



HAL
open science

Scaling up of resources in research and genomics for informed conservation management and geographic traceability of the illicit trade in pangolins

Sean Paul Heighton

► **To cite this version:**

Sean Paul Heighton. Scaling up of resources in research and genomics for informed conservation management and geographic traceability of the illicit trade in pangolins. Agricultural sciences. Université Paul Sabatier - Toulouse III, 2023. English. NNT : 2023TOU30069 . tel-04836611

HAL Id: tel-04836611

<https://theses.hal.science/tel-04836611v1>

Submitted on 13 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE

**En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE
Délivré par l'Université Toulouse 3 - Paul Sabatier**

**Présentée et soutenue par
Sean Paul HEIGHTON**

Le 28 février 2023

**État des lieux des connaissances et production de ressources
génomiques en appui à la conservation et la traçabilité du
commerce illégal des pangolins**

Ecole doctorale : **SEVAB - Sciences Ecologiques, Vétérinaires, Agronomiques et
Bioingénieries**

Spécialité : **Ecologie, biodiversité et évolution**

Unité de recherche :

EDB - Evolution et Diversité Biologique

Thèse dirigée par

Philippe GAUBERT et Jérôme MURIENNE

Jury

Mme Katharine ABERNETHY, Rapporteur

M. Rob OGDEN, Rapporteur

M. Emmanuel PARADIS, Examineur

M. Jérôme MURIENNE, Directeur de thèse

M. Philippe GAUBERT, Co-directeur de thèse

M. François-Xavier RICAUT, Président

M. Antoine FOUQUET, Invité

**Scaling up of resources in research and genomics for informed conservation
management and geographic traceability of the illicit trade in pangolins**

by

Sean Paul Heighton

Submitted and defended in fulfilment of the requirements for the degree

Doctor of Philosophy in **Écologie, Biodiversité et Évolution**

In the Laboratoire Évolution et Diversité Biologique

Institut de Recherche pour le Développement

SEVAB - Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénieries

UNIVERSITY OF TOULOUSE III – PAUL SABATIER



**UNIVERSITÉ
TOULOUSE III
PAUL SABATIER**



2023

Scaling up of resources in research and genomics for informed conservation management and geographic traceability of the illicit trade in pangolins

Student: Sean Heighton¹

Supervisors: Dr. Philippe Gaubert¹ and Dr. Jérôme Murienne¹

Affiliations: ¹ Laboratoire Evolution et Diversité Biologique (EDB), UPS/CNRS/IRD, Université Toulouse III, 118 route de Narbonne, Bât. 4R1, 31062 Toulouse, France

Degree: Doctor of Philosophy in Écologie, Biodiversité et Évolution

Summary

Pangolins (Pholidota) are a group of myrmecophagous mammals consisting of eight extant species, four in Asia and four in Africa. They are considered the most heavily trafficked mammals in the world, a label that along with that of an unfounded COVID-19 intermediate host, has placed pangolins in the public and conservation spotlight. Despite this, knowledge on pangolins is limited and there are yet to be resources and tools in place to guide efficient conservation action. Providing these resources to guide investment in research priority areas, delineate conservation units for management, as well as to identify trafficking source regions and routes is key to reducing the over-reliance in international regulation which has had limited effect on the pangolin trade. My aim is thus to provide an overview of knowledge of pangolins and generate resources to direct future conservation research through a systematic review, as well as to develop the basis for efficiently managing populations and tracing the pangolin trade using molecular techniques.

In the first chapter, I aimed to systematically review research gaps and dynamics on pangolins, the major drivers of public interest in them, and their commercial exploitation through

patented products. These three lines of investigation were then combined to suggest holistic research and conservation action guidelines for better conserving pangolins at a time when the world turned its attention to them. The second chapter revolved around understanding the global evolutionary framework among Pholidota, and the conservation implications thereof, by generating and analyzing whole-genomes of all extant species. I assessed pangolin phylogeny, divergence estimates, and ancient gene flow, along with per species demographic history and genome-wide diversity. I clarified the existence of an undescribed Asian pangolin species that potentially deserves to be labelled as a novel taxon, provided genomic resources for extant species of Pholidota and a list of orthologous genes ranked by diversity to aid in conservation genetic and forensic endeavors. The third chapter focused on identifying within-species populations and potential admixture across the geographical range for each pangolin species in order to upscale resources for conservation genetics and geographic traceability of the trade. Using a target sequencing (gene capture) approach of 671 loci, I identified distinct population units for each species and their underlying biogeographic drivers, variable levels of diversity with certain units requiring attention, potentially novel taxa, and warn of a globalized trade in major Asian cities. These clusters provide the first step to informed population management and a reference for geographic traceability once additional samples from seizures are obtained.

This work will provide significant resources to aid conservation researchers and practitioners in building on our knowledge of pangolins in priority areas and managing populations for all eight species. This work also forms the basis for a forensic DNA registry and for identifying the best suite of loci and SNPs for accurately separating populations, which in turn may be incorporated into versatile DNA-based ‘toolkits’ used for tracing the illicit pangolin trade in the future.

Keywords:

Pangolin, geographic traceability, conservation resources, wildlife trafficking, regulatory bodies, forensics, genomes, conservation genetics, population genomics, taxonomic revision, systematic review, population structure, population demography

Acknowledgements

My utmost gratitude to my supervisors that provided me with necessary knowledge and tools to complete this research. To Dr. Philippe Gaubert, thank you so much for not only believing in my capabilities, but for being willing to help me with all of my research and conservation endeavours, even those not related to the PhD. You have taught me the importance of collaborative networks, looking at the bigger picture, and how to really do gardening. Your passion and cheerfulness every day that we conducted this research was truly a motivating aspect of my PhD. I look forward to the exciting projects and interesting discussions around science and conservation to come. To Dr. Jérôme Murienne, you have been invaluable in showing me the importance of a work-life balance and the never say no attitude, even if you end up breaking bones. Your willingness to include me in various social and sports-related activities is truly something I appreciated. I have to thank you for showing me how bad I am at petanque and for teaching me how to scuba dive. My sincerest thanks go to the both of you for working around the clock to submit this dissertation under a difficult deadline. I cannot thank you enough.

To my family and friends, thank you for being there to share my downfalls and triumphs. Thank you for supporting me in following my passion far away from home for the past few years. To my French friends and colleagues, thank you for teaching me the culture and language of your beautiful country and for being so willing to keep me around longer. A special thanks to Dr. Jordi Salmona for your many hours explaining to me genomic concepts and for accepting me as part of your family. Thank you to Dr. Stanislas Zanvo for all the times we have enjoyed together and for including me in your research plans, I have made a friend and future research partner.

I would also like to thank the many collaborators that provided samples and genomic data for such a global approach to pangolin conservation genetics. Without your efforts, this study

would not have been possible. In particular, I would like to thank Philippe, Dr. Stanislas Zanvo, Din Dipita Alain, Koffi Jules Gossé, Dr. Darren Pietersen, Dr. Helen Nash, and Dr. Mukesh Thakur. I would also like to acknowledge the staff at the EDB laboratory and University of Toulouse III, particularly those that were invaluable in teaching me the lab techniques (Yves Cuenot, Amaia Iribar-Pelozuelo, and Sophie Courjal), and those integral for the administrative aspects of sample shipping and PhD thesis (Julie Branger and Dominique Pantalacci). A special thanks is needed for Catherine Stasiulis for always going out of your way to solve the unsolvable with French administration, you truly are a superhero.

A special thank you goes to Professor Katharine Abernethy and Professor Rob Ogden for being willing to take up the task of reviewing this thesis and for your intellectual contributions, both at the defence and in the thesis itself, that allowed me to grow as a scientist. I would like to also thank the other invited jury members of my PhD defence (Dr. François-Xavier Ricaut, Dr. Emmanuel Paradis, and Dr. Antoine Fouquet) for your time spent reading the thesis and the insightful discussion points or comments that you had provided during the defence.

I would like to acknowledge the Agence Nationale de la Recherche (ANR-17-CE02-0001 PANGO-GO), the Laboratoire Évolution et Diversité Biologique (UMR 5174), the Institut de Recherche pour le Développement (IRD) and SEVAB doctoral school for the support and funding of my thesis.

Table of contents

Scaling up of resources in research and genomics for informed conservation management and geographic traceability of the illicit trade in pangolins.....	1
Summary.....	ii
Acknowledgements.....	v
Table of contents.....	vii
List of figures.....	x
List of tables.....	xiii
Introductory Chapter.....	1
Rationale and background.....	1
<i>International wildlife trade.....</i>	<i>1</i>
<i>Relying on regulation – the perfect world scenario</i>	<i>3</i>
<i>The pangolin</i>	<i>5</i>
<i>Solving the pangolin problem.....</i>	<i>8</i>
Thesis Aim and Objectives.....	12
Thesis structure.....	13
Chapter 1: Systematic Review.....	14
Abstract.....	15
Introduction.....	16
Methods.....	19
<i>Online news and societal interest trends.....</i>	<i>21</i>
<i>Patents</i>	<i>22</i>
<i>Data analyses</i>	<i>22</i>
Results.....	23
<i>Pangolin science overview</i>	<i>23</i>
<i>Pangolin (focused) natural sciences trends.....</i>	<i>26</i>
<i>Pangolin natural sciences research dynamics</i>	<i>29</i>
<i>Pangolin popularization</i>	<i>32</i>
<i>Pangolin patents</i>	<i>33</i>
Discussion.....	33
<i>Pangolin research.....</i>	<i>33</i>

<i>Research dynamics</i>	39
<i>Pangolin popularization</i>	40
<i>Commercial interest in pangolins through patents</i>	42
<i>Conclusion of a holistic review on pangolins for their conservation</i>	43
Acknowledgments.....	47
Chapter 2: Species Genomics	48
Abstract.....	49
Introduction.....	50
Results and Discussion	53
Methods and Materials.....	69
<i>Sample acquisition</i>	69
<i>Genome sequencing and assembly</i>	69
<i>Single-copy orthologous gene dataset</i>	72
<i>Phylogenomics and forensic markers from genes</i>	75
<i>Mitochondrial genes and genomes</i>	79
<i>Genome-wide analyses</i>	80
<i>Data accessibility</i>	83
Acknowledgments.....	83
Author Contributions	84
Chapter 3: Population Genomics	85
Abstract.....	86
Introduction.....	87
Methods.....	91
<i>Bait design for gene-capture</i>	91
<i>Sample collection and metadata</i>	94
<i>DNA extraction, gene-capture and sequencing</i>	96
<i>Simple capture efficiency estimates to guide sequencing efforts</i>	98
<i>Gene-capture sequencing processing with EAGER</i>	98
<i>Published genome processing</i>	100
<i>Genotyping and filtering</i>	100
<i>Population stratification</i>	102
<i>Population genetic estimates</i>	103

<i>Historical relationships among regions</i>	104
Results.....	105
<i>Population stratification and structuring</i>	106
<i>Regional differentiation, migration and phylogenetics</i>	122
<i>Population genetic estimates</i>	128
<i>Species genetic distances and isolation by distance (IBD)</i>	131
Discussion.....	131
<i>Gene-capture design and efficiency</i>	132
<i>Potential novel/undescribed species</i>	134
<i>Population delimitation – building the database for conservation management</i>	137
<i>Major drivers for population structure</i>	152
<i>Tracing the trade</i>	154
Conclusion	157
Conclusion and Synthesis Chapter	159
Background.....	159
Major outputs and their future	160
<i>A pangolin information resource</i>	160
<i>Genomic resources for pangolin conservation</i>	162
<i>A basis from which to trace the trade</i>	164
The gaps	166
Conclusion: future for pangolin conservation.....	168
Appendices	170
<i>Chapter 1: Systematic review appendix</i>	170
<i>Chapter 2: Species genomics appendix</i>	177
<i>Chapter 3: Population genomics appendix</i>	198
References	211

List of figures

Introductory Chapter

Figure 1: The geographic distribution of the eight extant pangolin species in Africa and Asia (pg. 6).

Chapter 1

Figure 1: Global conservation status, taxonomy, and distribution of the eight extant pangolin species (pg. 17).

Figure 2: Pangolin research (literature focused and not focused entirely on pangolins) from 1865 to 2019 (pg. 25).

Figure 3: Heatmap of literature count in research themes and categories by species from literature with a pangolin focus and in the scientific domain “natural sciences” (pg. 28).

Figure 4: A co-occurrence network of terms and their links extracted with text mining algorithms in VOSviewer on pangolin literature (pg. 29).

Figure 5: Pangolin research dynamics by range-state, continent, and authorship through time (pg. 31).

Figure 6: Public interest in the form of societal interest and online news article trends (pg. 32).

Chapter 2

Figure 1: Phylogenomic relationships and reticulation amongst pangolins (pg. 56).

Figure 2: Divergence time estimates along the pangolin phylogenetic tree (pg. 61).

Figure 3: Recent demographic history of pangolin species and populations (pg. 64).

Figure 4: Proportion of genome-wide heterozygosity of pangolins and other mammals (pg. 67).

Chapter 3

Figure 1: The pipeline designed and used to create the final list of loci necessary for gene-capture (pg. 94).

Figure 2: Sampling effort for all eight known species of pangolin, along with their respective ranges (pg. 96).

Figure 3: Population structure and stratification of *Phataginus tricuspis* using single nucleotide polymorphisms (pg. 108).

Figure 4: Population structure and stratification of *Phataginus tetradactyla* using single nucleotide polymorphisms (pg. 111).

Figure 5: Population structure and stratification of *Smutsia gigantea* using single nucleotide polymorphisms (pg. 112).

Figure 6: Population structure and stratification of *Smutsia temminckii* using single nucleotide polymorphisms (pg. 113).

Figure 7: Population structure and stratification of *Manis javanica* using single nucleotide polymorphisms (pg. 116).

Figure 8: Population structure and stratification of *Manis pentadactyla* using single nucleotide polymorphisms (pg. 119).

Figure 9: Population structure and stratification of *Manis crassicaudata* using single nucleotide polymorphisms (pg. 120).

Figure 10: Population structure and stratification of *Manis culionensis* using single nucleotide polymorphisms (pg. 121).

Figure 11: F_{ST} heatmap and TreeMix maximum likelihood tree with most likely number of migration edges for African species (pg. 124).

Figure 12: F_{ST} heatmap and TreeMix maximum likelihood tree with most likely number of migration edges for Asian species (pg. 127).

Appendices

Chapter 1

Appendix G: Heatmap of literature count in research categories through time (pg. 172).

Appendix H: Heatmap of literature count per pangolin species through time (pg. 173).

Appendix K: Co-occurrence network of terms and their links extracted with text mining algorithms in VOSviewer from titles of patent records (pg. 176).

Chapter 2

Figure S1: Orthologous whole-gene marker pipeline and subsequent analyses (pg. 177).

Figure S2: Phylogenomic relationships of pangolins inferred from concatenated and coalescent trees (pg. 179).

Figure S3: Phylogenetic relationships of pangolins inferred from mitochondrial gene and genome trees (pg. 181).

Figure S4: Maximum pseudolikelihood phylogenetic networks (pg. 183).

Figure S5: Diversity of orthologous whole-gene markers (pg. 184).

Chapter 3

Supplementary Figure 1: Gene-capture summary statistics (pg. 198).

Supplementary Figure 2: Principle component analysis (PCA) of South-East Asian species for the top two axes (pg. 199).

Supplementary Figure 3: Principle component analysis (PCA) of *Phataginus tricuspis* zoomed in for the top two axes (pg. 200).

Supplementary Figure 4: Principle component analysis (PCA) of *Phataginus tricuspis* for the top four (pg. 201).

Supplementary Figure 5: *Phataginus tricuspis* ADMIXTURE bar graphs of all Ks (pg. 202).

Supplementary Figure 6: *Phataginus tetradactyla* ADMIXTURE bar graphs of all Ks (pg. 203).

Supplementary Figure 7: *Smutsia gigantea* ADMIXTURE bar graphs of all Ks (pg. 203).

Supplementary Figure 8: *Smutsia temminckii* ADMIXTURE bar graphs of all Ks (pg. 204).

Supplementary Figure 9: Principle component analysis (PCA) of *Manis javanica* for (A-C) top four axes of differentiation and (D) top two axes of differentiation in which points are coloured by previously identified clusters (pg. 205).

Supplementary Figure 10: *Manis javanica* ADMIXTURE bar graphs of all Ks (pg. 206).

Supplementary Figure 11: *Manis pentadactyla* ADMIXTURE bar graphs of all Ks (pg. 206).

Supplementary Figure 12: *Manis crassicaudata* ADMIXTURE bar graphs of all Ks (pg. 207).

Supplementary Figure 13: *Manis culionensis* ADMIXTURE bar graphs of all Ks (pg. 208).

Supplementary Figure 14: Boxplots and histograms representing pairwise Euclidian distances across all individuals from each of the eight pangolin species (pg. 209).

Supplementary Figure 15: Isolation by distance (IBD) plots comparing pairwise genetic divergence (Euclidean distance) with pairwise geographic distance (km) across all individuals from each of the eight pangolin species (pg. 210).

List of tables

Chapter 1

Table 1: Global recommendations to fill the knowledge gaps on pangolins for their conservation as identified through the review process on pangolin research, popularization and commercialization (pgs. 45–46).

Chapter 3

Table 1: Population genomic parameters for all eight pangolin species (pgs. 129–130).

Appendices

Chapter 1

Appendix B: The scientific domains (n= 6) assigned to literature of pangolins along with the assigned research themes (n= 9) and their categories (n= 30) under the domain “natural sciences” (pg. 170).

Appendix I: Representation of range-states in focused pangolin literature under the domain “natural sciences” (pg. 174).

Appendix J: Proportion of patents (n= 5 294) referring to each of the top 25 Derwent Class Codes from Derwent Innovations Index in Web of Science (pg. 175).

Chapter 2

Table S1: The samples sequenced and references used for Chapter 2 (pgs. 185–187).

Table S2: Output of the concordance analysis implemented in IQ-TREE along with the significance of the Chi-squared test of independence between the two discordant gene counts (pgs. 188–189).

Table S3: Dates used for soft bound fossil calibrations on specific nodes used as priors for the MCMCtree analysis of divergence estimates of pangolins (pg. 190).

Table S4: Estimated posterior mean or median divergence estimates and 95% Highest Posterior Density (HPD) interval of each node for this study and the one conducted by Gaubert et al. 2018 (pg. 191).

Table S5: Proportion of genome-wide heterozygosity for mammalian species ranked from least to most diverse (pgs. 192–196).

Table S6: Estimated generation times of each species of pangolins literature (pg. 197).

Introductory Chapter

Rationale and background

Humans share the earth with a countless number and diversity of species, many of which have and continue to shape our own existence (Costello et al. 2013). We, and the irreversible climatic changes that we have initiated, are now suggested to be the main drivers of biodiversity loss, triggering what is often considered the earth's sixth mass extinction event (Ceballos et al. 2020; Ceballos et al. 2010). Evidence for this is in the estimated extinction rates during this "Anthropocene", which are around 100–1000 times higher than the background norm (Pimm et al. 1995). With 515 vertebrate species at the brink of extinction and a 75% decline in flying insect biomass in the last 27 years (Ceballos et al. 2020; Hallmann et al. 2017), our ecosystems and the services that they provide are strained (Cardinale et al. 2012). An ecosystem functioning at reduced levels may have dire consequences for its wildlife and human inhabitants as ecosystem services are influenced by ecosystem health. Central to the anthropogenic activities that are considerably contributing to biodiversity loss is international trade in wildlife. In addition to the effects related to the establishment of invasive species, the international wildlife trade contributes to around 30% of global species threats (Lenzen *et al.*, 2012).

International wildlife trade

Wildlife trade refers to the sale or exchange of wild fauna and flora in an often legal, regulated manner (Oldfield 2003). The industry is estimated to be worth approximately US\$ 300–332.5 billion at a 5.5% annual increase, with timber and fishing industries accounting for 80–90% of the total (Roe 2007). It therefore forms an integral part of the global economy, one that is heavily

linked to the livelihoods of rural and indigenous communities from the developing world (Di Minin et al. 2019; Hughes 2021). However, there is a dark side of wildlife trade, in the sense that the harvest of fauna and flora can also be unsustainable, leading to it being illegal for them to be traded.

Illegal wildlife trade is a major part of trans-national crime, which includes art and antique smuggling to drug, weapon and even human trafficking (Hughes 2021; Warchol 2004). This trade is unregulated and clandestine, making estimates of the market value imprecise. Nevertheless, estimates are suggested to be between US\$ 7–23 billion annually (Nellemann et al. 2016; UNODC 2016). The trade itself is extensive, with 132 144 seizures across 120 countries recorded between 2004 and 2014 (Hughes 2021; Nellemann et al. 2016; UNODC 2016). Seizures represent only a fraction of the actual amount of trade occurring due to its clandestine nature and the inability to sometimes differentiate between legal and illegal sources (Heinrich et al. 2017; Stringham et al. 2021). The illicit trade in wildlife serves various purposes, including ethnomedicine (tiger bones), fashion (snake skins), specialised collections (cycads), nutrition (whale and bushmeat), and exotic pets (tortoises), all of which have been directly linked to examples of population declines (Goodrich et al. 2022; Mandimbahasina et al. 2020; Springer et al. 2003; Warchol 2004; Zhou and Jiang 2004). It is therefore a significant threat, to the point that it has played an integral role in the extinction in the wild of multiple taxa (Hammer and Watson 2012). These include the three subspecies of the iconic rhinoceros; western black (*Diceros bicornis longipes*), northern white (*Ceratotherium simum cottoni*), and the Javan (*Rhinoceros sondaicus annamiticus*) in Vietnam (Brook et al. 2012; Emslie 2020a, b).

Apart from population declines and species/population extirpation, the illicit trade has additional far-reaching consequences with regards to; disease epidemics in humans and wildlife

(Grace et al. 2012; O’Hanlon et al. 2018), invasive species (Westphal et al. 2008), and human exploitation and conflict (Brashares et al. 2014). These do not only negatively impact conservation (Lindsey et al. 2020), but have major economic implications which drastically affect peoples livelihoods globally (Chowdhury et al. 2021; Pimentel et al. 2005) – as was the case for COVID-19.

One major aspect of COVID-19 was involvement of the pangolin. At the early stages of the pandemic, pangolins were suggested to be intermediate host of SARS-CoV-2 which is responsible for the COVID-19 pandemic (Zhang et al. 2020c). This not only thrust pangolins into the public spotlight (Heighton and Gaubert 2021) but also made the world realize the dangers of illegally poaching and trading wildlife. Further research has provided evidence against the claim of the pangolins involvement (Frutos et al. 2020; Lee et al. 2020), but the question remains as to how to prevent future pandemics stemming from illegal wildlife trade (Borzée et al. 2020; Roe et al. 2020).

Relying on regulation – the perfect world scenario

International regulatory bodies/conventions were formulated by the growing consciousness of the overexploitation of wildlife, whereby their aim is to ensure that international trade of wildlife does not threaten the extinction of species. This is an important, but difficult task in the face of human population growth, climate change, and economic expansion (Chapman et al. 2022; Warchol 2004). Despite these challenges, regulation in unison with on-the-ground conservation action has worked for the conservation of certain species or populations such as the giant otter (*Pteronura brasiliensis*: Recharte Uscamaita and Bodmer 2010), southern white rhino (*Ceratotherium simum simum*: CITES 2022b), Cape mountain zebra (*Equus zebra zebra*: Smith et

al. 2008), giant panda (*Ailuropoda melanoleuca*: Swaisgood et al. 2018), and Vicuña (*Vicugna vicugna*: McNeill and Lichtenstein 2003) – however see Bennett et al. (2021).

Unfortunately, these successes are exceptions to the rule mainly due to the following reasons: (i) the absence of enforcement by regulatory bodies and stakeholders, (ii) non-compliance to regulatory processes, (iii) shortage of resources to devote to regulatory agreements, (iv) lack of knowledge of species and market forces, (v) political and emotional sway on regulatory decisions, and (vi) general faults with the design of the regulatory processes themselves (Berec et al. 2018; Challender et al. 2015; Oldfield 2003; Zain 2020). An example of some of these points can be found with CITES (the Convention on International Trade in Endangered Species of Wild Fauna and Flora), whereby its mandate is to provide recommendations for trade regulation, but does not enforce these regulations itself. It thus relies on Parties to the convention (government signatories) to comply with and undertake such suggestions, despite many often being underfunded and lacking capacity to do so. The effect of these inherent flaws is compounded by the over-reliance on regulatory bodies/conventions. Misconduct by these bodies/conventions and their stakeholders (governments, industry) can also undermine the regulatory process behind legally traded species, leading to illegal trade (Challender et al. 2015; Zain 2020). Misconduct has been noted in the fishing and timber industries with the mislabelling of species or sites (Ogden 2008; UNODC 2016).

Potential examples of misconduct related to pangolins include (i) a mismatch in the numbers sold vs. released in the Chinese national quota system whereby seized products are sold at legal auctions as legal products, (ii) a Traditional Chinese Medicine (TCM) hospital systematically prescribing pangolin scales to each patient even when not required, and (iii) the use of legal breeding farms to cover an illegal trade (Challender and Waterman 2017; Horner et al. 2020; Wang et al. 2020b; Xu et al. 2016). The latter is a result of a CITES regulation whereby

breeding farms in post Appendix I may legally trade the F2 generation, and onwards, if the founding population was sourced legally and prior to the amendment (CITES, 2017). This is in place despite the unlikely possibility of successful captive breeding of pangolins (Challender et al. 2019b; Hua et al. 2015) and evidence for breeding farms sourcing wild caught animals under the guise of captive breeding (Horner et al. 2020). It is therefore evident that complementary conservation approaches to international regulation should be implemented in order to reduce the over-reliance and thus impact of regulatory flaws.

The pangolin

Pangolins are predominantly nocturnal, myrmecophagous mammals with an outer covering of keratinous scales (Challender et al. 2020b). They form a monophyletic order, Pholidota, which diverged from its sister order Carnivora roughly 35–65 million years ago. The order is comprised of eight species from three genera, namely *Manis* (all four Asian species), *Phataginus* (two small, predominantly arboreal African species) and *Smutsia* (two large, terrestrial African species) which are found in tropical to inter-tropical zones of Africa and Asia (Figure 1; Challender et al. 2020b).

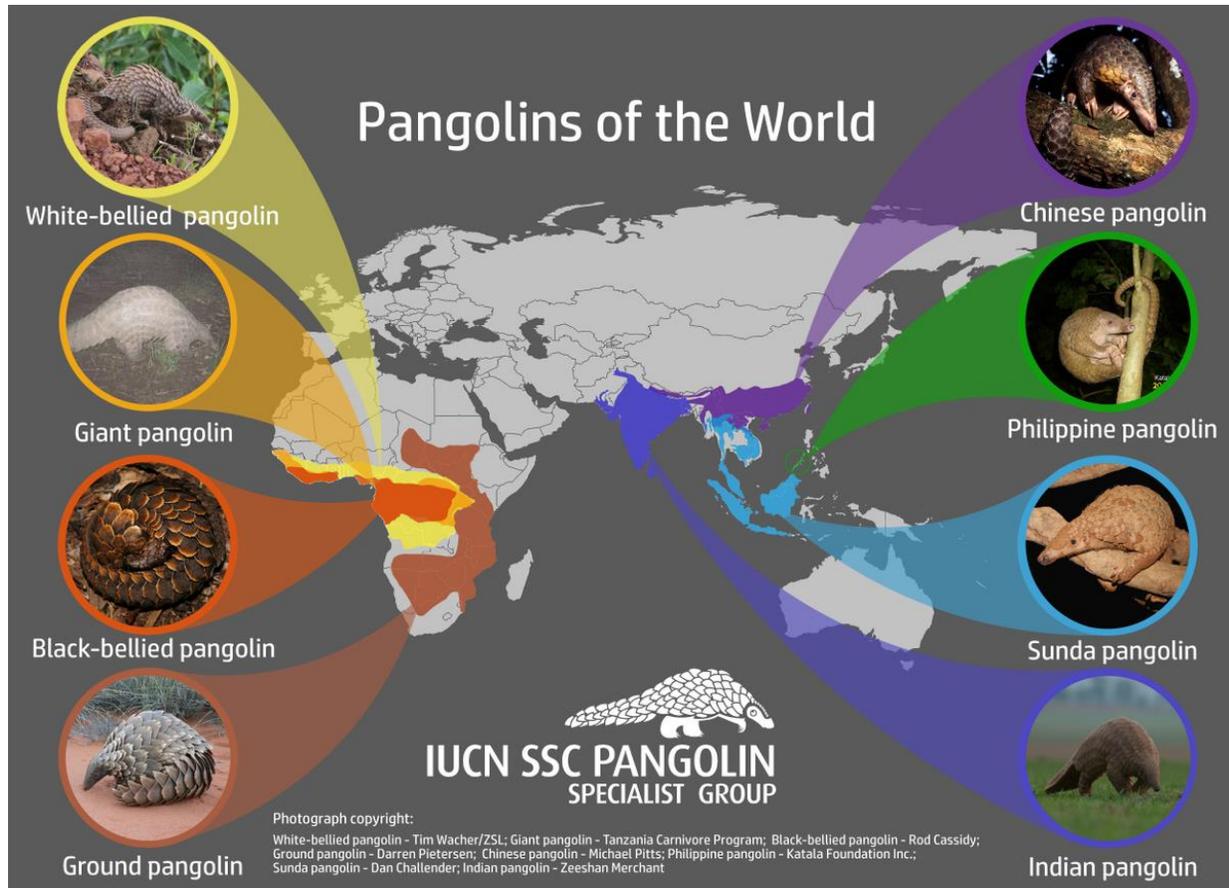


Figure 1: The geographic distribution of the eight extant pangolin species in Africa and Asia. The white-bellied (*Phataginus tricuspis* – Endangered), black-bellied (*P. tetradactyla* – Vulnerable), giant ground (*Smutsia gigantea* – Endangered) and Temminck’s ground (*S. temminckii* – Vulnerable) pangolins occur in Africa whilst the Chinese (*Manis pentadactyla* – Critically Endangered), Philippine (*M. culionensis* – Critically Endangered), Sunda (*M. javanica* – Critically Endangered) and Indian (*M. crassicaudata* – Endangered) pangolins occur in Asia. Image sourced from the IUCN Species Survival Commission Pangolin Specialist Group.

With the recent up-listing of all eight pangolin species from CITES Appendix II to Appendix I (CITES 2017), international trade, both legal and illicit, has been recognised as a serious threat to these species’ future survival. Despite these up-listings, pangolin trafficking estimates continue to rise to new records with no sign of slowing (Challender et al. 2020a; Emogor et al. 2021), which is a clear indication that regulatory bodies like CITES have had little effect. Estimates suggest that around one million pangolins were trafficked over 13 years, notably from Africa to Asia (Challender et al. 2020a; Heinrich et al. 2017), with pangolins constituting 5% of wildlife seizures documented in the World WISE database between 2005 and 2014, 2% more than

that of rhinoceroses (UNODC 2016). Further studies predict larger estimates through more finite, regional analyses (see: Ingram et al. 2018), such as a record high of up to 0.8 million African pangolins trafficked across Nigerian-linked trade routes alone between 2010–2021 (Emogor et al. 2021). This has resulted in pangolins being labelled the most heavily trafficked wild mammal in the world (Challender et al. 2014). The number of pangolins traded legally between 2007 and 2013 was 1 467, but the number seized from the illegal trade was 107 060 (UNODC 2016), which essentially equates to a maximum 1.37% of the total pangolin trade being legal prior to the 2017 CITES up-listing. This again indicates that conservation authorities have little control over the international pangolin trade through regulation, even prior to complete bans.

Following the original up-listings, further amendments have been requested at CoP19 (Conference of the Parties to CITES), one of the most notable was the suggestion for Parties to take necessary legislative, regulatory and enforcement measures to close their domestic markets in commercial trade of pangolins if they were contributing to the illegal trade (CITES 2022a). This comes as evidence is starting to mount of a “blurring of lines” between legal domestic markets and the illegal wildlife trade (Horner et al. 2020; Wang et al. 2020b; Xu et al. 2016), suggesting the need for a more in-depth look into the market dynamics and drivers.

Not only does the illegal trade pose a risk for the continued survival of pangolin populations and species, as evidenced in recent population declines (Nash et al. 2016; Zanzo et al. 2020), but it may affect many ethnic groups whereby pangolins are culturally significant (Soewu and Adekanola 2011; Suwal et al. 2022; Zanzo et al. 2021). The impact on ecological processes through their various ecological roles of burrowing and myrmecophagy (Chao et al. 2020; Sun et al. 2021b), as well as the risk of disease spread are also at stake (Daszak et al. 2020; Wang et al. 2020a). Recent studies on pangolins have already brought to light pathogens of health concern for

human, domestic animal, and wildlife populations (Chaber and Cunningham 2016; Liu et al. 2019; Wang et al. 2020a).

Solving the pangolin problem

Changing people's perceptions and cultural belief systems through education, whilst at the same time improving their circumstances through significant investment and capacity building is the key to mitigating the pangolin trade, but this is a long-term process (Chapman et al. 2022; Heighton and Gaubert 2021). Unfortunately, based on trade and population estimates, time is of the essence for pangolins as international regulatory bodies/conventions continue to have limited effect, thus more direct/fast-acting conservation approaches need to take priority in the interim (Heighton and Gaubert 2021). For pangolins, direct conservation measures include the investment in areas of conservation research priority, national law enforcement and policy, incentive programmes, and tools that can help us better understand, trace and mitigate its major threat – the illegal wildlife trade. As part of this PhD, I will delve into the investment of priority research areas for conservation as well as genomic resources for understanding pangolin populations and tracing the illegal trade.

Investment in research

One of the biggest failures in conservation is the research-implementation gap whereby researchers and conservation practitioners do not work in unison to undertake the research necessary for management plans and to produce management plans without the guidance of research (Cvitanovic et al. 2014; Knight et al. 2008). Around two-thirds of conservation assessments published in peer-reviewed scientific literature do not deliver conservation action

(Knight et al. 2008), whilst only 14% of citations in management plans for Marine Protected Areas were linked to scientific literature (Cvitanovic et al. 2014). A retrospective review of trends in scientific literature during the rhinoceroses' conservation crises indicated that even though there was an increase in research output, there was a large disconnect between the research needs in action plans compared to what was published (Linklater 2003). Furthermore, having gaps in critically important categories, species and geographic areas is often cited as an inhibitor to successful conservation management (Christie et al. 2020; Junker et al. 2020; Linklater 2003). Therefore, prioritising context-based research that is tailored to filling important gaps and for producing informed conservation actions is an integral part of conservation success (Christie et al. 2020; Linklater 2003). One way to do this is through systematic reviews as they can form a vital part of a conservation practitioners toolkit (Cook et al. 2013; Jacaban et al. 2022). This is especially true since evidence based action plans require a time-consuming review of literature that can often be inaccessible, difficult to interpret and is not synthesized across species and geography (Cook et al. 2013). On top of this, providing literature resources that can be accessible to researchers and conservation practitioners alike will reduce a significant barrier concerning the research-implementation gap and science in general (Christie et al. 2021; Cvitanovic et al. 2014).

Pangolins lack both these open-access resources for accessing literature related to them and systematic reviews for synthesising said literature. Currently, the only review regarding pangolin literature was a chapter in the first book on pangolins, which was aimed at synthesising research needs (Pietersen and Challender 2020). Although a great start, the review was not systematic in its approach nor did it delve into aspects of conservation not linked to research (popularisation or commercialisation). In terms of literature resources, the aforementioned book is a great resource as it delves into a number of important topics for conservation practitioners, but is not open-access

(Challender et al. 2020b). A systematic review on pangolins and an open-access resource in which literature can be obtained is thus of grave urgency for their conservation, particularly in light of the recent research and conservation interest from COVID-19 (Heighton and Gaubert 2021).

Geographic traceability through wildlife forensics

Another direct conservation approach is wildlife forensics, which has received increased attention in the past few years as evidenced by the formation of the International Consortium on Combating Wildlife Crime (ICWC) and various discussion points at recent Conferences of the Parties to CITES (Jota Baptista et al. 2022; Ogden and Linacre 2015; TRAFFIC 2016). Wildlife forensics is a discipline that integrates analytical techniques in order to investigate wildlife crime, encapsulating everything from pathology to ballistics in wildlife crime cases (Huffman and Wallace 2012; Jota Baptista et al. 2022). Although the discipline is implemented to solve various cases, like determining the time of death of the species that has been trafficked, the assignment of an individual to its origin (geographic traceability) is less refined but is becoming increasingly requested (Ogden and Linacre 2015). Geographic assignment helps determine if individuals are legally or illegally sourced, pinpoint poaching hotspots, determine trade routes, and even identify areas best-suited for repatriation of live individuals seized from trade (Fernandes and Caparroz 2013; Ogden and Linacre 2015).

Geographic traceability has been used to identify source populations of illegally traded products including African elephants (Wasser et al. 2022), Gopher tortoises (*Gopherus polyphemus*; Schwartz and Karl 2008), and African rhinoceroses (Harper et al. 2013). Research on pangolin traceability based on delimitation of populations (Gaubert et al. 2016), identification

of variation in trade seizures (Zhang et al. 2015b), and tracing the trade (Ewart et al. 2021; Nash et al. 2018) has had varying levels of success. The effectiveness of the method (assignment resolution and accuracy) is directly associated to the type and number of genetic markers used (single genes vs whole-genomes SNPs) as well as the range and resolution of population sampling efforts (Ogden and Linacre 2015).

The problem often faced with wildlife traceability of rare species is the undesirable effect that inbreeding and genetic drift have on the assignment resolution of neutral genetic markers (Arenas et al. 2017). Integrating markers that target geographically-linked traits (adaptive and fast evolving) can counter this, but these are not easily identifiable unless genome-wide studies have previously been conducted (Arenas et al. 2017; Hohenlohe et al. 2021). With less than 1% of the 13 505 species currently listed as threatened by the IUCN having published genomes, genome assemblies are an essential but rare resource that can be useful for conservation management (Brandies et al. 2019).

Molecular forensics of wildlife and their derivatives is the most well-established and widely applied forensic technique owing to the customary precedent of accepting DNA evidence in court cases and the ubiquity of DNA forensic facilities (TRAFFIC 2016). However, DNA samples from wildlife forensic cases are often degraded or at low-copy numbers which reduces the efficacy of obtaining sufficient DNA for geographic traceability using traditional techniques (Arenas et al. 2017; Natesh et al. 2019). Mitochondrial DNA can to some extent overcome this problem owing to its high copy number per cell as well as its circular genome structure which prevents degradation through exonucleases (Foran 2006). However, compared to that of an entire nuclear genome, relying solely on mitochondrial DNA dramatically reduces the number of potentially informative markers available, all of which have a maternally inherited route of

evolution. Microsatellite markers are also robust to degraded samples, but they require effort in development and testing, and they cannot be easily exchangeable across labs (Ogden and Linacre 2015).

Currently for pangolins, using the source and destination of seized shipments is often the main source of information for geographically tracing the trade (Heinrich et al. 2017). However, this method of tracking shipments is insufficient as traffickers are known to smuggle specimens across borders to exploit countries with more lax laws or law enforcement, or that have better infrastructure (especially harbours and airports) by which their consignments can be shipped (Heinrich et al. 2017; Mambeya et al. 2018).

In light of this, and the complementary conservation aspects of molecular traceability to conservation through international regulation, the up-scaling of resources necessary for robust traceability will aid in pangolin conservation.

Thesis Aim and Objectives

The aim of this PhD is to develop research and molecular resources to aid in future population management, conservation genetic, and forensic endeavors required for conserving pangolins, including tracing their illegal trade. This will be achieved by implementing the following objectives:

1. Identify research, commercial, and popularisation trends, together with major knowledge gaps through a systematic review. From here, I plan to propose holistic recommendations that may help conserve pangolins and mitigate their trade (Chapter 1).

2. Conduct the first full genome analysis on all eight pangolin species in order to determine evolutionary relationships, demographic trends, and key genetic parameters for species management (Chapter 2).
3. Delineate and characterize distinct population genetic units of all eight species in order to provide a reference for conservation planning and geographic traceability (Chapter 3).
4. Provide genomic resources for all eight species to form a basis for future conservation genetic/genomic efforts and tracing the illegal trade. This will include a list of markers and the completion of genome assemblies for Pholidota (Chapters 2 & 3).

Thesis structure

This thesis is structured in such a way that each data chapter is an article published (**Chapter 1**) or to be submitted (**Chapter 2 & Chapter 3**). This means that these chapters are in the format of their respective journals and thus do not follow the same structure of subsections or naming style of figures, tables and supplementary material. All supplementary material for these data chapters can be found in the global **Appendices** section whereby it is divided into the three data chapter subsections. The data chapters fall in-between the **Introductory Chapter** and **Conclusion and Synthesis Chapter** which are aimed, respectively, at introducing the thesis topic with its objectives, and providing a summary of its outputs in line with conservation actions to come. The data chapters are written with plural pronouns (i.e. we, our) for submission as articles whilst the introductory and conclusions and synthesis chapters are written with singular pronouns (i.e. I, my). All chapter references are found in a single, global **References** section.

Chapter 1:

A timely systematic review on pangolin research, commercialization, and popularization to identify knowledge gaps and produce conservation guidelines

This chapter has been published:

Heighton, S.P. and Gaubert, P., 2021. A timely systematic review on pangolin research, commercialization, and popularization to identify knowledge gaps and produce conservation guidelines. *Biological Conservation*, 256, p.109042. <https://doi.org/10.1016/j.biocon.2021.109042>.

Sean P. Heighton¹, Philippe Gaubert¹

¹ *Laboratoire Evolution et Diversité Biologique, UPS/CNRS/IRD, Université Paul Sabatier, 118 route de Narbonne, Bât. 4R1, 31062 Toulouse, France*

* Corresponding author: Sean P. Heighton – sean-paul.heighton@univ-tlse3.fr

Abstract

Conservationists can benefit from encompassing research, social, and economic aspects of conservation issues to produce holistic conservation guidelines. Pangolins have become a taxon of conservation eminence as they continue to be poached and trafficked to extinction, and have been mislabeled as potential intermediary hosts of SARS-CoV-2 which is responsible for the COVID-19 pandemic. Applying a systematic review approach, we extracted pangolin-related publications since 1865 from five research databases (814 publications), as well as data on 5 296 patents, online news trends (43 176 articles) and societal interest (Google Trends and Wikipedia Pageviews). Although we detected a significant increase in pangolin-related publications through time, we observed glaring knowledge gaps in contextually important categories including immunology, education, and implications of trade or poaching on populations. All eight species have literature knowledge gaps, however African species are less represented. Fifteen African range-states have no pangolin literature, while the number of publications with non-range-state lead authors increased from 8% to 42.9% since 2017. Pangolin media output and societal interest have remained low relative to other flagship species, however COVID-19 is shifting these dynamics. Pangolin patent production was linked to Traditional Chinese Medicine, which was neither driven by science nor traditional pharmacopoeia. To help conserve pangolins, we suggest increased effort in health and field-based conservation research, while directing more attention towards Africa. We

highlight the importance of maintaining range-author contributions, and of factors that may lead to increased public interest in pangolins. Our approach can be used to devise integrative conservation guidelines for other species.

Keywords: Literature database, patents, Pholidota, public interest, research dynamics, Traditional Chinese Medicine

Introduction

The illegal wildlife trade is a global conservation threat as it has accounted for 132 144 seizures in 120 countries between 2004 and 2014, and has an average of 100 million organisms traded annually (Harfoot et al. 2018; UNODC 2016). Its clandestine and multi-scaled nature involving complex interactions between conservation organizations, law enforcement agencies, economic wealth and society, requires a multi-disciplinary intervention approach (Akella and Allan 2012; UNODC 2016).

Pangolins (Pholidota, Mammalia), comprised of eight species in Africa and Asia (Fig. 1; Gaubert et al. 2020), have been given the unfortunate label of being the most heavily trafficked wild mammals on earth (Challender et al. 2014, 2020a; Heinrich et al. 2017). All eight species are listed on Appendix I of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES 2017), and either as Vulnerable, Endangered or Critically Endangered on the IUCN Red List of Threatened SpeciesTM, almost exclusively due to the illegal trade (Fig. 1; IUCN 2019). The most recent estimates suggest that around 895 000 pangolins were trafficked between August 2000 to July 2019, notably from Africa to Asia (Challender et al. 2020a), however this represents only a fraction of absolute trade volumes (Phelps and Webb 2015) while other

studies predict larger estimates through regional analyses (see: Ingram et al. 2018). The main driving force of this trade is the use in Traditional Chinese Medicine (TCM), but other traditional medicine, ornamental, and meat markets have been known to contribute (Baiyewu et al. 2018; Heinrich et al. 2017; Ingram et al. 2018; Shairp et al. 2016).

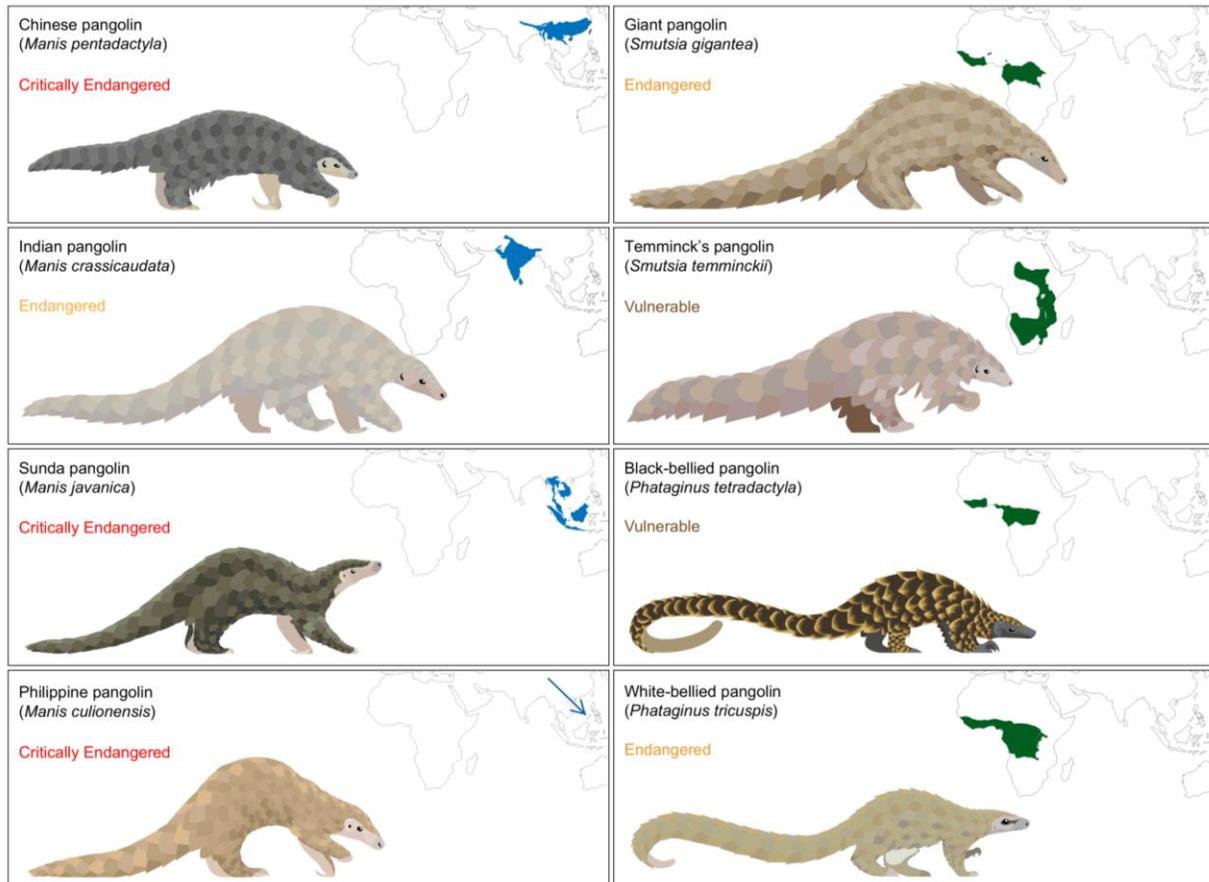


Figure 1: Global conservation status, taxonomy, and distribution of the eight extant pangolin species. Distribution maps from The IUCN Red List of Threatened Species™ 2019. Modified from Gaubert et al. (2020), illustrations by Sheila McCabe.

The illegal wildlife trade has made pangolins one of the most iconic animals amongst conservationists. Paradoxically, these nocturnal, myrmecophagous mammals have widely been considered understudied (Challender et al. 2014; Pietersen and Challender 2020). In the only review on pangolin research, Pietersen and Challender (2020) identified six study fields requiring

urgent attention, including trade networks, forensics, biology and ecology, genetics, husbandry and veterinary health, and the effects of climate change. Two reviews have been published on the public interest aspect of pangolin conservation, pointing to the importance of seizure records and accompanying ‘shocking’ imagery, conservation events and organizations, specific Google Doodles, documentaries, and celebrities in pangolin popularization (Harrington et al. 2018; Thomson and Fletcher 2020). However, four years have passed since the last systematic review data collection in 2016 (Harrington et al. 2018), and pangolins have recently been thrust into the public spotlight after being suggested as the intermediate host of SARS-CoV-2 which is responsible for the COVID-19 pandemic (Zhang et al. 2020c), despite no concrete evidence for this claim (Frutos et al. 2020). New peaks in interest and their determining factors have likely emerged, which could help conservationists and researchers to increase the public awareness regarding pangolins. Critical information on the economic aspects of the pangolin trade such as drivers and market dynamics are still lacking (‘t Sas-Rolfes and Challender 2020; Challender et al. 2015), while there is limited information on the commercialization of pangolin products by the TCM industry (Xu et al. 2016).

Conservation biology is part of a multifaceted system where solutions to conservation issues require a holistic, multidisciplinary approach integrating biological data (e.g. knowledge on the species), society, economics, and governance across multiple scales (Blair et al. 2017). Thus, establishing accurate conservation strategies relies on the availability of biological knowledge of conservation issues (Conenna et al. 2017; Williams et al. 2020), together with knowledge of public awareness and interest in a conservation problem (Olmedo et al. 2018; Phillis et al. 2013), and on the commercial determinants that result in species being exploited (e.g. Masters et al. 2020). For these reasons, we present an unprecedented systematic review of scientific knowledge of pangolins

since 1865, build on our knowledge of their popularization, and delve, for the first time, into their commercialization in the form of patent production. We identify the research, commercial, and popularization trends, together with the major knowledge gaps in natural sciences, from which we propose timely, holistic recommendations that may help conserve pangolins and mitigate their trade. We found no evidence of a review of this nature from the query ‘pangolin’ in the Conservation Evidence (<https://www.conservationevidence.com/>) database (query results: Heath and Coulson 1997; Shepherd et al. 2017). We hope these methods and insights may be useful for other species of conservation importance, particularly those affected by the illegal wildlife trade.

Methods

2.1. Pangolin scientific research database

To conduct this study we followed systematic review guidelines of conservation literature (Collaboration for Environmental Evidence 2018). Boolean search terms were used to broadly capture pangolin literature from five databases covering natural sciences, engineering, biomedical sciences, and humanities and social sciences. We extracted literature containing “pangolin” OR “scaly anteater” OR “spiny anteater” NOT “drosophila” (*pangolin* gene described in *drosophila*) NOT “Tachyglossus” (*echidna* genus, often referred to as “spiny anteater”) in their titles, abstracts or keywords from Web of Science (n= 1033; extracted 14/11/2019 from “all databases”), Scopus (n= 431; extracted 14/11/2019), Engineering Village (n= 40; extracted 6/12/2019), Project MUSE (n= 11; extracted 23/12/2019), and a personal dataset of older scientific literature compiled by one of us (n= 158; extracted 9/12/2019). For the humanities and social science database (Project MUSE), the default was to search through all literature content rather than just titles, abstracts and keywords, whilst only articles were included due to limited access to most books.

The five datasets were combined (n= 1 631), and duplicated literature by title were deleted with sed and awk Linux commands followed by manual inspection. The resultant global database (n= 997) included books, book chapters, reports, conference and workshop proceedings, and scientific articles (Appendix A). Titles and abstracts of this database were manually reviewed with 183 pieces of literature (from herein: publications) being omitted due to no mention of pangolins, whilst the remaining 814 publications were assigned a “Focus” (n= 561) or “No Focus” (n= 253) tag. “Focus” was assigned if the literature’s main focus was on pangolins or they were part of a direct comparison (*contra* “No Focus” literature). Additionally, a scientific domain was assigned to these 814 publications (Appendix B). As the “natural sciences” domain encapsulates the bulk of knowledge of pangolins (n= 662) and our main research interest, a theme (nine in total), corresponding primary and occasionally secondary category (30 in total; some publications had more than one category and corresponding theme), as well as whether it was a review or not, were additionally allocated to each publication (Appendix B). Themes and their corresponding categories were strictly defined by our list of definitions (Appendix C) which guided decision making. Definitions were outlined prior to selection of a subset of around 100 publications, which were then independently vetted by a second reviewer after which discrepancies were discussed and amended in the definitions. This process was repeated twice more with smaller subsets of data during the selection process. Focused, “natural sciences” literature (n= 467) were then manually reviewed to identify whether lead authors (i.e. first and/or last authors) were affiliated to range-states or not (n= 454), which species were under study (n= 466), and the range-states involved through study or specimen collection site (excluding museum collections) as a proxy (n= 335). The use of multiple databases allowed us to include literature from different domains, accessibility, and languages (51 languages are suggested to be covered between Scopus and Web of Science

alone), however there is still likely a bias towards English literature (Vera-Baceta et al. 2019). There is also a possibility of missing publications, particularly from rare languages, grey literature sources, books, and novel or cryptic journals, while not all published studies have equal merit. However, considering the volume of literature included in this review, we believe this may have not significantly altered our results.

Online news and societal interest trends

Daily, global online news article counts (both subject count and total monitored article count in over 100 different languages) were collected from the Online News Summary dashboard in GDELT Project (Global Database of Events, Language, and Tone; <https://www.gdeltproject.org/>) with '(pangolin OR pangolins)' as the query, starting between 01/01/2017–31/03/2020. We obtained a relative daily value of pangolin articles by dividing daily pangolin article counts by the daily number of monitored articles. The aforementioned process was repeated with the “Raw Timeline” (non-normalized) output type in “Comparison Visualization” to compare global online news coverage of pangolin(s), elephant(s), and lion(s). We extracted the same monthly animal comparison data from Wikipedia’s Pageviews Analysis (<https://tools.wmflabs.org/pageviews/>) between 01/07/2015–31/03/2020. Monthly societal interest was collected in the same manner as that of online news articles, however we used the global “Web Search”, “News Search” and “Image Search” in Google Trends (<https://www.google.com/trends>) of the “Animal” topic ‘pangolins’ from 01/01/2008–31/03/2020. Whilst Wikipedia provides raw counts of page views, Google Trends uses a relative search count of all searches by the chosen region and period of interest (Kämpf et al. 2015). Although there may be a slight geographic or language-specific bias in these tools, Google is still the largest search engine globally while Wikipedia has pages in over

304 languages (https://meta.wikimedia.org/wiki/List_of_Wikipedias).

Patents

Pangolin commercialization was represented in the form of patents that were mined from the Web of Science database (n= 5 296; extracted 14/11/2019) with the same Boolean search terms used for the extraction of scientific literature (Appendix D). Class codes from the Derwent Innovations Index on Web of Science were extracted for all patents except two (n= 5 294).

Data analyses

Negative binomial generalized linear models (best fit model by AIC; see Appendix E) with package “mass” (Ripley et al. 2013) were applied to investigate the total, focused and non-focused literature count trends through time. Due to some databases including literature for the beginning of 2020, this year was excluded for the aforementioned analyses as the incomplete count for this year would affect the trend. A likelihood ratio test with package “lmtest” (Hothorn et al. 2019) was utilized in order to test whether focused and non-focused literature slopes were significantly different. A Chi-squared goodness-of-fit and Chi-squared multi-pairwise-comparison test with Bonferroni correction were run with package “RVAideMemoire” (Hervé and Hervé 2020) to investigate whether domains (total literature) were significantly different from the expected equal distribution of literature count, and if so, which domains differed significantly from one another. From this point, only focused data of domain “natural sciences” were analyzed and the process above was repeated for themes, categories and species. However, for species, the expected values for the chi-squared goodness-of-fit analysis was weighted by the relative area of each species’ distribution, which was extracted from the IUCN Red List of Threatened SpeciesTM (IUCN 2019)

and transformed in QGIS v3.4.10 (QGIS Development Team 2016). Each piece of literature was treated as a separate data point except for focused “natural sciences” publication analyses by category (and corresponding theme; n= 112), species (n= 112), and range-state origin (n= 2) where each instance was treated as a separate data point as more than one instance may occur per publication.

Heatmaps were used to visualize trends in the “natural sciences” and to facilitate comparisons of multiple factors at once. Heatmaps of literature count for species *versus* category and theme, species *versus* year of publication, and category *versus* year of publication were created through the packages “ComplexHeatmap” (Gu 2015) and “RColorBrewer” (Neuwirth 2014). These heatmaps, along with all statistical analyses and their relevant assumptions, were considered or tested in the Rstudio interface of R v3.6.1 (Rstudio Inc., Massachusetts, U.S.A; Appendix E). To gain further insight into pangolin scientific knowledge, focused, “natural sciences” publications were scrutinized in VOSviewer v1.6.14 (van Eck and Waltman 2010) by the program constructing co-occurrence networks of important terms mined from the titles and abstracts. Instead of a relevance threshold as determined by VOSviewer, we used a manually curated thesaurus file based upon an initial network of 353 terms (Appendix F). This allowed us to combine multiple terms used for each species and decide which terms were irrelevant for the co-occurrence network. The process was repeated for patent records, but differed by extracting terms from the titles only and utilizing a relevance threshold of 60% of the most relevant terms instead of a thesaurus file.

Results

Pangolin science overview

The pangolin science trend encapsulating all publications (focused and non-focused; n= 814) is

significantly positive through time from 1865 onward (Estimate= 0.048, $z= 19.31$, $p < 0.001$, CI= 0.044–0.053; Fig. 2). The average growth rate for all pangolin-related publications between 1980 and 2012 is 19.97%, more than 6.75-fold the 2.96% publication growth rate of modern science estimated by Bornmann and Mutz (2015) during the same period. Publication count differed significantly according to domain ($\chi^2= 2456$, $df= 5$, $p < 0.001$), with “natural sciences” ($n= 662$) being significantly higher than all the other domains ($p < 0.001$ across all pairwise comparisons with other domains). For the other five domains ($n= 152$ combined), spikes in publications are observed in the early 1990’s (1990–1995) and middle 2000’s (2002–2008 with the exception of 2004; Fig. 2). During this second peak, “biomedical sciences” appears to be a predominant alternative domain which is followed by an increase in patent count per year (2004–2010; Fig. 2).

1. Systematic review

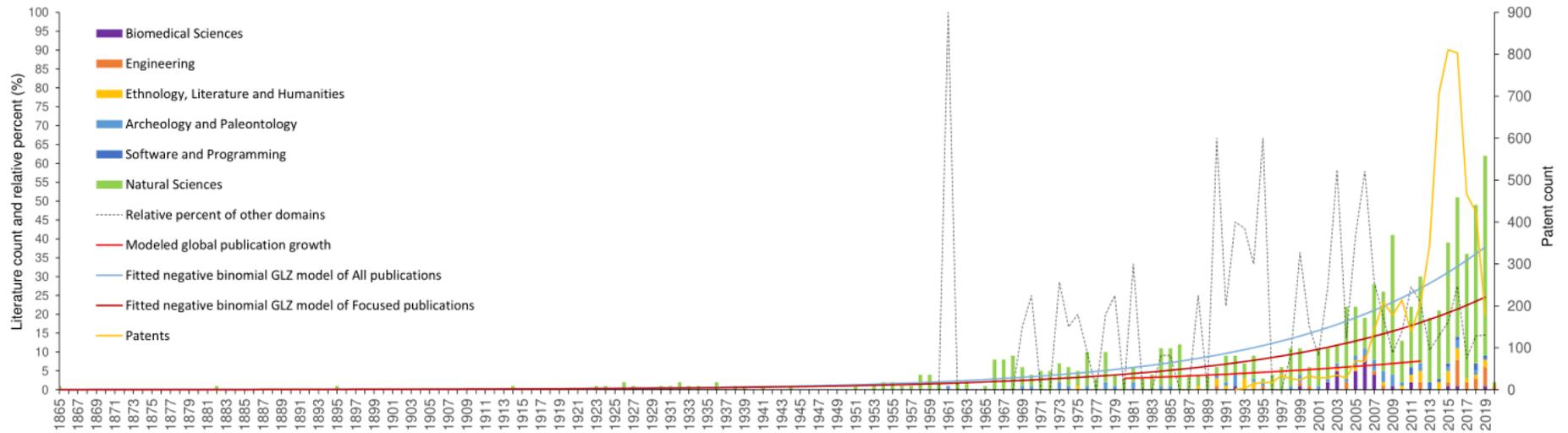


Figure 2: Pangolin research (literature focused and not focused entirely on pangolins) from 1865 to 2019. Each bar plot represents the literature count per year (1st y-axis) split into the six scientific domains. Trend lines from 1865 to 2019 are fitted values of the negative binomial generalized linear (GLZ) models for all publications (focused and non-focused literature) and focused publications (focused literature only) during this period. The line from 1980 to 2012 is the modeled global publication growth during that period, as estimated by Bornmann and Mutz (2015). The patents line is fitted to the patent count per year (2nd y-axis). The proportion of other scientific domains relative to “natural sciences” (1st y-axis) commences at 1961 as the first publication we could find from a domain other than “natural sciences” was from this year.

Pangolin (focused) natural sciences trends

We concentrated on literature that had a pangolin focus (vs. non-focus literature) for further analyses as they are more likely to be seen as literature aimed at increasing our knowledge of pangolins (see Appendix E for statistics on each trend and their difference). In the “natural sciences” domain, themes differed significantly from expected equal counts ($\chi^2= 345.04$, $df= 7$, $p< 0.001$), with the theme “conservation” ($n= 179$) being significantly higher than all other themes ($p< 0.05$), apart from “physiology and anatomy” ($n= 146$; $p= 1.00$). The themes “ethnzoology and society” ($n= 16$), “general description of pangolins” ($n= 16$), and “veterinary” ($n= 17$) grouped together and were significantly lower than other themes in research output ($p< 0.05$). “Veterinary” was less well distributed across species, including gaps in the literature for Philippine (*Manis culionensis*), giant (*Smutsia gigantea*) and black-bellied pangolins (*Phataginus tetradactyla*; Fig. 3).

Although categories differed significantly from expected equal counts ($\chi^2= 658.67$, $df= 26$, $p< 0.001$), no categories were significantly different from all other categories, or could be grouped together in the multiple pairwise comparison test ($p> 0.05$). “Anatomy” and “range/density/habitat/niche/behavior/predation” were joint highest ($n= 94$), while “volume or nature of trade and poaching” followed with 53 publications. However, unlike the other two, “anatomy” was heavily skewed towards the top three species by count, namely the Chinese pangolin (*M. pentadactyla*), Sunda pangolin (*M. javanica*), and in particular white-bellied pangolin (*P. tricuspis*; Fig. 3).

A total of 51.85% of categories had missing information for at least one of the eight species ($n= 27$ including “veterinary” and “general description of pangolins”) with “education” ($n= 2$) and “immunology” ($n= 3$) being the least represented across species and in literature output (Fig. 3).

Categories like “anthropogenic effects other than trade or poaching” (n= 6), “implications of trade or poaching for populations” (n= 8) and “people’s perception and awareness of pangolins” (n= 5) are significantly lower ($p < 0.05$) than the publication count of “volume or nature of trade and poaching” (n= 53). Publications on “captive breeding” (n= 29) are more than double that of “rehabilitation” (n= 13). Although many categories are not prominent in pangolin research, through time there has been an increase in the number of categories being studied (Appendix G). By publication count alone, African pangolin species are generally less represented (Fig 3). *Manis pentadactyla* (n= 161), *M. javanica* (n= 137), and *P. tricuspis* (n= 125) are more studied relative to the other five pangolin species, with the former two Asian species receiving good coverage across categories, missing three out of 27 categories (Fig. 3). *Manis culionensis* (n= 31) is the most understudied, whereas *S. gigantea* (n= 54), the second least studied, is close to *P. tetradactyla* (n= 62) and Temminck’s (*S. temminckii*; n= 69) pangolin publication counts (Fig. 3). The three species with the lowest publication counts also have the highest number of missing categories, each missing literature in 11 out of 27 categories (40.74%; Fig. 3). The two African species, *S. gigantea* and *P. tetradactyla*, share the same missing categories, whilst commonalities in missing categories can be observed for all species (Fig. 3). Through time, more research interest is being shared across all species rather than focused on historically prominent species like *M. pentadactyla* and *P. tricuspis* (Appendix H). By controlling for the area of distribution of each species, the Indian (*M. crassicaudata*; n= 94, $p > 0.05$) and *P. tetradactyla* (n= 62, $p > 0.05$) did not differ significantly from their expected publication counts. *Manis culionensis* (n= 31, $\chi^2 = 2651.382$, df= 1, $p < 0.001$), *M. javanica* (n= 137, $\chi^2 = 99.897$, df= 1, $p < 0.001$), and *M. pentadactyla* (n= 161, $\chi^2 = 115.801$, df= 1, $p < 0.001$) were significantly higher than expected after accounting for their area of distribution, whilst *S. gigantea* (n= 54, $\chi^2 = 93.2236$, df= 1, $p < 0.001$), *S. temminckii* (n= 69, $\chi^2 =$

130.846, $df = 1$, $p < 0.001$), and *P. tricuspis* ($n = 125$, $\chi^2 = 9.125$, $df = 1$, $p = 0.02$) were significantly lower.

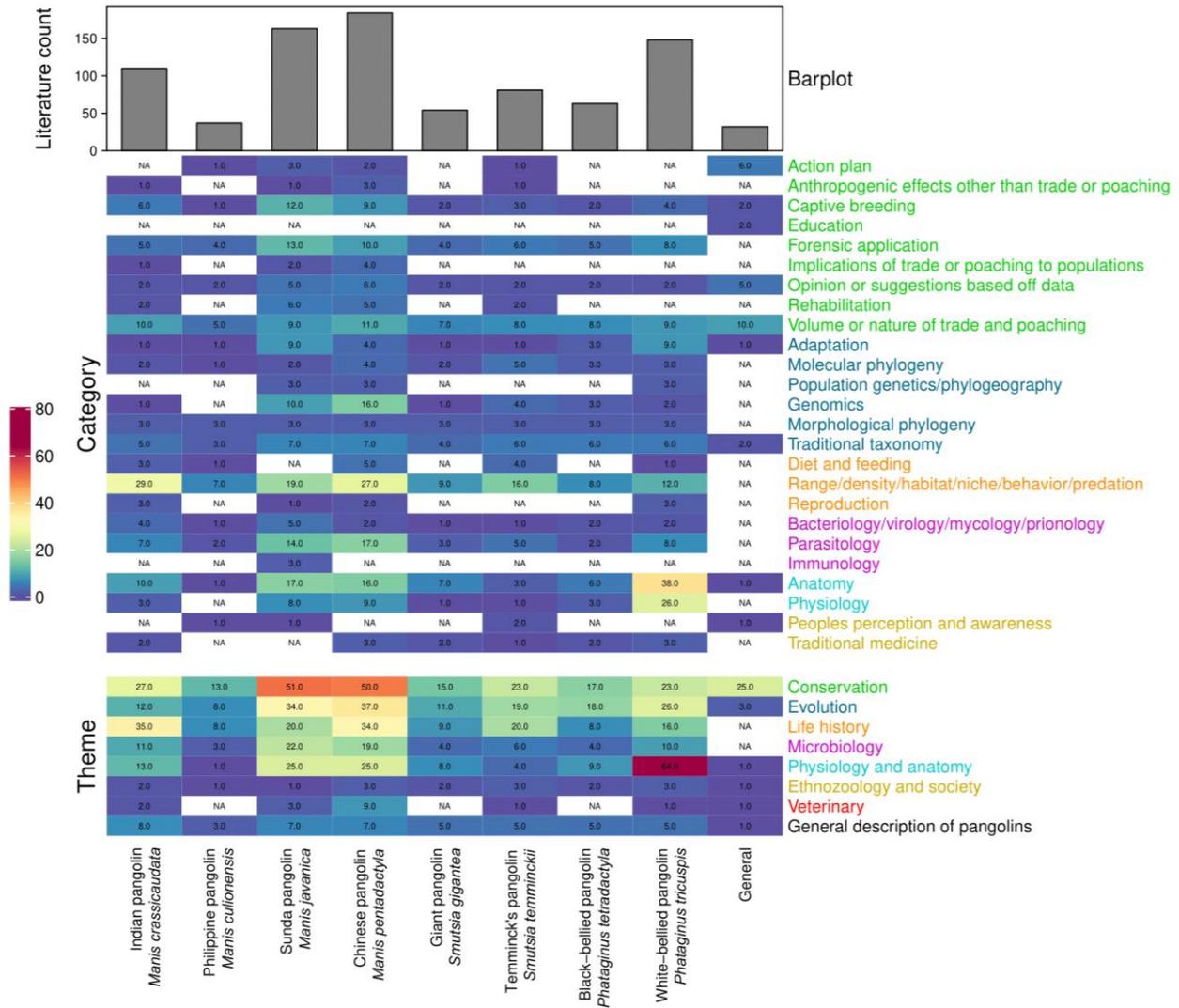


Figure 3: Heatmap of literature count in research themes and categories by species from literature with a pangolin focus and in the scientific domain “natural sciences”. Except for the “Veterinary” and “General description of pangolins” themes, which do not have categories, each set of categories are color coded to their corresponding research theme. An additional species slot named “General” is specified for literature that refers to pangolins in general and does not concentrate on a specific species. Each block in the heatmap contains the number of pieces of literature relating to a theme/category by species, whilst “NA” refers to absence of data.

We identified five clusters based on relationships between phrases from the titles and abstracts of focused, “natural sciences” publications with VOSviewer (Fig. 4). The first cluster

reflects rehabilitation and captive breeding with a large occurrence of terms relating to disease and parasites. The second cluster reflects anatomy and physiology, with *P. tricuspis* situated as a center term and heavily linked to comparative studies with bats and rats. The third relates to pangolin trade which is the largest and most densely populated cluster with a large proportion of links with other clusters (apart from the anatomy and physiology cluster). A fourth cluster relates to evolution and phylogeny with a subset of terms evidencing a molecular forensic link with the trade cluster (“identification”, “number” and “region” linking to “illegal trade”, “sample”, “protection” and “conservation”). Ecology, the fifth cluster, is the smallest, with *M. crassicaudata* being a prominent point. *Manis pentadactyla* and *M. javanica* are central to the entire cluster network.

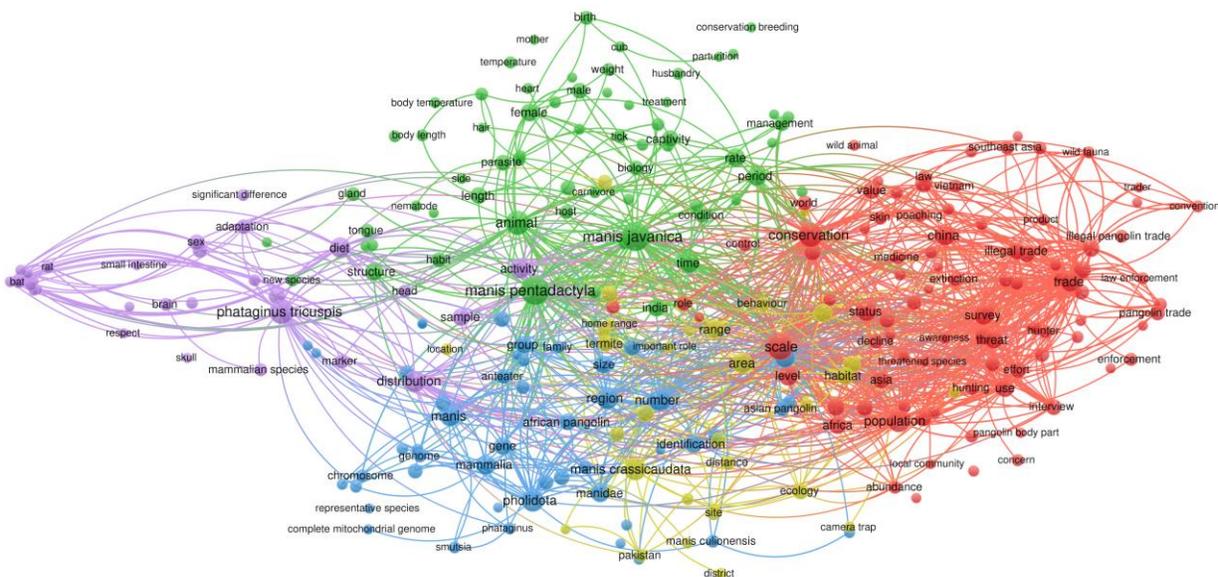


Figure 4: A co-occurrence network of terms and their links extracted with text mining algorithms in VOSviewer, from titles and abstracts of literature, with a pangolin focus and in the scientific domain “natural sciences”. A binary counting approach (presence or absence of a term per publication) was employed with a minimum term occurrence count of five and a manually curated thesaurus file to filter irrelevant terms (Appendix F). Phrase occurrence was used for weighting the network. The clusters are separated into rehabilitation and captive breeding (green), anatomy and physiology (purple), evolution and phylogeny (blue), ecology (yellow), and trade (red).

Pangolin natural sciences research dynamics

Publications with lead authors from range-states (i.e. either first or last authors from countries with

pangolins; n= 314) were 2.24-fold higher than publications without (n= 140; Fig. 5). In the last three years this trend is starting to change, with publications involving non-range-state lead authors increasing from 8% of the publications in 2017 (n= 2) to 42.86% in 2019 (n= 15; Fig. 5). Formulated on sample origin or study site, literature involving Asian range-states (n= 288) are 2.29-fold more than that of African range-states (n= 126), however there is also a shift in dynamics from 2017 (n= 5, 21.74% of publications from Africa) to 2019 (n= 11, 36.67% from Africa; Fig. 5). There are 15 African range-states and one Asian range-state (Bhutan) having no publication records of pangolins derived from sample collection or study sites (Fig. 5). Nigeria (n= 32) is the African country that is most heavily involved in pangolin research, accounting for 9.5% of all research by country (n= 337), placing it third behind China (n= 69, 20.47%) and India (n= 52, 15.43%; Appendix I).

