42.0	26.6	21.4	25	8.1	5.8	28.5	42.2
P. tetradactyla	41.8	26.0	22.0	25	6.4	5.2	29.9	43.4
S. gigantea	45.1	30.1	19.5	24	6.4	3.5	29.1	42.8
S. temminckii	43.4	30.2	20.5	24	7.2	3.5	28.9	42.4


(a)
Tet
Tric
Java
Penta
Tem
Gig
Dim 2 (21.93%)
Dim 1 (42.73%)
(b)
Tric
Tet
Penta
Java
Gig
Tem
Dim 2 (20.22%)
Dim 1 (43.19%)
(c)
Tri
Tet
Penta
Java
Gig
Tem
Dim 2 (19.07%)
Dim 1 (55.37%)



Supplementary Figure 3.2. Principal Component Analysis (PCA) of Relative Synonymous Codon Usage values (RSCU) for (a) *CoxI*; (b) *NadI*; (c) *NadIII*; (d) *AtpVIII*; (e) *NadIV*; (f) *NadVI* in Asian and African pangolins.

CHAPTER FOUR

PHYLOGEOGRAPHY OF TEMMINCK'S PANGOLIN (*SMUTSIA TEMMINCKII*) POPULATIONS IN SOUTHERN AFRICA USING A GEOREFERENCING APPROACH

4.1. INTRODUCTION

The principle of phylogeography comprises the geographical distribution (spatially and temporally) of a genealogical lineage (Avise 1998; Kidd & Ritchie 2006). Research in phylogeography has largely centred on endangered and invasive species with the genetic discordance mostly being associated with physical barriers, limited dispersal and climatic boundaries (Teske *et al*. 2011).

Several recent studies have explored mammalian phylogeographic patterns in the common eland antelope, (*Taurotragus oryx*), vervet monkeys (*Chlorocebus sensu lato*), white-bellied pangolin (*Phataginus tricuspis*) and Malayan pangolin (*Manis javanica*) using mtDNA markers (Lorenzen *et al*. 2010; Gaubert *et al*. 2016; Turner *et al*. 2016; Mason *et al*. 2019). These studies have reported that some species vary significantly from one geographical area to the next due to genetic differentiation and can sometimes even be considered as a sub-species.

Another approach to investigate the phylogeographic structure of populations relies on the use of nuclear markers, in particular STR markers (Pérez *et al*. 2002). The study of mammalian populations, using STRs, has revealed a link between genetic and geographic distance between populations signifying a flow of genetic material across the population continuum (Pérez *et al*. 2002). In recent years several investigations have used both STRs and mtDNA to elucidate the population structure of, for example, several mammalian species such as impala (*Aepyceros melampus*), roan (*Hippotragus equinus*), gemsbok (*Oryx gazella*), kob (*Kobus kob*), Grant's gazelle (*Nanger granti*), common warthog (*Phacochoerus africanus*), plains zebra

(*Equus quagga*), mountain zebra (*Equus zebra*) and giraffes (*Giraffa camelopardalis*) (Lorenzen *et al*. 2007a,b, 2008, 2012).

This study is the first attempt to utilise both mtDNA and STRs markers to determine the extent of genetic differentiation within and between Temminck's pangolin populations in southern Africa. We also aim to use the above markers along with additional samples and localities to determine the phylogeographic pattern between different distribution areas of Temminck's pangolin in southern Africa since previous studies were inconclusive when only using mtDNA markers. By adding more samples and sample localities as well as increasing the amount of markers, would provide a more holistic view of the Temminck's pangolin population structure. Furthermore, we will then attempt to construct a georeference distribution map to aid forensic investigations in determining the origin of confiscated samples.

4.2. MATERIALS AND METHODS

4.2.1. SAMPLE COLLECTION AND ISOLATION OF DNA

A total of 76 Temminck's pangolin samples, from 22 sampling sites in southern Africa, were collected by researchers in collaboration with the African Pangolin Working Group (APWG) and were stored at the National Biobank located at the South African National Biodiversity Institute (formally National Zoological Gardens (NZG)). The samples were stored under different conditions for each sample type to ensure optimum quality is retained. The blood samples were stored in Ethylene Diamine Tetra-acetic Acid **(**EDTA) tubes, scale samples were stored dry in envelopes, dried skin were stored in tubes, with fresh skin and tissue samples were preserved in absolute ethanol. The blood, tissue and fresh skin samples were kept at -80˚C for long term storage whereas scales and dried skin samples were stored at room temperature. For this study, we grouped the 76 samples, from 22 sampling sites, into 11 localities based on their proximity (Figure 4.1; Table 4.1).

DNA was extracted using the PrepFiler^{TM} Forensic DNA Extraction Kit (Thermo Scientific, Massachusetts, U.S.A.). The standard protocol was used for tissue, blood and skin samples and the hair protocol was followed for the scales samples. The

scale samples were grounded into a powder prior to extraction following the protocol as described in (Du Toit *et al.* 2016). All samples were eluted twice using 30 µl elution buffer. The DNA was quantified on a Qubit 3.0 Fluorometer (Thermo Fisher Scientific, Massachusetts, USA) and diluted to a quantity of 20 ng of genomic DNA. The extracted DNA was stored at -20˚C until further analysis.

Table 4.1. Geographical localities of *Smutsia temminckii* identified in southern Africa indicating the total number of individuals sampled per locality and the number of samples used for mitochondrial DNA (mtDNA) and Short Tandem Repeats (STRs) analysis from each locality.

Samples from known localities (Reference samples)					
Sampling site	**Phylogenetic tree labels**	**Total number of individuals**	**MtDNA**	**STRs**	**Geographic locality**
L01	KP306516.1	1	1	0	Arusha National Park, Tanzania
L02	ZAM	2	0	2	Sioma Ngwezi National Park, Zambia
L03	MOZ	1	1	0	Banhine National Park, Mozambique
L04	MOZ	1	1	1	M'Phingwe, Mozambique
L05	LIM	4	4	0	Mohololholo Wildlife Rehabilitation Centre, Limpopo Province, South Africa
L06	LIM	2	1	2	Waterberg region, Limpopo Province, South Africa
L07	MPU	7	7	7	Sabi Sands Game Reserve, Mpumalanga Province, South Africa
L08	KD	1	0	1	Kathu District, Northern Cape Province, South Africa
L09	KAL	29	21	20	Kalahari Oryx Game Farm, Northern Cape Province, South Africa
L10	TSW	10	8	9	Tswalu Kalahari Reserve, Northern Cape Province, South Africa
L11	KTP	2	1	2	Kgalagadi Transfrontier Park, Northern Cape Province, South Africa
L12	NAM	3	3	3	Mundulea Nature Reserve, Namibia
*L13	CON	1	1	0	Swakopmund region, Namibia
*L14	CON	1	1	0	Windhoek region, Namibia
*L15	CON	1	1	0	Otjiwarongo region, Namibia
*L16	CON	1	1	0	Zambezia (Caprivi Strip) region, Namibia
*L17	CON	1	1	0	South Africa
*L18	CON	1	1	1	Cape Town, Western Cape Province, South Africa
*L19	CON	1	1	0	Chimanimani Mountain Area, Zimbabwe
*L20	CON	1	1	0	Mutare, Zimbabwe
*L21	CON	1	1	1	Mudzi, Zimbabwe
*L22	CON	1	1	0	Chegutu, Zimbabwe
*L23	CON	4	4	2	Zimbabwe
*L24	MG196300.1	1	1	0	Unknown, southern Africa
Total no. of samples		**78**	**63**	**51**	

*Confiscated samples with unknown geographical origins. The region where they were confiscated has been listed as the sampling locality.

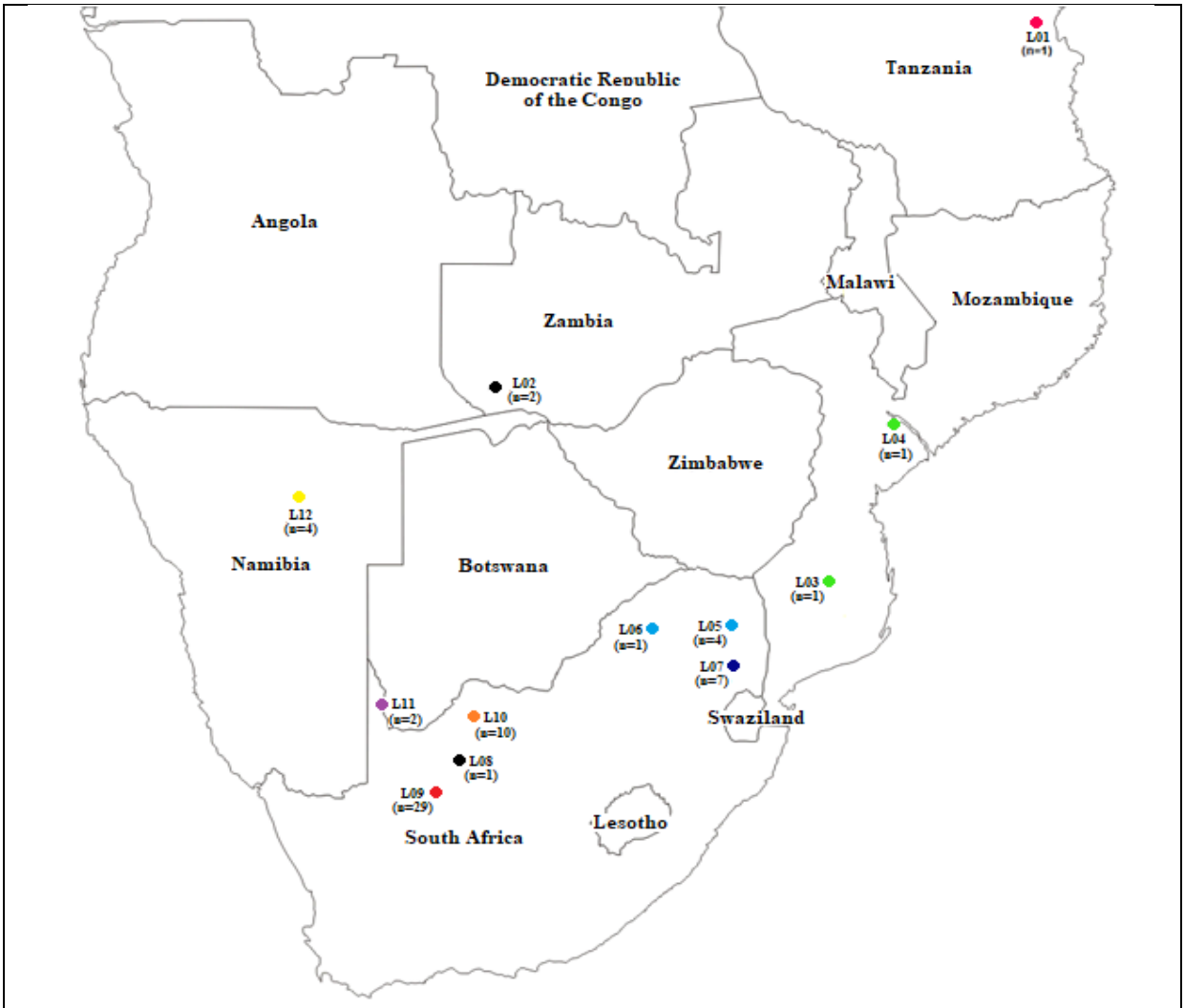


Figure 4.1. Map of southern Africa illustrating the 11 sampling localities and reference used for *Smutsia temminckii*. L = Locality number with the number of samples indicated in brackets.

4.2.2. LABORATORY ANALYSES: SANGER-SEQUENCING

Sanger-sequencing was conducted using species-specific mtDNA markers; *CoI, Cytb* and control region (Appendix A) (Du Toit 2014). Polymerase Chain Reaction (PCR) was performed in a final reaction volume of 25 µl consisting of DreamTaq™ green PCR mastermix (2x) (Thermo Fisher Scientific, Massachusetts, USA), 10 µM of each primer pair and 10ng of DNA. Amplification success were confirmed on 2% agarose gels and PCR product was purified using ExoSAP-IT™ cleanup reagent (Applied Biosystems, Foster City, California, USA), following the manufacturers protocol. Cycle sequencing was performed with the BigDye™ Terminator v3.1 cycle sequence kit (Applied Biosystems, Foster City, California, U.S.A.) followed by the

BigDye XTerminator™ Purification Kit (Applied Biosystems, Foster City, California, USA) before sequencing by capillary electrophoresis on an ABI 3500 genetic analyser (Applied Biosystems, Foster City, California, USA).

4.2.3. LABORATORY ANALYSES: SHORT TANDEM REPEATS (STRS)

A total of 30 polymorphic STR markers were assessed as described in Chapter 2. Fragment analysis was performed on an ABI 3500 genetic analyser (Applied Biosystems, Foster City, California, USA) using a Genescan™ 500 LIZ™ dye size standard (Applied Biosystems, Foster City, California, USA).

4.2.4. STATISTICAL ANALYSES: MTDNA

Sequence trace files of the samples were visually inspected and edited in MEGA v7 (Tamura *et al*. 2013). These sequences were then subjected to a nucleotide BLAST search on NCBI BLAST (www.ncbi.nm.nih.gov/blast) to verify their identity. Each locus was also assembled and trimmed using MEGA v7 prior to alignment using MAFFT v7 (Katoh & Standley 2013). Reference genomes from the six Temminck's pangolin mitogenomes available on Genbank were included. These mitogenomes were KP125951.1 (Du Toit *et al*. 2014), KP306516.1 (Hassanin *et al*. 2015), MF536685.1, MF536686.1, MF536687.1 (Chapter 3) and MG196300.1 (Gaubert *et al*. 2018). The reference sequence from *S. gigantea* (MF536684.1, Chapter 3) was used as an outgroup in the analysis.

Two separate datasets were used for all analysis performed. The first dataset included 43 sequences from known localities along with the six references sequences (total = 49 samples) to determine whether the geographical origin could be differentiated. The second dataset was set up with samples that included the 43 known localities, 14 unknown localities and six reference sequences (63 samples) to determine the potential origin of the unknown localities.

For each dataset, analysis was performed on a combined (all three loci combined) and individual basis (all three loci separate). Concatenation of the three mitochondrial loci is possible as all three originate within the mitochondria which is

inherited as a whole, without intermolecular recombination (Avise 1998). The individual locus based analysis was included as a control to assess how the loci perform on their own given their different mutation rates. The combined datasets will provide a more robust representation of the mtDNA genome as data from three different mitochondrial loci, with different mutation rates, were used.

All the datasets were subjected to jModeltest v2.1.7 (Guindon & Gascuel 2003; Darriba *et al*. 2012) to find the best-fit model of sequence evolution under the Akaike Information Criterion (AIC) (Akaike 1974). Phylogenetic analysis, using a Bayesian Inference approach, was implemented using MrBayes v3.2.6 (Ronquist *et al*. 2012). Bayesian Inference was run over 5 000 000 generations after which 25% were discarded as burn-in. A Maximum Likelihood (ML) approach was also performed using PhyML3.1 (Guindon *et al*. 2010). The ML analyses were carried out with 10 000 bootstrap replications and Bootstrap values below 70% support were excluded from the trees.

Haplotype and nucleotide diversities were estimated using DnaSP v5 (Librado & Rozas 2009). Haplotype networks were constructed using NETWORK v5 (Bandelt *et al*. 1999) to determine the distribution of haplotypes among localities. Spatial Analysis of Molecular Variance (SAMOVA) v1.0 (Excoffier & Lischer 2010) was performed to determine the level of variation for K = 2–4 locality groups. The SAMOVA was selected as it uses geographical coordinates of the localities to determine maximum genetic variance among groups without *a priori* inference (Wright 1950). These analyses were all conducted with 100 simulated annealing permutations to define genetic structure for these localities. Each analysis was performed twice to check for consistency between runs and results.

4.2.5. STATISTICAL ANALYSES: SHORT TANDEM REPEATS (STRS)

The STRs were visualized utilising the software GeneMapper v4 (Applied Biosystems, Inc., Foster City, CA, USA) and genotyping errors, allele dropout and null alleles were identified using MICRO-CHECKER (Van Oosterhout *et al*. 2004). The statistical program GenAlEx v6.5 (Peakall & Smouse 2006, 2012) was applied to

calculate the number of alleles per locus (*A*), expected heterozygosity (*He*), observed heterozygosity (*Ho*), and unbiased heterozygosity (*Hz*). To test for deviations from expected Hardy-Weinberg (*HW*) proportions of genotypes as well as estimate gametic disequilibrium between loci, the program ARLEQUIN v3.5.2 (Excoffier & Lischer 2010) was used. We also applied a *Bonferroni* adjustment with a significance level of 0,05 to assess and adjust the associated probability values in multiple comparisons.

CERVUS v3.0 (Kalinowski *et al.* 2007) was used to calculate the number of null alleles present in the datasets while STRUCTURE v2.3.4 (Pritchard *et al*. 2000) was used to determine the most probable clustering pattern without prior population information of the 51 individuals, consisting of multiple localities. The parameters used were 200 000 Markov Chain Monte Carlo (MCMC) repetitions following a burn-in of 20 000 repetitions. Analysis was run over 100 replicates for K = 1–22 and K = 1–5. STRUCTURE HARVESTER v0.6.94 (Earl & Von Holdt 2012) was used to determine the most likely K value by evaluating the greatest increase in posterior probability (ΔK, Evanno *et al*. 2005). Population differentiation using fixation index (F_{ST}) and Analysis of Molecular Variance (AMOVA) were performed in ARLEQUIN v3.5.2 (Excoffier & Lischer 2010) using 1 000 permutations. The groups that were tested in the AMOVA were based on the best cluster assignments identified by STRUCTURE.

4.3. RESULTS

4.3.1. PARTIAL MITOCHONDRIAL DNA (MTDNA) ANALYSIS

In this study, we successfully generated sequence data for 57 Temminck's pangolin individuals for the *Co1* (1 553 bp), *Cytb* (1 139 bp) and control region (1 220 bp) loci. From these individuals, 43 were from known localities whereas 14 were confiscated samples of unknown origin. The best-fit model of sequence evolution for each dataset, selected under AIC, as well as the invariable sites are represented in Table 4.2.

A total of 29 haplotypes were observed for *Co1* (Supp. Table 4.1), 16 for *Cytb* (Add. Table 4.2), 20 for the control region (Supp. Table 4.3) with 41 for the three loci combined (Supp. Table 4.4). The number of unique haplotypes compared to shared haplotypes between localities, were more pronounced in *Co1* and the combined dataset compared to *Cytb* and the control region. In general, haplotypes are only represented by one or two localities resulting in a high total haplotype diversity (H_D) of 96,93% for the combined dataset and between 82,54–94,73% for individual loci; thus haplotype frequencies would be potentially of limited use in determining population structure in pangolins if these are true estimates (Table 4.3). The nucleotide diversity was 0,146 for the combined loci and between 0,119–0,155 for the individual loci, which is low but within the range observed in mtDNA (Yu *et al*. 2003; Ray *et al*. 2004; Nabholz *et al*. 2008; Canino *et al*. 2010).

Table 4.2. Models of sequence evolution selected under the Akaike Information Criterion (AIC) for *CO1*, *Cytb* and control region datasets for *Smutsia temminckii*. The known-locality dataset (individuals with known sampling localities and reference samples) and overall dataset (individuals with known localities, unknown localities/confiscated samples and reference samples) were analysed separately to determine the best-fit model of sequence evolution. p-inv = Proportion of invariable sites; gamma = gamma shape.

Akaike Information Criterion (AIC)	***CO1***	***Cytb***	**Control Region**	**Combined Dataset**
Known localities	TIM2+I (p-inv = 0,710)	TIM2+G (gamma = 0,547)	TPM2uf+G (gamma = 0,106)	TIM2+I+G (p-inv+gamma = 0,788)
Known and Unknown localities	TPM3uf+I (p-inv = 0,934)	TPM3uf+I (p-inv = 0,917)	TVM+I (p-inv = 0,914)	TPM3uf+I (p-inv = 0.954)

Table 4.3. Haplotype summary for *Smutsia temminckii* showing number of haplotypes per locality (H_N), percentage haplotype diversity (H_D) and nucleotide diversity (N_D).

	CO1	***CytB***	**Control Region**	**Combined**
Number of Haplotypes (H_N)	29	16	20	41
Haplotype Diversity % (H_D)	94,73	82,54	87,92	96,93
Nucleotide Diversity (N_D)	0,155	0,119	0,149	0,146

4.3.1.1. KNOWN LOCALITIES DATASET

Both the ML and Bayesian trees display similar clustering across the individual trees. The samples show no distinctive geographical pattern according to localities, but indicate some genetic sub-structuring. Three clusters were observed consisting of various localities within all three loci and the combined dataset (Figure 4.2, Supp. Figure 4.1–4.3). Cluster I and III (red and green) are widespread and are made up of the majority of the individuals from South Africa and Namibia. Cluster II (blue) on the other hand is more restricted and is much smaller with the bulk of samples from all the countries included especially those from Mozambique and the northern area of the sampling map (including Tanzania).

Haplotype networks for all loci as well as the combined dataset have a star-shaped phylogeny with a common central ancestral haplotype from which low frequency haplotypes are derived by one to two mutations. These results suggest genetic sub-structuring with individual loci showing sub-clades consisting of multiple locations that are isolated from the rest of the network by three or more mutations (Supp. Figure 4.4). The same can be seen from the combined haplotype loci map where there are three sub-clades representing various localities (Figure 4.3). This is the same clustering pattern which is found in the phylogenetic trees (Figure 4.2, Supp. Figure 4.1–4.3).

The SAMOVA results show *K*=3 to be the groupings which are maximally differentiated from each other for all loci with the combined dataset (Table 4.4) showing 72,06% variation among groups and an F_{CT} value of 0,019 (±0,0). The individual loci (Supp. Table 4.5) also show *K*=3 with 69,65–74,56% variation among groups and an F_{CT} value of 0,01 (±0,0–0,003). This indicates highly significant support for the three groups observed in both the phylogenetic trees and haplotype networks.

