Immunogenetics of extinct woolly mammoths and extant elephants

Inaugural-Dissertation
to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry, Pharmacy
of Freie Universität Berlin

by

John Alexander Galindo Puentes
from Bogotá-Colombia

Berlin, 2023

Research presented in this dissertation was carried out under the supervision of Prof. Alex D. Greenwood at the Leibniz Institute for Zoo and Wildlife Research from 12.08.2015 until 28.02.2023 and it is submitted to the Department of Biology, Chemistry and Pharmacy of Freie Universität Berlin.

1st Reviewer:
Prof. Alex D. Greenwood, PhD.
Department of Veterinary Medicine, Freie Universität Berlin Berlin, Germany
and Leibniz Institute for Zoo and Wildlife Research, Department of Wildlife Diseases

2nd Reviewer:
Prof. Dr. Jens Rolff
Institute of Biology - Zoology, Freie Universität Berlin Berlin, Germany

Date of defense: 06.07.2023

Dedicated to

My Mother

María Teresa Puentes

ACKNOWLEDGEMENTS

This work would not have been possible without the generous help of professors, colleagues, family, and friends. First of all, I would like to thank Professor Alex D Greenwood for giving me the opportunity to embark on this fascinating journey. I am deeply grateful for his efforts in understanding my points of view, perspectives and for helping me to improve my English. Thank you very much for helping me overcome all the obstacles that were presented in this project. This thesis has been a fascinating and an incredible intellectual challenge that makes all the effort invested on worthwhile.

Thanks to all the scientific, technical, and administrative staff of the IZW, especially to Stephanie Volberg and Susanne Auls for all the support received with the bureaucratic procedures. Thanks to all the current and past members of the Department of Wildlife Diseases for all their contributions during this period. Thanks Karin Hönig and Katja Pohle for continuous assistance in the Lab and Rosie Nekzad and Carin Hoffmann for their administrative support. I am especially grateful to Dr. David Alquezar for teaching me in the particulars of working in the ancient DNA lab and for all his scientific contributions. A special thank you to Dr. Gayle McEwen for her invaluable assistance with the bioinformatic analyses. To the postdoctoral researches Mathias, Anisha, and Shannon and all my fellows from previous and current batches: Niccoló, Marina, Ulrike, Azza, Sanatana, Peter, Daniela, Saba, Juan, and Pau, thanks for your constant support, I learned a lot from all of you. Daniela it was a pleasure to share "my office" with you. Thank you so much for everything you have done for me, I still miss our little quarrels. All my admiration and respect to all doctoral students and scientific staff, especially to my friends Dorina Meneghini, Lorena Derezanin, Ksennia Kravchenko, and Rohit “*parcero*” Chakravarty for all the contagious passion that you imprint on your work.

I am grateful for all the help and support from various groups that were involved during the development of this work. The elephant experiment was conducted in Prof. Roca's laboratory in the Department of Animal Sciences at the University of Illinois at Urbana-Champaign. Thank you, Professor Alfred Roca and Dr. Yasuko Ishida, you made me feel like one of the team. Your technical expertise and intellectual contribution made

the elephant analysis possible. I also had the opportunity to visit the McMaster Centre for Ancient DNA at McMaster University directed by Professor Hendrik Poinar, where I had the opportunity to train in molecular paleogenomics techniques with the support of Dr. Maria Lembring and Emil Karpinski. Thank you for being such a welcoming lab and thank you Professor Poinar for all the comments and feedback for the analysis and your support in writing the mammoth manuscript. The bioinformatic analyses would not have been possible without the enormous help and expertise of Dr. Jazmín Ramos-Madrigal, thanks Jazmín for providing me with your time for discussion and supportive feedback through our meetings. I would like to thank the Center for Evolutionary Hologenomics, GLOBE Institute, University of Copenhagen for allowing me to use their server. Likewise, I want to thank its director, Professor Thomas P. Gilbert for his valuable comments on the mammoth manuscript.

My study was also possible thanks to the contribution of Dr. Kathryn L. Perrin and Dr. Mads F. Bertelsen from the Department of Veterinary Clinical Sciences, Faculty of Health and Medical Sciences, University of Copenhagen who provided me with valuable captive Asian elephant samples. My gratitude goes out to Dr. Paul D. Ling from the Department of Molecular Virology and Microbiology at the Baylor College of Medicine, Houston, Texas. I would also like to acknowledge the contribution of Dr. Christina Hvilsom from Copenhagen Zoo who kindly provided us with muskox samples. I also wish to thank Dr. Paula Campos from the CIIMAR–Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Jesper Stenderup and Professor Eske Willrslev from Centre of Excellence in GeoGenetics, University of Copenhagen. For allowing me to carry out muskox sampling from the Late Pleistocene.

Honestly, this thesis would not have been possible without the unconditional support of Susanne Auls. When I was at the point of no return and there was only darkness, I counted on her helping hand to keep going. Susanne, you have been extremely generous to me. Thank you so much for our incredible conversations, for teaching me German, -despite being a bad student-, for your spark, your humor, your kindness, intelligence and for all the patience you have shown in helping me solve my most absurd issues and bureaucratic messes.

During the last four years I have been lucky to be the captain of the "Tennis Classics Team". Alexander Badry and Miguel Mendes Veiga, tough rivals, great friends, thank you "*amigos*" for all the time we have shared together. When I bit the clay and found myself in one of the hardest moments of my life, you literally lifted me off the ground and gave me the boost I needed not to give up, to overcome this match point with a broken knee but with my morale intact. Without your help and generosity this thesis would not have been possible. Thank you very much, and get your passports ready, Colombia is waiting for us.

Berlin is a wild, rebellious and magical city all at the same time. Providentially, a friend appears to make your life easier. Juan Sebastian, *"hermanito*", what can I say? Thank you so much for these years of friendship. I really enjoy the ease with which we can move from the most bizarre conversations to highly complex topics. Your intelligence, your analytical skills help me to keep growing intellectually, but what is most surprising is your noble heart. Although a better world seems more and more distant, people like you make me believe that it is worth fighting for a better country, a better world. *Alegría! por encima de la tristeza.*

Jazmin thank you for all your help in solving this molecular puzzle. You are an amazing scientist. I really enjoy our meetings. I always learn a lot, I admire the sharpness and clarity of your ideas, maybe that's why I always want to extend our meetings longer. Thank you very much for your time, patience and for being the pole to ground my theories. I hope to continue learning from you.

Nicolas Fasel, "*amiguito*", in an infinite world of probabilities, I was lucky to have crossed paths with you. Thank you so much for having transformed our tandem into this frenetic friendship. And as everything in life can't always be a party, thanks for spending time explaining statistics to me and for your feedback, which has enriched my perspective during our discussions.

Paula, *sumercé*, thank you so much for everything! I miss our coffee in the mornings, our walks through Tierpark and all our conversations so much. I love the passion you have for your work, how smart and persevering you are. Keep working and fighting for your dreams, you are the best pathologist. To Daniel, thank you so much for hosting me in Berlin and for dealing with two rounds of PhD-student sorrows. Thank you for the Champions nights. I love you both and wish you all the best in the world.

My acknowledgements to the Evolutionary Immunology group of the Universidad Nacional de Colombia led by Professor Luis Fernando Cadavid. My time at this great school was incredible not only for the training that I received to become a scientist and a better human being, but also for how happy I was there. Thank you very much to everyone who came to visit me in Berlin.

Hoy concluye mi viaje con los mamuts y los elefantes, una larga travesía con estos majestuosos animales que me ha dejado hermosas enseñanzas. Admiro su prodigiosa inteligencia, su excepcional memoria y su carácter matriarcal y, además, descubrí que muchos de ellos son zurdos. Pero lo que más me impresionó fue su altruismo, compasión, sus ritos mortuorios y el duelo que sufren por sus seres queridos. A pesar de migrar miles de kilómetros, siguen conectados con su manada por los genes, por su corazón.

Ya solo me queda agradecer a mi propia manada de familia y amigos. Más de siete años y 8,336 kilómetros entre Bogotá y Berlín y a pesar de esas enormes distancias siempre los he sentido presentes. Mauricio, Andrea, Nicolás, Diana G, Oscar, Henry y Angie, muchas gracias por tantos años juntos y todo su apoyo durante este reto. Blanquita Schroeder y señora Gladys muchas gracias por sus mensajes, siempre fueron en los momentos más oportunos. Rita Baldrich que felicidad es hablar contigo, siempre me rio un montón y haces que las penas duelan menos.

Este proceso fue un cambio de vida no solo para mi sino para mi mami y nos llevó a los dos a migrar fuera del país con diferentes caminos, pero siempre unidos por el corazón. Gracias mami por todos los sacrificios que ha hecho por mí, sumercé es el cimiento y el cielo de mi mundo, y fue por quien me mantuve en pie batallando, perseverando todos estos años. Mil gracias, Elizabeth, Ofelia, tía Julia y Rachel por cuidar de mi mami, mil gracias por sus llamadas, por sus mimos y tanto cariño que recibo cada vez que las visito. Tía Luz muchas gracias por todo lo que hace por nosotros, por sus llamadas y por sus oraciones. Gracias, tío Norberto por tenerme siempre presente. Gracias a todos mis primos Andrés, Olga, Ofelia, Francisco, Eduardo, Martha, Sergio, Catalina y Carolina y a sus familias, por crecer juntos y por todo su generoso apoyo. A la siguiente generación Camila, Angélica, Ana María, Eric, Daniela, Victoria, Valentina, Vanessa, Gustavo Andrés, Camila C, Nicolás a mis Marías Ma Alejandra y Ma Pazzz, a Valeria, Julián, Bianca y a Gabriel, a todos los extraño y los quiero un montón. Este es también un pequeño tributo a la memoria de los familiares que han dejado este plano terrenal. A mi padre que me inculcó el amor por la lectura, a mis abuelas y abuelos, tías y tíos, a mi tío Miguel y a mi tío Custodio a quienes no pude darles el último adiós.

“𝄞 ¿Quién dijo que todo está perdido?
Yo vine a ofrecer mi corazón 𝄞 ”

(English Translation)

Today my journey with the mammoths and the elephants concludes, a long journey with these majestic animals that has left me with beautiful lessons. I admire their prodigious intelligence, exceptional memory, and matriarchal character, and I also discovered that many of them are left-handed. However, what impressed me most was their altruism, compassion, mortuary rituals, and the mourning they suffer for their loved ones. Despite migrating thousands of kilometers, they remain connected to their herd by their genes, by their hearts.

I now only have to thank my own herd of family and friends. More than seven years and 8,336 kilometers between Bogota and Berlin, and despite these enormous distances, I have always felt them present. Mauricio, Andrea, Nicolás, Diana G, Oscar, Henry, and Angie, thank you very much for so many years together and all your support during this challenge. Blanquita Schroeder and Mrs. Gladys, thank you very much for your messages; they were always at the most opportune moments. Rita Baldrich, what happiness it is to talk to you, I always laugh a lot, and you make the sorrows hurt less.

This process was a life-changing experience not only for me but for my mommy as well, and it led us both to migrate out of our country with different paths, but always united by the heart. Thank you Mom, for all the sacrifices you have made for me, *sumercé* is the foundation and the heaven of my world, and it was for you that I remained standing, battling, persevering all these years. Thank you, Elizabeth, Ofelia, Aunt Julia, and Rachel, for taking care of my mom, thank you for your calls, for your cuddles, and for so much love that I receive every time I visit you. Aunt Luz, thank you very much for everything you do for us, for your calls, and for your prayers. Thank you, Uncle Norberto, for always keeping me in mind. Thanks to all my cousins Andrés, Olga, Ofelia, Francisco, Eduardo, Martha, Sergio, Catalina, and Carolina, and their families, for growing up together and for all their generous support. To the next generation, Camila, Angélica, Ana María, Eric, Daniela, Victoria, Valentina, Vanessa, Gustavo Andrés, Camila C, Nicolás, to my Marias Ma Alejandra and Ma Pazzz, to Valeria, Julián, Bianca, and Gabriel, I miss you all and love you a lot. This is also a small tribute to the memory of family members who have left this earthly plane. To my father who instilled in me the love of reading, to my grandmothers and grandfathers, aunts and uncles, to my Uncle Miguel and my Uncle Custodio, to whom I could not say the last goodbye.

“𝄞 Who said that everything is lost?

I came to offer my heart 𝄞 ”

DECLARATION OF INDEPENDENCE

Herewith, I certify that I have prepared and written my thesis independently and that I have not used any sources and aids other than those indicated by me.

Intellectual property of other authors has been marked accordingly. I also declare that I have not applied for an examination procedure at any other institution and that I have not submitted the dissertation in this or any other form to any other faculty as a dissertation.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SUPPLEMENTARIES

ZUSAMMENFASSUNG

Anthropogene Aktivitäten und Klimawandel sind die Hauptursachen für den Verlust von Biodiversität. Es wird geschätzt, dass das Artenaussterben bis 2050 40% erreicht (Thomas et al. 2004). Der Verlust der genetischen Diversität, Inzuchtdepression, und die Anhäufung von Deletionsmutationen sind grundlegende Mechanismen, die das Risiko zum Aussterben erhöhen können und somit das Anpassungspotential von vom Aussterben bedrohter Arten reduziert. Paleogenomik trägt zum Verständnis von Populationsdynamiken und der Evolution der Megafauna ausgestorbener Arten bei, indem es den Prozess, sowie genomische Konsequenzen des Aussterbens wiederspiegelt. Diese Informationen können für den Schutz rezenter Arten von großer Bedeutung sein. Jedoch können die Ursachen für das Aussterben auch abhängig von der jeweiligen Art sein. In dieser Doktorarbeit werden die evolutionären, immuno-genetischen Populationsdynamiken des ausgestorbenen Wollhaarmammuts (Mammuthus primigenius) und der drei rezenten Elefanten-Arten (Loxodonta africana, L. cyclotis und Elephas maximus) untersucht. Mammuts und Elefanten bilden die Ordnung der Rüsseltiere (Proboscidea), eine Ordnung, die nicht nur durch eine schnelle Diversifizierung, sondern auch durch Aussterbensereignisse geprägt worden ist und somit fundamentale Einblicke in die genetischen Hintergründe des Aussterbens und Überlebens nahverwandter Arten gewähren kann. Die räumlich-zeitliche Immundiversität des Mammuts während des späten Pleistozäns wurde mittels „in-solution target hybridization capture “und„ next-generation sequencing” untersucht. Die immungenetische Diversität von TLRs (toll-like receptors) und MHC (major histocompatibility complex) der drei rezenten Elefantenarten wurde mit einer neuen Hoch-Durchsatz-Multilocus-Genotypisierungsmethode (high throughput multilocus genotyping) untersucht. Nachteilige Allele wurden während des späten Pleistozäns in den Mammutpopulationen weitervererbt und in hoher Zahl erhalten. Dahingegen zeigte die immungenetische Diversität von TLRs und MHC der drei rezenten Elefantenarten, dass Inzucht und Mangel an Heterozygotie die genetische Diversität mindern. Dies könnte eine abgeschwächte Immunantwort zur Folge haben. Eine ausbalancierte Selektion könnte aber auch dafür sorgen, die MHC-Diversität zu erhalten. Während einige der Ergebnisse spezifisch für Elefanten sind, könnten andere auch als allgemeine Risikoindikatoren für das Aussterben anderer Arten gelten und somit relevant für den Artenschutz sein.

SUMMARY

Anthropogenic pressure and climate change are the main causes of biodiversity loss. It is estimated that by 2050 the extinction of living species could reach 40%. Loss of genetic diversity, inbreeding depression, and mutation deleterious accumulation are the underlying mechanisms that might increase extinction risk reducing the adaptive potential of the endangered populations. Paleogenomics has helped to understand the population dynamics and evolution of extinct megafauna, however, causes of extinction may be species-specific. This PhD thesis is focused to understand the evolutionary dynamics of the immunity in populations of the extinct woolly mammoth and the three extant elephant species. Mammoths and elephants represent the Proboscidean order, a group that has undergone rapid diversification and extinction processes, thus providing fundamental insight into the genetic drivers of both extinction and conservation in closely related species. The approach consisted in analyze the spatiotemporal mammoths immunodiversity during the Late Pleistocene using in-solution target hybridization capture and next-generation sequencing. As a result, deleterious alleles segregate in mammoth populations and were maintained at considerable frequencies during the Late Pleistocene. On the other hand, immunogenetic diversity of TLRs and MHC was measured in the three living elephant populations by incorporation of a novel high throughput multilocus genotyping. Inbreeding and heterozygosity deficiency are decreasing genetic diversity in elephant populations that may implicate a depletion of the immune response, interesting, balancing selection is acting as rebound effect to maintain the MHC diversity. The findings found in this study provides important insight to quantify genetic threats in extinct and endangered species.

1. CHAPTER I INTRODUCTION

1.1. EXTINCTION

1.1.1. Megafauna extinction

At the global scale, biodiversity is being lost in the middle of a six mass extinction and scientific evidence suggests that climate change and anthropogenic pressure are the main causes (Urban 2015; Thomas et al. 2004). This six mass extinction started in the Late Quaternary and has been estimated that over 300 mammal species have disappeared since then (Davis, Faurby, and Svenning 2018). The Quaternary period began 2.6 million years ago, is divided into the Pleistocene and the Holocene, and was dominated by glacial-interglacial oscillations. The Pleistocene (2.6 My to 11,000 years before the present (BP)) is marked by long cold glacial intervals (glaciations) of up to 100,000 years followed by short warm interglacial intervals of 10,000 years. The Holocene, an interglacial period occurred 11,000 years ago during the last glaciation and where human civilizations emerged (Pillans and Gibbard 2012).

Towards the Mid Pleistocene transition (MPT) ~1.2-06 million years ago, oscillation of warming and cooling episode, coinciding with an enigmatic mass extinction in the deep oceans, altering the biomass equilibrium and consequent erosion of diversity in marine biota (Barnosky 2008; Kender et al. 2016). Post MPT, glacial−interglacial periodicity from 41,000 to ∼100,000 years cycle was modified, and extinctions became more pronounced in the last 800,000 years than during earlier glacial−interglacial cycles. Extinct vegetation and fauna were replaced by migration and evolution of similar species (Zhou et al. 2018; Anthony John Stuart 2015). During the Late Pleistocene occurred an accelerated defaunation in most parts of the planet knows as the Quaternary Megafauna Extinction (QME), which occurred during the last glacial period, becoming stronger towards 50,000 years ago at the end of the Pleistocene. Near 65 % of terrestrial megafauna (> 45 kg) became globally extinct, including several proboscidean (Barnosky et al. 2004; A. Cooper et al. 2015). Two highly controversial theories implicate environmental pressures caused by global climate change and

dispersal of prehistoric humans as the main drivers of several megafaunal species extinction (Barnosky et al. 2004; Elias and Schreve 2013).

Climate played a predominant role in the extinctions, but severity was highly variable among geographic regions, implying a major species loss in regions that experienced the most dramatic changes, with the most affected in North America, South America and Australasia, lesser extent in northern Eurasia, and southern Asia and sub-Saharan Africa were less affected (Nogués-Bravo et al. 2010; A. Cooper et al. 2015; Anthony John Stuart 2015). In North America, climate change reduced plant and animal diversity during the Younger Dryas cooling event (12,900 to 11,700 BP), in the transition to Holocene, plants recovered their diversity, while megafauna were decimated by extinction (Seersholm et al. 2020). In northern Europe, Siberia, and Alaska extinctions occurred in two climatic change pulses. The first coincided to increasingly glacial conditions between 45 and 21 kyr BP, including the Last Glacial Maximum (LGM) event (29-19 kyr BP), affecting warm-adapted species. During the second pulse of extinctions, between 21 and 3 kyr BP when temperatures rapidly increase in the Bølling–Allerød warming event (14.7 kyr BP), retreating glaciates and affecting cold-adapted species (Nogués-Bravo et al. 2010).

1.1.2. Prehistoric anthropogenic impact

Extinction of several species is often attributed to prehistorical humans through overkill, rapid overkill (blitzkrieg), fire, habitat fragmentation, and the introduction of exotic species and diseases (sitzkrieg) (Barnosky et al. 2004; Anthony John Stuart 2015). The arrival of humans in North America was simultaneous with climate change and archaeological evidence shows that Clovis people hunted large mammals and overkill was an additional factor in the extinction of certain species (Anthony John Stuart 2015; Seersholm et al. 2020). In certain areas of the Arctic region humans coexisted for a long period with mammoths without severely affecting their distribution (Wang et al. 2021). Human hunting appears to have been a minor factor in driving Pleistocene extinctions occurred in South America and Australia (Barnosky et al. 2004; Anthony John Stuart 2015; Barnosky and Lindsey 2010) and even though Neanderthals

developed advanced hunting techniques to select specific prey animals (Marín et al. 2017) there is no evidence of causing considerable damage on the megafauna.

An hyperdisease hypothesis proposes that with the arrived of aboriginal humans or their dogs, arrived potential disease vectors as rats or fleas with their respective parasites, and in this way, may have been possible to introduce one or more highly virulent diseases into native animals populations, being one of the possible drivers of extirpation or extinction of the Pleistocene megafauna (MacPhee, R.D.E. and Marx, P.A. 1997; MacPhee and Greenwood 2013; Kathleen Lyons et al. 2004). It has been proposed that the introduction of one Herpesviridae virus from humans to Neanderthals may have contributed to the extinction of Neanderthals (Wolff and Greenwood 2010). The case of Christmas Island rats for now is the only evidence that supports hyperdisease hypothesis, where fleas from invasive black rats may have infected the endemic Christmas Island rats with a trypanosome parasite, which could have caused their collapse (Wyatt et al. 2008).

Geographically and temporal patterns of extinction suggest a complex interplay where climate change and human effects appears to be the reasons for the extinction; however, megafaunal population dynamics and evolutionary scenarios are species-specific, making it difficult to have a complete picture causes of extinction (Lorenzen et al. 2011).

1.1.3. Anthropocene

Increasing human pressure is having profound impacts on natural environments and distributed species. Alarmingly, extinction rates have accelerated during the last century affecting severely large vertebrates, being one of the main causes the anthropogenic pressure (Ceballos et al. 2015; Ripple et al. 2019). Megafaunal decline can alter drastically ecological services, modifying ecosystems and producing a detriment in species interactions (Ripple et al. 2015). Endangered populations face a loss of genetic diversity associated with inbreeding depression that might affect the reproduction, reduce the fitness, and accumulate deleterious mutations causing a genomic erosion that may result in the subsequent extinction (Díez-del-Molino et al. 2018).

1.2. PROBOSCIDEAN

1.2.1. Proboscidean evolution and extinction

Proboscidean order including living and extinct elephants (Elephantidae) and their relatives. Its diversification exploded during the Miocene, and they spread outside Afro-Arabia reaching Eurasia and the Americas. Before the Quaternary collapse proboscidean reached a maximum of 33 species. The decline came in a first wave of extinctions approximately 7 Ma and intensified around 3 Main Eurasia and 2.4 in Africa. Climate change and over-hunting triggered the extirpation of some species in the Late Pleistocene (Cantalapiedra et al. 2021). Elephantidae extinction occurred at different times involving several species, the European straight tusked elephants (genus Palaeoloxodon) disappeared among ∼50,000 to 35,000 years (Anthony J. Stuart 2005), the North American Columbian mammoth (*Mammuthus columbi*) ∼11,000 years at the end of the last ice age (Enk et al. 2011), and the woolly mammoth (*Mammuthus primigenius*) is extirpated from the mainland ∼11,000 years, surviving isolated on small islands until ∼4,000 years into the Holocene (Haile et al. 2009; Nyström et al. 2010; Palkopoulou et al. 2015). Only three species of elephantids survive, the savanna elephant (*Loxodonta africana*) and the forest elephant (*Loxodonta cyclotis*) that are restricted to Africa, and one is endemic to Asia (*Elephas maximus*) (Roca et al. 2015). Elephants have suffered a remarkable decline due anthropogenic impacts, as fragmenting and reducing population sizes, disrupting social linkages, and decreasing genetic diversity (Goswami, Vasudev, and Oli 2014; Lee and Graham 2006).

1.2.2. Consequences of diversity loss

As a consequence, directly or indirectly, of human activities, ecosystems are experiencing a massive loss of diversity. For which, small and isolated populations have a high risk of extinction (Ceballos, Ehrlich, and Raven 2020) and this modern extinction crisis is prompting scientific efforts to understand the genetic and ecological interactions occurring across these small populations. One of the main threats of declining populations is to maintain high genetic diversity, otherwise they may

susceptible to extinction through mutational meltdown. Small populations with low connectivity have more chance to increase the inbreeding, reduce the genome-wide heterozygosity, accumulate and fix deleterious mutations that may result in a dramatic drop down of population size and reducing the adaptive fitness (Charmouh et al. 2022; Díez-del-Molino et al. 2018). This scenario coincides with the woolly mammoths experienced just prior its extinction. An isolated population on Wrangel Island that suffered a reduction of genetic diversity with accumulation and fixation of deleterious alleles, resulting in genomic meltdown and probably leading with functional consequences (Fry et al. 2020; Rogers and Slatkin 2017).

1.3. MEASURING THE IMMUNODIVERSITY

1.3.1. Innate and adaptive immune systems

The immune system of vertebrates has two powerful mechanisms, the innate and the adaptive immune systems, which act coordinated to recognize, control, and eliminate infectious agents and malignant cells (Buchmann 2014; Suckale, Sim, and Dodds 2005). Innate immunity is considered a nonspecific response because pattern-recognition is based on the recognition of a limited set of molecular structures that are evolutionary conserved and are maintained invariant among microorganisms, whereas the adaptive immune system provides high specificity despite having to respond to a wide number of targets (M. D. Cooper and Alder 2006; Ward and Rosenthal 2014). The adaptive immune system is composed of specialized T and B lymphocytes cells that express antigen-specific receptors on their cell surface, allowing the recognition of a unique microorganism by specific antigens, producing long-term immunological memory (Chaplin 2010).

1.3.2. Toll-Like-Receptors

Toll-like receptors (TLRs) are the first line of defense of innate immunity, sensing infectious microorganisms by recognizing a wide variety of pathogen-associated

molecular patterns (PAMPs), including glycolipids such as bacterial lipopolysaccharides (LPS), bacterial lipoproteins and lipoteichoic acids, flagellin, the unmethylated CpG DNA of bacteria and viruses, double-stranded RNA, and single-stranded viral RNA. The interactions of the TLRs trigger intracellular signaling pathways inducing the release of inflammatory cytokines (Iwasaki and Medzhitov 2004; Tang et al. 2012). All TLRs have a common structural framework including an extracelular LRR ectodomain, a single transmembrane helix, and an intracellular Toll/interleukin-1 receptor (TIR) domain (Botos, Segal, and Davies 2011). The ectodomain varies from 17 to 26 consecutive leucine-rich repeat motifs (LRR) that adopt a horseshoe-shaped solenoid structure with a wide versatility for the recognition of a variety of pathogens. The cytoplasmatic TIR domains are conserved across all TLRs and are involved in dimerization and initiate the signaling cascade and the transmembrane helix links the extracellular and intracellular portions and determines the subcellular localization of TLRs (Botos, Segal, and Davies 2011; Bella et al. 2008). TLRs can be located into endosomal compartments and there are effective against viruses, or into cell surface in which they detect bacteria, parasites and fungi (Kumar, Kawai, and Akira 2009).

1.3.3. The major histocompatibility complex

The major histocompatibility complex (MHC) is the most large multigenic and polymorphic region found in most vertebrates. MHC consists of two clusters of genes, MHC class I and MHC class II (Trowsdale 2011). Genes in the MHC locus code for molecules that present protein-derived peptides (antigens) from intracellular (e.g., viruses, cancer infected cells) and extracellular (e.g., bacteria and parasites) to T-cells. Through antigen presentation, MHC molecules are able to control sel/nonself recognition, autoimmunity, and to activate an immune response against the pathogens (Neefjes et al. 2011). A prominent characteristic of the MHC antigen-binding-recognition region (ABR), the pocket where antigens are bound, is the highly degree of polymorphism that allows it to recognize a high diversity of pathogens (Radwan et al. 2020). ABR in MHC I and II share a similar fold which is composed of two domains, with a single alpha-chain in the class I and alpha-chain and beta-chain in the class II, producing a wide plasticity to accommodate a broader range of antigens (Falk et al.

1991; Wieczorek et al. 2017; Manczinger et al. 2019). Several mechanisms of balancing selection have been proposed to explain the exceptional MHC polymorphism, the heterozygote advantage (HA) (overdominance), where the presence of two alleles per MHC gene increases the chance of pathogen recognition, similarly, negative frequency-dependent selection (NFDS) (rare allele advantage), posits that having highly diverged MHC alleles will increase the repertoire to bind more different antigens and in consequence triggering a specific immune response, and fluctuating selection (FS) (spatiotemporal selection), assumes an arms race between pathogen and host, with alleles are transient and replacing each other due to pathogen pressure (Radwan et al. 2020; Spurgin and Richardson 2010). Balancing selection may promoting trans-species polymorphisms (TSP), alleles that are more similar or identical in multiple species (Klein, Sato, and Nikolaidis 2007).

1.3.4. Immunodiversity loss

Demographic bottlenecks erode genetic diversity and may decreased fitness and adaptive potential. Immunodiversity is often substantially reduced during severe population bottlenecks and the depletion of variation at MHC loci has been associated with reduction to mount a protective immune response and potentially increase susceptibility to disease, being a major extinction risk factor for endangered species (Ejsmond and Radwan 2011). The evolution of TLR diversity is generally less well characterized, however, there is evidence showing species-specific pathogen recognition, which is important from the ecological point of view of the evolution of wild-type infectious diseases (Werling et al. 2009). TLRs with the MHC are associated with resistance to infectious disease and a reduction in the diversity of both could imply increased of susceptibility to pathogens, which in extreme, could affect species survival.

Neutral markers are useful for genetic profiling and to infer populations demography due to high rates of evolution of intra-species polymorphisms (Ishida et al. 2011). However, they are not under natural selection, they are inappropriate to investigate functional genetic variation behind episodes of pathogen exposure. On the other hand, diversity in the innate and adaptive immune systems can be of great use in determining

to a large extent the long-term survival and viability of populations (McCallum 2012; Smith, Sax, and Lafferty 2006). A combination of Neutral markers with immune genes could provide a better measure of diversity and population dynamics in wildlife species.

1.3.5. Technical approaches to measure the immunodiversity

Paleogenomics, the study of ancient DNA (aDNA), have revolutionized our understanding of many questions in anthropology, archaeology, ecology, and evolutionary biology reconstructing major prehistoric and historic events in a shorter timescale (Orlando et al. 2021). Studies in aDNA have improve from analysis of short DNA fragments (229 bp mitochondrial DNA in quagga) (Higuchi et al. 1984) till by genome-scale studies, reconstructing whole human population ancestry (Orlando et al. 2021; Hellenthal et al. 2014) and metagenomic and enviromental DNA analysis in complete ecoregions (Orlando et al. 2021; Fellows Yates et al. 2021; Kjær et al. 2022). Using a combination of aDNA in-solution target-enrichment techniques with next-generation sequencing (NGS) prove an invaluable tool for measuring diversity of immunological important nuclear DNA loci in the woolly mammoth and its distribution over time, specifically during the Late Pleistocene through the Holocene.

Genotyping genes with extensive polymorphism may result in a difficult task, especially in non-model vertebrates. Genome reference sequences are available in different databases; however, the genes annotation includes errors, notably, in immune genes that are commonly present at multiple copies, locus may be very divergent, or where domains are well conserved throughout many genes in other families, making methodological design of multilocus genotype challenging (Babik 2010; Sommer, Courtiol, and Mazzoni 2013). Genotyping has move from cloning/Sanger sequencing to NGS technologies reducing time and cost balance for larger sample size; however, some technical problems still be frequent like allelic dropout due to unbalanced amplification, alleles artefacts, and SNP calling (Babik 2010; Sommer, Courtiol, and Mazzoni 2013). Long-read sequencing platform (Chang et al. 2014) has the potential to produce longer reads that can facilitated massive genotyping in complex genomic organization regions, like immune genes, in poorly studied organisms such as wildlife elephant populations.

1.4. STUDY AIMS

The general objective of this study is to investigate the evolutionary dynamics of the immunity to extinct woolly mammoths and extant elephants. During most part the Pleistocene, Woolly mammoth populations dominated North America, Siberia, and Beringia, but abruptly collapsed on the mainland around 10,000 years ago at the end of the Late Pleistocene which may have been driven by climate warming and human predation. An isolated population survived on the Wrangel Islands until its final extinction about 4,000 years ago. Paleogenomic studies from one specimen on this Island have highlighted how the isolation reduced the genetic diversity and led to increase in the frequency of detrimental mutations. On the other hand, during the transition from the Pleistocene to the Holocene when the majority of megafaunal became extinct the three elephant species did not suffer population declines, however, due anthropogenic activities (i.e., poaching, ivory trading, territory fragmentation), elephants have suffered a remarkable loss of their diversity. This is potentially serious, especially in endangered species, because the immune response may be compromised, increasing susceptibility to infection.

To address these evolutionary aspects, two specific aims were established: the first aim was to test the effect of population collapse on genomes of Late Pleistocene mammoths. In **Chapter 2,** using in-solution target hybridization capture approach followed by next-generation sequencing, a large set of nuclear genes, covering primarily the immune system diversity were analyzed in mammoths over the Late Pleistocene and across different geographic regions. Our results suggest that during this period mammoths had a relatively high load of deleterious alleles.

The second aim was to measure the immunogenetic diversity of TLRs and MHC in the three living elephant populations. In **Chapter 3**, a multilocus genotyping method coupled with PacBio high throughput sequencing platform was implemented, a total of 263 wild and captive elephants were analyzed. Evidence of balancing selection acting on MHC diversity was found for populations of three elephant species, however, immunodiversity may going in detrimental as consequence of inbreeding and heterozygosity deficiency, which can increase susceptibility to infections.

REFERENCES

Babik, W. 2010. “Methods for MHC Genotyping in Non-model Vertebrates.” Molecular Ecology Resources 10 (2): 237–51. https://doi.org/10.1111/j.1755-0998.2009.02788.x.

Barnosky, Anthony D. 2008. “Megafauna Biomass Tradeoff as a Driver of Quaternary and Future Extinctions.” Proceedings of the National Academy of Sciences 105 (supplement_1): 11543–48. https://doi.org/10.1073/pnas.0801918105.

Barnosky, Anthony D., Paul L. Koch, Robert S. Feranec, Scott L. Wing, and Alan B. Shabel. 2004. “Assessing the Causes of Late Pleistocene Extinctions on the Continents.” Science 306 (5693): 70–75. https://doi.org/10.1126/science.1101476.

Barnosky, Anthony D., and Emily L. Lindsey. 2010. “Timing of Quaternary Megafaunal Extinction in South America in Relation to Human Arrival and Climate Change.” Quaternary International 217 (1–2): 10–29. https://doi.org/10.1016/j.quaint.2009.11.017.

Bella, J., K. L. Hindle, P. A. McEwan, and S. C. Lovell. 2008. “The Leucine-Rich Repeat Structure.” Cellular and Molecular Life Sciences 65 (15): 2307–33. https://doi.org/10.1007/s00018-008-8019-0.

Botos, Istvan, David M. Segal, and David R. Davies. 2011. “The Structural Biology of Toll-like Receptors.” Structure 19 (4): 447–59. https://doi.org/10.1016/j.str.2011.02.004.

Buchmann, Kurt. 2014. “Evolution of Innate Immunity: Clues from Invertebrates via Fish to Mammals.” Frontiers in Immunology 5 (September). https://doi.org/10.3389/fimmu.2014.00459.

Cantalapiedra, Juan L., Óscar Sanisidro, Hanwen Zhang, María T. Alberdi, José L. Prado, Fernando Blanco, and Juha Saarinen. 2021. “The Rise and Fall of Proboscidean Ecological Diversity.” Nature Ecology & Evolution 5 (9): 1266–72. https://doi.org/10.1038/s41559-021-01498-w.

Ceballos, Gerardo, Paul R. Ehrlich, Anthony D. Barnosky, Andrés García, Robert M. Pringle, and Todd M. Palmer. 2015. “Accelerated Modern Human–Induced Species Losses: Entering the Sixth Mass Extinction.” Science Advances 1 (5): e1400253. https://doi.org/10.1126/sciadv.1400253.

Ceballos, Gerardo, Paul R. Ehrlich, and Peter H. Raven. 2020. “Vertebrates on the Brink as Indicators of Biological Annihilation and the Sixth Mass Extinction.” Proceedings of the National Academy of Sciences 117 (24): 13596–602. https://doi.org/10.1073/pnas.1922686117.

Chang, Chia-Jung, Pei-Lung Chen, Wei-Shiung Yang, and Kun-Mao Chao. 2014. “A Fault-Tolerant Method for HLA Typing with PacBio Data.” BMC Bioinformatics 15 (1): 296. https://doi.org/10.1186/1471-2105-15-296.

Chaplin, David D. 2010. “Overview of the Immune Response.” Journal of Allergy and Clinical Immunology 125 (2): S3–23. https://doi.org/10.1016/j.jaci.2009.12.980.

Charmouh, Anders P., Jane M. Reid, Trine Bilde, and Greta Bocedi. 2022. “Eco-evolutionary Extinction and Recolonization Dynamics Reduce Genetic Load and Increase Time to Extinction in Highly Inbred Populations.” Evolution, September, evo.14620. https://doi.org/10.1111/evo.14620.

Cooper, Alan, Chris Turney, Konrad A. Hughen, Barry W. Brook, H. Gregory McDonald, and Corey J. A. Bradshaw. 2015. “Abrupt Warming Events Drove Late Pleistocene Holarctic Megafaunal Turnover.” Science 349 (6248): 602–6. https://doi.org/10.1126/science.aac4315.

Cooper, Max D., and Matthew N. Alder. 2006. “The Evolution of Adaptive Immune Systems.” Cell 124 (4): 815–22. https://doi.org/10.1016/j.cell.2006.02.001.

Davis, Matt, Søren Faurby, and Jens-Christian Svenning. 2018. “Mammal Diversity Will Take Millions of Years to Recover from the Current Biodiversity Crisis.” Proceedings of the National Academy of Sciences 115 (44): 11262–67. https://doi.org/10.1073/pnas.1804906115.

Díez-del-Molino, David, Fatima Sánchez-Barreiro, Ian Barnes, M. Thomas P. Gilbert, and Love Dalén. 2018. “Quantifying Temporal Genomic Erosion in Endangered Species.” Trends in Ecology & Evolution 33 (3): 176–85. https://doi.org/10.1016/j.tree.2017.12.002.

Ejsmond, Maciej Jan, and Jacek Radwan. 2011. “MHC Diversity in Bottlenecked Populations: A Simulation Model.” Conservation Genetics 12 (1): 129–37. https://doi.org/10.1007/s10592-009-9998-6.

Elias, S.A., and D.C. Schreve. 2013. “Vertebrate Records | Late Pleistocene Megafaunal Extinctions.” In Encyclopedia of Quaternary Science, 700–712. Elsevier. https://doi.org/10.1016/B978-0-444-53643-3.00245-4.

Enk, Jacob, Alison Devault, Regis Debruyne, Christine E King, Todd Treangen, Dennis O'Rourke, Steven L Salzberg, Daniel Fisher, Ross MacPhee, and Hendrik Poinar. 2011. "Complete Columbian Mammoth Mitogenome Suggests Interbreeding with Woolly Mammoths." Genome Biology 12 (5): R51. https://doi.org/10.1186/gb-2011-12-5-r51.

Falk, Kirsten, Olaf Rötzschke, Stefan Stevanović, Günther Jung, and Hans-Georg Rammensee. 1991. "Allele-Specific Motifs Revealed by Sequencing of Self-Peptides Eluted from MHC Molecules." Nature 351 (6324): 290–96. https://doi.org/10.1038/351290a0.

Fellows Yates, James A., Aida Andrades Valtueña, Åshild J. Vågene, Becky Cribdon, Irina M. Velsko, Maxime Borry, Miriam J. Bravo-Lopez, et al. 2021. "Community-Curated and Standardised Metadata of Published Ancient Metagenomic Samples with AncientMetagenomeDir." Scientific Data 8 (1): 31. https://doi.org/10.1038/s41597-021-00816-y.

Fry, Erin, Sun K Kim, Sravanthi Chigurapti, Katelyn M Mika, Aakrosh Ratan, Alexander Dammermann, Brian J Mitchell, Webb Miller, and Vincent J Lynch. 2020. "Functional Architecture of Deleterious Genetic Variants in the Genome of a Wrangel Island Mammoth." Edited by Charles Baer. Genome Biology and Evolution 12 (3): 48–58. https://doi.org/10.1093/gbe/evz279.

Goswami, Varun R., Divya Vasudev, and Madan K. Oli. 2014. "The Importance of Conflict-Induced Mortality for Conservation Planning in Areas of Human–Elephant Co-Occurrence." Biological Conservation 176 (August): 191–98. https://doi.org/10.1016/j.biocon.2014.05.026.

Haile, James, Duane G. Froese, Ross D. E. MacPhee, Richard G. Roberts, Lee J. Arnold, Alberto V. Reyes, Morten Rasmussen, et al. 2009. "Ancient DNA Reveals Late Survival of Mammoth and Horse in Interior Alaska." Proceedings of the National Academy of Sciences 106 (52): 22352–57. https://doi.org/10.1073/pnas.0912510106.

Hellenthal, G., G. B. J. Busby, G. Band, J. F. Wilson, C. Capelli, D. Falush, and S. Myers. 2014. "A Genetic Atlas of Human Admixture History." Science 343 (6172): 747–51. https://doi.org/10.1126/science.1243518.

Higuchi, Russell, Barbara Bowman, Mary Freiberger, Oliver A. Ryder, and Allan C. Wilson. 1984. "DNA Sequences from the Quagga, an Extinct Member of the Horse Family." Nature 312 (5991): 282–84. https://doi.org/10.1038/312282a0.

Ishida, Yasuko, Taras K. Oleksyk, Nicholas J. Georgiadis, Victor A. David, Kai Zhao, Robert M. Stephens, Sergios-Orestis Kolokotronis, and Alfred L. Roca. 2011. “Reconciling Apparent Conflicts between Mitochondrial and Nuclear Phylogenies in African Elephants.” Edited by William J. Murphy. PLoS ONE 6 (6): e20642. https://doi.org/10.1371/journal.pone.0020642.

Iwasaki, Akiko, and Ruslan Medzhitov. 2004. “Toll-like Receptor Control of the Adaptive Immune Responses.” Nature Immunology 5 (10): 987–95. https://doi.org/10.1038/ni1112.

Kathleen Lyons, S., Felisa A. Smith, Peter J. Wagner, Ethan P. White, and James H. Brown. 2004. “Was a ‘Hyperdisease’ Responsible for the Late Pleistocene Megafaunal Extinction? The Disease Hypothesis and West Nile Virus.” Ecology Letters 7 (9): 859–68. https://doi.org/10.1111/j.1461-0248.2004.00643.x.

Kender, Sev, Erin L. McClymont, Aurora C. Elmore, Dario Emanuele, Melanie J. Leng, and Henry Elderfield. 2016. “Mid Pleistocene Foraminiferal Mass Extinction Coupled with Phytoplankton Evolution.” Nature Communications 7 (1): 11970. https://doi.org/10.1038/ncomms11970.

Kjær, Kurt H., Mikkel Winther Pedersen, Bianca De Sanctis, Binia De Cahsan, Thorfinn S. Korneliussen, Christian S. Michelsen, Karina K. Sand, et al. 2022. “A 2-Million-Year-Old Ecosystem in Greenland Uncovered by Environmental DNA.” Nature 612 (7939): 283–91. https://doi.org/10.1038/s41586-022-05453-y.

Klein, Jan, Akie Sato, and Nikolas Nikolaidis. 2007. “MHC, TSP, and the Origin of Species: From Immunogenetics to Evolutionary Genetics.” Annual Review of Genetics 41 (1): 281–304. https://doi.org/10.1146/annurev.genet.41.110306.130137.

Kumar, Himanshu, Taro Kawai, and Shizuo Akira. 2009. “Toll-like Receptors and Innate Immunity.” Biochemical and Biophysical Research Communications 388 (4): 621–25. https://doi.org/10.1016/j.bbrc.2009.08.062.

Lee, P C, and M D Graham. 2006. “African Elephants and Human-Elephant Interactions: Implications for Conservation.” Int Zoo Yearb 40: 9–19. https://doi.org/10.1111/j.1748-1090.2006.00009.x.

Lorenzen, Eline D., David Nogués-Bravo, Ludovic Orlando, Jaco Weinstock, Jonas Binladen, Katharine A. Marske, Andrew Ugan, et al. 2011. “Species-Specific Responses of Late Quaternary Megafauna to Climate and Humans.” Nature 479 (7373): 359–64. https://doi.org/10.1038/nature10574.

MacPhee, R.D.E. and Marx, P.A. 1997. “The 40,000-Year Plaque: Humans, Hyperdisease and First Contact Extinctions.” In Natural Change and Human Impact in Madagascar (Goodman, S. and Patterson, B., Eds), Pp. 169–217, Smithsonia, 169–217.

MacPhee, Ross D. E., and Alex D. Greenwood. 2013. “Infectious Disease, Endangerment, and Extinction.” International Journal of Evolutionary Biology 2013 (January): 1–9. https://doi.org/10.1155/2013/571939.

Manczinger, Máté, Gábor Boross, Lajos Kemény, Viktor Müller, Tobias L. Lenz, Balázs Papp, and Csaba Pál. 2019. “Pathogen Diversity Drives the Evolution of Generalist MHC-II Alleles in Human Populations.” Edited by Andrew Fraser Read. PLOS Biology 17 (1): e3000131. https://doi.org/10.1371/journal.pbio.3000131.

Marín, Juan, Palmira Saladié, Antonio Rodríguez-Hidalgo, and Eudald Carbonell. 2017. “Neanderthal Hunting Strategies Inferred from Mortality Profiles within the Abric Romaní Sequence.” Edited by Michael D. Petraglia. PLOS ONE 12 (11): e0186970. https://doi.org/10.1371/journal.pone.0186970.

McCallum, Hamish. 2012. “Disease and the Dynamics of Extinction.” Philosophical Transactions of the Royal Society B: Biological Sciences 367 (1604): 2828–39. https://doi.org/10.1098/rstb.2012.0224.

Neefjes, Jacques, Marlieke L. M. Jongsma, Petra Paul, and Oddmund Bakke. 2011. “Towards a Systems Understanding of MHC Class I and MHC Class II Antigen Presentation.” Nature Reviews Immunology 11 (12): 823–36. https://doi.org/10.1038/nri3084.

Nogués-Bravo, David, Ralf Ohlemüller, Persaram Batra, and Miguel B. Araújo. 2010. “Climate Predictors of Late Quaternary Extinctions.” Evolution, April, no-no. https://doi.org/10.1111/j.1558-5646.2010.01009.x.

Nyström, Veronica, Love Dalén, Sergey Vartanyan, Kerstin Lidén, Nils Ryman, and Anders Angerbjörn. 2010. “Temporal Genetic Change in the Last Remaining Population of Woolly Mammoth.” Proceedings of the Royal Society B: Biological Sciences 277 (1692): 2331–37. https://doi.org/10.1098/rspb.2010.0301.

Orlando, Ludovic, Robin Allaby, Pontus Skoglund, Clio Der Sarkissian, Philipp W. Stockhammer, María C. Ávila-Arcos, Qiaomei Fu, et al. 2021. “Ancient DNA Analysis.” Nature Reviews Methods Primers 1 (1): 14. https://doi.org/10.1038/s43586-020-00011-0.

Palkopoulou, Eleftheria, Swapan Mallick, Pontus Skoglund, Jacob Enk, Nadin Rohland, Heng Li, Ayça Omrak, et al. 2015. “Complete Genomes Reveal Signatures of Demographic and Genetic Declines in the Woolly Mammoth.” Current Biology 25 (10): 1395–1400. https://doi.org/10.1016/j.cub.2015.04.007.

Pillans, B., and P. Gibbard. 2012. “The Quaternary Period.” In The Geologic Time Scale, 979–1010. Elsevier. https://doi.org/10.1016/B978-0-444-59425-9.00030-5.

Radwan, Jacek, Wiesław Babik, Jim Kaufman, Tobias L. Lenz, and Jamie Winternitz. 2020. “Advances in the Evolutionary Understanding of MHC Polymorphism.” Trends in Genetics 36 (4): 298–311. https://doi.org/10.1016/j.tig.2020.01.008.

Ripple, William J., Thomas M. Newsome, Christopher Wolf, Rodolfo Dirzo, Kristoffer T. Everatt, Mauro Galetti, Matt W. Hayward, et al. 2015. “Collapse of the World’s Largest Herbivores.” Science Advances 1 (4): e1400103. https://doi.org/10.1126/sciadv.1400103.

Ripple, William J., Christopher Wolf, Thomas M. Newsome, Matthew G. Betts, Gerardo Ceballos, Franck Courchamp, Matt W. Hayward, Blaire Valkenburgh, Arian D. Wallach, and Boris Worm. 2019. “Are We Eating the World’s Megafauna to Extinction?” Conservation Letters 12 (3). https://doi.org/10.1111/conl.12627.

Roca, Alfred L., Yasuko Ishida, Adam L. Brandt, Neal R. Benjamin, Kai Zhao, and Nicholas J. Georgiadis. 2015. “Elephant Natural History: A Genomic Perspective.” Annual Review of Animal Biosciences 3 (1): 139–67. https://doi.org/10.1146/annurev-animal-022114-110838.

Rogers, Rebekah L., and Montgomery Slatkin. 2017. “Excess of Genomic Defects in a Woolly Mammoth on Wrangel Island.” Edited by Gregory S. Barsh. PLOS Genetics 13 (3): e1006601. https://doi.org/10.1371/journal.pgen.1006601.

Seersholm, Frederik V., Daniel J. Werndly, Alicia Grealy, Taryn Johnson, Erin M. Keenan Early, Ernest L. Lundelius, Barbara Winsborough, et al. 2020. “Rapid Range Shifts and Megafaunal Extinctions Associated with Late Pleistocene Climate Change.” Nature Communications 11 (1): 2770. https://doi.org/10.1038/s41467-020-16502-3.

Smith, Katherine F., Dov F. Sax, and Kevin D. Lafferty. 2006. “Evidence for the Role of Infectious Disease in Species Extinction and Endangerment.” Conservation Biology 20 (5): 1349–57. https://doi.org/10.1111/j.1523-1739.2006.00524.x.

Sommer, Simone, Alexandre Courtiol, and Camila J Mazzoni. 2013. “MHC Genotyping of Non-Model Organisms Using next-Generation Sequencing: A New Methodology to Deal with Artefacts and Allelic Dropout.” BMC Genomics 14 (1): 542. https://doi.org/10.1186/1471-2164-14-542.

Spurgin, Lewis G., and David S. Richardson. 2010. “How Pathogens Drive Genetic Diversity: MHC, Mechanisms and Misunderstandings.” Proceedings of the Royal Society B: Biological Sciences 277 (1684): 979–88. https://doi.org/10.1098/rspb.2009.2084.

Stuart, Anthony J. 2005. “The Extinction of Woolly Mammoth (Mammuthus Primigenius) and Straight-Tusked Elephant (Palaeoloxodon Antiquus) in Europe.” Quaternary International 126–128 (January): 171–77. https://doi.org/10.1016/j.quaint.2004.04.021.

Stuart, Anthony John. 2015. “Late Quaternary Megafaunal Extinctions on the Continents: A Short Review.” Geological Journal 50 (3): 338–63. https://doi.org/10.1002/gj.2633.

Suckale, Jakob, Robert B. Sim, and Alister W. Dodds. 2005. “Evolution of Innate Immune Systems.” Biochemistry and Molecular Biology Education 33 (3): 177–83. https://doi.org/10.1002/bmb.2005.494033032466.

Tang, Daolin, Rui Kang, Carolyn B. Coyne, Herbert J. Zeh, and Michael T. Lotze. 2012. “PAMPs and DAMPs: Signal 0s That Spur Autophagy and Immunity.” Immunological Reviews 249 (1): 158–75. https://doi.org/10.1111/j.1600-065X.2012.01146.x.

Thomas, Chris D, Alison Cameron, Rhys E Green, Michel Bakkenes, Linda J Beaumont, Yvonne C Collingham, Barend F N Erasmus, et al. 2004. “Extinction Risk from Climate Change” 427: 4.

Trowsdale, John. 2011. “The MHC, Disease and Selection.” Immunology Letters 137 (1–2): 1–8. https://doi.org/10.1016/j.imlet.2011.01.002.

Urban, Mark C. 2015. “Accelerating Extinction Risk from Climate Change.” Science 348 (6234): 571–73. https://doi.org/10.1126/science.aaa4984.

Wang, Yucheng, Mikkel Winther Pedersen, Inger Greve Alsos, Bianca De Sanctis, Fernando Racimo, Ana Prohaska, Eric Coissac, et al. 2021. “Late Quaternary Dynamics of Arctic Biota from Ancient Environmental Genomics.” Nature 600 (7887): 86–92. https://doi.org/10.1038/s41586-021-04016-x.

Ward, Amanda E., and Benjamin M. Rosenthal. 2014. “Evolutionary Responses of Innate Immunity to Adaptive Immunity.” Infection, Genetics and Evolution 21 (January): 492–96. https://doi.org/10.1016/j.meegid.2013.12.021.

Werling, Dirk, Oliver C. Jann, Victoria Offord, Elizabeth J. Glass, and Tracey J. Coffey. 2009. “Variation Matters: TLR Structure and Species-Specific Pathogen Recognition.” Trends in Immunology 30 (3): 124–30. https://doi.org/10.1016/j.it.2008.12.001.

Wieczorek, Marek, Esam T. Abualrous, Jana Sticht, Miguel Álvaro-Benito, Sebastian Stolzenberg, Frank Noé, and Christian Freund. 2017. “Major Histocompatibility Complex (MHC) Class I and MHC Class II Proteins: Conformational Plasticity in Antigen Presentation.” Frontiers in Immunology 8 (March). https://doi.org/10.3389/fimmu.2017.00292.

