Leibniz Institut für Zoo und Wildtierforschung
Abteilung Evolutionsgenetik

Deciphering evolutionary histories of Southeast Asian Ungulates:

Comparative phylogeography in a Biodiversity Hotspot

Publikationsbasierte
DISSERTATION

zur Erlangung des akademischen Grades
"doctor rerum naturalium" (Dr. rer. nat.)
in der Wissenschaftsdisziplin "Evolutionsgenetik"

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Univesität Potsdam

von

Renata Filipa Ribeiro Martins

Potsdam, Juni 2017

Published online at the
Institutional Repository of the University of Potsdam:
URN urn:nbn:de:kobv:517-opus4-404669
http://nbn-resolving.de/urn:nbn:de:kobv:517-opus4-404669

This dissertation is based in the following manuscripts:

1 Renata F. Martins, Jörns Fickel, Minh Le, Thanh van Nguyen, Ha. M. Nguyen, Robert Timmins, Han Ming Gan, Jeffrine J. Rovie-Ryan, Dorina Lenz, Daniel W. Förster & Andreas Wilting (2017) **Phylogeography of red muntjacs reveals three distinct mitochondrial lineages**. *BMC Evolutionary Biology*, 17 (34). DOI: 10.1186/s12862-017-0888-0

2 Renata F. Martins, Anke Schmidt, Dorina Lenz, Andreas Wilting & Jörns Fickel. **Human mediated introduction of introgressed deer across Wallace's line: historical biogeography of *Rusa unicolor* and *R. timorensis***. Under review in *Journal of Biogeography*

3 Renata F. Martins, Dorina Lenz, Daniel W. Förster, Johanna von Seth, Love Dalén, Sen Nathan, Benoît Goossens, Peter van Coeverden de Groot, Jörns Fickel & Andreas Wilting. **The lost genetic legacy of Sumatran and Javan rhinoceros: phylogeography and genetic diversity of historical populations.** In preparation (formatted for submission)

Contents

Acknowledgments

This dissertation would not exist without help and support of many people whom I wish to thank first of all.

I would like to start by thanking Prof. Dr. Jörns Fickel for the opportunity of conducting this research project under his supervision at the Leibniz Institute for Zoo and Wildlife Research (IZW). I greatly appreciate his guidance, support and time throughout the entire duration of my work. Likewise, I would like to thank Dr. Andreas Wilting for his invaluable involvement and extraordinary knowledge about all parts of this project. Also, I thank him for the opportunity to participate in his ongoing collaborations. In consequence, I would like to thank both Dr. Douglas Yu from the Center for Evolution and Ecology of Kunming, in Kunming, China and Dr. Minh Le from the Center for Resources and Ecological Studies in Hanoi, Vietnam. I deeply appreciated the opportunity to visit their labs and museums during the course of this doctoral thesis, which have inspired me more than I anticipated. I am also thankful to Prof. Dr. Heribert Hofer for allowing me to conduct my PhD project at the IZW.

I want to thank as well collaborators to whom I am very grateful, which either provided samples or advice and, in many cases, both. Robert J. Timmins was an incredible teacher about the world of muntjacs. I thank also Peter J. von Coeverden de Groot for his precious help. Likewise, to all other co-authors who have made the manuscripts presented here possible I thank you from my heart. Finally, everyone from the museum collections, listed throughout the manuscripts, I wish to thank for allowing me to visit and sample, but especially for all the quick answers to all those urgent questions I had while analyzing the data.

At this time I would like to thank all the colleagues at the Department of Evolutionary Genetics, for their support throughout the years. I would specially like to thank Dr. Daniel W. Förster for all his input and ideas and for the help with the laboratory methodologies and manuscripts. Also, I would like to thank Riddhi Patel, my office mate, with whom I learned so much. Likewise, I thank both Anke Schmidt and Ramona Taubert for their precious help in the lab. To Dorina Lenz, thank you for the help with the bioinformatics, it would have been harder without you. I would also like to thank Dr. Marie-Louise Kampmann for all she has taught me, especially what a good lab book looks like. I thank the people who sat in the office(s) close to me throughout the last years for their patience with my random outbursts, you will not miss them. Special thanks go to Tanja Noventa, who more than a big help was a needed friend during this all process. And finally to Saskia Wutke, thank you for all the experiences and all the moments during this time, you are on your way to great things.

I want to extend thanks to all the PhD students at the IZW, Berlin - the seminars; the parties; our symposium, were a great influence to get inspired and a way of sharing the joys and sorrows of completing a PhD. However, a special thanks is also needed to Sónia Fontes without whom life in Berlin wouldn't have been what it was. And to David Lehmann whom I dearly missed in those last PhD moments.

A big thank you is owed to my family and my friends back home. Ten years (or more) of friendship cannot be summarized in a few sentences, as I cannot summarize what your friendship and support meant for the completion of this work. Sofia, Adriana, João, Susana, Francisco, Marta and Pio – thank you for everything. Our little emigrant moments in Portugal do more for my mental health than you can imagine and it is also because you believed in me that this work got done. To Ingo, thank you for entering my life when you did; for your support with my work and with my daily complaints. You did much more than you think. Lastly, obrigada mãe e pai por me encorajarem sempre a seguir em frente e a ir onde fosse preciso para atingir os meus objectivos. Obrigada pela força e inspiração, o vosso suporte foi e será sempre essencial.

To all, a heartfelt thank you,

Renata

List of Abbreviations

aDNA	archival/ancient DNA
bp	base pairs
BPP	Bayesian Posterior Probability
DNA	Deoxyribonucleic acid
gDNA	genomic DNA
HTS	High Throughput Sequencing
kb	kilobases (1000 bp)
kya	Thousand years ago
m	meters
MCMC	Markov Chain Monte Carlo
mtDNA	mitochondrial DNA
My	million years
Mya	Million years ago
nuDNA	nuclear DNA
NUMTs	Nuclear copies of mitochondrial DNA
NGS	Next Generation Sequencing
PCR	Polymerase chain reaction

Summary

During the course of millions of years, evolutionary forces have shaped the current distribution of species and their genetic variability, by influencing their phylogeny, adaptability and probability of survival. Southeast Asia is an extraordinary biodiverse region, where past climate events have resulted in dramatic changes in land availability and distribution of vegetation, resulting likewise in periodic connections between isolated islands and the mainland. These events have influenced the way species are distributed throughout this region but, more importantly, they influenced the genesis of genetic diversity. Despite the observation that a shared paleo-history resulted in very diverse species phylogeographic patterns, the mechanisms behind these patterns are still poorly understood.

In this thesis, I investigated and contrasted the phylogeography of three groups of ungulate species distributed within South and Southeast Asia, aiming to understand what mechanisms have shaped speciation and geographical distribution of genetic variability. For that purpose, I analysed the mitogenomes of historical samples, in order to account for populations from the entire range of species distributions – including populations that no longer exist. This thesis is organized in three manuscripts, which correspond to the three investigated groups: red muntjacs, Rusa deer and Asian rhinoceros.

Red muntjacs are a widely distributed species and occur in very different habitats. We found evidence for gene-flow among populations of different islands, indicative of their ability to utilize the available land corridors. However, we described also the existence of at least two dispersal barriers that created population differentiation within this group; one isolated Sundaic and Mainland populations and the second separated individuals from Sri Lanka.

Second, the two Rusa species investigated here revealed another consequence of the historical land connections. While the two species were monophyletic, we found evidence of hybridisation in Java, facilitated by the expansion of the widespread sambar, *Rusa unicolor*. Consequently, I found that all the individuals of Javan deer, *R. timorensis* which were transported to the east of Sundaland by humans, to be of hybrid descent.

In the last manuscript, we were able to include samples from the extinct mainland populations of both Sumatran and Javan rhinoceros. The results revealed a much higher genetic diversity of the historical populations than ever reported for the contemporaneous survivors. Their evolutionary histories revealed a close relationship to climatic events of the Pleistocene but, more importantly, point out the vast extent of genetic erosion within these two endangered species.

The specific phylogeographic history of the species showed some common patters of genetic differentiation that could be directly linked to the climatic and geological changes on the Sunda Shelf during the Pleistocene. However, by contrasting these results I discussed that the same geological events did not always result in similar histories. One obvious example was the different permeability of the land corridors of Sundaland, as the ability of each species to utilize this newly available land was directly related to their specific ecological requirements. Taken together, these results have an important contribution to the general understanding of evolution in this biodiversity hotspot and the main drivers shaping the distribution of genetic diversity, but could also have important consequences for taxonomy and conservation of the three investigated groups.

Zusammenfassung

Im Verlauf von Jahrmillionen gestalteten evolutionäre Kräfte die Verbreitung und genetische Variabilität von Arten, indem sie die Anpassungsfähigkeit und Überlebenswahrscheinlichkeit dieser Arten beeinflussten. Da Südostasien eine außerordentlich artenreiche Region darstellt, eignet sie sich besonders, um den Einfluss dieser Kräfte zu untersuchen. Historische Klimaveränderungen hatten dramatische Auswirkungen auf die Verfügbarkeit sowie die Verbreitung von Habitaten in Südostasien, weil hierdurch wiederholt das Festland mit sonst isolierten Inseln verbunden wurde. Dies beeinflusste nicht nur, wie Arten in dieser Region verbreitet sind, sondern ermöglichte auch eine zunehmende genetische Variabilität. Zwar ist es bekannt, dass Arten mit ähnlicher Evolutionsgeschichte unterschiedliche phylogeographische Muster aufweisen können. Die zugrundeliegenden Mechanismen sind jedoch nur gering verstanden.

Diese Dissertation behandelt die Phylogeographie von drei Gruppen von Huftieren, welche im Süden und Südosten Asiens vorkommen. Dabei war das vornehmliche Ziel, zu verstehen, wie es zur Ausbildung verschiedener Arten sowie zu einer regionalen Verteilung von genetischer Variabilität kam. Hierfür untersuchte ich die mitochondrialen Genome alter Proben. Dadurch war es möglich, Populationen des gesamten Verbreitungsgebietes der jeweiligen Arten zu untersuchen – auch solche Populationen, die heutzutage nicht mehr existieren. Entsprechend der einzelnen Huftiergruppen ist diese Arbeit in drei Kapitel unterteilt: Muntjaks (*Muntiacus* sp.), Hirsche der Gattung Rusa und asiatische Nashörner. Alle drei Gruppen weisen eine Aufteilung in unterschiedliche Linien auf, was jeweils direkt auf Ereignisse des Pleistozäns zurückgeführt werden kann.

Muntjaks sind eine weit verbreitete Art, die in verschiedensten Habitaten vorkommen kann. Ich wies nach, dass es in der Vergangenheit zu genetischem Austausch zwischen Populationen von verschiedenen Inseln des Sundalandes kam. Dies deutet auf die Fähigkeit von Muntjaks hin, sich an die ehemaligen Landbrücken anzupassen. Jedoch zeige ich auch, dass mindestens zwei Hindernisse bei ihrer Verbreitung existierten, wodurch es zu einer Differenzierung von Populationen kam: eine Barriere trennte Populationen des asiatischen Festlands von denen der Sundainseln, die andere isolierte sri-lankische von restlichen Muntjaks.

Die zwei untersuchten Rusa-Arten weisen ein anderes Muster auf, was wiederum eine weitere Folge der pleistozänen Landbrücken darstellt. Beide Arten sind ausschließlich monophyletisch. Allerdings gibt es Anzeichen für die Hybridisierung dieser Arten auf Java, was durch eine frühere Ausbreitung des sambar (*R. unicolor*) gefördert wurde. Aufgrund dessen fand ich zudem, dass all jene Individuen der anderen Art, *R. timorensis*, die durch den Menschen auf die östlichen Sundainseln gebracht wurden, in Wahrheit Hybride sind.

Für den dritten Teil war es mir möglich, Proben von Vertretern ausgestorbener Populationen vom asiatischen Festland des Sumatra- und des Java-Nashorns (*Dicerorhinus sumatrensis* und *Rhinoceros sondaicus*) zu analysieren. Die Ergebnisse meiner Arbeit belegen, dass die genetische Vielfalt dieser historischen Populationen bedeutend größer war als die der heutigen Nachkommen. Ihre jeweilige Evolutionsgeschichte korreliert stark mit pleistozänen Prozessen. Außerdem betonen meine Ergebnisse das enorme Ausmaß von verlorener genetischer Diversität dieser stark bedrohten Arten.

Jede Art besitzt eine individuelle phylogeographische Geschichte. Ebenso fand ich aber auch allgemeingültige Muster von genetischer Differenzierung in allen Gruppen, welche direkt mit Ereignissen des Pleistozäns assoziiert werden können. Vergleicht man jedoch die einzelnen Ergebnisse der Arten, wird deutlich, dass die gleichen geologischen Prozesse nicht zwangsläufig in gleiche evolutive Ergebnisse resultieren. Einer der Gründe hierfür könnte zum Beispiel die unterschiedliche Durchlässigkeit der entstandenen Landkorridore des Sundaschelfs sein. Die Möglichkeit diese neuen Habitate zu nutzen und somit auch zu passieren steht im direkten Bezug zu den spezifischen ökologischen Bedürfnissen der Arten.Zusammenfassend leisten meine Erkenntnisse einen wichtigen Beitrag, die Evolution und geographische Aufteilung der genetischen Vielfalt in diesem Hotspot an Biodiversität zu verstehen. Obendrein können sie aber auch Auswirkungen auf die Erhaltung und systematische Klassifikation der untersuchten Arten haben.

General Introduction

1 General Introduction

Spatial patterns of global biodiversity are the result of dynamic processes through which species overlap and replace each other over the Earth's surface. Current species ranges have been shaped by selective forces that influenced their evolutionary histories. Without a broad understanding of what were (and are) the drivers of evolution, we can't begin to understand and plan for future challenges.

1.1 *Historical biogeography and comparative phylogeography*

While ecology is focused on understanding *how* species are distributed, evolutionary biology focuses on the question of *why* species are distributed as we see them today (Gavin, 1991). What adaptations did they acquire through generations? What lineages got extinct throughout their histories? What is the ancestor of the observed populations? There are three main processes through which species respond to environmental and geographical challenges: extinction, dispersal and evolution (Ricklefs, 1987). These processes are innately linked and, thus, interplay in the patterns we observe.

Biogeography is the science that studies the spatial patterns of biological diversity (Ebach, 2015). Traditionally focussing on past and present distribution of organisms, modern biogeography includes the study of all geographic variation found in Nature – from genes to species and ecosystems. Biogeography also aims to understand which abiotic and biotic factors influence and limit species distribution; how morphological, genetic and physiological traits of individuals vary across the species' distribution ranges; how taxa distribution differs within evolutionary lineages and geographic regions; and how human interference influenced and will continue influencing the distribution, evolution and extinction of species (Wiens, 2012). With the onset of molecular biology scientists began to understand how the basic units of evolutionary response – genes – are distributed and have evolved within this biogeographical context. This discipline is called Phylogeography (Avise et al. 1987).

Researchers, now able to use larger genomic regions, gained access to genes under selection or with different mutation rates and have developed better statistical tools. This allows a more comprehensive analysis of coalescence processes (Ronquist & Sanmartin, 2011). Besides, in more recent years, techniques that allow access molecular data of fossil and archival material have developed quickly, leading to major improvements in inferring species evolutionary histories. Integrating fossil data to calibrate molecular evolution and as a reliable record of which organisms

occurred at what places during what times in the past is the key question of historical biogeography (Lieberman, 2003; Wiens, 2012). Historical biogeography has brought phylogeography to a temporal context far larger than before, a time frame of millions of years, through periods characterized by extreme climatic changes that affected both localization and composition of biomes.

1.2 *Biodiversity hotspots: natural experimental setups*

Biodiversity hotspots are, by definition, regions not only rich in total number of species, but especially in the number of endemic species (Myers, 1988; Myers et al. 2000). They are therefore the perfect setup to compare species phylogeography, as they allow us to understand how a shared geological and climatic history led to similarities and discordances among species distribution patterns.

In this thesis, I focused on South and Southeast Asia, home to five of earth's 25 biodiversity hotspots (Myers et al. 2000): Western Ghats and Sri Lanka; South-Central China; the Philippines; Wallacea; and Sundaland (Fig. 1). Taken together, these hotspots occupy 4 % of earth's surface but harbour almost one-quarter of all terrestrial species (Lohman et al. 2011). This extraordinary diversity is the result of 50 million years (My) of geological and vicariant history, of island formation and climatic cyclical changes. It has inspired Alfred Russel Wallace, like the Galapagos inspired Charles Darwin, to develop his ideas about the relationship between geological histories and animal distributions. Southeast Asia is in fact the birthplace of biogeography (Lohman et al. 2011).

In particular, I focused on the Sundaland biodiversity hotspot because of its quite dramatic geological history. Sundaland refers to the area of land that was exposed during glacial intervals, composed of the current islands of Borneo, Sumatra, Java and Peninsular Malaysia. This ensemble of islands, sitting on a shallow continental shelf is bordered on the east by one of the most striking species geographical barriers, the Wallace line (Bacon et al. 2013). Wallace noted that terrestrial species present in Sundaland had their distribution limits at the eastern border of Sundaland, while noting at the same time that the species compositions on Wallacea, to the east of this line, were very distinct from those of Sundaland (Wallace 1863). Later Wallace's line was modified to accommodate differences in faunal compositions between Palawan and the Philippines (Huxley, 1868), referred to as Huxley's line (Fig.1). To the north, the Thai-Malay peninsula represents a second transition zone, where many species from continental Southeast Asia (Mainland) have their southern distribution limits and many Sundaland species have their northern distribution limits (Woodruff & Turner, 2009). Generally, this transition zone is centred around the Isthmus of Kra,

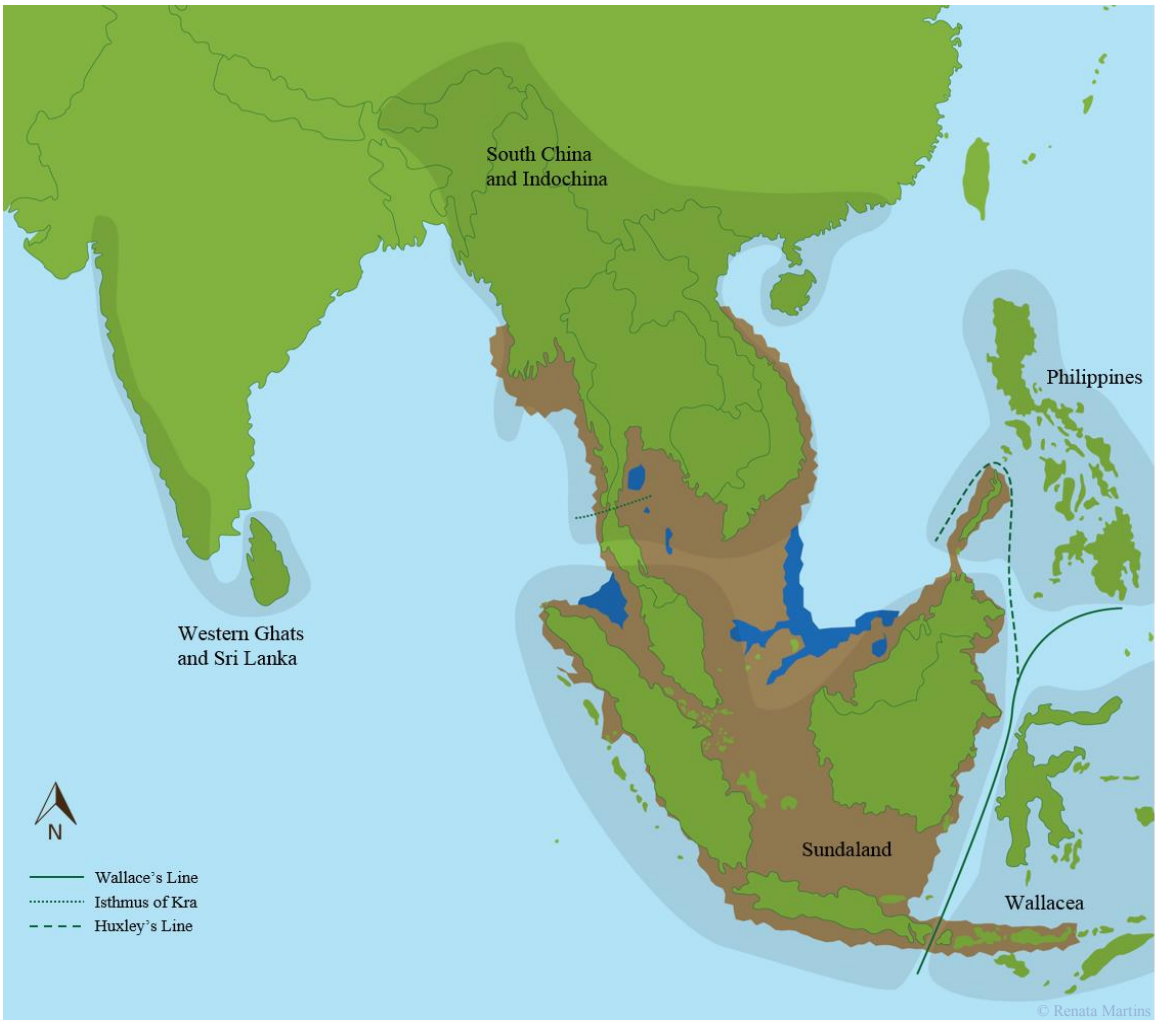

Figure 1. Map of South and Southeast Asia indicating in green the location of contemporary landmasses. Exposed shelf during times of low sea level is depicted in brown. Grey shades indicate the five, individually named, biodiversity hotspots which can be found in this region. Dark blue areas are a representation of the irrigated areas of Sundaland.

the narrowest part of the Peninsula (Fig. 1). However, this is a much more controversial biogeographical divide because the exact nature of this barrier is not fully understood (Parnell, 2013).

Another important feature of Sundaland's history are the major changes that occurred throughout past climatic cycles. While during most of the Pliocene (ca. 5 Mya) global and regional climate remained warm and humid, Pleistocene glacial cycles created dramatic changes in land availability and habitat distribution. The enormous amount of water fixed in glaciers caused sea level drops, at times 170 meters below the present level (van der Bergh et al. 1996) and modified land availability over much of Southeast Asian islands (Fig. 1). It is also likely that major rivers were draining from Thai-Malay Peninsula towards the South China Sea and other rivers flowed east of the Javan Sea (Voris, 2000).

Investigating this exciting geological, geographical and climatic history, led researchers to identify and compare species phylogeographical patterns within the region. While reconstructing the historical habitats present on Sundaland and their distribution has been so far controversial, with the advances of molecular biology, morphological and statistical inferences, some common cross-taxa patterns were found. Below I summarize some of the most common patterns described for species phylogeography histories, from where I draw the main questions addressed in the hypothesis of this thesis.

1.2.1 Sea level changes and land corridors

Low sea levels during the glacial intervals of the Pleistocene exposed the shallow Sunda Shelf, thereby increasing the available land area and connecting formerly isolated islands (Voris, 2000; Sathiamurthy & Voris, 2006). Species and populations, previously separated on the islands of the Sunda Shelf could cross and/or colonize these newly available landmasses, creating patterns of gene-flow among populations. Such gene-flow would have been facilitated by evergreen rainforest which was continuously distributed across the Sunda Shelf (Bird 2005; Raes et al. 2014; Latinne et al. 2015). However, other studies described that an area of low rainfall extended from Thai-Malay Peninsula to eastern Java, whose prevailing habitat type was either savannah or seasonal forest, often referred to as 'savannah corridor' (Gorog et al. 2004; Lim et al. 2010; Wurster et al. 2010). More recent analyses (Sheldon et al. 2015) proposed that one unifying scenario to describe Sundaland's vegetation cover might be too simplistic. They suggest rather several phases. In the early Pleistocene a dry-savannah corridor was present (more than once) in central Sundaland, allowing dry-habitat species to disperse southwards and reach Java. After this initial climate period, the last 1 My were characterized by the presence of ever-wet forest, which would have allowed populations of rainforest refugia to disperse and come in contact.

This notion that land corridors would promote gene-flow between island populations led early biogeographic theories to state that population genetic differentiation would be low between islands, as low sea level would reduce differentiation between metapopulations (Papadopoulou & Knowles, 2015). However, recent molecular studies found not only sharp breaks between lineages distributed in different islands, but mounting evidence also points to the existence of substantial intra-island genetic diversity (e.g. Leonard et al. 2015; Merckx et al. 2015; Demos et al. 2016).

Therefore, the interplay of vicariance and dispersal may have had a greater impact than what former studies reported and more analyses of rainforest dependent taxa are needed to understand the evolutionary history of Sundaland species. In this context, I drew my first question:

Question 1: Did the available landmasses of Sundaland allow for species/populations dispersal and thus do we find evidence of gene-flow between islands?

1.2.2 Isthmus of Kra as a biogeographic barrier

The northern border of Sundaland is located in the Thai-Malay Peninsula, around the Isthmus of Kra (Fig. 1; de Bruyn et al. 2005; Parnell, 2013). Several species have their distribution range limits around this region (e.g. crustaceans: de Bruyn et al. 2005; plants: Liao et al. 2009), so the Isthmus of Kra was considered as a strong biogeographic barrier to species dispersal (Parnell, 2013). Nevertheless, the exact latitude of the species northern and southern range limits was not always concordant, and so the Isthmus of Kra was defined as a species transition zone (Lohman et al. 2011). Molecular studies on the geographic distribution of genetic diversity around the Isthmus of Kra have been less extensive than those focusing on the phylogeography of the Sundaland, but many showed that populations from Sumatra and Java were closely related populations from Peninsula Malaysia (Moyle et al. 2012; Den Tex & Leonard, 2013; Luo et al. 2014).

Additionally, the reasons for how and why this region became a transition zone are also unclear. Some authors support strict vicariant scenarios that state that during high sea levels the peninsula would become an archipelago (Lohmann et al. 2011) creating less available area and consequently increase the risk of local population extinction (Woodruff & Turner, 2009). These sea transgressions occurred mostly in Miocene and Pliocene, but are not described during the Pleistocene (Dejtaradol et al. 2016). Therefore, other authors believe that the latitudinal distribution of forest (Baltzer, 2007), associated with species ecological adaptations on each side of this region is the most likely scenario to explain this phylogeographical pattern (Radchuck et al., *unpublished*).

To test for the presence of phylogeographic patterns of genetic differentiation at the Thai-Malay peninsula, it is necessary to investigate both Indochinese and Sundaic populations, as the existence of a true biogeographic divide in around the Isthmus of Kra would likely result in genetic differentiation north and south of this barrier. Therefore, I wanted to investigate:

Question 2: Is there evidence for differentiation between populations north and south of the transition zone around Isthmus of Kra?

1.2.3 Genetic differentiation among Mainland populations

While some species are truly Sundaland endemics, others are widespread throughout South and Southeast Asia. Populations of many mainland species were often considered under the same subspecies, but molecular studies revealed the existence of well differentiated lineages or even cryptic (sub-) species (e. g. dhole: Iyengar et al. 2005; birds: Fuchs et al. 2008; rodents: Lu et al. 2015). A study analyzing the molecular diversity among populations of a widespread forest-dwelling frog (*Limnonectes kuhlii*) reported a surprisingly high number of cryptic genetic lineages

across the species mainland distribution (McLeod, 2010). The author suggested the split of these lineages into different species and, in fact, claimed that there isn't any geographically widespread forest-dwelling frog species in all of Southeast Asia. Studies on mammal species have revealed similar patterns of genetic differentiation among mainland populations, e.g. the common palm civet, *Paradoxurus hermafroditus* (Patou et al. 2010), and three murine rodents species (Latinne et al. 2015). Both studies indicated that such divergence patterns were most likely a consequence of climatic cycles from the Pleistocene. The onset (~ 5 Mya) of a drier climate caused contractions of mainland wet zones, culminating in the disappearance of rainforests from most of Indochina (Karanth, 2003) and promoting the contraction of rainforest dependent species to refugia areas (likely mountain tops). Other vicariant processes, such as geographical barriers have also been appointed for patterns of genetic differentiation among mainland Southeast Asian lineages. For example, mountain chains, such as the Arakan mountain range of Myanmar or the Brahmaputra river (running between India and Indochina), likely created and maintained differentiation between populations of e.g. murine rodents and open-habitat vs. forest dwelling non-volant mammals respectively (Chingangbam et al. 2015; Meijaard & Groves, 2006).

The results from species with very different ecological requirements, point to the existence of several vicariant factors at play which affect both specialist and generalist species. This scenario directly leads to the third research question:

Question 3: Are there differentiated lineages within widely distributed species in mainland Southeast Asia?

1.3 *Selecting study species*

In this thesis I have selected to study the evolutionary patterns of Southeast Asian Ungulates. Ungulates (hoofed mammals) are a large, diverse, and highly successful group of terrestrial mammals. Almost all economically important domestic species are ungulates, including pigs, cows and horses. However, this group also includes some of the most conspicuous and iconic species (e.g. elephants and rhinos), and some of the most secretive and poorly understood ones (e.g. saola and aardvark). Ungulates in South and Southeast Asia are both diverse and ancient (Ahrestani & Sankaran, 2016) and occur in a wide variety of habitats. There are about 15 endemic ungulate populations in Sundaland and additional 84 ungulate species distributed throughout South and Southeast Asia (Ahrestani & Sankaran, 2016), that in their sum represent the highest number of threatened large herbivore species worldwide (Ripple et al. 2015). The ungulate species community on Southeast Asia is tightly connected to the geological history of the region. For example, the rapid divergence of three deer genera (*Axis*, *Cervus*, *Dama*) occurred in Southeast Asia in response to the

climate change from Miocene to Pliocene (Meijaard & Groves, 2004). The genus *Sus* (pigs) was likewise a very successful group, having colonized most of continental Eurasia and Southeast Asia, including all islands of Sundaland, Wallacea and the Philippines (Frantz et al. 2016). There, *Sus* diverged into multiple species, all restricted to very small ranges among these islands.

To answer the questions postulated above, I have focused on elucidating the phylogeographic patterns not of single species, but of species pairs, which share a common evolutionary history and which occur in South and Southeast Asia, not necessarily in sympatry. For this, I have selected three pairs of ungulate species that are described below in detail:

1.3.1 Red muntjacs

If there is a group of mammals contributing to the ever changing number of species being described in recent years, it is the muntjacs. These small deer belong to the tribe Muntiacini and concordance about the number of genera, species and subspecies has yet to be reached due to poorly informative morphological characters, extreme chromosomal number variations and overlapping distribution ranges.

Within the genus *Muntiacus*, the red muntjac (*sensu lato*) has one of the widest distributions and has also one of the most controversially discussed taxonomies. Because its distribution range, from South Asia, including Sri Lanka and mainland Southeast Asia, to all major and some of the lesser islands of the Sunda Shelf, some authors have split the red muntjacs into several species (Groves & Grubb, 2011). One of the most accepted taxonomic classification elevated mainland populations to species level (Groves, 2003), separating muntjacs from India, Indochina and Southern China as *Muntiacus vaginalis* from Sundaic muntjacs, *M. muntjak*.

In contrast to all other muntjacs, red muntjacs are *the* habitat generalists; they occur in a very wide range of habitats, from evergreen forest to secondary and highly-degraded forests, and from crop and savannah lands up to altitudes of 1500 m (Mattioli, 2011). The current wide distribution of red muntjacs throughout South and Southeast Asia indicates that these species have managed to survive the extreme geological and climatic changes of the Pleistocene. However, proposed taxonomic splits within red muntjacs have yet to be examined with molecular tools and underestimated genetic structuring might exist throughout their distribution ranges. Therefore, I investigated the phylogeography patterns of red muntjac populations from South and Southeast Asia in order to understand if any barriers to gene-flow were present among this widely distributed generalist species.

1.3.2 Rusa and Javan deer

Richard Lydekker (English naturlist, 1849-1915) said that "few groups of deer are more difficult to understand than the various kinds of sambar" (*Rusa sp*.) (Lydekker, 1898). The monophyly of *Rusa*, which was only recently accepted as a genus (Meijaard & Groves, 2004), is an ongoing debate of molecular biologists and morphologists alike. Four *Rusa* species are currently recognized – The Philippine deer (*Rusa marianna);* the Philippine spotted deer (R. *alfredi*), the Javan deer (*R. timorensis*); and the sambar (*R. unicolor*).

The two latter species, *Rusa unicolor* and *Rusa timorensis,* have a peculiar distribution on the islands of Sundaland. While the sambar is distributed throughout South and Southeast Asia, debatably being the most widespread species in the region (Ahrestani & Sankaran, 2016), the Javan deer is only native to Java and Bali. Its occurrence beyond the Sunda Shelf across the Wallace line (in Timor, Sulawesi and the Moluccas) is the result of historical human introductions (Heinsohn, 2002). Like the red muntjacs, the sambar currently occurs throughout most of Southeast Asia in a broad variety of habitats, indicating that this species would have been able to occupy a large distribution area even throughout the climatic and vegetational changes of the Pleistocene. Consequently, it is also likely that sambar were able to utilize the land corridors of Sundaland during the times of low sea levels. The dichotomic distributions of the sambar and the Javan deer then beg the question of what maintained genetic differentiation between the two closely related species.

Therefore, I describe the phylogeographic patterns of both species, including Javan deer from their introduction range, in order to understand what barriers kept the two species separated and if lineages within the widely distributed sambar, *R. unicolor* show signs of cryptic genetic structuring.

1.3.3 Sumatran and Javan rhinoceros

The family Rhinocerotidae, to which all living rhinoceros species belong, first emerged in the Late Eocene (56-33.9 Mya) in Eurasia (Hanson, 1989). Once very rich in species numbers and with a widespread geographical distribution, this family has collapsed to only five remaining species. Two of these species are native to Africa: the white rhino (*Ceratotherium simum*) and the black rhino (*Diceros bicornis*) and three to Southern Asia: the Indian rhino (*Rhinoceros unicornis*), the Javan rhino (*Rhinoceros sondaicus*) and the Sumatran rhino (*Dicerorhinus sumatrensis*). Despite their iconic status, the true phylogeny of the extant species of rhinoceros is still debated, a result from the various assessments obtained from different (often not truly diagnostic) markers and misconceptions regarding morphological traits and geographical distribution of the species (Dinerstein, 2011).

Due to habitat loss and poaching, Sumatran rhinos today are restricted to remote areas in Borneo and Sumatra mostly in upland or highland forests. It is unlikely that more than 100 individuals are

left in the wild (Ahmad Zafir et al. 2011). The Javan rhino experienced a similar dramatic population decline and today likely only 60 individuals are left in the wild in one isolated population in Java. Javan rhinos occur in lowland rain forest associated with large floodplains and wallows (van Strien et al. 2008).

Although both species are nowadays confined to very small remaining areas, their historical distribution was much larger. Sumatran and Javan rhino populations occurred from the Himalayan foothills in India, Indochina, South China, Peninsula Malaysia and the major islands of Sundaland. Investigating these species historical genetic diversity will not only reveal if the climate changes of the Pleistocene are linked to lineage differentiation, but it will also allow assessing if the low genetic diversity reported for modern populations was a result of genetic erosion through population extirpation.

1.4 Selection of methodologies

In order to have a more comprehensive view of the genetic patterns as outlined above, information from the complete distribution range of species is compulsory. Obtaining field samples in Southeast Asia is virtually impossible due to legal restrictions, habitat inaccessibility but, most importantly, the conservation status of the investigated species. Likewise, some species no longer occur in parts of their historical distribution and thus cannot be sampled there anymore. However, museum collections preserve an extraordinary number of individuals, populations and species – specimens that were collected throughout distribution ranges, as they were more than 100 years ago (Suarez & Tsutsui, 2004), before habitat fragmentation, over-hunting or local extinctions took their toll. Therefore, museum collections represent the historical distribution of species and individual pheno-/genotypes, acting as repositories of geographical and temporal variation (Austin & Melville, 2006).

For the purpose of this dissertation, I obtained permission to sample the mammal collections of several national and international institutes (Table 1).

Table 1. Summary of all collections from where samples were obtained. Details about the contact person for each collections and the collector of the samples are also provided.

	Country	City	Museum	Contact	Sample collector	Year
Europe	Germany	Berlin	Museum für Naturkunde	F. Mayer	This study	2013-2015
		Stuttgard	Staatliches Museum für Naturkunde	S. Merker	This study	2013
		Bonn	Zoological Research Museum Alexander Koenig	R. Hutterer	This study	2013
		Frankfurt	Naturmuseum Senckenberg	I. Ruff	This study	2014
		Munich	Zoologische Staatssammlung	M. Hiermeier	This study	2014
		Dresden	Senckenberg Naturhistorische Sammlungen	C. Stefan	This study	2014
	The Netherlands	Leiden	Naturalis	P. Kamminga	This study	2013-2015
		Utrecht	Universiteitsmuseum	P.H. Lambers	Dr. Peter J. vC. de Groot	2015
	France	Paris	Muséum National d'histoire naturelle	G. Veron	This study	2014
	Austria	Vienna	Naturhistorisches Museum	F. Zachos	This study	2014
	Denmark	Copenhagen	Zoologisk Museum	D.K. Johansson	This study	2015
	United Kingdom	London	Natural History Museum	R.P. Miguez	Dr. Peter J. vC. de Groot	2015
	Switzerland	Bern	Naturhistorisches Museum	S.T. Hertwig	Dr. Peter J. vC. de Groot	2015
North America	USA	Washington	Smithsonian Institution National Museum of Natural History	J.A. Coddington	Dr. Peter J. vC. de Groot	2015
	USA	New York	American Museum of Natural History	N.B. Simmons	Dr. Peter J. vC. de Groot	2015
Asia	Vietnam	Hanoi	Centre for Natural Resources and Environmental Studies	Dr. Minh Le	Dr. Minh Le	2015
	China	Kunming	Kunming Natural History Museum of Zoology	Dr. Douglas Yu	Dr. Douglas Yu	2015

Archival samples have inherent characteristics that in the past have impeded their inclusion in molecular studies. DNA from archival samples is highly degraded, often with fragments smaller than 100 bp (Rowe et al. 2011, Paijmans et al. 2012). This degradation can be caused by either

natural post-mortem chemical processes (Shved et al. 2014) or by exposure to the elements, UV light, handling and storage (Burrell et al. 2015). As a consequence, DNA fragments extracted from archival material are normally shorter than desired for classical laboratory techniques such as PCR. Long term exposure and storage might also contaminate museum material with foreign DNA. Besides that, even under the most controlled lab procedures, endogenous target DNA yield is generally low.

These characteristics of archival (museum) and ancient DNA (aDNA), however, became an advantage for more recent sequencing methodologies, such as High Throughput Sequencing (HTS). HTS is performed by creating DNA 'libraries', where DNA fragments are coupled on both ends with adapters of known sequence, thus allowing for the unknown target fragments to be read (Paijmans et al. 2012). And because most HTS sequencing platforms can only read short fragments, the naturally degraded aDNA samples can be readily transformed into sequencing libraries (e. g. Maricic et al. 2010). However, many archival samples contain less than 5 % endogenous DNA, rendering it necessary to enrich them for the targeted molecular markers. Hybridization capture methods enrich for the desired target DNA and ensure that downstream processes will become more efficient and yield better output data. These methods use DNA or RNA biotine-ligated baits with high sequence similarity to the target DNA, which are either immobilized onto a solid phase or are in solution. Baits and target DNA are then allowed to hybridize, while non-target DNA (having no sequence similarity to the baits) will be washed away before the final elution (Meyer & Kircher, 2010).

Traditional phylogeographic studies have used short mitochondrial gene fragments to estimate genetic variability through space and time. Mitochondrial DNA was the preferred marker due to its inherent characteristics: it is maternally inherited, therefore lacking recombination and its relatively high mutation rate compared to coding nuclear genes. Maternally inherited markers have lower effective population sizes (in mammals it is ¼ of 2n nuclear markers), which results in rapidly becoming reciprocally monophyletic and thus showing clear relationships between individuals (Paijmans et al. 2012). While it has been demonstrated that using short mitochondrial fragments only might lead to erroneous results regarding coalescence (Lari et al. 2011), analyzing the full mitogenome was reported to alleviate rooting effects in phylogeographic studies (Hirase et al. 2016). Finally, the complete mitogenome presents slower mutation rates than non-coding hyper-variable regions of the genome (like microsatellites), but higher rates than genes under purifying selection. Therefore, mitochondrial DNA is the most suitable marker to detect divergence events or geographical structuring among populations (Hofreiter & Stewart, 2009).

More specifically, mitogenomes are especially well suited as markers in studies employing aDNA, as they provide highly informative sequences even with small datasets and, most importantly, they occur in higher number of copies in the cell compared with the single nuclear

genome. This means that, for the majority of studies, it is the only reliable source of genetic information. Therefore, in my thesis I have specifically targeted the mitochondrial genome as the genetic marker for assessing the proposed phylogeographic patterns.

2 Aims of this study

Several past studies investigated the biogeography of terrestrial fauna of Southeast Asia and, nevertheless, the complex evolutionary patterns within this biodiversity hotspot are poorly understood. The goal of this thesis was to evaluate how a shared history of geological and climatic changes has shaped the evolutionary history of different species. I predicted that, by comparing species specific phylogeographic histories, contrasting patterns will be the result of species specific ecological requirements, history of local extinctions and their ability to disperse.