Table 4.4. Spatial analyses of molecular variance (SAMOVA) for *Smutsia temminckii* showing the results for the combined loci of 49 and 63 individuals, respectively. Analysis for *K* = 2–4 were run for each dataset.

	Source of Variation			
	N = 49		**N = 63**	
	% variation among groups	**p-value (± Standard deviation)**	**% variation among groups**	**p-value**
Two Groups (*K* = 2)	70,19	0,098 (0,0)	41,08	0,0 (0,0)
Three Groups (*K* = 3)	72,06	0,0186 (0,0)	50,66	0,022 (0,0)
Four Groups (*K* = 4)	64,64	0,0039 (0,002)	60,25	0,0 (0,0)

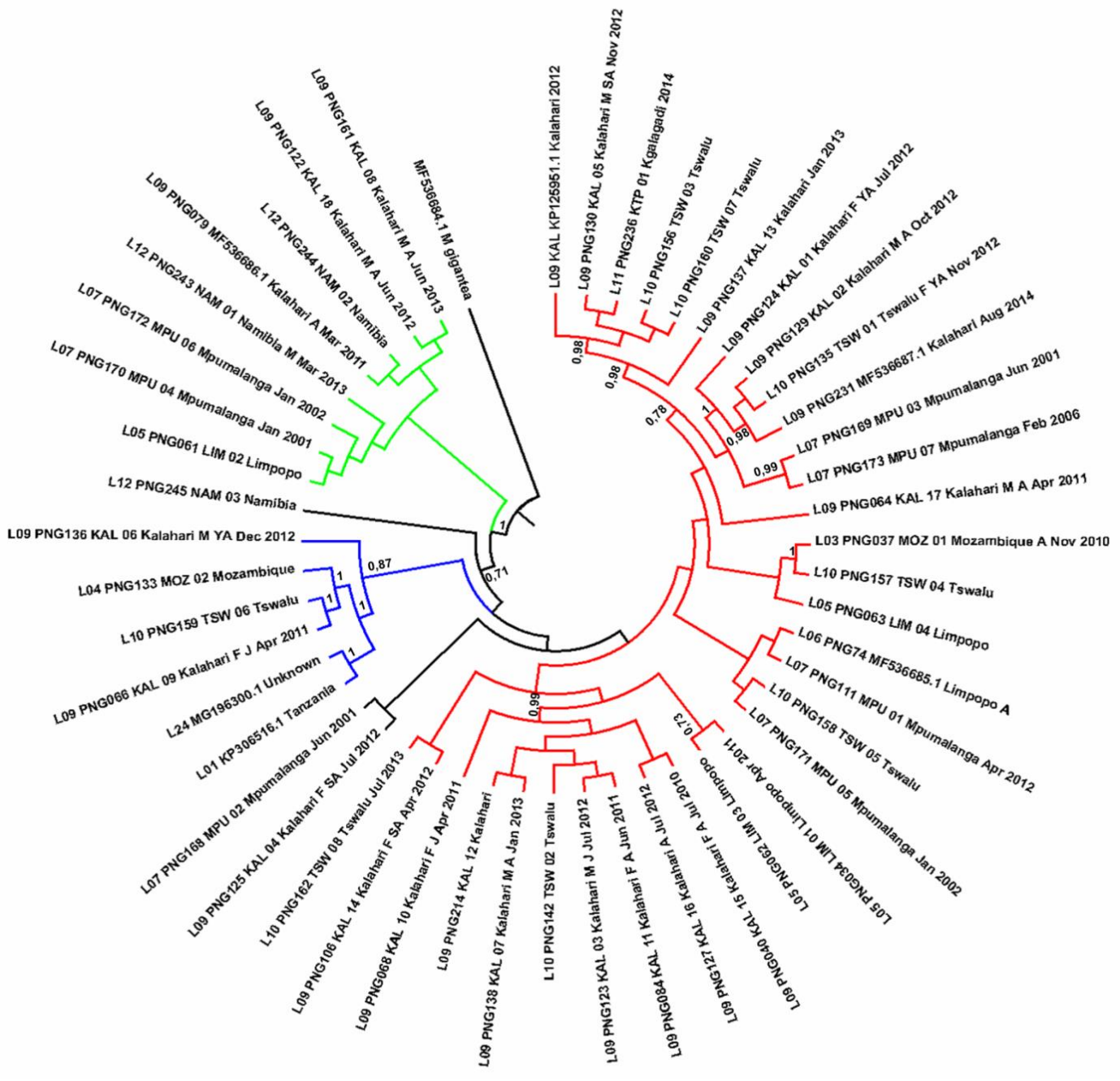


Figure 4.2. Consensus tree of concatenated loci of 49 *Smutsia temminckii* samples (known localities and references). Bayesian Inference (BI) posterior probability is noted on the branches with BI ≥ 0,7 indicating strong support.

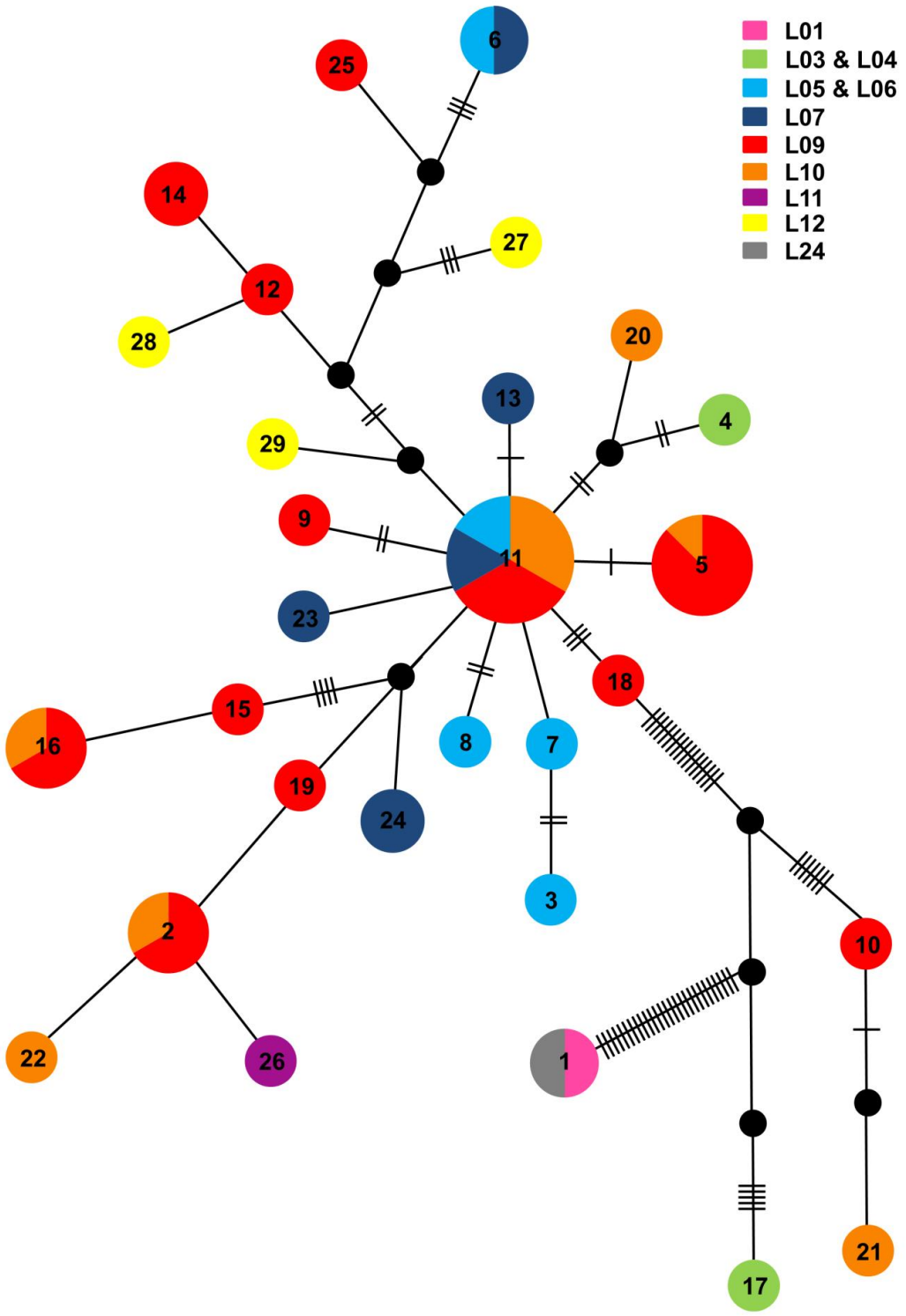


Figure 4.3. Haplotype network for the 49 combined *Smutsia temminckii* loci. Circles represent different haplotypes and the size of the circle indicates the number of individuals that share that haplotype. Colours indicate the sampling localities of the haplotype. The black bars indicate the number of mutations each haplotype underwent and the black circles indicate missing haplotypes.

4.3.1.2. KNOWN AND UNKNOWN LOCALITIES DATASET

The results for both ML and Bayesian Inference trees for the samples with known and unknown locations (63 individuals) displayed a similar topology in the three individual loci trees, in which there were three distinct clusters with no distinctive geographical pattern (Supp. Figure 4.5–4.7). The genetic sub-structuring of the concatenated dataset revealed four distinct clusters (Figure 4.4). The confiscated samples are interspersed among the different clusters in both the individual as well as the concatenated trees, suggesting that regional confiscations may have different points of origin. The four clusters consist of a variety of localities, similarly as seen in the 49 individual dataset. Clusters I, II and IV (red, green and pink) were widespread and consist of samples from South Africa, Namibia and southern Mozambique. Cluster III (blue) is more restricted and shows a mixture of all countries with the majority from the more north-eastern side of the sampling map.

Haplotype networks for the four datasets have a star-shaped phylogeny with a common central ancestral haplotype from which low frequency haplotypes are derived by one to two mutations. There is evidence of genetic sub-structuring with individual loci and four sub-clades consisting of multiple locations can be seen in the combined network (Figure 4.5; Supp. Figure 4.8). These clades correspond with the four clusters observed in the phylogenetic tree above.

The SAMOVA results show *K*=3 for the individual loci and *K*=4 to be the maximally differentiated groupings for the combined dataset. This is in accordance with the phylogenetic trees which show three clusters in the individual loci and four in the combined loci. The individual loci (Supp. Table 4.5) show a variation among groups of 68,58–74,0% and an F_{CT} value of 0,0 (±0,0) and the combined dataset (Table 4.4) show the variation among groups of 60,25% and an F_{CT} value of 0,0 (±0,0). This also shows high significant support for the four clusters observed in the phylogenetic tree and haplotype network.

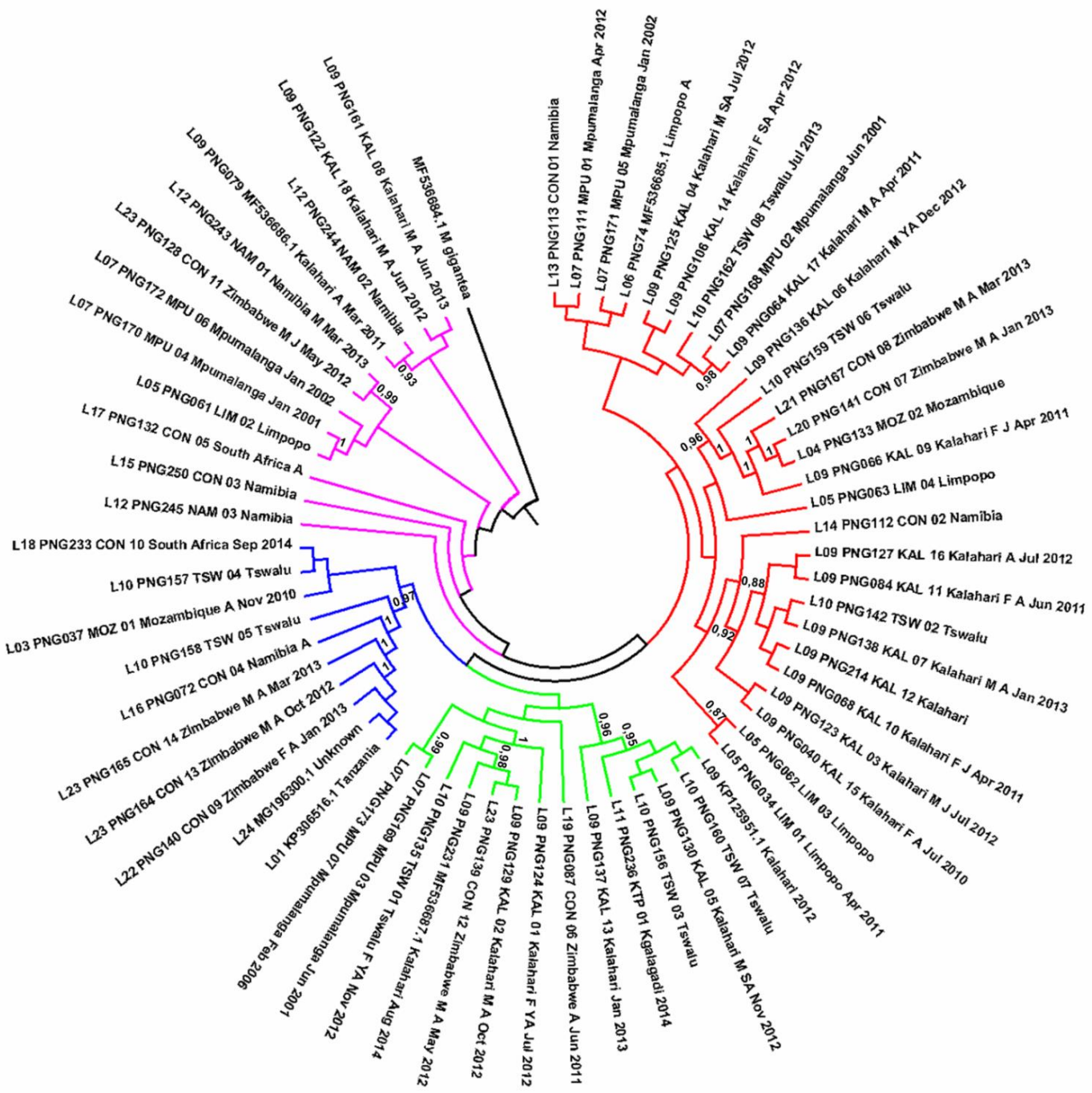


Figure 4.4. Consensus tree of the concatenated loci dataset for 63 *Smutsia temminckii* samples (known, unknown and reference localities). Bayesian Inference (BI) posterior probability is noted on the branches with BI ≥ 0,7 indicating strong support.

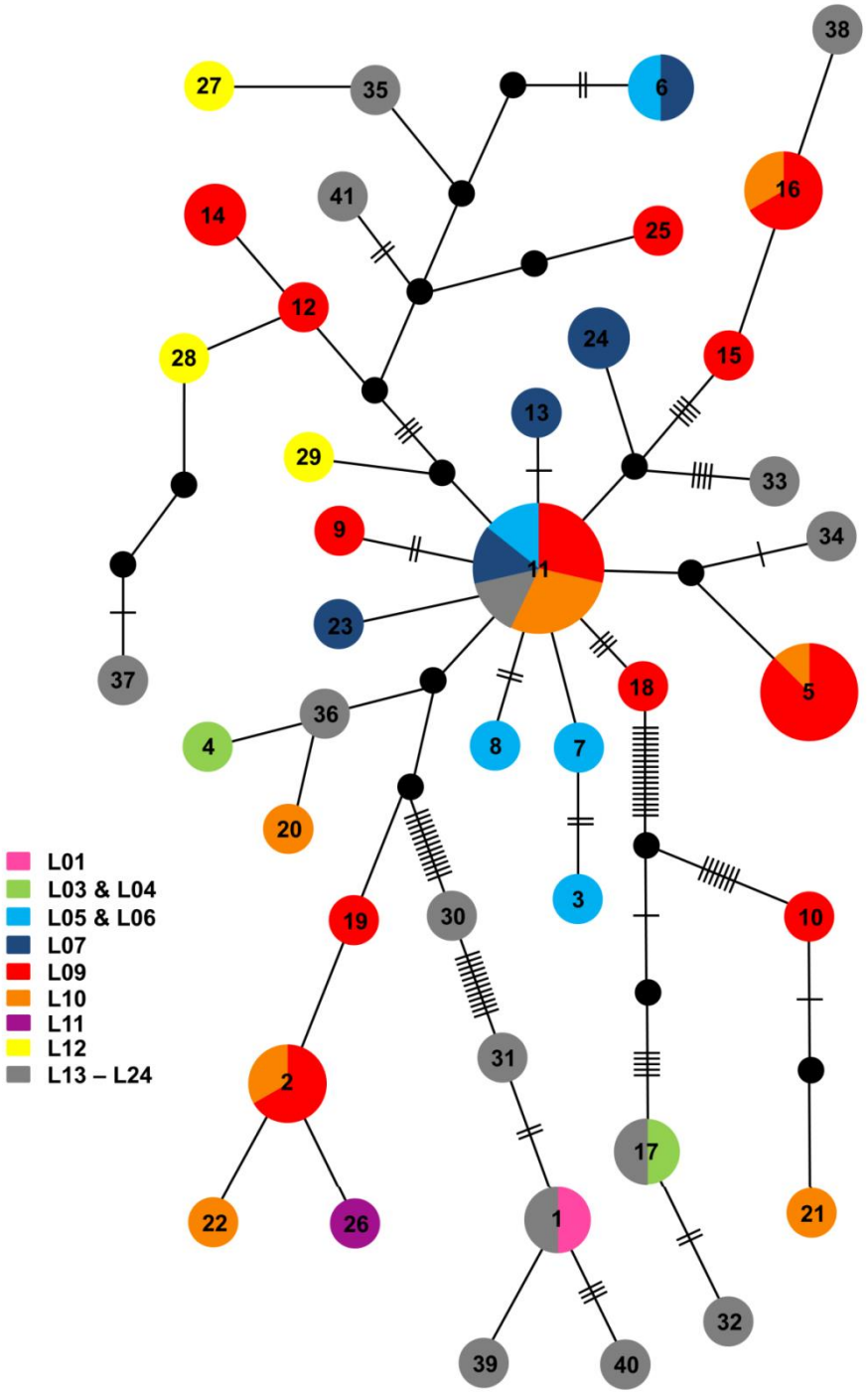


Figure 4.5. Haplotype network for the 63 combined *Smutsia temminckii* loci. Circles represent different haplotypes and the size of the circle indicates the number of individuals that share that haplotype. Colours indicate the sampling localities of the haplotype. The black bars indicate the number of mutations each haplotype underwent and the black circles indicate missing haplotypes.

4.3.2. SHORT TANDEM REPEATS (STRS) ANALYSIS

In this study, 30 STRs were used, for which the entire dataset was analysed for HWE and structure to infer possible populations. An AMOVA analysis was performed following the structure clustering.

The first analysis, consisting of 51 individuals and 30 STRs showed no evidence of linkage disequilibrium; however, nine markers (PAN01, PAN07, PAN08, PAN10, PAN17, PAN18, PAN20, PAN23 and PAN36) showed significant deviations from HWE. This could be as a result of null alleles or population substructure. Null alleles were calculated and found to be present in five of the nine markers ranging between 0,110–0,780. The average number of alleles per marker was 5,3 with a standard error (SE) value of 0,624. The expected (*H*e), observed (*Ho*) and unbiased heterozygosity (*uHe*) is 0,533; 0,521 and 0,538 respectively.

The second analysis consisted of 51 individuals and 21 STRs (excluding the nine markers which were out of HWE, as indicated above). The results show the average number of alleles to be 4,429 (*SE*=0,471) with *H*e 0,529; *Ho* 0,434 and *Hz* 0,535. The expected and unbiased heterozygosity is both slightly higher than the observed heterozygosity for both sets of results; which could indicate the presence of null alleles in the overall population. Null alleles were calculated and found in both datasets (30 and 21 markers). The values ranged between 0,110–0,783 (30 markers) and 0,111–0,140 (21 markers).

STRUCTURE analysis was run for both datasets (30 and 21 markers) using *K*=1–22 assuming no prior population clustering. STRUCTURE HARVESTER identified *K*=15 (30 markers) and K=2 (21 markers) to be the most likely true *K* value for the dataset submitted. When the figures for both datasets were created the results showed minimum distinctive patterns or clustering which could be used to differentiate between individuals. The results also indicated *K*=1–5 to be more evenly distributed in both datasets. Subsequently, another set of runs were performed for the 30 and 21 markers using *K*=1–5 as the prior. STRUCTURE HARVESTER showed the newly run, *K*=1–5 results to be *K*=3 and *K*=4 respectively.

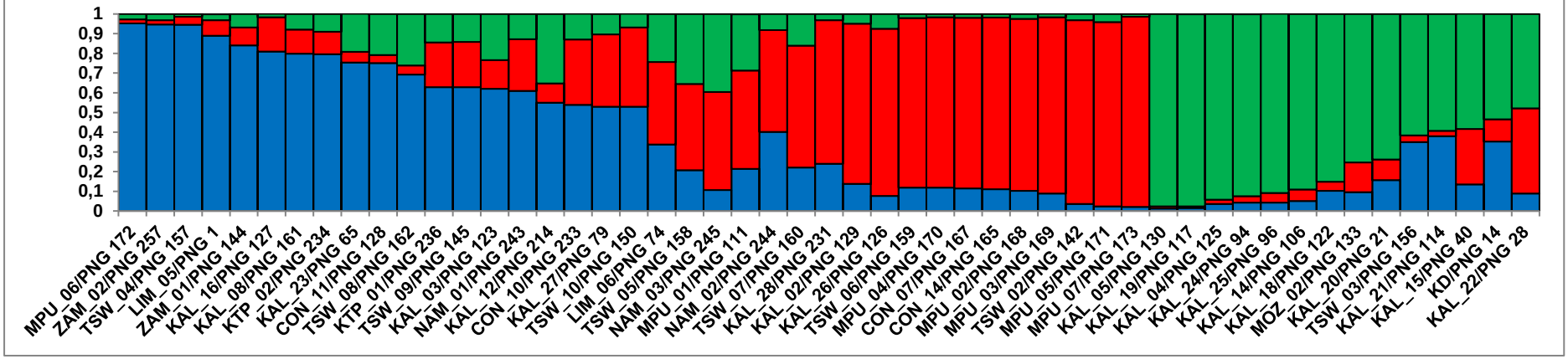


Figure 4.6. Structure analysis for 51 *Smutsia temminckii* microsatellite genotypes representing 30 markers. Each vertical bar represents a single individual and shows the proportional lengths, which determines the probability of certain individuals forming a cluster.