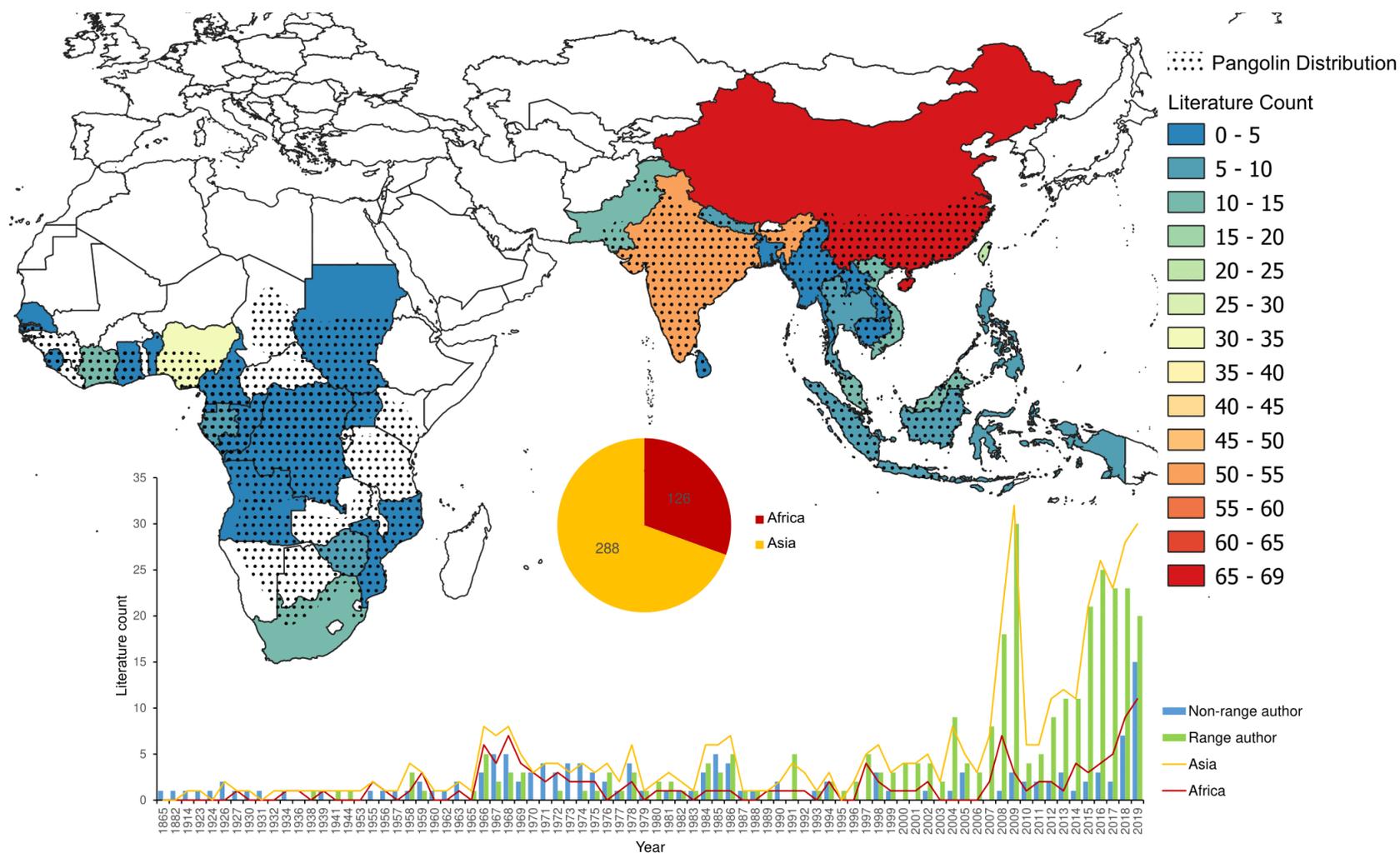


Figure 5: Pangolin research dynamics by range-state (map), continent static (pie chart), continent temporal (line graph), and authorship through time (bar plot). Publications used for the literature count are of a pangolin focus and in the scientific domain “natural sciences”. The map was created by adding the count data per country to a world map shapefile in QGIS v3.4.10 (QGIS Development Team 2016) by merging a CSV file containing data of count per country to the shapefile with the FIPS country code as a common merging string. Pangolin distribution is comprised of the IUCN Red List of Threatened Species™ (2019) distributions of the eight extant pangolin species. Values on the pie chart refer to literature count.

Pangolin popularization

We found four noticeable peaks of news and / or societal interest in pangolins since 2014, including April 2015 and February 2017 with similar sized peaks, May 2018 with the second-largest peak, and February to March 2020 with the largest peak (Fig. 6a). These peaks of interest surpassed those caused by the record-breaking seizures of trafficked pangolins made in February, April and December 2019, however the former two seizures received a comparatively large response by online news media (Fig. 6a). In fact, apart from the two largest peaks in public response, peaks of societal interest did not always appear to correlate with peaks in news media outputs. Although news articles mentioning pangolins did not surpass those of lions or elephants, Wikipedia page views did, with over 1 million views in March 2020 (Fig. 6b).

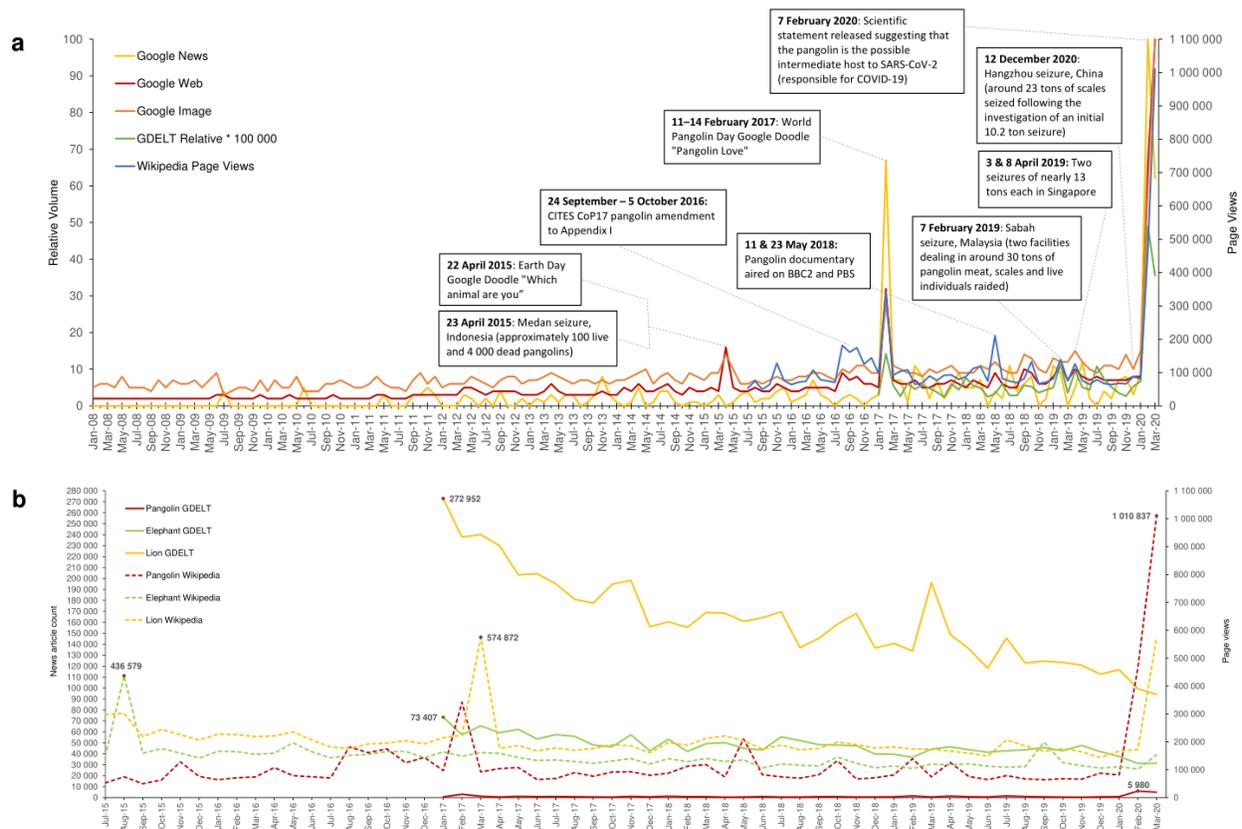


Figure 6. Public interest in the form of societal interest and online news article trends. a) Monthly societal interest is displayed by Google News, Web, and Image relative search volumes (1st y-axis) and Wikipedia page views (2nd y-

axis), while online news articles are displayed by GDELT relative values (1st y-axis). GDELT relative values were calculated by dividing the daily pangolin article count by the total number of articles monitored per day, and multiplying this value by 100 000 in order to fit on the relative volume axis (1st y-axis). Peaks of public interest and the largest seizures recorded are indicated in text boxes. b) GDELT article count (1st y-axis) and Wikipedia Page views (2nd y-axis) per month for lions, elephants and pangolins are displayed for comparative purposes between these iconic conservation species. The highest count for online news article and page views are displayed for each trend.

Pangolin patents

The most noticeable growth in patents per year occurred from 2011–2016, after which it dropped dramatically (Fig. 2). A total of 96.35% (n= 5 101) of patents are related to the Derwent Class Code B04 which are patents in pharmaceuticals containing natural products and polymers (Appendix J) and thus likely related to Traditional Chinese Medicine (see also: Appendix K). The first patent recorded was in 1993, a year after the first biomedical study we found containing pangolins in the study. The majority of top patent assignees were single authors with a few biomedical companies found across the upper part of the list, although the biomedical research author affiliations did not match with these companies (Appendix L).

Discussion

Pangolin research

Pangolin research has seen a significantly positive, exponential growth since 1865, 6.75-fold higher than that of the global research average. The knowledge garnered has been substantial enough to culminate in the first comprehensive book on pangolins (Challender et al. 2020b). “Natural sciences” is the most studied domain, although important knowledge gaps exist (see below).

High count research areas require direction

“Conservation” is the most studied theme (significantly different to all, except “physiology and

anatomy”), due to pangolins notoriety in the illegal wildlife trade. Prominence in “physiology and anatomy” may be attributed to it being comprised of comparatively older science categories (Habbal 2017; Pinter and Pinter 1993), as observed for pangolins (Appendix G). Additionally, pangolins – as ‘scaly anteaters’ – have unique anatomical and physiological features (Gaudin et al. 2020) that may explain why “anatomy” is the joint highest category for pangolins (with “range/density/habitat/niche/behavior/predation”), though it is skewed by species, particularly toward *P. tricuspis* (Fig. 4).

Pangolin ecological data (“range/density/habitat/niche/behavior/predation”) is well studied, with *M. crassicaudata* flagged as a well-represented species in the VOSviewer network. Nevertheless, it is suggested that there is a lack of population density and home-range estimates for some species (most notably *S. gigantea* and *P. tetradactyla*) and locations (e.g. *S. temminckii*’s eastern and northern ranges), as well as knowledge of factors affecting distribution and habitat preferences (Pietersen and Challender 2020). These are integral in determining baseline estimates for measuring the effects of habitat degradation and illegal trade on populations. Given the pangolins’ elusive nature, direct ecological monitoring studies may be challenging and costly, hence the use of proxy methods such as burrow counts, community interviews and molecular techniques have been suggested as alternatives (Ingram et al. 2019). Meta-analyses on the theme “checklist and monitoring” (36 publications in 19 countries under non-focused data) created in our review can be used as a new proxy method. Since literature were only screened by title, abstract and keywords, it is likely that more checklist or monitoring literature is available.

The third highest category, “volume or nature of trade and poaching”, has a high count as a result of the illegal trade being the principal conservation concern for pangolins (Challender et al. 2014). The number of links noted in the trade cluster of the VOSviewer analyses also

highlighted that China is frequently indicated in trade-related topics, likely as it is the main market driving demand (Heinrich et al. 2017). However, local consumption and bushmeat market trade is also a contributor to wild pangolin offtake and needs to be investigated further (Ingram et al. 2018). Tackling this illegal trade requires various approaches (Akella and Allan 2012), one of which could be molecular forensic identification of species and geographic origins of illegally traded specimens, as noted in the VOSviewer linking words (Gaubert et al. 2018; Zhang et al. 2020b). Despite its potential reach, there is a lack of georeferenced population genetic/phylogeographic inferences to trace the trade and determine evolutionary significant units (Zhang et al. 2015b), while there are currently only two species with published full genomes (Choo et al. 2016; Hu et al. 2020a), and limited development of population genetic markers (e.g. Aguillon et al. 2020; Luo et al. 2007). Forensic approaches (including stable isotopes and morphometrics) are a means of direct conservation interventions involving law enforcement and conservation organizations (Kotze et al. 2020; see also “forensic application” category in our database), and should be considered in more depth.

Although certain evidence-based outputs played an important role in pangolins receiving their conservation icon status through their trade (Heinrich et al. 2017; Pantel and Chin 2009; Xu et al. 2016), we still only have limited examples of accurate data pertaining to the role that the illegal wildlife trade plays in pangolin population declines and extinction (Irshad et al. 2015; Wu et al. 2004), genetic variation loss (Hu et al. 2020a), and likelihood of disease epidemics (e.g. Clark et al. 2009; Liu et al. 2019). These demographic (population structure, abundance and density), epidemiological (disease transmission and reservoirs), and evolutionary (population genetics) lines of evidence are integral for management decisions of populations (Cleaveland et al. 2007; Lande 1988).

Inequality in conservation research effort

The aforementioned examples of risks associated with illegal wildlife trade fall within “implications of trade or poaching to populations”, which is a substantially underrepresented conservation category, particularly when compared with “volume or nature of trade and poaching”. The same is noted for “anthropogenic effects other than trade or poaching” and “people’s perception and awareness of pangolins”. The extent of the former is limited to electric game-fencing (Pietersen et al. 2014), domestic dogs (Sun et al. 2019), road mortalities (Murthy and Mishra 2010), habitat destruction (Maddox 2007). Further research is required to elucidate the full spectrum of threats faced by pangolins (e.g. climate change). Research into “people’s perception and awareness of pangolins” is integral in determining the roots of demand, as well as awareness of pangolin legislation and their plight, which may better guide educational, awareness, and behavioral change programs (Baiyewu et al. 2018; Zanvo et al. 2020).

This contrast in research effort also applies to other categories, for instance references to “captive breeding” are double those of “rehabilitation”. Literature on best practices for rehabilitation and release is a necessity considering the high level of live pangolins trafficked (Heinrich et al. 2017) and the conditions under which they are trafficked, which leads to complications or mortalities during rehabilitation (Clark et al. 2009). Nonetheless, data on nutrition and welfare practices (Cabana et al. 2017; Zhang et al. 2017) in “captive breeding” are also important for rehabilitation.

The major knowledge gaps

The two lowest themes by publication count, “veterinary” and “ethnzoology and society”, are of conservation concern. The latter relates to “people’s perception and awareness of pangolins” but

also includes pangolin use in traditional medicine and food. More research in this theme would allow for better insight into demand through traditions and social norms, as well as when hunting pressure shifts from local to global, often unsustainable, commercial trade (Baiyewu et al. 2018; Ingram et al. 2018). Understanding the intricacies behind the demand from the perspective of the consumers may provide insights into reducing it (Wang et al. 2020b). Regarding the “veterinary” theme, the lack of literature on veterinary procedures and guidelines is concerning considering the aforementioned live trade and complications surrounding confiscated individuals. Research and conservation practices may also cause pangolin mortality and morbidity if the proper standards of anesthesia and veterinary care are not adhered to (see: Connelly et al. 2019). A recent chapter reviewing veterinary procedures for a variety of complications and diseases (Wicker et al. 2020) is an important step, but as noted by the authors and our research, there is limited information on *P. tricuspis*, *S. gigantea* and *M. culionensis*. *Phataginus tricuspis* requires particularly urgent attention in the veterinary field considering it has been the most traded pangolin species in recent years (Challender et al. 2020a).

Literature count of the category “education”, namely knowledge of the effectiveness or description of education programs, is alarmingly low, whereas such knowledge is critical to pangolin conservation and should be shared when possible (Ardoin et al. 2020). “Immunology” is an important research avenue that requires more effort as many pangolins in rehabilitation or captive breeding die from infections (Clark et al. 2009), diseases found through veterinary examination are likely linked to pathogens (Wicker et al. 2020), and innate immune defenses are suggested to be reduced in pangolins (Choo et al. 2016; Fischer et al. 2020). These, along with recent studies isolating pathogens of public health concern from pangolins (Liu et al. 2019; Wang et al. 2019) and high levels of trade highlight the urgent need for research into pangolin health,

immunity and epidemiology.

Species bias

Accounting for the area of distribution for each species produced different species bias results when compared with species literature count alone. Both of these measures are important, depending on the types of studies conducted. Ecological monitoring, phylogeographic, poaching, and epidemiological studies may benefit from taking into account species distributions as these are location-specific measures, whilst anatomical, veterinary, rehabilitation, and reproduction studies may only require species-specific measures.

By literature count alone, although African species are less represented, it is clear that *M. culionensis* is the most understudied species, likely as a result of it having the smallest distribution and subject to reduction in population sizes across its range (Schoppe et al. 2019). It may also be a result of the species only recently being described as distinct to the *M. javanica* (Gaubert and Antunes 2005; Gaubert et al. 2018). Data on ecological monitoring of *S. gigantea*, the second-most understudied species, suggest that population estimates of the species are likely low, and thus less likely to be studied (Bruce et al. 2018).

There is a difference between the number of categories covered by the top two species by count (11% missing for *M. pentadactyla* and *M. javanica*) and those by the bottom three species (41% missing for *M. culionensis*, *S. gigantea*, and *P. tetradactyla*). Furthermore, in the VOSviewer analysis, *M. pentadactyla* and *M. javanica* are central to the network, suggesting that these two historically prominent species are well represented in pangolin literature. This along with more than half of the 27 categories having missing literature for at least one species, suggests that there is a clear species bias. We suggest that targeted funding and research effort on underrepresented

species at a per category basis may be an accurate and cost-effective method in gaining the critical missing knowledge for each species (Williams et al. 2020). The commonality in categories missing for all species, particularly between the *S. gigantea* and *P. tetradactyla*, is possibly linked to the formation of novel categories in pangolin biological science (i.e., “population genetics/phylogeography”, “implications of trade or poaching to populations”; see: Appendix G) or difficulty in studying certain categories (i.e., “reproduction”, “immunology”, “diet and feeding”).

Research dynamics

With an increasing trend in non-range-state lead authors in recent years, probably due to increased global awareness of pangolins, range-state authors are playing less of a leading role. Although this trend is over a small timescale (since 2017) and we only analyzed lead author dynamics, conservation research is often conducted outside of biodiverse areas by foreign researchers, whilst researchers from countries of high conservation importance have little representation in international fora (Reboredo Segovia et al. 2020; Wilson et al. 2016). Collaboration with local researchers, experts, and institutions is needed to surpass some of the barriers encountered in developing research countries, in which the majority of pangolin distributions lie (Reboredo Segovia et al. 2020). It also provides research and conservation projects with local knowledge, access to local conservation actors, and a basis to initiate long-term studies, all of which are critical for successful conservation action (Kainer et al. 2009).

The dominance of research in Asian range-states relative to African range-states is clear through literature count per year and per country, as well as by number of range-states having no associated pangolin research. This could be due to Asian species being of greater conservation

concern through their historical dominance in the illegal trade for TCM, which only shifted to African species within the last decade (Challender et al. 2020a). Relative increase in proportion of studies in African range-states over the last few years suggests this to be true, but also may have been influenced by increased research effort of foreign researchers. Africa is also comparatively less populated, has a lower GDP per capita, higher incidence of poverty and civil war due to political instability, and lacks research resources (Elbadawi and Sambanis 2000; UN 2019). These influence its research and conservation potential, suggesting that increased pangolin research, collaboration and funding effort in Africa is required.

The high GDP, human population, and illegal pangolin trade may explain why China (main trafficking import country), India, and Nigeria (major trafficking source country) have the highest research output of all range-states (Heinrich et al. 2017; Meijaard et al. 2015). The two Asian countries are also the most populated countries in the world, whilst Nigeria is the most populated country in Africa (UN 2019). Due to the large distribution of pangolins, we did not analyze the species and category publication count per range-state trends in this review, however our database provides the means to do so for future, localized reviews (region/country).

Pangolin popularization

Determining underlying events that create large peaks in public interest is a systematic method that can be used to inform conservationists of how best to gain public support and funding (Phillis et al. 2013). Our findings concurred with Harrington et al. (2018) and Thomson and Fletcher (2020), namely: ‘shocking’ imagery receives greater societal interest than any other seizure information (April 2015), documentaries from large broadcasting platforms can create a large response while concurrently providing detailed awareness (May 2018), and Google Doodle’s have

had the largest influence on societal interest and news media coverage (April 2015 and February 2017). That is until scientists from South China Agriculture University released a statement (7/02/2020) that pangolins were a possible intermediate host to SARS-CoV-2, which is responsible for the COVID-19 pandemic. Regardless of there being no concrete evidence backing this claim thus far (Frutos et al. 2020), it resulted in pangolins having a higher societal interest by Wikipedia Pageviews than either elephants or lions for the first time since 2015. However, the peak in online media output was substantially lower than that of elephants and lions during the same time period. Our results also indicate that despite record-breaking seizures receiving a comparatively large online news output, societal interest was comparatively low and that the two do not always correlate. This could be related to the influence of social media which may amplify societal interest without the need for large news media output (Harrington et al. 2018). Whether these peaks of interest or media outputs are related to awareness raising and behavior change is unclear, however peaks in Wikipedia Pageviews corresponding to the latter two types of peaks suggest that some form of awareness raising is occurring.

The peak related to the claim that pangolins were possible intermediate hosts of SARS-CoV-2 may be considered bad publicity for pangolins. Whether it will result in the increased awareness of their plight or a new-found negative public perception is of great importance to their conservation (see: López-Baucells et al. 2018 for an example on bats). Although more costly and difficult, proactive programs creating new peaks of interest, and awareness or behavioral change programs aimed particularly at people in range-states should also be considered (Olmedo et al. 2018; Ploeg et al. 2011; Thomson and Fletcher 2020).

Commercial interest in pangolins through patents

The relationship between patent production and biomedical research trends in pangolins was not obvious in our study, and further superficial analyses did not provide concrete links by treatments or author affiliations (between patent assignee and biomedical researcher affiliation). Although the first pangolin patent was registered in 1993, a year after the first biomedical study testing the effect that pangolin scales have on lactation in mice (Nagasawa et al. 1992), the patent related to pangolin scales was used for cancer treatment (Chen, C: CN1062658-A). It is therefore likely that the 1992 revision of the “Patent Law of the People’s Republic of China” (since 1985), which allowed protection of TCM drugs and increase in grants, may have initiated patent production and biomedical study trends (Qiao and Zhu 2009). If patented treatments and medicines are being used without the guidance of science, they may pose a risk to patients and are a misuse of limited pangolin resources (Jin et al. 2021; Xu and Xia 2019).

Official Traditional Chinese Medicine pharmacopoeia only prescribe pangolin (*Chuan Shan Jia*) scales to promote lactation, improve circulation, reduce swelling, drain pus, and treat skin diseases (Chinese Pharmacopoeia Commission 2015). Nevertheless, we found examples of research and patents in treatment of hepatic fibrosis (Huang 2003; Li, X and Che, Y: CN103948877-A), kidney disease (Xiong et al. 2015; Zhao, Q and Wang, D: CN108245632-A), AIDS (Zhao, S: CN106362127-A) and cancer (Chen, C: CN1062658-A). Although no medicinal properties of pangolin scales have been identified through reliable evidence gathering (Jacobs et al. 2019; Jin et al. 2021), the argument of entrusting in TCM through thousands of years of practice is questioned if medicine and treatments are based on recent patents rather than traditional pharmacopoeia that are the embodiment of historical practices. However, we caution that our results are limited and require more in-depth analyses of these patents (major role-players,

medicines derived from pangolins, and ailments targeted), along with more evidence for the connection between the legal and illicit pangolin supply chains (Horner et al. 2020; Xu et al. 2016). This is necessary if profitability is driving pangolin-derived patented medicine, a possibility supported by the growth of the TCM industry, along with its profit margin and international reach, whilst large private corporations are suggested to be outcompeting the state-based industry (Tommaso et al. 2017; Xu and Xia 2019). Further investigating patented medicinal use may be needed to discredit illegal activities as recent evidence suggests that pangolins are still included in TCM Pharmacopoeia (EIA 2020), contrary to previous announcements (Liu 2020).

Although it is unclear what may have caused the substantial increase in the number of registered patents from 2011 onwards, the dramatic decrease in new patents being registered from 2016 onwards could be linked to the decision to up-list all eight pangolin species to Appendix I at the seventeenth meeting of the CITES Conference of the Parties in 2016. This decision prohibits the international trade of wild-caught pangolins and their derivatives for commercial purposes, and thus may have resulted in the creation of new patents for legal medicine production no longer being financially viable.

Conclusion of a holistic review on pangolins for their conservation

In this study, we provide a curated pangolin research database which could be used as a source of information for researchers and practitioners alike. From our in-depth review of pangolin research dynamics combined with an updated overview of public interest (as explored in: Harrington et al. 2018; Thomson and Fletcher 2020) and an unprecedented assessment of patents, we provide a series of recommendations to fill the knowledge gaps on pangolins for the benefit of their conservation (Table 1).

Although a multitude of conservation issues surround pangolins, the illegal poaching and trade is a major problem. Tackling this issue will require knowledge from a multitude of aspects (popularization, commercialization, and research) as well as conservation actions (Akella and Allan 2012). The latter is an aspect that we did not review directly, however from knowledge garnered from key publications through the review process, we suggest the following conservation actions to tackle the trade issues that are additional to those previously suggested by Challender et al. (2014): (1) horizon scanning analyses of patents for future trade trends (Masters et al. 2020), (2) clinical trials for pangolin-derived medicine through accredited, unbiased protocols (Jin et al. 2021), (3) forensic registers (DNA, isotope, morphology) to identify trade routes, sources, volumes, and legality (Kotze et al. 2020), (4) audits of breeding farms and stockpiles to prevent links between legal and illegal supply, and unethical treatment (Challender et al. 2019b; Horner et al. 2020), (5) multinational collaboration of law enforcement agencies utilizing a database of jurisdictions across the supply chain (see Legal Atlas: <https://www.legal-atlas.com/>), (6) monitoring of financial crime to trace the trade and provide evidence for prosecution (FATF 2020), and (7) pan-stakeholder working groups tasked specifically with generating and compiling knowledge for standard operating procedures of less readily published conservation action fields (rehabilitation and relocation, veterinary procedures, education and awareness raising, tourism, and law enforcement and seizure processing). This broad, multidisciplinary review comes at a time when public and scientific interest in pangolins is at its greatest, and is thus useful in concentrating foreseeable increased research and popularization for more effective conservation measures. Many of the methods employed and insights garnered throughout this study are likely to be applicable to the conservation of other species as well, notably those in the wildlife trade.

Table 1: Global recommendations to fill the knowledge gaps on pangolins for their conservation as identified through the review process on pangolin research, popularization and commercialization.

Issue	Expansion and Recommendation	Conservation aspect
Limited public interest and raising awareness	Given the lack of societal interest and news media outputs on pangolins compared to other iconic species, garnering public interest is a necessity for pangolin conservation. Conservationists should consider previously successful endeavors as a guiding factor, such as ‘shocking’ imagery, documentaries through large broadcasting platforms, and fun, interactive learning such as Google Doodles. Monitoring peaks of interest for reactive awareness raising (Clements 2013), while development of more behavior change and awareness raising programs, with response feedback loops should also be considered (Olmedo et al. 2018). For example, does the COVID-19 peak of interest result in the public knowing more about the plight of pangolins and how to help conserve them?	Popularization
COVID-19’s role in public perception	With the largest spike in public interest of pangolins following the news of their suggested link with COVID-19, gathering of global data on people’s perceptions of pangolins is important (increased awareness of their plight or a new-found ‘vermin-like’ persona). Conservation of pangolins is likely to rely heavily on how they are perceived.	Popularization
Influences of other large traditional medicine industries that use pangolin-derived products	Our results suggest that TCM is the sole contributor of patented pangolin medicines, however it is unclear what the influence of other traditional medicine industries may be on the pangolin trade.	Commercialization
Drivers of pangolin patents	With patents seemingly neither being driven by science nor TCM guideline texts, further insights into the main drivers of patent production is required, so as to prevent future peaks in patent production.	Commercialization
Limited information on patented products	More in-depth patent analyses should be undertaken to identify the major patented medicines/compositions in which pangolins derivatives are used, the major ailments treated, and the effect that these patents have on the demand for both legal and illicit products.	Commercialization

Major knowledge gaps in pangolin literature (natural sciences)	Researchers and funding bodies should target the large natural sciences gaps identified by research category (“veterinary”, “immunology”, and “education”), species (<i>M. culionensis</i> , <i>S. gigantea</i> and <i>P. tetradactyla</i> by publication count), and country-specific (mainly African range-states) levels for more efficient knowledge generation. We suggest using this study as a reference as it provides details of each level, including a level linking species by categories.	Research
Top three research areas by publication count are biased	Anatomical and physiological research mainly focused on <i>P. tricuspis</i> , ecological research on certain species in certain regions, and trade-related research on volumes and networks. Due to the limited amount of research on pangolins, biases are likely to be common. More research on pangolins will likely reduce this effect (as seen in: Supporting Information).	Research
Research areas of conservation importance are missing critical information	There are knowledge gaps in the implications of trade (populations and public health) and means to counter it (forensics, education, rehabilitation, and people’s perception and awareness of pangolins), as well as in population density and distributions, especially through time. Due to limited funding and time, conservation research on pangolins should focus on generating new data for these key areas requiring urgent attention. Meta-analyses on literature that performed checklist and monitoring analyses of conservation areas (see: Supporting Information) may also be an important tool for mining population density and distribution data.	Research
Limited knowledge on health risks surrounding pangolin trade	We suggest screening both natural pangolin populations (baseline estimates) and individuals in the trade (or derivatives thereof) to determine public and pangolin health risks associated with trafficking and releasing seized individuals. This is in light of the recent COVID-19 pandemic.	Research
Africa is trailing in research while lead authors from range-states are becoming less prominent	More conservation and research effort is needed on the African continent, particularly in its central and western states (where the <i>P. tricuspis</i> , <i>P. tetradactyla</i> and <i>S. gigantea</i> occur) since these have become a major source for traffickers in recent years (Heinrich et al. 2017). Additionally, the inclusion of local stakeholders and communities in range-states is important to successful conservation and research.	Research

Acknowledgments

This work and S.P.H are supported by the Agence Nationale de la Recherche (ANR-17-CE02-0001 PANGO-GO). We thank J. Murienne and LA. Poissonnier (Laboratoire Evolution et Diversité Biologique) for providing fruitful comments on the manuscript. We are grateful to three anonymous reviewers and V. Devictor (Editor-in-chief) for improving the first draft of this manuscript.

Chapter 2:

Pangolin genomes offer key insights and resources for the world's most trafficked mammals

*This chapter is available in [bioRxiv](#) and under review in *Molecular Biology and Evolution* (MBE)*

Sean P. Heighton^{1*}, **Rémi Allio**², **Jérôme Murienne**¹, **Jordi Salmons**¹, **Hao Meng**³, **Celine Scornavacca**², **Armanda D.S. Bastos**⁴, **Flobert Njiokou**⁵, **Darren W. Pietersen**⁴, **Marie-Ka Tilak**², **Shu-Jin Luo**^{3,9*}, **Frédéric Delsuc**^{2,9*}, **Philippe Gaubert**^{1,6,9*}.

¹ *Laboratoire Evolution et Diversité Biologique (EDB)— IRD-UPS-CNRS, Université Toulouse III, 118 route de Narbonne, Bât. 4R1, 31062 Toulouse, France.*

² *ISEM, Université de Montpellier, CNRS, IRD, Montpellier, France.*

³ *The State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China.*

⁴ *Mammal Research Institute, Department of Zoology & Entomology, University of Pretoria, Pretoria 0002, South Africa.*

⁴ *Laboratoire de Parasitologie et Ecologie, Université de Yaoundé I, Faculté des Sciences, Cameroon.*

⁶ *CIIMAR/CIMAR, Centro Interdisciplinar de Investigação Marinha e Ambiental, Universidade do Porto, Terminal de Cruzeiros do Porto de Leixões, Av. General Norton de Matos, s/n, 4450-208 Porto, Portugal.*

*Corresponding authors: sean-paul.heighton@univ-tlse3.fr, philippe.gaubert@univ-tlse3.fr, frederic.delsuc@umontpellier.fr, luo.shujin@pku.edu.cn

Abstract

Pangolins form a group of scaly, highly specialized mammals that are being trafficked at record numbers for their meat and supposed medicinal properties. Despite their conservation concern, our knowledge of their evolution is limited by incomplete taxonomic sampling and a paucity of genomic data. We aim to produce exhaustive genomic resources and provide a detailed assessment

of the evolution, demographic history, and genetic diversity among the eight African and Asian pangolin species using 3 238 orthologous genes and whole-genome polymorphisms. Robust orthologous gene-based phylogenies recovered the previously reported monophyly of the three genera of pangolins, but highlighted the existence of an undescribed species closely related to South-East Asian pangolins. Middle Miocene admixture between a ghost, possibly European, lineage and the ancestor of South-East Asian pangolins is suggested to have impacted the early evolution of the group. We detected contrasting demographic trajectories between continental vs. island populations and between biogeographic lineages of the same species, emphasizing prudence when analyzing demographic patterns from single genomes. Genome-wide heterozygosity estimates were variable across species and especially high in the white bellied pangolin, however, we warn that the effect of recent trafficking trends may not yet be realized in pangolin genomes. Finally, we produce the first reference genome for the large African pangolin (*Smutsia gigantea*) with long reads, draft genomes (~43x–77x) for four additional species, and a database of orthologous genes which totals over 3.4 million polymorphic sites. These genomic data will be integral for future conservation management, genetic and forensic endeavors required for conserving pangolins, including tracing their illegal trade.

Keywords: Pangolin, Pholidota, genomics, phylogeny, full genomes, genomic diversity, conservation, PSMC, ancient admixture, ghost lineage, divergence estimates, forensic markers, wildlife trade

Introduction

Genomics is becoming a standard in wildlife research as it provides genome-wide data for more

accurate inferences on species or population delimitation, demographic parameters, diversity, historical trajectories, and the adaptive capacity to global changes (Allendorf et al. 2010). Although transforming this research into conservation practice is yet to be common (Shafer et al. 2015), the gap is closing (Garner et al. 2016).

Pangolins are a group of mammals harboring eight extant species (four in Africa and Asia) under the order Pholidota (Gaubert et al. 2020) that have become a taxon of great public interest and conservation concern in recent years (Heighton and Gaubert 2021; Pietersen and Challender 2020). This is mainly due to them being the most trafficked wild mammals on Earth (Heinrich et al. 2017) and a recent, unclarified, suggestion that they may have been linked to the COVID-19 pandemic (Frutos et al. 2020; Lam et al. 2020; Lee et al. 2020). Despite their dire conservation circumstances, pangolins are considered to be understudied with major gaps in basic species or population research (Heighton and Gaubert 2021; Pietersen and Challender 2020). Even with a long-standing interest in their taxonomy, particularly in light of their convergent evolution with South American anteaters (Xenarthra: Miyamoto and Goodman 1986; Murphy et al. 2001; Wyss et al. 1987), phylogenetic studies focusing on pangolins have been incomplete. They have either been restricted to single markers or mitochondrial genomes, limited by taxon coverage, or clouded by incorrect taxonomic sampling (du Toit et al. 2017; 2014; Gaubert and Antunes 2015; Hassanin et al. 2015; Yu et al. 2011). The most comprehensive phylogenetic study to date contains 15 mitogenomes encompassing all eight species and nine nuclear genes (Gaubert et al. 2018).

As for genome-wide inferences of pangolins, the majority of studies are non-NGS (next generation sequencing) based through traditional methods like karyotyping (Che et al. 2008; Ray-Chaudhuri et al. 1969; Su et al. 1994; Zhihai et al. 2016). In fact, only the two species (*Manis pentadactyla* and *M. javanica*) having published full genomes have started to be the subject of

conservation genomic studies (Choo et al. 2016; Hu et al. 2020a). These have focused on demographic history to determine the consequences of recent population fluctuations due to climatic oscillations (Hu et al. 2020a; Wei et al. 2022). They have also delved into comparisons of diversity, inbreeding, mutational load amongst populations as well as population structuring for geographic assignment of illegally traded individuals. Despite the thoroughness of these studies, our understanding of the evolution of pangolins is still limited to two species, which effects the potential to use genetic markers for conservation and management purposes of the entire order (Allendorf et al. 2010; Kotze et al. 2020).

Producing genome-wide inferences of pangolins is a challenging task. First, their elusive behaviour and tropical distributions render genetic sampling to be time-consuming and costly. Second, the phylogenetic isolation of the group from other orders, the limited fossil records and the deep divergence within pangolins (Gaubert et al. 2018; Gaudin et al. 2009), pose methodological problems. Sister to Pholidota is order Carnivora which is estimated to have started to diverge from each other from around 76 Million years ago (Gaubert et al. 2018; Zhou et al. 2011), thus making it difficult to incorporate outgroup taxa for genomic inferences of divergence and introgression. Compounded by this is the deep divergence (± 37.9 Million years ago) of Asian and African pangolins (Gaubert et al. 2018) which can bias genomic inferences depending on which clade the reference genome is part of for alignment of the group (Bohling 2020; Günther and Nettelblad 2019). Therefore, alternative approaches to the simple variant-based phylogenomic analyses based on the alignment of all taxa to a single reference genome for pangolins need to be considered (Kapli et al. 2020).

The basis of conservation genomics relies upon generating genomic data such as reference genomes and datasets of markers (homologous / orthologous genes, SNPs) which can be used in

downstream species and population inferences (Allendorf et al. 2010). With pangolins in need of conservation action, compounded by limited genomic information on the group, a gap needs to be filled. We thus aim to conduct the first full genome analysis on all eight pangolin species in order to determine evolutionary relationships and demographic trends, identify key genetic parameters for species management, and develop a set of markers for future conservation genetic / genomic efforts. Additionally, we provide a good quality reference genome of the giant pangolin (*Smutsia gigantea*), an elusive fossorial species found in western and central Africa, which we hope will be a key reference for the *Smutsia* genus and the African species alike.

Results and Discussion

We sequenced and assembled the first reference genome for the genus *Smutsia* (giant pangolin; *S. gigantea*, ~87x) using long Nanopore reads and draft Illumina genomes for the black-bellied (*Phataginus tetradactyla*, ~43x), Temminck's (*Smutsia temminckii*, ~44x), Indian (*Manis crassicaudata*, ~53x), and Philippine (*Manis culionensis*, ~77x) pangolins (Table S1). These new genomic data along with previously published reference genomes of the remaining three species, namely the white-bellied (*Phataginus tricuspis*), Sunda (*M. javanica*), and Chinese (*M. pentadactyla*) pangolins (Table S1), provide a complete set of genomes for Pholidota. We also included a recently published genome of a pangolin of uncertain origin, seized from south-western China (Cao et al. 2021). In an effort to reduce ascertainment bias from mapping deeply divergent Asian and African species to a single reference (Albrechtsen et al. 2010; Lachance and Tishkoff 2013), species were mapped to a representative genome from their respective clade (Figure S1). The mapped genomes then underwent haploid and IUPAC ambiguity codes consensus assignment before extracting and aligning 3 238 entire single-copy orthologous genes (introns and exons) for

phylogenomic inferences and divergence time estimation.

Our partitioned concatenated (supermatrix), non-partitioned concatenated, and coalescent (summary tree) phylogenies based on orthologous, whole-gene markers show robust support for the previously reported clear dichotomy between African and Asian pangolins, and the three main clades in agreement with the three distinct pangolin genera (Figures 1 & S2; Gaubert et al. 2018). The seized individual from Sichuan China, either identified as *M. culionensis* in Cao et al. (2021) –based on mitogenomic data– or *M. crassicaudata* on NCBI (GCA_016801295), is sister to the South-East Asian pangolins (*M. javanica* and *M. culionensis*) and does not branch with the Indian pangolin (*M. crassicaudata*). In order to determine whether mitochondrial genes and nuclear genomes show a similar pattern for the *Manis* sp., we obtained the sample’s mitochondrial genome from the short read data and did phylogeny testing. The mitochondrial genomes show the same pattern as the orthologous full nuclear gene phylogeny (Figure S3a), while the *Cytb* (Figure S3b) and *COI* (Figure S3c) gene phylogenies suggest that this individual is forming a clade with two samples seized in Hong Kong. These two samples, with which this individual nests, have been suggested to be a potentially new Asian pangolin species based on *Cytb* and *COI* species delimitation (Hu et al. 2020b). Our results using orthologous whole-gene phylogenies with high bootstrap support and concordance analysis provide additional evidence that this individual likely represents a separate Asian lineage or species. However, with high levels of cryptic diversity across pangolin species (Gaubert et al. 2018; Gaubert et al. 2016), particularly within *M. culionensis* and *M. javanica* (Hu et al. 2020b; Nash et al. 2018; Zhang et al. 2015b), we cannot be certain that this previously undescribed taxon is indeed a new *Manis* species or part of the *M. javanica* species complex (including *M. culionensis*). Additional analyses using geo-referenced wild samples, and genome based species testing, including the use of morphological characters will provide further

insight into the status of this taxon and its distribution. These parameters will be vital to determining conservation priorities and management plans for this potentially new species.

2. Species genomics

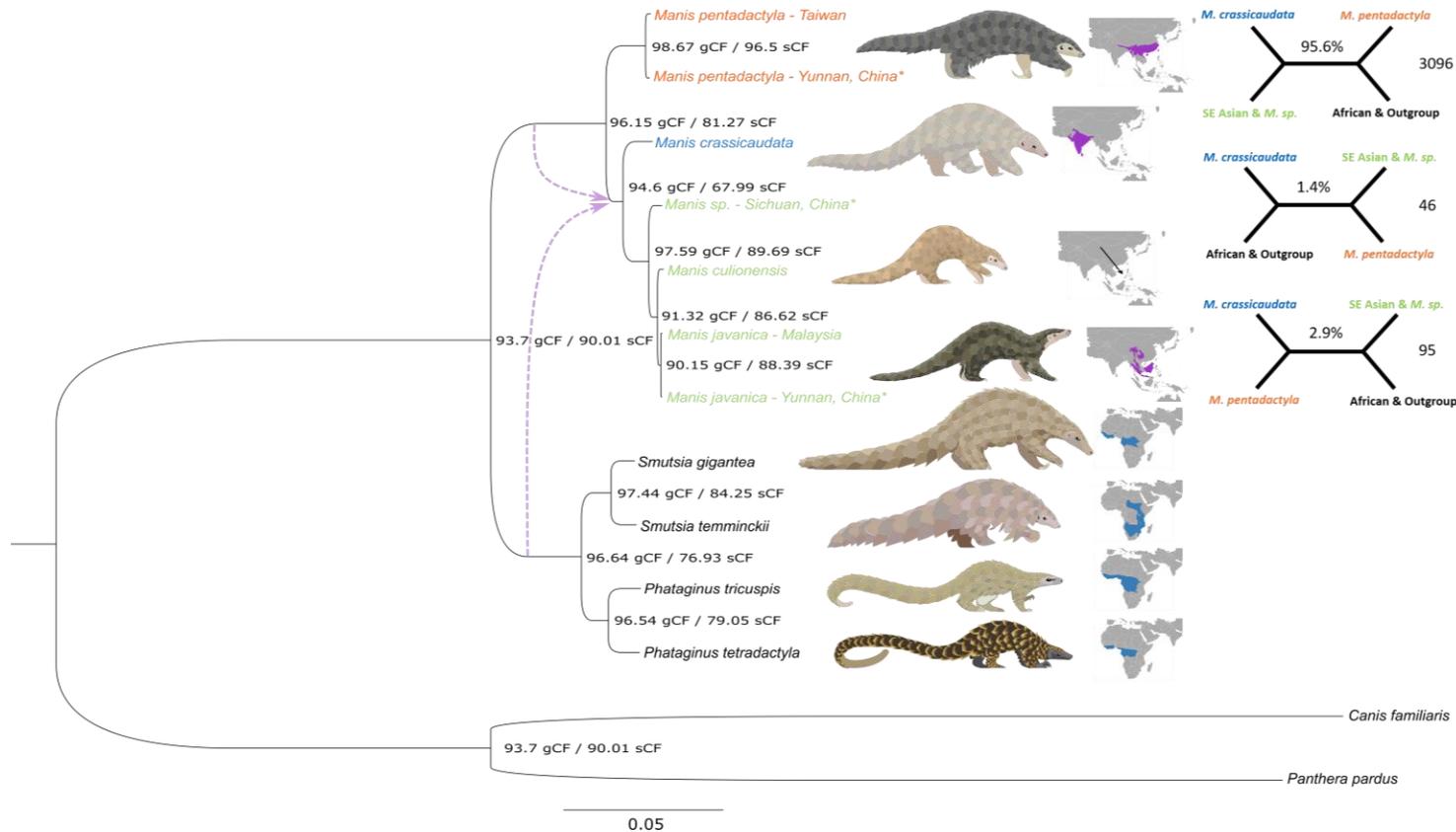


Figure 1: Phylogenomic relationships and reticulation amongst pangolins. Phylogenetic tree based on the alignment(s) of 58 724 014 bp from 2 238 IUPAC consensus whole-gene markers. The pangolin phylogeny consists of 13 individuals from all eight species and is rooted with two representatives of the sister order Carnivora (*Canis familiaris* and *Panthera pardus*). The same topology was derived from a partitioned concatenated (supermatrix), non-partitioned concatenated (Figure S2a) and coalescent (summary tree; Figure S2b) phylogenies, hence only the former is shown. All three phylogenies were derived from an orthologous whole-gene pipeline (Figure S1) and had full branch support for all nodes (100 for the concatenated phylogenies using 1 000 Felsenstein bootstrap replicates and 1 for the coalescent phylogeny using local posterior probability). Nodal numbers represent the concordance factors of genes (gCF) and sites (sCF), which indicate the proportion of genes and sites that fit the current topology. sCF values were based on 100 quartets randomly sampled around each internal branch. Dotted lines (purple) indicate a suggested reticulation event either from African or Asian basal nodes based on maximum pseudolikelihood phylonetworks testing (Figure S4). The main species tree quartet and the two alternative gene tree quartet topologies (from top to bottom on the top right) were identified from the coalescent analysis (Figure S2c). These quartets only implicate the internal branch around the four Asian pangolin species in corresponding colors in the main phylogeny. The value at the center of each quartet refers to the proportion of gene trees following this quartet while the value below the branch is the corresponding number of gene trees. Asterisks (*) indicate confiscated individuals whose origins could not be verified. Pangolin illustrations by Sheila McCabe.

Reticulation and incomplete lineage sorting in South-East Asian pangolins

To assess the level of alternative topologies across the phylogenetic species tree, we used a gene and site concordance analysis, along with the proportion of alternative quartets around each branch of the main topology as obtained from the coalescent species tree analysis (Minh et al. 2020a; Zhang et al. 2018). Overall, there was a low level of discordance and proportion of alternative topologies ($d = 0.981$; Figures 1 & S2c), indicating a global robustness of the pangolin tree topology (Minh et al. 2020a). The clade containing *M. javanica* and *M. culionensis* had the highest values for both gene discordance (90.6 gCF / 88.4 sCF) and alternative quartet topologies (8.81%; Figures 1 & S2c). *Manis javanica* and *M. culionensis* are currently regarded as distinct species based on five discriminant morphological characters (Gaubert and Antunes 2005) and phylogenetic testing (Gaubert et al. 2018), although mean pairwise genetic distances between the two were lower than any other species combinations, including estimates across the six *P. tricuspis* lineages. Thus, our results demonstrate the need for extensive population-level estimates across the distributions of South-East Asian pangolin species, including the undescribed *Manis* sp. taxon, to aid in species conservation planning and accurate post-seizure repatriation efforts of live individuals (Nash et al. 2018). The latter is particularly pertinent given the multiple South-East Asian islands from which lineages of *M. javanica* are sourced and distributed (Nash et al. 2018; Zhang et al. 2015b).

Given the relatively short branch length (as a result of recent divergence) possibly causing this uncertainty in separating *M. culionensis* from *M. javanica* through incomplete lineage sorting (ILS; Blom et al. 2016), a chi-squared test on the gene concordance factors was implemented. The results suggest that ILS is the sole cause of discordance as the alternative topologies are not significantly independent ($p < 0.05$; Table S2) regarding the frequency of gene trees supporting

each topology (Huson et al. 2005; Zheng and Janke 2018). This is however not the case for the branch leading to the *M. crassicaudata*/*M. sp.*/*M. culionensis*/*M. javanica* clade whereby alternative topologies are significantly independent (Figures 1 & Table S2). The clade also is an outlier for the lowest site concordance (94.6 gCF / 67.99 sCF), which points to site discrepancies for this clade. With introgression (ancient gene flow) and hybridization being a possible cause of unbalanced gene-tree discordance of this clade (Doyle 1992), we tested for this using a maximum pseudolikelihood network reticulation analysis across the entire topology (Than et al. 2008). This was conducted on all 3 238 gene trees. The results indicate that no more than one reticulation event is the most likely for pangolins (Figure 1), and that the timing and direction of this event is concurrent with that of the non ILS related bias in alternative topologies determined by the chi-squared analysis. The non-independence caused by the higher alternative topology (bottom quartet; Figure 1) points to introgression/hybridization between a lineage sister-group to the African (donating 31.9% of genes; Figure S4a) or Asian (donating 27.9% of genes; Figure S4b) pangolin clades and the ancestor of South-East Asian pangolins (including *M. sp.*). The difference between the two potential basal lineage donors is due to whether we grouped all individuals from a species as the same species (basal Asian donor), or kept them as separate evolutionary units (basal African donor). Based on our average divergence estimates (Figure 2; see below), this event likely occurred between 7.16–16.84 Ma during the Miocene.

An updated biogeographic scenario for pangolins

Given the uncertainty between the Asian and African basal branch donor and the timing of the reticulation event, we could hypothesize that this basal ghost lineage was likely European. It may have moved towards South-East Asia due to the shrinking of tropical environments in Europe

during middle Miocene climatic cooling (Jiménez-Moreno 2006), as seen with hominoids (Begun et al. 2012). This is plausible as pangolins are suggested to have occurred in Europe within the Miocene (*Necromanis*) (Alba et al. 2018) and into the Pleistocene (*S. olteniensis* ; 2.2–1.9 Ma) (Terhune et al. 2021). However, it may have also been widespread in Laurasia and diversified throughout the region, before having undergone mass extinction across parts of the continental mass. Additional population based analyses and fossils across the region will aid in determining whether these hypotheses hold, and whether this ancient admixture event was adaptive (Figueiró et al. 2017). Recent fossils of pangolins in South Africa (5 Ma, *S. gigantea*), India (Pleistocene, *M. lydekkeri*) and Java (42 000–47 000 ya, *M. paleojavanica*) provide further evidence of dramatic historical distribution changes (Gaudin et al. 2009; Terhune et al. 2021), and at the same time that the potential effect of ghost lineages on our results needs to be considered.

To delve further into the biogeographic scenario of pangolin diversification, the orthologous whole-genes were used to perform Bayesian estimation of divergence times guided by the coalescent tree and fossil calibrations (Table S3) (dos Reis and Yang 2019). At most nodes, our genome-wide results concur with the divergence estimates and resultant biogeographic scenario of diversification previously described with a few markers (Figure 2; Table S4) (Gaubert et al. 2018; Hu et al. 2020b). This is the case of the *Manis* sp. taxon which split from *M. culionensis* and *M. javanica* around the late Miocene (7.16 Ma; 4.73–10.42 Ma), similar to the times suggested by Hu et al. (2020b) using *COI* and *Cytb* markers (6.95 Ma; 4.64–9.85 Ma).

We found that the two South-East Asian pangolin species (*M. javanica* and *M. culionensis*) split during the Upper Miocene to Pliocene (mean = 2.71 Ma; 95% HPD = 1.70–4.21 Ma; Table S4), suggesting that the isolation by sea level rising (800–500 ka) of proto-Philippine pangolins coming from Borneo may have not been the cause of their divergence. The split occurs around the

same time as the *M. pentadactyla* population in Taiwan and the *M. pentadactyla* individual confiscated in Yunnan, China (2.82 Ma; 1.48–4.95 Ma), and overlaps with the split between the *M. javanica* population in Malaysia and the *M. javanica* individual confiscated in Yunnan, China (0.77 Ma; 0.46–1.24 Ma). These two novel, population-based estimates suggest an alternative hypothesis whereby divergences between lineages through relatively ancient population structuring for *M. javanica*, *M. culionensis*, and *M. pentadactyla* occurred before isolation of islands, as seen in some felid species (Luo et al. 2014). This is evidenced by the population divergence of *M. pentadactyla* having occurred even before the final separation of Taiwan from the continental mainland (Kawamura et al. 2016).

Our analysis indicates earlier divergence time estimates for Asian species than previously suggested (Table S4; Gaubert et al. 2018), particularly for the genus *Manis* (split between *M. pentadactyla* and the ancestor of *M. crassicaudata*/*M. sp.*/*M. culionensis*/*M. javanica*) which diverged during the Oligocene to Upper Miocene period (22.62 Ma; 17.07–28.23 Ma as opposed to 12.9 Ma; 10.3–15.6 Ma). The split of *M. pentadactyla* likely suggests a northern (*M. pentadactyla*; China) / southern (other *Manis* species; India, Indochina) Asian clade split, coincident with the Oligocene/Miocene boundary's second uplift of the Himalayan–Tibetan plateau in western China and extrusion of the Indochina block (Deng et al. 2021; Spicer et al. 2020). The biogeographic scenarios of other vertebrate species and current distribution limits of *M. javanica* and *M. pentadactyla* across the Ailao Shan–Red River shear zone (Tapponnier et al. 1990) coincide with this Indochina–China separation (Che et al. 2010; Xiang et al. 2021; Zhang et al. 2006). Overall, the paucity in the fossil record along with the possibility of multiple ghost lineages in pangolins makes it difficult to accurately calibrate and measure their divergence estimates (Gaudin et al. 2009; Gebo and Rasmussen 1985). More fossil evidence is required to

fully comprehend the evolution of pangolins.

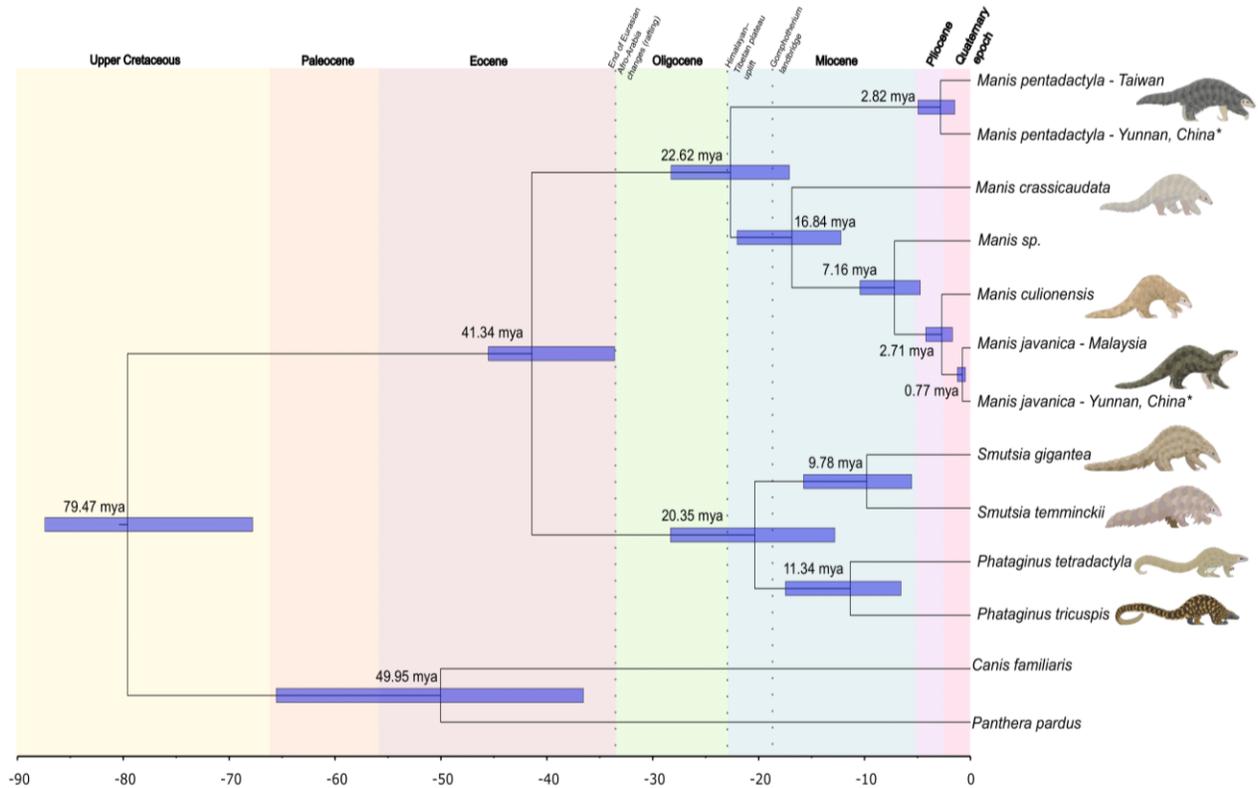


Figure 2: Divergence time estimates along the pangolin phylogenetic tree. Phylogeny with mean posterior divergence times in millions of years and 95% Highest Posterior Density (HPD) intervals (blue bars) for each node (see Table S4 for values). The divergence estimates were based on the unpartitioned, IUPAC concatenated tree dataset of 58 724 014 bp (Figure S3a) using Markov chain Monte Carlo (MCMC) sampling of 20 million generations (2 million generation burn-in) following an auto-correlated, log-normal relaxed clock model. Prior to this, we followed dos Reis and Yang (2019) by first identifying the substitution rate (rgene gamma) for the dataset using baseml and then doing the approximate likelihood calculation by obtaining gradient (g) and Hessian (H) values of branch lengths. We used fossil calibrations on the nodes of Ferae (upper bound from molecular data by Zhou et al. (2011)), Carnivora and Pholidota as priors for the analysis (Table S3). The various epochs and the Quaternary period are delineated by colored partitions with their corresponding labels above the phylogeny (dates taken from <https://stratigraphy.org/timescale/>). Three geological/biological events of importance in our study are also highlighted by dotted lines and labels above the phylogeny. Asterisks (*) indicate confiscated individuals whose origins could not be verified. Pangolin illustrations by Sheila McCabe.

Pangolin demographic history is shaped by lineage-specific biogeography

Inverse instantaneous coalescent rate (IICR) trajectories (Arredondo et al. 2021), which can be related to fluctuations of effective population size and connectivity, were estimated from the whole genome mapping data of each pangolin species using the pairwise sequentially Markovian coalescent (PSMC) model (Li and Durbin 2011). Although all species showed differing trends,

most trajectories started to decline between 1200–600 ka (Figure 3) during the Mid-Pleistocene Transition (Chalk et al. 2017). This period relates to the transition from the less extreme shorter 41 thousand year glacial-interglacial cyclicity to that of the more extreme 100 thousand year cycles (Chalk et al. 2017), which may have been too large of an environmental change for pangolins leading to population declines or structuring as their habitats shifted in size and connectivity (Hewitt 2000; Palkopoulou et al. 2015). For the Asian species particularly and *P. tetradactyla*, the most rapid phase of the declines occurred during a period that saw one of the two most extreme glaciation periods in the last 800 ka, the Marine Isotope Stage (MIS) 16 (676–621 ka; Lang and Wolff 2011).

Interestingly, the island species *M. culionensis* (Palawan Isl., Philippines) and the island population of *M. pentadactyla* (Taiwan) exhibited more recent declines (440–300 ka; Figure 3a) post-dating the last two major sea level lowstands of the past 500 ka (Robles et al. 2015; Rohling et al. 1998). These lowstand periods [440 ka (-139 m) and 340 ka (-122 m – -134 m)] align with fossil and phylogeographic evidence of faunal migration to the current Asian islands during the Middle Pleistocene, including between the strait connecting China to Taiwan (c. 70 m) and the strait connecting Borneo to Palawan (c. 145 m; Ali 2018; Hosoda et al. 2011; Kawamura et al. 2016; Tougard 2001). It is therefore possible that after these lowstands, pangolins on these islands were isolated with a decreasing landmass as sea levels rose. This difference between island vs. continental population fluctuations through time, and the influence of sea-levels for island populations have also been found in South-East island populations of *M. javanica* (Hu et al. 2020a). The absence of such a signature in the two *M. crassicaudata* individuals from the island population of Sri Lanka could likely be due to the strait connecting India to Sri Lanka being only a minimum of 20 m deep, allowing for extensive continental exchanges across the late quaternary

climatic fluctuations (Ali 2018; Bossuyt et al. 2004). This links with the previously suggested non-monophyly of Sri Lankan mitogenomes in the species (Gaubert et al. 2018).

The PSMC curves for the two *P. tricuspis* lineages showed a different trend; with the Western Central African (WCA) lineage indicating a stable IICR through time, while the Dahomey Gap (DG) lineage experienced a progressive decline (Figure 3b). These differences in demographic history are mirrored in their current populations whereby the DG lineage contains low levels of genetic diversity and high levels of inbreeding, while the WCA has the inverse (Figure 4, see below; Aguillon et al. 2020; Zanzo et al. 2022). Our results for two lineages out of six for *P. tricuspis* (Gaubert et al. 2016) indicate that the demographic history of *P. tricuspis* is complex and will require multiple individuals, however we note that the DG lineage sample had limited coverage (Table S1). It also indicates that these lineages that are now separated by strong biogeographic barriers should be treated as separated conservation units. Further deciphering of phylogeographic structure within pangolin species through population stratification methods (PCA and structure) before species wide estimates with the PSMC would be needed. Particularly since the PSMC is limited in its assumption of panmixia (Arredondo et al. 2021; Salmons et al. 2017), and pangolins have recently shown unexpected levels of cryptic diversity (Gaubert et al. 2016; Nash et al. 2018; Zhang et al. 2015b).

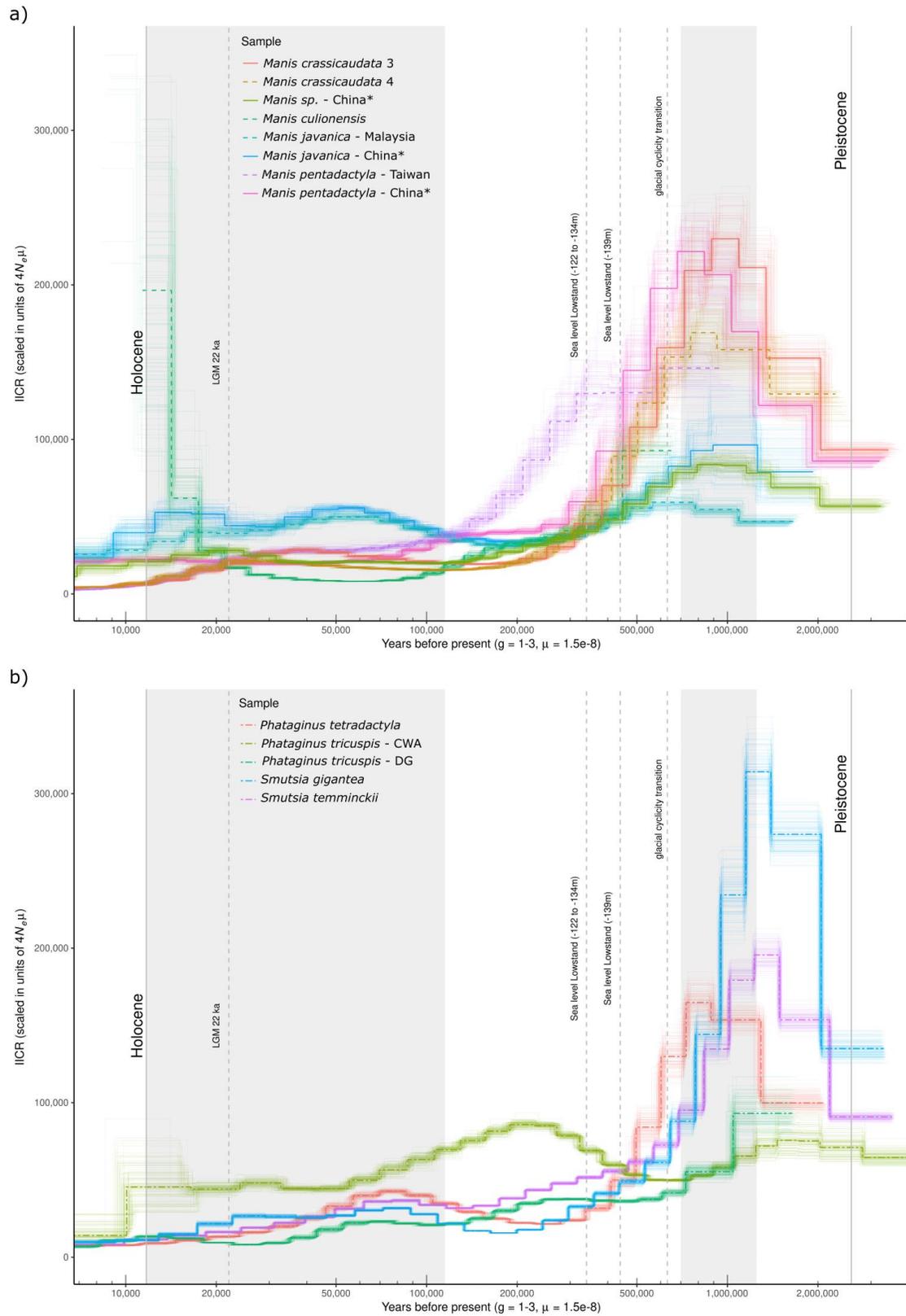


Figure 3: Recent demographic history of pangolin species and populations. Pairwise Sequentially Markovian

Coalescent (PSMC) model on diploid genomes provide demographic trajectories of each pangolin species/population for the (a) Asian and (b) African continental clades. Bootstrap support of 100 iterations for each species/population are indicated by the lighter lines of the corresponding species/population colour. Approximate inverse instantaneous coalescence rate (IICR) values are indicative of effective population size (N_e). Curves, however, may be influenced by changes in population size, connectivity and selection. Curves are scaled by a mutation rate of 1.5×10^{-8} substitutions per site per generation (μ) based on previous pangolin-related estimates (Choo et al. 2016). Generation time in years (g) per species was estimated from available literature (Table S6). Vertical dotted lines with labels indicate important climatic events for this study while the grey shaded areas represent the Last Glacial Maximum (left) and Miocene-Pleistocene Transition (right). This figure was produced through a modified version of the following R script: https://github.com/elhumble/SHO_analysis_2020. Asterisks (*) indicate confiscated individuals whose origins could not be verified.

Delayed loss in genome-wide diversity

We estimated genome-wide heterozygosity from genotype-likelihoods inferred from the site frequency spectrum of full genome mapping data (Korneliussen et al. 2014). We found large variation across pangolin species, up to a 6.65 fold divergence between the lowest and highest estimates (0.074 – 0.492% heterozygosity; Figure 4 & Table S5). The *Manis sp.* individual (0.161%) has an above average heterozygosity estimate for mammalian species (average = 0.156%). This taxon is still however comparable to species of conservation importance like the silvery gibbon (*Hylobates moloch*; Table S5). The lower end of the heterozygosity spectrum is occupied by the *P. tricuspis* DG lineage (0.074%), one *M. crassicaudata* individual from Sri Lanka (MCR4; 0.075%), and *M. pentadactyla* (0.085%) from Taiwan. The latter two populations are from islands, suggesting that isolation from continental populations may have reduced their levels of heterozygosity. Long-term small population size and isolation can result in genetic load due to the accumulation and fixation of deleterious variants (Robinson et al. 2016). The islands containing pangolins are comparatively large and therefore, we believe that international trade, habitat fragmentation and finite population structuring through geographical and anthropogenic barriers will probably play a larger role in decreasing heterozygosity levels of these island populations in the future, than their geographic isolation (Lane-deGraaf et al. 2014). Habitat fragmentation is the likely cause for the low heterozygosity of *P. tricuspis* DG lineage whereby high levels of

inbreeding can be caused by a steady IICR decline over the last 1 million years (Figure 3) as well as recent successive bottlenecks brought about by forest fragmentation through Holocene climatic oscillations (Demenou et al. 2016; Salzmann and Hoelzmann 2005; Zanvo et al. 2022). These fragmentations have likely acted in a similar way to islands, separating and isolating *P. tricuspis* lineages. Due to this climatic history, the Dahomey Gap is currently characterized as a savannah biome intermixed with forest patches, and thus influencing how the semi-arboreal *P. tricuspis* should be managed in this region. Along with *Manis culionensis* (0.492%), the other lineage of *P. tricuspis* (CWA lineage; 0.334%) is an outlier for high levels of heterozygosity (Figure 4), fitting alongside common species like the wild boar (*Sus scrofa*) and brown bear (*Ursus arctos*; Table S5). We nevertheless caution that our results may be influenced by genome quality or coverage (as is the case for *M. culionensis*; Mondol et al. 2013) and that they may not reflect species-wide diversity estimates as they are from only one or two representatives of each species. Given our results, we suggest conservation management plans should consider the genetic variability and demographic history within pangolin species. These heterozygosity estimates are likely not yet impacted by the recent boom in illicit pangolin trade (2016 for African species), the biggest contributor to population declines (Challender et al. 2020a; Mondol et al. 2013). This seems to be the case even for Asian pangolins for which the global trade started a few generations earlier (Challender et al. 2020a).

2. Species genomics

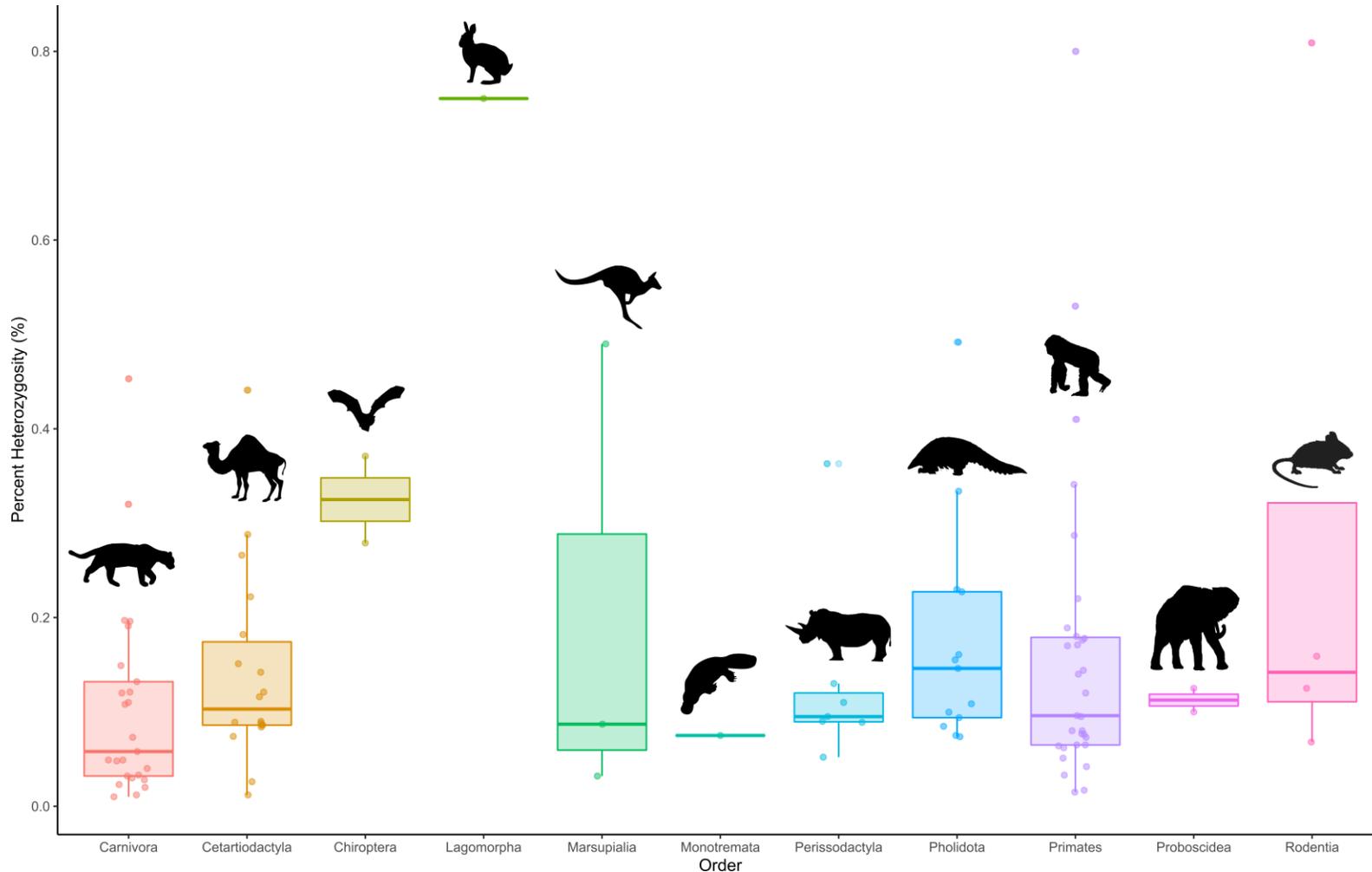


Figure 4: Proportion of genome-wide heterozygosity of pangolins (*Pholidota*) and other mammals. Estimates are grouped and colour-coded by mammalian order. Estimates are updated from the summary by Hu et al. (2020a), and displayed per species from smallest (least diverse) to largest (most diverse) in Table S5. Illustrations indicate the various taxonomic orders and are credited as follows (<https://creativecommons.org/licenses/by/3.0/>): *Pholidota*/*Cetartiodactyla* (Steven Traver), *Carnivora* (Gabriela Palomo-Munoz), *Chiroptera* (Margot Michaud), *Lagomorpha*/*Marsupialia*/*Monotremata* (Sarah Werning), *Perissodactyla* (Oscar Sanisidro), *Primates* (T. Michael Keesey), *Proboscidea* (Margot Michaud), *Rodentia* (Jiro Wada) sourced from PhyloPic (<http://phylopic.org/>).