Within the historical context described in this chapter, I aimed to investigate and contrast the phylogeographic patterns of three pairs of ungulate species distributed within South and Southeast Asia. The selected species pairs are composed of widespread species and more locally endemics, of forest dependent specialist species and true habitat generalists and also endangered species of flagship value. In addition to the three main general questions postulated throughout this chapter, investigating the phylogeography of these species will also aid to the ongoing taxonomic discussion, as many of the groups have not been assessed before through molecular methods. Therefore, I organized this dissertation in three manuscripts.

In manuscript I, I investigated the phylogeography of red muntjacs (*sensu lato*) to identify patterns of differentiation of their mitochondrial lineages across their distribution range and correlate these patterns with the aforementioned geological and climatic events.

In manuscript II I focused on the evolutionary histories of *Rusa unicolor* and *Rusa timorensis*, among their native and introduced distribution ranges. I aimed at understanding the effect of the available land corridors in maintaining speciation between the two geographically and phylogenetically close species.

Manuscript III dealt with the phylogeography of the endangered rhinoceros species *Dicerorhinus sumatrensis*, Sumatran rhino, and *Rhinoceros sondaicus*, Javan rhino, in order to understand if the dramatic changes of the Pleistocene created patterns of lineage differentiation or if, conversely, the available land corridors were permissive for these forest dependent species to disperse and generate gene-flow between populations.

Manuscript I

Phylogeography of red muntjacs

Phylogeography of red muntjacs reveals three distinct mitochondrial lineages

Renata F. Martins[1], Jörns Fickel[1,2], Minh Le[3,4], Thanh van Nguyen[4], Ha M. Nguyen[4,*], Robert Timmins[5], Han Ming Gan[6,7], Jeffrine J. Rovie Ryan[7], Dorina Lenz[1], Daniel W. Förster[1], Andreas Wilting[1]

[1] Leibniz Institute for Zoo and Wildlife Research (IZW), Alfred-Kowalke-Straße 17, 10315 Berlin, Germany;

[2] Potsdam University, Institute for Biochemistry and Biology, Karl-Liebknecht-Str. 22-24, 14476 Potsdam-Golm, Germany;

[3] Faculty of Environmental Science, Hanoi University of Science, Vietnam National University, 334 Nguyen Trai Road, Hanoi, Vietnam;

[4] Centre for Natural Resources and Environmental Studies, Vietnam National University, 19 Le Thanh Tong Street, Hanoi, Vietnam;

[5] 1123 Monroe Street, Evanston, Illinois, 60202, USA;

[6] School of Science, Monash University Malaysia, 47500 Bandar Sunway, Malaysia;

[7] Genomics Facility, Tropical Medicine and Biology Multidisciplinary Platform, Monash University Malaysia, 47500 Bandar Sunway, Malaysia;

[8] National Wildlife Forensic Laboratory (NWFL), Department of Wildlife and National Parks (DWNP) Peninsular Malaysia, 56100 Kuala Lumpur, Malaysia;

* Current address: U.S. Agency for International Development, Governance for Inclusive Growth Program, Chemonics International Inc, 115 Tran Hung Dao Street, Hanoi, Vietnam

Published in *BMC Evolutionary Biology*, 2017
(DOI: 10.1186/s12862-017-0888-0)

Martins *et al. BMC Evolutionary Biology* (2017) 17:34
DOI 10.1186/s12862-017-0888-0

BMC Evolutionary Biology

RESEARCH ARTICLE Open Access

Phylogeography of red muntjacs reveals three distinct mitochondrial lineages

Renata F. Martins[1*], Jörns Fickel[1,2], Minh Le[3,4], Thanh van Nguyen[4], Ha M. Nguyen[4,9], Robert Timmins[5], Han Ming Gan[6,7], Jeffrine J. Rovie-Ryan[8], Dorina Lenz[1], Daniel W. Förster[1] and Andreas Wilting[1]

Abstract

Background: The members of the genus *Muntiacus* are of particular interest to evolutionary biologists due to their extreme chromosomal rearrangements and the ongoing discussions about the number of living species. Red muntjacs have the largest distribution of all muntjacs and were formerly considered as one species. Karyotype differences led to the provisional split between the Southern Red Muntjac (*Muntiacus muntjak*) and the Northern Red Muntjac (*M. vaginalis*), but uncertainties remain as, so far, no phylogenetic study has been conducted. Here, we analysed whole mitochondrial genomes of 59 archival and 16 contemporaneous samples to resolve uncertainties about their taxonomy and used red muntjacs as model for understanding the evolutionary history of other species in Southeast Asia.

Results: We found three distinct matrilineal groups of red muntjacs: Sri Lankan red muntjacs (including the Western Ghats) diverged first from other muntjacs about 1.5 Mya; later northern red muntjacs (including North India and Indochina) and southern red muntjacs (Sundaland) split around 1.12 Mya. The diversification of red muntjacs into these three main lineages was likely promoted by two Pleistocene barriers: one through the Indian subcontinent and one separating the Indochinese and Sundaic red muntjacs. Interestingly, we found a high level of gene flow within the populations of northern and southern red muntjacs, indicating gene flow between populations in Indochina and dispersal of red muntjacs over the exposed Sunda Shelf during the Last Glacial Maximum.

Conclusions: Our results provide new insights into the evolution of species in South and Southeast Asia as we found clear genetic differentiation in a widespread and generalist species, corresponding to two known biogeographical barriers: The Isthmus of Kra and the central Indian dry zone. In addition, our molecular data support either the delineation of three monotypic species or three subspecies, but more importantly these data highlight the conservation importance of the Sri Lankan/South Indian red muntjac.

Keywords: Phylogeography, Archival DNA, Muntjac, Southeast Asia, Species complex

Background

The number of recognized deer species has increased in recent decades, and it continues to do so due to rare discoveries of new forms in the wild, increased molecular efforts and the careful reexamination of museum collections. For example, the genus Muntiacus has increased in the number of named species through discovery of the Gongshan muntjac (*Muntiacus gongshanensis*) from the wild in 1990 [1], the Putao muntjac (*Muntiacus putaoensis*) from Myanmar, described based on molecular comparisons [2] and the Bornean Yellow muntjac (*Muntiacus atherodes*) described from museum skulls and skins in 1982 following a reappraisal of the muntjac taxa described previously from Borneo [3].

The genus *Muntiacus* Rafinesque (1815) comprises an undefined number of species and subspecies all native to South, Southeast and East Asia. Although muntjacs are studied for their dramatic variation in chromosome numbers [4, 5], taxonomic concordance within this genus has not been achieved yet, due to lack of molecular studies combined with, in some cases, limited morphological or ecological differences. Although the genus

* Correspondence: martins@izw-berlin.de
[1]Leibniz Institute for Zoo and Wildlife Research (IZW), Alfred-Kowalke-Straße 17, 10315 Berlin, Germany
Full list of author information is available at the end of the article

is mainly composed of endemics and species with small regional distributions (independent of the taxonomic revision), there are two exceptions: Reeve's muntjac *M. reevesi*, which has a large native range largely confined to mainland China and even more striking, the red muntjacs. Even if different species/subspecies are considered, red muntjacs have the largest distributions of all muntjacs, ranging from the Indian subcontinent (Pakistan, India, Sri Lanka, Bangladesh, Nepal and Bhutan) to the Indochinese Peninsula (Vietnam, Laos, Myanmar, Thailand) and southeastern China and across the Malay Peninsula towards all the major islands of Sundaland (Borneo, Java, Sumatra) and Lesser Sunda Islands (Belitung, Bangka, Bali, Lombok, Lampung). In contrast to the majority of other muntjac species, red muntjacs occur in a wide range of habitats [6, 7]. They are generally forest-dwelling and occupy habitats from deciduous to evergreen forests and also occur in secondary forests and exotic commercial plantations [8, 9]. They have also been found in grasslands or shrubland savannahs, croplands and in altitudes commonly up to 1500 m asl (in extreme cases even up to 3500 m, [10]). Additionally, although habitat loss and hunting has already caused significant population declines throughout their range, in contrast to other ungulates in South and Southeast Asia, red muntjacs seem to be more resilient to habitat modifications and hunting [7].

Through most of the 20th century red muntjacs were generally classified as a single species *Muntiacus muntjak*, although it was noted by some authors that some forms might be better treated as distinct species (e.g. [11]). Groves in 2003 [12] elevated the mainland form to species level, designating it as *M. vaginalis*. As this study was solely based on a comparison between the karyotypes of one individual from Peninsular Malaysia and the karyotypes of Indochinese individuals, this classification has not been universally adopted. In 2008 *The IUCN/SSC Red List of Threatened Species* provisionally accepted this classification, but noted the need for corroborating studies. More recently, however, Groves and Grubb [13] described six 'good' species within the 'red muntjac group', based mainly on a few morphological characters and geographical distribution of populations (see Additional file 1: Table S1).

Here, we investigate the geographic distribution of mitochondrial lineages among red muntjac populations in order to address some of these taxonomic uncertainties, and discuss the results within the context of geological events that took place in Southeast Asia since the Pleistocene. This region's geography has been heavily impacted by the climatic fluctuations of the Pleistocene, where the low sea levels during the glacial periods repeatedly exposed the shallow continental shelf, creating land corridors between the islands and the mainland [14]. While phylogeographic patterns of some species can be explained by the presence of these land corridors, those of others cannot; they seem to be the result of other vicariant events that took place in Southeast Asia [15, 16]. Characterising phylogeographic patterns within this widely distributed species complex will allow a better understanding of how climatic variations affected generalist species and will therefore help to understand the mechanisms leading to the evolution and speciation of mammals in the biodiversity hotspot of South and Southeast Asia.

Methods

Samples and DNA extraction

Archival samples from 92 red muntjacs were collected at several natural history museums (See final dataset in Additional file 1: Table S2). We collected nasal bones and dry tissue from skulls and skeletons, or drilled antlers when only this material was available. Only samples with a known locality were collected and archival samples from zoos were only included if they had a known wild origin. In addition, we also opportunistically collected 24 contemporary samples from Vietnam, Laos and Peninsular Malaysia (Fig. 1). Contemporaneous samples were all collected from dead animals during routine field surveys or confiscated trophies from hunters. All molecular work with archival samples was performed in a separate facility, operated under controlled conditions in order to minimize contamination. For DNA extraction of archival samples, we followed the DNeasy Blood & Tissue kit (Qiagen, Hilden, Germany) protocol with some modifications: digestion with Proteinase K occurred overnight at 56 °C and elution was preceded by an incubation period of 20 min. at 37^0 C. DNA from fresh samples was extracted with the DNeasy Blood & Tissue kit (Qiagen) and the two samples from Peninsular Malaysia were extracted using conventional SDS/Proteinase K digestion followed by chloroform extraction [17]. Sham extractions were performed for archival and fresh samples and served as negative controls. They were processed like actual samples with every step and included in follow-up PCR reactions.

Library building and hybridization capture

All samples, including extraction negative controls, were processed into libraries for sequencing with either the Ion Torrent Personal Genome Machine (PGM; Thermo Fischer Scientific, USA) or the MiSeq (Illumina, San Diego, CA, USA). For PGM libraries we followed the manufacturer's protocol with modifications: we used the gDNA plus Fragment Library Kit (Thermo Fisher Scientific, USA) but all reactions were performed in a quarter of the suggested volume and blunt ending included a heat inactivation of the enzyme (20 min at 72 °C), so

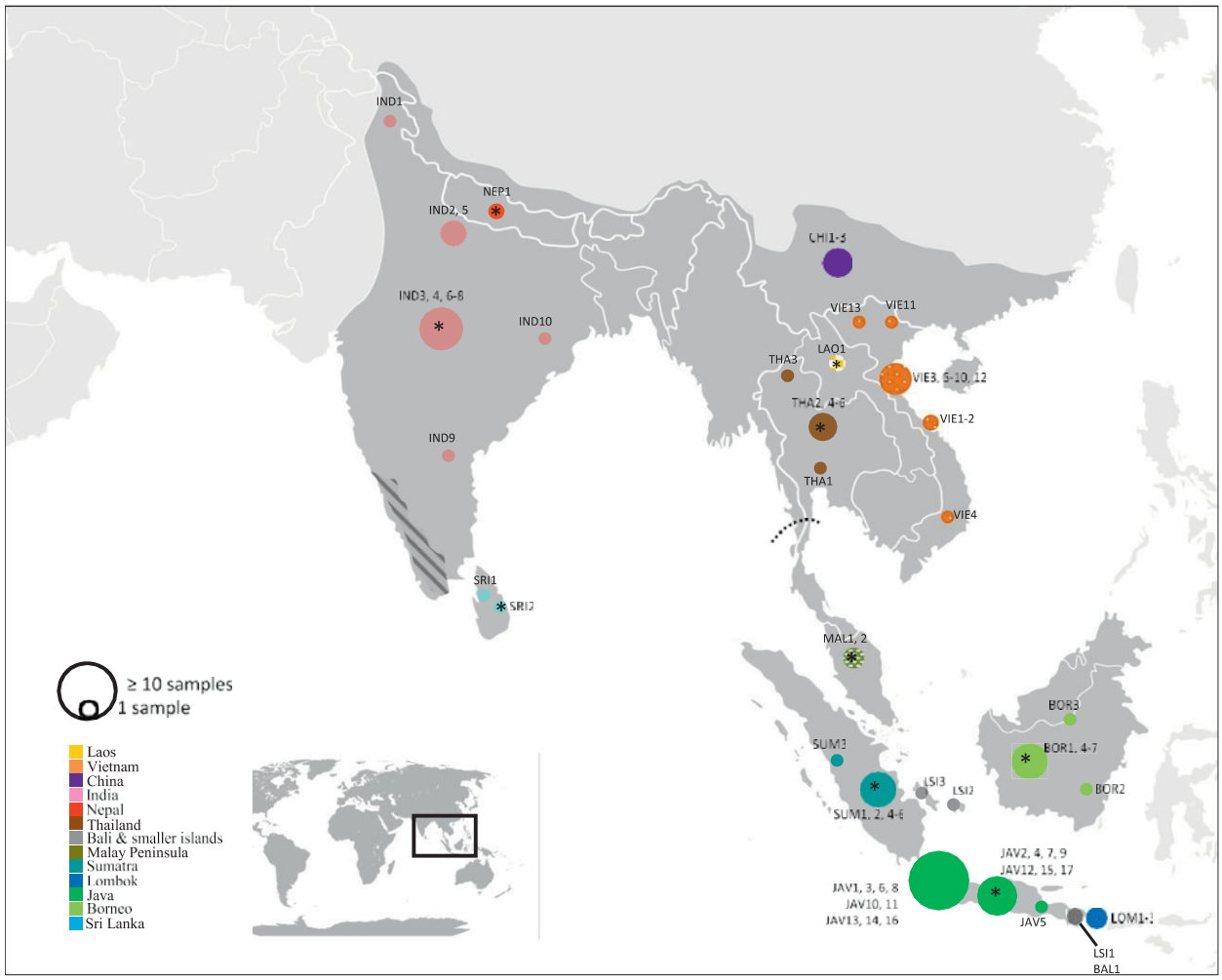

Fig. 1 Illustrative map depicting the combined range of all red muntjacs (*dark grey*). The *dashed line* indicates the relative position of the Isthmus of Kra and the area indicated by *dark grey lines* indicates the relative position of the Western Ghats. *Circles* indicate sample origin according to colour and size is relative to sample size. *Checkered patterns* indicate contemporaneous samples, *while solid colours* indicate position of museum samples. * indicate samples for which only country of origin is available (See full samples details in Additional file 1: Table S1)

that no intermediate purification step was necessary. Double-stranded sequencing libraries were also prepared for the Illumina MiSeq platform according to the protocol of Fortes and Paijmans [18] with single 8 nt indexing. For two samples from Peninsular Malaysia, purified gDNA was sheared to 500 bp using Covaris instrument (Woburn, MA) and subsequently prepared using NEB Next Ultra DNA library prep kit for Illumina (New England Biolabs, Ipswich, MA) with dual 8 nt indexing. The constructed libraries were sequenced on the MiSeq system located at the Monash University Malaysia Genomics Facility with the run configuration of 2 x 250 bp.

Because DNA extracted from museum samples is expected to be highly degraded and because external contamination from handling, storage and exposure may have introduced foreign DNA, we performed a hybridization capture step prior to sequencing in order to enrich the archival samples for their mitochondrial DNA (mtDNA). To capture target mtDNA, we created baits from a fresh sample of the closely related species *Muntiacus reevesi* (blood sample from Zoo Münster, Germany). The baits were generated by amplifying the whole mitochondrial DNA via long range PCR (primer sequences and PCR conditions described in Additional file 1: Table S3). Sheared and pooled long range PCR fragments were then prepared as bait [19]. Hybridization capture also followed the protocol described in [19]. For PGM sequencing the capture mixture was modified and included the blocking oligos for the PGM adapters.

Libraries from fresh samples and captured archival DNA libraries were pooled equimolarly, but never together in the same run, and sequenced with either the Ion PGM™ Sequencing 200 Kit v2 or the Illumina MiSeq

kit v3 (150-cycle), following the respective manufacturers' protocols.

Bioinformatic analyses

Sequence reads from Illumina were de-multiplexed according to the respective indexes with the Illumina software BCL2FASTQ v2.17 (Illumina, San Diego, CA, USA) and adapters were clipped from the sequence reads with the software CUTADAPT v1.3 [20]. Sequencing reads generated with the Ion Torrent PGM were first processed with the inbuilt software (de-multiplexing and adapter clipping). Quality trimming was done through a sliding window approach (10 bp; Q20) and quality filtering then proceeded by removing all reads shorter than 20 bp from the analyses. The complete mitochondrial genome sequence of *M. muntjak* (NCBI Accession Nr. NC_004563.1) was used as reference for mapping of the sequencing reads, which was done with the software BWA v.0.7.10 [21]. A following quality filtering step was performed to remove duplicate reads from the dataset, so that only unique reads were kept, with the software MARKDUPLICATES from PICARD-TOOLS v.1.106 (http://picard.sourceforge.net/). Indels were called with BCFTOOLS v.1.2 (http://github.com/samtools/bcftools) and variant calling was carried out in GATK v.1.6 [22], with N-masking of positions with less than 3 unique reads and ambiguous heterozygous positions. Variants were called based on the majority rule. Only sequences with 85% or more of the genome covered with 3x or greater depth were accepted for analysis (range of depth was between 6.87 and 294.49x, with a mean of 72.35x, see sequencing results in Additional file 1: Table S4) and missing data was included for analyses. Finally, potential presence of *numts* was investigated by searching for the presence of stop codons in coding genes and indels; and we found no evidence of *numts* in our dataset. All mitogenomes obtained were deposited in Genbank (KY052082-KY052156) and accession numbers for each sample are given in Additional file 1: Table S1.

Genetic diversity, divergence dating and population demography

All *Muntiacus* sequences obtained in this study were aligned using MAFFT v.7.245 [23] with the auto setting..We constructed a median joining (MJ) network with NETWORK v.4.6.1.4 [24], with missing/gaps sites and invariant sites removed from the dataset. Haplotype diversity and nucleotide diversity were calculated in DNASP v.5.10 [25]. Analysis of molecular variance (AMOVA) and population differentiation (F_{st}) were calculated with Arlequin v.3.5.1.2 [26]. Samples were grouped according to geographical origin into populations and populations belonging to different clades were considered groups, so that we had three different groups: Mainland, Sunda and Sri Lanka.

To estimate divergence times of different muntjac clades and population demographic changes over time, we inferred genealogies using a relaxed lognormal clock model as implemented in BEAST v.1.8.1 [27]. We performed this analysis by creating a dataset with ten other Cervidae species (at least one representative species from each tribe from the three subfamilies of Cervidae) and one Bovidae species (*Bos javanicus*) in order to estimate the divergence time and mutation rate for the muntjac clade. The root age (the most recent common ancestor [TMRCA] of Bovids and Cervids) was set to 16.6 million years (My) [28], with a normal prior distribution and standard deviation of 2. This date represents the minimum age of fossil evidence between Cervidae and Bovidae [29]. Another study reported similar divergence times [27] and dated the split between Reeve's muntjacs (*M. reevesi*) and red muntjac + black muntjac (*M. crinifrons*) clade at 3 Mya [30]. We subsequently used the estimated red muntjac split as tree prior for the calibrations within our dataset.

Using the complete mtDNA we employed RAXML GUI v.1.5 [31] to reconstruct a maximum likelihood (ML) phylogenetic tree of red muntjacs applying the substitution model determined by JMODELTEST v.2.1.7 (GTR + G + I, [32]). Both the ML tree and the MJ network revealed the presence of three monophyletic clades comprising individuals from three geographically distinct regions. Based on these results we created three different data sets and analyzed them independently with BEAST v.1.8: 1) full data set comprising all individuals from the three identified clades; 2) 'Mainland' clade only; and 3) 'Sunda' clade only. As analyses 2 and 3 describe distinct population groups, they are more suited for demographic analyses that assume a coalescent process. We did not generate a data set comprising only Sri Lankan samples (which constituted the third clade) due to the few individuals sampled. For analyses of the full dataset (dataset 1) we inferred the phylogeny with BEAST v1.8 using a lognormal clock model and a Yule speciation tree model (assuming a constant lineage birth rate for each branch), with GTR + G + I as substitution model determined by JMODELTEST (based on AIC). We set the root height to 1.5 Mya as inferred by the Cervidae/Bovidae phylogenetic tree and estimated the clock rate. For the demographic analysis (datasets 2 and 3) we assumed the same substitution model and the coalescent prior of Extended Bayesian Skyline as tree prior as implemented in BEAST v1.8. To test the hypothesis of a recent expansion of red muntjac populations we set the root height at 1.12 Mya, as inferred from the divergence dating phylogenetic tree. For

every analysis we performed two independent runs with 50 million iterations each, sampling one tree per 5000 iterations. Results of each run were visualized with TRACER v.1.6 (implemented in Beast v1.8). The first 10% per run (1000 trees) was discarded as burn-in and the remaining trees were combined with LOGCOMBINER v.1.8.1 (implemented in Beast v.1.8). Maximum credibility trees were obtained with TREEANNOTATOR, also distributed as part of the BEAST package. Skyline plots were generated using the R package GGPLOT2 [33].

As we did not obtain any samples from the Western Ghats, in Southern India, we included five cytochrome *b* (cyt*b*) sequences from NCBI (Accession numbers EU727189; FJ190162; FJ190160; JN861030; JN861033) to test their phylogenetic placement. We aligned these sequences with cyt*b* sequences extracted from the complete mitogenome dataset using MAFFT v.7.245 with specifications as before. RAxML GUI v.1.5 was used to construct the ML tree. We visualized and edited all trees with FIGTREE v.1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

Results

After quality filtering and mapping of all reads, we were able to obtain a final dataset of 16386 bp of the complete mitogenome for a total of 59 archival and 16 contemporary samples, which constituted a total of 65 different haplotypes. Number of variable and parsimony informative sites was high for all coding genes (Additonal file 1: Table S5).

Phylogeography and population genetic diversity

All three analytical approaches, MJ network (Fig. 2), Bayesian inference (Fig. 3) and maximum likelihood analysis (Additional file 1: Figure S1) revealed three well supported mtDNA clades Sequences from two major clades originated from (i) mainland South and Southeast Asia and China (henceforth referred to as Mainland) (ii) Peninsular Malaysia and the Sunda Islands (Sunda); while the third clade consisted exclusively of sequences originating from Sri Lanka (Sri Lanka). Haplotype and nucleotide diversity for each clade was similar (Table 1). F_{ST} values indicated substantial genetic structuring among the three clades: the highest differentiation was

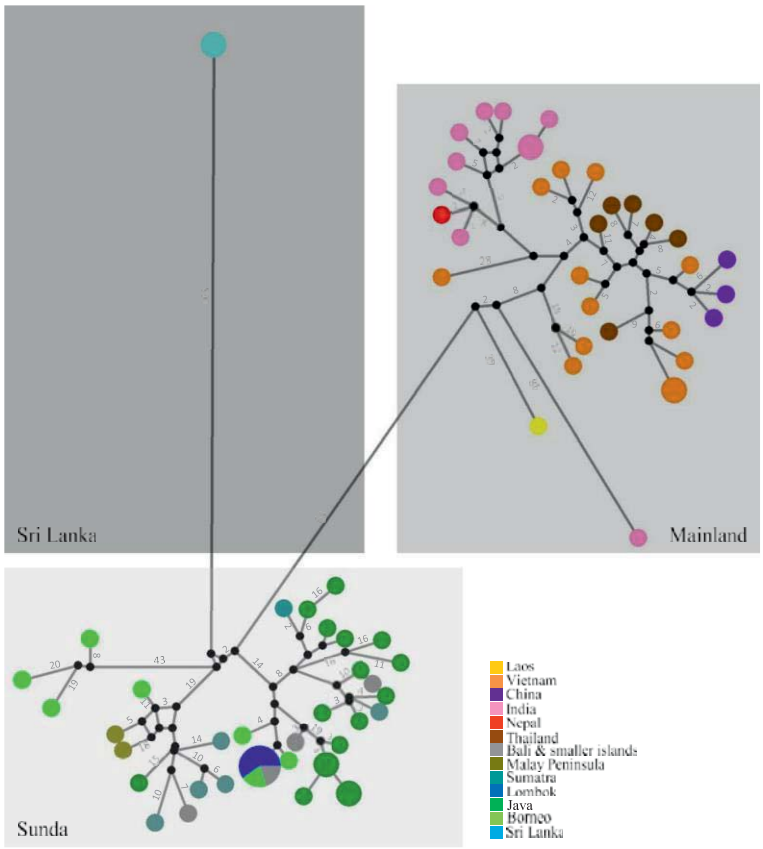

Fig. 2 Median joining network of full mitogenome of all archival and contemporaneous red muntjac samples. *Circle size* is proportional to haplotype frequencies; *fill color* denotes geographical origin; *lines* represent one mutational step, except when indicated otherwise with numbers. *Black circles* represent missing vectors. The three major clades are denoted in the different boxes and indicated by name

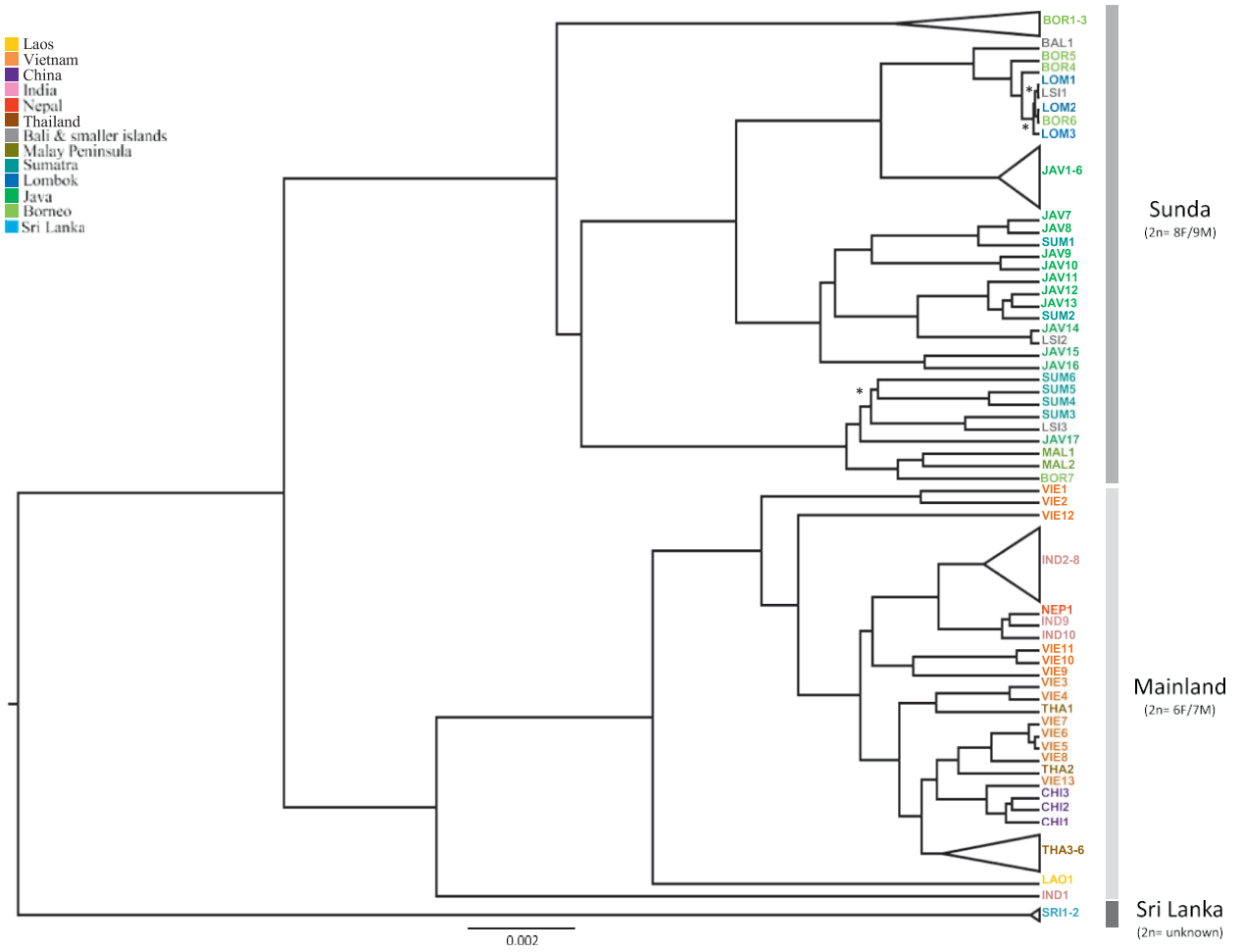

Fig. 3 Maximum credibility tree based on 59 archival and 16 contemporaneous *red muntjac* samples, spanning across *red muntjacs* combined distribution. All branches are supported with BPP>0.9 except the ones marked with * where BPP>0.8. This phylogenetic relationship was reconstructed with BEAST and the root age was constrained to 3 Mya as suggested by the divergence dating analyses. Clades of similar sequences have been collapsed and the detailed information about all individuals can be found in Table S1

found between Mainland and Sri Lanka (0.906, p <0.05), the second highest differentiation was between Sunda and Sri Lanka (0.848, p <0.001) and the lowest, yet still very high, between Mainland and Sunda (0.631, p <0.001).

Table 1 Summary of number of samples and haplotypes distributed in each clade and measurements of diversity indexes of the three major red muntjac clades. Haplotypic diversity (h) and nucleotide diversity (π) could not be measured for Sri Lanka clade due to the low sample number

	Mainland	Sunda	Sri Lanka
N	33	40	2
Haplotypes	31	33	1
Haplotype diversity (h)	0.996	0.982	—
Nucleotide diversity (π)	0.006	0.009	—

The Mainland clade comprised all samples from mainland South and Southeast Asia and China, excluding samples from Sri Lanka and Peninsular Malaysia. In general, samples with the same geographic origin formed cohesive branches, except individuals from Vietnam and Thailand which were placed sporadically together or clustered with the Chinese samples. Interestingly, one single individual sampled from Himachal Pradesh province in India (IND1) and one individual from Laos (LAO1) formed two basal branches in this clade (Fig. 3). Conversely, the Sunda clade comprised all red muntjacs from the Sunda Shelf, including individuals from Peninsular Malaysia, which were closely related to an individual from Borneo with no further indication of phylogeographic sub-structuring of samples from different land-masses in the Sunda Shelf. AMOVA results (Table 2) corroborate these findings, showing that variation is highest among the three groups tested (60.5%,

Table 2 AMOVA results among groups (Mainland, Sunda, Sri Lanka), among populations within groups (populations were defined based on country of origin) and within populations. Results show the majority of variation explained among groups, indicating differentiation between the three major clades

	d. f.	Variance components	Percentage of variation[a]
Among groups	2	94.1	60.5
Among population within groups	8	16.5	10.6
Within populations	64	45	28.9

[a]all *p*-values <0.001

p-value <0.001) and lowest among populations within groups (10.6%, p-value <0.001). The ML inference based on the cyt*b* gene only (Additional file 1: Figure S2), which included Western Ghats red muntjac samples, revealed a sister relationship between Western Ghats and Sri Lankan red muntjacs.

Divergence times and demographic changes

Based on the calibrated root age (TMRCA of Bovids and Cervids at 16.6 ± 2 Mya) our results suggested that the split between the red muntjacs and other *Muntiacus* species occurred in the late Pliocene around 3 Myr (95% High posterior density [HPD] = 1.7 – 5.7) (Fig. 4). Within red muntjacs, the split between the Sri Lankan matriline and the other red muntjacs was estimated to have occurred at around 1.5 Mya (95% HPD = 0.78 – 2.82), while the divergence time between the two other major clades was inferred to be around 1.12 Mya (95% HPD = 0.5 – 2.22). Divergence dates obtained were associated with relatively small 95% posterior density intervals.

The inferences of effective population sizes showed different demographic histories for each of the two

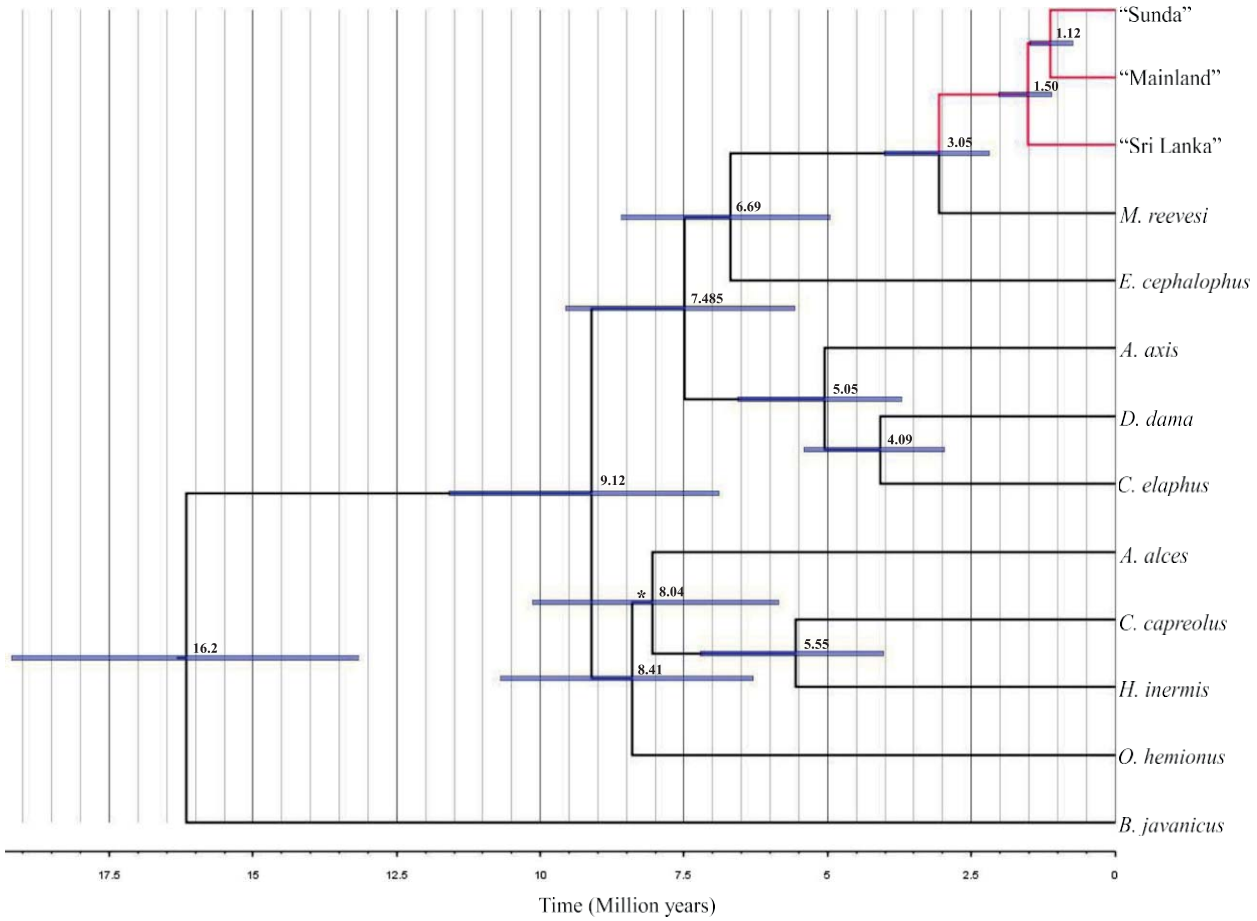

Fig. 4 Divergence dating maximum credibility tree obtain with BEAST. Root age was calibrated to 16.6 Myr and indicates split age between Bovidae and Cervidae (16.2 Mya) and between 10 different other Cervidae species. Branches highlighted in *red* indicate the clades analysed in this study

clades analysed (Fig. 5). Although both experienced a population expansion in the late Pleistocene (Mainland clade about 0.2 Mya, Sunda clade about 0.27 Mya), the increase in the mainland was maintained while the effective population size of Sundaic red muntjacs started to steeply decline about 25 thousand years ago (Fig. 5).

Discussion

Phylogeographic patterns and population structure

Investigating variation in complete mitochondrial genomes across the red muntjac distribution range enabled us to examine the evolutionary history of these populations, including phylogeographic structure and historical demography. The high genetic diversity within, and large divergence among the main clades indicate a long-term separation of these clades, which in turn implies that red muntjacs had spread across South and Southeast Asia shortly after their split from other muntjac species. We found, however, evidence for two geographic barriers that (i) separated populations from northern India and Sri Lanka + Western Ghats and (ii) separated Sundaic and Mainland populations.

The Sri Lanka clade split first from the other two populations around 1.5 Mya and our data provides the first molecular evidence for its distinction. Due to the low number of samples obtained from this region we were unable to further assess their evolutionary history and genetic diversity. Other authors, who regarded the population of Sri Lanka as a separate taxonomic unit (either as a full species *M. malabaricus* – [13] or as a subspecies *M. m. malabaricus* – [10]) have grouped the populations from the Indian Western Ghats together with the Sri Lankan red muntjacs. The results from the ML tree based on the cyt*b* gene support this inclusion and provide another example of the distinct populations in the 'Western Ghats - Sri Lanka biodiversity hotspot' [34]. Examples of a discontinuity in species distribution between Northern and Southern India are numerous and this disjunct distribution seems particularly pronounced in wet-zone species [35]. The dry zone of central India (with rainfall of just 50 – 100 cm per year) seems to be an unsuitable habitat for many species and has often led to speciation between populations isolated in more suitable forested habitats (refugia) in the northern and southern regions of the Indian subcontinent (e.g., the Asian fairy-bluebird *Irena puella* [36]; Flying lizards *Draco* genus [37]; Nilgiri tahr *Nilgiri tragushylocrius* [38]). Red muntjacs, however, are much more versatile and currently occur to the authors' knowledge generally throughout the 'dry zone'. This implies that the barrier to gene flow has existed in the past, presumably as a result of extreme dry climatic conditions caused by global ice advances. For red muntjacs, the isolation of populations in the southern wet-zones (Sri Lanka and Western Ghats) might have persisted even after the recolonization of the dry zone of central India, if in fact other barriers to gene flow, such as karyotype differences, had arisen in the interim, although at present no such evidence exists (see below).

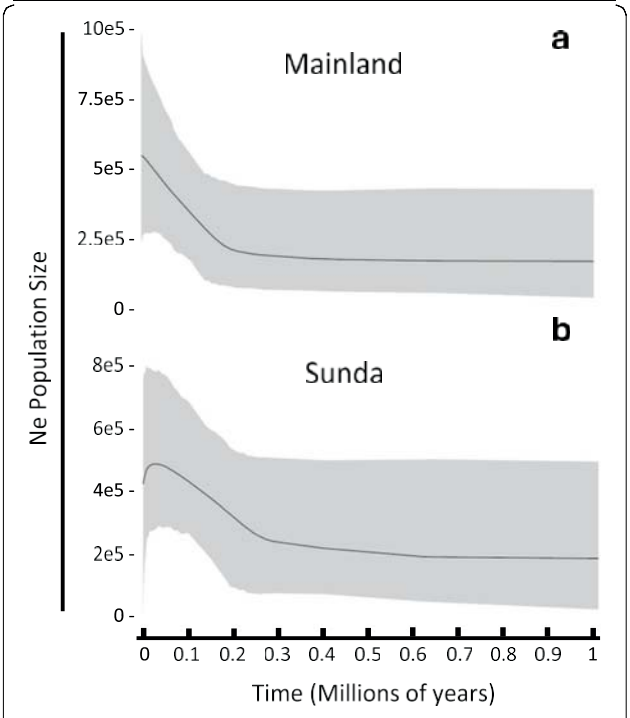

Fig. 5 Extended Bayesian Skyline Plots of female effective population size changes of two major clades (a) Mainland and (b) Sunda within *red muntjacs* through time. *Black line* represents the mean number of Ne changes through time and *grey area* denotes the 95% High Probability Density

The second major clade was composed of all Indian and Indochinese samples. The phylogenetic mtDNA tree showed geographical structuring and the basal placement of two individuals that came from Laos and from Himachal Pradesh Province in India. Their position (and the relatively long branch distance in the network analysis) suggests ancient isolation of localized populations during the Pleistocene. It is likely that populations in Indochina and India were repeatedly affected by climatic fluctuations of the Pleistocene, resulting in some areas of their current continuous range having become unsuitable (most probably too dry and supporting only very open habitat). Currently and presumably also during previous interglacials, when forest conditions predominated, red muntjacs would have expanded their range and reoccupied former distribution areas. An example of such phylogeographic pattern of a "colonization from the east" was observed in leopard cats (*Prionailurus bengalensis*), where populations became isolated in their refugia due to the drier and unsuitable habitat in the rest of the Indian subcontinent [39]. This effective population size expansion is reflected in the

Bayesian skyline plot and is likely responsible for the patterns of admixture we observed in our ML and BI trees, where samples from Vietnam, Thailand, China and India are closely related in the most derived branches of the trees. We estimated that this expansion started around 200 kya, a time marked by the beginning of an interglacial period (240 – 190 kya). This period succeeded a glacial time, with sea levels as low as −130 m below present [40] that would have provoke drier climate in continental areas.