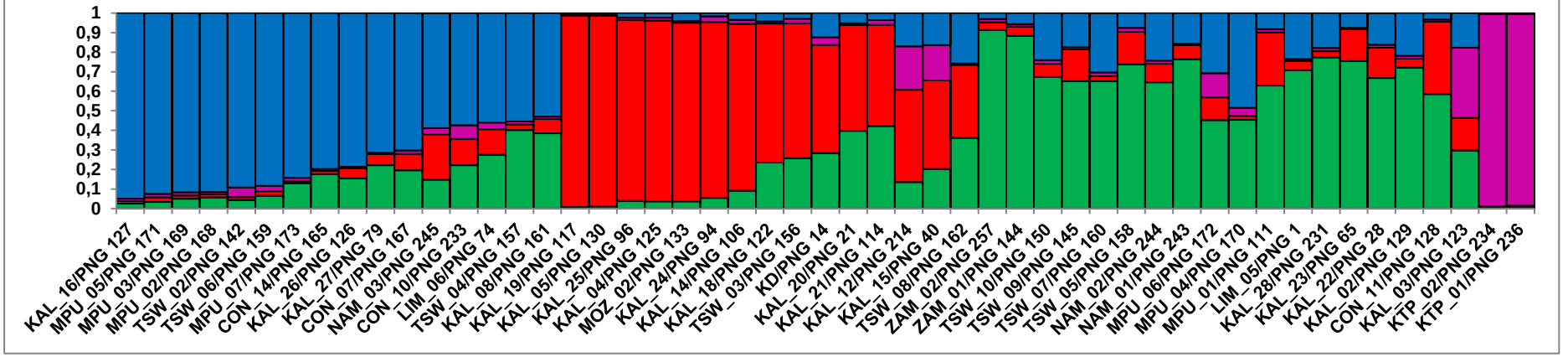


Figure 4.7. Structure analysis for 51 *Smutsia temminckii* microsatellite genotypes representing 21 markers. Each vertical bar represents a single individual and shows the proportional lengths, which determines the probability of certain individuals forming a cluster.

Both datasets show the same grouping patterns for the three clusters with the exception of three samples in the 21 markers dataset (contributing to the fourth cluster). Therefore, the three corresponding clusters in the both datasets (excluding the three samples in the 21 markers dataset) show more consistent grouping patterns as seen in the mtDNA results above (Figure 4.6, 4.7). The three clusters comprise of different localities each which is consistent with the results seen in the various mitochondrial DNA analyses for the dataset containing the 49 individuals (with known localities).

The AMOVA results (using the three clusters from the 30 marker dataset) indicate a pairwise comparison of 3,57% between the three clusters which is not significant for determining variation between the clusters. The results also show variation within each cluster to be 95,92% with FST 0,0408 (P = 0,0).

4.4. DISCUSSION

This is the first molecular geo-referencing study of Temminck's pangolin. Total haplotype diversity of Temminck's pangolin (with site–specific values ranging from 0,825–0,969) is high and within the range reported for the Malayan pangolin (*Manis javanica)* (0,947; Wirdateti 2013) for the *Cytb* locus. Furthermore, the haplotype diversity results obtained in this study were similar to those reported for antelope species, for example: Chinese water deer, *Hydropotes inermis inermis* (0,923; Hu *et al.* 2006), Indian muntjac, *Muntiacus muntjak* (0,862; Wu *et al.* 2006), kob, *Kobus kob* (0,72; Birungi & Arctander 2000) and the African buffalo, *Syncerus caffer* (0,91; Simonsen *et al.* 1998). The results are consistent across all analysis performed where no distinctive geographical pattern was observed between localities, however a clear clustering of different samples was observed which could be attributed to ancestral lineages present in the current population or females not dispersing great distances.

The results from the 49 individuals for the mtDNA (combined and individual loci) and 51 individuals for the STRs both showed three distinct clusters across all analysis performed. The three clusters were also observed when more individuals were added to the mtDNA dataset (63 individuals). However, when the three loci were

combined, the dataset containing the 63 individuals showed a clustering formation with four clusters instead of three. This is evident that the combined loci dataset have more data (>3 900 bp) to differentiate from compared to the individual loci which only have an average of 1 300 bp.

This study found low levels of genetic differentiation within the identified localities, suggesting gene flow between different sub-groups and the genetic results confirm that there are no obvious restrictions to gene flow within this species. It is a well-known phenomenon that individuals separated by geographic distance often display large genetic distances as observed in various African Bovids (Templeton & Georgiadis 1996).

Temminck's pangolins however are known to travel long distances as (Pietersen *et al.* 2014b) reported on a young male Temminck's pangolin which was tracked and found to have travelled 300 km in a window period of four months (Van Aarde *et al*. 1990), although the possibility that this might have been a human-assisted dispersal event cannot be entirely ruled out. If this was a *bona fide* dispersal event this could indicate that juveniles may disperse between the different localities, since the species' distributional range is continuous, thus maintaining gene flow between regions and sampling areas. Further evidence of possible dispersal between sampling localities was found, where juvenile/young adults dispersed minimum distances of 32–81 km in 20 days before contact was lost or the transmitters removed (Pietersen *et al*. 2014b). Furthermore, territorial behaviour and natal philopatry may play a role in genetic differentiation (Sugg *et al*. 1996; Piertney *et al*. 1998). Although the Temminck's pangolins are known as solitary animals with large territories and strong home range fidelity, little is currently known with regards to their behaviour. Thus a combination of habitat differences, substantial geographic distance and intrinsic factors may lead to a degree of gene isolation and genetic drift. A stepping stone model of migration could be suggested since the dispersal habitat of Temminck's pangolin is considered to be one continuous overlapping distribution across southern Africa.

The stepping-stone model suggests that migration is restricted to adjacent sub-populations as would be the case with the pangolins where the populations are not

isolated and neighbouring populations (especially juveniles) are able to disperse and exchange genetic material, thus maintaining a diverse gene pool (Kimura & Weisss 1964). This model might also be ascribed to ancestral traits which are shared between the different localities. The confiscated samples (or samples with unknown localities) are a good example of this model since their true origin could be ancestral to the clusters where they grouped. Most of the unknown localities are adult animals or juveniles which could have migrated to the areas where they were found or had a recent parent or ancestor which migrated before. Our sampling region was mostly represented by the western and eastern side of the distribution range with the central distribution area (Botswana and Zambia) being omitted. Therefore, additional sampling in the intervening regions is required to further investigate the origin of the different clustering patterns observed.

From this preliminary study and results, based on genotypic data from mtDNA and STRs, a general overview of the population structure for Temminck's pangolin was determined. Further studies should include ecological data in order to gather a clear understanding of the barriers that separate these animals in their distribution range as limited knowledge about their dispersal is known. Some of these barriers could include mountain ranges, rivers, habitat and access to food sources which could restrict movement between different countries or localities.

However this seems improbableas these impediments is not likely to disrupt the movements of panoglins. It is possible for pangolins to migrate around a mountain range since they are prone to travel long distances, expecially the juveniles. Rivers pose also little restriction as pangolins are good swimmers and able to cross large rivers. During the dry season, rivers diminish in size or dry up completely thus causing no restriction for crossing. The habitats for Temminck's pangolin share a certain level of composition and therefore they are fairly adaptable between different groups of pangolins, therefore it is unlikely to cause a true barrier between localities. Food sources are likely to be a true barrier between localities as pangolins are myrmecophages and feed on selected ant and termite species. Historical distribution ranges indicated that pangolins are absent in these habitats due to a disappearance of their preferred food sources.

Pertaining to the above, mtDNA and STRs are not robust enough to differentiate the different geographical populations of Temminck's pangolin in southern Africa and alternative methods might be developed such as SNPs or whole genome assemblies, which cover a broader range of the genome. In addition a more representative sampling strategy from the entire distribution range would shed light on possible genetic distances amongst locations and across various geographical areas. Although no distinctive georeferencing could be established for this species, using the identified localities and selected DNA markers, the low geographichal structure provides some indication from a conservation point of view that there is connectivity between most populations with gene flow being maintained.

4.5. SUPPLIMENTARY FIGURES AND TABLES

Supplementary Table 4.1. Haplotype layout for *CoI* loci illustrating the number of haplotypes shared between individuals as the first value; the proportion (reported to second decimal) of unique haplotypes as the second value; haplotype diversity and nucleotide diversity in the different localities observed during this study.

Haplotype	L01 (1)	L03 (1)	L04 (1)	L05 (4)	L06 (1)	L07 (7)	L09 (21)	L10 (8)	L11 (1)	L12 (3)	L13 (1)	L14 (1)	L15 (1)	L16 (1)	L17 (1)	L18 (1)	L19 (1)	L20 (1)	L21 (1)	L22 (1)	L23 (4)	L24 (1)
HAP 1	1/1,00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1/1,00	1/0,25	1/1,00
HAP 2	-	-	-	-	-	-	3/0,14	1/0,13	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 3	-	-	-	1/0,25	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 4	-	1/1,00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 5	-	-	-	-	-	-	7/0,33	1/0,13	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 6	-	-	-	1/0,25	-	1/0,14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 7	-	-	-	1/0,25	1/1,00	2/0,28	3/0,14	2/0,25	-	-	1/1,00	-	-	-	-	-	-	-	-	-	-	-
HAP 8	-	-	-	1/0,25	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 9	-	-	-	-	-	-	1/0,05	1/0,13	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 10	-	-	-	-	-	-	-	-	-	-	-	-	-	1/1,00	-	-	-	-	-	-	-	-
HAP 11	-	-	-	-	-	-	3/0,14	-	-	1/0,33	-	-	-	-	-	-	-	-	-	-	-	-
HAP 12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1/1,00	-	-	-	-	-
HAP 13	-	-	-	-	-	-	-	-	-	-	-	1/1,00	-	-	-	-	-	-	-	-	-	-
HAP 14	-	-	-	-	-	-	1/0,05	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 15	-	-	-	-	-	-	-	-	-	1/0,33	-	-	-	-	-	-	-	-	-	-	1/0,25	-
HAP 16	-	-	-	-	-	-	2/0,09	1/0,13	-	-	-	-	-	-	-	-	-	-	-	-	1/0,25	-
HAP 17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1/1,00	-	-	-	-	-	-	-
HAP 18	-	-	1/1,00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1/1,00	-	-	-	-
HAP 19	-	-	-	-	-	-	1/0,05	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 20	-	-	-	-	-	-	-	1/0,13	-	-	-	-	-	-	-	1/1,00	-	-	-	-	-	-
HAP 21	-	-	-	-	-	-	-	1/0,13	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 22	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1/0,25	-
HAP 23	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1/1,00	-	-	-
HAP 24	-	-	-	-	-	1/0,14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 25	-	-	-	-	-	2/0,28	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 26	-	-	-	-	-	1/0,14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 27	-	-	-	-	-	-	-	-	1/1,00	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 28	-	-	-	-	-	-	-	-	-	1/0,33	-	-	-	-	-	-	-	-	-	-	-	-
HAP 29	-	-	-	-	-	-	-	-	-	-	-	-	1/1,00	-	-	-	-	-	-	-	-	-
Haplotype Diversity (%)	0,0	0,0	0,0	100,0	0,0	90,5	85,2	96,4	0,0	100,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0
Nucleotide Diversity	0,0	0,0	0,0	0,114	0,0	0,069	0,089	0,129	0,0	0,061	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,288	0,0

Supplementary Table 4.2. Haplotype layout for *Cytb* loci illustrating the number of haplotypes shared between individuals as the first value; the proportion (reported to second decimal) of unique haplotypes as the second value; haplotype diversity and nucleotide diversity in the different localities observed during this study.

Haplotype	L01 (1)	L03 (1)	L04 (1)	L05 (4)	L06 (1)	L07 (7)	L09 (21)	L10 (8)	L11 (1)	L12 (3)	L13 (1)	L14 (1)	L15 (1)	L16 (1)	L17 (1)	L18 (1)	L19 (1)	L20 (1)	L21 (1)	L22 (1)	L23 (4)	L24 (1)
HAP 1	1/1,00	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	1/1,00
HAP 2	-	-	-	-	-	-	2/0,09	2/0,25	1/1,00	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 3	-	-	-	3/0,75	1/1,00	5/0,70	8/0,38	3/0,38	-	-	1/1,00	-	-	1/1,00	-	-	1/1,00	-	-	-	1/0,25	-
HAP 4	-	1/1,00	-	-	-	-	-	-	-	-	-	-	-	-	-	1/1,00	-	-	-	-	-	-
HAP 5	-	-	-	-	-	-	7/0,33	1/0,13	-	-	-	1/1,00	-	-	-	-	-	-	-	-	-	-
HAP 6	-	-	-	1/0,25	-	2/0,28	-	-	-	-	-	-	-	-	1/1,00	-	-	-	-	-	1/0,25	-
HAP 7	-	-	-	-	-	-	1/0,05	1/0,13	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 8	-	-	-	-	-	-	1/0,05	-	-	1/0,33	-	-	-	-	-	-	-	-	-	-	-	-
HAP 9	-	-	-	-	-	-	2/0,09	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 10	-	-	1/1,00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1/1,00	1/1,00	-	-	-
HAP 11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1/1,00	-	-
HAP 12	-	-	-	-	-	-	-	1/0,13	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1/0,25	-
HAP 14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1/0,25	-
HAP 15	-	-	-	-	-	-	-	-	-	1/0,33	-	-	1/1,00	-	-	-	-	-	-	-	-	-
HAP 16	-	-	-	-	-	-	-	-	-	1/0,33	-	-	-	-	-	-	-	-	-	-	-	-
Haplotype Diversity (%)	0,0	0,0	0,0	50,0	0,0	47,6	75,7	85,7	0,0	100,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0
Nucleotide Diversity	0,0	0,0	0,0	0,030	0,0	0,029	0,081	0,113	0,0	0,157	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,147	0,0

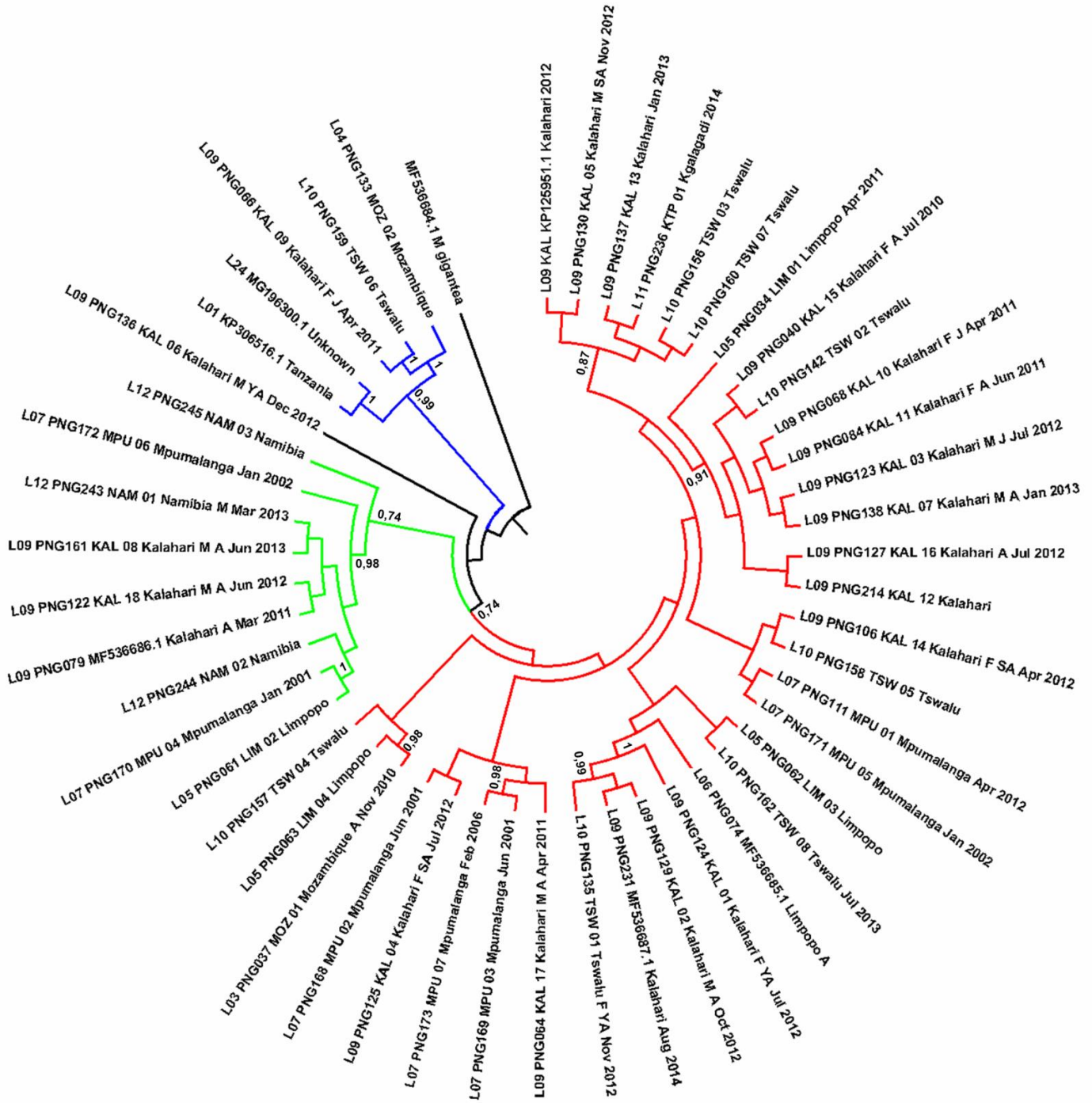
Supplementary Table 4.3. Haplotype layout for control region loci illustrating the number of haplotypes shared between individuals as the first value; the proportion (reported to second decimal) of unique haplotypes as the second value; haplotype diversity and nucleotide diversity in the different localities observed during this study.

Haplotype	L01 (1)	L03 (1)	L04 (1)	L05 (4)	L06 (1)	L07 (7)	L09 (21)	L10 (8)	L11 (1)	L12 (3)	L13 (1)	L14 (1)	L15 (1)	L16 (1)	L17 (1)	L18 (1)	L19 (1)	L20 (1)	L21 (1)	L22 (1)	L23 (4)	L24 (1)
HAP 1	1/1,00	-	-	-	-	-	-	-	-	-	-	-	-	1/1,00-	-	-	-	-	-	1/1,00	1/0,25	1/1,00
HAP 2	-	-	-	-	-	-	3/0,14	2/0,25	1/1,00	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 3	-	-	-	2/0,50	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 4	-	1/1,00	-	-	-	-	-	1/0,13	-	-	-	-	-	-	-	1/1,00	-	-	-	-	-	-
HAP 5	-	-	-	-	1/1,00	2/0,28	11/0,52	3/0,38	-	1/0,33	1/1,00	-	1/1,00	-	-	-	-	-	-	-	-	-
HAP 6	-	-	-	1/0,25	-	1/0,14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 7	-	-	-	1/0,25	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 8	-	-	-	-	-	-	1/0,05	1/0,13	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 9	-	-	-	-	-	-	3/0,14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1/1,00	-	-	-	-	-
HAP 11	-	-	-	-	-	1/0,14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 12	-	-	-	-	-	-	-	-	-	-	-	1/1,00	-	-	-	-	-	-	-	-	-	-
HAP 13	-	-	-	-	-	-	3/0,14	1/0,13	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 14	-	-	-	-	-	-	-	-	-	1/0,33	-	-	-	-	-	-	-	-	-	-	1/0,25	-
HAP 15	-	-	-	-	-	-	-	-	-	1/0,33	-	-	-	-	1/1,00	-	-	-	-	-	-	-
HAP 16	-	-	1/1,00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1/1,00	1/1,00	-	-	-
HAP 17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1/0,25	-
HAP 18	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1/0,25	-
HAP 19	-	-	-	-	-	2/0,28	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 20	-	-	-	-	-	1/0,14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Haplotype Diversity (%)	0,0	0,0	0,0	83,3	0,0	90,5	69,5	85,7	0,0	100,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0
Nucleotide Diversity	0,0	0,0	0,0	0,057	0,0	0,055	0,088	0,150	0,0	0,053	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,276	0,0

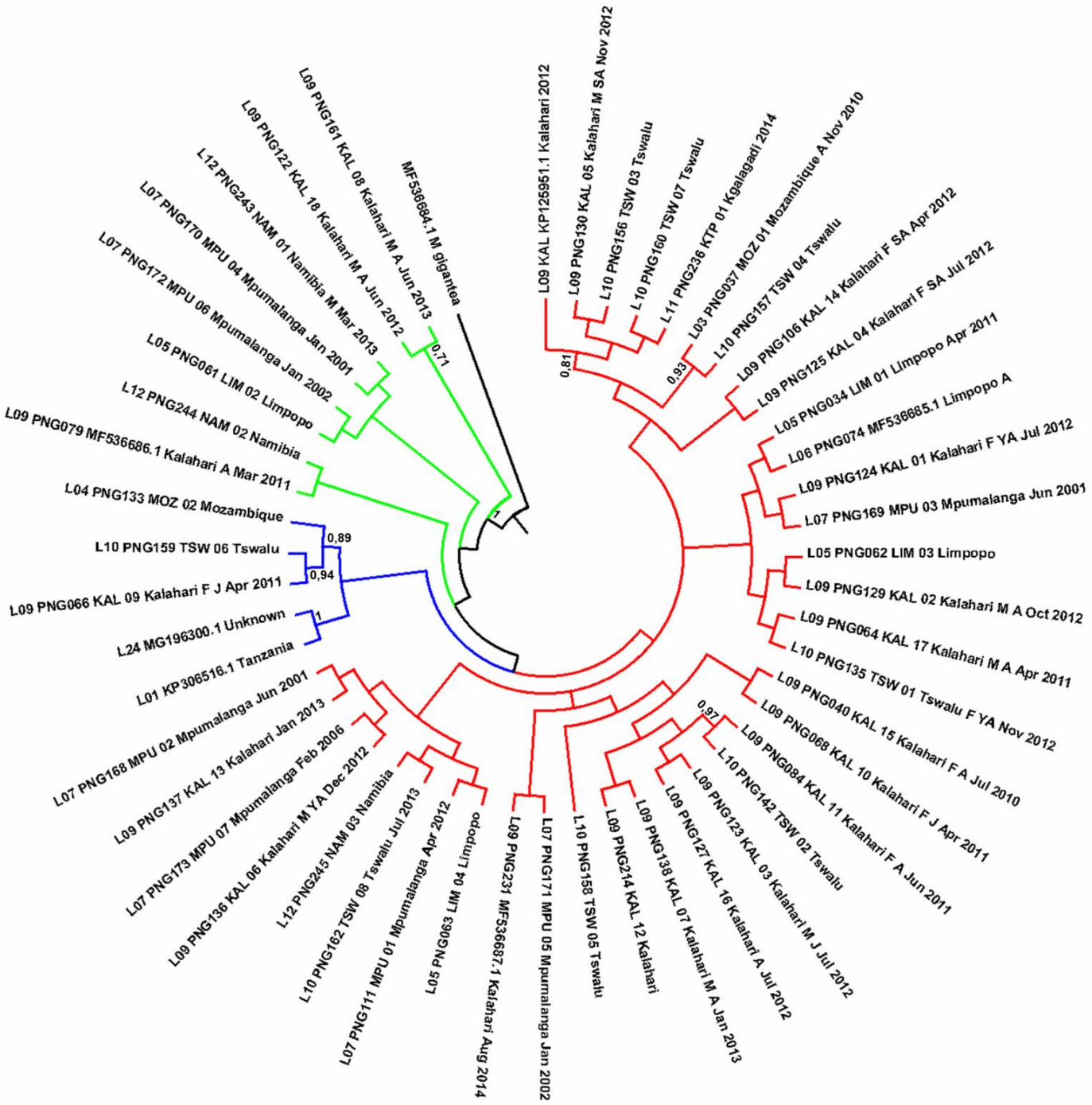
Supplementary Table 4.4. Haplotype layout for the three combined loci (*CO1*, *Cytb* and control region) illustrating the number of haplotypes shared between individuals as the first value; the proportion (reported to second decimal) of unique haplotypes as the second value; haplotype diversity and nucleotide diversity in the different localities observed during this study for *Smutsia temminckii*.

