GENOMIC PERSPECTIVES ON THE EVOLUTIONARY HISTORY OF NORTH AMERICAN CERVIDAE: ANCIENT DNA, DEMOGRAPHIC COLLAPSE AND HYBRIDISATION

A Thesis Submitted to the Committee on Graduate Studies in Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy in the Faculty of Arts and Science

TRENT UNIVERSITY

Peterborough, Ontario, Canada

Environmental and Life Sciences Ph.D. Graduate Program

January 2025

ABSTRACT

Genomic Perspectives on the Evolutionary History of North American Cervidae: Ancient DNA, Demographic Collapse and Hybridisation.

Camille Kessler

Since the mid-Pleistocene, evolutionary histories of North American species were shaped by extreme climatic oscillations involving major range and habitat shifts at a rapid pace. The peopling of the continent and the subsequent human colonisation waves created further pressures affecting North American faunal and plant populations. Cervidae (deer family) are a diverse group which arrived in North America approximately 5 million years ago, and is represented on the continent by several extant species. The overarching aim of my thesis is to understand how North America's dynamic history shaped the evolutionary trajectory of the region's Cervidae species. In Chapter 2, I investigated the speciation and hybridisation history of the sister species white-tailed (*Odocoileus virginianus*) and mule deer (*Odocoileus hemionus*) using whole genome data of individuals from across their ranges, including zones of sympatry and allopatry. I found negligible patterns of ancient gene-flow suggesting white-tailed and mule deer divergence occurred via drift and their hybridisation is the result of secondary contact. In Chapter 3, I furthered our understanding of the *Odocoileus* species pair evolutionary history through explicit demographic inference and selection analyses. I used deer samples from across North America and found strong signals of climate- and human-induced population declines. Particularly, this work highlights the impact of European settlers and patterns of conservation concerns for mule deer. For Chapter 4, I clarified the phylogenetic relationship of a contentious taxon, *Torontoceros*

hypogaeus, that went extinct during the late Pleistocene extinction event. I analysed 11k years old DNA of the single specimen representing the species, performed phylogenetic and divergence analyses, and found it belongs to the *Odocoileus* genus as *Odocoileus (Torontoceros) hypogaeus* which should be included in North America's late Pleistocene extinction list. Finally, in Chapter 5, I investigated genetic diversity over time in five North American Cervidae using contemporary and ancient DNA. I found patterns of change in genetic diversity that are consistent with known dispersal and demographic histories of our target species. Altogether, my thesis provides valuable insight into the evolutionary history of Northern American Cervidae, and on how they have been impacted by the continent's dynamic past.

Keywords: Caribou, Moose, Odocoileus, Wapiti, speciation genomics, demographic history, megafauna extinction, LGM, human impact, ancient DNA.

Preface

I have written my dissertation following a manuscript format as all data chapters are either published, submitted or being prepared for submission in a peer-reviewed journal. Specifically, Chapter 2 was published in *Molecular Ecology* in 2023, Chapter 3 was published in *Molecular Biology and Evolution* in 2024, Chapter 4 is under revision for *Current Biology*, and Chapter 5 is in preparation for submission. The style of each data chapter is specific to their respective journal formatting requirements, however, in the interest of space, I have made the choice to use a single citation format throughout with a single reference section at the end. Where collaborative work was involved, I have used the collective "we", and in addition to my own papers, I contributed to the following:

Cars, B. S., **Kessler, C.**, Hoffman, E. A., Côté, S. D., Koelsch, D., & Shafer, A. B. A. (2024). Island demographics and trait associations in white-tailed deer. *Heredity*, 1–10. https://doi.org/10.1038/s41437-024-00685-2

Acknowledgements

"I worked very hard, so… I deserve it"

Nobel prize Barbie - From Barbie by the spectacular Greta Gerwig (2023).