The surprising placement of the Himachal Pradesh Province (IND1) sample at a basal position allows us to speculate as well that there may exist a distinct 'high elevation Himalayan red muntjac' (also supported by RJT's unpublished morphological examinations), which may have evolved in refugial areas during these dry periods. Unfortunately, we could not include samples from Hainan Island, from high elevation Himalayan populations, from the Cardamom Mountains of Thailand and Cambodia, from the Southern Annamites of Vietnam, from Northern Myanmar, or from the Indian dry zones in our analyses (either samples had no proper location or yielded insufficient data); samples that from a phylogeographic perspective might provide further resolution or additional lineages. These regions correspond in part to the distribution of two currently described muntjac taxa: the sub-/species *M. (m.) nigripes*, described from Hainan Island and considered by some authors to occur also in northernmost Vietnam and Yunnan (China) and the sub-/species *M. (m.) aureus*, described to occur in northwestern India and considered to also occur in central India and disjunctly in Myanmar [13]. Thus, further genetic substructuring may exist within the Mainland clade, which could be unveiled with more intensive and extensive sampling.

The third clade, the Sunda clade, included all samples from Sundaland (including the lower Malay Peninsula). The very clear distinction of this clade indicated the existence of a long lasting migration barrier preventing gene flow between populations possibly north and south of the Isthmus of Kra, a recognized phyto- and zoogeographic boundary located on the Malay/Thai Peninsula around 10°30' N. Although the Isthmus of Kra separates the Indochinese and Sundaic subregions [41], studies on different taxa revealed that the Isthmus of Kra is not a clear geophysical barrier, but rather a transition zone ranging from 5° N to 13° N (e.g. bats and birds [42]; butterflies [43]; amphibians [44]; mammals [45]). As we had only two samples from Peninsular Malaysia, we could not address the exact latitude of this separation (or indeed whether there is in fact introgression as suggested by morphological samples – RJT unpublished data). Increased sampling efforts would therefore be required to identify the true nature and geographic location of the 'boundary' between northern and southern red muntjac populations. Nuclear DNA analyses are also needed to address questions of a presumed secondary contact zone and potential hybridization of northern and southern red muntjacs on the Malay Peninsula. Interestingly, a recent multi-species study on mammals showed that no clear physical barrier is needed to maintain the separation of the Sundaic and Indochinese faunas, but that instead a combination of different climatic conditions during the Pleistocene and species-specific life history traits are sufficient to result in the observed pattern (Radchuk, unpublished data). With respect to the red muntjacs, these findings could indicate that northern and southern red muntjacs evolved different ecological adaptations during the periods their ranges became restricted, which is indicated by the subsequent population expansions in the EBSPs of both clades.

Within the Sunda clade we found evidence of the effects of Toba super volcanic eruption in Northern Sumatra around 74 thousand years ago. The Toba super eruption is described as one of the greatest eruptions in the last two million years [46]. It created an ash cloud that would have covered the northern part of Sumatra and south of Peninsular Malaysia leading to changes in vegetation and possible local extinctions of mammal species [47–49]. Our results support such potential extinctions of red muntjacs in Peninsular Malaysia and Sumatra, as the analysed individuals all derived from Bornean or Java populations, which occupy all basal positions of the internal nodes. This pattern suggests that Sumatra and Peninsular Malaysia were colonized, more than once, from Bornean and Javan populations, potentially after the Toba eruption. Such a re-colonisation of the southern Malay Peninsula would have been facilitated by the existing land bridges throughout the Late Pleistocene between the larger Sunda Islands. These land bridges also explain the lack of geographical substructuring as seen both in the MJ network and in the gene trees within the Sunda clade, as the exposed shelf allowed gene flow among populations on all larger landmasses. Being habitat generalists red muntjacs could have easily colonized new habitats on the exposed shelf. After the Last Glacial Maximum, rising sea levels not only separated the larger landmasses but also drastically reduced the available land area in the Sunda Shelf. This reduction in land and thus habitat availability coincides with the observed decrease in population size observed in the skyline inferences. However inferred divergence times should still be considered rough estimates, because they depend on estimated mutation rates.

Taxonomic implications

Currently, up to six species of red muntjacs have been described. The most commonly accepted split within the

red muntjacs is the one separating the mainland and Sundaic forms in two species: Northern red muntjac *Muntiacus vaginalis* and Southern red muntjac *M. muntjak*. Because species delineation was based only on the karyotype of one individual from Peninsular Malaysia and on a few morphological traits, this classification is still applied with some reservations. The proposed additional division of mainland (Northern) red muntjacs into different species [13, 50] is likewise weak and seldom adopted. In that study [13], differences were mainly described based on morphological traits, but the described variations are more likely trait polymorphisms within a largely distributed species, rather than distinctive morphological characters [7]. The molecular data presented here does not support the delineation of six red muntjac species as we did not find six distinct matrilines. Instead, we found evidence for a deep split between Sundaic (southern) and mainland (northern) red muntjacs, which would be concordant with karyotypic evidence: the few examined individuals of Sundaic forms had 2n = 8 F/9 M, while the most frequently studied Mainland forms had 2n = 6 F/7 M [13]. The evidence presented here clearly shows the existence of a third distinct Sri Lankan + Western Ghats clade, for which the karyotype characterization is still lacking. Our molecular data supports likewise the recognition of the Sri Lankan population as a distinct taxonomic unit, as this lineage split first from all other red muntjacs at around 1.5 Mya. So far the Sri Lankan and Western Ghats populations have been recognized as *M. (m.) malabaricus* due to their smaller body size (compared to more northerly and eastern mainland red muntjacs) and due to some pelage coloration differences [13].

The three clades introduced here are distinguished by deep molecular splits and appear to be geographically separated. We believe that the results of this study provide a good basis for a future taxonomic reassessment. We refrain from assigning these clades species or subspecies rank because: a) the estimated divergence dates and the observed pairwise differences are in the range of both recognized species [5] and recognized subspecies [51] and thus are not decisive, b) we only analysed mtDNA and thus have no information on potential incomplete lineage sorting, and c) our analyses lacked samples from regions where the three clades must meet and potentially overlap (within central and southern India and probably within the transition zone of the Isthmus of Kra). Therefore, further sampling and analysis of nuclear DNA data, and potentially also of morphological and karyotype data is needed, since both characteristics are further indicators of barriers to gene flow.

Independent of their specific taxonomic assignment as species or subspecies, our data clearly advocate the distinct management of the populations in these clades. Despite the indiscriminate hunting pressure on larger mammals in Southeast Asia, particularly in Indochina [52], both Northern and Southern red muntjacs are still widespread and less threatened than other species in this region. Our data did not uncover any populations of greater conservation concern among Mainland or Sundaic red muntjacs, although more extensive sampling might reveal the existence of taxonomic units of conservation concern. However, our data do clearly reveal the distinctiveness of the Sri Lankan + Western Ghats populations. The spatial restriction of this clade and ongoing threats both in Sri Lanka and also the Western Ghats [53] highlight the conservation significance of red muntjacs in this region. Additional field surveys, including further molecular sampling, are important to assess and monitor the conservation status of *M.* (*m.*) *malabaricus*.

Conclusions

We found substantial genetic differentiation in a widespread species, corresponding to at least two biogeographical barriers located in the major biodiversity hotspot of South and Southeast Asia: the Isthmus of Kra in northern Peninsula Malaysia and the central Indian dry zone. However, within each of the three lineages we found evidence of mitochondrial admixture between populations which are now geographically isolated, suggesting that red muntjacs, as a generalist species, could utilize land corridors exposed during the low sea level periods of the Pleistocene. Our results, finally, indicate for the first time molecular evidence for the distinctiveness of the Sri Lanka and Western Ghats red muntjac populations.

Additional file

Additional file 1: Table S1. Comparison of number of accepted red muntjac species and subspecies among three different recent publications. Number of recognized species range from 1 to 6 and number of recognized subspecies range from none to 10. Table S2. Complete dataset used for analyses, with information on origin, location and contact. Table S3. Long-range PCR primer sequences and annealing temperatures used for bait development. Table S4. Sequencing results from all samples used in this study, with indication of sequencing platform, percentage of reads on target, average base coverage depth and percentage of reference genome covered with at least 3x coverage. Table S5. Number of Variable sites (V) and Parsimony informative sites (Pi) per coding gene, throughout the full mitogenomes obtained. Figure S1. Best tree obtained with RAxML through Maximum Likelihood Analysis. Figure S2. Maximum Likelihood tree of the Cytochrome B gene, including five individuals from Western Ghats, supporting the clade differentiation of Sri Lanka – Western Ghats from all other red muntjac populations. (PDF 519 kb)

Abbreviations

Asl: Above sea level; BPP: Bayesian posterior probability; gDNA: Genomic DNA; mtDNA: Mitochondrial DNA; Mya: Million years ago

Acknowledgements

The authors would like to heartily thank all people listed in Additional file 1: Table S2 for allowing us access to the museum collections and all the help

provided during the project and additionally to Dr. Clara Stefen from Museum of Zoology, Senckenberg Naturhistorische Sammlungen, Dresden and Dr. Geraldine Veron from Museum National d'Histoire Naturelle, Paris for also allowing sampling within the museum collections. RM would like to thank Ramona Taubert for the help with the laboratory work.

Funding
This project was funded by Leibniz grant SAW-2013-IZW-2. ML was supported by National Foundation for Science and Technology Development of Vietnam (NAFOSTED: Grant Nos. 106.152010.30) and the Partnerships for Enhanced Engagement in Research (PEER) (USAID-PEER-3-149).

Availability of data and material
All DNA sequences will be deposited in the Genbank Repository and Accession Numbers will be available in Additional file 1: Table S1. Alignment and Bayesian tree were deposited in TREEbase: http://purl.org/phylo/treebase/phylows/study/TB2:S20435.

Authors' contributions
RFM, JF, DWF and AW designed the study; RFM, TvN, HMG and JRRJ performed the experiments; RFM and DL analysed the data; ML, HMN, RJT and HMG provided samples and information about the ecology and distribution of the red muntjacs; RFM and AW wrote the first draft of the paper. All authors contributed to writing the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable.

Ethics approval and consent to participate
Not applicable.

Author details
[1]Leibniz Institute for Zoo and Wildlife Research (IZW), Alfred-Kowalke-Straße 17, 10315 Berlin, Germany. [2]Potsdam University, Institute for Biochemistry and Biology, Karl-Liebknecht-Str. 22-24, 14476 Potsdam-Golm, Germany. [3]Faculty of Environmental Science, Hanoi University of Science, Vietnam National University, 334 Nguyen Trai Road, Hanoi, Vietnam. [4]Centre for Natural Resources and Environmental Studies, Vietnam National University, 19 Le Thanh Tong Street, Hanoi, Vietnam. [5]1123 Monroe Street, Evanston, IL 60202, USA. [6]School of Science, Monash University Malaysia, 47500 Bandar Sunway, Malaysia. [7]Genomics Facility, Tropical Medicine and Biology Multidisciplinary Platform, Monash University Malaysia, 47500 Bandar Sunway, Malaysia. [8]Department of Wildlife and National Parks (DWNP) Peninsular Malaysia, National Wildlife Forensic Laboratory (NWFL), 56100 Kuala Lumpur, Malaysia. [9]Present address: U.S. Agency for International Development, Governance for Inclusive Growth Program, Chemonics International Inc, 115 Tran Hung Dao Street, Hanoi, Vietnam.

Received: 17 October 2016 Accepted: 17 January 2017
Published online: 26 January 2017

References

1. Ma SL, Wang YXSL. A new species of the genus muntiacus from Yunnan, China. Zool Res. 1990;11:47–52. Available from: http://en.cnki.com.cn/Article_en/CJFDTotal-DWXY199001009.htm.
2. Amato G, Egan MG, Rabinowitz a. A new species of mutjac, muntiacus putaoensis (artiodactila: cervidae). Anim Conserv. 1999;2:1–7.
3. Groves CP, Grubb P. The species of muntjac (genus muntiacus) in Borneo: unrecognised sympatry in tropical deer. Zool Meded. 1982;56:203–16.
4. Yang F, O'Brien PC, Wienberg J, Neitzel H, Lin CC, Ferguson-Smith M a. Chromosomal evolution of the Chinese muntjac (muntiacus reevesi). Chromosoma. 1997;106:37–43. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9169585.
5. Wang W, Lan H. Rapid and parallel chromosomal number reductions in muntjac deer inferred from mitochondrial DNA phylogeny. Mol Biol Evol. 2000;17:1326–33.
6. Nagarkoti A, Thapa T. Distribution pattern and habitat preference of barking deer (Muntiacus muntjac Zimmermann) in Nagarjun forest, Kathmandu. Himal J Sci. [Internet]. 2007;4:70–4. Available from: http://www.nepjol.info/nepal/index.php/HJS/article/viewFile/985/1021\n http://www.nepjol.info/index.php/HJS/article/viewArticle/985.
7. Timmins RJ, Steinmetz R, Samba Kumar N, Anwarul Islam Md SBH. Muntiacus vaginalis [Internet]. IUCN Red List Threat. Species. 2016 [cited 2016 Nov 28]. Available from: http://www.iucnredlist.org/details/136551/0.
8. Laidlaw RK. Effects of habitat disturbance and protected areas on mammals of peninsular Malaysia. Conserv Biol. 2000;14:1639–48. Available from: <Go to ISI>://000166035800012.
9. Francis C. Cervidae. In: Francis C, editor. A F. Guid. to Mamm. South-East Asia [Internet]. New Holland Publishers, Ltd.; 2008. p. 318–23. Available from: https://books.google.de/books?hl=en&lr=&id=zWeS8A6nunIC&oi=fnd&pg=PA8&ots=ekXxoMiB4w&sig=xpBuNyg1_gUrDfl9Cft0_eFnx9A#v=onepage&q&f=false.
10. Mattioli S. Family Cervidae (Deer). In: Wilson D, Mittermeier R, editors. Handb. Mamm. World - Vol. 2 Hoofed Mamm. [Internet]. Barcelona: Lynx Edicions; 2011. p. 350–443. Available from: http://www.lynxeds.com/hmw/handbook-mammals-world-volume-2.
11. Groves CP, Grubb P. Muntiacidae. In: Bubenik G, Bubenik A, editors. Horns, Pronghorns, Antlers - Evol. Morphol. Physiol. Soc. Significance [Internet]. New York: Springer; 1990. Available from: http://www.springer.com/de/book/9781461389682.
12. Groves C. Taxonomy of ungulates of the Indian subcontinent. Bombay Nat Hist Soc. 2003;100:341–61. Available from: http://hdl.handle.net/1885/76187.
13. Groves CP, Grubb P. Ungulate Taxonomy. Baltimore: The Jonhs Hopkins University Press; 2011.
14. Voris HK. Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. J Biogeogr. 2000;27:1153–67.
15. Gorog AJ, Sinaga MH, Engstrom MD. Vicariance or dispersal? historical biogeography of three sunda shelf marine rodents (maxomys surifer, leopoldamys sabanus and maxomys whiteheadi). Biol J Linn Soc. 2004;81:91–109.
16. Wilting A, Christiansen P, Kitchener AC, Kemp YJM, Ambu L, Fickel J. Geographical variation in and evolutionary history of the sunda clouded leopard (neofelis diardi) (mammalia: carnivora: felidae) with the description of a new subspecies from Borneo. Mol Phylogenet Evol. 2011;58:317–28. Available from: http://linkinghub.elsevier.com/retrieve/pii/S1055790310004306.
17. Sokolov EP. An improved method for DNA isolation from mucopolysaccharide-rich molluscan tissues. 2000. p. 573–5.
18. Fortes GG, Paijmans JLA. Analysis of whole mitogenomes from ancient samples. Whole Genome Amplif. Methods Protoc. 2015;179–95.
19. Maricic T, Whitten M, Pääbo S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. PLoS One. 2010;5:e14004. Available from: http://dx.plos.org/10.1371/journal.pone.0014004.
20. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal [Internet]. 2011;17:10. Available from: http://journal.embnet.org/index.php/embnetjournal/article/view/200/479.
21. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics. 2009;25:1754–60.
22. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303. Available from: http://genome.cshlp.org/cgi/doi/10.1101/gr.107524.110.
23. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
24. Bandelt HJ, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 1999;16:37–48.
25. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25:1451–2.
26. Excoffier L, Laval G, Schneider S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol. Bioinform. Online [Internet]. 2005;1:47–50. Available from: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=19325852&retmode=ref&cmd=prlinks\npapers2://publication/uuid/218B5AA9-3EAB-4B3D-87A2-4B144D33857B\n http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2658868&tool=pmcentrez&re.
27. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol. 2012;29:1969–73.
28. Bibi F. A multi-calibrated mitochondrial phylogeny of extant bovidae (artiodactyla, ruminantia) and the importance of the fossil record to systematics. BMC Evol Biol. 2013;13:166. Available from: http://www.biomedcentral.com/1471-2148/13/166.

29. Gentry A, Rössner G, Heizmann E. Suborder Ruminantia. In: Rössner G, Heissig K, editors. Miocene L. Mamm. Eur. [Internet]. Munich: Verlag Friedrich Pfeil. 1999; p. 225–58. Available from: http://www.pfeil-verlag.de/07pala/d2_50.html.
30. Pitra C, Fickel J, Meijaard E, Groves C. Evolution and phylogeny of old world deer. Mol Phylogenet Evol. 2004;33:880–95. Available from: http://linkinghub.elsevier.com/retrieve/pii/S1055790304002374.
31. Silvestro D, Michalak I. RaxmlGUI: a graphical front-end for RAxML. Org Divers Evol. 2012;12:335–7.
32. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2 : more models, new heuristics and high- performance computing. Nat Methods. 2015;9:6–9.
33. Wickham G. ggplot2: elegant graphics for data analysis. Baltimore: Springer New York; 2009.
34. Biju SD, Garg S, Mahony S, Wijayathilaka N, Senevirathne G, Meegaskumbura M. DNA barcoding, phylogeny and systematics of golden-backed frogs (hylarana, ranidae) of the western Ghats-Sri Lanka biodiversity hotspot, with the description of seven new species. Contrib to Zool. 2014;83:269–335.
35. Karanth KP. Evolution of disjunct distributions among wet-zone species of the Indian subcontinent: testing various hypotheses using a phylogenetic approach. Curr Sci. 2003;85:1276–83.
36. Zacharias VJ, Gaston AJ. The recent distribution of endemic, disjunct and globally uncommon birds in the forests of Kerala State, south-west India. Bird Conserv Int. 1999;9:191–225.
37. Honda M, Ota H, Kobayashi M, Nabhitabhata J, Yong H-S, Hikida T. Phylogenetic relationships of the flying lizards, genus Draco (reptilia, agamidae). Zoolog Sci. 1999;16:535–49. Available from: http://www.bioone.org/doi/abs/10.2108/zsj.16.535.
38. Ropiquet A, Hassanin A. Molecular evidence for the polyphyly of the genus hemitragus (mammalia, bovidae). Mol Phylogenet Evol. 2005;36:154–68.
39. Mukherjee S, Krishnan A, Tamma K, Home C, Navya R, Joseph S, et al. Ecology driving genetic variation: A comparative phylogeography of jungle cat (Felis chaus) and leopard cat (Prionailurus bengalensis) in India. PLoS One. 2010;5:e13724.
40. Rohling EJ, Fenton M, Jorissen FJ, Bertrand P, Ganssen G, Caulet JP. Magnitudes of sea-level lowstands of the past 500,000 years. Nature. 1998;394:162–5. Available from: http://www.nature.com/nature/journal/v394/n6689/full/394162a0.html.
41. Woodruff DS. Neogene marine transgressions, palaeogeography and biogeographic transitions on the Thai-Malay Peninsula. J Biogeogr. 2003;30:551–67. Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.606.8853&rep=rep1&type=pdf.
42. Hughes AC, Satasook C, Bates PJJ, Bumrungsri S, Jones G. Explaining the causes of the zoogeographic transition around the isthmus of Kra: using bats as a case study. J Biogeogr. 2011;38:2362–72.
43. Corbet AS, Pendlebury HM. The butterflies of the Malay peninsula. Fourth ed. Eliot JN, editor. Chapter IV Geographical Distribution. Kuala Lumpur: Malay Nature Society; 1992.
44. Inger RF. Distribution of amphibians in Southern Asia and adjacent islands. In: Duellman WE, editor. Patterns Distrib. Amphib. a Glob. Perspect. Baltimore: Johns Hopkins University Press; 1999. p. 445–82.
45. Woodruff DS, Turner LM. The Indochinese-Sundaic zoogeographic transition: a description and analysis of terrestrial mammal species distributions. J Biogeogr. 2009;36:803–21. Available from: http://doi.wiley.com/10.1111/j.1365-2699.2008.02071.x.
46. Williams M a J, Ambrose SH, van der Kaars S, Ruehlemann C, Chattopadhyaya U, Pal J, et al. Environmental impact of the 73 ka Toba super-eruption in South Asia. Palaeogeogr Palaeoclimatol Palaeoecol. 2009;284:295–314. Available from: http://dx.doi.org/10.1016/j.palaeo.2009.10.009.
47. Ambrose SH. Late Pleistocene human population bottlenecks, volcanic winter, and differentiation of modern humans. J Hum Evol. 1998;34:623–51.
48. Nater A, Nietlisbach P, Arora N, Van Schaik CP, Van Noordwijk MA, Willems EP, et al. Sex-biased dispersal and volcanic activities shaped phylogeographic patterns of extant orangutans (genus: Pongo). Mol Biol Evol. 2011;28:2275–88.
49. Wilting A, Sollmann R, Meijaard E, Helgen KM, Fickel J. Mentawai's endemic, relictual fauna: is it evidence for Pleistocene extinctions on Sumatra? J Biogeogr. 2012;39:1608–20. Available from: http://doi.wiley.com/10.1111/j.1365-2699.2012.02717.x.
50. Giao P, Tuoc D, Dung V, Wikramanayake E, Amato G, Arctander P, et al. Description of Muntiacus truongsonensis, a new species of muntjac (Arctiodactyla: Muntiacidae) from Central Vietnam, and implications for conservation. Anim. Conserv. 1998;61–8.
51. Lorenzini R, Garofalo L. Insights into the evolutionary history of *Cervus* (Cervidae, tribe Cervini) based on Bayesian analysis of mitochondrial marker sequences, with first indications for a new species. J Zool Syst Evol Res. 2015;53:340–9. Available from: http://doi.wiley.com/10.1111/jzs.12104.
52. Duckworth JW, Batters G, Belant JL, Bennett EL, Brunner J, Burton J, et al. Why south-east Asia should be the world's priority for averting imminent species extinctions, and a call to join a developing cross-institutional programme to tackle this urgent issue. Sapiens. 2012;5:76–95.
53. Jha CS, Dutt CBS, Bawa KS. Deforestation and land use changes in Western Ghats, India. Curr Sci. 1995;79:231–43.

Phylogeography of red muntjacs reveals three distinct mitochondrial lineages

Renata F. Martins, Jörns Fickel, Minh Le, Thanh van Nguyen, Ha M. Nguyen, Robert Timmins, Han Ming Gan, Jeffrine J. Rovie-Ryan, Dorina Lenz, Daniel W. Förster, Andreas Wilting

Additional Material 1

Table S1. Comparison of number of accepted red muntjac species and subspecies among three different recent publications. Number of recognized species range from 1 to 6 and number of recognized subspecies range from none to 10.

	Mattioli (2011)[10]	Grubb and Groves (2011)[13]	Timmins *et al*. (2016)[7]
Species	*M. muntjak*	*M. muntjak* *M. vaginalis* *M. malabaricus* *M. aureus* *M. nigripes* *'M. guongdongensis'**	*M. muntjak* *M. vaginalis*
Subspecies	*M. m. muntjak* *M. m. annamensis* *M. m. aureus* *M. m. curvostylis* *M. m. malabaricus* *M. m. mengalis* *M. m. montanus* *M. m. nigripes* *M. m. vaginalis* *M. m. yunnanensis*	*M. v. vaginalis* *M. v. curvostylis*	not applicable

* Identified as *Incertae sedis*, but referring to the different form described by Li & Xu, 1996

Table S2. Complete dataset used for analyses, with information on origin, location and contact.

Collection ID	Sample name	Name in Figures	Genbank	Source material	Geographic origin	Year	Contact
46238	MMU1	BAL 1	KY052120	nasal bone	West Bali		Dr. Frieder Mayer, Naturkundemuseum Berlin
46239	MMU2	JAV1 3	KY052121	nasal bone	West Java		Dr. Frieder Mayer, Naturkundemuseum Berlin
17534	MMU3	THA 2	KY052091	nasal bone	Thailand		Dr. Frieder Mayer, Naturkundemuseum Berlin
91131	MMU6	IND5	KY052092	nasal bone and tissue from skull	North India		Dr. Frieder Mayer, Naturkundemuseum Berlin
91143	MMU8	IND2	KY052093	nasal bone	North India		Dr. Frieder Mayer, Naturkundemuseum Berlin
40556	MMU10	CHI3	KY052094	nasal bone	Yunnan China		Dr. Frieder Mayer, Naturkundemuseum Berlin
4837	MMU12	IND1 0	KY052095	nasal bone	East India		Dr. Frieder Mayer, Naturkundemuseum Berlin
34074	MMU16	BOR 2	KY052127	nasal bone	Southeast Borneo		Dr. Frieder Mayer, Naturkundemuseum Berlin
92292	MMU17	LOM 2	KY052122	bone from skull	Lombok		Dr. Frieder Mayer, Naturkundemuseum Berlin
70927	MMU18	SRI1	KY052116	nasal bone	East Sri Lanka		Dr. Frieder Mayer, Naturkundemuseum Berlin
33459	MMU20	CHI1	KY052096	nasal bone	Yunnan China		Dr. Frieder Mayer, Naturkundemuseum Berlin
92296	MMU21	LSI1	KY052123	tissue from skull	Lesser Sunda Islands		Dr. Frieder Mayer, Naturkundemuseum Berlin
70928	MMU22	BOR 7	KY052124	nasal bone	Borneo		Dr. Frieder Mayer, Naturkundemuseum Berlin
ZMA 10490	MMU24	SUM 6	KY052125	nasal bone	Sumatra	1950	Pepijn Kamminga, Naturalis Leiden
ZMA 10491	MMU25	SUM 1	KY052126	nasal bone	Sumatra	1951	Pepijn Kamminga, Naturalis Leiden
16699	MMU27	JAV1 7	KY052128	nasal bone and tissue from skull	Java	1856	Dr. Stefan Merker, Staatliches Museum für Naturkunde Stuttgart
16701	MMU28	JAV1 5	KY052129	nasal bone	Java	1878	Dr. Stefan Merker, Staatliches Museum für Naturkunde Stuttgart
50998	MMU29	JAV5	KY052130	nasal bone	East Java	1950	Dr. Stefan Merker, Staatliches Museum für Naturkunde Stuttgart
15974	MMU30	BOR 5	KY052131	antler drill	Borneo	1902	Dr. Stefan Merker, Staatliches Museum für Naturkunde Stuttgart
15968	MMU31	BOR 6	KY052132	antler drill	Borneo	1902	Dr. Stefan Merker, Staatliches Museum für Naturkunde Stuttgart
15979	MMU32	BOR 4	KY052133	antler drill	Borneo	1902	Dr. Stefan Merker, Staatliches Museum für Naturkunde Stuttgart

97.679	MMU34	THA 6	KY052097	nasal bone	Thailand	1900	Dr. Rainer Hutterer, Zoologisches Forschungsmuseum Alexander Koenig Bonn
2886	MMU35	JAV6	KY052134	nasal bone	West Java	1930	Dr. Frank Zachos, Naturhistorisches Museum Wien
2887	MMU36	JAV1 1	KY052135	nasal bone	West Java	1930	Dr. Frank Zachos, Naturhistorisches Museum Wien
1468	MMU38	SUM 4	KY052136	nasal bone	Sumatra	1904	Dr. Frank Zachos, Naturhistorisches Museum Wien
2120	MMU40	SUM 2	KY052137	nasal bone	Sumatra	1852	Dr. Frank Zachos, Naturhistorisches Museum Wien
3022	MMU41	JAV1 4	KY052138	nasal bone	West Java	1858	Dr. Frank Zachos, Naturhistorisches Museum Wien
2931	MMU42	JAV1	KY052139	nasal bone	West Java	1929	Dr. Frank Zachos, Naturhistorisches Museum Wien
2119	MMU43	JAV1 6	KY052140	nasal bone	West Java	1858	Dr. Frank Zachos, Naturhistorisches Museum Wien
3178	MMU44	JAV3	KY052141	nasal bone	West Java	1929	Dr. Frank Zachos, Naturhistorisches Museum Wien
2052	MMU46	JAV8	KY052142	nasal bone	West Java	1858	Dr. Frank Zachos, Naturhistorisches Museum Wien
2117	MMU48	SUM 5	KY052143	tissue	Sumatra	1886	Dr. Frank Zachos, Naturhistorisches Museum Wien
B3399	MMU49	JAV7	KY052144	skin	Java	1919	Dr. Frank Zachos, Naturhistorisches Museum Wien
B3400	MMU50	JAV2	KY052145	skin	Java		Dr. Frank Zachos, Naturhistorisches Museum Wien
B3402	MMU51	JAV4	KY052146	skin	Java		Dr. Frank Zachos, Naturhistorisches Museum Wien
1497	MMU53	SUM 3	KY052147	bone drill	Sumatra	1904	Dr. Frank Zachos, Naturhistorisches Museum Wien
35585	MMU59	LOM 1	KY052148	tissue	Lombok		Dr. Irina Ruf, Senckenberg Forschungsinstitut Mammalogie Frankfurt
34878	MMU60	LOM 3	KY052149	tissue	Lombok		Dr. Irina Ruf, Senckenberg Forschungsinstitut Mammalogie Frankfurt
15953	MMU62	IND6	KY052098	tissue	India		Dr. Irina Ruf, Senckenberg Forschungsinstitut Mammalogie Frankfurt
M122	MMU64	JAV9	KY052150	tissue from skeleton	Java		Mogens Andersen, Natural History Museum of Denmark Copenhagen
M123	MMU65	NEP1	KY052099	nasal bone	Nepal	1838	Mogens Andersen, Natural History Museum of Denmark Copenhagen

M505	MMU67	LSI2	KY052151	nasal bone	Belitung Island	1887	Mogens Andersen, Natural History Museum of Denmark Copenhagen
M537	MMU68	IND7	KY052100	nasal bone and tissue from skull	India	1887	Mogens Andersen, Natural History Museum of Denmark Copenhagen
M1267	MMU72	THA 4	KY052101	tissue from skull	Thailand	1926	Mogens Andersen, Natural History Museum of Denmark Copenhagen
M125	MMU74	THA 1	KY052102	Tissue in alcohol	South Thailand	1877	Mogens Andersen, Natural History Museum of Denmark Copenhagen
46109	MMU75	IND9	KY052103	skin	South India		Dr. Frieder Mayer, Naturkundemuseum Berlin
91138	MMU81	IND3	KY052104	skin	India		Dr. Frieder Mayer, Naturkundemuseum Berlin
91130	MMU82	IND4	KY052105	nasal bone	India		Dr. Frieder Mayer, Naturkundemuseum Berlin
70927	MMU83	SRI2	KY052117	nasal bone	Sri Lanka		Dr. Frieder Mayer, Naturkundemuseum Berlin
11959	MMU86	THA 5	KY052106	nasal bone	Thailand		Dr. Frieder Mayer, Naturkundemuseum Berlin
91133	MMU87	IND8	KY052107	nasal bone	India		Dr. Frieder Mayer, Naturkundemuseum Berlin
RMNH.MAM.33818	MMU90	JAV10	KY052152	nasal bone	West Java	1930	Pepijn Kamminga, Naturalis Leiden
RMNH.MAM.36981	MMU91	THA 3	KY052108	tissue from skull	North Thailand	1989	Pepijn Kamminga, Naturalis Leiden
RMNH.MAM.33814	MMU95	LSI3	KY052153	tissue from skull	Bangka Island		Pepijn Kamminga, Naturalis Leiden
RMNH.MAM.4929	MMU100	JAV12	KY052154	nasal bone	Java		Pepijn Kamminga, Naturalis Leiden
RMNH.MAM.4916	MMU102	IND1	KY052109	nasal bone	North India		Pepijn Kamminga, Naturalis Leiden
RMNH.MAM.29758.a	MMU105	BOR 3	KY052155	tissue from skull	Central Borneo	1894	Pepijn Kamminga, Naturalis Leiden
RMNH.MAM.29754.a	MMU107	BOR 1	KY052156	tissue from skull	Borneo		Pepijn Kamminga, Naturalis Leiden
74035/009957	C.2	CHI2	KY052110	nasal bone	Yunnan China		Douglas Yu, The Ecology, Conservation, and Environment Center
RJT118	RJT118	LAO 1	KY052082		Laos		R. J. Timmins
M2.3	M2.3	VIE3	KY052083	tissue from skull	North C. Vietnam	2011	Dr. Minh Le, Center for Natural Resources & Environmental Studies
M2.9	M2.9	VIE7	KY052084	bone	North C. Vietnam	2010	Dr. Minh Le, Center for Natural Resources & Environmental Studies
M2.14	M2.14	VIE2	KY052085	bone	Central Vietnam	2011	Dr. Minh Le, Center for Natural Resources & Environmental Studies

M3.8	M3.8	VIE11	KY052086	bone	North E. Vietnam		Dr. Minh Le, Center for Natural Resources & Environmental Studies
M5.11	M5.11	VIE13	KY052087	dry skin	North Vietnam		Dr. Minh Le, Center for Natural Resources & Environmental Studies
M6.5	M6.5	VIE8	KY052088	bone	North C. Vietnam		Dr. Minh Le, Center for Natural Resources & Environmental Studies
M6.12	M6.12	VIE1	KY052089	bone	Central Vietnam		Dr. Minh Le, Center for Natural Resources & Environmental Studies
M6.17	M6.17	VIE4	KY052090	bone	South Vietnam	2012	Dr. Minh Le, Center for Natural Resources & Environmental Studies
x15	x15	VIE9	KY052111	bone	North C. Vietnam		Dr. Minh Le, Center for Natural Resources & Environmental Studies
x17	x17	VIE12	KY052112	bone	North C. Vietnam		Dr. Minh Le, Center for Natural Resources & Environmental Studies
x19	x19	VIE10	KY052113	bone	North C. Vietnam		Dr. Minh Le, Center for Natural Resources & Environmental Studies
x20	x20	VIE6	KY052114	bone	North C. Vietnam		Dr. Minh Le, Center for Natural Resources & Environmental Studies
x39	x39	VIE5	KY052115	bone	North C. Vietnam		Dr. Minh Le, Center for Natural Resources & Environmental Studies
mm13	mm13	MAL1	KY052118		Peninsular Malaysia		Monash University, Malaysia
mm20	mm20	MAL2	KY052119		Peninsular Malaysia		Monash University, Malaysia

Table S3. Long-range PCR primer sequences and annealing temperatures used for bait development.

	Position on ref.	Sequence 5'-3'	Annealing Temp.[a]
Fragment 1	2478	F: CGATTAAAGTCCTACGTGATCTGAG	58 °C
	6926	R: GTTATGATGTTGGCTTGAAACCAG	
Fragment 2	6506	F: GCTATYATRGGAGGATTTGTTCAC	58 °C
	12903	R: GATTAGGGCTGTTGTRGTAAATG	
Fragment 3	12339	F: TTACAAATCTTAACGCCTGAGACTTC	58 °C
	2548	R: TAGATAGAAACCGACCTGGATTACTC	

[a] PCR was done with MyFi™ Mix (Bioline GmbH, Germany) with 1x MyFi Mix, 0.4 µM each primer and water to the final volume of 50 µL for each of the fragments separately, in a total of 35 amplification cycles.

Table S4. Sequencing results from all samples used in this study, with indication of sequencing platform, percentage of reads on target, average base coverage depth and percentage of reference genome covered with at least 3x coverage.

Sample Name	Captured	Sequencing platform	Reads on target	Average read depth	% genome covered ≥ 3x

BAL1	yes	PGM	53.65%	109.24	100
JAV13	yes	PGM	45.29%	294.49	100
THA2	yes	PGM	20.92%	45.73	99.9
IND5	no	Illumina	0.54%	90.76	100
IND2	yes	PGM	18.15%	16.17	99.4
CHI3	no	Illumina	0.06%	11.13	92.4
IND10	yes	PGM	46.65%	257.54	100
BOR2	yes	Illumina	2.21%	14.07	93.7
LOM2	yes	Illumina	36.22%	124.63	100
SRI1	yes	PGM	20.46%	112.53	99.2
CHI1	yes	PGM	4.30%	186.14	100
LSI1	yes	Illumina	45.26%	172.8	100
BOR7	yes	PGM	10.17%	39.69	99.9
SUM6	yes	PGM	36.05%	96.74	100
SUM1	yes	PGM	47.57%	64.02	99.8
JAV17	yes	PGM	71.62%	65.65	99.8
JAV15	yes	PGM	65.44%	88.36	100
JAV5	yes	PGM	55.88%	163.68	100
BOR5	yes	PGM	6.24%	6.87	88.1
BOR6	yes	PGM	31.17%	7.97	93.9
BOR4	yes	PGM	18.20%	27.78	99.9
THA6	yes	PGM	26.33%	178.63	100
JAV6	yes	PGM	6.35%	187.35	100
JAV11	yes	PGM	40.60%	36.44	99
SUM4	yes	PGM	44.44%	278.49	100
SUM2	yes	PGM	61.77%	27.59	99
JAV14	yes	PGM	30.97%	144.77	100
JAV1	no	Illumina	0.92%	74.02	100
JAV16	yes	PGM	30.47%	36.33	99.5
JAV3	no	Illumina	0.18%	14.49	92.5
JAV8	yes	Illumina	4.88%	25.39	96.1
SUM5	no	Illumina	0.05%	100.91	100
JAV7	yes	Illumina	3.88%	79.14	98.2
JAV2	yes	Illumina	9.26%	89	99.6
JAV4	yes	PGM	50.03%	10.66	89.7
SUM3	yes	PGM	24.66%	72.7	100
LOM1	yes	PGM	28.73%	37.86	99.1
LOM3	yes	PGM	33.09%	35.93	99.3
IND6	yes	PGM	50.73%	47.47	99.9
JAV9	yes	PGM	27.96%	20.03	98.4
NEP1	yes	PGM	59.64%	31.76	100
LSI2	yes	PGM	16.87%	17.46	98.8
IND7	yes	PGM	47.52%	146.08	100
THA4	no	Illumina	0.03%	8.43	86.1
THA1	yes	Illumina	14.59%	92.01	100
IND9	yes	PGM	33.76%	7.29	92.6
IND3	yes	Illumina	1.76%	22.85	98.17
IND4	yes	Illumina	8.41%	56.76	99.65
SRI2	yes	PGM	61.03%	51.33	98.6
THA5	yes	Illumina	29.52%	155.25	100
IND8	yes	Illumina	64%	218.42	100
JAV10	no	Illumina	0.07%	78.4	99.7
THA3	yes	PGM	39.65%	17.13	99.8
LSI3	yes	Illumina	18.50%	147.87	100
JAV12	yes	Illumina	3.52%	34.46	99.4
IND1	yes	PGM	74.84%	108.69	100
BOR3	yes	Illumina	19.56%	88.28	97.9
BOR1	yes	Illumina	39.98%	115.74	99.3
CHI2	no	Illumina	0.19%	49.41	100
LAO1	no	Illumina	0.18%	53.8	99.05

VIE3	no	Illumina	0.04%	12.66	96.7
VIE7	no	Illumina	0.03%	11.42	95.4
VIE2	no	Illumina	0.07%	16.19	92.5
VIE11	no	Illumina	0.06%	21.06	97
VIE13	no	Illumina	0.08%	19.42	99.7
VIE8	no	Illumina	0.10%	29.03	96.8
VIE1	no	Illumina	0.03%	8.19	90.6
VIE4	no	Illumina	0.03%	11.25	94.1
VIE9	no	Illumina	0.12%	37.95	99.9
VIE12	no	Illumina	0.20%	70.42	100
VIE10	no	Illumina	0.07%	31.10	100
VIE6	no	Illumina	0.03%	9.38	94.2
VIE5	no	Illumina	0.06%	30.30	99.9
MAL1	no	Illumina	0.93%	66.63	100
MAL2	no	Illumina	0.58%	57.01	100

Table S5: Number of Variable sites (V) and Parsimony informative sites (Pi) per coding gene, throughout the full mitogenomes obtained.