Haplotype	L01 (1)	L03 (1)	L04 (1)	L05 (4)	L06 (1)	L07 (7)	L09 (21)	L10 (8)	L11 (1)	L12 (3)	L13 (1)	L14 (1)	L15 (1)	L16 (1)	L17 (1)	L18 (1)	L19 (1)	L20 (1)	L21 (1)	L22 (1)	L23 (4)	L24 (1)
HAP 1	1/1,00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1/1,00
HAP 2	-	-	-	-	-	-	2/0,09	1/0,13	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 3	-	-	-	1/0,25	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 4	-	1/1,00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 5	-	-	-	-	-	-	7/0,33	1/0,13	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 6	-	-	-	1/0,25	-	1/0,14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 7	-	-	-	1/0,25	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 8	-	-	-	1/0,25	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 9	-	-	-	-	-	-	1/0,05	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 10	-	-	-	-	-	-	1/0,05	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 11	-	-	-	-	-	-	-	-	-	-	-	-	-	1/1,00	-	-	-	-	-	-	-	-
HAP 12	-	-	-	-	1/1,00	1/0,14	2/0,09	2/0,25	-	-	1/1,00	-	-	-	-	-	-	-	-	-	-	-
HAP 13	-	-	-	-	-	-	1/0,05	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1/1,00	-	-	-	-	-
HAP 15	-	-	-	-	-	1/0,14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 16	-	-	-	-	-	-	-	-	-	-	-	1/1,00	-	-	-	-	-	-	-	-	-	-
HAP 17	-	-	-	-	-	-	2/0,09	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 18	-	-	-	-	-	-	1/0,05	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 19	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1/0,25	-
HAP 20	-	-	-	-	-	-	2/0,09	1/0,13	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 21	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1/1,00	-	-	-	-	-	-	-
HAP 22	-	-	1/1,00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1/1,00	-	-	-	-
HAP 23	-	-	-	-	-	-	1/0,05	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 24	-	-	-	-	-	-	1/0,05	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 25	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1/0,25	-
HAP 26	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1/1,00	-	-
HAP 27	-	-	-	-	-	-	-	1/0,13	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 28	-	-	-	-	-	-	-	1/0,13	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 29	-	-	-	-	-	-	-	1/0,13	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 30	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1/0,25	-
HAP 31	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1/0,25	-
HAP 32	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1/1,00	-	-	-
HAP 33	-	-	-	-	-	1/0,14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

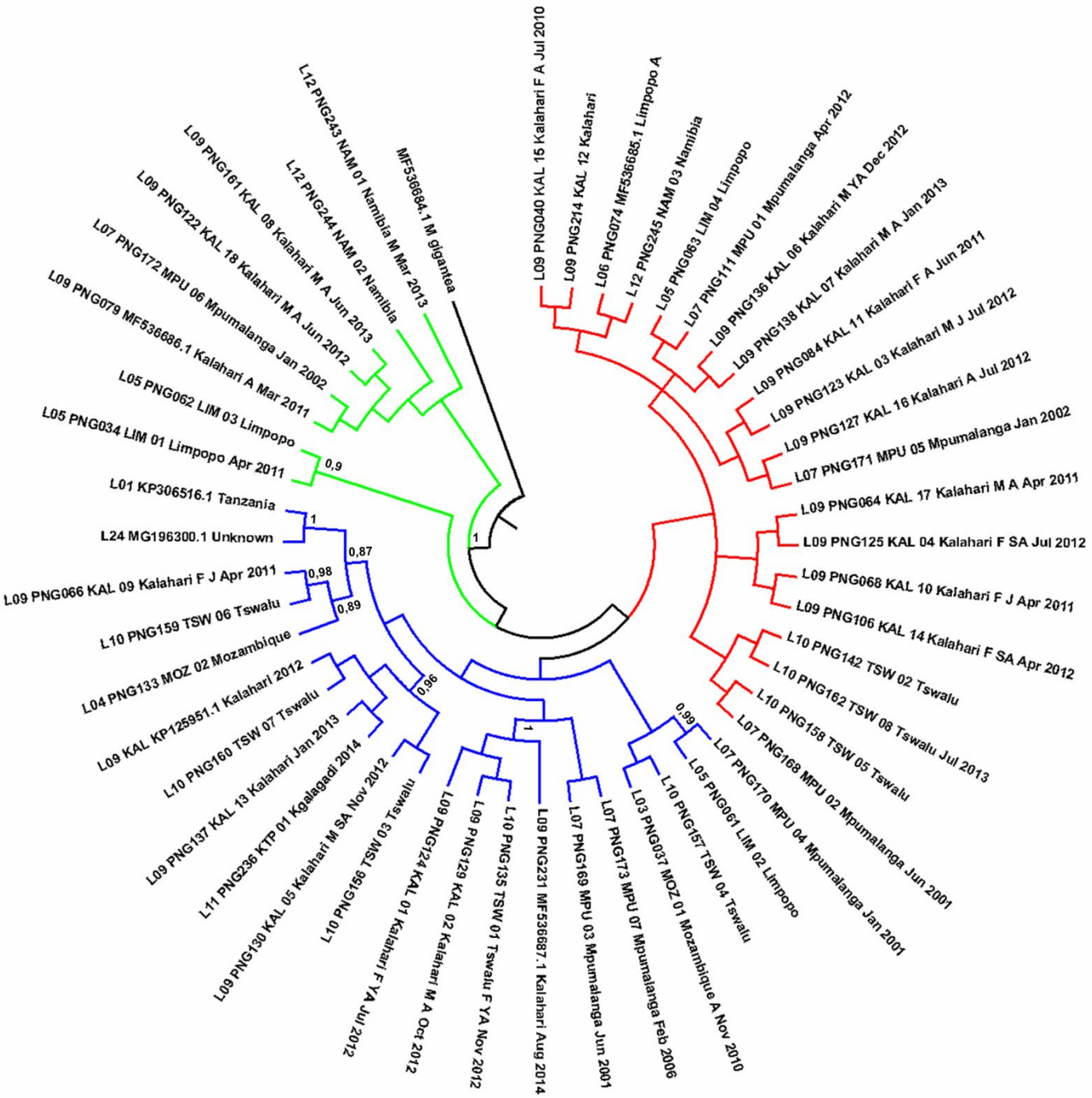
HAP 34	-	-	-	-	-	2/0,28	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 35	-	-	-	-	-	1/0,14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 36	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1/1,00	-	-	-	-	-	-
HAP 37	-	-	-	-	-	-	-	-	1/1,00	-	-	-	-	-	-	-	-	-	-	-	-	-
HAP 38	-	-	-	-	-	-	-	-	-	1/0,33	-	-	-	-	-	-	-	-	-	-	-	-
HAP 39	-	-	-	-	-	-	-	-	-	1/0,33	-	-	-	-	-	-	-	-	-	-	-	-
HAP 40	-	-	-	-	-	-	-	-	-	1/0,33	-	-	-	-	-	-	-	-	-	-	-	-
HAP 41	-	-	-	-	-	-	-	-	-	-	-	-	1/1,00	-	-	-	-	-	-	-	-	-
Haplotype Diversity (%)	0,0	0,0	0,0	100,0	0,0	95,2	88,1	96,4	0,0	100,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0
Nucleotide Diversity	0,0	0,0	0,0	0,077	0,0	0,057	0,087	0,135	0,0	0,074	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,260	0,0



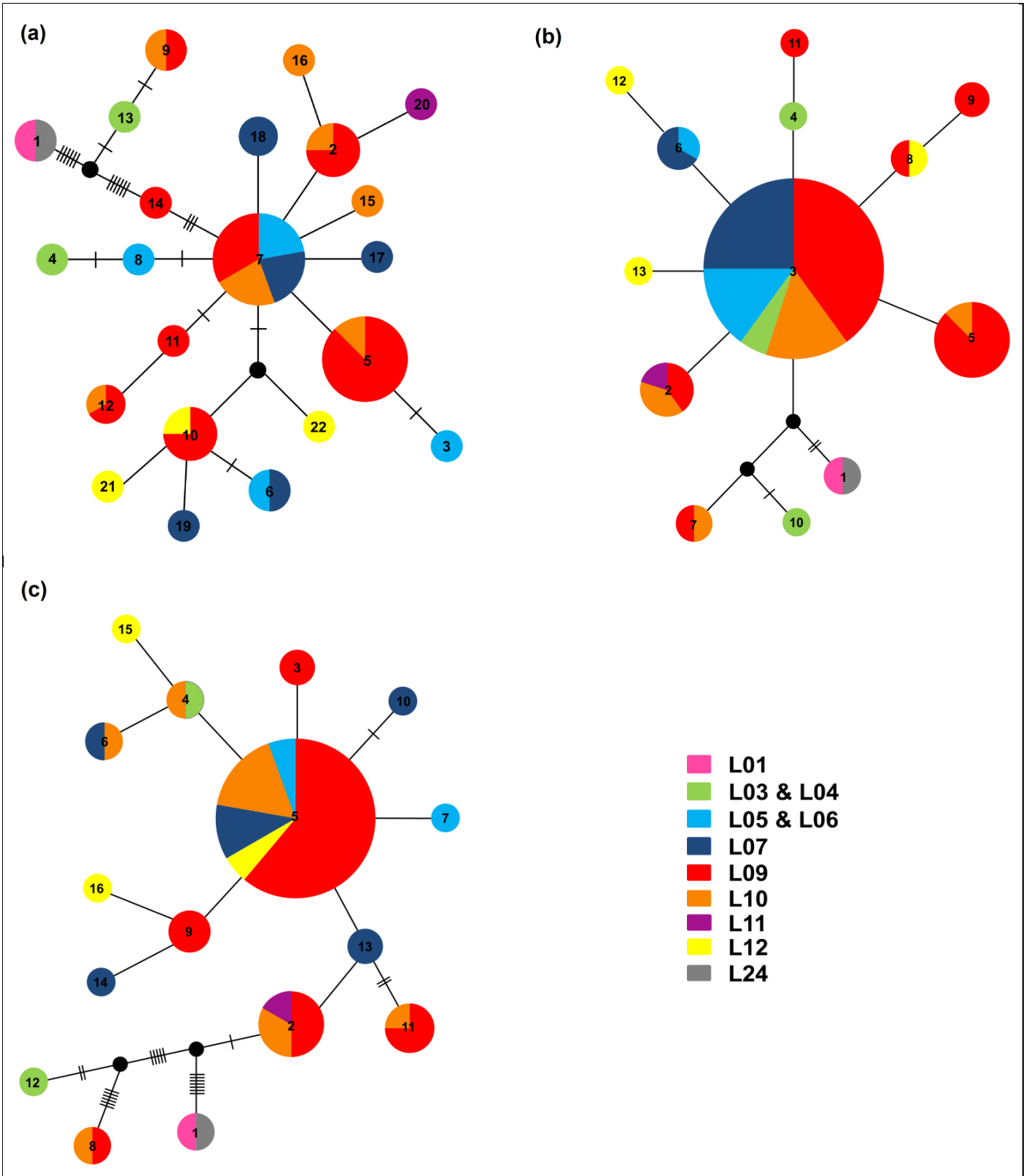
Supplementary Figure 4.1. Phylogenetic tree of *Co1* loci of 49 individuals, representing *Smutsia temminckii*. Bayesian Inference (BI) posterior probability is noted on the branches with BI ≥ 0.7 indicating strong support.



Supplementary Figure 4.2. Phylogenetic tree of *Cytb* loci of 49 individuals, representing *Smutsia temminckii*. Bayesian Inference (BI) posterior probability is noted on the branches with BI ≥ 0.7 indicating strong support.



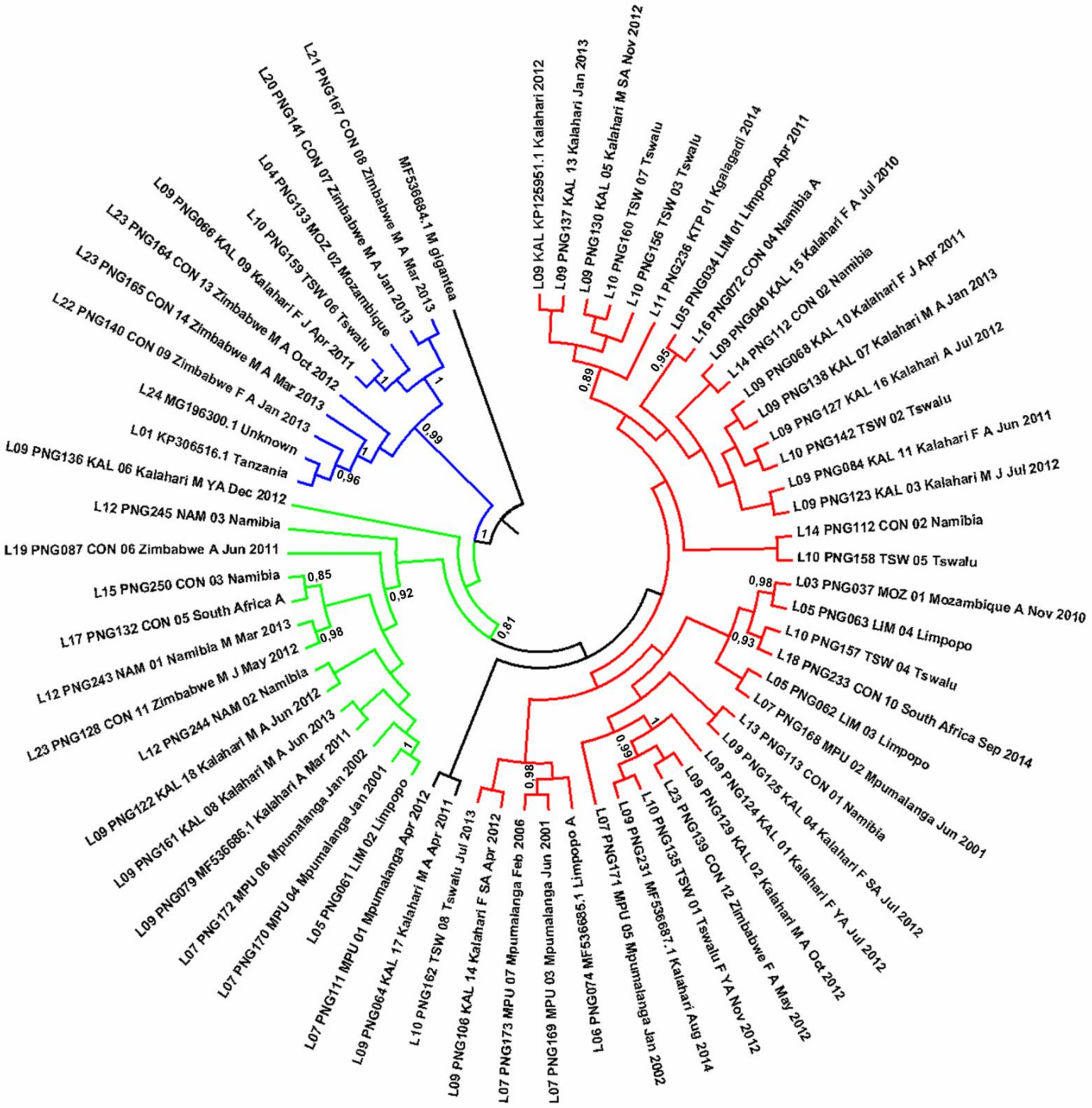
Supplementary Figure 4.3. Phylogenetic tree of control region loci of 49 individuals, representing *Smutsia temminckii*. Bayesian Inference (BI) posterior probability is noted on the branches with BI ≥ 0.7 indicating strong support.



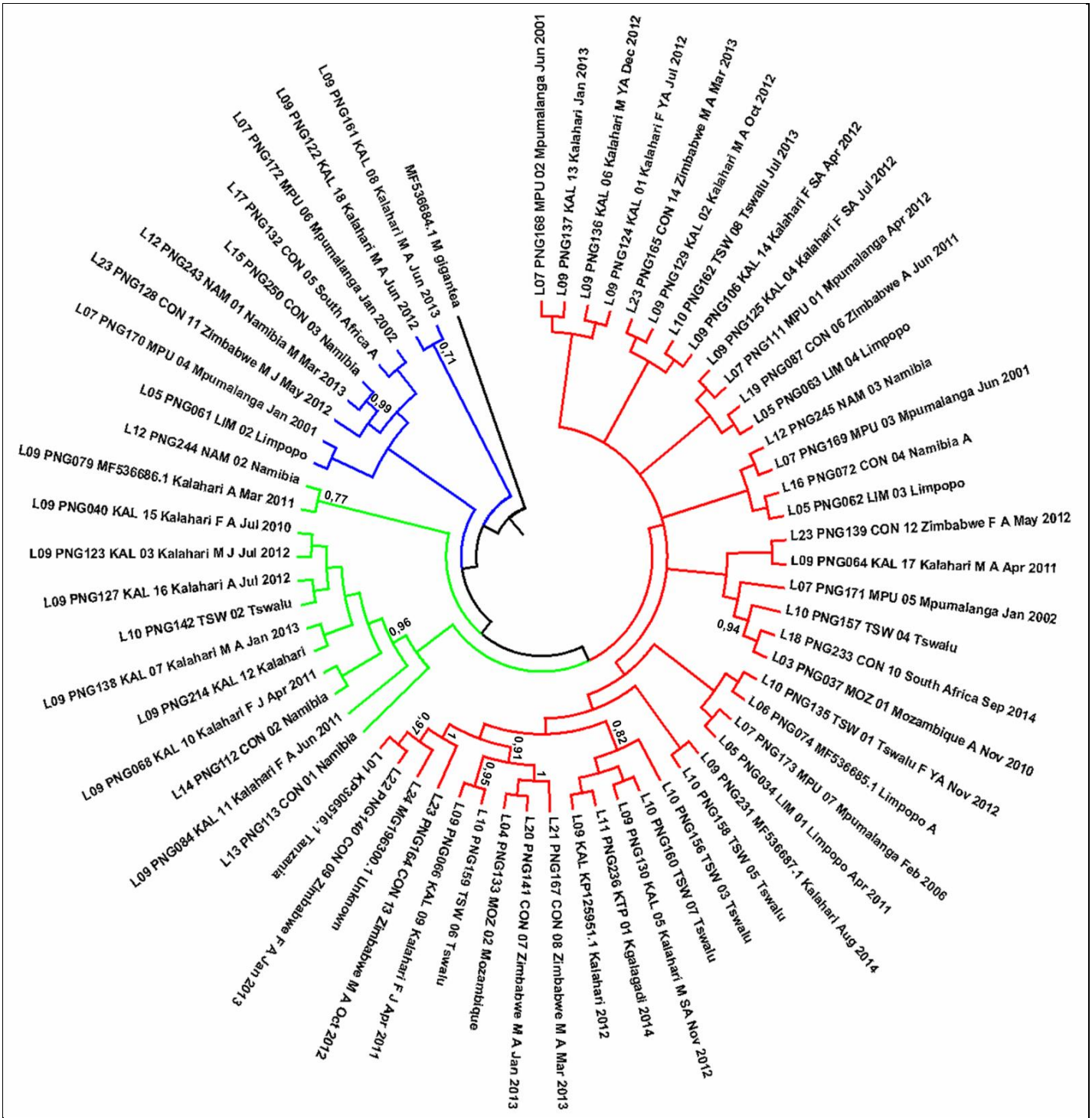
Supplementary Figure 4.4. Haplotype network for 49 samples of *Smutsia temminckii*: (a) *Co1*, (b) *Cytb* and (c) control region. Circles represent different haplotypes and the size of the circle indicates the number of individuals that share that haplotype. Colours indicate the sampling localities of the haplotype. The black bars indicate the amount of mutations each haplotype underwent and the black circles indicate missing haplotypes.