Wolff, Horst, and Alex D. Greenwood. 2010. “Did Viral Disease of Humans Wipe out the Neandertals?” Medical Hypotheses 75 (1): 99–105. https://doi.org/10.1016/j.mehy.2010.01.048.

Wyatt, Kelly B., Paula F. Campos, M. Thomas P. Gilbert, Sergios-Orestis Kolokotronis, Wayne H. Hynes, Rob DeSalle, Peter Daszak, Ross D. E. MacPhee, and Alex D. Greenwood. 2008. “Historical Mammal Extinction on Christmas Island (Indian Ocean) Correlates with Introduced Infectious Disease.” Edited by Niyaz Ahmed. PLoS ONE 3 (11): e3602. https://doi.org/10.1371/journal.pone.0003602.

Zhou, Xinying, Jilong Yang, Shiqi Wang, Guoqiao Xiao, Keliang Zhao, Yan Zheng, Hui Shen, and Xiaoqiang Li. 2018. “Vegetation Change and Evolutionary Response of Large Mammal Fauna during the Mid-Pleistocene Transition in Temperate Northern East Asia.” Palaeogeography, Palaeoclimatology, Palaeoecology 505 (September): 287–94. https://doi.org/10.1016/j.palaeo.2018.06.007.

2. CHAPTER II: WIDESPREAD GENOMIC DEGRADATION IN MAMMOTHS DURING THE LATE PLEISTOCENE

John A Galindo[1,2], Jazmín Ramos-Madrigal[3], Gayle McEwen[1], Jake Enk[4], Maria Lembring[5], Nicolas Fasel[6], M. Thomas P. Gilbert[3,7], Hendrik Poinar[5], Alex D. Greenwood[1,8, §]

[1] Department of Wildlife Diseases, Leibniz Institute for Zoo and Wildlife Research, Berlin, Germany

[2] Department of Biology, Chemistry and Pharmacy, Freie Universität Berlin, Berlin, Germany

[3] Center for Evolutionary Hologenomics, GLOBE Institute, University of Copenhagen. Copenhagen, Denmark

[4] Daicel Arbor Biosciences, Ann Arbor, Michigan, USA

[5] Department of Anthropology, Biology and Biochemestry McMaster University, Hamilton, Ontario, Canada

[6] Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland

[7] University Museum, Norwegian University of Science and Technology, Trondheim, Norway

[8] Department of Veterinary Medicine, Freie Universität Berlin, Berlin, Germany

§ correspondence: greenwood@izw-berlin.de

2.1. ABSTRACT

Mammoths (*Mammuthus sp.*) had a holoarctic distribution until the Pleistocene-Holocene transition approximately 10,000 years ago resulted in their population declining to isolated island populations and relict mainland populations. Among Island mammoth populations, genomic degradation has been observed which has been suggested to have been a driver of their eventual extinction. However, during the Pleistocene, mammoth populations exhibited a complex demographic history marked by population fluctuations, loss of genetic lineages, population replacement and interbreeding among populations. To test the genomic consequences of mammoth spatiotemporal dynamics, we performed in-solution target hybridization capture and next-generation sequencing on 90 mammoths samples temporally distributed across the Late Pleistocene and collected across the mammoth's range. Two hundred and fifty nine genes representing neutral, antiviral, arctic adaptation associated, cellular process associated, immune and sexual markers were analyzed. Loss of diversity was concentrated in innate immune genes. Potentially deleterious alleles were identified throughout the Pleistocene with little difference in frequency among different time points. Predictions of the functional consequences of observed non-synonymous variants suggest mammoths may have experienced genomic degradation, with dysregulation of the immune system as a major consequence, throughout the Pleistocene.

2.2. INTRODUCTION

Mammoths (*Mammuthus sp.*) dispersed from Africa to Asia approximately 5 million years ago gradually achieving a Holoartic distribution in the Early Pleistocene (1.9 Mya), including much of Europe and North America (~1.5 to 1.3 Mya) (Lister, 2001; Lister et al., 2005). The woolly mammoth (*Mammuthus primigenius*) and the Columbian mammoth (*Mammuthus columbi*) dispersed at different times during the Pleistocene though their range overlapped in North America (Enk et al., 2016). Mammoths populations collapsed on the mainland at the end of the Late Pleistocene around 10,000 years ago (Haile et al., 2009) while relic populations survived on islands such as St. Paul (5,600 yBP) (Graham et al., 2016) and Wrangel islands (4,000 yBP) until their final extinction (Nyström et al., 2010, 2012; Palkopoulou et al., 2015). Recent environmental DNA (eDNA) results from sediments suggests mammoths persisted in the Siberian mainland during the Holocene in northeast Siberia until ~7.3 kya, Taimyr Peninsula ~3.9 kya (Wang et al., 2021), North America ~8.6 kya (Wang et al., 2021), and eastern Beringia (Yukon) ~5.7 kya (Murchie et al., 2020, 2021) likely as relict populations.

Mammoths evolutionary history and demographic changes were inferred, originally, from mitochondrial DNA analysis. Around ~1.0-2.0 million years ago (Ma), three mitochondrial lineages diverged (Barnes et al., 2007; D. Chang et al., 2017; Debruyne et al., 2008; Palkopoulou et al., 2013). Clade I (haplogroups C, D, E, and F) originated in North America, and was confined there until migrated across the Beringian land bridge occurred circa 300 ka during the Middle Pleistocene. It dispersed throughout Eurasia around 66 kya and became the predominant clade during the Late Pleistocene (Barnes et al., 2007; Debruyne et al., 2008; Gilbert et al., 2008; Palkopoulou et al., 2013). Clade II (haplogroup A) originated in Siberia ~810-360 ka with a limited distribution to western Beringia and northern Siberia until its extinction. It coexisted in northeast Siberia with clade I during the Middle Pleistocene ~44 kya (Barnes et al., 2007; Debruyne et al., 2008; Gilbert et al., 2008; Palkopoulou et al., 2013). Clade III (haplogroup B) originated ~1.4-0.7 Ma in Europe and dispersed to North America in different waves (~1.3 Ma) and (~500-240 kya). Its extinction is explained by a competition replacement scenario with mtDNA clade I (D. Chang et al., 2017;

Debruyne et al., 2008; Enk et al., 2016; Fellows Yates et al., 2017; Palkopoulou et al., 2013). A new mitochondrial lineage was recently detected using eDNA (Wang et al., 2021) indicating that mammoth's phylogeographic history was far more complex than previously assumed and some evolutionary gaps remain un-resolved.

Although Columbian mammoths are considered to be a distinct species, the Columbian mammoth mitochondrial lineage is embedded within the woolly mammoth clade I, suggesting introgression occurred among mammoth species (Debruyne et al., 2008; Enk et al., 2011; Palkopoulou et al., 2018). Recent nuclear genomic analyses suggest a complex evolutionary history of mammoth species admixture and introgression across their range. Columbian mammoths appear to be the result of a hybrid speciation event between an ancient mammoth lineage and the woolly mammoth (van der Valk et al., 2021) followed by recent unidirectional gene flow from North American woolly mammoths (Palkopoulou et al., 2018; van der Valk et al., 2021) and woolly mammoths and straight-tusked elephants also interbred (Palkopoulou et al., 2018). These events are not surprising given their remarkable ability to disperse over long-distances, including extensive ranges covered during their lifetimes (Wooller et al., 2021).

The climate varied dramatically during the Pleistocene, from extremely cold and dry conditions until the last glacial maximum (LGM) (26.5-19 Kya) (Clark et al., 2009), followed by a rapid period of warming known as the Bølling–Allerød interstadial (~14.6-12.9 kya) (Rasmussen et al., 2006), and a subsequent return to colder temperatures during the Younger Dryas stadial (~12.9-11.7 kya) (Mangerud, 2021) which directly preceded the Pleistocene-Holocene transition. The warming event of the Bølling–Allerød interstadial drove a requisite shift in Arctic vegetation composition, leading to mammoth habitat loss, population decline, local extinctions and recolonization events (Gilbert et al., 2008; Kuzmin, 2010; MacDonald et al., 2012; Murchie et al., 2021; Nikolskiy, 2011; Nogués-Bravo et al., 2008; Stuart et al., 2004; Wang et al., 2021).

When compared to mainland populations of mammoths and Asian elephants, the genome of Wrangel Island mammoth exhibited an excess of homozygous substitutions and accumulation of detrimental mutations (Palkopoulou et al., 2015). The decreased population size and prolonged geographic isolation on Wrangel Island resulted in an

excess of the deleterious genetic variants, affecting genes associated with behavioral and developmental processes (Fry et al., 2020; Rogers & Slatkin, 2017). Although quantification of genetic factors such as population size, inbreeding depression, loss of genetic diversity, and mutational accumulation can help explain the extinction process (Frankham, 2005; M. Lynch et al., 1995), understanding their role as drivers of extinction is often challenging as species on the brink of extinction exhibit different evolutionary responses (Díez-del-Molino et al., 2018). For example, small-isolated populations of Iberian lynx (*Lynx pardinus*) (Abascal et al., 2016; Casas-Marce et al., 2017; Kleinman-Ruiz et al., 2022), Indian tigers (Panthera tigris tigris) (Khan et al., 2021), Indian lions (*Panthera leo leo*) (de Manuel et al., 2020), Channel Island foxes (*Urocyon littoralis*) (Robinson et al., 2016), Apennine bears (*Ursus arctos marsicanus*) (Benazzo et al., 2017), and Grauer's gorillas (*Gorilla beringei graueri*) have suffered differing genetic consequences in response to severe population decline.

The Iberian lynx suffered drastic population bottlenecks and severe genetic erosion, resulting in a loss of spatiotemporal genetic diversity, increased genetic differentiation between populations, and lineage extinction (Abascal et al., 2016; Casas-Marce et al., 2017). In the Indian tiger, purging has been efficiently removing the high load of deleterious recessive mutations; however, the remaining deleterious alleles occur at high frequencies and may have a fitness cost associated with inbreeding depression (Khan et al., 2021). The Indian lion experienced a low genetic diversity characterized with extreme reduction in heterozigosity with an accumulation of homozygous deleterious mutations (de Manuel et al., 2020), Channel Island foxes have had long term small population sizes resulting in increased homozygosity of deleterious alleles. However, the foxes may have compensated with phenotypic plasticity in regulatory and epigenetic mechanisms (Robinson et al., 2016). Apennine bears experienced inbreeding and low genetic diversity, with subsequent fixation of deleterious mutations by drift. However, their extinction risk remains low, due to dietary adaptations and maintenance of diversity of immune and olfactory receptors genes (Benazzo et al., 2017). A rapid population decline faced during the last 20 years by Grauer's gorilla decreased genome-wide diversity, increased genetic drift and inbreeding, which has increased the frequency of deleterious alleles affecting immunity and methylation genes with a possible impact in pathogens clearance (van der Valk et al., 2019).

Neanderthals and Denisovans may have suffered a rapid accumulation of deleterious alleles that contributed to lowered fitness and populations viability (Harris & Nielsen, 2016; Juric et al., 2016).

To test the effect of population collapse on genomes of Late Pleistocene mammoths, we performed an in-solution target hybridization capture approach on 90 mammoths samples distributed temporally across the Late Pleistocene and much of the known distribution. We focused on genome degradation process in a panel of 259 genes associated as antiviral (host antiviral proteins) (Tenthorey et al., 2020), neutral markers (neutral intron-markers) (Igea et al., 2010), arctic adaptation (genes associated to with arctic-climate adaptation) (V. J. Lynch et al., 2015), cellular process (genes involved in cell stability), immunity (genes associated with antigen recognition and immune response), and sexual markers (introns encode in X and Y chromosomes) (Roca et al., 2005). Our results suggest that accumulation of deleterious alleles was a persistent feature of mammoths throughout the Pleistocene.

2.3. RESULTS

2.3.1. Results Estimation of Endogenous content and coverage

We evaluated the success of our DNA capture-enrichment across 259 target gene regions, estimating the number of sites covered (≥1X) for mitochondrial and enriched nuclear DNA targets (Supplementary Figure 2.1). Mitochondrial DNA coverage was negligible. In two samples, 67 and 107 mtDNA sites were covered (≥1X), indicating that our nuclear DNA capture was specific enough to avoid mtDNA (Supplementary Figure 2.1). Additionally, 56 captured-enriched sequenced libraries with > 50,000 mapped reads were used to compare endogenous DNA, sites coverage, and read lengths for each geographic region Supplementary Figure 2.1 (Boxplot A-C). Sequences from the Alaska-Yukon region yielded more endogenous DNA and sites with coverage ≥ 1X. However, read lengths were longer in Eastern Siberia 78.43 to 87.26 base pairs (bp) and markedly shorter in North America 51.72 to 78.61 base pairs

(bp) Supplementary Figure 2.1 (Boxplot A-C). For all captured-enriched libraries, we calculated the percentage of target sites covered for each sample and each geographic region, finding that Alaska-Yukon and Eastern Siberian samples yielded the most coverage per target Supplementary Figure 2.1 (Boxplot D). Enrichment was more efficient in samples from high-latitude regions such as the Alaska-Yukon and Eastern Siberia regions and decreased for low latitude samples from North America samples. This result is consistent with studies of climate variables that determine aDNA preservation when comparing sub-arctic and arctic samples (Hofreiter et al., 2015; Schwarz et al., 2009; Smith et al., 2003).

2.3.2. Genetic affinity and dynamics of mammoths during the Late Pleistocene

To assess genetic relationships among the newly sequenced mammoths, we compared them to 24 extant and extinct proboscidean genomes and employed multidimensional scaling (MDS) and pairwise-distance analysis to understand the population dynamics throughout the Late-Pleistocene. MDS was used to capture genetic variation among all species and as a way to validate our low coverage data. A summary of type of analyses for each sample is reported in the Supplementary Table 2.1.

The first dimension demonstrated that 15.5% of the total genetic variation separated mammoths from elephants (Figure 2.1). The African elephant cluster retained the straight-tusked elephants suggesting genetic relatedness between African and European elephant species consistent with recent independent data (Palkopoulou et al., 2018). Woolly mammoths and Columbian mammoth also clustered consistent with previous analyses (Enk et al., 2016; Palkopoulou et al., 2018; van der Valk et al., 2021). A second MDS analysis included only Asian elephants and mammoths (Figure 2.1). Mammoths from both sides of the Bering Strait formed a central cluster, with individuals from the Taimyr peninsula, a distant region from Beringia, clustering separately. This suggests that while some geographic differentiation is present, it does not correspond to temporal or geographic proximity suggesting a complex population history across

the region and not a simple separation of two populations across the Bering strait. The genetic variation in these genes is not differentiating between the Columbian mammoth and mammoth cluster, thus suggesting close relatedness or lack of resolution at these markers (Figure 2.1). Comparing per-individual enriched mammoths against each reference genome make evident two groups could be distinguished: an woolly mammoth group and a woolly mammoth group closely genetically related to the Columbian mammoth.

Pairwise-distances were calculated for the 22 capture-enriched mammoth samples with sufficient coverage for analysis compared to a whole mammoth genome sequence panel which contained eight individuals from the Late Pleistocene (Supplementary Table 2.2). Of 22 captured-enriched mammoths, 10 samples were most closely related to genome H (Alaska) (Figure 2.1). Even though mammoths of our samples co-existed with the Oimyakon genome P during the Late Pleistocene, there were remotely related, suggesting that Oimyakon was an isolated lineage. In contrast, we found that five samples (SYU3, Ber5, Ber11, Ber20, and WR2) (Figure 2.1) were most closely related to genome Q, from Wrangel Island (Palkopoulou et al., 2015) despite being a non-contemporary (Figure 2.1) suggesting a temporal continuity of this lineage. We also found interrelationships with lineages distant from northern of Siberia that were dramatically reduced by local extirpation and widespread extinction. Samples 2006-001 from Yakutia (~41,300 yBP) and AM104 (42,764 yBP) from Cleary Creek (Alaska) could represent a small relict of a different lineage, both shared the same nuclear profile but with different mitochondrial haplotype. Mammoth 2002_472 (>48,800 yBP) was closely related to genome G (~31,500 yBP) from the Taymir Peninsula. Mammoth 17301 was related to genome S (from the Yamal Peninsula ~45,300 yBP, mtDNA clade III) despite being sampled in Yukon >45,400 years ago and bearing a North American mitochondrial haplotype (clade I). Consistent with the results of previous studies (Enk et al., 2016; Palkopoulou et al., 2018; van der Valk et al., 2021) Columbian mammoth mtDNA and nuclear genomes suggest an extensive interbreeding with woolly mammoths occurred across their overlapping ranges.

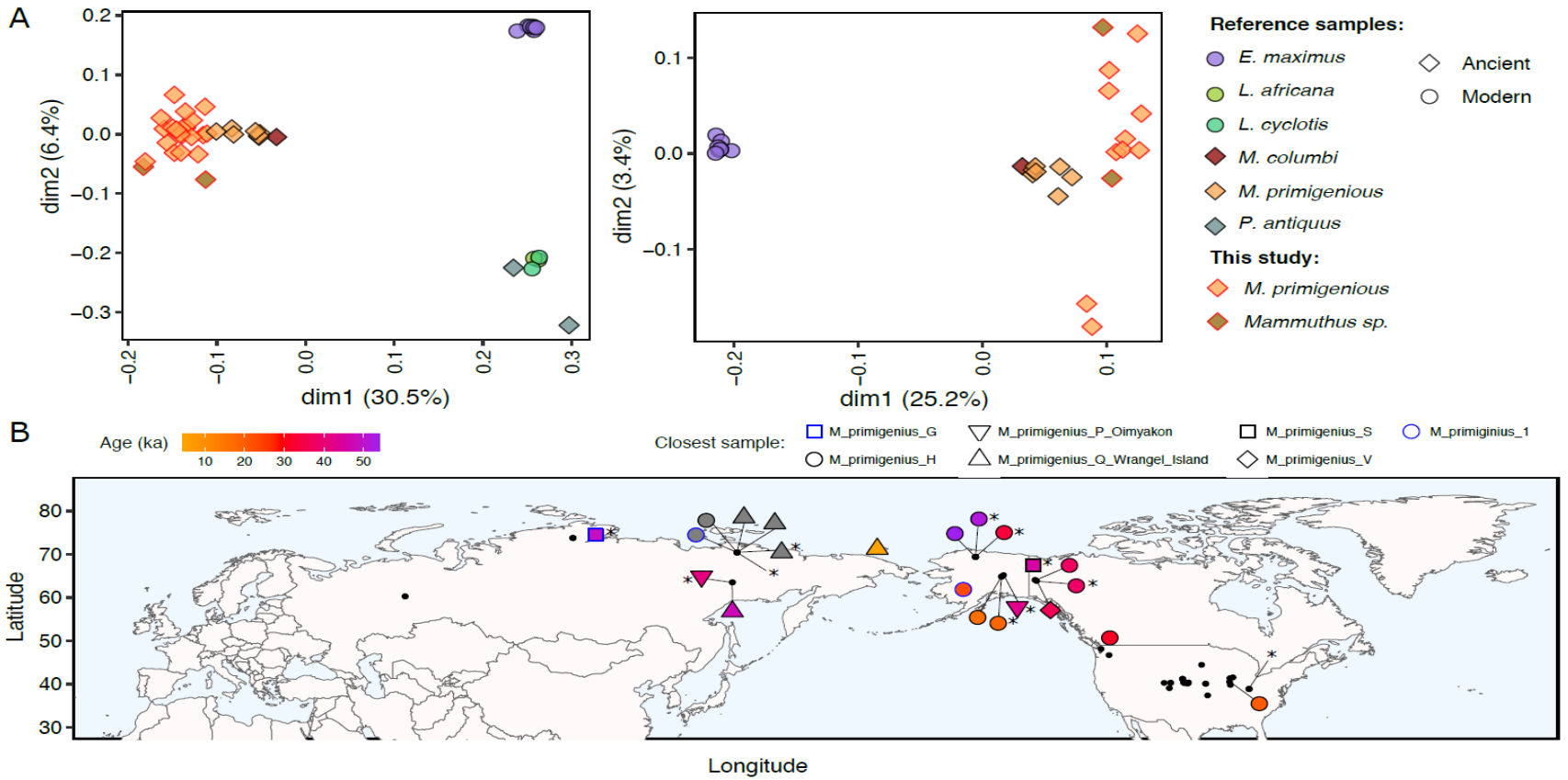


Figure 2.1.Genetic relationships among mammoths.

A. In the left panel, multidimensional scaling (MDS) of extant and extinct proboscidean species were projected onto the first two dimensions estimated (dim1 and dim2) using 15,566,481 transversion sites. In the right panel, MDS of mammoths and Asian elephants were projected. Modern samples are depicted as circles and ancient samples as diamonds. B. A map representing the genetic affinity among mammoths analyzed in this study against a whole mammoth genome panel is shown. Different nuclear lineages are represented by different geometric shapes as described in the figure. The color bar indicates mammoth temporal distributions, and small dots the geographic location. Black dots without lines correspond to samples of undetermined lineage**.**

2.3.3. Patterns of genetic diversity

Arctic adaptation-related genes showed a similar pattern of diversity to all other genes with analyzed with the exception of keratin 3 (*KRT3*), transient receptor potential cation channel subfamily A member 1 (*TRPA1*), and transient receptor potential cation channel subfamily V member 3 (*TRPV3*), which showed a reduction in diversity in mammoths compared to the Asian elephant (Figure 2.2). Similarly, *KRT3*, a keratin gene involved in hair development, and *TRPA1* and *TRPV3* which are temperature-sensitive transient receptor potential (thermoTRP) channels, involved in mammoth cold tolerance adaptation (Lynch et al., 2015) had reduced diversity.

Among immune related genes, mammoth MHC class I gene showed a four-fold increase in diversity compared to other mammoths immune genes (Figure 2.2). The comparison was based on differences in mean genetic distance values. *KIR3DL3* and *KIR3DX1* (Figure 2.2), members of the killer-cell immunoglobulin-like receptors (KIRs) exhibited a reduction in diversity compared to Asian elephants. KIRs include both activating and inhibitory receptors that interact with the MHC class I (Parham, 2005). The interferon-induced transmembrane 1 *IFITM1*-Like gene (Figure 2.2), a potent antiviral effector against human hepatitis C virus (HCV) (Narayana et al., 2015) and *IRGM* (Immunity-related GTPases M) involved in intracellular pathogen defense (Singh et al., 2010) also exhibited decreased diversity (Figure 2.2). *TLR2, TLR10*, and *TLR13* showed lower diversity than other mammoth TLRs (Figure 2.2). In the cellular gene category, MAGE family member H1 (*MAGEH1*) (Figure 2.2), which plays a role on tumorogenesis (Doyle et al., 2010) had reduced diversity compared to the Asian elephant.

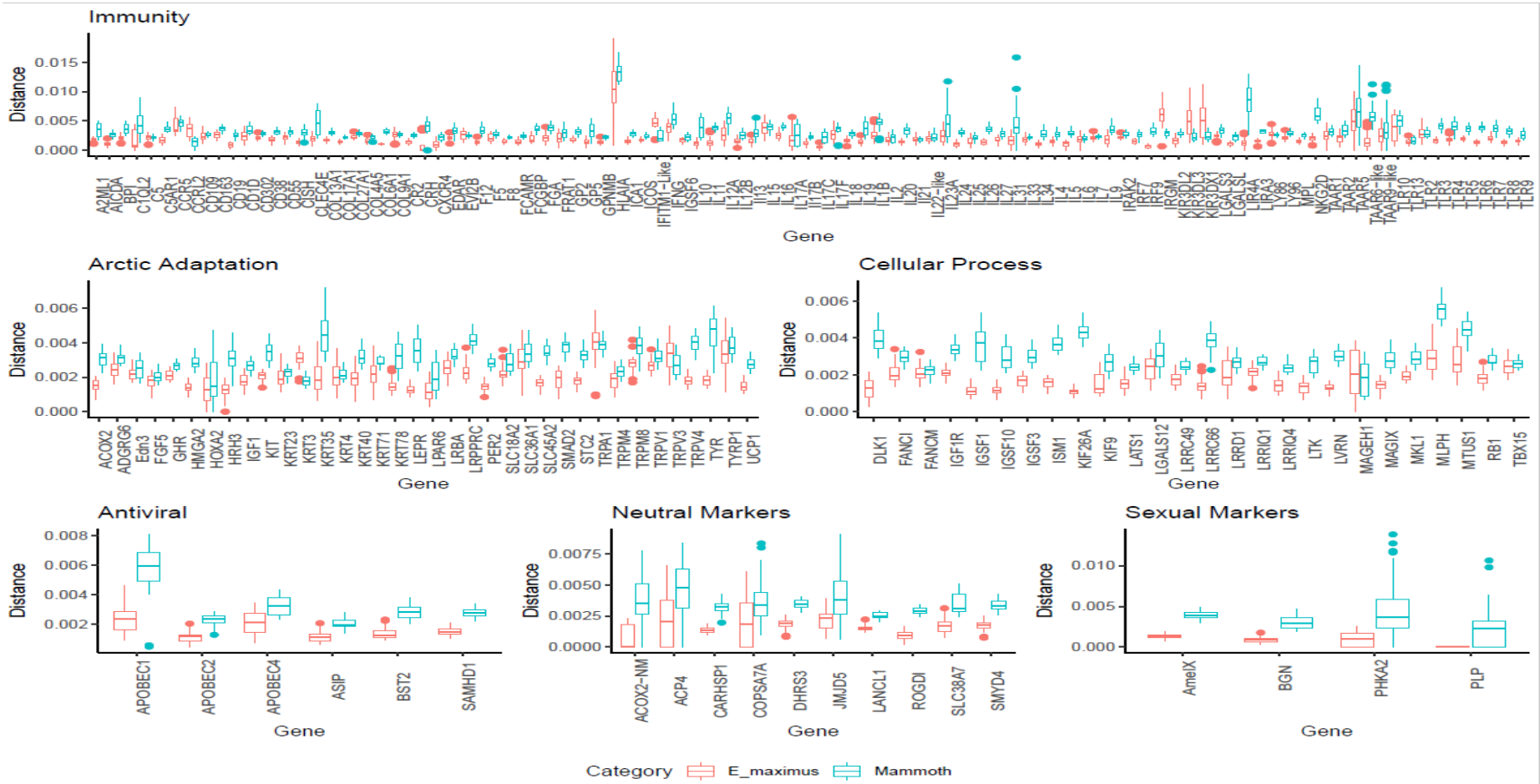


Figure 2.2.**Gene diversity.**

A boxplot compares the genetic distance between present-day Asian elephants and mammoths with genes grouped by function: arctic adaptation, cellular process, antiviral, neutral markers, and sexual markers are shown separately.

2.3.4. Adaptive diversity declines post climate change

We compared mammoth spatial and temporal genetic distances by fitting linear mixed models (LMMs) to different gene functions. Geographic distances were significantly correlated with the genetic distance in arctic adaptation genes, cellular process genes, immunity, and sexual markers, but was not correlated with the genetic distance in neutral markers (Supplementary Table 2.3). Based on gene function temporal genetic diversity estimation, LMMs showed that neutral markers, arctic adaptation, immunity related genes, and sexual markers diversity had a significant effect on mammoth population diversity throughout MIS3 (LGM 26.5-19 Kya). LMMs showed that the effect was maintained after genetic diversity decline occurred during MIS2 and MIS1 periods, in neutral markers, immunity, and sexual markers. In general, sexual markers showed reduced genetic diversity over time, whereas immunity and neutral markers was not linear with some lost diversity recovered over time.

2.3.5. Accumulation of detrimental mutations

For protein functional impact prediction, mammoth genetic variants were examined using the Ensembl predictor (VEP) (McLaren et al., 2016). When looking at the number of unique substitution per length of each gene unique substitutions and coverage per gene were not correlated (Supplementary Figure 2.2). One hundred and ninety three identified SNPs were classified into ten SNP types and variants were sorted into four categories of potential impacts (Supplementary Table 2.4). Fifty three mutations in 41 genes were classified as having low impact (minor coding effect), thirty six mutations in 25 genes as modifier impact, (non-coding variants, where impact prediction is not clear), ninety five mutations in 59 genes as moderate impact (a non-disturbing variant that might alter the protein function), and nine mutations in seven genes as high impact on gene function or loss-of-function (LoF) variants, that are predicted to disrupt gene function (Supplementary Table 2.4). The first group of mutations with predicted low impact, included 43 synonymous variants and ten intronic variants of which three appeared in a splice region in the complement C5 (*C5*), CD109 molecule (*CD109*), and RB transcriptional corepressor 1 (*RB1*) genes. In the second group with modifier

impact variants, 33 mutations occurred in introns, two in coding sequence in the complement C3d receptor 2 (*CR2*) and lymphocyte antigen 9 (*LY9*) genes, and one variant in the downstream region of the inositol-trisphosphate 3-kinase A (*ITPKA*) gene. The moderate impact group contained 94 missense variants, corresponding to non-synonymous single-nucleotide polymorphisms (nsSNPs), including two mutations that altered a splice region of laeverin (*LVRN*) and myeloid cell nuclear differentiation antigen (*MNDA*) genes. A single nsSNP introduced a proline in the T-box transcription factor 15 (*TBX15*) gene by frameshift.

LoF variants, included one premature stop in the C-reactive protein (*CRP*) (Sproston & Ashworth, 2018). Four frameshift deletion variants, two in the immunoglobulin superfamily member 10 (*IGSF10*) gene, one in a splice region of the Leucine rich repeat containing 9 (*LRRC9*) gene, and one in a splice donor site of the *LVRN* gene were observed. Three frameshift insertion variants in the *C5, CR2*, and *MNDA* genes were identified. The mutation in *CR2* occurred in a intron splice region and an additional splice donor variant in an intron of the *MNDA* gene was detected (Supplementary Table 2.4).

We examined whether there was a spatiotemporal association in accumulation of variants as previously observed for Wrangel Island (Fry et al., 2020; Rogers & Slatkin, 2017). We tested genetic variants distribution in the MIS3, MIS2, and MIS1 periods (Figure 2.3). Variants were evenly distributed spatiotemporally suggesting no specific selective pressure during various climatic upheaval.

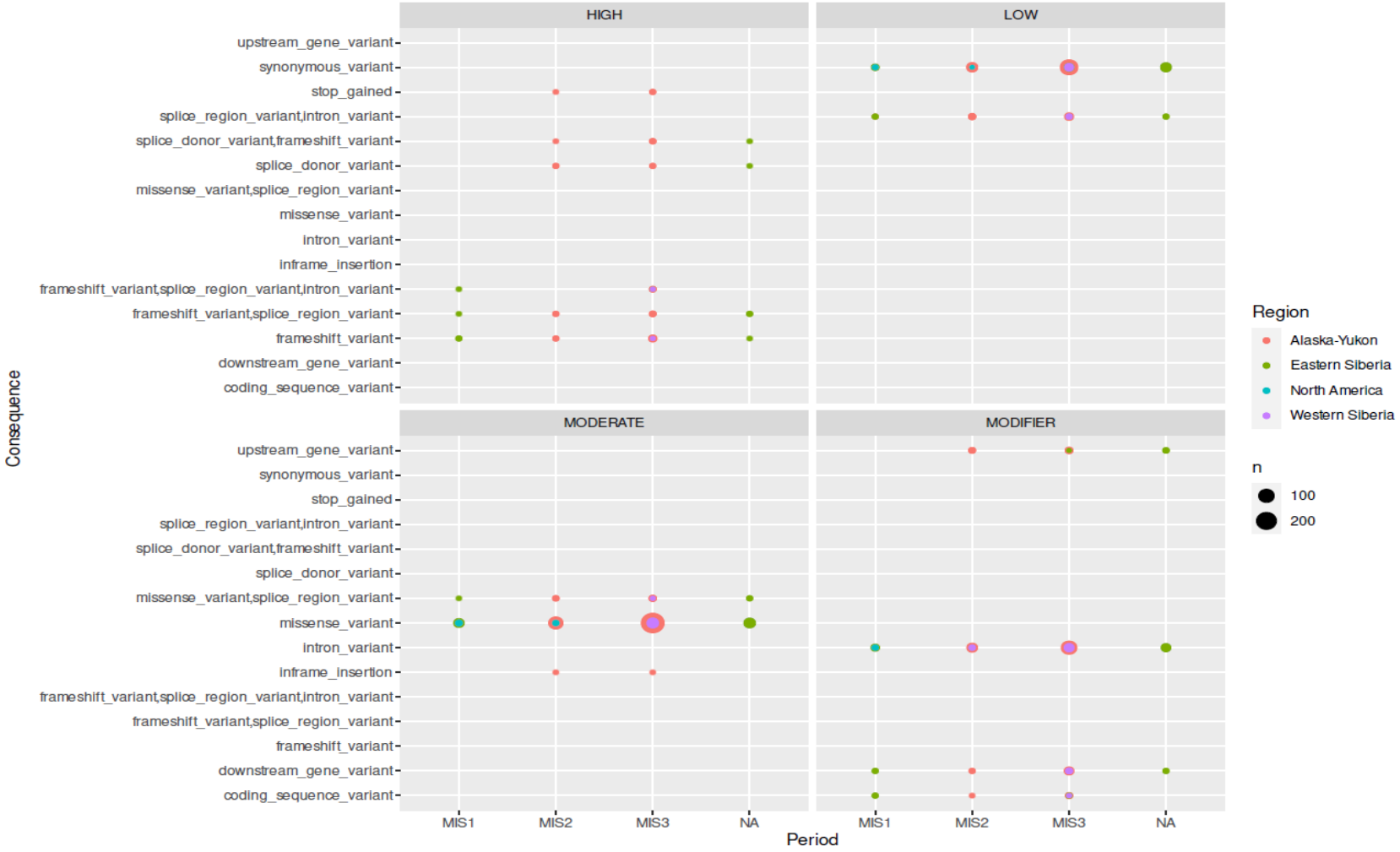


Figure 2.3. Temporal distribution of SNPs during the Late-Pleistocene.
Colored circles indicate the spatiotemporal distribution of genetic variants through the Late Pleistocene (periods MIS3, MIS2, and MIS1). Genetic variants were grouped according potential impacts as high, low, moderate, and modifier. The size of each circle represents the number of SNPs found for each type of variant within the mammoth population. No available information (NA).

2.3.6. Predicted functional impact of missense variants

PredictSNP identified 12 nsSNPs as deleterious substitutions on 9 genes and PROVEAN identified 10 nsSNPs as deleterious substitutions in 9 genes. In total, both methods predicted 17 nsSNPs as deleterious substitutions occurring in 14 genes, (summarized in Supplementary Table 2.9). Both methods classified 5 nsSNPs as deleterious in 4 genes: *MNDA* (S179Y and T228K), the atypical chemokine receptor 4 (*ACKR4*, also know as *CCRL1*) (R149G), the leptin receptor (*LEPR*) (D162V), and Toll-like receptor 13 (*TLR13*) (L142R). The results obtained from I-Muntant2.0 revealed that 74 nsSNPs have negative $\Delta\Delta$G values tending to decrease protein stability, while 20 nsSNPs increase protein stability (Supplementary Table 2.5).

Among the 7 LoF variants, *MNDA* and *IGSF10* carried additional deleterious nsSNPs. Three deleterious substitutions S179Y, T227K, and T228K occurred on the *MNDA* protein domain HIN-200 and may cause loss of modulating cell growth function (Bottardi et al., 2020). However, the $\Delta\Delta$G scores of the T227K and T228K variants show an increase in protein stability. With regard to IGSF10, the deleterious substitution N1893H, is within the immunoglobulin I-set domain which is involved in protein binding functions (Howard et al., 2016).

Three genes with deleterious nsSNPs belong to the Leucine-rich repeat superfamily. Leucine rich repeat and death domain containing 1 (*LRRD1*), an adapter protein whose domain architecture suggests a signaling function, (presumably in apoptosis) (Telliez et al., 2000), carried the L87I deleterious variant. The variant was outside of the repeat and death domains and is predicted to increase protein stability. Leucine rich repeat containing 66 (*LRRC66*), a protein that plays an important role in the development of innate immunity and the nervous system (Y. Chen et al., 2006; Ng et al., 2011), contained two deleterious variants R271P and D683E. The L142R mutation was detected in *TLR13*, a bacterial single-stranded RNA sensor of bacterial 23S rRNA (Oldenburg et al., 2012). *LRRC66* and *TLR13* deleterious nsSNPs are within the LRR structural motif and may produce structural and functional alterations.

Three genes involved in immune cell regulation processes each contained one deleterious nsSNP. L134Y, was observed in the lymphocyte antigen 86 (*LY86*) also known as myeloid differentiation 1 (*MD-1*), the mutation occurring in the MD-2-related lipid-recognition (ML) domain. Together with the radioprotective 105 (*RP105*) protein, LY86 forms a complex that plays a negative regulatory role in *TLR4* signaling (X. Chen et al., 2019). *R149G*, was detected in *ACKR4* a chemokine binding regulator of activated B cells differentiation (Kara et al., 2018) and tumor progression (Whyte et al., 2020) L1093M was identified in the hemoglobin (Hb) scavenger receptor, *CD163*, a macrophage-specific receptor involved in anti-inflammatory process (Etzerodt & Moestrup, 2013).

Among the nsSNPs classified as deleterious, there are 5 genes associated with arctic adaptation, two of these are important in hair growth. A396P substitution in collagen type XVII alpha 1 chain (*COL17A1*), a protein that participates in the assembly of hemidesmosomes and which has been associated as a niche for hair follicle stem cells (Natsuga et al., 2019). Q309H was found in the intermediate filament (IF) rod domain of Keratin 71 (*KRT71*), a protein involved in hair follicle morphogenesis (Fujimoto et al., 2012). I725M was found in the Lipopolysaccharide-responsive beige-like anchor protein (*LRBA*) which supports intracellular vesicular traffic, promotes *CTLA4* expression and associates to the subcellular targeting of hair bundle proteins in cochlear hair cells (Vogl et al., 2017). The D162V substitution occurred in the *LEPR* gene, which encodes a receptor for the fat cell-specific hormone leptin, which is involved in metabolic processes (Münzberg & Morrison, 2015). Trace amine associated receptors 1 and 3 (*TAAR1* and *TAAR3*) are vertebrate odorant G protein-coupled receptors that recognize amines and influence animal behavior (Dewan, 2021). T57S and M17R deleterious substitutions were detected in *TAAR1* and *TAAR3* respectively, at the 7 transmembrane receptor (rhodopsin family) domain suggesting that they may have had functional consequences.

2.4. DISCUSSION

Nuclear DNA sequences from mammoths revealed a central sequence cluster spanning the Bering Strait over much of the Pleistocene. Siberian and North American lineages overlapped, expanded, and migrated in both directions, maintaining mitogenome and nuclear genetic continuity for ~45000 years (Enk et al., 2016). Our data do not support a differentiation between populations of *Mammuthus primigenius* and *Mammuthus columbi*, which could be explained by extensive genetic exchange between populations and between species. This suggests that mammoths formed a largely panmitic population during the Pleistocene. However, they suffered continuous loss of diversity, including the complete loss of mitochondrial and nuclear DNA lineages up and until their restriction to Islands such as Wrangel.

The genetic distances of Arctic adaptation, cellular process, immunity and sexual markers genes were consistent with fine-scale genetic geographic adaptation (for example, microhabitats, refugia or discrete population herds), which was more pronounced throughout the MIS3 period, but became less defined after the LGM when a decrease in genetic diversity occurred. Woolly mammoth populations collapsed in the North dramatically (~25–20 kya) (Murchie et al., 2021; Wang et al., 2021). In contrast, populations increased in central and southern Siberia, suggesting an adaptive advantage for interior populations southward (MacDonald et al., 2012). We detected a resilience signal during the MIS1, thus, during a warm period, (the Bølling–Allerød interstadial, ~14.6-12.9 kya), woolly mammoths suffered a diversity decline, which was followed by a cold period, (the Younger Dryas stadial, ~12.9-11.7 kya), trended toward diversity recovery in neutral markers and immune genes but not sexual markers. The loss of diversity in sexual markers is consistent with lineages extinction. Proboscideans are characterized by matrilineal herds (Brand et al., 2020; Munshi-South, 2011; Pečnerová, Díez-del-Molino, et al., 2017; Schuttler et al., 2014). The patterns of diversity loss associated with adaptive matrilineal ancestry haves been observed in response to climate change in the American mastodon. Mastodons experienced mitochondrial clade extirpation in Northern populations, followed by southern expansion North, of small founder matriarchal

herds in response to climatic warming during interglaciations (Karpinski et al., 2020; Zazula et al., 2014).

The Asian elephant and woolly mammoth clades split from a common ancestor 2.5 Mya (Palkopoulou et al., 2018; Rohland et al., 2010). Compared to Asian elephants woolly mammoth genetic diversity was higher. Asian elephants suffered a sharp demographic decline ~120 kya (Palkopoulou et al., 2018) and genetic diversity has remained low (Fleischer et al., 2001; Vidya, 2016). Mammoths in contrast, despite losing lineages, remained genetically diverse until suffering a severe bottleneck after being isolated on Wrangel Island prior to their extinction (Palkopoulou et al., 2018). Additionally, woolly mammoths had a complex evolutionary history during the Pleistocene characterized by pervasive gene flow, including different episodes of interbreeding with the Columbian mammoth which could explain their relative higher diversity when compared to extant elephants (Enk et al., 2016; Palkopoulou et al., 2018; van der Valk et al., 2021). However, gene flow can introduce maladaptive alleles that can become problematic during periods of population decline. Deleterious alleles may be driven to high frequency, reducing fitness and increasing the genetic load and thus the extinction risk (Walters & Berger, 2019). This has been suggested to have occurred on Wrangel Island (Fry et al., 2020; Rogers & Slatkin, 2017) but according to our analysis likely plagued mammoths throughout the Pleistocene.

We detected low diversity in genes associated with arctic adaptation in mammoths when compared to elephants orthologous genes. *KRT3* is involved in hair growth, cornification, and keratinization and *TRPA1* is involved in thermal sensation. Selection for adaptation to cold temperatures may have reduced diversity in these genes before Columbian and woolly mammoth speciation (van der Valk et al., 2021). However, *TRPV3*, a hypomorphic hair growth and thermal sensation gene which results in reduced sensitivity to warm temperatures (V. J. Lynch et al., 2015) incorporated nonsynonymous changes over several hundreds of thousands of years (van der Valk et al., 2021) suggesting that climate-driven adaptation was local and clinal as suggested by our LMMs analyses. Local adaptation can be observed in humans, where cold temperatures in northern latitudes have selected for a specific

allele of *TRPM8*, a thermosensor that mediates sensitivity to cold temperatures (Key et al., 2018).

In contrast to arctic adaptation genes, *HLAIA* gene (MHC class I) exhibited more variation relative to neutral genomic regions analyzed in this study. This result suggests that the MHC I evolved under long-term balancing selection which is consistent with observed polymorphism in the MHC II DQA in mainland mammoth populations' compared to Wrangel's Island population (Pečnerová et al., 2016). Mammoth MHC diversity patterns were similar to those observed in archaic humans where MHC genes were more diverse than innate immune genes (Reher et al., 2019; Sullivan et al., 2017). This indicates that the MHC accumulated variants that might have been advantageous (de Filippo et al., 2016). However, contrasting with any potential advantages the adaptive immune system might have provided, the innate immune genes were characterized by a paucity of diversity. KIRs, *IFITM1-Like, IRGM, TLR2, TLR10, TLR13*, and *MAGEH1* all exhibited limited genetic diversity. The lack of diversity of these immune genes in mammoths compared to Asian elephants may indicate a reduced adaptive potential and vulnerability to infectious diseases (Morris et al., 2015). Reduced resistance to pathogens due to low diversity of innate immune genes (Castellano et al., 2014) and potentially damaging genetic variants associated with inborn errors of immunity (Zhou et al., 2022) had been proposed as one of the causes of Neanderthal extinction. Diseases as a driver of extinction would be more likely in megafauna particularly susceptible to pathogens (MacPhee, R.D.E. and Marx, P.A., 1997).

While most SNPs identified represented synonymous substitutions, a complex pattern of deleterious variation among mammoths populations was observed prior to restriction to Wrangel Island. Mammoths carried hundreds of deleterious mutations with different frequencies among Pleistocene and Wrangel Island populations (Fry et al., 2020; Rogers & Slatkin, 2017). We observed both loss of function frameshift insertion and splice mutations in genes involved in anti-tumor activity (*MNDA*), host defense and pathogen recognition (*CRP* and *CR2*), placentation (*LVRN*), and immunity (*IGSF10*) and nervous system development (*LRRC9*). In addition, mis-splicing in these genes might potentially compromise their

function thereby resulting in genetic disorders (Y. I. Li et al., 2016; Scotti & Swanson, 2016).

We additionally observed potentially deleterious nsSNPs in 14 genes, including the *MNDA* and *IGSF10* genes. Further nsSNPS in the *LRRD1, LRRC66*, and *TLR13* genes of the Leucine-rich repeat superfamily were identified, which can affect the stability of the LRR structural motif, a versatile structure able to bind diverse proteins and non-protein ligands and primarily involved in innate immunity and neural development (Dolan et al., 2007). This study also identified several deleterious mutations in *LY86, ACKR4*, and *CD163* genes that may have potentially caused dysregulation of immune function.

Pathogenesis in inflammatory bowel disease is associated with activation of *TLR4/NF-κB* signaling mediates by *LY86* suppression (X. Chen et al., 2019). Alteration of activated B cell differentiation and tumor immunity control is caused by chemokine *ACKR4* imbalance (Kara et al., 2018; Whyte et al., 2020). Decrease anti-inflammatory response is mediated by CD163 deficiency which occurs in vascular diseases such as atherosclerosis (Gutiérrez-Muñoz et al., 2020). We observed a high proportion of the deleterious variants occurring together with low diversity in at these genes, which suggest that immune response may have been impaired in mammoths.

The effect of nsSNPs in the *COL17A1, KRT71, LRBA*, and *LEPR* genes are less clear as they are involved in cold tolerance, adapting skin, long hair, and adipose metabolism to arctic conditions (V. J. Lynch et al., 2015). Therefore, the nsSNPs mutations may have been adaptive such as with *TRPV3*, since these genes are involved in similar processes. However, a missense mutation in the human *KRT71* gene has been associated with hypotrichosis, a hereditary hair disorder (Fujimoto et al., 2012). Furthermore, a deficiency in *LRBA* leads to sensorineural hearing loss and a deleterious mutation are associated with the autoimmune condition monogenic lupus (Liphaus et al., 2020). Homozygous loss of function of these genes would likely be maladaptive or lethal. However, functional studies would be

necessary to determine if they could have a potential beneficial phenotypic effect in heterozygous individuals.

Most TAARs genes are suggested to play a role in olfactory activity (Ferrero et al., 2011). The mammoth *TAAR1* and *TAAR3* genes carried deleterious substitutions. In humans, TAARs dysregulation has been associated with mental and metabolic disorders (Rutigliano & Zucchi, 2020). Functional validation of deleterious mutations demonstrated that in the Wrangel Island mammoth, the olfactory receptor *OR5A1* lost the capacity to detect ß-ionones and thus floral scents (Fry et al., 2020). Olfactory receptors suffered a high rate of pseudogenization in the Wrangel Island mammoth (Rogers & Slatkin, 2017). The hydrolethalus syndrome protein 1 (*HYLS1*) is a gene involved in behavioral and neurological function (Fry et al., 2020). Deleterious mutations in this gene may have had adverse neurological and behavioral consequences in mammoths, which likely worsened as the population declined. However, our data suggests such genomic degradation preceded restriction of the mammoths range to Wrangel Island and was a feature common to mammoths throughout the Pleistocene and that the accumulation of deleterious mutations may have been a determining factor in their extinction.

Spatiotemporal analysis suggests that most of the variants identified were present throughout much of the last 50K years and selection was unable to purge them as has been observed in island fox and Indian tiger populations that experienced a dramatic increase and fixation of deleterious variants (Khan et al., 2021; Robinson et al., 2016). At the same time, the accumulation of LoF variants may have occurred over a short time span, as has been demonstrated in Grauer's gorillas (van der Valk et al., 2019). Although genetic purging may remove deleterious alleles in populations with low genetic diversity or small population size (van der Valk, Tom et al., n.d.), an important mechanism to reduce the burden as has been observed in Iberian lynx (Kleinman-Ruiz et al., 2022), a rapid population decline may have a devastating effect even in species with high diversity (van der Valk, Tom et al., n.d.). This appears to be the case in the woolly rhinoceros (Coelodonta antiquitatis), which did not experience reduced genetic diversity, but its extinction, presumably, was driven by climate changes during the Bølling-Allerød interstadial (~14.6-12.9 Kya)

that eliminated its habitat (Lord et al., 2020). Mammoths had the plasticity to respond to this specific climatic event, but their genetic diversity was dramatically altered, probably accelerating the observed genetic erosion.

The low diversity of innate immune genes and the abundance of deleterious or loss of function mutations suggests mammoths suffered from long term genomic degradation throughout the Late Pleistocene. This was likely strongly exacerbated during the Pleistocene-Holocene transition when the population became restricted to islands or small relictual populations on the mainland. The high mutational load may have rendered mammoths less able to adapt to the drastic changes occurring across their range during the end Pleistocene compared to other sympatric species such as horse, reindeer, bison, and musk ox which exhibited a steady decline of genetic diversity before the Pleistocene-Holocene transition but suffered far less extensive subsequent loss (Lorenzen et al., 2011).

2.5. METHODS

2.5.1. Sampling and DNA extraction

Woolly and Columbian mammoths were sampled across much of their historical geographical range and temporal distribution throughout the Late Pleistocene. We combined samples previously included in two phylogeographic mtDNA studies of mammoths (Debruyne et al., 2008; Enk et al., 2016) of which 48 samples originated from Holartic permafrost (Debruyne et al., 2008), and 42 correspond to samples obtained across North America and previously analyzed by Enk et al. 2016 (Enk et al., 2016). In total, 90 samples from five regions including the Urals, the Taimyr Peninsula, Northeast Siberia, Alaska, and sub Arctic North America, with radiocarbon dating that ranged from 54,000 to 4,000 years ago were initially analyzed (Supplementary Table 2.6).

DNA extraction was performed in ancient DNA facilities at the McMaster ancient DNA Centre (McMaster University, Ontario, Canada) as described in (Debruyne et al., 2008; Enk et al., 2016) 11 double-stranded libraries were previously generated

by Enk et al. 2016 (Enk et al., 2016) and 79 double-stranded libraries were produced in the current study, according to Dabney & Meyer, 2012 and Kircher et al., 2012 (Dabney & Meyer, 2012; Kircher et al., 2012) with modifications detailed below. First, libraries were prepared from 42,5 µl of DNA extract adjusting the reaction to 50 µl according to the blunt-end repair protocol, which consisted of a blunt-end repair, adapter ligation and adapter fill-in step. For all purification steps, columns were incubated at 37ºC for 5 min in a heat block using the MinElute PCR purification kit (QIAGEN). All library building procedures were conducted in dedicated ancient DNA facilities at the Department of Wildlife Diseases of the Leibniz Institute for Zoo and Wildlife Research in Berlin, Germany.

2.5.2. Libraries

Double-indexed P5 and P7 specific combination were used for each library and each library was amplified in triplicate. A master mix PCR with Herculase II Fusion DNA Polymerase (Agilent Technologies) was adjusted to a 50 µl final reaction volume using 5 µl of the library as a template. The cycling conditions were a denaturation step of 95ºC for 5 min, followed by 12 cycles of 95ºC for 30 s, 60ºC for 30 s, 72ºC for 40 s, and a final extension of 72ºC for 7 min. Products were pooled and purified using the QiAquick PCR purification kit (QIAGEN) and eluted in 25 µl of TET Elution Buffer (buffer TE with 0.05% Tween-20). To ensure that library preparation and indexing worked, all indexed libraries were screened by qPCR using IS5 and IS6 primers (0,2 µM each) (Kircher et al., 2012) and KAPA SYBR® FAST qPCR Master Mix (2X) (Sigma-Aldrich). Each library was diluted 1:100 in EBT buffer, and standard serial dilutions from 10pM to 0.0625 pM of the PhiX Control V3 425bp-526bp (Illumina, San Diego, CA, USA) were employed as positive control. The final reaction was adjusted to 10 µl using 4 µl of template (libraries and Phix dilutions), and blanks (water and TET buffer). An activation step of 95ºC for 5 min, followed by 40 cycles of 95ºC for 30 s, and 60ºC for 45 s, with a melt curve of 60-95ºC, ending with an 8ºC hold for 30 s was employed.

2.5.3. Baits and Capture

An RNA oligonucleotide bait set of 259 targets was designed and synthesized by Daicel Arbor Biosciences (MyBaits kit) in a 12 K kit with a tilling density of 4 X. The genes of interest were extracted from the African elephant reference genome loxAfr3 from Ensembl (Broad/loxAfr3) using the name of the gene or employing a blast search with the human homologue (Supplementary Table 2.7). Additionally, the array included neutral intron-markers used in multi-locus phylogenetic analysis in mammals (Igea et al., 2010), sex specific markers used to determine African elephant dispersion patterns (Roca et al., 2005), and genes associated with adaptation to Arctic climates (V. J. Lynch et al., 2015).

The 90 mammoth libraries were enriched following the manufacture's instructions (MYcroarray, Mybaits Sequence Enrichment protocol v3.02). Briefly, two successive rounds of hybridization capture were performed for each library with baits at a final concentration 8,0 ng/µl, 55ºC hybridization temperature, 24h hybridization capture, and 20 µl of washed bead per reaction. Every round was washed 4 times at 55ºC, and the pellet was resuspended in EBT buffer, 11 µl in the first round of hybridization and 15 µl in the second round of hybridization. For each round of enrichment, the amplification was carried out in a final volume of 40 µl using the IS5 and IS6 primers (0,15 µM each) and the KAPA SYBR ® FAST qPCR Master Mix (2X) (Sigma-Aldrich). The cycling consisted of a 95˚C activation step for 5 min, followed by 12 cycles of 95ºC 30 s, and 60ºC 45 s, and a final step of 60ºC for 3 min. The final products were quantified using the same conditions described for qPCR. Of 90 libraries, 78 captured-enriched libraries were pooled in equimolar ratios and sequenced on an Illumina MiSeq plataform using paired-end sequencing at the Leibniz Institute for Zoo and Wildlife Research in Berlin, Germany.