Being Swiss comes with a lot of privileges, yes, but I only realised how ridiculously outrageous they are when I moved to Canada. I have been very lucky to be born in a country with a functional public transport system, where university is almost free, where PhD students are paid above the poverty line, and so much more. Never before have I experienced predatory behaviour, been taken advantage of, or felt threatened by my university, although none of those things should be a privilege. I am also white, straight, cis-gender and able; the best I can do is use those privileges and transfer their benefit to those without. I strive to educate myself, to recognise and unlearn my biases and do my best to be an ally towards other oppressed groups. There is always room for improvement, and I hope to continue cultivating a necessary criticism towards our society and its oppressive behaviours. I must also acknowledge the brutal colonial history of Canada and the Americas as a whole, a history on which I was largely ignorant. Four years on this land is an express visit compared to the millennia of care provided by Indigenous Peoples and the centuries of oppression imposed by colonialism; I am grateful to have had the opportunity to learn about this history, to live and benefit from this stolen land. Tent University is located on the territory of the Michi Saagiig Anishnaabeg, I would like to show my respect to the First Peoples and thank them for their care, stewardship, and teachings. Miigwetch.

Honestly, I thought of giving up on this PhD and leave the country several times, my main motivations for staying were an incredible project (read on, you'll see!), a supportive supervisor, great friends, and sheer stubbornness (or delusion perhaps). Now is therefore time for endless rounds of thank yous. First up: my family and friends whose help and support in everyday life were paramount and who alleviated much of the doubts and stress that sprinkled my PhD. On the Swiss side, énorme merci à maman, papa, Jérôme, Sonia, Charlotte et Chantal pour votre soutien constant et vos encouragements. On the Canadian side and, for the night-long rant and dance sessions, the protests, the hikes, the pumpkins smashing, the northern light chases, the movie nights, the racoon safaris and stakeouts (as I write these words, all such endeavours have proven unsuccessful and I still have not seen a live racoon in Peterborough), and so many more joyful occasions, thank you to Abigayle, Alberto, Allison, Chantelle, Heather, Irina, Kirsten, and Maegwin. The gift of your presence made my life in Canada memorable in a good way and I can only be grateful to have met you all. Enfin, Ewen, je n'ai pas de mots assez forts! Merci pour tout.

Then, Aaron (I'm honestly not sure if you're going to read this), there might have been times when you pushed for analyses I thought were overkill, or which weren't working, and yes, maybe that made me want to flee the country and never come back. But generally, your constant support mixed with funny unintelligible texts (e.g. "AMAAING, holow fuck") were essential for my motivation and mental health. Thank you for this project and for allowing me to make it mine. One can hear horror stories when it comes to PhD supervisors (quite a few at Trent), but if you were a story, Aaron, you'd be a children's book, where not everything makes sense, but where everyone is kind.

Talking of supportive supervisors, I would like to thank Daniel Croll who instilled in me the love of bioinformatics and population genetics during my master's thesis. From the same lab, thank you to Emilie Chanclud; to this day, I think of your words of encouragement, and I will be forever grateful.

Thank you to my committee members, Andy Foote and Paul Szpak, you have provided valuable feedback and guidance throughout those four years.

Finally, many thanks to Alan Cain, Steeve Côté, Catherine Cullingham, Anh Dao, Orrin Duvuvuei, Russell Easy, Marco Festa-Bianchet, Brad Fulk, Steve Griffin, Eric Hoffman, Levi Jaster, Lee Jeffers, Daniel Koelsch, Emily Latch, Joe Nocera, the NRDPFC, Brent Patterson, Charles Ruth, Jonathan Shaw, Don Stewart, David Walter, Geoff Williams, Jon Wheeler, Kevin White, Mark Wong, Kiana Young and Liana Zanette for providing white-tailed and mule deer samples. Additional thanks to Daniel Koelsch for having organised such a marvellous visit of Saint-Pierre et Miquelon in September 2023.

Trent University made almost 90,000 CAD off my back in tuition, excluding countless other fees (and for what, I'm still wondering but certainly not fancy food at Christmas parties!). I only paid a fraction of it and received funding from an International Graduate Scholarship, an Ontario Graduate Scholarship, a French American Charitable Trust Scholarship, a School of Graduate Studies Travel Bursary and an ENLS Conference and Travel Scholarship. This work was further supported by NSERC, ComputeCanada Resources for Research Groups and the Canadian Foundation for Innovation.