Supplementary Table 4.5. Spatial analyses of molecular variance (SAMOVA) for the three loci *CO1*, *Cytb* and control region using 49 and 63 individuals respectively for *Smutsia temminckii*. Analysis for *K* = 2–4 were run for each locus.

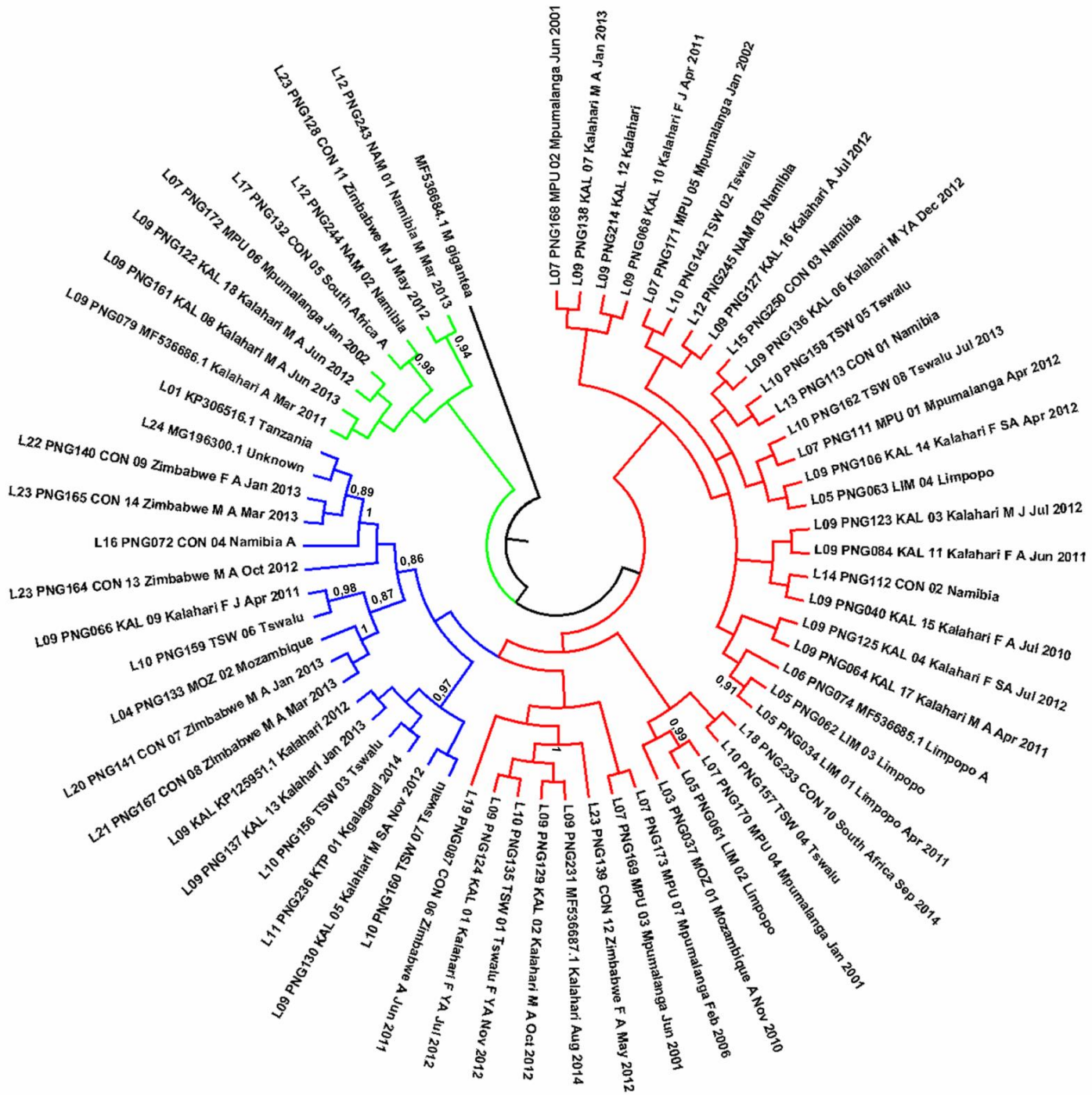
Source of variation						
N = 49						
	CO1		***Cytb***		**Control Region**	
	% variation among groups	**p-value (± Standard Deviation)**	**% variation among groups**	**p-value**	**% variation among groups**	**p-value (± Standard Deviation)**
Two groups (*K* = 2)	70,88	0,092 (0,008)	66,78	0,089 (0,009)	71,65	0,081 (0,0)
Three groups (*K* = 3)	70,63	0,010 (0,003)	69,65	0,010 (0,003)	74,56	0,010 (0,0)
Four groups (*K* = 4)	64,35	0,005 (0,002)	62,22	0,001 (0,001)	68,90	0,006 (0,0)
N = 63						
	CO1		***Cytb***		**Control Region**	
	% variation among groups	**p-value**	**% variation among groups**	**p-value**	**% variation among groups**	**p-value**
Two groups (*K* = 2)	64,83	0,005 (0,002)	68,57	0,001 (0,001)	68,79	0,003 (0,002)
Three groups (*K* = 3)	68,58	0,0 (0,0)	73,42	0,0 (0,0)	74,0	0,0 (0,0)
Four groups (*K* = 4)	68,15	0,0 (0,0)	73,04	0,0 (0,0)	72,84	0,0 (0,0)



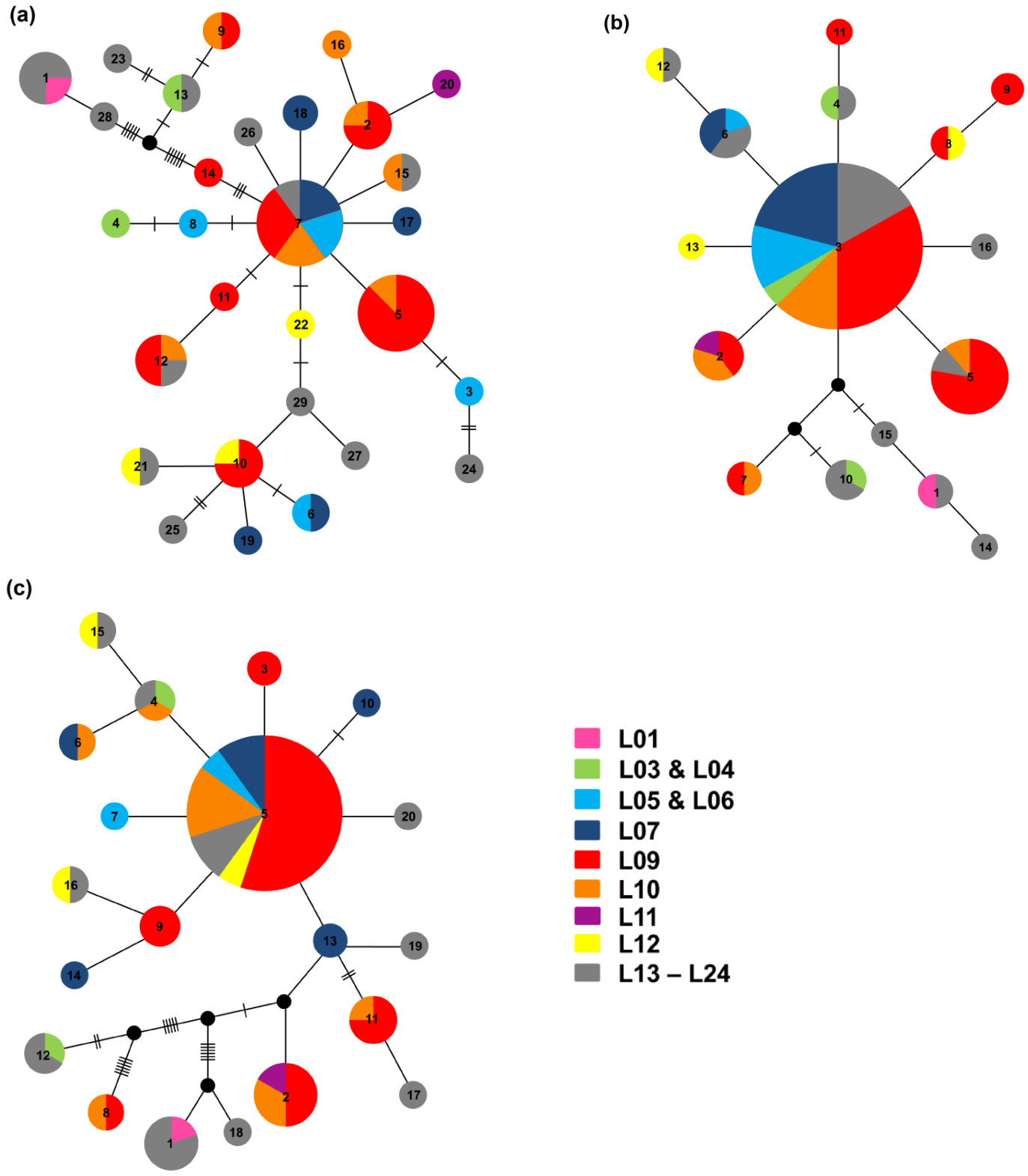
Supplementary Figure 4.5. Phylogenetic tree of *Co1* loci of 63 individuals, representing *Smutsia temminckii*. Bayesian Inference (BI) posterior probability is noted on the branches with BI ≥ 0.7 indicating strong support.



Supplementary Figure 4.6. Phylogenetic tree of *Cytb* loci of 63 individuals, representing *Smutsia temminckii*. Bayesian Inference (BI) posterior probability is noted on the branches with ≥ 0.7 BI indicating strong support.



Supplementary Figure 4.7. Phylogenetic tree of control region of 63 individuals, representing *Smutsia temminckii*. Bayesian Inference (BI) posterior probability is noted on the branches with BI ≥ 0.7 indicating strong support.



Supplementary Figure 4.8. Haplotype network for 63 samples of *Smutsia temminckii*: (a) *Co1*, (b) *Cytb* and (c) control region. Circles represent different haplotypes and the size of the circle indicates the number of individuals that share that haplotype. Colours indicate the sampling localities of the haplotype. The black bars indicate the amount of mutations each haplotype underwent and the black circles indicate missing haplotypes.

CHAPTER FIVE

GENERAL DISCUSSION, CONCLUSION AND RECOMMENDATIONS

5.1. INTRODUCTION

This thesis builds on previous work completed in 2014 which investigated population dynamics of Temminck's pangolin in southern Africa. The 2014 study consisted of two parts. Phylogenetic investigation of a full Temminck's pangolin mitochondrial genome sequence revealed that the order Pholidota was more closely related to the order Carnivora than to Xenarthra. Furthermore, the results also indicate a Laurasian origin (c.a 87 mya) and potential Paleocene migration to Africa (c.a 55 mya). The second part of the study analysed three mitochondrial genes (CO1, Cytb and D-loop) in 25 Temminck ground pangolin samples collected from Namibia, Zimbabwe, Mozambique and South Africa (Mpumalanga and the Northern Cape Provinces). Phylogeographic analyses provide support for two clusters; a Zimbabwe – Mozambique cluster and a South Africa – Nambia cluster. The differentiation of these clusters may be attributed to the formation of an ecological barrier, the Mega Kalahari Sand Sea.

The study presented here is an expansion of the previous study conducted in 2014 and presents the first in-depth analysis into the population structure of Temminck's pangolin in southern Africa using various molecular methods. Here, an extensive literature review was conducted (Chapter 1) on pangolins globally with specific reference to the African species which revealed critical gaps in our knowledge. This research showed that significant gaps such as SNPs and whole genome assembly were present with regards to the African pangolin species as they are the least studied compared to the Asian species. Therefore, the current study aimed to fill some of the void by generating the first set of Short Tandem Repeats for Temminck's pangolin with successful application for cross-species amplification in the remaining three African species (Chapter 2). Another pertinent void was to sequence the first accurate reference mitogenomes of all four African pangolin

species using NGS technologies and doing an in-depth comparison on molecular level between the different species (Chapter 3). This led to the first study to use both STRs and mtDNA markers to assess the population structure of Temminck's pangolin in southern Africa using a georeferencing approach and assess the application for determining the origin of confiscated samples which would aid forensic investigations (Chapter 4). The aims of this study were fulfilled for each research theme as detailed below.

5.2. DEVELOPMENT AND OPTIMIZATION OF STRS

The current study developed 30 STR markers using NGS technology for the Temmincks pangolin. The markers were tested on a population of 32 Temminck's, nine black-bellied, 10 white-bellied and 10 giant pangolins to optimize cross-species applicability for all African species. Of the 30 markers, 26 (87%) successfully amplified in the giant pangolin which was expected as both Temminck's and the giant pangolin are ground-dwelling species which are closely related based on various mitochondrial studies. In comparison, amplification of markers for the black and white-bellied pangolins were markedly lower (18 out of 30 or 60%). This would be expected as the above are both arboreal species and show a significant genetic distance to the ground-dwelling species as found in previous studies. A total of 11 markers were found to be polymorphic across all four African pangolin species. Developing species specific STRs for African pangolin species is important as these animals are a significant part of the illegal trade and large areas are being identified as traffic hotspots such as Nigeria, Uganda and Cameroon (Heinrich *et al.* 2017).

Prior to this study, STRs have only been developed for the Asian species and with limited success in African species (Luo *et al.* 2007; De Beer 2013). This created the opportunity to develop STRs for the African species especially Temminck's pangolin, the solitary pangolin species inhabiting southern Africa. During this process we also managed to optimize a large amount of the markers for cross-amplification in the remaining African species as described above. This would be beneficial for forensic application where species identification of confiscated African pangolin scale samples is crucial for the identification of traffic routes and hotspots in Africa. It can also assist in phylogeography/geo-referencing studies of the African species, in

particular Temminck's pangolin and to evaluate the phylogeographic distribution of this species in southern Africa.

5.3. MITOGENOME ANALYSIS OF AFRICAN PANGOLIN SPECIES

This study presents four mitogenomes from the four African pangolin species using NGS technologies. The results showed mitogenomes with the following lengths: Temminck's pangolin (16 558 bp), giant pangolin (16 540 bp), white-bellied pangolin (16 565 bp) and black-bellied pangolin (16 649 bp). The phylogenetic study, which included references from the available Asian species, indicated three lineages with high posterior probabilities which clearly supported the classification of the three genera, *Smutsia*, *Phataginus* and *Manis*.

Evaluating the mitogenomes of the four African pangolin species, the GC content percentages were found to be similar. When comparing the codon usage of the Asian and African pangolin species, the two groups varied significantly from each other. This could be a result of natural selection or mutation pressure since the Asian and African pangolins split thousands of years ago (Du Toit *et al*. 2014). The control region of the black-bellied pangolin comprised of two insertions, absent from the rest of the three African and two Asian species. This could be due to the black-bellied pangolin being under selection pressure from a recent mutation event.

Mitogenomes are seen as relevant research tools as it gives a glimpse into the maternal lineage of organisms in order to determine the species origin in accordance with fossil records, if any is available. It is a quick and affordable method to use in order to get access to genomic data of species which haven't been genotyped before, expanding the number of genetic loci available for phylogenetic analysis. This in turn will allow more genomic information to become available for the studied species which could determine the direction into which the follow up research questions could develop for the species.

Prior to this study, there were only two mitogenomes for Asian and one for the African pangolin species available (Arnason *et al.* 2002; Du Toit *et al.* 2014). The previous studies showed that the order Pholidota was closely related to the order

Carnivora rather than other Myrmecophagy orders such as Xenarthra. These findings revolutionized the evolutionary concept of the order Pholidota and confirmed the segregation of the order into three genera. The results presented above represent the first mitogenomes for three of the four African species, as Temminck's have been completed in a previous study. These mitogenomes, together with the available Asian genomes provided the first in-depth look on genomic level between all three genera of pangolin species, which showed that some of the species might have had a more recent evolutionary history compared to the rest of the species.

5.4. GEO-REFERENCING, GENETIC STRUCTURE AND DIVERSITY

Mitochondrial DNA can be adequate for discriminating between different populations within a species and can be used to ascribe new sub-species; however, in more closely related populations with high gene flow, it proves to be insufficient on its own. Short tandem repeats are also used for determining different heredities and population structure within species. A combination of both marker types was used to assess population structure of Temminck's pangolin in southern Africa.

The results show that there was some level of sub-structuring present with three distinctive clusters present across all the different datatypes. However, the clustering is not geographically informative, but rather supports overlap of different regions. This indicates a high level of gene flow between individuals across the distribution range. The low differentiation could also be inclusive of ancestral traits inherited by juveniles as they disperse to establish new home ranges. Another possibility could be due to insufficient data and low to no individual representation of the different localities. With a wider representation of individuals within the distribution range, it is expected that the different clustering and groupings could be better differentiated and a true representation of the population structure can be obtained. However, with the limited genetic differentiation currently shown, if this is a true representation of the general overall population structure between Temminck's pangolin in southern Africa, then the conservation strategies and plans for this species may be relatively more simplified.

This study contributed to the understanding of the genetic basis of extant populations highlighting the fact that the impact of inbreeding depression and bottleneck effects are limited, in the southern African distribution range of Temminck's pangolin populations This will be important for the implementation of future conservation strategies to identify diversity hotspots and the impact that reintroduction of a particular individual(s) may have on the extant population. It will also assist forensic cases where re-habilitated and protected areas can be identified for re-introductions of confiscated animals. This is the first study to combine both mitochondrial and STR data for a geo-referencing approach in the order Pholidota. Previous studies have only focussed on mitochondrial markers in order to asses population differentiation in the white-bellied and Temminck's pangolin species (Du Toit 2014; Gaubert *et al.* 2016).

5.5. CONSERVATION IMPLICATIONS

Pangolins are a group of mammals which are closely related but still show significant differentiation on a species level. This would suggest a species-level conservation and protection strategy compared to a global implementation. The large distribution ranges also presents a challenge in establishing a uniform conservation strategy among the different countries and areas. This suggests that each range-state should work on implementing their own strategy for conservation in correspondence to the bigger collaborative effort. It would also be beneficial when involving local communities to get involved in the conservation effort by creating group affiliations and awareness which highlights the importance of the species according to their different belief systems and allow the *muthi* markets to be more sustainable. They are also slow breeding mammals with only one pup born annually and sometimes bi-annually, which presents a greater challenge in the population growth and contributes to the decline of the population's numbers. Unfortunately, no meaningful structure and population assignments can be determined within Temminck's pangolins using microsatellite markers which impedes the assessment of hotspots used for illegal trade of individuals. This means that determination of the origin of confiscated samples is not possible currently using microsatellite markers alone.

However, as a result of the weak genetic structure observed in Temminck's pangolin in southern Africa, release and rehabilitation of individuals confiscated from illegal trades and markets is simplified. This indicates that the rehabilitated individuals can all be relocated to selected conservation areas which is protected from threats and also provides them with enough food and shelter needed to re-establish their home ranges. Recently, rescued pangolins have also been reintroduced in areas where they were extinct and are thriving since the habitat is similar to the current ones and adhering to all their basic needs such as food and shelter. This could revolutionise the conservation aspect of this species as more habitats could be made available for reintroductions based on historic distribution which would allow the populations to be belter monitored and protected from agriculture, deforestation, urbanisation and illegal trade.

5.6. LIMITATIONS OF THIS STUDY

As prominent with most research goals, there were some limitations observed during this study. Pangolins are a difficult species to study as they are secretive and elusive animals with large distribution ranges, especially Temminck's pangolin (distribution range from southern Africa into eastern Africa). This extensive range is also sparsely studied and limited surveys have been performed, which create data deficient gaps. Historic data from indigenous tribes and populations along with *muthi* markets indicate a wide distribution and occurrence across the distribution range; however, it is unknown what the true population statuses represent without the survey data.

In lieu of the speculative nature of the true population size caution should be taken when applying conservation strategies. It is strongly recommended that a wider sampling strategy needs to be implemented especially for the regions which have sparse representation as well as a larger and more representative sample set from existing areas. The usage of dried specimens such as museum samples were also difficult to obtain as very few specimens is available globally pertaining to the four different African pangolin species. The limited specimens which were available had insufficient locality and sampling information to use for geographic studies. In addition, one of the caveats was the quantity of sample material available, which may be mitigated by advances in extraction technology.

Metabarcoding (eDNA) is the study of environmental DNA to determine species composition for a specific region or area instead of individually collected samples from individuals. This could be considered as an alternative to study elusive or endangered animals especially in distribution ranges which are difficult to access. Since environmental samples are collected, it reduces the effort and sampling costs for projects and allows a variety of different species to be assessed. As a result, this could shed light on distribution patterns and behaviour of different species including pangolins. Successful identification of various small and medium mammal species have been determined in conjunction with camera and other traps. The use of e-DNA was successful in identifying additional mammal species which wasn't captured on the camera or other traps. This suggests that this method could be a valuable tool for future monitoring and assessing of pangolin populations without collecting individual samples for DNA extraction (Leempoel *et al.* 2020; Sales *et al.* 2020; Mena *et al.* 2021).