2.5.4. Sequence read processing

The reads were demultiplexed according to the respective indexes using bcl2fastq v. 2.17 (Illumina, San Diego, CA, USA). Adapter sequences, low quality stretches and leading/trailing N's were removed from the sequencing reads using AdapterRemoval 2.0 (Schubert et al., 2016). Reads shorter than 30 bp after trimming were discarded. Clean reads were then mapped to the African elephant reference genome loxAfr4 (Broad/loxAfr4) (Palkopoulou et al., 2018) using bwa-backtrack v0.7.15 algorithm (H. Li & Durbin, 2010). We disabled the seed parameter (l) to improve the mapping of reads with an excess of terminal substitutions, as expected from ancient DNA reads (Schubert et al., 2012). Non-mapping reads and reads with mapping quality lower than 30 were discarded. PCR duplicates were removed by using Picard-tools v2.6.0 (http://picard.sourceforge.net), and then realigned using GATK v3.6 (DePristo et al., 2011). We recalculated the MD-tag using Samtools version 1.9 (H. Li et al., 2009). Finally, we mapped the newly sequenced mammoths and reference samples to the loxAfr3 version of the African elephant reference genome using the same parameters described above. The loxAfr3 alignments were used exclusively for the Ensembl Variant Effect Predictor (VEP) (McLaren et al., 2016) analyses.

Sequencing and mapping statistics for the newly sequenced samples can be found in Supplementary Table 2.1. Sequencing reads of new and reference samples were processed using the same procedure. Next, we estimated genomic and on-target coverage, and endogenous content. Based on these results a second sequencing round was done for 20 samples (Supplementary Table 2.1) in order to increase the coverage.

2.5.5. Proboscidean genomes

For comparative purposes, we analyzed the captured-enriched data together with published genome data from proboscidean species (Dastjerdi et al., 2014; DePristo et al., 2011; Lynch et al., 2015; Meyer et al., 2017; Palkopoulou et al., 2015, 2018; Reddy et al., 2015; Yamagata et al., 2019). This dataset consisted of genome-wide sequencing data for 24 extant and extinct species. The extant species include eight Asian elephants (Elephas maximus), two forest elephants (*Loxodontha cyclotis*), two savanna elephants (*Loxodontha africana*) and the extinct species contain two straight-tusked elephants (*Elephas antiquus*), seven woolly mammoths (*Mammutus primiginius*), one Columbian mammoth (*Mammutus columbi*), and two American mastodons (*Mammut americanum*) (Supplementary Table 2.8).

2.5.6. aDNA damage

We used mapDamage2.0 (Jónsson et al., 2013) to evaluate the authenticity of aDNA sequencing data. MapDamage2.0 was run on the read alignments restricted to bases with a minimum Phred quality score of 20. Nucleotide substitution patterns and read length distribution are shown in Supplementary Figure 2.3. Most samples showed an increase of C to T and G to A substitutions at the end of the read, which is characteristic of authentic aDNA data.

2.5.7. Multidimensional Scaling Plots (MDS)

We used MDS analysis as an initial approach to evaluate the reliability of the data and compare the newly sequenced mammoths with reference samples. To do so, we built a dataset consisting of the newly sequenced samples and seven previously sequenced M. primigenius, one *M. columbi,* eight *E. maximus*, two *P. antiquus*, two *L. africana*, two *L. cyclotis* and two *M. americanum* (Dastjerdi et al., 2014; V. J. Lynch et al., 2015; Meyer et al., 2017; Palkopoulou et al., 2018; Reddy et al., 2015; Yamagata et al., 2019). Given the low coverage obtained for many of the samples, instead of calling genotypes, we followed a random sampling approach. For each

individual and each site, we sampled a random read after discarding reads with mapping quality below 30 and bases with quality below 20 using ANGSD v0.916 (Korneliussen et al., 2014). After discarding samples with less than 0.01% non-missing sites, invariable sites, transition sites and sites with more than 85% missing data we had a final dataset consisting of 46 samples (22 newly sequenced and 24 reference samples) and 15,566,481 transversion sites. We then used Plink2.0 (C. C. Chang et al., 2015) to estimate pairwise distances between samples and the *calmd* function in R to estimate the MDS. We performed an MDS analysis using the complete dataset (that included 12 of the newly sequenced samples that had at least 0.02% non-missing SNPs). Additionally, we used the pairwise-distance estimated using *Plink* to assess which of the reference mammoth samples was the closest to each of the newly sequenced mammoths.

2.5.8. MDS downsampling tests

To evaluate the effect of the missing data on the MDS analysis, we performed a subsampling experiment using the dataset described in the section above. We sub sampled the high coverage M-primigenius_V sample to varying depth of coverage (from 0.01% to 50% of the sites in the SNP dataset) and performed an MDS plot in each case (Supplementary Table 2.9). We used this experiment to set a threshold for the minimum coverage required to confidently place a given sample in the space of the MDS plot. Our results show that for the MDS plot built using the complete dataset (including *L. africana, L. cyclotis, P. antiquus, M. primgenius, M. columbi*, and *E. maximus*), we require a minimum of 0.01% non-missing sites (~1500), while for MDS plot restricting to *M. primigenius, M. columbi*, and *E. maximus* we require a minimum of 0.01-0.02% non-missing sites (~3000-4000). A total of 22 of the newly sequenced samples were within the first threshold and a total of 12 within the second.

2.5.9. Per-target diversity estimates

To evaluate changes in diversity over time and across different geographic regions on the captured genes, we estimated the pairwise-distance between samples for each of the captured genes. As describe previously, genes were categorized according to the main function in antiviral, neutral markers, Arctic adaptation, cellular process, immunity, and sexual markers. For each captured gene and each sample, we generated a consensus sequence using ANGSD v0.916 -dofasta 1 (Korneliussen et al., 2014). For a given pair of samples and a captured gene, we counted the fraction of differences between the pair across all sites in cases where both samples had no missing data (IBS pairwise-distance). We assessed the effect of different missingness thresholds in the percentage of non-missing sites in each paired comparison (including all samples, 20%, 50% and 75% coverage per gene and paired comparison) as well as the inclusion/exclusion of transition sites in the patterns observed. As expected, results including transitions showed on average a higher distance. Similarly, we observed a higher variance when we did not apply a minimum coverage filter. For the results presented here, we restricted the analysis to transversion sites and pairs with a minimum coverage of 50%. Given that the additional error derived from the aDNA damage is expected to artificially increase the diversity in ancient samples we focused on genes that showed a decrease in genetic diversity in the mammoths compared to the Asian elephant. After applying these filters, our comparisons involved a total of 18 mammoth samples, 11 of which were sequenced in this study, and 8 Asian elephant samples.

2.5.10. Linear mixed model

We build a linear mixed model to estimate the relationships of mammoths' genetic distance with geographic distance and temporal distribution. The different genes were analyzed separately, based on their genes function: neutral, antiviral, Arctic adaptation, cellular process, immunity, and sexual markers. Among the six models, gene identity was considered as a random factor. Fifty percent coverage per site

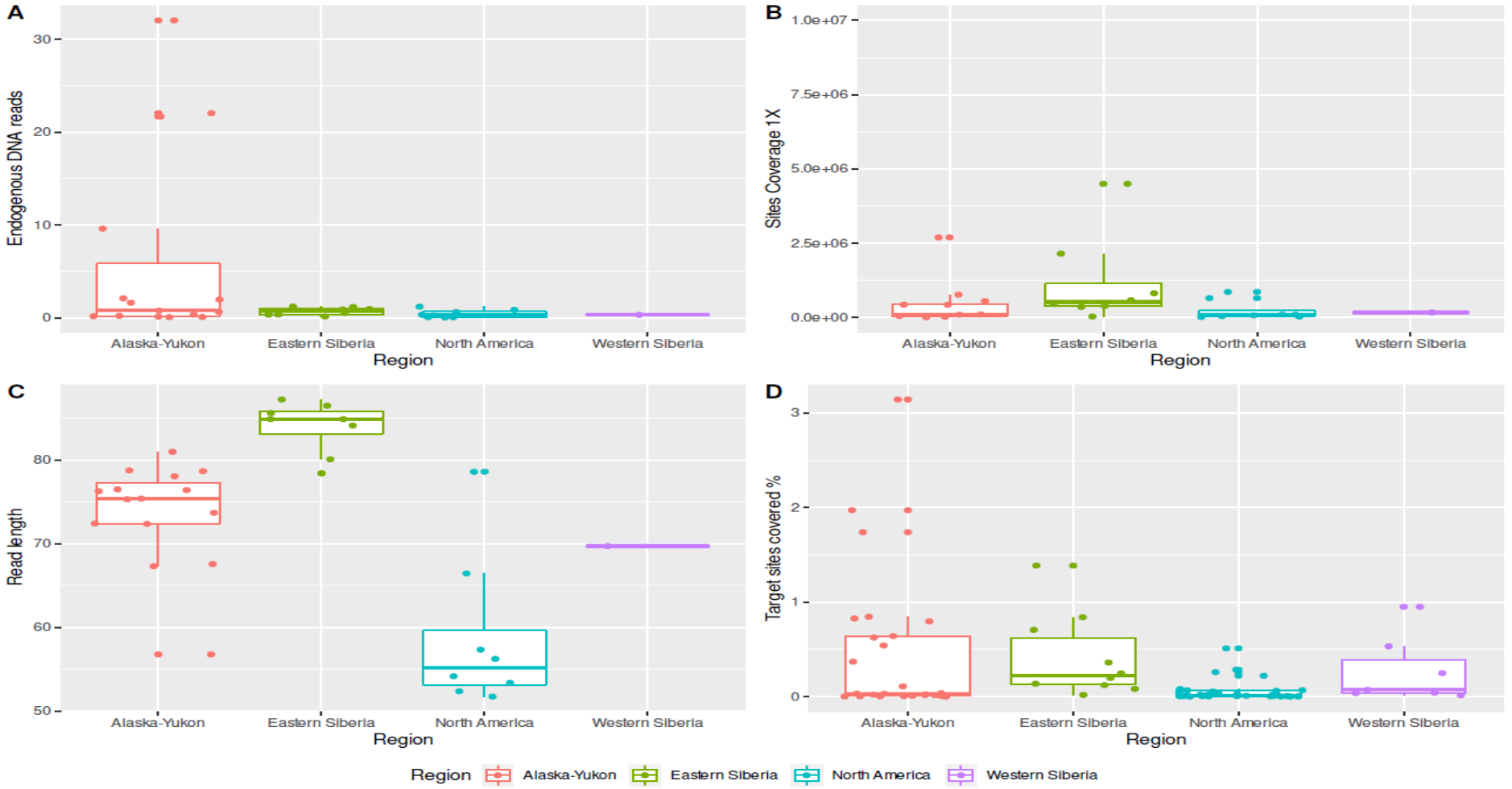
pairwise genetic distance was used for each individual. Due to the intermittent connection between Eurasian and North American mammoth population across the Bering land bridge in the Bering Strait, we forced the geographic distance measurement to go from east to west (through the Pacific Ocean). Samples were grouped according to Marine Isotope Stage 3 (MIS3, 27-60 Kya), Marine isotope stage 2 (MIS2, 15-27 Kya), and Marine isotope stage 1 (MIS1, 4-14 Kya). Because of the low coverage sites for samples of MIS1 and MIS2, the sample data were merged into MIS1, and all the missing data were ignored during the analysis. The linear mixed models (lmm) were performed using the “lme” function from the package “nlme”. All the analyses were performed in R 4.0.3 environment (R Core Team, 2020).

2.5.11. SNPs effect analysis

We used VEP (McLaren et al., 2016) to annotate the SNP substitutions on the target genes that showed contrasting diversity patterns between the mammoths and the African elephant. For this analysis we used the read alignments to the loxAfr3 of the reference genome. We first used GATK HaplotypeCaller to perform genotype-calling at the target genes and for the newly sequenced and reference mammoths. Bases with quality below 20 and reads with mapping quality below 30 were excluded. Additionally, we restricted the analysis to transversion SNPs in order to reduce the error derived from aDNA damage. We note that given the low coverage of our sequencing data we did not apply a depth of coverage filter, thus our data consists of mostly haploid calls and not proper diploid genotypes. We ran VEP to annotate the SNPs using the loxAfr3.0 database. Gene isoforms were removed, and variants detected in at least two samples were kept for the SNP analysis. We estimated if gene length and/or coverage have a correlation with the type of genetic variants and if there are associated with number of SNPs substitutions.

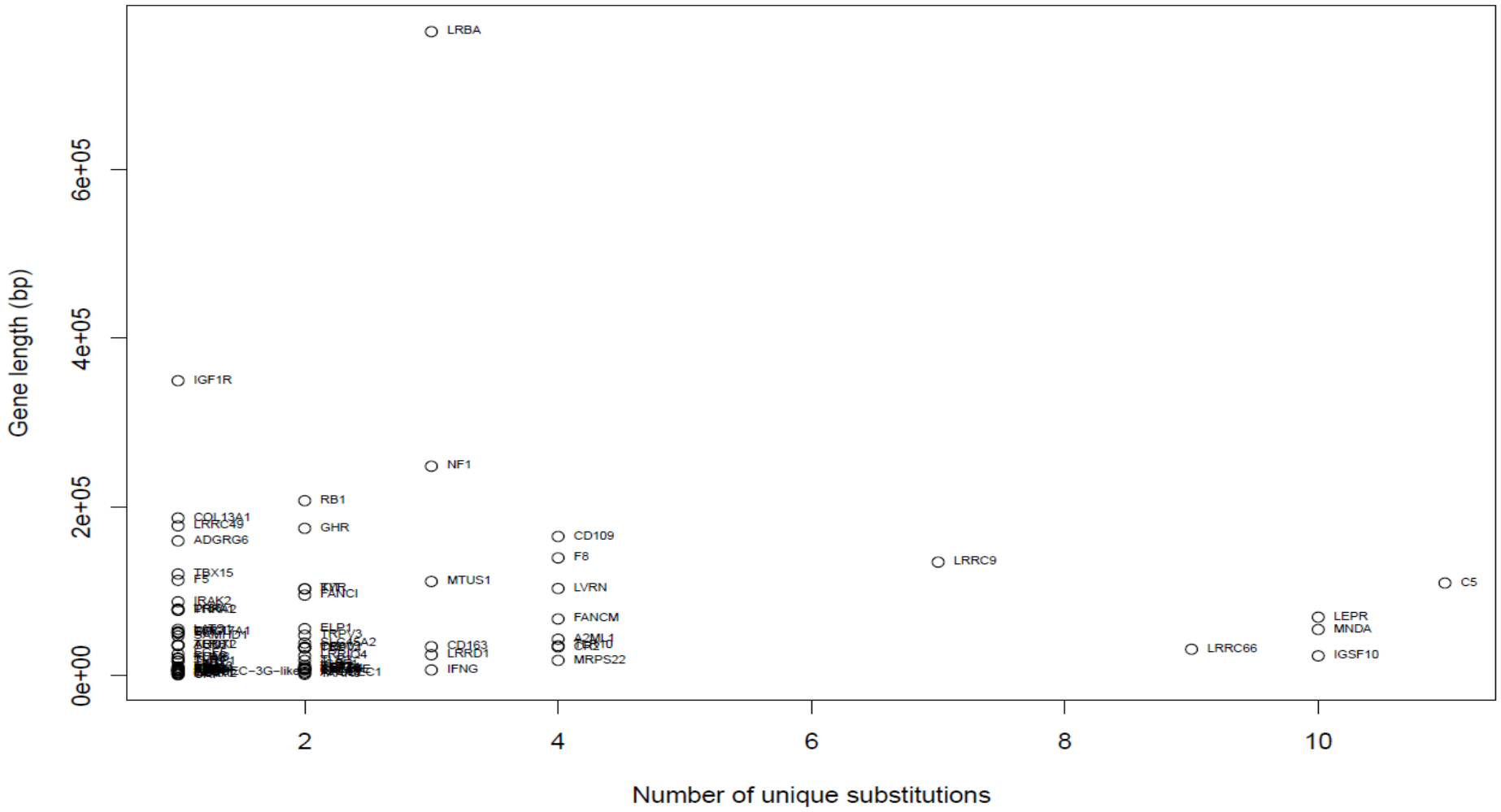
2.5.12. Missense variants analysis

Missense variant protein FASTA sequences were retrieved using Uniprot IDs (The UniProt Consortium et al., 2021) https://www.uniprot.org/. For distinguishing deleterious missense variants from neutral ones, two computational predictors tools were used. The program predictSNP (Bendl et al., 2014) https://loschmidt.chemi.muni.cz/predictsnp1/ is a machine learning consensus classifier that integrates six prediction tools (MAPP, PhD-SNP, PolyPhen-1, PolyPhen-2, SIFT, and SNAP). We also employed Protein Variation Effect Analyzer (PROVEAN) (Choi et al., 2012; Choi & Chan, 2015) http://provean.jcvi.org/seq_submit.php. This tool predicts the functional effect of a protein variant by searching for homolog proteins with BLAST for any organism, and providing a deleterious or neutral amino acid variation score. To investigate if missense mutations will affect the proteins stability, the Gibbs free energy change ($\Delta\Delta G$) was calculated using the support vector machine I-Mutant2.0 (Capriotti et al., 2005,



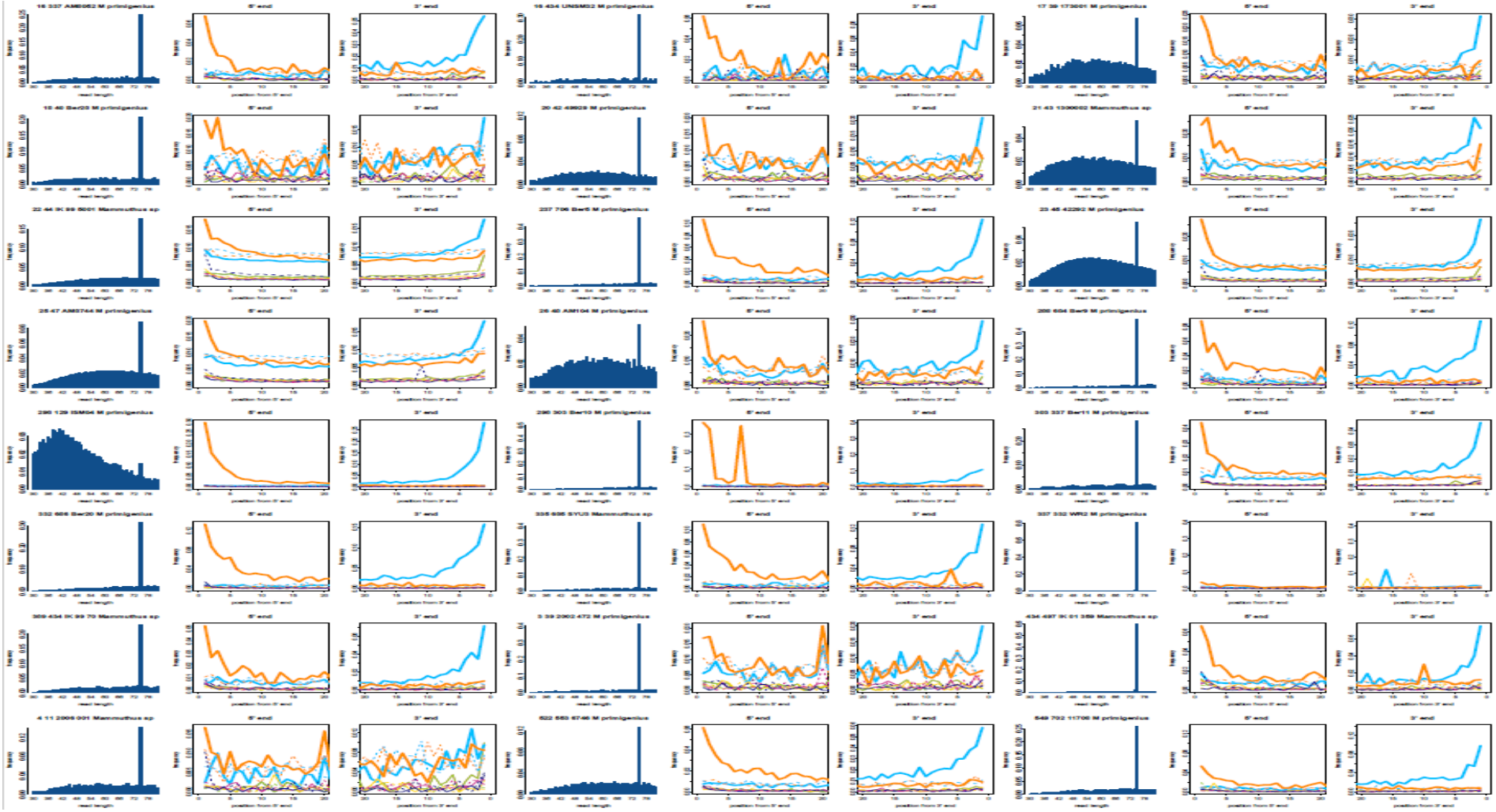
Supplementary Figure 2.1. DNA capture-enrichment quality.

A comparison of the quality of data obtained by capture-amplification experiment compared by region is shown. A. The proportion of endogenous reads recovered. B. correspond of sites coverage at least once aligned over the entire African elephant reference genome loxAfr4 (Broad/loxAfr4). C. Illustrates that read length distribution is longer in arctic regions (Alaska-Yukon and Eastern Siberia). D. Coverage estimation of the reads mapping to the target regions.



Supplementary Figure 2.2. Correlation analysis between gene length and nucleotide substitutions.

The *x-axis* represents the number of unique substitutions detected by gene and the y-*axis* the length in bp within different genes. The distribution of unique substitutions does not depend on gene length.



Supplementary Figure 2.3. MapDamage profile.

DNA damage pattern in the samples analyzed in this study. Length distribution of mapped reads (left graph). Substitutions C > T (5′) (central graph) and G > A (3′) (right graph) are shown in orange and blue, respectively, coincide with ancient endogenous DNA patterns.

Supplementary Table 2.1. Sequencing mapping statistics for mammoth libraries.

Summary statistics for captured-enrichment samples sequenced in this study. Samples included for each analysis are indicated in type of analysis and trimmed reads obtained following a second sequencing round is provided. Mapping statistics correspond to the total of reads remained after filtering and coverage for whole genome, mtDNA, and chromosome X are reported.

Individuals		Type of analysis					Second Sequencing	Reads									Whole-genome coverage			mtDNA coverage	Chr X coverage			
Sample	Specie	Endogenous DNA >50000 reads	MDS	ADMIXTURE	Diversity estimates	SNP effect	Trimmed reads	Total reads (pairs)	Trimmed reads	Mapped reads	Mapped + filt + dedup reads	Endogenous DNA (% mapped reads)	Clonality	Read length	GC content	Chr X reads	Number of sites covered (≥1X)	Coverage	Avg. Doc	mtDNA sites covere (≥1X)	chrX of sites covered	Chr X coverage	Chr X Avg. DoC	X/Y
Ber11	M primigenius	Yes	Yes	Yes	Yes	Yes	3817351	5246469	5571324	5286970	68273	1.225	0.985	78.43	0.477	1483	4498475	0.137	1.19	0	99634	0.0651	1.167	0.98
WR2	M primigenius	Yes	Yes	Yes	No	Yes	3282796	4564508	5375479	5045325	49064	0.913	0.988	80.1	0.492	1255	2149645	0.066	1.828	0	83152	0.0555	1.209	0.66
ISM04	M primigenius	Yes	Yes	Yes	No	Yes	1609423	2732446	2813210	1793205	17755	0.631	0.987	52.36	0.484	409	862694	0.026	1.078	0	19783	0.0086	1.083	1
IK-99-5001	Mammuthus sp	Yes	Yes	Yes	Yes	Yes	2874776	2544153	2885687	1775807	625465	21.675	0.548	78.06	0.432	23669	46565382	1.423	1.049	0	2E+06	11.546	1.04	0.99
IK-01-359	Mammuthus sp	Yes	Yes	Yes	No	Yes	252788	2357365	2718293	2515484	6282	0.231	0.997	81	0.456	156	433015	0.013	1.175	0	9069	0.0061	1.393	1.19
Ber5	M primigenius	Yes	Yes	Yes	No	Yes	520796	2103137	2240703	2052997	7755	0.346	0.995	84.9	0.473	308	583225	0.018	1.129	0	22551	0.0159	1.16	1.03
Ber9	M primigenius	Yes	Yes	Yes	No	Yes	359688	1137101	1270825	1149907	4546	0.358	0.995	87.26	0.461	174	357387	0.011	1.11	0	13904	0.0101	1.092	0.98
Ber20	M primigenius	Yes	Yes	Yes	No	Yes	458638	987361	1048085	932669	10105	0.964	0.987	84.91	0.468	400	816867	0.025	1.05	0	32553	0.023	1.043	0.99
IK-99-70	Mammuthus sp	Yes	Yes	Yes	Yes	Yes	No	982257	1027178	991028	8104	0.789	0.99	78.77	0.473	388	555147	0.017	1.15	0	25035	0.0164	1.221	1.06
11708	M primigenius	Yes	Yes	Yes	No	Yes	375489	976169	1044619	979386	9079	0.869	0.989	78.61	0.46	193	653028	0.02	1.093	0	14393	0.0094	1.054	0.96
AM8052	M primigenius	Yes	Yes	Yes	No	Yes	277976	954144	998134	917044	6656	0.667	0.991	76.42	0.45	222	430167	0.013	1.183	0	12711	0.0081	1.335	1.13
SYU3	Mammuthus sp	Yes	Yes	Yes	No	Yes	276322	821423	906956	773109	4918	0.542	0.992	84.14	0.468	176	390641	0.012	1.059	0	13599	0.0095	1.089	1.03
AM8744	M primigenius	Yes	Yes	Yes	Yes	Yes	713662	701992	720541	615325	158811	22.041	0.674	72.42	0.445	6781	10574283	0.323	1.088	107	433477	0.2615	1.133	1.04
6746	M primigenius	Yes	Yes	Yes	No	Yes	47203	647157	669539	626136	10967	1.638	0.979	73.7	0.498	253	769304	0.024	1.051	0	17030	0.0105	1.095	1.04
42292	M primigenius	Yes	Yes	Yes	Yes	Yes	632179	635052	639099	523098	204714	32.032	0.499	67.31	0.42	5237	11173374	0.342	1.233	68	245485	0.1376	1.436	1.16
Ber10	M primigenius	Yes	Yes	Yes	No	Yes	No	470508	519980	498839	6130	1.179	0.985	85.64	0.463	254	458036	0.014	1.146	0	19628	0.014	1.108	0.97
1300002	Mammuthus sp	Yes	Yes	Yes	No	Yes	422817	412821	424374	381056	40796	9.613	0.869	67.57	0.443	1527	2697118	0.082	1.022	0	99269	0.0559	1.039	1.02
30136	M primigenius	Yes	No	No	No	Yes	No	800136	912211	360282	780	0.086	0.998	75.32	0.475	14	53805	0.002	1.092	0	937	0.0006	1.125	1.03
2000-174	M primigenius	Yes	No	No	No	Yes	198281	765675	868137	789032	2830	0.326	0.996	69.71	0.439	107	170746	0.005	1.155	0	6301	0.0037	1.184	1.02

Individuals		Type of analysis					Second Sequencing	Reads									Whole-genome coverage			mtDNA coverage	Chr X coverage			
Sample	Specie	Endogenous DNA >50000 reads	MDS	ADMIXTURE	Diversity estimates	SNP effect	Trimmed reads	Total reads (pairs)	Trimmed reads	Mapped reads	Mapped + filt + dedup reads	Endogenous DNA (% mapped reads)	Clonality	Read length	GC content	Chr X reads	Number of sites covered (≥1X)	Coverage	Avg. Doc	mtDNA sites covere (≥1X)	chrX of sites covered	Chr X coverage	Chr X Avg. DoC	X/Y
T-02-110	Mammuthus sp	Yes	No	No	No	Yes	No	723109	752920	606565	1408	0.187	0.997	76.28	0.5	31	96172	0.003	1.117	0	2470	0.0016	0.957	0.86
UNSM23	Mammuthus sp	Yes	No	No	No	Yes	42168	585110	572922	141463	1541	0.269	0.98	66.46	0.459	74	95330	0.003	1.074	0	4007	0.0022	1.227	1.14
UNSM01	Mammuthus sp	Yes	No	No	No	Yes	No	576363	576169	279096	2005	0.348	0.991	51.72	0.485	78	101284	0.003	1.024	0	3862	0.0017	1.045	1.02
780001	M primigenius	Yes	No	No	No	Yes	No	469742	633758	604539	12564	1.982	0.974	75.41	0.294	16	54106	0.002	17.511	0	1331	0.0008	0.906	0.05
DMNS23	Mammuthus sp	Yes	No	No	No	Yes	No	75184	74823	56300	909	1.215	0.982	54.16	0.467	53	47259	0.001	1.042	0	2626	0.0012	1.093	1.05
DMNS28b	M columbi	Yes	No	No	No	No	1658298	2716450	3143043	419028	1348	0.043	0.995	57.33	0.463	30	72929	0.002	1.06	0	1739	0.0008	0.989	0.93
UNSM24	Mammuthus sp	Yes	No	No	No	No	No	858683	852178	168436	570	0.067	0.996	56.24	0.473	10	30965	0.001	1.035	0	545	0.0003	1.032	1
11340	M primigenius	Yes	No	No	No	No	No	407630	410119	377091	1448	0.353	0.996	78.68	0.466	73	104875	0.003	1.086	0	5413	0.0035	1.061	0.98
UNSM21	M columbi	Yes	No	No	No	No	160242	366551	423863	67496	325	0.077	0.986	53.38	0.449	22	16971	0.001	1.022	0	1107	0.0005	1.061	1.04
30133	M primigenius	Yes	No	No	No	No	No	304755	371382	112470	395	0.106	0.996	72.38	0.492	7	27518	0.001	1.039	0	442	0.0003	1.146	1.1
Ber7	M primigenius	Yes	No	No	No	No	No	275129	282015	270717	445	0.158	0.998	86.51	0.456	10	34216	0.001	1.125	0	802	0.0006	1.079	0.96
AM523	M primigenius	Yes	No	No	No	No	No	151377	150320	140435	207	0.138	0.998	56.77	0.504	6	11147	0	1.054	0	271	0.0001	1.257	1.19
42135	M primigenius	Yes	No	No	No	No	No	92481	106213	91841	2237	2.106	0.967	76.51	0.544	7	35169	0.001	4.866	0	568	0.0004	0.943	0.19
49929	M primigenius	No	Yes	Yes	No	Yes	No	20201	20501	17437	11543	56.305	0.14	70.09	0.415	362	647426	0.02	1.25	0	19173	0.0112	1.323	1.06
AM104	M primigenius	No	Yes	Yes	Yes	Yes	No	14992	14781	12861	9848	66.626	0.02	68.42	0.416	382	438313	0.013	1.537	0	14987	0.0085	1.744	1.13
173001	M primigenius	No	Yes	Yes	Yes	Yes	No	13536	13496	11775	8525	63.167	0.092	68.15	0.421	613	364539	0.011	1.594	0	19299	0.011	2.165	1.36
2002-472	M primigenius	No	Yes	Yes	Yes	Yes	No	12886	14727	11742	7380	50.112	0.233	84.04	0.419	248	402735	0.012	1.54	0	12891	0.009	1.617	1.05
2006-001	Mammuthus sp	No	Yes	Yes	Yes	Yes	No	12016	12335	10662	7963	64.556	0.054	73.23	0.392	304	354910	0.011	1.643	0	13445	0.0082	1.656	1.01
AM1208	M primigenius	No	No	No	No	Yes	No	24236	24335	22664	483	1.985	0.978	72.47	0.468	20	33901	0.001	1.033	0	1465	0.0009	0.989	0.96
UNSM32	M primigenius	No	No	No	Yes	Yes	7842	17066	19869	6552	1138	5.728	0.773	76.38	0.396	83	66647	0.002	1.304	0	5620	0.0036	1.128	0.86
DMNS47	M columbi	No	No	No	No	Yes	3266	15728	16618	15219	647	3.893	0.949	82.5	0.458	23	51874	0.002	1.029	0	1926	0.0013	0.985	0.96
URL1	Mammuthus sp	No	No	No	No	Yes	2923	15091	17400	16483	545	3.132	0.962	82.84	0.455	26	41459	0.001	1.089	0	2241	0.0015	0.961	0.88
Ber28	M primigenius	No	No	No	Yes	Yes	No	9325	9805	7898	5701	58.144	0.059	77.44	0.416	392	325837	0.01	1.355	0	18590	0.012	1.633	1.21

Individuals		Type of analysis					Second Sequencing	Reads									Whole-genome coverage			mtDNA coverage	Chr X coverage			
Sample	Specie	Endogenous DNA >50000 reads	MDS	ADMIXTURE	Diversity estimates	SNP effect	Trimmed reads	Total reads (pairs)	Trimmed reads	Mapped reads	Mapped + filt + dedup reads	Endogenous DNA (% mapped reads)	Clonality	Read length	GC content	Chr X reads	Number of sites covered (≥1X)	Coverage	Avg. Doc	mtDNA sites covere (≥1X)	chrX of sites covered	Chr X coverage	Chr X Avg. DoC	X/Y
2005-915	M primigenius	No	No	No	No	Yes	No	8856	10349	5281	3575	34.544	0.112	86.54	0.422	116	263533	0.008	1.174	0	8366	0.006	1.2	1.02
URL2	Mammuthus sp	No	No	No	No	Yes	No	1629	1632	1258	262	16.054	0.76	75.93	0.425	15	19268	0.001	1.032	0	902	0.0006	1.263	1.22
49562	M primigenius	No	No	No	No	Yes	No	627	686	407	203	29.592	0.389	72.91	0.407	9	14365	0	1.03	0	583	0.0004	1.126	1.09
2190003	M primigenius	No	No	No	No	No	No	1557061	1993207	1066	83	0.004	0.838	74.02	0.447	2	5571	0	1.103	0	150	0.0001	0.987	0.89
8572	M primigenius	No	No	No	No	No	No	1367824	1518727	43796	334	0.022	0.988	60.86	0.487	20	19535	0.001	1.041	0	1089	0.0006	1.118	1.07
ISM01	M primigenius	No	No	No	No	No	No	1069904	1178600	26289	196	0.017	0.99	58.79	0.437	4	10240	0	1.125	0	244	0.0001	0.964	0.86
WAST_01	Mammuthus sp	No	No	No	No	No	No	723151	717615	24248	570	0.079	0.964	48.01	0.493	13	25197	0.001	1.086	0	582	0.0002	1.072	0.99
UNSM27	M primigenius	No	No	No	No	No	No	685369	710721	31395	89	0.013	0.971	49.96	0.477	2	4329	0	1.027	0	78	0	1.281	1.25
ISM12	Mammuthus sp	No	No	No	No	No	No	544578	616653	28292	363	0.059	0.981	55.29	0.461	15	19671	0.001	1.02	0	929	0.0004	0.893	0.87
AK-323-V-I	M primigenius	No	No	No	No	No	No	480991	540155	542	43	0.008	0.104	71.53	0.409	2	2903	0	1.06	0	120	0.0001	1.192	1.13
UNSM09	Mammuthus sp	No	No	No	No	No	No	474118	527707	19911	168	0.032	0.972	57.51	0.45	4	9493	0	1.018	0	325	0.0002	0.708	0.7
UNSM07	Mammuthus sp	No	No	No	No	No	No	450600	440709	281	20	0.005	0.921	65.2	0.435	3	1295	0	1.007	0	220	0.0001	0.889	0.88
AM1187	M primigenius	No	No	No	No	No	No	366868	427254	448	26	0.006	0.907	55.62	0.45	1	1446	0	1	0	75	0	0.742	0.74
UNSM16	Mammuthus sp	No	No	No	No	No	No	297459	283273	4917	64	0.023	0.985	54.97	0.476	2	3415	0	1.03	0	82	0	1.341	1.3
UNSM29	M primigenius	No	No	No	No	No	No	200662	217565	162	35	0.016	0.698	63.83	0.427	1	2162	0	1.033	0	48	0	1.33	1.29
DMNS49	M columbi	No	No	No	No	No	No	149124	168459	4201	92	0.055	0.963	65.4	0.437	5	5722	0	1.052	0	346	0.0002	0.945	0.9
CMNH40031	Mammuthus sp	No	No	No	No	No	No	140753	148200	34326	144	0.097	0.996	49.04	0.485	1	6861	0	1.029	0	63	0	0.778	0.76
UCMP09	Mammuthus sp	No	No	No	No	No	No	37022	39077	128	27	0.069	0.767	65.22	0.425	3	1760	0	1.001	0	174	0.0001	1.125	1.12
GDY1	Mammuthus sp	No	No	No	No	No	No	20176	20055	7910	324	1.616	0.955	68.21	0.447	16	21813	0.001	1.013	0	1192	0.0007	0.916	0.9
AM1189	M primigenius	No	No	No	No	No	No	8647	10244	2042	12	0.117	0.964	65.5	0.483	0	786	0	1	0	0	0	NA	NA
UNSM14	Mammuthus sp	No	No	No	No	No	No	7154	7718	309	14	0.181	0.517	71	0.444	0	994	0	1	0	0	0	NA	NA
UNSM08	M columbi	No	No	No	No	No	No	5288	5149	65	12	0.233	0.745	41.33	0.489	1	496	0	1	0	50	0	0.827	0.83
UNSM02	Mammuthus sp	No	No	No	No	No	No	2052	2042	1324	89	4.358	0.919	79.27	0.43	6	6842	0	1.031	0	370	0.0002	1.285	1.25

Individuals		Type of analysis					Second Sequencing	Reads								Whole-genome coverage			mtDNA coverage	Chr X coverage			
Sample	Specie	Endogenous DNA >50000 reads	MDS	ADMIXTURE	Diversity estimates	SNP effect	Trimmed reads	Total reads (pairs)	Trimmed reads	Mapped reads	Mapped + filt + dedup reads	Endogenous DNA (% mapped reads)	Clonality	Read length	GC content	Chr X reads	Number of sites covered (≥1X)	Coverage	Avg. Doc	mtDNA sites covere (≥1X)	chrX of sites covered	Chr X coverage	Chr X Avg. DoC	X/Y
IK-99-322	Mammuthus sp	No	No	No	No	No	No	1055	1115	4	2	0.179	0.5	81	0.394	0	162	0	1	0	0	0	NA	NA
ISM15	Mammuthus sp	No	No	No	No	No	No	1004	950	461	35	3.684	0.891	65.8	0.414	0	2300	0	1.001	0	0	0	NA	NA
UNSM22	Mammuthus sp	No	No	No	No	No	No	657	443	1	0	0	NA	NA	NA	0	0	0	NA	0	0	NA	NA	NA
2000-173	M primigenius	No	No	No	No	No	No	652	729	674	62	8.505	0.858	84.1	0.439	2	5185	0	1.006	0	150	0.0001	1.121	1.12
DMNS08	Mammuthus sp	No	No	No	No	No	No	472	448	168	27	6.027	0.799	72.67	0.41	3	1962	0	1	0	196	0.0001	1.112	1.11
8139	M primigenius	No	No	No	No	No	No	446	492	167	34	6.911	0.714	63.68	0.418	2	2165	0	1	0	133	0.0001	0.958	0.96
UNSM15	M columbi	No	No	No	No	No	No	400	472	6	3	0.636	0	63.33	0.411	0	190	0	1	0	0	0	NA	NA
UNSM34	M columbi	No	No	No	No	No	No	364	262	61	35	13.359	0.186	60.54	0.387	1	2074	0	1.022	0	61	0	0.993	0.97
ISM07	M jeffersonii	No	No	No	No	No	No	213	245	67	20	8.163	0.6	60.2	0.418	2	1203	0	1.001	0	140	0.0001	0.86	0.86
AM1193	M primigenius	No	No	No	No	No	No	107	112	62	7	6.25	0.885	66.57	0.45	1	466	0	1	0	74	0	0.9	0.9
UNSM33	Mammuthus_sp	No	No	No	No	No	No	52	63	5	4	6.349	0	68.75	0.452	0	275	0	1	0	0	0	NA	NA
UNSM30	M_primigenius	No	No	No	No	No	No	48	54	7	3	5.556	0.5	48.67	0.31	0	146	0	1	0	0	0	NA	NA
AM2446	M_primigenius	No	No	No	No	No	No	17	15	0	0	0	NA	NA	NA	0	0	0	NA	0	0	NA	NA	NA

Supplementary Table 2.2. Whole mammoth genome sequence panel.

The panel includes eight mammoths across the Late-Pleistocene. Geographic distribution and mitochondrial clade were used for genetic affinity comparison with the samples enriched and sequenced in this study.

Genome ID	Name	Specie	Date, years ago	Geographic origin	Mitochondrial clade (Haplotype)
M. primigenius_G	Woolly mammoth	*Mammutus primiginius*	~31,500	Taimyr Peninsula, Russia	Clade I (E1)
M. primigenius_H	Woolly mammoth	*Mammutus primiginius*	~44,900	Alaska, USA	Clade I (C1)
M. primigenius_S	Woolly mammoth	*Mammutus primiginius*	~45,300	Yamal Peninsula, Russia	Clade III (B2)
M. primigenius_V	Woolly mammoth	*Mammutus primiginius*	~42,400	Wyoming, USA	Clade I (NA)
M. primigenius_Q	Woolly mammoth	*Mammutus primiginius*	~4,300	Wrangel Island, Russia	Clade I (NA)
M_primigenius_P	Woolly mammoth	*Mammutus primiginius*	~44,800	Oimyakon, Russia	Clade II (A1)
M primigenius -1 (Yuka)	Woolly mammoth	*Mammutus primiginius*	~28,140	Siberia, Russia	Clade I (NA)
M. columbi_U	Columbian mammoth	*Mammutus columbi*	~13400	Wyoming, USA	Clade I (NA)

Supplementary Table 2.3. Linear mixed models.

Results of LMMs relating mammoths' genetic and temporal distance with gene function. Standard error (SE), degree of freedom (DF), t-values, and p-values. Significant effects are highlighted in bold.

	Fixed Effects					
		Value	**SE**	**DF**	**t-value**	**p-value**
Neutral markers	Intercept	0.004636616	0.0004294222	439	10.797338	0.0000
	Per_diffMIS 1_MIS 3	-0.001126830	0.0003992227	439	-2.822560	**0.0050**
	Per_diffMIS 3_MIS 3	-0.000920003	0.0004050205	439	-2.271498	**0.0236**
	dist_geo	-0.000000039	0.0000000364	439	-1.084613	0.2787
Antiviral	Intercept	0.0029296655	0.0005622663	148	5.210459	0.0000
	Per_diffMIS 3_MIS 3	0.0002524799	0.0001830106	148	1.379591	0.1698
	dist_geo	0.0000000141	0.0000000461	148	0.305042	0.7608
Arctic Adaptation	Intercept	0.0027775829	0.0001334224	870	20.817963	0e+00
	Per_diffMIS 3_MIS 3	0.0001696302	0.00004633613	870	3.660862	**3e-04**
	dist_geo	0.0000000440	1.148000e-08	870	3.833384	**1e-04**
Cellular Process	Intercept	2.62629e-03	1.687788e-04	596	15.560542	0.0000
	Per_diffMIS 3_MIS 3	1.72294e-05	4.490433e-05	596	0.383692	0.7013
	dist_geo	9.15000e-08	1.125000e-08	596	8.139052	**0.0000**
Immunity	Intercept	0.004147007	0.0005103479	2812	8.125843	0.0000
	Per_diffMIS 1_MIS 3	-0.001161031	0.0004889756	2812	-2.374415	**0.0176**
	Per_diffMIS 3_MIS 3	-0.001107316	0.0004891371	2812	-2.263815	**0.0237**
	dist_geo	0.000000061	0.0000000130	2812	4.672039	**0.0000**
Sexual markers	Intercept	0.0019531151	0.0007376276	385	2.647833	0.0084
	Per_diffMIS 1_MIS 3	0.0018549161	0.0005817362	385	3.188586	**0.0015**
	Per_diffMIS 3_MIS 3	0.0028563642	0.0005799929	385	4.924826	**0.0000**
	dist_geo	-0.0000001816	0.0000000574	385	-3.162176	**0.0017**

Supplementary Table 2.4. Predicted SNPs effect.

Table describing the consequence and the impact for each variant; genome annotation and protein features based on African elephant reference genome *loxAfr3* for each gene are included. Accumulation of SNPs variants observed in the mammoth population are shown in the n.ind column and gene length is given.

	SYMBOL	Location	Gene	Consequence	IMPACT	Feature type	Protein position	Amino acids	Codons	Uniprot	UNIPARC	n. ind	Gene length
1	TAAR3	scaffold_0:27567772-27567772	ENSLAFG00000030742	missense_variant	MODERATE	Transcript	176	M/R	aTg/aGg	G3U118	UPI0001C5FC2A	2	1776
2	TAAR3	scaffold_0:27568229-27568229	ENSLAFG00000030742	synonymous_variant	LOW	Transcript	328	I	atA/atT	G3U118	UPI0001C5FC2A	10	1776
3	PMEL	scaffold_2:45328720-45328720	ENSLAFG00000014591	missense_variant	MODERATE	Transcript	324	E/V	gAg/gTg	G3TDI1	UPI0001C5E157	2	8274
4	TLR13	scaffold_24:23116114-23116114	ENSLAFG00000025903	missense_variant	MODERATE	Transcript	412	L/R	cTc/cGc	G3UEZ6	UPI0001C5AC8A	6	10266
5	TLR11	scaffold_256:178541-178541	ENSLAFG00000029363	missense_variant	MODERATE	Transcript	153	N/K	aaC/aaA	G3U832	UPI0001C5FDB2	3	7362
6	TLR11	scaffold_256:178706-178706	ENSLAFG00000029363	missense_variant	MODERATE	Transcript	208	S/R	agC/agG	G3U832	UPI0001C5FDB2	2	7362
7	MNDA	scaffold_33:7973961-7973961	ENSLAFG00000002774	missense_variant	MODERATE	Transcript	16	D/V	gAt/gTt	G3TRT8	UPI0001C5C21A	5	54935
8	MNDA	scaffold_33:7974506-7974506	ENSLAFG00000002774	missense_variant	MODERATE	Transcript	78	G/V	gGa/gTa	G3TRT8	UPI0001C5C21A	5	54935
9	MNDA	scaffold_33:7974525-7974525	ENSLAFG00000002774	splice_donor_variant	HIGH	Transcript	-	-	-	G3TRT8	UPI0001C5C21A	5	54935
10	MNDA	scaffold_33:7974970-7974970	ENSLAFG00000002774	missense_variant	MODERATE	Transcript	111	V/L	Gta/Tta	G3TRT8	UPI0001C5C21A	3	54935
11	MNDA	scaffold_33:7981639-7981639	ENSLAFG00000002774	frameshift_variant	HIGH	Transcript	201	E/EX	gaa/gAaa	G3TRT8	UPI0001C5C21A	5	54935
12	MNDA	scaffold_33:7981648-7981648	ENSLAFG00000002774	missense_variant	MODERATE	Transcript	204	V/F	Gtt/Ttt	G3TRT8	UPI0001C5C21A	2	54935
13	MNDA	scaffold_33:7981667-7981667	ENSLAFG00000002774	missense_variant	MODERATE	Transcript	210	E/V	gAg/gTg	G3TRT8	UPI0001C5C21A	9	54935
14	MNDA	scaffold_33:7981718-7981718	ENSLAFG00000002774	missense_variant	MODERATE	Transcript	227	T/K	aCa/aAa	G3TRT8	UPI0001C5C21A	4	54935
15	MNDA	scaffold_33:7981721-7981721	ENSLAFG00000002774	missense_variant, splice_region_variant	MODERATE	Transcript	228	T/K	aCg/aAg	G3TRT8	UPI0001C5C21A	4	54935
16	MNDA	scaffold_33:8102089-8102089	ENSLAFG00000002774	missense_variant	MODERATE	Transcript	179	D/Y	Gat/Tat	G3TRT8	UPI0001C5C455	6	54935
17	FGA	scaffold_51:1096901-1096901	ENSLAFG00000032360	missense_variant	MODERATE	Transcript	253	K/T	aAg/aCg	G3U772	UPI0001C5CAAD	4	10479
18	APOBEC-3G-like	scaffold_73:6503283-6503283	ENSLAFG00000020906	missense_variant	MODERATE	Transcript	219	P/T	Cct/Act	G3TLG1	UPI0001C5DA6D	2	4519
19	A2ML1	scaffold_15:54008684-54008684	ENSLAFG00000017288	synonymous_variant	LOW	Transcript	326	T	acG/acC	G3TIL7	UPI0001C5E533	5	43464
20	A2ML1	scaffold_15:54015995-54015995	ENSLAFG00000017288	missense_variant	MODERATE	Transcript	485	E/D	gaG/gaC	G3TIL7	UPI0001C5E533	6	43464
21	A2ML1	scaffold_15:54033251-54033251	ENSLAFG00000017288	synonymous_variant	LOW	Transcript	1297	L	ctC/ctG	G3TIL7	UPI0001C5E533	2	43464
22	A2ML1	scaffold_15:54039703-54039703	ENSLAFG00000017288	synonymous_variant	LOW	Transcript	1433	S	tcG/tcC	G3TIL7	UPI0001C5E533	4	43464
23	ACKR4	scaffold_103:3874007-3874007	ENSLAFG00000014708	missense_variant	MODERATE	Transcript	128	Q/H	caG/caC	G3TDP3	UPI0001C5FFA8	2	8165
24	ACKR4	scaffold_103:3874068-3874068	ENSLAFG00000014708	missense_variant	MODERATE	Transcript	149	R/G	Cga/Gga	G3TDP3	UPI0001C5FFA8	2	8165
25	ACOX2	scaffold_12:8774662-8774662	ENSLAFG00000027265	missense_variant	MODERATE	Transcript	649	S/A	Tct/Gct	G3U1U7	UPI0001C5ADAF	2	36552

	SYMBOL	Location	Gene	Consequence	IMPACT	Feature type	Protein position	Amino acids	Codons	Uniprot	UNIPARC	n. ind	Gene length
26	ACP4	scaffold_4:10627502-10627502	ENSLAFG00000001738	intron_variant	MODIFIER	Transcript	-	-	-	G3SP54	UPI0001C5CFEF	5	6290
27	ADGRG6	scaffold_0:16673868-16673868	ENSLAFG00000005050	synonymous_variant	LOW	Transcript	870	S	tcC/tcG	G3SVG3	UPI0001C5B35A	3	159924
28	APOBEC1	scaffold_15:53748325-53748325	ENSLAFG00000030837	missense_variant	MODERATE	Transcript	108	H/Q	caT/caA	G3U0R4	UPI0001C5E4DC	2	3139
29	APOBEC1	scaffold_15:53748854-53748854	ENSLAFG00000030837	synonymous_variant	LOW	Transcript	158	P	ccG/ccC	G3U0R4	UPI0001C5E4DC	2	3139
30	ARNTL	scaffold_21:36176422-36176422	ENSLAFG00000026673	intron_variant	MODIFIER	Transcript	-	-	-	G3TVX8	UPI0001C5C69E	2	35840
31	BPI	scaffold_19:37485057-37485057	ENSLAFG00000005285	missense_variant	MODERATE	Transcript	71	N/K	aaT/aaG	G3SVY2	UPI0001C5F262	2	50445
32	C5	scaffold_6:38197001-38197001	ENSLAFG00000013034	synonymous_variant	LOW	Transcript	16	T	acT/acA	G3TAK1	UPI0001C5F4F7	7	109981
33	C5	scaffold_6:38197052-38197052	ENSLAFG00000013034	intron_variant	MODIFIER	Transcript	-	-	-	G3TAK1	UPI0001C5F4F7	3	109981
34	C5	scaffold_6:38210839-38210839	ENSLAFG00000013034	intron_variant	MODIFIER	Transcript	-	-	-	G3TAK1	UPI0001C5F4F7	2	109981
35	C5	scaffold_6:38210841-38210841	ENSLAFG00000013034	splice_region_variant, intron_variant	LOW	Transcript	-	-	-	G3TAK1	UPI0001C5F4F7	2	109981
36	C5	scaffold_6:38210850-38210850	ENSLAFG00000013034	frameshift_variant	HIGH	Transcript	140-141	-/X	-/T	G3TAK1	UPI0001C5F4F7	3	109981
37	C5	scaffold_6:38210954-38210954	ENSLAFG00000013034	intron_variant	MODIFIER	Transcript	-	-	-	G3TAK1	UPI0001C5F4F7	2	109981
38	C5	scaffold_6:38219964-38219964	ENSLAFG00000013034	intron_variant	MODIFIER	Transcript	-	-	-	G3TAK1	UPI0001C5F4F7	3	109981
39	C5	scaffold_6:38228465-38228465	ENSLAFG00000013034	missense_variant	MODERATE	Transcript	556	I/M	atT/atG	G3TAK1	UPI0001C5F4F7	6	109981
40	C5	scaffold_6:38240914-38240914	ENSLAFG00000013034	intron_variant	MODIFIER	Transcript	-	-	-	G3TAK1	UPI0001C5F4F7	2	109981
41	C5	scaffold_6:38270043-38270043	ENSLAFG00000013034	intron_variant	MODIFIER	Transcript	-	-	-	G3TAK1	UPI0001C5F4F7	2	109981
42	C5	scaffold_6:38282836-38282836	ENSLAFG00000013034	intron_variant	MODIFIER	Transcript	-	-	-	G3TAK1	UPI0001C5F4F7	3	109981
43	CD109	scaffold_0:92051216-92051216	ENSLAFG00000006753	synonymous_variant	LOW	Transcript	1179	S	tcT/tcA	G3SYT3	UPI0001C5D8AE	3	165168
44	CD109	scaffold_0:92066103-92066103	ENSLAFG00000006753	intron_variant	MODIFIER	Transcript	-	-	-	G3SYT3	UPI0001C5D8AE	2	165168
45	CD109	scaffold_0:92073370-92073370	ENSLAFG00000006753	missense_variant	MODERATE	Transcript	804	Q/E	Cag/Gag	G3SYT3	UPI0001C5D8AE	5	165168
46	CD109	scaffold_0:92082093-92082093	ENSLAFG00000006753	splice_region_variant, intron_variant	LOW	Transcript	-	-	-	G3SYT3	UPI0001C5D8AE	2	165168
47	CD163	scaffold_15:52345224-52345224	ENSLAFG00000002941	missense_variant	MODERATE	Transcript	1093	L/M	Ctg/Atg	G3SRE6	UPI0001C5E7ED	4	34603
48	CD163	scaffold_15:52352414-52352414	ENSLAFG00000002941	intron_variant	MODIFIER	Transcript	-	-	-	G3SRE6	UPI0001C5E7ED	2	34603
49	CD163	scaffold_15:52352474-52352474	ENSLAFG00000002941	missense_variant	MODERATE	Transcript	700	V/F	Gtt/Ttt	G3SRE6	UPI0001C5E7ED	4	34603
50	CD302	scaffold_3:68193526-68193526	ENSLAFG00000010149	intron_variant	MODIFIER	Transcript	-	-	-	G3T527	UPI0001C5F046	2	33788
51	CD302	scaffold_3:68194644-68194644	ENSLAFG00000010149	missense_variant	MODERATE	Transcript	117	D/E	gaT/gaA	G3T527	UPI0001C5F046	9	33788
52	CLEC4E	scaffold_15:52572775-52572775	ENSLAFG00000022194	missense_variant	MODERATE	Transcript	82	K/N	aaG/aaC	G3TN20	UPI0001C5E843	2	7027
53	CLEC4E	scaffold_15:52572886-52572886	ENSLAFG00000022194	synonymous_variant	LOW	Transcript	119	L	ctG/ctT	G3TN20	UPI0001C5E843	5	7027
54	COL13A1	scaffold_10:51508063-51508063	ENSLAFG00000015605	intron_variant	MODIFIER	Transcript	-	-	-	G3TFI4	UPI0001C5F476	2	187098
55	COL17A1	scaffold_10:14403043-14403043	ENSLAFG00000011870	missense_variant	MODERATE	Transcript	396	A/P	Gcc/Ccc	G3TZN0	UPI0001C5F092	3	52171
56	CR2	scaffold_13:12923243-12923243	ENSLAFG00000009563	missense_variant	MODERATE	Transcript	58	L/V	Ttg/Gtg	G3T3Z4	UPI0001C5C5B6	6	34190