Providing genomic resources for pangolin conservation

In light of the drastic loss in genomic diversity likely to come (Dufresnes et al. 2018; van der Valk et al. 2019) and no evidence of a reduction in trade (Challender et al. 2020a), the need for genomic data and genetic markers for pangolin conservation is evident. We therefore provide a database of > 3 200 genes ranked by levels of mean pairwise identity (Database S1: <https://doi.org/10.5281/zenodo.7517409>), which could be used to design forensic markers for trade monitoring and prosecution by providing evidence of possible geographic origin of poaching from seized samples (Ewart et al. 2021; Gaubert et al. 2018; Kotze et al. 2020; Zhang et al. 2015b). Most genes have the ability to discriminate between, and likely within, the eight extant pangolin species, thus providing an opportunity to create typing tools using the most discriminant genes for forensic and population genomic inferences. These tools may include SNP panels containing a SNP from each gene or chromosome and mini-barcoding primers targeting highly discriminant regions of genes. After removing genes that we determined to be outliers (610 genes based on the mean and deviation of mean pairwise identity; Figure S5), the database has 2 623 orthologous genes totaling 3 410 610 segregating/polymorphic sites. The top gene (*TMEM38B*) has a mean pairwise identity of $79.51 \pm 13.56\%$ (min = 65.95%; max = 93.07%) and 20.86% parsimony informative sites. The outlier test is a conservative and basic measure of determining the top genes and so we provide the full list of genes for more in-depth forensic marker discovery. We also provide a good quality reference genome for *S. gigantea* and the draft genomes of pangolin species for which virtually no current genomic data exist. These genomic resources and the resultant inferences we made will be key for better conservation management, further conservation genetic research, and combating the illegal trade of pangolins (Allendorf et al. 2010; Kardos 2021; Kotze et al. 2020).

Methods and Materials

Sample acquisition

Pangolin samples that were used for DNA extraction were collected in the field or from museum repositories (see Table S1 for more information). Tissue samples were taken from two deceased *Manis crassicaudata* individuals that were confiscated from the trade in Sri Lanka. Tissue samples from *Phataginus tetradactyla* and *Smutsia gigantea* were collected from the Yaoundé bushmeat market, Cameroon. A spleen sample of *Smutsia temminckii* was collected from an individual that had succumbed to its injuries on an electrified game fence in the Kalahari Oryx Game Farm, South Africa. Finally, we collected a skin sample from *Manis culionensis* at the Field Museum of Natural History, Chicago (FMNH 62919), that had originated from Casuyan, Palawan Island, Philippines. Since the museum sample showed signs of degradation, we took precautions during the mapping and consequent filtering of its short read data as well as interpretation of the results surrounding this species.

Genome sequencing and assembly

Reference genome of the giant pangolin (*Smutsia gigantea*)

High molecular weight DNA suitable for Oxford Nanopore Technologies (ONT) long read sequencing was extracted from an ethanol-preserved muscle sample of a giant pangolin collected in Cameroon (CAM011) using the protocol optimized by Tilak et al. (2020). Long read sequencing was performed with the ONT MinION instrument using libraries prepared with the ONT Ligation Sequencing kit SQK-LSK109 on four flow cells (FLO-MIN-106). This generated a total of 31.7 Gb of raw long read data representing a genome depth of coverage of about 13x. ONT raw signal FAST5 files were then base-called with the high accuracy mode of Guppy v3.2.4

(<https://github.com/nanoporetech>) on a GPU computing server of the Montpellier Bioinformatics Biodiversity platform (<https://mbb.univ-montp2.fr/MBB>) and subsequently cleaned using Porechop v0.2.3 (<https://github.com/rrwick/Porechop>). Complementary Illumina 150PE short reads were produced on an Illumina HiSeq 3000 sequencing system at the GeT-PlaGe sequencing platform at Genotoul (<https://www.genotoul.fr>). After cleaning using Trimmomatic v0.33 (options -phred33 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:50; (Bolger et al. 2014)), we obtained about 185 Gb of short read data representing a genome depth-of-coverage of about 87x (Table S1). A *de novo* hybrid genome assembly was performed using MaSuRCA v3.2.9 (Zimin et al. 2013), which first assembles short reads into “super-reads” before assembling them guided by the long ONT reads. To evaluate genome quality, traditional measures like the BUSCO score (Waterhouse et al. 2018), the number of scaffolds and contig N50, and mean and maximum lengths were computed.

De novo genome annotation was performed for our *S. gigantea* hybrid assembly and the DNA Zoo chromosome-length assembly of *P. tricuspis* following the approach used in Allio et al. (2021). First, repetitive elements were annotated and masked to avoid producing false evidence for gene annotations (Yandell and Ence 2012). This annotation was first performed for each genome independently using RepeatModeler v2.0.2 (Smit and Hubley 2008). To improve the accuracy of these *de novo* annotations, the libraries obtained from the different genomes were cleaned by removing protein-like sequences and were clustered for further analyses. The second step was to identify repeat elements by similarity search against publicly available libraries of mammalian repeats (DFAM (Wheeler et al. 2012)) using RepeatMasker v4.1.2-p1 (Tarailo-Graovac and Chen 2009). The annotations resulting from these two steps were synthesized in a GFF file to be fed into MAKER v3 (Holt and Yandell 2011).

To improve the gene annotation, we relied on transcriptomic information using publicly available RNAseq data for *M. javanica* (Table S1). To do so, raw reads from 25 transcriptomes were downloaded from SRA, cleaned using fastp v0.20.0 (Chen et al. 2018) using default parameters, and assembled with Trinity v2.9.0 (Grabherr et al. 2011). The resulting transcriptome assemblies were annotated with an adapted version of assembly2ORF (<https://github.com/ellefeg/assembly2orf>), which was specifically designed to annotate transcriptomes. This pipeline relies on evidence-based gene predictions to extract and annotate gene CDSs from transcriptome assemblies. The CDSs resulting from the annotation were concatenated and clustered by similarity with CD-HIT v4.8.1 (Li and Godzik 2006) to improve the efficiency of the subsequent MAKER annotation. The CDSs obtained from RNA-seq data were fed into MAKER v3 to help the evidence-based gene prediction. Additionally, the manually annotated, non-redundant protein sequence database Uniprot/SWISSPROT (Bairoch and Apweiler 2000; The UniProt Consortium 2009) was provided to MAKER v3 for the annotation.

To improve the annotation, three runs of MAKER v3 were performed iteratively. In the first run, evidence-based gene predictions using sequence similarities with the CDSs extracted from both the Swiss-prot database and the transcriptomes were computed. Then, two additional runs were performed, with SNAP v2006-07-28 (Korf 2004) and Augustus v3.2.3 ((Stanke et al. 2006); via BUSCO v3, (Waterhouse et al. 2018)) as implemented within MAKER v3 to help create more sound gene models. In doing this, MAKER v3 uses the annotations from the two prediction programs in addition to the evidence-based gene predictions (similarities with reference CDSs) when constructing its models.

Draft genomes for other pangolin species

Paired-end Illumina short read (150 bp) sequencing was conducted on *Manis crassicaudata*, *M. culionensis*, *S. temminckii* and *P. tetradactyla* samples ranging from 44x to 76x sequencing depth (Table S1). SOAPdenovo2 vr240 (Luo et al. 2012) was used to assemble the genomes of all sequences generated in this study. The best kmer length assembly for each species was identified using SeqKit v0.9.3 (Shen et al. 2016) statistics on the scaffolds (N50, mean and maximum length) and the assemblies were gap-closed with GapCloser v1.12 (SOAPdenovo2).

Additional genomic data

Short read data from two previously published genomes of *M. javanica* and *M. pentadactyla* (Choo et al. 2016; Hu et al. 2020a), an uncertain *Manis* species (labeled as *M. crassicaudata* on NCBI but published as *M. culionensis*) (Cao et al. 2021), and one current draft genome of *P. tricuspis* were extracted from the Sequence Read Archive (SRA) on NCBI using the SRA toolkit v2.9.6 (Table S1; Leinonen et al. 2010). We also obtained chromosome-scale genome assemblies of *P. tricuspis* (https://www.dnazoo.org/assemblies/Phataginus_tricuspis; Choo et al. 2016) and *M. javanica* (https://www.dnazoo.org/assemblies/Phataginus_tricuspis) from DNA Zoo (Dudchenko et al. 2017; Dudchenko et al. 2018). The former also included short read data that we used in our analyses (Table S1).

Single-copy orthologous gene dataset

To build a comprehensive phylogenomic dataset for all extant pangolin species, we relied on the OrthoMaM v10 database (Scornavacca et al. 2019), which is composed of 14 509 single-copy orthologous genes for 116 mammal species, including 13 403 CDSs (coding DNA sequences) for

M. javanica. We implemented a pipeline in which we used a dual strategy to reduce the effect on the high level of divergence between pangolin species (particularly between Asian and African clades) if one reference were used. First, we used the *de novo* annotations to conduct single-copy orthologous gene extraction of the DNA Zoo reference genome assemblies for each pangolin clade (*Manis javanica* for Asian clade and *Phataginus tricuspis* for African clade). This provided us with what we termed reference CDSs and their IDs. This was followed by subsequent mapping of Illumina reads of all species on their clade-specific reference genome assemblies and using the IDs from the CDS references to extract the relevant orthologous genes (Figure S1).

CDS extraction from *de novo* reference assemblies

To extract the CDSs specifically corresponding to the single-copy orthologs of the OrthoMaM database, for each orthologous gene alignment, a HMM profile was created via *hmmbuild* of the HMMER toolkit v3.1b2 (Eddy 2011). Then, all HMM profiles were concatenated and summarized using *hmmcompress* to construct a HMM database. Finally, for each CDS newly annotated by MAKER v3, *hmmsearch* was used on the HMM database to retrieve the best hits among the orthologous gene alignments. For each orthologous gene alignment, the most similar sequences for each species were detected via *hmmsearch*. Outputs from *hmmsearch* and *hmmsearch* were discarded if the first hit score was not substantially better than the second ($\text{hit } 2 < 0.9 \text{ hit } 1$). This ensured that our orthology predictions for the newly annotated CDSs were robust.

From this and previous work, we had obtained orthologous CDS information from *Manis* (*M. javanica* from GenBank annotation and CDS extraction from OMM), *Phataginus* (*P. tricuspis* from DNAZoo, annotation and CDS extraction from this study), and *Smutsia* (*S. gigantea*, annotation and CDS extraction from this study).

Genome-wide mapping and consensus

Short read data from all species were cleaned with fastp v0.19.4 with a base pair quality threshold (>20 Phred) and the paired-end base correction function using overlapped reads (-c option). The cleaned reads were then mapped to their respective clade-specific reference genome assemblies (Asian species: *M. javanica* from GenBank; *Phataginus*: *P. tricuspis* from DNA Zoo) using the default settings of the BWA-MEM algorithm of BWA v0.7.15 (Li and Durbin 2009) after testing mismatch, clipping and gap penalties. However, due to the lower quality of the sequencing data of *Manis culionensis*, we applied a more stringent mapping approach (seed = 23, mismatch penalty = 7). We used SAMtools v1.10 (Li et al. 2009) *view* to keep only reads mapped into a proper pair (-f 2) and mapped reads with a high mapping quality (>30 Phred). These filters were tested using SAMtools v1.10 *flagstat* before and after filtering, and using Qualimap 2 v0.7.1 (Okonechnikov et al. 2015) after filtering. Resulting bam files were sorted with SAMtools v1.10 *sort*, duplicates marked with Picard v2.20.7 *MarkDuplicates*, as well as genome-wide mean depth and proportion covered at a depth of >1x and 10x that were calculated using SAMtools v1.10 *depth* and a custom script (Custom script 1). ANGSD v0.933 (Korneliussen et al. 2014) *dofasta* was used to obtain consensus fasta sequences files for each bam file using genotype likelihoods. Both option 3 (most common allele chosen to make haploid consensus) and option 4 (multiple alleles chosen to make IUPAC consensus) were chosen in conjunction with command *docounts* 1. Reads were filtered if they were of bad quality (-remove_bads), not paired (-only_proper_pairs), below a mapping quality of Phred 30 (-minmapq), and had multiple mappings (-uniqueOnly). We also filtered for bases below Phred 20 (-minQ; except *M. culionensis* which had Phred 30) and a minimum sequence depth of 10x (-setMinDepth) and a maximum of 2 times that of the average depth per individual (-setMaxDepth). Outgroup taxa *Panthera pardus* (GCF_001857705.1) (Kim et al.

2016) and *Canis familiaris* (https://www.dnazoo.org/assemblies/Canis_lupus_familiaris_Basenji), which are part of the sister order Carnivora (Liu et al. 2017), underwent the same process with each being mapped to their respective Hi-C reference genomes from DNAZoo (Dudchenko et al. 2017; Dudchenko et al. 2018).

Obtaining orthologous full gene markers using orthologous CDS IDs

Using the orthologous CDS gene IDs from OrthoMaM for each pangolin reference (*M. javanica* and *P. tricuspis*), whole gene annotations containing these IDs were extracted from the annotation files of the references (*M. javanica*, *P. tricuspis* and the two outgroup species; Custom script 2). Annotations with duplicates and those that were not found in both pangolin references were removed. This left us with gff3 annotation files of 5 660 orthologous full genes. BEDtools v2.29.0 (Quinlan and Hall 2010) *getfasta* was used to extract these orthologous full genes from the full genome IUPAC and haploid consensus, after being indexed with SAMtools v1.10 *faidx*.

Phylogenomics and diagnostic markers from genes

Multi-gene fasta files of each species were converted into multi-species fasta files per gene marker using SeqKit v0.9.3 (Shen et al. 2016) *split* (--by-id option) and subsequent concatenation using the *cat* command. These consisted of all eight species (including two representatives each of *M. javanica* and *M. pentadactyla*) and the outgroup taxa from the Order Carnivora. The dataset was then split into two parts, one with haploid consensus genes of only one representative of the eight known species, and the second with IUPAC consensus genes of all aforementioned pangolin individuals and the outgroup taxa. The first was used to determine per-gene diversity estimates across pangolins and the second was used for phylogenetic testing, divergence time estimation,

and identifying reticulation events.

Genetic diversity estimates from whole genes

Obtaining genetic diversity estimates for pangolins from genes requires the use of one representative per species and the removal of outgroup taxa as these would bias the level of diversity per gene. Hence, the two Carnivora outgroup taxa, together with *M. javanica*, *M. pentadactyla*, and *Manis* sp. [from Cao et al. (2021); Choo et al. (2016)] were removed from the gene dataset before undergoing alignment. Sequences were aligned using MAFFT v7.313 (Kato and Standley 2013) --auto option and Transitive Consistency Score (TCS) (Chang et al. 2014) for each alignment were assessed using T-COFFEE v11.00.8 (Notredame et al. 2000) *evaluate* with the `clustalw2_msa` method. Multiple Sequence Alignments (MSAs) with TCS scores under 80 were removed based on the possibility of paralogy, repetitive regions or mis-annotated genes between the two pangolin reference genomes. This resulted in 3 238 gene MSAs of high quality for further analyses. PhyKIT v1.1.3 (Steenwyk et al. 2021) was used to obtain statistics on the pairwise proportional identity of each marker (command `pairwise_identity`) and parsimony informative sites (command `parsimony_informative_sites`). Due to PhyKit calculating alignment gaps as informative sites, all gaps were removed prior to this analysis using TrimAL v1.4.1 (Capella-Gutiérrez et al. 2009) with the option `--nogaps`. DnaSP v6 (Rozas et al. 2017) was used to conduct the “DNA Polymorphism” analysis in batch mode to obtain the rest of the diversity estimates. These outputs were then merged with a custom script (Custom script 3) to obtain a range of diversity statistics per gene. Using the package “robustbase” v0.93 (Maechler et al. 2021) in R v3.6.1 (Rstudio Inc., Massachusetts, U.S.A), we identified possible outlying genes with low levels of mean proportional pairwise identity (or high levels of diversity between species) by filtering

values over two Qn deviations from the median of both the mean (487 outliers) and standard deviations (566 outliers) of pairwise identities (Figure S5) (Rousseeuw and Croux 1993). This was done to prevent inconsistent markers being used in forensic screening approaches due in part to possible biological reasons (paralogy, repetitive elements, inconsistency across taxa, etc.).

Phylogenomic tree building and concordance analyses

From the results on TCS alignments scores above, the same 3 238 cleaned, orthologous genes with all pangolin individuals and outgroup taxa were aligned and trimmed following the aforementioned protocol (MAFFT v7.313 and TrimAL v1.4.1), however the `-gappyout` option was used for trimming. Additionally, these alignments stem from IUPAC consensus and they were again evaluated with T-COFFEE v11.00.8 in order to remove outgroup taxa that had a TCS score of lower than 90 (using SeqKit v0.9.3 *grep*). Alignment statistics of each MSA file per gene were obtained and were then concatenated into a single MSA using AMAS v0.98 *summary* and *concat* (Borowiec 2016) with `--part-format` set to RAXML in order to provide a partitioning file by gene. ModelTest-NG v0.1.5 (Darriba et al. 2019) was used to identify the best model of sequence evolution for each gene-partition (input partition file from AMAS) as well as the entire alignment (no partition). We set the `-h` option to `uigf` (tests rate heterogeneity: Uniform, +I, +G, and +I & +G) and `-T` option (model test template) to `phyml` (all 11 models of evolution) in order to test 88 DNA models. Partitioned and non-partitioned concatenated IUPAC phylogenies (supermatrix) with 1 000 Felsenstein bootstrap replicates were inferred from the concatenated MSAs using RAXML-NG v0.9.0 (Kozlov et al. 2019) following the best-fitted DNA model for each partition under the BIC test. A multiple species coalescent summary tree was also inferred from the MSAs by first using RAXML-NG v0.9.0 to build a gene tree per gene whilst accounting for the best-

fitting DNA model (BIC) for each gene (extracted from the aforementioned ModelTest-NG analysis), followed by summarizing these gene trees with ASTRAL-III v5.6.3 (Zhang et al. 2018). A polytomy test (-t10) using these data was conducted in order to test whether each branch is a potential polytomy (Sayyari and Mirarab 2018), but none were suggested as a polytomy. Finally, gene (--gcf) and site (--scf with 100 quartets randomly sampled around each internal branch) concordance factor analyses based on the aforementioned gene trees and ASTRAL-III output (coalescent species tree) were performed with IQ-TREE 2 v2.0.6 (Minh et al. 2020b). This provides a full analysis of the raw data in terms of genes or sites that may be disagreeing with the species tree phylogeny. This was followed by a Chi-squared test of independence in R v3.6.1 (Rstudio Inc., Massachusetts, U.S.A) based on a script designed by Robert Lanfear (http://www.robertlanfear.com/blog/files/concordance_factors.html). This tests the frequencies of two alternative quartet topologies of each internal branch whereby significance indicates the possibility of something other than incomplete lineage sorting causing gene tree discordance. We did not use this test on site concordance frequencies, as this test does not take linkage disequilibrium between sites into account, which can be assumed to play a large role for multiple sites in a single gene.

Reticulate evolutionary relationships

We used PhyloNet v3.8.2 (Than et al. 2008) to run two separate analyses in order to determine possible reticulation events through maximum pseudolikelihood estimates. The first was to consider each individual as a separate species and the second was to merge multiple individuals of the same species as one. Both analyses were conducted by inputting the RAxML gene trees whilst the best number of reticulation events (0-4) were determined by comparing the change in

maximum pseudolikelihood scores between them. PhyloPlot in Julia was used to visualize the networks.

Divergence time estimates

We used MCMCTree in the PAML-4.9h package (Yang 2007) to estimate divergence times on the non-partitioned RAxML-NG (supermatrix) phylogeny inferred from the IUPAC consensus gene dataset. We used Ferae (root), Pholidota and Carnivora fossils as upper and lower bound calibrations (Table S3). We followed dos Reis and Yang (2019) by first identifying the substitution rate (rgene gamma) for the IUPAC gene dataset using baseml and then doing the approximate likelihood calculation by obtaining gradient (g) and Hessian (H) values for branch lengths. The MCMC sampling from the posterior distribution of times and rates using MCMCTree was run for 20 million generations with a burn-in of 2 million generations and was run twice to test for convergence in Tracer from BEAST 2 (Bouckaert et al. 2014). An auto-correlated, log-normal relaxed clock model was implemented. Finally, we ran the analysis by sampling from the prior of times and rates (usedata = 0) to test the soundness of the prior and whether our calibrations may influence the results.

Mitochondrial genes and genomes

To obtain mitochondrial genomes of each species in this study as well as those from the undetermined *Manis sp.* (Cao et al. 2021) and the two unpublished *P. tricuspis* reference genomes, 10 million reads were extracted from fastq files and mapped to a published mitochondrial genome reference of each species using the default options of Geneious mapper in Geneious v9.1.8 (Kearse et al. 2012). These were then consensus called and aligned to the mitochondrial genome dataset

from Gaubert et al. (2018) with MUSCLE (10 iterations) in Geneious. The alignment underwent Neighbour-Joining (NJ) phylogeny testing in MEGA X (Kumar et al. 2018) with Kimura 2-parameter model of evolution and 1000 bootstrap replicates. We also aligned the *COI* and *Cytb* genes from two samples of the potential new Asian species suggested by Hu et al. (2020b) to the mitochondrial genome dataset and extracted these regions for an additional NJ phylogeny of these two genes.

Genome-wide analyses

The aforementioned cleaned short read data for each pangolin species were mapped to their closest reference per genus (*Phataginus*: *P. tricuspis* DNA Zoo; *Manis*: *M. javanica* DNA Zoo; *Smutsia*: *S. gigantea* this study) using BWA-MEM from BWA v0.7.15. Resulting bam files were filtered using SAMtools v1.10 *view* (-f 2; >30 Phred), sorted with SAMtools v1.10 *sort*, had duplicates marked with Picard v2.20.7 *MarkDuplicates*, as well as had genome-wide mean depth and proportion covered at a depth of >1x and 10x that were calculated using SAMtools v1.10 *depth* and a custom script (Custom script 1).

Comparative genome-wide heterozygosity

ANGSD v0.933 (Korneliussen et al. 2014) was used to calculate genome-wide heterozygosity per sample using a script adapted from de Jager et al. (2021) following the ANGSD workflow (<http://www.popgen.dk/angsd/index.php/Heterozygosity>). To do this, the command *doSaf* was implemented on the alignment bam file of each individual from genome-wide mapping to determine its folded site allele frequency likelihood. We used the genus-specific reference as the ancestral state and the SAMtools method for genotype likelihoods (-GL 1) as it considers possible

sequencing errors and performs necessary corrections. The option `-C 50` was used to adjust mapping quality for excessive mismatches, and reads were removed if they were of bad quality (`-remove_bads`), not paired (`-only_proper_pairs`), below a mapping quality of Phred 30 (`-minmapq`), and had multiple mappings (`-uniqueOnly`). Bases below Phred 20 were also removed (`-minQ`), with the exception of *M. culionensis* which had Phred 30. After *doSaf*, we determined the site frequency spectrum with the ANGSD sub-program *realSFS*, and finally, the genome-wide heterozygosity by dividing the number of heterozygous sites by the total number of sites per genome. We updated the list of genome-wide heterozygosity estimates for mammalian species developed by (Hu et al. 2020a), many of which are of conservation importance, by including our estimates as well as additional examples we found since the conception of the table.

Demographic history reconstruction with PSMC

Each alignment file (bam) from the whole-genome pipeline output was converted to a diploid consensus fastq file by first determining genotype likelihoods using BCFtools v1.8 *mpileup* (minimum base and mapping quality of Phred 30), calling the genotypes with BCFtools v1.8 *call* (`-c` option for consensus caller), and then filtering and converting the vcf to a fastq file using the *vcftutils.pl vcf2fq* script (SAMtools v1.10). Filtering with *vcftutils.pl* included a minimum mapping quality of Phred 30 (`-Q`), a minimum coverage threshold of 10x (`-d`) and maximum coverage threshold of double the mean genome-wide coverage of each individual (`-D`) as determined from the whole genome pipeline. For all genomes, the mean genome-wide depth was higher than the recommended PSMC cutoff of $\geq 18x$ and the percentage of missing data per genome at $\geq 10x$ depth was lower than the recommended PSMC cutoff of 25% (Nadachowska-Brzyska et al. 2016). This was the case except for the one of two *M. crassicaudata* (MCras 4; depth = 9x and

37% of reference covered at $\geq 10x$) and *P. tricuspis* (From DNA Zoo; depth = 10.8x and 45% of reference covered at $\geq 10x$) genomes. These were only used for PSMC and heterozygosity estimates and should be taken with caution.

Using the program PSMC v0.6.5 (Li and Durbin 2011), the fastq consensus files with diploid variant information were then converted to fasta-like files (psmcfa) containing information of whether there is at least one homozygote in a bin of 100 bp. From there, we ran the PSMC analysis following the default settings described by Li and Durbin (2011) for human populations (-N 25, -t 15, -r 5, -p "4+25*2+4+6") and used on other mammalian species (Chen et al. 2019; de Jager et al. 2021; Westbury et al. 2018). By randomly sampling subsections of the psmcfa file, 100 bootstrap replicate analyses were performed in order to estimate the variance in the approximate inverse instantaneous coalescence rate (IICR). This is the inverse of the rate at which coalescence events take place through time, which was originally used as a function of effective population size in a panmitic model (Li and Durbin 2011). However, inferences must take into account the potential of a non-panmitic model of demographic history and the confounding effects of natural selection and population structure, which can also affect coalescence (Arredondo et al. 2021; Johri et al. 2021). A mutation rate of 1.5×10^{-8} years/site was implemented (Choo et al. 2016), while a generation time in years was chosen for each species depending on available literature (Table S6). The PSMC figures combining the species from each continental clade were constructed in R v5.1 (Rstudio Inc., Massachusetts, U.S.A) following de Jager et al. (2021) by editing a script (found at: https://github.com/elhumble/SHO_analysis_2020) that utilizes the plotPsmcR function (Liu and Hansen 2017).

Data accessibility

Draft genomes (*Manis culionensis*, *M. crassicaudata*, *Phataginus tetradactyla*, *Smutsia temminckii*) and a reference genome with associated metadata (*S. gigantea*) have been deposited in Genbank and are publicly available as of the date of publication. The associated sequence read data have also been deposited in Genbank (SRA) for the aforementioned genomes (except for *S. temminckii*). This paper also analyzes existing, publicly available genomic data (*M. pentadactyla*, *M. javanica*, *P. tricuspis*). The accession numbers or links for the aforementioned genomic data are listed in Table S1.

A database containing the list orthologous, whole-genes ranked by diversity amongst all eight pangolin species has been deposited at Zenodo and is publicly available (Database S1: <https://doi.org/10.5281/zenodo.7517409>). All original code in the form of custom scripts for processing the genomics data in this study have been deposited at Zenodo and are publicly available (<https://doi.org/10.5281/zenodo.7517409>).

Acknowledgments

We thank the team at PANGO-GO (<https://anr.fr/Project-ANR-17-CE02-0001>) for the guidance, provision of material and financial resources to conduct this research. Our gratitude to the following institutions and staff that provided the necessary equipment, resources and guidance: B2M staff at EDB, Get-PlaGe, Genotoul, Tshwane University of Technology, the University of Pretoria, and Kalahari Oryx Private Game Farm. We thank Jayanthi Alahakoon for providing samples from Sri Lanka, and Yan Zhuang, Xin Sun, Ke Liu, and Yeyizhou Fu for technical assistance. Our gratitude to Jeremy Johson and the BROAD Institute for providing us with early access to the *Phataginus tricuspis* raw genomic data (GCA_004765945.1). We thank the DNA

Zoo (<http://www.dnazoo.org>) for making their pangolin genomic assemblies publically available. Many thanks to Sheila McCabe for creating and providing illustrations of each pangolin species. This work has been supported by grants from the Agence Nationale de la Recherche (PANGO-GO project: ANR-17-CE02-0001), European Research Council (ConvergeAnt project: ERC-2015-CoG-683257), Mohamed bin Zayed Species Conservation Fund (project: 0925713), the National Research Foundation of South Africa (grant: 71454) and the National Natural Science Foundation of China (NSFC 32070598).

Author Contributions

P.G., and S.P.H. conceived the study. S.-J.L., F.N., P.G. and D.W.P. collected the samples. P.G., H.M, S.P.H., and M.-K.T., fulfilled laboratory procedures. P.G., S.-J.L., A.D.S.B. and F.D. undertook and funded the genome sequencing. R.A. assembled the *S. gigantea* genome. R.A. and C.S. undertook the annotation and the extraction of CDSs for the three reference genomes. S.P.H. and R.A. undertook draft genome assemblies. S.P.H. performed mapping and variant calling, creation of consensus genes and their alignments, and all the genomic analyses of the study. R.A., M.-K.T. and F.D. wrote the methods section for the *S. gigantea* assembly and annotation. S.P.H. wrote the rest of the manuscript and its supplementary information as well as created the figures and tables, with input from all authors.

Chapter 3:

Population genomics of pangolins: defining units for conservation management and tracing the trade

This chapter will be prepared for submission to Nature Ecology & Evolution after the inclusion of 3D morphometric data.

Abstract

Defining and characterizing distinct populations through their genetics is an imperative step toward informed conservation management of species. Defined populations can also form a reference for the geographic assignment of individuals from seizures to their origins, in-turn tracing the illegal trade. Population genetics studies are limited with regards to pangolins, thus preventing essential insight for conservation management of the most heavily trafficked mammals. For these population-based management plans to be effective, DNA from geo-referenced samples are required from across the species range as it allows for accurate distinction of populations. However, pangolins are elusive animals, thus obtaining exhaustive sample sets from the field remains a difficult task for geneticists. We therefore have designed and implemented a gene-capture approach to target genes of likely high discriminative value (high evolutionary rates and locally adaptive) capable to cope with various sample types –including museum samples– from across all eight pangolin species. Due to unprecedented coverage across species ranges, we provide valuable population designations of each species and their population genetic characteristics (diversity and inbreeding). Major drivers of structure for pangolins were likely linked to large water bodies in unison with Pleistocene climatic changes that altered forest and savannah refugia across their distribution. Based on a range of evidence from analyses targeting population stratification, migrations, population differentiation and maximum likelihood phylogenies, we propose re-assessed taxonomic boundaries –including potentially novel species and taxa– that require specific attention. Additionally, we test the efficacy of the method both in terms of capture success with highly divergent and degraded samples, but also in geographic traceability with a few examples. This provides a baseline from which further research into adaptive capacity of these populations and the geographic traceability of pangolin seizures can be thoroughly tested to aid

authorities and practitioners in mitigating future population declines.

Keywords: conservation genetics, population genomics, pangolin, geographic traceability, targeted sequencing, museum, admixture, structure, inbreeding, biogeography.

Introduction

A central interest in population genetics/genomics is its applicability for identifying and characterizing distinct populations or units for conservation (Allendorf 2017; Hohenlohe et al. 2021). These studies provide conservation practitioners with insights on how to best split a species into management units, the size of these units, their demographic trajectories, how and where they are connected to each other, and how they will adapt to future conditions. Insights like these allow for more informed management policies, preventing cases of inbreeding depressions, hybridization, accumulation of deleterious variants, and loss of genetic diversity and adaptive capacity through management practices, or lack thereof, like genetic rescue through translocation and captive breeding (Allendorf and Hard 2009; Musmann et al. 2017; Willoughby and Christie 2019).

Population genomics additionally provides the means for a newly emerging tool named geographic traceability, which has the potential to drastically change how we combat wildlife crime (Ogden and Linacre 2015). The method is based on the ability to assign a sample of unknown geographic location to its genetic population of origin and has been applied to ivory trafficking (Wasser et al. 2015), fishery fraud (Martinsohn and Ogden 2009), and timber trade (Ng et al. 2022). Although neutral loci are often considered key to population genetics due to their assumptions associated with neutrality, there has been a recent drive for the use of fast evolving and adaptive

loci in both conservation genetics and geographic traceability (Funk et al. 2012; Ogden and Linacre 2015). The reason for this is that they can allow for more finite traceability by teasing apart populations under local environmental selection (Ogden and Linacre 2015), but that they can also provide vital information on adaptive variation and connectivity, which may alter management plans based solely on neutral loci (Funk et al. 2012; Hohenlohe et al. 2021). One of the biggest restrictions to geographic traceability, and population genomics as a whole, is the requirement of good quality DNA from samples that have been geographically referenced with a certain degree of confidence, especially those meant for forensic traceability (Hohenlohe et al. 2021; Ogden and Linacre 2015). Acquiring this as well as a sufficient amount of ethically sourced samples from the field, especially from elusive or endangered species, is often resource intensive and time consuming.

Museums can help counter this restriction as they provide a sufficient amount of samples from a geographically wide range, without these logistical and cost related burdens (Nakahama 2021). They can additionally provide added insight into the relationships of species or populations already extinct, causes of temporal transitions of genetic diversity and selection, and the construction of conservation units with the inclusion of samples of endangered species prior to bottlenecks or genetic drift (Nakahama 2021). DNA degradation and high levels of contamination (exogenous DNA) of these archival specimens can however make it difficult and expensive to obtain the required material for accurate insights, even with next generation sequencing (NGS) technologies (Andrews et al. 2021). Samples from bushmeat markets and seizures often share similar characteristics of degradation and contamination, especially from products that underwent processing like grinding into powders, cooking, smoking or fixing (Eaton et al. 2010). Gene-capture provides a solution to this through the design of small RNA probes that can enrich small

fragments of degraded, endogenous DNA, thus increasing the relative DNA yield of the target species for NGS (Derkarabetian et al. 2019; White et al. 2019). When undertaking a capture experiment, the level of degradation and contamination to be encountered as well as the divergence between the reference used to design probes and the target species needs to be accounted for (Cruz-Dávalos et al. 2017; Paijmans et al. 2016; Schott et al. 2017).

Pangolins are myrmecophagous mammals that form a highly divergent but monophyletic order called Pholidota. The order is composed of eight extant species, but possibly more (see: Heighton et al. unpublished – Chapter 2), that diverged into Asian and African continental clades from around 40mya (Gaubert et al. 2018; Gaubert et al. 2020). They are particularly difficult to study due to their typically low population densities, usually nocturnal activity patterns and elusive behavior (Challender and Waterman 2017; Willcox et al. 2019). These factors have contributed to the pangolin being relatively unheard-of before COVID-19 and heavily understudied in many facets of its biology, particularly those that require field samples (Heighton and Gaubert 2021). In a systematic review on pangolin literature between 1865 and 2019, only nine publications targeting three species were linked to the population genetics and phylogeography of pangolins (Heighton and Gaubert 2021). The most comprehensive in terms of geographic range and sample number was on the white-bellied pangolin (*Phataginus tricuspis*), whereby the authors implemented a multi-locus approach to identify six cryptic lineages (Gaubert et al. 2016). Since then, these lineages have been used as a reference for geographic traceability of large seizures (Ewart et al. 2021; Zhang et al. 2020b), and an in-depth population genetic characterization and localised test of traceability for one of the lineages (Zanvo et al. 2022). Being based on multi-locus and microsatellites markers, these studies lacked the resolution that can be achieved with large amounts of single nucleotide polymorphisms (SNPs) associated with genomic data (population genomics).

Population genomics studies thus far have been on the Sunda (*Manis javanica*) and Chinese (*Manis pentadactyla*) pangolins, in which they provided much needed estimates of inter-population comparisons of diversity, inbreeding, demography, and mutational load as well as population structuring for geographic traceability (Hu et al. 2020a; Nash et al. 2018; Wang et al. 2022; Wei et al. 2022). However, due to the lack of reference samples of known localities and the limited area of the species ranges in which these studies occurred, assignment was limited. The lack of a reference database can not only hinder any geographic assignment of seized samples but also reduce the usefulness of management plans without a geographic context to these populations and their genetic characteristics, as was the case for a study solely on seizures of *M. javanica* (Zhang et al. 2015b).

These sampling gaps and the deficiency in research on the population structuring and characterization of all species of pangolin has been highlighted as a concern for the conservation of the group (Heighton and Gaubert 2021; Kotze et al. 2020). This is especially true for pangolins with levels of trafficking trumping that of any wild mammal (Heinrich et al. 2017) and recent evidence of population declines (Nash et al. 2016; Zanzo et al. 2020). Using both museum and freshly collected samples across an unprecedented geographical range for all eight species, we aim to identify population units, characterize them through baseline population genetic statistics, and identify the biogeographical mechanisms that resulted in their formation and characteristics. To achieve this, we have designed and implemented a gene-capture approach that takes into account the deep divergences of Pholidota, DNA degradation, contamination, and the effect of neutral and adaptive loci. Although not the main aim of this study, we also provide examples of its utility for geographic traceability. This will provide a baseline on which future management plans and enforcement through forensics of pangolins can be based (Allendorf et al. 2010; Hohenlohe et al.

2021; Ogden and Linacre 2015).

Methods

Bait design for gene-capture

The aim of the bait design procedure was to generate RNA probes capable of working on all eight pangolin species from highly contaminated and degraded DNA. It also needed to target genomic regions that would provide high resolution structuring and answer both neutral and adaptive evolution based questions. To do this, a bioinformatics workflow was designed and implemented in order to produce a short-list of loci accounting for relative evolutionary rates (RER; see below) and biological relevance (BR; see below) using Geneious 9.1.8 (<https://www.geneious.com>; Figure 1). This resulted in a custom myBaits® kit of 673 loci with 40K probes at 3x tiling (every 27nt) and 70mer baits (*ca.* 1.2mb). The kit was designed off a reference species from each continent with a 3:1 ratio of the theoretically more neutral RER to more adaptive BR loci. This meant that the probes would be able to target species from both continental clades (the two references), be useful for highly degraded museum samples (70mer baits), have limited gaps in coverage across the loci of interest (3x tiling), and be useful for both neutral and adaptive based questions (RER and BR) concerning pangolins.

Loci forming the RER data were extracted from *M. javanica* in the OrthoMaM v10 database (http://www.orthomam.univ-montp2.fr/orthomam_v10/) of one-to-one orthologous genes encompassing 116 mammalian taxa (Scornavacca et al. 2019), and were separated into three different categories tagged as: relative evolutionary rates of 2.5–3, 3.1–4 and 4.1–5 (maximum for v10 database). These RER estimates are based on a Super Distance Matrix procedure on highest-likelihood trees in OrthoMaM pipeline whereby higher figures relate to faster evolutionary rates

(Ranwez et al. 2007). Genes of biological relevance (BR; n = 112) were selected through a literature review of other animal models that exhibited adaptive signatures which could be associated to pangolin specific phenotypic or geoclimatic variation amongst species and populations. These themes were tagged on gene names and included climatic variability (temperature, moisture and altitude), external appearance [mainly scale, skin and hair colour but also keratin genes related to scale formation; see Choo et al. (2016)], external reception (vision and taste), size and tail length, digestive pathway (chitinases and stomach acids), immunity, skull and rostral morphology [based on variation found in pangolins by Ferreira-Cardoso et al. (2019)], digging, circadian rhythm, behaviour, and pseudogenes.

These loci were blasted against the *M. javanica* RefSeq genome assembly (145x coverage - GCF_001685135.1; Choo et al. 2016) and the *P. tricuspis* genome assembly (30x coverage - GCA_004765945.1; unpublished) using Geneious 9.1.8 in order to extract 100bp flanking regions on either side of each locus (intronic and intergenic regions). It also accounted for missing data found in OrthoMaM's database due to the internal filtering step using trimAl (Capella-Gutiérrez et al. 2009) to remove false sequences and/ or codons, post alignment (Scornavacca et al. 2019) as these data were extracted from the exonic database. Initial filtering of loci was based on a size selection criterion whereby loci outside the 400bp – 2 500bp range were removed. Loci that were too small would not fit library preparation criteria and those that were too large would reduce the number of loci we could target (take up more probes). Genomic data from *M. pentadactyla* (GCA_000738955.1; Choo et al. 2016), *P. tetradactyla* (this study) and *S. gigantea* (this study) were mapped to the reference loci using Geneious 9.1.8 with default parameters. Polymorphisms were called using *bcftools mpileup* (parameter *-d/--max-depth* increased to 100 million) piped into *bcftools call (-c/--consensus-caller)* commands in BCFtools v1.8 (Danecek et al. 2021). This was

only done for mapped data on the *M. javanica* reference loci owing to a high coverage of the reference genome, and thus more accurate polymorphism estimation. Pairwise divergence values were calculated based on the number of polymorphisms per base pair length of loci, thus accounting for the relationship between locus size and number of polymorphisms. In the second filtering step, loci were removed based on too high coverage (>2x coverage; tandem or interspersed repeats) as well as pairwise divergence of lower than 1% between *M. javanica* reference loci and the three genera mapped to them (likely un-informative). Some loci were re-added if they were of high biological importance (two or more studies observing relevance), had above the average variance, and under 3 000bp.

Due to slight discrepancies in locus size between *P. tricuspis* and *M. javanica* reference loci, owing mainly to the low coverage of the *P. tricuspis* reference genome, these references were compared in order to rectify any missing loci deleted or added in *P. tricuspis* reference loci (these loci were double-checked against the coverage and variant criteria). One locus from the original *P. tricuspis* reference loci dataset was added, another was taken from the consensus of the mapped *P. tetradactyla* genomic data to *M. javanica* reference loci whilst the rest were flagged, but kept in the final *M. javanica* reference loci list (n= 8). Lastly, genomic data of *S. temminckii*, *M. culionensis* and *M. crassicaudata* were mapped against the *M. javanica* and *P. tricuspis* reference loci for any additional loci to be flagged based on upper coverage (around 2x or more) due to tandem and interspersed repeats.

To reach the required 1.2mb of a 40K myBaits® custom kit, we added an additional 49 loci that fitted the filtering requirements, but were outside the initial sizing requirements (between 350-400bp or between 2000-5000bp). This list was sent to myBaits® for final probe design whereby baits for loci, or parts therefore, that were within a 2% divergence between the African

(*P. tricuspis*) and Asian (*M. javanica*) continental clades were merged into a single bait. Baits that matched a part of the genome that had >25% repetitive elements were removed.

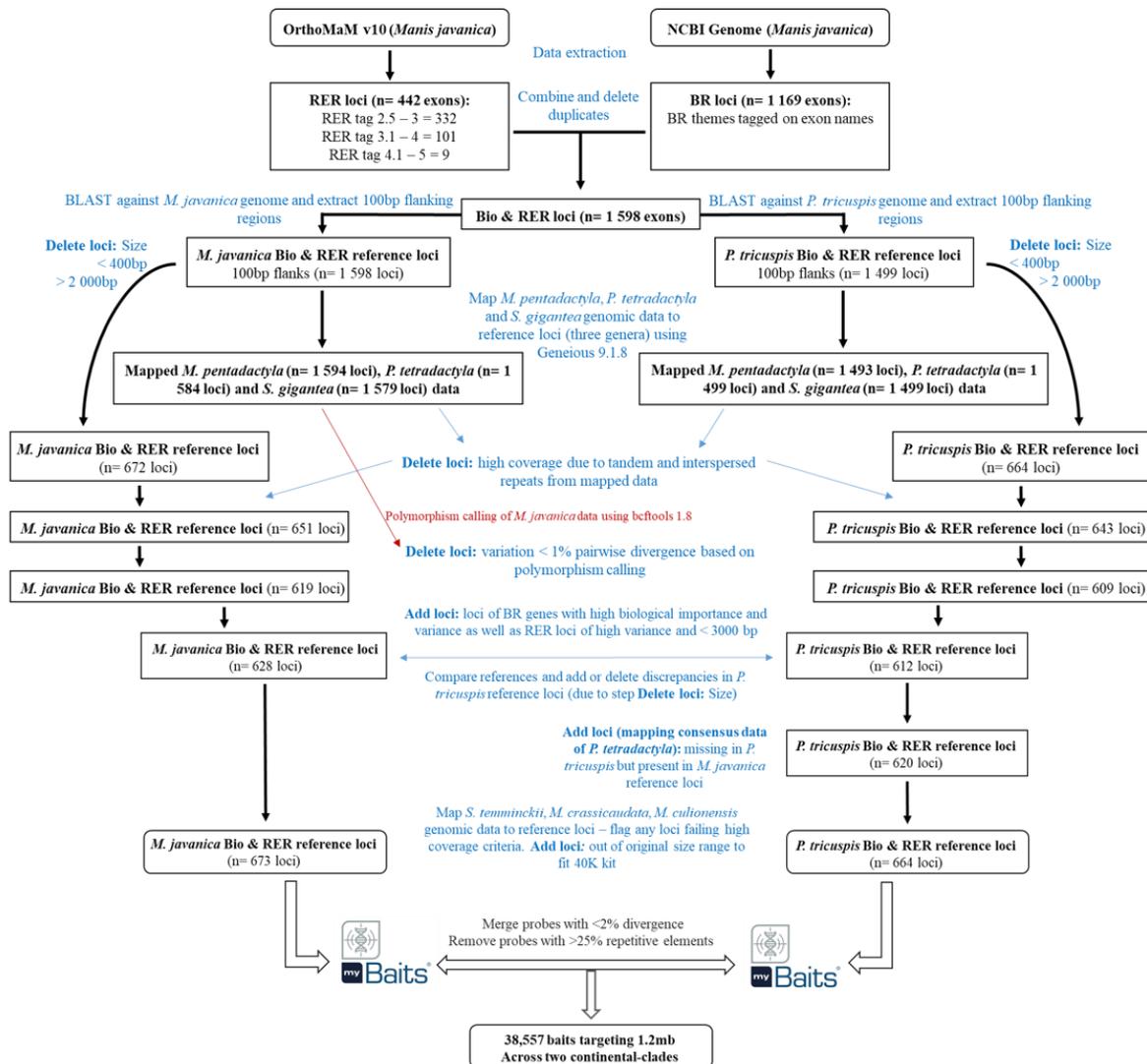


Figure 1: The pipeline designed and implemented to create the final list of loci necessary for gene-capture. Based on the *Manis javanica* (GCF_001685135.1) and *Phataginus tricuspis* (GCA_004765945.1) reference assemblies as well as OrthoMaM v10 database of orthologous genes.

Sample collection and metadata

A total of 842 tissue / DNA samples were collected across all eight extant species from 16 museums (n=189) and fresh sample collections (n=653; Figure 2). We curated a database of per-

sample metadata that included all information related to each sample from its source. Sample localities were crosschecked by first reading through the geographic notes, locality names or coordinates provided by collectors, and then finding these locations with Google Maps. Once sampling localities were identified, additional metadata was added: additional/new names of locality, decimal co-ordinates (unless already provided), accuracy of locality information (0=continent/sub-continent, 1=country, 2=north/south/east/west of country, 3=region/district/province, 4=town/city, 5=accurate field site description, 6=field site with co-ordinates given), museum/fresh sample, and site-type (field site, short-range market, long-range market, seizure, unknown). Site-type was determined either by the author's knowledge of the sample location or by using information surrounding the locality with Google maps (roads, infrastructure, and terrain). If the latter was used this is how the site-types were determined:

- Field site – an isolated locality not near a town, city or major route. These are supposed to be linked to animals found in the field, but could sometimes be small villages where the sample most likely came from a walk-able distance in the surrounding field.
- Restricted-range market – a site that forms part of a small town not near major crossroads or routes. These samples likely came from the surrounding field or villages nearby, but not regional or cross-country.
- Long-range market – a city market, large city, crossroads or town linked by major routes. These samples are likely part of a larger regional to cross-country commercial trade linked by major routes and within major cities, whereby restricted-range markets are likely feeding these markets.
- Seizure – samples either stated as seizures by authorities or those collected outside home-range countries whereby they formed part of an international trade.

- Unknown – samples that did not have accurate enough sampling localities to be given a field-site type (accuracy of 0–3 or continental–regional).

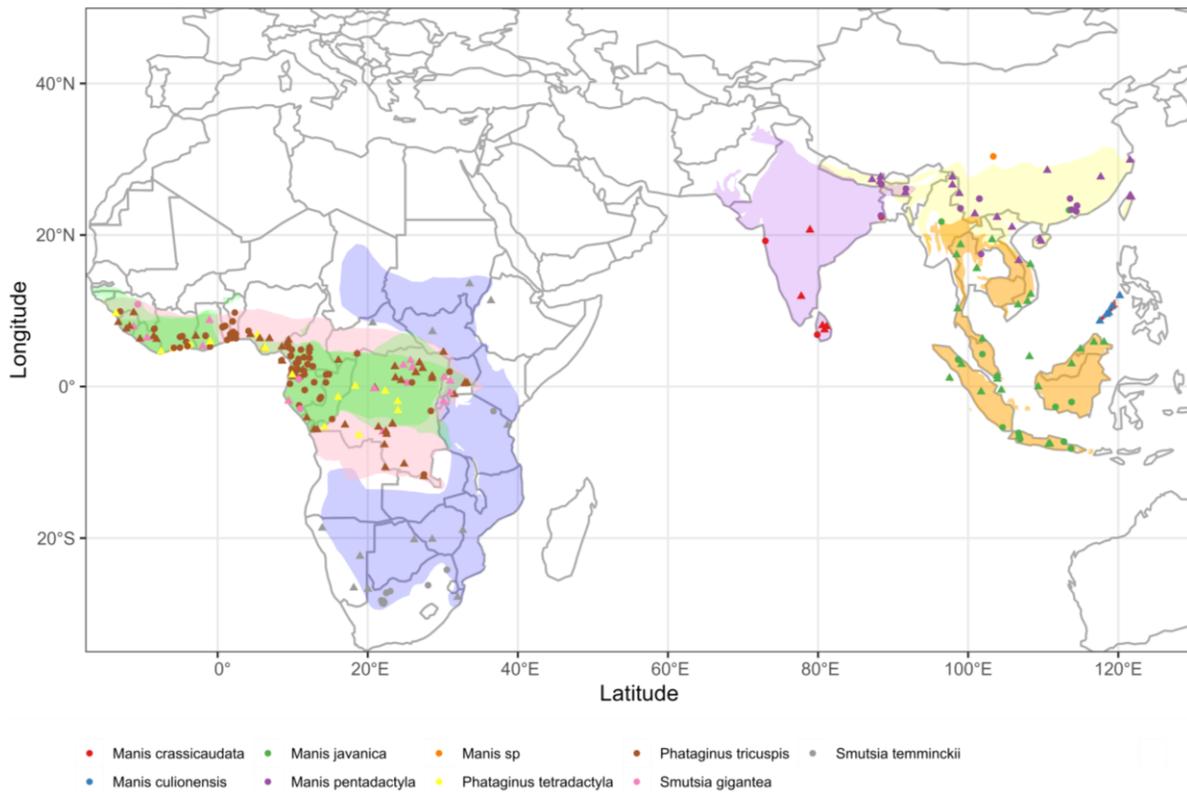


Figure 2: Sampling effort for all eight known species of pangolin and a potentially new taxon (*Manis sp.*), along with their respective ranges. Points are coloured per species with shapes indicating whether they come from museum specimens (filled triangle) or freshly sampled individuals (filled circle).

DNA extraction, gene-capture and sequencing

DNA was extracted using Qiagen DNeasy Blood & Tissue Kit, an in-house bone extraction protocol (M. Pruvost; PACEA, Université de Bordeaux) with the QIAquick Gel Purification Kit, or an in-house phenol chloroform protocol depending on sample type and whether they were fresh or came from museums. We used Nanodrop (spectrophotometric quantification of DNA), Qubit 2.0™ (fluorometric quantitation of dsDNA), and Agilent D5400 Fragment Analyzer (DNA size

profiling) to measure DNA quality and yield after extractions for all samples.

Genomic DNA libraries were prepared using the Illumina Truseq Nano HT kit, whereby each sample was indexed using Illumina unique dual indexes. Variations in the bead-based sizing and purification as well as the number of cycles for enrichment PCR steps were dependent on the level of DNA degradation. However, we aimed for a minimum DNA fragment length of 70bp, which fitted the bait size (70bp baits) and allowed for the inclusion of heavily degraded libraries with small fragment sizes. Qubit 2.0™ and Agilent D5400 Fragment Analyzer were again used to measure library yield and complexity before pooling samples at equimolar concentrations by 4–17 samples per pool for gene-capture. Samples that were heavily degraded were pooled in pools of 1–3 samples. A preliminary gene-capture run with four museum samples was used to test the best bait hybridisation and washing temperatures. The temperatures were 63° C (10–15% bait-target divergence) and 65° C (10% bait-target divergence), whereby 65° C was chosen based on a higher capture efficiency (83.57% compared to 64.57% average capture efficiency – percent of reads sequenced to reads mapped to the target region).

We then used the myBaits® v5 standard protocol to capture our regions of interest at the aforementioned bait hybridisation and washing temperate. Incubation times for bait hybridisation (16–24 hours) and capture PCR amplification cycles (12x–16x) varied depending on library concentrations. Post-capture PCR enrichment was performed with KAPA HiFi HotStart ReadyMix and xGen® Library Amplification Primers (IDT), whereby each post-capture pool was aliquoted into two parts for enrichment and then re-pooled after enrichment in order to increase DNA yields per capture reaction without high levels of PCR duplication (White et al. 2019). Enriched, post-captured pools were purified using cleanNGS (CleanNA) purification beads at 1:1 bead-sample ratio. Qubit 2.0™ and KAPA SYBR FAST qPCRs were used to measure post-capture

pool yield before pooling captured libraries by equimolar concentrations for paired-end sequencing with Illumina MiSeq Nano / MiSeq 2 next generation sequencers (150bp/250bp read lengths depending on initial sample quality).

Simple capture efficiency estimates to guide sequencing efforts

In order to guide further laboratory processing and sequencing efforts we employed a simple pipeline to clean newly sequenced raw reads (fastq), map them to the *M. javanica* baits reference (BWA) and undergo post-mapping cleaning with summary statistics (SAMtools, Qualimap and custom scripts). These summary statistics included capture efficiency, average per-site-depth, average percentage of the target regions covered with at least 10x depth, and mean Phred mapping quality. Capture efficiency was calculated by the percentage of reads mapped (post mapping quality cleaning) compared to those that were sequenced (post sequence quality cleaning).

Gene-capture sequencing processing with EAGER

For the following bioinformatics analyses, we used the bait design loci of *M. javanica* and *P. tricuspis* as fasta references whereby each locus is a “chromosome”. A mitochondrial genome was added to each reference (KT445979 for *M. javanica* and KP306514 for *P. tricuspis*) as an additional “chromosome” in order to obtain mitogenomes from mtDNA amplicons that were sequenced as a by-product (Hari and Siew Woh 2015; Hassanin et al. 2015). These references were indexed with SAMtools v1.10 *faidx* (Li et al. 2009).

For sequencing processing, we employed the nf-core platform’s EAGER v2.4.5 pipeline (Ewels et al. 2020; Peltzer et al. 2016), a comprehensive bioinformatics pipeline for read pre-processing, mapping, and genotyping that is tailored to ancient DNA (aDNA). We applied the

pipeline to the point prior to genotyping after which we followed our own genotyping protocols (see below). Pipelines were run separately for Asian (*M. javanica* baits reference) and African (*P. tricuspis* baits reference) species and differed depending on whether samples were from museums or not. We used a TSV (Tab Separated Values) input for each run, which meant the merging of sequencing data across lanes, sequencing runs, sequencing configuration types, and samples where necessary. Each of the steps performed are described briefly below.

Sequencing quality was assessed with FastQC v0.119 before and after undergoing cleaning and read merging (collapsing) with fastp v0.20.1 (Chen et al. 2018) and AdapterRemoval v2.3.2. Cleaned reads were mapped to their respective bait reference using the BWA-MEM algorithm of BWA v0.7.17 (Li and Durbin 2009), followed by filtering with SAMtools v1.12 at a mapping quality threshold of >30 Phred. Removal of PCR amplicon duplicates was completed by Picard v2.26.0 *MarkDuplicates* and library complexity was estimated with Preseq v3.1.1 (Daley and Smith 2013). To assess DNA damage, DamageProfiler v0.4.9 (Neukamm et al. 2021) was implemented, whereby we set the length filter option (-l) to 70bp due to our lower sizing bounds of our sequencing libraries. For museum samples, we included the option *run_mapdamage_rescaling* which implements mapDamage2 v2.2.1 (Ginolhac et al. 2011) BAM rescaling (probabilistically replaces T's back to C's depending on the likelihood of damage induced reference-sample mismatch). SAMtools v1.12 *flagstat* was implemented before and after mapping filtering while Qualimap v2.2.2 (Okonechnikov et al. 2015) was used to extract mapping statistics post-filtering and cleaning.

MTNucRatioCalculator v0.7 was used to calculate mtDNA to nDNA ratios per de-duplicated library, whilst EndorS.py v0.04 provided levels of endogenous DNA off flagstat outputs (Peltzer et al. 2016). Finally, all aforementioned quality estimates and pipeline statistics were

combined per run using MultiQC v1.13 to help guide the influences of contamination and degradation on genomic insights (Ewels et al. 2016).

Published genome processing

In order to include all available genomic data on pangolins, short read data from previously published sources were downloaded. This included genomic data (n=79 individuals) from the five novel species genomes of Heighton et al. (unpublished – Chapter 2), *M. javanica* and *M. pentadactyla* representatives from previously identified populations (Choo et al. 2016; Hu et al. 2020a; Wang et al. 2022) and an unknown species of pangolin (see Chapter 2: Cao et al. 2021).

Short read data from all species were cleaned with fastp v0.19.4 using default parameters and then mapped to the respective clade baits reference (Asian or African) using BWA v0.7.15. We used SAMtools v1.10 *view* to discard unpaired reads (flag -f 1), mapped reads with a low mapping quality (>30 Phred using -q30), and unmapped, secondary and supplementary mapped reads (flag -F 2308). The reason for implementing this filtering method is due to us mapping whole genome data on a reduced representation of the genome (baits loci). This can result in the possibility of reads half-mapping to the end of a locus boundary and then being inaccurately mapped to other parts of this same locus (supplementary alignments) and reads inaccurately mapping to more than one locus (secondary alignments). Removal of PCR amplicon duplicates was completed by Picard v2.26.0 *MarkDuplicates*.

Genotyping and filtering

With the possibility of species mislabeling of samples and duplicate samples per individual, we conducted a pre-phylogeny and principle components analysis (PCA) per continent as well as

kinship analyses to tend to these anomalies. In order to do this we used BCFtools v1.8 to obtain genotype likelihoods per continent (African and Asian species), and filtered them by depth (8x-350x), site quality (>20 Phred), site missingness (<30%), and variant type (removed indels) using VCFtools v0.1.15 (Danecek et al. 2011). After excluding the mitochondrial genomes with VCFtools v0.1.15, vcf files were converted to fasta alignment files with *vcf2phylip.py* (Ortiz 2019) before undergoing Neighbour-Joining phylogeny testing in MEGA X (Kumar et al. 2018) using the Kimura 2-parameter model of evolution. To aid in the identification of mislabelled samples, a PCA was conducted on each continental clade using PLINK v2.0 (Chang et al. 2015) in which individuals above 95% missing data were removed. Pairwise relatedness estimates based on the original vcf call files were determined using the KING algorithm (Manichaikul et al. 2010) which was run with *relatedness2* in VCFtools v0.1.15. An individual from each pairwise relatedness comparison of above 0.354 (duplicate/ monozygotic twin) was removed from further analyses, whilst mislabeled individuals were correctly assigned to their species based on the continent specific phylogenies and PCAs.

Once duplicates and mislabeling was corrected for, BCFtools v1.8 was again used to determine per species genotype likelihoods from all cleaned bam files using *mpileup*, call for variants using *call*, and normalization of those variants using *norm*. Each species-specific resultant vcf/SNP file was indexed with BCFtools v1.8, before extracting various variant statistics (mean depth per site, allelic frequencies, site missingness, individual missingness) with VCFtools v0.1.15 (Danecek et al. 2011). These estimates were plotted in ggplot2 (Wickham 2011) in order to guide filtering with VCFtools v0.1.15. This included removing indels, sites under 8x average depth (except *S. gigantea*, *M. crassicaudata* and *M. culionensis* that was 5x), a species-specific maximum site depth (35x-75x), sites with more than 30% missing data, variant quality below

Phred 20, and a minor allele count of 3 (except *M. culionensis* due to limited samples). Variants under selection (non-neutral) from the resultant per-species filtered SNP files were identified using Tajima's D estimates (>2 & >-2) in VCFtools v0.1.15. These files were then split into three partitions: nDNA loci with neutral SNP's only, nDNA loci with all SNPs (neutral and non-neutral), and the mitogenome. Both the neutral and full nDNA SNP files for each species underwent linkage disequilibrium pruning with PLINK v2.0 (Chang et al. 2015) by removing SNPs within a 50 SNP sliding window, shifting 10 SNPs per step, that had $r^2 > 0.2$. During this step, all individuals with more than 75% missing data were excluded, except one from *M. crassicaudata* due to the limited sample count.

Population stratification

The linkage pruned files containing all the nDNA SNPs (neutral and non-neutral) of each species were used to create PCAs using PLINK v2.0 and ggplot2. All individuals with more than 75% missing data were excluded, except one from a *M. crassicaudata* individual as it was a vital reference sample. All SNPs were used as the idea was to stratify the populations as much as possible by including both neutral genetic distance (neutral SNPs) and local adaptation (non-neutral SNPs).

The linkage pruned files containing solely bi-allelic (filtered with PLINK v2.0), neutral nDNA SNPs of each species were used for population structuring with ADMIXTURE v1.3 (Alexander et al. 2009). Here each K (1-10) was computed 10 times, with a cross-validation of 10 iterations, in order to test for variability amongst the replicates surrounding each K with CLUMPAK (Kopelman et al. 2015). CLUMPAK found no genuinely different modes for each K of each species and the best K per species was determined from cross-validation estimates and

geographic scenarios. Admixture/structure plots were created in Rstudio v4.2.1 following a custom R script after merging sample metadata with ADMIXTURE's qmatrix. The following map layers and their sources were used for mapping the genotypes: global ecological land units (We used Biomes: Sayre et al. 2014), global elevation (GlobalSolarAtlas-v2; <https://globalsolaratlas.info>), and 1:50m scaled water bodies and country boundaries (Natural Earth - <https://www.naturalearthdata.com/>; South 2022).

Population genetic estimates

Samples were assigned to ancestral population clusters based on the best K qmatrix estimates for each species from ADMIXTURE. Samples that had more than 80% of their genotypes linked to a defined cluster were placed in that cluster while those that had a certain level of admixture (less than 80% aligned to a cluster) were discarded. This was similar to what has been done before (Nørgaard et al. 2017) based on the idea that clustering methods are not perfect. We calculated mean values of standard genomic parameters using the “gstudio” (Dyer 2014) and “poppr” (Kamvar et al. 2014) packages for each of these aforementioned clusters following von Takach et al. (2022). These included the number of alleles (A), effective number of alleles (A_E), observed heterozygosity (H_O), expected heterozygosity (H_E), Wright's inbreeding coefficient (F_{IS}), and number of private alleles (A_p). A small sample size correction was used to account for the biases resulting from differences in sample size between clusters. All estimates, apart from private alleles, were averaged across loci whilst a standard error (within population variation) was calculated for each metric.

To quantify the level of genomic differentiation between countries, we calculated estimates of F_{ST} (Weir and Cockerham 1984) and their significance levels using the *stampFst* function of

the “StAMPP” package (Pembleton et al. 2013). The function calculates significance values by bootstrapping (n=1000 replicates) across loci. As these were per-country estimates, all individuals were used (introgressed or not), except those that came from seizures due to the probability of them not coming from the country in which they were seized, and thus influencing the differentiation between countries.

Patterns of isolation-by-distance (IBD) were visualised through scatterplots of scaled Euclidean distance against geographical distance for all individual-based pairwise comparisons. Geographical distances were calculated from their co-ordinates with the “fossil” package (Vavrek 2011). To test the significance of geographical distance on genetic differentiation, a Mantel test was implemented using 999 permutations. As these estimates included geographic distances, all individuals were used (introgressed or not), except those that came from seizures and long-range markets due to the probability of them not having local co-ordinates and thus influencing the geographic distances between individuals.

All calculations were conducted in R v4.2.1 whereby vcf files were first converted to genind objects using the “dartR” package (Gruber et al. 2018), and then subsequently manipulated using “adegenet” (Jombart and Ahmed 2011). Prior to these calculations, sampling duplicates, locus and individual completeness as well as tests of whether loci were in Hardy-Weinberg Equilibrium (HWE) were investigated using “adegenet”, “poppr” (Kamvar et al. 2014) and “pegas” (Paradis 2010) packages. We found no indications of failed assumptions in HWE or duplicated sampling.

Historical relationships among regions

TreeMix v1.13 (Pickrell and Pritchard 2012) was implemented using the same dataset used for F_{ST}

calculations (seized samples excluded) per country. Countries showing the highest level of differentiation with pairwise F_{ST} calculations were placed as an outgroup for building of the maximum-likelihood phylogeny. To determine the optimal number of migration edges for the phylogeny with OptM using the default method (Fitak 2021), each number of migration edges allowed (0-10) was computed 10 times with a random seed per iteration.

Results

We sequenced 842 samples at an average sequencing depth of 35.86x with an average of 64.33% of the target region covered by at least 10x depth (after excluding failed samples - lower than 3x depth). Capture efficiency was 18.78% (percentage of reads mapped compared to those that were sequenced) and was equivalent across species (Supplementary Fig 1a). Capture efficiency was seemingly higher for museum samples than that of fresh samples (Supplementary Fig 1d). Except for capture pools of 16–17 individuals, capture efficiency was seemingly not influenced by the number of individuals in capture pools (Supplementary Fig 1c). The sequencing depth was low for pools of 16–17 individuals, which probably influenced the latter results (Supplementary Fig 1d).

Through initial continental-specific phylogenies and PCAs, as well as relatedness estimates, we identified 15 individuals that had incorrect species labels and 47 individuals that had duplicate samples or were part of a pair of monozygotic twins. The *Manis* species identified as a possibly new taxon was separated from other South-East Asian species using an initial PCA and thus was excluded from further analyses (Supplementary Fig 2; Also see Heighton et al. unpublished – Chapter 2).

Population stratification and structuring

The PCA for *P. tricuspis* separates into three major clusters (western/central/eastern part of its range) where axis 1 stratifies across central-western Africa (central species range) with central Africa (eastern range), while axis 2 stratifies across western Africa (western range; Figure 3a). Looking at each major cluster in more depth five clusters can also be noticeable (Figure 3a; Supplementary Fig 3a). These include western-central (Cameroon, Equatorial Guinea, Gabon, Nigeria; Supplementary Fig 3a), eastern-central (Republic of Congo, Gabon, Central African Republic; Supplementary Fig 3b), eastern (Republic of Congo, Democratic Republic of Congo, Uganda; Supplementary Fig 3b), Dahomey Gap (Nigeria, Benin, Togo; Supplementary Fig 3c), and western (Ghana, Ivory Coast, Liberia, Sierra Leone, Guinea; Supplementary Fig 3c) parts of the species range. One Ghanaian individual from a field site could be an outlier from the western range cluster (Figure 3a; Supplementary Fig 3c). Nigeria, Gabon and Republic of Congo are found in more than one of the five clusters, however, the only Nigerian individual on the eastern part of the country falls within western-central range cluster. Likewise, axes 3 and 4 had an influence on population stratification, indicating the five minor clusters with the merging of the potential Ghanaian outlier to the western range cluster (Supplementary Fig 4).

ADMIXTURE results suggest eight most likely clusters ($K=8$; Figure 3b & c), separated in a similar fashion to the five minor clusters of the PCA, except that there are four clusters in western-central range (around Cameroon) which show high levels of admixture across the region. These four are not easily split geographically either (northwest has more light blue, northeast has more turquoise, central has more purple, and south has more light green), though, there seems to be a partition into two across the Sanaga River (Figure 3c). Yaoundé market is made up of a mixture of all four clusters, albeit very little from northwest (light blue), while Douala market is

exclusively linked to the two clusters found north of the Sanaga River (light blue and turquoise; Figure 3c). The Ghana outlier individual is again merged with its western range in West Africa.

At $K=3$, the three major clusters of western, central and eastern parts of the species range is noticeable (Supplementary Fig 5). At $K=2$, central is a separate cluster to the eastern and western parts of the species range. At $K=4$, there is a split between northeast and southwest of Cameroon, across the Sanaga River. At $K=6$ and $K=7$, the western (West Africa) and central-eastern (Central Africa) parts of the species range are split into two at the same points that can be found at $K=8$. At $K=10$, structure can be found in West Africa with the separation of Ghana and Ivory Coast (admixed with far west) as well as within Dahomey Gap suggesting finer levels of structure may still be found at finer geographic scales. The latter has gradual shift (genetic cline) between two clusters on east and west of the gap with high levels of admixture (Supplementary Fig 5).

With limited seizures for the species, only one individual seized in Central African Republic could be linked to the eastern range cluster, the rest were already correctly assigned through seizure reports.

3. Population genomics

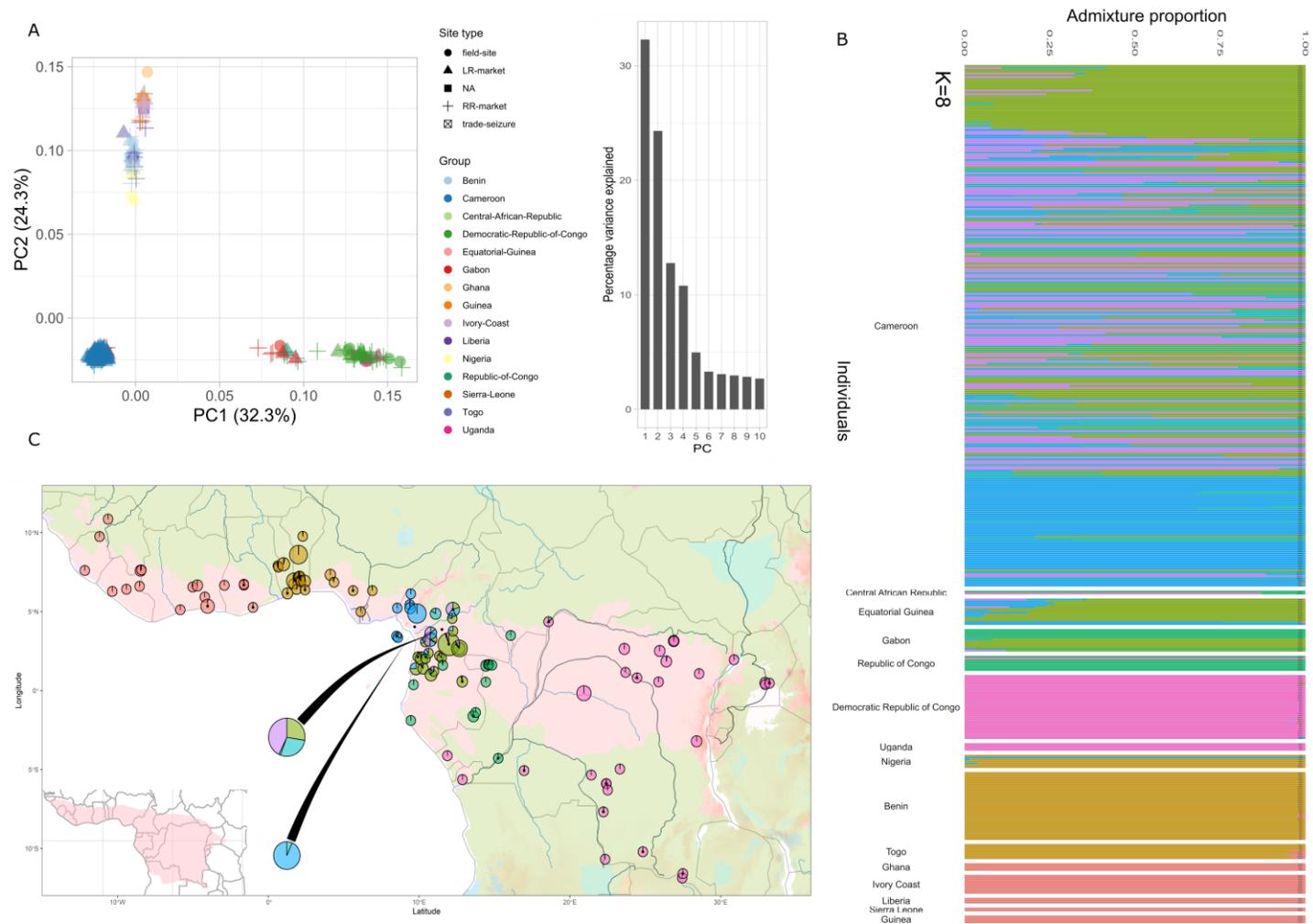


Figure 3: Population structure and stratification of *Phataginus tricuspis* using single nucleotide polymorphisms (SNPs). (A) The Principle component analysis (PCA; along with a histogram of variance explained per axis) for the top two axes of differentiation in which points are coloured by country and shaped by type of site from which the sample was collected (8 520 SNPs, 427 individuals). (B) ADMIXTURE bar graph for the most likely K-value in which samples belonging to countries are clumped and each site type is labelled (8 312 SNPs, 456 individuals). Samples are ordered (from top down) from South-East to North-West in each country. (C) The geographic distribution of these ADMIXTURE clusters in the form of pie-charts that are size-scaled by the number of samples and whereby the admixture proportions of samples within 1.1 km were averaged. The two major urban bushmeat markets in Cameroon (Yaoundé - top and Doula - bottom) are placed in the sea due to large sample size. The larger the water bodies the darker the blue, while pies with symbols are either long-range markets (filled circle), trade seizures (filled triangle), or unknown (filled diamond).

The PCA for *P. tetradactyla* also separates into three major clusters (western, central and eastern parts of the species range) where axis 1 stratifies across east-west and axis 2 stratifies across north-south (Figure 4a). Individuals from the Democratic Republic of Congo are found in both the central and eastern clusters. The one individual at the centre of the country that aligns to the central range cluster (Western Central Africa) is of unknown locality within the country while the one southwest of the country is known to stem from that region, but we could not find an exact locality (both labelled with an unknown site type). ADMIXTURE results validate the same cluster trend at the most likely K (K=3; Figure 4b &c). At K=2, a similar pattern is indicated but the eastern part of the species range is admixed between the other two clusters (Supplementary Fig 6). At higher Ks there is a separation of Nigeria from the central part of the species range (Western Central Africa) at K=4 onward, and a split in the western range (West Africa) at K=8 between Liberia and Ivory Coast (Supplementary Fig 6).