TABLE OF CONTENTS

List of Figures

LIST OF TABLES

List of Abbreviations and Symbols

BLS Balancing selection

CCR Cross-coalescent rate

DMI Dobzhansky-Muller incompatibilities

DwGF Divergence with gene flow

GO Gene ontology

Kya Thousand years ago

iHS Integrated haplotype score

ILS Incomplete lineage sorting

LD Linkage disequilibrium

LGM Last Glacial Maximum

m Migration rate

μ Mutation rate

MD Mule deer

MHC Major histocompatibility complex

MSMC Multiple sequentially Markovian coalescent

Mya Million years ago

$\mathbf{N_c}$ Census size

$\mathbf{N_e}$ Effective population size

NGS Next generation sequencing

PCA Principal component analysis

π Genetic diversity

RDA Redundancy analysis

ρ Recombination rate

ROH Runs of homozygosity

SFS Site frequency spectrum

SNP Single nucleotide polymorphism

WGS Whole genome sequencing

WTD White-tailed deer

CHAPTER 1 : GENERAL INTRODUCTION

Two million years of change: a (very) brief history of North America

The Quaternary period (2.6 mya - present) was characterised by glacial cycles of great intensity, with ice sheets covering most of the northern hemisphere during the coldest periods (Batchelor et al. 2019; Gowan et al. 2021). Along with those climatic oscillations came global environmental changes and vegetation shifts (Allen et al. 2020; Marean et al. 2020; Mottl et al. 2021). The vegetation changes were likely similar after each cycle (Cheddadi et al. 2005): the landscape was open and herbaceous during the cold periods, and the forest colonisation process started with deglaciation dynamics (Cheddadi et al. 2005; Allen et al. 2020; Mottl et al. 2021; Hanberry 2023). Altogether, these changes strongly affected the fauna on the northern hemisphere and led to demographic fluctuations, population divergence, range shifts and extinctions (Lorenzen et al. 2011; Da Silva Coelho et al. 2023; Hold et al. 2024).

In North America, at the end of the Pleistocene and the early Holocene, the deglaciation dynamics heavily influenced regional climate through changes in moisture levels, temperatures, and the albedo effect (e.g. Shuman and Marsicek 2016; Zhang et al. 2022). These climatic changes triggered particularly fast environmental shifts (Shuman 2002; Mottl et al. 2021; Benfield et al. 2023; Hanberry 2023) which, likely combined with the arrival of humans ~15 kya, contributed to a large megafaunal extinction event approximately 11 kya (Broughton and Weitzel 2018; Meltzer 2020; Stewart et al. 2021; Bergman et al. 2023). Over 30 large mammal genera went extinct at that time (Stuart 2015; Meltzer 2020; Stewart et al. 2021), in turn affecting the environment and plant composition through loss of ecosystem functions (Malhi et al. 2016; Malhi et al. 2022).

Climatic oscillations continued throughout the Holocene at lower intensity, and 7 - 5 kya was a period of overall warming which was followed by a cooling episode culminating approximately 2 kya (Shuman and Marsicek 2016; Thompson et al. 2022; Zhang et al. 2022). Finally, since the 16th century, human activities have been impacting the environment and the climate at the global scale (Ceballos et al. 2020; IPCC 2022; Ceballos and Ehrlich 2023). In the Americas, European colonisers brought new farming practices, intense logging and hunting, among other things, and generally drastically changed the land which heavily affected local human and wildlife populations (Lindo et al. 2016; Smith et al. 2021; Brain and Prosser 2022). An iconic example of coloniser impact on wildlife is the extinction of the passenger pigeon (*Ectopistes migratorius*) in 1914; while its flocks used to contain millions of individuals and take days to fly overhead in the 19th century, habitat fragmentation and intense overhunting drove passenger pigeons to extinction in mere decades (Schorger 1955).

Northern American Cervidae

The highly diverse deer family (Cervidae) is represented by over 50 species native to all continents but Antarctica and Oceania (Smith-Jones 2022). Deer are antler-bearing ruminants that can be found from the Alaskan tundra to the tropical forests of Myanmar, and can weight from 9 to over 800 kg (Smith-Jones 2022). Northern America (Canada & USA) is home to five extant Cervidae species: caribou (*Rangifer tarandus*), moose (*Alces alces*), mule deer (*Odocoileus hemionus*), wapiti (*Cervus canadensis*) and white-tailed deer (*Odocoileus virginianus*).