	SYMBOL	Location	Gene	Consequence	IMPACT	Feature type	Protein position	Amino acids	Codons	Uniprot	UNIPARC	n. ind	Gene length
57	CR2	scaffold_13:12925579-12925579	ENSLAFG00000009563	synonymous_variant	LOW	Transcript	301	V	gtT/gtG	G3T3Z4	UPI0001C5C5B6	2	34190
58	CR2	scaffold_13:12930824-12930824	ENSLAFG00000009563	frameshift_variant, splice_region_variant, intron_variant	HIGH	Transcript	738-739	-/X	-/T	G3T3Z4	UPI0001C5C5B6	7	34190
59	CR2	scaffold_13:12930825-12930825	ENSLAFG00000009563	coding_sequence_variant	MODIFIER	Transcript	-	-	-	G3T3Z4	UPI0001C5C5B6	7	34190
60	CRP	scaffold_33:9027520-9027520	ENSLAFG00000029698	stop_gained	HIGH	Transcript	110	L/*	tTg/tAg	G3UIV1	UPI0001C5C36F	3	1158
61	ELP1	scaffold_6:50803877-50803877	ENSLAFG00000000736	synonymous_variant	LOW	Transcript	698	R	cgG/cgC	G3SMA4	UPI0001C5F784	2	55915
62	ELP1	scaffold_6:50810895-50810895	ENSLAFG00000000736	missense_variant	MODERATE	Transcript	931	S/T	aGt/aCt	G3SMA4	UPI0001C5F784	7	55915
63	F5	scaffold_33:24844509-24844509	ENSLAFG00000001975	missense_variant	MODERATE	Transcript	814	L/R	cTg/cGg	G3SPK7	UPI0001C5D06D	3	113234
64	F8	scaffold_120:1033757-1033757	ENSLAFG00000015274	missense_variant	MODERATE	Transcript	60	R/L	cGc/cTc	G3TEU8	UPI0001C5A7EC	2	139793
65	F8	scaffold_120:1058023-1058023	ENSLAFG00000015274	synonymous_variant	LOW	Transcript	305	S	tcT/tcG	G3TEU8	UPI0001C5A7EC	3	139793
66	F8	scaffold_120:1105775-1105775	ENSLAFG00000015274	missense_variant	MODERATE	Transcript	961	L/I	Tta/Ata	G3TEU8	UPI0001C5A7EC	6	139793
67	F8	scaffold_120:1106011-1106011	ENSLAFG00000015274	missense_variant	MODERATE	Transcript	1039	D/E	gaT/gaG	G3TEU8	UPI0001C5A7EC	6	139793
68	FANCI	scaffold_28:17697830-17697830	ENSLAFG00000016287	splice_region_variant, intron_variant	LOW	Transcript	-	-	-	G3TGQ9	UPI0001C5CD39	4	95707
69	FANCI	scaffold_28:17698249-17698249	ENSLAFG00000016287	intron_variant	MODIFIER	Transcript	-	-	-	G3TGQ9	UPI0001C5CD39	2	95707
70	FANCM	scaffold_9:20799327-20799327	ENSLAFG00000001136	synonymous_variant	LOW	Transcript	910	A	gcC/gcG	G3SN12	UPI0001C5B14D	8	67397
71	FANCM	scaffold_9:20806441-20806441	ENSLAFG00000001136	missense_variant	MODERATE	Transcript	1400	L/V	Tta/Gta	G3SN12	UPI0001C5B14D	4	67397
72	FANCM	scaffold_9:20806518-20806518	ENSLAFG00000001136	missense_variant	MODERATE	Transcript	1425	K/N	aaA/aaT	G3SN12	UPI0001C5B14D	3	67397
73	FANCM	scaffold_9:20829255-20829255	ENSLAFG00000001136	missense_variant	MODERATE	Transcript	1782	C/S	tGt/tCt	G3SN12	UPI0001C5B14D	3	67397
74	FGF5	scaffold_30:16583108-16583108	ENSLAFG00000016173	intron_variant	MODIFIER	Transcript	-	-	-	G3TGG7	UPI0001C5EDB5	2	25853
75	GHR	scaffold_7:47730850-47730850	ENSLAFG00000025524	synonymous_variant	LOW	Transcript	107	P	ccT/ccA	G3UMD7	UPI0001C5C8EA	5	174786
76	GHR	scaffold_7:47751697-47751697	ENSLAFG00000025524	synonymous_variant	LOW	Transcript	555	A	gcG/gcT	G3UMD7	UPI0001C5C8EA	4	174786
77	GP5	scaffold_25:14960380-14960380	ENSLAFG00000002812	missense_variant	MODERATE	Transcript	542	L/I	Ctc/Atc	G3SR71	UPI0001C5BED6	4	3649
78	HOXA2	scaffold_5:31368220-31368220	ENSLAFG00000001565	missense_variant	MODERATE	Transcript	176	G/A	gGg/gCg	G3SNS8	UPI0000E32225	2	2433
79	IFNG	scaffold_2:58253558-58253558	ENSLAFG00000004484	missense_variant	MODERATE	Transcript	109	S/A	Tcc/Gcc	G3SUC6	UPI0001C5E889	2	6995
80	IFNG	scaffold_2:58253691-58253691	ENSLAFG00000004484	missense_variant	MODERATE	Transcript	64	D/E	gaC/gaA	G3SUC6	UPI0001C5E889	3	6995
81	IFNG	scaffold_2:58253839-58253839	ENSLAFG00000004484	synonymous_variant	LOW	Transcript	46	A	gcG/gcC	G3SUC6	UPI0001C5E889	3	6995
82	IGF1R	scaffold_28:31235087-31235087	ENSLAFG00000011658	synonymous_variant	LOW	Transcript	432	I	atA/atC	G3UCL7	UPI0001C5CF66	3	349784
83	IGSF1	scaffold_100:978106-978106	ENSLAFG00000011426	synonymous_variant	LOW	Transcript	1209	I	atT/atA	G3T7G2	UPI0001C5B0B5	4	12333
84	IGSF1	scaffold_100:986469-986469	ENSLAFG00000011426	missense_variant	MODERATE	Transcript	301	L/V	Ctc/Gtc	G3T7G2	UPI0001C5B0B5	4	12333
85	IGSF10	scaffold_41:8433532-8433532	ENSLAFG00000000038	missense_variant	MODERATE	Transcript	177	R/S	Cgc/Agc	G3SKZ8	UPI0001C5AEDB	2	23666
86	IGSF10	scaffold_41:8438232-8438232	ENSLAFG00000000038	synonymous_variant	LOW	Transcript	286	S	tcA/tcT	G3SKZ8	UPI0001C5AEDB	3	23666
87	IGSF10	scaffold_41:8439597-8439597	ENSLAFG00000000038	missense_variant	MODERATE	Transcript	729	S/R	agC/agG	G3SKZ8	UPI0001C5AEDB	2	23666

	SYMBOL	Location	Gene	Consequence	IMPACT	Feature type	Protein position	Amino acids	Codons	Uniprot	UNIPARC	n. ind	Gene length
88	IGSF10	scaffold_41:8440774-8440774	ENSLAFG00000000038	missense_variant	MODERATE	Transcript	1122	F/V	Ttc/Gtc	G3SKZ8	UPI0001C5AEDB	5	23666
89	IGSF10	scaffold_41:8441082-8441082	ENSLAFG00000000038	synonymous_variant	LOW	Transcript	1224	T	acA/acC	G3SKZ8	UPI0001C5AEDB	3	23666
90	IGSF10	scaffold_41:8441090-8441091	ENSLAFG00000000038	frameshift_variant	HIGH	Transcript	1227	P/X	ccA/cc	G3SKZ8	UPI0001C5AEDB	3	23666
91	IGSF10	scaffold_41:8441135-8441136	ENSLAFG00000000038	frameshift_variant	HIGH	Transcript	1241	N/X	Aat/at	G3SKZ8	UPI0001C5AEDB	3	23666
92	IGSF10	scaffold_41:8444034-8444034	ENSLAFG00000000038	missense_variant	MODERATE	Transcript	1893	N/H	Aac/Cac	G3SKZ8	UPI0001C5AEDB	2	23666
93	IGSF10	scaffold_41:8444054-8444054	ENSLAFG00000000038	missense_variant	MODERATE	Transcript	1899	R/S	agG/agC	G3SKZ8	UPI0001C5AEDB	3	23666
94	IGSF10	scaffold_41:8449819-8449819	ENSLAFG00000000038	synonymous_variant	LOW	Transcript	1992	S	tcG/tcC	G3SKZ8	UPI0001C5AEDB	3	23666
95	IL17B	scaffold_1:68809228-68809228	ENSLAFG00000021716	synonymous_variant	LOW	Transcript	73	L	ctG/ctC	G3TP06	UPI0001C5D721	2	6614
96	IL1B	scaffold_50:1818140-1818140	ENSLAFG00000004059	synonymous_variant	LOW	Transcript	207	R	Cga/Aga	G3STJ8	UPI0001C5D732	2	6832
97	IL20	scaffold_13:12069514-12069514	ENSLAFG00000018715	missense_variant	MODERATE	Transcript	34	V/L	Gtg/Ctg	G3TLC9	UPI0001C5BBF2	2	3667
98	IL24	scaffold_13:12120705-12120705	ENSLAFG00000013062	splice_region_variant, intron_variant	LOW	Transcript	-	-	-	G3TAL5	UPI0001C5BC3C	8	6805
99	IRAK2	scaffold_12:36765895-36765895	ENSLAFG00000002546	synonymous_variant	LOW	Transcript	602	A	gcC/gcG	G3SQP2	UPI0001C5E246	9	87751
100	ITPKA	scaffold_64:11105276-11105276	ENSLAFG00000017766	downstream_gene_variant	MODIFIER	Transcript	-	-	-	G3TJI5	UPI000195307D	2	8640
101	KDM8	scaffold_65:9508841-9508841	ENSLAFG00000018540	intron_variant	MODIFIER	Transcript	-	-	-	G3TL07	UPI0001C5FBDA	2	21325
102	KIT	scaffold_38:6340250-6340250	ENSLAFG00000022078	intron_variant	MODIFIER	Transcript	-	-	-	G5E7C2	UPI0001C5F0F0	3	103034
103	KIT	scaffold_38:6342974-6342974	ENSLAFG00000022078	synonymous_variant	LOW	Transcript	962	S	tcC/tcA	G5E7C2	UPI0001C5F0F0	5	103034
104	KRT24	scaffold_31:22629111-22629111	ENSLAFG00000000455	missense_variant	MODERATE	Transcript	267	S/T	aGt/aCt	G3SLR9	UPI0001C5E489	4	6491
105	KRT3	scaffold_2:39213772-39213772	ENSLAFG00000015682	missense_variant	MODERATE	Transcript	561	G/R	Ggc/Cgc	G3TFN2	UPI0001C5DB4B	2	5714
106	KRT35	scaffold_31:23306338-23306338	ENSLAFG00000004691	missense_variant	MODERATE	Transcript	103	V/F	Gtc/Ttc	G3SUR3	UPI0001C5E395	2	4435
107	KRT35	scaffold_31:23306345-23306345	ENSLAFG00000004691	synonymous_variant	LOW	Transcript	100	T	acA/acC	G3SUR3	UPI0001C5E395	2	4435
108	KRT4	scaffold_2:39241653-39241653	ENSLAFG00000005514	intron_variant	MODIFIER	Transcript	-	-	-	G3SWD3	UPI0001C5DC3C	2	7216
109	KRT71	scaffold_2:38928404-38928404	ENSLAFG00000017618	splice_region_variant, intron_variant	LOW	Transcript	-	-	-	G3TJ84	UPI0001C5DEC2	2	8790
110	KRT71	scaffold_2:38928468-38928468	ENSLAFG00000017618	missense_variant	MODERATE	Transcript	309	Q/H	caG/caT	G3TJ84	UPI0001C5DEC2	2	8790
111	KRT76	scaffold_2:39193923-39193923	ENSLAFG00000005508	splice_region_variant, intron_variant	LOW	Transcript	-	-	-	G3TJ84	UPI0001C5DAA4	4	8648
112	KRT76	scaffold_2:39195585-39195585	ENSLAFG00000005508	splice_region_variant, intron_variant	LOW	Transcript	-	-	-	G3SWD0	UPI0001C5DAA4	2	8648
113	LATS1	scaffold_0:8479143-8479143	ENSLAFG00000005083	intron_variant	MODIFIER	Transcript	-	-	-	G3SVI4	UPI0000E336EA	4	55134
114	LEPR	scaffold_17:28469778-28469778	ENSLAFG00000003592	synonymous_variant	LOW	Transcript	801	G	ggA/ggC	G3SSQ3	UPI0001C5DDDF	6	69544
115	LEPR	scaffold_17:28469928-28469928	ENSLAFG00000003592	synonymous_variant	LOW	Transcript	751	G	ggG/ggT	G3SSQ3	UPI0001C5DDDF	4	69544
116	LEPR	scaffold_17:28484540-28484540	ENSLAFG00000003592	splice_region_variant, intron_variant	LOW	Transcript	-	-	-	G3SSQ3	UPI0001C5DDDF	3	69544
117	LEPR	scaffold_17:28487875-28487875	ENSLAFG00000003592	missense_variant	MODERATE	Transcript	319	N/K	aaT/aaA	G3SSQ3	UPI0001C5DDDF	2	69544
118	LEPR	scaffold_17:28488140-28488140	ENSLAFG00000003592	synonymous_variant	LOW	Transcript	268	V	gtG/gtC	G3SSQ3	UPI0001C5DDDF	3	69544

	SYMBOL	Location	Gene	Consequence	IMPACT	Feature type	Protein position	Amino acids	Codons	Uniprot	UNIPARC	n. ind	Gene length
119	LEPR	scaffold_17:28489623-28489623	ENSLAFG00000003592	intron_variant	MODIFIER	Transcript	-	-	-	G3SSQ3	UPI0001C5DDDF	3	69544
120	LEPR	scaffold_17:28489630-28489630	ENSLAFG00000003592	intron_variant	MODIFIER	Transcript	-	-	-	G3SSQ3	UPI0001C5DDDF	5	69544
121	LEPR	scaffold_17:28492185-28492185	ENSLAFG00000003592	missense_variant	MODERATE	Transcript	162	D/V	gAt/gTt	G3SSQ3	UPI0001C5DDDF	3	69544
122	LEPR	scaffold_17:28494568-28494568	ENSLAFG00000003592	synonymous_variant	LOW	Transcript	112	T	acT/acG	G3SSQ3	UPI0001C5DDDF	8	69544
123	LEPR	scaffold_17:28494592-28494592	ENSLAFG00000003592	synonymous_variant	LOW	Transcript	104	S	tcC/tcA	G3SSQ3	UPI0001C5DDDF	8	69544
124	LRBA	scaffold_51:5835553-5835553	ENSLAFG00000014125	synonymous_variant	LOW	Transcript	619	T	acT/acG	G3TCN9	UPI0001C5C880	5	763450
125	LRBA	scaffold_51:5853439-5853439	ENSLAFG00000014125	missense_variant	MODERATE	Transcript	725	I/M	atC/atG	G3TCN9	UPI0001C5C880	6	763450
126	LRBA	scaffold_51:5883295-5883295	ENSLAFG00000014125	synonymous_variant	LOW	Transcript	1377	G	ggT/ggA	G3TCN9	UPI0001C5C880	5	763450
127	LRRC49	scaffold_28:2397727-2397727	ENSLAFG00000009974	missense_variant	MODERATE	Transcript	120	Q/P	cAg/cCg	G3T4Q3	UPI0001C5AAF5	3	177694
128	LRRC66	scaffold_38:3251265-3251265	ENSLAFG00000008892	missense_variant	MODERATE	Transcript	709	T/N	aCc/aAc	G3T2R1	UPI0001C5F1F0	3	31550
129	LRRC66	scaffold_38:3251342-3251342	ENSLAFG00000008892	missense_variant	MODERATE	Transcript	683	D/E	gaC/gaG	G3T2R1	UPI0001C5F1F0	2	31550
130	LRRC66	scaffold_38:3251653-3251653	ENSLAFG00000008892	missense_variant	MODERATE	Transcript	580	V/L	Gta/Tta	G3T2R1	UPI0001C5F1F0	3	31550
131	LRRC66	scaffold_38:3251665-3251665	ENSLAFG00000008892	missense_variant	MODERATE	Transcript	576	E/Q	Gag/Cag	G3T2R1	UPI0001C5F1F0	3	31550
132	LRRC66	scaffold_38:3252110-3252110	ENSLAFG00000008892	missense_variant	MODERATE	Transcript	427	D/E	gaC/gaA	G3T2R1	UPI0001C5F1F0	2	31550
133	LRRC66	scaffold_38:3254233-3254233	ENSLAFG00000008892	missense_variant	MODERATE	Transcript	271	R/P	cGc/cCc	G3T2R1	UPI0001C5F1F0	8	31550
134	LRRC66	scaffold_38:3258303-3258303	ENSLAFG00000008892	intron_variant	MODIFIER	Transcript	-	-	-	G3T2R1	UPI0001C5F1F0	3	31550
135	LRRC66	scaffold_38:3269782-3269782	ENSLAFG00000008892	synonymous_variant	LOW	Transcript	124	I	atA/atT	G3T2R1	UPI0001C5F1F0	7	31550
136	LRRC66	scaffold_38:3270064-3270064	ENSLAFG00000008892	missense_variant	MODERATE	Transcript	30	I/M	atT/atG	G3T2R1	UPI0001C5F1F0	3	31550
137	LRRC9	scaffold_9:35586494-35586494	ENSLAFG00000016138	intron_variant	MODIFIER	Transcript	-	-	-	G3TGF2	UPI0001C5C3C8	3	134814
138	LRRC9	scaffold_9:35604484-35604484	ENSLAFG00000016138	intron_variant	MODIFIER	Transcript	-	-	-	G3TGF2	UPI0001C5C3C8	3	134814
139	LRRC9	scaffold_9:35605622-35605622	ENSLAFG00000016138	missense_variant	MODERATE	Transcript	328	L/V	Ctt/Gtt	G3TGF2	UPI0001C5C3C8	8	134814
140	LRRC9	scaffold_9:35612992-35612992	ENSLAFG00000016138	missense_variant	MODERATE	Transcript	418	N/K	aaC/aaA	G3TGF2	UPI0001C5C3C8	4	134814
141	LRRC9	scaffold_9:35638373-35638373	ENSLAFG00000016138	missense_variant	MODERATE	Transcript	642	I/L	Ata/Tta	G3TGF2	UPI0001C5C3C8	12	134814
142	LRRC9	scaffold_9:35638440-35638440	ENSLAFG00000016138	intron_variant	MODIFIER	Transcript	-	-	-	G3TGF2	UPI0001C5C3C8	4	134814
143	LRRC9	scaffold_9:35648492-35648493	ENSLAFG00000016138	frameshift_variant, splice_region_variant	HIGH	Transcript	760	P/X	Ccc/cc	G3TGF2	UPI0001C5C3C8	4	134814
144	LRRD1	scaffold_5:57325540-57325540	ENSLAFG00000030816	missense_variant	MODERATE	Transcript	87	L/I	Tta/Ata	G3UN87	UPI0001C5CBF2	7	24982
145	LRRD1	scaffold_5:57325861-57325861	ENSLAFG00000030816	missense_variant	MODERATE	Transcript	194	V/L	Gta/Tta	G3UN87	UPI0001C5CBF2	7	24982
146	LRRD1	scaffold_5:57326866-57326866	ENSLAFG00000030816	missense_variant	MODERATE	Transcript	529	Q/E	Caa/Gaa	G3UN87	UPI0001C5CBF2	7	24982
147	LRRIQ4	scaffold_42:10499636-10499636	ENSLAFG00000014769	missense_variant	MODERATE	Transcript	416	Y/F	tAc/tTc	G3TDT7	UPI0001C5D090	2	23749
148	LRRIQ4	scaffold_42:10509002-10509002	ENSLAFG00000014769	missense_variant	MODERATE	Transcript	26	H/Q	caT/caA	G3TDT7	UPI0001C5D090	9	23749
149	LVRN	scaffold_1:48084969-48084969	ENSLAFG00000002068	intron_variant	MODIFIER	Transcript	-	-	-	G3SPR9	UPI0001C5CA43	4	103511

	SYMBOL	Location	Gene	Consequence	IMPACT	Feature type	Protein position	Amino acids	Codons	Uniprot	UNIPARC	n. ind	Gene length
150	LVRN	scaffold_1:48087837-48087837	ENSLAFG00000002068	synonymous_variant	LOW	Transcript	523	L	ctC/ctG	G3SPR9	UPI0001C5CA43	4	103511
151	LVRN	scaffold_1:48121103-48121103	ENSLAFG00000002068	missense_variant, splice_region_variant	MODERATE	Transcript	276	R/G	Cgt/Ggt	G3SPR9	UPI0001C5CA43	5	103511
152	LVRN	scaffold_1:48121126-48121127	ENSLAFG00000002068	splice_donor_variant, frameshift_variant	HIGH	Transcript	268	A/X	Gcc/cc	G3SPR9	UPI0001C5CA43	5	103511
153	LY86	scaffold_54:9844957-9844957	ENSLAFG00000000519	missense_variant	MODERATE	Transcript	134	L/V	Ctg/Gtg	G3SLV3	UPI0001C5C03F	2	79203
154	LY9	scaffold_33:10338073-10338073	ENSLAFG00000023009	missense_variant	MODERATE	Transcript	125	E/A	gAg/gCg	G3TMS7	UPI0001C5C9F2	3	34074
155	LY9	scaffold_33:10350906-10350906	ENSLAFG00000023009	coding_sequence_variant	MODIFIER	Transcript	-	-	-	G3TMS7	UPI0001C5C9F2	5	34074
156	MRPS22	scaffold_76:74347-74347	ENSLAFG00000010630	synonymous_variant	LOW	Transcript	344	T	acA/acT	G3T5Z5	UPI0001C5B290	6	18382
157	MRPS22	scaffold_76:89782-89782	ENSLAFG00000010630	synonymous_variant	LOW	Transcript	126	P	ccA/ccC	G3T5Z5	UPI0001C5B290	4	18382
158	MRPS22	scaffold_76:91150-91150	ENSLAFG00000010630	synonymous_variant	LOW	Transcript	79	T	acC/acA	G3T5Z5	UPI0001C5B290	2	18382
159	MRPS22	scaffold_76:91203-91203	ENSLAFG00000010630	missense_variant	MODERATE	Transcript	62	Q/K	Cag/Aag	G3T5Z5	UPI0001C5B290	2	18382
160	MTUS1	scaffold_22:38535888-38535888	ENSLAFG00000009785	synonymous_variant	LOW	Transcript	609	V	gtC/gtG	G3T4C7	UPI0001C5FAD6	7	111850
161	MTUS1	scaffold_22:38536469-38536469	ENSLAFG00000009785	missense_variant	MODERATE	Transcript	416	H/N	Cat/Aat	G3T4C7	UPI0001C5FAD6	6	111850
162	MTUS1	scaffold_22:38536601-38536601	ENSLAFG00000009785	missense_variant	MODERATE	Transcript	372	S/A	Tct/Gct	G3T4C7	UPI0001C5FAD6	2	111850
163	NF1	scaffold_31:864661-864661	ENSLAFG00000000258	intron_variant	MODIFIER	Transcript	-	-	-	G3U9S1	UPI0001C5B6AE	5	248450
164	NF1	scaffold_31:864671-864671	ENSLAFG00000000258	intron_variant	MODIFIER	Transcript	-	-	-	G3U9S1	UPI0001C5B6AE	5	248450
165	NF1	scaffold_31:864754-864754	ENSLAFG00000000258	intron_variant	MODIFIER	Transcript	-	-	-	G3U9S1	UPI0001C5B6AE	8	248450
166	PHKA2	scaffold_39:12666388-12666388	ENSLAFG00000009425	intron_variant	MODIFIER	Transcript	-	-	-	G3T3Q0	UPI0001C5D662	3	77577
167	RB1	scaffold_23:29420373-29420373	ENSLAFG00000013579	intron_variant	MODIFIER	Transcript	-	-	-	G3UL03	UPI0001C5D688	2	207581
168	RB1	scaffold_23:29465404-29465404	ENSLAFG00000013579	splice_region_variant, intron_variant	LOW	Transcript	-	-	-	G3UL03	UPI0001C5D688	2	207581
169	SAMHD1	scaffold_19:36086800-36086800	ENSLAFG00000001940	intron_variant	MODIFIER	Transcript	-	-	-	G3SPI6	UPI0001C5F0B8	2	47337
170	SLC45A2	scaffold_7:37902693-37902693	ENSLAFG00000017087	intron_variant	MODIFIER	Transcript	-	-	-	G3TI69	UPI0001C5BDA9	2	38882
171	SLC45A2	scaffold_7:37915256-37915256	ENSLAFG00000017087	missense_variant	MODERATE	Transcript	276	G/A	gGa/gCa	G3TI69	UPI0001C5BDA9	5	38882
172	TAAR1	scaffold_0:27518319-27518319	ENSLAFG00000008959	missense_variant	MODERATE	Transcript	57	T/S	Act/Tct	G3T2U3	UPI0001C5FC0B	7	1727
173	TBX15	scaffold_11:65831490-65831490	ENSLAFG00000018452	inframe_insertion	MODERATE	Transcript	546-547	-/P	-/CCT	G3TKU5	UPI0001C5FE56	2	120723
174	TLR10	scaffold_18:41571276-41571276	ENSLAFG00000008544	missense_variant	MODERATE	Transcript	800	D/A	gAc/gCc	G3TSH8	UPI0001C5F920	2	35638
175	TLR10	scaffold_18:41571826-41571826	ENSLAFG00000008544	missense_variant	MODERATE	Transcript	617	N/H	Aac/Cac	G3TSH8	UPI0001C5F920	3	35638
176	TLR10	scaffold_18:41572528-41572528	ENSLAFG00000008544	missense_variant	MODERATE	Transcript	383	Q/K	Caa/Aaa	G3TSH8	UPI0001C5F920	9	35638
177	TLR10	scaffold_18:41572565-41572565	ENSLAFG00000008544	missense_variant	MODERATE	Transcript	370	H/Q	caC/caA	G3TSH8	UPI0001C5F920	12	35638
178	TLR2	scaffold_51:2370508-2370508	ENSLAFG00000005069	missense_variant	MODERATE	Transcript	407	T/N	aCt/aAt	G3UL46	UPI0001C6019D	4	10029
179	TLR2	scaffold_51:2371020-2371020	ENSLAFG00000005069	missense_variant	MODERATE	Transcript	236	F/L	ttT/ttG	G3UL46	UPI0001C6019D	11	10029
180	TLR3	scaffold_22:42068418-42068418	ENSLAFG00000017558	missense_variant	MODERATE	Transcript	274	E/Q	Gag/Cag	G3TJ47	UPI0001C5FC38	2	18315

	SYMBOL	Location	Gene	Consequence	IMPACT	Feature type	Protein position	Amino acids	Codons	Uniprot	UNIPARC	n. ind	Gene length
181	TLR3	scaffold_22:42073534-42073534	ENSLAFG00000017558	synonymous_variant	LOW	Transcript	119	T	acC/acA	G3TJ47	UPI0001C5FC38	2	18315
182	TLR4	scaffold_6:41870027-41870027	ENSLAFG00000006775	missense_variant	MODERATE	Transcript	330	Y/F	tAc/tTc	G3SYU3	UPI0001C5F551	8	17536
183	TLR6	scaffold_18:41635735-41635735	ENSLAFG00000032091	missense_variant	MODERATE	Transcript	482	K/Q	Aaa/Caa	G3ULC1	UPI0001C5BF3A	3	36640
184	TLR7	scaffold_39:18946361-18946361	ENSLAFG00000001156	synonymous_variant	LOW	Transcript	225	S	tcT/tcG	G3ULI5	UPI0001C5D68F	3	51481
185	TLR8	scaffold_39:18879315-18879315	ENSLAFG00000011755	missense_variant	MODERATE	Transcript	30	T/S	Act/Tct	G3T843	UPI0001C5D922	3	20340
186	TRPA1	scaffold_8:18951396-18951396	ENSLAFG00000010128	missense_variant	MODERATE	Transcript	1031	R/T	aGa/aCa	G3T518	UPI0001C5B834	6	77873
187	TRPV1	scaffold_47:7229191-7229191	ENSLAFG00000013211	synonymous_variant	LOW	Transcript	585	V	gtG/gtC	G3T518	UPI0001C5C74C	3	32815
188	TRPV1	scaffold_47:7240388-7240388	ENSLAFG00000013211	missense_variant	MODERATE	Transcript	325	K/T	aAg/aCg	G3T518	UPI0001C5C74C	2	32815
189	TRPV3	scaffold_47:7190668-7190668	ENSLAFG00000013204	missense_variant	MODERATE	Transcript	333	N/K	aaC/aaA	G3TAW1	UPI0001C5C6EF	2	48154
190	TRPV3	scaffold_47:7198477-7198477	ENSLAFG00000013204	synonymous_variant	LOW	Transcript	114	R	Cgg/Agg	G3TAW1	UPI0001C5C6EF	3	48154
191	TYR	scaffold_62:12547534-12547534	ENSLAFG00000016209	missense_variant	MODERATE	Transcript	164	R/T	aGa/aCa	G3TGJ7	UPI0000E32C8D	3	103233
192	TYR	scaffold_62:12547745-12547745	ENSLAFG00000016209	synonymous_variant	LOW	Transcript	234	P	ccG/ccC	G3TGJ7	UPI0000E32C8D	2	103233
193	TYRP1	scaffold_6:86260142-86260142	ENSLAFG00000001901	missense_variant	MODERATE	Transcript	486	S/A	Tct/Gct	G3SPG1	UPI0001C5FF09	7	17047

Supplementary Table 2.5. Deleterious missense variants.

Deleterious effect of missense variants were estimated using three predictors (PredictSNP, PROVEAN, and I-Mutant2.0). Deleterious aminoacid substitutions with predictions scores are underlined in bold.

Gene	Uniprot ID	Amino acid substitution	PredictSNP		PROVEAN		I-Mutant2.0	
			Predicted Effect	Accuracy %	Predicted Effect	Cutoff=-2.5	Predicted Effect	ΔΔG
A2ML1	G3TIL7	E485D	Neutral	74	Neutral	-0.785	Increase Stability	0.18
ACKR4	G3TDP3	Q128H	Neutral	63	Neutral	-0.792	Decrease Stability	-1.70
ACKR4	G3TDP3	**R149G**	**Deleterious**	**61**	**Deleterious**	**-3.206**	Decrease Stability	-1.00
ACOX2	G3U1U7	S649A	Neutral	83	Neutral	2.483	Increase Stability	0.13
APOBEC-3G-like	G3TLG1	P219T	Neutral	74	Neutral	0.615	Decrease Stability	-1.43
APOBEC1	G3U0R4	H108Q	Neutral	83	Neutral	-1.601	Decrease Stability	-0.21
BPI	G3SVY2	N71K	Neutral	74	Neutral	2.476	Decrease Stability	-1.60
C5	G3TAK1	I556M	Neutral	83	Neutral	0.271	Decrease Stability	-0.75
CD109	G3SYT3	Q804E	Neutral	73	Neutral	-1.222	Increase Stability	0.18
CD163	G3SRE6	V700F	Neutral	63	Neutral	-1.017	Decrease Stability	-2.63
CD163	G3SRE6	**L1093M**	**Deleterious**	**51**	Neutral	-0.426	Decrease Stability	-0.26
CD302	G3T527	D117E	Neutral	83	Neutral	0.124	Increase Stability	0.46
CLEC4E	G3TN20	K82N	Neutral	74	Neutral	-0.799	Decrease Stability	-0.64
COL17A1	G3TZN0	**A396P**	**Deleterious**	**51**	Neutral	-1.759	Decrease Stability	-1.37
CR2	G3T3Z4	L58V	Neutral	83	Neutral	1.204	Decrease Stability	-1.18
ELP1	G3SMA4	S931T	Neutral	83	Neutral	0.293	Decrease Stability	-0.46
F5	G3SPK7	L814R	Neutral	75	Neutral	-2.267	Decrease Stability	-1.96
F8	G3TEU8	R60L	Neutral	75	Neutral	1.020	Decrease Stability	-0.70

Gene	Uniprot ID	Amino acid substitution	PredictSNP		PROVEAN		I-Mutant2.0	
			Predicted Effect	Accuracy %	Predicted Effect	Cutoff=-2.5	Predicted Effect	ΔΔG
F8	G3TEU8	L961I	Neutral	74	Neutral	-0.205	Decrease Stability	-0.26
F8	G3TEU8	D1039E	Neutral	83	Neutral	-0.779	Decrease Stability	-0.34
FANCM	G3SN12	L1400V	Neutral	83	Neutral	-0.654	Increase Stability	0.28
FANCM	G3SN12	K1425N	Neutral	83	Neutral	-0.796	Decrease Stability	-0.57
FANCM	G3SN12	C1782S	Neutral	83	Neutral	0.763	Decrease Stability	-2.58
FGA	G3U772	K253T	Neutral	83	Neutral	0.877	Decrease Stability	-0.26
GP5	G3SR71	L542I	Neutral	83	Neutral	0.570	Decrease Stability	-0.40
HOXA2	G3SNS8	G176A	Neutral	83	Neutral	3.557	Decrease Stability	-1.23
IFNG	G3SUC6	D64E	Neutral	75	Neutral	-0.901	Decrease Stability	-0.51
IFNG	G3SUC6	S109A	Neutral	74	Neutral	-1.512	Decrease Stability	-2.39
IGSF1	G3T7G2	L301V	Neutral	83	Neutral	-0.264	Decrease Stability	-0.64
IGSF10	G3SKZ8	R177S	Neutral	75	Neutral	0.941	Decrease Stability	-2.75
IGSF10	G3SKZ8	S729R	Neutral	83	Neutral	4.633	Increase Stability	0.51
IGSF10	G3SKZ8	F1122V	Neutral	75	Neutral	-1.512	Decrease Stability	-2.12
IGSF10	G3SKZ8	**N1893H**	Neutral	60	**Deleterious**	**-3.930**	Decrease Stability	-1.34
IGSF10	G3SKZ8	R1899S	Neutral	63	Neutral	1.369	Decrease Stability	-3.10
IL20	G3TLC9	V34L	Neutral	83	Neutral	-0.589	Decrease Stability	-0.50
KRT24	G3SLR9	S267T	Neutral	83	Neutral	0.363	Decrease Stability	-0.94
KRT3	G3TFN2	G561R	Neutral	83	Neutral	-2.183	Increase Stability	0.11
KRT35	G3SUR3	V103F	Neutral	75	Neutral	4.723	Decrease Stability	-1.18
KRT71	G3TJ84	Q309H	Neutral	63	**Deleterious**	**-3.216**	Decrease Stability	-0.98

Gene	Uniprot ID	Amino acid substitution	PredictSNP		PROVEAN		I-Mutant2.0	
			Predicted Effect	Accuracy %	Predicted Effect	Cutoff=-2.5	Predicted Effect	ΔΔG
LEPR	G3SSQ3	**D162V**	**Deleterious**	**61**	**Deleterious**	**-4.132**	Decrease Stability	-0.27
LEPR	G3SSQ3	N319K	Neutral	83	Neutral	-0.522	Decrease Stability	-0.49
LRBA	G3TCN9	I725M	Neutral	63	Deleterious	-2.574	Decrease Stability	-1.49
LRRC49	G3T4Q3	Q120P	Neutral	83	Neutral	-0.045	Decrease Stability	-0.45
LRRC66	G3T2R1	I30M	Neutral	60	Neutral	-0.468	Decrease Stability	-0.34
LRRC66	G3T2R1	**R271P**	**Deleterious**	**87**	Neutral	0.305	Decrease Stability	-2.26
LRRC66	G3T2R1	D427E	Neutral	83	Neutral	-0.793	Decrease Stability	-0.16
LRRC66	G3T2R1	E576Q	Neutral	83	Neutral	-0.317	Decrease Stability	-1.01
LRRC66	G3T2R1	V580L	Neutral	74	Neutral	0.850	Increase Stability	0.05
LRRC66	G3T2R1	**D683E**	**Deleterious**	**51**	Neutral	-1.367	Decrease Stability	-0.36
LRRC66	G3T2R1	T709N	Neutral	83	Neutral	1.517	Decrease Stability	-0.41
LRRC9	G3TGF2	L328V	Neutral	74	Neutral	0.203	Decrease Stability	-0.85
LRRC9	G3TGF2	N418K	Neutral	83	Neutral	-0.480	Decrease Stability	-1.35
LRRC9	G3TGF2	I642L	Neutral	83	Neutral	-0.598	Decrease Stability	-0.44
LRRD1	G3TSC9	**L87I**	**Deleterious**	**51**	Neutral	-0.124	Increase Stability	0.45
LRRD1	G3TSC9	V194L	Neutral	83	Neutral	1.158	Decrease Stability	-1.47
LRRD1	G3TSC9	Q529E	Neutral	74	Neutral	-0.863	Increase Stability	0.18
LRRIQ4	G3TDT7	H26Q	Neutral	83	Neutral	1.388	Decrease Stability	-1.75
LRRIQ4	G3TDT7	Y416F	Neutral	83	Neutral	-1.225	Decrease Stability	-0.46
LVRN	G3SPR9	R276G	Neutral	83	Neutral	1.177	Decrease Stability	-2.37
LY86	G3SLV3	**L134V**	**Deleterious**	**55**	Neutral	-0.054	Decrease Stability	-1.71

Gene	Uniprot ID	Amino acid substitution	PredictSNP		PROVEAN		I-Mutant2.0	
			Predicted Effect	Accuracy %	Predicted Effect	Cutoff=-2.5	Predicted Effect	ΔΔG
LY9	G3TMS7	E125A	Neutral	74	Neutral	2.444	Decrease Stability	-0.57
MNDA	G3SR44	D16V	Neutral	83	Neutral	1.337	Decrease Stability	-1.28
MNDA	G3SR44	G78V	Neutral	63	Neutral	-1.390	Decrease Stability	-0.65
MNDA	G3SR44	V111L	Neutral	74	Neutral	-0.162	Decrease Stability	-1.02
MNDA	G3SR44	**S179Y**	**Deleterious**	**51**	**Deleterious**	**-2.835**	Decrease Stability	-0.12
MNDA	G3SR44	V204F	Neutral	83	Neutral	-2.231	Decrease Stability	-1.77
MNDA	G3SR44	E210V	Neutral	75	Neutral	2.355	Increase Stability	0.73
MNDA	G3SR44	**T227K**	**Deleterious**	**51**	Neutral	1.047	Increase Stability	0.42
MNDA	G3SR44	**T228K**	**Deleterious**	**87**	**Deleterious**	**-4.626**	Increase Stability	0.35
MRPS22	G3T5Z5	Q62K	Neutral	83	Neutral	0.563	Decrease Stability	-0.59
MTUS1	G3T4C7	S372A	Neutral	74	Neutral	-0.750	Increase Stability	0.40
MTUS1	G3T4C7	H416N	Neutral	75	Neutral	1.210	Decrease Stability	-2.57
PMEL	G3TDI1	E324V	Neutral	83	Neutral	1.284	Increase Stability	1.12
SLC45A2	G3TI69	G276A	Neutral	83	Neutral	-0.970	Decrease Stability	-0.73
TAAR1	G3T2U3	**T57S**	Neutral	63	**Deleterious**	**-3.166**	Decrease Stability	-0.69
TAAR3	G3U118	**M176R**	Neutral	60	**Deleterious**	**-3.006**	Decrease Stability	-1.69
TLR10	G3TSH8	H370Q	Neutral	83	Neutral	1.184	Decrease Stability	-0.27
TLR10	G3TSH8	Q383K	Neutral	83	Neutral	0.942	Decrease Stability	-0.33
TLR10	G3TSH8	N617H	Neutral	83	Neutral	3.308	Increase Stability	0.05
TLR10	G3TSH8	D800A	Neutral	74	Neutral	0.266	Decrease Stability	-0.98
TLR11	G3U832	N153K	Neutral	83	Neutral	-0.592	Decrease Stability	-1.88

Gene	Uniprot ID	Amino acid substitution	PredictSNP		PROVEAN		I-Mutant2.0	
			Predicted Effect	Accuracy %	Predicted Effect	Cutoff=-2.5	Predicted Effect	ΔΔG
TLR11	G3U832	S208R	Neutral	74	Neutral	-0.310	Decrease Stability	-0.70
TLR13	G3UEZ6	**L412R**	**Deleterious**	**61**	**Deleterious**	**-3.961**	Decrease Stability	-1.21
TLR2	G3UL46	F236L	Neutral	83	Neutral	-1.452	Decrease Stability	-1.87
TLR2	G3UL46	T407N	Neutral	74	Neutral	-1.622	Decrease Stability	-0.58
TLR3	G3TJ47	E274Q	Neutral	83	Neutral	0.266	Decrease Stability	-0.06
TLR4	G3SYU3	Y330F	Neutral	83	Neutral	0.286	Increase Stability	0.73
TLR6	G3ULC1	K482Q	Neutral	83	Neutral	1.189	Decrease Stability	-0.04
TLR8	G3T843	T30S	Neutral	83	Neutral	0.388	Increase Stability	0.19
TRPA1	G3T518	R1031T	Neutral	83	Neutral	0.522	Decrease Stability	-1.64
TRPV1	G3TAW7	K325T	Neutral	83	Neutral	1.585	Increase Stability	0.59
TRPV3	G3TAW1	N333K	Neutral	83	Neutral	1.753	Decrease Stability	-0.95
TYR	G3TGJ7	R164T	Neutral	83	Neutral	0.591	Decrease Stability	-1.47
TYRP1	G3SPG1	S486A	Neutral	83	Neutral	1.210	Increase Stability	0.58

Supplementary Table 2.6. Samples.

This table contains detailed information on each mammoth sample, including geographic distribution, calibrated radiocarbon ages (14C date/(yBP)), and mitochondrial haplotypes.

Taxon-MADC#	Taxon	Area	Locality	Lat.	Long.	14cdate /(yBP)	Mitochondrial Haplotype	Study
URL1	Mammuthus sp	Ural mountains	nd	nd	nd	nd	nd	(Debruyne et al., 2008)
URL2	Mammuthus sp	Ural mountains	nd	nd	nd	24430	Clade I (E3)	
GDY1	Mammuthus sp	Western Siberia	Gydan peninsula	nd	nd	15160	Clade I (E18)	
2000-173	M. primigenius	Taimyr Peninsula	Arilakh	73.75	102	11900	Clade I (E11)	
2000-174	M. primigenius	Taimyr Peninsula	Arilakh	73.75	102	28210	Clade I (E9)	
2002-473	M. primigenius	Taimyr Peninsula	Arilakh	73.75	102	46700	Clade I (D4)	
Ber5	M. primigenius	Eastern Siberia	Berelekh	70.4	143.95	nd	Clade I (D18)	
Ber7	M. primigenius	Eastern Siberia	Berelekh	70.4	143.95	nd	Clade I (D1)	
Ber9	M. primigenius	Eastern Siberia	Berelekh	70.4	143.95	nd	Clade I (D14)	
Ber10	M. primigenius	Eastern Siberia	Berelekh	70.4	143.95	nd	Clade I (D17)	
Ber11	M. primigenius	Eastern Siberia	Berelekh	70.4	143.95	nd	Clade I (E6)	
Ber20	M. primigenius	Eastern Siberia	Berelekh	70.4	143.95	nd	Clade I (E6)	
SYU3	Mammuthus sp	Eastern Siberia	Sanga-Yuriakh	63.5	142.75	>47800	Clade I (E7)	
WR2	M. primigenius	Wrangel Island	nd	71.24	-179.78	4420	Clade I (E5)	
IK-99-70	Mammuthus sp	Alaskan North Slope	Upper Ikpikpuk River	69.37	-154.67	>51900	Clade I (C1)	
IK-99-322	Mammuthus sp	Alaskan North Slope	Upper Ikpikpuk River	69.37	-154.67	>52000	Clade I (C13)	
IK-01-359	Mammuthus sp	Alaskan North Slope	Upper Ikpikpuk River	69.37	-154.67	>54000	Clade I (C12)	
T-02-110	Mammuthus sp	Alaskan North Slope	Upper Ikpikpuk River	69.37	-154.67	26410	Clade I (D19)	
AK-323-V-I	M. primigenius	Alaskan North Slope	Upper Ikpikpuk River	69.37	-154.67	31100	Clade I (C19)	
8572	M. primigenius	Alaska	South East AK	nd	nd	18090	Clade I (E14)	
6746	M. primigenius	Alaska	Tanana	61.858	-157.788	23150	Clade I (C4)	
AM1187	M. primigenius	Alaska	Inglutalik Cr	65	165	>41081	Clade I (C20)	
AM1189	M. primigenius	Alaska	Inglutalik Cr	65	165	31360	Clade I (C18)	
AM1193	M. primigenius	Alaska	Inglutalik Cr	65	165	16319	Clade I (C6)	
AM1208	M. primigenius	Alaska	Sullivan Creek	65.10	151	12677	Clade I (C7)	
AM2446	M. primigenius	Alaska	Cripple Creek	64.60	148	26022	Clade I (C2)	
11340	M. primigenius	Alaska	Cripple Hill	61.858	-157.788	37800	Clade I (C22)	
AM523	M. primigenius	Alaska	Cleary Creek	64.83	-148	43239	Clade I (C28)	
AM8052	M. primigenius	Alaska	Cleary Creek	64.83	-148	18379	Clade I (C1)	
42135	M. primigenius	Alaska	Eldorado Cr	nd	nd	30000	Clade I (C8)	
780001	M. primigenius	Yukon	Last Chance Creek	nd	nd	>48800	Clade I (C1)	
50069	M. primigenius	Yukon	Hunker Creek	63.82	-139.03	>47500	Clade I (C2)	
290248	M. primigenius	Yukon	Hunker Creek	63.82	-139.03	43500	Clade I (C16)	
46308	M. primigenius	Yukon	Hunker Creek	63.82	-139.03	35800	Clade I (C16)	
20007	M. primigenius	Yukon	Hunker Creek	63.82	-139.03	27540	Clade I (C3)	
1330021	Mammuthus sp	Yukon	Whitman Gulch	63.43	138.38	34180	Clade I (C23)	
1360005	M. primigenius	Yukon	Sulphur Creek	63.44	138.50	nd	Clade I (C5)	
30019	M. primigenius	Yukon	Finning	63.50	138.15	44700	Clade I (C25)	
30133	M. primigenius	Yukon	Finning	63.50	138.15	29030	Clade I (C21)	
30134	M. primigenius	Yukon	Finning	63.50	138.15	nd	Clade I (C18)	
30136	M. primigenius	Yukon	Finning	63.50	138.15	29170	Clade I (C21)	
30229	M. primigenius	Yukon	Finning	63.50	138.15	nd	Clade I (C27)	
2190003	M. primigenius	Yukon	Finning	63.30	137.15	31740	Clade I (C29)	
43124	M. primigenius	Yukon	Dawson area	64.05	-139.42	36600	Clade I (C17)	

Taxon-MADC#	Taxon	Area	Locality	Lat.	Long.	14cdate /(yBP)	Mitochondrial Haplotype	Study
49927	M. primigenius	Yukon	Dawson area	64.05	-139.42	46600	Clade I (C10)	
8139	M. primigenius	British Columbia	Dominion Cr	50.68	-120.34	39500	Clade I (C26)	
11708	M. primigenius	British Columbia	Dominion Cr	50.68	-120.34	31600	Clade I (C23)	
49562	M. primigenius	NW Territories	Parson`s Lake	69.00	133.30	>47200	Clade I (C2)	
2005-915	M. primigenius	Taimyr Peninsula	Baikura-Turku	73,75	102	27740	Clade I (E1)	(Enk et al., 2016)
2002-472	M. primigenius	Taimyr Peninsula	Arilakh	74,42	107,75	>48800	Clade I (D6)	
Ber28	M. primigenius	Eastern Siberia	Berelekh	70,4	143,95	12125	Clade I (E12)	
2006-001	Mammuthus sp	Yakutia	Oymyakon	63,5	142,75	41300	Clade II (A1)	
IK-99-5001	Mammuthus sp	Alaskan North Slope	Upper Ikpikpuk River	69,37	-154,67	33530	Clade I (E17)	
AM104	M. primigenius	Alaska	Cleary Creek	65,17	-147,5	42764	Clade I (C28)	
AM8744	M. primigenius	Alaska	Ester Creek	64,83	-148	16789	Clade I (D19)	
173001	M. primigenius	Yukon	Ch'ijees's Bluff	67,48	-139,92	>45400	Clade I (C2)	
1300002	Mammuthus sp	Yukon	Hunker Creek	63,82	-139,03	36690	Clade I (D1)	
42292	M. primigenius	Yukon	Dawson area	64,05	-139,42	37920	Clade III (B2)	
49929	M. primigenius	Yukon	Dawson area	64,05	-139,42	38600	C1	
CMNH40031	Mammuthus sp	YU	Old Crow	68,06	-139,78	nd	nd	
DMNS23	Mammuthus sp	CO	nd	39,07	-105,13	14661	nd	
DMNS28b	M. columbi	NE	La Sena	40,38	-100,23	18440	nd	
DMNS47	M. columbi	CO	Dent	40,3	-104,8	10990	nd	
DMNS49	M. columbi	CO	Dent	40,3	-104,8	10990	nd	
ISM01	M. primigenius	IL	Near Pekin	40,52	-89,72	17510	nd	
ISM04	M. primigenius	IL	Gravel Pit near Clear	39,82	-89,53	20550	nd	
ISM07	M. jeffersonii	IL	Wyanet	41,37	-89,65	15947	nd	
ISM12	Mammuthus sp	IL	North LaSalle Country	41,55	-88,87	12495	nd	
ISM15	M. sp (intermediate)	SD	Near Brookings	44,46	-96,88	12490	nd	
UNSM01	Mammuthus sp	NE	Red Willow Fauna	40,22	-100,37	17070	nd	
UNSM02	Mammuthus sp	NE	Red Willow Fauna	40,22	-100,37	12130	nd	
UNSM08	M. columbi	NE	Red Willow Fauna	40,22	-100,37	16160	nd	
UNSM09	Mammuthus sp	NE	Little Sand Pit	40,2	-100,5	11585	nd	
UNSM14	Mammuthus sp	NE	Trento Reservoir	40,17	-101,07	33670	nd	
UNSM15	M. columbi	NE	Richardson Co	40,12	-95,87	nd	nd	
UNSM16	M. sp (intermediate)	NE	Crappie Hole	41,2	-101,75	23590	nd	
UNSM21	M. columbi	NE	South Fork Big Nemah	40,07	-95,82	13850	nd	
UNSM22	Mammuthus sp	NE	Palisade Sand Pit	40,35	-101,42	nd	nd	
UNSM07	Mammuthus sp	NE	Red Willow Fauna	40,22	-100,37	nd	nd	
UNSM23	Mammuthus sp	NE	Palisade Sand Pit	40,35	-101,42	nd	nd	
UNSM24	Mammuthus sp	NE	Palisade Sand Pit	40,35	-101,42	nd	nd	
UNSM27	M. primigenius	KY	Big Bone Lick	38,88	-84,75	13985	nd	
UNSM29	M. primigenius	KY	Big Bone Lick	38,88	-84,75	12930	nd	
UNSM30	M. primigenius	KY	Big Bone Lick	38,88	-84,75	13215	nd	
UNSM32	M. primigenius	KY	Big Bone Lick	38,88	-84,75	13860	nd	
UNSM33	M. sp (intermediate)	NE	Crappie Hole	41,2	-101,75	nd	nd	
UNSM34	M. columbi	NE	Crappie Hole	41,2	-101,75	23670	nd	
WAST_01	Mammuthus sp	WA	Wenas Creek	46,7	-120,55	13398	nd	
DMNS08	Mammuthus sp	CO	Badger Creek	40,29	-106,45	nd	nd	
UCMP09	Mammuthus sp	WA	Whidbey Island	48,12	-122,58	19200	nd	

Supplementary Table 2.7. Targets.