	ND1	ND2	Cox1	Cox2	Cox3	ND3	ND4L	ND4	ND5	ND6	Cytb	D-loop
V	126	147	163	76	108	43	38	170	277	64	182	157
Pi	102	101	111	54	79	31	27	132	191	45	127	120

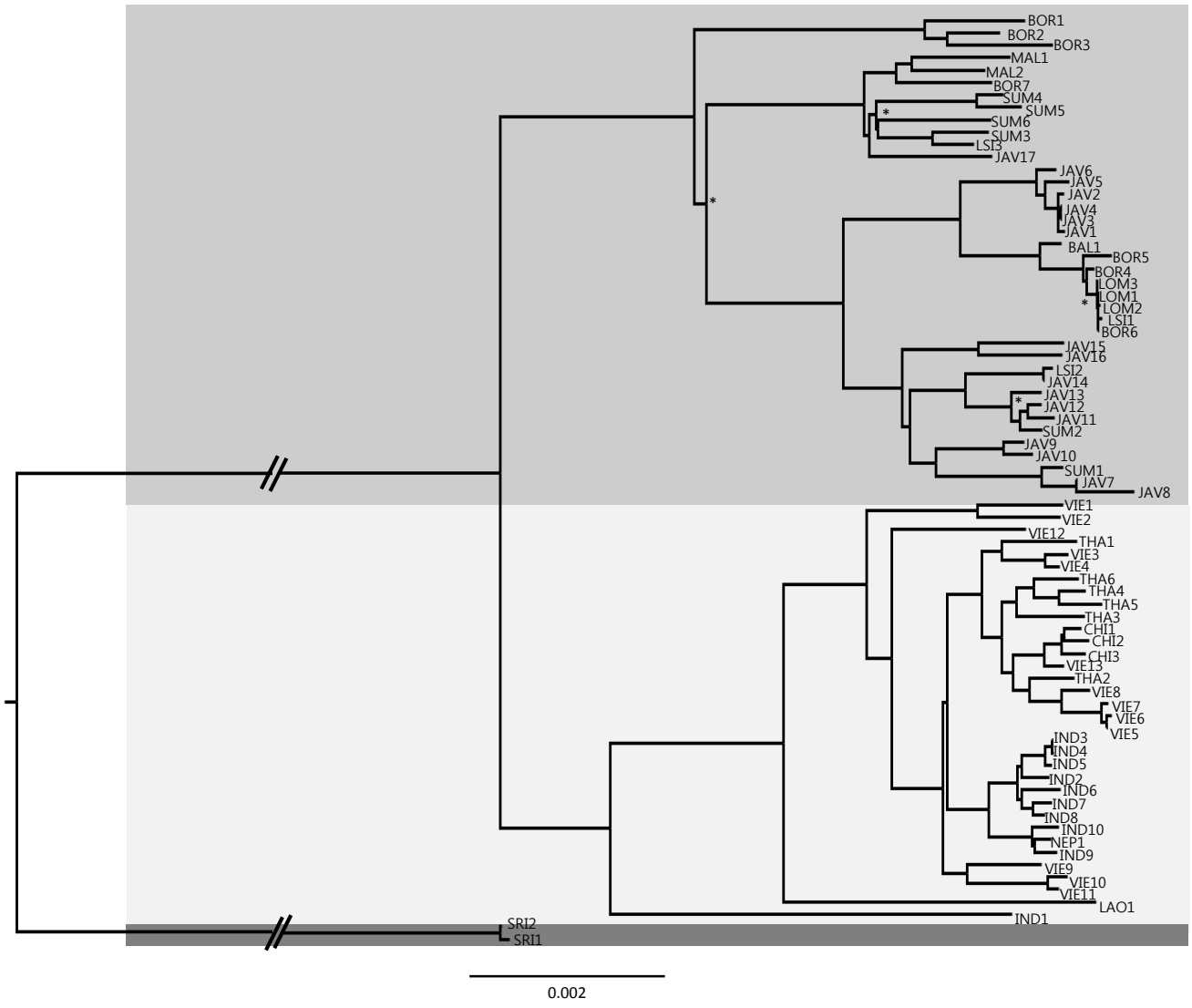

Figure S1. Best tree obtained with Maximum Likelihood analyses with RAxML. The same topology is recovered as with Bayesian inferences, with three distinct mitochondrial clades. All branches are supported with at least 90% bootstrap support, except when indicated otherwise with *.

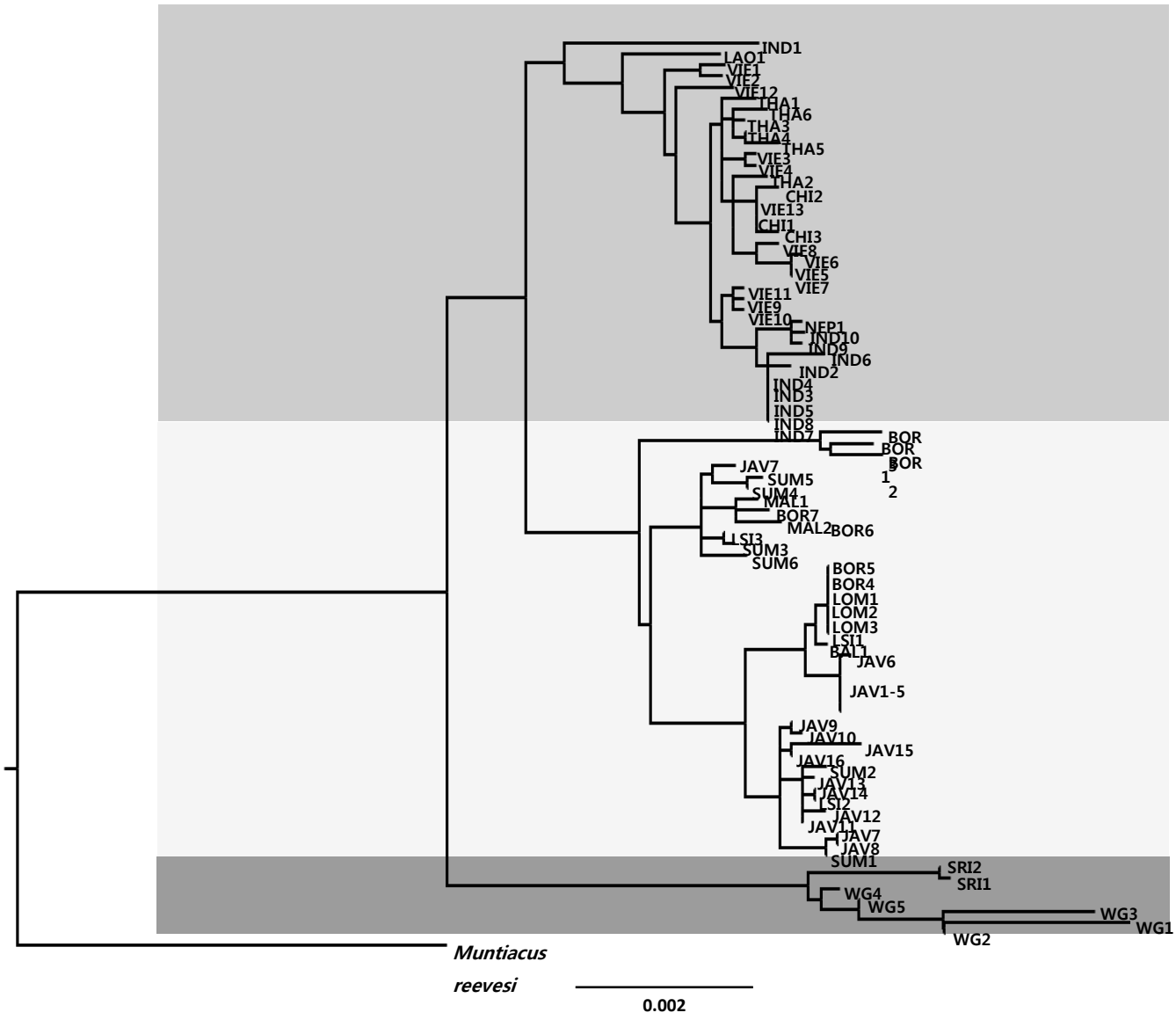

Figure S2. Best tree obtained with Maximum Likelihood analyses of the Cytochrome B gene of all red muntjac samples included in this study and five additional Genbank sequences from the Western Ghats. This tree was obtained with RAxML and shows similar topology as the trees with full mitogenome, although bootstrap values are generally lower. This result shows the relationship between Sri Lanka and Western Ghats populations, as they cluster together in the same clade.

Manuscript II

Introgression in Rusa deer

Human mediated introduction of introgressed deer across Wallace's line: historical biogeography of *Rusa unicolor* and *R. timorensis*

Renata F. Martins,[1*] Anke Schmidt,[1] Dorina Lenz,[1] Andreas Wilting,[1] Jörns Fickel[1,2]

[1] Department of Evolutionary Genetics, Leibniz Institute for Zoo and Wildlife Research, Alfred-Kowalke-Str. 17, 10365 Berlin, Germany

[2] Institute for Biochemistry and Biology, University of Potsdam, Karl-Liebknecht-Str. 24–25, 14476 Potsdam, Germany.

* Corresponding author: Renata F. Martins, martins@izw-berlin.de

Under review in *Journal of Biogeography*

Human mediated introduction of introgressed deer across Wallace's line: historical biogeography of *Rusa unicolor* and *R. timorensis*

Running title: Historical hybridization in Rusa deer

Authors: Renata F. Martins,[1*] Anke Schmidt,[1] Dorina Lenz,[1] Andreas Wilting,[1] Jörns Fickel[1,2]

Affiliations: [1] Department of Evolutionary Genetics, Leibniz Institute for Zoo and Wildlife Research, Alfred-Kowalke-Str. 17, 10365 Berlin, Germany. [2] Institute for Biochemistry and Biology, University of Potsdam, Karl-Liebknecht-Str. 24–25, 14476 Potsdam, Germany.

* Corresponding author: Renata F. Martins, martins@izw-berlin.de

Acknowledgements: The authors would like to thank all curators listed in Supplemental Table S1 for providing access to their collections. This project was funded by the Leibniz competitive fund SAW-2013-IZW-2.

Abstract:

Aim We compared the phylogeographic patterns of two Rusa species, *Rusa unicolor* and *Rusa timorensis*, in order to understand what drove and maintained differentiation between these two geographically and genetically close species and investigated the route of introduction of individuals to the islands outside of the Sunda Shelf.

Location Southeast Asia mainland, Sundaland and islands of Wallacea

Methods We analysed full mitogenomes from 56 archival samples from the distribution areas of the two species and 18 microsatellite loci in a subset of 16 individuals to generate the phylogeographic patterns of both species. Bayesian inference with fossil calibration was used to estimate the age of each species and major divergence events.

Results Our results indicated that the split between the two species took place during the Pleistocene, ~1.8 Mya, possibly driven by adaptations of *R. timorensis* to the drier climate found on Java compared to the other islands of Sundaland. Although both markers identified two well differentiated clades, there was a largely discrepant pattern between mitochondrial and nuclear markers. nDNA separated the individuals into the two species, but mtDNA revealed that all *R. timorensis* sampled to the east of the Sunda shelf carried haplotypes from *R. unicolor*, while one *Rusa unicolor* from South Sumatra carried a *R. timorensis* haplotype.

Main conclusions Our results show that hybridisation occurred between these two sister species in Sundaland during the Late Pleistocene and resulted in the first reported of early human mediated

introduction of hybrid descendants in all islands outside Sundaland, constituting the first report of early hybrid translocation events mediated by humans.

Key Words: Cervidae; hybridisation; human introduction; Phylogeography; Sundaland; Wallace's line

Introduction

Biogeographic barriers interrupt migration and reproduction among separated populations, and thus are responsible for driving and maintaining genetic differentiation potentially leading to speciation. Sundaland, a Southeast Asian biodiversity hotspot, is bordered in the east by one of the best known faunal boundaries – the Wallace line (Bacon *et al.,* 2013). This barrier is responsible for a sharp break between faunal compositions of Sunda and Wallacea. At the southern border, the Wallace line runs between the islands of Bali and Lombok, and at the northern edge it separates fauna and flora of Borneo and the Philippines from that of Sulawesi (Fig. 1). Although some species and populations have naturally crossed this barrier (e.g. Plantain squirrel *Callosciurus notatus* in Lombok), the presence of the majority of Sundaic mammal species occurring past the Wallace line into the eastern islands of Wallacea is associated with human transportation (Groves, 1983; Heinsohn, 2003; Veron *et al*., 2014). Contrary to natural dispersal of individuals, human mediated introductions occurred within a short time frame, too short for introduced individuals to become genetically differentiated from their source population.

Sundaland's dynamic geologic and climatic history, especially during the Plio-Pleistocene epochs, resulted in sea level changes that repeatedly exposed the continental shelf connecting the major islands

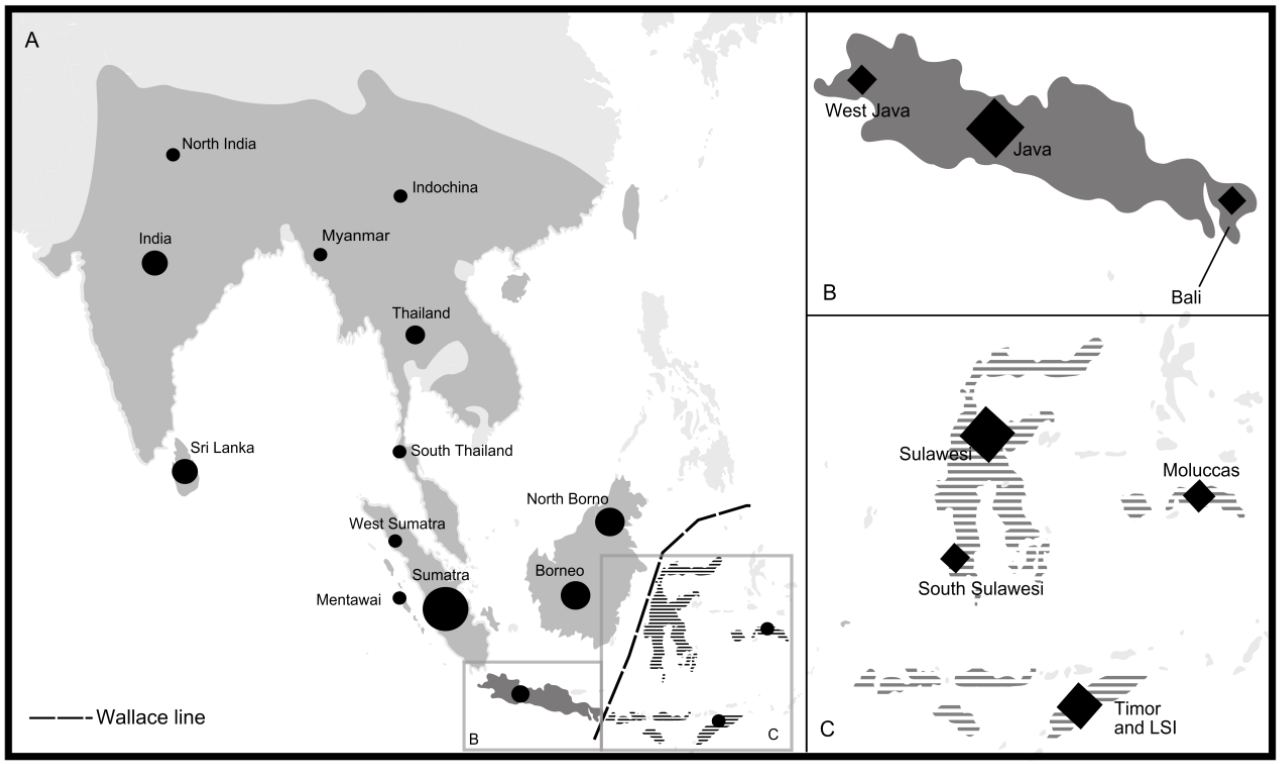

Figure 1 - Distribution map of both species and sampling location. Light grey indicates the distribution range of *Rusa unicolor* (A). Dark grey indicates the native distribution of *Rusa timorensis* (B) and dashed dark grey areas indicate introduction range of *R. timorensis* (C). Filled circles and diamonds indicate *R. unicolor* and *R. timorensis*, respectively, and size is proportional to the number of samples. For detailed information about all samples see Supplemental Table S1.

of this archipelago (Voris, 2000). It is generally believed that these available land corridors would allow populations previously isolated on single islands to disperse, creating a large panmitic population within this whole system. However, geographical barriers like the Wallace line would stay in place, thus creating patterns of genetic divergence between taxa on both sides of these barriers.

Here we investigated the phylogeographic patterns of two Rusa species: the sambar, *Rusa unicolor* and the Javan deer, *R. timorensis*. While *R. unicolor* is widespread throughout South and Southeast Asia (from India and Sri Lanka, Southern China and most of Indochina to Borneo and Sumatra, the two largest Sunda Islands), *R. timorensis* has its native range on Java and Bali only (Fig. 1). Its presence on islands east of the Wallace line (e.g. Lesser Sunda Islands, Sulawesi and the Moluccas) is described to be the result of pre-historic to historical human-mediated introductions during the Holocene (less than 10 thousand years ago; Heinsohn, 2003; Groves & Grubb, 2011).

The aim of this study was to compare phylogeographic patterns, genetic diversity and evolutionary history of these two related species in order to answer the following three questions: 1) In the presence of land bridges connecting islands of the Sunda Shelf, what speciation drivers kept the two species as separate units or is there evidence of admixture between them? 2) Do populations of the widely distributed *R. unicolor* show signs of genetic structuring corresponding to known geographical barriers or do they show variation along a cline? 3) Does *Rusa timorensis* show genetic signature of non-natural dispersal and what is the most likely founder population of the introduced populations east of the Wallace line?

Given the different (past and current) climatic conditions on Sumatra and Java (Leonard *et al*., 2015) we hypothesize that adaptation to these different local climate drove speciation between the two sister species, but that in the presence of permissive land corridors, individuals from previously isolated populations could disperse and admix. To test this hypothesis we sampled historical populations of both species from their native range and from the introduction range of *R. timorensis*. We analysed full mitogenomes and a set of 18 microsatellite loci to test for admixture between the two species.

Materials and Methods

Sampling and DNA extraction

We sampled 110 individuals labelled as *Rusa unicolor* (RUN) and *Rusa timorensis* (RTI) from European museums by collecting either turbinal bones from the nasal cavity, skin, dry tissue from skeletons, and antler drills exclusively from individuals with known locality. All molecular work, including DNA extraction and constructing sequencing libraries, was conducted in a facility dedicated to work with archival samples to reduce the risk of contamination. DNA extraction followed the DNeasy Tissue and Blood kit protocol (Qiagen, Hilden, Germany), with overnight digestion of samples in Lysis buffer and Proteinase K at 56 ^{0}C and a pre-elution incubation for 20 min. at 37 ^{0}C.

Mitochondrial genome

All extractions, including negative controls, were built into individual sequencing libraries with single 8-nt indexes (Fortes & Paijmans, 2015), which were then sequenced on an Illumina MiSeq to assess sample DNA quantity and quality (150 cycles v3 kit, Illumina, CA, USA). Samples with low quality DNA were consequently enriched for their mitochondrial DNA, using an in-solution target hybridization capture technique (Maricic *et al*., 2010). Baits for hybridisation were obtained by amplifying three overlapping mitochondrial fragments from one fresh sample of *R. unicolor* (from the IZW archive) which were consequently prepared into capture baits (Maricic *et al*., 2010; primers and PCR conditions as described in Martins *et al*., 2017). After hybridization capture, libraries were amplified for no more than 18 cycles and sequenced again on the Illumina MiSeq platform.

Bioinformatics

Sequencing reads were first de-multiplexed into respective samples with BCL2FASTQ v2.17 (Illumina, CA, USA). CUTADAPT v1.3 (Martin, 2011) was used to find and remove adapter sequences from the sequenced reads. Adapter-clipped reads were then quality trimmed through a sliding window approach of 10 bp for a phred score of at least Q20. Finally, reads shorter than 20 bp were removed from further analyses. Mapping of quality filtered reads was done in two phases: a first mapping run was performed with BWA v.0.7.10 (Li & Durbin, 2009), using a genome reference from *R. unicolor dejeani* (NCBI accession no. NC_031835). Clonal reads were removed from the mapped reads using MARKDUPLICATES v1.106 (http://picard.sourceforge.net/picard-tools). SAMTOOLS MPILEUP v1.1 and BCFTOOLSv1.2(http://github.com/samtools/bcftools) were used for variant calling (SNPs and InDels). A consensus sequence was then generated for each sample using a threshold of minimum 3× coverage and majority rule (> 50%) for base calling. The second mapping step used the newly generated consensus sequence as reference for each sample, in order to increase mapping quality and base coverage. BWA and MARKDUPLICATES were used as before, but GATK v1.6 (McKenna *et al*., 2010) was applied for variant calling of the final consensus sequence. Positions with coverage lower than 5× were N-masked, as were ambiguous heterozygous positions. A final quality filtering step was performed to remove all samples with less than 80% of their mitogenome covered at least with 5× depth. Due to the limitations of the sequencing method (reads no longer than 75 bp) we were not able to clearly resolve the repeat region of the d-loop. Therefore, we trimmed all sequences, by removing 460 bp of the d-loop region. Mitogenomes obtained were deposited in Genbank (for accession numbers see Table S1).

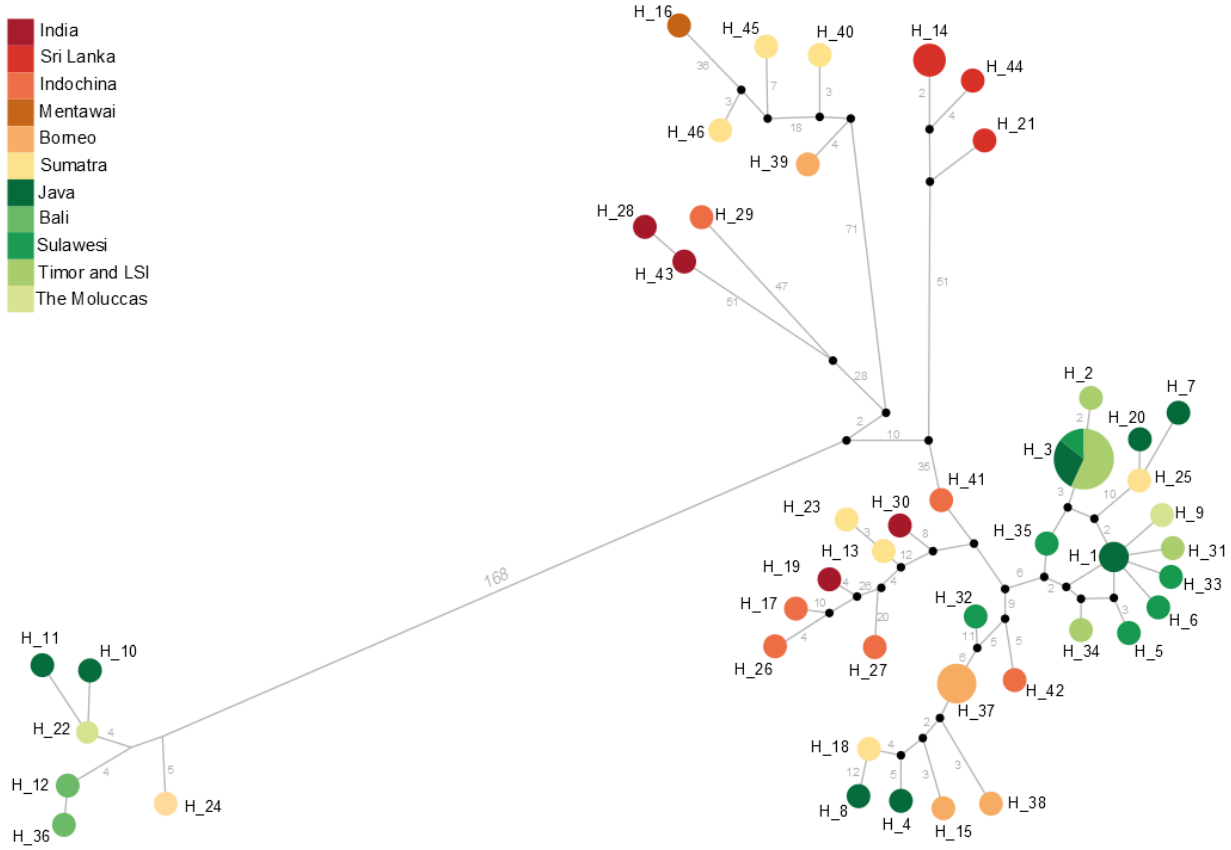

Figure 2 – Haplotypic network for all 46 haplotypes shared among the two species. Circle size is in accordance with frequency and colour represents sampling location. Small black circles represent median vectors. All branches represent one mutation step, except when indicated otherwise by numbers on branches. Haplotypes are described in Supplemental Table S1.

Microsatellite DNA

Microsatellite genotyping was achieved by amplification of 18 loci on those samples for which mitochondrial DNA was obtained as described above (microsatellite loci and references in Table S2). All samples were amplified for all loci through PCR with the Type-it Microsatellite PCR kit (Qiagen, Hilden, Germany), with 1 µM of each primer. Annealing temperatures followed a gradient from 63 °C to 55 °C in 2 °C steps and final amplification occurred for 40 cycles at 55 ^{0}C. Allele sizes were determined on an ABI3130xl Genetic Analyser using GeneScan™ 500 ROX (both Thermo Fischer Scientific Darmstadt, Germany) as internal size standard. Alleles were scored with the software GeneMapper v.4.0 (Applied Biosystems, Germany).

Genetic diversity, phylogeography and differentiation times

All mitochondrial sequences obtained were aligned using the auto setting as implemented in MAFFT v7.245 (Katoh & Standley, 2013). The relationship among all haplotypes was reconstructed by a median joining (MJ) network using the software NETWORK v. 4.6.1.4 (Bandelt *et al.*, 1999). Haplotypes were generated by removing non-informative sites and positions with gaps or missing data.

Haplotypic and nucleotide diversity for the full dataset and for each species were assessed with the software DNASP v.5.10 (Librado & Rozas, 2009). We estimated genetic differentiation through F_{ST} as implemented in ARLEQUIN v.3.5.12 (Excoffier *et al*., 2005). For this analysis we created two datasets: i) two populations corresponding to species as determined by the museum identification and ii) populations corresponding to major haplotype clades.

The best fitting substitution model for the full mitogenome dataset (GTR +G +I) was obtained by the hierarchical likelihood ratio test as implemented in JMODELTEST v2.1.7 (Darriba *et al*., 2015). We reconstructed phylogenetic relationships through Maximum Likelihood (ML) with RAXML GUI v1.5 (Silvestro & Michalak, 2012) and Bayesian Inference (BI) as implemented in MRBAYES v3.2.6 (Ronquist & Huelsenbeck, 2003), applying the determined substitution model. Both approaches were congruent with the haplotypic network and with each other, revealing the existence of separated mitochondrial clades, which we dated based on fossil information for the Cervidae stem of the Arctiodactyl family (18.4 Mya; Bibi, 2013). We used the calibrated time of the split between Cervus and Rusa [2.1 My; Highest Posterior Density (HPD) = 1 – 2.8 My] to be set as a prior for the root height of the gene tree of the Rusa samples obtained in this study, as implemented in BEAST v.1.8.1 (Drummond *et al*., 2012). We ran MCMC chains with 50 million, with a lognormal uncorrelated clock and a Yule speciation tree as further estimation priors. Trace results were analysed with TRACER v.1.6 (implemented in BEAST v.1.8), for parameter convergence and ESS values above 200. TREEANNOTATOR v.1.8.1 (BEAST software package) was used to annotate all trees, after a burn-in of 10% of the trees. All topologies were visualised and edited with the software FIGTREE v.1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

Population structure analyses

Analyses of microsatellite data proceeded by removing all individuals with missing data at more than two loci. We estimated the probability for the presence of null alleles on our dataset with the software FREENA (Chapouis & Estoup, 2007). Tests for the presence of linkage disequilibria and an exact test for deviations from Hardy-Weinberg Equilibrium (HWE) were performed with ARLEQUIN, applying the Bonferroni correction for multiple tests. Levels of observed (HO) and expected (HE) heterozygosity and inbreeding coefficient (F_{IS}) were calculated in GENETIX v.4.05.2 (Belkhir *et al*., 1996-2004).

A Bayesian approach was used to test for population structure with the software STRUCTURE 2.3.4 (Pritchard *et al*., 2000). The λ value was estimated by running a prospective run of K=1 with 10 iterations and a burn-in of 10% after 15×10^4 generations. A second MCMC simulation was run for 20×10^4 generations, with a 10% burn-in. The likelihoods were estimated for K values from 1 to 6 (because 6 was higher than the maximum number of mitochondrial clades obtained in our analyses). The admixture model was applied with correlated allele frequencies and $\lambda = 3$. STRUCTURE HARVESTER v.0.6.94 (http://taylor0.biology.ucla.edu/structureHarvester/; Earl & von Holdt, 2012)

was used to estimate the most likely number of K by using the ΔK method (Evanno *et al*., 2005). Population differentiation was calculated with the software ARLEQUIN by estimating F_{ST} both among the clusters identified by ΔK and by species.

Results

Mitochondrial genome analyses

The final dataset consisted of 56 individuals for which the mitogenome of 16064 bp was obtained. Of these, 23 were labelled in the museum collections as *R. timorensis* and 33 individuals were labelled as *R. unicolor* (Table S1). The origin of three specimens labelled as *R. unicolor* (RUN37 Moluccas, RUN39 Timor and RUN61 Java) suggests that these specimens are actually Javan deer, *R. timorensis*. All other museum labels matched the geographic distribution ranges of the two species. The 56 deer shared 46 haplotypes with an overall haplotype diversity of H = 0.983 (SD: 0.010) and a very low nucleotide diversity of π = 0.00941 (SD: 0.00118). Within individuals labelled as RTI, we found 18 haplotypes with H =0.972 (SD: 0.026) and π = 0.0085 (SD: 0.0024). Within RUN labelled individuals,

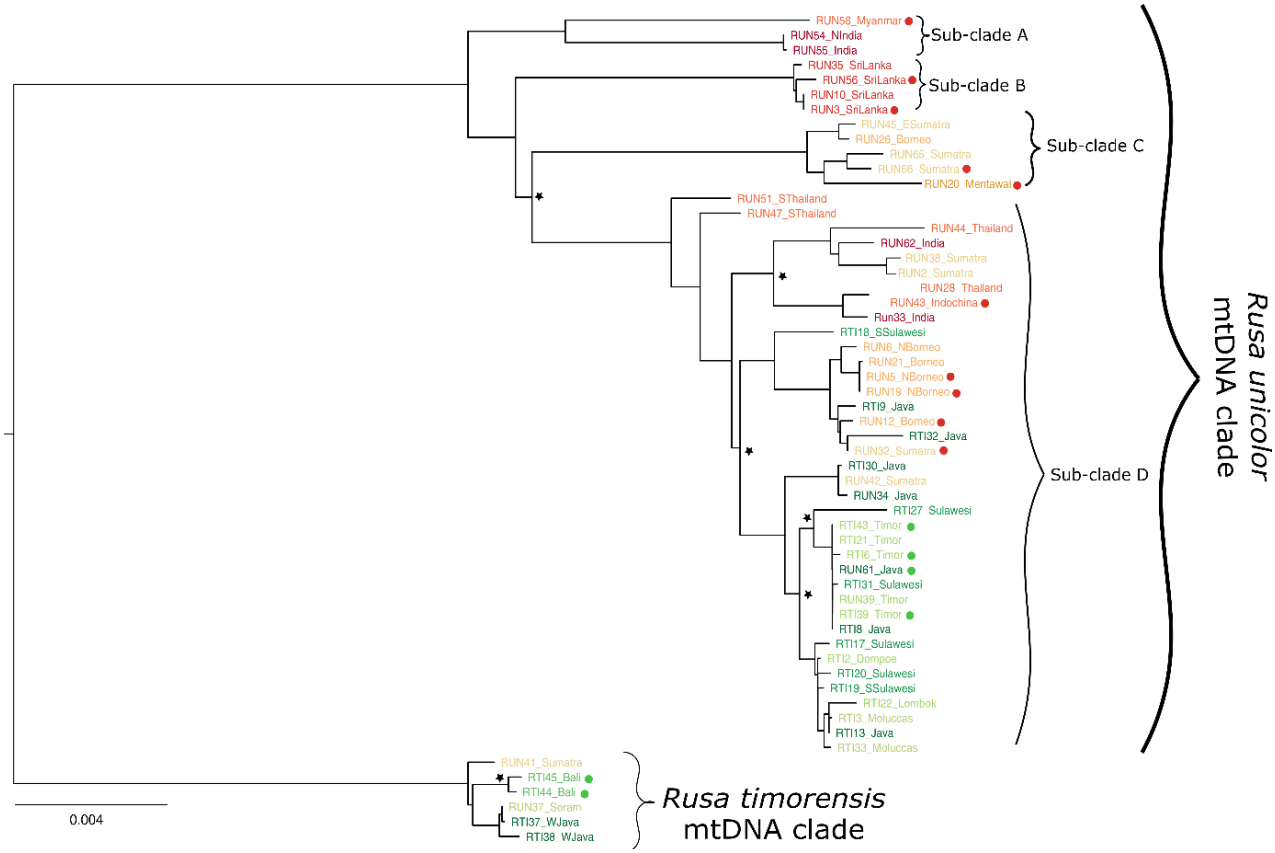

Figure 3 – Mitogenome Maximum Likelihood tree of both species. Colours on tips represent sampling location (as in Fig. 2) and stars represent split events with bootstrap values/Bayesian posterior probabilities lower than 90/0.95 (but bigger than 50/0.5). Red and green dots represent samples for which we obtained nDNA; red dot: assigned to the Rusa unicolor genotypic cluster, green dot: assigned to the Rusa timorensis genotypic cluster. Major mtDNA clades and sub-clades are labelled with curved brackets. Scale bar indicates number of substitutions per position.

there were 28 unique haplotypes and one haplotype was shared with RTI (H_3). Overall, the 29 haplotypes had H = 0.991 (SD: 0.011) and π = 0.011 (SD: 0.0012).

The full mtDNA haplotype network separated two major clades by a minimum of 168 mutational steps (Fig. 2). The smaller clade comprised six individuals from Java, Bali, Moluccas, and South Sumatra (Fig. 2) of which four had been labeled in the museum collections as *R. timorensis* (Javan deer, RTI) and two as *R. unicolor* (sambar, RUN; RUN41: South Sumatra; RUN37: Moluccas, see above). The second major clade comprised all other samples, including samples labelled as RTI from Java and from the introduction range of Javan deer (Lesser Sunda Islands, Sulawesi and the Moluccas). Both Maximum Likelihood (ML) and Bayesian Inference (BI) tree topologies were concordant with the overall pattern recovered by the haplotypic network, also revealing the existence of two well differentiated clades (Fig. 3). Generally, genetic diversity of haplotypes showed geographical structure only in some parts of the tree (sub-clades A, B, C) but none in others; e.g. samples from the Moluccas, Sulawesi and Timor (outside Sundaland) were present in more than one branch of the phylogenetic tree (sub-clade D).

To be able to determine the age of the most recent common ancestor of RUN and RTI we first estimated divergence dates for a subset of Cervidae species. The resulting node ages were similar to those reported in other studies (Table 1). Using those, we then estimated the divergence between the two main clades to have started in the early Pleistocene, about 1.8 million years ago (Mya) (HPD = 0.95 – 3.1). The position of all clades was similar to the topology recovered by ML, with the exception of the Sri Lankan clade position (sub-clade B, Fig.3 & 4), which was more recent in the BI tree. This sub-clade was also accompanied by low Bayesian Posterior Probability values. Nevertheless, our divergence estimates indicated that sub-clade A diverged first within the RUN mitogenome clade about 1.4 Mya (HPD = 0.7 – 2.3); sub-clade C diverged 1.185 Mya (HPD = 0.6 – 2) and sub-clade D at around 1.13 Mya (HPD = 0.6 – 2) (Fig.4). F_{ST} showed significant population differentiation only when populations were based on mtDNA clade assignment, but not when they were based on species assignment from the museum collections (Table 2).

Table 1 – Calibrated divergence dates estimated for the Cervidae tree. Ages [in million years (My)] represent the median obtained for each of the described split. Values in bold represent fossil-based calibrations.

	Age	minimum	maximum	Calibration
Bovidae/Cervidae	16.8	10.7	23.3	**18.4** (Bibi, 2013)
Muntiacus/Cervus	7.2	4.1	10.2	7.5 (Martins et al., 2017)
Axis/Cervus/Dama	4.6	2.6	5.4	**6** (Di Stefano & Petronio, 2002)
Cervus/Rusa	2.1	1	2.8	**3.4** (Pitra et al. 2004)
R. unicolor /timorensis	1.4	0.7	2.2	---

Microsatellite analyses

Of all archival samples for which mitogenomes were obtained, 16 individuals (~ 29%) could be successfully genotyped at 18 loci. These individuals were distributed relatively well across the clades of the mitogenome tree (Fig. 3). Linkage disequilibria were found at 10% of all pairwise loci combinations, yet without any consistent pattern, and percentage of null alleles was 0.18. Therefore we retained all loci for further analyses. We detected significant deviations from Hardy-Weinberg Equilibrium, which indicated population structure within our dataset. Expected and observed heterozygosities at each locus ranged from 0.23 (locus Mu_4D) to 0.92 (locus Mu_1_51) and from 0 (locus Mu_4D) to 0.61 (locus Roe09), respectively (Table S2). Number of alleles varied among loci, with the highest number found at loci Mu_1_51 and NVHRT48 (16 alleles) and the lowest found at locus Mu_4D with only two alleles (Table S2).

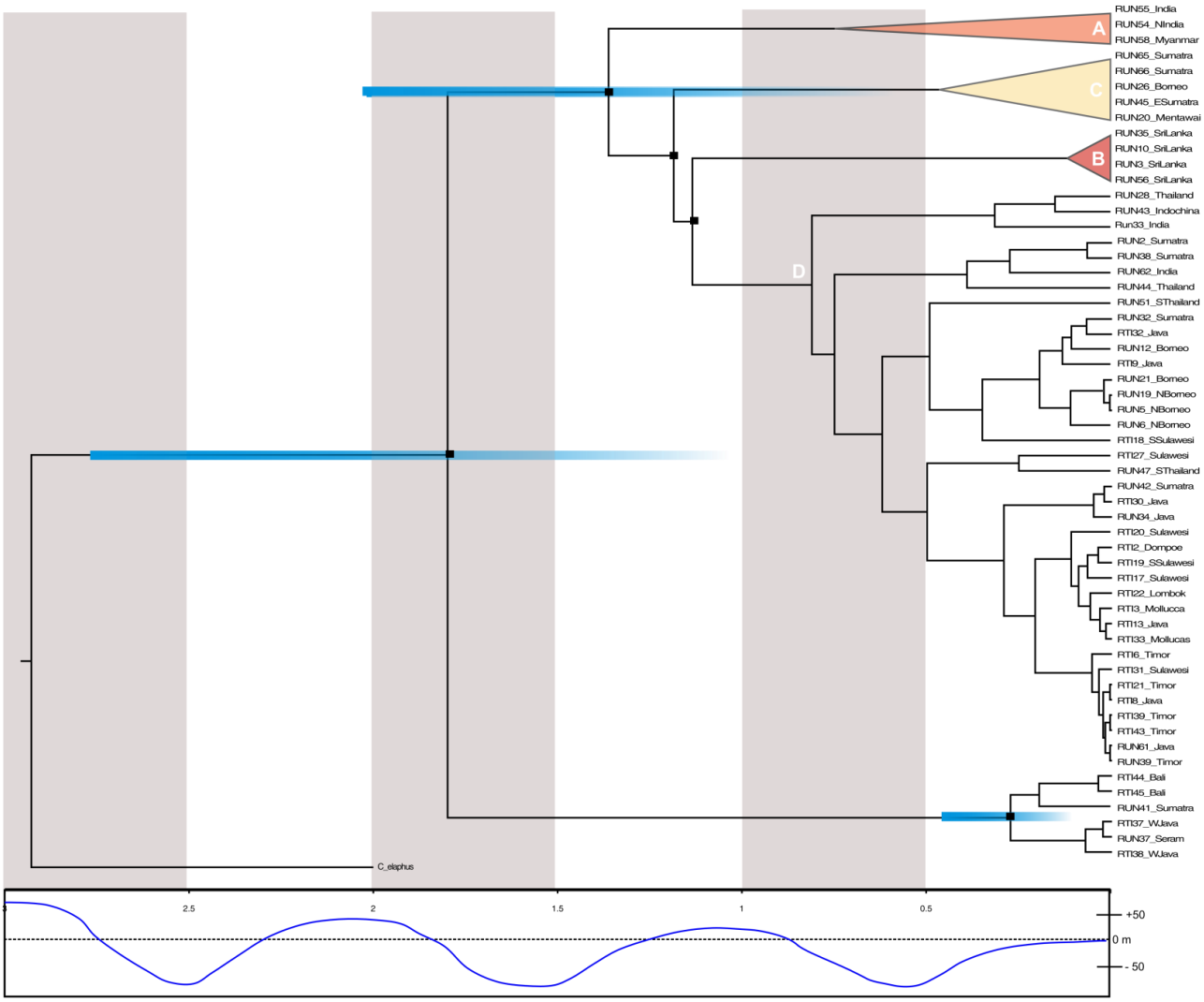

Figure 4 – Dated tree according to BEAST analyses, from 3 Mya to present. Blue bars represent associated deviations for the most important splits. A time scale in millions of years and a rough estimate of sea level changes through time (adapted from 54) are presented below.