5.7. FUTURE RESEARCH AREAS

As previously indicated, forensic investigations could benefit significantly from the STR panel developed. This panel contains markers that can be used for inter- and intra-species differentiation. This panel can be validated by establishing a reference database of known localities and species among all four African pangolin species. This in turn can be used to identify confiscated pangolin scale samples obtained from recent illegal seizures and determine the area from where they have been sourced providing valuable insights into poaching and trafficking hotspots. Geo-referencing studies of Temminck's and other pangolin species can be further enhanced by developing SNP markers to represent a larger part of the genome. This in parallel to whole genome sequencing will provide a more holistic approach to determining different population groups. Developing SNPs and assembling the genomes of pangolins have already been implemented for some of the Asian species (Choo *et al.* 2016; Li *et al.* 2017; Hu *et al.* 2020).

In addition, identifying genes related to specific traits such as adaptation to climate change or evolutionary adaptions (loss of teeth and hair) could shed light on distinct evolutionary paths of African pangolin species (Meredith *et al.* 2014; Choo *et al.*

2016). A recent article published by Heighton & Gaubert (2021) compared all the available literature and data for the order Pholidota across a variety of study fields such as ecology, parasitology, behaviour, distribution, molecular and conservation to mention but a few. The authors determined what studies have been performed for each of the eight extant pangolin species as well as which areas are still deficient in data. In addition, future research areas were also identified that will facilitate a better understanding of this order and be able to implement an appropriate conservation plan for these highly trafficked mammals (Heighton & Gaubert 2021).

5.8. CONCLUDING REMARKS

The results reported here provided valuable insights into the population dynamics of Temminck's pangolin as well as the relationship between the four African species. This understanding will provide significant insight in developing and advancing conservation strategies to curb illegal trading activities and comprehend the effects of the *muthi* markets on different pangolin populations as the high genetic diversity permits sufficient gene flow to progress between populations of Temminck's pangolin, preventing inbreeding bottlenecks. This in turn allows the species to be rehabilitated and reintroduced into conservation areas which is fit for providing the necessary protection they need. Using this as a baseline for all the countries to support their own similar conservation plan along with the assistance of local tribes would have a significant impact on the protection of this wonderful introverted awkward animal to be preserved for all future generations.

SUMMARY

The order Pholidota includes the most trafficked mammal species globally. The family Maninae with recognized species are categorized as *Endangered* and *Critically Endangered* on the IUCN Red List of Threatened Species. Due to the sharp decline in Asian pangolin numbers, the shift was towards the African pangolin species, to meet the demand in Asia and in Africa. Subsequently, the status of Smutsiinae has been revised by the IUCN and varies between *Vulnerable* and *Endangered*. This study contributes to clarify the genetic status of the four African pangolin species using molecular technologies.

The first part of the study aimed to assemble mitogenomes of the four African pangolin species using NGS technologies. Whole mitogenomes were assembled and compared with published mitogenomes available in online databases. Mitogenomes were only available for six of the eight pangolin species. Our results showed different genome lengths for the four species: Temminck's pangolin (16 558 bp), giant pangolin (16 540 bp), white-bellied pangolin (16 565 bp) and black-bellied pangolin (16 649 bp). Three distinctive clusters were observed supporting three genera. The first cluster consisted of the Asian pangolins (*Manis*), the second cluster was the African tree pangolins (*Phataginus*) and the third cluster represented the African ground pangolins (*Smutsia*). There were also two insertions found in the control region of the black-bellied pangolin which could indicate a recent mutation or selection event.

The second part of the study focused on the development of species-specific STRs for Temminck's pangolin and in combination with previous developed mtDNA markers the aim was to assess the population structure across the distribution range in southern Africa. Three mtDNA loci were analysed (*Co1, Cytb*, control region) with 30 STRs. A total of 62 individuals from six countries were included in the study. The results showed little to no geographical differentiation across the identified sampling locations. This indicates high levels of gene flow between the populations. Three distinctive clusters were observed within the different sampling areas. This could indicate inherited ancestral traits in juveniles with dispersal into new home ranges

following a stepping stone model. Temminck's pangolin has been reported to cover vast distances in search of new home ranges or territories.

This study provides the first account of a combination of mtDNA and STR markers to investigate possible geo-referencing of Temminck's pangolin in southern Africa. It is evident that the mtDNA and STRs markers used in this study, are not robust enough to discriminate between different sampling locations or populations. Further in-depth investigations, including whole genome analysis and SNPs are recommended. A more comprehensive sampling strategy is required to include all areas within the distribution range as well as more individuals for an in-depth genome analyses.

KEYWORDS: mitochondrial genomes, microsatellites, geo-referencing, African pangolins

REFERENCES

Aguillon S, Dipita AD, Lecompte E, Missoup AD, Tindo M, Gaubert P. 2020. Development and characterization of 20 polymorphic microsatellite markers for the white-bellied pangolin *Phataginus tricuspis* (Mammalia, Pholidota). *Molecular Biology Reports* **47**:4827–4833.

Akaike H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**:716–723.

Akpona HA, Djagoun CAMS, Sinsin B. 2008. Ecology and ethnozoology of the three-cusped pangolin *Manis tricuspis* (Mammalia, Pholidota) in the Lama forest reserve, Benin. *Mammalia* **72**:198–202.

Ali OA, O'Rourke SM, Amish SJ, Meek MH, Luikart G, Jeffres C, Miller MR. 2016. RAD capture (Rapture): flexible and efficient sequence-based genotyping. *Genetics* **202**:389–400.

Allen GM, Coolidge HJ. 1940. Mammal collections of the Asiatic primate expeditions. *Bulletin of the Museum of Comparative Zoology, Harvard* **97**:131–166.

Alves RRN, Rosa IL. 2005. Why study the use of animal products in traditional medicines? *Journal of Ethnobiology and Ethnomedicine* **1**:5.

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Available from http://dnacore.missouri.edu/PDF/FastQC_Manual.pdf (accessed July 18, 2016).

Angelici FM, Grimod I, Politano E. 1999. Mammals of the Eastern Niger Delta (Rivers and Bayelsa States, Nigeria): An environment affected by a gas-pipeline. *Folia Zoologica* **48**:249–264.

APWG. 2020a. Temminck's ground pangolin (*Smutsia temminckii*). Available from https://africanpangolin.org/discover/temmincks-ground-pangolin/ (accessed September 13, 2020).

APWG. 2020b. Giant ground pangolin (*Smutsia gigantea*). Available from https://africanpangolin.org/discover/giant-ground-pangolin/ (accessed September 13, 2020).

APWG. 2020c. Black-bellied pangolin (*Phataginus tetradactyla*). Available from https://africanpangolin.org/discover/black-bellied-pangolin/ (accessed September 13, 2020).

APWG. 2020d. White-bellied pangolin (*Phataginus tricuspis*). Available from https://africanpangolin.org/discover/white-bellied-pangolin/ (accessed September 13, 2020).

Aquadro CF, Greenberg BD. 1983. Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. *Genetics* **103**:287–312.

Arnason U, Adegoke JA, Bodin K, Born EW, Esa YB, Gullberg A, Nilsson M, Short R V, Xu X, Janke A. 2002. Mammalian mitogenomic relationships and the root of the Eutherian tree. *Proceedings of the National Academy of Sciences of the United States of America* **99**:8151–8156.

Arnason U, Gullberg A, Janke A, Kullberg M, Lehman N, Petrov EA, Väinölä R. 2006. Pinniped phylogeny and a new hypothesis for their origin and dispersal. *Molecular Phylogenetics and Evolution* **41**:345–354.

Avise JC. 1998. The history and purview of phylogeography: a personal reflection. *Molecular Ecology* **7**:371–379.

Azhar B, Lindenmayer D, Wood J, Fischer J, Manning A, McElhinny C, Zakaria M. 2012. Contribution of illegal hunting, culling of pest species, road accidents and feral dogs to biodiversity loss in established oil-palm landscapes. *Wildlife Research* **40**:1–9.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* **3**:e3376.

Bandelt HJ, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**:37–48.

Beck A. 2008. Electric fence induced mortality in South Africa. University of the Witwatersrand, South Africa.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* **27**:573–580.

Bentzen P, Leggett WC, Brown GG. 1988. Length and restriction site heteroplasmy in the mitochondrial DNA of american shad (*Alosa sapidissima*). *Genetics* **118**:509–518.

Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, Pütz J, Middendorf M, Stadler PF. 2013. MITOS: Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution* **69**:313–319.

Bi SG, Sokouri DP. 2015. Nucleotide variation and selective pressure in the mitochondrial genome of African Elephants. *International Journal of Science and Research* **4**:1733–1742.

Birstein VJ, DeSalle R. 1998. Molecular phylogeny of Acipenserinae. *Molecular Phylogenetics and Evolution* **9**:141–155.

Birungi J, Arctander P. 2000. Large sequence divergence of mitochondrial DNA genotypes of the control region within populations of the African antelope, kob (*Kobus kob*). *Molecular Ecology* **9**:1997–2008.

Boakye MK, Kotzé A, Dalton DL, Jansen R. 2016. Unravelling the pangolin bushmeat commodity chain and the extent of trade in Ghana. *Human Ecology* **44**:257–264.

Boakye MK, Pietersen DW, Kotzé A, Dalton D-L, Jansen R. 2015. Knowledge and uses of African pangolins as a source of traditional medicine in Ghana. *PLOS ONE* **10**:e0117199.

Boakye MK, Pietersen DW, Kotzé A, Dalton DL, Jansen R. 2014. Ethnomedicinal use of African pangolins by traditional medical practitioners in Sierra Leone. *Journal of Ethnobiology and Ethnomedicine* **10**:76–85.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**:2114–2120.

Bon C, Berthonaud V, Maksud F, Labadie K, Poulain J, Artiguenave F, Wincker P, Aury J-M, Elalouf J-M. 2012. Coprolites as a source of information on the genome and diet of the cave hyena. *Proceedings of the Royal Society B: Biological Sciences* **279**:2825–2830.

Botha J, Gaudin T. 2007. An early pliocene pangolin (Mammalia; Pholidota) from Langebaanweg, South Africa. *Journal of Vertebrate Paleontology* **27**:484–491.

Bräutigam A, Howes J, Humphreys T, Hutton J. 1994. Recent information on the status and utilization of African Pangolins. *TRAFFIC Bulletin* **15**:15–22.

Brown JR, Beckenbach K, Beckenbach AT, Smith MJ. 1996. Length variation, heteroplasmy and sequence divergence in the mitochondrial DNA of four species of sturgeon (*Acipenser*). *Genetics* **142**:525–535.

Brown WM, George M, Wilson AC. 1979. Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America* **76**:1967–1971.

Burgener M, Hübner P. 1998. Mitochondrial DNA enrichment for species identification and evolutionary analysis. *Zeitschrift fur Lebensmittel-Untersuchung und -Forschung A* **207**:261–263.

Burger PA, Steinborn R, Walzer C, Petit T, Mueller M, Schwarzenberger F. 2004. Analysis of the mitochondrial genome of cheetahs (*Acinonyx jubatus*) with neurodegenerative disease. *Gene* **338**:111–119.

Buroker NE, Brown JR, Gilbert TA, O'Hara PJ, Beckenbach AT, Thomas WK, Smith MJ. 1990. Length heteroplasmy of sturgeon mitochondrial DNA: an illegitimate elongation model. *Genetics* **124**:157–163.

Canino MF, Spies IB, Lowe SA, Grant WS. 2010. Highly discordant nuclear and mitochondrial DNA diversities in Atka mackerel. *Marine and Coastal Fisheries: Dynamics, Management, and Ecosystem Science* **2**:375–387.

Cano RJ, Poinar HN, Pieniazek NJ, Poinar A, Poinar GO. 1993. Amplification and sequencing of DNA from a 120–135-million-year-old weevil. *Nature* **363**:536–538.

Cao Y, Adachi J, Janke A, Pääbo S, Hasegawa M. 1994. Phylogenetic relationships among eutherian orders estimated from inferred sequences of mitochondrial proteins: instability of a tree based on a single gene. *Journal of Molecular Evolution* **39**:519–527.

Chakravorty J, Meyer-Rochow VB, Ghosh S. 2011. Vertebrates used for medicinal purposes by members of the Nyishi and Galo tribes in Arunachal Pradesh (North-East India). *Journal of Ethnobiology and Ethnomedicine* **7**:13–26.

Challender D, Nash H, Waterman C. 2019a. *Pangolins: Science, Society and Conservation*. (Nyhus PJ, editor), 1st edition. Academic Press, Elsevier, United Kingdom.

Challender D, Willcox DHA, Panjang E, Lim N, Nash H, Heinrich S, Chong J. 2019b. *Manis javanica*. Available from https://www.iucnredlist.org/species/12763/123584856#conservation-actions (accessed September 13, 2020).

Challender D, Wu S, Kaspal P, Khatiwada A, Ghose A, Sun NC-M, Mohapatra RK, Suwal TL. 2019c. *Manis pentadactyla. Available from https://dx.doi.org/10.2305/IUCN.UK.2019-3.RLTS.T12764A168392151.en (accessed September 13, 2020).*

Chao J. 2002. General status of Formosan pangolin, *Manis pentadactyla pentadactyla*. Taipei, Taiwan.

Choo SW et al. 2016. Pangolin genomes and the evolution of mammalian scales and immunity. *Genome Research* **26**:1312–1322.

CITES. 1995. Amendments to Appendices I and II of the Convention. Available from https://cites.org/sites/default/files/eng/cop/09/E9-Amend-to-AppI-II.pdf (accessed April 18, 2017).

CITES. 2016. Convention on International Trade in Endngered Species of Wild Fauna and Flora. Pages 1–28 Seventeenth Meeting of the Conference of the Parties (CoP17) (Johannesburg, South Africa), September 24 – October 5, 2016. Available from https://cites.org/sites/default/files/eng/cop/17/prop/CAR_etc_pangolins_E.pdf (accessed October 3, 2016).

CITES. 2017a. Conference of the Parties. Available from https://cites.org/eng/cop/index.php (accessed April 18, 2017).

CITES. 2017b. Appendices I, II and III. Available from https://www.cites.org/eng/app/appendices.php (accessed March 10, 2017).

Coggins C. 2003. The tiger and the pangolin: nature, culture, and conservation in China. *The Geographical Review* **0**:258–260.

Collins A, Lonjou C, Morton NE. 1999. Genetic epidemiology of single-nucleotide polymorphisms. *Proceedings of the National Academy of Sciences* **96**:15173–15177.

Collura R V, Stewart CB. 1995. Insertions and duplications of mtDNA in the nuclear genomes of Old World monkeys and hominoids. *Nature* **378**:485–489.

Corbet GB, Hill JE. 1991. *A World List of Mammalian Species*, 3rd edition. Natural History Museum Publications & Oxford University Press, London.

Coulson I. 1989. The pangolin (*Manis temminckii* Smuts, 1832) in Zimbabwe. African *Journal of Ecology* **27**:149–155.

Crozier RH. 1997. Preserving the information content of species: genetic diversity, phylogeny, and conservation worth. *Annual Review of Ecology and Systematics* **28**:243–268.

Cummings MP, Otto SP, Wakeley J. 1995. Sampling properties of DNA sequence data in phylogenetic analysis. *Molecular Biology and Evolution* **12**:814–822.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**:772–775.

Davey JW, Cezard T, Fuentes-Utrilla P, Eland C, Gharbi K, Blaxter ML. 2013. Special features of RAD Sequencing data: implications for genotyping. *Molecular Ecology* **22**:3151–3164.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* **12**:499–510.

Davies E. 2014. 'Shocking' scale of pangolin smuggling revealed. Available from http://www.bbc.co.uk/nature/26549963 (accessed February 24, 2017).

Davit-Béal T, Tucker AS, Sire J-Y. 2009. Loss of teeth and enamel in tetrapods: fossil record, genetic data and morphological adaptations. *Journal of Anatomy* **214**:477–501.

De Beer C. 2013. Insights into the genetics of the ground pangolin (*Smutsia temminckii*). University of the Free State, South Africa.

De Elera C. 1895. Cata´logo sistema´tico de toda la fauna de Filipinas: conocida hasta el presente, y a´ la vez el de la coleccio´n zoolo´gica del Museo de PP. ominicos del Colegio–Universidad de Sto. Toma´s de Manila, escrito con motivo de la Exposici, 1st edition. Imprenta del Colegio de Santo TomÁs, Manila, Philippines.

Desmarest AG. 1822. Mammalogie ou description des especes de mammiferes. Seconde partie, contenant les ordres de Rongeurs, des Edentes, des Pachydermes, des Ruminans et des Cetaces.

Development Core Team R. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Dickman CR, Richer RA. 2001. *Pangolins*. Pages 800–801 in D. W. MacDonald, editor. The Encyclopedia of Mammals, 1st edition. Facts on File, New York, USA.

Dixon M, Weiskotten C. 2016. CITES CoP17: Victory Today for Pangolins. Available from https://newsroom.wcs.org/News-Releases/articleType/ArticleView/articleId/9303/CITES-CoP17-Victory-Today-for-Pangolins.aspx (accessed October 20, 2016).

Doran GA, Allbrook DB. 1973. The tongue and associated structures in two species of African pangolins, *Manis gigantea* and *Manis tricuspis*. *Journal of Mammalogy* **54**:887–899.

Dou H, Zhang Y, Feng L. 2016. Complete mitochondrial genome of the Himalayan serow (*Capricornis thar*) and its phylogenetic status within the genus Capricornis. *Biochemical Systematics and Ecology* **65**:115–123.

Du Toit Z. 2014. Population genetic structure of the ground pangolin based on mitochondrial genomes. University of the Free State, South Africa.

Du Toit Z, Grobler JP, Kotzé A, Jansen R, Brettschneider H, Dalton D-L. 2014. The complete mitochondrial genome of Temminck's ground pangolin (*Smutsia temminckii*; Smuts, 1832) and phylogenetic position of the Pholidota (Weber, 1904). *Gene* **551**:49–54.

Du Toit Z, Grobler JP, Kotzé A, Jansen R, Dalton D-L. 2016. Scale samples from Temminck's ground pangolin (*Smutsia temminckii*): a non-invasive source of DNA. *Conservation Genetics Resources* **9**:1–4.

Earl DA, Von Holdt BM. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**:359–361.

Ellegren H. 2004. Microsatellites: simple sequences with complex evolution. *Nature Reviews: Genetics* **0**:435–445.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **6**:e19379.

Elsner J, Deschler-Erb S, Stopp B, Hofreiter M, Schibler J, Schlumbaum A. 2016. Mitochondrial d-loop variation, coat colour and sex identification of Late Iron Age horses in Switzerland. *Journal of Archaeological Science: Reports* **6**:386–396.

Emry RJ. 1970. A North American Oligocene pangolin and other additions to the Pholidota. *Bulletin of the American Museum of Natural History* **142**:459–510.

Esposti MD, De Vries S, Crimi M, Ghelli A, Patarnello T, Meyer A. 1993. Mitochondrial cytochrome b: evolution and structure of the protein. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1143**:243–271.

Esselstyn JA, Widmann P, Heaney LR. 2004. The mammals of Palawan Island, Philippines. *Proceedings of the Biological Society of Washington* **117**:271–302.

Etter PD, Bassham S, Hohenlohe PA, Johnson EA, Cresko WA. 2011. SNP discovery and genotyping for evolutionary genetics using RAD sequencing. *Methods in Molecular Biology* **772**:297–317.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology* **14**:2611–2620.

Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**:564–567. Blackwell Publishing Ltd.

Fang LX. 1981. Investigation on pangolins by following their trace and observing their cave. *Nature* **3**:64–66.

Fang LX, Wang S. 1980. A preliminary survey on the habits of pangolin. *Memoirs of Beijing Natural History Museum* **7**:1–6.

Foran DR, Hixson JE, Brown WM, Arbor A. 1988. Comparisons of ape and human sequences that regulate mitochondrial DNA transcription and D-loop DNA synthesis. *Nucleic Acids Research* **16**:5841–5861.

Friedman H. 2016. Do or die: deciding the pangolin's fate at CoP17. Available from http://untoldafrica.com/pangolins-no-more-trading-of-the-most-trafficked-mammal-we-have-never-heard-of/ (accessed March 10, 2017).

Gaubert P. 2011. *Family Manidae (Pangolins)*. Pages 82–103 in D. E. Wilson and R. A. Mittermeier, editors. Handbook of the Mammals of the World - Hoofed Mammals, 2nd edition. Lynx Edicions, Barcelona, Spain.

Gaubert P, Njiokou F, Olayemi A, Pagani P, Dufour S, Danquah E, Nutsuakor MEK, Ngua G, Missoup AD, Tedesco PA, Dernat R, Antunes A. 2015. Bushmeat genetics: Setting up a reference framework for the DNA typing of African forest bushmeat. *Molecular Ecology Resources* **15**:633–651.

Gaubert P, Njiokou F, Ngua G, Afiademanyo K, Dufour S, Malekani J, Bi SG, Tougard C, Olayemi A, Danquah E, Djagoun CAMS, Kaleme P, Mololo CN, Stanley W, Luo S-J, Antunes A. 2016. Phylogeography of the heavily poached African common pangolin (Pholidota, *Manis tricuspis*) reveals six cryptic lineages as traceable signatures of Pleistocene diversification. *Molecular Ecology* **25**:5975–5993.

Gaubert P, Antunes A, Meng H, Miao L, Peigné S, Justy F, Njiokou F, Dufour S, Danquah E, Alahakoon J, Verheyen E, Stanley WT, O'Brien SJ, Johnson WE, Luo S-J. 2018. The complete phylogeny of pangolins: Scaling up resources for the molecular tracing of the most trafficked mammals on earth. *Journal of Heredity* **109**:347–359.

Gaubert P, Antunes A. 2005. Assessing the taxonomic status of the Palawan pangolin *Manis Culionensis* (Pholidota) using discrete morphological characters. Journal of Mammalogy **86**:1068–1074.

Gaudin TJ, Emry RJ, Pogue B. 2006. A new genus and species of pangolin (Mammalia, Pholidota) from the Late Eocene of Inner Mongolia, China. *Journal of Vertebrate Paleontology* **26**:146–159.