List of target genes analyzed in this study with corresponding proteins names and biological role.

Gene	Protein	Group
APOBEC1	Apolipoprotein B mRNA editing enzyme catalytic subunit 1	Antiviral
APOBEC2	Apolipoprotein B mRNA editing enzyme catalytic subunit 2	Antiviral
APOBEC3A-like	DNA dC->dU-editing enzyme APOBEC-3G-like	Antiviral
APOBEC3B	DNA dC->dU-editing enzyme APOBEC-3G-like	Antiviral
APOBEC3C	DNA dC->dU-editing enzyme APOBEC-3G-like	Antiviral
Apobec3G-Like	CMP/dCMP-type deaminase domain-containing protein	Antiviral
APOBEC4	Apolipoprotein B mRNA editing enzyme catalytic polypeptide like 4	Antiviral
BST2	Bone marrow stromal cell antigen 2	Antiviral
CYPa	Peptidylprolyl Isomerase A	Antiviral
SAMHD1	SAM and HD domain containing deoxynucleoside triphosphate triphosphohydrolase 1	Antiviral
Trim5	Tripartite motif-containing protein 5-like	Antiviral
ACOX2	Acyl-CoA oxidase 2, branched chain	Arctic Adaptation
ADGRG6	Adhesion G protein-coupled receptor G	Arctic Adaptation
ASIP	Agouti signaling protein	Arctic Adaptation
BMAL1	Aryl hydrocarbon receptor nuclear translocator like	Arctic Adaptation
CART1	ALX homeobox 1	Arctic Adaptation
Edn3	Endothelin 3	Arctic Adaptation
FGF5	Fibroblast growth factor 5	Arctic Adaptation
GHR	Growth hormone receptor	Arctic Adaptation
HMGA2	Histamine receptor H1	Arctic Adaptation
HOXA2	Homeobox A2	Arctic Adaptation
HR1	Histamine H1 receptor	Arctic Adaptation
HRH3	Histamine H3 receptor-like	Arctic Adaptation
IGF1	Insulin like growth factor 1	Arctic Adaptation
KIT	KIT proto-oncogene receptor tyrosine kinase	Arctic Adaptation
KRT23	Keratin 23	Arctic Adaptation
KRT24	Keratin 24	Arctic Adaptation
KRT3	Keratin 3	Arctic Adaptation
KRT35	Keratin 35	Arctic Adaptation
KRT4	Keratin 4	Arctic Adaptation
KRT40	Keratin 40	Arctic Adaptation
KRT71	Keratin 71	Arctic Adaptation
KRT73	Keratin 73	Arctic Adaptation
KRT76	Keratin 76	Arctic Adaptation
KRT78	Keratin 78	Arctic Adaptation
KRTAP17-1	Keratin-associated protein 17-1	Arctic Adaptation
KRTAP8-1	Keratin associated protein 8-1	Arctic Adaptation
LEPR	Leptin receptor	Arctic Adaptation

Gene	Protein	Group
LPAR6	Lysophosphatidic Acid Receptor	Arctic Adaptation
LRBA	LPS Responsive Beige-Like Anchor Protein	Arctic Adaptation
LRPPRC	Leucine Rich Pentatricopeptide Repeat Containing	Arctic Adaptation
LRTOMT	Leucine rich transmembrane and O-methyltransferase domain containing	Arctic Adaptation
MCR1	Melanocortin-1 receptor gene	Arctic Adaptation
MPRS22	Mitochondrial ribosomal protein S22	Arctic Adaptation
PER2	Period circadian regulator 2	Arctic Adaptation
PMEL17	Premelanosome protein	Arctic Adaptation
SLC18A2	Solute carrier family 18 member A2	Arctic Adaptation
SLC36A1	Solute carrier family 36 member 1	Arctic Adaptation
SLC45A2	Solute carrier family 45 member 2	Arctic Adaptation
SMAD2	SMAD family member 2	Arctic Adaptation
STC2	Stanniocalcin 2	Arctic Adaptation
TRPA1	Transient receptor potential cation channel subfamily A member 1	Arctic Adaptation
TRPM4	Transient receptor potential cation channel subfamily M member 4	Arctic Adaptation
TRPM8	Transient receptor potential cation channel subfamily A member 8	Arctic Adaptation
TRPV1	Transient receptor potential cation channel subfamily V member 1	Arctic Adaptation
TRPV3	Transient receptor potential cation channel subfamily V member 3	Arctic Adaptation
TRPV4	Transient receptor potential cation channel subfamily V member 4	Arctic Adaptation
TYR	Tyrosinase	Arctic Adaptation
TYRP1	Tyrosinase related protein 1	Arctic Adaptation
UCP1	Uncoupling protein 1	Arctic Adaptation
DLK1	Delta like non-canonical Notch ligand 1	Cellular Process
TP53	Tumor Protein P53	Cellular Process
TP53RTG1	Tumor Protein P53 retrogene 1	Cellular Process
TP53RTG2	Tumor Protein P53 retrogene 2	Cellular Process
TP53RTG3	Tumor Protein P53 retrogene 3	Cellular Process
TP53RTG4	Tumor Protein P53 retrogene 4	Cellular Process
TP53RTG5	Tumor Protein P53 retrogene 5	Cellular Process
TP53RTG6	Tumor Protein P53 retrogene 6	Cellular Process
TP53RTG16	Tumor Protein P53 retrogene 16	Cellular Process
TP53RTG17	Tumor Protein P53 retrogene 17	Cellular Process
TP53RTG18	Tumor Protein P53 retrogene 18	Cellular Process
TP53RTG19	Tumor Protein P53 retrogene 19	Cellular Process
FANCI	Fanconi anemia complementation group I	Cellular Process
FANCM	Fanconi anemia complementation group M	Cellular Process
IGF1R	Insulin like growth factor 1 receptor	Cellular Process
IGSF1	Immunoglobulin Superfamily Member 1	Cellular Process
IGSF10	Immunoglobulin Superfamily Member 10	Cellular Process
IGSF3	Immunoglobulin Superfamily Member 3	Cellular Process
ISM1	Isthmin 1	Cellular Process

Gene	Protein	Group
KIF26A	Kinesin family member 26A	Cellular Process
KIF9	Kinesin family member 9	Cellular Process
LATS1	Large tumor suppressor kinase 1	Cellular Process
LGALS12	Galectin 12	Cellular Process
LRRC49	Leucine rich repeat containing 49	Cellular Process
LRRC66	Leucine rich repeat containing 66	Cellular Process
LRRC9	Leucine rich repeat containing 9	Cellular Process
LRRD1	Leucine rich repeats and death domain containing 1	Cellular Process
LRRIQ1	Leucine rich repeats and IQ motif containing 1	Cellular Process
LRRIQ4	Leucine rich repeats and IQ motif containing 4	Cellular Process
LTK	Leukocyte receptor tyrosine kinase	Cellular Process
LVRN	Laeverin	Cellular Process
MAGEH1	MAGE family member H1	Cellular Process
MAGIX	MAGI family member, X-linked	Cellular Process
MKL1	Myocardin related transcription factor A	Cellular Process
MLPH	Melanophilin	Cellular Process
MTUS1	Microtubule associated scaffold protein 1	Cellular Process
RB1	RB transcriptional corepressor 1	Cellular Process
TBX15	T-box transcription factor 15	Cellular Process
A2ML1	Alpha-2-Macroglobulin-Like 1	Immunity
AICDA	Activation induced cytidine deaminase	Immunity
BPI	Bactericidal Permeability Increasing Protein	Immunity
C1QL2	Complement C1q like 2	Immunity
C5	Complement C5	Immunity
C5AR1	Complement C5a receptor 1	Immunity
CCR5	C-C motif chemokine receptor 5	Immunity
CCRL1	Atypical chemokine receptor 4	Immunity
CCRL2	C-C motif chemokine receptor like 2	Immunity
CD109	CD109 molecule	Immunity
CD163	CD163 molecule	Immunity
CD19	CD19 molecule	Immunity
CD1D	CD1d molecule	Immunity
CD302	CD302 molecule	Immunity
CD38	CD38 molecule	Immunity
CD55	CD55 molecule	Immunity
CD94	Natural killer cells antigen CD94-like	Immunity
CISH	Cytokine inducible SH2 containing protein	Immunity
CLEC4E	C-type lectin domain family 4 member E	Immunity
CLEC4G	C-type lectin domain family 4 member G	Immunity
COL13A1	Collagen type XIII alpha 1 chain	Immunity
COL17A1	Collagen type XVII alpha 1 chain	Immunity

Gene	Protein	Group
COL27A1	Collagen type XXVII alpha 1 chain	Immunity
COL4A5	Collagen type IV alpha 5 chain	Immunity
COL6A1	Collagen type VI alpha 1 chain	Immunity
COL9A1	Collagen type IX alpha 1 chain	Immunity
CR2	Complement C3d receptor 2	Immunity
CRH	Corticotropin releasing hormone	Immunity
CRP	C-reactive protein	Immunity
CXCR4	C-X-C motif chemokine receptor 4	Immunity
DQA	HLA class II histocompatibility antigen DQA	Immunity
DQB	HLA class II histocompatibility antigen, DQ beta 1 chain-like	Immunity
DRA	HLA class II histocompatibility antigen, DR alpha chain-like	Immunity
DRB	HLA class II histocompatibility antigen, DQ beta 1 chain-like	Immunity
EDAR	Ectodysplasin A receptor	Immunity
EVI2B	Ecotropic viral integration site 2B	Immunity
F12	Coagulation factor XII	Immunity
F5	Coagulation factor V	Immunity
F8	Coagulation factor VIII	Immunity
FAIM3	Fc fragment of IgM receptor	Immunity
FCAMR	Fc fragment of IgA and IgM receptor	Immunity
FCGBP	Fc fragment of IgG binding protein	Immunity
FGA	Fibrinogen alpha chain	Immunity
FRAT1	WNT signaling pathway regulator	Immunity
GP2	Glycoprotein 2	Immunity
GP5	Glycoprotein V platelet	Immunity
GPNMB	Glycoprotein nmb	Immunity
HBB/HBD	Hemoglobin subunit beta/delta hybrid	Immunity
HLAIA	HLA class I histocompatibility antigen, A-11 alpha chain-like	Immunity
HLAIB	HLA class I histocompatibility antigen, A-11 alpha chain-like	Immunity
ICA1	Islet cell autoantigen 1	Immunity
ICOS	Inducible T cell costimulator	Immunity
IFITM1-Like	Interferon-induced transmembrane protein 1-like	Immunity
IFITM3-Like	Interferon-induced transmembrane protein 3-like	Immunity
IFNG	Interferon gamma	Immunity
IGSF6	Immunoglobulin Superfamily Member 6	Immunity
IKBKAP	Elongator complex protein 1	Immunity
IL10	Interleukin 10	Immunity
IL11	Interleukin 11	Immunity
IL12A	Interleukin 12A	Immunity
IL12B	Interleukin 12B	Immunity
Il13	Interleukin 13	Immunity
IL15	Interleukin 15	Immunity

Gene	Protein	Group
IL16	Interleukin 16	Immunity
IL17A	Interleukin 17A	Immunity
Il17B	Interleukin 17B	Immunity
IL17C	Interleukin 17C	Immunity
IL17F	Interleukin 17F	Immunity
IL18	Interleukin 18	Immunity
IL19	Interleukin 19	Immunity
IL1B	Interleukin 1 beta	Immunity
IL2	Interleukin 2	Immunity
IL20	Interleukin 20	Immunity
Il21	Interleukin 21	Immunity
IL22-like	Interleukin 22-like	Immunity
IL23A	Interleukin 23 subunit alpha	Immunity
IL24	Interleukin 24	Immunity
IL25	Interleukin 25	Immunity
IL26	Interleukin 26	Immunity
IL27	Interleukin 27	Immunity
IL31	Interleukin 31	Immunity
IL33	Interleukin 33	Immunity
IL34	Interleukin 34	Immunity
IL36A	Interleukin 36 alpha	Immunity
IL4	Interleukin 4	Immunity
IL5	Interleukin 5	Immunity
IL6	Interleukin 6	Immunity
IL7	Interleukin 7	Immunity
IL8	C-X-C motif chemokine ligand 8	Immunity
IL9	Interleukin 9	Immunity
IRAK2	Interleukin 1 receptor associated kinase 2	Immunity
IRF3	Interferon regulatory factor 3	Immunity
IRF7	Interferon regulatory factor 7	Immunity
IRF9	Interferon regulatory factor 9	Immunity
IRGM	Immunity-related GTPase family M protein-like	Immunity
KIR3DL2	Killer cell immunoglobulin-like receptor-like protein KIR3DX1	Immunity
KIR3DL3	Killer cell immunoglobulin-like receptor 3DS1	Immunity
KIR3DX1	Killer cell immunoglobulin-like receptor like protein KIR3DP1	Immunity
LGALS3	Galectin 3	Immunity
LGALSL	Galectin Like	Immunity
LIR4A	Leukocyte immunoglobulin-like receptor subfamily A member 4	Immunity
LIRA3	Leukocyte immunoglobulin-like receptor subfamily A member 3	Immunity
LIRA6	Leukocyte immunoglobulin-like receptor subfamily A member 6	Immunity
LIRB2	Leukocyte immunoglobulin-like receptor subfamily B member 3	Immunity

Gene	Protein	Group
LIRB3	Leukocyte immunoglobulin-like receptor?	Immunity
LIRB4	Leukocyte immunoglobulin-like receptor subfamily B member 3	Immunity
LRRC33	Negative regulator of reactive oxygen species	Immunity
LY86	Lymphocyte antigen 86	Immunity
LY9	Lymphocyte antigen 9	Immunity
LY96	Lymphocyte antigen 96	Immunity
MAGB6	MAGE Family Member B6	Immunity
MNDA	Myeloid cell nuclear differentiation antigen	Immunity
MPL	MPL proto-oncogene, thrombopoietin receptor	Immunity
NKG2A/B	NKG2-A/NKG2-B type II integral membrane protein-like	Immunity
NKG2D	NKG2D Natural killer cells antigen receptor	Immunity
NKG2F	NKG2-F type II integral membrane protein-like	Immunity
TAAR1	Trace amine associated receptor 1	Immunity
TAAR2	Trace amine associated receptor 2	Immunity
TAAR3	Trace amine associated receptor 3	Immunity
TAAR4-like	Trace amine-associated receptor 4-like	Immunity
TAAR5	Trace amine associated receptor 5	Immunity
TAAR6-like	Trace amine-associated receptor 6-like	Immunity
TAAR8	Trace amine associated receptor 8	Immunity
TAAR9-like	Trace amine-associated receptor 6-like	Immunity
TLR1	Toll like receptor 1	Immunity
TLR10	Toll like receptor 10	Immunity
TLR11	Toll like receptor 11	Immunity
TLR12	Toll like receptor 12	Immunity
TLR13	Toll like receptor 13	Immunity
TLR2	Toll like receptor 2	Immunity
TLR3	Toll like receptor 3	Immunity
TLR4	Toll like receptor 4	Immunity
TLR5	Toll like receptor 5	Immunity
TLR6	Toll like receptor 6	Immunity
TLR7	Toll like receptor 7	Immunity
TLR8	Toll like receptor 8	Immunity
TLR9	Toll like receptor 9	Immunity
ACOX2-NM	Acyl-CoA oxidase 2, branched chain	Nuclear Marker
ACP4	Acid phosphatase 4	Nuclear Marker
CARHSP1	Calcium-regulated heat stable protein 1	Nuclear Marker
COPSA7A	COP9 signalosome complex subunit 7a	Nuclear Marker
DHRS3	Short-chain dehydrogenase/reductase 3	Nuclear Marker
JMJD5	Jumonji domain containing protein 5	Nuclear Marker
LANCL1	LanC-like protein 1	Nuclear Marker
ROGDI	Rogdi atypical leucine zipper	Nuclear Marker

Gene	Protein	Group
SLC38A7	Solute carrier family 38 member 7	Nuclear Marker
SMYD4	SET and MYND domain-containing protein 4	Nuclear Marker
AmelX	Amelogenin, X-linked	Sexual marker
AMELY	Amelogenin, Y-linked (ChrY)	Sexual marker
BGN	Biglycan (ChrX)	Sexual marker
PHKA2	Phosphorylase kinase alpha subunit (ChrX)	Sexual marker
PLP	Proteolipid protein 1 (ChrX)	Sexual marker

Supplementary Table 2.8. Proboscidean genomes.

Genomes of extant and extinct proboscidean species and geographic origin for species included in this analysis. Sequencing raw data are available for download at the European Nucleotide Archive (ENA).

Specimen ID	Name	Specie	Date, years ago	Geographic origin	ENA Accession	Study
L. africana_B	Savanna elephant	*Loxodontha africana*	Modern	Kenya	ERX2312222	(Palkopoulou et al., 2015)
L. africana_C	Savanna elephant	*Loxodontha africana*	Modern	South Africa	ERX2312223	(Palkopoulou et al., 2018)
L.cyclotis_A	Forest elephant	*Loxodontha cyclotis*	Modern	Central African Republic	ERX2312221	(Palkopoulou et al., 2018)
L. cyclotis_F	Forest elephant	*Loxodontha cyclotis*	Modern	Sierra Leone	ERX2312226	(Palkopoulou et al., 2018)
E. maximus_D	Asian elephant	*Elephas maximus*	Modern	Myanmar	ERX2312224	(Palkopoulou et al., 2018)
E. maximus_E	Asian elephant	*Elephas maximus*	Modern	Malaysia (Borneo)	ERX2312225	(Reddy et al., 2015)
E.maximus_Z (Jayaprakash)	Asian elephant	*Elephas maximus*	Modern	Karnataka, India	SRR2912975	(V. J. Lynch et al., 2015)
E. maximus_L (Parvarty)	Asian elephant	*Elephas maximus*	Modern	India	SRX1015604	(V. J. Lynch et al., 2015)
E. maximus_M (Asha)	Asian elephant	*Elephas maximus*	Modern	India	SRX1015603	(V. J. Lynch et al., 2015)
E maximus_Y	Asian elephant	*Elephas maximus*	Modern	Assam, India	SRX1015606	(Dastjerdi et al., 2014)
E. maximus (Emelia)	Asian elephant	*Elephas maximus*	Modern	Captivity	ERX334764	(Dastjerdi et al., 2014)
E. maximus (Raman)	Asian elephant	*Elephas maximus*	Modern	Captivity	ERX334765	(Dastjerdi et al., 2014)
P. antiquus_N	Straight-tusked elephant	*Elephas antiquus*	~120,000	Germany	ERX2312230	(Meyer et al., 2017)
P. antiquus_O (NEU2A)	Straight-tusked elephant	*Elephas antiquus*	~120,000	Germany	ERR1753653	(Meyer et al., 2017)
M. primigenius_G	Woolly mammoth	*Mammutus primiginius*	~31,500	Taimyr Peninsula, Russia	ERX2312227	(Palkopoulou et al., 2018)
M. primigenius_H	Woolly mammoth	*Mammutus primiginius*	~44,900	Alaska, USA	ERX2312228	(Palkopoulou et al., 2018)
M. primigenius_S	Woolly mammoth	*Mammutus primiginius*	~45,300	Yamal Peninsula, Russia	ERX2312231	(Palkopoulou et al., 2018)

Specimen ID	Name	Specie	Date, years ago	Geographic origin	ENA Accession	Study
M. primigenius_V	Woolly mammoth	*Mammutus primiginius*	~42,400	Wyoming, USA	ERX2312233	(Palkopoulou et al., 2018) (Palkopoulou et al., 2018)
M. primigenius_Q	Woolly mammoth	*Mammutus primiginius*	~4,300	Wrangel Island, Russia	ERX935618	(Palkopoulou et al., 2015)
M_primigenius_P	Woolly mammoth	*Mammutus primiginius*	~44,800	Oimyakon, Russia	ERX931666	(Palkopoulou et al., 2015)
M primigenius _1 (Yuka)	Woolly mammoth	*Mammutus primiginius*	~28,140	Siberia, Russia	DRX053291	(Yamagata et al., 2019)
M. columbi_U	Columbian mammoth	*Mammutus columbi*	~13400	Wyoming, USA	ERX2312232	(Palkopoulou et al., 2018)
M. americanum_X	American mastodon	*Mammut americanum*	~13400	Gulf of Maine, USA	ERX2312234	(Palkopoulou et al., 2018)
M. americanum_I	American mastodon	*Mammut americanum*	> 50000	Alaska, USA	ERX2312229	(Palkopoulou et al., 2018)

Supplementary Table 2.9. Summary statistics of missing data.

The cut-off threshold of non-missing sites to implement the the MDS full panel and MDS subset are highlighted in color. The highlighted cells represents the percentage estimated per sample that were included in each MDS analysis.

Total sites	11724049	
Total targets length	8219510	
Min.% non-missing MDS full panel	0.01%	≥ 0.01% genome
Min.% non-missing MDS subset	1%	≥ 0.1% targets cov.

Sample	Specie	non-missing sites	% non-missing sites	Number of sites covered (≥1X)	% Genome covered	Avg. DoC In region covered	% target sites covered	Number of target sites covered
AM1208	*M_primigenius*	348	0.003	33901	0.001	1.033	0.032	2620
URL1	*Mammuthus_sp*	434	0.004	41459	0.001	1.089	0.041	3411
49562	*M_primigenius*	318	0.003	14365	0	1.03	0.062	5060
2005-915	*M_primigenius*	3258	0.028	263533	0.008	1.174	0.533	43830
2000-173	*M_primigenius*	67	0.001	5185	0	1.006	0.015	1241
AM523	*M_primigenius*	113	0.001	11147	0	1.054	0.005	390
2000-174	*M_primigenius*	2066	0.018	170746	0.005	1.155	0.249	20496
DMNS23	*Mammuthus_sp*	589	0.005	47259	0.001	1.042	0.081	6634
AM8052	*M_primigenius*	5054	0.043	430167	0.013	1.183	0.626	51449
UNSM32	*M_primigenius*	1196	0.01	66647	0.002	1.304	0.285	23458
173001	*M_primigenius*	4821	0.041	364539	0.011	1.594	0.828	68072
Ber28	*M_primigenius*	4054	0.035	325837	0.01	1.355	0.707	58090
UNSM30	*M_primigenius*	0	0	146	0	1	0.001	46
CMNH40031	*Mammuthus_sp*	74	0.001	6861	0	1.029	0.002	168

Sample	Specie	non-missing sites	% non-missing sites	Number of sites covered (≥1X)	% Genome covered	Avg. DoC In region covered	% target sites covered	Number of target sites covered
URL2	*Mammuthus_sp*	320	0.003	19268	0.001	1.032	0.072	5894
49929	*M_primigenius*	6895	0.059	647426	0.02	1.25	0.797	65536
1300002	*Mammuthus_sp*	22722	0.194	2697118	0.082	1.022	0.541	44452
IK-99-5001	*Mammuthus_sp*	348297	2.971	46565382	1.423	1.049	3.145	258508
42292	*M_primigenius*	97018	0.828	11173374	0.342	1.233	1.973	162155
DMNS28b	*M_columbi*	726	0.006	72929	0.002	1.06	0.062	5105
42135	*M_primigenius*	284	0.002	35169	0.001	4.866	0.005	415
UNSM33	*Mammuthus_sp*	8	0	275	0	1	0	0
DMNS47	*M_columbi*	782	0.007	51874	0.002	1.029	0.067	5539
UNSM3	*M_columbi*	44	0	2074	0	1.022	0.004	349
Ber5	*M_primigenius*	4765	0.041	583225	0.018	1.129	0.137	11281
AM8744	*M_primigenius*	86932	0.741	10574283	0.323	1.088	1.74	143012
Ber7	*M_primigenius*	321	0.003	34216	0.001	1.125	0.018	1488
780001	*M_primigenius*	491	0.004	54106	0.002	17.511	0.021	1757
DMNS49	*M_columbi*	56	0	5722	0	1.052	0.005	433
AM104	*M_primigenius*	5374	0.046	438313	0.013	1.537	0.845	69479
ISM01	*M_primigenius*	147	0.001	10240	0	1.125	0.028	2274
Ber9	*M_primigenius*	2910	0.025	357387	0.011	1.11	0.122	10011
ISM04	*M_primigenius*	9125	0.078	862694	0.026	1.078	0.511	41972
Ber10	*M_primigenius*	4089	0.035	458036	0.014	1.146	0.197	16155
WAST-01	*Mammuthus_sp*	330	0.003	25197	0.001	1.086	0.035	2870

Sample	Specie	non-missing sites	% non-missing sites	Number of sites covered (≥1X)	% Genome covered	Avg. DoC In region covered	% target sites covered	Number of target sites covered
GDY1	*Mammuthus_sp*	315	0.003	21813	0.001	1.013	0.038	3107
2002-472	*M_primigenius*	5343	0.046	402735	0.012	1.54	0.951	78128
UNSM24	*Mammuthus_sp*	281	0.002	30965	0.001	1.035	0.021	1685
11340	*M_primigenius*	1041	0.009	104875	0.003	1.086	0.038	3085
DMNS08	*Mammuthus_sp*	42	0	1962	0	1	0.005	382
Ber11	*M_primigenius*	39501	0.337	4498475	0.137	1.19	1.387	113971
ISM07	*M_jeffersonii*	24	0	1203	0	1.001	0.003	286
ISM12	*Mammuthus_sp*	254	0.002	19671	0.001	1.02	0.03	2498
Ber20	*M_primigenius*	6823	0.058	816867	0.025	1.05	0.246	20233
ISM15	*Mammuthus_sp*	46	0	2300	0	1.001	0.009	778
SYU3	*Mammuthus_sp*	3412	0.029	390641	0.012	1.059	0.082	6760
WR2	*M_primigenius*	16200	0.138	2149645	0.066	1.828	0.361	29662
UNSM01	*Mammuthus_sp*	1085	0.009	101284	0.003	1.024	0.068	5623
UCMP09	*Mammuthus_sp*	38	0	1760	0	1.001	0.009	734
UNSM02	*Mammuthus_sp*	133	0.001	6842	0	1.031	0.021	1765
IK-99-70	*Mammuthus_sp*	5837	0.05	555147	0.017	1.15	0.64	52640
2006-001	*Mammuthus_sp*	4626	0.039	354910	0.011	1.643	0.84	69079
30133	*M_primigenius*	198	0.002	27518	0.001	1.039	0.01	857
IK-99-322	*Mammuthus_sp*	5	0	162	0	1	0.001	79
UNSM08	*M_columbi*	4	0	496	0	1	0.001	50
IK-01-359	*Mammuthus_sp*	3784	0.032	433015	0.013	1.175	0.108	8887

Sample	Specie	non-missing sites	% non-missing sites	Number of sites covered (≥1X)	% Genome covered	Avg. DoC In region covered	% target sites covered	Number of target sites covered
UNSM09	*Mammuthus_sp*	116	0.001	9493	0	1.018	0.02	1603
30136	*M_primigenius*	482	0.004	53805	0.002	1.092	0.03	2491
T-02-110	*Mammuthus_sp*	742	0.006	96172	0.003	1.117	0.025	2037
UNSM14	*Mammuthus_sp*	21	0	994	0	1	0.003	268
UNSM07	*Mammuthus_sp*	6	0	1295	0	1.007	0.006	509
UNSM15	*M_columbi*	0	0	190	0	1	0.001	75
AK-323-V-I	*M_primigenius*	9	0	2903	0	1.06	0.018	1481
UNSM16	*Mammuthus_sp*	27	0	3415	0	1.03	0.004	310
8572	*M_primigenius*	191	0.002	19535	0.001	1.041	0.015	1254
2190003	*M_primigenius*	13	0	5571	0	1.103	0.017	1359
UNSM23	*Mammuthus_sp*	1337	0.011	95330	0.003	1.074	0.221	18178
UNSM21	*M_columbi*	186	0.002	16971	0.001	1.022	0.055	4519
6746	*M_primigenius*	7468	0.064	769304	0.024	1.051	0.37	30390
AM1187	*M_primigenius*	20	0	1446	0	1	0.003	265
AM1189	*M_primigenius*	6	0	786	0	1	0	0
8139	*M_primigenius*	37	0	2165	0	1	0.008	668
UNSM27	*M_primigenius*	48	0	4329	0	1.027	0.005	426
AM1193	*M_primigenius*	6	0	466	0	1	0.002	128
UNSM29	*M_primigenius*	32	0	2162	0	1.033	0.009	721
11708	*M_primigenius*	6027	0.051	653028	0.02	1.093	0.259	21282
E_maximus_Y	*E_maximus*	11172035	95.292					

Sample	Specie	non-missing sites	% non-missing sites	Number of sites covered (≥1X)	% Genome covered	Avg. DoC In region covered	% target sites covered	Number of target sites covered
E_maximus _M (Asha)	*E_maximus*	11677732	99.605					
E_maximus_D	*E_maximus*	5193988	44.302					
E_maximus_E	*E_maximus*	7071203	60.314					
E_maximus_Emelia	*E_maximus*	11488108	97.988					
E_maximus_Z (Jayaprakash)	*E_maximus*	11640409	99.287					
E_maximus _L (Parvarty)	*E_maximus*	11672444	99.56					
E_maximus (Raman)	*E_maximus*	10755665	91.74					
L_africana_B	*L_africana*	1471357	12.55					
L_africana_C	*L_africana*	2454624	20.937					
L_africana_A	*L_africana*	5585521	47.642					
L_cyclotis_F	*L_cyclotis_F*	2637087	22.493					
M_americanum_I	*M_americanum*	11025116	94.038					
M_americanum_X	*M_americanum*	6139442	52.366					
M_columbi_U	*M_columbi*	8992181	76.699					
M_primigenius_G	*M_primigenius*	6676357	56.946					
M_primigenius_H	*M_primigenius*	5683630	48.478					
M_primigenius_P (Oimyakon)	*M_primigenius*	11710271	99.882					

Sample	Specie	non-missing sites	% non-missing sites	Number of sites covered (≥1X)	% Genome covered	Avg. DoC In region covered	% target sites covered	Number of target sites covered
M_primigenius_Q (Wrangel Island)	*M_primigenius*	11718248	99.951					
M_primigenius_S	*M_primigenius*	6821603	58.185					
M_primigenius_V	*M_primigenius*	11348006	96.793					
M_primiginius_1 (Yuka)	*M_primigenius*	11040170	94.167					
P_antiquus_N	P_antiquus	11543784	98.462					
P_antiquus_O (NEU2A)	P_antiquus	1950009	16.633					

ACKNOWLEDGMENTS

The authors thank Emil Karpinski and Karin Hönig for technical support. JAG and ADG were supported by the Deutsche Forschungsgeminschaft (DFG, grant GR 3924/7-1). JRM was funding by Danish National Research Foundation award 143. ML was supported by the Sweden-America Foundation.

REFERENCES

Abascal, F., Corvelo, A., Cruz, F., Villanueva-Cañas, J. L., Vlasova, A., Marcet-Houben, M., Martínez-Cruz, B., Cheng, J. Y., Prieto, P., Quesada, V., Quilez, J., Li, G., García, F., Rubio-Camarillo, M., Frias, L., Ribeca, P., Capella-Gutiérrez, S., Rodríguez, J. M., Câmara, F., … Godoy, J. A. (2016). Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx. Genome Biology, 17(1), 251. https://doi.org/10.1186/s13059-016-1090-1

Barnes, I., Shapiro, B., Lister, A., Kuznetsova, T., Sher, A., Guthrie, D., & Thomas, M. G. (2007). Genetic Structure and Extinction of the Woolly Mammoth, Mammuthus primigenius. Current Biology, 17(12), 1072–1075. https://doi.org/10.1016/j.cub.2007.05.035

Benazzo, A., Trucchi, E., Cahill, J. A., Maisano Delser, P., Mona, S., Fumagalli, M., Bunnefeld, L., Cornetti, L., Ghirotto, S., Girardi, M., Ometto, L., Panziera, A., Rota-Stabelli, O., Zanetti, E., Karamanlidis, A., Groff, C., Paule, L., Gentile, L., Vilà, C., … Bertorelle, G. (2017). Survival and divergence in a small group: The extraordinary genomic history of the endangered Apennine brown bear stragglers. Proceedings of the National Academy of Sciences, 114(45), E9589–E9597. https://doi.org/10.1073/pnas.1707279114

Bendl, J., Stourac, J., Salanda, O., Pavelka, A., Wieben, E. D., Zendulka, J., Brezovsky, J., & Damborsky, J. (2014). PredictSNP: Robust and Accurate Consensus Classifier for Prediction of Disease-Related Mutations. PLoS Computational Biology, 10(1), e1003440. https://doi.org/10.1371/journal.pcbi.1003440

Bottardi, S., Guieze, R., Bourgoin, V., Fotouhi-Ardakani, N., Dougé, A., Darracq, A., Lakehal, Y. A., Berger, M. G., Mollica, L., Bay, J.-O., Omichinski, J. G., & Milot, E. (2020). MNDA controls the expression of MCL-1 and BCL-2 in chronic lymphocytic leukemia cells. Experimental Hematology, 88, 68-82. e5. https://doi.org/10.1016/j.exphem.2020.07.004

Brand, C. M., Johnson, M. B., Parker, L. D., Maldonado, J. E., Korte, L., Vanthomme, H., Alonso, A., Ruiz-Lopez, M. J., Wells, C. P., & Ting, N. (2020). Abundance, density, and social structure of African forest elephants (Loxodonta cyclotis) in a human-modified landscape in southwestern Gabon. PLOS ONE, 15(4), e0231832. https://doi.org/10.1371/journal.pone.0231832

Capriotti, E., Fariselli, P., & Casadio, R. (2005). I-Mutant2.0: Predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Research, 33(Web Server), W306–W310. https://doi.org/10.1093/nar/gki375

Casas-Marce, M., Marmesat, E., Soriano, L., Martínez-Cruz, B., Lucena-Perez, M., Nocete, F., Rodríguez-Hidalgo, A., Canals, A., Nadal, J., Detry, C., Bernáldez-Sánchez, E., Fernández-Rodríguez, C., Pérez-Ripoll, M., Stiller, M., Hofreiter, M., Rodríguez, A., Revilla, E., Delibes, M., & Godoy, J. A. (2017). Spatiotemporal Dynamics of Genetic Variation in the Iberian Lynx along Its Path to Extinction Reconstructed with Ancient DNA. Molecular Biology and Evolution, 34(11), 2893–2907. https://doi.org/10.1093/molbev/msx222

Castellano, S., Parra, G., Sanchez-Quinto, F. A., Racimo, F., Kuhlwilm, M., Kircher, M., Sawyer, S., Fu, Q., Heinze, A., Nickel, B., Dabney, J., Siebauer, M., White, L., Burbano, H. A., Renaud, G., Stenzel, U., Lalueza-Fox, C., de la Rasilla, M., Rosas, A., … Paabo, S. (2014). Patterns of coding variation in the complete exomes of three Neandertals. Proceedings of the National Academy of Sciences, 111(18), 6666–6671. https://doi.org/10.1073/pnas.1405138111

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. GigaScience, 4(1), 7. https://doi.org/10.1186/s13742-015-0047-8

Chang, D., Knapp, M., Enk, J., Lippold, S., Kircher, M., Lister, A., MacPhee, R. D. E., Widga, C., Czechowski, P., Sommer, R., Hodges, E., Stümpel, N., Barnes, I., Dalén, L., Derevianko, A., Germonpré, M., Hillebrand-Voiculescu, A., Constantin, S., Kuznetsova, T., … Shapiro, B. (2017). The evolutionary and phylogeographic history of woolly mammoths: A comprehensive mitogenomic analysis. Scientific Reports, 7(1), 44585. https://doi.org/10.1038/srep44585

Chen, X., Pan, H., Li, J., Zhang, G., Cheng, S., Zuo, N., Zhao, Q., & Peng, Z. (2019). Inhibition of myeloid differentiation 1 specifically in colon with antisense oligonucleotide exacerbates dextran sodium sulfate-induced colitis. Journal of Cellular Biochemistry, 120(10), 16888–16899. https://doi.org/10.1002/jcb.28947

Chen, Y., Aulia, S., Li, L., & Tang, B. L. (2006). AMIGO and friends: An emerging family of brain-enriched, neuronal growth modulating, type I transmembrane proteins with leucine-rich repeats (LRR) and cell adhesion molecule motifs. Brain Research Reviews, 51(2), 265–274. https://doi.org/10.1016/j.brainresrev.2005.11.005

Choi, Y., & Chan, A. P. (2015). PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics, 31(16), 2745–2747. https://doi.org/10.1093/bioinformatics/btv195

Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., & Chan, A. P. (2012). Predicting the Functional Effect of Amino Acid Substitutions and Indels. PLoS ONE, 7(10), e46688. https://doi.org/10.1371/journal.pone.0046688

Clark, P. U., Dyke, A. S., Shakun, J. D., Carlson, A. E., Clark, J., Wohlfarth, B., Mitrovica, J. X., Hostetler, S. W., & McCabe, A. M. (2009). The Last Glacial Maximum. Science, 325(5941), 710–714. https://doi.org/10.1126/science.1172873

Dabney, J., & Meyer, M. (2012). Length and GC-biases during sequencing library amplification: A comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. BioTechniques, 52(2). https://doi.org/10.2144/000113809

Dastjerdi, A., Robert, C., & Watson, M. (2014). Low coverage sequencing of two Asian elephant (Elephas maximus) genomes. GigaScience, 3(1), 12. https://doi.org/10.1186/2047-217X-3-12

de Filippo, C., Key, F. M., Ghirotto, S., Benazzo, A., Meneu, J. R., Weihmann, A., NISC Comparative Sequence Program, Parra, G., Green, E. D., & Andrés, A. M. (2016). Recent Selection Changes in Human Genes under Long-Term Balancing Selection. Molecular Biology and Evolution, 33(6), 1435–1447. https://doi.org/10.1093/molbev/msw023

de Manuel, M., Barnett, R., Sandoval-Velasco, M., Yamaguchi, N., Garrett Vieira, F., Zepeda Mendoza, M. L., Liu, S., Martin, M. D., Sinding, M.-H. S., Mak, S. S. T., Carøe, C., Liu, S., Guo, C., Zheng, J., Zazula, G., Baryshnikov, G., Eizirik, E., Koepfli, K.-P., Johnson, W. E., … Gilbert, M. T. P. (2020). The evolutionary history of extinct and living lions. Proceedings of the National Academy of Sciences, 117(20), 10927–10934. https://doi.org/10.1073/pnas.1919423117

Debruyne, R., Chu, G., King, C. E., Bos, K., Kuch, M., Schwarz, C., Szpak, P., Gröcke, D. R., Matheus, P., Zazula, G., Guthrie, D., Froese, D., Buigues, B., de Marliave, C., Flemming, C., Poinar, D., Fisher, D., Southon, J., Tikhonov, A. N., … Poinar, H. N. (2008). Out of America: Ancient DNA Evidence for a New World Origin of Late Quaternary Woolly Mammoths. Current Biology, 18(17), 1320–1326. https://doi.org/10.1016/j.cub.2008.07.061

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics, 43(5), 491–498. https://doi.org/10.1038/ng.806

Dewan, A. (2021). Olfactory signaling via trace amine-associated receptors. Cell and Tissue Research, 383(1), 395–407. https://doi.org/10.1007/s00441-020-03331-5

Díez-del-Molino, D., Sánchez-Barreiro, F., Barnes, I., Gilbert, M. T. P., & Dalén, L. (2018). Quantifying Temporal Genomic Erosion in Endangered Species. Trends in Ecology & Evolution, 33(3), 176–185. https://doi.org/10.1016/j.tree.2017.12.002

Dolan, J., Walshe, K., Alsbury, S., Hokamp, K., O'Keeffe, S., Okafuji, T., Miller, S. F., Tear, G., & Mitchell, K. J. (2007). The extracellular Leucine-Rich Repeat superfamily; a comparative survey and analysis of evolutionary relationships and expression patterns. BMC Genomics, 8(1), 320. https://doi.org/10.1186/1471-2164-8-320

Doyle, J. M., Gao, J., Wang, J., Yang, M., & Potts, P. R. (2010). MAGE-RING Protein Complexes Comprise a Family of E3 Ubiquitin Ligases. Molecular Cell, 39(6), 963–974. https://doi.org/10.1016/j.molcel.2010.08.029

Enk, J., Devault, A., Debruyne, R., King, C. E., Treangen, T., O'Rourke, D., Salzberg, S. L., Fisher, D., MacPhee, R., & Poinar, H. (2011). Complete Columbian mammoth mitogenome suggests interbreeding with woolly mammoths. Genome Biology, 12(5), R51. https://doi.org/10.1186/gb-2011-12-5-r51

Enk, J., Devault, A., Widga, C., Saunders, J., Szpak, P., Southon, J., Rouillard, J.-M., Shapiro, B., Golding, G. B., Zazula, G., Froese, D., Fisher, D. C., MacPhee, R. D. E., & Poinar, H. (2016). Mammuthus Population Dynamics in Late Pleistocene North America: Divergence, Phylogeography, and Introgression. Frontiers in Ecology and Evolution, 4. https://doi.org/10.3389/fevo.2016.00042

Etzerodt, A., & Moestrup, S. K. (2013). CD163 and Inflammation: Biological, Diagnostic, and Therapeutic Aspects. Antioxidants & Redox Signaling, 18(17), 2352–2363. https://doi.org/10.1089/ars.2012.4834

Fellows Yates, J. A., Drucker, D. G., Reiter, E., Heumos, S., Welker, F., Münzel, S. C., Wojtal, P., Lázničková-Galetová, M., Conard, N. J., Herbig, A., Bocherens, H., & Krause, J. (2017). Central European Woolly Mammoth Population Dynamics: Insights from Late Pleistocene Mitochondrial Genomes. Scientific Reports, 7(1), 17714. https://doi.org/10.1038/s41598-017-17723-1

Ferrero, D. M., Lemon, J. K., Fluegge, D., Pashkovski, S. L., Korzan, W. J., Datta, S. R., Spehr, M., Fendt, M., & Liberles, S. D. (2011). Detection and avoidance of a carnivore odor by prey. Proceedings of the National Academy of Sciences, 108(27), 11235–11240. https://doi.org/10.1073/pnas.1103317108

Fleischer, R. C., Perry, E. A., Muralidharan, K., Stevens, E. E., & Wemmer, C. M. (2001). Phylogeography of the Asian Elephant (Elephas maximus) Based on Mitochondrial DNA. Evolution, 55(9), 1882–1892. https://doi.org/10.1111/j.0014-3820.2001.tb00837.x

Frankham, R. (2005). Genetics and extinction. Biological Conservation, 126(2), 131–140. https://doi.org/10.1016/j.biocon.2005.05.002

Fry, E., Kim, S. K., Chigurapti, S., Mika, K. M., Ratan, A., Dammermann, A., Mitchell, B. J., Miller, W., & Lynch, V. J. (2020). Functional Architecture of Deleterious Genetic Variants in the Genome of a Wrangel Island Mammoth. Genome Biology and Evolution, 12(3), 48–58. https://doi.org/10.1093/gbe/evz279

Fujimoto, A., Farooq, M., Fujikawa, H., Inoue, A., Ohyama, M., Ehama, R., Nakanishi, J., Hagihara, M., Iwabuchi, T., Aoki, J., Ito, M., & Shimomura, Y. (2012). A Missense Mutation within the Helix Initiation Motif of the Keratin K71 Gene Underlies Autosomal Dominant Woolly Hair/Hypotrichosis. Journal of Investigative Dermatology, 132(10), 2342–2349. https://doi.org/10.1038/jid.2012.154

Gilbert, M. T. P., Drautz, D. I., Lesk, A. M., Ho, S. Y. W., Qi, J., Ratan, A., Hsu, C.-H., Sher, A., Dalen, L., Gotherstrom, A., Tomsho, L. P., Rendulic, S., Packard, M., Campos, P. F., Kuznetsova, T. V., Shidlovskiy, F., Tikhonov, A., Willerslev, E., Iacumin, P., … Schuster, S. C. (2008). Intraspecific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes. Proceedings of the National Academy of Sciences, 105(24), 8327–8332. https://doi.org/10.1073/pnas.0802315105

Graham, R. W., Belmecheri, S., Choy, K., Culleton, B. J., Davies, L. J., Froese, D., Heintzman, P. D., Hritz, C., Kapp, J. D., Newsom, L. A., Rawcliffe, R., Saulnier-Talbot, É., Shapiro, B., Wang, Y., Williams, J. W., & Wooller, M. J. (2016). Timing and causes of mid-Holocene mammoth extinction on St. Paul Island, Alaska. Proceedings of the National Academy of Sciences, 113(33), 9310–9314. https://doi.org/10.1073/pnas.1604903113

Gutiérrez-Muñoz, C., Méndez-Barbero, N., Svendsen, P., Sastre, C., Fernández-Laso, V., Quesada, P., Egido, J., Escolá-Gil, J. C., Martín-Ventura, J. L., Moestrup, S. K., & Blanco-Colio, L. M. (2020). CD163 deficiency increases foam cell formation and plaque progression in atherosclerotic mice. The FASEB Journal, 34(11), 14960–14976. https://doi.org/10.1096/fj.202000177R

Haile, J., Froese, D. G., MacPhee, R. D. E., Roberts, R. G., Arnold, L. J., Reyes, A. V., Rasmussen, M., Nielsen, R., Brook, B. W., Robinson, S., Demuro, M., Gilbert, M. T. P., Munch, K., Austin, J. J., Cooper, A., Barnes, I., Möller, P., & Willerslev, E. (2009). Ancient DNA reveals late survival of mammoth and horse in interior Alaska. Proceedings of the National Academy of Sciences, 106(52), 22352–22357. https://doi.org/10.1073/pnas.0912510106

Harris, K., & Nielsen, R. (2016). The Genetic Cost of Neanderthal Introgression. Genetics, 203(2), 881–891. https://doi.org/10.1534/genetics.116.186890

Hofreiter, M., Paijmans, J. L. A., Goodchild, H., Speller, C. F., Barlow, A., Fortes, G. G., Thomas, J. A., Ludwig, A., & Collins, M. J. (2015). The future of ancient DNA: Technical advances and conceptual shifts: Prospects & Overviews. BioEssays, 37(3), 284–293. https://doi.org/10.1002/bies.201400160

Howard, S. R., Guasti, L., Ruiz-Babot, G., Mancini, A., David, A., Storr, H. L., Metherell, L. A., Sternberg, M. J., Cabrera, C. P., Warren, H. R., Barnes, M. R., Quinton, R., Roux, N., Young, J., Guiochon-Mantel, A., Wehkalampi, K., André, V., Gothilf, Y., Cariboni, A., & Dunkel, L. (2016). IGSF 10 mutations dysregulate gonadotropin-releasing hormone neuronal migration resulting in delayed puberty. EMBO Molecular Medicine, 8(6), 626–642. https://doi.org/10.15252/emmm.201606250

Igea, J., Juste, J., & Castresana, J. (2010). Novel intron markers to study the phylogeny of closely related mammalian species. BMC Evolutionary Biology, 10(1), 369. https://doi.org/10.1186/1471-2148-10-369

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., & Orlando, L. (2013). mapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics, 29(13), 1682–1684. https://doi.org/10.1093/bioinformatics/btt193

Juric, I., Aeschbacher, S., & Coop, G. (2016). The Strength of Selection against Neanderthal Introgression. PLOS Genetics, 12(11), e1006340. https://doi.org/10.1371/journal.pgen.1006340

Kara, E. E., Bastow, C. R., McKenzie, D. R., Gregor, C. E., Fenix, K. A., Babb, R., Norton, T. S., Zotos, D., Rodda, L. B., Hermes, J. R., Bourne, K., Gilchrist, D. S., Nibbs, R. J., Alsharifi, M., Vinuesa, C. G., Tarlinton, D. M., Brink, R., Hill, G. R., Cyster, J. G., … McColl, S. R. (2018). Atypical chemokine receptor 4 shapes activated B cell fate. Journal of Experimental Medicine, 215(3), 801–813. https://doi.org/10.1084/jem.20171067

Karpinski, E., Hackenberger, D., Zazula, G., Widga, C., Duggan, A. T., Golding, G. B., Kuch, M., Klunk, J., Jass, C. N., Groves, P., Druckenmiller, P., Schubert, B. W., Arroyo-Cabrales, J., Simpson, W. F., Hoganson, J. W., Fisher, D. C., Ho, S. Y. W., MacPhee, R. D. E., & Poinar, H. N. (2020). American mastodon mitochondrial genomes suggest multiple dispersal events in response to Pleistocene climate oscillations. Nature Communications, 11(1), 4048. https://doi.org/10.1038/s41467-020-17893-z

Key, F. M., Abdul-Aziz, M. A., Mundry, R., Peter, B. M., Sekar, A., D'Amato, M., Dennis, M. Y., Schmidt, J. M., & Andrés, A. M. (2018). Human local adaptation of the TRPM8 cold receptor along a latitudinal cline. PLOS Genetics, 14(5), e1007298. https://doi.org/10.1371/journal.pgen.1007298

Khan, A., Patel, K., Shukla, H., Viswanathan, A., van der Valk, T., Borthakur, U., Nigam, P., Zachariah, A., Jhala, Y. V., Kardos, M., & Ramakrishnan, U. (2021). Genomic evidence for inbreeding depression and purging of deleterious genetic variation in Indian tigers. Proceedings of the National Academy of Sciences, 118(49), e2023018118. https://doi.org/10.1073/pnas.2023018118

Kircher, M., Sawyer, S., & Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Research, 40(1), e3–e3. https://doi.org/10.1093/nar/gkr771

Kleinman-Ruiz, D., Lucena-Perez, M., Villanueva, B., Fernández, J., Saveljev, A. P., Ratkiewicz, M., Schmidt, K., Galtier, N., García-Dorado, A., & Godoy, J. A. (2022). Purging of deleterious burden in the endangered Iberian lynx. Proceedings of the National Academy of Sciences, 119(11), e2110614119. https://doi.org/10.1073/pnas.2110614119

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics, 15(1), 356. https://doi.org/10.1186/s12859-014-0356-4

Kuzmin, Y. V. (2010). Extinction of the woolly mammoth (Mammuthus primigenius) and woolly rhinoceros (Coelodonta antiquitatis) in Eurasia: Review of chronological and environmental issues. Boreas, 39(2), 247–261. https://doi.org/10.1111/j.1502-3885.2009.00122.x

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics, 26(5), 589–595. https://doi.org/10.1093/bioinformatics/btp698

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics, 25(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, Y. I., van de Geijn, B., Raj, A., Knowles, D. A., Petti, A. A., Golan, D., Gilad, Y., & Pritchard, J. K. (2016). RNA splicing is a primary link between genetic variation and disease. Science, 352(6285), 600–604. https://doi.org/10.1126/science.aad9417

Liphaus, B. L., Caramalho, I., Rangel-Santos, A., Silva, C. A., Demengeot, J., & Carneiro-Sampaio, M. M. S. (2020). LRBA deficiency: A new genetic cause of monogenic lupus. Annals of the Rheumatic Diseases, 79(3), 427–428. https://doi.org/10.1136/annrheumdis-2019-216410

Lister, A. M. (2001). The Origin and Evolution of the Woolly Mammoth. Science, 294(5544), 1094–1097. https://doi.org/10.1126/science.1056370

Lister, A. M., Sher, A. V., van Essen, H., & Wei, G. (2005). The pattern and process of mammoth evolution in Eurasia. Quaternary International, 126–128, 49–64. https://doi.org/10.1016/j.quaint.2004.04.014

Lord, E., Dussex, N., Kierczak, M., Díez-del-Molino, D., Ryder, O. A., Stanton, D. W. G., Gilbert, M. T. P., Sánchez-Barreiro, F., Zhang, G., Sinding, M.-H. S., Lorenzen, E. D., Willerslev, E., Protopopov, A., Shidlovskiy, F., Fedorov, S., Bocherens, H., Nathan, S. K. S. S., Goossens, B., van der Plicht, J., … Dalén, L. (2020). Pre-extinction Demographic Stability and Genomic Signatures of Adaptation in the Woolly Rhinoceros. Current Biology, 30(19), 3871-3879.e7. https://doi.org/10.1016/j.cub.2020.07.046

Lorenzen, E. D., Nogués-Bravo, D., Orlando, L., Weinstock, J., Binladen, J., Marske, K. A., Ugan, A., Borregaard, M. K., Gilbert, M. T. P., Nielsen, R., Ho, S. Y. W., Goebel, T., Graf, K. E., Byers, D., Stenderup, J. T., Rasmussen, M., Campos, P. F., Leonard, J. A., Koepfli, K.-P., … Willerslev, E. (2011). Species-specific responses of Late Quaternary megafauna to climate and humans. Nature, 479(7373), 359–364. https://doi.org/10.1038/nature10574

Lynch, M., Conery, J., & Burger, R. (1995). Mutation Accumulation and the Extinction of Small Populations. The American Naturalist, 146(4), 489–518. https://doi.org/10.1086/285812

Lynch, V. J., Bedoya-Reina, O. C., Ratan, A., Sulak, M., Drautz-Moses, D. I., Perry, G. H., Miller, W., & Schuster, S. C. (2015). Elephantid Genomes Reveal the Molecular Bases of Woolly Mammoth Adaptations to the Arctic. Cell Reports, 12(2), 217–228. https://doi.org/10.1016/j.celrep.2015.06.027

MacDonald, G. M., Beilman, D. W., Kuzmin, Y. V., Orlova, L. A., Kremenetski, K. V., Shapiro, B., Wayne, R. K., & Van Valkenburgh, B. (2012). Pattern of extinction of the woolly mammoth in Beringia. Nature Communications, 3(1), 893. https://doi.org/10.1038/ncomms1881

MacPhee, R.D.E. and Marx, P.A. (1997). The 40,000-year plaque: Humans, hyperdisease and first contact extinctions. In Natural Change and Human Impact in Madagascar (Goodman, S. and Patterson, B., eds), pp. 169–217, Smithsonia (pp. 169–217).