Smutsia gigantea is also split into three (western, western-central, and eastern species range) with the PCA, where axis 1 stratifies the western-central/eastern range clusters and axis 2 is for the western cluster (Figure 5a). Although these two axes explain the majority of variance, there is no major decline in variance explained from axes 3 to 10. ADMIXTURE results at the most likely K (K=4) suggest an additional cluster splitting the eastern cluster into central and central eastern (Figure 5b &c). At K=2, east of the species range is split from the central and western parts around Democratic Republic of Congo and Central African Republic (Supplementary Fig 7). There are no meaningful structure patterns at higher Ks, apart from the aligning of the far eastern Democratic Republic of Congo individual with the rest of the samples from this country, rather than it being aligned to individuals from Uganda and Rwanda at the most likely K (Supplementary Fig 7).

By PCA, *S. temminckii* is suggested to have three major clusters (Sudanese, eastern, and southern species range) whereby axis 1 separates the Sudanese cluster from the other two and axis 2 segregates

individuals in a distinctive north-south pattern (Figure 6a). A fourth, minor cluster may be possible for Zimbabwe and Botswana (as well as the individuals of an unknown locality and a seizure). Axis 1 is dominant in its level of explained variance. ADMIXTURE also indicates three clusters (most likely: K=3) but rather that the eastern range cluster extends further south until Zimbabwe and Botswana, after which a new cluster (south-western) can be noted in South Africa and Namibia (Figure 6b & c). The latter cluster has a large proportion of individuals exhibiting admixture with a slight trend of south-west (less eastern cluster) to north-east (more eastern cluster). At K=2, the individuals from Sudan and South Sudan are separate to the rest of the localities, a pattern found throughout the values of K (Supplementary Fig 8). At K=4, Ethiopia and Tanzania form their own cluster with some levels of admixture in Zimbabwean individuals. An individual of unknown locality for the species can be linked to Botswana through the PCA, and shares 100% of the eastern range cluster in ADMIXTURE. Two seized samples from Johannesburg, South Africa can be linked to South Africa/Namibia and Zimbabwe, respectively with the PCA. The former has a similar admixture pattern to individuals from Namibia and South Africa, while the latter shares 100% of the eastern range cluster.

3. Population genomics

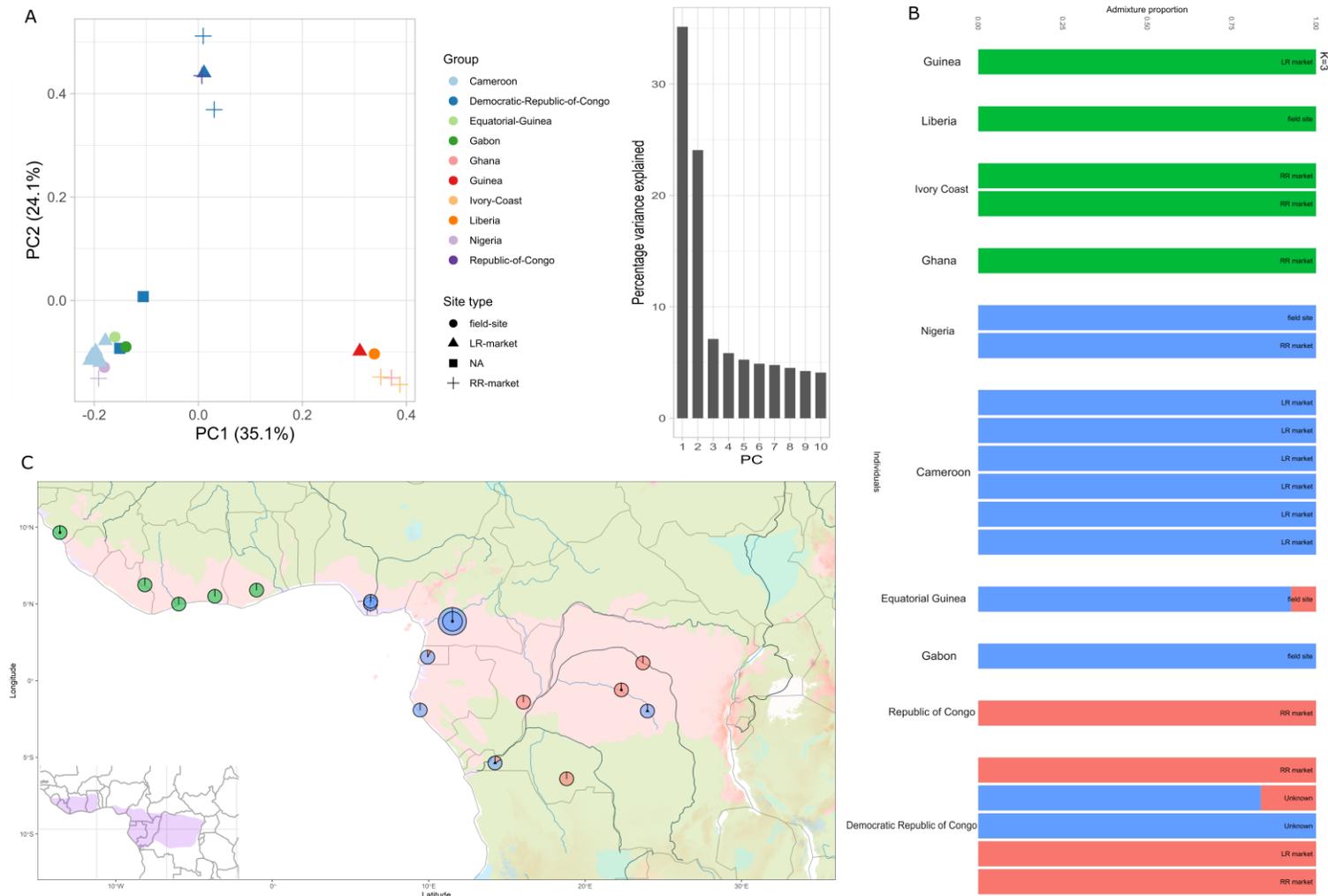


Figure 4: Population structure and stratification of *Phataginus tetradactyla* using single nucleotide polymorphisms (SNPs). (A) The Principle component analysis (PCA; along with a histogram of variance explained per axis) for the top two axes of differentiation in which points are coloured by country and shaped by type of site from which the sample was collected (798 SNPs, 21 individuals). (B) ADMIXTURE bar graph for the most likely K-value in which samples belonging to countries are clumped and each site type is labelled (793 SNPs, 21 individuals). Samples are ordered (from top down) from South-East to North-West in each country. (C) The geographic distribution of these ADMIXTURE clusters in the form of pie-charts that are size-scaled by the number of samples and whereby the admixture proportions of samples within 1.1 km were averaged. The larger the water bodies the darker the blue, while pies with symbols are either long-range markets (filled circle) or unknown (filled triangle). The outlier individual in central Democratic Republic of Congo had an unknown sampling locality apart from the country itself, hence it was placed in the centre of the country.

3. Population genomics

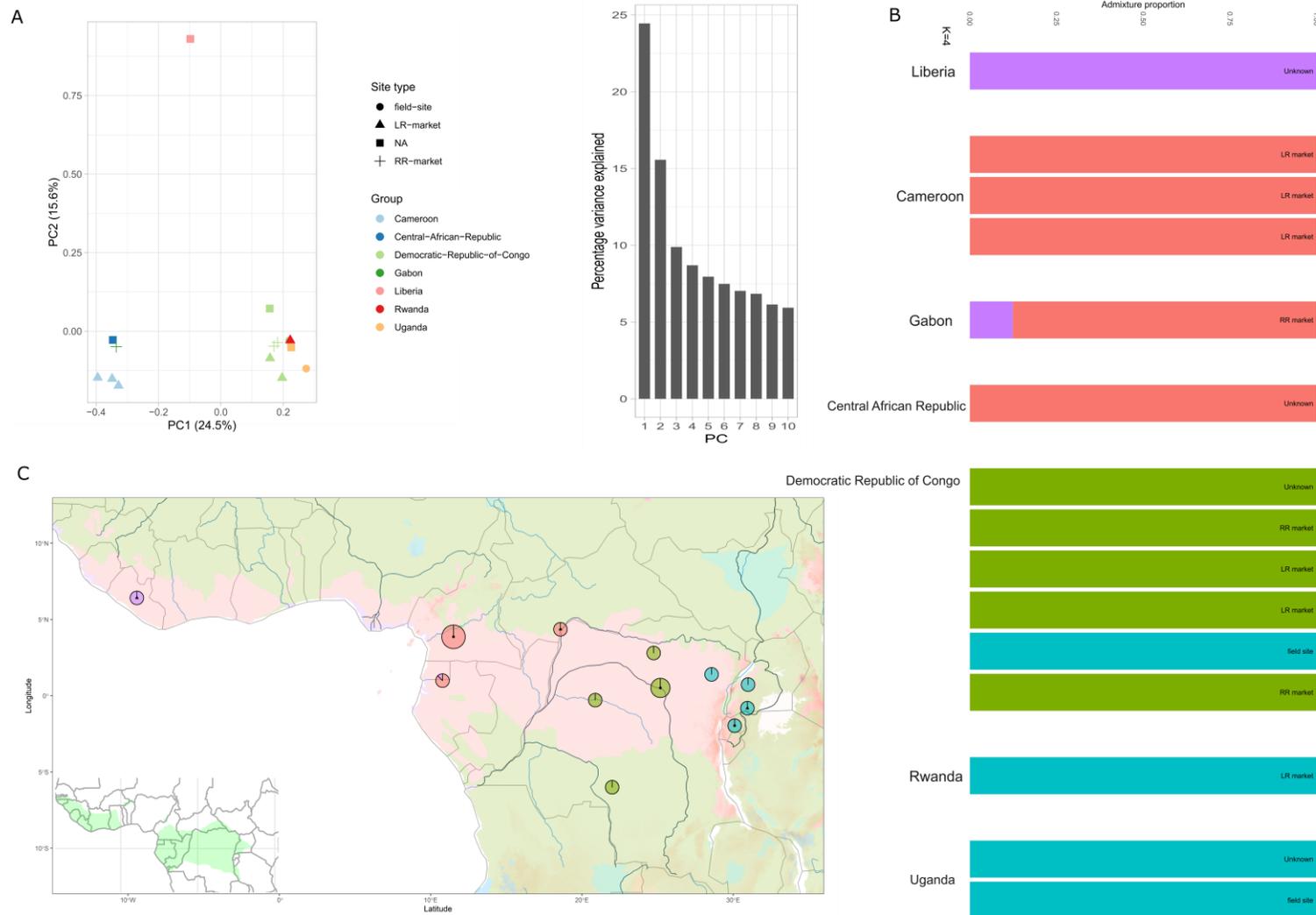


Figure 5: Population structure and stratification of *Smutsia gigantea* using single nucleotide polymorphisms (SNPs). (A) The Principle component analysis (PCA; along with a histogram of variance explained per axis) for the top two axes of differentiation in which points are coloured by country and shaped by type of site from which the sample was collected (794 SNPs, 16 individuals). (B) ADMIXTURE bar graph for the most likely K-value in which samples belonging to countries are clumped and each site type is labelled (768 SNPs, 15 individuals). Samples are ordered (from top down) from South-East to North-West in each country. (C) The geographic distribution of these ADMIXTURE clusters in the form of pie-charts that are size-scaled by the number of samples and whereby the admixture proportions of samples within 1.1 km were averaged. The larger the water bodies the darker the blue, while pies with symbols are either long-range markets (filled circle), trade seizures (filled triangle), or unknown (filled diamond).

3. Population genomics

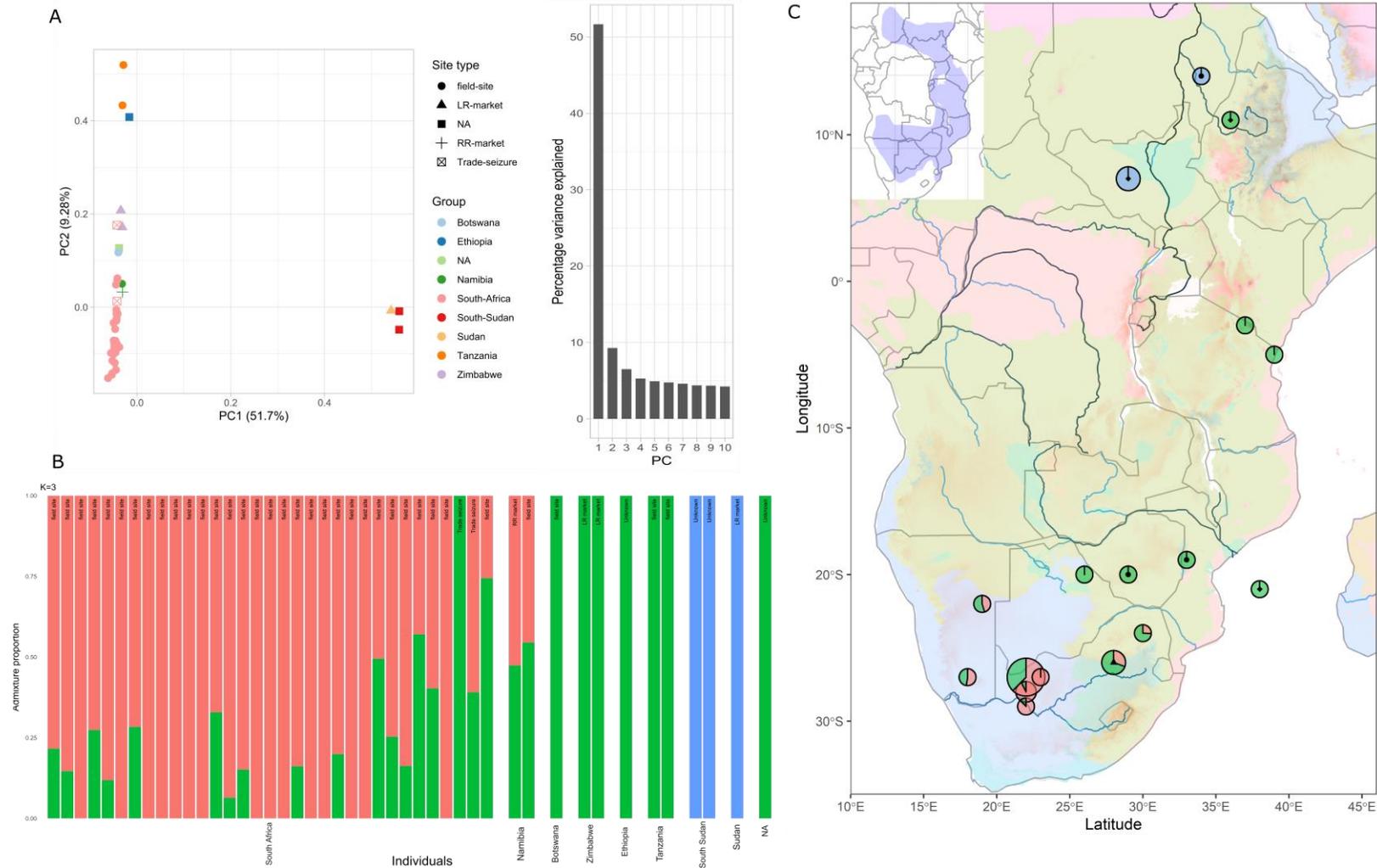


Figure 6: Population structure and stratification of *Smutsia temminckii* using single nucleotide polymorphisms (SNPs). (A) The Principle component analysis (PCA; along with a histogram of variance explained per axis) for the top two axes of differentiation in which points are coloured by country and shaped by type of site from which the sample was collected (1 516 SNPs, 45 individuals). (B) ADMIXTURE bar graph for the most likely K-value in which samples belonging to countries are clumped and each site type is labelled (1 506 SNPs, 45 individuals). Samples are ordered (from top down) from South-West to North-East in each country. (C) The geographic distribution of these ADMIXTURE clusters in the form of pie-charts that are size-scaled by the number of samples and whereby the admixture proportions of samples within 111 km were averaged. One individual of unknown sampling locality is placed in the sea. The larger the water bodies the darker the blue, while pies with symbols are either long-range markets (filled circle), trade seizures (filled triangle), or unknown (filled diamond)

With the inclusion of seizures, the PCA for *M. javanica* does not suggest any major population stratification (Figure 7a), especially when looking at a country scale (Supplementary Fig 5a-c). When excluding these seizures (boxes), *M. javanica* separates into four major clusters where axis 1 stratifies with a north-south pattern, while axis 2 stratification is less clear (Figure 7a). These clusters include: South-East Asian islands (Sundaland: Sabah, Borneo, Brunei, Java), Sumatra/South-East Asian mainland east (Sumatra, Singapore, mainland Malaysia, southern Thailand, Vietnam, Laos), Indochina west (northern Thailand, southern Myanmar), and western part of the species range (Sino-Burmese border, Yunnan-China, Myanmar). There were no western Javan samples with known locality (only seizures), therefore we do not place this in a cluster, although they are probably within the South-East Asian islands clusters. Including clusters by seizures, there are seemingly two additional clusters with one closer to Sumatra/South-East Asian mainland east (some individuals from MJB2) and the other closer to Indochina west (some individuals from MJB3) clusters (Figure 7a; Supplementary Fig 9d). Axes 3 and 4 also had an influence on population stratification, including the separation of a few outlier individuals from Java Island, Indonesia (Supplementary Fig 9d). Previously identified clusters did not necessarily separate as previous publications found, particularly for the Borneo and Java clusters as well as the MJB clusters (MJB1-3; Supplementary Fig 9c).

ADMIXTURE results suggest four clusters as the most likely ($K=4$) with a similar pattern to that of the PCA, except that Java has its own cluster (purple) that differentiates it from the other South-East Asian islands, and the Indochina west cluster is no longer visible (Admixed individuals; Figure 7b & c). This method therefore separates the previously identified clusters (Java and Borneo; Figure 7b). The results also suggest admixed individuals seized in Java Island, Thailand, Myanmar, and the two major cities of China (Guangzhou and Yunnan; Figure 7b).

At $K=2$ there is a distinct separation of mainland (Indochina northward) vs. South-East Asian islands (Sundaland) populations, whereby Sumatra is admixed between the two (Supplementary Fig 10). After that

at $K=3$, individuals from Sumatra merge with the individuals from Thailand, Singapore and Laos. At higher values of K (from $K=5$ onward), Thailand shows admixture patterns in a gradient (genetic cline) from south (Malaysia Peninsula) to north, all of which show similar patterns to the seizures in Yunnan and Guangzhou China. At $K=6$, the Java cluster is split into two, although it is unclear where these clusters lie.

Seized samples belonging to clusters outside those in which they were seized are only visible for individuals stemming from Sumatra/South-East Asian mainland east cluster (seized in Javan cities, Yunnan China, Guangzhou China) and Borneo cluster (seized in Javan cities). Higher values of K show the seizure MJB1 cluster (Yunnan and Guangzhou) to be aligned with Malaysian Peninsula and southern Thailand, MJB2 (Yunnan) to be with Laos, MJB3 (Yunnan and Guangzhou) with southern Myanmar and northern Thailand, and MJA clusters (MJA1 and MJA2 in Sino-Burmese border and Yunnan) to be aligned with central to northern Myanmar (Supplementary Fig 10).

3. Population genomics

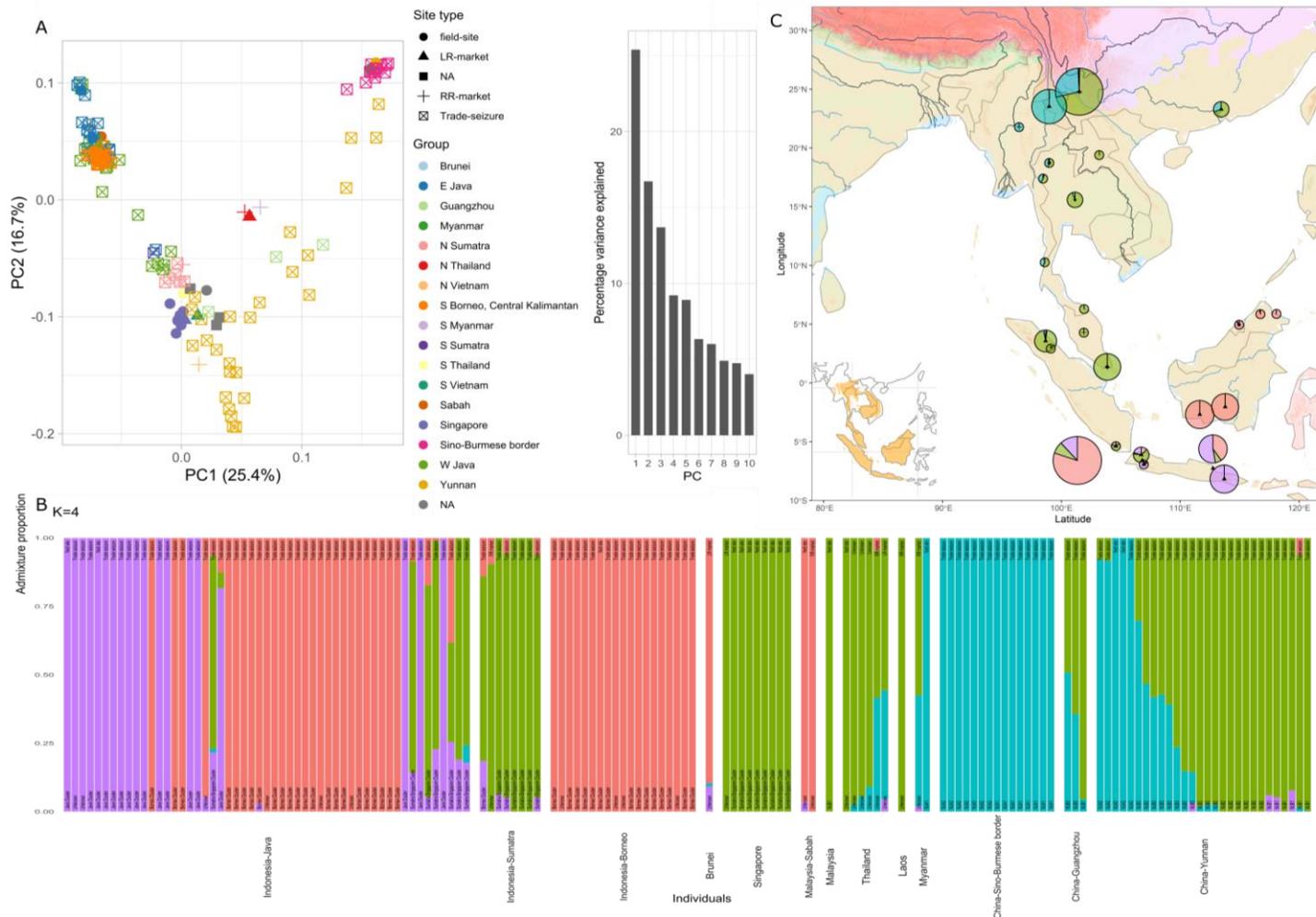


Figure 7: Population structure and stratification of *Manis javanica* using single nucleotide polymorphisms (SNPs). (A) The Principle component analysis (PCA; along with a histogram of variance explained per axis) for the top two axes of differentiation in which points are coloured by island and shaped by type of site from which the sample was collected (4 418 SNPs, 149 individuals). (B) ADMIXTURE bar graph for the most likely K-value in which samples belonging to islands are clumped and each site type is labelled (4 162 SNPs, 148 individuals). Samples are ordered (from top down) from South-East to North-West in each island. (C) The geographic distribution of these ADMIXTURE clusters in the form of pie-charts that are size-scaled by the number of samples and whereby the admixture proportions of samples within 1.1 km were averaged. The two major seizures in eastern Java (Bogor - left and Surabaya - right) are placed in the sea on either side of the island due to large sample size. The larger the water bodies the darker the blue, while pies with symbols are either long-range markets (filled circle), trade seizures (filled triangle), unknown (filled diamond), or field sites with regional accuracy (plus).

Five clusters are suggested for *M. pentadactyla* with the PCA, where axis 1 stratifies with an east-west pattern, while axis 2 stratification is north-south (Figure 8a). These clusters include: Taiwanese (with individuals from unknown locality), central (southern China, Thailand, Vietnam), north-western, (Yunnan-China), western (India, Nepal, Myanmar) parts of the species range, and an unknown location (seizures from Yunnan-China). ADMIXTURE suggests only three clusters as the most likely (K=3), two of which fit the PCA clusters; Taiwanese and central range clusters (Figure 8b & c). The third, western, is a combination of north-western range, western range and unknown location clusters. A number of individuals in the southern range cluster show levels of admixture to the Taiwanese cluster, particularly those closer to that region (Figure 8c). At K=2, the western part of the species range (India, Nepal, Myanmar) is separated from the rest (Supplementary Fig 11). Seizures from China can be linked to the western (seized in Yunnan-China and Sino-Burmese border) and central range clusters (seized in cities in Guangdong and Yunnan provinces of China), whilst those collected/seized in an unknown location were linked to the Taiwanese cluster (Figure 8).

Manis crassicaudata had a low sample size with a large proportion of low depth/quality genomes. Four clusters can be identified with PCA (Figure 9a), these include eastern range (Maharashtra samples), western (West Bengal), unknown-seizures (two seizures in Sri Lanka), and unknown (three individuals from unknown localities in Sri Lanka and India). ADMIXTURE suggests three clusters as the most likely, though there is no distinct geographical pattern with high proportion of admixed individuals (Figure 9b & c). This is also true at higher and lower values of K (Supplementary Fig 12).

Manis culionensis too had a low sample size, nonetheless the three clusters based on the three islands are discernible with the PCA (Figure 10a). Only two clusters were identified with ADMIXTURE in which the two smaller islands surrounding Palawan (Boayan and Bosuanga Islands) were merged (Figure 10b & c). However, one individual from a restricted-range market in in the Palawan cluster was admixed. This

admixed individual is seemingly linked with the individual from Bosuanga Island at $K=3$, and the two satellite islands split from each other at $K=4$ (Supplementary Fig 13).

3. Population genomics

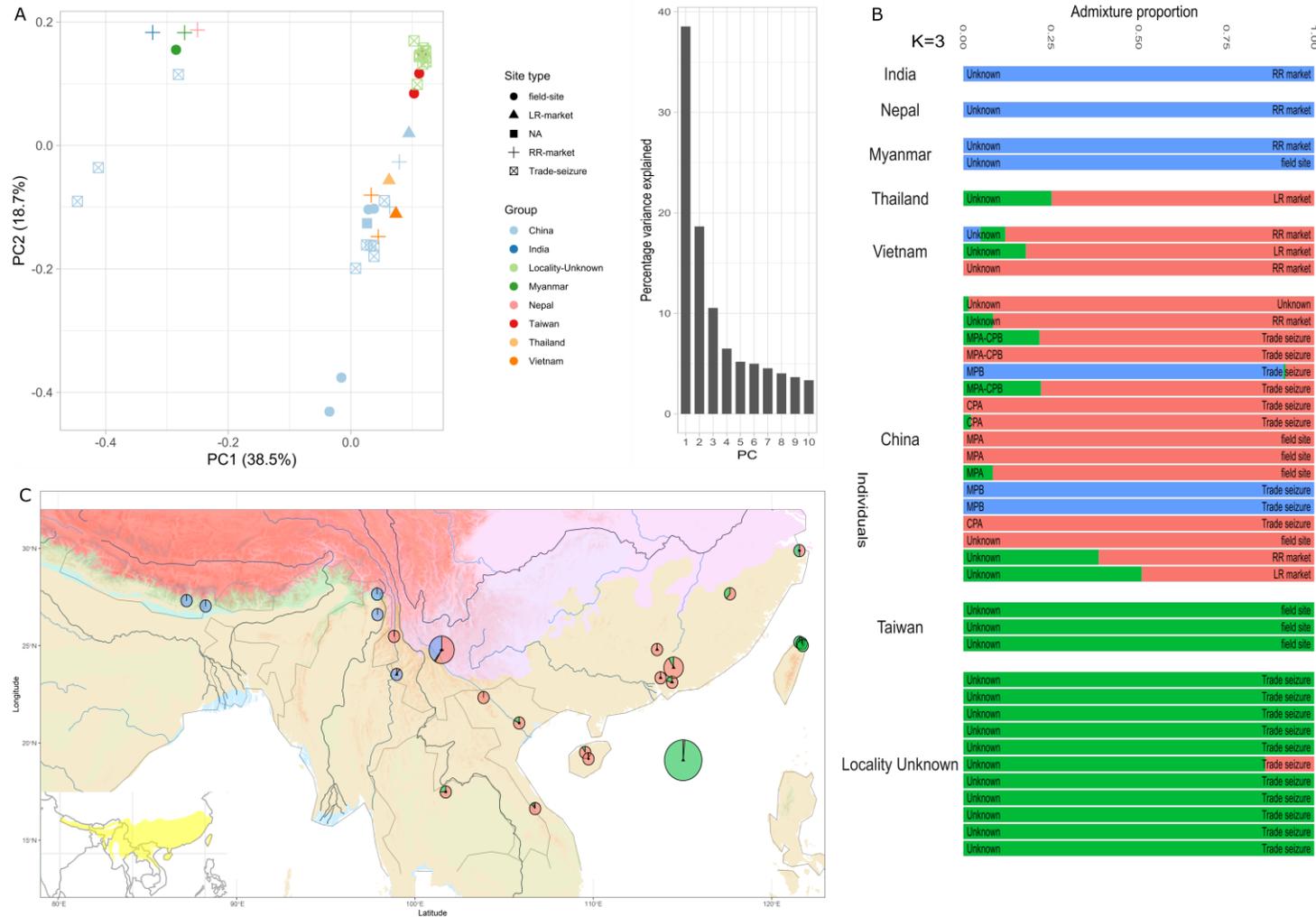


Figure 8: Population structure and stratification of *Manis pentadactyla* using single nucleotide polymorphisms (SNPs). (A) The Principle component analysis (PCA; along with a histogram of variance explained per axis) for the top two axes of differentiation in which points are coloured by country and shaped by type of site from which the sample was collected (840 SNPs, 39 individuals). (B) ADMIXTURE bar graph for the most likely K-value in which samples belonging to countries are clumped and each site type is labelled (816 SNPs, 39 individuals). Samples are ordered (from top down) from South-East to North-West in each country. (C) The geographic distribution of these ADMIXTURE clusters in the form of pie-charts that are size-scaled by the number of samples and whereby the admixture proportions of samples within 1.1km were averaged. Samples coming from an unknown locality were placed in the sea. The larger the water bodies the darker the blue, while pies with symbols are either long-range markets (filled circle), trade seizures (filled triangle), unknown (filled diamond), or field sites with regional accuracy (plus).

3. Population genomics

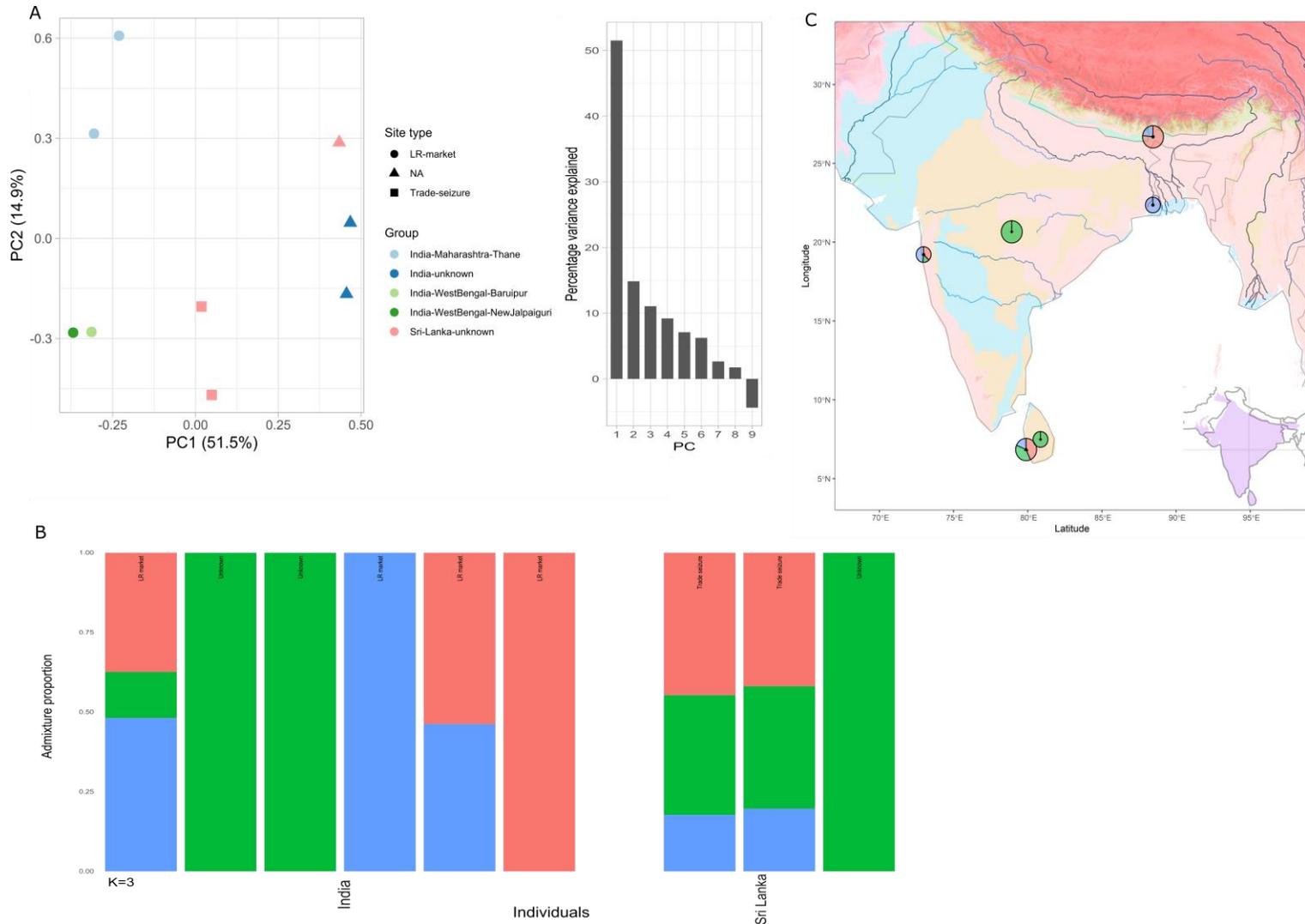


Figure 9: Population structure and stratification of *Manis crassicaudata* using single nucleotide polymorphisms (SNPs). (A) The Principle component analysis (PCA; along with a histogram of variance explained per axis) for the top two axes of differentiation in which points are coloured by site and shaped by type of site from which the sample was collected (77 SNPs, 9 individuals). (B) ADMIXTURE bar graph for the most likely K-value in which samples belonging to countries are clumped and each site type is labelled (73 SNPs, 9 individuals). Samples are ordered (from top down) from South-East to North-West in each country. (C) The geographic distribution of these ADMIXTURE clusters in the form of pie-charts that are size-scaled by the number of samples and whereby the admixture proportions of samples within 1.1km were averaged. The larger the water bodies the darker the blue, while pies with symbols are either long-range markets (filled circle), trade seizures (filled triangle), or unknown (filled diamond).

3. Population genomics

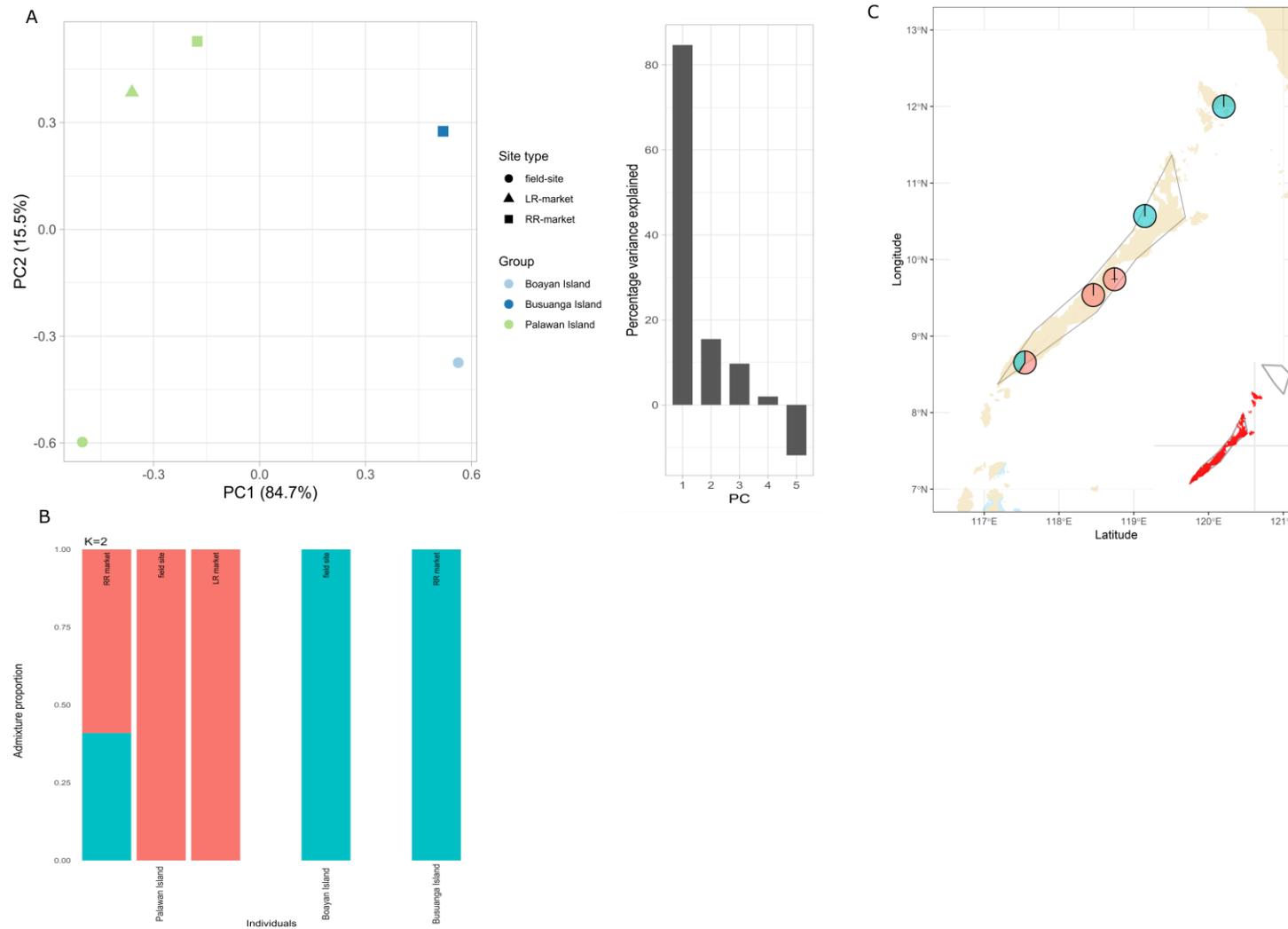


Figure 10: Population structure and stratification of *Manis culionensis* using single nucleotide polymorphisms (SNPs). (A) The Principle component analysis (PCA; along with a histogram of variance explained per axis) for the top two axes of differentiation in which points are coloured by country and shaped by type of site from which the sample was collected (39 SNPs, 5 individuals). (B) ADMIXTURE bar graph for the most likely K-value in which samples belonging to countries are clumped and each site type is labelled (39 SNPs, 5 individuals). Samples are ordered (from top down) from South-West to North-East in each island. (C) The geographic distribution of these ADMIXTURE clusters in the form of pie-charts that are size-scaled by the number of samples and whereby the admixture proportions of samples within 1.1km were averaged. The larger the water bodies the darker the blue, while pies with symbols are long-range markets (plus sign).

Regional differentiation, migration and phylogenetics

Pairwise F_{ST} estimates per country for *S. gigantea* and *S. temminckii* are on average, lower than the *P. tricuspis* and *P. tetradactyla* estimates (Figure 11).

F_{ST} estimates per country for *P. tricuspis* follow a similar pattern to that found in population structuring (Figure 3 & 11a). Uganda is largely differentiated from the rest of the countries, together with the other countries in the eastern range cluster (Democratic Republic of Congo and Republic of Congo), whilst Cameroon and Equatorial Guinea show no differentiation (Figure 11a). The western range cluster has the least amount of differentiation between countries, except for Ghana. Central African Republic is also largely differentiated from other countries despite its centralised geographic positioning (Figure 11a), but is comprised of only one individual. The TreeMix maximum likelihood phylogeny suggests the five major clusters with Ghana and Ivory Coast being slightly differentiated from the rest of the western range countries (West Africa) further west (Figure 11a). The most likely amount of migration edges is two, the largest is between Gabon and Cameroon/Equatorial-Guinea, and the second between Equatorial-Guinea and Nigeria.

Phataginus tetradactyla F_{ST} estimates suggest large differentiation between countries in the western range cluster (Ghana, Guinea, Liberia, Ivory Coast) to that of the central range cluster (Cameroon, Gabon, Equatorial-Guinea; Figure 11b). The Republic of Congo (eastern cluster) is also differentiated from the former two, however the Democratic Republic of Congo does not display this pattern (Figure 11b) as a result of outlier individuals (unknown locality) linked to the central cluster supposedly being sampled from this country (Figure 3b&c). This is why Gabon and the Democratic Republic of Congo have the lowest pairwise F_{ST} estimates. TreeMix results confirm the three major clusters with Nigeria showing high levels of drift from its own cluster (Figure 11a). Migration is common between countries that are within clusters ($n=2$ edges) but also

between countries from different clusters (n=3 edges), all of which are stemming from the central cluster into the two clusters on either side of it. There was a large level of heterogeneity across estimates for the best number of edges.

Liberia is the most differentiated from other countries for *S. gigantea*, whereby Uganda and Liberia show the highest level of pairwise differentiation (Figure 11c). Three major clusters are confirmed through the TreeMix phylogeny, however the Democratic Republic of Congo (majority of samples making up the fourth) does differentiate from Uganda and Rwanda (Figure 11c). There are a number of migration events (n=5), two of which are probably misidentified as they stem from a single sample in Liberia into the eastern part of the species range.

For *S. temminckii* F_{ST} estimates, Sudan and South Sudan are largely differentiated from the other countries with estimates higher than most across all species (Figure 11d). Individuals from South Africa and Namibia have the least amount of differentiation. TreeMix confirms the three major clusters, except that Tanzania is linked to Namibia rather than Ethiopia (Figure 11d). Instead, there is a high migration weight between Ethiopia and Tanzania suggesting a strong link.

Pairwise F_{ST} estimates per country/region for *M. javanica* and *M. crassicaudata* are on average, lower than the *M. culionensis* and *M. pentadactyla* estimates (Figure 12).

Population structure is less clearly identified through F_{ST} estimates for *M. javanica*. However, in general, estimates are closer to zero the closer the countries/islands are from each other (Figure 12a). Individuals from Yunnan-China show the highest level of differentiation when compared with other countries/islands. Islands of Sumatra and Java have lower F_{ST} estimates between them than between them and Borneo Island. Contrast to that, TreeMix phylogeny suggests that individuals from Java Island are sister to Borneo rather than to Sumatra (Figure 12a). This is in line with ADMIXTURE results (Figure 7b & c). Myanmar and Yunnan-China are largely differentiated from the rest of the countries/islands that fall further south. One migration event is suggested to have occurred from the branch connecting the eastern cluster (Myanmar and Yunnan-China) and all South-East Asian countries/islands, to Thailand.

F_{ST} estimates show a clear east-west pattern of differentiation with Taiwan and India having the highest estimates between them (Figure 12b). There was no differentiation between Vietnam and Thailand individuals and limited differentiation between Taiwan and China. TreeMix results indicate high levels of migration ($n=5$), however there was a large level of heterogeneity across estimates for the best number of edges, probably due to low sample sizes and admixed individuals.

We found below zero F_{ST} values between India and Sri Lanka for *M. crassicaudata* (Figure 12c). Despite their separation geographically as seen with PCAs (Figure 8a), the two known locations in India were sister two each other with TreeMix (as seen with ADMIXTURE). Migration was evident between Sri Lankan and Indian individuals.

High F_{ST} estimates are suggested between Palawan and the two smaller islands of Boayan

and Busuanga, the former comparison having the highest estimate across all pairwise comparisons of all eight species (Figure 12d). Despite multiple attempts, OptM could not determine the most likely number of edges for these three islands (probably due to there only being three populations to test, two of which consisted of one sample), therefore we did not include any in the phylogeny. The phylogeny suggests high levels of standard error in tree drift estimates; nonetheless Palawan individuals were more closely related to the closer island of Boayan, contrasting F_{ST} estimates.

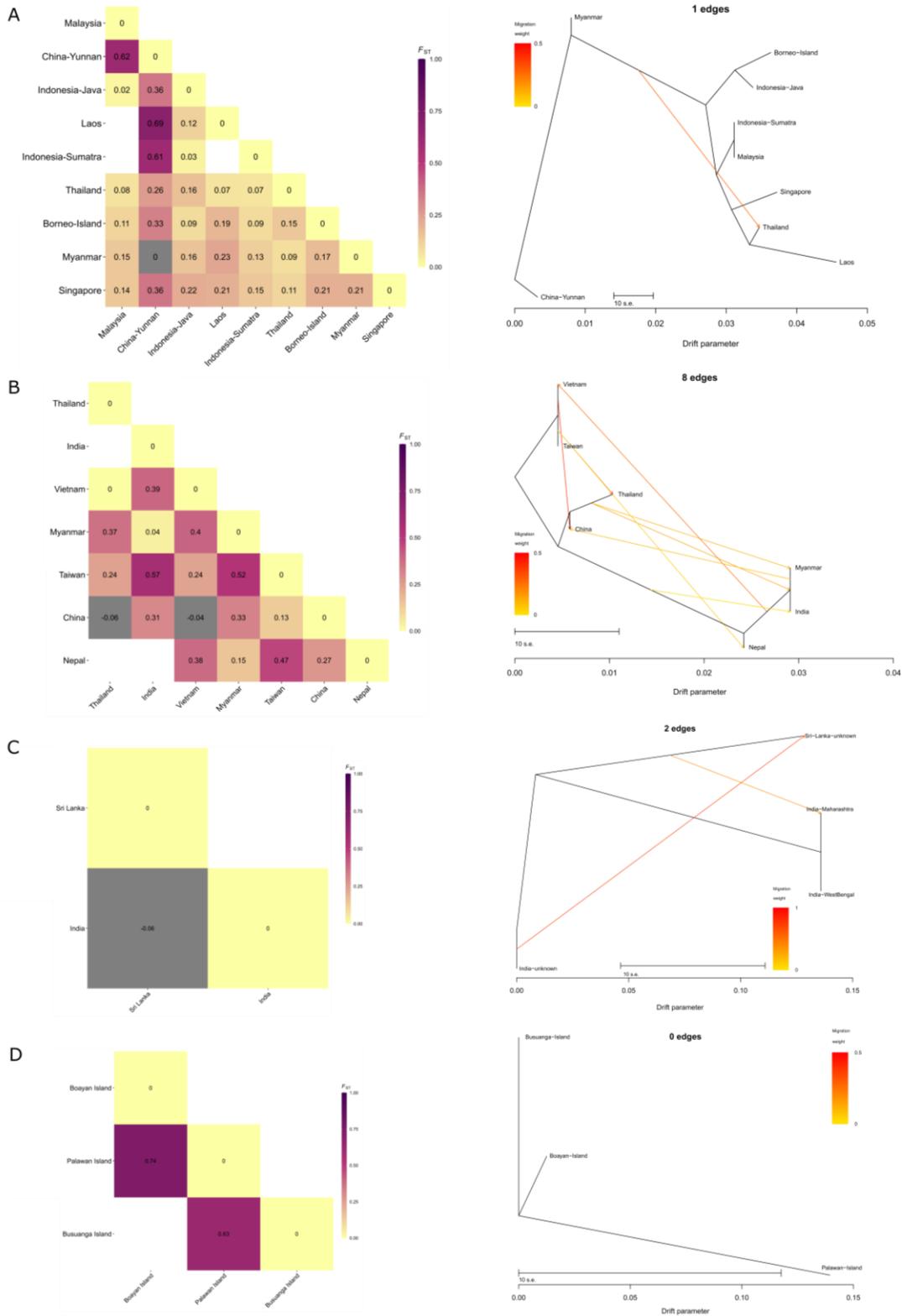


Figure 12: F_{ST} heatmap (left) and TreeMix maximum likelihood tree with most likely number of migration edges (right) for Asian species. (A) *Manis javanica*, (B) *Manis pentadactyla*, (C) *Manis crassicaudata*, and (D) *Manis culionensis*.

Population genetic estimates

Diversity estimates (A_e : effective number of alleles, H_o : observed heterozygosity, H_E : expected heterozygosity, and F_{IS} : Wright's inbreeding coefficient) varied dramatically across species and clusters ($A_{E\text{mean:range}}=1.213:0.707$, $H_{E\text{mean:range}}=0.143:0.375$, $H_{O\text{mean:range}}=0.197:0.681$, and $F_{IS\text{mean:range}}=-0.166:0.415$) with all clusters having a larger observed heterozygosity than what would be expected with random mating (hence the negative F_{IS} values; Table 1). All clusters of *M. crassicaudata* had the highest level of genetic diversity while *P. tricuspis* clusters from Dahomey Gap and western species range had the lowest, followed by the *M. javanica* western species range cluster (A_E & H_O ; Table 1). The islands of Busuanga and Boayan (Philippines) for *M. culionensis* had the highest inbreeding coefficient ($F_{IS}=0$, $n=2$), despite this still being seen as random mating. No private alleles were found for some *M. crassicaudata* clusters and the western species range cluster of *S. gigantea* (due to it containing only one individual). The clusters with the largest amount of private alleles per species after accounting for sampling effort were as follows: *S. temminckii* (Sudanese $A_{p/\text{sample}}=291.33$, $n=3$), *P. tricuspis* (eastern species range $A_{p/\text{sample}}=162$, $n=40$), *P. tetradactyla* (western $A_{p/\text{sample}}=117.4$, $n=5$), *M. pentadactyla* (western $A_{p/\text{sample}}=106.43$, $n=7$), *M. javanica* (Borneo Cluster $A_{p/\text{sample}}=80.37$, $n=49$), *S. gigantea* (central $A_{p/\text{sample}}=46$, $n=5$), and *M. culionensis* (Palawan $A_{p/\text{sample}}=41$, $n=3$).

3. Population genomics

Table 1: Population genomic parameters for all eight pangolin species. These included the number of alleles (A), effective number of alleles (A_E), observed heterozygosity (H_O), expected heterozygosity (H_E), Wright's inbreeding coefficient (F_{IS}), and number of private alleles (A_p), along with their respective standard errors (between individual variability). Cluster names are based on the area of the species distributions.

Species	Clusters	Sample number	A_p	A_p/Sample	A	A_e	H_e	H_o	F_{IS}	ASE	AeSE	HeSE	HoSE	FisSE
<i>P. tricuspis</i>	Eastern	40	6508	162.7	1.361	1.061	0.047	0.054	-0.072	0.005	0.001	0.001	0.001	0.002
<i>P. tricuspis</i>	Western Central - green cluster	75	154	2.053333	1.411	1.054	0.042	0.048	-0.061	0.005	0.001	0.001	0.001	0.002
<i>P. tricuspis</i>	Western Central - light blue cluster	49	232	4.734694	1.322	1.053	0.041	0.047	-0.073	0.005	0.001	0.001	0.001	0.002
<i>P. tricuspis</i>	Dahomey-Gap	50	4657	93.14	1.151	1.031	0.022	0.028	-0.105	0.004	0.001	0.001	0.001	0.005
<i>P. tricuspis</i>	Eastern Central	14	860	61.42857	1.281	1.076	0.058	0.067	-0.096	0.005	0.002	0.001	0.002	0.003
<i>P. tricuspis</i>	Western Africa	23	2929	127.3478	1.154	1.039	0.029	0.035	-0.113	0.004	0.001	0.001	0.001	0.004
<i>P. tricuspis</i>	Western Central - purple cluster	68	163	2.397059	1.404	1.056	0.044	0.05	-0.063	0.005	0.001	0.001	0.001	0.002
<i>P. tricuspis</i>	Western Central - turquoise cluster	39	88	2.25641	1.316	1.054	0.041	0.048	-0.073	0.005	0.001	0.001	0.001	0.002
<i>P. tetradactyla</i>	Central	12	1246	103.8333	1.625	1.301	0.2	0.265	-0.23	0.017	0.012	0.007	0.01	0.011
<i>P. tetradactyla</i>	Eastern	4	250	62.5	1.382	1.27	0.185	0.245	-0.261	0.018	0.013	0.009	0.012	0.016
<i>P. tetradactyla</i>	Western	5	587	117.4	1.301	1.183	0.123	0.165	-0.261	0.016	0.011	0.007	0.01	0.018
<i>S. gigantea</i>	Central	5	233	46.6	1.544	1.359	0.246	0.322	-0.236	0.019	0.014	0.009	0.013	0.012
<i>S. gigantea</i>	DRC	5	84	16.8	1.634	1.362	0.248	0.321	-0.222	0.017	0.013	0.008	0.012	0.011
<i>S. gigantea</i>	Eastern	4	37	9.25	1.5	1.316	0.226	0.282	-0.194	0.019	0.013	0.009	0.012	0.012
<i>S. gigantea</i>	Western	1	0	0	0.964	1.16	0.133	0.16	0	0.02	0.015	0.012	0.015	0
<i>S. temminckii</i>	Eastern	8	65	8.125	1.558	1.254	0.176	0.224	-0.187	0.013	0.008	0.005	0.007	0.008
<i>S. temminckii</i>	Sudanese	3	874	291.3333	1.132	1.116	0.078	0.108	-0.299	0.01	0.008	0.005	0.007	0.019
<i>S. temminckii</i>	South-Western	34	1599	47.02941	1.75	1.291	0.192	0.257	-0.22	0.011	0.008	0.004	0.007	0.007
<i>M. javanica</i>	Borneo	49	3938	80.36735	1.614	1.123	0.092	0.109	-0.094	0.008	0.003	0.002	0.002	0.002
<i>M. javanica</i>	Java	19	891	46.89474	1.34	1.089	0.066	0.079	-0.112	0.007	0.003	0.002	0.002	0.004
<i>M. javanica</i>	Western	21	271	12.90476	1.147	1.051	0.035	0.046	-0.176	0.006	0.003	0.002	0.002	0.01
<i>M. javanica</i>	Sumatra-Thai-Malay Peninsula	41	1428	34.82927	1.574	1.113	0.085	0.101	-0.092	0.008	0.003	0.002	0.002	0.003

3. Population genomics

<i>M. pentadactyla</i>	Central	13	265	20.38462	1.522	1.245	0.16	0.22	-0.249	0.017	0.011	0.007	0.01	0.013
<i>M. pentadactyla</i>	Taiwanese	14	152	10.85714	1.316	1.194	0.115	0.18	-0.415	0.016	0.012	0.007	0.011	0.022
<i>M. pentadactyla</i>	Western	7	745	106.4286	1.358	1.217	0.144	0.198	-0.276	0.017	0.012	0.008	0.011	0.017
<i>M. crassicaudata</i>	North-Eastern	2	0	0	1.247	1.581	0.356	0.556	-0.162	0.099	0.058	0.045	0.059	0.039
<i>M. crassicaudata</i>	Sri Lanka	3	0	0	0.89	1.529	0.277	0.52	-0.109	0.101	0.076	0.05	0.075	0.041
<i>M. crassicaudata</i>	Unknown	2	5	2.5	1.548	1.493	0.379	0.479	-0.275	0.059	0.055	0.043	0.055	0.04
<i>M. crassicaudata</i>	Western	2	0	0	1.397	1.738	0.397	0.709	-0.319	0.101	0.048	0.036	0.05	0.035
<i>M. culionensis</i>	<i>satellite</i> Islands	2	45	22.5	0.923	1.123	0.09	0.113	0	0.093	0.053	0.041	0.05	0
<i>M. culionensis</i>	Palawan Island	3	123	41	1.436	1.328	0.243	0.299	-0.201	0.08	0.064	0.046	0.061	0.063

Species genetic distances and isolation by distance (IBD)

Pairwise Euclidean distances of individuals were variable across species, where *P. tetradactyla* had the highest median distances and *M. culionensis* had the lowest (Supplementary Fig 14). The bar graph of *P. tricuspis* designates five distinct peaks, which may be associated to the five clusters (Supplementary Fig 14a). *Smutsia temminckii* and *M. javanica* have distinct clusters of outliers (Supplementary Fig 14d&e).

Using the Mantel test, geographic distance and genetic divergence were significantly, positively correlated for all species, except *M. culionensis* ($p > 0.05$). *Smutsia gigantea* exhibited the strongest correlation ($r=0.813\pm 0.075$) followed by *P. tricuspis* ($r=0.731\pm 0.01$), *S. temminckii* ($r=0.711\pm 0.018$) and *M. javanica* ($r=0.680\pm 0.011$; Supplementary Figure 15). *Manis crassicaudata* had the lowest, significant correlation ($r=0.290\pm 0.024$) followed by *M. pentadactyla* ($r=0.508\pm 0.01$) and *P. tetradactyla* ($r=0.584\pm 0.013$). A clear stratification in pairwise measures is visible for *S. temminckii* due to the Sudanese cluster (Supplementary Fig 15).

Discussion

To our knowledge, this is the first population genomics study on all eight species of pangolins, with previous studies only focusing on *M. pentadactyla* and *M. javanica* (Hu et al. 2020a; Nash et al. 2018; Wang et al. 2022; Wei et al. 2022). Nonetheless, sample counts for *M. crassicaudata* and *M. culionensis* are limited in this study, which resulted in limited inferences for both species. *Manis culionensis* has a restricted range being found only in Palawan (Philippines) and surrounding islands, whilst *M. crassicaudata* is restricted to a few countries in the Indian sub-continent (Mahmood et al. 2019; Schoppe et al. 2019) whereby India (largest portion of the species range) has limitations in sample exchanges.

Gene-capture design and efficiency

Gene-capture is becoming an increasingly important tool in genomics (White et al. 2019). Using this approach, we were able to readily include an additional 72 samples from previously published articles (Cao et al. 2021; Choo et al. 2016; Hu et al. 2020a; Wang et al. 2022) by reducing large amounts of genomic data to the target regions of interest (Karin et al. 2019). The majority of these genomes were from seizures of *M. pentadactyla* and *M. javanica*, which firstly permitted us to gather further insight into each species but also to gain access to often-inaccessible seizures from major consumer and trade hotspots. This reduced the unnecessary doubling of resources, procurement of required sampling permits, and processing times of laboratory work and bioinformatics analyses if whole-genomes would have been used.

The method we employed for designing baits allowed for efficient gene capture, regardless of sample type (museum vs. fresh). The higher capture efficiency for museum samples is likely a by-product of the more stringent protocols of library preparation and capture. It has been noted that museum samples showed higher capture efficiency at a single hybridisation temperature, whereas fresh samples work better with a touchdown approach (Paijmans et al. 2016), a method we did not employ in this study. Additionally, there is likely to be a large proportion of fresh samples that are degraded and contaminated, particularly those sourced from long-range markets. Despite deep divergences of around 40 million years ago (mya) in Pholidota (Heighton et al unpublished - Chapter 2; Gaubert et al. 2018), capture efficiency was not affected by target species. Evolutionary distance between references for bait design and species targeted can influence capture efficiency (Portik et al. 2016; Schott et al. 2017). These cannot be too distantly related, as seen with less than 5% of CDSs captured for gnathostome vertebrates of divergences of above 200 mya (0% for 298.6 mya) using standard protocols (Li et al. 2013a). A 10% divergence related to a

90% capture completeness (measured by percent of targeted region covered by at least one read) and 20% divergence to 85% completeness for 16 squamate reptile species (Schott et al. 2017). We employed a similar method to Schott et al. (2017) regarding the manual curation of exons of interest rather than bulk download, and a multi-reference species approach to bait design and mapping of raw reads.

Our capture efficiency (18.78%), the percentage of on-target reads, was lower than that found in a study on chimpanzee (*Pan troglodytes schweinfurthii*) faecal samples (67.1%) that were designed off the human genome (diverged around 13mya: White et al. 2019). Capture efficiency was also lower than that found on Anura in general (47.3%), but closer to what was found for the outgroup genera of that study (Microhylidae: 23%) that diverged from the target species around 103mya (Portik et al. 2016). Therefore, the large difference in capture efficiency may be explained by the deep divergence of Pholidota, but also the lower quality and completeness of the reference genomes and a significant proportion of bushmeat (high levels of exogenous DNA) samples in our study (Paijmans et al. 2016; White et al. 2019). Our study was also designed to target rapidly evolving genes and regions (intronic and inter-genic) rather than only conserved exonic regions (Portik et al. 2016; White et al. 2019) which may influence capture efficiency (Karin et al. 2019). This could explain why all clusters had a larger observed heterozygosity than what would be expected with random mating (Zhan et al. 2015). This is in line with our estimates of heterozygosity being higher than those previously published with whole genome sequencing (WGS) for the same populations of *M. javanica* and *M. pentadactyla* (Hu et al. 2020a; Wang et al. 2022). Other causes of this heterozygosity excess could be the (i) size of the geographical range of these clusters that prevents HWE (finer population structure within each cluster), (ii) these loci are under selection despite our filtering efforts (Beaumont and Nichols 1996), (iii) samples are of

a genetic-time series across many generations (failing HWE), or (iv) there have been major declines in effective population size (Cornuet and Luikart 1996). The latter is of importance to the conservation of pangolins globally, which does fit with the declines in PSMCs for all species (Heughton et al. unpublished – Chapter 2). Additional analyses on inbreeding indices, including runs of homozygosity from whole-genomes, will be required to further determine the underlying causes.

We included loci of potentially adaptive importance as well as were able to capture the mitochondrial genomes as a by-product of gene capture sequencing. These regions were not fully utilised in our study due to the assumptions surrounding many of the population genomics analyses, however there is space for delving into the adaptive evolution of all eight species across geographic and climate scales (Hohenlohe et al. 2021). From further research on our data, we will be able to identify genes or variants responsible for inbreeding depression or adaptation to altering environments. This can be useful for determining the adaptive capacity of populations in response to environmental changes, but also provide practitioners with the information on how to manage for adaptive variation and inbreeding (Hohenlohe et al. 2021). Mitochondrial DNA, in combination with our nDNA insights, will provide stronger evidence for population delimitation indices.

Potential novel/undescribed species

Through a population based PCA of South-East Asian pangolins, we add further evidence to the possibility of a previously undescribed *Manis* species using a previously published genome (Heughton et al. unpublished - Chapter 2; Cao et al. 2021; Hu et al. 2020b). We, however, were unable to find additional samples linked to this novel species that could lead us to determining its distribution as all samples associated with this species have been obtained solely from seizures in

Sichuan and Hong-King in China from previous studies (Heighton et al. unpublished - Chapter 2; Cao et al. 2021; Hu et al. 2020b). In order to identify this species' likely range, we recommend obtaining geographically referenced samples in areas/regions that we have not yet covered. Major gaps in our sampling effort (as well as those of previously published) that are within the South-East Asian species ranges are Cambodia, parts of Myanmar, and smaller islands between the major islands of Malaysia and Indonesia (i.e. Riau and Bali Islands). China (particularly central), Bangladesh, India, and Pakistan may also be viable areas to search, despite their distance from South-East Asia.

Smutsia temminckii samples from Sudan and South Sudan (Sudanese cluster) are considerably differentiated from other *S. temminckii* populations through multiple lines of evidence (population differentiation indices with TreeMix, PCA, private alleles, IBD, F_{ST} and ADMIXTURE). This region has been suggested to harbour a species, *Phatages hedenborgii*, originally said to be distinct from *S. temminckii* (Fitzinger 1872), but after which has been described as a synonym (Allen 1939; Jentink 1882). This is despite a lack of evidence and reasoning for it being synonymised as authors discussions revolved around a generalised and basic morphological similarity as the main argument. Through our study, clear differentiation in their genetics provides insight against a synonym classification. Especially since we do not only have the genomic data for the type-synonym (*P. hedenborgii*; Sudan), but we also have genomic data of 33 samples around 100km from Kuruman (South Africa), the type-locality of *S. temminckii* (Heath 1992). Morphologically, *P. hedenborgii* is suggested to differ from *S. temminckii* predominantly by a much longer tail and a lower number of longitudinal rows of dorsal scales; 11 instead of 13 (Fitzinger 1872). Fitzinger (1872) indicated that the type-synonym was obtained in Sennaar (Sudan) but provided evidence of possible specimens from other parts of south-eastern

Sudan (Bayuda Desert, Kordofan, White Nile) and Taka of Somalia. Our results confirm Sudan and South Sudan as range states for the cluster, but disagree with the Somali location as a sample from Ethiopia (in between Somalia and Sudan) is affiliated to *S. temminckii* and there is currently no pangolins found in Somalia (Pietersen et al. 2019). We believe this previously described taxon may occupy the northern portion of the current *S. temminckii* range (Pietersen et al. 2019) which extends northeast of Ethiopia through South Sudan, Sudan, northern Central African Republic, and Chad. A likely large geographical barrier separating the two taxa is the White Nile, particularly the 40,000–90,000 km² Sudd Wetland of South Sudan. The river and its wetland has been suggested as a key geographical barrier of gene-flow for giraffes (Hassanin et al. 2007), so much so that the two subspecies of the northern giraffe, Kordofan (*Giraffa camelopardalis antiquorum*) and Nubian (*G. c. camelopardalis*), lie on either side of this barrier (Coimbra et al. 2021; Fennessy et al. 2016).

There are many instances in our study that show estimates of pangolin population differentiation and distributions of population clusters that mirror that of species and subspecies boundaries for felids (Luo et al. 2014), ungulates (Garcia-Erill et al. 2022; Lorenzen et al. 2012), and primates (Telfer et al. 2003). These higher taxonomic delimitations were already noted as a possibility for two clusters of *M. pentadactyla* across China (Wang et al. 2022), the cryptic diversity of *M. javanica* in seizures (Zhang et al. 2015b) and the six cryptic lineages defined for *P. tricuspis* (Gaubert et al. 2016).

Before we can conclude that these are indeed novel or redefined species or subspecies, a thorough taxonomic investigation is required. Robust phylogeny and species delimitation testing with nuclear genes as well as mitochondrial genomes captured in our study will provide genetic evidence (Carstens et al. 2013). Combining this with morphological analyses of 3D morphometrics

of skulls (Ferreira-Cardoso et al. 2019) could provide sufficient evidence for these claims.

Population delimitation – building the database for conservation management

Clear population delimitation is an essential precursor to better manage pangolin populations as distinct conservation units and trace the illegal trade (Allendorf et al. 2010; Hohenlohe et al. 2021). Delimitation of six cryptic lineages of *P. tricuspis* by Gaubert et al. (2016) provided the necessary geographic database to which individuals from seizures could be traced (Ewart et al. 2021; Zhang et al. 2020b). Here we provide the genetic clusters of each species through multiple lines of evidence and the discussion points surrounding their anomalies and genetic characteristics. The naming of these clusters follow, to some extent, the biogeographic regions of Africa determined by Droissart et al. (2018) and Asia determined by Woodruff and Turner (2009). If biogeographic regions could not be linked to clusters, naming followed the part of the species range (i.e. eastern) or the country/island that it was found in.

The three western to central African species that have similar ranges (*P. tricuspis*, *P. tetradactyla* and *S. gigantea*) have the same three major clusters across a number of lines of evidence (TreeMix, PCA, F_{ST} and ADMIXTURE), apart from ADMIXTURE for *S. gigantea*. These are defined as Western Africa (Upper Guinea bioregion: Guinea to Ghana), Central-Western Africa (Lower Guinea bioregion: Nigeria to Gabon), and Central-Eastern Africa (Congolia bioregion: Central African Republic to Rwanda). This suggests that these species were likely influenced by the same historical changes and biogeographical barriers, as is often the case in comparative phylogeography for species sharing similar habitats and modes of life (Lorenzen et al. 2012; Luo et al. 2014). Sampling gaps for *P. tetradactyla* and *S. gigantea* (especially in West Africa) would need to be filled and divergence time estimates would need to be conducted before

these can be fully tested.

White-bellied pangolin - *Phataginus tricuspis*

We recommend three major and five minor clusters:

- Upper Guinea Major (western species range)
 1. Upper Guinea (far west species range) – Guinea, Sierra Leone, Liberia, Ivory Coast, Ghana. Grouping Western Africa and Ghana lineages (Gaubert et al. 2016).
 2. Dahomey Gap – Togo, Benin, Nigeria (west of Niger River). Originally Dahomey Gap lineage (Gaubert et al. 2016).
- Lower Guinea Major (central species range)
 3. Lower Guinea – Nigeria (east of Niger River), Cameroon, Equatorial Guinea, Gabon (northern). Originally Western Central Africa lineage (Gaubert et al. 2016).
- Congolia Major (eastern species range)
 4. Western Congolia – Gabon (central & southern), Republic of Congo (northern & western), Central African Republic. Originally Gabon lineage (Gaubert et al. 2016).
 5. Eastern Congolia – Republic of Congo (southern & eastern), Democratic Republic of Congo, Uganda. Originally Central Africa lineage (Gaubert et al. 2016).

Major clustering was supported at $K=3$ whereby the species range can be split into three major sections at central Nigeria (Upper Guinea Major and Lower Guinea Major), and central-western Gabon through to Central African Republic (Lower Guinea Major and Congolia). Minor clustering was similar to that of the evolutionary significant units by Gaubert et al. (2016) with mitogenomes, however we recommend five clusters instead of six and suggest some changes in the positioning of certain clusters. We identified a new cluster called Western Congolia that includes Gabon, Central African Republic and Republic of Congo. This has resulted in the renaming of the Central African lineage (Gaubert et al. 2016) as the Eastern Congolia cluster, especially with the inclusion of samples from Uganda (eastern part of the species range). Both ADMIXTURE and PCA confirm

that Gabon lineage is no longer a separate cluster but rather split between Lower Guinea (northern part of country) and this new Western Congolia cluster (the rest of Gabon grouped with the Republic of Congo and Central African Republic). The ambiguity of the Gabonese lineage was foreshadowed by Gaubert et al. (2016) due to sampling gaps, which we can now confirm. As with Gaubert et al. (2016), we cannot easily explain a biogeographic scenario that would separate this cluster with Lower Guinea at the point of contact in northern Gabon. The Ogooué River may have restricted gene flow between these clusters as it did in mandrills (*Mandrillus sphinx*) and gorillas (Gorilla gorilla: Anthony et al. 2007; Telfer et al. 2003), however we do not find clear geographical differentiation or admixture at the zone of contact (the river). We therefore agree with previous authors that putative Pleistocene forest refugia in southern (Massif du Chaillu and Monts Doudou) and northern (Monts de Cristal) Gabon (Dauby et al. 2014), in combination with secondary dispersal from central Africa, may have influenced population structuring (Gaubert et al. 2016; Hardy et al. 2013; Jacquet et al. 2014). This contact area may be an important hybridization zone between the two clusters, suggesting high levels of admixture (Bonnald et al. 2021), however we did not find this pattern of admixture in our study, leaving the hypothesis of reproductive incompatibility (post refugia) or sympatric speciation as a future research prospect.

Although we identified a Ghanaian outlier with the PCA, complementary analyses do not concur and most samples from Ghana group with the Upper Guinea cluster, suggesting that the Ghana lineage is likely no longer a separate cluster (Gaubert et al. 2016). However, we do find evidence of some separation of the Ghana cluster ($K=10$), which is further evidenced in F_{ST} results as Ghana was the most differentiated across Upper Guinea. At $K=10$, the Ghanaian cluster is however also grouped with the admixed Ivory Coast population (Supplementary Fig 8), implying admixture between what was thought to be two previously defined lineages (Gaubert et al. 2016).

We hypothesize that the Sassandra River may be an influential barrier of gene flow, however not enough to result in significant differentiation of clusters (Jacquet et al. 2014).

Through ADMIXTURE, Cameroon was the zone in which four clusters were identified. These were not easily distinguishable across a geographical scale, apart from two clusters that could be separated by the Sanaga River. This river has mixed evidence of being a significant biogeographic barrier probably as a result of it being a putative Pleistocene forest refuge which may have allowed for gene-flow on either side of the banks (Dauby et al. 2014; Hardy et al. 2013), however from $K=4$ onwards there was seemingly evidence of its effect. Other population stratification analyses did not confirm these clusters, but rather one single cluster as previously described (Gaubert et al. 2016). We therefore believe that these clusters are likely an artefact of high sample numbers in Cameroon compared to the rest of the species range (273 out of 420), thus influencing ADMIXTURE analyses (Lawson et al. 2018). We also cannot rule out the influence of IBD on these results within this cluster (Perez et al. 2018).

Suggested migrations between countries are directly associated with the countries that contained individuals in multiple minor clusters, like Nigeria and Gabon with Lower Guinea cluster. These are therefore likely not true migration events but rather artefacts of partitioning by country for TreeMix analyses. However, countries can form an easy priori which is often used for testing population genetics estimates (see: Garcia-Erill et al. 2022).

Black-bellied pangolin – *Phataginus tetradactyla*

We recommend three clusters:

1. Upper Guinea (western species range) – Guinea, Liberia, Ivory Coast, Ghana.
2. Lower Guinea (central species range) – Nigeria, Cameroon, Equatorial Guinea, Gabon.
3. Congolia (eastern species range) – Republic of Congo, Democratic Republic of Congo.

These clusters fit with the major clusters found in *P. tricuspis*, an expected finding considering their similar life history traits of being small and arboreal with a close association to the tropical forest biome. Two individuals from the Democratic Republic of Congo seemingly fit with the Lower Guinea cluster. The individual in the southwest of the country exhibits some level of admixture, likely suggesting that this region is between the two described clusters (Lower Guinea and Congolia). The second, which was placed in the centre of the country because of its unknown sampling locality, is likely stemming from a region near southwest of the country or was mislabelled by museum collectors due to the countries synonymous name with the Republic of Congo.

Significant levels of drift for the Nigerian samples (within the Niger River delta) compared to other countries of its own cluster (Lower Guinea) indicates that the Sanaga River (Cameroon) or the delta itself could be important geographic barriers. Both have been shown to be important for structuring in *P. tricuspis*, the Sanaga River having less of an influence. Evidence of migration between Nigeria and the Congolia cluster confirms a similar possibility for the latter for *P. tetradactyla*. Recent studies have indicated the importance of the Niger River for both African lion (*Panthera leo*) and buff-spotted woodpecker (*Campethera nivosus*) population structuring (Bertola et al. 2016; Fuchs and Bowie 2015).