Gaudin TJ, Emry RJ, Wible JR. 2009. The phylogeny of living and extinct pangolins (mammalia, pholidota) and associated taxa: A morphology based analysis. *Journal of Mammalian Evolution* **16**:235–305.

Gaudin TJ, Wible JR. 1999. The entotympanic of pangolins and the phylogeny of the Pholidota (Mammalia). *Journal of Mammalian Evolution* **6**:39–65.

Gebo DL, Rasmussen DT. 1985. The earliest fossil pangolin (Pholidota : Manidae) from Africa. *Journal of Mammalogy* **66**:538–541.

Geoffroy É. 1803. Catalogue des mammiferes du Museum national d'Histoire naturelle. Paris.

Gibb GC, Kardailsky O, Kimball RT, Braun EL, Penny D. 2007. Mitochondrial genomes and avian phylogeny: complex characters and resolvability without explosive radiations. *Molecular Biology and Evolution* **24**:269–280.

Graham CF et al. 2015. Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). *Molecular Ecology Resources* **15**:1304–1315.

Grassé PP. 1955. *Ordre de Pholidotes*. Pages 1267–1282 in P. P. Grassé, editor. Traité de Zoologie: Mammifères, 17th edition. Masson et Cie, Paris.

Gray JE. 1865. Revision of the genera and species of entomophagous Edentata, founded on the examination of the specimens in the British Museum. *Proceedings of the Zoological Society of London* **33**:359–386.

Grubb P, Jones TS, Davies AG, Edberg E, Starin ED, Hill JE. 1998. *Mammals of Ghana, Sierra Leone and the Gambia*. Trendrine Press, Zennor, St Ives, Cornwall, UK.

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate Maximum-Likelihood Phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* **59**:307–321.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52**:696–704.

Gurung KK, Singh R. 1996. Field guide to the mammals of the Indian subcontinent. Academic Press, San Diego, California, USA.

Hance J. 2014. Over a million pangolins slaughtered in the last decade. Available from https://news.mongabay.com/2014/07/over-a-million-pangolins-slaughtered-in-the-last-decade/ (accessed March 10, 2017).

Hari R, Paterson IC, Choo SW. 2016. A new complete mitogenome of the critically endangered Chinese pangolin Manis pentadactyla. *Conservation Genetics Resources* **8**:423–426.

Hari R, Siew Woh C. 2015. Complete mitogenome of *M. javanica* and *M. pentadactyla* obtained from next generation sequencing data. Unpublished. Available from https://www.ncbi.nlm.nih.gov/nuccore/KT445979.1.

Hassanin A, Hugot J-P, Jansen van Vuuren B. 2015. Comparison of mitochondrial genome sequences of pangolins (Mammalia, Pholidota). *Comptes Rendus Biologies* **338**:260–265.

Hayashi J, Tagashira Y, Yoshida MC. 1985. Absence of extensive recombination between inter- and intraspecies mitochondrial DNA in mammalian cells. *Experimental Cell Research* **160**:387–395.

Heaney LR et al. 1998. A Synopsis of the Mammalian Fauna of the Philippine Islands. *Fieldiana: Zoology* (New Series) **88**:1–61.

Heath ME. 1992a. Mammalian species: *Manis temminckii*. *The American Society of Mammalogists* **415**:1–5.

Heath ME. 1992b. Mammalian Species: *Manis pentadactyla*.

Heath ME. 1995. Mammalian Species: *Manis crassicaudata*.

Heath ME, Coulson IM. 1997a. Preliminary studies on relocation of Cape pangolins *Manis temminckii*. *South African Journal of Wildlife Research* **27**:51–56.

Heath ME, Coulson IM. 1997b. Home range size and distribution in a wild population of Cape pangolins, *Manis temminckii*, in north-west Zimbabwe. *African Journal of Ecology* **35**:94–109.

Heath ME, Coulson IM. 1998. Measurements of length and mass in a wild population of Cape pangolins (*Manis temminckii*) in north-west Zimbabwe. *African Journal of Ecology* **36**:267–270.

Heath ME, Vanderlip SL. 1988. Biology, husbandry, and veterinary care of captive Chinese pangolins (*Manis pentadactyla*). *Zoo Biology* **7**:293–312.

Hebert PDN, Ratnasingham S, DeWaard JR. 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London. Series B* **270**:S96-99.

Heighton SP, Gaubert P. 2021. A timely systematic review on pangolin research, commercialization, and popularization to identify knowledge gaps and produce conservation guidelines. *Biological Conservation* **256**:109042.

Heinrich S, Wittman TA, Ross J V., Shepherd CR, Challender DWS, Cassey P. 2017. *The global trafficking of pangolins: seizures and trafficking routes from 2010–2015*. Pages 1–50 Wildlife Trade Report, Southeast Asia Regional Office. TRAFFIC, Petaling Jaya, Selangor, Malaysia. Available from https://www.traffic.org/publications/reports/the-global-trafficking-of-pangolins/ (accessed September 6, 2020).

Heinrich S, Wittmann TA, Prowse TAA, Ross J V, Delean S, Shepherd CR, Cassey P. 2016. Where did all the pangolins go? International CITES trade in pangolin species. *Global Ecology and Conservation* **8**:241–253.

Herklots GAC. 1937. The pangolin or scaly anteater. *The Hong Kong Naturalist* **8**:78–83.

Hirose K, Kawano S, Konishi S, Ichikawa M. 2011. Bayesian Information Criterion and selection of the number of factors in factor analysis models. *Journal of Data Science* **9**:243–259.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics* **6**:e1000862.

Hsieh H-M, Lee JC-I, Wu J-H, Chen C-A, Chen Y-J, Wang G-B, Chin S-C, Wang L-C, Linacre A, Tsai L-C. 2011. Establishing the pangolin mitochondrial D-loop sequences from the confiscated scales. *Forensic Science International: Genetics* **5**:303–307.

Hsieh HM, Chiang HL, Tsai LC, Lai SY, Huang NE, Linacre A, Lee JCI. 2001. Cytochrome b gene for species identification of the conservation animals. *Forensic Science International* **122**:7–18.

Hu J-Y, Hao Z-Q, Frantz L, Wu S-F, Chen W, Jiang Y-F, Wu H, Kuang W-M, Li H, Zhang Y-P, Yu Li. 2020. Genomic consequences of population decline in critically endangered pangolins and their demographic histories. *National Science Review* **7**:798–814.

Hu J, Fang S-G, Wan Q-H. 2006. Genetic diversity of Chinese water deer (*Hydropotes inermis inermis*): Implications for conservation. *Biochemical Genetics* **44**:161–172.

Illiger. 1815. *Manis gigantea*

Ingman M, Kaessmann H, Pääbo S, Gyllensten U. 2000. Mitochondrial genome variation and the origin of modern humans. *Nature* **408**:708–713.

Ingram DJ, Shirley MH, Pietersen D, Ichu IG, Sodeinde O, Moumbolou C, Hoffmann M, Gudehus M, Challender D. 2019. *Phataginus tetradactyla*. Available from https://www.iucnredlist.org/species/12766/123586126 (accessed September 13, 2020).

Irwin DM, Kocher TD, Wilson AC. 1991. Evolution of the cytochromeb gene of mammals. *Journal of Molecular Evolution* **32**:128–144.

IUCN-SSC. 2020a. Chinese Pangolin (*Manis pentadactyla*). Available from https://www.pangolinsg.org/pangolins/chinese-pangolin/ (accessed September 13, 2020).

IUCN-SSC. 2020b. Sunda Pangolin (*Manis javanica*). Available from https://www.pangolinsg.org/pangolins/sunda-pangolin/ (accessed September 13, 2020).

IUCN-SSC. 2020c. Philippine Pangolin (*Manis culionensis*). Available from https://www.pangolinsg.org/pangolins/philippine-pangolin/ (accessed September 13, 2020).

IUCN-SSC. 2020d. Indian Pangolin (*Manis crassicaudata*). Available from https://www.pangolinsg.org/pangolins/indian-pangolin/ (accessed September 13, 2020).

IUCN-SSC. 2020e. Giant Pangolin (*Smutsia gigantea*). Available from https://www.pangolinsg.org/pangolins/giant-pangolin/ (accessed September 13, 2020).

IUCN-SSC. 2020f. Black-bellied Pangolin (*Phataginus tetradactyla*). Available from https://www.pangolinsg.org/pangolins/black-bellied-pangolin/ (accessed September 13, 2020).

IUCN-SSC. 2020g. Temminck's Pangolin (*Smutsia temminckii*). Available from https://www.pangolinsg.org/pangolins/temmincks-pangolin/ (accessed September 13, 2020).

IUCN-SSC. 2020h. White-bellied Pangolin (*Phataginus tricuspis*). Available from https://www.pangolinsg.org/pangolins/white-bellied-pangolin/ (accessed September 13, 2020).

IUCN. 2016. What does the new trade ban mean for pangolin conservation? Available from http://www.iucnredlist.org/news/what-does-the-new-trade-ban-mean-for-pangolin-conservation (accessed October 20, 2016).

Jacobsen NHG, Newbery RE, Dewet MJ, Viljoen PC, Pietersen E. 1991. A contribution of the ecology of the Steppe Pangolin Manis temminckii in the Transvaal. *Zeitschrift Fur Saugetierkunde* **56**:94–100.

Jarman SN, Elliott NG. 2000. DNA evidence for morphological and cryptic cenozoic speciations in the Anaspididae, 'living fossils' from the triassic. *Journal of Evolutionary Biology* **13**:624–633.

Jentink FA. 1882. Revision of the Manidae in the Leyden Museum. *Notes From The Leyden Museum* **4**:193–209.

Kalinowski ST, Taper ML, Marshall TC. 2007. Revising how the computer program cervus accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* **16**:1099–1106.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**:772–780.

Kidd DM, Ritchie MG. 2006. Phylogeographic information systems: putting the geography into phylogeography. *Journal of Biogeography* **33**:1851–1865.

Kimura M, Weisss GH. 1964. The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* **49**:561–576. v

King JL, LaRue BL, Novroski NM, Stoljarova M, Seo SB, Zeng, X, Warshauer DH, Davis CP, Parson W, Sajantila A, Budowle B. 2014. High-quality and high-throughput massively parallel sequencing of the human mitochondrial genome using the Illumina MiSeq. *Forensic Science International: Genetics* **12**:128–135.

Kingdon J. 1997. *The Kingdon field guide of African mammals*. Academic Press, London.

Kingdon J, Hoffmann M. 2013. *Carnivores, pangolins, equids and rhinoceroses.* Page 560 Mammals of Africa, 1st edition. Bloomsbury Natural History.

Kingdon JS. 1971. *East African Mammals: An Atlas of Evolution in Africa*. Academic Press, London, UK.

Knowlton N. 1993. Sibling Species in the Sea. *Annual Review of Ecology and Systematics* **24**:189–216.

Koblmüller S, Vilà C, Lorente-Galdos B, Dabad M, Ramirez O, Marques-Bonet T, Wayne RK, Leonard JA. 2016. Whole mitochondrial genomes illuminate ancient intercontinental dispersals of grey wolves (*Canis lupus*).

Koenigswald W. 1999. *Order Pholidota*. Pages 75–80 in G. E. Rössner and K. Heissig, editors. The Miocene Land Mammals of Europe. Verlag Dr. Friedrich Pfeil, Munich, Germany.

Kress WJ, Erickson DL. 2008. DNA barcodes: genes, genomics, and bioinformatics. *Proceedings of the National Academy of Sciences of the United States of America* **105**:2761–2762.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: An information aesthetic for comparative genomics. *Genome Research* **19**:1639–1645.

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

Kumar VP, Rajpoot A, Mukesh, Shukla M, Kumar D, Goyal SP. 2016b. Illegal trade of Indian pangolin (*Manis crassicaudata*): genetic study from scales based on mitochondrial genes. *Egyptian Journal of Forensic Sciences* **6**:524–533.

Kumar VP, Rajpoot A, Shukla M, Nigam P, Goyal SP. 2018. Inferring the molecular affinity of Indian pangolin with extant Manidae species based on mitochondrial genes: a wildlife forensic perspective. *Mitochondrial DNA Part B* **3**:640–644.

Kwok P-Y, Deng Q, Zakeri H, Taylor SL, Nickerson DA. 1996. Increasing the information content of STS-based genome maps: identifying polymorphisms in mapped STSs. *Genomics* **31**:123–126.

Lanave C, Preparata G, Saccone C, Serio G. 1984. A new method for calculating evolutionary substitution rates. *Journal of Molecular Evolution* **20**:86–93.

Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. 2017. PartitionFinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution* **34**:772–773.

Larizza A, Pesole G, Reyes A, Sbisà E, Saccone C. 2002. Lineage specificity of the evolutionary dynamics of the mtDNA D-loop region in rodents. *Journal of Molecular Evolution* **54**:145–155.

Lê S, Josse J, Husson F. 2008. FactoMineR: an R package for multivariate analysis. *Journal of Statistical Software* **25**:1–18.

Lee T-Y, Chang W-C, Hsu JB-K, Chang T-H, Shien D-M. 2012. GPMiner: an integrated system for mining combinatorial cis-regulatory elements in mammalian gene group. *BMC Genomics* **13**:S3–S14.

Leempoel K, Hebert T, Hadly EA. 2020. A comparison of eDNA to camera trapping for assessment of terrestrial mammal diversity. *Proceedings of the Royal Society B* **287**:20192353.

Lei R, Shore GD, Brenneman RA, Engberg SE, Sitzmann BD, Bailey CA, Kimmel LM, Randriamampionona R, Ranaivoarisoa JF, Louis EE. 2010. Complete sequence and gene organization of the mitochondrial genome for Hubbard's sportive lemur (*Lepilemur hubbardorum*). *Gene* **464**:44–49.

Li L, Wei L, Ma J, Chen J. 2017. Development and characterization of 21 SNP markers in Manis javanica based on high-throughput sequencing. *Conservation Genetics Resources* **10**:351–353.

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

Linnaeus C. 1758. Systema naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. Page (Tomus I, editor). Holmiae, Laurentii Salvii.

Linnaeus C. 1766. Systema naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. Page (Tomus I, editor). Holmiae, Laurentii Salvii.

Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M. 2012. Comparison of next-generation sequencing systems. *Journal of Biomedicine & Biotechnology* **2012**:251–364.

Lorenzen ED, Arctander P, Siegismund HR. 2007a. Three reciprocally monophyletic mtDNA lineages elucidate the taxonomic status of Grant's gazelles. *Conservation Genetics* **9**:593–601.

Lorenzen ED, Arctander P, Siegismund HR. 2008. High variation and very low differentiation in wide ranging plains zebra (*Equus quagga*): insights from mtDNA and microsatellites. *Molecular Ecology* **17**:2812–2824.

Lorenzen ED, De Neergaard R, Arctander P, Siegismund HR. 2007b. Phylogeography, hybridization and Pleistocene refugia of the kob antelope (*Kobus kob*). *Molecular Ecology* **16**:3241–3252.

Lorenzen ED, Heller R, Siegismund HR. 2012. Comparative phylogeography of African savannah ungulates. *Molecular Ecology* **21**:3656–3670.

Lorenzen ED, Masembe C, Arctander P, Siegismund HR. 2010. A long-standing Pleistocene refugium in southern Africa and a mosaic of refugia in East Africa: Insights from mtDNA and the common eland antelope. *Journal of Biogeography* **37**:571–581.

Luczon AU, Ong PS, Quilang JP, Fontanilla IKC. 2016. Determining species identity from confiscated pangolin remains using DNA barcoding. *Mitochondrial DNA Part B* **1**:763–766.

Luo SJ, Cai QX, David VA, Zhang L, Martelli P, Lim NTL, Ferrand N, Chin SC, Gaubert P, Ramos MJ, O'Brien SJ, Antunes A, Johnson WE. 2007. Isolation and characterization of microsatellite markers in pangolins (Mammalia, Pholidota, *Manis* spp.). *Molecular Ecology Notes* **7**:269–272.

Ma J-E, Li L-M, Jiang H-Y, Zhang X-J, Li J, Li G-Y, Yuan L-H, Wu J, Chen J-P. 2017. Transcriptomic analysis identifies genes and pathways related to myrmecophagy in the Malayan pangolin (*Manis javanica*). *PeerJ* **5**:e4140. PeerJ Inc. Available from https://peerj.com/articles/4140.

Mahmood T, Challender D, Khatiwada A, Andleeb S, Perera P, Trageser S, Ghose A, Mohapatra R. 2019. *Manis crassicaudata*. Available from https://www.iucnredlist.org/species/12761/123583998 (accessed September 13, 2020).

Martin AP, Palumbi SR. 1993. Body size, metabolic rate, generation time, and the molecular clock. *Proceedings of the National Academy of Sciences of the United States of America* **90**:4087–4091.

Mason VC, Helgen KM, Murphy WJ. 2019. Comparative phylogeography of forest-dependent mammals reveals paleo-forest corridors throughout sundaland. *Journal of Heredity* **110**:158–172.

Mckenna MC, Bell SK. 1997. *Classification of Mammals Above the Species Level*. Columbia University Press, New York, USA.

Meganathan PR, Pagan HJT, McCulloch ES, Stevens RD, Ray DA. 2012. Complete mitochondrial genome sequences of three bats species and whole genome mitochondrial analyses reveal patterns of codon bias and lend support to a basal split in Chiroptera. *Gene* **492**:121–129.

Mena JL, Yagui H, Tejeda V, Bonifaz E, Bellemain E, Valentini A, Tobler MW, Sánchez-Vendizú P, Lyet A. 2021. Environmental DNA metabarcoding as a useful tool for evaluating terrestrial mammal diversity in tropical forests. *Ecological Applications* **31**:e02335.

Meredith RW, Zhang G, Gilbert MTP, Jarvis ED, Springer MS. 2014. Evidence for a single loss of mineralized teeth in the common avian ancestor. *Science* (New York, N.Y.) **346**:1254390.

Meyer W, Liumsiricharoen M, Suprasert A, Fleischer LG, Hewicker-Trautwein M. 2013. Immunohistochemical demonstration of keratins in the epidermal layers of the Malayan pangolin (*Manis javanica*), with remarks on the evolution of the integumental scale armour. *European Journal of Histochemistry* **57**:e27.

Mignotte F, Gueride M, Champagne A-M, Mounolou J-C. 1990. Direct repeats in the non-coding region of rabbit mitochondrial DNA. Involvement in the generation of intra- and inter-individual heterogeneity. *European Journal of Bochemistry* **194**:561–571.

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research* **17**:240–248.

Mindell DP, Knight A, Baer C, Huddleston CJ. 1996. Slow rates of molecular evolution in birds and the metabolic rate and body temperature hypotheses. *Molecular Biology and Evolution* **13**:422–426.

Minin V, Abdo Z, Joyce P, Sullivan J. 2003. Performance-based selection of likelihood models for phylogeny estimation. *Systematic Biology* **52**:674–683.

Mooers AØ, Holmes EC. 2000. The evolution of base composition and phylogenetic inference. *Trends in Ecology & Evolution* **15**:365–369.

Moore WS. 1995. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. *Evolution* **49**:718–726.

Moritz C, Brown WM. 1987. Tandem duplications in animal mitochondrial DNAs: variation in incidence and gene content among lizards. *Proceedings of the National Academy of Sciences of the United States of America* **84**:7183–7187.

Mwale M, Dalton DL, Jansen R, Roelofse M, Pietersen DW, Mokgokong PS, Kotze A. 2017. The forensic application of DNA barcoding for identification of illegally traded African pangolin scales. *Genome* **60**:272–284.

Nabholz B, Mauffrey J-F, Bazin E, Galtier N, Glemin S. 2008. Determination of mitochondrial genetic diversity in mammals. *Genetics* **178**:351–361.

Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA. 2013. Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology* **22**:2841–2847.

Nielsen R, Paul JS, Albrechtsen A, Song YS. 2011. Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics* **12**:443–451.

Nixon S, Pietersen D, Challender D, Hoffmann M, Ichu IG, Bruce T, Ingram DJ, Matthews N, Shirley MH. 2019. *Smutsia gigantea*. Available from https://www.iucnredlist.org/species/12762/123584478 (accessed September 13, 2020).

Nowak RM. 1999. *Walker's mammals of the world*, 6th edition. Johns Hopkins University Press.

Ohashi J, Tokunaga K. 2003. Power of genome-wide linkage disequilibrium testing by using microsatellite markers. *Journal of Human Genetics* **48**:487–491.

Olayemi A, Oyeyiola A, Antunes A, Bonillo C, Cruaud C, Gaubert P. 2011. Contribution of DNA-typing to bushmeat surveys : assessment of a roadside market in south-western Nigeria. *Wildlife Research* **38**:696–716.

Opazo JC, Zavala K, Krall P, Arias RA. 2017. Evolution of gremlin 2 in cetartiodactyl mammals : gene loss. *PeerJ* **5**:e2901.

Pagés E. 1970. Sur l'ecologie et les adaptations de l'orycterope et des pangolins sympatrique du Gabon. *Biologia Gabonica* **6**:27–92.

Pagés E. 1972. Comportement aggressif et sexual chez les pangolins arboricoles (*Manis tricuspis* et *M. longicaudata*). *Biologia Gabonica* **8**:1–62.

Pagés E. 1975. Etude éco-éthologique de *Manis tricuspis* par radio-tracking. *Mammalia* **39**:613–641.

Pamilo P, O'Neill RJW. 1997. Evolution of the Sry genes. *Molecular Biology and Evolution* **14**:49–55.

Papeş M, Gaubert P. 2007. Modelling ecological niches from low numbers of occurrences: assessment of the conservation status of poorly known viverrids (Mammalia, Carnivora) across two continents. *Diversity and Distributions* **13**:890–902.

Pappin S. 2012. Meet the species: long-tailed pangolin. Available from http://www.pangolins.org/2012/02/16/meet-the-species-long-tailed-pangolin/ (accessed October 3, 2016).

Peakall R, Smouse PE. 2006. GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**:288–295.

Peakall R, Smouse PE. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* **28**:2537–2539.