Mangerud, J. (2021). The discovery of the Younger Dryas, and comments on the current meaning and usage of the term. Boreas, 50(1), 1–5. https://doi.org/10.1111/bor.12481

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., Flicek, P., & Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome Biology, 17(1), 122. https://doi.org/10.1186/s13059-016-0974-4

Meyer, M., Palkopoulou, E., Baleka, S., Stiller, M., Penkman, K. E. H., Alt, K. W., Ishida, Y., Mania, D., Mallick, S., Meijer, T., Meller, H., Nagel, S., Nickel, B., Ostritz, S., Rohland, N., Schauer, K., Schüler, T., Roca, A. L., Reich, D., … Hofreiter, M. (2017). Palaeogenomes of Eurasian straight-tusked elephants challenge the current view of elephant evolution. ELife, 6, e25413. https://doi.org/10.7554/eLife.25413

Morris, K. M., Wright, B., Grueber, C. E., Hogg, C., & Belov, K. (2015). Lack of genetic diversity across diverse immune genes in an endangered mammal, the Tasmanian devil (*Sarcophilus harrisii)*. Molecular Ecology, 24(15), 3860–3872. https://doi.org/10.1111/mec.13291

Munshi-South, J. (2011). Relatedness and Demography of African Forest Elephants: Inferences from Noninvasive Fecal DNA Analyses. Journal of Heredity, 102(4), 391–398. https://doi.org/10.1093/jhered/esr030

Münzberg, H., & Morrison, C. D. (2015). Structure, production and signaling of leptin. Metabolism, 64(1), 13–23. https://doi.org/10.1016/j.metabol.2014.09.010

Murchie, T. J., Kuch, M., Duggan, A. T., Ledger, M. L., Roche, K., Klunk, J., Karpinski, E., Hackenberger, D., Sadoway, T., MacPhee, R., Froese, D., & Poinar, H. (2020). Optimizing extraction and targeted capture of ancient environmental DNA for reconstructing past environments using the PalaeoChip Arctic-1.0 bait-set. Quaternary Research, 1–24. https://doi.org/10.1017/qua.2020.59

Murchie, T. J., Monteath, A. J., Mahony, M. E., Long, G. S., Cocker, S., Sadoway, T., Karpinski, E., Zazula, G., MacPhee, R. D. E., Froese, D., & Poinar, H. N. (2021). Collapse of the mammoth-steppe in central Yukon as revealed by ancient environmental DNA. Nature Communications, 12(1), 7120. https://doi.org/10.1038/s41467-021-27439-6

Narayana, S. K., Helbig, K. J., McCartney, E. M., Eyre, N. S., Bull, R. A., Eltahla, A., Lloyd, A. R., & Beard, M. R. (2015). The Interferon-induced Transmembrane Proteins, IFITM1, IFITM2, and IFITM3 Inhibit Hepatitis C Virus Entry. Journal of Biological Chemistry, 290(43), 25946–25959. https://doi.org/10.1074/jbc.M115.657346

Natsuga, K., Watanabe, M., Nishie, W., & Shimizu, H. (2019). Life before and beyond blistering: The role of collagen XVII in epidermal physiology. Experimental Dermatology, 28(10), 1135–1141. https://doi.org/10.1111/exd.13550

Ng, A. C. Y., Eisenberg, J. M., Heath, R. J. W., Huett, A., Robinson, C. M., Nau, G. J., & Xavier, R. J. (2011). Human leucine-rich repeat proteins: A genome-wide bioinformatic categorization and functional analysis in innate immunity. Proceedings of the National Academy of Sciences, 108(Supplement_1), 4631–4638. https://doi.org/10.1073/pnas.1000093107

Nikolskiy, P. A. (2011). Last straw versus Blitzkrieg overkill: Climate-driven changes in the Arctic Siberian mammoth population and the Late Pleistocene extinction problem. Quaternary Science Reviews, 20.

Nogués-Bravo, D., Rodríguez, J., Hortal, J., Batra, P., & Araújo, M. B. (2008). Climate Change, Humans, and the Extinction of the Woolly Mammoth. PLoS Biology, 6(4), e79. https://doi.org/10.1371/journal.pbio.0060079

Nyström, V., Dalén, L., Vartanyan, S., Lidén, K., Ryman, N., & Angerbjörn, A. (2010). Temporal genetic change in the last remaining population of woolly mammoth. Proceedings of the Royal Society B: Biological Sciences, 277(1692), 2331–2337. https://doi.org/10.1098/rspb.2010.0301

Nyström, V., Humphrey, J., Skoglund, P., McKEOWN, N. J., Vartanyan, S., Shaw, P. W., Lidén, K., Jakobsson, M., Barnes, I., Angerbjörn, A., Lister, A., & Dalén, L. (2012). Microsatellite genotyping reveals end-Pleistocene decline in mammoth autosomal genetic variation: MICROSATELLITE VARIATION IN WOOLLY MAMMOTHS. Molecular Ecology, 21(14), 3391–3402. https://doi.org/10.1111/j.1365-294X.2012.05525.x

Oldenburg, M., Kruger, A., Ferstl, R., Kaufmann, A., Nees, G., Sigmund, A., Bathke, B., Lauterbach, H., Suter, M., Dreher, S., Koedel, U., Akira, S., Kawai, T., Buer, J., Wagner, H., Bauer, S., Hochrein, H., & Kirschning, C. J. (2012). TLR13 Recognizes Bacterial 23S rRNA Devoid of Erythromycin Resistance-Forming Modification. Science, 337(6098), 1111–1115. https://doi.org/10.1126/science.1220363

Palkopoulou, E., Dalén, L., Lister, A. M., Vartanyan, S., Sablin, M., Sher, A., Edmark, V. N., Brandström, M. D., Germonpré, M., Barnes, I., & Thomas, J. A. (2013). Holarctic genetic structure and range dynamics in the woolly mammoth. Proceedings of the Royal Society B: Biological Sciences, 280(1770), 20131910. https://doi.org/10.1098/rspb.2013.1910

Palkopoulou, E., Lipson, M., Mallick, S., Nielsen, S., Rohland, N., Baleka, S., Karpinski, E., Ivancevic, A. M., To, T.-H., Kortschak, R. D., Raison, J. M., Qu, Z., Chin, T.-J., Alt, K. W., Claesson, S., Dalén, L., MacPhee, R. D. E., Meller, H., Roca, A. L., … Reich, D. (2018). A comprehensive genomic history of extinct and living elephants. Proceedings of the National Academy of Sciences, 115(11), E2566–E2574. https://doi.org/10.1073/pnas.1720554115

Palkopoulou, E., Mallick, S., Skoglund, P., Enk, J., Rohland, N., Li, H., Omrak, A., Vartanyan, S., Poinar, H., Götherström, A., Reich, D., & Dalén, L. (2015). Complete Genomes Reveal Signatures of Demographic and Genetic Declines in the Woolly Mammoth. Current Biology, 25(10), 1395–1400. https://doi.org/10.1016/j.cub.2015.04.007

Parham, P. (2005). MHC class I molecules and kirs in human history, health and survival. Nature Reviews Immunology, 5(3), 201–214. https://doi.org/10.1038/nri1570

Pečnerová, P., Díez-del-Molino, D., Dussex, N., Feuerborn, T., von Seth, J., van der Plicht, J., Nikolskiy, P., Tikhonov, A., Vartanyan, S., & Dalén, L. (2017). Genome-Based Sexing Provides Clues about Behavior and Social Structure in the Woolly Mammoth. Current Biology, 27(22), 3505-3510.e3. https://doi.org/10.1016/j.cub.2017.09.064

Pečnerová, P., Díez-del-Molino, D., Vartanyan, S., & Dalén, L. (2016). Changes in variation at the MHC class II DQA locus during the final demise of the woolly mammoth. Scientific Reports, 6(1), 25274. https://doi.org/10.1038/srep25274

R Core Team. (2020). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. https://www.R-project.org/

Rasmussen, S. O., Andersen, K. K., Svensson, A. M., Steffensen, J. P., Vinther, B. M., Clausen, H. B., Siggaard-Andersen, M.-L., Johnsen, S. J., Larsen, L. B., Dahl-Jensen, D., Bigler, M., Röthlisberger, R., Fischer, H., Goto-Azuma, K., Hansson, M. E., & Ruth, U. (2006). A new Greenland ice core chronology for the last glacial termination. Journal of Geophysical Research, 111(D6), D06102. https://doi.org/10.1029/2005JD006079

Reddy, P. C., Sinha, I., Kelkar, A., Habib, F., Pradhan, S. J., Sukumar, R., & Galande, S. (2015). Comparative sequence analyses of genome and transcriptome reveal novel transcripts and variants in the Asian elephant Elephas maximus. Journal of Biosciences, 40(5), 891–907. https://doi.org/10.1007/s12038-015-9580-y

Reher, D., Key, F. M., Andrés, A. M., & Kelso, J. (2019). Immune Gene Diversity in Archaic and Present-day Humans. Genome Biology and Evolution, 11(1), 232–241. https://doi.org/10.1093/gbe/evy271

Robinson, J. A., Ortega-Del Vecchyo, D., Fan, Z., Kim, B. Y., vonHoldt, B. M., Marsden, C. D., Lohmueller, K. E., & Wayne, R. K. (2016). Genomic Flatlining in the Endangered Island Fox. Current Biology, 26(9), 1183–1189. https://doi.org/10.1016/j.cub.2016.02.062

Roca, A. L., Georgiadis, N., & O'Brien, S. J. (2005). Cytonuclear genomic dissociation in African elephant species. Nature Genetics, 37(1), 96–100. https://doi.org/10.1038/ng1485

Rogers, R. L., & Slatkin, M. (2017). Excess of genomic defects in a woolly mammoth on Wrangel island. PLOS Genetics, 13(3), e1006601. https://doi.org/10.1371/journal.pgen.1006601

Rohland, N., Reich, D., Mallick, S., Meyer, M., Green, R. E., Georgiadis, N. J., Roca, A. L., & Hofreiter, M. (2010). Genomic DNA Sequences from Mastodon and Woolly Mammoth Reveal Deep Speciation of Forest and Savanna Elephants. PLoS Biology, 8(12), e1000564. https://doi.org/10.1371/journal.pbio.1000564

Rutigliano, G., & Zucchi, R. (2020). Molecular Variants in Human Trace Amine-Associated Receptors and Their Implications in Mental and Metabolic Disorders. Cellular and Molecular Neurobiology, 40(2), 239–255. https://doi.org/10.1007/s10571-019-00743-y

Schubert, M., Ginolhac, A., Lindgreen, S., Thompson, J. F., AL-Rasheid, K. A., Willerslev, E., Krogh, A., & Orlando, L. (2012). Improving ancient DNA read mapping against modern reference genomes. BMC Genomics, 13(1), 178. https://doi.org/10.1186/1471-2164-13-178

Schubert, M., Lindgreen, S., & Orlando, L. (2016). AdapterRemoval v2: Rapid adapter trimming, identification, and read merging. BMC Research Notes, 9(1), 88. https://doi.org/10.1186/s13104-016-1900-2

Schuttler, S., Whittaker, A., Jeffery, K., & Eggert, L. (2014). African forest elephant social networks: Fission-fusion dynamics, but fewer associations. Endangered Species Research, 25(2), 165–173. https://doi.org/10.3354/esr00618

Schwarz, C., Debruyne, R., Kuch, M., McNally, E., Schwarcz, H., Aubrey, A. D., Bada, J., & Poinar, H. (2009). New insights from old bones: DNA preservation and degradation in permafrost preserved mammoth remains. Nucleic Acids Research, 37(10), 3215–3229. https://doi.org/10.1093/nar/gkp159

Scotti, M. M., & Swanson, M. S. (2016). RNA mis-splicing in disease. Nature Reviews Genetics, 17(1), 19–32. https://doi.org/10.1038/nrg.2015.3

Singh, S. B., Ornatowski, W., Vergne, I., Naylor, J., Delgado, M., Roberts, E., Ponpuak, M., Master, S., Pilli, M., White, E., Komatsu, M., & Deretic, V. (2010). Human IRGM regulates autophagy and cell-autonomous immunity functions through mitochondria. Nature Cell Biology, 12(12), 1154–1165. https://doi.org/10.1038/ncb2119

Smith, C. I., Chamberlain, A. T., Riley, M. S., Stringer, C., & Collins, M. J. (2003). The thermal history of human fossils and the likelihood of successful DNA amplification. Journal of Human Evolution, 45(3), 203–217. https://doi.org/10.1016/S0047-2484(03)00106-4

Sproston, N. R., & Ashworth, J. J. (2018). Role of C-Reactive Protein at Sites of Inflammation and Infection. Frontiers in Immunology, 9, 754. https://doi.org/10.3389/fimmu.2018.00754

Stuart, A. J., Kosintsev, P. A., Higham, T. F. G., & Lister, A. M. (2004). Pleistocene to Holocene extinction dynamics in giant deer and woolly mammoth. Nature, 431(7009), 684–689. https://doi.org/10.1038/nature02890

Sullivan, A. P., de Manuel, M., Marques-Bonet, T., & Perry, G. H. (2017). An evolutionary medicine perspective on Neandertal extinction. Journal of Human Evolution, 108, 62–71. https://doi.org/10.1016/j.jhevol.2017.03.004

Telliez, J.-B., Bean, K. M., & Lin, L.-L. (2000). LRDD, a novel leucine rich repeat and death domain containing protein. Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology, 1478(2), 280–288. https://doi.org/10.1016/S0167-4838(00)00029-7

Tenthorey, J. L., Young, C., Sodeinde, A., Emerman, M., & Malik, H. S. (2020). Mutational resilience of antiviral restriction favors primate TRIM5α in host-virus evolutionary arms races. ELife, 9, e59988. https://doi.org/10.7554/eLife.59988

The UniProt Consortium, Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi, E., Bowler-Barnett, E. H., Britto, R., Bursteinas, B., Bye-A-Jee, H., Coetzee, R., Cukura, A., Da Silva, A., Denny, P., Dogan, T., Ebenezer, T., Fan, J., … Teodoro, D. (2021). UniProt: The universal protein knowledgebase in 2021. Nucleic Acids Research, 49(D1), D480–D489. https://doi.org/10.1093/nar/gkaa1100

van der Valk, T., Díez-del-Molino, D., Marques-Bonet, T., Guschanski, K., & Dalén, L. (2019). Historical Genomes Reveal the Genomic Consequences of Recent Population Decline in Eastern Gorillas. Current Biology, 29(1), 165-170.e6. https://doi.org/10.1016/j.cub.2018.11.055

van der Valk, T., Pečnerová, P., Díez-del-Molino, D., Bergström, A., Oppenheimer, J., Hartmann, S., Xenikoudakis, G., Thomas, J. A., Dehasque, M., Sağlıcan, E., Fidan, F. R., Barnes, I., Liu, S., Somel, M., Heintzman, P. D., Nikolskiy, P., Shapiro, B., Skoglund, P., Hofreiter, M., … Dalén, L. (2021). Million-year-old DNA sheds light on the genomic history of mammoths. Nature. https://doi.org/10.1038/s41586-021-03224-9

van der Valk, Tom, de Manuel, Marc, Marques-Bonet, Tomas, & Guschanski, Katerina. (n.d.). Estimates of genetic load suggest frequent purging of deleterious alleles in small populations. 30. https://doi.org/10.1101/696831

Vidya, T. N. C. (2016). Evolutionary History and Population Genetic Structure of Asian Elephants in India. Indian Journal of History of Science, 51(2.2). https://doi.org/10.16943/ijhs/2016/v51i2.2/48453

Vogl, C., Butola, T., Haag, N., Hausrat, T. J., Leitner, M. G., Moutschen, M., Lefèbvre, P. P., Speckmann, C., Garrett, L., Becker, L., Fuchs, H., de Angelis, M. H., Nietzsche, S., Kessels, M. M., Oliver, D., Kneussel, M., Kilimann, M. W., & Strenzke, N. (2017). The BEACH protein LRBA is required for hair bundle maintenance in cochlear hair cells and for hearing. EMBO Reports, 15.

Walters, R. J., & Berger, D. (2019). Implications of existing local (mal)adaptations for ecological forecasting under environmental change. Evolutionary Applications, 12(7), 1487–1502. https://doi.org/10.1111/eva.12840

Wang, Y. (n.d.). Late Quaternary dynamics of Arctic biota from ancient environmental genomics. 19. https://doi.org/10.1038/s41586-021-04016-x

Wang, Y., Pedersen, M. W., Alsos, I. G., De Sanctis, B., Racimo, F., Prohaska, A., Coissac, E., Owens, H. L., Merkel, M. K. F., Fernandez-Guerra, A., Rouillard, A., Lammers, Y., Alberti, A., Denoeud, F., Money, D., Ruter, A. H., McColl, H., Larsen, N. K., Cherezova, A. A., … Willerslev, E. (2021). Late Quaternary dynamics of Arctic biota from ancient environmental genomics. Nature, 600(7887), 86–92. https://doi.org/10.1038/s41586-021-04016-x

Whyte, C. E., Osman, M., Kara, E. E., Abbott, C., Foeng, J., McKenzie, D. R., Fenix, K. A., Harata-Lee, Y., Foyle, K. L., Boyle, S. T., Kochetkova, M., Aguilera, A. R., Hou, J., Li, X.-Y., Armstrong, M. A., Pederson, S. M., Comerford, I., Smyth, M. J., & McColl, S. R. (2020). ACKR4 restrains antitumor immunity by regulating CCL21. Journal of Experimental Medicine, 217(6), e20190634. https://doi.org/10.1084/jem.20190634

Wooller, M. J., Bataille, C., Druckenmiller, P., Erickson, G. M., Groves, P., Haubenstock, N., Howe, T., Irrgeher, J., Mann, D., Moon, K., Potter, B. A., Prohaska, T., Rasic, J., Reuther, J., Shapiro, B., Spaleta, K. J., & Willis, A. D. (2021). Lifetime mobility of an Arctic woolly mammoth. Science, 373(6556), 806–808. https://doi.org/10.1126/science.abg1134

Yamagata, K., Nagai, K., Miyamoto, H., Anzai, M., Kato, H., Miyamoto, K., Kurosaka, S., Azuma, R., Kolodeznikov, I. I., Protopopov, A. V., Plotnikov, V. V., Kobayashi, H., Kawahara-Miki, R., Kono, T., Uchida, M., Shibata, Y., Handa, T., Kimura, H., Hosoi, Y., … Iritani, A. (2019). Signs of biological activities of 28,000-year-old mammoth nuclei in mouse oocytes visualized by live-cell imaging. Scientific Reports, 9(1), 4050. https://doi.org/10.1038/s41598-019-40546-1

Zazula, G. D., MacPhee, R. D. E., Metcalfe, J. Z., Reyes, A. V., Brock, F., Druckenmiller, P. S., Groves, P., Harington, C. R., Hodgins, G. W. L., Kunz, M. L., Longstaffe, F. J., Mann, D. H., McDonald, H. G., Nalawade-Chavan, S., & Southon, J. R. (2014). American mastodon extirpation in the Arctic and Subarctic predates human colonization and terminal Pleistocene climate change. Proceedings of the National Academy of Sciences, 111(52), 18460–18465. https://doi.org/10.1073/pnas.1416072111

Zhou, Z., Swagemakers, S. M. A., Lourens, M. S., Suratannon, N., van der Spek, P., Dalm, V. A. S. H., Dik, W. A., IJspeert, H., & van Hagen, P. M. (2022). Did variants in inborn errors of immunity genes contribute to the extinction of Neanderthals? [Preprint]. Evolutionary Biology. https://doi.org/10.1101/2022.10.19.512108

3. CHAPTER III: MEASURING IMMUNOGENETIC DIVERSITY IN EXTANT ELEPHANTS

John A. Galindo[1,2], Gayle McEwen[1], Yasuko Ishida[3], Kathryn L. Perrin[4,5], Jazmín Ramos-Madrigal[6], Mads F. Bertelsen[4], Paul D. Ling[7], Alfred Roca[8], Alex D. Greenwood[1,9]

[1] Department of Wildlife Diseases, Leibniz Institute for Zoo and Wildlife Research, Berlin, Germany

[2] Department of Biology, Chemistry and Pharmacy, Freie Universität Berlin, Berlin, Germany

[3] Department of Animal Sciences, University of Illinois at Urbana-Champaign, Urbana, IL, United States

[4] Department of Veterinary Clinical Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Frederiksberg, Denmark

[5] San Diego WildlifenAlliance, Escondido, CA, United States

[6] Center for Evolutionary Hologenomics, GLOBE Institute, University of Copenhagen. Copenhagen, Denmark

[7] Molecular Virology and Microbiology, Baylor College of Medicine, Houston, Texas, United States

[8] Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL, United States

[9] Department of Veterinary Medicine, Freie Universität Berlin, Berlin, Germany

3.1. ABSTRACT

Immunogenetic variability is associated with the ability of populations to resist pathogens. Understanding the association is important to understanding the long-term survival and adaptation of vulnerable species. Unlike neutral markers, immune genes may retain high levels of polymorphism due to pathogen pressure. Elephants have undergone a dramatic decline in genetic diversity due anthropogenic activities, which may have genetic consequences for their immune gene diversity and the potential of elephants to respond to infection. We utilized a novel multilocus genotyping method coupled with the PacBio long read high throughput sequencing platform on toll-like receptors (TLRs) representing innate immunity and the Major Histocompatibility Complex (MHC) representing adaptive immunity from 263 samples of savanna elephants (*Loxodonta Africana*), forest elephants (*Loxodonta cyclotis*), and Asian elephants (*Elephas maximus*). Our analysis demonstrates high levels of inbreeding, heterozygosity deficiency and low genetic diversity in the three elephant populations. However, balancing selection appears to counteract the low levels of diversity in the MHC genes, conserving trans-species polymorphisms over the long-term.

3.2. INTRODUCTION

Genomic regions encoding for immune genes in vertebrates are among the most rapidly evolving (Barreiro & Quintana-Murci, 2010; Flajnik & Kasahara, 2010; Litman et al., 2005). The Major Histocompatibility Complex (MHC) is part of the adaptive immune system with a key function in specific pathogen recognition. MHC-pathogen interactions impose strong selective pressure on immune genes making these loci extremely polymorphic compared to the rest of the genome (Horrocks et al., 2015; Norman et al., 2017; The MHC sequencing consortium, 1999). Three mechanisms of balancing selection have been proposed to maintain MHC polymorphism. Heterozygote advantage (HA) (overdominance), negative frequency-dependent selection (NFDS) (rare allele advantage), and fluctuating selection (FS) (spatiotemporal selection) all maintain MHC polymorphism (reviewed in (Radwan et al., 2020; Spurgin & Richardson, 2010)). The diversity of innate immunity has evolved to defend organisms against pathogens (Buchmann, 2014; Messier-Solek et al., 2010). Toll-like receptors (TLRs) recognize a wide range of molecular patterns characteristic of pathogens (Akira et al., 2006; Wang et al., 2016). Vertebrate lineage-specific diversification of the TLRs has been associated with a co-evolutionary host-pathogen arms race (Khan et al., 2019; G. Liu et al., 2020). However, some TLR variants may result in susceptibility to infectious and inflammatory diseases (Netea et al., 2012; Skevaki et al., 2015). Both the innate and adaptative immune genes are influenced by the pathogens encountered by the host and by host demographic history and evolution.

The proboscideans were once among the most diverse and widespread megafaunal distributed across Africa, Eurasia and the Americas. However, the late Quaternary extinctions (3 Ma- 2.4 Ma) resulted in progressive megafaunal extinctions (Cantalapiedra et al., 2021; Stuart, 2015) culminating in the Pleistocene-Holocene transition (Elias & Schreve, 2013) when many of the world's largest herbivores became extinct (Ripple et al., 2015). Proboscideans were reduced to two genera with three living species.

The African elephants (Loxodonta) are restricted to Africa and are represented by the African savanna elephant, *Loxodonta africana* and the African forest elephant,

Loxodonta cyclotis. In Asia the Asian elephant, *Elephas maximus* (Shoshani, 1998) is the only other remaining post Pleistocene elephant lineage. Currently, the International Union for Conservation of Nature (IUCN) Red List classifies elephants as endangered or critically endangered. African savanna elephants have suffered a remarkable decline as a result of poaching during the 20th century (Lee & Graham, 2006). Habitat fragmentation has led to the loss of populations of forest (Maisels et al., 2013) and Asian elephants (Goswami et al., 2014; Leimgruber et al., 2003).

The effects of this long and complex demographic history on the immunogenetics of elephants is unclear. Previous research on MHC diversity in 30 African savanna and three Asian elephants on the DQA locus, demonstrated moderate diversity with balancing selection detected (Archie et al., 2010). Other than the DQA locus, MHC and TLR evolution has not been characterized in elephants.

In this study we implemented a multilocus genotyping method to measure the immunogenetic diversity of TLRs and MHC in populations of the three extant elephant species comparing immunogenetic evolutionary patterns with those of neutrally evolving markers. Our results suggest that although there is balancing selection, particularly in the MHC of elephants, the general loss of heterozygosity in shrinking elephant populations is leading to a general loss of elephant immune gene variability.

3.3. METHODS

3.3.1. A Multilocus genotyping method for genetic diversity analysis

To measure genetic diversity of elephants, we developed a multilocus genotyping method designed for amplifying long-fragment PCR products coupled with the PacBio high throughput sequencing platform. The assay covered an extensive set of genetic markers, including three genes (*BGN*, *PHKA2*, and *PLP*) on the elephant X-chromosome and *AMELY* encoded on the elephant Y-chromosome (Roca et al., 2005) (Table 3.1). We designed elephant specific primers for nine mammalian neutral intron markers (Igea et al., 2010) (Table 3.1). We included TLRs 1 to 13 to cover genes

involved in innate Immunity and the MHC (class I and II) to cover genes representing adaptive immunity. TLR primers were designed to flank the coding sequence and for MHC class I (HLA I) and MHC class II (*DQA*, *DQB*, *DRA*, and *DRB*), primes were designed to amplify the antigen-binding-recognition region. DQA primers only covered the transmembrane domain (exon 5). Additionally, to amplify the complete mitochondrial genome, primer pairs for eight overlapping PCR fragments were designed.

Table 3.1. Genetic markers, PCR fragments, and primers used in this study.

	Genetic Markers	PCR fragment	Amplicon	Primers	Sequence 5'-3'	Reference
Mitochondrial genome	Fragment 1 (srRNA -ND1)	MtGenE.F1	2,485 bp	mtGenE.F1-F	GCGGCCATACGATTAGTCCA	This study
				mtGenE.F1-R	GCCTAAGGCCTTTCGTTCAACT	
	Fragment 2 (tRNA-leu-COX1)	MtGenE.F2	2,972 bp	mtGenE.F2-F	CGGAGGTTCAACTCCTCTTCT	This study
				mtGenE.F2-R	TATACGGTCCAACCAGTGCCT	
	Fragment 3 (tRNA-Trp- COX2)	MtGenE.F3	2,585 bp	mtGenEp.F3-F	ACCAAGAGCCTTCAAAGCCC	This study
				mtGenE.F3-R	CGTCCTGGAATTGCATCTGT	
	Fragment 4 (COX2- ND4L)	MtGenE.F4	2,842 bp	mtGenE.F4-F	TTAATTGCCCTGCCCTCT	This study
				mtGenE.F4-R	GCTTCACAGGCTGCGAATACT	
	Fragment 5 (ND3- ND5)	MtGenE.F5	2,652 bp	mtGenE.F5-F	GAATGCGGCTTTGATCCA	This study
				mtGenE.F5-R	TGTGTTAGCGTCTGTTCGTC	
	Fragment 6 (tRNA- ND6)	MtGenE.F6	2,338 bp	mtGenE.F6-F	GCTACCCATTGGTCTTAGGCA	This study
				mtGenE.F6-R	ATGAGTGTGCTTTATGTGGTAGGT	
	Fragment 7 (ND6- Dloop)	MtGenE.F7	2,440 bp	mtGenE.F7-F	TCACCCAGCCATAGCCAAAA	This study
				mtGenE.F7-R	ATGTCCTCCGAGCATTGACT	
	Fragment 8 (Dloop- srRNA)	MtGenE.F8	1,338 bp	mtGenE.F8-F	TTCAGCTATGGCCGTCTGAG	This study
				mtGenE.F8-R	AGGGCTAGGCATAGTGAGGT	
Nuclear markers	Acyl-coenzyme A oxidase 2, peroxisomal (ACOX2)	ACOX2	510 bp	ACOX2-F	GGGCTCAGATGAGCAGATTG	This study
				ACOX2-R	GTCTCCAAGCCCTGAAGGTA	
	Acid phosphatase 4 (ACP4)	ACP4	480 bp	ACP4-F	ATTTTGACCGGACACTGGAG	This study
				ACP4-R	TCTCGAAGCAGCTCATGGTA	

	Genetic Markers	PCR fragment	Amplicon	Primers	Sequence 5'-3'	Reference
	Calcium-regulated heat stable protein 1 (CARHSP1)	CARHSP1	905 bp	CARHSP1-F	CTCTTCGTGGCAATGTGGT	This study
				CARHSP1-R	CCTTGGATCGACAGAAGCAT	
	COP9 signalosome complex subunit 7a (COPS7A)	COPS7A	840 bp	COPSA7A-F	CTGTGTACGCTGACGTGCTT	This study
				COPSA7A-R	GCTTACTTGCTCCTCGATGC	
	Short-chain dehydrogenase/reductase 3 (DHRS3)	DHRS3	630 bp	DHRS3-F	CCCTCCTTAAGTCCCAGCAT	This study
				DHRS3-R	CTCATGCCCTGGAACATCTC	
	jumonji domain containing protein 5 (JMJD5)	JMJD5	1,345 bp	JMJD5-F	GCCATGCATGAAGAAGTGG	This study
				JMJD5-R	TTTGCTGATGAACTCGCTGA	
	LanC-like protein 1 (LANCL1)	LANCL1	762 bp	LANCL1-F	TGAAATGCTGTACGGACGAA	This study
				LANCL1-R	AATTCCTGCTAGGCCATGAG	
	Leucine zipper domain protein (ROGDI)	ROGDI	481 bp	ROGDI-F	ATTGCTGCTAGTGGCCTCAC	This study
				ROGDI-R	TGGTACATGGTCAGGCAGAG	
	Amino acid transporter (SLC38A7)	SLC38A7	1,048 bp	SLC38A7-F	ATCGGCAAGGTCATCTCAGT	This study
				SLC38A7-R	TGGCTTGACTTCCTCCATCT	
	SET and MYND domain-containing protein 4 (SMYD4)	SMYD4	818 bp	SMYD4-F	GTCAGCCTCCTGAACCATTC	This study
				SMYD4-R	CTCAGCTTCTGCTGCCTTTC	
Tool-like receptors	Toll-like receptor 1 (TLR1)	TLR1	2,780 bp	TLR1-E-F	TTCCCCAGGATCTGTATCTGC	This study
				TLR1-E-R	GTGTGGAGTTTTCTAACATTGACC	
	Toll-like receptor 2 (TLR2)	TLR2	3,030 bp	TLR2-E-F	GGTGCAAGGCAGGTTGGTGA	This study
				TLR2-E-R	TAAAGACCAGAACTAGGCCCAAA	

	Genetic Markers	PCR fragment	Amplicon	Primers	Sequence 5'-3'	Reference
	Toll-like receptor 3 (TLR3)	TLR3a	2,900 bp	TLR3a-E-F	TGAAGACGGTGGAAGGAGTT	This study
				TLR3a-E-R	AGAAAGGGACTCCCCACACT	
		TLR3b	2,430 bp	TLR3b-E-F	AAGCTGTGCAAGGGTTATACG	This study
				TLR3b-E-R	TCTAATTTACCGGGGAGCTTT	
	Toll-like receptor 4 (TLR4)	TLR4a	400 bp	TLR4a-E-F	GCTGCCACTCTCACTTCCTC	This study
				TLR4a-E-R	GTCACACAGTCAGCCAGTCA	
		TLR4b	350 bp	TLR4b-E-F	ATTCTAGGAGGAAGGGAGTTGG	This study
				TLR4b-E-R	TGAGTGGAGATTGAGACTCTACC	
		TLR4c	2,650 bp	TLR4c-E-F	CTGGATCTGTGGGGACTTCT	This study
				TLR4c-E-R	GCAGCCTGCTTCTGTAAAGTG	
	Toll-like receptor 5 (TLR5)	TLR5	3,040 bp	TLR5-E-F	CCTGCGGGAACAGACACATCA	This study
				TLR5-E-R	TCTGTAACGTCTCCCAGGGTG	
	Toll-like receptor 6 T(LR6)	TLR6	2,570 bp	TLR6-E-F	TTGAGATAGCCACTGCAACA	This study
				TLR6-E-R	AGCCCAATTCTCCAGCTCAG	
	Toll-like receptor 7 (TLR7)	TLR7a	2,670 bp	TLR7a-E-F	GGCTCTTCCATTGTAGCAT	This study
				TLR7a-E-R	ACCAGACAAACCACACAGCA	
		TLR7b	1,829 bp	TLR7b-E-F	GAGACTCCTGTTTTGCCCTTT	This study
				TLR7b-E-R	TACAGGCAATCCCAGCTCTT	
	Toll-like receptor 8 (TLR8)	TLR8a	1,720 bp	TLR8a-E-F	CCGAGAAGCTGAACACATTG	This study
				TLR8a-E-R	CAGATTTAAACAGGCGATGTCA	
		TLR8b	2,280 bp	TLR8b-E-F	GGGCTATCAACCTGGGCATT	This study

	Genetic Markers	PCR fragment	Amplicon	Primers	Sequence 5'-3'	Reference
				TLR8b-E-R	CACATGGCAAGGAGCTGAGA	
	Toll-like receptor 9 (TLR9)	TLR9	3,305 bp	TLR9a-E-F	ATCCCTGGGAAGTGGAGTG	This study
				TLR9a-E-R	TGAGTGCCTGCAATATGCCA	
	Toll-like receptor 10 (TLR10)	TLR10	2,740 bp	TLR10-E-F	AGGCAAAAGCCCACGTAAAGGA	This study
				TLR10-E-R	TGTCTGTAGGGCCCAATTTC	
	Toll-like receptor 11 (TLR11)	TLR11	2,938 bp	TLR11-E-F	GCTGGTATTGTTGGCTTTGG	This study
				TLR11-E-R	GCATTGCCTAGAGACCCATC	
	Toll-like receptor 12 (TLR12)	TLR12	2,810 bp	TLR12-E-F	CTCTCTGTGGACCAGGCTT	This study
				TLR12-E-R	TGCCAAGCTGGCTTGCTTTT	
	Toll-like receptor 13 (TLR13)	TLR13	2,340 bp	TLR13-E-F	AGTGAGCATCGCACATGTGA	This study
				TLR13-E-R	AGCAGACCTTCCCCAGATCA	
Major Histocompatibility Complex	HLA I (MHC Class I)	HLA I	2,020bp	HLA-IA-E-F	TCTCCCTAGAACCCCAGTACC	This study
				HLA-IA-E-R	GTGATCTCCGCMGGGTAGAA	
	DQA (MHC Class II)	DQA	670 bp	DQA-E-F	AGCCCCAGTACAGGAGGAAAGA	This study
				DQA-E-R	CCACAGATATAGGGCTTAGGATT	
	DQB (MHC Class II)	DQB	890 bp	DQB-E-F	GGGATTTCATGCGAGAATGTTCTG	This study
				DQB-E-R	ATCTCCACTCGCTGCCTCAGTCTT	
	DRA (MHC Class II)	DRA	1,210 bp	DRA-E-F	ACAGTGTCCTGAGAAACACATGCC	This study
				DRA-E-R	GTGGGCACATGACAGAAATTAGGC	
	DRB (MHC Class II)	DRB	1,200 bp	DRB-E-F	CCTTTGAGGTTTCCAGGAGT	This study
				DRB-E-R	TAGCAGAGGACTAGCACTACTAGG	

	Genetic Markers	PCR fragment	Amplicon	Primers	Sequence 5'-3'	Reference
Sexual markers	X-Linked Chromosome	BGN	900bp	BGN-F11	CCGGAACACGAACTGCAT	(Roca et al., 2005).
				BGN-R11	GCTAATTACCTGCCCCCTCATGTCT	
		PHK	1,002bp	PHK-F3	CGCCTATAAGCACAGGTATGAA	(Roca et al., 2005).
				PHK-R3	AGGTGACCAGCTGTCCTGTT	
		PLP	479bp	PLP-F3	CCAGGACTATGAGTATCTCATCAATGT	(Roca et al., 2005).
				PLP-R3	CTGACCCTTCAGAGATGCTACCT	
	Y-Linked Chromosome	AMELY	2,500bp	AMELY-F	AGCGTTTCTCAAATCGTTCAAT	(Roca et al., 2005).

3.3.2. Genome scaffold screening

Each target gene was extracted from the genomes of *Loxodonta africana* (Broad/loxAfr3), manatee *Trichechus manatus latirostris* (Broad v1.0/triMan1), and Rock hyrax *Procavia capensis* (Broad/proCap1), UCSC Genome Browser http://genome.ucsc.edu/, using BLAT (Kent, 2002) with the human orthologous used as query sequences. To amplify the complete mitogenome, whole mitochondrial genomes from African savanna elephant *L africana* (NC_000934), African forest elephant *L cyclotis* (NC_020759), woolly mammoth *Mammuthus primigenius* (NC_007596), and Columbian mammoth *Mammuthus columbi* (NC_015529) were downloaded from NCBI and aligned using Clustal W (Larkin et al., 2007) integrated in Geneious 9.0.5 https://www.geneious.com/.

3.3.3. Primer design

From aligned scaffolds, primers were designed based on highly conserved gene regions using Primer 3 (Untergasser et al., 2012) integrated in Geneious 9.0.5. During *in silico* primer design, target location, primer size (20-25 bp), annealing temperatures (52-60°C), melting temperatures, secondary structure and dimer formation risk were adjusted. Primers pairs were evaluated, adjusted manually and specificity PCR tested in silico using the African forest elephant genome http://genome.ucsc.edu/ before being synthesized and applied in the laboratory.

3.3.4. PCR Optimization

PCR standardization was performed using wild African elephant samples. Total genomic DNA from blood was extracted using a QIAamp DNA Mini kit (QIAGEN, Germany) and checked for quality and quantity on a Qubit fluorometer and frozen at -20 °C. Before multilocus analysis, all primer pairs were tested individually to achieve optimal PCR conditions e.g Magnesium chloride ($MgCl_2$) and DNA-template concentrations and balance annealing temperature for all the PCR reactions. Amplicon size was verified by electrophoresis on agarose gels (1-2 %). Purified products were sequenced in both directions by LGC Genomics (Berlin, Germany) and all sequences were confirmed by Blast searching the sequenced products. Validated primers for each of the targets are listed in Table 3.1.

Primers were synthesized with a dual conserved sequence on the 5´ ends. All forward primers had conserved sequence one (CS1), forward-CS1 ACACTGACGACATGGTTCTACA and all reverse primers had conserved sequence two (CS2), reverse-CS2 TACGGTAGCAGACTTGGTCT attached to the 5' end (Ison et al., 2016). PacBio Index sequences were synthesized with a corresponding complementary sequence for CS1 or CS2 on the 3' end to allow them to tag PCR products during the index PCR (Figure 3.1).

3.3.5. Multilocus-multiplex PCR

To avoid biased results favoring amplification of short-fragments, amplification targets were distributed into two short fragment pools (350-1,345 bp) and five large fragment pools (890-3,305 bp). The TLR3b fragment (2,430 bp) was amplified in a separate master-mix. For the mitochondrial genome, primers were pooled in two master-mixes avoiding amplification of short and overlapping fragments (Table 3.2). The length of the amplicons could be differentiated by electrophoresis and all fragments were amplified to a similar degree resulting in balanced multiplex PCR reactions. An overview of the method is presented in Figure 3.1.

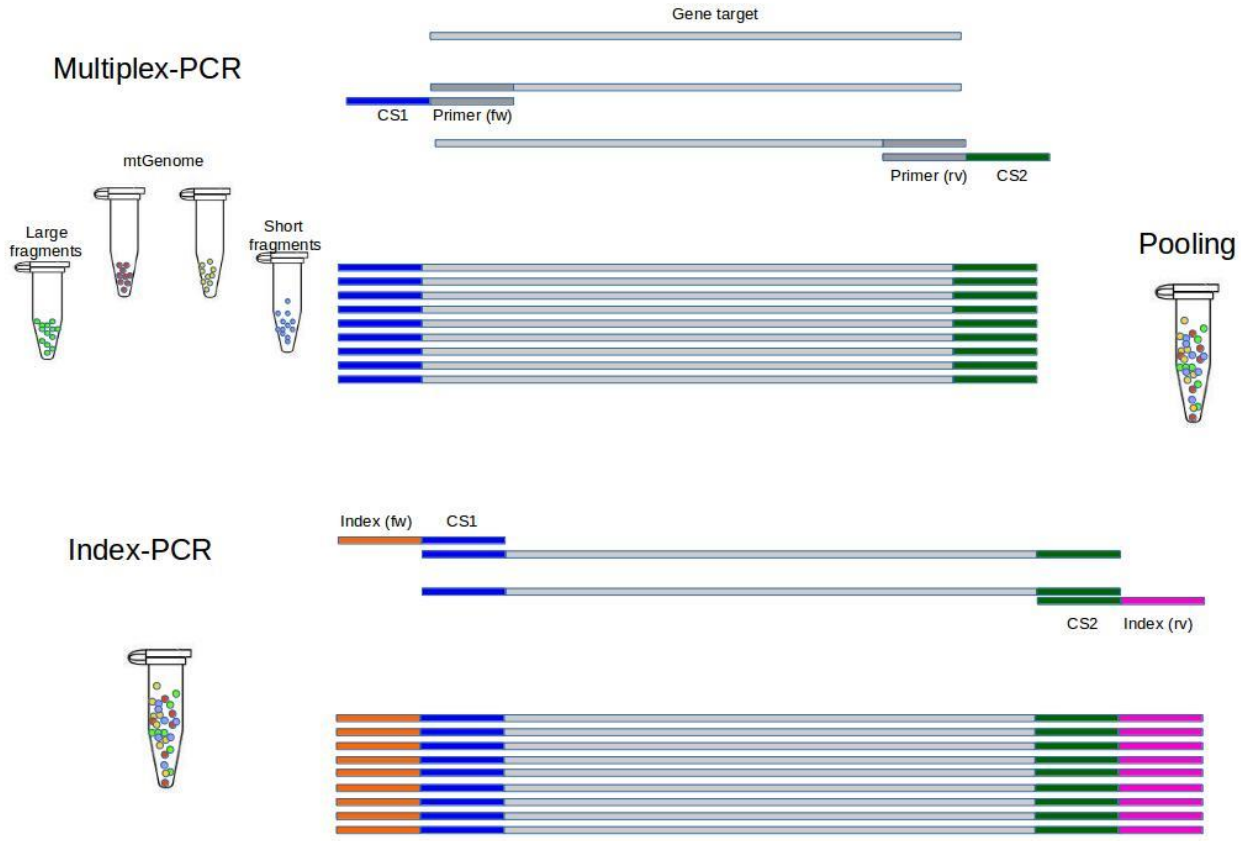


Figure 3.1. Multiplex-PCR approach.

Conserved sequence, CS1 (in blue), and CS2 (in green) were ligated to the 5´ end of forward and reverse primers respectively. Multiplex-PCRs were performed in pools according to size and avoiding short and unspecific fragments. Pools were tagged using an asymmetric barcoding combination.

Multiplex PCRs were performed in a final volume of 25 µl of 2X QIAGEN Multiplex-PCR Master Mix (QIAGEN, Germany), using stock primers pooled at 10 µM and 0.5 µl per reaction, and 1 µl of DNA template at 10 ng/µl. The thermal profile consisted of an initial activation step of 95ºC for 15 min, followed by 45 cycles of denaturation 94ºC, 30 s, annealing 59ºC 90 s, extension 72º C for 2 min and 10 s, and a final extension 72ºC for 10 min. For large fragments PCRs, the extension time was increased to 3 min. The fragment TLR3b did not work in a multiplex reaction, therefore it was amplified in a single PCR in a final volume of 25 µl of Multiplex PCR 1X MyTaq HS (BIOLINE, Germany), with a final primer concentration of 0.2 µM and 1 µl of DNA template. The thermal profile was an initial denaturation of 94ºC, 3 min, followed by 35 cycles of denaturation 95ºC, 15 s, annealing 52ºC 30 s, extension 72º C for 90 s, and a final extension 72ºC for 3 min.

PCRs were confirmed by electrophoresis and products were pooled in a single tube. One hundred µl of pooled sample was purified using the MSB Spin PCRapace kit (STRATEC Molecular GmbH) according to manufacturer recommendations. One µl of purified product was used as a template for index PCR. The index PCR was carried in a final volume of 20 µl using the Multiplex PCR 1X MyTaq HS (BIOLINE, Germany) with 0.2 uM primers final concentration and 1 µl of the purified product as template. The barcoding consisted of an asymmetric primer combination of the PacBio primers https://rawgit.com/PacificBiosciences/Bioinformatics-Training/master/scripts/generateAsymmetric_newBarcodes.html and was carried out using the following thermal profile: an initial denaturation at 95ºC, 10 min, followed by 30 cycles of denaturation at 95ºC, 15 s, at annealing 52ºC 30 s, extension at 72º C for 90 s, and a final extension at 72ºC for 3 min.

Table 3.2. Multiplex PCR pools.
Distribution of PCR mixes according to fragments, pool of targets with their respectively sequence length (bp) are indicated.

Genetic Markers	PCR pool	Targets
Short fragments	PCRS1	TLR4b (350 bp), TLR4a (400 bp), ACP4 (480 bp), ACOX2 (510 bp), LANCL1(762 bp), SMYD4 (818 bp), BGN (900 bp), CARHSP1 (905 bp), DRB (1,200 bp), DRA (1,210 bp).
	PCRS2	PLP (479 bp), ROGD1 (481 bp), DHRS (630 bp), DQA (670 bp), COPSA7 (840 bp), DQB (890 bp), PHK (1,002 bp), SLC38A7 (1,048 bp), JMJD5 (1,345 bp).
Large fragments	PCRL1	DQB (890 bp), DRB (1,200 bp), TLR13 (2,340 bp), TLR10 (2,740 bp).
	PCRL2	TLR7b (1,829 bp), HLAI (2,020 bp), TLR4c (2,650 bp), TLR12 (2,810 bp), TLR5 (3,040 bp).
	PCRL3	TLR8a (1,720 bp), TLR7a (2,670 bp), TLR11 (2,938 bp), TLR9 (3,305 bp).
	PCRL4	TLR8b (2,280 bp), AMELY (2,500 bp), TLR10 (2,740 bp), TLR3a (2,900 bp).
	PCRL5	TLR13 (2,340 bp), TLR6 (2,570 bp), TLR1 (2,780 bp), TLR2 (3,030 bp).
Single PCR	TLR3b	TLR3b (2,430 bp)
Mitochondrial genome	mtG1	PCR1 (2,485 bp), PCR3 (2,585 bp), PCR5 (2,652 bp), PCR7 (2,440 bp).
	mtG2	PCR2 (2,972 bp), PCR4 (2,842 bp), PCR6 (2,338 bp), PCR8 (1,338 bp).

3.3.6. Genotyping of elephant samples

A total of 263 elephants samples (87 savanna elephants (*L. africana*), 10 forest elephants (*L. cyclotis*) (Table 3.3) (Ishida et al., 2013), and 166 Asian elephants (*E. maximus*)) representing the three extant elephant species were examined in this study. Asian elephant samples come from North American collections and from European zoos. 63 Asian elephants were born in the wild, 88 in captivity, and 15 are of unknown origin (Table 3.4). Multilocus multiplex PCR experiments for samples from African elephants and Asian elephants from North American zoos were carried out in the Institute for Genomic Biology, University Illinois at Urbana-Champaign and for captive Asian elephants collected in Europe, the experiments were carried at the Leibniz Institute for Zoo and Wildlife Research in Berlin, Germany, following the protocol described above.

Table 3.3. African elephant samples.
Sample distribution for forest and savanna wild elephants.

Sample	Specie	Population	Country	Locallity	Loc	Reservoir
DS1506	*Loxodonta cyclotis*	Forest	Central African Republic	Dzanga Sangha	DS	Dzanga-Sangha National Park
DS1509	*Loxodonta cyclotis*	Forest	Central African Republic	Dzanga Sangha	DS	Dzanga-Sangha National Park
DS1543	*Loxodonta cyclotis*	Forest	Central African Republic	Dzanga Sangha	DS	Dzanga-Sangha National Park
DS1548	*Loxodonta cyclotis*	Forest	Central African Republic	Dzanga Sangha	DS	Dzanga-Sangha National Park
DS1557	*Loxodonta cyclotis*	Forest	Central African Republic	Dzanga Sangha	DS	Dzanga-Sangha National Park
GR0013	*Loxodonta cyclotis*	Forest	Democratic Republic of the Congo	Garamba	GR	Garamba National Park
GR0020	*Loxodonta cyclotis*	Forest	Democratic Republic of the Congo	Garamba	GR	Garamba National Park
GR0035	*Loxodonta cyclotis*	Forest	Democratic Republic of the Congo	Garamba	GR	Garamba National Park
LO3511	*Loxodonta cyclotis*	Forest	Gabon	Lope	LO	Lope Natonal Park
SL0001	*Loxodonta cyclotis*	Forest	Sierra Leone	Sierra Leone	SL	Outamba Kilimi National Park
WA4003	*Loxodonta africana*	Savanna	Cameroon	Waza	WA	No Park
WA4006	*Loxodonta africana*	Savanna	Cameroon	Waza	WA	No Park
WA4010	*Loxodonta africana*	Savanna	Cameroon	Waza	WA	No Park
WA4021	*Loxodonta africana*	Savanna	Cameroon	Waza	WA	No Park
BE3049	*Loxodonta africana*	Savanna	Cameroon	Benoue	BE	Benoue
MALI0003	*Loxodonta africana*	Savanna	Mali	Mali	Mali	Gourma Rharous
AB4542	*Loxodonta africana*	Savanna	Kenya	Aberdares	AB	Aberdares Forest
AM0003	*Loxodonta africana*	Savanna	Kenya	Amboseli	AM	Ambonseli National Park
AM0012	*Loxodonta africana*	Savanna	Kenya	Amboseli	AB	Ambonseli National Park

Sample	Specie	Population	Country	Locallity	Loc	Reservoir
BA0001	*Loxodonta africana*	Savanna	Ethiopia	Bale	BA	Bale Mountains National Park
BA0002	*Loxodonta africana*	Savanna	Ethiopia	Bale	BA	Bale Mountains National Park
BA0003	*Loxodonta africana*	Savanna	Ethiopia	Bale	BA	Bale Mountains National Park
BA0011	*Loxodonta africana*	Savanna	Ethiopia	Bale	BA	Bale Mountains National Park
KE4504	*Loxodonta africana*	Savanna	Kenya	Laikipia	KE	Laikipia Reservoir
KE4511	*Loxodonta africana*	Savanna	Kenya	Laikipia	KE	Laikipia Reservoir
KE4515	*Loxodonta africana*	Savanna	Kenya	Laikipia	KE	Laikipia Reservoir
KE4517	*Loxodonta africana*	Savanna	Kenya	Laikipia	KE	Laikipia Reservoir
NG2178	*Loxodonta africana*	Savanna	Tanzania	Ngorongoro	NG	Ngorongoro Conservation Area
NG2180	*Loxodonta africana*	Savanna	Tanzania	Ngorongoro	NG	Ngorongoro Conservation Area
NG2181	*Loxodonta africana*	Savanna	Tanzania	Ngorongoro	NG	Ngorongoro Conservation Area
NG2182	*Loxodonta africana*	Savanna	Tanzania	Ngorongoro	NG	Ngorongoro Conservation Area
NG2191	*Loxodonta africana*	Savanna	Tanzania	Ngorongoro	NG	Ngorongoro Conservation Area
SE2051	*Loxodonta africana*	Savanna	Tanzania	Serengeti	SE	Serengeti National Park
TA1138	*Loxodonta africana*	Savanna	Tanzania	Tarangire	TA	Tarangire National Park
TA1429	*Loxodonta africana*	Savanna	Tanzania	Tarangire	TA	Tarangire National Park
TA1434	*Loxodonta africana*	Savanna	Tanzania	Tarangire	TA	Tarangire National Park
TA1435	*Loxodonta africana*	Savanna	Tanzania	Tarangire	TA	Tarangire National Park
TA1439	*Loxodonta africana*	Savanna	Tanzania	Tarangire	TA	Tarangire National Park
TA1440	*Loxodonta africana*	Savanna	Tanzania	Tarangire	TA	Tarangire National Park
TA1441	*Loxodonta africana*	Savanna	Tanzania	Tarangire	TA	Tarangire National Park

Sample	Specie	Population	Country	Locallity	Loc	Reservoir
TA1443	*Loxodonta africana*	Savanna	Tanzania	Tarangire	TA	Tarangire National Park
TA1449	*Loxodonta africana*	Savanna	Tanzania	Tarangire	TA	Tarangire National Park
TA1450	*Loxodonta africana*	Savanna	Tanzania	Tarangire	TA	Tarangire National Park
TA1454	*Loxodonta africana*	Savanna	Tanzania	Tarangire	TA	Tarangire National Park
TA1457	*Loxodonta africana*	Savanna	Tanzania	Tarangire	TA	Tarangire National Park
TA1462	*Loxodonta africana*	Savanna	Tanzania	Tarangire	TA	Tarangire National Park
TA1463	*Loxodonta africana*	Savanna	Tanzania	Tarangire	TA	Tarangire National Park
TA1464	*Loxodonta africana*	Savanna	Tanzania	Tarangire	TA	Tarangire National Park
TA1465	*Loxodonta africana*	Savanna	Tanzania	Tarangire	TA	Tarangire National Park
TA1466	*Loxodonta africana*	Savanna	Tanzania	Tarangire	TA	Tarangire National Park
TA1467	*Loxodonta africana*	Savanna	Tanzania	Tarangire	TA	Tarangire National Park
CH0878	*Loxodonta africana*	Savanna	Botswana	Chobe	CH	Chobe National Park
CH0879	*Loxodonta africana*	Savanna	Botswana	Chobe	CH	Chobe National Park
CH0881	*Loxodonta africana*	Savanna	Botswana	Chobe	CH	Chobe National Park
CH0882	*Loxodonta africana*	Savanna	Botswana	Chobe	CH	Chobe National Park
CH0885	*Loxodonta africana*	Savanna	Botswana	Chobe	CH	Chobe National Park
CH0886	*Loxodonta africana*	Savanna	Botswana	Chobe	CH	Chobe National Park
CH0895	*Loxodonta africana*	Savanna	Botswana	Chobe	CH	Chobe National Park
CH0907	*Loxodonta africana*	Savanna	Botswana	Chobe	CH	Chobe National Park
CH0908	*Loxodonta africana*	Savanna	Botswana	Chobe	CH	Chobe National Park
CH0932	*Loxodonta africana*	Savanna	Botswana	Chobe	CH	Chobe National Park

Sample	Specie	Population	Country	Locallity	Loc	Reservoir
CH0934	*Loxodonta africana*	Savanna	Botswana	Chobe	CH	Chobe National Park
CH0935	*Loxodonta africana*	Savanna	Botswana	Chobe	CH	Chobe National Park
HW0059	*Loxodonta africana*	Savanna	Zimbabwe	Hwange	HW	Hwange National Park
HW0081	*Loxodonta africana*	Savanna	Zimbabwe	Hwange	HW	Hwange National Park
HW0083	*Loxodonta africana*	Savanna	Zimbabwe	Hwange	HW	Hwange National Park
HW0087	*Loxodonta africana*	Savanna	Zimbabwe	Hwange	HW	Hwange National Park
HW0093	*Loxodonta africana*	Savanna	Zimbabwe	Hwange	HW	Hwange National Park
HW0097	*Loxodonta africana*	Savanna	Botswana	Chobe	CH	Chobe National Park
HW0151	*Loxodonta africana*	Savanna	Zimbabwe	Hwange	HW	Hwange National Park
KR0007	*Loxodonta africana*	Savanna	South Africa	Kruger	KR	Kruger National Park
KR0008	*Loxodonta africana*	Savanna	South Africa	Kruger	KR	Kruger National Park
KR0027	*Loxodonta africana*	Savanna	South Africa	Kruger	KR	Kruger National Park
KR0057	*Loxodonta africana*	Savanna	South Africa	Kruger	KR	Kruger National Park
KR0071	*Loxodonta africana*	Savanna	South Africa	Kruger	KR	Kruger National Park
MA0803	*Loxodonta africana*	Savanna	Botswana	Mashatu	MA	No Park
MA0811	*Loxodonta africana*	Savanna	Botswana	Mashatu	MA	No Park
MA0815	*Loxodonta africana*	Savanna	Botswana	Mashatu	MA	No Park
MA0816	*Loxodonta africana*	Savanna	Botswana	Mashatu	MA	No Park
NA4658	*Loxodonta africana*	Savanna	Northern Namibia	Etosha	ET	Etosha National Park
NA4667	*Loxodonta africana*	Savanna	Northern Namibia	Etosha	ET	Etosha National Park
NA4671	*Loxodonta africana*	Savanna	Namibia	Etosha	ET	Etosha National Park

Sample	Specie	Population	Country	Locallity	Loc	Reservoir
NA4697	*Loxodonta africana*	Savanna	Northern Namibia	Etosha	ET	Etosha National Park
NA4703	*Loxodonta africana*	Savanna	Namibia	Etosha	ET	Etosha National Park
NA4704	*Loxodonta africana*	Savanna	Namibia	Etosha	ET	Etosha National Park
NA4707	*Loxodonta africana*	Savanna	Northern Namibia	Etosha	ET	Etosha National Park
NA4708	*Loxodonta africana*	Savanna	Northern Namibia	Etosha	ET	Etosha National Park
NA4710	*Loxodonta africana*	Savanna	Northern Namibia	Etosha	ET	Etosha National Park
NA5205	*Loxodonta africana*	Savanna	Northern Namibia	Etosha	ET	Etosha National Park
NA5206	*Loxodonta africana*	Savanna	Northern Namibia	Etosha	ET	Etosha National Park
NA5207	*Loxodonta africana*	Savanna	Northern Namibia	Etosha	ET	Etosha National Park
NA5208	*Loxodonta africana*	Savanna	Northern Namibia	Etosha	ET	Etosha National Park
SA0995	*Loxodonta africana*	Savanna	Botswana	Savuti	SA	Savuti National Park
SW0905	*Loxodonta africana*	Savanna	Zimbabwe	Sengwa	SW	No Park
SW0911	*Loxodonta africana*	Savanna	Zimbabwe	Sengwa	SW	No Park
ZZ0149	*Loxodonta africana*	Savanna	Zimbabwe	Zambezi	ZZ	Zambezi National Park
ZZ0157	*Loxodonta africana*	Savanna	Zimbabwe	Zambezi	ZZ	Zambezi National Park

Table 3.4. Asian elephant samples.
Sample distribution for Asian elephants.