Migration was frequent for the species, with the Lower Guinea cluster seemingly being a source of migration leading into the two clusters on either side. Migration between Lower Guinea and Congolia is more plausible with the continuation of the forest belt, even though the last glacial maximum rainforest refugia (Gaubert et al. 2016; Hardy et al. 2013). Between Upper Guinea (west) and the other two clusters, migration was less likely due to the Dahomey Gap splitting the forest and the species range into two. However, based on previous divergence estimates by Gaubert

et al. (2018), this supposed migration event may have occurred before the gap formed around 3.4–4.5 kya (Salzmann and Hoelzmann 2005). There was a significant amount of noise when choosing the most likely number of edges and so these results should be taken with caution, especially considering the limited dispersal of this fossorial species (Gudehus et al. 2020). This limited dispersal capability and small home ranges of the species is likely why it had the highest pairwise Euclidean distances compared to other species.

Giant pangolin – *Smutsia gigantea*

We recommend four clusters:

1. Upper Guinea (western species range) – Liberia.
2. Lower Guinea (central-western species range) – Cameroon, Equatorial Guinea, Central African Republic.
3. Congolia (central-eastern species range) – Democratic Republic of Congo.
4. Albertine Rift (eastern species range) – Democratic Republic of Congo (eastern outlier), Uganda, Rwanda.

The PCA revealed the same three major clusters as the previous two species, but ADMIXTURE suggests an additional cluster on the eastern part of the species range. The significant levels of drift for the samples from Democratic Republic of Congo provide further evidence for this separate cluster (Congolia). The Nile River and Great Lakes of eastern Africa (Albertine rift) are well known biogeographic barriers for vertebrates (Hassanin et al. 2007), which may support the Albertine Rift cluster separating from the Congolia cluster. This area is also known to be a hybridisation zone for African savannah (*Loxodonta africana*) and forest (*L. cyclotis*) elephants (Bonnald et al. 2021) likely as a result of the two biomes converging in this region as a forest-savannah transition zone known as the Lake Victoria Regional Mosaic (White 1979). This mosaic

may have added to the differentiation of these individuals from this region through local adaptation. With this region being a zone overlapping that of *S. temminckii*'s range (sister species), it could be a potential driver of population segregation of *S. gigantea* through introgression / hybridization, complicating its recognition as a conservation unit (Gagnaire et al. 2011; Taylor et al. 2020). However, additional studies on inter-species introgression at this zone would be required to provide evidence insight into this. One outlier individual of the Albertine Rift cluster that was sampled from a field site (Democratic Republic of Congo) was found to merge with the Congolia cluster at higher values of K, suggesting it may require further investigation. Particularly since there is no biographical evidence to support this Albertine Rift outlier at its exact location, nor is this structure found in other studies in this region (Stanton et al. 2014). However, forest refugia in the Great Lakes area may again have played a role, as noted for Gabon in *P. tricuspis*.

Substantial differentiation between samples from Liberia and Uganda follow an expected pattern of isolation by distance as both are on either side of the species range (Nixon et al. 2019). This, along with measures of stratification, contradicts suggested migration events between the Upper Guinea cluster of Liberia and other clusters. However, this cluster constitutes a single sample which can affect the population based phylogenetic assumptions of TreeMix (Pickrell and Pritchard 2012).

Temminck's pangolin – *Smutsia temminckii*

We recommend three clusters with a possibility of a fourth:

1. South Western – Namibia, South Africa
2. Central Eastern – Botswana, Zimbabwe, Ethiopia, Tanzania. Possible split between former and latter two countries.
3. Sudanese – South Sudan, Sudan.

The incongruence between PCA and ADMIXTURE in the partitioning of the two clusters, whereby the former suggests Botswana and Zimbabwe are grouped with the South Western instead of the Central Eastern could be due to the addition of SNPs under selection. Adaptive vs neutral loci are likely to differentiate depending on spatial patterns of gene flow, drift, and selection (Funk et al. 2012). It is encouraged to use multiple lines of evidence to determine conservation units that are more in line with management across evolutionary and adaptive scales (Funk et al. 2012). We therefore hypothesize that with additional samples across the range, there are likely to be Eastern and South Eastern clusters that split Central Eastern into two: Ethiopia and Tanzania; Botswana and Zimbabwe. Although IBD may be the causal factor, even with additional samples. The contact zone between these clusters is likely to be located around the Zambezi River and its Pleistocene flood plains, mirroring that of common warthog (*Phacochoerus africanus*) subspecies (Garcia-Erill et al. 2022; Lorenzen et al. 2012) and bushveld gerbil (*Gerbilliscus leucogaster*) populations (McDonough et al. 2015). However, we cannot dispute a potential segregating zone around central Tanzania, mirroring that of common eland (*Taurotragus oryx*) populations and sable (*Hippotragus niger*) lineages (Lorenzen et al. 2012).

The differentiation between South Western and Central Eastern clusters is confirmed by du Toit (2014) with mitogenomes of the species, whereby the authors correlated the timing of this split (South Africa/Namibia with Mozambique/Zimbabwe) with the Mega Kalahari Sand Sea phenomenon forming a barrier between the populations (Stokes et al. 1997). The Limpopo River may be forming a potential barrier as well, however there is limited evidence to support this river as a noteworthy barrier (Lorenzen et al. 2012).

TreeMix results indicating a grouping of Tanzania with Namibia rather than Ethiopia, can be influenced by the single samples of Ethiopia and Tanzania, thus influencing the strength of

support for the population based phylogeny (Pickrell and Pritchard 2012). However, there is evidence of substantial migration between Ethiopia and Tanzania in this phylogeny, suggesting some connection. Additional samples would provide further insight into this.

Sunda pangolin – *Manis javanica*

We recommend five clusters:

1. Borneo – Borneo Island (Sabah and Sarawak of Malaysia, Kalimantan of Indonesia, Brunei). As with previously published Borneo Cluster (Nash et al. 2018).
2. Java – Java Island. As with previously published Java Cluster (Nash et al. 2018).
3. Sumatra-Thai-Malay Peninsula – Sundaland north (Sumatra, Singapore, Peninsular Malaysia) and Indochina east (southern Thailand, Vietnam, Laos). As with previously published Sumatra-Singapore Cluster (Nash et al. 2018), and MJB1 and MJB2 (Hu et al. 2020a).
4. Indochina North – northern Thailand, southern Myanmar. Originally MJB3 (Hu et al. 2020a).
5. Western – Sino-Burmese border, Yunnan-China, Myanmar. Originally MJA1 and MJA2 (Hu et al. 2020a).

Country scale analyses were not feasible for this species based on the structuring of islands and samples from the same country falling on either side of the Isthmus of Kra (Thai-Malay Peninsula). With the exception of two samples from two markets in Laos (restricted-range) and Vietnam (long-range), the Isthmus of Kra is likely an important biogeographic barrier for *M. javanica*. This barrier, and the region in which it lies is well known for separation of Indochinese and Sundaic lineages and subspecies of fauna (de Bruyn et al. 2005; Inger and Voris 2001; Luo et al. 2014; Patel et al. 2017) through sea level rises during the Plio-Pleistocene that narrowed the piece of land between these regions (Woodruff and Turner 2009). If the two samples from Vietnam and Laos markets are to be true origins, this hypothesis does not hold, going against most literature

(Although see the leopard: Luo et al. 2014). Under this assumption, the Mekong River would likely have played an important role as a biogeographic barrier once there was a Sundaic population expansion across the Gulf of Thailand during the last glacial maximum (Bird et al. 2005).

Three clusters fit with that of Nash et al. (2018) whereby their Sumatra-Singapore cluster is expanded to the north through additional samples (Sumatra-Thai-Malay Peninsula – Malaysia Peninsula, southern Thailand, and possibly Vietnam and Laos). These fit with the clusters of MJB1 (likely Peninsular Malaysia and southern Thailand) and MJB2 (likely Vietnam and Laos) by Hu et al. (2020a), whereas MJB3 is more closely aligned to the Indochina North cluster (northern Thailand, southern Myanmar). The Western cluster fits with both the MJA1 and MJA2 clusters (Hu et al. 2020a), however we could not differentiate between these two like the authors could. Although we provide more samples of known localities for this species, thus providing better geographic assignment, we could lack the differentiating power found in WGS (Hu et al. 2020a).

As previously, we also identified a notable number of admixed individuals (Hu et al. 2020a; Nash et al. 2018), particularly those in the Sumatra-Thai-Malay Peninsula cluster that were seized in cities in western Java and the Indochina North cluster that were seized in Yunnan and Guangzhou. Western Java is said to be an important redistribution hub for songbirds that are collected across Indonesia (Chng et al. 2015). These seized individuals therefore could be originating from hybrid zones between the island clusters, such as satellite islands that were once linked during the LGM (Bird et al. 2005). The Indochina North cluster individuals seized in China (MJB3) are progressively admixed with individuals from the Western cluster, which could be a potential IBD gradient from southeast Indochina to northwest Myanmar. These levels of admixture and IBD patterns, along with within country structure across the Isthmus of Kra (Thailand and Myanmar) may explain the limited differential power of F_{ST} estimates by countries. Despite this,

Yunnan China was largely differentiated from the rest of the localities due to its extreme north location, isolation from other localities by rivers (drainage basin), and higher elevation (Meijaard and Groves 2006; Zhang et al. 2020a).

Individuals from the islands of Sumatra and Java are closely associated through F_{ST} estimates, contradicting that of close association between Java and Borneo in the ML phylogeny. The latter is a result of Sumatra forming a close association with other samples in its cluster from Peninsular Malaysia. Regardless, both these scenarios do not fit with the close association of Sumatra and Borneo for passerines, with Java often being the more divergent for species and population comparisons, probably as are result of its different climate and vegetation (Leonard et al. 2015). However, there are exceptions to these (the red spiny rat – *Maxomys surifer* and the Asian palm civet – *Paradoxurus hermaphroditus*) with Sumatra and Java being more closely associated through proximity (Leonard et al. 2015). Differentiation between these islands is therefore likely to follow these mammalian species examples.

The suggested migration event from base of the clusters in Indochina and South-East Asia to Thailand is likely related to the country being at the contact point between these clusters (Isthmus of Kra), harboring clusters from both sides of the barrier. This indicates the importance of correctly selecting genetic partitions for tests of introgression.

Chinese pangolin – *Manis pentadactyla*

We recommend three clusters but potentially four:

1. Western – India, Nepal, Myanmar. Originally MPB cluster (Hu et al. 2020a), CPC cluster and one individual (MP16) from the CPB cluster (Wang et al. 2022). There are potentially two separate clusters based on a seized sample at the Sino-Burmese border, China (Western, MPB, CPB) and two seized samples in Yunnan Province, China

- (Unknown cluster 1, MPB, CPB).
2. Central – China, Thailand, Vietnam. Originally MPA (Hu et al. 2020a), CPA and CPB clusters (Wang et al. 2022).
 3. Taiwanese – Taiwan, unknown locality.

With more samples of known origin, our study confirms and expands on the previously defined separation of western (Western cluster – MPB/CPC) and central (Central cluster – MPA/CPA/CPB) regions of the species range (Hu et al. 2020a; Wang et al. 2022). Although they could not confirm it, both studies suggested that the MPB (Hu et al. 2020a) and CPC (Wang et al. 2022) clusters were likely originating from Myanmar, which we can not only confirm but can extend it to Nepal and India. Hu et al. (2020a) also suggested that their MPA cluster originated from southern China and Thailand, which we can also confirm, but again can include Vietnam, and north-western and eastern provinces of China (Yunnan, Fujian, Zhejiang). The Taiwanese population (*M. pentadactyla pentadactyla*) is often considered to be a subspecies to that of the mainland (*M.p.aurita*) and Hainan Island (*M.p.pusila*) populations through basic morphological characteristics (Allen 1938; Allen and Owston 1906), and recently off a mitochondrial genome phylogeny of the former two populations (Sun et al. 2021a). We confirm the potential for a subspecies classification of *M.p.pentadactyla* and *M.p.aurita* that likely started to diverge around 1.48–4.95 mya (Heighton *et al.* unpublished – Chapter 2), but caution that the patterns of admixture and low F_{ST} estimates between these populations in this study may contradict said classification (see: Wang et al. 2022). The likely cause of this separation is the west-east isolation by distance that we identified for the species (hence some shared Taiwanese ancestry in eastern China), followed by subsequent and recent isolation of the island. We also provide no evidence of a monophyly for *M.p.pusila* as samples from Hainan Island clustered within the Central cluster due to its recent isolation from the mainland, as shown in other mammals (Balakrishnan et al. 2003;

Wenhua et al. 2012).

Our results correlated partially with that of Wang et al. (2022), whereby the authors suggest three clusters (CPC – probably Myanmar, CPB – Yunnan, western Guangdong and Taiwan, and CPA – eastern Guangdong) instead of the two we found. Wang et al. (2022) concluded that CPA is a new cluster originating from eastern Guangdong and separate from CPB and CPC [MPA and MPB of Hu et al. (2020a)], however, we group CPA and CPB with MPA as part of the Central Cluster. The authors suggested that the two clusters we merge (CPB and CPA), diverged roughly 2–4 kya and those that we separate (CPA/CPB and CPC) diverged around 25–40 kya (Wang et al. 2022). Our lower number of variants (targeted approach instead of WGS) may not be sufficient to separate clusters by a few thousand years. However anomalies in Wang et al. (2022) that contradict this study and Hu et al. (2020a) would need to be relooked at. These include the grouping of Taiwan to western Guangdong and Yunnan (CPB), instead of the more proximal eastern Guangdong (CPA), as well as a sample MP16 that they grouped with CPB (Central – MPA) instead of CPC (Western – MPB). It is difficult to identify the reasoning behind possible differences between these studies as the previous studies (Hu et al. 2020a; Wang et al. 2022) provided limited information on site type (field site or market locality) and no accurate geographic locations (only regional information was provided) for each sample.

We do additionally identify potential structure within these major clusters using the PCA, something not identified in previous studies (Hu et al. 2020a; Wang et al. 2022). We identified a potential cluster of unknown origin from seizures in Yunnan (Unknown cluster 1 separating under Western cluster). Two field site samples in Yunnan (Central) from Hu et al. (2020a) are suggested to be differentiated in our study through PCA, in line with their ADMIXTURE results at $K=3$, conversely an additional two samples from this province do not align with this separation, therefore

requiring additional sampling effort or sampling information from previous publication authors.

Indian pangolin – *Manis crassicaudata*

We do not recommend any clusters based on the lack of consensus between population stratification techniques, high levels of suggested migration and admixture, and low levels of population differentiation.

PCA was the only structuring method that could distinguish some of the populations. We found no private alleles within populations and the highest levels of genetic diversity for all pangolin species. The Indian subcontinent however does have examples of limited structure across species ranges within this region. Phylogeography of the jungle cat (*Felis chaus*) exhibited no convincing structure despite high genetic variation (in line with our diversity estimates) whereby the authors attributed this to population growth (range expansion) during the mid-Pleistocene (Mukherjee et al. 2010). Jana and Karanth (2022) too attributed population expansion to their lack of structure using multiple markers for the blackbuck (*Antelope cervicapra*). This species in our study does not have enough samples or high quality data (majority of samples were low coverage whole genome resequencing – c. 10x) to provide insightful recommendations, but range expansions could be playing a role.

Our results indicate no clear structuring, differentiation or migration between Sri Lankan and Indian samples fitting with the non-monophyly from previous phylogenetic studies (Heighton et al. unpublished - Chapter 2; Gaubert et al. 2018). This is likely as a result of the extensive continental exchanges across late quaternary climatic fluctuations (Ali 2018; Bossuyt et al. 2004).

Palawan pangolin – *Manis culionensis*

We recommend two clusters:

1. Palawan – Palawan Island (southwest of range).
2. *satellite* islands – Boayan Island, Busuanga Island (northeast of range).

Although we suggest two clusters, the PCA alluded to the potential separation of the *satellite* islands cluster into the two islands that it constitutes. Population differentiation (F_{ST}) between these islands were the highest amongst any other pairwise comparisons for pangolins in this study, indicating low levels of gene flow between these islands. These highly differentiated populations across islands spanning small geographical distances, and the low sample numbers, is why the species is the only one to not have a significant IBD curve (Hinten et al. 2003; Robinson et al. 2016). We hypothesize that the smaller islands were likely drivers of rapid population separation through inbreeding rather than a gradual differentiation due to lack of gene flow (Robinson et al. 2016), particularly since these islands were all connected during the recent Pleistocene glaciation events (15,000 and 13,000 yr: Voris 2000). This is confirmed by these islands having the highest number of private alleles for the species and the highest levels of inbreeding for pangolins, although the latter was still in the range of random breeding. The small size of Palawan, where the majority of samples were collected, could have resulted in pairwise Euclidean distances being the lowest for pangolins as a result of higher inbreeding and limited structure (Hinten et al. 2003; Robinson et al. 2016). Long-term small population size and isolation can result in genetic load due to the accumulation and fixation of deleterious variants (Robinson et al. 2016). With the current critically endangered status of the species, more proactive management plans are likely to follow. We therefore suggest further analyses on the adaptive capacity of these small and isolated satellite island populations as well as the potential problems that may arise through trans-island accidental

migration or genetic rescue programs (outbreeding depression) in the future (Hohenlohe et al. 2021).

ML Phylogeny (TreeMix) suggested the proximate island as the sister island to Palawan individuals, despite a higher F_{ST} estimate. However, we suggest caution with our results as standard errors were high for this analysis. Additional, and fresh, samples across these islands are required for more informed insights.

Major drivers for population structure

We found some major water bodies and Pleistocene climatic refugia rather than shifts in elevation, eco-regions or biomes to seemingly have had an effect on population structuring for pangolins.

Large rivers are well documented biogeographic barriers across various species of fauna (Anthony et al. 2007; Fuchs and Bowie 2015; Telfer et al. 2003), leading to restricted gene flow and eventually population differentiation and allopatric speciation (riverine barrier hypothesis: Wallace 1854). Despite them having an affinity to water and being able swimmers that can traverse up to 40m wide rivers (Pagès 1970; Yang et al. 2007), we found examples of the influence of rivers and lakes across all African pangolin species and *M. pentadactyla*. Rivers and lakes that seemed to have had an influence on one or multiple species were the Niger River, Lake Volta, Congo River, White Nile (Sudd Wetland), and possibly the Great Lakes of Eastern Africa. Interestingly, different rivers had varying effects on structure likely as a result of historical changes in hydrology and water courses, and the age of river systems (Meijaard and Groves 2006). The size and dispersal ability of the different species may have also played a role in species-specific effects of rivers (Harcourt and Wood 2012). Similarly, but less unexpected, we found geographic structuring across South East Asian islands for *M. javanica* and *M. culionensis*, despite reports of the former species

ability to swim between proximate islands (Chong et al. 2020). Therefore, although dispersal is probable across these bodies of water, they may have acted as a significant enough barrier for pangolin populations. Nonetheless, additional sampling efforts on either side of the banks of these major rivers for formal hypothesis testing and more in-depth knowledge of their hydrological histories, are required to confirm these results.

Pleistocene (2.5mya – 11.7kya) climatic fluctuations fits with estimates of divergence times for intra-species populations or lineages of pangolins in Africa and Asia (Heighton et al. Unpublished - Chapter 2: Gaubert et al. 2018; Hu et al. 2020a). These fluctuations not only influenced island connectivity in Sundaland (*M. javanica*) through oscillations in sea level (Leonard et al. 2015), but also the contraction and expansion of flora which resulted in the formation of forest and savannah refugia across all species ranges (Pleistocene refuge hypothesis: Haffer 1969). Gaubert et al. (2016) concluded that forest refugia was likely a major driver for *P. tricuspis* population structuring.

With additional samples, species and markers, we concur that these refugia (forest and savannah) probably played a role, but likely in tandem with some major bodies of water acting as barriers in moister periods, once forests and savannahs reconnected neighboring refugia through their expansion. This superimposition of effects follows a mixed refugia-river barrier hypothesis that has been suggested for other species (Jackson and Austin 2010; Nicolas et al. 2011), but would require formal testing in the future for pangolins. Potential exceptions to this hypothesis include the Dahomey Gap for *P. tricuspis* and the Lake Victoria Regional Mosaic for *S. gigantea*, both of which are shifts in biomes across the species range. The Dahomey Gap is a savannah corridor that started to establish during the drying period of the Holocene around 3.4–4.5 kya (Salzmann and Hoelzmann 2005). This has been an important biographic barrier for fauna and flora across the

Western tropical forest block, separating Upper and Lower Guinea biogeographic regions (Demenou et al. 2018; Demenou et al. 2016; Fuchs and Bowie 2015). The Lake Victoria Regional Mosaic is considered a transition zone between savannah and forest biomes, where Pleistocene climatic oscillations resulted in shifts in these biomes and the species related to them (Faith et al. 2016).

These forest-savannah mosaics show the lowest level diversity for *S. gigantea* (Eastern cluster) and for pangolins as a whole (*P. tricuspis* – Dahomey Gap). Previous studies on *P. tricuspis* show a similar trend for the species which could be linked to the reduction of its effective population size and founder effects of the Dahomey Gap lineage (Zanvo et al. 2022), possibly over extended time periods (Heighton et al. Unpublished – Chapter 2). These unsuitable, fragmented habitats have likely influenced population connectivity and size, thus resulting in higher levels of inbreeding and bottlenecks. Genetic diversity is often considered to be positively correlated with fitness and adaptive capacity, implicating a conservation importance for these populations (Frankham 2005; Habel and Schmitt 2012; Hohenlohe et al. 2021). More in-depth analyses on these population's adaptive capacity with our loci of biological relevance should be conducted for more informed management in the face of further anthropogenic driven fragmentation (Hohenlohe et al. 2021).

Tracing the trade

With limited samples stemming from major seizures, especially those known to harbour multiple species sourced from multiple countries (Ewart et al. 2021; Heinrich et al. 2017; Zhang et al. 2020b), our geographic traceability inferences for seizures are limited. Despite these caveats, we do provide some insights into the potential of our approach for the geographic traceability of

pangolins.

Many seizures were regionally sourced (i.e., from the same cluster), with the exception of a few, mainly Asian seizures in cities in Java (Indonesia), Yunnan and Guangzhou (China), and possibly Brazzaville (Republic of the Congo) and Johannesburg (South Africa). The cause of these more globalised seizures could be from them being part of major consumer markets (Yunnan and Guangzhou China) or important distribution points along the international supply chain (Like cities in Java: Chng et al. 2015; Heinrich et al. 2017). The two seized samples of *S. temminckii* in Johannesburg were assigned to their respective countries of South Africa and Zimbabwe and have been confirmed through interviews with poachers that were arrested (Darren Pietersen, pers. obs.). We confirm that Java is major sink for *M. javanica* populations in the region (Nash et al. 2018), as it is for other species (Chng et al. 2015; Nijman et al. 2016a), with more individuals being sourced from Borneo Island than Sumatra/Thai-Malay Peninsula. Chinese seizures of *M. javanica* were mainly comprised of individuals outside of its borders (Myanmar, Indochina and Sumatra-Thai-Malay Peninsula), except for potentially western China. Myanmar, and the border between it and China, is an important transit route for pangolins, often sourcing individuals from neighbouring countries of Thailand and India (Nijman et al. 2016b). It is possible that this route is feeding the markets in Yunnan province since it borders Myanmar and is comprised of potentially two (this study) to three (Hu et al. 2020a) *M. javanica* population clusters exterior to it. Additionally, we found a similar influence of Yunnan for *M. pentadactyla*, although seizures are seemingly more localised for this species in China, possibly as a result of its distribution throughout China (Challender et al. 2019a). Interestingly we did not find *M. javanica* individuals from Java and Borneo in these Chinese seizures, likely as a result of limited seizures and alternative supply chains to different cities. The method can also be used to assign museum samples of

inaccurate or unknown locality, as was the case for the *S. temminckii* sample that probably originated from Botswana and the *M. pentadactyla* samples that originated from Taiwan.

This dataset is unique in the sense that we differentiate long-range, urban markets to that of samples with more accurate geographic origins. This gives us the basis on which we can test whether these markets are regional (sourcing from their own populations) or more global (sourcing from other populations). We found that major long-range markets in both continents are exclusively regionally sourced. An expected outcome given that these often openly tolerated markets are in place to sell bushmeat (animal protein) to local clientele often for sustenance (Wilkie et al. 2016), limiting the financial incentives for transportation over vast geographic distances. Regional sourcing of pangolins solely across the Dahomey Gap were identified in long-range traditional medicine and bushmeat markets in Benin (Zanvo et al. 2022), thus confirming our hypothesis. Testing this finer scale of traceability for densely sampled clusters from our capture data using empirical assignment methods (Battey et al. 2020; Kuismin et al. 2020), may be the next step in determining its assignment accuracy. More in-depth dynamics between these long-range markets and their local sources (field sites, restricted range markets) within clusters will provide insight for population and hunting management at regional to country-wide scales (Finch et al. 2020; McNamara et al. 2015).

Now that we provide a genetic baseline linked to the geographic locality and site type of samples (from field sites to long-range markets), future investigations will not only determine the geographic assignment of seizures but also the dynamics between long-range and restricted-range markets.

Conclusion

Our study provides an essential and unprecedented insight into the global structure of pangolin populations for all eight species, as well as their connectivity and estimates of diversity. Additional samples across the ranges of species will however improve the accuracy of population delimitation, particularly for species with major sampling gaps like *M. crassicaudata*, *P. tetradactyla*, *S. gigantea* and *S. temminckii*. Furthermore, our inferences on phylogeny, differentiation indices (F_{ST}), and introgression that were based by grouping samples per country, could benefit from individual-based estimates. This is especially true when countries formed part of two clusters. This approach has been used before (Garcia-Erill et al. 2022) as it provides a global overview of differences across an entire continent, but also inadvertently provides country-specific estimates which can be more useful to management authorities that are less likely to implement trans-border species conservation policies (Bakhtiari et al. 2018; Ferreira 2004). Although we did identify IBD across the range in most species (apart from *M. culionensis*), and even potential evidence of IBD clines within certain clusters (in *M. javanica*), we did not fully explore its potential effect on population structuring. Isolation by distance can influence the inferences of population clustering methods, therefore additional analyses using spatial clustering programs (i.e. GENELAND) will be beneficial to conservation management plans (Perez et al. 2018).

With further sampling and evidence of evolutionary and ecological distinctiveness of the populations from this study's baseline, the creation of conservation units (CUs) is the next step to institutionalised conservation and protection (Funk et al. 2012; Moritz 1994; Weckworth et al. 2018). We suggest that in-depth species testing of the potentially novel species, *Manis sp.* and *Phatages hedenborgii*, should be a priority as these may require different management plans. An important aspect considering that it is unclear as to where the former is distributed, while there is

limited pangolin research in the areas in which we suspect the latter to be distributed (Heighton and Gaubert 2021; Ingram et al. 2020).

The inclusion of adaptive connectivity and capacity of these populations through variants and loci under selection, like those selected for this study, will be an integral part of delineating CUs (Funk et al. 2012). As we did not focus on this in our study, future studies can also benefit from these data to answer questions of evolutionary significance for pangolins using these adaptive loci (Figueiró et al. 2017). Including the variants under selection from our study will likely provide more finite estimates on geographic traceability (as seen in our PCAs), an aspect highly relevant to pangolin conservation considering their levels of trafficking. We hope that this study will provide an important reference for future efforts regarding the tracing of pangolins, or derivatives thereof, from seizures and large commercial markets in order to help mitigate their declines in the wild (Martinson and Ogden 2009; Milliken et al. 2016).

Conclusion and Synthesis Chapter

Background

Conservation is multifaceted in its nature where solutions to issues require a holistic, multidisciplinary approach integrating biological data (e.g. knowledge on the species), resources for conservation action, political will, and socio-economic circumstances across multiple scales (Blair et al. 2017). Thus, conservation requires investment. Investment in the capacity of the people that research and conserve wildlife, the local people that live in and around areas of high biodiversity, and the infrastructure used for research and conservation activities (Chapman et al. 2022). Although a priority, we are likely to see the effects of these large investments in the long-term (Chapman et al. 2022). For certain areas and species under excessive strain, conservation action is time-sensitive whereby short-term solutions need to be put in place in the interim. To provide conservation practitioners with the ability to make informed management plans as soon as possible we need to focus our research and development of resources in the priority areas (Christie et al. 2020; Sheil 2001). The aim of this PhD thesis was to do exactly this for pangolins by delving into the research priorities by category, species and geography (Chapter 1), providing resources for conservation genomics (Chapter 2 & 3), and developing a geographic traceability baseline for countering the pangolins greatest threat – the illegal wildlife trade (Chapter 3). From both these genomic and research resources I give important insights through a conservation lens, which I hope will aid in informative conservation action plans for each species and priority populations. In this chapter, I will provide details on (i) the major PhD outputs and how they can or are already contributing to pangolin conservation, (ii) a brief overview of the gaps that need prioritization, and (iii) a conclusion on the future for pangolin conservation.

Major outputs and their future

A pangolin information resource

Establishing accurate conservation strategies relies on the availability of biological knowledge (Conenna et al. 2017; Williams et al. 2020), together with insights on public awareness and interest in a conservation problem (Olmedo et al. 2018; Phillis et al. 2013), and in some cases the commercial determinants that result in a species being exploited (e.g. Masters et al. 2020). I therefore aimed for the first time, to provide an unprecedented synthesis of the literature, public interest and commercialization of pangolins through a systematic review (PhD objective 1). During this synthesis I linked my findings to a series of recommendations aimed at filling important knowledge gaps directly related to pangolin conservation. I also provided a list of holistic conservation actions based on the multidisciplinary review process as a means to update the current, outdated action plan for pangolins (Challender et al. 2014). These research and conservation action recommendations are an important resource for practitioners and researchers alike – allowing them to know where and when to prioritize their efforts (Blair et al. 2017; Linklater 2003). One of these recommendations was the development of pan-stakeholder working groups tasked specifically with generating and compiling knowledge for standard operating procedures of less readily published conservation action fields (rehabilitation and relocation, veterinary procedures, education and awareness raising, tourism, and law enforcement and seizure processing). Following this publication, the IUCN Species Survival Commission’s Pangolin Specialist Group (from hereafter PSG) formed thematic working groups, some of which are tasked with tackling the aforementioned conservation action fields (i.e. the Outreach and Education Working Group or Veterinary Focal Point: <https://www.pangolinsg.org/about/>). Although I am unsure if the publication’s recommendations were linked to the formation of the PSG working

groups, the fact that these are being created and utilized in pangolin conservation spheres is a step in the right direction.

As part of this research, I also provided two databases as appendices that can be, and have been, implemented for research and conservation prioritization of pangolins. The first is the literature database, which is a culmination of literature metadata (scientific, grey, books, reports, proceedings) related to pangolins from 1865 to 2019 that I sourced and cleaned from five public literature databases. The database allows researchers and practitioners to search for pangolin literature by author, country, species, year, and/or research topic, making it relatively easy to identify important pieces of literature specific to a given context. An example is how van Boom et al. (2022) used the database, and my synthesis of it, to identify the lack of studies that have investigated pangolin morphology (notably those of African species), providing reasoning for their research on the morphology of *S. temminckii*. The literature database is currently being used to guide PSG members in creating an online library of pangolin literature that is accessible to anyone (IUCN SSC PangolinSG's literature library: <https://www.pangolinsg.org/resources/library/>). The patent database that was produced from this PhD is another source that can be useful to investigators as it provides a list of registered patents linked to pangolins, the reasoning for the patent, and the individuals or companies that registered them. Analyses of patents, particularly those linked to wildlife trade can provide an important tool for horizon scanning in order to predict future trade trends in terms of products and consumer markets (Masters et al. 2020). For pangolins I found ethnozoological medicine from China (TCM) linked mainly to scales driving global patent registrations, an expected outcome. However, the hope is that the database will be used for further studies that look at trends in types of medicines being patented and the organizations registering them which can provide insight for behavior change and education programs for consumers

(Masters et al. 2020).

In addition, this publication and its appendices are open-access (Heighton and Gaubert 2021), preventing access barriers for researchers and practitioners.

Genomic resources for pangolin conservation

Genomics is becoming a standard in conservation research as it provides genome-wide data for more accurate inferences on species or population delimitation and assignment, demographic parameters, diversity, historical trajectories, and adaptive capacity to global changes (Allendorf et al. 2010; Hohenlohe et al. 2021). Insights like these allow for more informed management decisions, thus preventing cases of in/outbreeding depressions, hybridization, accumulation of deleterious variants and loss of genetic diversity and adaptive capacity (Allendorf and Hard 2009; Mussmann et al. 2017; Willoughby and Christie 2019). Although transforming this research into conservation practice is yet to be a common practice (Shafer et al. 2015), the gap is closing (Garner et al. 2016). Central to conservation genomics/genetics is the development of resources that can aid in answering the questions above.

One such resource is genome assemblies as they form the backbone to genomics by providing improved identification of loci, a way to link markers to candidate genes, reduce ascertainment biases and the effects these have on important conservation genomic outputs (Brandies et al. 2019; Hohenlohe et al. 2021). The importance of genomes is evidenced by its successful use in Tasmanian devil (*Sarcophilus harrisii*) conservation (Brandies et al. 2019) and the large collaborative projects aimed at sequencing and assembling the tree of life such as the Earth BioGenome Project (Lewin et al. 2018). Through this research, I have been able to complete the whole-genome assemblies for Pholidota by producing five genomes that complemented the

three species with already available genomes (PhD objective 4). One genome (*S. gigantea*) is a high quality, fully annotated reference assembly using a combination of Oxford Nanopore long read and Illumina short read sequencing. It is the first reference assembly for the *Smutsia* genus and will become the first peer-review published genome for African species, thus a significant resource for African pangolin genomics. The other genomes are draft genomes from Illumina short read sequencing at depths of between 43x–77x for *M. culionensis*, *M. crassicaudata*, *S. temminckii*, and *P. tetradactyla*. These species are amongst the least studied species of pangolin, often related to their elusiveness and low population densities (Heighton and Gaubert 2021). Although these four are not annotated, they still provide a basis for population genomic analyses (mapping short read data to these genomes) but also a basis from which annotations can follow, especially once RNA sequencing efforts are undertaken (Choo et al. 2016). Through the genome assemblies of *M. javanica* and *M. pentadactyla* (Choo et al. 2016; Hu et al. 2020a), subsequent studies have yielded estimates of population diversity, inbreeding, demography and mutational load as well as population structuring for geographic traceability (Hu et al. 2020a; Wang et al. 2022; Wei et al. 2022). Subsequent population genomic studies using the whole genome assemblies from this PhD will follow suit, providing much needed genomic estimates for all pangolin species.

Through these genomes and gene-capture experiments, I also provide a set of markers that have been designed for species and population delimitation (PhD objective 4). The identification of at least two potentially novel or reassessed pangolin species (*Manis* sp. from Asia and *Phatages hedenborgii* from Sudan) is an important by-product of my research efforts. I plan to follow on this research with proper species testing using nuclear and mitochondrial DNA as well as morphological characteristics (3D morphometrics). If confirmed, these species will require tailored conservation plans and policies, not to mention IUCN Red List™ assessments, updated

visual identification guides (Cota-Larson 2017), and research priorities. I warn that this is time-sensitive as trade has been noted in South Sudan with potential Sudanese species (Ingram et al. 2020), and the only information we have for the new *Manis* species is that samples have been acquired from multiple seizures (Cao et al. 2021; Hu et al. 2020b). These markers, as well as the population clusters that they were derived from, also provide the baseline for conservation management strategies through the assignment of conservation units (CUs) such as evolutionary significant units (ESUs: Funk et al. 2012; Hohenlohe et al. 2021). Although my PhD did not integrate ecological distinctiveness, morphological differentiation and full adaptive structuring, the population clusters provide a starting point from which CUs can be fully developed for all eight pangolin species (PhD objective 3). The management of pangolin ESUs can maximize evolutionary / adaptive potential of each species under future population declines while legal protection in many countries can be granted for these pangolin ESUs (Funk et al. 2012).

A basis from which to trace the trade

Developing forensic registers (DNA, isotope, morphology) to identify trade routes, sources, volumes, and legality is a key goal for the most heavily trafficked mammals, pangolins (Heighton and Gaubert 2021; Kotze et al. 2020). This is a particularly important goal considering the inadequate effect international regulation has had on curbing their trade (Challender et al. 2020a). The reason for this is that geographic traceability can provide a means to directly tackle the current problem linked to pangolins by determining poaching hotspots, syndicate networks or even identify areas best-suited for repatriation of live individuals seized from trade (Fernandes and Caparroz 2013; Ogden and Linacre 2015; Wasser et al. 2022). These allow for direct, knock-on pangolin conservation efforts in a more efficient manner by implementing community behavioral

change and education programs in hotspots, upscaling law enforcement along important points in the supply chain, and building more compelling cases for the prosecution of criminals.

The aforementioned genomic resources, and the species and population clusters that they delineated through this PhD is the baseline required for achieving this goal (Ogden and Linacre 2015). The applicability for this has already been tested in this PhD through simple population structuring-based assignment of seizures and unknown samples. However, the method would need to be further tested with additional seizures and more robust assignment methods (Battey et al. 2020). Successful assignment from large pangolin seizures in previous studies that were based on baselines similar to this study suggest that the database stemming from this PhD will be an important point of reference to come (Ewart et al. 2021; Nash et al. 2018; Wang et al. 2022; Zhang et al. 2020b). What makes this study significant is the fact that it provides the resources for all eight pangolin species across the majority of their ranges, thus dramatically scaling up the current resources for tracing the trade. However, if this is to be viable for forensic investigation, there is a great deal of effort and investment required in sampling procedures, transport through chain-of-custody, marker validation studies, and forensic laboratory accreditation (Linacre et al. 2011).

What the PhD lacks is the identification of highly informative loci across populations and the subsequent development of efficient typing tools (SNP arrays, primers for mini-fragments, and DNA fingerprinting) that can be used in tracing the trade. There has been major effort in this aspect for pangolins with the development of STR/microsatellite markers (Aguillon et al. 2020; du Toit et al. 2020), forensic mini-barcodes (Xie et al. 2021), and even fluorescent HyBeacon® probes (George et al. 2019). However, the latter two were designed off mitochondrial Cytochrome Oxidase I (COI) sequences which allowed for simple species level delimitation while STR markers have their difficulties with inter-lab reproducibility. With its genomic patterns across most

populations for all eight pangolin species, this PhD is a vital starting point for the aforementioned identification of informative loci and subsequent development of the typing tools. I hope that these will form part of the already designed markers to produce similar DNA forensic ‘toolkits’ to those found to work for fisheries (Martinson and Ogden 2009) and rhinoceroses (Harper et al. 2018).

There are a number of groups currently working on geographic traceability of pangolins, however I have not seen major success in terms of collaboration, which is a major limiting factor if we are to produce robust outputs for tracing the trade and converting it into forensic and conservation action. What is required in my opinion is an organization or government with a track record and the resources to combine the efforts of the various research groups, drive the development of the necessary tools and build legislation or regulation from which they can work – thus going beyond the scientific publication. Successful examples on which we can model include FishPopTrace for commercial fishing which is now implemented by the government of the United Kingdom (Martinson and Ogden 2009), RhODIS (Rhino DNA Index System) for rhino trafficking which is now implemented by the South African National Police Service (Harper et al. 2013), and ETIS (Elephant Trade Information System – not DNA traceability related) for elephant trafficking which is now managed by TRAFFIC (Milliken et al. 2016). We are currently working closely with the PSG to start discussions with all interested parties and role-players as we aim to follow suit.

The gaps

Throughout this research I have identified important gaps that will need to be taken precedent if we are to provide the best chance for efficient pangolin conservation. In terms of species, African species are underrepresented in literature compared to their Asian counterparts, however *M.*

culionensis (Palawan Isl. and satellite islands) is still the most underrepresented species. Due to limited good quality samples, I was not able to provide much insight in *M. crassicaudata* and *M. culionensis* population delimitation and population genetics, thus providing an opportunity to build on this research. The potentially newly identified taxa will also require additional sampling and increased efforts to identify their distributions, population delimitations, and current levels of genetic fitness.

In terms of geography, there are major literature gaps across the African continent, particularly across the eastern and central parts. Conversely, sampling efforts for population genomics was limited in Asia, particularly the east (China) and west (India and Pakistan) of the continent and various satellite islands across Sundaland (e.g. Belitung, RIAU, Bali, or Lombok Islands). However, there were noticeable gaps in western Africa for *S. gigantea* and central-eastern Africa for *S. temminckii*.

Research fields requiring additional effort are those related to the implications of trade or poaching on populations, people's perceptions and awareness of pangolins, and veterinary and rehabilitation research. I provided a general overview of patents but did not go into depth of their link with the pangolin trade, thus requiring more research effort. Major gaps related to the PhD's population genomics efforts are the use of mitogenomes and non-neutral loci and how these help delimitate populations but also provide an idea of the adaptive capacities of the designated population clusters. Further research delving into this with the inclusion of more seizures will strengthen the identification of CUs for these clusters but also provide further evidence of their effectiveness in tracing the trade.

Conclusion: future for pangolin conservation

More rapid initiatives may need to take preference while we wait for the major investments in communities and areas of pangolin conservation importance to arrive and take effect. Conservation genetics can play an integral part in more rapid, direct conservation action by providing the necessary tools to help governments and organizations develop population management plans and combat the illegal trade. The hope is that this PhD's geographic reference database and the markers that were used to make it will form part of a multifaceted approach to traceability research, conservation and prosecution. The formation of these traceability tools for identifying hotspots and routes, and the subsequent direct conservation actions that can follow in these areas may help the pangolin trade predicament by alleviating the burden that governments are facing through more efficient conservation and the complementing of flaws inherent to regulatory conservation approaches (Challender et al. 2015; Oldfield 2003). This can only help reduce the possible impact that the pangolin trade may have on ecological processes (Cooney et al., 2015), the risk of disease epidemics (Grace et al. 2012; O'Hanlon et al. 2018) and the resultant economic costs (Chowdhury et al. 2021; Pimentel et al. 2005).

When delving into the trade statistics so often cited in pangolin research, the future for pangolins looks bleak. For the case of the pangolin, international regulation alone has not been successful in managing the legal trade prior to CITES Appendix I and curbing the illegal trade post up-listing (Challender et al. 2020a; UNODC 2016). However, there has been a spike in public interest and research in recent years as pangolins undoubtedly turn into a conservation icon (Heighton and Gaubert 2021). The increased public interests, particularly through COVID-19, can correlate to public awareness, which will lead to more investment in pangolin conservation and research. I therefore foresee a substantial increase in research and the development of action plans,

policies, and legislation as governments are made aware of the plight of pangolins and are put under increasing public pressure. This PhD provides the necessary resources to help direct and build on this newly found interest, both in the form of research synthesis and conservation genomics. It is an exciting time to be a part of the upscaling of resources and effort for pangolin conservation, a worthwhile endeavor for such a unique creature.

Appendices

Chapter 1: Systematic review appendix

Too large for thesis, can be found at <https://doi.org/10.1016/j.biocon.2021.109042>

Appendix A: The research database of all publications of pangolin literature, after curation and inclusion of relevant information for further analyses. Red highlighted publications are irrelevant pieces of literature (no mention of pangolin) and were excluded from all analyses. Irrelevant columns for our analyses were kept, but not included under our classification, and thus their column headings were left as Web of Science acronyms. For the “Species” column, “general” is specified for literature that refer to pangolins in general and do not concentrate on specific species, and “unknown” is specified when the authors of the publications that we reviewed could not determine the species under study. Titles under the “Title” column that have square brackets ([]) enclosing them are English translations of original titles which can often be found in the column “Title in original language”.

Appendix B: The scientific domains (n= 6) assigned to literature of pangolins along with the assigned research themes (n= 9) and their categories (n= 30) under the domain “natural sciences”.

Scientific Domain	Theme – categories
Natural sciences	Ethnozoology and society – people’s perception and awareness, traditional food ^a , traditional medicine Life history – diet and feeding, reproduction, range/density/habitat/niche/behavior/predation Physiology and anatomy – physiology, anatomy Evolution – genomics, molecular phylogeny, population genetics/phylogeography, morphological phylogeny, traditional taxonomy, adaptation Conservation – education, rehabilitation, captive breeding, anthropogenic effects other than trade or poaching, action plan, opinion or suggestions based off data, forensic application, volume or nature of trade and poaching, implications of trade or poaching to populations Microbiology – bacteriology/virology/mycology/prionology, parasitology, immunology Checklist and monitoring ^a – checklist ^a , monitoring ^a Veterinary ^b General description of pangolins ^b
Biomedical sciences	
Engineering	
Software and programming	
Archeology and paleontology	
Ethnology, literature and humanities	

^a Themes or categories including only non-focused literature, in contrast to the others which include both focused (with a pangolin focus or where pangolins were directly compared with other species) and non-focused literature. As a result, these marked themes or categories did not form part of analyses with a pangolin focus.

^b These are considered as both themes and categories, as there are no discrete categories.

Too large for thesis, can be found at <https://doi.org/10.1016/j.biocon.2021.109042>

Appendix C: Definitions of themes and their categories under the domain “natural sciences” used for pangolin literature theme and category demarcation.

Too large for thesis, can be found at <https://doi.org/10.1016/j.biocon.2021.109042>

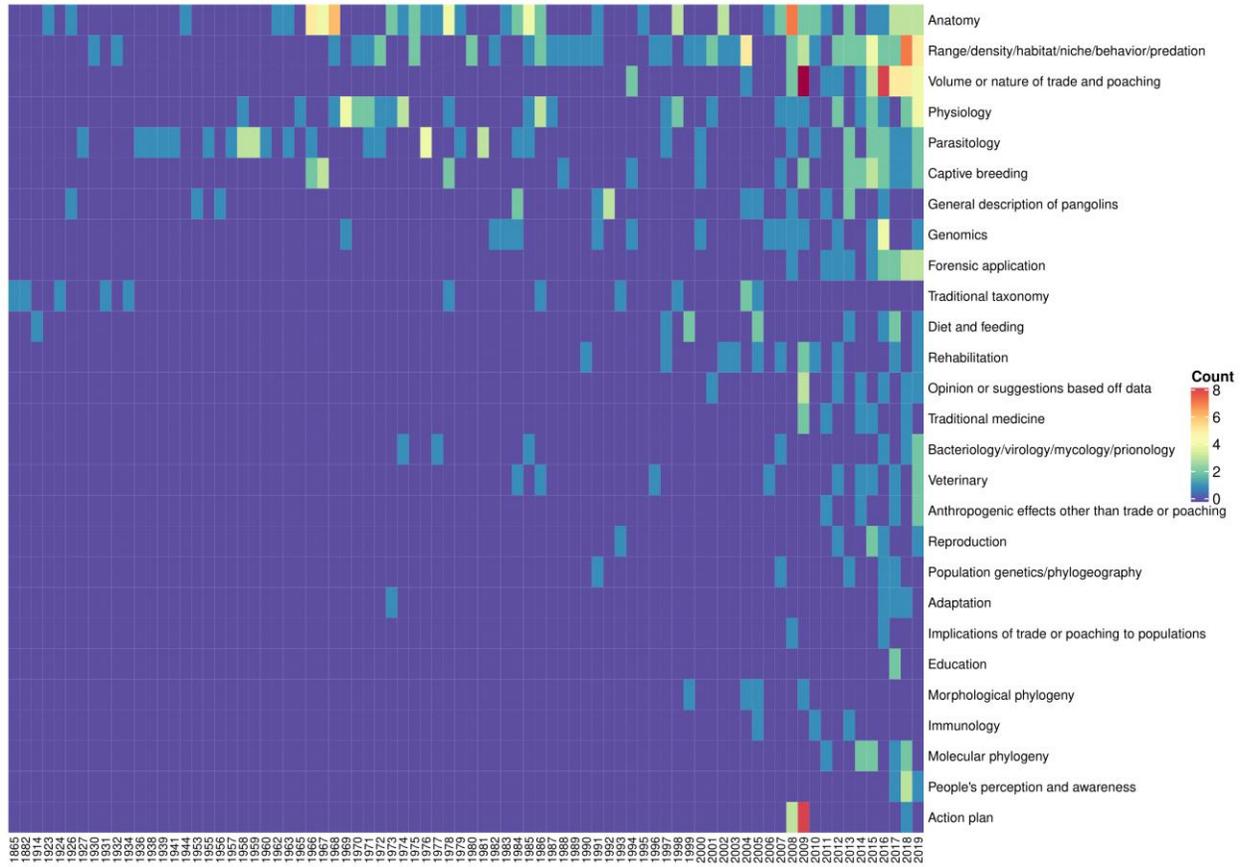
Appendix D: All 5 926 patents, along with their relevant information from Web of Science.

Too large for thesis, can be found at <https://doi.org/10.1016/j.biocon.2021.109042>

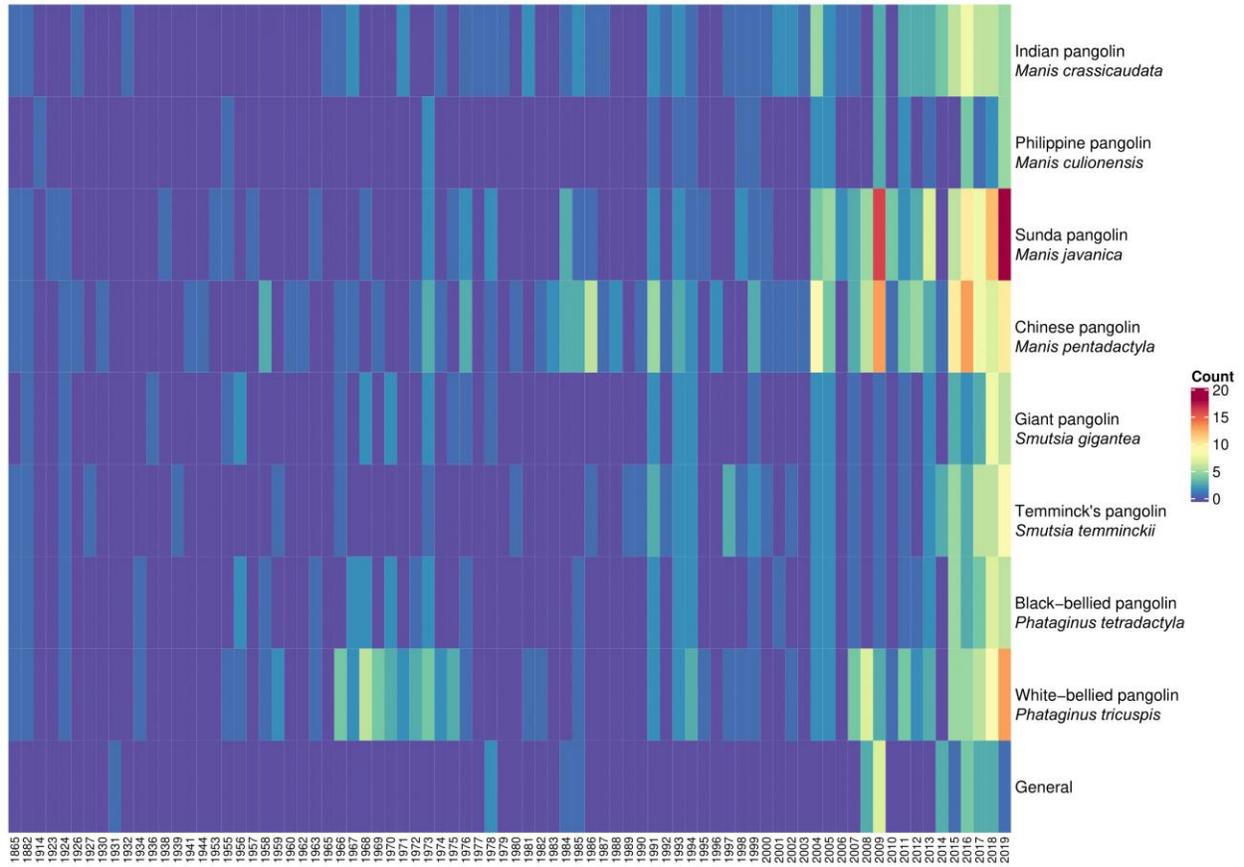
Appendix E: R-script of analyses conducted along with the details of the assumption tests for each analysis.

Too large for thesis, can be found at <https://doi.org/10.1016/j.biocon.2021.109042>

Appendix F: Manually curated VOSviewer thesaurus used to merge similar terms and ignore negligible terms. Terms that should be merged contain the term in column “label” and the term it will be merged or changed to in the column “replace by”. Terms that should be ignored contain the term in the column “label” but have no corresponding term in the column “replace by”.



Appendix G: Heatmap of literature count in research categories through time from literature with a pangolin focus and in the scientific domain “natural sciences”. Time (in years) is not continuous, a year is only displayed if a piece of literature was published during that year. Dark blue blocks represent absence of literature for a specific category for a specific year.



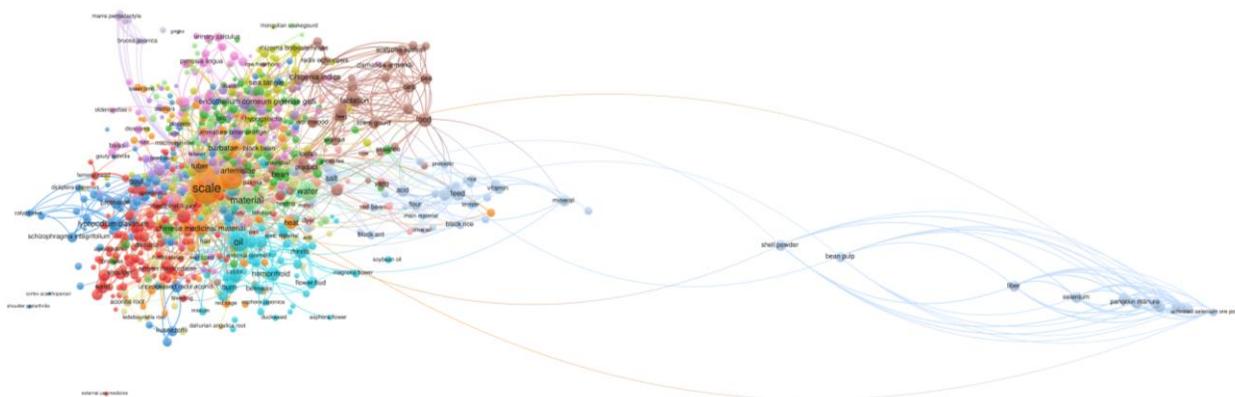
Appendix H: Heatmap of literature count per pangolin species through time from literature with a pangolin focus and in the scientific domain “natural sciences”. Time (in years) is not continuous – a year is only displayed if a piece of literature was published during that year. Dark blue blocks represent no literature for a specific species for a specific year. An additional species slot named “General” is specified for literature that refers to pangolins in general and does not concentrate on a specific species.

Appendix I: Representation of range-states in focused pangolin literature under the domain “natural sciences” (n= 377; range-states not referred to in the literature are excluded).

Range-state	Literature count	Proportion (%)
China	69	20,47
India	52	15,43
Nigeria	32	9,50
Taiwan	24	7,12
Malaysia	13	3,86
Pakistan	13	3,86
South Africa	13	3,86
Ivory Coast	11	3,26
Vietnam	11	3,26
Indonesia	9	2,67
Philippines	9	2,67
Thailand	9	2,67
Zimbabwe	9	2,67
Gabon	8	2,37
Singapore	8	2,37
Nepal	7	2,08
Cambodia	5	1,48
Sri Lanka	5	1,48
Bangladesh	3	0,89
Ghana	3	0,89
Lao PDR	3	0,89
Myanmar	3	0,89
Benin	2	0,59
Congo Republic	2	0,59
Democratic Republic of the Congo	2	0,59
Mozambique	2	0,59
Uganda	2	0,59
Angola	1	0,30
Brunei	1	0,30
Cameroon	1	0,30
Equatorial Guinea	1	0,30
Senegal	1	0,30
Sierra Leone	1	0,30
Singapore	1	0,30
South Sudan	1	0,30

Appendix J: Proportion of patents (n= 5 294) referring to each of the top 25 Derwent Class Codes from Derwent Innovations Index in Web of Science.

Derwent Class Code	Code definition	Patent count	Proportion (%)
B04	Pharmaceuticals - Natural products and polymers	5101	96,35
B07	Pharmaceuticals - General	1480	27,96
B05	Pharmaceuticals - Other organics	787	14,87
B06	Pharmaceuticals - Inorganics	462	8,73
D16	Food and Fermentation - Fermentation industry	314	5,93
D13	Food and Fermentation - Other foodstuffs and treatment	285	5,38
A96	Applications - Medical, dental, veterinary, cosmetic	284	5,37
C03	Agricultural Chemicals - Other organic compounds, inorganic compounds and multicomponent mixtures. Polymers and proteins	188	3,55
D22	Cosmetics, Disinfectants and Detergents - Sterilising, bandages, dressing and skin protection agents	155	2,93
D21	Cosmetics, Disinfectants and Detergents - Preparations for dental or toilet purposes	52	0,98
P34	Health, Amusement - Sterilising, syringes	47	0,89
B02	Pharmaceuticals - Fused ring heterocyclics	46	0,87
P33	Health, Amusement - Medical aids, oral administration	39	0,74
S05	Instrumentation, Measuring and Testing - Electrical Medical Equipment	32	0,60
A97	Applications - Miscellaneous goods not specified elsewhere	22	0,42
P21	Personal, Domestic - Wearing apparel	21	0,40
C04	Agricultural Chemicals - Fertilisers	18	0,34
F04	Textiles, Paper, Cellulose - Braiding, knitting	16	0,30
B03	Pharmaceuticals - Other heterocyclics	11	0,21
P14	Agriculture, Food, Tobacco - Animal management and care	11	0,21
P32	Health, Amusement - Dentistry, bandages, veterinary, prosthesis	10	0,19
C02	Agricultural Chemicals - Heterocyclic	9	0,17
B01	Pharmaceuticals - Steroids	7	0,13
P22	Personal, Domestic - Footwear	7	0,13
S03	Instrumentation, Measuring and Testing - Scientific Instrumentation	7	0,13



Appendix K: Co-occurrence network of terms and their links extracted with text mining algorithms in VOSviewer from titles of patent records. A binary counting approach (presence or absence of a term per publication) was conducted with a minimum term occurrence count of five and 60% term relevance threshold. Total linkage strength of each term was used for weighting the network.

Too large for thesis, can be found at <https://doi.org/10.1016/j.biocon.2021.109042>

Appendix L: List of publications in domain “biomedical science” with the author affiliations and funders of research (red), along with a list of patent assignees (red) and their patent count. Additional information on treatment and species have been added to the biomedical sciences data.

Chapter 2: Species genomics appendix

Database S1. Genes ranked by diversity. Too large for the thesis. Available at Zenodo (<https://doi.org/10.5281/zenodo.7517409>)

Custom scripts 1-3. Available at Zenodo (<https://doi.org/10.5281/zenodo.7517409>)

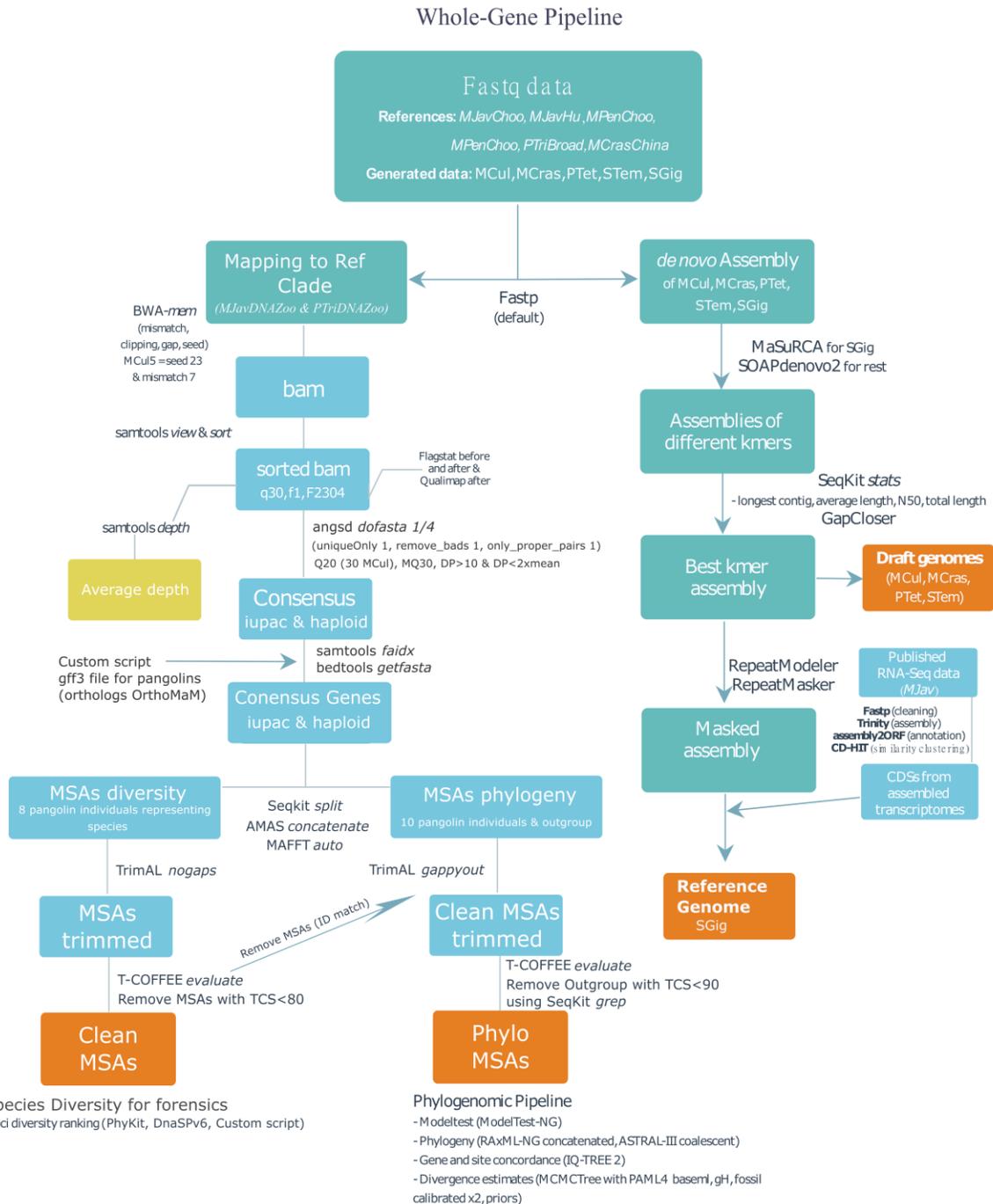


Figure S1: Orthologous whole-gene marker pipeline and subsequent analyses.

This pipeline was formulated and used to create the whole-gene markers (introns and exons) required for further analyses. During this process, the reference genome (*Smutsia gigantea*) and draft genomes for four other pangolin species were created. First, the *de novo* assembly side of the pipeline (right side) was conducted before the whole genome and subsequent whole-gene extraction side of the pipeline could be conducted (left side). Custom scripts can be found at Zenodo (<https://doi.org/10.5281/zenodo.7517409>). We used shorthand notations to indicate the genomes used for each species as follows: MJavChoo - *Manis javanica* from Malaysia (Choo et al. 2016), MJavHu - *M. javanica* confiscated in China (Hu et al. 2020a), MPenChoo - *Manis pentadactyla* from Taiwan (Choo et al. 2016), MPenHu - *M. pentadactyla* confiscated in China (Hu et al. 2020a), MCrasChina - *Manis sp.* confiscated in China (Cao et al. 2021), PTriBROAD - *Phataginus tricuspis* (unpublished BROAD institute; GCA_004765945.1), PTriDNAZoo - *P. tricuspis* (unpublished DNA Zoo; https://www.dnazoo.org/assemblies/Phataginus_tricuspis), MCul - *M. culionensis* (this study), MCras - *M. crassicaudata* (this study), PTet - *P. tetradactyla* (this study), SGig - *Smutsia gigantea* (this study), STem - *S. temminckii* (this study). Further information on these genomes and samples can be found in Table S1.

-
- a) Non-partitioned concatenated phylogeny based on 58 724 014 bp from 2 238 IUPAC consensus whole-gene markers. The pangolin phylogeny consists of 13 individuals from all eight species and is rooted with two representatives of the sister Order Carnivora (*Canis familiaris* and *Panthera pardus*). Nodal values indicate bootstrap support from 1 000 Felsenstein replicates. The model used for the maximum likelihood tree search was the TVM+FO+G4m based on a best-fit DNA model test.
 - b) Multiple species coalescent phylogeny based on 2 238 gene trees. Branch lengths are in coalescent units while nodal values indicate local posterior probability support (1=complete support).
 - c) Multiple species coalescent phylogeny based on 2 238 gene trees. Branch lengths are in coalescent units. Nodal values indicate the level of gene tree conflict, which is calculated through the number of alternative gene tree quartets that agree with the main species tree quartet topology (this figure) at each internal branch. The normalized quartet score for the entire phylogeny is 0.981. Using a polytomy test we did not identify any polytomies in the tree (which could influence these results). q_1 , q_2 , and q_3 refer to the proportion of quartets in the gene trees that agree with a branch (quartet support) for the main topology (LR|SO), first alternative (RS|LO) and second alternative (RO|LS), respectively. f_1 , f_2 , and f_3 refer to the same as above but are the raw number of quartet trees instead of the proportion. QC is the total number of quartets possible around each branch and EN is the number of effective genes for each branch.

Asterisks (*) indicate confiscated individuals whose origins could not be verified.

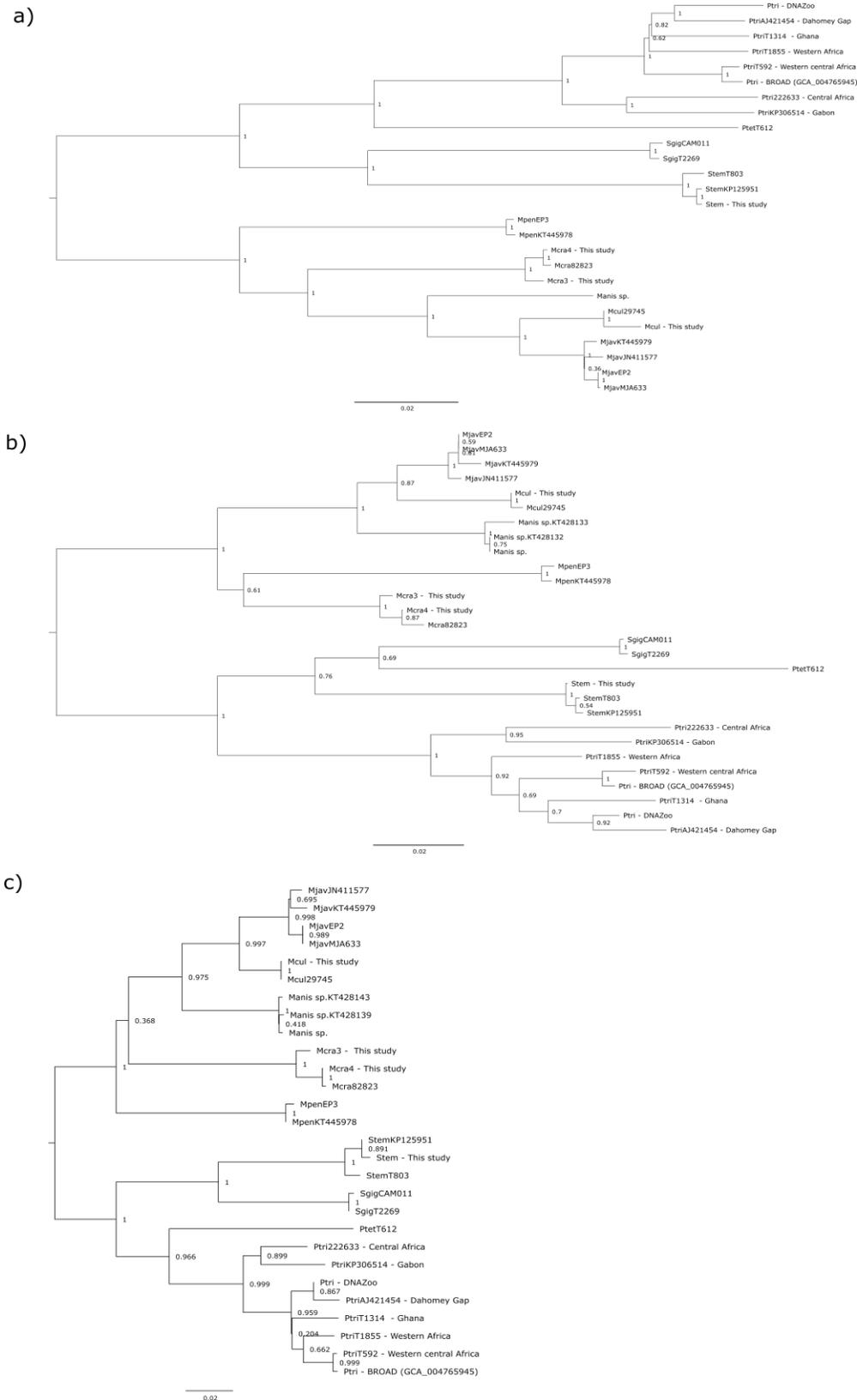


Figure S3: Phylogenetic relationships of pangolins inferred from mitochondrial gene and genome trees

- a) Full mitochondrial genome phylogeny based on 16 437 bp using the Neighbor-Joining method in MEGA X (Kumar et al. 2018). The pangolin phylogeny consists of 26 individuals from all eight species and six cryptic lineages of *Phataginus tricuspis*. Nodal values relate to bootstrap support (1000 replicates). The DNA model of evolution was Kimura 2-parameter method with all positions with less than 90% site coverage eliminated (partial deletion option).
- b) Full cytochrome b (*Cytb*) gene phylogeny based on 399 bp using the Neighbor-Joining method in MEGA X (Kumar et al. 2018). The pangolin phylogeny consists of 28 individuals, including that from all eight species, the six cryptic lineages of *Phataginus tricuspis*, and the two samples suggested as a possibly new *Manis* species (Hu et al. 2020b). Nodal values relate to bootstrap support (1000 replicates). The DNA model of evolution was Kimura 2-parameter method with all positions with less than 90% site coverage eliminated (partial deletion option).
- c) Full cytochrome oxidase subunit I (*COI*) gene phylogeny based on 600 bp using the Neighbor-Joining method in MEGA X (Kumar et al. 2018). The pangolin phylogeny consists of 28 individuals, including that from all eight species, the six cryptic lineages of *Phataginus tricuspis*, and the two samples suggested as a possibly new *Manis* species (Hu et al. 2020b). Nodal values relate to bootstrap support (1000 replicates). The DNA model of evolution was Kimura 2-parameter method with all positions with less than 90% site coverage eliminated (partial deletion option).

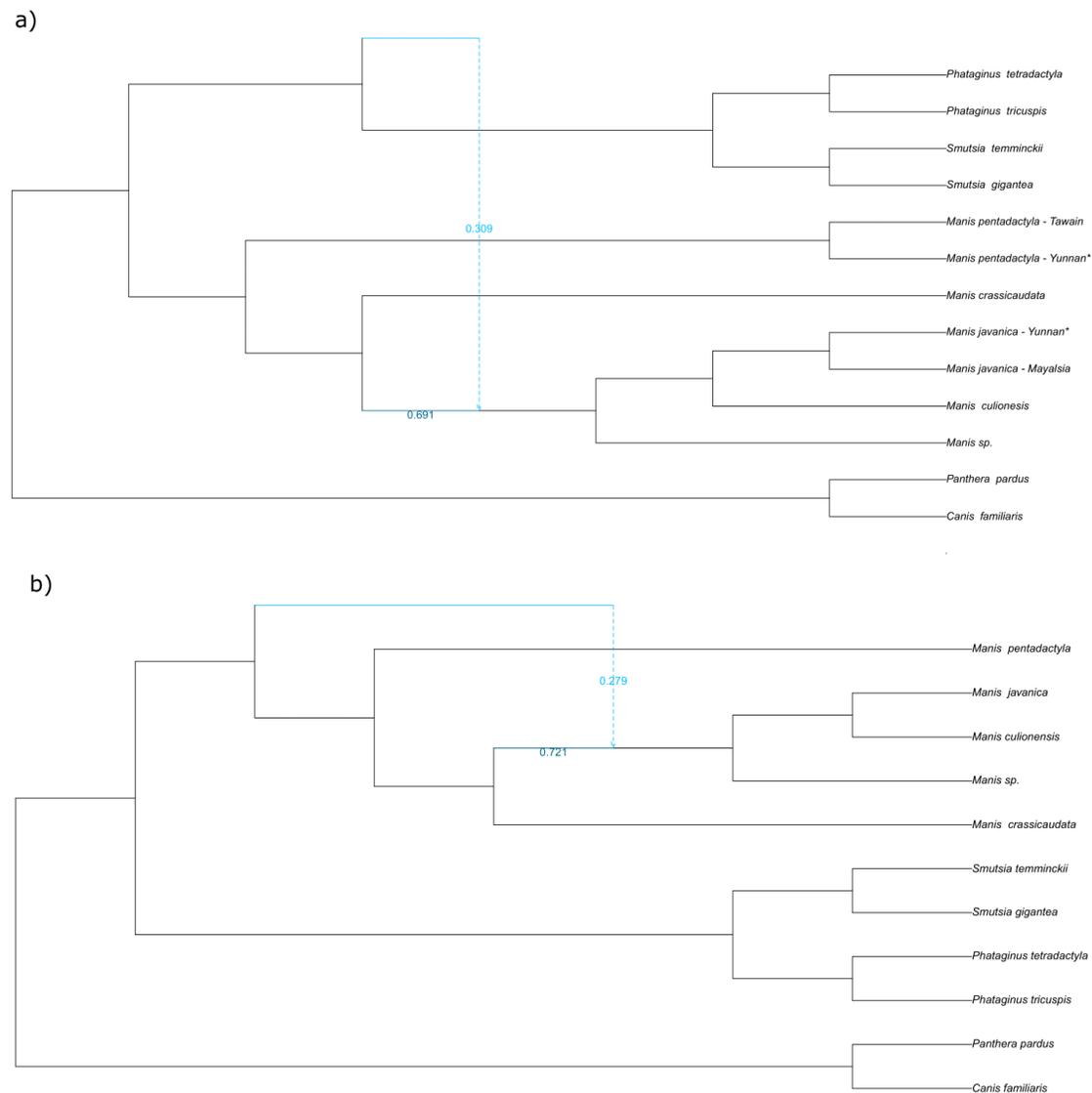


Figure S4: Maximum pseudolikelihood phylogenetic networks. Phylogenetic networks, using maximum pseudolikelihood estimates, depicted as rooted phylogenies to infer reticulation (introgression/hybridization) events within pangolins. Dotted blue lines indicate connection and direction of gene-flow between the donor and recipient taxa. Solid blue lines indicate ancestry of donor and recipient taxa. Numbers correspond to the proportion of genes shared between recipient and donor (light blue) and recipient and ancestor (dark blue). Analyses were run twice, (a) first by using each individual as a separate evolutionary unit and then (b) by indicating that multiple individuals of the same species were the same species. Both analyses indicated 1 reticulation event as the most likely outcome. Networks were drawn using the PhyloPlot function in Julia. Asterisks (*) indicate confiscated individuals whose origins could not be verified.