The *Odocoileus* genus was among the first Cervidae genera to arise in North America about 5 mya (Webb 2000; Heffelfinger 2011). In the region, the genus is currently represented by mule and white-tailed deer which appear to have diverged

between 750 kya and 4.3 mya (Baccus et al. 1983; Douzery and Randi 1997). Caribou originated in Beringia between 1.5 and 2 mya (Webb 2000; Weckworth et al. 2012), but the North American lineage diverged from its Euro-Beringian counterpart approximately 70 kya (Flagstad and Røed 2003; Taylor et al. 2021; Hold et al. 2024). These three species survived in different refugia during the glacial cycles of the Quaternary; white-tailed and mule deer stayed essentially south of the ice sheets except for the black-tailed deer subspecies (*O. h colombianus* & *sitkensis*) which likely subsisted on the Pacific Northwest coast (Ellsworth et al. 1994; Greenslade 1998; Latch et al. 2009; Latch et al. 2014; Klicka et al. 2023). Caribou populations remained in Beringia and south of the ice, with the southern populations showing signs of introgression from several migration waves from Beringia (Flagstad and Røed 2003; Weckworth et al. 2012; Taylor et al. 2021). Glacial cycles had a strong demographic impact on all three species with warm periods linked to declines in caribou (Taylor et al. 2021), and cold periods to declines in white-tailed and mule deer (Combe et al. 2021; Lamb et al. 2021).

After the LGM, moose, wapiti, and humans, entered North America through Beringia (Hundertmark et al. 2002; Meiri et al. 2014; Meiri et al. 2020; Willerslev and Meltzer 2021; Mackiewicz et al. 2022a). These range expansions were closely followed by the late Pleistocene extinction event where four Cervidae genera went extinct: *Bretzia, Cervalces, Navahoceros* and *Torontoceros*. Some ambiguity surrounds those taxa, particularly *Torontoceros hypogaeus* which is represented by a single museum specimen (Churcher and Peterson 1982). There is little information on Indigenous impact on deer populations through the Holocene, although it was likely minor (Wolverton et al. 2008; Weitzel 2021; Groves et al. 2022), whereas all species were heavily depleted after the European colonisation in the 19th and 20th centuries (Gill 1999; O'Gara

and Dundas 2002; Schaefer 2003; McDonald et al. 2004; DeCesare et al. 2014; Jensen et al. 2023; Rosenblatt et al. 2023). Taken together, Northern American Cervidae exhibit a gradient of extinct, declining, stable and expanding species, which provides an ideal framework to study differential evolutionary history in a changing environment.

Genomics as a tool to study evolution

Technological advancement and decrease in cost of DNA sequencing methods has allowed to move from markers of limited resolution, such as microsatellites, to next generation sequencing (NGS) methods (Allendorf et al. 2010). NGS methods allow for answering complex questions in greater detail using analyses such as demographic modelling, and selection scans (Allendorf et al. 2010; Pérez-González et al. 2023). As changes in population size, hybridisation, and other evolutionary processes leave discernible patterns in the genome, we can retrace a species' evolutionary history using genomics as a tool. For example, the genome-wide analysis of population divergence (F_{ST}), sequence divergence (d_{xy}) and nucleotide diversity (π) allows to decipher patterns of selection and speciation (Cruickshank and Hahn 2014; Burri 2017; Ravinet et al. 2017; Booker et al. 2021; Shang et al. 2021). With an analysis of genetic load, inbreeding, Tajima's D and site frequency spectrum (SFS), it is possible to infer effective population size (N_e) through time, understand extinction processes, and detect patterns of conservation concern (Frankham 2005; Charlesworth 2009; Hahn 2018; Wilder et al. 2023). Finally, using a time-sequence of ancient and modern samples, we can compare present-day patterns of selection, adaptation, and demography to those of ancient populations, gaining a better understanding of a species' history.

The aim of my thesis was to understand and clarify the evolutionary history of Northern American Cervidae, how they were impacted by the continent's dynamic past,

and whether they require further conservation measures. Specifically, in Chapters 2 and 3, I investigated the speciation and demographic history of the *Odocoileus* sister species using range-wide low coverage whole genome resequencing data. Chapter 2 clarifies the hybridisation history of white-tailed and mule deer and suggests that gene-flow is the result of secondary contact and was not involved in the species divergence. In Chapter 3, I inferred white-tailed and mule deer demographic history through time and highlight the potential role of the glacial cycles in deer declines. Most strikingly, I uncovered strong signals of coloniser impact with a dramatic decline in both species followed by a recovery in white-tailed deer, the lack of such recovery in mule deer is a pattern of potential conservation concern. In Chapter 4, I clarified the taxonomic relationship of the contentious taxon *Torontoceros hypogaeus* using ancient DNA analysis. I discovered that the specimen belongs to the genus *Odocoileus* and introduce the new species *Odocoileus (Torontoceros) hypogaeus* which disappeared during the late Pleistocene extinction event. Finally, in Chapter 5, I explored changes in genetic diversity over time using chronological sampling of five extant North American Cervidae and found patterns consistent with known dispersal and demographic histories of our target species.