According to the ΔK approach, the most likely number of genotypic nDNA clusters was K = 2 (Fig. 5A). However, there was indication of further sub-structuring (at K=5; Fig. 5B). These two main clusters corresponded well to the two species, sambar (R. unicolor, green cluster) and Javan deer (R. timorensis, red cluster). RUN61 from Java was genotypically assigned to RTI (see above). Population differentiation (F_{ST}) was always significant, independent of the grouping method (museum collections or STRUCTURE analyses, Table 2).

Discussion

The mitochondrial genomes and the nDNA loci of sambar and Javan deer investigated here revealed an intriguing (and surprising) pattern of genetic diversity and population differentiation between the two species. Although monophyly of *R. timorensis* and *R. unicolor* remained undisputed, our results point to a more complex history of (potentially reciprocal) hybridisation between species and multiple human-mediated introductions outside the Sunda Shelf.

The presence of two divergent matrilineages clearly indicates molecular differentiation between two groups of Rusa deer, which we interpret as the historical separation between *R. timorensis* and *R. unicolor*. Our fossil calibrated estimates are corroborated by recent studies (Escobedo-Morales *et al*., 2016; Bibi, 2013; Table 1) and are in accordance with the dates suggested by other authors for the age of the genus Rusa (e.g. 2 - 2.5 Mya; Di Stefano & Petronio, 2002). Some authors proposed that Rusa unicolor presents the most ancestral-like antler morphology, so it may have been the first species to diverge from the Rusa common ancestor (Pitra *et al*., 2004). The separation of the two deer species investigated here had been challenged in the past (mentioned in van Bemmel, 1949), yet subsequent studies found robust support for their distinctiveness (morphological: van Bemmel, 1994; Meijaard & Groves, 2004; and molecular: Emerson & Tate, 1993; Pitra *et al*., 2004). Our comprehensive molecular study corroborated these findings, but also provides evidence for a much more complex evolutionary history of the Rusa deer. It has been proposed that Rusa-like deer have appeared in Northern India, around 2.5 Mya, where they adapted to dense forest habitats with some open grass

Table 2 – Population differentiation estimates (F_{ST}) according to marker and grouping. Estimations were performed both for the mtDNA and nDNA. Populations were generated by either museum ID (both for mtDNA and nDNA) and by assignment of individuals to one of the two major clades (mtDNA) or to one of the genotypic clusters (nDNA). Statistically significant comparisons are indicated by bold p-values.

		Fst	p-value
mtDNA	Museum ID	0.085	> 0.001
	Results (2 clades)	0.75	**< 0.001**
nDNA	Museum ID	0.12	**< 0.001**
	Results (K=2)	0.14	**< 0.001**

vegetation (Geist, 1998). However, during the Pleistocene, sub-tropical forest migrated southwards, completely disappearing from China (Meijaard & Groves, 2004). This would have shifted the distribution area of sub-tropical forest-adapted Rusa (or Rusa-like) species southwards too. When low sea levels allowed, Rusa deer could reach Sundaic islands, including Java. Sea levels remained low until 1.4 Mya, maintaining landmasses connected by the emerged continental shelf (van der Bergh *et al*., 1996). By 1 Mya, sea levels had risen again and had reached a highstand at +5 m compared to present day (Zazo, 1999), thereby flooding all land connections between islands. At this time, Rusa populations of Java and Sumatra (clade D) would have become isolated and habitat availability for forest dependent species would have been reduced.

Our data indicate as well a second wave of colonization to Sundaic islands by *Rusa unicolor*, likely from Thailand (Mainland). This second wave would have likely occurred during the Late Pleistocene, with drops in sea levels and once again cooler and drier climates. This southward expansion brought previously isolated Rusa unicolor in contact with Rusa timorensis from Java, facilitated by the presence of the emerged Sunda strait, a strait that submerged just ~10 kya (Sathiamurthy & Voris, 2006). This fact raises the obvious question of what then maintained differentiation between the two species and what kept hybridisation contained to a small secondary contact region. We see an explanation in the different ecological niches (vegetation) on Java and Sumatra. Speciation may have resulted from the ecological adaptation of Javan deer *Rusa timorensis* to the prevailing vegetation type on Java, separating it from its sister species, the sub-tropical forest-adapted sambar *Rusa unicolor*. Java and Bali, although part of Sundaland, had (and still have) different climatic conditions than Sumatra and Borneo and thus allowed differentiation of species based on evolved ecological adaptations (Leonard *et al*., 2015). The climate on Java is characterized by a West-East gradient, a transition from a slightly seasonal climate in the West to a strongly seasonal one in the East. Central and East Java are characterised by drier, cooler climate (climate-data.org) and the vegetation has more grass areas than on the surrounding islands (Heaney, 1991; Mishra *et al*., 2010). Therefore, it is likely that *R. timorensis*, being better adapted to drier climate would have crossed the dry central Sundaland during Pleistocene

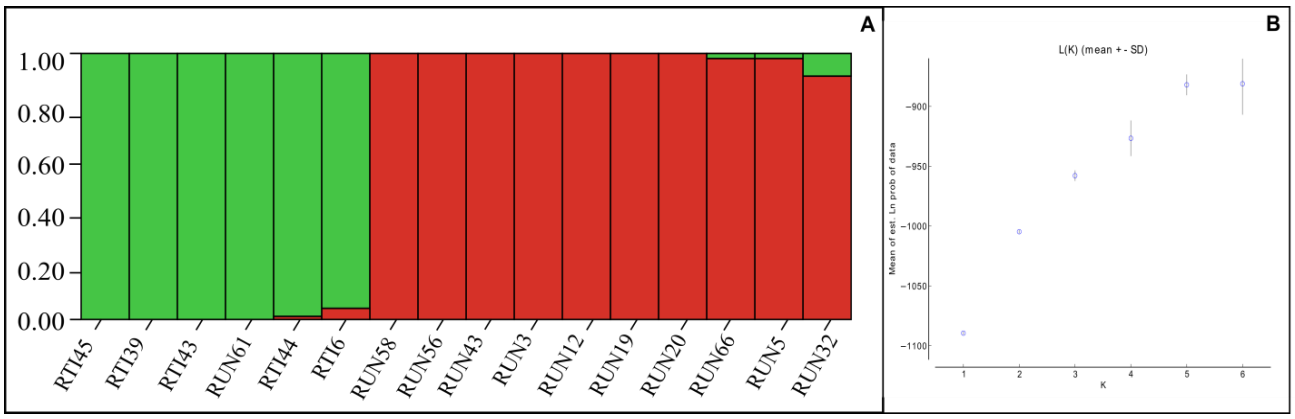

Figure 5 – Genotyping results from 16 individuals. Panel A shows a Structure plot for K = 2, with *R. timorensis* samples in green and *R. unicolor* individuals in red. Panel B shows the likelihood values (y-axis) for K values from 1 to 6, showing a peak at K =5

glacials to colonize east Java (Sheldon *et al*. 2015), where it stayed isolated from its sister species. After this initial separation, we find evidence of range expansion, likely during consequent drops in sea levels, demonstrated by the evidence of introgression in Java and possibly South Sumatra. At the time of LGM, West Java presented drier and cooler climates in the lowlands (Sun *et al*., 2000), therefore facilitating the dispersal of *R. timorensis*.

Phylogeography and taxonomy of Rusa timorensis

Within *R. timorensis*, individuals from Bali and West Java showed some genetic divergence both at mtDNA and nDNA. This substructure indicated limited gene-flow during parts of the Late Pleistocene, which might corroborate the classification of Bali populations as *R. t. renschi*, with genetic isolation most likely being the result of a 'small population effect' and a limited/interrupted gene flow to Javan populations. However, because we only had two samples from Bali and no samples from East Java, our assessment has to be viewed cautiously. The remaining Javan individuals clustered within the *R. unicolor* mtDNA clade, showing genetic variability between introgressed individuals. The genetic distances and number of mutational steps between these individuals indicate that the introgression to have occurred during the Late Pleistocene.

The mtDNA of RTI-labelled samples from islands beyond the Wallace line clustered with *R. unicolor* mtDNA but did not show any clear geographical distribution pattern. These RTI hybrids shared haplotypes with RUN-labelled samples from Sumatra and Borneo (Fig. 2 & 3). Genetic distances among RTI hybrid haplotypes and to other RUN haplotypes were very low, indicating a recent, thus human-mediated, introduction to these Wallacea islands. Quite recently it had been suggested to classify *R. timorensis* into seven subspecies according to their occurrence on islands within and outside of the Sunda shelf (Mattioli, 2011; Hedges et al., 2015). However, our data does not support such suggestion, as all samples from the introduction range shared haplotypes among themselves, indicating lack of differentiation among individuals from these Wallacean islands. Furthermore, the few samples from Java and Timor that could be genotyped showed genetic similarity, again indicating the lack of differentiation.

Interestingly, one sambar individual from South Sumatra was found to carry RTI mtDNA. This indicates the possibility that individuals of *R. timorensis* also migrated to at least South Sumatra, where they hybridised with *R. unicolor*. Such a range expansion would have been enabled by the drier and cooler climates and the emerged land corridor between Sumatra and Java during the Late Pleistocene. Because we did not obtain the genotype of this sample, more intensive sampling of South Sumatran populations would be required to conclude that these results reflect evidence of reciprocal hybridisation between two sister species of deer.

Phylogeography and taxonomy of Rusa unicolor

Sambar is currently subdivided into five subspecies: *R. u. unicolor* (India, Nepal, Bangladesh, and Sri Lanka), *R. u. brookei* (Borneo), *R. u. cambojensis* (mainland SE Asia, from South China/Hainan and Myanmar to Peninsula Malaysia), *R. u. equine* (Sumatra, Mentawai), and *R. u. swinhoei* (Taiwan) (Mattioli, 2011). The mitogenome structure recovered here, however, did not support any of the described subspecies, as it indicated gene-flow between all populations. Especially among populations of Sundaic islands we found a lack of genetic structure that would correspond to isolated islands, evidenced by the presence of individuals distributed throughout the tree topologies and haplotypic inferences. Likewise, we found evidence of at least three deep split (>1 My) sub-clades within *Rusa unicolor* haplotypes which were not in accordance with the current subspecies assignment. Sub-clade A (the oldest) comprised haplotypes from Myanmar and India, with an age of about 1.36 My; the second clade included Sundaic populations from Sumatra, Mentawai and Borneo (clade C) and was dated to be ~1.18 My old, the third one included all haplotypes from Sri Lanka (clade B) and split from the remaining populations ~1.13 Mya. These sub-clades would represent centres of sambar distribution, which would have remained in place during times of warmer and wetter conditions, which contracted sambar populations to sub-tropical refugia. From these centres, we observed waves of expansion. The branching order of these three old sub-clades indicated colonization from northern Indochina southwards to Sri Lanka and to the Sunda Shelf respectively. The 'younger' individuals within sub-clade D from India, as well as from Sumatra and Borneo appear then to be descendants from a second/third natural dispersion wave (possibly from Thailand) during glacial periods of the Pleistocene, when low sea levels again exposed the shallow Sunda shelf connecting all major islands (Voris, 2000; Bird *et al*., 2005). During glacial periods, climate was drier and cooler in tropical regions (Gorog *et al*. 2004). However, species having retained a broad ecological niche such as *Rusa unicolor*, would have been able to utilize the newly emerged habitats, resulting in the haplotypic distribution pattern we observed here. During glacial periods, Sundaland was also connected to Southeast Asian mainland, allowing secondary admixture between formerly separated populations, thus generating the patterns we observe between haplotypes from Thailand, India and Sundaland.

There is, however, some support for the recognition of Sri Lankan sambar as being distinct. This support comes both from morphological assessments (Groves & Grubb, 2011) and karyotype differences (2n = 56 in Sri Lankan sambar vs. 2n = 58 in Indian and 2n = 62 in Chinese and Malaysian sambar; Leslie Jr., 2011). Sri Lankan populations are often more genetically related to the Western Ghats than to other Indian regions. Very recently, a 40 bp insertion was detected in the control region of the mitochondrial DNA in samples from the Western Ghats (Gupta *et al*., 2015), whose presence we, however, were unable to verify due to method limitations.

Introductions past the Wallace line

It is generally accepted that the presence of Rusa deer on islands beyond Sundaland (excluding Philippines) was the result of human interference (Long, 2003; Groves & Grubb, 2011; Hedges et al., 2015). However, until now these individuals were assumed to have been pure *R. timorensis*, collected and transported for venison and as game species by humans from the islands of Java and Bali during the Holocene (Heinsohn, 2003). While the nDNA data clearly separated the two species — being highly concordant with their description from the museum collections — the mitochondrial genomes point to a more surprising pattern of past hybridisations and human introductions.

All mtDNA haplotypes of samples labelled in the museum collections as *R. timorensis* and sampled from Wallacean islands (Sulawesi, Lesser Sunda Islands and The Moluccas) were of *R. unicolor* origin, rendering these individuals hybrid descendants. Although we cannot prove that hybridisation occurred prior to translocation, the most parsimonious explanation for the molecular patterns obtained in this study is that hybridisation occurred on Java (centre of star like pattern in the haplotypic network), with natural dispersion of female sambar. These immigrated individuals were likely from Sumatra and/or the Thai-Malay Peninsula, as indicated by the basal position of the two Southern Thailand individuals (RUN51 and RUN57), and they would have used the connecting land bridges. After the introgression of Sambar haplotypes into Javan populations, humans would then have transported hybrids (and their introgressed descendants) from Java to Timor, the Moluccas and Sulawesi. Despite evidence for multiple independent introductions (see discussion below) almost all introduced individuals carried sambar haplotypes (except RUN37 from Seram, the Moluccas). This indicates that either humans selected for individuals to be introduced (e.g. carrying a particular trait only found in introgressed Javan deer); that most introductions occurred from a single region (e.g. West Java) where RUN haplotypes got fixed; or that the introgressed individuals had a higher surviving probability after their introduction.

Introduction of Javan deer to Timor seems to have occurred only once and, presumably, with very few founders because of the lack of mtDNA diversity found among all individuals. In fact, populations recently introduced to Australia and New Caledonia from a low, known number of individuals from Timor have been shown to have very low genomic diversity, which would be the expected result after an introduction of individuals that had come from an already genetically impoverished population (Webley *et al*., 2004; de Garine-Wichatitsky *et al*., 2009). Although we obtained only very few samples from the Moluccas and other Lesser Sunda Islands (Dompoe and Lombok), they did not share haplotypes, indicating either multiple introductions or a higher number of founders. Sulawesi had by far the most genetically diverse Javan deer population of all the Wallacean islands. Its haplotypes were present in almost all younger clades of the mtDNA tree. This pattern indicated that Rusa deer reached Sulawesi multiple times. One sample (RTI18) was closer related to Bornean populations than the other haplotypes from Sulawesi. Although natural dispersal from Borneo to Sulawesi over the Makassar Strait is conceivable it is highly unlikely as the last possible connection between these two land masses

was during the late Pliocene/early Pleistocene ~2.5 Mya, a date that by far predates the emergence of this mtDNA lineage. The most likely scenario is a human mediated introduction of Bornean sambar to Sulawesi where it hybridised with the introduced Javan deer. If true, this represents an earlier hybridisation event on Sulawesi (compared to the Late Pleistocene hybridisation in Java and South Sumatra) and further studies on Sulawesi Javan deer would be required to test this hypothesis. The remaining haplotypes from Sulawesi individuals were closely related to the haplotypes from individuals introduced to the Moluccas and Timor Islands, indicating that all of them have been introduced in one wave.

This is the first report of historical hybridisations between sambar (*R. unicolor*) and Javan deer (*R. timorensis*). Occurrence of such hybridisation had been assumed before, namely between *R. timorensis* individuals introduced to Borneo with the local Bornean *R. unicolor* (West Kalimantan, now possibly extinct, Hedges *et al*., 2015) and attained through husbandry before (Leslie Jr., 2011). Hybridisations with fertile offspring have also been reported to occur between other deer species, among others between sambar and red deer *Cervus elaphus* (Muir *et al*.,1997) and red deer and Sika *Cervus nippon* (Smith *et al*., 2014).

This fact has potentially important conservation implications for the two Rusa species analysed in this study. Despite being one of the most widespread deer species in southern Asia, *R. unicolor* is no longer abundant throughout most of its native range (Timmins *et al*., 2015). Likewise, *R. timorensis* is currently considered as pest species in areas where it has recently been introduced (e.g. Australia) but has, however, decreased largely in population numbers in native and historical introduction regions (Java, Lesser Sunda Islands, Sulawesi and the Moluccas). Both Rusa species studied here are now considered as Vulnerable by the IUCN/Red List of Threatened Species (Hedges *et al*. 2015; Timmins *et al*., 2015). Therefore, genetic monitoring of individuals, both at mtDNA and nuclear genomes is necessary to access if pure RTI and RUN individuals are being introduced (or reproductively assisted in their native ranges) and not introgressed individuals. Moreover, more intensive and extensive sampling of *R. timorensis* on their native range is necessary to discern if pure RTI populations still remain in Java and Bali or if they are composed in their majority by hybrid individuals.

Conclusion

In addition to representing the first comprehensive phylogeographical study on *R. unicolor* and *R. timorensis*, this study revealed surprising evolutionary histories of these two sister species. While climate adaptations were likely responsible for maintaining species monophyly, Pleistocene climate changes were responsible for secondary contact and consequent hybridisation between sambar and Javan deer. We recovered a pattern of (likely reciprocal) introgressions between the two species, facilitated by the presence of land corridors during periods of low sea levels in Sundaland. The introgressed populations of Javan deer on Java were then the source of all human-mediated introductions to the islands east of the Wallace line, as we found that all *R. timorensis* individuals

carried *R. unicolor* haplotypes. Our results thus constitute evidence for one of the earliest hybrid translocation events in human history.

References

Bacon, C.D., Henderson, A.J., Mckenna, M.J., Milroy, A.M. & Simmons, M.P. (2013) Geographic and Taxonomic disparities in species diversity: dispersal and diversification rates across Wallace's line. *Evolution*, **67**, 2058–2071.

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

Belkhir, K., Borsa, P., Chikki, L., Raufaste, N. & Bonhomme, F. (1996-2004) Genetix 4.05, logiciel sous Windows TM pour la genetique des populations. *CNRS Umr* 5000.

Van Bemmel, A.C.V. (1949) Revision of the Rusine Deer. *Treubia*, **20**.

Van Den Bergh, G.D., De Vos, J. & Sondaar, P.Y. (2001) The Late Quaternary palaeogeography of mammal evolution in the Indonesian Archipelago. *Palaeogeography, Palaeoclimatology and Palaeoecology*, **171**, 385–408.

Bibi, F. (2013) A multi-calibrated mitochondrial phylogeny of extant Bovidae (Artiodactyla, Ruminantia) and the importance of the fossil record to systematics. *BMC Evolutionary Biology*, **13**, 166.

Bird, M.I., Taylor, D. & Hunt, C. (2005) Palaeoenvironments of insular Southeast Asia during the Last Glacial Period: a savanna corridor in Sundaland? *Quaternary Scientific Reviews*, **24**, 2228–2242.

Chapuis, M.P. & Estoup, A. (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution*, **24**, 621–631.

Darriba, D., Taboada, G.L., Doallo, R. & Posada, D. (2015) jModelTest 2 : more models, new heuristics and high-performance computing. *Nature Methods*, **9**, 6–9.

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

Earl, D.A. & von Holdt, B.M. (2012) STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetic Resources*, **4**, 359–361.

Emerson, B.C. & Tate, M.L. (1993) Genetic Analysis of Evolutionary Relationships Among Deer (Subfamily Cervinae). *Journal of Heredity*, **84**, 266–273.

Escobedo-Morales, L.A., Mandujano, S., Eguiarte, L.E., Rodriguez-Rodriguez, M.A. & Maldonado, J.E. (2016) First phylogenetic analysis of Mesoamerican brocket deer *Mazama pandora* and *Mazama temama* (Cetartiodactyla: Cervidae) based on mitochondrial sequences: Implications on Neotropical deer evolution. *Mammalian Biology*, **81**, 303–313.

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

Excoffier, L., Laval, G. & Schneider, S. (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.

Fortes, G.G. & Paijmans, J.L.A. (2015) Analysis of whole mitogenomes from ancient samples. *Whole Genome Amplification Methods Protocols*, 179–195.

de Garine-Wichatitsky, M., de Meeus, T., Chevillon, C., Berthier, D., Barre, N., Thevenon, S. & Maillard, J.-C. (2009) Population genetic structure of wild and farmed rusa deer (*Cervus timorensis russa*) in New-Caledonia inferred from polymorphic microsatellite loci. *Genetica*, **137**, 313–323.

Geist, V. (1998) Deer of the world. Their evolution, behaviour, and ecology. Stackpole Books, Mechanicsburg.

Gorog, A.J., Sinaga, M.H. & Engstrom, M.D. (2004) Vicariance or dispersal? Historical biogeography of three Sunda shelf marine rodents (*Maxomys surifer*, *Leopoldamys sabanus* and *Maxomys whiteheadi*). *Biological Journal of the Linnean Society*, **81**, 91–109.

Groves, C.P. (1983) Pigs East of the Wallace Line. *Journal de la Societe des Oceanistes*, **39**, 105–119.

Groves, C.P., Grubb, P. (2011) Ungulate Taxonomy. The Jonhs Hopkins University Press.

Gupta, S.K., Kumar, A., Gaur, A. & Hussain, S.A. (2015) Detection of 40 bp insertion‐deletion (INDEL) in mitochondrial control region among sambar (*Rusa unicolor*) populations in India. *BMC Research Notes*, 1–7.

Heany, L.R. (1991) A synopsis of climate and vegetational change in Southeast Asia. *Climatic Change*, **19**, 53–61.

Hedges, S., Duckworth, J.W., Timmins, R.J., Semiadi, G. & Dryden, G. (2015) *Rusa timorensis*. IUCN Red List Threatened Species.

Heinsohn, T. (2003) Animal translocation: Long-term human influences on the vertebrate zoogeography of Australasia (natural dispersal versus ethnophoresy). *Australian Zoologist*, **32**, 351–376.

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.

Leonard, J.A., den Tex, R.-J., Hawkins, M.T.R., Munoz-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.

Leslie, Jr., D.M. (2011) *Rusa unicolor* (Artiodactyla: Cervidae). *Mammalian Species*, **43**, 1–30.

Li, H. & Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.

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

Long, J.L. (2003) Introduced mammals of the world. Their history, distribution and influence. CSIRO Publishing, Australia.

Maricic, T., Whitten, M. & Pääbo, S. (2010) Multiplexed DNA Sequence Capture of Mitochondrial Genomes Using PCR Products. *PLoS One*, **5**, e14004.
Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, **17**, 10.
Martins, R.F., Fickel, J., Le, M., van Nguyen, T., Nguyen, H.M., Timmins, R., Gan, H.M., Rovie-Ryan, J.J., Lenz, D., Förster, D.W. & Wilting, A. (2017) Phylogeography of red muntjacs reveals three distinct mitochondrial lineages. *BMC Evolutionary Biology*, **17**, 34.
Mattioli, S. (2011) Cervidae in Handbook of the Mammals of the World - Volume 2 Hoofed Mammals (D. Wilson & R. Mittermeier, Eds.), Lynx Edicions, Barcelona, 350–443.
McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 1297–1303.
Meijaard, E. & Groves, C.P. (2004) Morphometrical relationships between South-east Asian deer (Cervidae, tribe Cervini): evolutionary and biogeographic implications. *Journal of the Zoological Society of London*, **263**, 179–196.
Mishra, S., Gaillard, C., Hertler, C., Moigne, A.M. & Simanjuntak, T. (2010) India and Java: Contrasting records, intimate connections. *Quaternary International*, **223–224**, 265–270.
Muir, P., Semiadi, G., Asher, G.W., Broad, T.E., Tate, M.L. & Barry, T.N. (1997) Sambar Deer (C*ervus unicolor*) x Red Deer (*C. elaphus*) interspecies Hybrids. *Journal of Heredity*, **88**, 366–372.
Patou, M.-L., Wilting, A., Gaubert, P., Esselstyn, J.A., Cruaud, C., Jennings, A.P., Fickel, J. & Veron, G. (2010) Evolutionary history of the Paradoxurus palm civets - a new model for Asian biogeography. *Journal of Biogeography*, **37**, 2077-2097.
Pitra, C., Fickel, J., Meijaard, E. & Groves, C. (2004) Evolution and phylogeny of old world deer. *Molecular Phylogenetics and Evolution*, **33**, 880–895.
Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
Ronquist, F & Huelsenbeck, J.P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.
Sathiamurthy, E. & Voris, K.H. (2006) Maps of Holocene Sea Level Transgression and Submerged Lakes on the Sunda Shelf. *The Natural History Journal of Chulalongkorn University*, **Suppl. 2**, 1-43.
Sheldon, F.H., Lim, H.C. & Moyle, R.G. (2015) Return to the Malay Archipelago: the biogeography of Sundaic rainforest birds. *Journal of Ornithology*, **156**, 90–113.
Smith, S.L., Carden, R.F., Coad, B., Birkitt, T. & Pemberton, J.M. (2014) A survey of the hybridisation status of Cervus deer species on the island of Ireland. *Conservation Genetics*, **15**, 823–835.
Silvestro, D. & Michalak, I. (2012) RaxmlGUI: A graphical front-end for RAxML. *Organisms Diversity & Evolution*, **12**, 335–337.
Di Stefano, G. & Petronio, C. (2002) Systematics and evolution of the Eurasian Plio-Pleistocene tribe Cervini (Artiodactyla, Mammalia). *Geologica Romana*, **36**, 311–334.
Sun, X., Li, X., Luo, Y. & Chen, X. (2000) The vegetation and climate at the last glaciation on the emerged continental shelf of the South China Sea. *Palaeogeography, Palaeoclimatology and Palaeoecolog*y, **160**, 301–316.
Timmins, R.J., Kawanishi, K., Giman, B., Lynam, A., Chan, B., Steinmetz, R., Sagar Baral, H. & Samba Kumar, N. (2015) *Rusa unicolor*. IUCN Red List Threatened Species, e.T41790A85628124.
Veron, G., Willsch, M., Dacosta, V., Patou, M.-L., Seymour, A., Bonillo, C., Couloux, A., Te Wong, S., Jennings, A.P., Fickel, J. & Wilting, A. (2014) The distribution of the Malay civet *Viverra tangalunga* (Carnivora: Viverridae) across Southeast Asia: Natural or human-mediated dispersal? *Zoological Journal of the Linnean Society*, **170**, 917–932.
Voris, H.K. (2000) Maps of Pleistocene sea levels in Southeast Asia: Shorelines, river systems and time durations. *Journal of Biogeography*, **27**, 1153–1167.
Webley, L.S., Zenger, K.R., English, A.W. & Cooper, D.W. (2004) Low levels of genetic variation within introduced Javan rusa deer (*Cervus timorensis russa*) in Australia. *European Journal of Wildlife Research*, **50**, 137–140.
Zazo, C. (1999) Interglacial sea levels. *Quaternary International*, **55**, 101–113.

Biosketch

Renata F. Martins focuses her research on biogeography of Southeast Asia, in particular of Sundaland. This work constitutes one component of the PhD work at the IZW, Berlin on the phylogeography patterns of several ungulate species from Southeast Asia.

Author contributions: RFM, AW and JF designed the study; RFM and AS performed the laboratory procedures; RFM and DL performed the bioinformatic analyses; RFM, AW and JF wrote the manuscript. All authors contributed equally for the final version of this manuscript.

Human mediated introduction of introgressed deer across Wallace's line: historical biogeography of *Rusa unicolor* and *R. timorensis*

Renata Martins, Anke Schmidt, Dorina Lenz, Andreas Wilting and Jörns Fickel

Supplemental Material

Table S1 – Complete dataset details. Samples are indicated according to the species assignment from the museum collections. mtDNA haplotype and genotypic cluster assignment is provided according to the results obtained.

Species	Sample ID	Origin	MtDNA haplotype	Accession Number	nDNA assignment	Museum Collection	Museum ID	Curator
Rusa unicolor	RUN2	Sumatra	H_13	MF176993		Berlin NHM	75150	F. Mayer
Rusa unicolor	RUN3	Sri Lanka	H_14	MF177018	RUN	Berlin NHM	75133	F. Mayer
Rusa unicolor	RUN5	North Borneo	H_37	MF177019	RUN	Berlin NHM	11259	F. Mayer
Rusa unicolor	RUN6	North Borneo	H_38	MF177020		Berlin NHM	11261	F. Mayer
Rusa unicolor	RUN10	Sri Lanka	H_14	MF176994		Berlin NHM	12368	F. Mayer
Rusa unicolor	RUN12	Sarawak, Borneo	H_15	MF176995	RUN	Berlin NHM	14702	F. Mayer
Rusa unicolor	RUN19	North Borneo	H_37	MF177021	RUN	Berlin NHM	103292	F. Mayer
Rusa unicolor	RUN20	Mentawai	H_16	MF176996	RUN	Naturalis	139-50	P. Kamminga
Rusa unicolor	RUN21	Borneo	H_37	MF177022		Stuttgart NHM	16900	S. Merker
Rusa unicolor	RUN26	Borneo	H_39	MF177023		Stuttgart NHM	15900	S. Merker
Rusa unicolor	RUN28	Thailand	H_17	MF176997		Bonn NHM	ZFMK_MAM_2013.618	R. Hutterer
Rusa unicolor	RUN32	Sumatra	H_18	MF176998	RUN	Vienna NHM	7333	F. Zachos
Rusa unicolor	RUN33	India	H_19	MF176999		Vienna NHM	40867	F. Zachos
Rusa unicolor	RUN34	Java	H_20	MF177000		Seckenberg Frankfurt	15937	I. Ruf
Rusa unicolor	RUN35	Sri Lanka	H_21	MF177001		Paris NHM	1877-10	G. Veron
Rusa unicolor	RUN37	Burum, Moluccas	H_22	MF279250		Munich NHM	1906/3021	M. Hiermeyer
Rusa unicolor	RUN38	Sumatra	H_23	MF177002		Munich NHM	1905/46	M. Hiermeyer
Rusa unicolor	RUN39	Timor	H_3	MF177003		Munich NHM	1911/2147	M. Hiermeyer
Rusa unicolor	RUN41	South Sumatra	H_24	MF279251		Munich NHM	1908/465	M. Hiermeyer
Rusa unicolor	RUN42	Sumatra	H_25	MF177004		Munich NHM	1908/466	M. Hiermeyer
Rusa unicolor	RUN43	Indochina	H_26	MF177005	RUN	Munich NHM	1962/235	M. Hiermeyer

Rusa unicolor	RUN44	South Thailand	H_27	MF177006		Copenhagen NHM	M1594	D.K. Johansson
Rusa unicolor	RUN45	East Sumatra	H_40	MF177024		Copenhagen NHM	M1650	D.K. Johansson
Rusa unicolor	RUN47	South Thailand	H_41	MF177025		Copenhagen NHM	M1593	D.K. Johansson
Rusa unicolor	RUN51	South Thailand	H_42	MF177026		Copenhagen NHM	M1583	D.K. Johansson
Rusa unicolor	RUN54	Bhootan Dooars, India	H_43	MF177027		Copenhagen NHM	M534	D.K. Johansson
Rusa unicolor	RUN55	Bhootan Dooars, India	H_28	MF177007		Copenhagen NHM	M533	D.K. Johansson
Rusa unicolor	RUN56	Sri Lanka	H_44	MF177028	RUN	Copenhagen NHM	M1236	D.K. Johansson
Rusa unicolor	RUN58	South Myanmar	H_29	MF177008	RUN	Naturalis	RMNH.MAM.693.a	P. Kamminga
Rusa unicolor	RUN61	Central Java	H_3	MF177009	RTI	Naturalis	RMNH.MAM.33832	P. Kamminga
Rusa unicolor	RUN62	Bengala, India	H_30	MF177010		Naturalis	RMNH.MAM.51460	P. Kamminga
Rusa unicolor	RUN65	South Sumatra	H_45	MF177029		Naturalis	RMNH.MAM.51453	P. Kamminga
Rusa unicolor	RUN66	West Sumatra	H_46	MF177030	RUN	Naturalis	RMNH.MAM.1033.b	P. Kamminga
Rusa timorensis	RTI2	Lesser Sunda Islands	H_31	MF177011		Berlin NHM	92305	F. Mayer
Rusa timorensis	RTI3	Moluccas	H_1	MF176981		Berlin NHM	30840	F. Mayer
Rusa timorensis	RTI6	Timor	H_2	MF176982	RTI	Stuttgart NHM	15878	S. Merker
Rusa timorensis	RTI8	Java	H_3	MF176983		Stuttgart NHM	15892	S. Merker
Rusa timorensis	RTI9	Java	H_4	MF176984		Stuttgart NHM	15897	S. Merker
Rusa timorensis	RTI13	Java	H_1	MF176985		Stuttgart NHM	15891	S. Merker
Rusa timorensis	RTI17	South Sulawesi	H_5	MF176986		Bonn NHM	58.109	R. Hutterer
Rusa timorensis	RTI18	South Sulawesi	H_32	MF177012		Bonn NHM	58.75	R. Hutterer
Rusa timorensis	RTI19	South Sulawesi	H_33	MF177013		Bonn NHM	58.105	R. Hutterer
Rusa timorensis	RTI20	South Sulawesi	H_6	MF176987		Bonn NHM	58.112	R. Hutterer
Rusa timorensis	RTI21	Timor	H_3	MF177014		Vienna NHM	231	F. Zachos
Rusa timorensis	RTI22	Lesser Sunda Islands	H_34	MF177015		Vienna NHM	2049	F. Zachos
Rusa timorensis	RTI27	Sulawesi	H_35	MF177016		Vienna NHM	7334	F. Zachos
Rusa timorensis	RTI30	Java	H_7	MF176988		Dresden NHM	B730	C. Stefen
Rusa timorensis	RTI31	Sulawesi	H_3	MF176989		Dresden NHM	B2686	C. Stefen
Rusa timorensis	RTI32	Java	H_8	MF176990		Senckenberg Frankfurt	15462	I. Ruf
Rusa timorensis	RTI33	Moluccas	H_9	MF176991		Senckenberg Frankfurt	5562	I. Ruf
Rusa timorensis	RTI37	West Java	H_10	MF279247		Naturalis	RMNH.MAM.5679	P. Kamminga

Rusa timorensis	RTI38	West Java	H_11	MF279248		Naturalis	RMNH.MAM.5680	P. Kamminga
Rusa timorensis	RTI39	Timor	H_3	MF176992	RTI	Naturalis	RMNH.MAM.51419	P. Kamminga
Rusa timorensis	RTI43	Timor	H_3	MF177017	RTI	Naturalis	RMNH.MAM.51417	P. Kamminga
Rusa timorensis	RTI44	Bali	H_12	MF279249	RTI	Naturalis	RMNH.MAM.33828	P. Kamminga
Rusa timorensis	RTI45	Bali	H_36	MF279252	RTI	Naturalis	RMNH.MAM.33827	P. Kamminga

Table S2 – Microsatellite loci details and results. Allelic range, number of alleles (N_a), expected and observed heterozygosity (H_E and H_O) and inbreeding coefficient (F_{IS}) is given for each loci, as calculated for the 16 individuals genotyped.

Locus	Allelic Range	N_a	H_E	H_O	F_{IS} (W&C)	Reference
Haut14	116-130	6	0.66	0.25	0.628	Kühn et al., 1996
VH110	101-153	12	0.89	0.44	0.516	Talbot et al., 1996
NVHRT48	71-120	12	0.8	0.31	0.615	Røed & Midthjell, 1998
BM757	167-199	12	0.92	0.5	0.424	Slate et al., 1998
CSSM39	158-200	12	0.87	0.37	0.575	Slate et al., 1998
CSSM41	120-146	11	0.85	0.37	0.566	Slate et al., 1998
FSHB	171-206	13	0.92	0.5	0.426	Slate et al., 1998
Roe09	167-199	12	0.9	0.63	0.31	Fickel & Reinsch, 2000
CSSM14	131-161	10	0.82	0.13	0.851	Kühn et al., 2003
T115	139-184	9	0.81	0.44	0.467	Meredith et al., 2005
C143	154-174	4	0.71	0.19	0.741	Meredith et al., 2005
C180	138-160	6	0.67	0.37	0.448	Meredith et al., 2005
INRA6	103-141	10	0.78	0.37	0.529	Senn & Pemberton, 2009
Mu_4D	111-113	2	0.23	0	1.000	Schröder et al., 2016
Mu_1_51	114-152	16	0.95	0.75	0.216	Schröder et al., 2016
C183	119-133	5	0.65	0.25	0.62	Schröder et al., 2016
Mu_1_25	98-110	6	0.78	0.37	0.529	Schröder et al., 2016
Mu_1_550	139-173	14	0.94	0.37	0.608	Schröder et al., 2016

References

Fickel, J. & Reinsch, A. (2000) Microsatellite markers for the European Roe deer (*Capreolus capreolus*). *Molecular Ecology,* **9**, 993-1011.

Kühn, R., Anastassiadis, C. & Pirchner, F. (1996) Transfer of bovine microsatellites to the cervine (*Cervus elaphus*). *Animal Genetics,* **27**, 199-201.

Kühn, R., Schröder, W., Pirchner, F. & Rottmann, O. (2003) Genetic diversity, gene flow and drift in Bavarian red deer populations (*Cervus elaphus*). *Conervation Genetics,* **4**, 157-166.

Meredith, E.P., Rodzen, J.A., Levine, K.F. & Banks, J.D. (2005) Characterization of an additional 14 microsatellite loci in California Elk (*Cervus elaphus*) for use in forensic and population applications. *Conservation Genetics,* **6**, 151-153.

Røed, K.H. & Midthjell, L. (1998) Microsatellites in reindeer, *Rangifer tarandus*, and their use in other cervids. *Molecular Ecology,* **7**, 1771-1788.

Schröder, O., Lieckfeldt, D., Lutz, W., Rudloff, C., Frölich, K. & Ludwig, A. (2016) Limited hybridisation between domestic and the European mouflon in Western Germany. *European Journal of Wildlife Research,* **62**, 307-314.

Senn, H.V. & Pemberton, J.M. (2009) Variable extent of hybridization between invasive sika (*Cervus nippon*) and native red deer (*C. elaphus*) in a small geographical area. *Molecular Ecology,* **18**, 862-876.

Slate, J., Coltman, D.W., Goodman, S.J., MacLean, I., Pemberton, J.M. & Williams, J.L. (1998) Bovine microstellite loci are highly conserved in red deer (*Cervus elaphus*), sika deer (*Cervus nippon*) and Soay sheep (*Ovis aries*). *Animal Genetics,* **29**, 307-315.

Manuscript III

Historical phylogeography of Asian rhinos

The lost genetic legacy of Sumatran and Javan rhinoceros: phylogeography and genetic diversity of historical populations

Renata F. Martins[1]*, Dorina Lenz[1], Daniel Förster[1], Johanna von Seth[2,3], Love Dalén[2], Sen Nathan[4], Benoît Goossens[5], Peter van Coeverden de Groot[6], Jörns Fickel[1] and Andreas Wilting[1]

[1] Department of Evolutionary Genetics, Leibniz Institute for Zoo and Wildlife Research (IZW), Alfred-Kowalke-Straße 17, 10315 Berlin, Germany

[2] Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Box 50007, SE-104 05 Stockholm, Sweden

[3] Department of Zoology, Stockholm University, SE-10691 Stockholm, Sweden

[4] Sabah Wildlife Department, 88000, Kota Kinabalu, Sabah, Malaysia

[5] Organisms and Environment Division, Cardiff School of Biosciences, Cardiff University, Cardiff, UK.