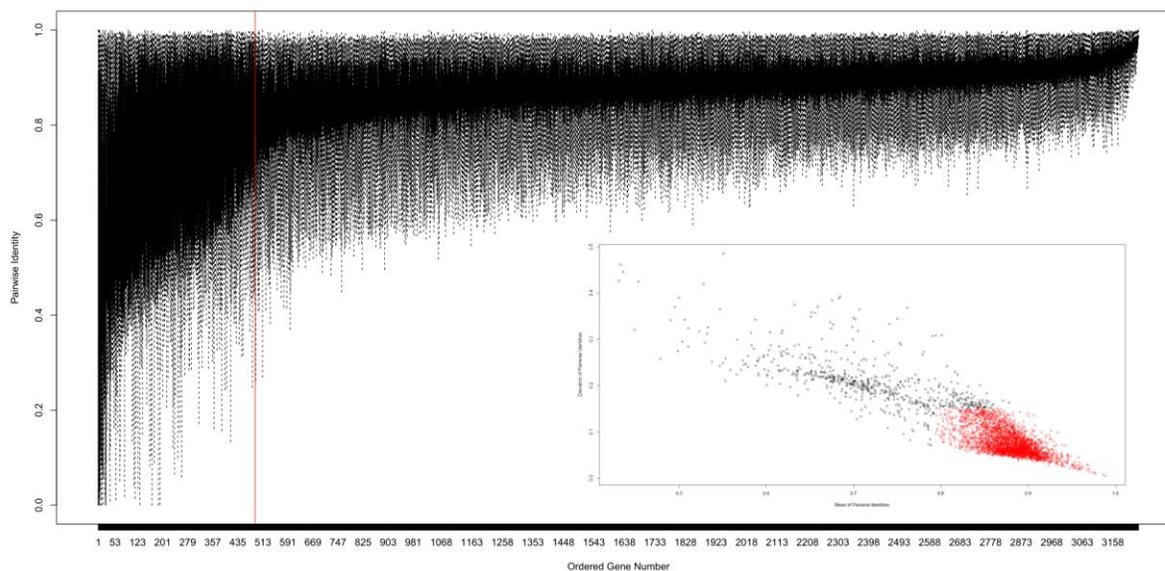


Figure S5: Diversity of orthologous whole-gene markers. Boxplot of whole-genes ranked from lowest (most diverse) to highest (least diverse) mean pairwise identity and the deviation of these markers (inner scatter plot). The average pairwise identity was calculated from the multiple sequence alignments of the eight pangolin species (using a single representative of each species only), by obtaining the pairwise identity for each combination of pairs and averaging it. This indicates the level of similarity (1=100% mean pairwise identity/similarity) which can be interpreted as pairwise diversity on the inverse (1=0% diverse). The vertical red line on the boxplot indicates the point (marker 566) at which genes to the left of it are likely outliers (values >2 Qn deviations from the median of the mean pairwise identities) as they are too diverse to be reliable (possible influence of paralogy, repetitive regions, bad alignment, bad gene annotation, etc.). This cutoff, along with another (values >2 Qn deviations from the median of the standard deviations of mean pairwise identities) can be viewed in the scatter plot whereby black points are the outliers (first 610 markers with the lowest mean pairwise identity or highest standard deviation of mean pairwise identity) and red points are likely reliable markers. These outlier markers have been highlighted in red in the sheet “removed genes” the diversity database (Database S1: available at <https://doi.org/10.5281/zenodo.7517409>). The rest make up 3 410 polymorphisms from 2 623 orthologous whole-genes.

Table S1: The samples sequenced and references used in this study. BUSCO v5 scores were obtained from mammalian orthologues by uploading the assemblies to gVolante (<https://gvolante.riken.jp/>). Unpublished genome assemblies and sequencing data for DNA Zoo versions of *Manis pentadactyla*, *M. javanica*, and *Phataginus tricuspis* are used with permission from the DNA Zoo Consortium (dnazoo.org). These DNA Zoo draft assemblies were created and reviewed following the Hi-C method (Dudchenko et al. 2017; Dudchenko et al. 2018).

Individual	Genbank accession number	Isolate; origin	Study	Collection year; sample type; collector	Sequencing platform	BUSCO scores (completed_S; completed_D; Fragmented; Missing)	Estimated coverage. (1) Lander/Waterman equation (2) mapped to <i>P. tricuspis</i>	Notes
Chinese pangolin (<i>Manis pentadactyla</i>)	GCA_000738955.1	MPE899; Taiwan	(Choo et al. 2016)	N/A	Illumina HiSeq	70.5; 0.7; 10; 18.8	59x	Unpublished DNA Zoo version with additional Hi-C and NovaSeq data can be found here: https://www.dnazoo.org/assemblies/Manis_pentadactyla
Chinese pangolin (<i>Manis pentadactyla</i>)	GCA_014570555.1	MP20; Confiscated in Yunnan, China	(Hu et al. 2020a)	N/A	Illumina HiSeq and 10X genomic	93.1; 1.3; 1.2; 4.4	281.6x	Used 77.19 Gb of the 180.66 Gb available data for mapping
Sunda pangolin (<i>Manis javanica</i>)	GCA_001685135.1	MP_PG03-UM; Malaysia	(Choo et al. 2016)	N/A	Illumina HiSeq	74.6; 0.6; 6; 8.8	145.7x	Used 71.33 Gb of the 169.4 Gb available data for mapping Unpublished DNA Zoo version with additional Hi-C and HiSeq data can be found here: https://www.dnazoo.org/assemblies

Appendices

Sunda pangolin (<i>Manis javanica</i>)	GCA_014 570535.1	MJ74; Confiscated in Yunnan, China	(Hu et al. 2020a)	N/A	Illumina HiSeq and 10X genomic	92.3; 0.9; 2; 4.8	411.8x	Manis javanica Used 112.07 Gb of the 316.2 Gb available data for mapping
Philippine pangolin (<i>Manis culionensis</i>)		MCUP0005; Casuyan, Palawan Isl., Philippines	This study	1945; museum skin; H.H. Hoogstraal / Field Museum of Natural History, Chicago (FMNH 62919)	Illumina HiSeq X Ten	6.2; 0; 12.5; 81.3	(1) 76.7x / (2) 19.9x	
Indian pangolin (3) (<i>Manis crassicaudata</i>)		MCR3; Confiscated but died in captivity, Sri Lanka	This study	2008; tissue from dead individual; Jayanthi Alahakoon / Colombo Zoo	Illumina HiSeq X Ten	40.9; 0.2; 16.9; 42	(1) 53.1x / (2) 33.2x	
Indian pangolin (4) (<i>Manis crassicaudata</i>)		MCR4; Confiscated but died in captivity, Sri Lanka	This study	2008; tissue from dead individual; Jayanthi Alahakoon / Colombo Zoo	Illumina HiSeq X Ten	N/A	(1) 16.2x / (2) 9.1x	Only used in Heterozygosity and PSMC analyses due to limited coverage
<i>Manis sp.</i>	GCA_016 801295.1	Confiscated in Sichuan, China	(Cao et al. 2021)	N/A	Illumina HiSeq	54.5; 0.3; 13.9; 31.3	44x	Indicated at <i>M. crassicaudata</i> on NCBI
Black-bellied pangolin (<i>Phataginus tetradactyla</i>)		CAM085; Yaoundé bushmeat market, Cameroon	This study	2007; tissue from dead individual; F. Njiokou & P. Gaubert	Illumina HiSeq 3000	55.9; 0.4; 15.3; 28.4	(1) 43.4x / (2) 30.1x	
White-bellied pangolin (<i>Phataginus tricuspis</i>)	GCA_004 765945.1	BS60	Unpublished - BROAD institute	O. Ryder / San Diego Zoo Institute for Conservation	Illumina HiSeq	65.5; 0.8; 12.4; 21.3	30.2x	

Appendices

White-bellied pangolin (<i>Phataginus tricuspis</i>)	Jaziri	Unpublished - DNA Zoo	Research Pittsburgh Zoo & PPG Aquarium	Illumina NovaSeq and Hi-C	87.4; 1.5; 4.2; 6.9	(2) 10.8x	The assembly can be found here courtesy of DNA Zoo: https://www.dnazoo.org/assemblies/Phataginus_tricuspis
Temminck's pangolin (<i>Smutsia temminckii</i>)	STEM 81; Kalahari Oryx Game Farm, South Africa	This study	2012; tissue from dead individual; D.W. Pietersen	Illumina HiSeq X Ten	46.4; 0.4; 17.2; 36	(1) 43.9x / (2) 21.9x	
Giant pangolin (<i>Smutsia gigantea</i>)	CAM011; Yaoundé bushmeat market, Cameroon	This study	2007; tissue from dead individual; F. Njiokou & P. Gaubert	Illumina HiSeq 3000 and Oxford Nanopore	79.8; 1.3; 7; 11.9	(1) 100.4x / (2) 57.4x	Mapping coverage estimate is from short read data only

Table S2: Output of the concordance analysis implemented in IQ-TREE along with the significance of the Chi-squared test of independence between the two discordant gene counts (gDF1_N and gDF2_N). This was used to test whether incomplete lineage sorting (ILS) may be the sole cause of discordance for a branch whereby a significant p-value (*) indicates the contrary. Branch Clade refers to the clade onto which the branch opens in the phylogenetic tree in Figure 1. The rest of the column ID's follow that of IQ-TREE; ID: Branch ID, gCF: Gene concordance factor (=gCF_N/gN %), gCF_N: Number of trees concordant with the branch, gDF1: Gene discordance factor for NNI-1 branch (=gDF1_N/gN %), gDF1_N: Number of trees concordant with NNI-1 branch, gDF2: Gene discordance factor for NNI-2 branch (=gDF2_N/gN %), gDF2_N: Number of trees concordant with NNI-2 branch, gDFP: Gene discordance factor due to polyphyly (=gDFP_N/gN %), gDFP_N: Number of trees decisive but discordant due to polyphyly, gN: Number of trees decisive for the branch, sCF: Site concordance factor averaged over 100 quartets (=sCF_N/sN %), sCF_N: sCF in absolute number of sites, sDF1: Site discordance factor for alternative quartet 1 (=sDF1_N/sN %), sDF1_N: sDF1 in absolute number of sites, sDF2: Site discordance factor for alternative quartet 2 (=sDF2_N/sN %), sDF2_N: sDF2 in absolute number of sites, sN: Number of informative sites averaged over 100 quartet, gEF_p: p-value of the Chi-squared test of independence for genes.

Branch ID	Branch Clade	gCF	gCF_N	gD F1	gDF 1_N	gD F2	gD F2_N	gD FP	gDFP_N	gN	sCF	sCF_N	sDF 1	sDF 1_N	sDF 2	sDF 2_N	sN	Branch -length	gEF_p
16	<i>M. culionensis</i> / <i>M. javanica</i> Malaysia/ <i>M. javanica</i> China	91.3 2	295 7	3.1 8	103	3.68	119	1.82	59	323 8	86.62	7792 1.4	7.18 11	5998.	6.21	531 8.77	89238. 28	2.1391 2	0.2773535 21
18	<i>M. sp.</i> / <i>M. culionensis</i> / <i>M. javanica</i> Malaysia/ <i>M. javanica</i> China	97.5 9	316 0	0.1 2	4	0.06	2	2.22	72	323 8	89.69	2479 31.1	5.21 1248 7.84	5.1	121 77.4 3	272596 .4	3.9126 2	0.4193165 78	
19	<i>M. crassicaudata</i> / <i>M. sp.</i> / <i>M. culionensis</i> / <i>M. javanica</i> Malaysia/ <i>M. javanica</i> China	94.6 3	306 3	2.5	81	1.14	37	1.76	57	323 8	67.99	1669 50	16.9 1 0.42	3931	15.1	356 05.5 3	241866	2.7166 4	4.98E-05*
20	<i>M. pentadactyla</i> Taiwan/ <i>M. pentadactyla</i> China	98.6 7	319 5	0.3 7	12	0.37	12	0.59	19	323 8	96.5	3101 96.3	1.59 4394. 39	1.91	519 4.61 .3	319785	4.1602 5	1	
21	<i>Manis</i> (Asian	96.1	237	1.2	30	1.22	30	1.42	35	246	81.27	3913	8.93	4293	9.79	469	481221	3.0453	1

Appendices

	pangolins)	5	3	2						8		09.6		6.17		76.1	.9	4	
22	Pholidota & Carnivora	93.7 1	107 2	2.8	32	3.15	36	0.35	4	114 4	91.01	1235 125	4.27	5791 9.86	4.72	640 50.9	135709 6	2.3800 6	0.6269431 82
23	African pangolins	96.6 4	238 5	1.0 1	25	1.26	31	1.09	27	246 8	76.93	3455 20.9	11.8 2	5308 3.68	11.2 5	505 36.6	449141 .2	3.1279 5	0.4097803 82
24	<i>Phataginus</i>	96.5 4	312 7	1.2 7	41	1.08	35	1.11	36	323 9	79.05	2485 59.4	10.2 8	2990 4.15	10.6 7	311 91.6	309655 .2	3.1357 1	0.4801612 24
25	<i>Smutsia</i>	97.4 4	315 5	0.6 5	21	0.8	26	1.11	36	323 8	84.25	2804 67	8.01	2445 1.21	7.74	237 95.5	328713 .7	3.5615 2	0.4790559 21
17	<i>M. javanica</i> Malaysia/ <i>M. javanica</i> China	90.1 5	291 9	3.7 4	121	4.39	142	1.73	56	323 8	88.39	2320 7.07	5.95	1502. 69	5.66	144 4.31	26154. 07	2.0209 7	0.1944997 4

Table S3: Dates used for soft bound fossil calibrations on specific nodes to be used as priors for the MCMCtree analysis of divergence estimates of pangolins. These calibrations are based on both dated fossils and molecular phylogeny estimates with reasoning provided for each. The calibration of Pholidota is the most recent calibration node possible due to the scarcity of fossils for genus/species-based estimates of the group.

Node	Date	Fossil	Reference	Notes
Ferae	66–87 Ma	Min UALVP 50993 and 50994 (Oldest stem- carnivores - miacids, viverravids)	(Fox et al. 2010)	Due to no upper estimates of Ferae, we used a molecular dated calibration which has been used in previous studied
		Max Molecular estimate	(Gaubert et al. 2018; Zhou et al. 2011)	
Carnivora	37.3–66 Ma	Min <i>Daphoenus</i> & <i>Hesperocyon</i>	(Benton et al. 2015)	<i>Tapocyon</i> may be an even older caniform; (46– 43 Ma). However, it is placed outside Carnivora (Wesley-Hunt and Flynn 2005). The oldest feliforms may be the nimravids, but this too is uncertain (Hunt 2004).
		Max UALVP 50993 and 50994 (Oldest stem- carnivores - miacids, viverravids)	(Fox et al. 2010)	
Pholidota	31–45 Ma	Min <i>Manidae</i> (oldest manidae fossil)	(Gebo and Rasmussen 1985)	Messel deposits (Germany). Found with <i>Eurotamandua</i> <i>joresi</i> which has been debated as to whether it should be included in the Pholidota or whether it predates this order but Gaudin et al. (2009) places it under Pholidota as sister to <i>Eomanis</i> and <i>Euromanis</i>
		Max <i>Euromanis krebsi</i> (oldest) <i>Eurotamandua joresi</i> <i>Eomanis waldi</i>	(Gaudin et al. 2009; Rose 2012)	

Table S4: Estimated posterior mean or median divergence estimates and 95% Highest Posterior Density (HPD) interval of each node for this study and the one conducted by Gaubert et al. (2018). The latter study involved mitogenomes and nine nuclear genes, included more individuals and fossil calibrations within the order Carnivora, and used the program BEAST (Bouckaert et al. 2014) to obtain time to most recent common ancestor (TMRCA) estimates. \$ refers to nodes with fossil priors for this study. * refers to nodes where the 95% HPD of divergence estimates do not overlap in the two studies.

Node	This study		Gaubert et al. (2018)	
	Mean	95% HPD	Median	95% HPD
Ferae\$	79.47	67.66 – 87.24	78.9	69.6 – 87.0
Carnivora\$	49.95	36.49 – 65.44	50.8	44.9 – 57.4
Caniformia			41.6	38.0 – 46.0
<i>Mustela – Ailuropoda</i>			35.4	30.5 – 40.9
Felidae			11.3	10.0 – 13.3
<i>Acinonyx – Felis</i>			7.2	3.3 – 10.2
Pholidota\$	41.34	33.54 – 45.45	37.9	31.4 – 44.6
<i>Smutsia - Phataginus</i>	20.35	12.79 – 28.29	22.9	18.7 – 27.2
<i>Manis</i>	22.62	17.07 – 28.23*	12.9	10.3 – 15.6
<i>M. crassicaudata – (M. sp, M. javanica, M. culionensis)</i>	16.84	12.21 – 22.00*	9.1	6.6 – 11.4
<i>M. sp – (M. javanica – M. culionensis)</i>	7.16	4.73 – 10.42		
<i>M. javanica – M. culionensis</i>	2.71	1.70 – 4.21	1.7	0.4 – 2.5
<i>M. javanica</i> China – <i>M. javanica</i> Malaysia	0.77	0.46 – 1.24		
<i>M. pentadactyla</i> China – <i>M. pentadactyla</i> Taiwan	2.82	1.48 – 4.95		
<i>Smutsia</i>	9.78	5.55 – 15.74	9.8	5.6 – 13.2
<i>Phataginus</i>	11.34	6.55 – 17.45	13.3	9.3 – 16.5
<i>P. tricuspis</i>			2.7	0.8 – 4.6
<i>P. tricuspis</i> West Africa – Western Central Africa			1.1	0.0 – 2.4

Table S5: Proportion of genome-wide heterozygosity for mammalian species ranked from least to most diverse. As depicted in Figure 4, each mammalian order is colour-coded (Pholidota = dark red), and citations are given for the studies in which the value was originally provided for each species. Heterozygosity values for pangolins (Pholidota) include those calculated in our study as well as Hu et al. (2020a), who used a different method of calculation (using VCFtools on autosomal SNPs with 50 kb non-overlapping sliding windows). This is an edited and updated version of the table created by Hu et al. (2020a).

Species	Heterozygosity (%)	Mammalian order	Sources
Iberian lynx (<i>Lynx pardinus</i>)	0.010	Carnivora	(Abascal et al. 2016)
Domestic cat (<i>Felis catus</i>)	0.012	Carnivora	(Cho et al. 2013)
Baiji (<i>Lipotes vexillifer</i>)	0.012	Certartiodactyla	(Zhou et al. 2013)
Myanmar snub-nosed monkey (<i>Rhinopithecus strykeri</i>)	0.015	Primates	(Zhang et al. 2016)
Altai Neanderthal (<i>Homo sapiens</i>)	0.017	Primates	(Prüfer et al. 2014)
Cheetah (<i>Acinonyx jubatus</i>)	0.020	Carnivora	(Dobrynin et al. 2015)
Snow leopard (<i>Panthera uncia syn</i>)	0.023	Carnivora	(Cho et al. 2013)
Yangtze river dolphin (<i>Lipotes vexillifer</i>)	0.026	Certartiodactyla	(Zhou et al. 2013)
Eurasian lynx (<i>Lynx lynx</i>)	0.028	Carnivora	(Abascal et al. 2016)
Siberian tiger (<i>Panthera tigris altaica</i>)	0.030	Carnivora	(Dobrynin et al. 2015)
Domestic dog (<i>Canis familiaris</i>)	0.032	Carnivora	(Lindblad-Toh et al. 2005)
Tasmanian devil (<i>Sarcophilus harrisi</i>)	0.032	Marsupialia	(Cho et al. 2013)
Black snub-nosed monkey (<i>Rhinopithecus bieti</i>)	0.033	Primates	(Zhou et al. 2016)
Island fox (<i>Urocyon littoralis</i>) - San Miguel	0.033	Carnivora	(Robinson et al. 2016)
Bengal tiger (<i>Panthera tigris tigris</i>)	0.040	Carnivora	(Dobrynin et al. 2015)
Golden snub-nosed monkey (<i>Rhinopithecus roxellana</i>)	0.042	Primates	(Zhou et al. 2016)
White lion (<i>Panthera leo</i>)	0.048	Carnivora	(Cho et al. 2013)
Amur tiger (<i>Panthera tigris altaica</i>)	0.049	Carnivora	(Cho et al. 2013)
Island fox (<i>Urocyon littoralis</i>) - San Nicolis	0.049	Carnivora	(Robinson et al. 2016)
Aye-Aye (<i>Daubentonia madagascariensis</i>)	0.051	Primates	(Perry et al. 2011)
Wild horse (<i>Equus ferus przewalskii</i>)	0.052	Perissodactyla	(Huang et al. 2014)
African lion (<i>Panthera leo</i>)	0.058	Carnivora	(Cho et al. 2013)
Grey snub-nosed monkey (<i>Rhinopithecus brelichi</i>)	0.062	Primates	(Zhou et al. 2016)
Eastern lowland gorilla (<i>Gorilla beringei graueri</i>)	0.064	Primates	(Xue et al. 2015)
Bornean orangutan (<i>Pongo pygmaeus</i>)	0.065	Primates	(Locke et al. 2011)
Mountain gorilla (<i>Gorilla beringei beringei</i>)	0.065	Primates	(Xue et al. 2015)
Naked mole rat (<i>Heterocephalus glaber</i>)	0.068	Rodentia	(Kim et al. 2011)
Pileated gibbon (<i>Hylobates pileatus</i>)	0.073	Primates	(Carbone et al.

			2014)
White tiger (<i>Panthera tigris tigris</i>)	0.073	Carnivora	(Cho et al. 2013)
White-bellied pangolin (<i>Phataginus tricuspis</i>) - DG	0.074	Pholidota	Sequencing data from DNAZoo
Dromedary (<i>Camelus dromedarius</i>)	0.074	Certartiodactyla	(Wu et al. 2014)
Platypus (<i>Ornithorhynchus anatinus</i>)	0.075	Monotremata	(Warren et al. 2008)
Indian pangolin (<i>Manis crassicaudata</i>) – MCR4	0.075	Pholidota	This study
Eastern lowland gorilla (<i>Gorilla beringei graueri</i>)	0.076	Primates	(Scally et al. 2012)
Human Han (<i>Homo species</i>)	0.077	Primates	(Meyer et al. 2012)
West African chimpanzees (<i>Pan troglodytes verus</i>)	0.080	Primates	(Mikkelsen et al. 2005)
African green monkey (<i>Chlorocebus aethiops aethiops</i>)	0.080	Primates	(Warren et al. 2015)
Wild bactrian camel (<i>Camelus bactrianus ferus</i>)	0.084	Certartiodactyla	(Wang et al. 2012)
Chinese pangolin (<i>Manis pentadactyla</i>) - Taiwan	0.085	Pholidota	Sequencing data from Choo et al. (2016). Analysis from this study
Sunda pangolin (<i>Manis javanica</i>) - Average over 74 individuals	0.085	Pholidota	Samples in Hu et al. (2020a). Analysis in their study
Minke whale (<i>Balaenoptera acutorostrata</i>)	0.086	Certartiodactyla	(Yim et al. 2014)
Finless porpoise (<i>Neophocaena phocaenoides</i>)	0.086	Certartiodactyla	(Yim et al. 2014)
Koala (<i>Phascolarctos cinereus</i>)	0.087	Marsupialia	(Johnson et al. 2018)
Tibetan antelope (<i>Pantholops hodgsonii</i>)	0.088	Certartiodactyla	(Ge et al. 2013)
Yak (<i>Bos grunniens</i>)	0.089	Certartiodactyla	(Qiu et al. 2012)
Mongolian horse (<i>Equus ferus caballus</i>)	0.089	Perissodactyla	(Huang et al. 2014)
Domestic bactrian camel (<i>Camelus bactrianus</i>)	0.090	Certartiodactyla	(Wang et al. 2012)
Southern white rhinoceros (<i>Ceratotherium simum simum</i>)	0.090	Perissodactyla	(Tunstall et al. 2018)
Indian pangolin (<i>Manis crassicaudata</i>) – MCR3	0.094	Pholidota	This study
Common chimpanzee (<i>Pan troglodytes</i>)	0.095	Primates	(Mikkelsen et al. 2005)
Domestic horse (<i>Equus ferus caballus</i>)	0.095	Perissodactyla	(Wade et al. 2009)
Cross River gorilla (<i>Gorilla gorilla diehli</i>)	0.096	Primates	(Xue et al. 2015)
Black-bellied pangolin (<i>Phataginus tetradactyla</i>)	0.100	Pholidota	This study
Wrangel woolly mammoth (<i>Mammuthus primigenius</i>)	0.100	Proboscidea	(Palkopoulou et al. 2015)
Polar bear (<i>Ursus maritimus</i>)	0.108	Carnivora	(Liu et al. 2014)
Chinese pangolin (<i>Manis pentadactyla</i>) – Yunnan, China confiscation	0.109	Pholidota	Sequencing data from Hu et al. (2020a). Analysis

Island fox (<i>Urocyon littoralis</i>)- San Clemente	0.110	Carnivora	from this study (Robinson et al. 2016)
Northern white rhinoceros (<i>Ceratotherium simum cottoni</i>)	0.110	Perissodactyla	(Tunstall et al. 2018)
Chinese pangolin (<i>Manis pentadactyla</i>)	0.114	Pholidota	Sample MP20 in Hu et al. (2020a). Analysis in their study
Bactrian camel (<i>Camelus bactrianus</i>)	0.116	Certartiodactyla	(Wu et al. 2014)
Sumatran orangutan (<i>Pongo abelii</i>)	0.120	Primates	(Xue et al. 2015)
Gray fox (<i>Urocyon cinereoargenteus</i>)	0.120	Carnivora	(Robinson et al. 2016)
Brown hyena (<i>Parahyaena brunnea</i>)	0.121	Carnivora	(Westbury et al. 2018)
Cow (<i>Bos taurus</i>)	0.121	Certartiodactyla	(Corbett-Detig et al. 2015)
Rat (<i>Rattus norvegicus</i>)	0.125	Rodentia	(Leffler et al. 2012)
Oimyakon woolly mammoth (<i>Mammuthus primigenius</i>)	0.125	Proboscidea	(Palkopoulou et al. 2015)
Chinese pangolin (<i>Manis pentadactyla</i>) - Average over 23 individuals	0.127	Pholidota	Samples in Hu et al. (2020a). Analysis in their study
Sumatran rhinoceros (<i>Dicerorhinus sumatrensis</i>)	0.130	Perissodactyla	(Mays et al. 2018)
Giant panda (<i>Ailuropoda melanoleuca</i>)	0.132	Carnivora	(Li et al. 2010)
Siamang (<i>Symphalangus syndactylus</i>)	0.140	Primates	(Carbone et al. 2014)
Bottlenose dolphin (<i>Tursiops truncatus</i>)	0.142	Certartiodactyla	(Yim et al. 2014)
Western lowland gorilla (<i>Gorilla gorilla gorilla</i>)	0.144	Primates	(Xue et al. 2015)
Giant pangolin (<i>Smutsia gigantea</i>)	0.146	Pholidota	This study
Gray wolf (<i>Canis lupus</i>)	0.149	Carnivora	(Corbett-Detig et al. 2015)
Fin whale (<i>Balaenoptera physalus</i>)	0.151	Certartiodactyla	(Yim et al. 2014)
Sunda pangolin (<i>Manis javanica</i>)	0.152	Pholidota	Sample MJ74 in Hu et al. (2020a). Analysis in their study
Temminck's pangolin (<i>Smutsia temminckii</i>)	0.155	Pholidota	This study
Chinese hamster (<i>Cricetulus griseus</i>)	0.159	Rodentia	(Lewis et al. 2013)
Asian pangolin sp. (<i>Manis</i> sp.) - Sichuan, China confiscated	0.161	Pholidota	Sequencing data from Cao et al. (2021). Analysis in this study.
Silvery gibbon (<i>Hylobates moloch</i>)	0.17	Primates	(Carbone et al. 2014)

Malaysian cynomolgus macaque (<i>Macaca fascicularis</i>)	0.171	Primates	(Higashino et al. 2012)
Central African chimpanzees (<i>Pan troglodytes troglodytes</i>)	0.176	Primates	(Mikkelsen et al. 2005)
Western lowland gorilla (<i>Gorilla gorilla gorilla</i>)	0.178	Primates	(Scally et al. 2012)
Vervet monkey (<i>Chlorocebus aethiops pygerythrus</i>)	0.180	Primates	(Warren et al. 2015)
Wild boar (<i>Sus scrofa</i>) - Tibetan	0.182	Certartiodactyla	(Li et al. 2013b)
Olive baboon (<i>Papio anubis</i>)	0.189	Primates	(Corbett-Detig et al. 2015)
Island fox (<i>Urocyon littoralis</i>) - San Rosa	0.191	Carnivora	(Robinson et al. 2016)
Island fox (<i>Urocyon littoralis</i>) - San Catalina	0.196	Carnivora	(Robinson et al. 2016)
Island fox (<i>Urocyon littoralis</i>) - San Cruz	0.197	Carnivora	(Robinson et al. 2016)
Northern white-cheeked gibbon (<i>Nomascus leucogenys</i>)	0.220	Primates	(Carbone et al. 2014)
Bighorn sheep (<i>Ovis canadensis</i>)	0.222	Certartiodactyla	(Corbett-Detig et al. 2015)
Sunda pangolin (<i>Manis javanica</i>) - Malaysia	0.224	Pholidota	Sequencing data from Choo et al. (2016). Analysis from this study
Sunda pangolin (<i>Manis javanica</i>) - Yunnan, China confiscation	0.230	Pholidota	Sequencing data from sample MJ74 in Hu et al. (2020a). Analysis from this study
Alpaca (<i>Vicugna pacos</i>)	0.266	Certartiodactyla	(Wu et al. 2014)
David's myotis (<i>Myotis davidii</i>)	0.279	Chiroptera	(Zhang et al. 2013)
Rhesus macaque (<i>Macaca mulatta</i>)	0.287	Primates	(Corbett-Detig et al. 2015)
Cape buffalo (<i>Syncerus caffer caffer</i>)	0.288	Certartiodactyla	(de Jager et al. 2021)
Brown bear (<i>Ursus arctos</i>)	0.320	Carnivora	(Liu et al. 2014)
White-bellied pangolin (<i>Phataginus tricuspis</i>) - CWA	0.334	Pholidota	Sequencing data from Genbank (GCA_00476594.5.1). Analysis from this study
Common marmoset (<i>Callithrix jacchus</i>)	0.341	Primates	(Consortium 2014)
Przewalski's horse (<i>Equus ferus przewalskii</i>)	0.363	Perissodactyla	(Corbett-Detig et al. 2015)
Brandt's bat (<i>Myotis brandtii</i>)	0.371	Chiroptera	(Seim et al. 2013)

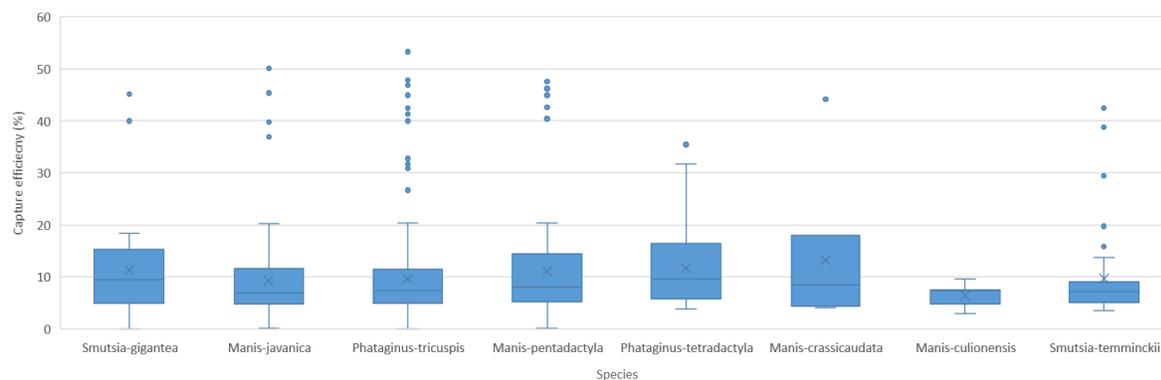
Chinese rhesus macaque (<i>Macaca mulatta lasiota</i>)	0.410	Primates	(Yan et al. 2011)
Wild boar (<i>Sus scrofa</i>)	0.441	Certartiodactyla	(Corbett-Detig et al. 2015)
Black flying fox (<i>Pteropus alecto</i>)	0.453	Carnivora	(Zhang et al. 2013)
Opossum (<i>Monodelphis domestica</i>)	0.490	Marsupialia	(Mikkelsen et al. 2007)
Philippine pangolin (<i>Manis culionensis</i>)	0.492	Pholidota	This study
Crab-eating macaque (<i>Macaca fascicularis</i>)	0.530	Primates	(Yan et al. 2011)
Rabbit (<i>Oryctolagus cuniculus</i>)	0.750	Lagomorpha	(Carneiro et al. 2014)
Eastern hoolock gibbon (<i>Hoolock leuconedys</i>)	0.800	Primates	(Carbone et al. 2014)
House mouse (<i>Mus musculus castaneus</i>)	0.809	Rodentia	(Corbett-Detig et al. 2015)

Table S6: Estimated generation times of each species of pangolins based on the currently available literature. These times include a sum of the gestation period (passing on genomic information to the next generation at conception) and time until sexual maturity (when genomic information can be passed on to the next generation again). These estimates were used in the PSMC (pairwise sequentially Markovian coalescent) model analysis on pangolins in order to get the timing of changes in IICR (inverse instantaneous coalescence rate) as accurate as possible. We defined generation time as from the conception of an individual to the time until the first born of that individual is conceived.

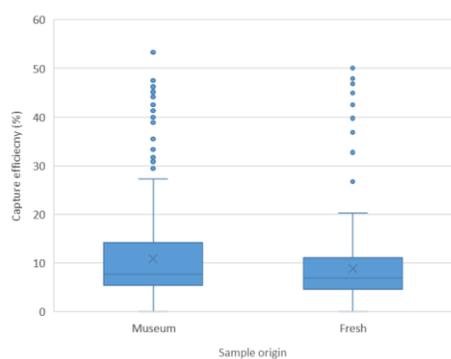
Species	Estimated generation time	Evidence	Citations
Chinese pangolin (<i>Manis pentadactyla</i>)	2 years	Reaches sexual maturity at 1–1.5 years old, but may be as low as 6–7 months; gestation is 6–7 months (captive data)	(Chin et al. 2012; Zhang et al. 2016)
Sunda pangolin (<i>Manis javanica</i>)	1.5 years	1 year (supposed sexual maturity of males is 1.5 years based off sperm analysis – A. Kurniawan, unpublished) until sexual maturity but may be as low as 6–7 months; gestation is 6 months (captive and seizure data)	(Zhang et al. 2015a; Zhang et al. 2017)
Philippine pangolin (<i>Manis culionensis</i>)	1.5 years	No data, but likely similar to the Sunda pangolin.	
Indian pangolin (<i>Manis crassicaudata</i>)	3 years	Up to 3 years until sexual maturity (unpublished data), gestation is around 6–8 months (135–251 days; captive data).	(Mahmood et al. 2020; Mohapatra et al. 2018)
Black-bellied pangolin (<i>Phataginus tetradactyla</i>)	2.5 years	2 years (unpublished suggestion – book chapter), gestation is 5 months (150 days; wild data).	(Gudehus et al. 2020; Pagès 1972a)
White-bellied pangolin (<i>Phataginus tricuspis</i>)	2.5 years	No data for maturity (probably similar to Black-bellied pangolin), gestation is 5 months (140–150 days; wild data).	(Pagès 1972b)
Temminck’s pangolin (<i>Smutsia temminckii</i>)	2.5 years	Sexual maturity is likely around 2 years but no home range is established for a few more years (assuming home range is important, then 3–7 years), gestation is 3.5–4.5 months (105–140 days; both captive and wild data).	(Pietersen et al. 2020; van Ee 1966)
Giant pangolin (<i>Smutsia gigantea</i>)	3 years	No data available, but likely similar, if not longer than, Temminck’s pangolin due to larger size.	

Chapter 3: Population genomics appendix

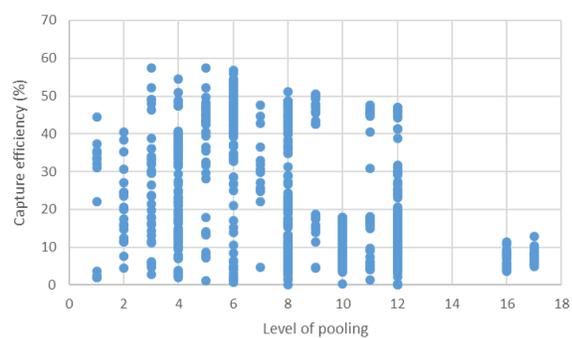
A



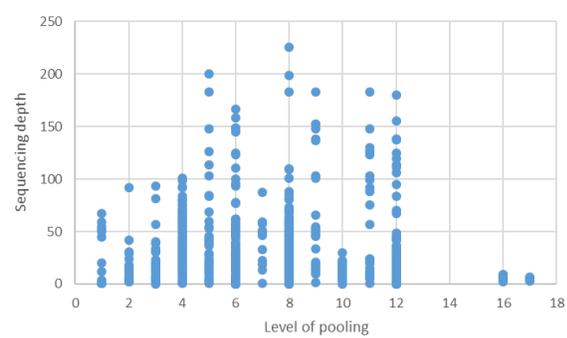
B



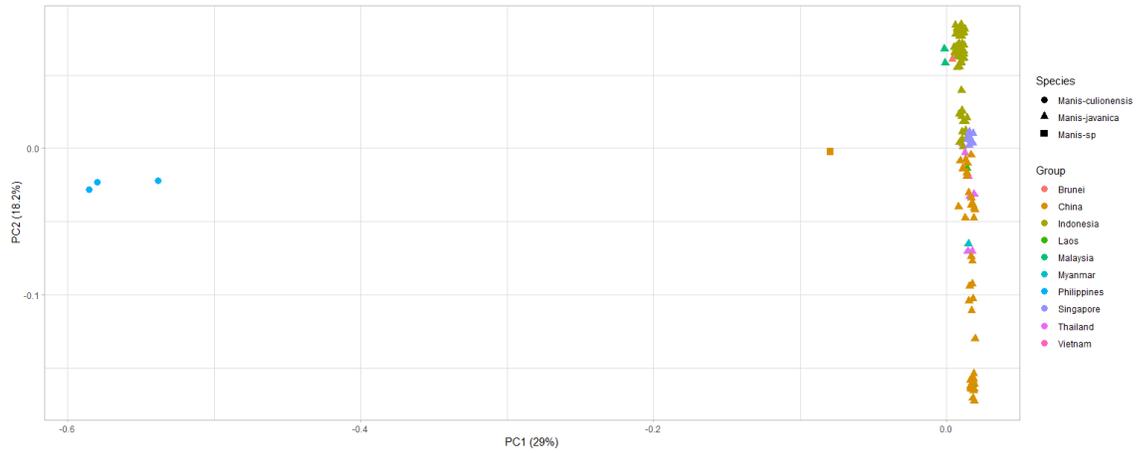
C



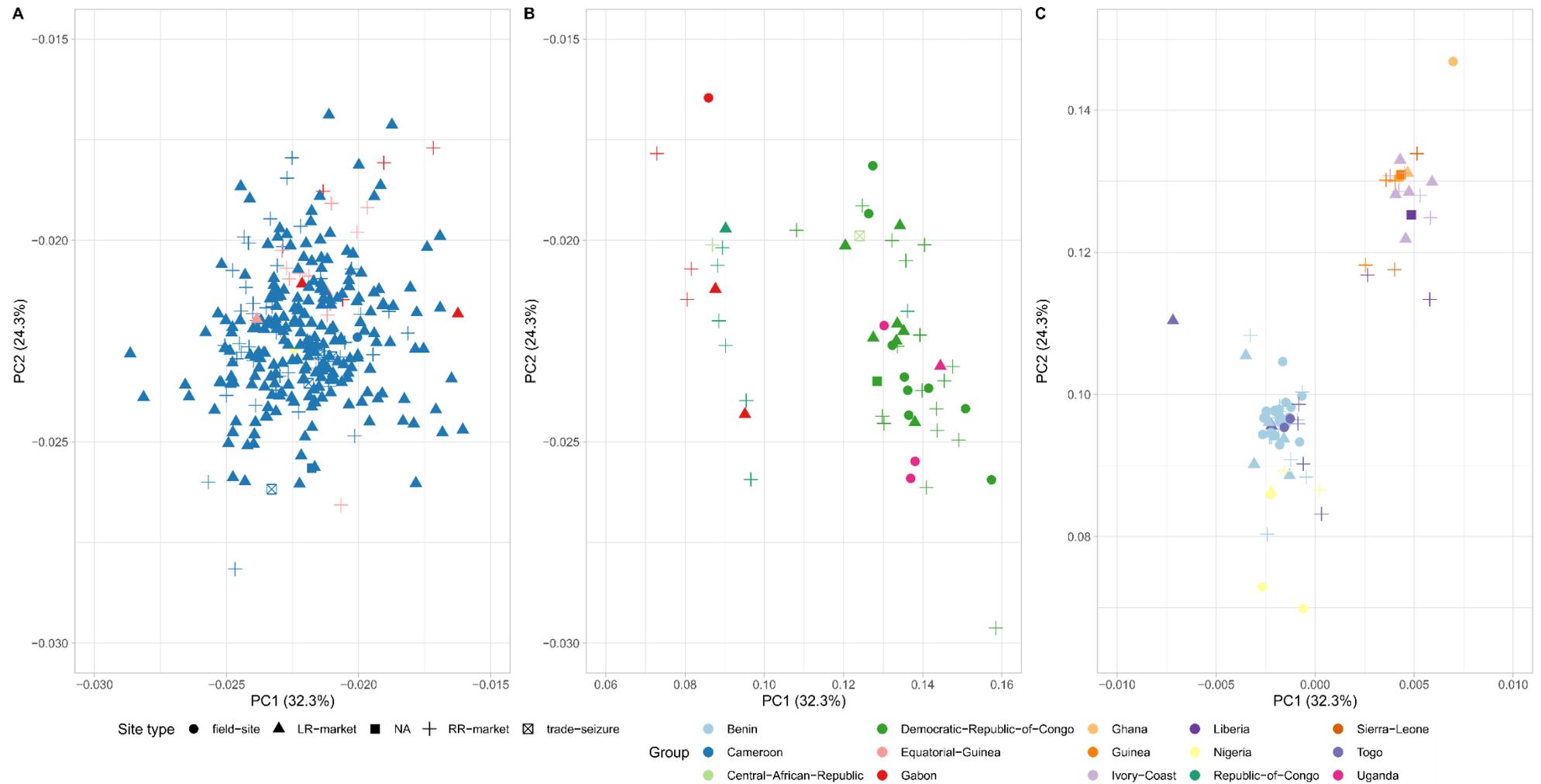
D



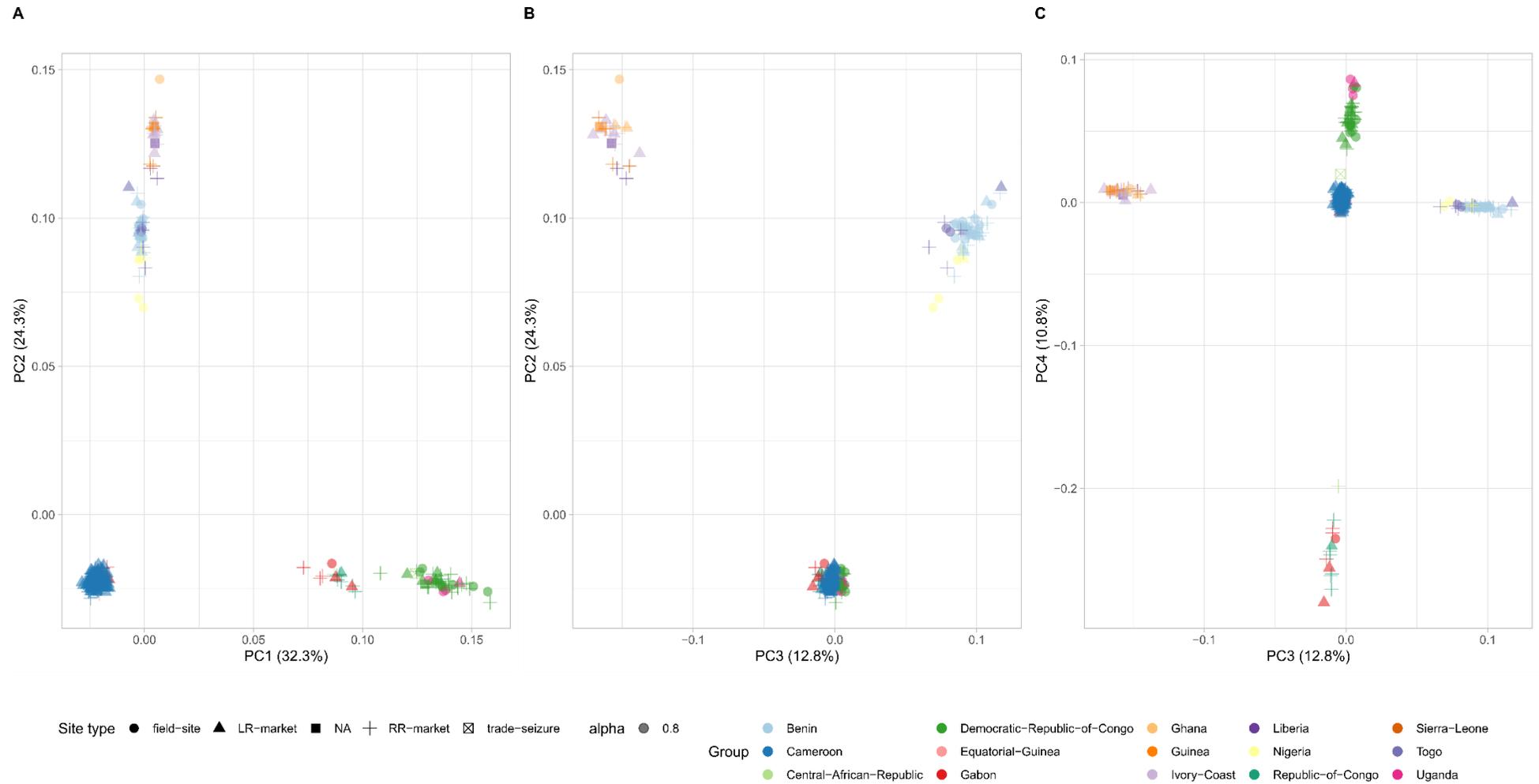
Supplementary Figure 1: Gene-capture summary statistics. Box and whisker plots of capture efficiency across (A) species and (B) quality (museum vs fresh) with X representing mean and outliers shown as points. Scatter plots comparing amount of individuals pooled in a single capture reaction to (C) capture efficiency and (D) sequencing depth.



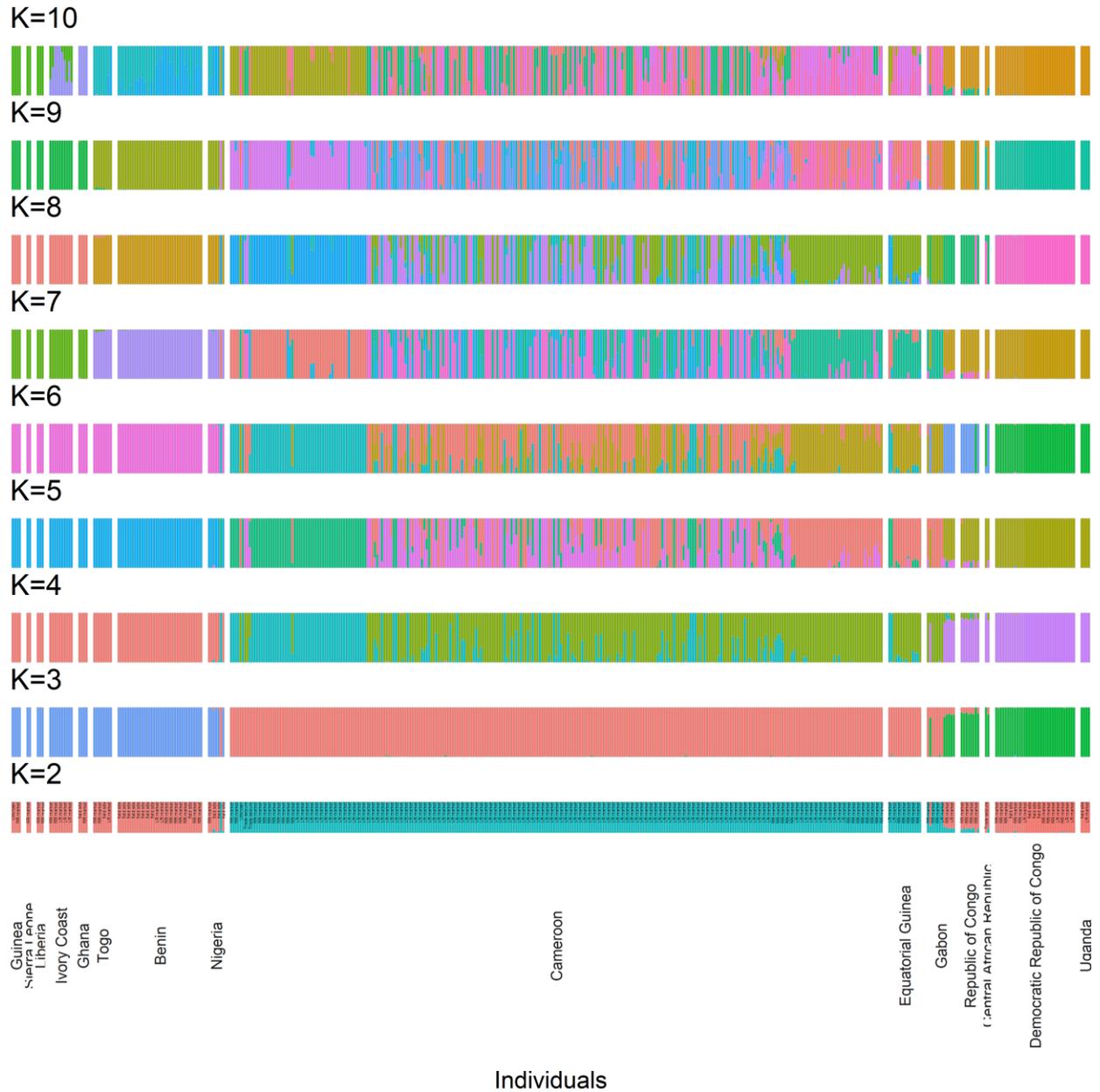
Supplementary Figure 2: Principle component analysis (PCA) of South-East Asian species for the top two axes of differentiation in which points are coloured by country and shaped by species.



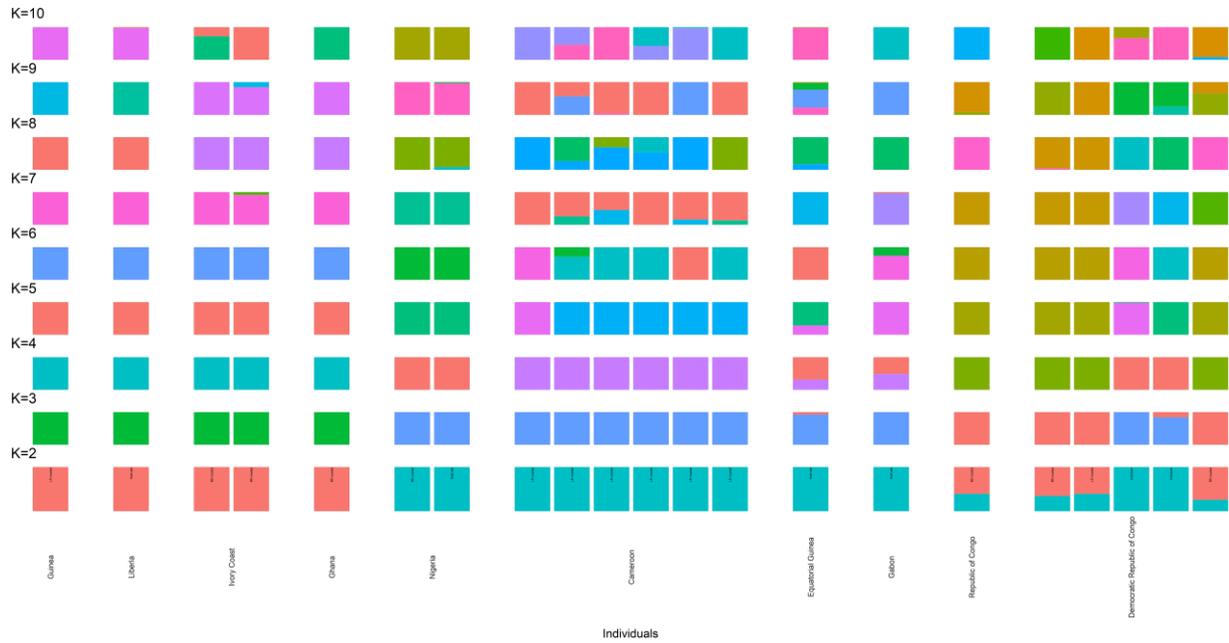
Supplementary Figure 3: Principle component analysis (PCA) of *Phataginus tricuspis* zoomed in for the top two axes of differentiation in which points are coloured by country and shaped by type of site from which the sample was collected (8 520 SNPs, 427 individuals). These represent the three major clades of the species (A) central-western Africa, (B) central Africa, and (C) western Africa.



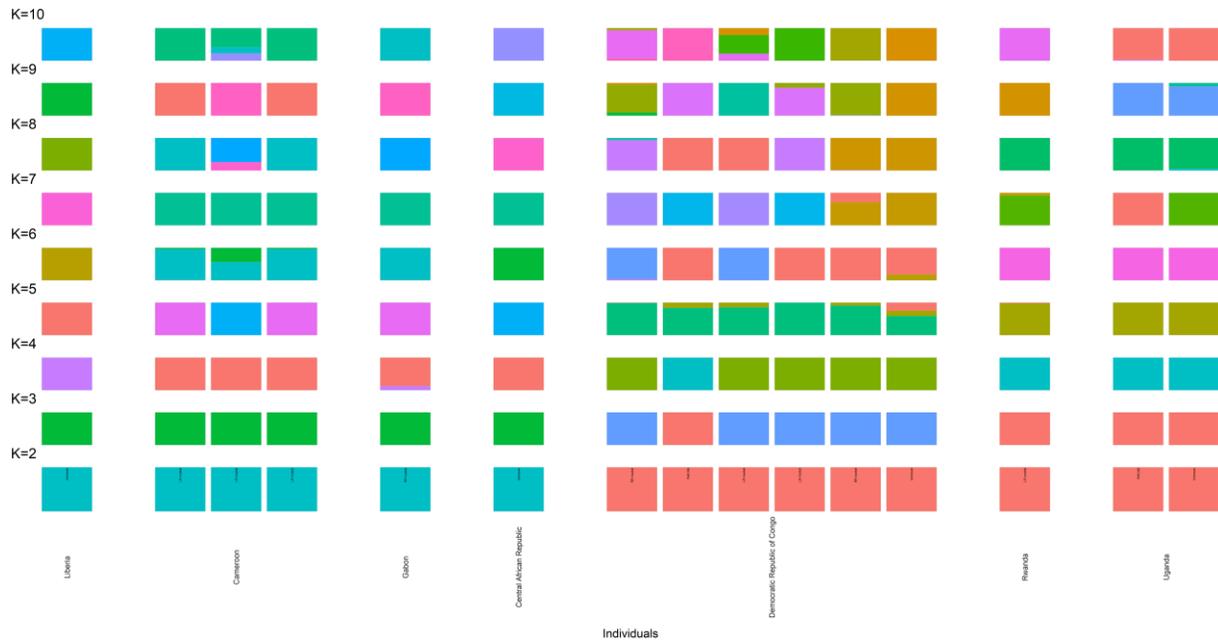
Supplementary Figure 4: Principle component analysis (PCA) of *Phataginus tricuspis* for the top four axes of differentiation in which points are coloured by country and shaped by type of site from which the sample was collected (8 520 SNPs, 427 individuals).



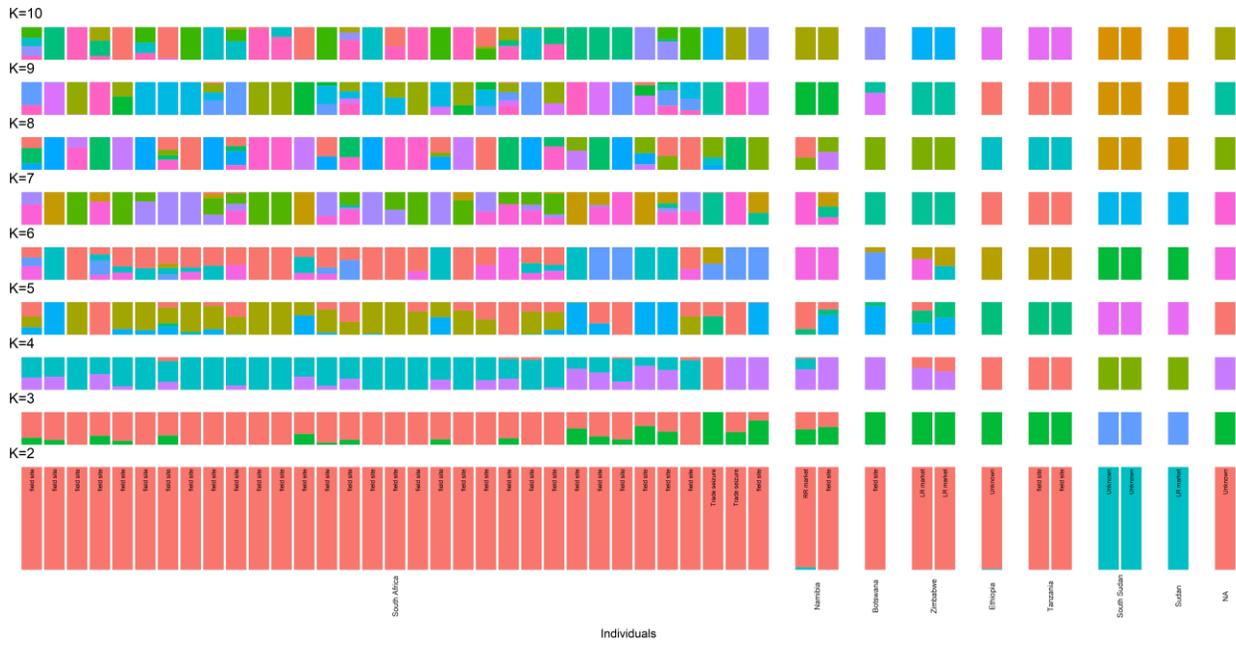
Supplementary Figure 5: *Phataginus tricuspidis* ADMIXTURE bar graphs of all Ks (K=2-10) in which samples belonging to countries are clumped and each site type is labelled (8 312 SNPs, 456 individuals). Samples are ordered from South-East to North-West in each country.



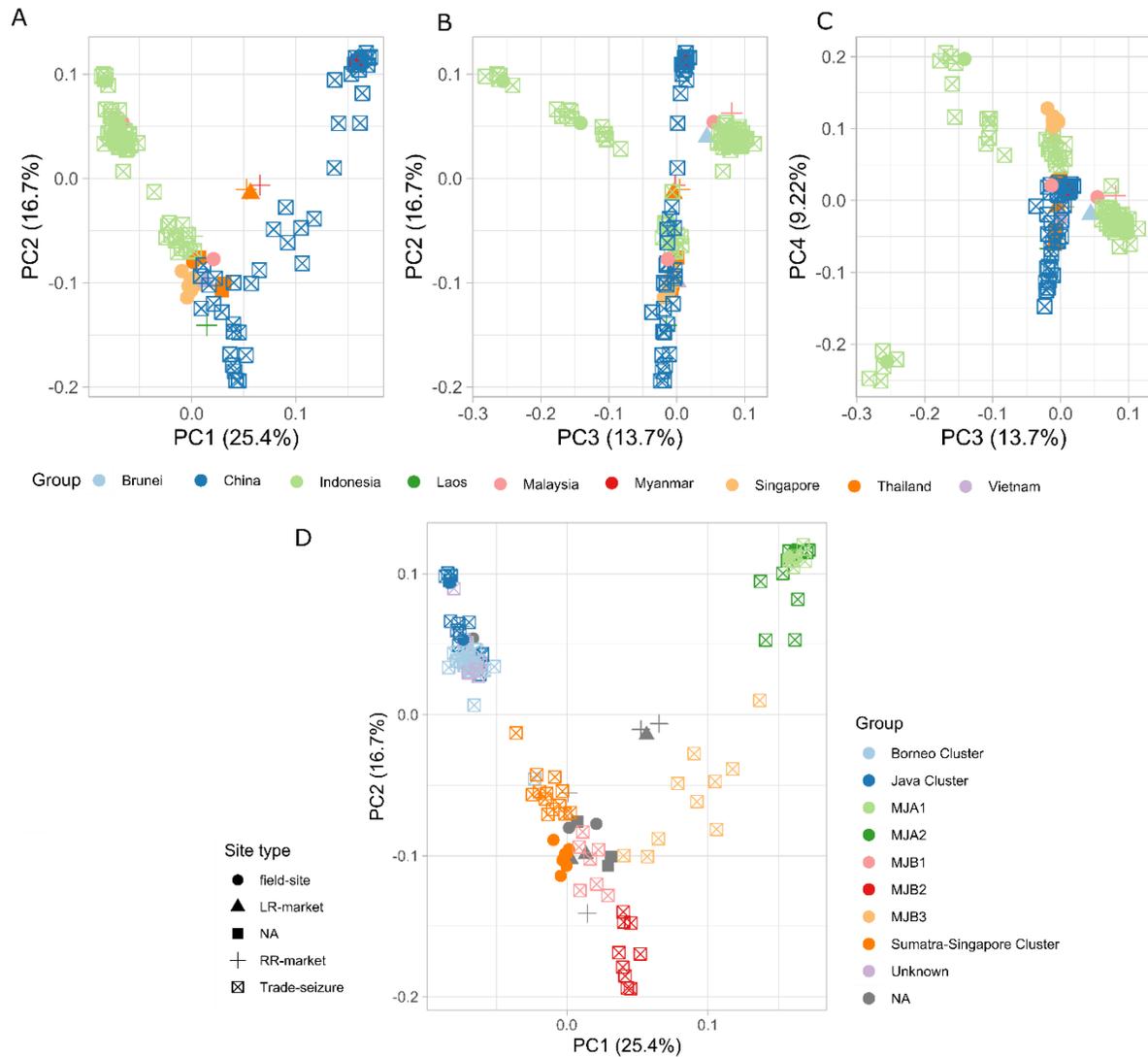
Supplementary Figure 6: *Phataginus tetradactyla* ADMIXTURE bar graphs of all Ks (K=2-10) in which samples belonging to countries are clumped and each site type is labelled (793 SNPs, 21 individuals). Samples are ordered from South-East to North-West in each country.



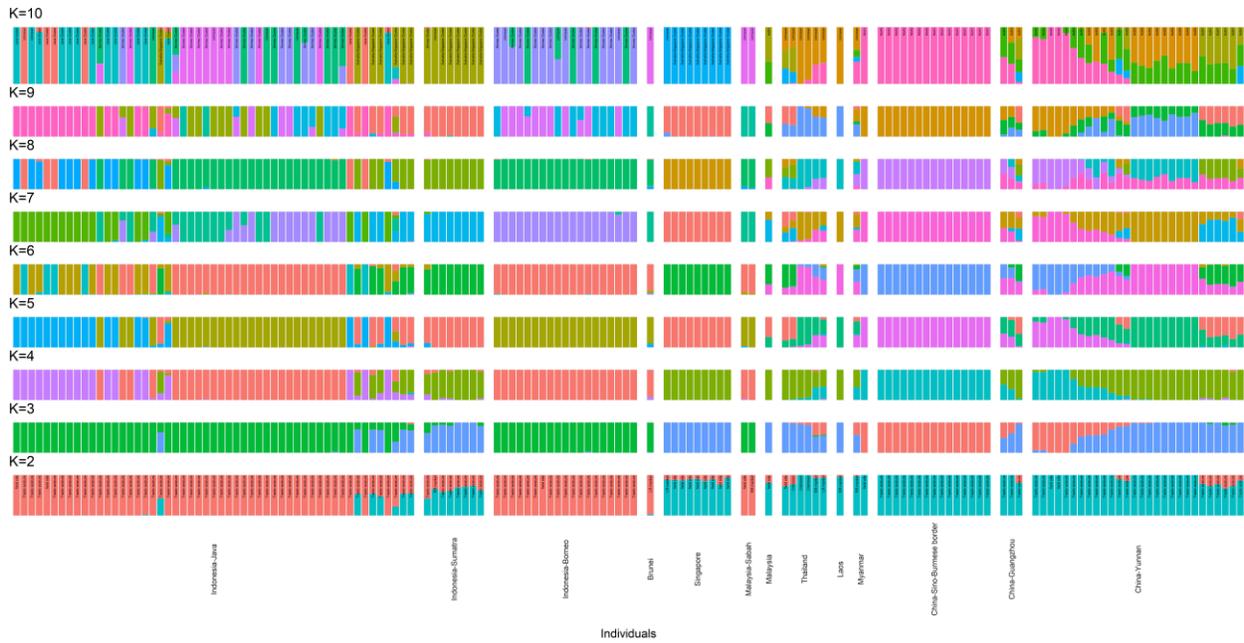
Supplementary Figure 7: *Smutsia gigantea* ADMIXTURE bar graphs of all Ks (K=2-10) in which samples belonging to countries are clumped and each site type is labelled (768 SNPs, 15 individuals). Samples are ordered from South-East to North-West in each country.



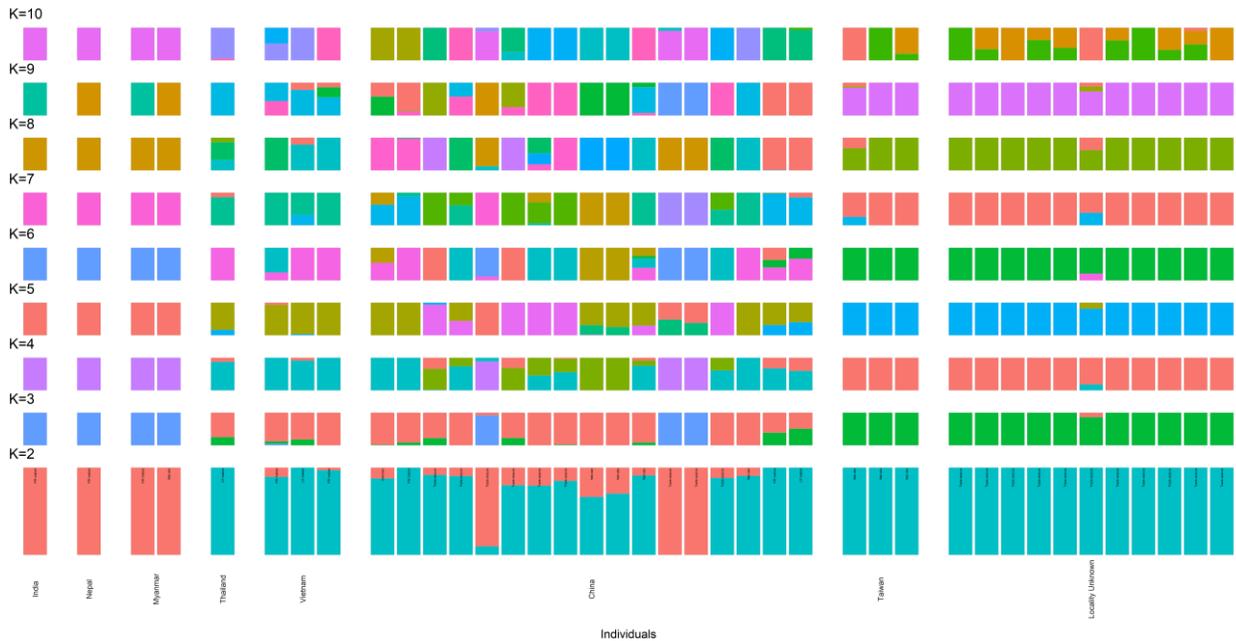
Supplementary Figure 8: *Smutsia temminckii* ADMIXTURE bar graphs of all Ks (K=2-10) in which samples belonging to countries are clumped and each site type is labelled (1 506 SNPs, 45 individuals). Samples are ordered from South-West to North-East in each country.



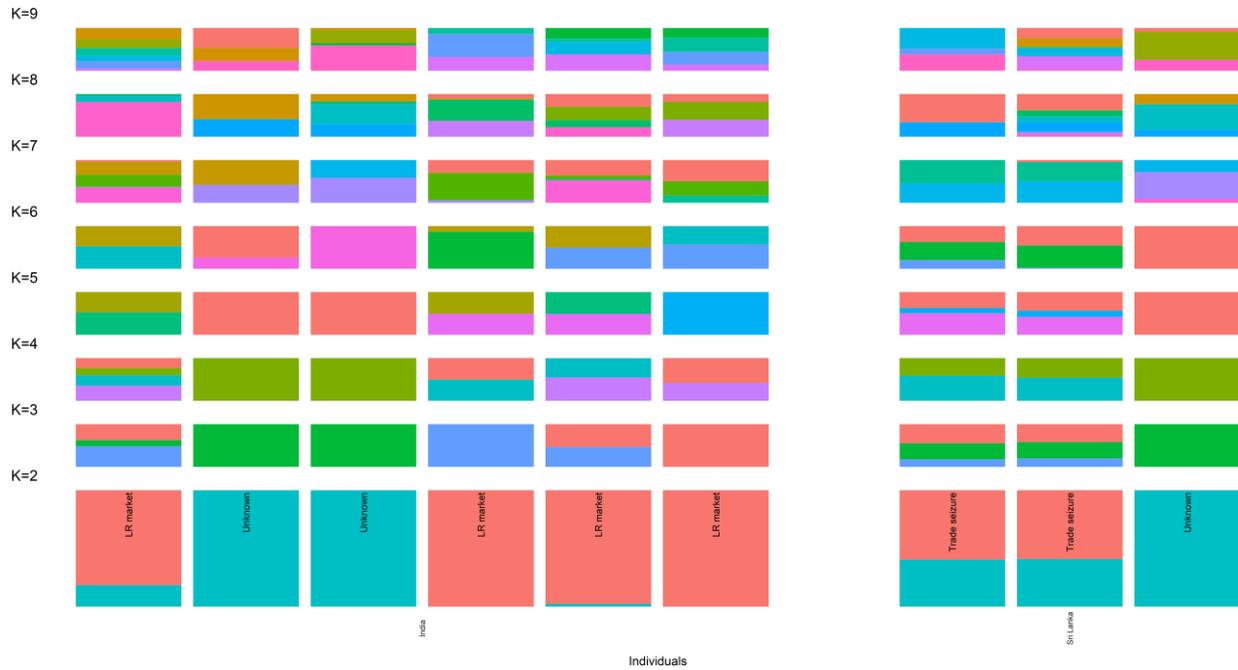
Supplementary Figure 9: Principle component analysis (PCA) of *Manis javanica* (194 SNPs, 16 individuals) for the (A-C) top four axes of differentiation in which points are coloured by country and shaped by type of site from which the sample was collected. (D) PCA with the top two axes of differentiation in which points are coloured by previously identified clusters (Hu et al. 2020a; Nash et al. 2018) and shaped by type of site from which the sample was collected (NA represent samples that come from this study and thus did not yet have a cluster).



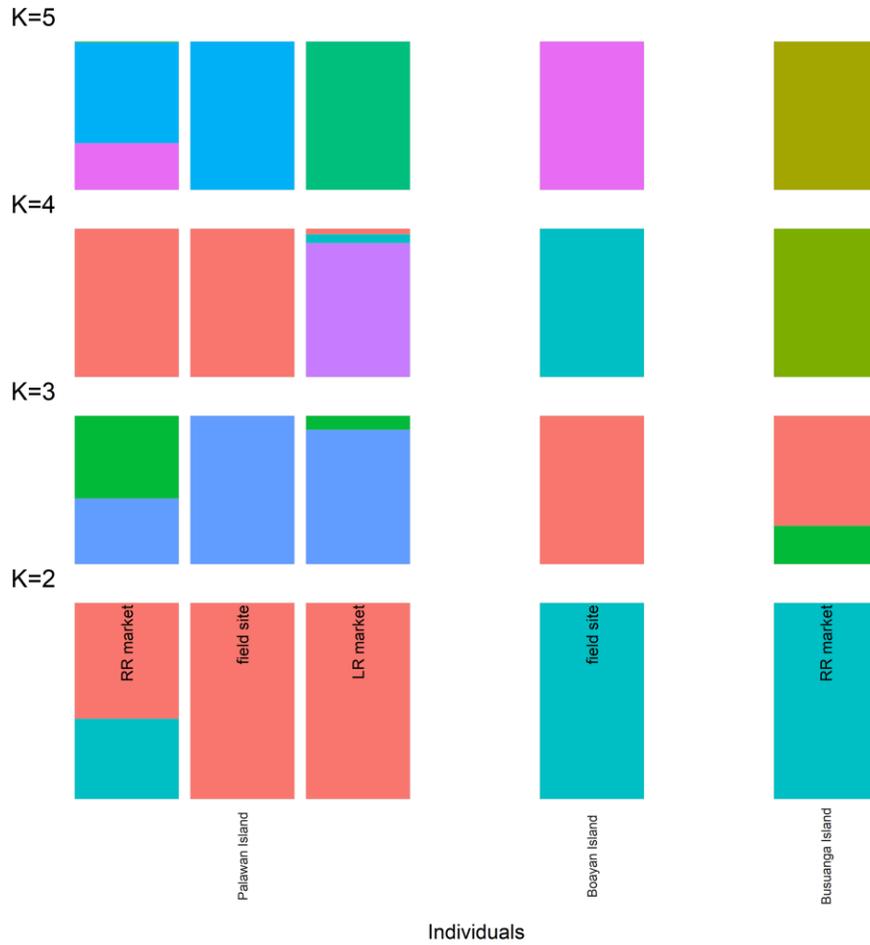
Supplementary Figure 10: *Manis javanica* ADMIXTURE bar graphs of all Ks (K=2-10) in which samples belonging to countries are clumped and each site type is labelled (4 162 SNPs, 148 individuals). Samples are ordered from South-East to North-West in each country.