Pendharkar V. 2016. Will ban on trade of pangolins give 'the world's most trafficked animal' a chance to survive? Available from http://www.firstpost.com/living/will-ban-on-trade-of-pangolins-give-the-worlds-most-trafficked-animal-a-chance-to-survive-3041520.html (accessed March 10, 2017).

Pérez T, Albornoz J, Domínguez A. 2002. Phylogeography of chamois (*Rupicapra* spp.) inferred from microsatellites. *Molecular Phylogenetics and Evolution* **25**:524–534.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE* **7**:e37135.

Piertney SB, MacColl AD, Bacon PJ, Dallas JF. 1998. Local genetic structure in red grouse (*Lagopus lagopus scoticus*): evidence from microsatellite DNA markers. *Molecular Ecology* **7**:1645–54.

Pietersen D, Jansen R, Connelly E. 2019a. *Smutsia temminckii*. Available from https://www.iucnredlist.org/species/12765/123585768 (accessed September 13, 2020).

Pietersen D, Moumbolou C, Ingram DJ, Soewu D, Jansen R, Sodeinde O, Keboy Mov Linkey Iflankoy C, Challender D, Shirley MH. 2019b. *Phataginus tricuspis*. Available from https://www.iucnredlist.org/species/12767/123586469 (accessed September 13, 2020).

Pietersen DW. 2013. Behavioural ecology and conservation biology of ground pangolins *Smutsia temminckii* in the Kalahari Desert. University of Pretoria, South Africa.

Pietersen DW, Jansen R, Swart J, Kotzé A. 2016a. *A conservation assessment of Smutsia temminckii*. Pages 1–11 in M. F. Child, L. Roxburgh, E. Do Linh San, D. Raimondo, and H. T. Davies-Mostert, editors. The Red List of Mammals of South Africa, Swaziland and Lesotho. South African National Biodiversity Institute and Endangered Wildlife Trust, South Africa.

Pietersen DW, McKechnie AE, Jansen R. 2014a. A review of the anthropogenic threats faced by Temminck's ground pangolin, *Smutsia temminckii*, in southern Africa. *South African Journal of Wildlife Research* **44**:167–178.

Pietersen DW, McKechnie AE, Jansen R. 2014b. Home range, habitat selection and activity patterns of an arid-zone population of Temminck's ground pangolins, *Smutsia temminckii*. *African Zoology* **49**:265–276.

Pietersen DW, Symes CT, Woodborne S, McKechnie AE, Jansen R. 2016b. Diet and prey selectivity of the specialist myrmecophage, Temminck's ground pangolin. *Journal of Zoology* **298**:198–208.

Pocock RI. 1924. The external characters of the pangolins (Manidae). *Proceedings of the Zoological Society of London* **94**:707–718.

Poland JA, Rife TW. 2012. Genotyping-by-Sequencing for plant breeding and genetics. *The Plant Genome Journal* **5**:92.

Posada D. 2003. Using MODELTEST and PAUP* to select a model of nucleotide substitution. *Current Protocols in Bioinformatics* **00:6.5**:6.5.1–6.5.14.

Prater SH. 1971. *The Book of Indian Animals*, 3rd edition. Bombay Natural History Society, India.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**:945–959.

Qin X-M, Dou S-R, Guan Q-X, Qin P-S, She Y. 2012. Complete mitochondrial genome of the *Manis pentadactyla* (Pholidota, Manidae): Comparison of *M. pentadactyla* and *M. tetradactyla*. *Mitochondrial DNA* **23**:37–38.

Rafinesque CS. 1821. Sur le genre *Manis* et description d'une nouvelle espèce: *Manis ceonyx*. Pages 214–215 in B. De Saint-Vincent, P. A. J. Drapiez, and J.-B. Van Mons, editors. Annales Générales des Sciences Physiques, 7th edition.

Rahm U. 1956. Notes on Pangolins of the Ivory Coast. *Journal of Mammalogy* **37**:531–537.

Rand DM. 1994. Thermal habit, metabolic rate and the evolution of mitochondrial DNA. *Trends in Ecology and Evolution* **9**:125–131.

Rand DM, Harrison RG. 1989. Ecological genetics of a mosaic hybrid zone : mitochondrial , nuclear , and reproductive differentiation of crickets by soil type. *Evolution* **43**:432–449.

Ray DA et al. 2004. Low levels of nucleotide diversity in *Crocodylus moreletii* and evidence of hybridization with C. acutus. *Conservation Genetics* **5**:449–462.

Raz T, Kapranov P, Lipson D, Letovsky S, Milos PM, Thompson JF. 2011. Protocol dependence of sequencing-based gene expression measurements. PLoS ONE **6**:e19287.

Richer R, Coulson I, Heath M. 1997. Foraging behaviour and ecology of the Cape pangolin (*Manis temminckii*) in north-western Zimbabwe. *African Journal of Ecology* **35**:361–369.

Roberts TJ. 1977. *The Mammals of Pakistan*. Ernest Benn, London, UK.

Robin ED, Wong R. 1988. Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. *Journal of Cellular Physiology* **136**:507–513.

Rocha D, Gut I, Jeffreys AJ, Kwok P-Y, Brookes AJ, Chanock SJ. 2006. Seventh international meeting on single nucleotide polymorphism and complex genome analysis: 'ever bigger scans and an increasingly variable genome'. *Human Genetics* **119**:451–456.

Rodríguez F, Oliver JL, Marín A, Medina JR. 1990. The general stochastic model of nucleotide substitution. *Journal of Theoretical Biology* **142**:485–501.

Rohr B. 2016. CITES Decision a Huge Win for Pangolins. Available from http://www.worldwildlife.org/press-releases/cites-decision-a-huge-win-for-pangolins (accessed February 2, 2017).

Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. Mrbayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**:539–542.

Saitou N, Ueda S. 1994. Evolutionary rates of insertion and deletion in noncoding nucleotide sequences of primates. *Molecular Biology and Evolution* **11**:504–512.

Sales NG, Kaizer M da C, Coscia I, Perkins JC, Highlands A, Boubli JP, Magnusson WE, Silva MNF Da, Benvenuto C, Mcdevitt AD. 2020. Assessing the potential of environmental DNA metabarcoding for monitoring Neotropical mammals: a case study in the Amazon and Atlantic Forest, Brazil. *Mammal Review* **50**:221–225.

Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R. 2005. Towards writing the encyclopedia of life: an introduction to DNA barcoding. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **360**:1805–1811.

Schlitter DA. 2005. *Order Pholidota.* Pages 530–531 in D. E. Wilson and D. M. Reeder, editors. Mammal Species of the World, 3rd edition. Johns Hopkins University Press, Baltimore.

Schoppe S, Katsis L, Lagrada L. 2019. *Manis culionensis.* Available from https://www.iucnredlist.org/species/136497/123586862 (accessed September 13, 2020).

Seo SB, Zeng X, King JL, Larue BL, Assidi M, Al-Qahtani MH, Sajantila A, Budowle B. 2015. Underlying data for sequencing the mitochondrial genome with the massively parallel sequencing platform Ion Torrent™ PGM™. *BMC Genomics* **16**:S4.

Shastry BS. 2002. SNP alleles in human disease and evolution. *Journal of Human Genetics* **47**:561–566.

Shi J, Wang Y, Huang W. 2009. Development and application of genotyping technologies. *Science in China Series C: Life Sciences* **52**:17–23.

Simonsen BT, Siegismund HR, Arctander P. 1998. Population structure of African buffalo inferred from mtDNA sequences and microsatellite loci: high variation but low differentiation. *Molecular Ecology* **7**:225–237.

Skinner JD, Chimimba CT. 2005. *The mammals of the southern African subregion,* 3rd edition. Cambridge University Press, Cambridge, NY.

Smith DR. 2012. Not seeing the genomes for the DNA. *Briefings in Functional Genomics* **11**:289–290.

Smith DR. 2013. RNA-Seq data: a goldmine for organelle research. *Briefings in Functional Genomics* **12**:454–456.

Smith DR. 2015. The past, present and future of mitochondrial genomics: have we sequenced enough mtDNAs? *Briefings in Functional Genomics* **15**:1–8.

Smuts. 1832. *Manis temminckii*.

Sodeinde OA, Adedipe SR. 1994. Pangolins in south-west Nigeria – current status and prognosis. *Oryx* **28**:43–50.

Soewu DA. 2008. Wild animals in ethnozoological practices among the Yorubas of southwestern Nigeria and the implications for biodiversity conservation. *African Journal of Agricultural Research* **3**:421–427.

Soewu DA, Adekanola TA. 2011. Traditional-medical knowledge and perception of pangolins (*Manis* sps) among the Awori people, Southwestern Nigeria. *Journal of Ethnobiology and Ethnomedicine* **7**:25–35.

Soewu DA, Ayodele IA. 2009. Utilisation of pangolin (*Manis* sps) in traditional Yorubic medicine in Ijebu province, Ogun State, Nigeria. *Journal of Ethnobiology and Ethnomedicine* **5**:39–49.

Springer MS, DeBry RW, Douady C, Amrine HM, Madsen O, De Jong WW, Stanhope MJ. 2001. Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. *Molecular Biology and Evolution* **18**:132–143.

Srinivasulu C, Srinivasulu B. 2012. *South Asian Mammals: Their Diversity, Distribution, and Status*, 1st edition. Springer-Verlag, New York.

Steyn C, Soley JT, Crole MR. 2018. Osteology and radiological anatomy of the thoracic limbs of Temminck's ground pangolin (*Smutsia temminckii*). *The Anatomical Record* **301**:624–635.

Streisinger G, Okada Y, Emrich J, Newton J, Tsugita A, Terzaghi E, Inouye M. 1966. Frameshift mutations and the genetic code. *Cold Spring Harbor Symposia on Quantitative Biology* **31**:77–84.

Stuart CT. 1980. The distribution and status of *Manis temminckii* Pholidota Manidae. *Saugetierk Mitt* **28**:123–129.

Sugg DW, Chesser RK, Dobson FS, Hoogland JL. 1996. Population genetics meets behavioural ecology. *Trends in Ecology and Evolution* **11**:338–342.

Sun NC-M, Chang S-P, Lin J-S, Tseng Y-W, Pei KJ-C, Hung K-H. 2020. The genetic structure and mating system of a recovered Chinese pangolin population (*Manis pentadactyla* Linnaeus, 1758) as inferred by microsatellite markers. *Global Ecology and Conservation* **23**:e01195.

Sutter JD. 2014a. The most trafficked mammal you've never heard of. Available from http://edition.cnn.com/interactive/2014/04/opinion/sutter-change-the-list-pangolin-trafficking/ (accessed March 10, 2017).

Sutter JD. 2014b. The pangolin trade must stop now. Available from http://edition.cnn.com/2014/07/31/opinion/sutter-pangolin-iucn-update/ (accessed June 15, 2016).

Swails B, McKenzie D. 2016. Race to save rare pangolin from tragic end. Available from http://edition.cnn.com/2016/10/05/africa/pangolin-extinction-threat/ (accessed February 2, 2017).

Swart JM. 1996. Foraging behaviour of the Cape Pangolin *Manis temminckii* in the Sabi Sand Wildtuin. University of Pretoria, South Africa.

Swart JM, Richardson PRK, Ferguson JWH. 1999. Ecological factors affecting the feeding behaviour of pangolins (*Manis temminckii*). *Journal of Zoology* **247**:281–292.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* **30**:2725–2729.

Tan TK, Tan KY, Hari R, Yusoff AM, Wong GJ, Siow CC, Mutha NVR, Rayko M, Komissarov A, Dobrynin P, Krasheninnikova K, Tamazian G, Paterson IC, Warren WC, Johnson WE, O'Brien SJ, Choo SW. 2016. PGD: a pangolin genome hub for the research community. *Database* **2016**:1–10.

Templeton AR, Georgiadis NJ. 1996. *A landscape approach to conservation genetics: conserving evolutionary processes in African Bovidae.* Pages 398–430 in J. C. Avise and Hamrick J L, editors. Conservation Genetics: Case Histories from Nature. Chapman & Hall, New York.

Teske PR, Von der Heyden S, Mcquaid CD, Barker NP. 2011. A review of marine phylogeography in southern Africa. *South African Journal of Science* **107**:43–53.

Tikader BK. 1983. *Threatened Animals of India.* Zoological Survey of India, Calcutta, India.

Tong J, Lü T-B, Ma Y-H, Wang H-K, Ren L-Q, Arnell RD. 2007. Two-body abrasive wear of the surfaces of pangolin scales. *Journal of Bionic Engineering* **4**:77–84.

Tong J, Ren L-Q, Chen B-C. 1995. Chemical constitution and abrasive wear behaviour of pangolin scales. *Journal of Materials Science Letters* **14**:1468–1470.

Trebbi D, Maccaferri M, de Heer P, Sørensen A, Giuliani S, Salvi S, Sanguineti MC, Massi A, van der Vossen EAG, Tuberosa R. 2011. High-throughput SNP discovery and genotyping in durum wheat (*Triticum durum* Desf.). *Theoretical and Applied Genetics* **123**:555–569.

Turner TR, Coetzer WG, Schmitt CA, Lorenz JG, Freimer NB, Grobler JP. 2016. Localized population divergence of vervet monkeys (*Chlorocebus* spp.) in South Africa: evidence from mtDNA. *American Journal of Physical Anthropology* **159**:17–30.

Uddin A, Chakraborty S. 2014. Analysis of codon usage pattern in mitochondrial ATPase 6 in some mammalian species. *International Journal of Recent Scientific Research* **5**:883–888.

Van Aarde RJ, Richardson RK, Pietersen E. 1990. Report on the behavioural ecology of the Cape Pangolin (*Manis temminckii*). University of Pretoria.

Van Ee CA. 1966. A note on breeding the Cape pangolin *Manis temniincki* at Bloemfontein zoo. *International Zoo Yearbook* **6**:163–164.

Van Ee CA. 1978. Pangolins can't be bred in captivity. *African Wildlife* **32**:24–25.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**:535–538.

Van Tassell CP, Smith TPL, Matukumalli LK, Taylor JF, Schnabel RD, Lawley CT, Haudenschild CD, Moore SS, Warren WC, Sonstegard TS. 2008. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nature Methods* **5**:247–252.

Whitfort A. 2012. Evaluating China's draft animal protection law. *The Sydney Law Review* **34**:347– 370.

Whiting MJ, Williams VL, Hibbitts TJ. 2011. Animals traded for traditional medicine at the Faraday market in South Africa: species diversity and conservation implications. *Journal of Zoology* **284**:84–96.

Widman P, Lachenmaier K, Widmann IL, Schoppe S, Dumalag RM, Matillano JD, Villafuerte DF, Diaz DH, Cervancia M. 2004. Wirbeltiergemeinschaften in Rotsteisskakadu-Habitaten in nord-Palawan. *ZGAP Mitteilungen ember* **20**:3–7.

Wilkinson GS, Mayer F, Kerth G, Petri B. 1997. Evolution of repeated sequence arrays in the D-loop region of bat mitochondrial DNA. *Genetics* **146**:1035–1048.

Wilson DE, Reeder DM. 2005. *Mammal species of the world: a taxonomic and geographic reference*, 3rd edition. Johns Hopkins University Press, Baltimore.

Wirdateti GSY. 2013. Identifikasi trenggiling (*Manis javanica*) menggunakan penanda cytochrome B mitokondria DNA (IDENTIFICATION OF PANGOLIN (*MANIS JAVANICA* DESMAREST, 1822) USING CYTOCHROME B mtDNA MARKER). *Jurnal Veteriner* **14**:467–474.

Wright S. 1950. The genetic structure of populations. *Annals of Eugenics* **15**:323–354.

Wu H-L, Wan Q-H, Fang S-G. 2006. Population structure and gene flow among wild populations of the black muntjac (*Muntiacus crinifrons*) based on mitochondrial DNA control region sequences. *Zoological Science* **23**:333–340.

Wu S-B, Wang Y-X, Feng Q. 2005. A new record of Mammalia in China - *Manis javanica*. *Acta Zootaxonomica Sinica* **30**:440–443.

Yang CW, Chen S, Chang C-Y, Lin MF, Block E, Lorentsen R, Chin JSC, Dierenfeld ES. 2007. History and dietary husbandry of pangolins in captivity. *Zoo Biology* **26**:223–230.

Yang L, Tan Z, Wang D, Xue L, Guan M-X, Huang T, Li R. 2014. Species identification through mitochondrial rRNA genetic analysis. *Scientific Reports* **4**:4089–4099.

You FM, Huo N, Gu YQ, Luo M, Ma Y, Hane D, Lazo GR, Dvorak J, Anderson OD. 2008. BatchPrimer3: A high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics* **9**:253–265.

Yu H-T, Ma G-C, Lee D-J, Chin S-C, Tsao H-S, Wu S-H, Shih S-Y, Chen M. 2011. Molecular delineation of the Y-borne Sry gene in the Formosan pangolin (*Manis pentadactyla pentadactyla*) and its phylogenetic implications for Pholidota in extant mammals. *Theriogenology* **75**:55–64.

Yu N, Jensen-Seaman MI, Chemnick L, Kidd JR, Deinard AS, Ryder O, Kidd KK, Li W-H. 2003. Low nucleotide diversity in chimpanzees and bonobos. *Genetics* **164**:1511–1518.

Yusoff AM, Tan TK, Hari R, Koepfli K-P, Wee WY, Antunes A, Sitam FT, Rovie-Ryan JJ, Karuppannan KV, Wong GJ, Lipovich L, Warren WC, O'Brien SJ, Choo SW. 2016. De novo sequencing, assembly and analysis of eight different transcriptomes from the Malayan pangolin. *Scientific Reports* **6**:1–11.

Zardoya R, Meyer A. 1996. Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. *Molecular Biology and Evolution* **13**:933–942.

Zhang H, Ades G, Miller MP, Yang F, Lai K wai, Fischer GA. 2020. Genetic identification of African pangolins and their origin in illegal trade. *Global Ecology and Conservation* **23**:e01119.

Zhang H, Miller MP, Yang F, Chan HK, Gaubert P, Ades G, Fischer GA. 2015. Molecular tracing of confiscated pangolin scales for conservation and illegal trade monitoring in southeast Asia. *Global Ecology and Conservation* **4**:414–422.

Zhihai H, Jiang X, Shuiming X, Baosheng L, Yuan G, Chaochao Z, Xiaohui Q, Wen X, Shilin C. 2016. Comparative optical genome analysis of two pangolin species: *Manis pentadacty* and *Manis javanica*. *Giga Science* **5**:1–5.

APPENDIX A

SPECIES-SPECIFIC MARKERS FOR EACH LOCUS DERIVED FROM *SMUTSIA TEMMINCKII*

Species-specific markers for each locus derived from *Smutsia temminckii*. The expected fragment sizes of the sequences are indicated in base pairs (bp) along with the corresponding annealing temperatures. The thermal cycling conditions are indicated below the table (Du Toit 2014).

Gene region	Primer name	Primer sequence	Fragment size (bp)	Annealing temperature (˚C)
CO1	Pan_6A	F: 5'- TCA GCC ACC TTA CCT ATG TTC -3'	380–400	50
		R: 5'- GAT GAG ATA CCC GCT AAA TG -3'		
	Pan_6B	F: 5'- GCT GGA ACT GGC TGA ACT GTA -3'	400–440	53
		R: 5'- TTA CGC TGC CTC CAT GTA AGG -3'		
	Pan_6C	F: 5'- GCT GGA ACT GGC TGA ACT GT-3'	90–100	50
		R: 5'- TGA CTT ATA GCG GGA GGT -3'		
	Pan_6D	F: 5'- GGC TTT ATC GTT TGA GCA CAT C -3'	90–100	50
		R: 5'- AAG CCC TAA TGC TCA TAG TAG AG -3'		
	Pan_7A	F: 5'- CAC ACG AGC CTA CTT TAC ATC AGC -3'	300–330	50
		R: 5'- TGC CCG AAA GAC CAA GGA AGT GTT G -3'		
	Pan_7B	F: 5'- CAA CGA CAC ATG AGC AAA AG -3'	280–300	50
		R: 5'- CAT AGG TAT GAT ATT GGC TTG -3'		
Cyt*b*	Pan_14A	F: 5'- CCA TAA ATA GGA GAA GGC TTA GAA G -3'	230–250	50
		R: 5'- GTG TTC TGC TGT GTA GTG TAT TGC -3'		
	Pan_14B	F: 5'- CCC TCC AAC ATC TCA GCA T -3'	570–600	50
		R: 5'- AGG GGT CGG AAT ATC ATA GTG CGT T -3'		
	Pan_14C	F: 5'- CCC CTC CAA CAT CTC AGC AT -3'	130–150	50
		R: 5'- CCG TAA TAT AAG CCT CGT CC -3'		
	Pan_14D	F: 5'- AAC CCC CTA AGC ACA CCT CCC CAT A -3'	350–380	50
		R: 5'- GGG CTT TAG TCT CCT TCC TGA GTC -3'		
D-Loop	Pan_15N	F: 5'- AAG GAG ATT CTA ACC TCC CC -3'	330–360	50
		R: 5'- CCT TCA GTG GAG GTG ATA CG -3'		
	Pan_15A	F: 5'- CCA ACG GGC AAA TAC GCT ATG -3'	480–500	50
		R: 5'- GTC CTG CGA CCA TTG ACT GAA-3'		
	Pan_15B	F: 5'- GAC TGT GGG GTA GTT ATA GGA GAA -3'	450–480	50
		R: 5'- TTA GAG CGG GCA GAA AAC TG -3'		
	Pan Pre-1	F: 5'- CAT CTT GTC AAA CCC CAA AAG C -3'	500–530	53
		R: 5'- GGC ACG AGA TTT ACC AAC CCA T -3'		

Thermal cycling was conducted as follows: initial denaturation at 95°C for 5 min, 45 cycles of denaturation at 95°C for 30 sec, annealing at 50–53°C for 1 min, extension at 72°C for 45 sec, followed by final extension at 72°C for 7 min.