[6] Department of Biology, Queens University, 116 Barrie Street, Kingston, Ontario, Canada

* Corresponding author: Renata F. Martins, martins@izw-berlin.de

Manuscript formatted for submission

The lost genetic legacy of Sumatran and Javan rhinoceros: phylogeography and genetic diversity of historical populations

Abstract Javan and Sumatran rhinoceros are two of the most charismatic and the most endangered mega-herbivores on earth. Once widespread throughout Southeast Asia, both species now occur only in small isolated populations. The few molecular studies on these species revealed very low levels of genetic diversity, yet were focused only on contemporary populations. Here we investigated the historical phylogeography of the two species, by including samples from their former distribution ranges, analysed for their full mitogenome. Our aim was to understand if historical genetic diversity had always been low as indicated by modern population studies or if, instead, the collapse of effective population sizes also led to a collapse in genetic diversity of both species. Additionally, by including historical populations we also wanted to evaluate the subspecific taxonomy of Javan and Sumatran rhinos. We analysed the complete mitogenome of archival samples and dated the splits between genetic lineages. Our findings indicated that the climatic events of the Pleistocene influenced the distribution of genetic divergence of both species, revealing that each species' ecological needs likely led to different speciation events. Divergence estimates indicate colonization of Sundaland during a) a glacial period, when sea levels were low and the habitat of central Sundaland likely was dry; and b) interglacial period around 1 Mya when rainforests were present throughout Southeast Asia. These results are consistent with the Javan rhino being able to disperse through vegetation zones of the available land corridors, while the Sumatran rhino would likely confined to montane forest refugia in times of low sea level. More importantly, our results also revealed a much higher genetic diversity within both species than ever reported, proving that with the decline of these species also unique genetic lineages, equalling subspecies were lost.

Introduction

Compared with their African counterparts, Asian rhinoceros have been much less studied, both in ecological and genetic terms. Sumatran and Javan rhinos are two of the most iconic, yet most endangered large mammals in the world, with less than a hundred individuals left in small patches of their former distribution. Their historical distribution spanned a significant proportion of Southeast Asia (Fig. 1). The Sumatran rhino was distributed from Bhutan and North-East India to Southern China, through Indochina and Peninsular Malaysia and onto the islands of Borneo and Sumatra, while the Javan rhino was distributed in India and Bangladesh, Indochina and possibly Southern China, throughout Peninsular Malaysia to Sumatra and Java. The extreme distribution and population decline of Sumatran and Javan rhinos resulted from a combined impact of overhunting and habitat loss (Groves & Leslie, Jr., 2011; Havmøller et al. 2016). After the Sumatran rhino has

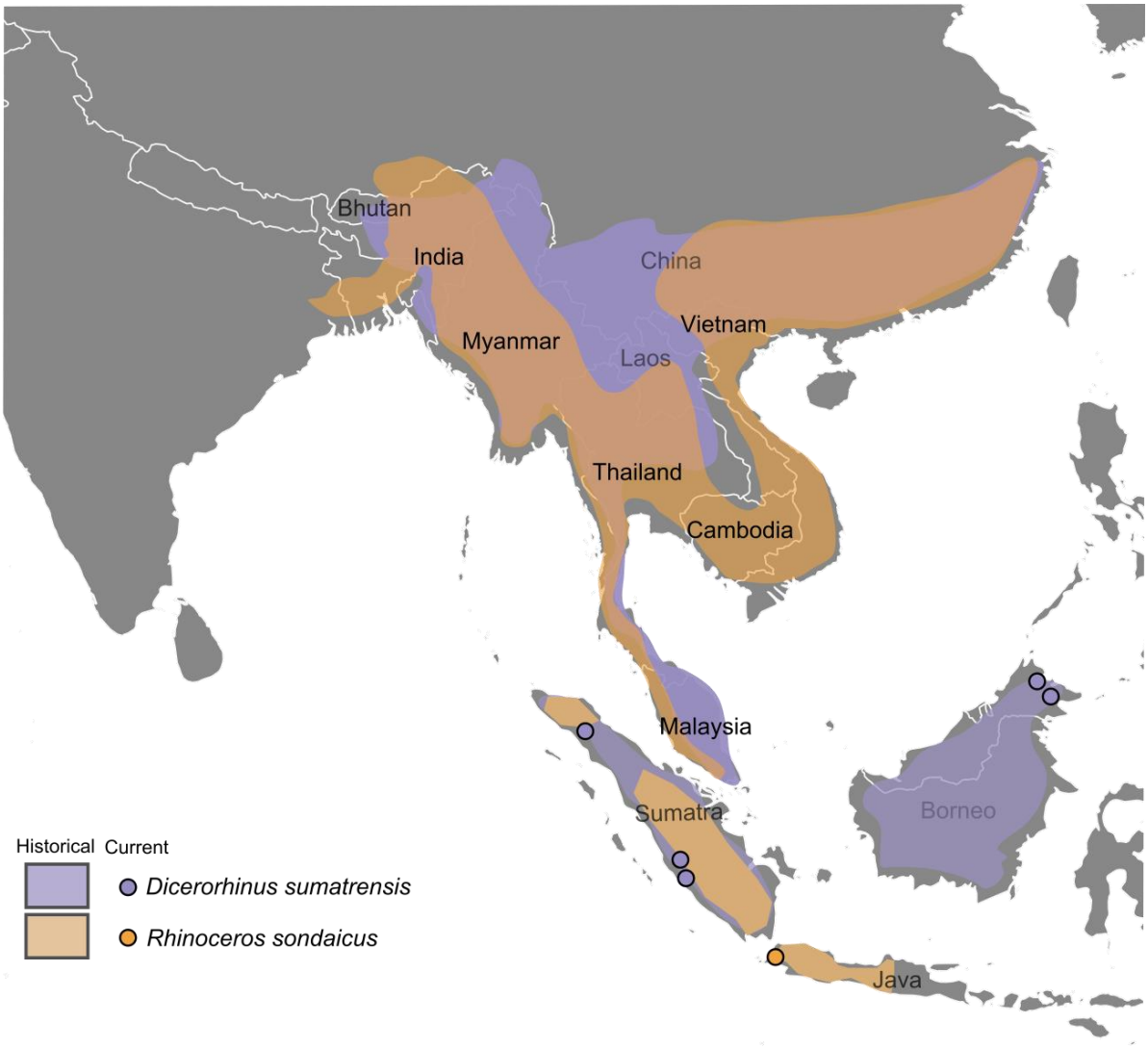

Figure 1. Map of the historical and current distribution of Sumatran (purple) and Javan (orange) rhinos.

been declared to be extinct in Sabah, Malaysia Borneo recently (Lee, 2015) it most likely only occurs with a very small population in Kalimantan, Borneo and in three reserves in Sumatra (Kretzschmar et al. 2016). Javan rhinos are extinct in all of their former distribution, except for a small population in Ujung Kulon, in west Java. Until recently it also survived in Cat Tien National park (Vietnam), where possibly the last individual was shot for its valuable horn in 2010 (Brook et al. 2012).

There are surprisingly few molecular studies on these species, most likely due to the extinctions of populations before the use of molecular techniques became a standard procedure in conservation planning. The few studies that have specifically looked at genetic diversity within contemporaneous Sumatran and Javan rhino populations and were, therefore, hampered by low sample numbers and exceedingly low geographic coverage; it is thus unsurprising that they reported very low genetic diversity for both species (Sumatran rhino: Amato et al. 1995, Morales et al. 1997; Javan rhino:

Fernando et al. 2006). The dramatic decline of rhinos in Southeast Asia also prevented in depth ecological studies, resulting as well in poor knowledge about their ecology. However, from what it is known, both species seem to depend on the occurrence of forested habitats. Javan Rhino has been reported mostly from lowland tropical rainforest, often in the vicinity of water sources (van Strien et al. 2008a). Sumatran rhinos were found both in lowland primary and secondary tropical rainforest and at higher altitudes in moss montane forest and are dependent on salt licks (van Strien et al. 2008b). Although these species' historical distribution ranges overlapped, it is unlikely that both Sumatran and Javan rhinos evolved in sympatry. Their specific ecological requirements might have led to different evolutionary histories, especially in response to the dramatic Pleistocene events that greatly shaped land availability and vegetational distribution throughout South and Southeast Asia (Cannon et al. 2009, Lohman et al. 2011). Land corridors, which cyclically connected the land masses of the Sundaland during glacial periods, allowed species to disperse and re-establish gene-flow between populations (e.g. Malay civet, Veron et al. 2014; red muntjacs, Martins et al. 2017). However, for forest dependent species newly available habitat may not have allowed for expansion, as this is believed to have been composed mainly of dry savannah corridors (Bird et al. 2005, Wurster et al. 2010).

Here, we investigated the historical phylogeography of both species using complete mitochondrial genomes (mitogenomes) from archival samples collected across their historical ranges. Our main aim was to address the amount of genetic diversity which was lost through population extinction, but also to clarify the distribution of lineages within/among their subspecies.

Material and Methods

Samples and DNA extraction

Archival samples from 36 Javan (RSO) and 28 Sumatran rhinos (DSU) were collected at several natural history museums (for final dataset see Table S1) and only samples with a known locality were included in the analysis. All molecular procedures on archival samples were performed in a separate facility, dedicated to work with archival material and operated under controlled conditions in order to minimize any potential cross contamination. DNA extraction followed the Qiagen DNeasy Blood & Tissue kit (Qiagen, Hilden, Germany) protocol with modifications (Martins *et al.* 2017). One fresh sample was collected from the semi-captive Sumatran rhinoceros 'Tam', which was processed in a separate facility.

Library building and Hybridization capture

All samples, including extraction negative controls, were processed into Illumina sequencing libraries with single 8 nt indexing (Fortes & Paijmans, 2015). Amplification cycles were optimized for each sample to ensure minimum PCR-bias and to minimize read clonality. The extracted DNA

proved to be highly degraded, as expected from archival samples (Paijmans et al. 2012). Therefore, we performed a hybridization capture step prior to sequencing, in order to increase the relative amount of endogenous mitochondrial DNA (mtDNA) in the samples. To capture target mtDNA, we created baits from a fresh sample of Indian rhinoceros (*Rhinoceros unicornis*) from IZW's genome resource bank (www.izw-berlin.de/genomic-tissue-bank.html). This was done by amplifying *R. unicornis* complete mitogenome via long-range PCR (for primer sequences and PCR conditions see Table S2). Preparation of baits and hybridisation capture conditions followed Maricic *et al.* (2010). Libraries were then pooled equimolarly and sequenced with either the Illumina MiSeq v3 150-cycle kit or Illumina NextSeq 500/550 high throughput kit, following manufacturers' protocols. DNA from the fresh sample was extracted using a Kingfisher Duo robot, and was subsequently converted into 180bp and 650bp insert libraries using the TruSeq DNA PCR-free Library Prep Kit. The resulting libraries were sequenced on one lane each on the HiSeq 2500 platform at the National Genomics Infrastructure in Stockholm, using the 2x125 bp setup in HighOutput mode.

Bioinformatic analyses

Sequence reads were sorted according to their respective indexes with the Illumina software bcl2fastq v2.17 (Illumina, San Diego, CA, USA) and adapters were clipped from the sequence reads with the software cutadapt v1.3 (Martin, 2011). All reads shorter than 20 bp were removed from the analyses and reads were trimmed using a sliding window approach (10 bp; Q20). The complete mitogenome sequences of *D. sumatrensis* and of *R. sondaicus* (respective NCBI Accession Nos. NC_012684 and NC_012683) were used as mapping references. Mapping was performed in two phases with the software BWA (Li & Durbin, 2009). Duplicate reads were removed with the software MarkDuplicates from picard-tools v.1.106 (http://picard.sourceforge.net/). The consensus sequence resulting from this first mapping phase was then used as reference for a second round of mapping, to increase the number of mapped sequences and hence improve sequence coverage. We used bcftools v.1.2 to call InDels (http://github.com/samtools/bcftools) and variant calling was carried out with GATK v.1.6 (McKenna *et al.*, 2010). Positions with less than 3 unique sequences and heterozygous positions were N-masked. Variants were called based on the majority rule. Only sequences with 70% or more of the genome covered with 3× or greater depth were accepted for analysis (final sequencing statistics in Table S3).

Genetic diversity, divergence dating and population demography

Mitogenome sequences obtained for the two species were aligned separately using MAFFT v.7.245 (Katoh & Standley, 2013) with the auto setting. Haplotype diversity and nucleotide diversity for each species were calculated in DNASP v.5.10 (Librado & Rozas, 2009). We constructed median joining (MJ) networks with the software NETWORK v.4.6.1.4 (Bandelt et al. 1999) for each species,

with invariant and non-informative sites removed. Phylogenetic relationships between all sampled individuals of both species were assessed through Maximum Likelihood (ML) implemented in RAxML GUI v.1.5 (Silvestro and Michalak, 2012), applying the substitution model determined by jModelTest v.2.1.7. (GTR+G+I; Darriba *et al.*, 2015). Analyses of molecular variance (AMOVA) were calculated with Arlequin v.3.5.1.2 (Excoffier *et al.*, 2005). Populations were defined based on the results from the MJ and ML analyses. We tested three groups within Javan rhinos: India, Indochina and Sunda (Sumatra and Java); and two groups of Sumatran rhinos: Mainland (Laos and Myanmar) and Sunda (Sumatra and Borneo).

Phylogeographic and phylogenetic analyses revealed the presence of distinct lineages within both target species. To estimate the divergence times of different clades within each of the species, we inferred genealogies using a Yule speciation coalescence method as implemented in BEAST v.1.8.1 (Drummond *et al.*, 2012). We performed this analysis by creating a mitogenome dataset with all extant rhino species (*Rhinoceros sondaicus*: NC_012683; *Dicerorhinus sumatrensis*: NC_012684; *Rhinoceros unicornis*: X97336.1; *Ceratothrium simum*: Y07726.1; *Diceros bicornis*: FJ905814.1), one extinct rhino species (woolly rhinoceros, *Coelodonta antiquitatis*: FJ905813.1) and one sequence from the Malayan tapir (*Tapirus indicus:* AJ428947.1) as an outgroup. We also included the mitogenome of one sequenced individual from each of the clades of Sumatran and the Javan rhinoceros. We used three independent fossil calibration points to estimate the clock rate and divergence times: 1) a root age (t_{MRCA} of Rhinos and Tapirs) was set to 50 ± 2 million years ago (Mya) (see Tougard et al. 2001); 2) an age range of 26-13 Mya for the split between Dicerorhinus and the subtribe Rhinocerontina; 3) the origin of subtribe Dicerontina (Black and White rhinoceros) at 13 Mya. We subsequently used the estimated clock rate (0.003 substitutions/site/Myear) as a tree prior for the calibrations within both Sumatran and Javan rhinoceros datasets and estimated the root height. We performed two parallel MCMC runs of 1×10^7 generations for each analysis and checked for parameter convergence and ESS values>200 with the software Tracer v.1.6 (implemented in BEAST package). Runs and trees were combined with the software Logcombiner v.1.8.1 (BEAST package) and all tree topologies were visualized and edited with the software Figtree v.1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

Results

The dataset for Javan rhinos (*Rhinoceros sondaicus*, RSO) included 15 individuals with known locality of origin and consisted of sequences of 16.424 bp length. The dataset for the Sumatran rhinos (*Dicerorhinus sumatrensis*, DSU) included 12 individuals with known origin, with sequences of 16.471 bp length. We identified 11 distinct haplotypes within *R. sondaicus* (RSO)

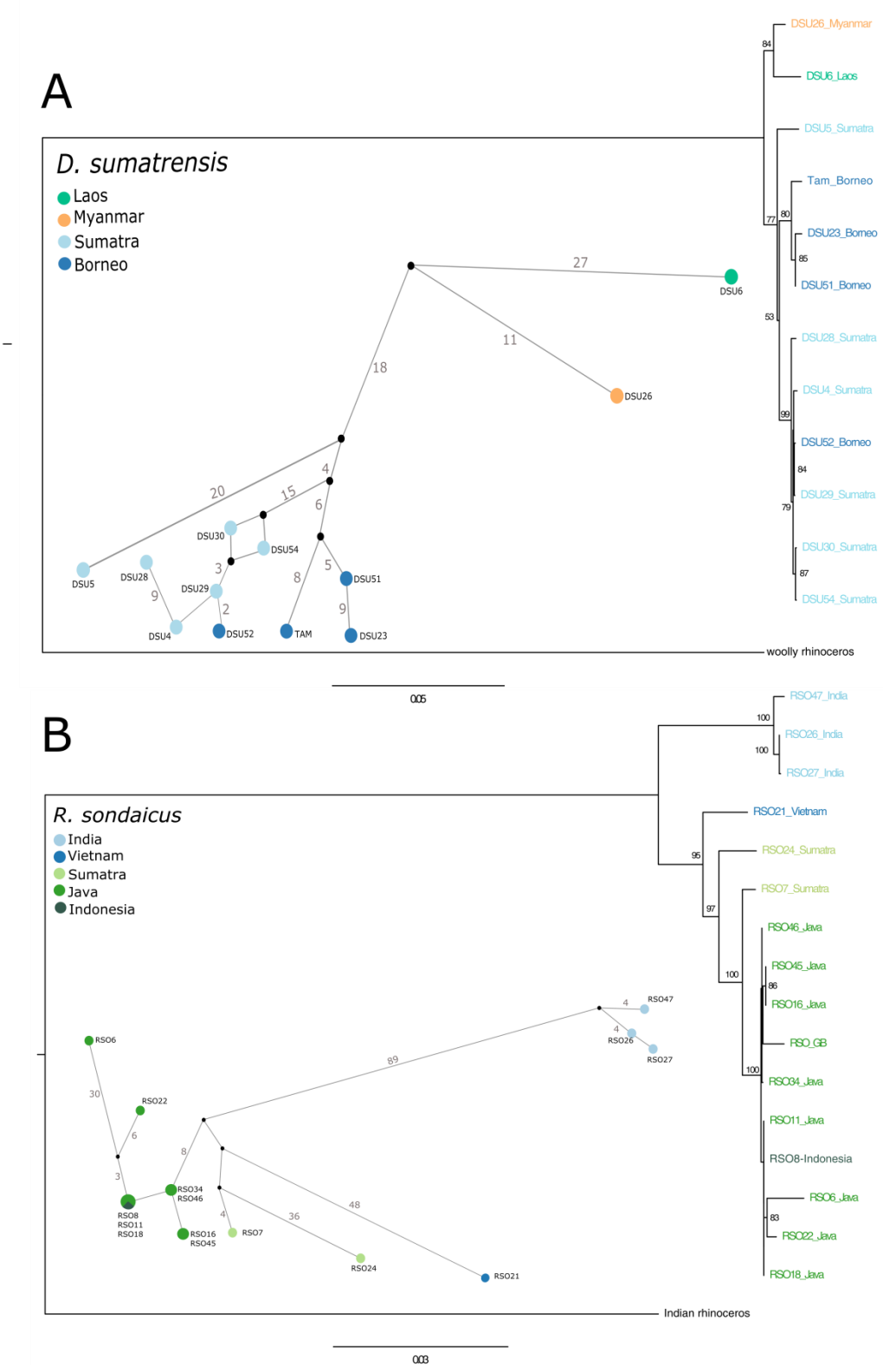

Figure 2 - ML gene tree and MJ network for Sumatran (A) and Javan (B) rhinos. Size of circles is proportional to the frequency of the haplotype and connecting vectors represent one mutational step except when indicated otherwise with numbers. Colours on haplotypes and branches represent geographical origin of samples.

and 12 within *D. sumatrensis* (DSU), respectively. Both species had a high haplotypic diversity, with Hd = 0.908 (SD = 0.0016) in RSO and Hd = 1 (SD = 0.034) in DSU, but a very low nucleotide diversity with $\pi = 0.0057$ (SD = 1.5×10^{-6}) in RSO and $\pi = 0.0035$ (SD = 0.0006) in DSU.

Phylogeography and genetic diversity

For Javan rhino, MJ network and ML topology were congruent (Fig. 2a) and showed the existence of three well defined clades, supported by high bootstrap values: one Indian, one Vietnamese (Indochinese) and a Sundaic one. Within the Sundaic clade all Javan rhinos were closely related and formed a distinct sub-clade from the Sumatran individuals. The haplotypic network separated the Indian clade from the remaining populations, by a minimum of 89 mutation steps. The Vietnamese individual was separated from all others by at least 48 mutations. The AMOVA results showed differentiation between the three clades, as the highest percentage of variation was explained by the differences between groups (India, Indochina and Sunda; Table 1).

The Sumatran rhino populations investigated here showed genetic structuring both in both the ML coalescent analysis and in the haplotypic network, albeit with smaller bootstrap support (Fig. 2b). Our data revealed two well differentiated clades: one constituted by the two Mainland individuals from Laos and Myanmar and the second Sundaic clade comprised all Sumatran and Bornean individuals. The haplotypic network separated the Indochinese clade from the Sundaic clade by at least 45 mutation steps, whereas the Myanmar and Laos sample were separated by 38 mutation steps. Within the Sundaic clade, there was evidence for two sub-clades; one with three individuals from Borneo and one with all individuals from Sumatra. The fourth individual from Borneo

Table 1 – AMOVA results for both species. Rhinoceros sondaicus (RSO) groups were India, Vietnam and Sunda; Dicerorhinus sumatrensis (DSU) groups were Mainland and Sunda.

		df	Sum of squares	% Variance	p-value
RSO	Among groups	1	204	42.09	0.16
	Among pop. within groups	2	112.1	43.8	<0.05
	Within populations	11	73.4	14.13	<0.001
DSU	Among groups	1	57	45.29	0.09
	Among pop. within groups	2	45.5	16.92	<0.001
	Within populations	8	79.3	37.78	<0.001

(DSU52) clustered within the Sumatran sub-clade. The two Sundaic sub-clades were separated at

least by 27 mutation steps. AMOVA results showed that the highest percentage of variation between the two groups (Mainland and Sunda), despite population differentiation being high between Laos and Myanmar and between Bornean and Sumatran sub-clades (Table 1).

Divergence dating and demographic changes

Fossil calibrated analyses revealed the presence of three well differentiated groups within the Rhinoceros phylogeny: Sumatran rhino + woolly rhino; Indian rhino + Javan rhino; white rhino + black rhino (Fig. 3). Bayesian Posterior Probabilities (BPP) were very high for all species splits (1), except for the split between the African rhinos and the other Rhinoceros species (BPP = 0.82). Our data suggested that the Sumatran + woolly rhino group split from all other species occurred first approximately 25 Mya (95% High posterior density [HPD] = 21.4 – 28.4); the Indian + Javan rhino clade diverged from their African sister clade shortly after, around 23.3 Mya (95% HPD = 20.4 – 26.4). Within Asian clades, our results suggested that the split between Sumatran and the extinct woolly rhino occurred around 15.8 Mya (95% HPD = 13.1 – 18.6), while the split between Javan rhino and Indian rhino happened around 10 Mya (95% HPD = 8.3 – 12.1). This analysis estimated the intraspecific split for lineages in the Sumatran rhino to have occurred around 900k years ago (HPD = 0.66 – 1.14 Mya), and the intraspecific divergence of lineages in the Javan rhinoceros to have occurred around 1.8 Mya (HPD = 1.4 – 2.23).

Specifically for each species, we calculated gene trees with the sets of complete mitogenomes using the clock rate as a prior. Root heights for both species were consistent with the values obtained in the phylogeny tree, although associated with larger 95% CI but branches were well supported by BPP values (Fig. 4). Tree topology for the Sumatran rhino was similar to that recovered by ML, revealing the presence of two well differentiated clades. For the Javan rhino, BI tree topology differed. The BI analysis placed the Vietnamese sample with the Indian clade, from which it would have split ca. 880 kya.

Discussion

The use of archival samples, in addition to one DSU contemporaneous samples, allowed us to investigate the historical genetic diversity of both highly endangered rhinoceros species, while shedding further light onto the evolution of all extant and the one extinct rhino species.

Reconstructing the true relationship between the five extant rhinoceros species remains controversial to this day. Classic studies based on morphology tended to group the two-horned species together (Sumatran + white + black rhinoceros), disregarding the geographical distribution of species (Loose 1975), a topology that was also supported by a molecular study (mtDNA, Steiner & Ryder 2011). However, the majority of recent molecular studies support the position of both

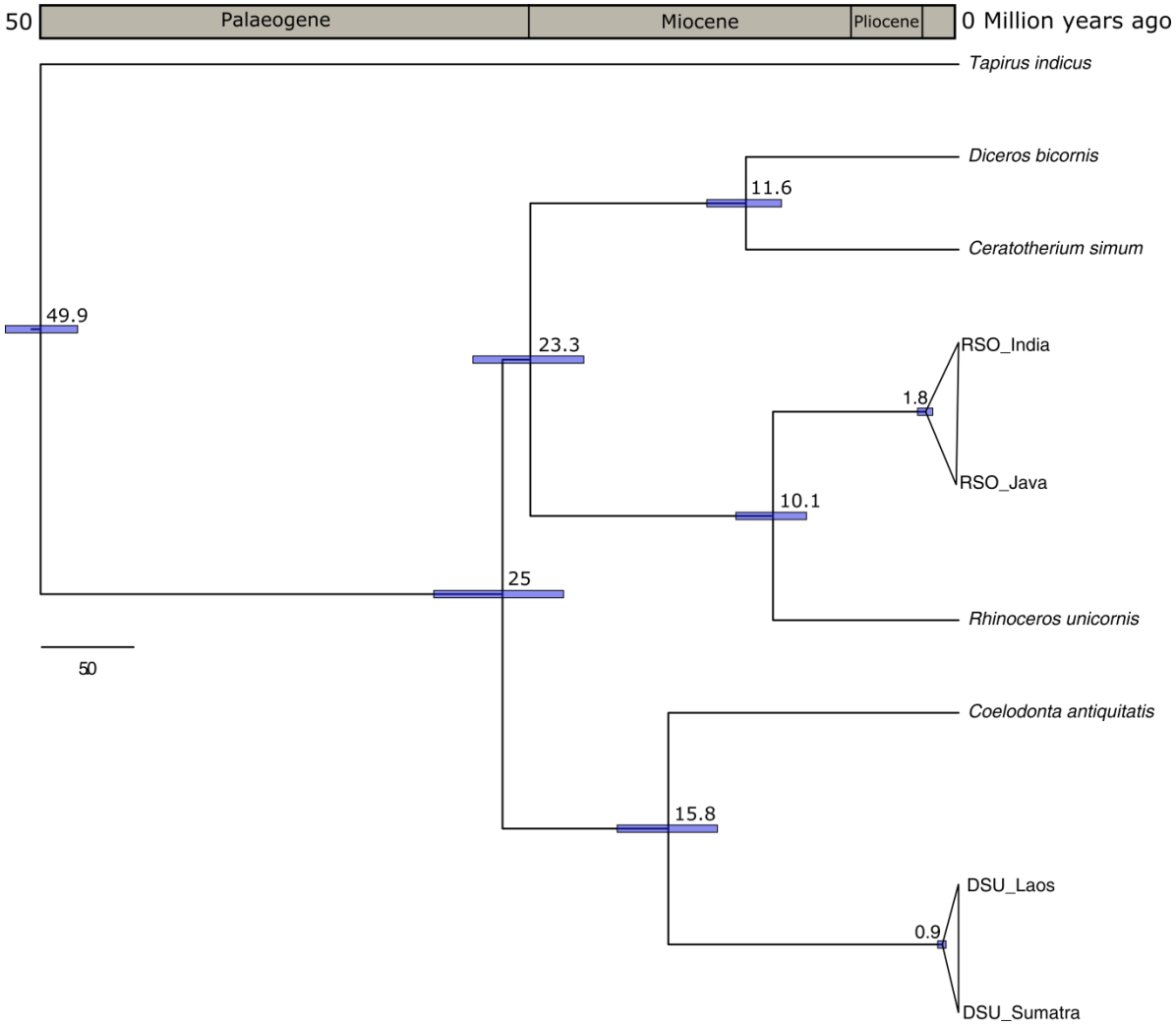

Figure 3. - Dated rhinoceros phylogeny with Bayesian Inference. Blue bars represent 95%HPD interval and numbers on nodes represent estimated dates.

African rhinos in a more basal position than both groups of Asian rhinos (Orlando et al. 2003, Welker et al. 2017). These conflicting phylogenies are likely a result of the fact that single gene trees were reported to provide different topologies for the rhinoceros phylogeny (Willerslev *et al.*, 2009). The full mitogenome based phylogeny reported here supports (with high probability) the basal position of Sumatran and woolly rhinos, which diverged from all other rhino species about 25 Mya. Such a basal position was also reported by other molecular studies employing single mitochondrial markers (12s and D-loop, Fernando et al. 2006), full mitogenomes (Harley et al. 2016) and nine nuclear genes (Steiner & Ryder, 2011). The split of the Javan and Indian rhinoceros from their African sister group was dated ca. 23.3 Mya, a date that largely overlaps with the confidence interval of the split between the Sumatran + woolly rhino and all other rhino species.

During the Miocene, ~15 Mya, the woolly rhinoceros started to diverge from the Sumatran rhino, and, ~10 Mya the two sister species of Rhinoceros (Javan and Indian) diverged as well, continuing

to inhabit the tropical regions of Mainland Southeast Asia. Our data suggested that only much later, during the Middle Pleistocene, both Javan and Sumatran rhinos expanded their range from Indochina to the Sunda Shelf. Although this pattern was consistent between the two species, we found that population divergence events occurred at different times, therefore corresponding to different climatic conditions. The differing timing of deep splits in the two species, e.g. between the Sundaic and the Indochinese clade (Sumatran rhino) or clades (Javan rhino) indicate that historically both species likely had different ecological requirements, resulting in the different evolutionary histories we observed.

Colonization of Sundaland by the Javan rhino started ca. 1.5 Mya, when sea levels are described to be as low as 100 m below present levels after the onset of a major global cooling event (Dennel, 2004). The shallow continental shelf was exposed and connected all land masses of Sundaland when sea levels were low. Reconstruction of the paleo-climate of these land corridors has been controversial, but many authors agree that forests were not distributed equally across the Sunda Shelf, and that most of the area had been covered by grassland or savannah (Meijaard, 2003; Wurster et al. 2010). Javan rhinos, alongside other more open-habitat, deciduous forest species (e.g. *Rusa timorensis*, Martins et al. 2017 *unpublished*), would have been able to utilize these available corridors and disperse towards Java. Land connection remained until ~1 Mya, when sea levels rose to 5 meters above present levels (Zazo, 1999). During the interglacial periods, climate was warmer and wetter throughout the region, which likely facilitated the expansion of topical evergreen rainforests. It is conceivable that these rainforests actually became barriers for Javan rhinos and thus resulting in the genetic differentiation between the Javan rhino populations from India, Vietnam and the Sunda islands. Our data revealed further genetic differentiation among Sundaic populations. Interestingly, Sumatran individuals were distinct from each other and from the monophyletic Javan population. This basal position of Sumatran haplotypes is further emphasized by the long branches. However, due to the lack of more Sumatran samples and samples from Peninsular Malaysia, the evolutionary history of the Javan rhinos from Sumatra could not be further examined.

Inclusion of samples from the historical continental Southeast Asia distribution of Javan rhinoceros, suggested that historical mainland populations presented similarly high levels of genetic diversity, especially illustrated here within the Indian lineage. Our data also suggested that the Indian and Indochinese populations represent distinct lineages. Such a differentiation between

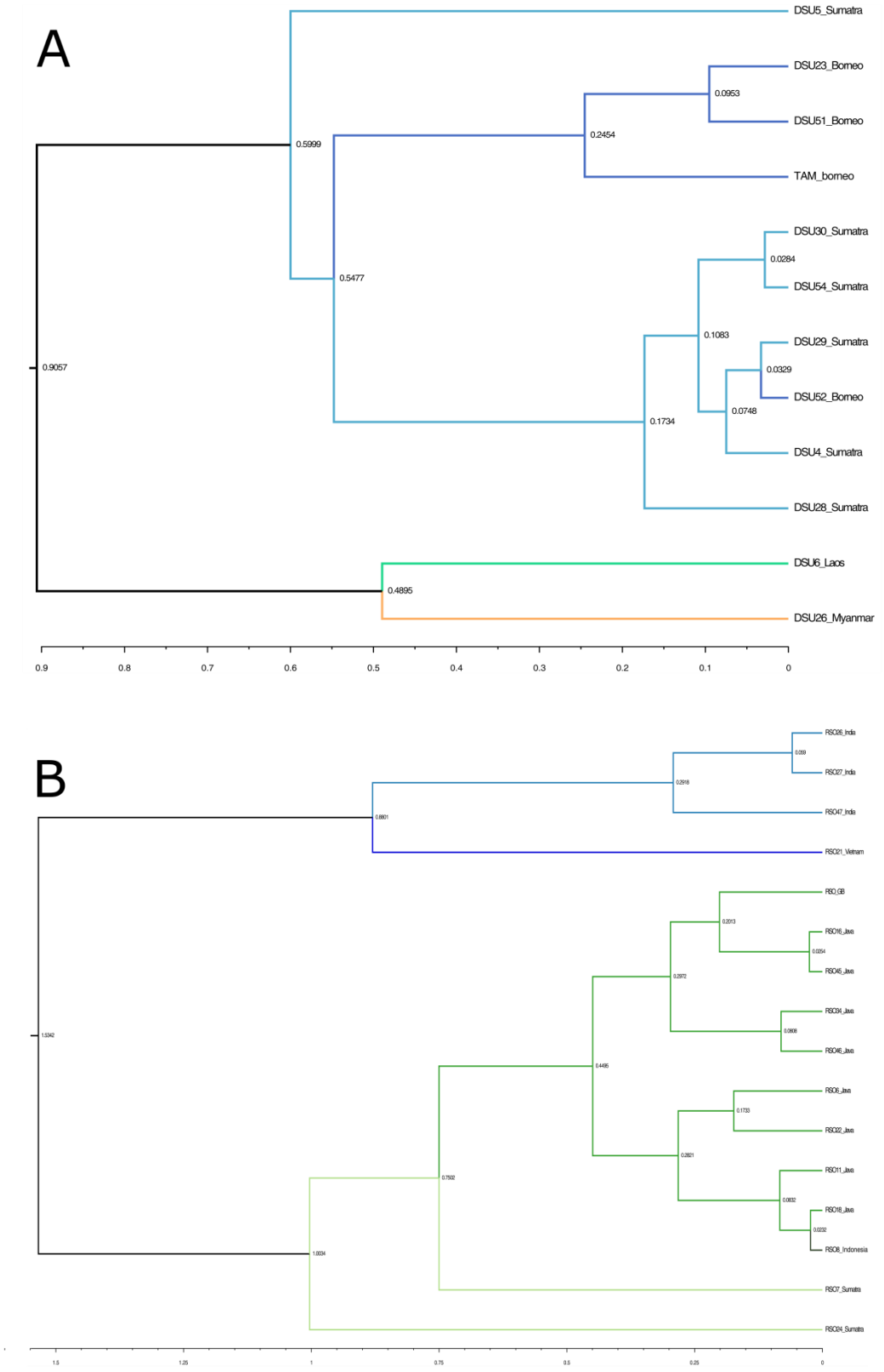

Figure 4. - Dated gene trees for Suamtran (A) and Javan (B) rhinos, through Bayesian inference. Scale is in millions of years before present.

Indian and Indochinese populations has been found in other species (e.g. mongooses: Veron et al. 2007; spiny frogs: Che et al. 2010; geckos: Wood Jr. 2012; murine rodents: Chingangbam et al. 2015), where the presence of the Arakan mountain range, which runs north-west through Myanmar, was appointed as a major barrier to gene-flow. However, because we do not have more samples from mainland Javan rhinos we cannot exclude that this is the result of sampling bias. Samples from the remaining historical distribution, especially from Bangladesh, Myanmar, Thailand and Cambodia would be crucial to investigate if genetic structure existed here as well.

In contrast to the Javan rhinoceros, southwards colonization of Sundaland by the Sumatran rhinos likely occurred during an interglacial period about 1 Mya, when the sea levels were elevated. It is likely that for this forest dependent species dispersal would have been facilitated by the warmer and wetter climate during this period. While during this highstand (+5 m; Zazo, 1999) some areas of the Sunda Shelf would have been inaccessible, shortly thereafter sea levels subsided supporting the expansion of Sumatran rhinos to Borneo and Sumatra. Later, with the onset of a glacial period around 800 kya lasting until 450 kya (Rohling et al. 1998; van der Bergh et al. 2001), we found evidence for population differentiation within both Indochinese and Sundaic populations. The cooler and, most importantly, drier climates during these glacial periods likely pushed rainforest-dependent species to more upland refugia or to lowland areas for example in north-eastern Borneo, which remained forested even during the glacial periods (Meijaard, 2003). We believe that this restriction of suitable habitats during the glacial periods caused the genetic differentiation we observed between the Sumatran rhino populations from Myanmar and Laos and between populations from the Sunda Shelf. Surprisingly, our Sumatran samples did not form a close cluster, rather showing high differentiation among Sumatran individuals. Speciation between West and East Sumatran populations of Sumatran rhinoceros has been proposed before (Morales 1997), and there is mounting evidence for within-island speciation processes in other vertebrate species (Demos et al. 2016). However, due to the lack of samples from Peninsular Malaysia and southern Borneo it remains uncertain if these differences are a result of speciation processes within Sumatra or if they rather are the result of more recent immigrations from Middle Pleistocene refugial populations in Peninsular Malaysia and Borneo. The position of sample DSU52 labelled to be from Borneo suggests such a gene flow between the major landmasses of the Sunda Shelf (Sumatra to Borneo), although further evidence, including nuDNA markers and more comprehensive sampling would be needed to confirm this hypothesis.

Genetic diversity within historical populations

Genetic diversity reported here for both Sumatran and Javan rhinos was higher than in previous studies, where only contemporaneous populations had been investigated.

One of the few studies on the phylogeography of the Javan rhinoceros (Fernando et al. 2006) found a similar divergence between Javan and Indochinese populations of Javan rhinos, concluding that there was sufficient genetic divergence between the two populations to confer their subspecies status, as *R. sondaicus sondaicus* (Java) and *R. s. annamiticus* (Indochina). However, that study did not include populations from India, whose morphological traits (e.g. body mass, footprint; Groves & Grubb 2011) led to the distinction of a third subspecies, *R. s. inermis*. Here, we found evidence to support the Indian subspecies of Javan rhino. Because we could not include samples from other Mainland locations (Thailand, Laos and Cambodia), the existence of further (potentially now lost) unique lineages cannot be excluded. Additionally to these three subspecies, some authors have also argued for the distinction of Sumatran and Javan populations of the Javan rhino into different subspecies (*R. s. floweri* and *R. s. sondaicus* respectively), a definition not widely accepted. Although we found no evidence of gene flow between the two populations, the genetic distance of the two Javan rhino samples from Sumatra to each other was similar, or even greater, to the distance of the Sumatran individuals to the Javan clade. Therefore our data suggests high genetic variability among *R. s. sondaicus*, rather than population differentiation within this subspecies.

Genetic variability among the small fragments investigated before had been observed to be very low among Sumatran rhino populations from Peninsular Malaysia, Sumatra and Borneo (Amato et al. 1995; Morales et al. 1997). These studies did not include any samples from the mainland subspecies *D. s. lasiotis* (generally considered to be larger than their Sundaic relatives Groves and Grubb, 2011). Our data confirmed that these populations were very distinct to Sumatran rhinos from the Sunda Shelf, supporting their subspecific status. Within the Sunda Shelf, two Sumatran rhino subspecies have been described: *D. s. sumatrensis* occurring in Peninsular Malaysia and Sumatra and *D. s. harrisoni* from Borneo. These subspecies have been controversially discussed; whereas both Amato et al. (1995) and Morales et al. (1997) found genetic differences between the populations, Amato and colleagues argue that these were not sufficient to justify distinct conservation units, Morales et al. 1997 supported the distinction into two subspecies. Goossens et al. 2013 reviewed these papers and referred to unpublished data by Rovie-Ryan et al. concluding that the observed differences are not sufficient to justify a separate management of Sumatran rhinos from Borneo and Sumatra. Our data confirm this, as although three Bornean samples formed a monophyletic cluster, they were nested within the Sumatran clade and there might be some evidence for recent gene-flow from Sumatra to Borneo (DSU52, if the labelling is correct).

The loss of subspecies, genetic erosion and the future of Asian rhinos

Although the exact historical distribution of Javan and Sumatran rhinos is not known, both species have occurred throughout large areas of Southeast Asia. Javan rhinos once occurred from India and Bangladesh, where it was the most abundant rhinoceros species, to Vietnam, Laos, Cambodia,

Myanmar and Thailand up to the southern border of China (Groves and Leslie, Jr., 2011). While it had already become rare in Myanmar by the mid-1800s, it was still present in Thailand, Laos and Cambodia until the 1930s. Presumably, the last Javan rhino in Vietnam (*R. s. annamiticus*) was shot in 2010 (Brook et al. 2012). Our data confirmed that this death likely meant the extinction of an entire subspecies, i.e. *R. s. annamiticus*. Although it is unknown when the last Javan rhino was killed in India, our data showed that this loss also constituted the loss of a subspecies, *R. s. inermis*. Only *R. s. sondaicus* survives until today in Ujon Kulon, West Java with a subpopulation no larger than 50 individuals (van Stein et al. 2008a). But also for this subspecies genetic diversity became greatly reduced during the population decline on Java and the extinction of *R. s. sondaicus* on Sumatran and Peninsular Malaysia (not included in this study) in the 1930s (Groves and Lesli, Jr., 2011).