Sample	Specie	Sex	Region	Born	Captive location
EMA-SN126	*Elephas maximus*	Male	Thailand	Wild	Houston Zoo
EMA-SN127	*Elephas maximus*	Female	Thailand	Wild	Houston Zoo
EMA-SN130	*Elephas maximus*	Female	Thailand	Wild	St. Louis Zoo
EMA-SN15	*Elephas maximus*	Female	Asia	Wild	Albuquerque Biological Park
EMA-SN159	*Elephas maximus*	Female	Asia	Wild	Tulsa Zoo & living museum
EMA-SN160	*Elephas maximus*	Male	Asia	Wild	Albuquerque Biological Park
EMA-SN167	*Elephas maximus*	Female	Thailand	Wild	Fort Worth Zoo
EMA-SN179	*Elephas maximus*	Female	Thailand	Wild	Fort Worth Zoo
EMA-SN182	*Elephas maximus*	Female	India	Wild	Ringling Bros.-FL
EMA-SN184	*Elephas maximus*	Female	India	Wild	Ringling Bros.-FL
EMA-SN187	*Elephas maximus*	Female	India	Wild	Ringling Bros.-FL
EMA-SN195	*Elephas maximus*	Female	Thailand	Wild	Ringling Bros.-FL
EMA-SN196	*Elephas maximus*	Female	Burma	Wild	Ringling Bros.-FL
EMA-SN198	*Elephas maximus*	Female	Burma	Wild	Ringling Bros.-FL
EMA-SN199	*Elephas maximus*	Female	Burma	Wild	Ringling Bros.-FL
EMA-SN203	*Elephas maximus*	Male	India	Wild	Fort Worth Zoo
EMA-SN216	*Elephas maximus*	Female	Asia	Wild	Columbus Zoo and Aquarium
EMA-SN234	*Elephas maximus*	Female	India	Wild	St. Louis Zoo
EMA-SN235	*Elephas maximus*	Female	Asia	Wild	St. Louis Zoo
EMA-SN239	*Elephas maximus*	Female	Asia	Wild	Tulsa Zoo & living museum
EMA-SN245	*Elephas maximus*	Female	Thailand	Wild	Oklahoma City Zoo
EMA-SN246	*Elephas maximus*	Female	Unknown	Unknown	Oklahoma City Zoo
EMA-SN247	*Elephas maximus*	Female	Thailand	Wild	St. Louis Zoo
EMA-SN249	*Elephas maximus*	Female	Asia	Wild	Ringling Bros.-FL
EMA-SN251	*Elephas maximus*	Female	Asia	Wild	Ringling Bros.-FL
EMA-SN252	*Elephas maximus*	Female	Asia	Wild	Ringling Bros.-FL
EMA-SN254	*Elephas maximus*	Female	Asia	Wild	Ringling Bros.-FL
EMA-SN255	*Elephas maximus*	Female	Asia	Wild	Ringling Bros.-FL
EMA-SN260	*Elephas maximus*	Male	Unknown	Captive	Ringling Bros.-FL
EMA-SN263	*Elephas maximus*	Male	Karnataka	Wild	Oklahoma City Zoo
EMA-SN270	*Elephas maximus*	Male	India	Wild	Oregon Zoo
EMA-SN276	*Elephas maximus*	Male	Unknown	Captive	Columbus Zoo and Aquarium

Sample	Specie	Sex	Region	Born	Captive location
EMA-SN288	*Elephas maximus*	Female	Malaysia	Wild	Houston Zoo
EMA-SN302	*Elephas maximus*	Female	India	Wild	Fort Worth Zoo
EMA-SN308	*Elephas maximus*	Female	Unknown	Captive	Houston Zoo
EMA-SN311	*Elephas maximus*	Female	Unknown	Captive	Houston Zoo
EMA-SN337	*Elephas maximus*	Female	Unknown	Captive	Albuquerque Biological Park
EMA-SN339	*Elephas maximus*	Female	Unknown	Captive	St. Louis Zoo
EMA-SN342	*Elephas maximus*	Female	Unknown	Captive	Ringling Bros.-FL
EMA-SN353	*Elephas maximus*	Female	Unknown	Captive	Columbus Zoo and Aquarium
EMA-SN365	*Elephas maximus*	Female	Unknown	Captive	Tulsa Zoo & living museum
EMA-SN379	*Elephas maximus*	Female	Unknown	Captive	Ringling Bros.-FL
EMA-SN380	*Elephas maximus*	Female	Unknown	Captive	Ringling Bros.-FL
EMA-SN382	*Elephas maximus*	Female	Unknown	Captive	Ringling Bros.-FL
EMA-SN385	*Elephas maximus*	Female	Unknown	Captive	Tulsa Zoo & living museum
EMA-SN386	*Elephas maximus*	Female	Unknown	Captive	St. Louis Zoo
EMA-SN411	*Elephas maximus*	Female	Asia	Wild	Ringling Bros.-FL
EMA-SN413	*Elephas maximus*	Female	Asia	Wild	Ringling Bros.-FL
EMA-SN416	*Elephas maximus*	Female	Asia	Wild	Ringling Bros.-FL
EMA-SN419	*Elephas maximus*	Female	Asia	Wild	Ringling Bros.-FL
EMA-SN423	*Elephas maximus*	Female	Asia	Wild	Ringling Bros.-FL
EMA-SN503	*Elephas maximus*	Male	Asia	Wild	Fort Worth Zoo
EMA-SN514	*Elephas maximus*	Male	Unknown	Captive	Albuquerque Biological Park
EMA-SN515	*Elephas maximus*	Female	Unknown	Captive	Fort Worth Zoo
EMA-SN516	*Elephas maximus*	Male	Unknown	Captive	Albuquerque Biological Park
EMA-SN519	*Elephas maximus*	Female	Malaysia	Wild	Oregon Zoo
EMA-SN523	*Elephas maximus*	Female	Asia	Wild	Albuquerque Biological Park
EMA-SN534	*Elephas maximus*	Female	Unknown	Captive	Ringling Bros.-FL
EMA-SN537	*Elephas maximus*	Female	Unknown	Captive	Ringling Bros.-FL
EMA-SN539	*Elephas maximus*	Female	Unknown	Captive	Ringling Bros.-FL
EMA-SN540	*Elephas maximus*	Female	Unknown	Captive	Ringling Bros.-FL
EMA-SN546	*Elephas maximus*	Male	Unknown	Captive	Ringling Bros.-FL
EMA-SN632	*Elephas maximus*	Male	Unknown	Captive	Houston Zoo
EMA-SN633	*Elephas maximus*	Female	Unknown	Captive	Ringling Bros.-FL
EMA-SN634	*Elephas maximus*	Male	Unknown	Captive	Ringling Bros.-FL
EMA-SN642	*Elephas maximus*	Female	Unknown	Captive	St. Louis Zoo

Sample	Specie	Sex	Region	Born	Captive location
EMA-SN645	*Elephas maximus*	Male	Unknown	Unknown	Houston Zoo
EMA-SN646	*Elephas maximus*	Female	Unknown	Captive	St. Louis Zoo
EMA-SN657	*Elephas maximus*	Male	Unknown	Captive	Columbus Zoo and Aquarium
EMA-SN671	*Elephas maximus*	Male	Unknown	Captive	Houston Zoo
EMA-SN70	*Elephas maximus*	Female	Unknown	Captive	Oregon Zoo
EMA-SN71	*Elephas maximus*	Female	Unknown	Captive	Oregon Zoo
EMA-SN735	*Elephas maximus*	Female	Unknown	Captive	Houston Zoo
EMA-SN736	*Elephas maximus*	Female	Unknown	Captive	Oklahoma City Zoo
EMA-SN760	*Elephas maximus*	Male	Unknown	Captive	Houston Zoo
EMA-SN761	*Elephas maximus*	Male	Unknown	Captive	Houston Zoo
EMA-SN762	*Elephas maximus*	Unknown	Unknown	Captive	Houston Zoo
EMA776212	*Elephas maximus*	Unknown	Unknown	Unknown	Oklahoma City Zoo
EMA776515	*Elephas maximus*	Unknown	Unknown	Unknown	Oklahoma City Zoo
2389	*Elephas maximus*	Female	Burma	Wild	Chester Zoo
24865	*Elephas maximus*	Female	Unknown	Wild	Chester Zoo
25897	*Elephas maximus*	Female	Unknown	Captive	Chester Zoo
28459	*Elephas maximus*	Male	Unknown	Captive	Chester Zoo
5677	*Elephas maximus*	Male	Unknown	Captive	Twycross Zoo
6482	*Elephas maximus*	Female	Unknown	Captive	Twycross Zoo
Acrra	*Elephas maximus*	Female	Unknown	Unknown	Le Pal
Angele	*Elephas maximus*	Female	Unknown	Captive	Budapest
Asha	*Elephas maximus*	Female	Unknown	Captive	Budapest
Assam	*Elephas maximus*	Male	Unknown	Captive	Budapest
Azizah	*Elephas maximus*	Female	Malaysia	Wild	ZSL Whipsnade
Brahma	*Elephas maximus*	Male	Unknown	Captive	Kolmården Zoo
Bua	*Elephas maximus*	Female	Unknown	Captive	Kolmården Zoo
Buba	*Elephas maximus*	Male	Unknown	Unknown	Selwo-Madrid Zoos
Califa	*Elephas maximus*	Female	Unknown	Captive	Hamburg Zoo
Chandrika	*Elephas maximus*	Female	Unknown	Captive	Woburn
Chang	*Elephas maximus*	Male	Unknown	Captive	Plankendael
Chikki	*Elephas maximus*	Female	Unknown	Wild	Selwo-Madrid Zoos
Damini	*Elephas maximus*	Female	Unknown	Captive	Woburn
Donkey	*Elephas maximus*	Female	Unknown	Wild	Kolmården Zoo
Drumbo1	*Elephas maximus*	Female	Unknown	Wild	Berlin Zoo

Sample	Specie	Sex	Region	Born	Captive location
Elmaoi	*Elephas maximus*	Unknown	Unknown	Unknown	Givskud
Emilia	*Elephas maximus*	Female	Unknown	Captive	ZSL Whipsnade
Emmet	*Elephas maximus*	Male	Unknown	Captive	ZSL Whipsnade
Felix	*Elephas maximus*	Male	Unknown	Captive	Hamburg Zoo
Geetha	*Elephas maximus*	Female	Unknown	Captive	ZSL Whipsnade
Ghandi	*Elephas maximus*	Male	Unknown	Captive	Copenhagen Zoo
Hari	*Elephas maximus*	Male	Unknown	Captive	Chester Zoo
Ida	*Elephas maximus*	Female	India	Wild	Copenhagen Zoo
Inda	*Elephas maximus*	Female	Unknown	Wild	Copenhagen Zoo
Jade	*Elephas maximus*	Female	Unknown	Captive	Le Pal
Jangoli	*Elephas maximus*	Female	Unknown	Wild	Selwo-Madrid Zoos
Jula	*Elephas maximus*	Female	Unknown	Captive	Copenhagen Zoo
Karishma	*Elephas maximus*	Female	Unknown	Captive	ZSL Whipsnade
Kavely	*Elephas maximus*	Female	India	Wild	Le Pal
Kaylee	*Elephas maximus*	Female	Myanmar	Wild	ZSL Whipsnade
Kewa	*Elephas maximus*	Female	Myanmar	Wild	Berlin-TP
KhaoSok	*Elephas maximus*	Male	Unknown	Captive	Copenhagen Zoo
Kungrao	*Elephas maximus*	Female	Thailand	Captive	Copenhagen Zoo
Louise	*Elephas maximus*	Female	Unknown	Wild	Berlin-TP
Lucha	*Elephas maximus*	Unknown	Unknown	Captive	ZSL Whipsnade
Ludra	*Elephas maximus*	Unknown	Unknown	Captive	Hamburg Zoo
Lyoti	*Elephas maximus*	Unknown	Unknown	Unknown	Berlin Zoo
Manari	*Elephas maximus*	Female	Unknown	Wild	Hamburg Zoo
Max	*Elephas maximus*	Male	Unknown	Captive	Le Pal
Max-62	*Elephas maximus*	Male	Unknown	Captive	ZSL Whipsnade
Maya	*Elephas maximus*	Female	Unknown	Wild	ZSL Whipsnade
MumbaArtis	*Elephas maximus*	Female	Unknown	Captive	Amsterdan
Mya	*Elephas maximus*	Unknown	Unknown	Unknown	ZSL Whipsnade
Ned	*Elephas maximus*	Male	Unknown	Captive	ZSL Whipsnade
Nikolai	*Elephas maximus*	Male	Unknown	Captive	Hamburg Zoo
Nina	*Elephas maximus*	Female	Unknown	Captive	Le Pal
Nomsai	*Elephas maximus*	Male	Unknown	Captive	Kolmården Zoo
PangPha1	*Elephas maximus*	Female	Unknown	Captive	Berlin Zoo
Pantha	*Elephas maximus*	Female	Unknown	Captive	Berlin-TP

Sample	Specie	Sex	Region	Born	Captive location
Pepa	*Elephas maximus*	Female	India	Wild	Selwo-Madrid Zoos
Pequena	*Elephas maximus*	Female	Unknown	Wild	Selwo-Madrid Zoos
Plaisak	*Elephas maximus*	Male	Thailand	Wild	Copenhagen Zoo
Punjab	*Elephas maximus*	Male	Unknown	Captive	Copenhagen Zoo
Raja	*Elephas maximus*	Male	Unknown	Captive	Woburn
Riddle	*Elephas maximus*	Male	Unknown	Captive	ZSL Whipsnade
Saba	*Elephas maximus*	Female	Unknown	Unknown	Kolmården Zoo
Sam	*Elephas maximus*	Male	Unknown	Captive	ZSL Whipsnade
Sammy	*Elephas maximus*	Female	India	Wild	Selwo-Madrid Zoos
Sanci	*Elephas maximus*	Unknown	Unknown	Unknown	Unknown
Saphira	*Elephas maximus*	Unknown	Unknown	Unknown	Hamburg Zoo
Sayang	*Elephas maximus*	Female	Unknown	Wild	Hamburg Zoo
Scott	*Elephas maximus*	Male	Unknown	Captive	ZSL Whipsnade
Shanti	*Elephas maximus*	Unknown	Unknown	Unknown	Hamburg Zoo
Sitara	*Elephas maximus*	Female	Unknown	Captive	Hamburg Zoo
Sundai	*Elephas maximus*	Female	Vietnam	Wild	Kolmården Zoo
Surin	*Elephas maximus*	Female	Thailand	Wild	Copenhagen Zoo
Surin (calf)	*Elephas maximus*	Unknown	Unknown	Captive	Copenhagen Zoo
Synneni	*Elephas maximus*	Female	Unknown	Wild	Givskud
Tahli	*Elephas maximus*	Female	Unknown	Captive	Woburn
Tanja	*Elephas maximus*	Female	Unknown	Wild	Berlin Zoo
Taru	*Elephas maximus*	Unknown	Unknown	Unknown	Hamburg Zoo
ThonaThai	*Elephas maximus*	Unknown	Unknown	Unknown	Amsterdan
Thuza	*Elephas maximus*	Female	Unknown	Captive	Berlin-TP
Tima	*Elephas maximus*	Female	Unknown	Captive	Selwo-Madrid Zoos
Tom	*Elephas maximus*	Male	Unknown	Captive	Le Pal
Tonsak	*Elephas maximus*	Male	Unknown	Captive	Copenhagen Zoo
Victor1	*Elephas maximus*	Male	Unknown	Captive	Berlin Zoo
W09M_083	*Elephas maximus*	Female	Unknown	Captive	ZSL Whipsnade
WinThida	*Elephas maximus*	Female	Unknown	Captive	Copenhagen Zoo
YuZin	*Elephas maximus*	Female	Unknown	Captive	Woburn
Yumi	*Elephas maximus*	Female	Unknown	Captive	Hamburg Zoo

3.3.7. PacBio sequencing

PacBio sequences were demultiplexed using the Pacific Biosciences SMRT Tools Lima command (PacBio SMRTLink v6.01) Circular consensus sequence (CCS) reads were generated using the SMRT Tools ccs command with minimum predicted accuracy of 90% and converted to fastq files using the SMRT Tools bam2fastq command. Reads from the same individual from different PacBio runs were combined. The reads were aligned to the African elephant (*L. africana*) reference genome (Loxafr3.0) using BWA mem (bwa mem -x pacbio) (Li & Durbin, 2010).

Longshot was used to call SNPs on the alignments (Edge & Bansal, 2019). SNPs were initially called for each individual and then the combined list of SNPs from all individuals was used to recall variants at variable positions. SNPs were combined using GATK (Genome Analysis Toolkit). SNPs were filtered using VCFtools (v0.1.15) (Danecek et al., 2011) to remove SNPs where the read depth was < 20x, and the number of missing individuals was < 60 % and PCR ambiguities were removed. We discarded 17 samples, as they did not pass the quality filters. In order to avoid bias in the genetic diversity estimation due to maternal exclusive inheritance of the mitochondria and paternal exclusive inheritance of the Y chromosome, mitochondrial PCR fragments and the male marker *AMELY* were removed from the dataset. High-quality SNP loci with high coverage across individuals were kept for subsequent analysis and vcf files from African elephants and Asian elephants were merged using bcftools (Li, 2011).

3.3.8. Genetic diversity analysis

The vcf SNP data was imported into R-4.1.3 (R Core Team, 2020) using the vcfR package (Knaus & Grünwald, 2017) and transformed to a genlight object using the Adegenet package (Jombart, 2008). Diversity indices such as observed heterozygosity (H_o), expected heterozygosity (H_e), and the inbreeding coefficient (F_{is}) were calculated using the function gl.report.heterozygosity correcting for sample size and missing values. To test the statistical significance of differences in heterozygosity between

elephants populations we implemented a pairwise comparisons using the gl.test.heterozygosity function with a re-randomization of 10,000 replicates. Deviations from Hardy-Weinberg equilibrium (HWE) were assessed by an exact test using gl.report.hwe together with gl.diagnostics.hwe in dartR (Gruber et al., 2018). The *L cyclotis* samples were excluded from the HWE analysis due to insufficient sample size (n individuals < 15). A number of private alleles (alleles unique to a population) and the count of fixed alleles were estimated by pairwise comparison between populations (Gruber et al., 2018). Genetic differentiation between populations was calculated using the fixation index (F_{st}) employing the gl.fst.pop function with a bootstrap of 1,000 in dartR (Gruber et al., 2018).

Assuming that diversity varies extensively across the genome, we computed additional statistics within and between the three elephant species to measure heterogeneity among the neutral markers, sex markers, MHC class II, and TLRs. We calculated the fixation index (F_{st}) (Weir & Cockerham, 1984) between each of three possible species pairs, the absolute genetic divergence (D_{xy}) and the nucleotide diversity (π) (Nei & Li, 1979), using a 100 bp sliding window. Specifically, we used pixy (Korunes & Samuk, 2021) to minimize bias generated by missing data. We used Pearson correlations to calculate the covariation for F_{st}, D_{xy}, and π for each group of genes within and between populations.

3.4. RESULTS

3.4.1. Multiplex Amplicon sequencing

We tested a novel genotyping method that combined long amplicons in a multiplex elephant specific PCR system coupled with PacBio sequencing technology. This multilocus genotyping method resulted in successful genotyping of 93.5 % (246 of 263 samples) and amplification of 31 nuclear genes fragments with a success rate of 83.8 % (31 of 37 PCRs). Ten neutral markers, three X linked markers, four MHC class II amplicons, and 14 TLR PCR fragments were represented among the amplicons. After

filtering a total of 2,729 loci, 5,421 alleles were obtained for 246 individuals, (83 *L africana*, 9 *L cyclotis*, and 154 *E. maximus*).

3.4.2. Genetic diversity among elephant species

We found a pronounced deficiency of heterozygosity in all elephant species indicated by values of observed and expected heterozygosity (Figure 3.2). All populations showed that observed heterozygosity (H_o) was lower than expected heterozygosity (H_e), *E. maximus* (H_o = 0.072 and H_e = 0.104), *L. africana* (H_o = 0.107 and H_e = 0.159), and *L. cyclotis* (H_o = 0.148 and H_e = 0.190). Interpopulation heterozygosity comparisons indicated significant differences in genetic diversity among *E. maximus* and *L. cyclotis* (diff = -0.086, p = 0.000) and among *L. africana* and *L. cyclotis* (diff = -0.031, p = 0.000), but was not significant among *E. maximus* and *L. africana* (diff = -0.055, p = 0.455). Inbreeding coefficients were similar across the three elephant populations (*E. maximus*: $_{Fis}$ = 0.306, *L. africana*: $_{Fis}$= 0.327, and *L. cyclotis*: $_{Fis}$ = 0.265) (Figure 3.2 and Table 3.5).

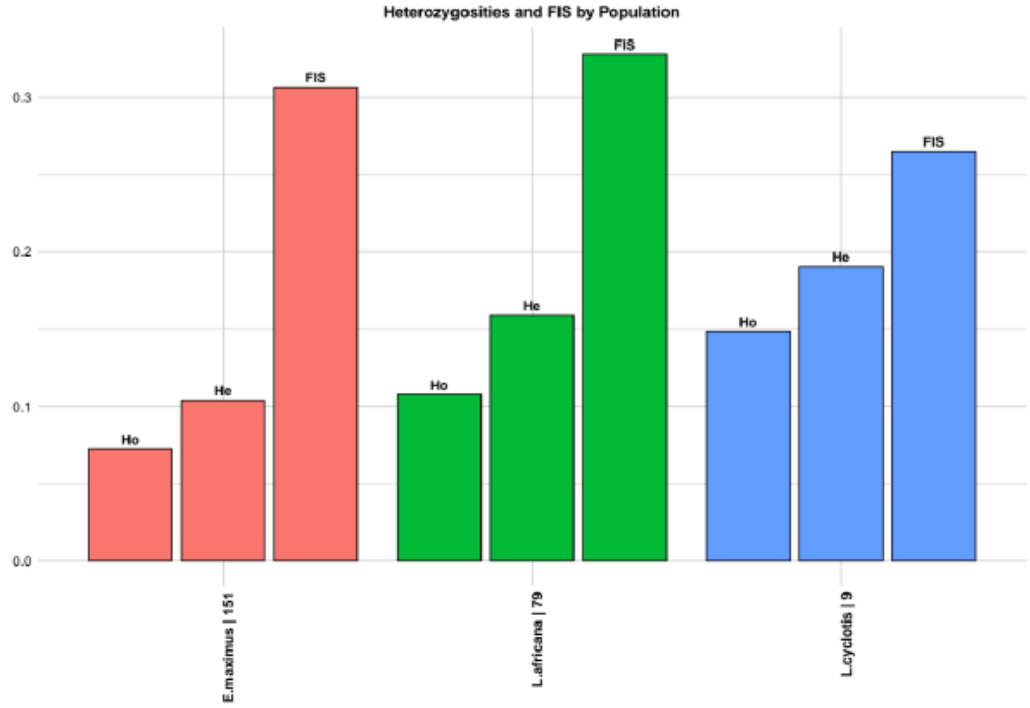


Figure 3.2. Elephants genetic diversity.
Barplots show the differences among observed (Ho) and expected heterozygosity (He) and Inbreeding coefficient ($_{Fis}$) across elephant populations. The values were calculated separately for each population and corrected for sample size.

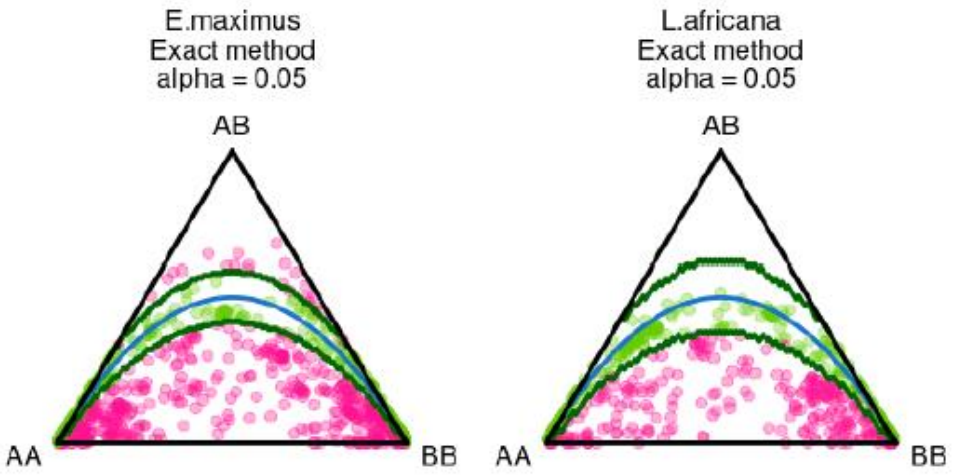
Table 3.5. Mean observed and expected heterozygosities and by population

Population	Ind	Loc	Ho	HoSD	He	HeSD	uHe	uHeSD	FIS
E Maximus	150.804	2380	0.072	0.108	0.104	0.138	0.104	0.139	0.306
L Africana	78.970	1002	0.107	0.134	0.159	0.172	0.160	0.173	0.327
L cyclotis	8.814	1002	0.148	0.172	0.190	0.171	0.201	0.181	0.265

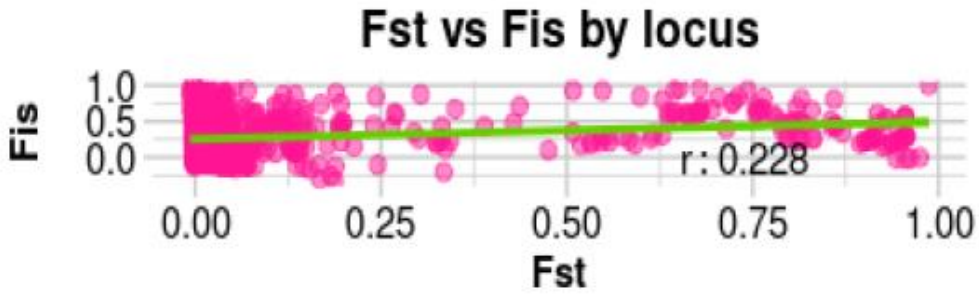
Ind, Individuals; Loc, locus; Ho, observed heterozygosity, HoSD, observed heterozygosity standard desviation; He, expected heterozygosity; HeSD, expected heterozygosity standard desviation; unbiased expected heterozygosity uHe; unbiased expected heterozygosity standard desviation; uHeSD; FIS, Inbreeding coefficient.

Ternary Hardy-Weinberg equilibrium plots showed a deviation from HWE proportions in both Asian elephants and savanna elephants. In Asian elephants, deviations from HWE were predominantly toward heterozygote deficiency and to a lesser extent toward heterozygote excess. In contrast, deviation from HWE in savanna elephants was exclusively a result of heterozygote deficiency (Figure 3.3A). A positive correlation between $_{Fst}$ and $_{Fis}$ (R = 0.228) obtained by HWE analysis was observed suggesting that heterozygosity deficiency could result from mating of close relatives, and to a lesser degree, selection (Figure 3.3B).

A



B



C

Population	Significant	Expected	Deficiency	Excess	Proportion Deficiency	Chisquare	pvalue
E Maximus	572	136.45	547	25	0.9562937	23431.592	0.00000e+00
E Africana	287	136.45	287	0	1.0000000	7632.677	3.286579e-77

Std Err Fis 3.366046e-06.
Std Err Fst 6.167369e-06.

Figure 3.3. Heterozygote deficiency.
Hardy-Weinberg Equilibrium (HWE) analysis of Asian (*E. maximus*) and African (*L. africana*) elephant populations. In panel A, ternary plots show loci following HWE in the middle curve (green dots) and loci departing from HWE (pink dots) in the lower curve (heterozygote deficiency) and in the upper curve (heterozygote excess). Alpha corresponds to levels of statistical significance. In panel B, a scatterplot linear regression shows the correlation between the fixation index ($_{Fst}$) and inbreeding coefficient ($_{Fis}$). The Spearman correlation coefficient is shown in panel C. A summary of the HWE tests is shown by each population including expected and observed values. The standard error for $_{Fis}$ and $_{Fst}$ are shown.

We quantified the number of private alleles by pairwise comparison across species. Private alleles are expected to increase with the divergence time between populations (Szpiech & Rosenberg, 2011). However, we found that there are less private alleles among African elephants and Asian elephants than between African elephants (savanna vs forest) (Table 3.6). The largest number of private alleles is carried by savanna-forest elephants (African region) (Table 3.6) which are phylogenetically closer. Suggesting that are higher trans-species alleles among Asian-African lineages and these variants have persisted for long-term due to ancient balancing selection. Intrapopulation differences showed that Asian elephants had 3 fold more private alleles than savanna elephants and 7 fold more than forest elephant populations (Table 3.6), suggesting population sub-structure in Asian elephants. Pairwise genetic differentiation ($_{Fst}$) showed significant levels of genetic differentiation between populations, ranging from intermediate ($_{Fst}$ = 0.151; *L. africana* vs *L. cyclotis*), high (F_{st} = 0.293; *L. cyclotis* vs *E. maximus*) to relatively highly ($_{Fst}$ = 0.412; *L. africana* vs *E. maximus*) supporting the expected taxonomy of these species.

Table 3.6. Private alleles and fixed allelic differences.
A pairwise comparison for all populations. Each row shows, for each pair of populations (Pop 1, Pop 2) the number of individuals (N1, N2) included by population. The number of fixed loci and private alleles differences for both populations are indicated.

Pop 1	Pop 2	N 1	N 2	Fixed	Priv 1	Priv 2	Total Priv
E Maximus	*L Africana*	154	83	0	65	22	87
E Maximus	*L cyclotis*	154	9	0	157	14	171
L Africana	*L cyclotis*	83	9	2	272	179	451

3.4.3. Differential patterns of diversity and divergences at neutral and immune loci

Based on comparison of differentiation ($_{Fst}$), divergence ($_{Dxy}$), and diversity (π) between 31 genetic regions among the three elephant species, we found that neutral genes were divergent but exhibited low diversity for all the species (Figure 3.4). The same pattern was observed using $_{Fst}$ and D_{xy} statistics. X-linked genes also exhibited this pattern (Figure 3.4). The $_{Fst}$, $_{Dxy}$, π means calculated per gene are displayed in Table 3.7 to Table 3.9 respectively.

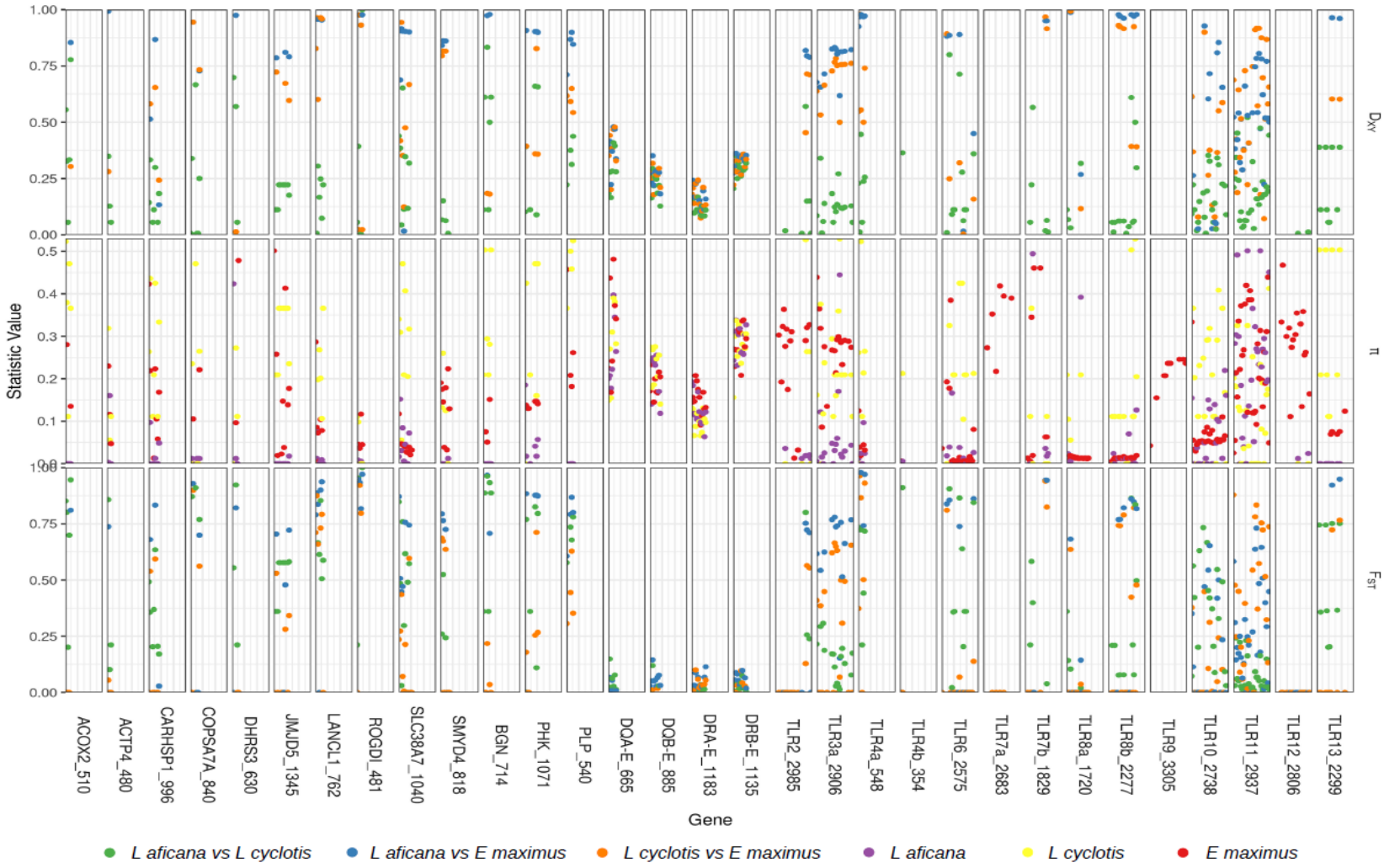


Figure 3.4. Landscape of three summary statistics compared within and between elephant species.

Nucleotide diversity (π), genetic differentiation (F_{st} and D_{xy}) for neutral markers, X-linked genes, MHC class II and TLRs are shown. The x-axis is shown in windows of 100 bp per gene.

Genetic differentiation among TLRs (F_{st} and D_{xy}) was moderate to high, indicating interlineage differentiation (Table 3.7 and Table 3.8). These patterns were broadly similar for both neutral loci and X-linked genes, consistent with divergence associated with elephant speciation. TLR genes exhibited an intermediate level of diversity comparable to neutral and sexual markers all much lower than the MHC region and without evidence for balancing selection. However, each TLR had a different pattern of divergence and diversity (Table 3.9 and Figure 3.5). TLR6, TLR8a, TLR8b, and TLR10 were more divergent in Asian than African elephants (Figure 3.5). However, diversity was considerably lower (mean diversity 0.012, 0,014, 0,012, and 0.058 respectively) compared to other species and to the other TLR loci, suggesting that the specific loci may have been under purifying selection.

Table 3.7. Evolutionary divergence fixation index (F_{st}).
Mean of F_{st} calculated by gene among populations. Number of SNPs included in the estimation are indicated. Not available (NA) values.

pop1	pop2	Gene	Mean F_{st}	SNPs
L.africana	*L.cyclotis*	*ACOX2_510*	0.85099598950276	9
L.africana	*E. maximus*	*ACOX2_510*	0.647394148122384	9
L.cyclotis	*E. maximus*	*ACOX2_510*	-0.005957509687722	9
L.africana	*L.cyclotis*	*ACTP4_480*	0.535141063731232	13
L.africana	*E. maximus*	*ACTP4_480*	0.44214445890856	13
L.cyclotis	*E. maximus*	*ACTP4_480*	0.039058887724631	13
L.africana	*L.cyclotis*	*CARHSP1_996*	0.43507516524507	16
L.africana	*E. maximus*	*CARHSP1_996*	0.660694463751772	16
L.cyclotis	*E. maximus*	*CARHSP1_996*	0.351080042354325	16
L.africana	*L.cyclotis*	*COPSA7A_840*	0.840681479858671	10
L.africana	*E. maximus*	*COPSA7A_840*	0.767065301232077	10
L.cyclotis	*E. maximus*	*COPSA7A_840*	0.455011692181621	10
L.africana	*L.cyclotis*	*DHRS3_630*	0.780094061984478	7
L.africana	*E. maximus*	*DHRS3_630*	0.45271891242311	7
L.cyclotis	*E. maximus*	*DHRS3_630*	-0.000510427714239	7
L.africana	*L.cyclotis*	*JMJD5_1345*	0.550388698666366	28
L.africana	*E. maximus*	*JMJD5_1345*	0.498099053663717	28
L.cyclotis	*E.maximus*	*JMJD5_1345*	0.294517593537	28
L.africana	*L.cyclotis*	*LANCL1_762*	0.718429846531389	21
L.africana	*E.maximus*	*LANCL1_762*	0.880337740943093	21

pop1	pop2	Gene	Mean $_{Fst}$	SNPs
L.cyclotis	*E.maximus*	*LANCL1_762*	0.670791792495204	21
L.africana	*L.cyclotis*	*ROGDI_481*	0.966455991910747	11
L.africana	*E.maximus*	*ROGDI_481*	0.915498612729566	11
L.cyclotis	*E.maximus*	*ROGDI_481*	0.685519173677546	11
L.africana	*L.cyclotis*	*SLC38A7_1040*	0.634705973232916	197
L.africana	*E.maximus*	*SLC38A7_1040*	0.521420464041233	197
L.cyclotis	*E.maximus*	*SLC38A7_1040*	0.239843508871799	197
L.africana	*L.cyclotis*	*SMYD4_818*	0.40546762047896	17
L.africana	*E.maximus*	*SMYD4_818*	0.698506178118132	17
L.cyclotis	*E.maximus*	*SMYD4_818*	0.608176869395709	17
L.africana	*L.cyclotis*	*BGN_714*	0.913112514064878	15
L.africana	*E.maximus*	*BGN_714*	0.749941042661158	15
L.cyclotis	*E.maximus*	*BGN_714*	0.024588590147677	15
L.africana	*L.cyclotis*	*PHK_1071*	0.685346823743537	10
L.africana	*E.maximus*	*PHK_1071*	0.845245173274426	10
L.cyclotis	*E.maximus*	*PHK_1071*	0.435049669669425	10
L.africana	*L.cyclotis*	*PLP_540*	0.71984425701613	6
L.africana	*E.maximus*	*PLP_540*	0.771002408877054	6
L.cyclotis	*E.maximus*	*PLP_540*	0.426646453730363	6
L.africana	*L.cyclotis*	*DQA-E_665*	0.054032643348356	50
L.africana	*E.maximus*	*DQA-E_665*	0.027718773777741	50
L.cyclotis	*E.maximus*	*DQA-E_665*	-0.014120759537336	50
L.africana	*L.cyclotis*	*DQB-E_885*	0.025921513690599	265
L.africana	*E.maximus*	*DQB-E_885*	0.071434385731156	265
L.cyclotis	*E.maximus*	*DQB-E_885*	-0.004796431618717	265
L.africana	*L.cyclotis*	*DRA-E_1183*	0.002797404428345	291
L.africana	*E.maximus*	*DRA-E_1183*	0.045437888272747	291
L.cyclotis	*E.maximus*	*DRA-E_1183*	0.033845364703332	291
L.africana	*L.cyclotis*	*DRB-E_1135*	0.02542301933269	178
L.africana	*E.maximus*	*DRB-E_1135*	0.062559677299092	178
L.cyclotis	*E.maximus*	*DRB-E_1135*	0.020374447457682	178
L.africana	*L.cyclotis*	*TLR2_2985*	0.598230955853731	31
L.africana	*E.maximus*	*TLR2_2985*	0.40204895636758	31
L.cyclotis	*E.maximus*	*TLR2_2985*	0.221805171517553	31
L.africana	*L.cyclotis*	*TLR3a_2906*	0.144809041507625	53
L.africana	*E.maximus*	*TLR3a_2906*	0.590466237894171	53
L.cyclotis	*E.maximus*	*TLR3a_2906*	0.441254537152012	53

pop1	pop2	Gene	Mean $_{Fst}$	SNPs
L.africana	*L.cyclotis*	*TLR4a_548*	0.593744960046733	14
L.africana	*E.maximus*	*TLR4a_548*	0.879138103679672	14
L.cyclotis	*E.maximus*	*TLR4a_548*	0.744736724977296	14
L.africana	*L.cyclotis*	*TLR4b_354*	0.911164429042668	4
L.africana	*E.maximus*	*TLR4b_354*	0	4
L.cyclotis	*E.maximus*	*TLR4b_354*	0	4
L.africana	*L.cyclotis*	*TLR6_2575*	0.747606017900977	196
L.africana	*E.maximus*	*TLR6_2575*	0.584128783986237	196
L.cyclotis	*E.maximus*	*TLR6_2575*	0.193830277030928	196
L.africana	*L.cyclotis*	*TLR7a_2683*	NA	8
L.africana	*E.maximus*	*TLR7a_2683*	0	8
L.cyclotis	*E.maximus*	*TLR7a_2683*	0	8
L.africana	*L.cyclotis*	*TLR7b_1829*	0.366954559476223	14
L.africana	*E.maximus*	*TLR7b_1829*	0.523464893268928	14
L.cyclotis	*E.maximus*	*TLR7b_1829*	0.48653324211706	14
L.africana	*L.cyclotis*	*TLR8a_1720*	0.062309010056535	331
L.africana	*E.maximus*	*TLR8a_1720*	0.175946280765451	331
L.cyclotis	*E.maximus*	*TLR8a_1720*	0.171520410584558	331
L.africana	*L.cyclotis*	*TLR8b_2277*	0.586429270408473	281
L.africana	*E.maximus*	*TLR8b_2277*	0.601736846842363	281
L.cyclotis	*E.maximus*	*TLR8b_2277*	0.429467232374831	281
L.africana	*L.cyclotis*	*TLR9_3305*	NA	14
L.africana	*E.maximus*	*TLR9_3305*	NA	14
L.cyclotis	*E.maximus*	*TLR9_3305*	NA	14
L.africana	*L.cyclotis*	*TLR10_2738*	0.338104241112499	404
L.africana	*E.maximus*	*TLR10_2738*	0.22435709570813	404
L.cyclotis	*E.maximus*	*TLR10_2738*	0.093454632156512	404
L.africana	*L.cyclotis*	*TLR11_2937*	0.056449266168391	132
L.africana	*E.maximus*	*TLR11_2937*	0.305692703463239	132
L.cyclotis	*E.maximus*	*TLR11_2937*	0.410983081434818	132
L.africana	*L.cyclotis*	*TLR12_2806*	-0.025910599083025	30
L.africana	*E.maximus*	*TLR12_2806*	0	30
L.cyclotis	*E.maximus*	*TLR12_2806*	0	30
L.africana	*L.cyclotis*	*TLR13_2299*	0.701844644201885	16
L.africana	*E.maximus*	*TLR13_2299*	0.870210564213107	16
L.cyclotis	*E.maximus*	*TLR13_2299*	0.315140317023572	16

Table 3.8. Absolute genetic divergence (D_{xy}).
Mean of D_{xy} calculated by gene among populations. Number of sites included in the estimation are indicated. Not available (NA) values.

pop1	pop2	Gene	Mean D_{xy}	Sites
L.africana	*L.cyclotis*	*ACOX2_510*	0.38878842676311	7
L.africana	*E.maximus*	*ACOX2_510*	0.853896103896104	1
L.cyclotis	*E.maximus*	*ACOX2_510*	0.303391053391053	1
L.africana	*L.cyclotis*	*ACTP4_480*	0.217536813922356	9
L.africana	*E.maximus*	*ACTP4_480*	0.993506493506493	1
L.cyclotis	*E.maximus*	*ACTP4_480*	0.280663780663781	1
L.africana	*L.cyclotis*	*CARHSP1_996*	0.202034624308406	9
L.africana	*E.maximus*	*CARHSP1_996*	0.615935779816514	5
L.cyclotis	*E.maximus*	*CARHSP1_996*	0.593755790253845	5
L.africana	*L.cyclotis*	*COPSA7A_840*	0.207050423917894	9
L.africana	*E.maximus*	*COPSA7A_840*	0.80163224961932	3
L.cyclotis	*E.maximus*	*COPSA7A_840*	0.804878048780488	3
L.africana	*L.cyclotis*	*DHRS3_630*	0.492771084337349	5
L.africana	*E.maximus*	*DHRS3_630*	0.975194897236003	1
L.cyclotis	*E.maximus*	*DHRS3_630*	0.013071895424837	1
L.africana	*L.cyclotis*	*JMJD5_1345*	0.183830560336584	14
L.africana	*E.maximus*	*JMJD5_1345*	0.796153846153846	3
L.cyclotis	*E.maximus*	*JMJD5_1345*	0.662535612535613	3
L.africana	*L.cyclotis*	*LANCL1_762*	0.184304532261665	18
L.africana	*E.maximus*	*LANCL1_762*	0.943434307823834	9
L.cyclotis	*E.maximus*	*LANCL1_762*	0.867633910428351	9
L.africana	*L.cyclotis*	*ROGDI_481*	0.280036813922356	8
L.africana	*E.maximus*	*ROGDI_481*	0.965630114566285	4
L.cyclotis	*E.maximus*	*ROGDI_481*	0.489361702127659	4
L.africana	*L.cyclotis*	*SLC38A7_1040*	0.29989604989605	19
L.africana	*E.maximus*	*SLC38A7_1040*	0.777332419541304	12
L.cyclotis	*E.maximus*	*SLC38A7_1040*	0.477821708778649	12
L.africana	*L.cyclotis*	*SMYD4_818*	0.083878565820466	6
L.africana	*E.maximus*	*SMYD4_818*	0.851303504578052	4
L.cyclotis	*E.maximus*	*SMYD4_818*	0.806173647469459	4
L.africana	*L.cyclotis*	*BGN_714*	0.488888888888889	10
L.africana	*E.maximus*	*BGN_714*	0.977272727272727	3
L.cyclotis	*E.maximus*	*BGN_714*	0.181818181818182	3
L.africana	*L.cyclotis*	*PHK_1071*	0.298319327731092	8

pop1	pop2	Gene	Mean D_{xy}	Sites
L.africana	*E.maximus*	*PHK_1071*	0.90335485682201	4
L.cyclotis	*E.maximus*	*PHK_1071*	0.487663693300436	4
L.africana	*L.cyclotis*	*PLP_540*	0.3270911360799	4
L.africana	*E.maximus*	*PLP_540*	0.829229015954811	4
L.cyclotis	*E.maximus*	*PLP_540*	0.600652173913043	4
L.africana	*L.cyclotis*	*DQA-E_665*	0.269869073419826	46
L.africana	*E.maximus*	*DQA-E_665*	0.363448191890905	26
L.cyclotis	*E.maximus*	*DQA-E_665*	0.360749255002128	26
L.africana	*L.cyclotis*	*DQB-E_885*	0.22514927730233	202
L.africana	*E.maximus*	*DQB-E_885*	0.268904662927958	128
L.cyclotis	*E.maximus*	*DQB-E_885*	0.262937848279735	128
L.africana	*L.cyclotis*	*DRA-E_1183*	0.113765286230953	216
L.africana	*E.maximus*	*DRA-E_1183*	0.166976269811158	168
L.cyclotis	*E.maximus*	*DRA-E_1183*	0.156604915587644	168
L.africana	*L.cyclotis*	*DRB-E_1135*	0.291420154359311	165
L.africana	*E.maximus*	*DRB-E_1135*	0.323266592402889	149
L.cyclotis	*E.maximus*	*DRB-E_1135*	0.308495362285093	149
L.africana	*L.cyclotis*	*TLR2_2985*	0.101361573373676	6
L.africana	*E.maximus*	*TLR2_2985*	0.801211662249148	3
L.cyclotis	*E.maximus*	*TLR2_2985*	0.625727989037341	3
L.africana	*L.cyclotis*	*TLR3a_2906*	0.156949806949807	31
L.africana	*E.maximus*	*TLR3a_2906*	0.765203849252415	17
L.cyclotis	*E.maximus*	*TLR3a_2906*	0.701492537313433	17
L.africana	*L.cyclotis*	*TLR4a_548*	0.204515029816235	11
L.africana	*E.maximus*	*TLR4a_548*	0.964584619831562	6
L.cyclotis	*E.maximus*	*TLR4a_548*	0.678117048346056	6
L.africana	*L.cyclotis*	*TLR4b_354*	0.364123159303882	2
L.africana	*E.maximus*	*TLR4b_354*	NA	0
L.cyclotis	*E.maximus*	*TLR4b_354*	NA	0
L.africana	*L.cyclotis*	*TLR6_2575*	0.22389851743579	14
L.africana	*E.maximus*	*TLR6_2575*	0.58991263022154	6
L.cyclotis	*E.maximus*	*TLR6_2575*	0.280660377358491	6
L.africana	*L.cyclotis*	*TLR7a_2683*	NA	0
L.africana	*E.maximus*	*TLR7a_2683*	NA	0
L.cyclotis	*E.maximus*	*TLR7a_2683*	NA	0
L.africana	*L.cyclotis*	*TLR7b_1829*	0.143199233716475	7
L.africana	*E.maximus*	*TLR7b_1829*	0.950531286894923	2

pop1	pop2	Gene	Mean D_{xy}	Sites
L.cyclotis	*E.maximus*	*TLR7b_1829*	0.941558441558442	2
L.africana	*L.cyclotis*	*TLR8a_1720*	0.104194556001785	6
L.africana	*E.maximus*	*TLR8a_1720*	0.627836019402285	2
L.cyclotis	*E.maximus*	*TLR8a_1720*	0.554834054834055	2
L.africana	*L.cyclotis*	*TLR8b_2277*	0.16474324566112	15
L.africana	*E.maximus*	*TLR8b_2277*	0.973144959606143	6
L.cyclotis	*E.maximus*	*TLR8b_2277*	0.745842371655821	6
L.africana	*L.cyclotis*	*TLR9_3305*	NA	0
L.africana	*E.maximus*	*TLR9_3305*	NA	0
L.cyclotis	*E.maximus*	*TLR9_3305*	NA	0
L.africana	*L.cyclotis*	*TLR10_2738*	0.187166497804796	41
L.africana	*E.maximus*	*TLR10_2738*	0.484597118633263	16
L.cyclotis	*E.maximus*	*TLR10_2738*	0.337757902706795	16
L.africana	*L.cyclotis*	*TLR11_2937*	0.203663640591197	86
L.africana	*E.maximus*	*TLR11_2937*	0.531990968776443	46
L.cyclotis	*E.maximus*	*TLR11_2937*	0.569436426612787	46
L.africana	*L.cyclotis*	*TLR12_2806*	0.009036144578313	2
L.africana	*E.maximus*	*TLR12_2806*	NA	0
L.cyclotis	*E.maximus*	*TLR12_2806*	NA	0
L.africana	*L.cyclotis*	*TLR13_2299*	0.238625017829126	10
L.africana	*E.maximus*	*TLR13_2299*	0.961873638344227	3
L.cyclotis	*E.maximus*	*TLR13_2299*	0.602638586298717	3

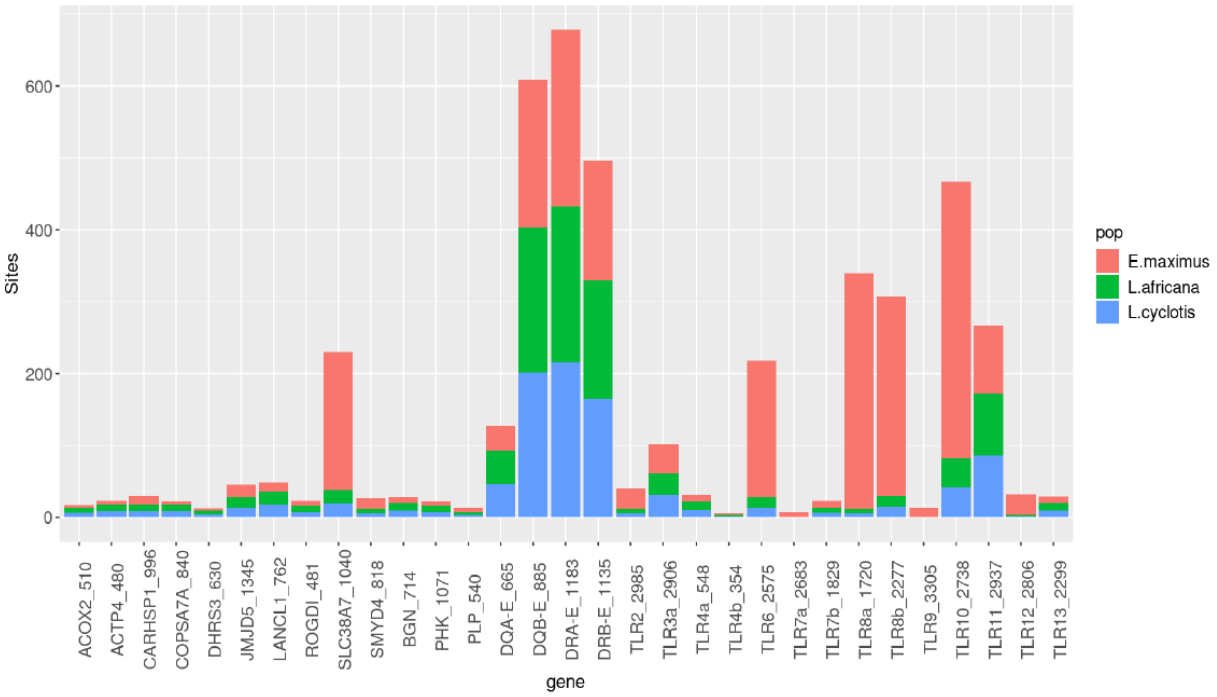


Figure 3.5. Polymorphic sites.
Distribution of sites used to calculated the diversity (π) per genetic marker across elephant populations are shown.