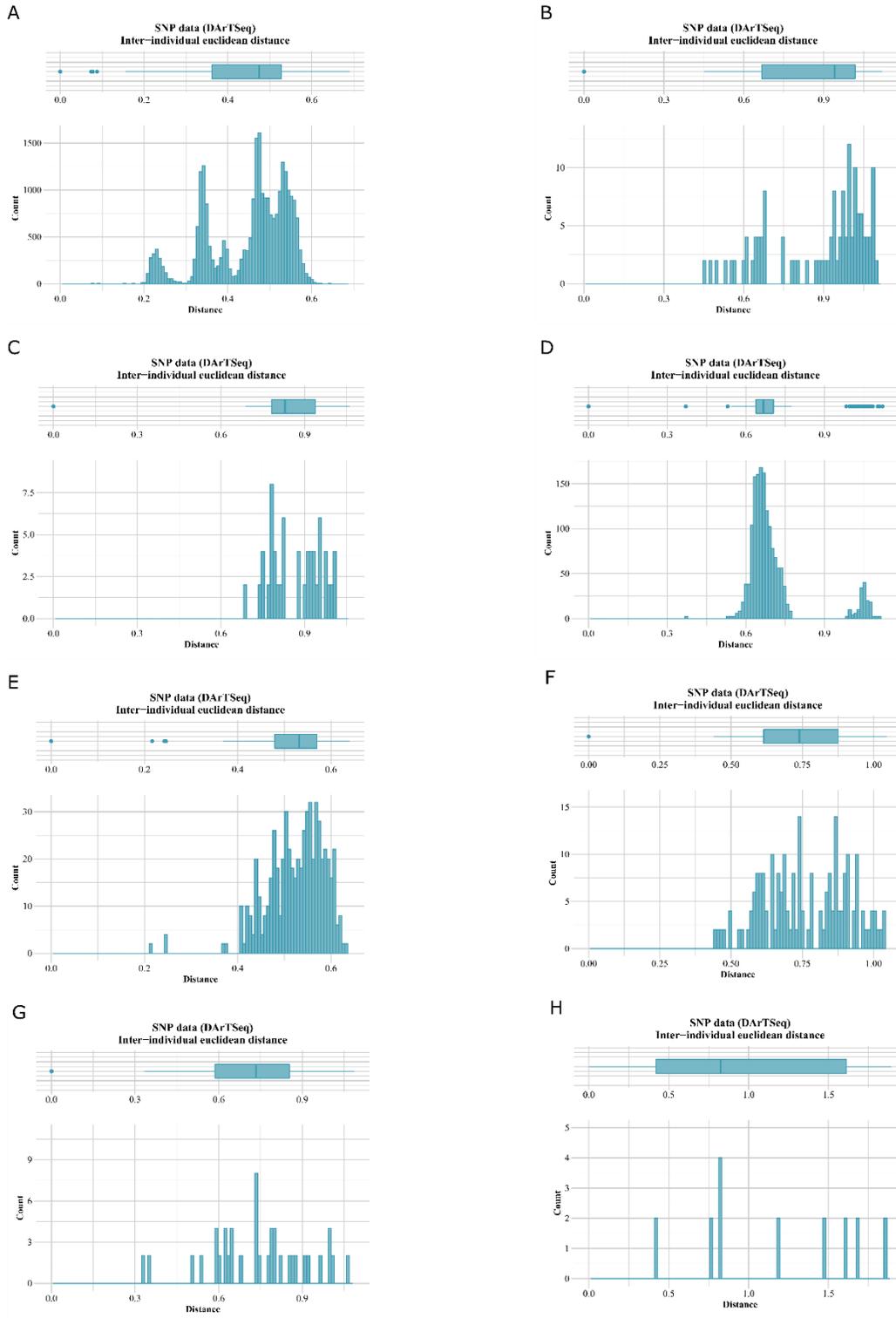
Supplementary Figure 11: *Manis pentadactyla* ADMIXTURE bar graphs of all Ks (K=2-10) in which samples belonging to countries are clumped and each site type is labelled (816 SNPs, 39 individuals). Samples are ordered from South-East to North-West in each country.



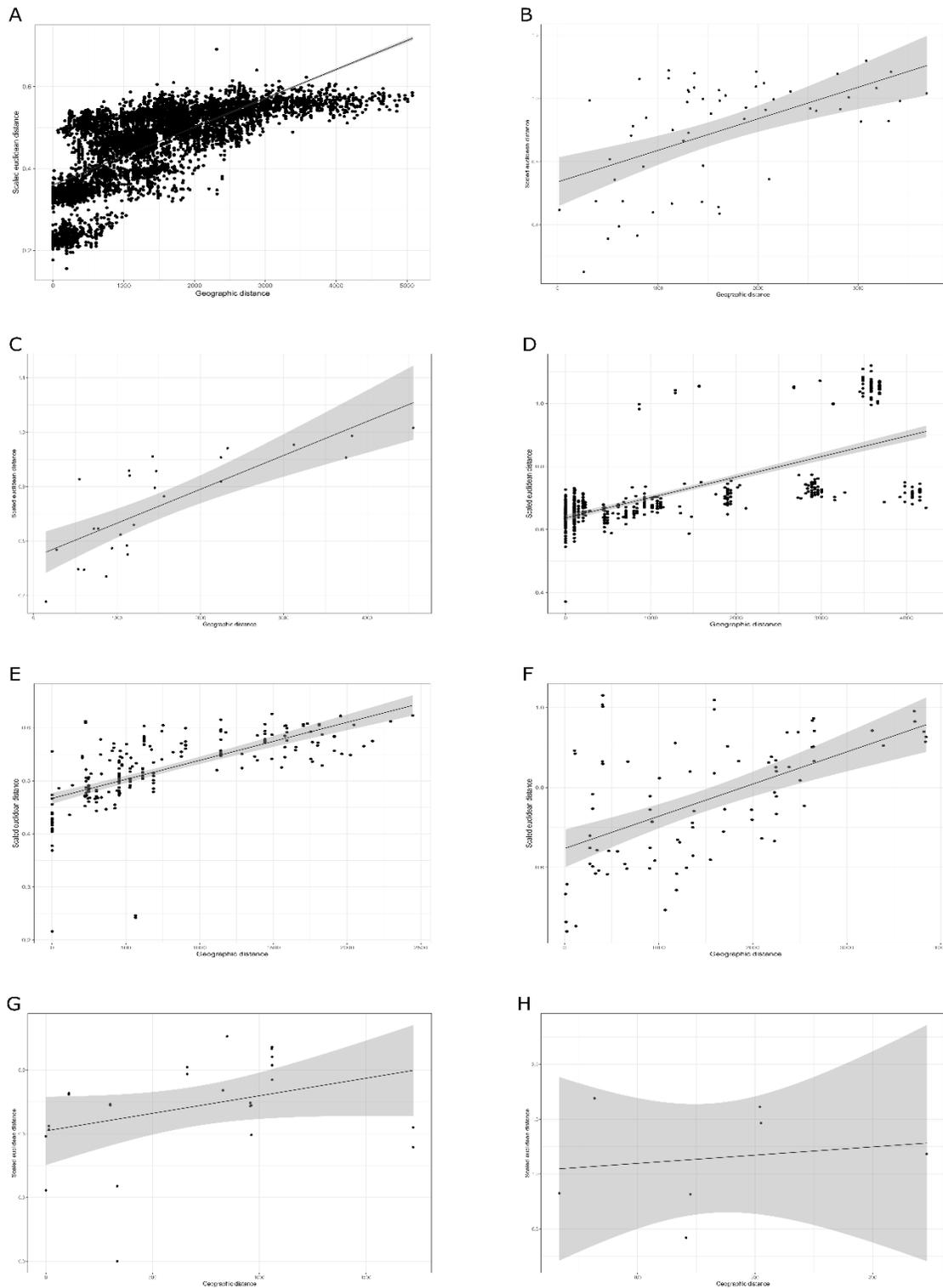
Supplementary Figure 12: *Manis crassicaudata* ADMIXTURE bar graphs of all Ks (K=2-9) in which samples belonging to countries are clumped and each site type is labelled (73 SNPs, 9 individuals). Samples are ordered from South-East to North-West in each country.



Supplementary Figure 13: *Manis culionensis* ADMIXTURE bar graphs of all Ks (K=2-5) in which samples belonging to countries are clumped and each site type is labelled (39 SNPs, 5 individuals). Samples are ordered from South-West to North-East in each island.



Supplementary Figure 14: Boxplots and histograms representing pairwise Euclidian distances across all individuals from each of the eight pangolin species. (A) *Phataginus tricuspis*, (B) *Phataginus tetradactyla*, (C) *Smutsia gigantea*, (D) *Smutsia temminckii*, (E) *Manis javanica*, (F) *Manis pentadactyla*, (G) *Manis crassicaudata*, and (H) *Manis culionensis*.



Supplementary Figure 15: Isolation by distance (IBD) plots comparing pairwise genetic divergence (Euclidean distance) with pairwise geographic distance (km) across all individuals from each of the eight pangolin species. (A) *Phataginus tricuspis*, (B) *Phataginus tetradactyla*, *Smutsia gigantea*, (D) *Smutsia temminckii*, (E) *Manis javanica*, (F) *Manis pentadactyla*, (G) *Manis crassicaudata*, and (H) *Manis culionensis*.

References

- 't Sas-Rolfes, M., Challender, D.W.S., 2020. Evaluating the impact of pangolin farming on conservation, In *Pangolins: Science, Society and Conservation*. eds D.W.S. Challender, H.C. Nash, C. Waterman, pp. 517-527. Academic Press, London, UK.
- Abascal, F., Corvelo, A., Cruz, F., Villanueva-Cañas, J.L., Vlasova, A., Marcet-Houben, M., Martínez-Cruz, B., Cheng, J.Y., Prieto, P., Quesada, V., Quilez, J., Li, G., García, F., Rubio-Camarillo, M., Frias, L., Ribeca, P., Capella-Gutiérrez, S., Rodríguez, J.M., Câmara, F., Lowy, E., Cozzuto, L., Erb, I., Tress, M.L., Rodríguez-Ales, J.L., Ruiz-Orera, J., Reverter, F., Casas-Marce, M., Soriano, L., Arango, J.R., Derdak, S., Galán, B., Blanc, J., Gut, M., Lorente-Galdos, B., Andrés-Nieto, M., López-Otín, C., Valencia, A., Gut, I., García, J.L., Guigó, R., Murphy, W.J., Ruiz-Herrera, A., Marques-Bonet, T., Roma, G., Notredame, C., Mailund, T., Albà, M.M., Gabaldón, T., Alioto, T., Godoy, J.A., 2016. Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx. *Genome biology* 17, 251.
- Aguillon, S., Din Dipita, A., Lecompte, E., Missoup, A.D., 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.
- Akella, A.S., Allan, C., 2012. Dismantling wildlife crime, p. 15. TRAFFIC, Washington DC, USA.
- Alba, D.M., Hammond, A.S., Vinuesa, V., Casanovas-Vilar, I., 2018. First record of a Miocene Pangolin (Pholidota, Manoidea) from the Iberian Peninsula. *Journal of Vertebrate Paleontology* 38.
- Albrechtsen, A., Nielsen, F.C., Nielsen, R., 2010. Ascertainment Biases in SNP Chips Affect Measures of Population Divergence. *Molecular Biology and Evolution* 27, 2534-2547.
- Alexander, D.H., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19, 1655-1664.
- Ali, J.R., 2018. Islands as biological substrates: Continental. *Journal of Biogeography* 45, 1003-1018.
- Allen, G.M., 1938. The mammals of China and Mongolia, In *Central Asiatic Expeditions, 1921-1930*. American Museum of Natural History.
- Allen, G.M., 1939. A checklist of African mammals. [S.n.]. Boston, Mass.
- Allen, J.A., Owston, A., 1906. Mammals from the island of Hainan, China. order of the Trustees, American Museum of Natural History.
- Allendorf, F.W., 2017. Genetics and the conservation of natural populations: allozymes to genomes. *Molecular Ecology* 26, 420-430.
- Allendorf, F.W., Hard, J.J., 2009. Human-induced evolution caused by unnatural selection through harvest of wild animals. *Proceedings of the National Academy of Sciences* 106, 9987-9994.
- Allendorf, F.W., Hohenlohe, P.A., Luikart, G., 2010. Genomics and the future of conservation genetics. *Nature Reviews Genetics* 11, 697-709.
- Allio, R., Tilak, M.-K., Scornavacca, C., Avenant, N.L., Kitchener, A.C., Corre, E., Nabholz, B., Delsuc, F., 2021. High-quality carnivoran genomes from roadkill samples enable comparative species delineation in aardwolf and bat-eared fox. *Elife* 10, e63167.
- Andrews, K.R., De Barba, M., Russello, M.A., Waits, L.P., 2021. Advances in Using Non-invasive, Archival, and Environmental Samples for Population Genomic Studies, In

- Population Genomics: Wildlife. eds P.A. Hohenlohe, O.P. Rajora, pp. 63-99. Springer International Publishing, Cham.
- Anthony, N.M., Johnson-Bawe, M., Jeffery, K., Clifford, S.L., Abernethy, K.A., Tutin, C.E., Lahm, S.A., White, L.J.T., Utley, J.F., Wickings, E.J., Bruford, M.W., 2007. The role of Pleistocene refugia and rivers in shaping gorilla genetic diversity in central Africa. *Proceedings of the National Academy of Sciences* 104, 20432-20436.
- Ardoin, N.M., Bowers, A.W., Gaillard, E., 2020. Environmental education outcomes for conservation: A systematic review. *Biological Conservation* 241, 108224.
- Arenas, M., Pereira, F., Oliveira, M., Pinto, N., Lopes, A.M., Gomes, V., Carracedo, A., Amorim, A., 2017. Forensic genetics and genomics: Much more than just a human affair. *PLoS Genetics* 13, e1006960.
- Arredondo, A., Mourato, B., Nguyen, K., Boitard, S., Rodríguez, W., Noûs, C., Mazet, O., Chikhi, L., 2021. Inferring number of populations and changes in connectivity under the n-island model. *Heredity* 126, 896-912.
- Bairoch, A., Apweiler, R., 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Research* 28, 45-48.
- Baiyewu, A.O., Boakye, M.K., Kotzé, A., Dalton, D.L., Jansen, R., 2018. Ethnozoological survey of traditional uses of Temminck's ground pangolin (*Smutsia temminckii*) in South Africa. *Society and Animals* 26, 306-325.
- Bakhtiari, F., Jacobsen, J.B., Thorsen, B.J., Lundhede, T.H., Strange, N., Boman, M., 2018. Disentangling Distance and Country Effects on the Value of Conservation across National Borders. *Ecological Economics* 147, 11-20.
- Balakrishnan, C.N., Monfort, S.L., Gaur, A., Singh, L., Sorenson, M.D., 2003. Phylogeography and conservation genetics of Eld's deer (*Cervus eldi*). *Molecular Ecology* 12, 1-10.
- Battey, C.J., Ralph, P.L., Kern, A.D., 2020. Predicting geographic location from genetic variation with deep neural networks. *Elife* 9, e54507.
- Beaumont, M.A., Nichols, R.A., 1996. Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 263, 1619-1626.
- Begun, D.R., Nargolwalla, M.C., Kordos, L., 2012. European Miocene Hominids and the Origin of the African Ape and Human Clade. *Evolutionary Anthropology: Issues, News, and Reviews* 21, 10-23.
- Bennett, E.L., Underwood, F.M., Milner-Gulland, E.J., 2021. To Trade or Not to Trade? Using Bayesian Belief Networks to Assess How to Manage Commercial Wildlife Trade in a Complex World. *Frontiers in Ecology and Evolution* 9.
- Benton, M.J., Donoghue, P.C., Asher, R.J., Friedman, M., Near, T.J., Vinther, J., 2015. Constraints on the timescale of animal evolutionary history. *Palaeontol Electron* 18, 1-106.
- Berec, M., Vršecká, L., Šetlíková, I., 2018. What is the reality of wildlife trade volume? CITES Trade Database limitations. *Biological Conservation* 224, 111-116.
- Bertola, L.D., Jongbloed, H., van der Gaag, K.J., de Knijff, P., Yamaguchi, N., Hooghiemstra, H., Bauer, H., Henschel, P., White, P.A., Driscoll, C.A., Tende, T., Ottosson, U., Saidu, Y., Vrieling, K., de Iongh, H.H., 2016. Phylogeographic Patterns in Africa and High Resolution Delineation of Genetic Clades in the Lion (*Panthera leo*). *Scientific Reports* 6, 30807.

- Bird, M.I., Taylor, D., Hunt, C., 2005. Palaeoenvironments of insular Southeast Asia during the Last Glacial Period: a savanna corridor in Sundaland? *Quaternary Science Reviews* 24, 2228-2242.
- Blair, M.E., Le, M.D., Sethi, G., Thach, H.M., Nguyen, V.T.H., Amato, G., Birchette, M., Sterling, E.J., 2017. The importance of an interdisciplinary research approach to inform wildlife trade management in Southeast Asia. *BioScience* 67, 995-1003.
- Blom, M.P.K., Bragg, J.G., Potter, S., Moritz, C., 2016. Accounting for uncertainty in gene tree estimation: Summary-coalescent species tree inference in a challenging radiation of Australian lizards. *Systematic Biology* 66, 352-366.
- Bohling, J., 2020. Evaluating the effect of reference genome divergence on the analysis of empirical RADseq datasets. *Ecology and Evolution* 10, 7585-7601.
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114-2120.
- Bonnald, J., Utge, J., Kuhner, M.K., Wasser, S.K., Asalu, E., Okimat, J.P., Krief, S., 2021. Who are the elephants living in the hybridization zone? How genetics may guide conservation to better protect endangered elephants. *Global Ecology and Conservation* 32, e01917.
- Bornmann, L., Mutz, R., 2015. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology* 66, 2215-2222.
- Borowiec, M.L., 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* 4, e1660-e1660.
- Borzée, A., McNeely, J., Magellan, K., Miller, J.R.B., Porter, L., Dutta, T., Kadinjappalli, K.P., Sharma, S., Shahabuddin, G., Aprilinayati, F., Ryan, G.E., Hughes, A., Abd Mutalib, A.H., Wahab, A.Z.A., Bista, D., Chavanich, S.A., Chong, J.L., Gale, G.A., Ghaffari, H., Ghimirey, Y., Jayaraj, V.K., Khatiwada, A.P., Khatiwada, M., Krishna, M., Lwin, N., Paudel, P.K., Sadykova, C., Savini, T., Shrestha, B.B., Strine, C.T., Sutthacheep, M., Wong, E.P., Yeemin, T., Zahirudin, N.Z., Zhang, L., 2020. COVID-19 highlights the need for more effective wildlife trade legislation. *Trends in ecology & evolution*.
- Bossuyt, F., Meegaskumbura, M., Beenaerts, N., Gower, D.J., Pethiyagoda, R., Roelants, K., Mannaert, A., Wilkinson, M., Bahir, M.M., Manamendra-Arachchi, K., Ng, P.K.L., Schneider, C.J., Oommen, O.V., Milinkovitch, M.C., 2004. Local endemism within the Western Ghats-Sri Lanka biodiversity hotspot. *Science* 306, 479-481.
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: A software platform for bayesian evolutionary analysis. *PLOS Computational Biology* 10, e1003537.
- Brandies, P., Peel, E., Hogg, C.J., Belov, K., 2019. The Value of Reference Genomes in the Conservation of Threatened Species. *Genes* 10, 846.
- Brashares, J.S., Abrahms, B., Fiorella, K.J., Golden, C.D., Hojnowski, C.E., Marsh, R.A., McCauley, D.J., Nuñez, T.A., Seto, K., Withey, L., 2014. Wildlife decline and social conflict. *Science* 345, 376-378.
- Brook, S., de Groot, P.v.C., Scott, C., Boag, P., Long, B., Ley, R., Reischer, G., Williams, A., Mahood, S., Hien, T.M., 2012. Integrated and novel survey methods for rhinoceros populations confirm the extinction of *Rhinoceros sondaicus annamiticus* from Vietnam. *Biological Conservation* 155, 59-67.
- Bruce, T., Kamta, R., Mbobda, R.B.T., Kanto, S.T., Djibrilla, D., Moses, I., Deblauwe, V., Njabo, K., Lebreton, M., Ndjassi, C., Barichievy, C., Olson, D., 2018. Locating giant ground

- pangolins (*Smutsia gigantea*) using camera traps on burrows in the Dja Biosphere Reserve, Cameroon. *Tropical Conservation Science* 11, 1-5.
- Cabana, F., Plowman, A., Van Nguyen, T., Chin, S.C., Wu, S.L., Lo, H.Y., Watabe, H., Yamamoto, F., 2017. Feeding Asian pangolins: An assessment of current diets fed in institutions worldwide. *Zoo Biology* 36, 298-305.
- Cao, P., Dai, Q., Deng, C., Zhao, X., Qin, S., Yang, J., Ju, R., Wang, Z., Lu, G., Gu, X., Yang, Z., Zhu, L., 2021. Genome-wide signatures of mammalian skin covering evolution. *Science China Life Sciences* 64, 1765-1780.
- Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972-1973.
- Carbone, L., Alan Harris, R., Gnerre, S., Veeramah, K.R., Lorente-Galdos, B., Huddleston, J., Meyer, T.J., Herrero, J., Roos, C., Aken, B., Anaclerio, F., Archidiacono, N., Baker, C., Barrell, D., Batzer, M.A., Beal, K., Blancher, A., Bohrsen, C.L., Brameier, M., Campbell, M.S., Capozzi, O., Casola, C., Chiatante, G., Cree, A., Damert, A., de Jong, P.J., Dumas, L., Fernandez-Callejo, M., Flicek, P., Fuchs, N.V., Gut, I., Gut, M., Hahn, M.W., Hernandez-Rodriguez, J., Hillier, L.W., Hubley, R., Ianc, B., Izsvák, Z., Jablonski, N.G., Johnstone, L.M., Karimpour-Fard, A., Konkel, M.K., Kostka, D., Lazar, N.H., Lee, S.L., Lewis, L.R., Liu, Y., Locke, D.P., Mallick, S., Mendez, F.L., Muffato, M., Nazareth, L.V., Nevenon, K.A., O'Bleness, M., Ochis, C., Odom, D.T., Pollard, K.S., Quilez, J., Reich, D., Rocchi, M., Schumann, G.G., Searle, S., Sikela, J.M., Skollar, G., Smit, A., Sonmez, K., Hallers, B.t., Terhune, E., Thomas, G.W.C., Ullmer, B., Ventura, M., Walker, J.A., Wall, J.D., Walter, L., Ward, M.C., Wheelan, S.J., Whelan, C.W., White, S., Wilhelm, L.J., Woerner, A.E., Yandell, M., Zhu, B., Hammer, M.F., Marques-Bonet, T., Eichler, E.E., Fulton, L., Fronick, C., Muzny, D.M., Warren, W.C., Worley, K.C., Rogers, J., Wilson, R.K., Gibbs, R.A., 2014. Gibbon genome and the fast karyotype evolution of small apes. *Nature* 513, 195-201.
- Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P., Narwani, A., Mace, G.M., Tilman, D., Wardle, D.A., 2012. Biodiversity loss and its impact on humanity. *Nature* 486, 59-67.
- Carneiro, M., Rubin, C.-J., Di Palma, F., Albert, F.W., Alföldi, J., Martinez Barrio, A., Pielberg, G., Rafati, N., Sayyab, S., Turner-Maier, J., Younis, S., Afonso, S., Aken, B., Alves, J.M., Barrell, D., Bolet, G., Boucher, S., Burbano, H.A., Campos, R., Chang, J.L., Duranthon, V., Fontanesi, L., Garreau, H., Heiman, D., Johnson, J., Mage, R.G., Peng, Z., Queney, G., Rogel-Gaillard, C., Ruffier, M., Searle, S., Villafuerte, R., Xiong, A., Young, S., Forsberg-Nilsson, K., Good, J.M., Lander, E.S., Ferrand, N., Lindblad-Toh, K., Andersson, L., 2014. Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science* 345, 1074-1079.
- Carstens, B.C., Pelletier, T.A., Reid, N.M., Satler, J.D., 2013. How to fail at species delimitation. *Molecular Ecology* 22, 4369-4383.
- Ceballos, G., Ehrlich, P.R., Raven, P.H., 2020. Vertebrates on the brink as indicators of biological annihilation and the sixth mass extinction. *Proceedings of the National Academy of Sciences* 117, 13596-13602.
- Ceballos, G., García, A., Ehrlich, P.R., 2010. The sixth extinction crisis: Loss of animal populations and species. *Journal of Cosmology* 8, 31.
- Chaber, A.-L., Cunningham, A., 2016. Public health risks from illegally imported african bushmeat and smoked fish. *EcoHealth* 13, 135-138.

- Chalk, T.B., Hain, M.P., Foster, G.L., Rohling, E.J., Sexton, P.F., Badger, M.P.S., Cherry, S.G., Hasenfratz, A.P., Haug, G.H., Jaccard, S.L., Martínez-García, A., Pälke, H., Pancost, R.D., Wilson, P.A., 2017. Causes of ice age intensification across the Mid-Pleistocene Transition. *Proceedings of the National Academy of Sciences* 114, 13114-13119.
- Challender, D., Waterman, C., 2017. Implementation of CITES Decisions 17.239 B) and 17.240 on Pangolins (*Manis spp.*), In Prepared by IUCN for the CITES Secretariat. SC69 Doc 57.
- Challender, D., Waterman, C., Baillie, J., 2014. Scaling up pangolin conservation. IUCN SSC pangolin specialist group conservation action plan, Zoological Society of London, London, UK.
- Challender, D., Wu, S., Kaspal, P., Khatiwada, A., Ghose, A., Ching-Min Su, N., Laxmi Suwal, T., 2019a. *Manis pentadactyla* (errata version published in 2020). The IUCN Red List of Threatened Species 2019, e.T12764A168392151.
- Challender, D.W., Ades, G.W., Chin, J.S., Sun, N.C.-M., Iian Chong, J., Connelly, E., Hywood, L., Luz, S., Mohapatra, R.K., de Ornellas, P., 2019b. Evaluating the feasibility of pangolin farming and its potential conservation impact. *Global Ecology and Conservation* 20, e00714.
- Challender, D.W., Heinrich, S., Shepherd, C.R., Katsis, L.K., 2020a. International trade and trafficking in pangolins, 1900–2019, In *Pangolins: Science, Society and Conservation*. eds D. Challender, H. Nash, C. Waterman, pp. 259-276. Academic Press, London, UK.
- Challender, D.W., Nash, H.C., Waterman, C., 2020b. *Pangolins: Science, Society and Conservation*. Academic Press, London, UK.
- Challender, D.W.S., Harrop, S.R., MacMillan, D.C., 2015. Understanding markets to conserve trade-threatened species in CITES. *Biological Conservation* 187, 249-259.
- Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4.
- Chang, J.-M., Di Tommaso, P., Notredame, C., 2014. TCS: A New Multiple Sequence Alignment Reliability Measure to Estimate Alignment Accuracy and Improve Phylogenetic Tree Reconstruction. *Molecular Biology and Evolution* 31, 1625-1637.
- Chao, J.-T., Li, H.-F., Lin, C.-C., 2020. Chapter 3 - The role of pangolins in ecosystems, In *Pangolins*. eds D.W.S. Challender, H.C. Nash, C. Waterman, pp. 43-48. Academic Press.
- Chapman, C.A., Abernathy, K., Chapman, L.J., Downs, C., Effiom, E.O., Gogarten, J.F., Golooba, M., Kalbitzer, U., Lawes, M.J., Mekonnen, A., Omeja, P., Razafindratsima, O., Sheil, D., Tabor, G.M., Tumwesigye, C., Sarkar, D., 2022. The future of sub-Saharan Africa's biodiversity in the face of climate and societal change. *Frontiers in Ecology and Evolution* 10.
- Che, J., Wang, J., Su, W., Ye, J., Wang, Y., Nie, W., Yang, F., 2008. Construction, characterization and FISH mapping of a bacterial artificial chromosome library of Chinese pangolin (*Manis pentadactyla*). *Cytogenetic and Genome Research* 122, 55-60.
- Che, J., Zhou, W.-W., Hu, J.-S., Yan, F., Papenfuss, T.J., Wake, D.B., Zhang, Y.-P., 2010. Spiny frogs (Paini) illuminate the history of the Himalayan region and Southeast Asia. *Proceedings of the National Academy of Sciences* 107, 13765-13770.
- Chen, L., Qiu, Q., Jiang, Y., Wang, K., Lin, Z., Li, Z., Bibi, F., Yang, Y., Wang, J., Nie, W., Su, W., Liu, G., Li, Q., Fu, W., Pan, X., Liu, C., Yang, J., Zhang, C., Yin, Y., Wang, Y., Zhao, Y., Zhang, C., Wang, Z., Qin, Y., Liu, W., Wang, B., Ren, Y., Zhang, R., Zeng, Y., da Fonseca, R.R., Wei, B., Li, R., Wan, W., Zhao, R., Zhu, W., Wang, Y., Duan, S., Gao, Y., Zhang, Y.E., Chen, C., Hvilsom, C., Epps, C.W., Chemnick, L.G., Dong, Y., Mirarab, S.,

- Siegismund, H.R., Ryder, O.A., Gilbert, M.T.P., Lewin, H.A., Zhang, G., Heller, R., Wang, W., 2019. Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. *Science* 364, eaav6202.
- Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884-i890.
- Chin, S.C., Lien, C.Y., Chan, Y.T., Chen, C.L., Yang, Y.C., Yeh, L.S., 2012. Monitoring the gestation period of rescued Formosan pangolin (*Manis pentadactyla pentadactyla*) with progesterone radioimmunoassay. *Zoo Biology* 31, 479-489.
- Chinese Pharmacopoeia Commission, 2015. Pharmacopoeia of the People's Republic of China 2015. Medical Science and Technology Press, Beijing, China.
- Chng, S.C., Eaton, J.A., Krishnasamy, K., Shepherd, C.R., Nijman, V., 2015. In the Market for Extinction: An inventory of Jakarta's bird markets. Traffic Report, 31.
- Cho, Y.S., Hu, L., Hou, H., Lee, H., Xu, J., Kwon, S., Oh, S., Kim, H.-M., Jho, S., Kim, S., Shin, Y.-A., Kim, B.C., Kim, H., Kim, C.-u., Luo, S.-J., Johnson, W.E., Koepfli, K.-P., Schmidt-Küntzel, A., Turner, J.A., Marker, L., Harper, C., Miller, S.M., Jacobs, W., Bertola, L.D., Kim, T.H., Lee, S., Zhou, Q., Jung, H.-J., Xu, X., Gadhvi, P., Xu, P., Xiong, Y., Luo, Y., Pan, S., Gou, C., Chu, X., Zhang, J., Liu, S., He, J., Chen, Y., Yang, L., Yang, Y., He, J., Liu, S., Wang, J., Kim, C.H., Kwak, H., Kim, J.-S., Hwang, S., Ko, J., Kim, C.-B., Kim, S., Bayarlkhagva, D., Paek, W.K., Kim, S.-J., O'Brien, S.J., Wang, J., Bhak, J., 2013. The tiger genome and comparative analysis with lion and snow leopard genomes. *Nature Communications* 4, 2433.
- Chong, J.L., Panjang, E., Willcox, D., Nash, H.C., Semiadi, G., Sodsai, W., Lim, N.T.L., Fletcher, L., Kurniawan, A., Cheema, S., 2020. Chapter 6 - Sunda pangolin *Manis javanica* (Desmarest, 1822), In Pangolins. eds D.W.S. Challender, H.C. Nash, C. Waterman, pp. 89-108. Academic Press.
- Choo, S.W., Rayko, M., Tan, T.K., Hari, R., Komissarov, A., Wee, W.Y., Yurchenko, A.A., Kliver, S., Tamazian, G., Antunes, A., Wilson, R.K., Warren, W.C., Koepfli, K.P., Minx, P., Krashennnikova, K., Kotze, A., Dalton, D.L., Vermaak, E., Paterson, I.C., Dobrynin, P., Sitam, F.T., Rovie-Ryan, J.J., Johnson, W.E., Yusoff, A.M., Luo, S.J., Karuppanan, K.V., Fang, G., Zheng, D., Gerstein, M.B., Lipovich, L., O'Brien, S.J., Wong, G.J., 2016. Pangolin genomes and the evolution of mammalian scales and immunity. *Genome Research* 26, 1312-1322.
- Chowdhury, E.K., Khan, I.I., Dhar, B.K., 2021. Catastrophic impact of Covid-19 on the global stock markets and economic activities. *Business and Society Review* n/a.
- Christie, A.P., Amano, T., Martin, P.A., Petrovan, S.O., Shackelford, G.E., Simmons, B.I., Smith, R.K., Williams, D.R., Wordley, C.F.R., Sutherland, W.J., 2020. Poor availability of context-specific evidence hampers decision-making in conservation. *Biological Conservation* 248, 108666.
- Christie, A.P., White, T.B., Martin, P.A., Petrovan, S.O., Bladon, A.J., Bowkett, A.E., Littlewood, N.A., Mupepele, A.-C., Rocha, R., Sainsbury, K.A., Smith, R.K., Taylor, N.G., Sutherland, W.J., 2021. Reducing publication delay to improve the efficiency and impact of conservation science. *PeerJ* 9, e12245.
- CITES, 2017. Amendments to the conservation of and trade in pangolins, In Seventeenth meeting of the Conference of the Parties Johannesburg, South Africa, 24 September – 5 October 2016. CITES Secretariat.

- CITES, 2022a. Amendments to resolution Conf. 17.10. , In Nineteenth meeting of the Conference of the Parties Panama City (Panama), 14 – 25 November 2022. CITES Secretariat.
- CITES, 2022b. Consideration of proposals for amendment of Appendices I and II - CoP19 Prop. 2, In Nineteenth meeting of the Conference of the Parties Panama (Panama City), 14 – 25 November, 2022. CITES Secretariat.
- Clark, L., Van Thai, N., Phuong, T.Q., 2009. A long way from home: the health status of Asian pangolins confiscated from the illegal wildlife trade in Viet Nam, In Proceedings of the workshop on trade and conservation of pangolins native to south and southeast Asia: 30 June-2 July 2008. p. 111. TRAFFIC Southeast Asia, Singapore Zoo, Singapore.
- Cleaveland, S., Mlengeya, T., Kaare, M., Haydon, D., Lembo, T., Laurenson, M.K., Packer, C., 2007. The conservation relevance of epidemiological research into carnivore viral diseases in the Serengeti. *Conservation Biology* 21, 612-622.
- Clements, C.F., 2013. Public interest in the extinction of a species may lead to an increase in donations to a large conservation charity. *Biodiversity and Conservation* 22, 2695-2699.
- Coimbra, R.T.F., Winter, S., Kumar, V., Koepfli, K.-P., Gooley, R.M., Dobrynin, P., Fennessy, J., Janke, A., 2021. Whole-genome analysis of giraffe supports four distinct species. *Current Biology* 31, 2929-2938.e2925.
- Collaboration for Environmental Evidence, 2018. Guidelines and Standards for Evidence synthesis in Environmental Management. Version 5.0, eds A. Pullin, G. Frampton, B. Livoreil, G. Petrokofsky.
- Conenna, I., Rocha, R., Russo, D., Cabeza, M., 2017. Insular bats and research effort: a review of global patterns and priorities. *Mammal Review* 47, 169-182.
- Connelly, E., Donaldson, M.H., Hywood, L., Van Der Meer, E., 2019. Case study: Anesthesia and veterinary care of rescued Temminck's ground pangolins (*Smutsia temminckii*). *Journal of Wildlife Rehabilitation* 39, 13-16.
- Consortium, T.M.G.S.a.A., 2014. The common marmoset genome provides insight into primate biology and evolution. *Nature Genetics* 46, 850.
- Cook, C.N., Possingham, H.P., Fuller, R.A., 2013. Contribution of systematic reviews to management decisions. *Conservation Biology* 27, 902-915.
- Corbett-Detig, R.B., Hartl, D.L., Sackton, T.B., 2015. Natural selection constrains neutral diversity across a wide range of species. *PLoS biology* 13, e1002112.
- Cornuet, J.M., Luikart, G., 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* 144, 2001-2014.
- Costello, M.J., May, R.M., Stork, N.E., 2013. Can We Name Earth's Species Before They Go Extinct? *Science* 339, 413-416.
- Cota-Larson, R., 2017. Pangolin species identification guide: a rapid assessment tool for field and desk. Prepared for the United States Agency for International Development. USAID Wildlife Asia Activity, Bangkok. USAID Wildlife Asia Activity. Available from. <http://www.usaidwildlifeasia>
- Cruz-Dávalos, D.I., Llamas, B., Gaunitz, C., Fages, A., Gamba, C., Soubrier, J., Librado, P., Seguin-Orlando, A., Pruvost, M., Alfarhan, A.H., Alquraishi, S.A., Al-Rasheid, K.A.S., Scheu, A., Beneke, N., Ludwig, A., Cooper, A., Willerslev, E., Orlando, L., 2017. Experimental conditions improving in-solution target enrichment for ancient DNA. *Molecular ecology resources* 17, 508-522.
- Cvitanovic, C., Fulton, C.J., Wilson, S.K., van Kerkhoff, L., Cripps, I.L., Muthiga, N., 2014. Utility of primary scientific literature to environmental managers: An international case

- study on coral-dominated marine protected areas. *Ocean & Coastal Management* 102, 72-78.
- Daley, T., Smith, A.D., 2013. Predicting the molecular complexity of sequencing libraries. *Nature Methods* 10, 325-327.
- Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., Group, G.P.A., 2011. The variant call format and VCFtools. *Bioinformatics* 27, 2156-2158.
- Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., Li, H., 2021. Twelve years of SAMtools and BCFtools. *GigaScience* 10.
- Darriba, D., Posada, D., Kozlov, A.M., Stamatakis, A., Morel, B., Flouri, T., 2019. ModelTest-NG: A new and scalable tool for the selection of DNA and protein evolutionary models. *Molecular Biology and Evolution* 37, 291-294.
- Daszak, P., Olival, K.J., Li, H., 2020. A strategy to prevent future epidemics similar to the 2019-nCoV outbreak. *Biosafety and Health* 2, 6-8.
- Dauby, G., Duminil, J., Heuertz, M., Koffi, G.K., Stévant, T., Hardy, O.J., 2014. Congruent phylogeographical patterns of eight tree species in Atlantic Central Africa provide insights into the past dynamics of forest cover. *Molecular Ecology* 23, 2299-2312.
- de Bruyn, M., Nugroho, E., Hossain, M.M., Wilson, J.C., Mather, P.B., 2005. Phylogeographic evidence for the existence of an ancient biogeographic barrier: the Isthmus of Kra Seaway. *Heredity* 94, 370-378.
- de Jager, D., Glanzmann, B., Möller, M., Hoal, E., van Helden, P., Harper, C., Bloomer, P., 2021. High diversity, inbreeding and a dynamic Pleistocene demographic history revealed by African buffalo genomes. *Scientific Reports* 11, 4540.
- Deméno, B.B., Doucet, J.-L., Hardy, O.J., 2018. History of the fragmentation of the African rain forest in the Dahomey Gap: insight from the demographic history of *Terminalia superba*. *Heredity* 120, 547-561.
- Deméno, B.B., Piñeiro, R., Hardy, O.J., 2016. Origin and history of the Dahomey Gap separating West and Central African rain forests: insights from the phylogeography of the legume tree *Distemonanthus benthamianus*. *Journal of Biogeography* 43, 1020-1031.
- Deng, T., Wu, F., Wang, S., Su, T., Zhou, Z., 2021. Major turnover of biotas across the Oligocene/Miocene boundary on the Tibetan Plateau. *Palaeogeography, Palaeoclimatology, Palaeoecology* 567, 110241.
- Derkarabetian, S., Benavides, L.R., Giribet, G., 2019. Sequence capture phylogenomics of historical ethanol-preserved museum specimens: Unlocking the rest of the vault. *Molecular ecology resources* 19, 1531-1544.
- Di Minin, E., Brooks, T.M., Toivonen, T., Butchart, S.H.M., Heikinheimo, V., Watson, J.E.M., Burgess, N.D., Challender, D.W.S., Goettsch, B., Jenkins, R., Moilanen, A., 2019. Identifying global centers of unsustainable commercial harvesting of species. *Science Advances* 5, eaau2879.
- Dobrynin, P., Liu, S., Tamazian, G., Xiong, Z., Yurchenko, A.A., Krashennikova, K., Kliver, S., Schmidt-Küntzel, A., Koepfli, K.-P., Johnson, W., Kuderna, L.F.K., García-Pérez, R., Manuel, M.d., Godinez, R., Komissarov, A., Makunin, A., Brukhin, V., Qiu, W., Zhou, L., Li, F., Yi, J., Driscoll, C., Antunes, A., Oleksyk, T.K., Eizirik, E., Perelman, P., Roelke, M., Wildt, D., Diekhans, M., Marques-Bonet, T., Marker, L., Bhak, J., Wang, J., Zhang,

- G., O'Brien, S.J., 2015. Genomic legacy of the African cheetah, *Acinonyx jubatus*. *Genome biology* 16, 277-277.
- dos Reis, M., Yang, Z., 2019. Bayesian molecular clock dating using genome-scale datasets, In *Evolutionary Genomics: Statistical and Computational Methods*. ed. M. Anisimova, pp. 309-330. Springer New York, New York, NY.
- Doyle, J.J., 1992. Gene trees and species trees: Molecular systematics as one-character taxonomy. *Systematic Botany* 17, 144-163.
- Droissart, V., Dauby, G., Hardy, O.J., Deblauwe, V., Harris, D.J., Janssens, S., Mackinder, Barbara A., Blach-Overgaard, A., Sonké, B., Sosef, M.S.M., Stévant, T., Svenning, J.-C., Wieringa, J.J., Couvreur, T.L.P., 2018. Beyond trees: Biogeographical regionalization of tropical Africa. *Journal of Biogeography* 45, 1153-1167.
- du Toit, Z., 2014. Population genetic structure of the ground pangolin based on mitochondrial genomes., In *Department of Genetics*. p. 149. University of the Free State, Bloemfontein, South Africa.
- du Toit, Z., Dalton, D.L., du Plessis, M., Jansen, R., Grobler, J.P., Kotzé, A., 2020. Isolation and characterization of 30 STRs in Temminck's ground pangolin (*Smutsia temminckii*) and potential for cross amplification in other African species. *Journal of Genetics* 99, 20.
- du Toit, Z., du Plessis, M., Dalton, D.L., Jansen, R., Paul Grobler, J., Kotzé, A., 2017. Mitochondrial genomes of African pangolins and insights into evolutionary patterns and phylogeny of the family Manidae. *BMC Genomics* 18, 1-13.
- Du Toit, Z., Grobler, J.P., 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.
- Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P., Aiden, E.L., 2017. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* 356, 92-95.
- Dudchenko, O., Shamim, M.S., Batra, S.S., Durand, N.C., Musial, N.T., Mostofa, R., Pham, M., Glenn St Hilaire, B., Yao, W., Stamenova, E., Hoeger, M., Nyquist, S.K., Korchina, V., Pletch, K., Flanagan, J.P., Tomaszewicz, A., McAloose, D., Pérez Estrada, C., Novak, B.J., Omer, A.D., Aiden, E.L., 2018. The Juicebox Assembly Tools module facilitates *de novo* assembly of mammalian genomes with chromosome-length scaffolds for under \$1000. *bioRxiv*, 254797.
- Dufresnes, C., Miquel, C., Remollino, N., Biollaz, F., Salamin, N., Taberlet, P., Fumagalli, L., 2018. Howling from the past: historical phylogeography and diversity losses in European grey wolves. *Proceedings of the Royal Society B: Biological Sciences* 285, 20181148.
- Dyer, R.J., 2014. *gstudio: An package for the spatial analysis of population genetic marker data*. Virginia Commonwealth University, Richmond.
- Eaton, M.J., Meyers, G.L., Kolokotronis, S.-O., Leslie, M.S., Martin, A.P., Amato, G., 2010. Barcoding bushmeat: molecular identification of Central African and South American harvested vertebrates. *Conservation Genetics* 11, 1389-1404.
- Eddy, S.R., 2011. Accelerated Profile HMM Searches. *PLOS Computational Biology* 7, e1002195.
- EIA, 2020. Despite the headlines, China's Government still promotes pangolin scales in traditional medicines. Environmental Investigation Agency, UK.
- Elbadawi, E., Sambanis, N., 2000. Why are there so many civil wars in Africa? Understanding and preventing violent conflict. *Journal of African Economies* 9, 244-269.

- Emogor, C.A., Ingram, D.J., Coad, L., Worthington, T.A., Dunn, A., Imong, I., Balmford, A., 2021. The scale of Nigeria's involvement in the trans-national illegal pangolin trade: Temporal and spatial patterns and the effectiveness of wildlife trade regulations. *Biological Conservation* 264, 109365.
- Emslie, R., 2020a. *Ceratotherium simum ssp. cottoni*. In The IUCN Red List of Threatened Species. p. e.T4183A45813838.
- Emslie, R., 2020b. *Diceros bicornis ssp. longipes*. The IUCN Red List of Threatened Species 2020, e.T39319A45814470.
- Ewart, K.M., Lightson, A.L., Sitam, F.T., Rovie-Ryan, J., Nguyen, S.G., Morgan, K.I., Luczon, A., Anadon, E.M.S., De Bruyn, M., Bourgeois, S., Ouitavon, K., Kotze, A., Bakar, M.S.A., Salgado-Lynn, M., McEwing, R., 2021. DNA analyses of large pangolin scale seizures: Species identification validation and case studies. *Forensic Science International: Animals and Environments* 1, 100014.
- Ewels, P., Magnusson, M., Lundin, S., Källér, M., 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047-3048.
- Ewels, P.A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M.U., Di Tommaso, P., Nahnsen, S., 2020. The nf-core framework for community-curated bioinformatics pipelines. *Nature biotechnology* 38, 276-278.
- Faith, J.T., Tryon, C.A., Peppe, D.J., 2016. Environmental Change, Ungulate Biogeography, and Their Implications for Early Human Dispersals in Equatorial East Africa, In *Africa from MIS 6-2: Population Dynamics and Paleoenvironments*. eds S.C. Jones, B.A. Stewart, pp. 233-245. Springer Netherlands, Dordrecht.
- FATF, 2020. Money Laundering and the Illegal Wildlife Trade. FATF, Paris, France.
- Fennessy, J., Bidon, T., Reuss, F., Kumar, V., Elkan, P., Nilsson, Maria A., Vamberger, M., Fritz, U., Janke, A., 2016. Multi-locus Analyses Reveal Four Giraffe Species Instead of One. *Current Biology* 26, 2543-2549.
- Fernandes, G.A., Caparroz, R., 2013. DNA sequence analysis to guide the release of blue-and-yellow macaws (*Ara ararauna*, Psittaciformes, Aves) from the illegal trade back into the wild. *Molecular Biology Reports* 40, 2757-2762.
- Ferreira-Cardoso, S., Billet, G., Gaubert, P., Delsuc, F., Hautier, L., 2019. Skull shape variation in extant pangolins (Pholidota: Manidae): allometric patterns and systematic implications. *Zoological Journal of the Linnean Society* 188, 255-275.
- Ferreira, S., 2004. Problems associated with tourism development in Southern Africa: The case of Transfrontier Conservation Areas. *GeoJournal* 60, 301-310.
- Figueiró, H.V., Li, G., Trindade, F.J., Assis, J., Pais, F., Fernandes, G., Santos, S.H.D., Hughes, G.M., Komissarov, A., Antunes, A., Trinca, C.S., Rodrigues, M.R., Linderoth, T., Bi, K., Silveira, L., Azevedo, F.C.C., Kantek, D., Ramalho, E., Brassaloti, R.A., Villela, P.M.S., Nunes, A.L.V., Teixeira, R.H.F., Morato, R.G., Loska, D., Saragüeta, P., Gabaldón, T., Teeling, E.C., O'Brien, S.J., Nielsen, R., Coutinho, L.L., Oliveira, G., Murphy, W.J., Eizirik, E., 2017. Genome-wide signatures of complex introgression and adaptive evolution in the big cats. *Science Advances* 3, e1700299.
- Finch, K.N., Cronn, R.C., Ayala Richter, M.C., Blanc-Jolivet, C., Correa Guerrero, M.C., De Stefano Beltrán, L., García-Dávila, C.R., Honorio Coronado, E.N., Palacios-Ramos, S., Paredes-Villanueva, K., Jones, F.A., 2020. Predicting the geographic origin of Spanish Cedar (*Cedrela odorata* L.) based on DNA variation. *Conservation Genetics* 21, 625-639.

- Fischer, H., Tschachler, E., Eckhart, L., 2020. Pangolins lack IFIH1/MDA5, a cytoplasmic RNA sensor that initiates innate immune defense upon coronavirus infection. *Frontiers in Immunology* 11, 939.
- Fitak, R.R., 2021. OptM: estimating the optimal number of migration edges on population trees using Treemix. *Biology Methods and Protocols* 6.
- Fitzinger, L.J., 1872. In *Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften. Mathematisch-Naturwissenschaftliche Classe, CI., LXV, Abth. I.* pp. 9-83.
- Foran, D.R., 2006. Relative Degradation of Nuclear and Mitochondrial DNA: An Experimental Approach*. *Journal of Forensic Sciences* 51, 766-770.
- Fox, R.C., Scott, C.S., Rankin, B.D., 2010. New early carnivoran specimens from the Puercan (Earliest Paleocene) of Saskatchewan, Canada. *Journal of Paleontology* 84, 1035-1039.
- Frankham, R., 2005. Genetics and extinction. *Biological Conservation* 126, 131-140.
- Frutos, R., Serra-Cobo, J., Chen, T., Devaux, C.A., 2020. COVID-19: Time to exonerate the pangolin from the transmission of SARS-CoV-2 to humans. *Infection, Genetics and Evolution* 84, 104493.
- Fuchs, J., Bowie, R.C.K., 2015. Concordant genetic structure in two species of woodpecker distributed across the primary West African biogeographic barriers. *Molecular Phylogenetics and Evolution* 88, 64-74.
- Funk, W.C., McKay, J.K., Hohenlohe, P.A., Allendorf, F.W., 2012. Harnessing genomics for delineating conservation units. *Trends Ecol Evol* 27, 489-496.
- Gagnaire, P.-A., Minegishi, Y., Zenboudji, S., Valade, P., Aoyama, J., Berrebi, P., 2011. Within-population structure highlighted by differential introgression across semipermeable barriers to gene flow in *Anguilla marmorata*. *Evolution* 65, 3413-3427.
- Garcia-Erill, G., Jørgensen, C.H.F., Muwanika, V.B., Wang, X., Rasmussen, M.S., de Jong, Y.A., Gaubert, P., Olayemi, A., Salmons, J., Butynski, T.M., Bertola, L.D., Siegismund, H.R., Albrechtsen, A., Heller, R., 2022. Warthog Genomes Resolve an Evolutionary Conundrum and Reveal Introgression of Disease Resistance Genes. *Molecular Biology and Evolution* 39.
- Garner, B.A., Hand, B.K., Amish, S.J., Bernatchez, L., Foster, J.T., Miller, K.M., Morin, P.A., Narum, S.R., O'Brien, S.J., Roffler, G., 2016. Genomics in conservation: case studies and bridging the gap between data and application. *Trends in ecology & evolution* 31, 81-83.
- 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.
- Gaubert, P., Antunes, A., 2015. What's behind these scales? Comments to "The complete mitochondrial genome of Temminck's ground pangolin (*Smutsia temminckii*; Smuts, 1832) and phylogenetic position of the Pholidota (Weber, 1904)". *Gene* 563, 106-108.
- Gaubert, P., Antunes, A., Meng, H., Miao, L., Peigné, S., Justy, F., Njiokou, F., Dufour, S., Danquah, E., Alahakoon, J., Verheyen, E., Stanley, W.T., O'Brien, S.J., Johnson, W.E., 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., Njiokou, F., Ngua, G., Afiademanyo, K., Dufour, S., Malekani, J., Bi, S.G., Tougard, C., Olayemi, A., Danquah, E., Djagoun, C.A.M.S., Kaleme, P., Mololo, C.N., 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., Wible, J.R., Heighton, S.P., Gaudin, T.J., 2020. Phylogeny and systematics, In *Pangolins: Science, Society and Conservation*. eds D. Challender, H. Nash, C. Waterman, pp. 25-39. Academic Press, London, UK.
- Gaudin, T.J., Emry, R.J., Wible, J.R., 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, T.J., Gaubert, P., Billet, G., Hautier, L., Ferreira-Cardoso, S., Wible, J.R., 2020. Evolution and morphology, In *Pangolins: Science, Society and Conservation*. eds D.W.S. Challender, H.C. Nash, C. Waterman, pp. 5-23. Academic Press, London, UK.
- Ge, R.-L., Cai, Q., Shen, Y.-Y., San, A., Ma, L., Zhang, Y., Yi, X., Chen, Y., Yang, L., Huang, Y., He, R., Hui, Y., Hao, M., Li, Y., Wang, B., Ou, X., Xu, J., Zhang, Y., Wu, K., Geng, C., Zhou, W., Zhou, T., Irwin, D.M., Yang, Y., Ying, L., Bao, H., Kim, J., Larkin, D.M., Ma, J., Lewin, H.A., Xing, J., Platt, R.N., Ray, D.A., Auvil, L., Capitanu, B., Zhang, X., Zhang, G., Murphy, R.W., Wang, J., Zhang, Y.-P., Wang, J., 2013. Draft genome sequence of the Tibetan antelope. *Nature Communications* 4, 1858.
- Gebo, D.L., Rasmussen, D.T., 1985. The earliest fossil pangolin (Pholidota: Manidae) from Africa. *Journal of Mammalogy* 66, 538-541.
- George, K., Masters, A., Dawnay, N., 2019. Development of HyBeacon® probes for the forensic detection of *Panthera*, rhinoceros, and pangolin species. *Molecular and Cellular Probes* 48.
- Ginolhac, A., Rasmussen, M., Gilbert, M.T.P., Willerslev, E., Orlando, L., 2011. mapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics* 27, 2153-2155.
- Goodrich, J., Wibisono, H., Miquelle, D., Lynam, A., Sanderson, E., Chapman, S., Gray, T., Chanchani, P., Harihar, A., 2022. *Panthera tigris*. The IUCN Red List of Threatened Species 2022: e. T15955A214862019.
- Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome, In *Nature biotechnology*. pp. 644-652.
- Grace, D., Mutua, F., Ochungo, P., Kruska, R., Jones, K., Brierley, L., Lapar, L., Said, M., Herrero, M., Phuc, P., Thao, N., 2012. Mapping of poverty and likely zoonoses hotspots. Zoonoses project 4. Report to the UK department for international development, p. 119. International Livestock Research Institute, Nairobi, Kenya.
- Gruber, B., Unmack, P.J., Berry, O.F., Georges, A., 2018. dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Molecular ecology resources* 18, 691-699.
- Gu, Z., 2015. Complexheatmap: Making complex heatmaps, In R package version 2.2.0.
- Gudehus, M., Pietersen, D.W., Hoffmann, M., Cassidy, R., Cassidy, T., Sodeinde, O., Lapuente, J., Assovi, B.G.-M., Shirley, M.H., 2020. Black-bellied pangolin *Phataginus tetradactyla* (Linnaeus, 1766), In *Pangolins: Science, Society and Conservation*. eds D.W.S. Challender, H.C. Nash, C. Waterman, pp. 123-138. Academic Press.
- Günther, T., Nettelblad, C., 2019. The presence and impact of reference bias on population genomic studies of prehistoric human populations. *PLoS Genetics* 15, e1008302.

- Habbal, O., 2017. The science of anatomy: a historical timeline. Sultan Qaboos University medical journal 17, e18-e22.
- Habel, J.C., Schmitt, T., 2012. The burden of genetic diversity. *Biological Conservation* 147, 270-274.
- Haffer, J., 1969. Speciation in Amazonian Forest Birds. *Science* 165, 131-137.
- Hallmann, C.A., Sorg, M., Jongejans, E., Siepel, H., Hofland, N., Schwan, H., Stenmans, W., Müller, A., Sumser, H., Hörrén, T., Goulson, D., de Kroon, H., 2017. More than 75 percent decline over 27 years in total flying insect biomass in protected areas. *PLoS ONE* 12, e0185809.
- Hammer, S., Watson, R., 2012. The challenge of managing Spix Macaws (*Cyanopsitta spixii*) at Qatar - an eleven-year retrospection. *Der Zoologische Garten* 81, 81-95.
- Harcourt, A.H., Wood, M.A., 2012. Rivers as Barriers to Primate Distributions in Africa. *International Journal of Primatology* 33, 168-183.
- Hardy, O.J., Born, C., Budde, K., Dainou, K., Dauby, G., Duminil, J., Ewédjé, E.-E.B.K., Gomez, C., Heuertz, M., Koffi, G.K., Lowe, A.J., Micheneau, C., Ndiade-Bourobou, D., Piñeiro, R., Poncet, V., 2013. Comparative phylogeography of African rain forest trees: A review of genetic signatures of vegetation history in the Guineo-Congolian region. *Comptes Rendus Geoscience* 345, 284-296.
- Harfoot, M., Glaser, S.A., Tittensor, D.P., Britten, G.L., McLardy, C., Malsch, K., Burgess, N.D., 2018. Unveiling the patterns and trends in 40 years of global trade in CITES-listed wildlife. *Biological Conservation* 223, 47-57.
- Hari, R., Siew Woh, C., 2015. Complete mitogenome of *M. javanica* and *M. pentadactyla* obtained from next generation sequencing data. Unpublished.
- Harper, C., Ludwig, A., Clarke, A., Makgopela, K., Yurchenko, A., Guthrie, A., Dobrynin, P., Tamazian, G., Emslie, R., van Heerden, M., Hofmeyr, M., Potter, R., Roets, J., Beytell, P., Otiende, M., Kariuki, L., du Toit, R., Anderson, N., Okori, J., Antonik, A., Koepfli, K.-P., Thompson, P., O'Brien, S.J., 2018. Robust forensic matching of confiscated horns to individual poached African rhinoceros. *Current Biology* 28, R13-R14.
- Harper, C.K., Vermeulen, G.J., Clarke, A.B., de Wet, J.I., Guthrie, A.J., 2013. Extraction of nuclear DNA from rhinoceros horn and characterization of DNA profiling systems for white (*Ceratotherium simum*) and black (*Diceros bicornis*) rhinoceros. *Forensic Science International: Genetics* 7, 428-433.
- Harrington, L.A., D'Cruze, N., Macdonald, D., 2018. Rise to fame: events, media activity and public interest in pangolins and pangolin trade, 2005–2016. *Nature Conservation* 30, 107-133.
- Hassanin, A., Hugot, J.P., Van Vuuren, B.J., 2015. Comparison of mitochondrial genome sequences of pangolins (Mammalia, Pholidota). *Comptes Rendus - Biologies* 338, 260-265.
- Hassanin, A., Ropiquet, A., Gourmand, A.-L., Chardonnet, B., Rigoulet, J., 2007. Mitochondrial DNA variability in *Giraffa camelopardalis*: consequences for taxonomy, phylogeography and conservation of giraffes in West and central Africa. *Comptes Rendus Biologies* 330, 265-274.
- Heath, M.E., 1992. *Manis temminckii*. *Mammalian species*, 1-5.
- Heath, M.E., Coulson, I.M., 1997. Preliminary studies on relocation of Cape pangolins *Manis temminckii*. *African Journal of Wildlife Research* 27, 51-56.

- Heighton, S.P., 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, T., Ross, J., Shepherd, C., Challender, D., Cassey, P., 2017. The global trafficking of pangolins: A comprehensive summary of seizures and trafficking routes from 2010–2015, p. 50. TRAFFIC Southeast Asia, Selangor, Malaysia.
- Hervé, M., Hervé, M.M., 2020. Package ‘RVAideMemoire’, In R package version 0.9-74.
- Hewitt, G., 2000. The genetic legacy of the Quaternary ice ages. *Nature* 405, 907-913.
- Higashino, A., Sakate, R., Kameoka, Y., Takahashi, I., Hirata, M., Tanuma, R., Masui, T., Yasutomi, Y., Osada, N., 2012. Whole-genome sequencing and analysis of the Malaysian cynomolgus macaque (*Macaca fascicularis*) genome. *Genome biology* 13, R58.
- Hinten, G., Harriss, F., Rossetto, M., Braverstock, P.R., 2003. Genetic variation and island biogeography: Microsatellite and mitochondrial DNA variation in island populations of the Australian bush rat, *Rattus fuscipes greyii*. *Conservation Genetics* 4, 759-778.
- Hohenlohe, P.A., Funk, W.C., Rajora, O.P., 2021. Population genomics for wildlife conservation and management. *Molecular Ecology* 30, 62-82.
- Holt, C., Yandell, M., 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12, 491.
- Horner, F., Thorne, D., Shaver, A., 2020. Tipping the scales: Exposing the growing trade of african pangolins into China’s traditional medicine industry, p. 60. C4ADS.
- Hosoda, T., Sato, J.J., Lin, L.-K., Chen, Y.-J., Harada, M., Suzuki, H., 2011. Phylogenetic history of mustelid fauna in Taiwan inferred from mitochondrial genetic loci. *Canadian Journal of Zoology* 89, 559-569.
- Hothorn, T., Zeileis, A., Farebrother, R.W., Cummins, C., Millo, G., Mitchell, D., Zeileis, M.A., 2019. Package ‘lmtree’, In R package version 0.9-37.
- 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, L., 2020a. Genomic consequences of population decline in critically endangered pangolins and their demographic histories. *National Science Review* 7, 798-814.
- Hu, J., Roos, C., Lv, X., Kuang, W., Yu, L., 2020b. Molecular genetics supports a potential fifth Asian pangolin species (Mammalia, Pholidota, *Manis*). *Zoological Science* 37, 538-543, 536.
- Hua, L., Gong, S., Wang, F., Li, W., Ge, Y., Li, X., Hou, F., 2015. Captive breeding of pangolins: Current status, problems and future prospects. *ZooKeys* 507, 99-114.
- Huang, J., Zhao, Y., Shiraigol, W., Li, B., Bai, D., Ye, W., Daidiikhuu, D., Yang, L., Jin, B., Zhao, Q., Gao, Y., Wu, J., Bao, W., Li, A., Zhang, Y., Han, H., Bai, H., Bao, Y., Zhao, L., Zhai, Z., Zhao, W., Sun, Z., Zhang, Y., Meng, H., Dugarjaviin, M., 2014. Analysis of horse genomes provides insight into the diversification and adaptive evolution of karyotype. *Scientific Reports* 4, 4958.
- Huang, Y., 2003. Clinical studies of Yexiazhu Decoction in the treatment of hepatic fibrosis caused by hepatitis B. *Chinese Journal of Information on Traditional Chinese Medicine* 10, 55-56.
- Huffman, J.E., Wallace, J.R., 2012. *Wildlife forensics: methods and applications*. John Wiley & Sons, Sussex, UK.
- Hughes, A.C., 2021. Wildlife trade. *Current Biology* 31, R1218-R1224.

- Hunt, R.M., 2004. Global climate and the evolution of large mammalian carnivores during the later Cenozoic in North America. *Bulletin of the American Museum of Natural History* 2004, 139-156.
- Huson, D.H., Klöpper, T., Lockhart, P.J., Steel, M.A., 2005. Reconstruction of reticulate networks from gene trees, pp. 233-249. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Inger, R.F., Voris, H.K., 2001. The biogeographical relations of the frogs and snakes of Sundaland. *Journal of Biogeography* 28, 863-891.
- Ingram, D.J., Awol, P.P., Ding, K.A., Schenk, A., 2020. First records of pangolin trafficking in South Sudan. *African Journal of Ecology* 58, 133-137.
- Ingram, D.J., Coad, L., Abernethy, K.A., Maisels, F., Stokes, E.J., Bobo, K.S., Breuer, T., Gandiwa, E., Ghiurghi, A., Greengrass, E., 2018. Assessing Africa-wide pangolin exploitation by scaling local data. *Conservation Letters* 11, e12389.
- Ingram, D.J., Willcox, D., Challender, D.W.S., 2019. Evaluation of the application of methods used to detect and monitor selected mammalian taxa to pangolin monitoring. *Global Ecology and Conservation* 18, e00632.
- Irshad, N., Mahmood, T., Hussain, R., Nadeem, M.S., 2015. Distribution, abundance and diet of the Indian pangolin (*Manis crassicaudata*). *Animal Biology* 65, 57-71.
- IUCN, 2019. The IUCN Red List of Threatened Species.
- Jacaban, E., Rytwinski, T., Taylor, J.J., Young, N., Nguyen, V.M., Cooke, S.J., 2022. Do environmental systematic reviews impact policy and practice? Author perspectives on the application of their work. *Environmental Science & Policy* 129, 159-167.
- Jackson, N.D., Austin, C.C., 2010. The combined effects of rivers and refugia generate extreme cryptic fragmentation within the common ground skink (*Scincella lateralis*). *Evolution: International Journal of Organic Evolution* 64, 409-428.
- Jacobs, R.L., McClure, P.J., Baker, B.W., Espinoza, E.O., 2019. Myth debunked: Keratinous pangolin scales do not contain the analgesic tramadol. *Conservation Science and Practice* 1, e82.
- Jacquet, F., Nicolas, V., Colyn, M., Kadjo, B., Hutterer, R., Decher, J., Akpatou, B., Cruaud, C., Denys, C., 2014. Forest refugia and riverine barriers promote diversification in the West African pygmy shrew (*Crocidura obscurior* complex, Soricomorpha). *Zoologica Scripta* 43, 131-148.
- Jana, A., Karanth, K.P., 2022. Not all is black and white: phylogeography and population genetics of the endemic blackbuck (*Antelope cervicapra*). *Conservation Genetics*.
- Jentink, F., 1882. Revision of the Manidae in the Leyden Museum. *Notes from the Leyden Museum* 4, 193-209.
- Jiménez-Moreno, G., 2006. Progressive substitution of a subtropical forest for a temperate one during the middle Miocene climate cooling in Central Europe according to palynological data from cores Tengelic-2 and Hidas-53 (Pannonian Basin, Hungary). *Review of Palaeobotany and Palynology* 142, 1-14.
- Jin, X., Chua, H.Z., Wang, K., Li, N., Zheng, W., Pang, W., Yang, F., Pang, B., Zhang, M., Zhang, J., 2021. Evidence for the medicinal value of *Squama Manitis* (pangolin scale): A systematic review. *Integrative medicine research* 10, 100486-100486.
- Johnson, R.N., O'Meally, D., Chen, Z., Etherington, G.J., Ho, S.Y.W., Nash, W.J., Grueber, C.E., Cheng, Y., Whittington, C.M., Dennison, S., Peel, E., Haerty, W., O'Neill, R.J., Colgan, D., Russell, T.L., Alquezar-Planas, D.E., Attenbrow, V., Bragg, J.G., Brandies, P.A., Chong, A.Y.-Y., Deakin, J.E., Di Palma, F., Duda, Z., Eldridge, M.D.B., Ewart, K.M.,

- Hogg, C.J., Frankham, G.J., Georges, A., Gillett, A.K., Govendir, M., Greenwood, A.D., Hayakawa, T., Helgen, K.M., Hobbs, M., Holleley, C.E., Heider, T.N., Jones, E.A., King, A., Madden, D., Graves, J.A.M., Morris, K.M., Neaves, L.E., Patel, H.R., Polkinghorne, A., Renfree, M.B., Robin, C., Salinas, R., Tsangaras, K., Waters, P.D., Waters, S.A., Wright, B., Wilkins, M.R., Timms, P., Belov, K., 2018. Adaptation and conservation insights from the koala genome. *Nature Genetics* 50, 1102-1111.
- Johri, P., Riall, K., Becher, H., Excoffier, L., Charlesworth, B., Jensen, J.D., 2021. The impact of purifying and background selection on the inference of population history: Problems and prospects. *Molecular Biology and Evolution* 38, 2986-3003.
- Jombart, T., Ahmed, I., 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 27, 3070-3071.
- Jota Baptista, C., Seixas, F., Gonzalo-Orden, J.M., Oliveira, P.A., 2022. Wildlife Forensic Sciences: A Tool to Nature Conservation towards a One Health Approach. *Forensic Sciences* 2, 808-817.
- Junker, J., Petrovan, S.O., Arroyo-Rodríguez, V., Boonratana, R., Byler, D., Chapman, C.A., Chetry, D., Cheyne, S.M., Cornejo, F.M., Cortés-Ortiz, L., Cowlshaw, G., Christie, A.P., Crockford, C., Torre, Stella De L., De Melo, F.R., Fan, P., Grueter, C.C., Guzmán-Caro, D.C., Heymann, E.W., Herbinger, I., Hoang, M.D., Horwich, R.H., Humle, T., Ikemeh, R.A., Imong, I.S., Jerusalinsky, L., Johnson, S.E., Kappeler, P.M., Kierulff, M.C.M., KonÉ, I., Kormos, R., Le, K.Q., Li, B., Marshall, A.J., Meijaard, E., Mittermeier, R.A., Muroyama, Y., Neugebauer, E., Orth, L., Palacios, E., Papworth, S.K., Plumptre, A.J., Rawson, B.M., Refisch, J., Ratsimbazafy, J., Roos, C., Setchell, J.M., Smith, R.K., Sop, T., Schwitzer, C., Slater, K., Strum, S.C., Sutherland, W.J., Talebi, M., Wallis, J., Wich, S., Williamson, E.A., Wittig, R.M., KÜhl, H.S., 2020. A Severe Lack of Evidence Limits Effective Conservation of the World's Primates. *BioScience* 70, 794-803.
- Kainer, K.A., DiGiano, M.L., Duchelle, A.E., Wadt, L.H., Bruna, E., Dain, J.L., 2009. Partnering for greater success: local stakeholders and research in tropical biology and conservation. *Biotropica* 41, 555-562.
- Kämpf, M., Tessenow, E., Kenett, D.Y., Kantelhardt, J.W., 2015. The detection of emerging trends using Wikipedia traffic data and context networks. *PLoS ONE* 10, e0141892.
- Kamvar, Z.N., Tabima, J.F., Grünwald, N.J., 2014. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2, e281.
- Kapli, P., Yang, Z., Telford, M.J., 2020. Phylogenetic tree building in the genomic age. *Nature Reviews Genetics* 21, 428-444.
- Kardos, M., 2021. Conservation genetics. *Current Biology* 31, R1185-R1190.
- Karin, B.R., Gamble, T., Jackman, T.R., 2019. Optimizing Phylogenomics with Rapidly Evolving Long Exons: Comparison with Anchored Hybrid Enrichment and Ultraconserved Elements. *Molecular Biology and Evolution* 37, 904-922.
- Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* 30, 772-780.
- Kawamura, A., Chang, C.-H., Kawamura, Y., 2016. Middle Pleistocene to Holocene mammal faunas of the Ryukyu Islands and Taiwan: An updated review incorporating results of recent research. *Quaternary International* 397, 117-135.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012.

- Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647-1649.
- Kim, E.B., Fang, X., Fushan, A.A., Huang, Z., Lobanov, A.V., Han, L., Marino, S.M., Sun, X., Turanov, A.A., Yang, P., Yim, S.H., Zhao, X., Kasaikina, M.V., Stoletzki, N., Peng, C., Polak, P., Xiong, Z., Kiezun, A., Zhu, Y., Chen, Y., Kryukov, G.V., Zhang, Q., Peshkin, L., Yang, L., Bronson, R.T., Buffenstein, R., Wang, B., Han, C., Li, Q., Chen, L., Zhao, W., Sunyaev, S.R., Park, T.J., Zhang, G., Wang, J., Gladyshev, V.N., 2011. Genome sequencing reveals insights into physiology and longevity of the naked mole rat. *Nature* 479, 223-227.
- Kim, S., Cho, Y.S., Kim, H.-M., Chung, O., Kim, H., Jho, S., Seomun, H., Kim, J., Bang, W.Y., Kim, C., An, J., Bae, C.H., Bhak, Y., Jeon, S., Yoon, H., Kim, Y., Jun, J., Lee, H., Cho, S., Uphyrkina, O., Kostyria, A., Goodrich, J., Miquelle, D., Roelke, M., Lewis, J., Yurchenko, A., Bankevich, A., Cho, J., Lee, S., Edwards, J.S., Weber, J.A., Cook, J., Kim, S., Lee, H., Manica, A., Lee, I., O'Brien, S.J., Bhak, J., Yeo, J.-H., 2016. Comparison of carnivore, omnivore, and herbivore mammalian genomes with a new leopard assembly. *Genome biology* 17, 211.
- Knight, A.T., Cowling, R.M., Rouget, M., Balmford, A., Lombard, A.T., Campbell, B.M., 2008. Knowing but not doing: selecting priority conservation areas and the research–implementation gap. *Conservation Biology* 22, 610-617.
- Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A., Mayrose, I., 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular ecology resources* 15, 1179-1191.
- Korf, I., 2004. Gene finding in novel genomes. *BMC Bioinformatics* 5, 59.
- Korneliussen, T.S., Albrechtsen, A., Nielsen, R., 2014. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* 15, 356.
- Kotze, A., Ogden, R., Gaubert, P., Ahlers, N., Ades, G., Nash, H.C., Dalton, D.L., 2020. Understanding illegal trade in pangolins through forensics: applications in law enforcement, In *Pangolins: Science, Society and Conservation*. eds D. Challender, H. Nash, C. Waterman, pp. 321-332. Academic Press, London, UK.
- Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., Stamatakis, A., 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 35, 4453-4455.
- Kuismin, M., Saatoglu, D., Niskanen, A.K., Jensen, H., Sillanpää, M.J., 2020. Genetic assignment of individuals to source populations using network estimation tools. *Methods in Ecology and Evolution* 11, 333-344.
- Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution* 35, 1547-1549.
- Lachance, J., Tishkoff, S.A., 2013. SNP ascertainment bias in population genetic analyses: Why it is important, and how to correct it. *BioEssays* 35, 780-786.
- Lam, T.T.-Y., Shum, M.H.-H., Zhu, H.-C., Tong, Y.-G., Ni, X.-B., Liao, Y.-S., Wei, W., Cheung, W.Y.-M., Li, W.-J., Li, L.-F., Leung, G.M., Holmes, E.C., Hu, Y.-L., Guan, Y., 2020. Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. *Nature* 538, 282-285.
- Lande, R., 1988. Genetics and demography in biological conservation. *Science* 241, 1455-1460.