Similar to the Javan rhino, Sumatran rhinos experienced a dramatic population decline and likely became extinct throughout Indochina and India, resulting in the disappearence of *D. s. lasiotis*. The second subspecies, *D. s. sumatrensis,* found in the Sunda Shelf, also went extinct throughout most of its range. Despite being once relatively abundant in Peninsular Malaysia, its decline started in the 1930s and the last the last sighting of a Sumatran Rhino in the region was in 1994 (Ahmad Zafir et al. 2011). Similarly the species was once widely distributed in Borneo, but after it was declared extinct in the wild in Sabah in 2015 (Lee, 2015), it is probable that there is only one remaining population in Kalimantan. Thus, almost all wild Sumatran rhinos today can be found in three protected areas on Sumatra (Havmøller et al. 2016). It is very likely that these small populations only represent a fraction of the genetic diversity of the species, as samples included in this study were partly over 100 years old and thus these museum specimens represent an already lost diversity.

Conclusion

Saving Javan and Sumatran rhinos will be challenging. The surviving individuals likely retain only a fraction of the species' historical genetic diversity. The wild populations are very small and in the case of the Sumatran rhino also isolated from each other. A reduction of mating encounters, in turn, results in severe reproductive tract pathologies that hinder non-assisted reproduction of these rhino species (Hermes et al. 2004; see further discussion in Kretzschmar et al. 2016). Earlier studies already highlighted the danger of keeping the remaining populations of Sumatran rhinos from Borneo and Sumatra separated (Goossens et al. 2013) and others have inferred that the outcome of outbreeding in Indian rhinoceros may have less negative consequences than previously believed (Zschokke, 2016). Our molecular data can only implicate the possibility of recent gene-flow between Sumatran rhinos from Sumatra and Borneo. However, due to the urgency of the situation of Southeast Asian rhinos, we suggest that urgent analyses of adaptive nuclear variation are necessary to assess the probability of outbreeding depression or genetic rescue, respectively, of

joint management. It is likely that joint management of populations of Sumatran rhino will be the only way that the little genetic diversity which is left can be used to save these iconic species from extinction.

Acknowledgements

The authors would like to thank all the contacts listed in Supplementary material for allowing us access to the museum collections. RM would like to thank Ramona Taubert and Anke Schmidt for the help with the laboratory work. This project was funded by SAW-2013-IZW-2. The authors also acknowledge support from Science for Life Laboratory, the Knut and Alice Wallenberg Foundation, the National Genomics Infrastructure funded by the Swedish Research Council, and Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure.

References

Ahmad Zafir, A.W., Payne, J., Mohamed, A., Lau, C.F., Sharma, D.S.K., Alfred, R., Williams, A.C., Nathan, S., Ramono, W.S., Clements, G.R. (2011) Now or never: what will it take to save the Sumatran rhinoceros *Dicerorhinus sumatrensis* from extinction? *Oryx*, 45, 225-233.

Amato, G., Wharton, D., Zainuddin, Z.Z. & Powell, J.R. (1995) Assessment of conservation units for the Sumatran rhinoceros (*Dicerorhinus sumatrensis*). *Zoo Biology*, **14**, 395-402.

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

Bird, M.I., Taylor, D., Hunt, C. (2005) Paleoenvironments of insular Southeast Asia during the Last Glacial Period: a savannah corridor in Sundaland? *Quaternary Science Reviews*, 24, 2228-2242.

Brook, S.M., van Coeverden de Groot, P., Scott, C., Boag, P., Long, B., Ley, R.E., Reischer, G.H., Williams, A.C. et al. (2012) Integrated and novel survey methods for rhinoceros populations confirm the extinction of *Rhinoceros sondaicus annamiticus* from Vietnam. *Biological Conservation*, **155**, 59-67.

De Bruyn, M., Nugroho, E., Mokarrom Hossain, Md., 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.

Cannon, C.H., Morley, R.J., Bush, A.B.G. (2009) The current refugial rainforests of Sundaland are unrepresentative of their biogeographic past and highly vulnerable to disturbance. *PNAS*, **106**, 11188-11193.

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. *PNAS*, **107**, 13765-13770.

Chingangbam, D.S., Laishram, J.M., Suzuki, H. (2015) Molecular phylogenetic characterization of common murine rodents from Manipur, Northeast India. *Genes & genetic systems*, **90**, 21-30.

Cranbrook, E. & Piper, P.J. (2007) The Javan rhinoceros *Rhinoceros sondaicus* in Borneo. *The Raffles Bulletin of Zoology*, **55**, 217-220.

Darriba, D., Taboada, G.L., Doallo, R. & Posada, D. (2015) jModelTest 2 : more models , new heuristics and high-performance computing. *Nature Methods*, **9**, 6–9.

Demos, T.C., Achmadi, A.S., Giarla, T.C., handika, H., Amsi, M., Rowe, K.C., Esselstyn, J.A. (2016) Local endemism and within-island diversification of shrews illustrate the importance of speciation in building Sundaland mammal diversity. *Molecular Evolution*, 25, 5158-5173.

Dennel, R.W. (2004) Hominid dispersals and Asian biogeography during the lower and Early Middle Pleistocene, c. 2.0-0.5 Mya. *Asian Perspectives*, **43**.

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

Excoffier, L., Laval, G. & Schneider, S. (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online,* **1**, 47–50.

Fernando, P., Polet, G., Foead, N., Ng, L.S., Pastorini, J., Melnick, D.J. (2006) genetic diversity, phylogeny and conservation of the Javan rhinoceros (*Rhinoceros sondaicus*). *Conservation Genetics*, **7**, 439-448.

Fortes, G.G. & Paijmans, J.L.A. (2015) Analysis of whole mitogenomes from ancient samples. *Whole Genome Amplification Methods Protocols*, 179–195.

Goossens, B., Salgado-Lynn, M., Rovie-Ryan, J.J., Ahmad, A.H., Payne, J., Zainuddin, Z.Z., Nathan, S.K.S.S., Ambu, L.N. (2013) Genetics and the last stand of the Sumatran rhinoceros Dicerorhinus sumatrensis. *Oryx*, **47**, 340.

Groves, C.P., Grubb, P. (2011) Ungulate Taxonomy. The Jonhs Hopkins University Press.

Groves, C.P. & Leslie, Jr., D.M. (2011) *Rhinoceros sondaicus* (Perissodactyla: Rhinocerotidae). *Mammalian Species*, **43**, 190-208.

Harley, E.H., de Waal, M., Murray, S., O'Ryan, C. (2016) Comparison of whole mitochondrial genome sequences of northern and southern white rhinoceroses (*Ceratotherium simum*): the conservation consequences of species definitions. *Conservation Genetics*, **17**, 1285-1291.

Havmøller, R.G., Payne, J., Ramono, W., Yoganand, S.E.K., Long, B., Dinerstein, E., Williams, A.C., Putra, R.H. et al. (2016) Will current conservation responses save the critically endangered Sumatran rhinoceros *Dicerorhinus sumatrensis*? *Oryx*, **50**, 355-359.

Hermes, R., Hildebrandt, T., Göritz, F. (2004) Reproductive problems directly attributable to long-term captivity-asymmetric reproductive ageing. *Animal Reproduction Science*, **82–83**, 49–60.

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.

Kretzschmar, P., Kramer-Schadt, S., Ambu, L., Bender, J., Bohm, T., Ernsing, M., Göritz, F., Hermes, R., Payne, J., Schaffer, N., Thayaparan, S.T., Zainal, Z.Z., Hildebrandt, T.B., Hofer, H. (2016) The catastrophic decline of the Sumatran rhino (*Dicerorhinus sumatrensis harrissoni*) in Sabah: Historic exploitation, reduced female reproductive performance and population viability. *Global Ecology and Conservation*, **6**, 257-275.

Li, H. & Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.

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

Lohmann, D.J., de Bruyn, M., Page, T., von Rintelen, K., Hall, R., Ng, P.K.L., Shih, H.-T., Carvalho, G.R., von Rintelen, T. (2016) Biogeography of the Indo-Australian Archipelago. *Annual Reviews of Ecology, Evolution and Systematics*, **42**, 205-226.

Loose, H. (1975) Pleistocene Rhinocerotidae of W. Europe with reference to the recent two-horned species of Africa and S.E. Asia. *Scripta Geologica*, **33**.

Maricic, T., Whitten, M. & Pääbo, S. (2010) Multiplexed DNA Sequence Capture of Mitochondrial Genomes Using PCR Products. *PLoS One,* **5**, e14004.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, **17**, 10.

Martins, R.F., Fickel, J., Le, M., van Nguyen, T., Nguyen, H.M., Timmins, R., Gan, H.M., Rovie-Ryan, J.J., Lenz, D., Förster, D.W. & Wilting, A. (2017) Phylogeography of red muntjacs reveals three distinct mitochondrial lineages. *BMC Evolutionary Biology*, **17**, 34.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 1297–1303.

Meijaard, E. (2003) Mammals of South-east Asian islands and their Late Pleistocene environments. *Journal of Biogeography*, **30**, 1245-1257.

Morales, J.C., Andau, P.M., Supriatna, J., Zainuddin, Z.Z. & Melnick, D.J. (1997) Mitochondrial DNA variability and conservation genetics of the Sumatran rhinoceros. *Conservation Biology*, **11**, 539-543.

Orlando, L., leonard, J.A., thenot, A., Laudet, V., Guerin, C., Hänni, C. (2003) Ancient DNA analysis reveals woolly rhino evolutionary relationships. *Molecular Phylogenetics and Evolution*, **28**, 485-499.

Paijmans, J.L.A., Gilbert, M.T.P., Hofreiter, M. (2012) Mitogenomic analysis from ancient DNA. *Molecular Phylogenetics and Evolution*, **69**, 404-416.

Parnell, J. (2013) The biogeography of the Isthmus of Kra region: a review. *Nordic journal of Botany*, **31**, 001-015.

Patou, M.-L., Wilting, A., Gaubert, P., Esselstyn, J.A., Cruaud, C., Jennings, A.P., Fickel, J. & Veron, G. (2010) Evolutionary history of the Paradoxurus palm civets - a new model for Asian biogeography. *Journal of Biogeography*, **37**, 2077-2097.

Silvestro, D. & Michalak, I. (2012) RaxmlGUI: A graphical front-end for RAxML. *Organisms Diversity & Evolution*, **12**, 335–337.

Steiner, C.C. & Ryder, O.A. (2011) Molecular phylogeny and evolution of the Perissodactyla. *Zoological Journal of the Linnean Society*, **163**, 1289-1303.

van Strien, N.J., Steinmetz, R., Manullang, B., Sectionov, Han, K.H., Isnan, W., Rookmaaker, K., Sumardja, E., Khan, M.K.M. & Ellis, S. (2008a) *Rhinoceros sondaicus*. The IUCN Red List of Threatened Species 2008: e.T19495A8925965.

van Strien, N.J., Manullang, B., Sectionov, Isnan, W., Khan, M.K.M, Sumardja, E., Ellis, S., Han, K.H., Boeadi, Payne, J. & Bradley Martin, E. (2008b) *Dicerorhinus sumatrensis*. The IUCN Red List of Threatened Species 2008: e.T6553A12787457.Tougard, C., Delefosse, T., Hänni, C., Mongelard, C. (2001) Phylogenetic relationships of the five extant rhinoceros species (Rhinocerotidae, Perissodactyla) based on mitochondrial cytochrome *b* ad 12s rRNA genes. *Molecular Phylogenetics and Evolution*, **19**, 34-44.

Veron, G., Patou, M.L., Pothet, G., Simberloff, D., Jennings, A.P. (2007) Systematic status and biogeography of the Javan and small Indian mongooses (Herpestidae, Carnivora). *Zoologica Scripta*, **36**, 1-10.

Veron, G., Willsch, M., Dacosta, V., Patou, M.-L., Seymour, A., Bonillo, C., Couloux, A., Wong, S.T., Jennings, A.P., Fickel, J., Wilting, A. (2014) The distribution of the Malay civet Viverra tangalunga (Carnivora: Viverridae) across Southeast Asia: natural or human-mediated dispersal? Zoological Journal of the Linnean Society, **170**, 917–932.

Welker, F., Smith, G.M., Hutson, J.M., Kindler, L., garcia-moreno, A., Villaluenga, A., Turner, E., Gaudzinski-Windheuser, S. (2017) Middle Pleistocene protein sequences from the rhinoceros genus Stephanorhinus and the phylogeny of extant and extinct Middle/Late Pleistocene Rhinocerotidae. *PeerJ*, **5**, e3033.

Willerslev, E., gilbert, M.T.P., Binladen, J., Ho, S.Y.W., Campos, P.F., Ratan, A., Tomsho, L.P., da Fonseca, R.R., Sher, A. et al. (2009) Analysis of complete mitochondrial genomes from extinct and extant rhinoceros reveals lack of phylogenetic resolution. *BMC Evolutionary Biology*, **9**.

Wilting, A., Sollmann, R., Meijaard, E., Helgen, K.M., Fickel, J. (2012) Mentawai's endemic, relictual faun: is it evidence for Pleistocene extinctions on Sumatra? *Journal of Biogeography*, **39**, 1608-1620.

Wood, P.L., Heinicke, M.P., Jackman, T.R., Bauer, A.M. (2012) Phylogeny of bent-toed geckos (Cyrtodactylus) reveals a west to east pattern of diversification. *Molecular phylogenetics and evolution*, **65**, 992-1003.

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.

Wurster, C.M., Bird, M.I., Bull, I.D., Creed, F., Bryant, C., Dungait, J.A.J. (2010) Forest contraction in north equatorial Southeast Asia during the Last Glacial Period. *PNAS*, **107**, 15508-15511.

Zazo, C. (1999) Interglacial sea levels. *Quaternary International*, **55**, 101–113.

Zschokke, S. (2016) Genetic structure of the wild populations of the Indian Rhinoceros (*Rhinoceros unicornis*). *Indian Journal of History of Science*, **51**, 380-389.

Supplemental Material

Table S1 – Complete sampling list with details on sample origin and contact

Species	Study ID	Locality	Date collected	Collection	catalog #	Contact
Dicerorhinus sumatrensis	DSU4	Sumatra	1860	Naturalis	19594	P.Kaminga
Dicerorhinus sumatrensis	DSU5	Sumatra	1880	Naturalis	19596	P. Kaminga
Dicerorhinus sumatrensis	DSU6	Laos	1904	Vienna NHM	8173	F. Zachos
Dicerorhinus sumatrensis	DSU23	Borneo	1908	Munich NHM	1908/571	M. Hiermeyer
Dicerorhinus sumatrensis	DSU26	Myanmar		AMNH	54764	N. B. Simmons
Dicerorhinus sumatrensis	DSU28	Sumatra		Stud book	25	J. Christman
Dicerorhinus sumatrensis	DSU29	Sumatra		Stud book	28	J. Christman
Dicerorhinus sumatrensis	DSU30	Sumatra		Stud book	29	J. Christman
Dicerorhinus sumatrensis	DSU51	Borneo		Smithsonian	199551	J.A. Coddington
Dicerorhinus sumatrensis	DSU52	Borneo		Smithsonian	102076	J.A. Coddington
Dicerorhinus sumatrensis	DSU54	Sumatra	1893	Copenhagen NHM	ZMUC617	D.K. Johansson
Dicerorhinus sumatrensis	TAM	Borneo	fresh	na	na	S. Nathan
Rhinoceros sondaicus	RSO6	Java		Naturalis	19587	P. Kaminga
Rhinoceros sondaicus	RSO7	Palembang, Sumatra	1928	Naturalis	ZMA7681	P. Kaminga
Rhinoceros sondaicus	RSO8	India		Naturalis	ZMA 512	P. Kaminga
Rhinoceros sondaicus	RSO11	Java	1852	Stuttgard NHM	32015	S. Merker
Rhinoceros sondaicus	RSO16	Java		Berlin NHM	ZMB-MAM-37533	F. Meyer
Rhinoceros sondaicus	RSO18	Java		Berlin NHM	ZMB-MAM-37535	F. Meyer
Rhinoceros sondaicus	RSO21	Vietnam	1896	Paris NHM	1896-2003	G. Veron
Rhinoceros sondaicus	RSO22	Java		Paris NHM	A-7971	G. Veron
Rhinoceros sondaicus	RSO24	Sumatra	1930	Munich NHM	1930/352	M. Hiermeyer

Rhinoceros sondaicus	RSO26	Bengala, India		Berlin NHM	ZMB-MAM-1958	F. Meyer
Rhinoceros sondaicus	RSO27	Bengala, India		Berlin NHM	ZMB-MAM-1957	F. Meyer
Rhinoceros sondaicus	RSO34	West Java		MCZ Harvard Uni.	MCZ 27234	H. Hoekstra
Rhinoceros sondaicus	RSO45	Java		Copenhagen NHM	ZMUC26	D.K. Johansson
Rhinoceros sondaicus	RSO46	Java		Copenhagen NHM	ZMUC29	D.K. Johansson
Rhinoceros sondaicus	RSO47	Bhootan Doras, India	1887	Copenhagen NHM	ZMUC524	D.K. Johansson

Table S2. Long-range PCR primer sequences and annealing temperatures used for bait development.

	Position on ref.	Fragment size	Sequence 5'-3'	Annealing Temp.[a]
Fragment 1	10600	7 kb	TGAGGTAAYCAAACAGAACGACTAA	58^{0}C
	1030		GTACGCTTACCTTGTTACGACTTG	
Fragment 2	623	6 kb	GACCCTGTTCCATAAYCGATAAAC	58^{0}C
	6751		TATAACTGCTGTGAGTGAGATGAAGG	
Fragment 3	5387	3 kb	CTTTCTCAGCCAYTTTACCTATGTTC	58^{0}C
	7225		TATAGTRCTTGTGTGKGTTAGTTTG	
Fragment 4	7250	3 kb	CAAACTAACMCACACAAGYACTATA	58^{0}C
	11504		GTGTTGTGATTARTATGTAGAGGGA	

[a] PCR was done with MyFi™ Mix (Bioline GmbH, Germany) with 1x MyFi Mix, 0.4 µM each primer and water to the final volume of 50 µL for each of the fragments separately, in a total of 35 amplification cycles.

Table S3. Sequencing results from all samples used in this study, with indication of number of reads on target (mtDNA), de-duplicated reads and the percentage of genome covered with more than 3 unique reads.

Species	Sample ID	Quality trimmed reads	Selection efficiency	Unique mapped reads	average depth coverage	genome coverage ≥ 3×
Dicerorhinus sumatrensis	DSU4	8,309,903	0.45%	3,380	16.28	95.7
Dicerorhinus sumatrensis	DSU5	796,845	65.16%	25,824	158.19	100
Dicerorhinus sumatrensis	DSU6	5,183,162	0.10%	4,419	18.49	98.7
Dicerorhinus sumatrensis	DSU23	8,961,283	0.45%	1,581	8.12	93.0
Dicerorhinus sumatrensis	DSU26	6,832,044	0.05%	1,825	9.24	72.4
Dicerorhinus sumatrensis	DSU28	8,610,206	0.03%	2,318	10.09	93.6
Dicerorhinus sumatrensis	DSU29	10,085,661	0.10%	5,313	23.58	99.8
Dicerorhinus sumatrensis	DSU30	3,506,185	0.09%	2,691	11.40	98.7
Dicerorhinus sumatrensis	DSU51	5,424,384	3.01%	15,171	81.81	99.7
Dicerorhinus sumatrensis	DSU52	10,138,888	0.07%	5,209	21.70	99.9
Dicerorhinus sumatrensis	DSU54	5,942,995	0.02%	786	4.29	71.4
Dicerorhinus sumatrensis	TAM	41,198	87.91%	21,282	222.15	99.7
Rhinoceros sondaicus	RSO6	65,385,334	0.01%	1,726	8.33	84.6
Rhinoceros sondaicus	RSO7	5,393,018	0.05%	2,413	10.51	96.6
Rhinoceros sondaicus	RSO8	60,171,224	0.09%	7,326	36.86	99.9
Rhinoceros sondaicus	RSO11	8,949,220	0.02%	1,589	6.59	87.1
Rhinoceros sondaicus	RSO16	783,174	7.20%	8,265	40.20	98.2
Rhinoceros sondaicus	RSO18	5,019,421	0.51%	8,104	37.48	98.1
Rhinoceros sondaicus	RSO21	7,467,226	0.01%	6,121	24.31	99.3
Rhinoceros sondaicus	RSO22	17,138,923	0.46%	1,166	5.53	80.9
Rhinoceros sondaicus	RSO24	7,859,344	0.93%	2,826	12.18	96.5
Rhinoceros sondaicus	RSO26	48,702,837	0.48%	12,994	59.50	98.7
Rhinoceros sondaicus	RSO27	413,658	6.44%	1,977	7.27	72.4
Rhinoceros sondaicus	RSO34	7,670,955	0.03%	1,980	8.24	92.4
Rhinoceros sondaicus	RSO45	1,020,882	28.78%	16,327	75.39	99.0
Rhinoceros sondaicus	RSO46	52,225,425	0.58%	23,770	141.79	100.0
Rhinoceros sondaicus	RSO47	6,188,725	1.39%	4,422	24.15	99.0

General Discussion

General Discussion

The aim of this thesis was to investigate what factors influenced the current distribution of species genetic diversity in several biodiversity hotspots in South and Southeast Asia. To achieve that, I set out to describe and compare phylogeographic patterns of ungulate species distributed in this region. In particular, I focused at understanding the effects of the dramatic climate changes of the Pleistocene and the consequent changes in land and habitat availability.

In summary I found very diverse phylogeographic patterns among all species investigated. By comparing the results within and among each species, I found some common patterns of genetic diversity that correlated well to known geographical boundaries, but others were more related to the specific ecological requirements of each species. Taken together, the results obtained in this dissertation highlight the importance of comparing patterns of genetic diversity between different species to understand how climate variations of the past have affected the distribution, dispersal and survival of populations. In addition, species specific phylogeographic patterns could also constitute an important basis for future conservation and taxonomic decisions.

1 Comparative phylogeography

1.1 Land corridors of Sundaland

Land corridors connecting the Sundaic islands among each other and to the mainland have been repeatedly exposed when sea levels dropped throughout the glacial periods of the Pleistocene (Voris, 2000). However, many contradictory studies have proposed that either this newly available landmass was covered by dry savannah (Gorog et al. 2004; Wurster et al. 2010), while others argued that forests expanded on the exposed shelf (Cannon et al. 2009; Raes, 2014). Other authors rather argued that the distribution of rainforest and savannah habitat throughout the emerged shelf was patchier, with the survival of forest covered islands among the surrounding open habitat (Meijaard, 2003) or that a dry central corridor was present throughout Sundaland in early Pleistocene, giving place more pre-humid conditions in the late glacial periods of Late Pleistocene (Sheldon et al. 2015).

When compared, the results obtained throughout this dissertation show a distinct response from generalists and specialist species (Table 1). As shown, land corridors provided different levels of permissiveness to different species. More specifically, the more generalist species (i.e. that occur in different habitats) showed evidence of gene-flow between island populations, which we interpreted as the consequence of dispersal through the available corridors.

Table 1 - Summary of evidence for the use of land corridors from all investigated species

Manuscript	Species	Ecology	used land corridor	Divergence date (Mya)
I	*Muntiacus muntjak*	generalist	yes	< 1.12
II	*Rusa unicolor*	generalist	yes	< 1.18
	Rusa timorensis	open-habitat forest	at least once	
III	*Rhinoceros sondaicus*	lowland/open-habitat forest	at least once	1.5
	Dicerorhinus sumatrensis	evergreen rainforest	no	1

Open-forest species, like the Javan deer, *Rusa timorensis*, showed evidence of limited gene flow among individuals on Java and Bali, its native distribution range. However, we argued that this result may have been due to a reduced sample size from Bali and could actually represent a 'small population' effect. Additionally, we found at least one individual carrying a *R. timorensis* haplotype in Sumatra, which might indicate that indeed some Javan deer individuals were able to cross the available land corridor between Java and Sumatra. As this species represents an open-forest specialist, the available land corridors should have been quite permissive for population dispersal, so here this expansion might have been hindered by the sister-species *Rusa unicolor*, which was already present in high numbers in the newly available habitat ('priority effect', De Meester et al. 2016).

A similar result was shown for the Javan rhinoceros, which is a lowland and open-habitat forest species. We showed that expansion of Javan rhinos through Sundaland was facilitated by the low sea levels of the glacial period, ca. 1.5 Mya. Conversely, the Sumatran rhino, a truly forest dependent species showed very little evidence of utilizing these available landmasses. Indeed, we found evidence that expansion of this species southwards happened rather during the interglacial period, ca. 1 Mya (van der Bergh et al. 1996) and dating of lineage differentiation of island populations always matched periods of low sea levels, thus supporting the idea that the land corridors were not permissive for large scale dispersal of this forest-dependent species.

For those species that showed evidence of gene-flow, the tempi of dispersal events was consistent. All events of gene flow described for the red muntjacs, the sambar and the Javan rhinoceros occurred at glacial periods (Table 1). At 1.5 Mya sea levels are reported to have been about 100 m below present levels (Dennel, 2004) and later, around 800 kya, sea levels would have dropped to a maximum of -170 m, maintaining a land connection between islands until ~ 450 kya, with the onset of the next interglacial interval (Rohling et al. 1998).

The results of my dissertation indicate therefore that species specific ecological needs do indeed influence their ability to disperse across the land corridors during the Pleistocene. Forest dependent species showed limited evidence of gene-flow between island populations, whereas lowland and open-habitat species used the exposed shelf and could, therefore, expand their range to Java. We

were unable to determine if the more open habitat species (*Rusa timorensis* and *Rhinoceros sondaicus*) utilized the land connections more than once, after their arrival on Java. However, red muntjacs and the sambar, both species well adapted to very different kinds of habitat, showed evidence of repeated gene-flow between the islands of Sundaland, as we found evidence of complete lack of geographical structuring among these populations. Our data thus support the hypothesis that the available land corridors of the Sundaland, during the glacial periods of the Pleistocene, would be rather covered by open and dry habitat and that rainforests were likely restricted.

1.2 Vicariance beyond the presence of land corridors

Even species that showed extensive gene-flow between populations from different islands, showed also some patterns of genetic sub-structuring. These findings were not easily explained just by the different permissiveness of the land corridors and pointed to other vicariant processes at play, within and among islands.

We found three major matrilineages of red muntjacs: a Sri Lankan lineage, a mainland and a Sundaic lineage. Within the Sundaic lineage, we revealed that all Sumatran muntjacs had a derived position in the coalescence analysis. We attributed this finding to the effects of the super-eruption of Toba, in North Sumatra, around 74 kya. The eruption of Toba is described as one of the greatest volcanic eruptions of the last two million years and thought to have caused local extinctions in northern Sumatra and Peninsular Malaysia, due to the prolonged presence of an ash cloud (Williams et al. 2009; Wilting et al. 2012). Thus, one of the most parsimonious explanation for the derived position of all Sumatran individuals is a wipe-out of ancestral Sumatran populations of red muntjacs by the Toba eruption, followed by re-colonization from other Sundaic islands.

For the Sumatran rhino, we found likewise an interesting genetic sub-structure within Sumatran populations, as one individual was much older than all remaining samples from that island. The position of that haplotype could be the result of either one of two scenarios, both suggesting within-island mechanisms of genetic differentiation: 1) the haplotype represents a survivor of an ancestral population that went through a severe bottleneck (Toba eruption) or 2) the haplotype indicates differentiation between eastern and western populations of Sumatran rhino possibly by the Bukit Barisan volcanic mountain range (running from North to South Sumatra). The latter scenario had actually been proposed before in a study of short fragments of the mtDNA control region of Sumatran populations (Morales et al. 1997).

Additionally I found that both species which occur in Java (Javan deer and Javan rhinoceros) show well differentiated lineages to populations/species on the surrounding islands. We attributed this result to the very distinct climate on Java, which presents more open-habitat and

grasslands than the neighbouring islands (Mishra et al. 2010). Local adaptations to this climate seem to have acted as a driver for generating and maintaining genetic differentiation. This pattern was observed previously in other sister-species pairs where populations distributed in Java and Bali rather represented endemic species (in one or both of the mentioned islands; e.g. barbets *M. armillaris* and *M. corvine*, Den Tex & Leonard, 2013). In fact, in a recent study, which compared 28 non-migratory and forest-dependent vertebrate species, the authors found that Javan populations, although related to their Sumatran conspecifics, represented the most divergent lineages (Leonard et al. 2015).

Taken together, these considerations contrast with the traditional expectation that species would be widespread throughout Sundaland and with low levels of genetic differentiation. They rather point toward more stochastic and local drivers of differentiation, such as within-island barriers to dispersal and adaptations to local climate.

1.3 Isthmus of Kra as a geographical barrier

The Thai-Malay peninsula is a narrow landmass, located at the junction between the Indochinese and Sundaic zoogeographic regions, and very rich in endemic species (Woodruff 2010). The narrowest part of this peninsula is located around the Isthmus of Kra, which has been recognized as a biogeographic barrier to the dispersal of many species (de Bruyn et al. 2005), impeding gene-flow between populations north and south of this barrier.

I tested for the existence of a sharp genetic break around the Isthmus of Kra in all widely distributed species included in this work. The red muntjacs unmistakeably showed a pattern of genetic differentiation around the Isthmus of Kra. For this group, the dataset included samples from Peninsula Malaysia which presented similar haplotypes with Bornean individuals, indicating the presence of a geographical barrier limiting gene-flow in the Thai-Malay peninsula. Karyotype studies had already indicated that red muntjac forms from Sundaland differed from those in mainland Southeast Asia (2n=8F/9M vs. 2n=6F/7M; Groves & Grubb, 2011), also pointing to some morphological differences between individuals. Although these inferences have been contested before (Timmins et al. 2016), the results presented in Manuscript I seem to support the separation between the two forms into (sub-) species status (see discussion below in section 2).

A similar pattern was found for the Sumatran rhino, where two individuals from Myanmar and Laos clustered together in a mainland clade, well differentiated from all Sundaic populations. However, because we had no samples from Peninsula Malaysia, we could not confirm the position of this barrier. Within the Javan rhino we did not see such a clear split between Mainland and Sunda

lineages, because samples from the mainland were equally distant to each other and to the Sundaic lineages. We interpreted this as lack of a sharp genetic break around the Thai-Malay peninsula.

In Manuscript II I investigated the phylogeography of the widespread sambar (*Rusa unicolor*) by combining mitogenome and nDNA analysis. Results did not indicate the presence of a biogeographic divide among sambar populations. While we noted the existence of three older clades in our phylogenetic analyses, the younger and much larger clade included samples from both mainland and Sundaland, without any apparent geographic sub-structure.

Thus, comparing the results obtained for each group, it is clear that specialist and generalist species alike showed evidence of having crossed the Thai-Malay Peninsula at least once, both colonizing the landmasses of Sundaland and creating patterns of genetic admixture among the island populations. Our results render it rather unlikely that a strong physical barrier to dispersal existed at the Isthmus of Kra, at least during the Pleistocene interglacial intervals (Parnell, 2013). The conflicting results about the presence of a sharp dispersal barrier around the Isthmus of Kra obtained in this study are representative of the ongoing discussion about the nature of the Thai-Malay transition zone. Previous comparative studies of species distributions have identified differences in the position of this barrier. Woodruff and Turner (2009) investigated the southern limits of 152 species from Indochina and the northern range limits of 147 Sundaic species and concluded that there was no evidence for a faunal transition at the Isthmus of Kra. They, however, suggested the current species distribution patterns – latitudinal diversity of species compositions and a reduced number of species in central Thai-Malay peninsula – to be the result of prolonged sea level rises during the last 5 My, which reduced the habitat availability, compressed species and thus caused local extirpations in the narrowest areas of the peninsula. During the Pleistocene climate changes, habitat availability and the presence of rainforest in this region might have been the most differentiating force. The nature of this transition zone remains, therefore, largely unresolved and indicates the need for further studies, especially those that identify the paleo-habitat present in the region throughout the Pleistocene climate changes, as it seems to have the biggest influence on the distribution of terrestrial species (Parnell, 2013; Dejtaradol et al. 2016).

1.4 Genetic differentiation among mainland lineages

Inclusion of widespread Southeast Asian species in this dissertation allowed inferring the phylogeography of mainland populations. Previous phylogeographic studies, which included mainland populations, revealed intriguing patterns of lineage divergence and indicated that the cyclic changes of the Pleistocene had major impacts on the climate and vegetation of mainland Southeast Asia (Yuan et al. 2016).

Red muntjacs showed evidence of gene-flow between mainland populations, but the large genetic distance that two samples (one from India and one from Laos) had to the other mainland haplotypes, indicated the presence of so far unknown refugial lineages among mainland populations. The validity of a 'high elevation Himalayan' red muntjac, which was suggested by morphological examinations (Timmins, pers. comm. 2015) is therefore strongly supported. In fact, endemic (cryptic) red muntjac species are assumed to exist in the Cardamom Mountains in Thailand and Cambodia and the Annamite Mountains of Vietnam and Laos. This, however, can only be confirmed (or rejected) by analysing appropriate datasets from those regions.

We also recovered a pattern of complete differentiation of South Indian (Western Ghats) and Sri Lankan populations from the rest of the Indian red muntjacs, which formed a monophyletic clade at a basal position. A very similar pattern was recovered by the phylogeography of the sambar, *Rusa unicolor*. Sri Lankan individuals were genetically differentiated from the remaining Indian populations. We did not obtain any samples from the Western Ghats, but a recent study had shown a 40 bp insertion in the control region of the mitochondrial DNA of Western Ghats populations, also supporting the uniqueness of these lineages (Gupta et al. 2015). We argued that differentiation of Sri Lankan populations of both species was the result of isolation in the wet-zones that prevailed in these regions during the Pleistocene glacial periods. Although Indochina was not glaciated during the Pleistocene glacial periods, except for high altitude mountains, it experienced drier climates during the Pleistocene (Li et al. 2004). Consequently, rainforests and other humidity dependent vegetation zones (e.g. wetlands) contracted their expansion range (Gong et al. 2015). Thus, mainland species/populations experienced likewise cyclic habitat expansions and contractions and would have retracted into small refugia during glacial periods (Fuchs et al. 2008; McLeod, 2010). We suggested that if isolation of lineages in refugia resulted in the generation of other barriers to gene-flow (e.g. ecological, behavioural, reproductive), isolation would have been maintained even after the re-colonization of central India.

It is not possible to infer the presence of cryptic lineages within the two rhino species investigated here. Even though this was the first time samples from historical mainland populations had been included in a phylogeographic study, our sampling across the mainland was nevertheless very poor. Among Javan rhino samples, we observed that Indian and Indochina populations were genetically distant, supporting their subspecies status. Also within the Sumatran rhino the two samples from Myanmar and Laos showed genetic differentiation. Thus, these results indicate a vicariant history, where populations likely had their range periodically contracted, accompanying the retraction of forest during the glacial periods of the Pleistocene (Sumatran rhino) or the presence of other barriers to dispersal, like mountain ranges (Javan rhino; e.g. Arakan mountains). Sadly, there are no Sumatran or Javan rhinos left in mainland Southeast Asia, so these results also illustrate the great genetic diversity loss in Southeast Asian megaherbivores.

2 Taxonomic and conservation implications of this study

One of the immediate values of phylogeographic studies is their benefit for identifying taxonomic units and lineages of conservation concerns (Wiens, 2012; Losos et al. 2013). Southeast Asia has the highest relative rate of deforestation of any major tropical region (Sodhi et al. 2004), therefore understanding what species/lineages occur where is essential to conservation measures as they aim at mitigating the loss of biodiversity. Within biodiversity hotspots it is possible that 'widespread species actually represent multiple species of smaller geographical ranges' (Stuart et al. 2006). Likewise, not knowing what taxonomic units are distributed within a conservation priority area biases our understanding of biodiversity within 'Biodiversity Hotspot' (Bland et al. 2015), especially when the number of threatened species or lienages might already surpass the available conservation resources (Myers et al. 2000)

In this dissertation, first and foremost the use of archival samples allowed us to uncover how much of the former genetic variability of Javan and Sumatran rhinoceros has already been lost. For these two species, not only were we able to show the extraordinary amount of genetic erosion that current populations have undergone, but also the uniqueness of the already extinct mainland populations. This study also results in some intriguing thoughts about the sub-species status of some of the populations. The genetic differentiation within Javan rhinos supported the distinction of the three subspecies (*Rhinoceros sondaicus inermis* from India; *R. s. annamiticus* from Indochina and *R. s. sondaicus* from Sumatra and Java). However, genetic differentiation between Sumatran rhino populations rather argues for cryptic taxonomic units among mainland populations and the re-evaluation of *Dicerorhinus sumatrensis lasiotis*. More importantly, we argued that urgent studies on the Sumatran and Bornean populations of this species are needed to confirm the presence of gene-flow between these two islands. If confirmed, it should lead to consider both Sumatran and Bornean popualtions under the subspecies, *D. s. sumatrensis*, having important conservation implications as joint management (already proposed as the only solution for this species recovery, Goossens et al. 2013) would be supported as well by molecular evidence.

While the Asian rhinoceros are undoubtely two of the most endangered large mammals in the world, the red muntjacs are among some of the most resilient. In Manuscript I, we reported evidence for the division of the Sundaic and mainland populations into different taxonomic units. These two clades are currently considered as different species (Sundaland: *Muntiacus muntjak* and Mainland: *M. vaginalis*; see The IUCN Red List of Threatened Species), which have now clear molecular support. Similarly, in Manuscript II, we mentioned the taxonomic implications of our results, as the seven subspecies that are currently described for both sambar and the Javan deer (Mattioli, 2011) were not supported by molecular evidence. The only exception was the Javan deer subspecies *Rusa timorensis renschi* that occurs only on Bali (Hedges et al. 2015), for which we

found clear evidence both from mitogenome and nuDNA analyses. Most importantly, we also discussed that introgression between the two Rusa species resulted in all individuals in the islands where Javan deer were introduced to be of hybrid descend. This result has strong taxonomic and conservation implications, as it drastically reduces the population of true *Rusa timorensis*. Finally, both red muntjacs and the sambar showed genetically differentiated Sri Lankan (and Western Ghats lineages, if I consider previous findings). Therefore, I suggest these populations (Sri Lanka and Western Ghats) to be managed differently, as human pressure on Sri Lankan habitats (Jha et al. 1995) might be enough to grant them specific conservation policies.

3 Suitability of the methodology

While methods for the development of sequencing libraries from archival samples were established and improved for Illumina sequencing platforms (Fortes & Paijmans, 2015), they were still lacking for alternative technologies. As briefly mentioned in Manuscript I, some samples were not sequenced on an Illumina platform but with the Personal Genome Machine (PGM, Thermo Fisher scientific, CA, USA). The PGM platform offers a set of advantages, like scalability of the input data and, more importantly, the possibility to sequence very low diversity fragments (as the PGM does not work by cluster recognition). I have, therefore, modified the commercial protocol provided by the manufacturer to build sequencing libraries adapted for low quality and low quantity DNA. The rationale of the changes performed was to account for the low starting DNA concentration and, more importantly, to minimize the loss of small DNA fragments, the main component of archival DNA samples. Our results provided a modified and cost effective protocol that can be used both on archival and fresh samples, without compromising quality of sequencing libraries. The modified protocol is *in press* in Methods in Molecular Biology (Springer protocols) and can be found in Annex 1.

Archival samples proved to be an invaluable source of information. Their DNA allowed insights into the genetic histories of populations, which are not accessible from field surveys and, more importantly, which may already have gone extinct. Quantity and quality of endogenous DNA was low, as expected, and fragments were no longer than 150 bp in all cases. However, it was possible to observe different sample 'behaviours' that are worth being mentioned. Samples from different tissues yielded generally different amounts of extracted DNA. The best performing samples were nasal bones and tissue collected from skeletons. From those sample types I regularly obtained full mitogenomes even without performing the hybridisation capture step. However, for the vast majority of samples, hybridisation capture was absolutely necessary in order to obtain sufficient

sequence information at a quality, which passed our quality filters. An illustrative comparison for Rusa samples is in Annex 2.

For both Manuscript I and III, I analysed only the mitochondrial DNA, which was sufficient to describe the relationships between genetic lineages within the species investigated. However, in Manuscript II the results indicated the presence of possible hybrid individuals. Their existence was only confirmed by the inclusion of microsatellite loci. We specifically selected for loci whose amplification product was no longer than 100 bp and that were already successfully amplified in other Cervidae or Bovidae species.