Table 3.9. Nucleotide diversity (π).
Mean of π calculated by gene in each population. Number of sites included in the estimation are indicated. Not available (NA) values.

pop	gene	Mean π	Sites
L.africana	*ACOX2_510*	0	7
L.cyclotis	*ACOX2_510*	0.372549019607843	7
E.maximus	*ACOX2_510*	0.181724007561437	3
L.africana	*ACTP4_480*	0.072556894243641	9
L.cyclotis	*ACTP4_480*	0.178649237472767	9
E.maximus	*ACTP4_480*	0.13395236685139	5
L.africana	*CARHSP1_996*	0.029550368686589	9
L.cyclotis	*CARHSP1_996*	0.297029702970297	9
E.maximus	*CARHSP1_996*	0.16070513185624	12
L.africana	*COPSA7A_840*	0.009338363555231	9
L.cyclotis	*COPSA7A_840*	0.163398692810458	9
E.maximus	*COPSA7A_840*	0.191630740239607	4
L.africana	*DHRS3_630*	0.091902154070829	5
L.cyclotis	*DHRS3_630*	0.18562091503268	5

pop	gene	Mean π	Sites
E.maximus	*DHRS3_630*	0.223665130897532	3
L.africana	*JMJD5_1345*	0.002550461586606	14
L.cyclotis	*JMJD5_1345*	0.302521008403361	14
E.maximus	*JMJD5_1345*	0.203843422459893	17
L.africana	*LANCL1_762*	0.021277027877862	18
L.cyclotis	*LANCL1_762*	0.17356572258533	18
E.maximus	*LANCL1_762*	0.095906506445659	12
L.africana	*ROGDI_481*	0.004499817451625	8
L.cyclotis	*ROGDI_481*	0.050653594771242	8
E.maximus	*ROGDI_481*	0.061712672848444	7
L.africana	*SLC38A7_1040*	0.051816172484257	19
L.cyclotis	*SLC38A7_1040*	0.296374516015487	19
E.maximus	*SLC38A7_1040*	0.039057879492528	192
L.africana	*SMYD4_818*	0.002287917024876	6
L.cyclotis	*SMYD4_818*	0.150594451783355	6
E.maximus	*SMYD4_818*	0.148676658291256	15
L.africana	*BGN_714*	0	10
L.cyclotis	*BGN_714*	0.334640522875817	10
E.maximus	*BGN_714*	0.110159735595414	8
L.africana	*PHK_1071*	0.045675003163089	8
L.cyclotis	*PHK_1071*	0.286314021830395	8
E.maximus	*PHK_1071*	0.138614700119333	6
L.africana	*PLP_540*	0	4
L.cyclotis	*PLP_540*	0.456140350877193	4
E.maximus	*PLP_540*	0.285533983074777	6
L.africana	*DQA-E_665*	0.245366662994309	46
L.cyclotis	*DQA-E_665*	0.260385438972163	46
E.maximus	*DQA-E_665*	0.309177198180053	35
L.africana	*DQB-E_885*	0.206238090804298	202
L.cyclotis	*DQB-E_885*	0.229007633587786	202
E.maximus	*DQB-E_885*	0.194875948738336	205
L.africana	*DRA-E_1183*	0.119792951890132	216
L.cyclotis	*DRA-E_1183*	0.104975246011857	216
E.maximus	*DRA-E_1183*	0.151203670442629	246
L.africana	*DRB-E_1135*	0.284755373693246	165
L.cyclotis	*DRB-E_1135*	0.281798375916023	165
E.maximus	*DRB-E_1135*	0.279786061827856	166

pop	gene	Mean π	Sites
L.africana	TLR2_2985	0.019873307663644	6
L.cyclotis	TLR2_2985	0.131147540983607	6
E.maximus	TLR2_2985	0.247201410107944	28
L.africana	TLR3a_2906	0.058217739313414	31
L.cyclotis	TLR3a_2906	0.228395061728395	31
E.maximus	TLR3a_2906	0.274010929848272	40
L.africana	TLR4a_548	0.037226592319692	11
L.cyclotis	TLR4a_548	0.221628045157457	11
E.maximus	TLR4a_548	0.061195762045599	9
L.africana	TLR4b_354	0.006024096385542	2
L.cyclotis	TLR4b_354	0.212418300653595	2
E.maximus	TLR4b_354	0.006493506493506	2
L.africana	TLR6_2575	0.025495170319307	14
L.cyclotis	TLR6_2575	0.189788053949904	14
E.maximus	TLR6_2575	0.012840293759403	190
L.africana	TLR7a_2683	NA	0
L.cyclotis	TLR7a_2683	NA	0
E.maximus	TLR7a_2683	0.331497336896257	8
L.africana	TLR7b_1829	0.084449840897672	7
L.cyclotis	TLR7b_1829	0.099906629318394	7
E.maximus	TLR7b_1829	0.19729401976959	9
L.africana	TLR8a_1720	0.075222100523305	6
L.cyclotis	TLR8a_1720	0.123093681917211	6
E.maximus	TLR8a_1720	0.014407942621042	327
L.africana	TLR8b_2277	0.036529412340172	15
L.cyclotis	TLR8b_2277	0.161655773420479	15
E.maximus	TLR8b_2277	0.012845941065206	278
L.africana	TLR9_3305	NA	0
L.cyclotis	TLR9_3305	NA	0
E.maximus	TLR9_3305	0.187480737038671	14
L.africana	TLR10_2738	0.082758844085714	41
L.cyclotis	TLR10_2738	0.199904351984696	41
E.maximus	TLR10_2738	0.058002427026507	385
L.africana	TLR11_2937	0.233133586537763	86
L.cyclotis	TLR11_2937	0.137737853762195	86
E.maximus	TLR11_2937	0.217592466413229	95
L.africana	TLR12_2806	0.017999269806499	2

pop	gene	Mean π	Sites
L.cyclotis	TLR12_2806	0	2
E.maximus	TLR12_2806	0.285643058745547	28
L.africana	TLR13_2299	0	10
L.cyclotis	TLR13_2299	0.336601307189542	10
E.maximus	TLR13_2299	0.084151612703817	9

The strongest F_{st} depression was observed for four genes (DQA, DQB, DRA, and DRB) located in the MHC class II region in contrast with neutral or X-linked markers (mean per gene from -0.004 to 0.062). These patterns were concordant with the divergence (Dxy) means. The MHC region exhibited the lowest divergence for the three species (Figure 3.4, Table 3.7 and Table 3.8). Comparing MHC diversity per gene measured by the π, no major variation was observed between the transmembrane domain region of DQA, and the antigen-binding-recognition region sequence of the DQB, DRA, and DRB genes. Despite the MHC class II genes having the highest number of polymorphic sites observed (Table 3.9 and Figure 3.5), evolution of the MHC genomic region across species is consistent with trans-species polymorphisms mediated by long-term balancing selection, maintaining shared variants among elephants lineages which separated millions of years ago.

Pearson correlations between summary statistics of neutral markers, X chromosome genes, MHC class II and TLRs of species and intra-species are shown in Figure 3.6. The MHC demonstrated a strong positive correlation between nucleotide diversity (π) and genetic divergence (D_{xy}) ranging from 0.887-0.996 for all interspecies combinations. By contrast, nucleotide diversity (π) and the fixation index (F_{st}) were not correlated between *L. africana* and *E. maximus* and was negatively correlated between *L. cyclotis* and *E. maximus*. This further supports that the MHC class II region polymorphisms are maintained by balancing selection. No correlations were found between the other genetic markers.

Mean nucleotide diversity per marker, estimated among elephant species showed much higher genetic variation for forest elephants, followed by Asian elephants, then

savanna elephants. Despite the savanna elephants being from different geographical regions, nucleotide diversity was low.

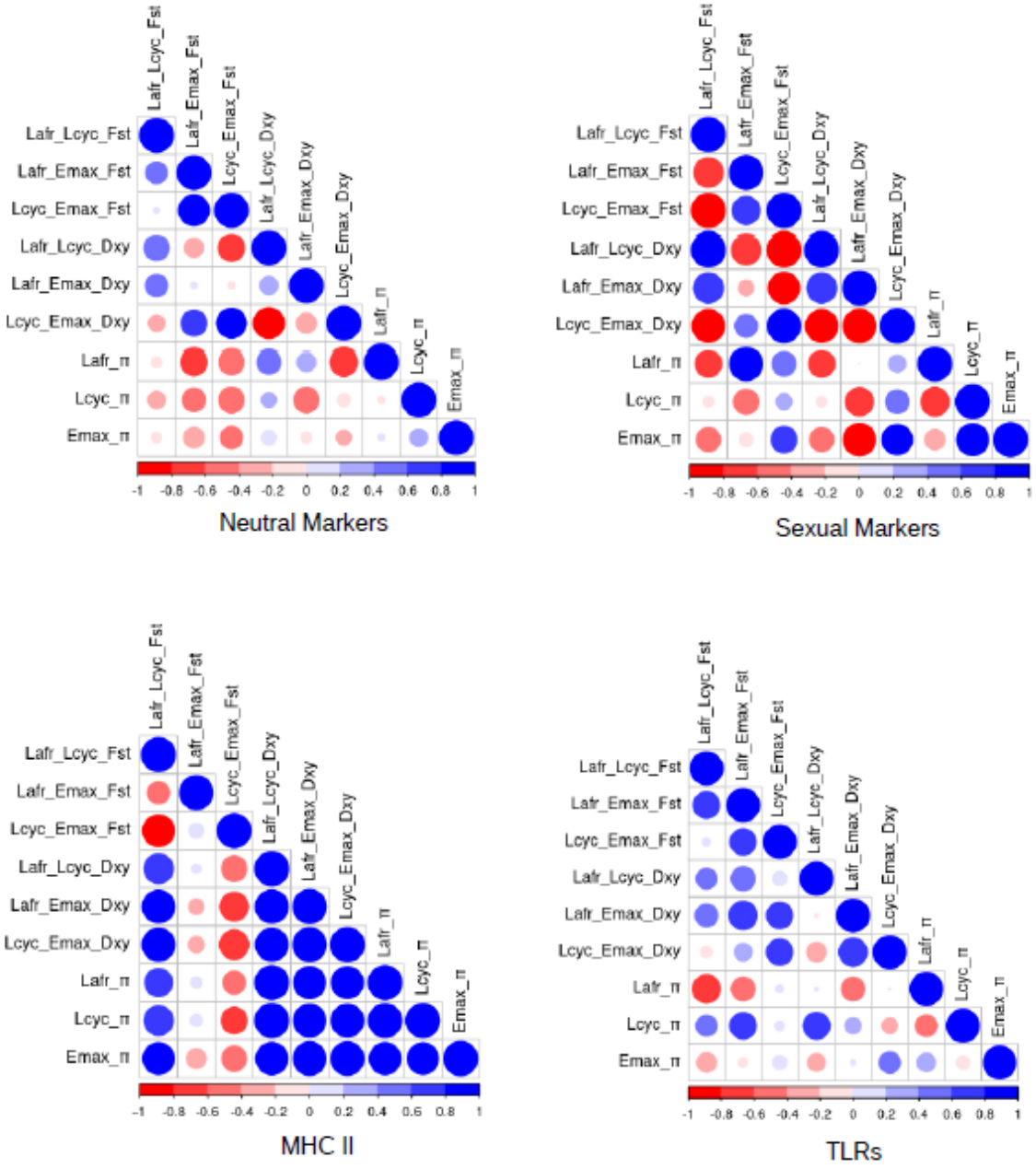


Figure 3.6. Summary statistics by marker across a 100 bp window per gene.
Inter-population and between population correlations of fixation index F_{st}, absolute divergence D_{xy} and nucleotide diversity π are shown. Circle proportions represent the Spearman's correlation coefficient. Blue indicates a positive correlation and red a negative one.

3.5. DISCUSSION

Multilocus genotyping developed in this study to measure immune genetic diversity in elephants was evaluated in populations of the three extant elephant species. Genotyping of 93.5 % of samples and an amplicon success rate of 83.8 % for 31 unlinked genes distributed across the genome of closely related species was achieved. Most of the markers, due to their polymorphic nature make them difficult to amplify in multiple species. However, the employed genotyping method was successful for a wide range of genetic markers and provided a high resolution for specie differentiation, that would have been hard to achieve by other means for example, with the use of microsatellites which are biased against coding regions (Väli et al., 2008).

We found the highest levels of genetic diversity in forest elephants relative to savanna and Asian elephants. This pattern is consistent with data from whole genome sequencing of Proboscideans (Palkopoulou et al., 2018). The significantly lower heterozygosity in savanna elephants than forest elephants is associated with a founder effect or a bottleneck in recent savanna elephant history (Comstock et al., 2002; Roca et al., 2001). Savanna elephants from the Kavango-Zambesi Transfrontier Conservation Area in southern Africa, a population without significant structure and with sustained gene flow, exhibited heterozygote deficiency for nuclear DNA markers (de Flamingh et al., 2015). In forest elephants, habitat fragmentation has resulted in increasing isolation by distance, reduction of gene flow and loss of genetic connectivity between elephants in the western and eastern regions, altering population structure and possible increasing the effects of genetic drift and inbreeding (Ishida et al., 2018). Over time the demographic challenges faced by African elephants will likely result in a continuous loss of diversity.

Asian elephants display reduction in genetic diversity compared to African elephants. They are subdivided into three subspecies (*Elephas maximus maximus*, *E. m. indicus*, and *E. m. sumatranus*). Molecular analyses based on mtDNA indicates that two allopatric clades, which probably diverged 1.35 Mya exist. Genetic diversity of Asian elephants has been shaped by geographic isolation and bottlenecks that occurred in the southern Indian population (Chakraborty et al., 2014; Gray et al., 2014; Vidya et

al., 2005). Asian elephant populations in China face a decrease of genetic diversity and scarcity of reproductive females making them prone to inbreeding depression (He et al., 2020; Sun et al., 2021).

The observed departure from HWE can be primarily attributed to heterozygote deficiency. In elephants, the matrilineal social structure and Wahlund effect (population substructure), in principle can lead to departure from HWE and generate heterozygosity deficiency (Garnier-Géré & Chikhi, 2013; Wittemyer et al., 2009). However, our analysis suggested that structure had marginal effects on deviations from HWE. By contrast, the F_{st} and F_{is} correlation does not support that null alleles or Wahlund effect explains the heterozygosity deficiency. In savanna elephants gene flow mediated by male elephant migration homogenizes populations (Roca et al., 2001). Although no HWE deviation has been detected in forest elephants (Gugala et al., 2016; Ishida et al., 2018), our results indicated they also experience heterozygosity deficiency and effects of inbreeding on genetic diversity. These patterns provide new evidence that the three elephant species are undergoing a decline in genetic diversity, and inbreeding may increase the genetic load (lethal mutations). Low heterozygosity and inbreeding may produce different effects in wild populations, compromising their long-term survival (Bosse & van Loon, 2022; Keller, 2002; O'Grady et al., 2006). Rhinoceroses also exhibit low heterozygosity where anthropogenic pressure in recent decades has accelerated inbreeding rates, leading to genetic diversity loss and declining population density (S. Liu et al., 2021; Sánchez-Barreiro et al., 2021). Low heterozygosity is implicated in fitness decrease in Iberian red deer (*Cervus elaphus hispanicus*) a fitness decrease where hunting pressure has reduced male antler size which affect mate selection (Pérez-González et al., 2010). Immunity has been compromised in East African Shorthorn Zebu (crossbred *Bos taurus* x *Bos indicus*) where inbreeding depression is associated with vulnerability to infectious disease (Murray et al., 2013). The Tasmania devil (*Sarcophilus harrisii*) has lost most of its immunogenetic diversity and is confronted with serious health challenges to the species (Morris et al., 2015). Low diversity and inbreeding may lead to accumulation of deleterious mutations, and if not purged quickly, may lead to severe inbreeding depression and rapid extinction (Kyriazis et al., 2021). Whether loss of heterozygosity

and diversity in elephants presents a current or future challenge to fitness in some or all extant species remains to be determined.

Quantification of $_{Fst}$ and $_{Dxy}$ for neutral markers showed high genetic divergence between the African elephant lineages but low diversity (measured by π). This analysis is consistent with the deep divergence between African elephant species revealed using nuclear introns (Roca et al., 2001). Generally, Intronic neutral markers are located in genome regions described as speciation islands where divergence between populations is high with patterns of low genetic diversity and without selection pressure (Wolf & Ellegren, 2017) consistent with our analysis.

The observed patterns of the X chromosome markers (*BGN*, *PHK*, and *PLP*) were similar to those of the neutral markers but with greater divergence and lower diversity. In savanna elephants the loss of diversity was almost complete (*BGN*, and *PLP* had a zero value of π), consistent with observations in North American captive African elephants where low haplotype diversity was detected using the same markers (Lei et al., 2009).

We obtained a comprehensive picture of genetic variability and diversity patterns in TLRs from the three elephant species. TLR nucleotide diversity were generally comparable to nuclear neutral and sex-linked genes levels exhibiting the same pattern of diversity among elephants. Despite the fact that we found higher overall levels of genetic diversity in Asian elephants, our results show a striking reduction in the genetic diversity in the TLR6, TLR8, and TLR10 genes suggesting that they are under strong purifying selection. This may suggest Asian elephants have adapted to pathogens not shared with African elephants. Asian elephants have a higher susceptibility to infections such as elephant endotheliotropic herpesvirus (EEHV) and tuberculosis, as well as a higher prevalence of malignant cancers (Tollis et al., 2021). In primates intra-species studies, purifying selection has been restricted to endosomal TLRs in humans and to endosomal and cell surface in chimpanzees and gorillas. However, the strongest effect occurred in African great apes, particularly in gorillas, where global nucleotide diversity has dropped dramatically (Quach et al., 2013; Wlasiuk & Nachman, 2010). TLRs recognize conserved molecular patterns and this might

constraint the evolution of the domain involved in pathogen recognition producing structural adaptations as has occurred throughout vertebrate TLR evolution (Botos et al., 2011; Roach et al., 2005; Wang et al., 2016). Our analyses indicate that TLRs are experiencing species specific molecular evolution in elephants. Differential selective pressures could be a result of pathogen exposures and environmental factors which are species specific.

In contrast to neutral and sexual markers, an accumulation of polymorphic sites in the MHC genes was observed, reflecting a higher diversity in this genomic region but with unusually low F_{st} values among the three elephant species. This signature of selection on the MHC class II indicates long-term balancing selection together with possible episodes of trans-species polymorphism (TSP). Balancing selection may maintain MHC variants by long-term overdominance, creating reservoirs of shared polymorphisms among populations with a resulting decrease in genetic differentiation, but which could be important in mediating subsequent adaptation (Brandt et al., 2018; de Filippo et al., 2016; Lenz, 2011). Distribution of private alleles indicates that there are more alleles shared between Asian elephants with African elephants than between African elephants (savanna-forest elephants), suggesting that old alleles have transcended species boundaries. Among savanna and forest elephants, the decline in common alleles is probably due to loss of diversity or local pathogen pressure, considering that private allele uniqueness may indicate population-specific selection (Kalinowski, 2004). TSP in the MHC DQA has been found between savanna and Asian elephants (Archie et al., 2010) and between elephants and mammoths (Pečnerová et al., 2016). MHC class II ancestral alleles retention has also occurred during primate diversification in macaques (*Macaca mulatta*) MHC-DQB genes (Yao et al., 2014, p. 1) and in lemurs MHC-DRB genes (Go et al., 2002). Our study suggests that balancing selection and TSP are not only acting on DQA but are occurring throughout the MHC II region (*DQA*, *DRA*, and *DRB*) in all three extant elephant species.

In a previous study, the peptide-binding region (PBR) of the DQA exhibited high diversity in elephants (Archie et al., 2010). In our analysis the DQA diversity is also high, reflected in the number of sites and higher π values, despite that we only sequenced the transmembrane domain region. However, our sample size was larger.

Similar patterns of balancing selection were found in the PBR of *DQA*, *DRA*, and *DRB* regions. The polymorphisms persistence and low $_{Fst}$ values among elephants MHC suggest that some MHC supertypes may be conserved in the elephant lineage. Supertypes are clusters of MHC alleles characterized by similar physicochemical properties in amino acids at the peptide-binding region (PBR) which are under positive selection and may potentially have unique immunological functions (Doytchinova & Flower, 2005; Lighten et al., 2017; Trachtenberg et al., 2003). It is thought supertypes play an important role in pathogen resistance (Trachtenberg et al., 2003), however their long-term persistence is controversial (Ejsmond et al., 2018). During a bottleneck, genetic drift may result in depletion of allelic diversity and balancing selection may protect advantageous alleles from extinction by retaining some MHC supertypes (Consuegra et al., 2013; Ejsmond et al., 2018; Lighten et al., 2017). Heterozygosity in the mammoth MHC DQA locus dropped drastically during Pleistocene-Holocene transition bottleneck coinciding with isolation on Wrangel Island where balancing selection failed to maintain MHC DQA diversity (Pečnerová et al., 2016). This is also the scenario for the endangered Indo-Pacific humpback dolphin (*Sousa chinensis*) where both microsatellites and MHC class II (*DQB* and *DRB*) heterozygosity deficiency and HWE deviation are observed (Zhang et al., 2016). Balancing selection might not be sufficient to mitigate the low level of genetic diversity in the MHC, increasing potential disease susceptibility which may have a major impact on this small population (Zhang et al., 2016). Whether extant elephants will suffer a similar fate will depend on the success of conservation efforts for all three species.

In summary, high levels of inbreeding and heterozygosity deficiency among elephant populations have resulted in a loss of genetic diversity. We found that neutral markers, X-linked, and TLRs reflected genetic differentiation between elephant species. TLRs Comparison revealed purifying selection of TLR6, TLR8, and TLR10 genes on Asian elephants, suggesting pathogen lineage-specific adaptation. A key novelty of this work is the evidence that balancing selection has maintained trans-specific MHC class II alleles among three elephant species. However, if inbreeding prevails over balancing selection, a decline in the MHC heterozygosity advantage may compromise the ability to eliminate pathogens and exacerbating the endangered situation, reducing the chances of long-term survival.

REFERENCES

Akira, S., Uematsu, S., & Takeuchi, O. (2006). Pathogen Recognition and Innate Immunity. *Cell*, *124*(4), 783–801. https://doi.org/10.1016/j.cell.2006.02.015

Archie, E. A., Henry, T., Maldonado, J. E., Moss, C. J., Poole, J. H., Pearson, V. R., Murray, S., Alberts, S. C., & Fleischer, R. C. (2010). Major histocompatibility complex variation and evolution at a single, expressed DQA locus in two genera of elephants. *Immunogenetics*, *62*(2), 85–100. https://doi.org/10.1007/s00251-009-0413-8

Barreiro, L. B., & Quintana-Murci, L. (2010). From evolutionary genetics to human immunology: How selection shapes host defence genes. *Nature Reviews Genetics*, *11*(1), 17–30. https://doi.org/10.1038/nrg2698

Bosse, M., & van Loon, S. (2022). Challenges in quantifying genome erosion for conservation. *Frontiers in Genetics*, *13*, 960958. https://doi.org/10.3389/fgene.2022.960958

Botos, I., Segal, D. M., & Davies, D. R. (2011). The Structural Biology of Toll-like Receptors. *Structure*, *19*(4), 447–459. https://doi.org/10.1016/j.str.2011.02.004

Brandt, D. Y. C., César, J., Goudet, J., & Meyer, D. (2018). The Effect of Balancing Selection on Population Differentiation: A Study with HLA Genes. *G3 Genes|Genomes|Genetics*, *8*(8), 2805–2815. https://doi.org/10.1534/g3.118.200367

Buchmann, K. (2014). Evolution of Innate Immunity: Clues from Invertebrates via Fish to Mammals. *Frontiers in Immunology*, *5*. https://doi.org/10.3389/fimmu.2014.00459

Cantalapiedra, J. L., Sanisidro, Ó., Zhang, H., Alberdi, M. T., Prado, J. L., Blanco, F., & Saarinen, J. (2021). The rise and fall of proboscidean ecological diversity. *Nature Ecology & Evolution*, *5*(9), 1266–1272. https://doi.org/10.1038/s41559-021-01498-w

Chakraborty, S., Boominathan, D., Desai, A. A., & Vidya, T. N. C. (2014). Using genetic analysis to estimate population size, sex ratio, and social organization in an Asian elephant population in conflict with humans in Alur, southern India. *Conservation Genetics*, *15*(4), 897–907. https://doi.org/10.1007/s10592-014-0587-y

Comstock, K. E., Georgiadis, N., Pecon-Slattery, J., Roca, A. L., Ostrander, E. A., O'Brien, S. J., & Wasser, S. K. (2002). Patterns of molecular genetic variation among African elephant populations. *Molecular Ecology*, *11*(12), 2489–2498. https://doi.org/10.1046/j.1365-294X.2002.01615.x

Consuegra, S., Ellison, A., Allainguillaume, J., Pachebat, J., Peat, K. M., & Wright, P. (2013). Balancing selection and the maintenance of MHC supertype variation in a selfing vertebrate. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1754), 20122854. https://doi.org/10.1098/rspb.2012.2854

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

de Filippo, C., Key, F. M., Ghirotto, S., Benazzo, A., Meneu, J. R., Weihmann, A., NISC Comparative Sequence Program, Parra, G., Green, E. D., & Andrés, A. M. (2016). Recent Selection Changes in Human Genes under Long-Term Balancing Selection. *Molecular Biology and Evolution*, *33*(6), 1435–1447. https://doi.org/10.1093/molbev/msw023

de Flamingh, A., Sole, C. L., & van Aarde, R. J. (2015). Genetic evidence for spatial structuring in a continuous African elephant (Loxodonta africana) population. *Conservation Genetics*, *16*(3), 613–623. https://doi.org/10.1007/s10592-014-0686-9

Doytchinova, I. A., & Flower, D. R. (2005). In Silico Identification of Supertypes for Class II MHCs. *The Journal of Immunology*, *174*(11), 7085–7095. https://doi.org/10.4049/jimmunol.174.11.7085

Edge, P., & Bansal, V. (2019). *Longshot: Accurate variant calling in diploid genomes using single-molecule long read sequencing* [Preprint]. Bioinformatics. https://doi.org/10.1101/564443

Ejsmond, M. J., Phillips, K. P., Babik, W., & Radwan, J. (2018). The role of MHC supertypes in promoting trans-species polymorphism remains an open question. *Nature Communications*, *9*(1), 4362. https://doi.org/10.1038/s41467-018-06821-x

Elias, S. A., & Schreve, D. C. (2013). Vertebrate Records | Late Pleistocene Megafaunal Extinctions. In *Encyclopedia of Quaternary Science* (pp. 700–712). Elsevier. https://doi.org/10.1016/B978-0-444-53643-3.00245-4

Flajnik, M. F., & Kasahara, M. (2010). Origin and evolution of the adaptive immune system: Genetic events and selective pressures. *Nature Reviews Genetics*, *11*(1), 47–59. https://doi.org/10.1038/nrg2703

Garnier-Géré, P., & Chikhi, L. (2013). Population Subdivision, Hardy–Weinberg Equilibrium and the Wahlund Effect. In John Wiley & Sons, Ltd (Ed.), *ELS* (1st ed.). Wiley. https://doi.org/10.1002/9780470015902.a0005446.pub3

Go, Y., Satta, Y., Kawamoto, Y., Rakotoarisoa, G., Randrianjafy, A., Koyama, N., & Hirai, H. (2002). Mhc-DRB genes evolution in lemurs. *Immunogenetics*, *54*(6), 403–417. https://doi.org/10.1007/s00251-002-0480-6

Goswami, V. R., Vasudev, D., & Oli, M. K. (2014). The importance of conflict-induced mortality for conservation planning in areas of human–elephant co-occurrence. *Biological Conservation*, *176*, 191–198. https://doi.org/10.1016/j.biocon.2014.05.026

Gray, T. N. E., Vidya, T. N. C., Potdar, S., Bharti, D. K., & Sovanna, P. (2014). Population size estimation of an Asian elephant population in eastern Cambodia through non-invasive mark-recapture sampling. *Conservation Genetics*, *15*(4), 803–810. https://doi.org/10.1007/s10592-014-0579-y

Gruber, B., Unmack, P. J., Berry, O. F., & Georges, A. (2018). DART: An R package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Molecular Ecology Resources*, *18*(3), 691–699. https://doi.org/10.1111/1755-0998.12745

Gugala, N. A., Ishida, Y., Georgiadis, N. J., & Roca, A. L. (2016). Development and characterization of microsatellite markers in the African forest elephant (Loxodonta cyclotis). *BMC Research Notes*, *9*(1), 364. https://doi.org/10.1186/s13104-016-2167-3

He, C., Du, J., Zhu, D., & Zhang, L. (2020). Population viability analysis of small population: A case study for Asian elephant in China. *Integrative Zoology*, *15*(5), 350–362. https://doi.org/10.1111/1749-4877.12432

Horrocks, N. P. C., Hegemann, A., Ostrowski, S., Ndithia, H., Shobrak, M., Williams, J. B., Matson, K. D., & Tieleman, B. I. (2015). Environmental proxies of antigen exposure explain variation in immune investment better than indices of pace of life. *Oecologia*, *177*(1), 281–290. https://doi.org/10.1007/s00442-014-3136-y

Igea, J., Juste, J., & Castresana, J. (2010). Novel intron markers to study the phylogeny of closely related mammalian species. In *BMC Evolutionary Biology* (Vol. 10, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/1471-2148-10-369

Ishida, Y., Georgiadis, N. J., Hondo, T., & Roca, A. L. (2013). Triangulating the provenance of African elephants using mitochondrial DNA. *Evolutionary Applications*, *6*(2), 253–265. https://doi.org/10.1111/j.1752-4571.2012.00286.x

Ishida, Y., Gugala, N. A., Georgiadis, N. J., & Roca, A. L. (2018). Evolutionary and demographic processes shaping geographic patterns of genetic diversity in a keystone species, the African forest elephant (*Loxodonta cyclotis*). *Ecology and Evolution*, *8*(10), 4919–4931. https://doi.org/10.1002/ece3.4062

Ison, S. A., Delannoy, S., Bugarel, M., Nagaraja, T. G., Renter, D. G., den Bakker, H. C., Nightingale, K. K., Fach, P., & Loneragan, G. H. (2016). Targeted amplicon sequencing for single-nucleotide-polymorphism genotyping of attaching and effacing Escherichia coli O26:H11 cattle strains via a high-throughput library preparation technique. *Applied and Environmental Microbiology*, *82*(2), 640–649. https://doi.org/10.1128/AEM.03182-15

Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Kalinowski, S. T. (2004). Counting Alleles with Rarefaction: Private Alleles and Hierarchical Sampling Designs. *Conservation Genetics*, *5*(4), 539–543. https://doi.org/10.1023/B:COGE.0000041021.91777.1a

Keller, L. (2002). Inbreeding effects in wild populations. *Trends in Ecology & Evolution*, *17*(5), 230–241. https://doi.org/10.1016/S0169-5347(02)02489-8

Kent, J. W. (2002). BLAT—The BLAST-Like Alignment Tool. *Genome Research*, *12*, 656–664. https://doi.org/10.1101/gr.229202

Khan, I., Maldonado, E., Silva, L., Almeida, D., Johnson, W. E., O'Brien, S. J., Zhang, G., Jarvis, E. D., Gilbert, M. T. P., & Antunes, A. (2019). The Vertebrate TLR Supergene Family Evolved Dynamically by Gene Gain/Loss and Positive Selection Revealing a Host–Pathogen Arms Race in Birds. *Diversity*, *11*(8), 131. https://doi.org/10.3390/d11080131

Knaus, B. J., & Grünwald, N. J. (2017). VCFR: A package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, *17*(1), 44–53. https://doi.org/10.1111/1755-0998.12549

Korunes, K. L., & Samuk, K. (2021). PIXY: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Molecular Ecology Resources*, *21*(4), 1359–1368. https://doi.org/10.1111/1755-0998.13326

Kyriazis, C. C., Wayne, R. K., & Lohmueller, K. E. (2021). Strongly deleterious mutations are a primary determinant of extinction risk due to inbreeding depression. *Evolution Letters*, *5*(1), 33–47. https://doi.org/10.1002/evl3.209

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, *23*(21), 2947–2948. https://doi.org/10.1093/bioinformatics/btm404

Lee, P. C., & Graham, M. D. (2006). African elephants and human-elephant interactions: Implications for conservation. *Int Zoo Yearb*, *40*, 9–19. https://doi.org/10.1111/j.1748-1090.2006.00009.x

Lei, R., Brenneman, R. A., Schmitt, D. L., & Louis, E. E. (2009). Detection of cytonuclear genomic dissociation in the North American captive African elephant collection. *Journal of Heredity*, *100*(6), 675–680. https://doi.org/10.1093/jhered/esp069

Leimgruber, P., Gagnon, J. B., Wemmer, C., Kelly, D. S., Songer, M. A., & Selig, E. R. (2003). Fragmentation of Asia's remaining wildlands: Implications for Asian elephant conservation. *Animal Conservation*, *6*(4), 347–359. https://doi.org/10.1017/S1367943003003421

Lenz, T. L. (2011). Computational prediction of MHC II-antigen binding supports divergent allele advantage and explains trans-species polymorphism. *Evolution*, *65*(8), 2380–2390. https://doi.org/10.1111/j.1558-5646.2011.01288.x

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*, *26*(5), 589–595. https://doi.org/10.1093/bioinformatics/btp698

Lighten, J., Papadopulos, A. S. T., Mohammed, R. S., Ward, B. J., G. Paterson, I., Baillie, L., Bradbury, I. R., Hendry, A. P., Bentzen, P., & van Oosterhout, C. (2017). Evolutionary genetics of immunological supertypes reveals two faces of the Red Queen. *Nature Communications*, *8*(1), 1294. https://doi.org/10.1038/s41467-017-01183-2

Litman, G. W., Cannon, J. P., & Dishaw, L. J. (2005). Reconstructing immune phylogeny: New perspectives. *Nature Reviews Immunology*, *5*(11), 866–879. https://doi.org/10.1038/nri1712

Liu, G., Zhang, H., Zhao, C., & Zhang, H. (2020). Evolutionary History of the Toll-Like Receptor Gene Family across Vertebrates. *Genome Biology and Evolution*, *12*(1), 3615–3634. https://doi.org/10.1093/gbe/evz266

Liu, S., Westbury, M. V., Dussex, N., Mitchell, K. J., Sinding, M.-H. S., Heintzman, P. D., Duchêne, D. A., Kapp, J. D., von Seth, J., Heiniger, H., Sánchez-Barreiro, F., Margaryan, A., André-Olsen, R., De Cahsan, B., Meng, G., Yang, C., Chen, L., van der Valk, T., Moodley, Y., … Gilbert, M. T. P. (2021). Ancient and modern genomes unravel the evolutionary history of the rhinoceros family. *Cell*, *184*(19), 4874-4885.e16. https://doi.org/10.1016/j.cell.2021.07.032

Maisels, F., Strindberg, S., Blake, S., Wittemyer, G., Hart, J., Williamson, E. A., Aba'a, R., Abitsi, G., Ambahe, R. D., Amsini, F., Bakabana, P. C., Hicks, T. C., Bayogo, R. E., Bechem, M., Beyers, R. L., Bezangoye, A. N., Boundja, P., Bout, N., Akou, M. E., … Warren, Y. (2013). Devastating Decline of Forest Elephants in Central Africa. *PLoS ONE*, *8*(3), e59469. https://doi.org/10.1371/journal.pone.0059469

Messier-Solek, C., Buckley, K. M., & Rast, J. P. (2010). Highly diversified innate receptor systems and new forms of animal immunity. *Seminars in Immunology*, *22*(1), 39–47. https://doi.org/10.1016/j.smim.2009.11.007

Morris, K. M., Wright, B., Grueber, C. E., Hogg, C., & Belov, K. (2015). Lack of genetic diversity across diverse immune genes in an endangered mammal, the Tasmanian devil (*Sarcophilus harrisii*). *Molecular Ecology*, *24*(15), 3860–3872. https://doi.org/10.1111/mec.13291

Murray, G. G., Woolhouse, M. E., Tapio, M., Mbole-Kariuki, M. N., Sonstegard, T. S., Thumbi, S. M., Jennings, A. E., van Wyk, I., Chase-Topping, M., Kiara, H., Toye, P., Coetzer, K., deC Bronsvoort, B. M., & Hanotte, O. (2013). Genetic susceptibility to infectious disease in East African Shorthorn Zebu: A genome-wide analysis of the effect of heterozygosity and exotic introgression. *BMC Evolutionary Biology*, *13*(1), 246. https://doi.org/10.1186/1471-2148-13-246

Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, *76*(10), 5269–5273. https://doi.org/10.1073/pnas.76.10.5269

Netea, M. G., Wijmenga, C., & O'Neill, L. A. J. (2012). Genetic variation in Toll-like receptors and disease susceptibility. *Nature Immunology*, *13*(6), 535–542. https://doi.org/10.1038/ni.2284

Norman, P. J., Norberg, S. J., Guethlein, L. A., Nemat-Gorgani, N., Royce, T., Wroblewski, E. E., Dunn, T., Mann, T., Alicata, C., Hollenbach, J. A., Chang, W., Shults Won, M., Gunderson, K. L., Abi-Rached, L., Ronaghi, M., & Parham, P. (2017). Sequences of 95 human *MHC* haplotypes reveal extreme coding variation in genes other than highly polymorphic *HLA class I* and *II*. *Genome Research*, *27*(5), 813–823. https://doi.org/10.1101/gr.213538.116

O'Grady, J. J., Brook, B. W., Reed, D. H., Ballou, J. D., Tonkyn, D. W., & Frankham, R. (2006). Realistic levels of inbreeding depression strongly affect extinction risk in wild populations. *Biological Conservation*, *133*(1), 42–51. https://doi.org/10.1016/j.biocon.2006.05.016

Palkopoulou, E., Lipson, M., Mallick, S., Nielsen, S., Rohland, N., Baleka, S., Karpinski, E., Ivancevic, A. M., To, T.-H., Kortschak, R. D., Raison, J. M., Qu, Z., Chin, T.-J., Alt, K. W., Claesson, S., Dalén, L., MacPhee, R. D. E., Meller, H., Roca, A. L., … Reich, D. (2018). A comprehensive genomic history of extinct and living elephants. *Proceedings of the National Academy of Sciences*, *115*(11). https://doi.org/10.1073/pnas.1720554115

Pečnerová, P., Díez-del-Molino, D., Vartanyan, S., & Dalén, L. (2016). Changes in variation at the MHC class II DQA locus during the final demise of the woolly mammoth. *Scientific Reports*, *6*(1), 25274. https://doi.org/10.1038/srep25274

Pérez-González, J., Carranza, J., Torres-Porras, J., & Fernández-García, J. L. (2010). Low Heterozygosity at Microsatellite Markers in Iberian Red Deer with Small Antlers. *Journal of Heredity*, *101*(5), 553–561. https://doi.org/10.1093/jhered/esq049

Quach, H., Wilson, D., Laval, G., Patin, E., Manry, J., Guibert, J., Barreiro, L. B., Nerrienet, E., Verschoor, E., Gessain, A., Przeworski, M., & Quintana-Murci, L. (2013). Different selective pressures shape the evolution of Toll-like receptors in human and African great ape populations. *Human Molecular Genetics*, *22*(23), 4829–4840. https://doi.org/10.1093/hmg/ddt335

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Radwan, J., Babik, W., Kaufman, J., Lenz, T. L., & Winternitz, J. (2020). Advances in the Evolutionary Understanding of MHC Polymorphism. *Trends in Genetics*, *36*(4), 298–311. https://doi.org/10.1016/j.tig.2020.01.008

Ripple, W. J., Newsome, T. M., Wolf, C., Dirzo, R., Everatt, K. T., Galetti, M., Hayward, M. W., Kerley, G. I. H., Levi, T., Lindsey, P. A., Macdonald, D. W., Malhi, Y., Painter, L. E., Sandom, C. J., Terborgh, J., & Van Valkenburgh, B. (2015). Collapse of the world's largest herbivores. *Science Advances*, *1*(4), e1400103. https://doi.org/10.1126/sciadv.1400103

Roach, J. C., Glusman, G., Rowen, L., Kaur, A., Purcell, M. K., Smith, K. D., Hood, L. E., & Aderem, A. (2005). The evolution of vertebrate Toll-like receptors. *Proceedings of the National Academy of Sciences*, *102*(27), 9577–9582. https://doi.org/10.1073/pnas.0502272102

Roca, A. L., Georgiadis, N., & Brien, S. J. O. (2005). *Cytonuclear genomic dissociation in African elephant species*. *37*(1), 96–100. https://doi.org/10.1038/ng1485

Roca, A. L., Georgiadis, N., Pecon-slattery, J., & Brien, S. J. O. (2001). *Genetic Evidence for Two Species of Elephant in Africa*. *293*(August), 1473–1477.

Sánchez-Barreiro, F., Gopalakrishnan, S., Ramos-Madrigal, J., Westbury, M. V., de Manuel, M., Margaryan, A., Ciucani, M. M., Vieira, F. G., Patramanis, Y., Kalthoff, D. C., Timmons, Z., Sicheritz-Pontén, T., Dalén, L., Ryder, O. A., Zhang, G., Marquès-Bonet, T., Moodley, Y., & Gilbert, M. T. P. (2021). Historical population declines prompted significant genomic erosion in the northern and southern white rhinoceros (Ceratotherium *simum*). *Molecular Ecology*, *30*(23), 6355–6369. https://doi.org/10.1111/mec.16043

Shoshani, J. (1998). Understanding proboscidean evolution: A formidable task. *Trends in Ecology & Evolution*, *13*(12), 480–487. https://doi.org/10.1016/S0169-5347(98)01491-8

Skevaki, C., Pararas, M., Kostelidou, K., Tsakris, A., & Routsias, J. G. (2015). Single nucleotide polymorphisms of Toll-like receptors and susceptibility to infectious diseases. *Clinical and Experimental Immunology*, *180*(2), 165–177. https://doi.org/10.1111/cei.12578

Spurgin, L. G., & Richardson, D. S. (2010). How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1684), 979–988. https://doi.org/10.1098/rspb.2009.2084

Stuart, A. J. (2015). Late Quaternary megafaunal extinctions on the continents: A short review. *Geological Journal*, *50*(3), 338–363. https://doi.org/10.1002/gj.2633

Sun, Y., Chen, Y., Díaz-Sacco, J. J., & Shi, K. (2021). Assessing population structure and body condition to inform conservation strategies for a small isolated Asian elephant (Elephas maximus) population in southwest China. *PLOS ONE*, *16*(3), e0248210. https://doi.org/10.1371/journal.pone.0248210

Szpiech, Z. A., & Rosenberg, N. A. (2011). On the size distribution of private microsatellite alleles. *Theoretical Population Biology*, *80*(2), 100–113. https://doi.org/10.1016/j.tpb.2011.03.006

The MHC sequencing consortium. (1999). Complete sequence and gene map of a human major histocompatibility complex. *Nature*, *401*(6756), 921–923. https://doi.org/10.1038/44853

Tollis, M., Ferris, E., Campbell, M. S., Harris, V. K., Rupp, S. M., Harrison, T. M., Kiso, W. K., Schmitt, D. L., Garner, M. M., Aktipis, C. A., Maley, C. C., Boddy, A. M., Yandell, M., Gregg, C., Schiffman, J. D., & Abegglen, L. M. (2021). Elephant Genomes Reveal Accelerated Evolution in Mechanisms Underlying Disease Defenses. *Molecular Biology and Evolution*, *38*(9), 3606–3620. https://doi.org/10.1093/molbev/msab127

Trachtenberg, E., Korber, B., Sollars, C., Kepler, T. B., Hraber, P. T., Hayes, E., Funkhouser, R., Fugate, M., Theiler, J., Hsu, Y. S., Kunstman, K., Wu, S., Phair, J., Erlich, H., & Wolinsky, S. (2003). Advantage of rare HLA supertype in HIV disease progression. *Nature Medicine*, *9*(7), 928–935. https://doi.org/10.1038/nm893

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3-new capabilities and interfaces. *Nucleic Acids Research*, *40*(15), 1–12. https://doi.org/10.1093/nar/gks596

Väli, Ü., Einarsson, A., Waits, L., & Ellegren, H. (2008). To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? *Molecular Ecology*, *17*(17), 3808–3817. https://doi.org/10.1111/j.1365-294X.2008.03876.x

Vidya, T. N. C., Fernando, P., Melnick, D. J., & Sukumar, R. (2005). Population differentiation within and among Asian elephant (Elephas maximus) populations in southern India. *Heredity*, *94*(1), 71–80. https://doi.org/10.1038/sj.hdy.6800568

Wang, J., Zhang, Z., Liu, J., Zhao, J., & Yin, D. (2016). Ectodomain Architecture Affects Sequence and Functional Evolution of Vertebrate Toll-like Receptors. *Scientific Reports*, *6*(1), 26705. https://doi.org/10.1038/srep26705

Weir, B. S., & Cockerham, C. C. (1984). Estimating F -Statistics for the analysis of population structure. *Evolution*, *38*(6), 1358–1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x

Wittemyer, G., Okello, J. B. A., Rasmussen, H. B., Arctander, P., Nyakaana, S., Douglas-Hamilton, I., & Siegismund, H. R. (2009). Where sociality and relatedness diverge: The genetic basis for hierarchical social organization in African elephants. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1672), 3513–3521. https://doi.org/10.1098/rspb.2009.0941

Wlasiuk, G., & Nachman, M. W. (2010). Adaptation and Constraint at Toll-Like Receptors in Primates. *Molecular Biology and Evolution*, *27*(9), 2172–2186. https://doi.org/10.1093/molbev/msq104

Wolf, J. B. W., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, *18*(2), 87–100. https://doi.org/10.1038/nrg.2016.133

Yao, Y.-F., Dai, Q.-X., Li, J., Ni, Q.-Y., Zhang, M.-W., & Xu, H.-L. (2014). Genetic diversity and differentiation of the rhesus macaque (Macaca mulatta) population in western Sichuan, China, based on the second exon of the major histocompatibility complex class II DQB (MhcMamu-DQB1) alleles. *BMC Evolutionary Biology*, *14*(1), 130. https://doi.org/10.1186/1471-2148-14-130

Zhang, X., Lin, W., Zhou, R., Gui, D., Yu, X., & Wu, Y. (2016). Low Major Histocompatibility Complex Class II Variation in the Endangered Indo-Pacific Humpback Dolphin (*Sousa chinensis*): Inferences About the Role of Balancing Selection. *Journal of Heredity*, *107*(2), 143–152. https://doi.org/10.1093/jhered/esv138

4. CHAPTER IV: CONCLUDING REMARKS

This dissertation aims to fill an important gap to understanding evolution of immunity of extinct woolly mammoths and extant elephants. The first part of this work focused on determinate the effect of progressive loss of nuclear genetic diversity in mammoth populations during the Late Pleistocene through the Holocene (a period ~50,000 years), before to their isolation, collapse and extinction. It is known that in small and isolated populations, inbreeding increases leading to the accumulation of deleterious mutations. The research presented in **Chapter 2** has traced the spatiotemporal accumulation of detrimental mutations over the Late Pleistocene in woolly mammoths. The results reinforce a view of selection was unable to purge deleterious variants and that, instead, these were maintained over the long term for much of the last 50,000 years. Mammoths formed a largely panmictic population, and climate change drove the loss of diversity, affecting among others innate immune genes. Predicting functional impact of missense variants revealed a potential dysregulation of immune system and alterations in the neural development. It is therefore possible that this mutational load reduced adaptive potential and increased the vulnerability to infectious diseases in woolly mammoths. The findings found in this study pave the way for deeper investigations into the origin, consequence, and evolution of genomic erosion, as well as to track population genomic diversity changes through time, and to understand the extinction of mammoths in more detail.

The second part of this work, **Chapter 3**, focused on investigating the diversity of TLRs and MHC, innate and adaptive immunity respectively, in extant elephants (savanna, forest, and Asian elephants). These three elephant populations have been different and complex demographic trajectories, coinciding in a decline of genetic diversity due human activities. TLRs and MHC genes are crucial to defence against pathogens, therefore, diversity in these genes is essential for the long-term survival of wildlife species, and consequently, of primary interest in conservation genetics of endangered species. The viability of elephant populations may be affected due to increased inbreeding, which in turn leads to loss of genetic diversity, and heterozygosity deficiency, and which may be resulting in accumulation of deleterious alleles. On the other hand, a differential selective pressure is occurring in the TLRs implying species-

specific responses to pathogens. In contrast, trans-specific MHC class II alleles among three elephant species has been maintained by balancing selection suggesting a long-term persistence of polymorphisms driven by host-parasite co-evolution. Inbreeding prevalence plus loss of diversity may ensuing in a depletion of immunodiversity, compromising the response to infections.

Understanding how the extinction occurred in woolly mammoths provides fundamental insight into the history, dynamics and conservation of contemporary elephant populations. Consequences of loss of diversity together with missense variants accumulation remain poorly understood, however can lead to a broad breakdown of the immune system, including impairments in pathogen recognition, unbalance of immune homeostasis, as well as neurodevelopment alterations. This work provides a computational prediction of deleterious variants impact but falls short in understanding the functional implications, thus, new large-scale screening methodologies are needed to evaluate in vitro the implications of each mutation to have an idea how occurred the genome meltdown in mammoths. Remarkably, results of genetic diversity in elephants reinforce the idea that elephants have undergone diversity decline and may increase their extinction risk via decreased the capacity to mount an immunity response. Comprehension of long-term persistence of MHC polymorphisms is certainly incomplete, for addressing this aspect, it would be interesting to ascertain the adaptive advantages and to estimate the time of origin and evolution of MHC supertypes.