- Lane-deGraaf, K.E., Fuentes, A., Hollocher, H., 2014. Landscape genetics reveal fine-scale boundaries in island populations of Indonesian long-tailed macaques. *Landscape Ecology* 29, 1505-1519.
- Lang, N., Wolff, E.W., 2011. Interglacial and glacial variability from the last 800 ka in marine, ice and terrestrial archives. *Clim. Past* 7, 361-380.
- Lawson, D.J., van Dorp, L., Falush, D., 2018. A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communications* 9, 3258.
- Lee, J., Hughes, T., Lee, M.-H., Field, H., Rovie-Ryan, J.J., Sitam, F.T., Sipangkui, S., Nathan, S.K.S.S., Ramirez, D., Kumar, S.V., Lasimbang, H., Epstein, J.H., Daszak, P., 2020. No evidence of Coronaviruses or other potentially zoonotic viruses in Sunda pangolins (*Manis javanica*) entering the wildlife trade via Malaysia. *EcoHealth* 17, 406-418.
- Leffler, E.M., Bullaughey, K., Matute, D.R., Meyer, W.K., Ségurel, L., Venkat, A., Andolfatto, P., Przeworski, M., 2012. Revisiting an old riddle: What determines genetic diversity levels within species? *PLoS biology* 10, e1001388.
- Leinonen, R., Sugawara, H., Shumway, M., 2010. The Sequence Read Archive. *Nucleic Acids Research* 39, D19-D21.
- Leonard, J.A., den Tex, R.-J., Hawkins, M.T.R., Muñoz-Fuentes, V., Thorington, R., Maldonado, J.E., 2015. Phylogeography of vertebrates on the Sunda Shelf: a multi-species comparison. *Journal of Biogeography* 42, 871-879.
- Lewin, H.A., Robinson, G.E., Kress, W.J., Baker, W.J., Coddington, J., Crandall, K.A., Durbin, R., Edwards, S.V., Forest, F., Gilbert, M.T.P., Goldstein, M.M., Grigoriev, I.V., Hackett, K.J., Haussler, D., Jarvis, E.D., Johnson, W.E., Patrinos, A., Richards, S., Castilla-Rubio, J.C., van Sluys, M.-A., Soltis, P.S., Xu, X., Yang, H., Zhang, G., 2018. Earth BioGenome Project: Sequencing life for the future of life. *Proceedings of the National Academy of Sciences* 115, 4325-4333.
- Lewis, N.E., Liu, X., Li, Y., Nagarajan, H., Yerganian, G., O'Brien, E., Bordbar, A., Roth, A.M., Rosenbloom, J., Bian, C., Xie, M., Chen, W., Li, N., Baycin-Hizal, D., Latif, H., Forster, J., Betenbaugh, M.J., Famili, I., Xu, X., Wang, J., Palsson, B.O., 2013. Genomic landscapes of Chinese hamster ovary cell lines as revealed by the *Cricetulus griseus* draft genome. *Nature biotechnology* 31, 759-765.
- Li, C., Hofreiter, M., Straube, N., Corrigan, S., Naylor, G.J., 2013a. Capturing protein-coding genes across highly divergent species. *Biotechniques* 54, 321-326.
- Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754-1760.
- Li, H., Durbin, R., 2011. Inference of human population history from individual whole-genome sequences. *Nature* 475, 493-496.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-2079.
- Li, M., Tian, S., Jin, L., Zhou, G., Li, Y., Zhang, Y., Wang, T., Yeung, C.K.L., Chen, L., Ma, J., Zhang, J., Jiang, A., Li, J., Zhou, C., Zhang, J., Liu, Y., Sun, X., Zhao, H., Niu, Z., Lou, P., Xian, L., Shen, X., Liu, S., Zhang, S., Zhang, M., Zhu, L., Shuai, S., Bai, L., Tang, G., Liu, H., Jiang, Y., Mai, M., Xiao, J., Wang, X., Zhou, Q., Wang, Z., Stothard, P., Xue, M., Gao, X., Luo, Z., Gu, Y., Zhu, H., Hu, X., Zhao, Y., Plastow, G.S., Wang, J., Jiang, Z., Li, K., Li, N., Li, X., Li, R., 2013b. Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nature Genetics* 45, 1431-1438.

- Li, R., Fan, W., Tian, G., Zhu, H., He, L., Cai, J., Huang, Q., Cai, Q., Li, B., Bai, Y., Zhang, Z., Zhang, Y., Wang, W., Li, J., Wei, F., Li, H., Jian, M., Li, J., Zhang, Z., Nielsen, R., Li, D., Gu, W., Yang, Z., Xuan, Z., Ryder, O.A., Leung, F.C.-C., Zhou, Y., Cao, J., Sun, X., Fu, Y., Fang, X., Guo, X., Wang, B., Hou, R., Shen, F., Mu, B., Ni, P., Lin, R., Qian, W., Wang, G., Yu, C., Nie, W., Wang, J., Wu, Z., Liang, H., Min, J., Wu, Q., Cheng, S., Ruan, J., Wang, M., Shi, Z., Wen, M., Liu, B., Ren, X., Zheng, H., Dong, D., Cook, K., Shan, G., Zhang, H., Kosiol, C., Xie, X., Lu, Z., Zheng, H., Li, Y., Steiner, C.C., Lam, T.T.-Y., Lin, S., Zhang, Q., Li, G., Tian, J., Gong, T., Liu, H., Zhang, D., Fang, L., Ye, C., Zhang, J., Hu, W., Xu, A., Ren, Y., Zhang, G., Bruford, M.W., Li, Q., Ma, L., Guo, Y., An, N., Hu, Y., Zheng, Y., Shi, Y., Li, Z., Liu, Q., Chen, Y., Zhao, J., Qu, N., Zhao, S., Tian, F., Wang, X., Wang, H., Xu, L., Liu, X., Vinar, T., Wang, Y., Lam, T.-W., Yiu, S.-M., Liu, S., Zhang, H., Li, D., Huang, Y., Wang, X., Yang, G., Jiang, Z., Wang, J., Qin, N., Li, L., Li, J., Bolund, L., Kristiansen, K., Wong, G.K.-S., Olson, M., Zhang, X., Li, S., Yang, H., Wang, J., Wang, J., 2010. The sequence and de novo assembly of the giant panda genome. *Nature* 463, 311-317.
- Li, W., Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658-1659.
- Linacre, A., Gusmão, L., Hecht, W., Hellmann, A.P., Mayr, W.R., Parson, W., Prinz, M., Schneider, P.M., Morling, N., 2011. ISFG: Recommendations regarding the use of non-human (animal) DNA in forensic genetic investigations. *Forensic Science International: Genetics* 5, 501-505.
- Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., Karlsson, E.K., Jaffe, D.B., Kamal, M., Clamp, M., Chang, J.L., Kulbokas, E.J., Zody, M.C., 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438, 803-819.
- Lindsey, P., Allan, J., Brehony, P., Dickman, A., Robson, A., Begg, C., Bhammar, H., Blanken, L., Breuer, T., Fitzgerald, K., Flyman, M., Gandiwa, P., Giva, N., Kaelo, D., Nampindo, S., Nyambe, N., Steiner, K., Parker, A., Roe, D., Thomson, P., Trimble, M., Caron, A., Tyrrell, P., 2020. Conserving Africa's wildlife and wildlands through the COVID-19 crisis and beyond. *Nature Ecology & Evolution* 4, 1300-1310.
- Linklater, W.L., 2003. Science and management in a conservation crisis: a case study with rhinoceros. *Conservation Biology* 17, 968-975.
- Liu, L., Zhang, J., Rheindt, F.E., Lei, F., Qu, Y., Wang, Y., Zhang, Y., Sullivan, C., Nie, W., Wang, J., Yang, F., Chen, J., Edwards, S.V., Meng, J., Wu, S., 2017. Genomic evidence reveals a radiation of placental mammals uninterrupted by the KPg boundary. *Proceedings of the National Academy of Sciences of the United States of America* 114, E7282-E7290.
- Liu, P., Chen, W., Chen, J.P., 2019. Viral metagenomics revealed Sendai virus and coronavirus infection of Malayan pangolins (*Manis javanica*). *Viruses* 11.
- Liu, S., Hansen, M.M., 2017. PSMC (pairwise sequentially Markovian coalescent) analysis of RAD (restriction site associated DNA) sequencing data. *Molecular ecology resources* 17, 631-641.
- Liu, S., Lorenzen, Eline D., Fumagalli, M., Li, B., Harris, K., Xiong, Z., Zhou, L., Korneliussen, Thorfinn S., Somel, M., Babbitt, C., Wray, G., Li, J., He, W., Wang, Z., Fu, W., Xiang, X., Morgan, Claire C., Doherty, A., O'Connell, Mary J., McInerney, James O., Born, Erik W., Dalén, L., Dietz, R., Orlando, L., Sonne, C., Zhang, G., Nielsen, R., Willerslev, E., Wang, J., 2014. Population Genomics Reveal Recent Speciation and Rapid Evolutionary Adaptation in Polar Bears. *Cell* 157, 785-794.

- Liu, Y., 2020. 穿山甲已从《中国药典》“除名”，专家：其药用功效可被替代, In Xinhua Newspaper. Jiangsu Xinhua Newspaper Media Group, Yiwu Shi, China. Accessed at http://www.xhby.net/index/202006/t20200609_6681027.shtml.
- Locke, D.P., Hillier, L.W., Warren, W.C., Worley, K.C., Nazareth, L.V., Muzny, D.M., Yang, S.-P., Wang, Z., Chinwalla, A.T., Minx, P., Mitreva, M., Cook, L., Delehaunty, K.D., Fronick, C., Schmidt, H., Fulton, L.A., Fulton, R.S., Nelson, J.O., Magrini, V., Pohl, C., Graves, T.A., Markovic, C., Cree, A., Dinh, H.H., Hume, J., Kovar, C.L., Fowler, G.R., Lunter, G., Meader, S., Heger, A., Ponting, C.P., Marques-Bonet, T., Alkan, C., Chen, L., Cheng, Z., Kidd, J.M., Eichler, E.E., White, S., Searle, S., Vilella, A.J., Chen, Y., Flicek, P., Ma, J., Raney, B., Suh, B., Burhans, R., Herrero, J., Haussler, D., Faria, R., Fernando, O., Darré, F., Farré, D., Gazave, E., Oliva, M., Navarro, A., Roberto, R., Capozzi, O., Archidiacono, N., Valle, G.D., Purgato, S., Rocchi, M., Konkel, M.K., Walker, J.A., Ullmer, B., Batzer, M.A., Smit, A.F.A., Hubley, R., Casola, C., Schrider, D.R., Hahn, M.W., Quesada, V., Puente, X.S., Ordoñez, G.R., López-Otín, C., Vinar, T., Brejova, B., Ratan, A., Harris, R.S., Miller, W., Kosiol, C., Lawson, H.A., Taliwal, V., Martins, A.L., Siepel, A., RoyChoudhury, A., Ma, X., Degenhardt, J., Bustamante, C.D., Gutenkunst, R.N., Mailund, T., Dutheil, J.Y., Hobolth, A., Schierup, M.H., Ryder, O.A., Yoshinaga, Y., de Jong, P.J., Weinstock, G.M., Rogers, J., Mardis, E.R., Gibbs, R.A., Wilson, R.K., 2011. Comparative and demographic analysis of orang-utan genomes. *Nature* 469, 529-533.
- López-Baucells, A., Rocha, R., Fernández-Llamazares, Á., 2018. When bats go viral: negative framings in virological research imperil bat conservation. *Mammal Review* 48, 62-66.
- Lorenzen, E.D., Heller, R., Siegmund, H.R., 2012. Comparative phylogeography of African savannah ungulates. *Molecular Ecology* 21, 3656-3670.
- Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D.W., Yiu, S.-M., Peng, S., Xiaoqian, Z., Liu, G., Liao, X., Li, Y., Yang, H., Wang, J., Lam, T.-W., Wang, J., 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* 1.
- Luo, S.-J., Zhang, Y., Johnson, W.E., Miao, L., Martelli, P., Antunes, A., Smith, J.L.D., O'Brien, S.J., 2014. Sympatric Asian felid phylogeography reveals a major Indochinese–Sundaic divergence. *Molecular Ecology* 23, 2072-2092.
- Luo, S.J., Cai, Q.X., David, V.A., Zhang, L., Martelli, P., Lim, N.T.L., Ferrand, N., Chin, S.C., Gaubert, P., Ramos, M.J., O'Brien, S.J., Antunes, A., Johnson, W.E., 2007. Isolation and characterization of microsatellite markers in pangolins (Mammalia, Pholidota, Manis spp.): Primer note. *Molecular Ecology Notes* 7, 269-272.
- Maddox, T., 2007. The conservation of tigers and other wildlife in oil palm plantations: Jambi province, Sumatra, Indonesia (October 2007), In ZSL Conservation Report No. 7. Zoological Society of London, London, UK.
- Maechler, M., Rousseeuw, P., Croux, C., Todorov, V., Ruckstuhl, A., Salibian-Barrera, M., Verbeke, T., Koller, M., Conceicao, E.L., di Palma, M.A., 2021. Package ‘robustbase’, <https://robustbase.r-forge.r-project.org/>.
- Mahmood, T., Challender, D., Khatiwada, A., Andleeb, S., Perera, P., Trageser, S., Ghose, A., Mohapatra, R., 2019. Manis crassicaudata. The IUCN Red List of Threatened Species 2019, e.T12761A123583998.
- Mahmood, T., Mohapatra, R.K., Perera, P., Irshad, N., Akrim, F., Andleeb, S., Waseem, M., Sharma, S., Panda, S., 2020. Indian pangolin Manis crassicaudata (Geoffroy, 1803), In

- Pangolins: Science, Society and Conservation. eds D.W.S. Challender, H.C. Nash, C. Waterman, pp. 71-88. Academic Press.
- Mambeya, M.M., Baker, F., Momboua, B.R., Koumba Pambo, A.F., Hega, M., Okouyi Okouyi, V.J., Onanga, M., Challender, D.W.S., Ingram, D.J., Wang, H., Abernethy, K., 2018. The emergence of a commercial trade in pangolins from Gabon. *African Journal of Ecology* 56, 601-609.
- Mandimbihasina, A.R., Woolaver, L.G., Concannon, L.E., Milner-Gulland, E., Lewis, R.E., Terry, A.M., Filazaha, N., Rabetafika, L.L., Young, R.P., 2020. The illegal pet trade is driving Madagascar's ploughshare tortoise to extinction. *ORYX* 54, 188-196.
- Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., Chen, W.-M., 2010. Robust relationship inference in genome-wide association studies. *Bioinformatics* 26, 2867-2873.
- Martinsohn, J.T., Ogden, R., 2009. FishPopTrace—Developing SNP-based population genetic assignment methods to investigate illegal fishing. *Forensic Science International: Genetics Supplement Series* 2, 294-296.
- Masters, S., van Andel, T., de Boer, H.J., Heijungs, R., Gravendeel, B., 2020. Patent analysis as a novel method for exploring commercial interest in wild harvested species. *Biological Conservation* 243, 108454.
- Mays, H.L., Hung, C.-M., Shaner, P.-J., Denvir, J., Justice, M., Yang, S.-F., Roth, T.L., Oehler, D.A., Fan, J., Rekulapally, S., Primerano, D.A., 2018. Genomic Analysis of Demographic History and Ecological Niche Modeling in the Endangered Sumatran Rhinoceros *Dicerorhinus sumatrensis*. *Current Biology* 28, 70-76.e74.
- McDonough, M.M., Šumbera, R., Mazoch, V., Ferguson, A.W., Phillips, C.D., Bryja, J., 2015. Multilocus phylogeography of a widespread savanna–woodland-adapted rodent reveals the influence of Pleistocene geomorphology and climate change in Africa's Zambezi region. *Molecular Ecology* 24, 5248-5266.
- McNamara, J., Kusimi, J.M., Rowcliffe, J.M., Cowlishaw, G., Brenyah, A., Milner-Gulland, E.J., 2015. Long-term spatio-temporal changes in a West African bushmeat trade system. *Conservation Biology* 29, 1446-1457.
- McNeill, D., Lichtenstein, G., 2003. Local Conflicts and International Compromises: The Sustainable Use of Vicuña in Argentina. *Journal of International Wildlife Law & Policy* 6, 233-253.
- Meijaard, E., Cardillo, M., Meijaard, E.M., Possingham, H.P., 2015. Geographic bias in citation rates of conservation research. *Conservation Biology* 29, 920-925.
- Meijaard, E., Groves, C.P., 2006. The Geography of Mammals and Rivers in Mainland Southeast Asia, In *Primate Biogeography: Progress and Prospects*. pp. 305-329. Springer US, Boston, MA.
- Meyer, M., Kircher, M., Gansauge, M.-T., Li, H., Racimo, F., Mallick, S., Schraiber, J.G., Jay, F., Prüfer, K., de Filippo, C., Sudmant, P.H., Alkan, C., Fu, Q., Do, R., Rohland, N., Tandon, A., Siebauer, M., Green, R.E., Bryc, K., Briggs, A.W., Stenzel, U., Dabney, J., Shendure, J., Kitzman, J., Hammer, M.F., Shunkov, M.V., Derevianko, A.P., Patterson, N., Andrés, A.M., Eichler, E.E., Slatkin, M., Reich, D., Kelso, J., Pääbo, S., 2012. A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338, 222-226.
- Mikkelsen, T., Hillier, L., Eichler, E., Zody, M., Jaffe, D., Yang, S.-P., Enard, W., Hellmann, I., Lindblad-Toh, K., Altheide, T., 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437, 69-87.

- Mikkelsen, T.S., Wakefield, M.J., Aken, B., Amemiya, C.T., Chang, J.L., Duke, S., Garber, M., Gentles, A.J., Goodstadt, L., Heger, A., Jurka, J., Kamal, M., Mauceli, E., Searle, S.M.J., Sharpe, T., Baker, M.L., Batzer, M.A., Benos, P.V., Belov, K., Clamp, M., Cook, A., Cuff, J., Das, R., Davidow, L., Deakin, J.E., Fazzari, M.J., Glass, J.L., Grabherr, M., Grealley, J.M., Gu, W., Hore, T.A., Huttley, G.A., Kleber, M., Jirtle, R.L., Koina, E., Lee, J.T., Mahony, S., Marra, M.A., Miller, R.D., Nicholls, R.D., Oda, M., Papenfuss, A.T., Parra, Z.E., Pollock, D.D., Ray, D.A., Schein, J.E., Speed, T.P., Thompson, K., VandeBerg, J.L., Wade, C.M., Walker, J.A., Waters, P.D., Webber, C., Weidman, J.R., Xie, X., Zody, M.C., Baldwin, J., Abdouelleil, A., Abdulkadir, J., Abebe, A., Abera, B., Abreu, J., Acer, S.C., Aftuck, L., Alexander, A., An, P., Anderson, E., Anderson, S., Arachi, H., Azer, M., Bachantsang, P., Barry, A., Bayul, T., Berlin, A., Bessette, D., Bloom, T., Blye, J., Boguslavskiy, L., Bonnet, C., Boukhgalter, B., Bourzgui, I., Brown, A., Cahill, P., Channer, S., Cheshatsang, Y., Chuda, L., Citroen, M., Collymore, A., Cooke, P., Costello, M., D'Aco, K., Daza, R., De Haan, G., DeGray, S., DeMaso, C., Dhargay, N., Dooley, K., Dooley, E., Doricent, M., Dorje, P., Dorjee, K., Dupes, A., Elong, R., Falk, J., Farina, A., Faro, S., Ferguson, D., Fisher, S., Foley, C.D., Franke, A., Friedrich, D., Gadbois, L., Gearin, G., Gearin, C.R., Giannoukos, G., Goode, T., Graham, J., Grandbois, E., Grewal, S., Gyaltsen, K., Hafez, N., Hagos, B., Hall, J., Henson, C., Hollinger, A., Honan, T., Huard, M.D., Hughes, L., Hurhula, B., Husby, M.E., Kamat, A., Kanga, B., Kashin, S., Khazanovich, D., Kisner, P., Lance, K., Lara, M., Lee, W., Lennon, N., Letendre, F., LeVine, R., Lipovsky, A., Liu, X., Liu, J., Liu, S., Lokyitsang, T., Lokyitsang, Y., Lubonja, R., Lui, A., MacDonald, P., Magnisalis, V., Maru, K., Matthews, C., McCusker, W., McDonough, S., Mehta, T., Meldrim, J., Meneus, L., Mihai, O., Mihalev, A., Mihova, T., Mittelman, R., Mlenga, V., Montmayeur, A., Mulrain, L., Navidi, A., Naylor, J., Negash, T., Nguyen, T., Nguyen, N., Nicol, R., Norbu, C., Norbu, N., Novod, N., O'Neill, B., Osman, S., Markiewicz, E., Oyono, O.L., Patti, C., Phunkhang, P., Pierre, F., Priest, M., Raghuraman, S., Rege, F., Reyes, R., Rise, C., Rogov, P., Ross, K., Ryan, E., Settipalli, S., Shea, T., Sherpa, N., Shi, L., Shih, D., Sparrow, T., Spaulding, J., Stalker, J., Stange-Thomann, N., Stavropoulos, S., Stone, C., Strader, C., Tesfaye, S., Thomson, T., Thoulutsang, Y., Thoulutsang, D., Topham, K., Topping, I., Tsamla, T., Vassiliev, H., Vo, A., Wangchuk, T., Wangdi, T., Weiland, M., Wilkinson, J., Wilson, A., Yadav, S., Young, G., Yu, Q., Zembek, L., Zhong, D., Zimmer, A., Zwirko, Z., Jaffe, D.B., Alvarez, P., Brockman, W., Butler, J., Chin, C., Gnerre, S., MacCallum, I., Graves, J.A.M., Ponting, C.P., Breen, M., Samollow, P.B., Lander, E.S., Lindblad-Toh, K., Broad Institute Genome Sequencing, P., Broad Institute Whole Genome Assembly, T., 2007. Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* 447, 167-177.
- Milliken, T., Burn, R., Sangalaku, L., 2016. The Elephant Trade Information System (ETIS) and the Illicit Trade in Ivory: A report to the 17th meeting of the Conference of the Parties to CITES. CoP17 Doc 56.
- Minh, B.Q., Hahn, M.W., Lanfear, R., 2020a. New methods to calculate concordance factors for phylogenomic datasets. *Molecular Biology and Evolution* 37, 2727-2733.
- Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R., 2020b. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 37, 1530-1534.

- Miyamoto, M.M., Goodman, M., 1986. Biomolecular systematics of eutherian mammals: phylogenetic patterns and classification. *Systematic Biology* 35, 230-240.
- Mohapatra, R., Panda, S., Sahu, S., 2018. On the gestation period of Indian pangolins (*Manis crassicaudata*) in captivity. *Biodiversity Int J* 2, 559-560.
- Mondol, S., Bruford, M.W., Ramakrishnan, U., 2013. Demographic loss, genetic structure and the conservation implications for Indian tigers. *Proceedings of the Royal Society B: Biological Sciences* 280, 20130496.
- Moritz, C., 1994. Defining 'evolutionarily significant units' for conservation. *Trends in ecology & evolution* 9, 373-375.
- Mukherjee, S., Krishnan, A., Tamma, K., Home, C., R, N., Joseph, S., Das, A., Ramakrishnan, U., 2010. Ecology Driving Genetic Variation: A Comparative Phylogeography of Jungle Cat (*Felis chaus*) and Leopard Cat (*Prionailurus bengalensis*) in India. *PLoS ONE* 5, e13724.
- Murphy, W.J., Eizirik, E., Johnson, W.E., Zhang, Y.P., Ryder, O.A., O'Brien, S.J., 2001. Molecular phylogenetics and the origins of placental mammals. *Nature* 409, 614-618.
- Murthy, K.L.N., Mishra, S., 2010. A note on road killing of Indian pangolin *Manis crassicaudata* Gray at Kambalakonda Wildlife Sanctuary of Eastern Ghat Ranges. *Small Mammal Mail Bi-Annual Newsletter for CCINSA & RISCINSA* 2, 8-10.
- Musmann, S.M., Douglas, M.R., Anthonysamy, W.J., Davis, M.A., Simpson, S.A., Louis, W., Douglas, M.E., 2017. Genetic rescue, the greater prairie chicken and the problem of conservation reliance in the Anthropocene. *R Soc Open Sci* 4, 160736.
- Nadachowska-Brzyska, K., Burri, R., Smeds, L., Ellegren, H., 2016. PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white *Ficedula* flycatchers. *Molecular Ecology* 25, 1058-1072.
- Nagasawa, H., Sakamoto, S., Mori, T., Wu, G., Piuo, H., Li, Y., Bai, S., Jin, C., Yamamuro, Y., 1992. Effects of some natural products as components of Chinese herbal medicines on mammary gland growth and function in mice. *In vivo (Athens, Greece)* 6, 135-140.
- Nakahama, N., 2021. Museum specimens: An overlooked and valuable material for conservation genetics. *Ecological Research* 36, 13-23.
- Nash, H.C., Wirdateti, Low, G.W., Choo, S.W., Chong, J.L., Semiadi, G., Hari, R., Sulaiman, M.H., Turvey, S.T., Evans, T.A., Rheindt, F.E., 2018. Conservation genomics reveals possible illegal trade routes and admixture across pangolin lineages in Southeast Asia. *Conservation Genetics* 19, 1083-1095.
- Nash, H.C., Wong, M.H., Turvey, S.T., 2016. Using local ecological knowledge to determine status and threats of the Critically Endangered Chinese pangolin (*Manis pentadactyla*) in Hainan, China. *Biological Conservation* 196, 189-195.
- Natesh, M., Taylor, R.W., Truelove, N.K., Hadly, E.A., Palumbi, S.R., Petrov, D.A., Ramakrishnan, U., 2019. Empowering conservation practice with efficient and economical genotyping from poor quality samples. *Methods in Ecology and Evolution* 10, 853-859.
- Nellemann, C., Henriksen, R., Kreilhuber, A., Stewart, D., Kotsovou, M., Raxter, P., Mrema, E., Barrat, S., 2016. The rise of environmental crime: a growing threat to natural resources, peace, development and security. *United Nations Environment Programme (UNEP) and RHIPTO Rapid Response – Norwegian Center for Global Analyses*.
- Neukamm, J., Peltzer, A., Nieselt, K., 2021. DamageProfiler: fast damage pattern calculation for ancient DNA. *Bioinformatics* 37, 3652-3653.
- Neuwirth, E., 2014. RColorBrewer: ColorBrewer palettes, In R package version 1.1-2.

- Ng, C.H., Ng, K.K.S., Lee, S.L., Zakaria, N.-F., Lee, C.T., Tnah, L.H., 2022. DNA databases of an important tropical timber tree species *Shorea leprosula* (Dipterocarpaceae) for forensic timber identification. *Scientific Reports* 12, 9546.
- Nicolas, V., Missouf, A.D., Denys, C., Kerbis Peterhans, J., Katuala, P., Couloux, A., Colyn, M., 2011. The roles of rivers and Pleistocene refugia in shaping genetic diversity in *Praomys misonnei* in tropical Africa. *Journal of Biogeography* 38, 191-207.
- Nijman, V., Spaan, D., Nekaris, K.A.-I., 2016a. Large-Scale Trade in Legally Protected Marine Mollusc Shells from Java and Bali, Indonesia. *PLoS ONE* 10, e0140593.
- Nijman, V., Zhang, M.X., Shepherd, C.R., 2016b. Pangolin trade in the Mong La wildlife market and the role of Myanmar in the smuggling of pangolins into China. *Global Ecology and Conservation* 5, 118-126.
- Nixon, S., Pietersen, D., Challender, D., Hoffmann, M., Godwill Ichu, I., Bruce, T., Ingram, D.J., Matthews, N., Shirley, M.H., 2019. *Smutsia gigantea*. The IUCN Red List of Threatened Species 2019, e.T12762A123584478.
- Nørgaard, L.S., Mikkelsen, D.M.G., Elmeros, M., Chriél, M., Madsen, A.B., Nielsen, J.L., Pertoldi, C., Randi, E., Fickel, J., Brygida, S., Ruiz-González, A., 2017. Population genomics of the raccoon dog (*Nyctereutes procyonoides*) in Denmark: insights into invasion history and population development. *Biological Invasions* 19, 1637-1652.
- Notredame, C., Higgins, D.G., Heringa, J., 2000. T-coffee: a novel method for fast and accurate multiple sequence alignment. Edited by J. Thornton. *Journal of Molecular Biology* 302, 205-217.
- O'Hanlon, S.J., Rieux, A., Farrer, R.A., Rosa, G.M., Waldman, B., Bataille, A., Kosch, T.A., Murray, K.A., Brankovics, B., Fumagalli, M., Martin, M.D., Wales, N., Alvarado-Rybak, M., Bates, K.A., Berger, L., Böll, S., Brookes, L., Clare, F., Courtois, E.A., Cunningham, A.A., Doherty-Bone, T.M., Ghosh, P., Gower, D.J., Hintz, W.E., Höglund, J., Jenkinson, T.S., Lin, C.-F., Laurila, A., Loyau, A., Martel, A., Meurling, S., Miaud, C., Minting, P., Pasmans, F., Schmeller, D.S., Schmidt, B.R., Shelton, J.M.G., Skerratt, L.F., Smith, F., Soto-Azat, C., Spagnoletti, M., Tessa, G., Toledo, L.F., Valenzuela-Sánchez, A., Verster, R., Vörös, J., Webb, R.J., Wierzbicki, C., Wombwell, E., Zamudio, K.R., Aanensen, D.M., James, T.Y., Gilbert, M.T.P., Weldon, C., Bosch, J., Balloux, F., Garner, T.W.J., Fisher, M.C., 2018. Recent Asian origin of chytrid fungi causing global amphibian declines. *Science* 360, 621-627.
- Ogden, R., 2008. Fisheries forensics: the use of DNA tools for improving compliance, traceability and enforcement in the fishing industry. *Fish and Fisheries* 9, 462-472.
- Ogden, R., Linacre, A., 2015. Wildlife forensic science: A review of genetic geographic origin assignment. *Forensic Science International: Genetics* 18, 152-159.
- Okonechnikov, K., Conesa, A., García-Alcalde, F., 2015. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* 32, 292-294.
- Oldfield, S., 2003. *The trade in wildlife: regulation for conservation*. Routledge, New York.
- Olmedo, A., Sharif, V., Milner-Gulland, E.J., 2018. Evaluating the design of behavior change interventions: a case study of rhino horn in Vietnam. *Conservation Letters* 11, e12365.
- Ortiz, E., 2019. vcf2phylip v2. 0: convert a VCF matrix into several matrix formats for phylogenetic analysis. URL <https://doi.org/105281/zenodo.2540861>.
- Pagès, E., 1970. Sur l'écologie et les adaptations de l'oryctérope et des pangolins sympatriques d'Afrique. *Biologia Gabonica* 6, 27-92.

- Pagès, E., 1972a. Comportement agressif et sexuel chez les pangolins arboricoles (*Manis tricuspis* et *M. longicauda*). *Biol. Gabonica* 1, 3-62.
- Pagès, E., 1972b. Comportement maternel et développement du jeune chez un pangolin arboricole (*M. tricuspis*). *Biol. Gabon* 8, 63-120.
- Paijmans, J.L., Fickel, J., Courtiol, A., Hofreiter, M., Förster, D.W., 2016. Impact of enrichment conditions on cross-species capture of fresh and degraded DNA. *Molecular ecology resources* 16, 42-55.
- Palkopoulou, E., Mallick, S., Skoglund, P., Enk, J., Rohland, N., Li, H., Omrak, A., Vartanyan, S., Poinar, H., Götherström, A., Reich, D., Dalén, L., 2015. Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. *Current Biology* 25, 1395-1400.
- Pantel, S., Chin, S.Y., 2009. Proceedings of the workshop on trade and conservation of pangolins native to south and southeast Asia: 30 June-2 July 2008, Singapore Zoo, Singapore. TRAFFIC Southeast Asia, Selangor, Malaysia.
- Paradis, E., 2010. pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* 26, 419-420.
- Patel, R.P., Wutke, S., Lenz, D., Mukherjee, S., Ramakrishnan, U., Veron, G., Fickel, J., Wilting, A., Förster, D.W., 2017. Genetic Structure and Phylogeography of the Leopard Cat (*Prionailurus bengalensis*) Inferred from Mitochondrial Genomes. *Journal of Heredity* 108, 349-360.
- Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., Nieselt, K., 2016. EAGER: efficient ancient genome reconstruction. *Genome biology* 17, 60.
- Pembleton, L.W., Cogan, N.O.I., Forster, J.W., 2013. StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular ecology resources* 13, 946-952.
- Perez, M.F., Franco, F.F., Bombonato, J.R., Bonatelli, I.A.S., Khan, G., Romeiro-Brito, M., Fegies, A.C., Ribeiro, P.M., Silva, G.A.R., Moraes, E.M., 2018. Assessing population structure in the face of isolation by distance: Are we neglecting the problem? *Diversity and Distributions* 24, 1883-1889.
- Perry, G.H., Reeves, D., Melsted, P., Ratan, A., Miller, W., Michelini, K., Louis, E.E., Jr, Pritchard, J.K., Mason, C.E., Gilad, Y., 2011. A genome sequence resource for the Aye-Aye (*Daubentonia madagascariensis*), a nocturnal lemur from Madagascar. *Genome biology and evolution* 4, 126-135.
- Phelps, J., Webb, E.L., 2015. "Invisible" wildlife trades: Southeast Asia's undocumented illegal trade in wild ornamental plants. *Biological Conservation* 186, 296-305.
- Phillis, C.C., O'Regan, S.M., Green, S.J., Bruce, J.E., Anderson, S.C., Linton, J.N., Derby, E.O.R., Favaro, B., 2013. Multiple pathways to conservation success. *Conservation Letters* 6, 98-106.
- Pickrell, J.K., Pritchard, J.K., 2012. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genetics* 8, e1002967.
- Pietersen, D., Jansen, R., Connelly, E., 2019. *Smutsia temminckii*. The IUCN Red List of Threatened Species 2019, e.T12765A123585768.
- Pietersen, D.W., Challender, D.W., 2020. Research needs for pangolins, In *Pangolins: Science, Society and Conservation*. eds D. Challender, H. Nash, C. Waterman, pp. 537-543. Academic Press, London, UK.

- Pietersen, D.W., Jansen, R., Swart, J., Panaino, W., Kotze, A., Rankin, P., Nebe, B., 2020. Temminck's pangolin *Smutsia temminckii* (Smuts, 1832), In Pangolins: Science, Society and Conservation. eds D.W.S. Challender, H.C. Nash, C. Waterman, pp. 175-193. Academic Press.
- Pietersen, D.W., McKechnie, A.E., Jansen, R., 2014. A review of the anthropogenic threats faced by Temminck's ground pangolin, *Smutsia temminckii*, in southern Africa. *African Journal of Wildlife Research* 44, 167-178.
- Pimentel, D., Zuniga, R., Morrison, D., 2005. Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics* 52, 273-288.
- Pimm, S.L., Russell, G.J., Gittleman, J.L., Brooks, T.M., 1995. The Future of Biodiversity. *Science* 269, 347-350.
- Pinter, G.G., Pinter, V., 1993. Is physiology a dying discipline? *Physiology* 8, 94-95.
- Ploeg, J., Cauilan-Cureg, M., Weerd, M.v., de Groot, W., 2011. Assessing the effectiveness of environmental education: mobilizing public support for Philippine crocodile conservation. *Conservation Letters* 4, 313-323.
- Portik, D.M., Smith, L.L., Bi, K., 2016. An evaluation of transcriptome-based exon capture for frog phylogenomics across multiple scales of divergence (Class: Amphibia, Order: Anura). *Molecular ecology resources* 16, 1069-1083.
- Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G., Sudmant, P.H., De Filippo, C., 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505, 43-49.
- QGIS Development Team, 2016. QGIS geographic information system. Open source geospatial Foundation project.
- Qiao, Y., Zhu, X., 2009. Patent protection of the Traditional Chinese Medicine and its impact on the related industries in China. *Journal of international biotechnology law* 6, 99-108.
- Qiu, Q., Zhang, G., Ma, T., Qian, W., Wang, J., Ye, Z., Cao, C., Hu, Q., Kim, J., Larkin, D.M., Auville, L., Capitanu, B., Ma, J., Lewin, H.A., Qian, X., Lang, Y., Zhou, R., Wang, L., Wang, K., Xia, J., Liao, S., Pan, S., Lu, X., Hou, H., Wang, Y., Zang, X., Yin, Y., Ma, H., Zhang, J., Wang, Z., Zhang, Y., Zhang, D., Yonezawa, T., Hasegawa, M., Zhong, Y., Liu, W., Zhang, Y., Huang, Z., Zhang, S., Long, R., Yang, H., Wang, J., Lenstra, J.A., Cooper, D.N., Wu, Y., Wang, J., Shi, P., Wang, J., Liu, J., 2012. The yak genome and adaptation to life at high altitude. *Nature Genetics* 44, 946-949.
- Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841-842.
- Ranwez, V., Delsuc, F., Ranwez, S., Belkhir, K., Tilak, M.-K., Douzery, E.J.P., 2007. OrthoMaM: A database of orthologous genomic markers for placental mammal phylogenetics. *BMC Evolutionary Biology* 7, 241.
- Ray-Chaudhuri, S.P., Sharma, T., Pathak, S., Singh, L., 1969. Chromosomes and the karyotype of the Pangolin, *Manis pentadactyla* L. (Pholidota-Mammalia). *Experientia* 25, 1167-1168.
- Reboredo Segovia, A.L., Romano, D., Armsworth, P.R., 2020. Who studies where? Boosting tropical conservation research where it is most needed. *Frontiers in Ecology and the Environment* 18, 159-166.
- Recharte Uscamaita, M., Bodmer, R., 2010. Recovery of the Endangered giant otter *Pteronura brasiliensis* on the Yavarí-Mirín and Yavarí Rivers: a success story for CITES. *ORYX* 44, 83-88.

- Ripley, B., Venables, B., Bates, D.M., Hornik, K., Gebhardt, A., Firth, D., Ripley, M.B., 2013. Package 'mass', In R package version 7.3-51.4.
- Robinson, Jacqueline A., Ortega-Del Vecchyo, D., Fan, Z., Kim, Bernard Y., vonHoldt, Bridgett M., Marsden, Clare D., Lohmueller, Kirk E., Wayne, Robert K., 2016. Genomic flatlining in the endangered Island Fox. *Current Biology* 26, 1183-1189.
- Robles, E., Piper, P., Ochoa, J., Lewis, H., Paz, V., Ronquillo, W., 2015. Late Quaternary sea-level changes and the palaeohistory of Palawan Island, Philippines. *The Journal of Island and Coastal Archaeology* 10, 76-96.
- Roe, D., 2007. Trading nature: a report, with case studies, on the contribution of wildlife trade management to sustainable livelihoods and the Millennium Development Goals, p. 83. IUCN: International Union for Conservation of Nature, Cambridge.
- Roe, D., Dickman, A., Kock, R., Milner-Gulland, E.J., Rihoy, E., 't Sas-Rolfes, M., 2020. Beyond banning wildlife trade: COVID-19, conservation and development. *World Development* 136, 105121.
- Rohling, E.J., Fenton, M., Jorissen, F.J., Bertrand, P., Ganssen, G., Caulet, J.P., 1998. Magnitudes of sea-level lowstands of the past 500,000 years. *Nature* 394, 162-165.
- Rose, K.D., 2012. The importance of Messel for interpreting Eocene Holarctic mammalian faunas. *Palaeobiodiversity and Palaeoenvironments* 92, 631-647.
- Rousseeuw, P.J., Croux, C., 1993. Alternatives to the Median Absolute Deviation. *Journal of the American Statistical Association* 88, 1273-1283.
- Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.E., Sánchez-Gracia, A., 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution* 34, 3299-3302.
- Salmona, J., Heller, R., Lascoux, M., Shafer, A., 2017. Inferring demographic history using genomic data, In *Population Genomics: Concepts, Approaches and Applications*. ed. O.P. Rajora, pp. 511-537. Springer International Publishing, Cham.
- Salzmann, U., Hoelzmann, P., 2005. The Dahomey Gap: an abrupt climatically induced rain forest fragmentation in West Africa during the late Holocene. *The Holocene* 15, 190-199.
- Sayre, R., Dangermond, J., Frye, C., Vaughan, R., Aniello, P., Breyer, S., Cribbs, D., Hopkins, D., Nauman, R., Derrenbacher, W., 2014. A new map of global ecological land units—an ecophysiological stratification approach. Washington, DC: Association of American Geographers 46.
- Sayyari, E., Mirarab, S., 2018. Testing for polytomies in phylogenetic species trees using quartet frequencies. *Genes* 9, 132.
- Scally, A., Dutheil, J.Y., Hillier, L.W., Jordan, G.E., Goodhead, I., Herrero, J., Hobolth, A., Lappalainen, T., Mailund, T., Marques-Bonet, T., McCarthy, S., Montgomery, S.H., Schwalie, P.C., Tang, Y.A., Ward, M.C., Xue, Y., Yngvadottir, B., Alkan, C., Andersen, L.N., Ayub, Q., Ball, E.V., Beal, K., Bradley, B.J., Chen, Y., Clee, C.M., Fitzgerald, S., Graves, T.A., Gu, Y., Heath, P., Heger, A., Karakoc, E., Kolb-Kokocinski, A., Laird, G.K., Lunter, G., Meader, S., Mort, M., Mullikin, J.C., Munch, K., O'Connor, T.D., Phillips, A.D., Prado-Martinez, J., Rogers, A.S., Sajjadian, S., Schmidt, D., Shaw, K., Simpson, J.T., Stenson, P.D., Turner, D.J., Vigilant, L., Vilella, A.J., Whitener, W., Zhu, B., Cooper, D.N., de Jong, P., Dermitzakis, E.T., Eichler, E.E., Flicek, P., Goldman, N., Mundy, N.I., Ning, Z., Odom, D.T., Ponting, C.P., Quail, M.A., Ryder, O.A., Searle, S.M., Warren, W.C., Wilson, R.K., Schierup, M.H., Rogers, J., Tyler-Smith, C., Durbin, R., 2012. Insights into hominid evolution from the gorilla genome sequence. *Nature* 483, 169-175.

- Schoppe, S., Katsis, L., Lagrada, L., 2019. *Manis culionensis*, In The IUCN Red List of Threatened Species. p. e.T136497A123586862.
- Schott, R.K., Panesar, B., Card, D.C., Preston, M., Castoe, T.A., Chang, B.S., 2017. Targeted capture of complete coding regions across divergent species. *Genome biology and evolution* 9, 398-414.
- Schwartz, T.S., Karl, S.A., 2008. Population Genetic Assignment of Confiscated Gopher Tortoises. *The Journal of Wildlife Management* 72, 254-259.
- Scornavacca, C., Belkhir, K., Lopez, J., Dernas, R., Delsuc, F., Douzery, E.J.P., Ranwez, V., 2019. OrthoMaM v10: Scaling-up orthologous coding sequence and exon alignments with more than one hundred mammalian genomes. *Molecular Biology and Evolution* 36, 861-862.
- Seim, I., Fang, X., Xiong, Z., Lobanov, A.V., Huang, Z., Ma, S., Feng, Y., Turanov, A.A., Zhu, Y., Lenz, T.L., Gerashchenko, M.V., Fan, D., Hee Yim, S., Yao, X., Jordan, D., Xiong, Y., Ma, Y., Lyapunov, A.N., Chen, G., Kulakova, O.I., Sun, Y., Lee, S.-G., Bronson, R.T., Moskalev, A.A., Sunyaev, S.R., Zhang, G., Krogh, A., Wang, J., Gladyshev, V.N., 2013. Genome analysis reveals insights into physiology and longevity of the Brandt's bat *Myotis brandtii*. *Nature Communications* 4, 2212.
- Shafer, A.B.A., Wolf, J.B.W., Alves, P.C., Bergström, L., Bruford, M.W., Brännström, I., Colling, G., Dalén, L., De Meester, L., Ekblom, R., Fawcett, K.D., Fior, S., Hajibabaei, M., Hill, J.A., Hoebel, A.R., Höglund, J., Jensen, E.L., Krause, J., Kristensen, T.N., Krützen, M., McKay, J.K., Norman, A.J., Ogden, R., Österling, E.M., Ouborg, N.J., Piccolo, J., Popović, D., Primmer, C.R., Reed, F.A., Roumet, M., Salmons, J., Schenker, T., Schwartz, M.K., Segelbacher, G., Senn, H., Thaulow, J., Valtonen, M., Veale, A., Vergeer, P., Vijay, N., Vilà, C., Weissensteiner, M., Wennerström, L., Wheat, C.W., Zieliński, P., 2015. Genomics and the challenging translation into conservation practice. *Trends in ecology & evolution* 30, 78-87.
- Shairp, R., Verissimo, D., Fraser, I., Challender, D., MacMillan, D., 2016. Understanding urban demand for wild meat in Vietnam: implications for conservation actions. *PLoS ONE* 11, e0134787.
- Sheil, D., 2001. Conservation and biodiversity monitoring in the tropics: realities, priorities, and distractions. *Conservation Biology* 15, 1179-1182.
- Shen, W., Le, S., Li, Y., Hu, F., 2016. SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS ONE* 11, e0163962.
- Shepherd, C.R., Connelly, E., Hywood, L., Cassey, P., 2017. Taking a stand against illegal wildlife trade: The Zimbabwean approach to pangolin conservation. *ORYX* 51, 280-285.
- Smit, A.F., Hubley, R., 2008. RepeatModeler Open-1.0. Available from <http://www.repeatmasker.org>.
- Smith, R.K., Marais, A., Chadwick, P., Lloyd, P.H., Hill, R.A., 2008. Monitoring and management of the endangered Cape mountain zebra *Equus zebra zebra* in the Western Cape, South Africa. *African Journal of Ecology* 46, 207-213.
- Soewu, D.A., Adekanola, T.A., 2011. Traditional-medical knowledge and perception of pangolins (*manis* spp) among the awori people, Southwestern Nigeria. *Journal of Ethnobiology and Ethnomedicine* 7.
- South, A., 2022. rnaturalearth: World Map Data from Natural Earth. <https://docs.ropensci.org/rnaturalearth>.

- Spicer, R.A., Su, T., Valdes, P.J., Farnsworth, A., Wu, F.-X., Shi, G., Spicer, T.E.V., Zhou, Z., 2020. Why ‘the uplift of the Tibetan Plateau’ is a myth. *National Science Review* 8, nwaa091.
- Springer, A.M., Estes, J., Van Vliet, G.B., Williams, T., Doak, D., Danner, E., Forney, K., Pfister, B., 2003. Sequential megafaunal collapse in the North Pacific Ocean: An ongoing legacy of industrial whaling? *Proceedings of the National Academy of Sciences* 100, 12223-12228.
- Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., Morgenstern, B., 2006. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Research* 34, W435-W439.
- Stanton, D.W.G., Hart, J., Galbusera, P., Helsen, P., Shephard, J., Kümpel, N.F., Wang, J., Ewen, J.G., Bruford, M.W., 2014. Distinct and Diverse: Range-Wide Phylogeography Reveals Ancient Lineages and High Genetic Variation in the Endangered Okapi (*Okapia johnstoni*). *PLoS ONE* 9, e101081.
- Steenwyk, J.L., Buida, T.J., Labella, A.L., Li, Y., Shen, X.-X., Rokas, A., 2021. PhyKIT: a broadly applicable UNIX shell toolkit for processing and analyzing phylogenomic data. *Bioinformatics* 37, 2325–2331.
- Stokes, S., Thomas, D.S.G., Washington, R., 1997. Multiple episodes of aridity in southern Africa since the last interglacial period. *Nature* 388, 154-158.
- Stringham, O.C., García-Díaz, P., Toomes, A., Mitchell, L., Ross, J.V., Cassey, P., 2021. Live reptile smuggling is predicted by trends in the legal exotic pet trade. *Conservation Letters* 14, e12833.
- Su, B., Liu, R.-Q., Wang, Y.-x., Shi, L.-m., 1994. Genetic diversity in the Chinese pangolin (*Manis pentadactyla*) inferred from protein electrophoresis. *Biochemical Genetics* 32, 343-349.
- Sun, N.C., Huang, C.C., Tseng, Y.W., Laxmi Suwal, T., Chi, M.J., Jang-Liaw, N.H., Hung, K.H., 2021a. Complete mitochondrial genome of *Manis pentadactylapentadactyla* (Mammalia: Pholidota), an endemic subspecies of Chinese pangolin: mitogenome characterisation and phylogenetic implications. *Biodivers Data J* 9, e77961.
- Sun, N.C.M., Arora, B., Lin, J.S., Lin, W.C., Chi, M.J., Chen, C.C., Pei, C.J.C., 2019. Mortality and morbidity in wild Taiwanese pangolin (*Manis pentadactyla pentadactyla*). *PLoS ONE* 14, e0212960.
- Sun, S., Dou, H., Wei, S., Fang, Y., Long, Z., Wang, J., An, F., Xu, J., Xue, T., Qiu, H., 2021b. A review of the engineering role of burrowing animals: Implication of Chinese Pangolin as an Ecosystem Engineer. *Journal of Zoological Research* 3.
- Suwal, T.L., Gurung, S., Shrestha, M.B., Ingram, D.J., Pei, K.J.-C., 2022. Human Dimensions of Pangolin Conservation: Indigenous and Local Knowledge, Ethnozoological Uses, and Willingness of Rural Communities to Enhance Pangolin Conservation in Nepal. *Journal of Ethnobiology* 42, 1-18, 18.
- Swaigood, R.R., Wang, D., Wei, F., 2018. Panda Downlisted but not Out of the Woods. *Conservation Letters* 11, e12355.
- Tapponnier, P., Lacassin, R., Leloup, P.H., Schärer, U., Dalai, Z., Haiwei, W., Xiaohan, L., Shaocheng, J., Lianshang, Z., Jiayou, Z., 1990. The Ailao Shan/Red River metamorphic belt: Tertiary left-lateral shear between Indochina and South China. *Nature* 343, 431-437.
- Tarailo-Graovac, M., Chen, N., 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics* 5, 4.10.11-14.10.14.

- Taylor, R.S., Manseau, M., Horn, R.L., Keobouasone, S., Golding, G.B., Wilson, P.J., 2020. The role of introgression and ecotypic parallelism in delineating intraspecific conservation units. *Molecular Ecology* 29, 2793-2809.
- Telfer, P.T., Souquière, S., Clifford, S.L., Abernethy, K.A., Bruford, M.W., Disotell, T.R., Sterner, K.N., Roques, P., Marx, P.A., Wickings, E.J., 2003. Molecular evidence for deep phylogenetic divergence in *Mandrillus sphinx*. *Molecular Ecology* 12, 2019-2024.
- Terhune, C.E., Gaudin, T., Curran, S., Petculescu, A., 2021. The youngest pangolin (Mammalia, Pholidota) from Europe. *Journal of Vertebrate Paleontology* 41, e1990075.
- Than, C., Ruths, D., Nakhleh, L., 2008. PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics* 9, 322.
- The UniProt Consortium, 2009. The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Research* 38, D142-D148.
- Thomson, P., Fletcher, L., 2020. No longer a forgotten species: history, key events, and lessons learnt from the rise of pangolin awareness, In *Pangolins: Science, Society and Conservation*. eds D. Challender, H. Nash, C. Waterman, pp. 335-347. Academic Press, London, UK.
- Tilak, M.-K., Allio, R., Delsuc, F., 2020. An Optimized Protocol for Sequencing Mammalian Roadkill Tissues with Oxford Nanopore Technology (ONT) V1.
- Tommaso, M.R.D., Bonnini, S., Qi, Y., 2017. Focusing on the Chinese health industry: an empirical enquiry on the TCM listed firms. Is large and private beautiful? *International Journal of Healthcare Technology and Management* 16, 77-94.
- Tougaard, C., 2001. Biogeography and migration routes of large mammal faunas in South-East Asia during the Late Middle Pleistocene: focus on the fossil and extant faunas from Thailand. *Palaeogeography, Palaeoclimatology, Palaeoecology* 168, 337-358.
- TRAFFIC, 2016. Out of Sequence: is wildlife forensics delivering as an illegal trade enforcement tool? *Wildlife DNA Forensics. Traffic Bulletin* 28, 56-58.
- Tunstall, T., Kock, R., Vahala, J., Diekhans, M., Fiddes, I., Armstrong, J., Paten, B., Ryder, O.A., Steiner, C.C., 2018. Evaluating recovery potential of the northern white rhinoceros from cryopreserved somatic cells. *Genome Research* 28, 780-788.
- UN, 2019. World population prospects 2019: Highlights. United Nations, Department of Economic and Social Affairs, Population Division. New York, USA.
- UNODC, 2016. World wildlife crime report: Trafficking in protected species, p. 101. United Nations Office on Drugs and Crime (UNODC), Vienna, Austria.
- van Boom, K.M., Breed, D., Hughes, A., Blackhurst, D., Kohn, T.A., 2022. A novel description of the *Vastus lateralis* morphology of the Temminck's ground pangolin (*Manis temminckii*). *The Anatomical Record* 305, 3463-3471.
- van der Valk, T., Díez-del-Molino, D., Marques-Bonet, T., Guschanski, K., Dalén, L., 2019. Historical Genomes Reveal the Genomic Consequences of Recent Population Decline in Eastern Gorillas. *Current Biology* 29, 165-170.e166.
- van Eck, N., Waltman, L., 2010. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* 84, 523-538.
- van Ee, C.A., 1966. A note on breeding the Cape pangolin *Manis temminckii* at Bloemfontein Zoo. *International Zoo Yearbook* 6, 163-164.
- Vavrek, M.J., 2011. Fossil: palaeoecological and palaeogeographical analysis tools. *Palaeontologia electronica* 14, 16.

- Vera-Baceta, M.-A., Thelwall, M., Kousha, K., 2019. Web of Science and Scopus language coverage. *Scientometrics* 121, 1803-1813.
- von Takach, B., Ranjard, L., Burrige, C.P., Cameron, S.F., Cremona, T., Eldridge, M.D.B., Fisher, D.O., Frankenberg, S., Hill, B.M., Hohnen, R., Jolly, C.J., Kelly, E., MacDonald, A.J., Moussalli, A., Ottewell, K., Phillips, B.L., Radford, I.J., Spencer, P.B.S., Trewella, G.J., Umbrello, L.S., Banks, S.C., 2022. Population genomics of a predatory mammal reveals patterns of decline and impacts of exposure to toxic toads. *Molecular Ecology* 31, 5468-5486.
- Voris, H.K., 2000. Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. *Journal of Biogeography* 27, 1153-1167.
- Wade, C., Giulotto, E., Sigurdsson, S., Zoli, M., Gnerre, S., Imsland, F., Lear, T., Adelson, D., Bailey, E., Bellone, R., 2009. Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* 326, 865-867.
- Wallace, A.R., 1854. On the Monkeys of the Amazon. *Annals and Magazine of Natural History* 14, 451-454.
- Wang, Q., Lan, T., Li, H., Sahu, S.K., Shi, M., Zhu, Y., Han, L., Yang, S., Li, Q., Zhang, L., Deng, Z., Liu, H., Hua, Y., 2022. Whole-genome resequencing of Chinese pangolins reveals a population structure and provides insights into their conservation. *Communications Biology* 5, 821.
- Wang, S.L., Tu, Y.C., Lee, M.S., Wu, L.H., Chen, T.Y., Wu, C.H., Tsao, E.H.S., Chin, S.C., Li, W.T., 2020a. Fatal Canine Parvovirus-2 (CPV-2) infection in a rescued free-ranging Taiwanese pangolin (*Manis pentadactyla pentadactyla*). *Transboundary and Emerging Diseases* 67, 1074-1081.
- Wang, X., Chen, W., Xiang, R., Li, L., Chen, J., Zhong, R., Xiang, H., Chen, J., 2019. Complete genome sequence of Parainfluenza Virus 5 (Piv5) from a Sunda pangolin (*Manis javanica*) in China. *Journal of Wildlife Diseases* 55, 947-950.
- Wang, Y., Turvey, S.T., Leader-Williams, N., 2020b. Knowledge and attitudes about the use of pangolin scale products in Traditional Chinese Medicine (TCM) within China. *People and Nature* 2, 903-912.
- Wang, Z., Ding, G., Chen, G., Sun, Y., Sun, Z., Zhang, H., Wang, L., Hasi, S., Zhang, Y., Li, J., Shi, Y., Xu, Z., He, C., Yu, S., Li, S., Zhang, W., Batmunkh, M., Ts, B., Narenbatu, Unierhu, Bat-Ireedui, S., Gao, H., Baysgalan, B., Li, Q., Jia, Z., Turigenbayila, Subudenggerile, Narenmanduhu, Wang, Z., Wang, J., Pan, L., Chen, Y., Ganerdene, Y., Dabxilt, Erdemt, Altansha, Altansukh, Liu, T., Cao, M., Aruuntsever, Bayart, Hosblig, He, F., Zha-ti, A., Zheng, G., Qiu, F., Sun, Z., Zhao, L., Zhao, W., Liu, B., Li, C., Chen, Y., Tang, X., Guo, C., Liu, W., Ming, L., Temuulen, Cui, A., Li, Y., Gao, J., Li, J., Wurentaodi, Niu, S., Sun, T., Zhai, Z., Zhang, M., Chen, C., Baldan, T., Bayaer, T., Li, Y., Meng, H., 2012. Genome sequences of wild and domestic bactrian camels. *Nature Communications* 3, 1202-1202.
- Warchol, G.L., 2004. The Transnational Illegal Wildlife Trade. *Criminal Justice Studies* 17, 57-73.
- Warren, W.C., Hillier, L.W., Graves, J.A.M., Birney, E., Ponting, C.P., Grützner, F., Belov, K., Miller, W., Clarke, L., Chinwalla, A.T., 2008. Genome analysis of the platypus reveals unique signatures of evolution. *Nature* 453, 175.

- Warren, W.C., Jasinska, A.J., García-Pérez, R., Svardal, H., Tomlinson, C., Rocchi, M., Archidiacono, N., Capozzi, O., Minx, P., Montague, M.J., 2015. The genome of the vervet (*Chlorocebus aethiops sabaues*). *Genome Research* 25, 1921-1933.
- Wasser, S.K., Brown, L., Mailand, C., Mondol, S., Clark, W., Laurie, C., Weir, B.S., 2015. Genetic assignment of large seizures of elephant ivory reveals Africa's major poaching hotspots. *Science* 349, 84-87.
- Wasser, S.K., Wolock, C.J., Kuhner, M.K., Brown, J.E., Morris, C., Horwitz, R.J., Wong, A., Fernandez, C.J., Otiende, M.Y., Hoareau, Y., Kaliszewska, Z.A., Jeon, E., Han, K.-L., Weir, B.S., 2022. Elephant genotypes reveal the size and connectivity of transnational ivory traffickers. *Nature Human Behaviour* 6, 371-382.
- Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E.V., Zdobnov, E.M., 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution* 35, 543-548.
- Weckworth, B.V., Hebblewhite, M., Mariani, S., Musiani, M., 2018. Lines on a map: conservation units, meta-population dynamics, and recovery of woodland caribou in Canada. *Ecosphere* 9, e02323.
- Wei, S., Sun, S., Dou, H., An, F., Gao, H., Guo, C., Hua, Y., 2022. Influence of Pleistocene climate fluctuations on the demographic history and distribution of the critically endangered Chinese pangolin (*Manis pentadactyla*). *BMC Zoology* 7, 50.
- Weir, B.S., Cockerham, C.C., 1984. Estimating F-Statistics for the Analysis of Population Structure. *Evolution* 38, 1358-1370.
- Wenhua, Y., Chen, Z., Li, Y., Wu, Y., 2012. Phylogeographic Relationships of *Scotophilus kuhlii* between Hainan Island and Mainland China. *Mammal Study* 37, 139-146, 138.
- Wesley-Hunt, G.D., Flynn, J.J., 2005. Phylogeny of the carnivora: Basal relationships among the carnivoramorphans, and assessment of the position of 'miacoidea' relative to carnivora. *Journal of Systematic Palaeontology* 3, 1-28.
- Westbury, M.V., Hartmann, S., Barlow, A., Wiesel, I., Leo, V., Welch, R., Parker, D.M., Sicks, F., Ludwig, A., Dalén, L., Hofreiter, M., 2018. Extended and continuous decline in effective population size results in low genomic diversity in the world's rarest hyena species, the brown hyena. *Molecular Biology and Evolution* 35, 1225-1237.
- Westphal, M.I., Browne, M., MacKinnon, K., Noble, I., 2008. The link between international trade and the global distribution of invasive alien species. *Biological Invasions* 10, 391-398.
- Wheeler, T.J., Clements, J., Eddy, S.R., Hubley, R., Jones, T.A., Jurka, J., Smit, A.F.A., Finn, R.D., 2012. Dfam: a database of repetitive DNA based on profile hidden Markov models. *Nucleic Acids Research* 41, D70-D82.
- White, F., 1979. The Guineo-Congolian Region and Its Relationships to Other Phytochoria. *Bulletin du Jardin botanique national de Belgique / Bulletin van de National Plantentuin van België* 49, 11-55.
- White, L.C., Fontseré, C., Lizano, E., Hughes, D.A., Angedakin, S., Arandjelovic, M., Granjon, A.C., Hans, J.B., Lester, J.D., Rabanus-Wallace, M.T., 2019. A roadmap for high-throughput sequencing studies of wild animal populations using noninvasive samples and hybridization capture. *Molecular ecology resources* 19, 609-622.
- Wicker, L.V., Lourens, K., Hai, L.K., 2020. Veterinary health of pangolins, In *Pangolins: Science, Society and Conservation*. eds D. Challender, H. Nash, C. Waterman, pp. 461-493. Academic Press, London, UK.
- Wickham, H., 2011. ggplot2. *WIREs Computational Statistics* 3, 180-185.

- Wilkie, D.S., Wieland, M., Boulet, H., Le Bel, S., van Vliet, N., Cornelis, D., BriacWarnon, V., Nasi, R., Fa, J.E., 2016. Eating and conserving bushmeat in Africa. *African Journal of Ecology* 54, 402-414.
- Willcox, D., Nash, H.C., Trageser, S., Kim, H.J., Hywood, L., Connelly, E., Ichu, G.I., Nyumu, J.K., Moumbolou, C.L.M., Ingram, D.J., 2019. Evaluating methods for detecting and monitoring pangolin populations (Pholidota: Manidae). *Global Ecology and Conservation*, e00539.
- Williams, D.R., Balmford, A., Wilcove, D.S., 2020. The past and future role of conservation science in saving biodiversity. *Conservation Letters* 13, e12720.
- Willoughby, J.R., Christie, M.R., 2019. Long-term demographic and genetic effects of releasing captive-born individuals into the wild. *Conservation Biology* 33, 377-388.
- Wilson, K.A., Auerbach, N.A., Sam, K., Magini, A.G., Moss, A.S.L., Langhans, S.D., Budiharta, S., Terzano, D., Meijaard, E., 2016. Conservation research is not happening where it is most needed. *PLoS biology* 14, e1002413-e1002413.
- Woodruff, D.S., Turner, L.M., 2009. The Indochinese–Sundaic zoogeographic transition: a description and analysis of terrestrial mammal species distributions. *Journal of Biogeography* 36, 803-821.
- Wu, H., Guang, X., Al-Fageeh, M.B., Cao, J., Pan, S., Zhou, H., Zhang, L., Abutarboush, M.H., Xing, Y., Xie, Z., Alshanqeeti, A.S., Zhang, Y., Yao, Q., Al-Shomrani, B.M., Zhang, D., Li, J., Manee, M.M., Yang, Z., Yang, L., Liu, Y., Zhang, J., Altammami, M.A., Wang, S., Yu, L., Zhang, W., Liu, S., Ba, L., Liu, C., Yang, X., Meng, F., Wang, S., Li, L., Li, E., Li, X., Wu, K., Zhang, S., Wang, J., Yin, Y., Yang, H., Al-Swailem, A.M., Wang, J., 2014. Camelid genomes reveal evolution and adaptation to desert environments. *Nature Communications* 5, 5188.
- Wu, S., Liu, N., Zhang, Y., Ma, G., 2004. Assessment of threatened status of Chinese Pangolin (*Manis pentadactyla*). *Chinese Journal of Applied and Environmental Biology* 10, 456-461.
- Wyss, A.R., Novacek, M.J., McKenna, M.C., 1987. Amino acid sequence versus morphological data and the interordinal relationships of mammals. *Molecular Biology and Evolution* 4, 99-116.
- Xiang, K.-L., Erst, A.S., Yang, J., Peng, H.-W., Ortiz, R.d.C., Jabbour, F., Erst, T.V., Wang, W., 2021. Biogeographic diversification of *Eranthis* (Ranunculaceae) reflects the geological history of the three great Asian plateaus. *Proceedings of the Royal Society B: Biological Sciences* 288, 20210281.
- Xie, X., Ye, H., Cai, X., Li, C., Li, F., Tian, E., Chao, Z., 2021. DNA Mini-Barcodes, a Potential Weapon for Conservation and Combating Illegal Trade of Pangolin. *Tropical Conservation Science* 14, 19400829211017361.
- Xiong, L., Yang, H., Wang, Y., 2015. Professor HUANG Wen-zheng's experience on the treatment of kidney disease applying Taohe Chengqi Decoction. *China Journal of Traditional Chinese Medicine and Pharmacy* 30, 2811-2813.
- Xu, J., Xia, Z., 2019. Traditional Chinese Medicine (TCM) – Does its contemporary business booming and globalization really reconfirm its medical efficacy & safety? *Medicine in Drug Discovery* 1, 100003.
- Xu, L., Guan, J., Lau, W., Xiao, Y., 2016. An overview of pangolin trade in China. *TRAFFIC Briefing Paper*. Traffic International, Cambridge, UK.

- Xue, Y., Prado-Martinez, J., Sudmant, P.H., Narasimhan, V., Ayub, Q., Szpak, M., Frandsen, P., Chen, Y., Yngvadottir, B., Cooper, D.N., 2015. Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. *Science* 348, 242-245.
- Yan, G., Zhang, G., Fang, X., Zhang, Y., Li, C., Ling, F., Cooper, D.N., Li, Q., Li, Y., van Gool, A.J., Du, H., Chen, J., Chen, R., Zhang, P., Huang, Z., Thompson, J.R., Meng, Y., Bai, Y., Wang, J., Zhuo, M., Wang, T., Huang, Y., Wei, L., Li, J., Wang, Z., Hu, H., Yang, P., Le, L., Stenson, P.D., Li, B., Liu, X., Ball, E.V., An, N., Huang, Q., Zhang, Y., Fan, W., Zhang, X., Li, Y., Wang, W., Katze, M.G., Su, B., Nielsen, R., Yang, H., Wang, J., Wang, X., Wang, J., 2011. Genome sequencing and comparison of two nonhuman primate animal models, the cynomolgus and Chinese rhesus macaques. *Nature biotechnology* 29, 1019-1023.
- Yandell, M., Ence, D., 2012. A beginner's guide to eukaryotic genome annotation. *Nature Reviews Genetics* 13, 329-342.
- Yang, C.W., Chen, S., Chang, C.Y., Lin, M.F., Block, E., Lorentsen, R., Chin, J.S., Dierenfeld, E.S., 2007. History and dietary husbandry of pangolins in captivity. *Zoo Biology: Published in affiliation with the American Zoo and Aquarium Association* 26, 223-230.
- Yang, Z., 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution* 24, 1586-1591.
- Yim, H.-S., Cho, Y.S., Guang, X., Kang, S.G., Jeong, J.-Y., Cha, S.-S., Oh, H.-M., Lee, J.-H., Yang, E.C., Kwon, K.K., Kim, Y.J., Kim, T.W., Kim, W., Jeon, J.H., Kim, S.-J., Choi, D.H., Jho, S., Kim, H.-M., Ko, J., Kim, H., Shin, Y.-A., Jung, H.-J., Zheng, Y., Wang, Z., Chen, Y., Chen, M., Jiang, A., Li, E., Zhang, S., Hou, H., Kim, T.H., Yu, L., Liu, S., Ahn, K., Cooper, J., Park, S.-G., Hong, C.P., Jin, W., Kim, H.-S., Park, C., Lee, K., Chun, S., Morin, P.A., O'Brien, S.J., Lee, H., Kimura, J., Moon, D.Y., Manica, A., Edwards, J., Kim, B.C., Kim, S., Wang, J., Bhak, J., Lee, H.S., Lee, J.-H., 2014. Minke whale genome and aquatic adaptation in cetaceans. *Nature Genetics* 46, 88-92.
- 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.
- Zain, S., 2020. Corrupting trade: An overview of corruption issues in illicit wildlife trade. *Targeting Natural Resource Corruption*, 10.
- Zanvo, S., Djangoun, C.A.M.S., Azihou, A.F., Djossa, B., Afiademanyo, K., Olayemi, A., Agbangla, C., Sinsin, B., Gaubert, P., 2022. Can DNA help trace the local trade of pangolins? Conservation genetics of white-bellied pangolins from the Dahomey Gap (West Africa). *BMC Ecology and Evolution* 22, 16.
- Zanvo, S., Djangoun, S.C.A.M., Azihou, F.A., Djossa, B., Sinsin, B., Gaubert, P., 2021. Ethnozoological and commercial drivers of the pangolin trade in Benin. *Journal of Ethnobiology and Ethnomedicine* 17, 18.
- Zanvo, S., Gaubert, P., Djangoun, C.A.M.S., Azihou, A.F., Djossa, B., Sinsin, B., 2020. Assessing the spatiotemporal dynamics of endangered mammals through local ecological knowledge combined with direct evidence: The case of pangolins in Benin (West Africa). *Global Ecology and Conservation* 23, e01085.
- Zhan, X., Dixon, A., Batbayar, N., Bragin, E., Ayas, Z., Deutchova, L., Chavko, J., Domashevsky, S., Dorosencu, A., Bagyura, J., Gombobaatar, S., Grlica, I.D., Levin, A., Milobog, Y., Ming, M., Prommer, M., Purev-Ochir, G., Ragyov, D., Tsurkanu, V., Vetrov, V., Zubkov,

- N., Bruford, M.W., 2015. Exonic versus intronic SNPs: contrasting roles in revealing the population genetic differentiation of a widespread bird species. *Heredity* 114, 1-9.
- Zhang, C., Rabiee, M., Sayyari, E., Mirarab, S., 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19, 153.
- Zhang, D.-R., Hui, H., Yu, G.-H., Song, X.-Q., Liu, S., Yuan, S.-Q., Xiao, H., Rao, D.-Q., 2020a. Shared response to changes in drainage basin: Phylogeography of the Yunnan small narrow-mouthed frog, *Glyphoglossus yunnanensis* (Anura: Microhylidae). *Ecology and Evolution* 10, 1567-1580.
- Zhang, F., Wu, S., Yang, L., Zhang, L., Sun, R., Li, S., 2015a. Reproductive parameters of the sunda pangolin *Manis javanica*. *Folia Zoologica* 64, 129-135.
- Zhang, F., Wu, S., Zou, C., Wang, Q., Li, S., Sun, R., 2016. A note on captive breeding and reproductive parameters of the Chinese pangolin, *Manis pentadactyla* Linnaeus, 1758. *ZooKeys* 2016, 129-144.
- Zhang, F., Yu, J., Wu, S., Li, S., Zou, C., Wang, Q., Sun, R., 2017. Keeping and breeding the rescued Sunda pangolins (*Manis javanica*) in captivity. *Zoo Biology* 36, 387-396.
- Zhang, G., Cowled, C., Shi, Z., Huang, Z., Bishop-Lilly, K.A., Fang, X., Wynne, J.W., Xiong, Z., Baker, M.L., Zhao, W., 2013. Comparative analysis of bat genomes provides insight into the evolution of flight and immunity. *Science* 339, 456-460.
- Zhang, H., Ades, G., Miller, M.P., Yang, F., Lai, K.-w., Fischer, G.A., 2020b. Genetic identification of African pangolins and their origin in illegal trade. *Global Ecology and Conservation* 23, e01119.
- Zhang, H., Miller, M.P., Yang, F., Chan, H.K., Gaubert, P., Ades, G., Fischer, G.A., 2015b. Molecular tracing of confiscated pangolin scales for conservation and illegal trade monitoring in Southeast Asia. *Global Ecology and Conservation* 4, 414-422.
- Zhang, P., Chen, Y.-Q., Zhou, H., Liu, Y.-F., Wang, X.-L., Papenfuss, T.J., Wake, D.B., Qu, L.-H., 2006. Phylogeny, evolution, and biogeography of Asiatic Salamanders (Hynobiidae). *Proceedings of the National Academy of Sciences* 103, 7360-7365.
- Zhang, T., Wu, Q., Zhang, Z., 2020c. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. *Current Biology* 30, 1346-1351.e1342.
- Zheng, Y., Janke, A., 2018. Gene flow analysis method, the D-statistic, is robust in a wide parameter space. *BMC Bioinformatics* 19, 10.
- 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 pentadactyla* and *Manis javanica*. *GigaScience* 5.
- Zhou, X., Meng, X., Liu, Z., Chang, J., Wang, B., Li, M., Wengel, P.O.-t., Tian, S., Wen, C., Wang, Z., 2016. Population genomics reveals low genetic diversity and adaptation to hypoxia in snub-nosed monkeys. *Molecular Biology and Evolution* 33, 2670-2681.
- Zhou, X., Sun, F., Xu, S., Fan, G., Zhu, K., Liu, X., Chen, Y., Shi, C., Yang, Y., Huang, Z., Chen, J., Hou, H., Guo, X., Chen, W., Chen, Y., Wang, X., Lv, T., Yang, D., Zhou, J., Huang, B., Wang, Z., Zhao, W., Tian, R., Xiong, Z., Xu, J., Liang, X., Chen, B., Liu, W., Wang, J., Pan, S., Fang, X., Li, M., Wei, F., Xu, X., Zhou, K., Wang, J., Yang, G., 2013. Baiji genomes reveal low genetic variability and new insights into secondary aquatic adaptations. *Nature Communications* 4, 2708.
- Zhou, X., Xu, S., Xu, J., Chen, B., Zhou, K., Yang, G., 2011. Phylogenomic analysis resolves the interordinal relationships and rapid diversification of the laurasiatherian mammals. *Systematic Biology* 61, 150-150.

- Zhou, Z., Jiang, Z., 2004. International Trade Status and Crisis for Snake Species in China. *Conservation Biology* 18, 1386-1394.
- Zimin, A.V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S.L., Yorke, J.A., 2013. The MaSuRCA genome assembler. *Bioinformatics* 29, 2669-2677.