4 Concluding Remarks and Outlook

Understanding the phylogeography and evolutionary histories of species contained within biodiversity hotspots provides a powerful tool to comprehend the influence of past events in shaping the current distribution of genes, populations, species and communities. Moreover, determining the genetic relationships among species or populations allows us to identify their unique responses, providing an historical context to how adaptations arose (Losos, et al. 2013). These regions, and thus the plethora of ecological niches, endemic species and unique genetic diversity they comprise, are also the most susceptible to human made changes (Sechrest et al. 2002; Mittermeier et al. 2011). In particular, while Sundaland is among the most biodiverse areas on earth it is also one of the most anthropogenically disturbed (Sodhi et al. 2004).

In this dissertation, I have investigated and compared the phylogeographical patterns of three groups of ungulate species throughout their distribution in South and Southeast Asia. The results reported in the manuscripts and discussed here show how a shared history has shaped different responses in species with different ecological needs. I have identified that land corridors, exposed during past glacial periods, were not equally permissive to all species. Open-habitat adapted species and generalists were able to utilize these available connections, at least once, as gene flow between populations occurring in now isolated Sundaic islands was observed for the red muntjacs, Rusa deer and for the Javan rhino. Conversely, forest dwelling species showed little evidence of dispersal during the drier and cooler climate of the Pleistocene glacial periods in central Sundaland. This result is especially important when considering the human impact of logging and deforestation in Southeast Asia, which will result in diminishing rainforests and more open-habitat. Additionally, I have found genetic lineages of conservation importance (as discussed in point 2 above), as they represent units of unique genetic diversity and point to the importance of separate management.

However, the power of a comparative phylogeographic framework increases with the increasing number of patterns that can be compared, our understanding of species adaptability and a realistic number of unique lineages of conservation importance. Therefore, other species that have their distribution tightly linked to the geological history of the Sundaland should be included, in particular ungulate species, as they provide ecological services which will disappear with further biodiversity loss. Finally, the inclusion of nuclear genes was absolutely necessary for resolving the intriguing phylogeographic patterns of the Rusa deer investigated in Manuscript II. As prospective future work with the development of sequencing techniques, more nuclear loci should become accessible from archival samples. For example, the development of RAD-seq and Genotyping By Sequencing techniques have made it possible to access the nuclear genomes of non-model species using archival samples (e.g. Tin et al, 2014). The study of nuclear genes will have important implications in understanding direct adaptations of species to their changing environment. Combining these techniques will provide researchers with a powerful historical framework on which to make informed decisions about habitat and species conservations. Still, without direct translation of the scientific findings to the policy makers and species specialists groups, these efforts will be in vain.

Bibliography

Ahmad Zafir, A.W., Payne, J., Mohamed, A., Lau, C.F., Sharma, D.S.K., Alfred, R., Williams, A.C., Nathan, S., Ramono, W.S., Clements, G.R. (2011) Now or never: what will it take to save the Sumatran rhinoceros *Dicerorhinus sumatrensis* from extinction? *Oryx*, **45**, 225-233.

Ahrestani, F.S. and Sankaran, M. (2016) Chapter I: The large herbivores of South and Southeast Asia – A prominent but neglected guild. *in* The Ecology of Large Herbivores in South and Southeast Asia (Vol. 225). Springer.

Austin, J.J., Melville, J. (2006) Incorporating historical museum specimens into molecular systematic and conservation genetics research. *Molecular Ecology Notes*, **6**, 1089-1092.

Avise, J.C., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E., Reeb, C.A. and Saunders, N.C. (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual review of ecology and systematics*, **18**, 489-522.

Bacon, C.D., Michonneau, F., Henderson, A.J., McKenna, M.J., Milroy, A.M. and Simmons, M.P. (2013) Geographic and taxonomic disparities in species diversity: dispersal and diversification rates across Wallace's Line. *Evolution*, **67**, 2058-2071.

Baltzer, J.L., Davies, S.J., Noor, N.S.M., Kassim, A.R. and LaFrankie, J.V. (2007) Geographical distributions in tropical trees: can geographical range predict performance and habitat association in co-occurring tree species? *Journal of Biogeography*, **34**, 1916-1926.

van den Bergh, G.D., de Vos, J. Sondaar, P.Y. (1996) The Late Quaternary palaeogeography of mammal evolution in the Indonesian Archipelago. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **171**, 385-408.

Bird, M.I., Taylor, D., Hunt, C. (2005) Paleoenvironments of insular Southeast Asia during the Last Glacial Period: a savannah corridor in Sundaland? *Quaternary Science Reviews*, **24**, 2228-2242.

Bland, L.M., Collen, B.E.N., Orme, C.D.L. and Bielby, J.O.N. (2015) Predicting the conservation status of data-deficient species. *Conservation Biology*, 29, 250-259.

De Bruyn, M., Nugroho, E., Mokarrom Hossain, Md., 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.

Burrell, A.S., Disotell, T.R. and Bergey, C.M. (2015) The use of museum specimens with high-throughput DNA sequencers. *Journal of human evolution*, **79**, 35-44.

Cannon, C.H., Morley, R.J., Bush, A.B.G. (2009) The current refugial rainforests of Sundaland are unrepresentative of their biogeographic past and highly vulnerable to disturbance. *PNAS*, **106**, 11188-11193.

Chingangbam, D.S., Laishram, J.M. and Suzuki, H. (2015) Molecular phylogenetic characterization of common murine rodents from Manipur, Northeast India. *Genes & genetic systems*, **90**, 21-30.

Dejtaradol, A., Renenr, S.C., Karapan, S., Bates, P.J., Moyle, R.G. and Päckert, M. (2016) Indochinese-Sundaic faunal transition and phylogeographical divides north of the Isthmus of Kra in Southeast Asian Bulbus (Aves: Pycnonotidae). *Journal of Biogeography*, **43**, 471-483.

Demos, T.C., Achmadi, A.S., Giarla, T.C., Handika, H., Rowe, K.C. and Esselstyn, J.A. (2016) Local endemism and within-island diversification of shrews illustrate the importance of speciation in building Sundaland mammal diversity. *Molecular Ecology*, **25**, 5158-5173.

Dennel, R.W. (2004) Hominid dispersals and Asian biogeography during the lower and Early Middle Pleistocene, c. 2.0-0.5 Mya. *Asian Perspectives*, **43**.

Dinerstein, E. (2011) Family Rhinocerotidae (Rinoceroses). *in* Handbook Mammals of the World - Vol. 2 Hoofed Mammals, Barcelona, Lynx Edicions, 144-181.

Ebach, M.C. (2015) Origins of biogeography: the role of biological classification in early plant and animal geography (Vol. 13). Springer.

Fortes, G.G. & Paijmans, J.L.A. (2015) Analysis of whole mitogenomes from ancient samples. *Whole Genome Amplification Methods Protocols*, 179–195.

Frantz, L., Meijaard, E., Gongora, J., Haile, J., Groenen, M.A. and Larson, G. (2016) The evolution of Suidae. *Annual review of animal biosciences*, **4**, 61-85.

Fuchs, J., Ericson, P.G. and Pasquet, E. (2008) Mitochondrial phylogeographic structure of the white-browed piculet (*Sasia ochracea*): cryptic genetic differentiation and endemism in Indochina. *Journal of Biogeography*, **35**, 565-575.

Gavin, T.A. (1991) Why ask" why": the importance of evolutionary biology in wildlife science. *The Journal of wildlife management*, 760-766.

Gong, W., Liu, W., Gu, L., Kaneko, S., Koch, M.A. and Zhang, D. (2016) From glacial refugia to wide distribution range: demographic expansion of *Loropetalum chinense* (Hamamelidaceae) in Chinese subtropical evergreen broadleaved forest. *Organisms Diversity & Evolution*, **16**, 23-38.

Goossens, B., Salgado-Lynn, M., Rovie-Ryan, J.J., Ahmad, A.H., Payne, J., Zainuddin, Z.Z., Nathan, S.K.S.S. and Ambu, L.N. (2013) Genetics and the last stand of the Sumatran rhinoceros *Dicerorhinus sumatrensis*. *Oryx*, **47**, 340.

Gorog, A.J., Sinaga, M.H. and Engstrom, M.D. (2004) Vicariance or dispersal? Historical biogeography of three Sunda shelf murine rodents (*Maxomys surifer*, *Leopoldamys sabanus* and *Maxomys whiteheadi*). *Biological Journal of the Linnean Society*, **81**, 91-109.

Groves, C. (2003) Taxonomy of Ungulates of the Indian Subcontinent. *Bombay Natural History Society*, **100**, 341–61.

Groves, C.P., Grubb, P. (2011) Ungulate Taxonomy. The Jonhs Hopkins University Press.

Gupta, S.K., Kumar, A., Gaur, A. and Hussain, S.A. (2015) Detection of 40 bp insertion-deletion (INDEL) in mitochondrial control region among sambar (*Rusa unicolor*) populations in India. *BMC research notes*, **8**, 581.

Hanson, C.B. (1989) *Teletaceras radinskyi*, a new primitive rhinocerotid from the late Eocene Clarno Formation, Oregon. *in* The evolution of Perissodactyls, 379-398.

Hedges, S., Duckworth, J.W., Timmins, R.J., Semiadi, G., Dryden, G. (2015) *Rusa timorensis*. IUCN Red List Threatened Species, e.T41789A22156866.

Heinsohn, T. (2002) Animal translocation: long-term human influences on the vertebrate zoogeography of Australasia (natural dispersal versus ethnophoresy). *Australian Zoologist*, **32**, 351-376.

Hirase, S., Takeshima, H., Nishida, M. and Iwasaki, W. (2016) Parallel mitogenome sequencing alleviates random rooting effect in phylogeography. *Genome biology and evolution*, **8**, 1267-1278.

Hofreiter, M. and Stewart, J. (2009) Ecological change, range fluctuations and population dynamics during the Pleistocene. *Current biology*, **19**, R584-R594.

Huxley, T.H. (1868) On the Classification and Distribution of the Alectoromorphae and Heteromorphae. *Proceedings of the Zoological Society of London*, 294–319.

Iyengar, A., Babu, V.N., Hedges, S., Venkataraman, A.B., Maclean, N. and Morin, P.A. (2005) Phylogeography, genetic structure, and diversity in the dhole (*Cuon alpinus*). *Molecular Ecology*, **14**, 2281-2297.

Jha C.S., Dutt C.B.S., Bawa K.S. (1995) Deforestation and land use changes in Western Ghats , India. *Current Science*, **79**, 231–43.

Karanth K.P. (2003) Evolution of disjunct distributions among wet-zone species of the Indian subcontinent: Testing various hypotheses using a phylogenetic approach. *Current Science*, **85**, 1276–83.

Lari, M., Rizzi, E., Mona, S., Corti, G., Catalano, G., Chen, K., Vernesi, C., Larson, G., Boscato, P., De Bellis, G. and Cooper, A. (2011) The complete mitochondrial genome of an 11,450-year-old aurochsen (*Bos primigenius*) from Central Italy. *BMC evolutionary biology*, **11**, p.32.

Latinne, A., Meynard, C.N., Herbreteau, V., Waengsothorn, S., Morand, S., Michaux, J.R. (2015) Influence of past and future climate changes on the distribution of three Southeast Asian murine rodents. *Journal of Biogeography*, **42**, 1714-1726.

Leonard, J.A., Tex, R.J., Hawkins, M.T., Muñoz-Fuentes, V., Thorington, R. and Maldonado, J.E. (2015) Phylogeography of vertebrates on the Sunda Shelf: A multi-species comparison. *Journal of Biogeography*, **42**, 871-879.

Li, J.J., Shu, Q., Zhou, S.Z., Zhao, Z.J. and Zhang, J.M., 2004. Review and prospects of Quaternary glaciation research in China. *Journal of Glaciology and Geocryology*, **26**, 235-243.

Liao, P.C., Chiang, Y.C., Huang, S. and Wang, J.C. (2009) Gene flow of *Ceriops tagal* (Rhizophoraceae) across the kra isthmus in the Thai Malay peninsula. *Botanical Studies*, **50**, 193-204.

Lieberman, B.S., 2003. Paleobiogeography: the relevance of fossils to biogeography. *Annual Review of Ecology, Evolution, and Systematics*, 34, 51-69.

Lim, H.C., Rahman, M.A., Kim, S.L.H. Moyle, R.G., Sheldon, F.H. (2010) Revisiting Wallace's haunt: coalescent simulations and comparative niche modelling reveal historical mechanisms that promoted avian population divergence in the Malay Archipelago. *Evolution*, **65**, 321-334.

Lohman, D.J., de Bruyn, M., Page, T., von Rintelen, K., Hall, R., Ng, P.K.L., Shih, H.-T., Carvalho, G.R., von Rintelen, T. (2016) Biogeography of the Indo-Australian Archipelago. *Annual Reviews of Ecology, Evolution and Systematics*, **42**, 205-226.

Losos, J.B., Arnold, S.J., Bejerano, G., Brodie III, E.D., Hibbett, D., Hoekstra, H.E., Mindell, D.P., Monteiro, A., Moritz, C., Orr, H.A. and Petrov, D.A. (2013) Evolutionary biology for the 21st century. *PLoS Biology*, **11**, e1001466.

Lu, L., Ge, D., Chesters, D., Ho, S.Y., Ma, Y., Li, G., Wen, Z., Wu, Y., Wang, J., Xia, L. and Liu, J. (2015) Molecular phylogeny and the underestimated species diversity of the endemic white-bellied rat (Rodentia: Muridae: Niviventer) in Southeast Asia and China. *Zoologica Scripta*, **44**, 475-494.

Luo, S.-J., Zhang, Y., Johnson, W.E., Miao, L., Martelli, P., Antunes, A., Smith, J.L.D. and O'Brien, S.J. (2014) Sympatric Asian felid phylogeography reveals a major Indochinese-Sundaic divergence. *Molecular Ecology*, **23**, 2072–2092.

Lydekker, R. (1898) The deer of all lands: a history of the family Cervidæ living and extinct. Rowland Ward, Limited, London, United Kingdom.

Maricic, T., Whitten, M. & Pääbo, S. (2010) Multiplexed DNA Sequence Capture of Mitochondrial Genomes Using PCR Products. *PLoS One,* **5**, e14004.

Mattioli S. (2011) Family Cervidae (Deer) *in* Handbook Mammals of the World - Vol. 2 Hoofed Mammals, Barcelona, Lynx Edicions, 350–443

McLeod, D.S. (2010) Of least concern? Systematics of a cryptic species complex: *Limnonectes kuhlii* (Amphibia: Anura: Dicroglossidae). *Molecular Phylogenetics and Evolution*, **56**, 991-1000.

De Meester, L., Vanoverbeke, J., Kilsdonk, L.J., Urban, M.C. (2016) Evolving perspectives on Monopolization and Priority effects. *Trends in Ecology and Evolution*, **31**, 136-146.

Meijaard, E. (2003) Mammals of southeast Asian islands and their Late Pleistocene environments. *Journal of Biogeography*, **30**, 1245-1257.

Meijaard, E. and Groves, C.P. (2004) Morphometrical relationships between South-east Asian deer (Cervidae, tribe Cervini): Evolutionary and biogeographic implications. *Journal of Zoology*, **263**, 179-196.

Meijaard, E. and Groves, P. (2006) The geography of Mammals and rivers in mainland Southeast Asia *in* Primate Biogeography (eds. S.M. Lehman and J.G. Fleagle). Springer, New York.

Merckx, V., Hendriks, K.P., Beentjes, K.K. et al. (2015) Evolution of endemism on a young tropical mountain. *Nature*, **524**, 347-350.

Meyer, M., Kircher, M. (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *in* Cold Spring Harbor Protocols, pdb-prot5448.

Mishra, S., Gaillard, C., Hertler, C., Moigne, A.M. and Simanjuntak, T. (2010) India and Java: contrasting records, intimate connections. *Quaternary International*, **223**, 265-270.

Mittermeier, R.A., Turner, W.R., Larsen, F.W., Brooks, T.M. and Gascon, C. (2011) Global biodiversity conservation: the critical role of hotspots. *in* Biodiversity hotspots (3-22). Springer Berlin Heidelberg.

Morales, J.C., Andau, P.M., Supriatna, J., Zainuddin, Z.Z. & Melnick, D.J. (1997) Mitochondrial DNA variability and conservation genetics of the Sumatran rhinoceros. *Conservation Biology*, **11**, 539-543.

Moyle, R.G., Andersen, M.J., Oliveros, C.H., Steinheimer, F. and Reddy, S. (2012) Phylogeny and biogeography of the core babblers (Aves: Timaliidae). *Systematic Biology*, **61**, 63–651.

Myers, N. (1988) Threatened biotas:" hot spots" in tropical forests. *Environmentalist*, **8**, 187-208.

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

Paijmans, J.L.A., Gilbert, M.T.P., Hofreiter, M. (2012) Mitogenomic analysis from ancient DNA. *Molecular Phylogenetics and Evolution*, **69**, 404-416.

Papadopoulou, A. and Knowles, L. (2015) Species-specific responses to island connectivity cycles: refined models for testing phylogeographic concordance across a Mediterranean Pleistocene Aggregate Island Complex. *Molecular Ecology*, **24**, 4252-4268.

Parnell, J. (2013) The biogeography of the Isthmus of Kra region: a review. *Nordic Journal of Botany*, **31**, 001-015.

Patou, M.-L., Wilting, A., Gaubert, P., Esselstyn, J.A., Cruaud, C., Jennings, A.P., Fickel, J. & Veron, G. (2010) Evolutionary history of the Paradoxurus palm civets - a new model for Asian biogeography. Journal of Biogeography, **37**, 2077-2097.

Raes, N., Cannon, C.H., Hijmans, R.J., Piessens, T., Saw, L.G., van Welzen, P.C., Silk, W.F. (2014) Historical distribution of Sundaland's Dipterocarp rainforests at Quaternary glacial maxima. *PNAS*, **111**.

Ricklefs, R.E. (1987) Community diversity: relative roles of local and regional processes. *Science*, **235**, 167-171.

Ripple, W.J., Newsome, T.M., Wolf, C., Dirzo, R., Everatt, K.T., Galetti, M., Hayward, M.W., Kerley, G.I., Levi, T., Lindsey, P.A. and Macdonald, D.W. (2015) Collapse of the world's largest herbivores. *Science Advances*, **1**, e1400103.

Rohling, E.J., Fenton, M., Jorissen, F.J., Bertrand, P., Ganssen, G. and Caulet, J.P. (1998) Magnitudes of sea-level lowstands of the past 500,000 years. *Nature*, **394**, 162-165.

Ronquist, F. and Sanmartín, I. (2011) Phylogenetic methods in biogeography. *Annual Review of Ecology, Evolution, and Systematics*, **42**, 441-464.

Rowe, K.C., Singhal, S., Macmanes, M.D., Ayroles, J.F., Morelli, T.L., Rubidge, E.M., Bi, K.E. and Moritz, C.C. (2011) Museum genomics: low-cost and high-accuracy genetic data from historical specimens. *Molecular Ecology Resources*, **11**, 1082-1092.

Sathiamurthy, E.V.H.K. and Voris, H.K. (2006) Maps of Holocene sea level transgression and submerged lakes on the Sunda Shelf. *The Natural History Journal of Chulalongkorn University*, Supplement, **2**, 1-43.

Sechrest, W., Brooks, T.M., da Fonseca, G.A., Konstant, W.R., Mittermeier, R.A., Purvis, A., Rylands, A.B. and Gittleman, J.L.(2002) Hotspots and the conservation of evolutionary history. *Proceedings of the National Academy of Sciences*, **99**, 2067-2071.

Sheldon, F.H., Lim, H.C. and Moyle, R.G. (2015) Return to the Malay Archipelago: the biogeography of Sundaic rainforest birds. *Journal of Ornithology*, **156**, 91-113.

Sodhi, N.S., Koh, L.P., Brook, B.W. and Ng, P.K. (2004) Southeast Asian biodiversity: an impending disaster. *Trends in Ecology & Evolution*, **19**, 654-660.

van Strien, N.J., Steinmetz, R., Manullang, B., Sectionov, Han, K.H., Isnan, W., Rookmaaker, K., Sumardja, E., Khan, M.K.M., Ellis, S. (2008) *Rhinoceros sondaicus*. The IUCN Red List of Threatened Species, e.T19495A8925965.

Stuart, B.L., Inger, R.F. and Voris, H.K. (2006) High level of cryptic species diversity revealed by sympatric lineages of Southeast Asian forest frogs. *Biology Letters*, **2**, 470-474.

Suarez, A.V. and Tsutsui, N.D. (2004) The value of museum collections for research and society. *BioScience*, **54**, 66-74.

Shved, N., Haas, C., Papageorgopoulou, C., Akguel, G., Paulsen, K., Bouwman, A., Warinner, C. and Rühli, F. (2014) Post mortem DNA degradation of human tissue experimentally mummified in salt. *PloS one*, **9**, e110753.

den Tex, R.-J., Leonard, J.A. (2013) A molecular phylogeny of Asian barbets: Speciation and extinction in the tropics. *Molecular Phylogenetics and Evolution*, **68**, 1-13.

Timmins, R.J., Duckworth, J.W., Hedges, S. (2016) *Muntiacus muntjak*. The IUCN Red List of Threatened Species, e.T42190A56005589.

Tin, M.M.Y., Economo, E.P. and Mikheyev, A.S. (2014) Sequencing degraded DNA from non-destructively sampled museum specimens for RAD-tagging and low-coverage shotgun phylogenetics. *PloS one*, **9**, e96793.

Tougard, C., Delefosse, T., Hänni, C. and Montgelard, C. (2001) Phylogenetic relationships of the five extant rhinoceros species (Rhinocerotidae, Perissodactyla) based on mitochondrial cytochrome b and 12S rRNA genes. *Molecular phylogenetics and evolution*, **19**, 34-44.

Voris, H.K. (2000) Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. *Journal of Biogeography*, **27**, 1153-1167.

Wallace, A.R. (1863) On the physical geography of the Malay Archipelago. *The Journal of the Royal Geographical Society of London*, **33**, 217-234.

Wiens, J.J.(2012) Perspective: why biogeography matters: historical biogeography vs. phylogeography and community phylogenetics for inferring ecological and evolutionary processes. *Frontiers of Biogeography*, **4**.

Williams, M.A., Ambrose, S.H., van der Kaars, S., Ruehlemann, C., Chattopadhyaya, U., Pal, J. and Chauhan, P.R. (2009) Environmental impact of the 73ka Toba super-eruption in South Asia. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **284**, 295-314.

Wilting, A., Sollmann, R., Meijaard, E., Helgen, K.M., Fickel, J. (2012) Mentawai's endemic, relictual fauna: is it evidence for Pleistocene extinctions on Sumatra? *Journal of Biogeography*, **39**, 1608-1620.

Woodruff, D.S. (2010) Biogeography and conservation in Southeast Asia: how 2.7 million years of repeated environmental fluctuations affect today's patterns and the future of the remaining refugial-phase biodiversity. *Biodiversity and Conservation*, **19**, 919-941.

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.

Wurster, C.M., Bird, M.I., Bull, I.D., Creed, F., Bryant, C., Dungait, J.A.J. (2010) Forest contraction in north equatorial Southeast Asia during the Last Glacial Period. *PNAS*, **107**, 15508-15511.

Yuan, Z.Y., Suwannapoom, C., Yan, F., Poyarkov, N.A., Nguyen, S.N., Chen, H.M., Chomdej, S., Murphy, R.W. and Che, J. (2016) Red River barrier and Pleistocene climatic fluctuations shaped the genetic structure of *Microhyla fissipes* complex (Anura: Microhylidae) in southern China and Indochina. *Current Zoology*, **42**.

Appendix

Annex 1

Protocol for sequencing libraries for non-Illumina platforms

In press *Methods in Molecular Biology: aDNA methods and Protocols.*

Library preparation from degraded samples for non-Illumina platforms

Renata Martins[1], Marie-Louise Kampmann[1,2] and Daniel W. Förster[1]

[1] Leibniz Institute for Zoo and Wildlife Research (IZW), Department of Evolutionary Genetics, 10315 Berlin, Germany

[2] Current address: Section of Forensic Genetics, Department of Forensic Medicine, University of Copenhagen, 2100 Copenhagen, Denmark

Corresponding author: Renata Martins

Running head: Library preparation for Ion Torrent PGM

Abstract

Efficient methods for building genetic libraries from degraded DNA are in place for Illumina sequencing platforms for some years now, but are still lacking for other sequencing platforms like the Ion Torrent PGM. In this chapter, we provide a protocol for building sequencing libraries from degraded DNA (archival or ancient sample material) for the Ion Torrent™ high-throughput sequencing platforms. Aside from reduced time and costs in comparison to commercial kits, this protocol removes purification steps before amplification, an important consideration for work involving historical samples. Sequencing libraries prepared using this method are ready for shotgun sequencing or enrichment methods.

Key Words: archival DNA; degraded samples; Ion Torrent™ PGM; non-Illumina platforms; Sequencing libraries

1. Introduction

DNA obtained from archival or ancient samples is often highly degraded due to post-mortem fragmentation of DNA strands, rendering more traditional PCR methods ineffective for obtaining sequence data (1). High-throughput sequencing technologies, on the other hand, can make use of such short DNA fragments. However, most of the protocols developed for this purpose focus on

the Illumina sequencing platforms and methods suitable for ancient DNA (aDNA) samples have yet to be developed for other high-throughput sequencing platforms.

The Ion Torrent™ Personal Genome Machine (PGM) system is a high-throughput sequencing platform that carries out sequencing by synthesis (SBS) using real-time measurement of hydrogen ions released during nucleotide incorporation (2). The chip is flooded with a single nucleotide solution at a time; each nucleotide added during DNA replication results in a pH change which is then measured by sensors and directly converted into a DNA sequence. An advantage of this sequencing platform is its scalability – the throughput can be adjusted by choosing different capacity chips. For example, a 'smaller' chip like the Ion 314™ generates less data (100 Mb of sequencing data) and is therefore well suited for a low number of samples, while 'large' chips like the Ion 318™ generate up to 2 Gb of sequence data, suitable, for example, for larger sample sets.

Here, we present a protocol to build sequencing libraries for the Ion Torrent™ PGM, suitable for degraded DNA samples. The priority for modifying the commercial protocols for use with aDNA was to reduce template loss. This was achieved by the removal of all purification steps before PCR amplification of final libraries, as the increased length of added adapters ought to be sufficient to pass the cut-off limits of purification kits (3). The protocol can also be applied to modern DNA samples, with the advantage of lower costs and time of library preparation compared to the commercial protocol provided by the manufacturer.

2. Materials

To avoid any external contamination of archival and ancient samples, sequencing libraries should be prepared in a laboratory dedicated to the work with such material (clean lab). Amplification of the sequencing libraries should thus be performed outside the clean lab, in the modern DNA lab (1). Unless stated otherwise, reagents should be kept at -20 °C until usage and reactions should be set up on ice.

2.1 Oligonucleotides and amplification primers as adapted from the commercial kits:

Step	Component	Sequence[a]	Size (nt)	Remarks
P1 adapter	P1 adapter 5'-3'	CCA CTA CGC CTC CGC TTT CCT CTC TAT GGG CAG TCG GTG AT	41	HPLC purification during synthesis
	P1 adapter 3'-5'	ATC ACC GAC TGC CCA TAG AGA GGA AAG CGG AGG CGT AGT GG*T*T	43	HPLC purification during synthesis

Barcoded Adapters	PGM Barcode 1 5'-3'	CCA TCT CAT CCC T*G*C GTG TCT CCG ACT CAG **CTA AGG TAA C**GAT	43	HPLC purification during synthesis
	PGM Barcode 1 3'-5'	ATC GTT ACC TTA GCT GAG TCG GAG ACA CG*C	30	HPLC purification during synthesis
	PGM Barcode 2 5'-3'	CCA TCT CAT CCC T*G*C GTG TCT CCG ACT CAG **TAA GGA GAA C**GA T	43	HPLC purification during synthesis
	PGM Barcode 2 3'-5'	ATC GTT CTC CTT ACT GAG TCG GAG ACA CG*C	30	HPLC purification during synthesis
	PGM Barcode 3 5'-3'	CCA TCT CAT CCC T*G*C GTG TCT CCG ACT CAG **AAG AGG ATT C**GA T	43	HPLC purification during synthesis
	PGM Barcode 3 3'-5'	ATC GAA TCC TCT TCT GAG TCG GAG ACA CG*C	30	HPLC purification during synthesis
	PGM Barcode 4 5'-3'	CCA TCT CAT CCC T*G*C GTG TCT CCG ACT CAG **TAC CAA GAT C**GAT	43	HPLC purification during synthesis
	PGM Barcode 4 3'-5'	ATC GAT CTT GGT ACT GAG TCG GAG ACA CG*C	30	HPLC purification during synthesis
	PGM Barcode 5 5'-3'	CCA TCT CAT CCC T*G*C GTG TCT CCG ACT CAG **CAG AAG GAA C**GAT	43	HPLC purification during synthesis
	PGM Barcode 5 3'-5'	ATC GTT CCT TCT GCT GAG TCG GAG ACA CG*C	30	HPLC purification during synthesis
Amplification	Primer A_amp	CCA TCT CAT CCC TGC GTG TC	20	
	Primer P1_amp	CCA CTA CGC CTC CGC TTT CCT CTC TAT G	28	

* phosphorotioate bond in the sequence

[a] Bases in bold indicate the barcode sequence; for barcode 6 - 48 please see reference 2

2.2 Reagents

1. Molecular grade water (ddH_2O)

2. Ion Xpress™ Plus Fragment Library 200 Kit (Ion Xpress™ kit; Thermo Fischer)

3. AmpliTaq Gold (Thermo Fischer, supplied with 10x AmpliTaq buffer and 25 mM $MgCl_2$)

4. Oligo Hybridisation Buffer (10X): 500 mM NaCl, 10 mM Tris-HCl (pH 8.0), 1 mM EDTA (pH 8.0), stored at 4 °C (4)

5. Bovine Serum Albumin (BSA)

6. SPRIselect magnetic beads (Beckman & Coulter), stored at 4 ° C

7. Molecular Grade Ethanol (100%), stored at room temperature

8. Ion PGM™ Template OT2 200 kit for template preparation

9. Ion PGM™ IC 200 kit for library sequencing

10. Ion PGM™ Chips, which can vary in capacity depending on target size and number of samples, stored at room temperature

11. (*Optional*) Pippin Prep 2% Agarose Gel cassette (Sage Science)

2.3 Equipment

1. Thermocyclers, one in the clean lab and one in the modern DNA lab.

2. Microcentrifuge, in the clean lab.

3. Magnetic rack, one in the clean lab and one in the modern DNA lab.

4. Real-time PCR (qPCR) thermal cycler in the modern DNA lab.

5. Ion Torrent™ template preparation equipment: either OneTouch™ or Ion Chef™.

3. Methods

Before library preparation, check for successful DNA extraction. It is strongly encouraged to incorporate library blanks during each step of the protocol.

3.1 Adapter and barcode preparation

1. Prepare *P1 adapter mix* as follows (200 µM):

Oligo Hybridization Buffer 10x 10 µL

P1 adapter 5'-3' 40 µL

P1 adapter 3'-5' 40 µL

ddH_2O 10 µL

2. Prepare *Barcoded adapter mix* as follows (200 µM):

Oligo Hybridization Buffer 10x 10 µL

PGM barcode X 5'-3' 40 µL

PGM barcode X 3'-5' 40 µL

ddH_2O 10 µL

3. Incubate each reaction mix for 10 sec at 95 °C and then ramp down from 95 °C to 12 °C at a rate of 0.1 °C/sec.

4. Create "ready to use" solutions of 25 µM of *P1 adapter* and of *each barcoded* adapter by diluting 2.5 µL of product with 17.5 µL of ddH_2O.

3.2 Blunt-ending DNA fragments

For starting DNA concentration and volume modifications, please see Notes 1 and 2.

1. Set up master mix for blunt-end repair. Include per reaction:

- 5.0 µL of 5x End Repair Buffer (Ion Xpress™ kit)
- 0.25 µL of Blunt Ending Enzyme (Ion Xpress™ kit)

2. Distribute 5.25 µL of the mix in each labelled PCR tube.

3. Add 19.5 µL of DNA extraction in each individual PCR tube (or bring final volume to 25 µL by adding ddH_2O, according with initial template volume).

4. Incubate the mixture in a thermocycler for 20 min at 20°C and then heat-inactivate the enzyme by incubating for 20 min at 72°C (see Note 3).

3.3 Adapter ligation

1. Remove tube with blunt-ended DNA from the thermocycler and keep on ice.

2. Prepare Adapter ligation reaction mix, including per reaction:

- 2.5 µL of 10x Ligase Buffer (Ion Xpress™ kit)
- 1.0 µL of P1 adapter
- 0.5 µL of dNTP Mix
- 0.5 µL of DNA Ligase
- 2.0 µL of Nick Repair Polymerase (Ion Xpress™ kit)

3. Distribute 6.5 µL of the mix into a new PCR tube.

4. Add 1.0 µL of each of the chosen PGM Barcode X into PCR tube of each respective sample.

5. Add 20.5 µL of blunt-ended DNA from step 3.2.

6. Incubate the mixture in a thermocycler for 15 min at 25°C, followed by 5 min at 72°C (see Note 3).

3.4 Bead purification

This purification step can be performed at room temperature (see Note 4).

1. Prepare a fresh 70% Ethanol solution, with a final volume of at least 500 µL × number of samples.

2. Add 40 µL of SPRIselect beads to the product from step 3.3, for a final ratio of 1.8x (broad range of recovered fragment size and upper limit set at 200 bp).

3. Mix thoroughly by pipetting and let the tubes incubate for 1 min in the magnetic rack.

4. Remove supernatant without disturbing the magnetic beads.

5. Add 500 µL of freshly prepared 70% ethanol to each sample, mix well, and let the tubes incubate until all beads are captured by the magnetic rack.

6. Remove supernatant without disturbing the magnetic beads and let them air dry, but for a maximum of 5 minutes.

7. Add 20 µL of ddH_2O to elute DNA, mix well and let the tubes rest for another minute in the magnetic rack.

8. Remove elution and transfer it into a new (labelled) tube, keep on ice until next reaction.

3.5. Amplification

1. Dilute primers by adding 10 µL of primer stock solution (at 100 µM) in 90 µL of ddH_2O.

2. For the amplification of libraries, set up a master mix, with per parallel reaction:

- 7.8 µL of ddH_2O
- 2.0 µL of AmpliTaqBuffer 10x
- 0.2 µL of dNTPs (25 mM each)
- 1.6 µL of MgCl2 (25 mM)
- 0.2 µL of BSA (10mg/mL)
- 1.5 µL of each primer (10 µM each)
- 0.2 µL of AmpliTaq Gold (5U/µL)

3. Distribute 15 µL in each PCR tube and add 5.0 µL of the purified product from step 3.4 (see Note 5).

4. In the modern DNA lab set up and run the thermocycler with the following protocol: denaturation for 10 min at 94 °C, followed by 15 cycles of 30 sec at 94 °C, 45 sec at 60 °C and 45 sec at 72 °C. Final extension occurs at 72 °C for 5 min (see Note 6).

5. Pool all parallel amplifications for the same sample and repeat step 3.4 with a 1.0x ratio, eluting the sequencing libraries in 20 µL (see Note 7).

6. Visualize library concentration and fragment size distribution with e.g. Tapestation HS D1000 or 3% Agarose gel (see Note 8).

3.6 Sequencing

1. Pool all libraries in equimolar amounts and quantify the molarity of the complete range of pooled fragments, with e.g. Tapestation HS1000 (See Note 9).

2. Dilute the library pool to 20 - 23 pM and prepare the chips either using OneTouch™ or Ion Chef™. Sequence the loaded chips within 24 hours.

4. Notes

1. Commercial protocols for library building provided by Thermo Fischer are optimized for only two initial DNA concentrations (1 µg and 100ng). Often such high concentrations cannot be obtained from low quality samples, so we suggest starting with 100 ng or the highest possible volume of extracted DNA.

2. Although starting DNA concentration can vary, do not dilute the initial DNA extraction, but rather add water accordingly to bring the mix up to the final volume.

3. Utilizing heat-inactivation of enzymes allows for the elimination of purification steps prior to amplification. This avoids unnecessary loss of short DNA fragments during purification.

4. We have tested silica column based purifications (e.g. MinElute, Qiagen) and have concluded that column based methods are less permissive in retaining smaller fragments, because the columns are optimized for a cut-off of 80 bp. The advantage of bead purification is the possibility to adapt the ratio of beads to use, to have a larger range of fragments recovered.

5. The volumes suggested here are based on four parallel amplifications per sample. If you adjust the number of parallel reactions, adjust likewise volumes of template DNA.

6. Estimate the optimum number of cycles for each sample by qPCR using primers A_amp and P1_amp. It is noteworthy that increasing the number of cycles much above the suggested 15 cycles increases fragment clonality, which in turn decreases the amount of data obtained.

7. After pooling and purifying each library there should be a final amount of product with around 10 nM, although less final molarity is also acceptable due to the low starting amount necessary for loading each sequencing chip. If an adapter dimer peak is visible even after purification, include it in the calculations for concentration of the final sequencing pool.

8. Sequencing libraries prepared using this method are ready to be sequenced. However, if the fragment size range after pooling is too wide or an adapter/primer dimer is still present in the sequencing pool, it is suggested to include a size selection step (e.g. using Pippin Prep or a gel extraction). This step, while costly, increases sequencing success.

9. Do not dilute each individual library to final concentration, as small volumes are prone to pipetting errors. Dilute each sample to 4 or 2 nM and pool them in equimolar amounts. After measuring the region molarity, do serial dilutions of the pool until the final concentration of 20-23 pM.

References

1. Pääbo S., Poinar H., Serre D., et al. (2004) Genetic analyses from ancient DNA. *Annual Review of Genetics* **38**(1), 645-679

2. Ion torrent by Thermo Fischer. (2016) Ion Xpress Plus gDNA Fragment Library Preparation – User Guide. *Thermo Fischer* ***MAN0009847***

3. Taylor P.G. (1996) Reproducibility of ancient DNA sequences from extinct Pleistocene fauna. *Molecular Ecology and Evolution* **13**,283-285

4. Maricic T., Whitten M., Pääbo S. (2010) Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *Plos ONE* **5**(11): e14004

Annex 2

Table A2 - Comparison of sequencing results of archival samples with and without (shotgun) sample enrichment, through hybridization capture.

Quality trimmed reads are the total number of reads after trimming of adapters and low quality sliding windows; Mapped unique reads refer to the reads that mapped against the given sample, after removal of clonal reads; Selection efficiency is calculated by the relation of number of mapped reads per total of quality trimmed reads; Average depth indicates the mean depth of coverage of all bases; and finally % genome coverage is the total of the reference genome that was covered at least by five reads for each position.

	Quality trimmed reads		**Mapped unique reads**		**Selection efficiency**		**Average depth**		**% Genome coverage ≥ 5×**	
	Shotgun	Capture	Shotgun	Capture	Shotgun	Capture	Shotgun	Capture	Shotgun	Capture
RTI9	7,796,492	18,590,435	281	11,634	0.00%	23.35%	1.21	48.84	1.6	99.1
RTI13	10,483,040	58,507,539	417	6,488	0.01%	32.28%	1.85	26.21	8.9	97.6
RTI17	1,584,419	3,139,741	249	2,824	0.02%	51.40%	1.03	13.80	1.1	87.8
RTI18	42,506,828	5,597,083	1,011	5,893	0.00%	0.14%	4.16	21.98	42.4	91.2
RTI19	9,529,717	14,540,861	216	7,436	0.00%	13.91%	0.86	31.54	0.8	98.2
RTI20	4,564,949	4,291,035	1,028	7,836	0.03%	56.76%	4.07	41.32	37.3	97.9
RTI22	13,775,749	8,889,794	117	3,851	0.00%	11.28%	0.44	15.27	0	90.0
RTI30	1,243,809	4,359,511	903	5,660	0.09%	68.72%	2.92	21.77	22.2	80.1
RTI31	5,575,896	15,569,516	3,567	16,983	0.11%	19.98%	12.81	71.1	65.8	97.3
RTI32	10,252,226	53,432,886	751	7,443	0.01%	33.37%	2.34	27.07	16	94.9
RTI33	8,010,994	21,178,023	1,516	13,241	0.03%	29.36%	4.7	47.98	32	97.3
RTI37	11,242,537	2,166,717	61	6,037	0.00%	0.82%	0.25	31.77	0	96.8
RTI38	3,408,683	25,259,084	328	31,748	0.01%	50.49%	1.36	142.98	3.1	99

As shown for the samples illustrated in this example, the complete mitogenome was not recovered at our quality filters in any of the shotgun sequence runs. 'Captured' samples show an increase of selection efficiency by an average of 29% after performing enrichment capture. Even in the cases where efficiency was below 1% for captured samples, this increase was sufficient to obtain complete mitogenomes with a good average coverage depth.

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen oder indirekt übernommenen Daten sind unter Angabe der Quelle gekennzeichnet. Die Arbeit wurde bisher weder im In- noch im Ausland an einer anderen Hochschule vorgelegt.

Renata Martins

Potsdam, Juni 2017