CHAPTER 2 : SPECIATION WITHOUT GENE-FLOW IN HYBRIDISING DEER

Authors:

Camille Kessler, Eric Wootton, and Aaron B. A. Shafer

A version of this chapter was published in *Molecular Ecology*

Contributions:

CK and ABAS conceived the study, CK performed bioinformatic analyses with contribution from ABAS, CK and EW performed the molecular laboratory work, CK and ABAS wrote the manuscript. All authors commented on the text.

Complete citation:

Kessler, C., Wootton, E., & Shafer, A. B. A. (2023). Speciation without gene-flow in hybridizing deer. *Molecular Ecology*, 32(5), 1117–1132.

https://doi.org/10.1111/mec.16824

Abstract

Under the ecological speciation model, divergent selection acts on ecological differences between populations, gradually creating barriers to gene flow and ultimately leading to reproductive isolation. Hybridisation is part of this continuum and can both promote and inhibit the speciation process. Here, we used white-tailed (*Odocoileus virginianus*) and mule deer (*O. hemionus*) to investigate patterns of speciation in hybridising sister species. We quantified genome-wide historical introgression and performed genome scans to look for signatures of four different selection scenarios. Despite ample modern evidence of hybridisation, we found negligible patterns of ancestral introgression and no signatures of divergence with gene flow, rather localised patterns of allopatric and balancing selection were detected across the genome. Genes under balancing selection were related to immunity, MHC and sensory perception of smell, the latter of which is consistent with deer biology. The deficiency of historical gene-flow suggests that white-tailed and mule deer were spatially separated during the glaciation cycles of the Pleistocene and genome wide differentiation accrued via genetic drift. Dobzhansky-Muller incompatibilities and selection against hybrids are hypothesised to be acting, and diversity correlations to recombination rates suggests these sister species are far along the speciation continuum.

Keywords: Speciation genomics, hybridisation, allopatric speciation, secondary contact, *Odocoileus*.

Introduction

Hybridisation is a widespread phenomenon that occurs at variable rates (Mallet 2005; Ragavan et al. 2017; Iacolina et al. 2019; Taylor and Larson 2019; Adavoudi and Pilot 2021). The prevalence of hybridisation suggests speciation follows a continuum as opposed to the more discrete and allopatric view introduced by Mayr (1942) (Mallet 2005; Stankowski and Ravinet 2021). However, both ideas are not incompatible as the speciation continuum ultimately leads to reproductive isolation with hybridisation being a natural outcome (Stankowski and Ravinet 2021). Under the ecological speciation model (sensu Darwin, 1859), hybrids are common and can even facilitate speciation (Nosil 2012). Ecological speciation works when ecological differences lead to divergent selection between populations; the underlying genetic mechanisms (i.e. causative loci) gradually create barriers to gene flow that accumulate which leads to reproductive isolation (Rundle and Nosil 2005; Nosil 2012; Shafer and Wolf 2013).

The genetic mechanisms involved in ecological speciation can take the form of barrier loci which act as a restraint to gene flow in different ways: for example, such loci can be under divergent selection, involved in mate choice, contribute to assortative mating, or reduce hybrid fitness (Abbott et al. 2013; Ravinet et al. 2017). Divergent selection should produce genomic islands in this context, also referred to as genomic islands of speciation, that are regions where the differentiation (e.g. F_{ST}) is higher than the neutral genomic background (Ravinet et al. 2017; Campbell et al. 2018). Such loci and islands have been observed in several species underpinning a wide array of speciation and divergence processes (Poelstra et al. 2014; Momigliano et al. 2017; Lavretsky et al. 2019; Marques et al. 2019). Scanning the genome for just F_{ST} peaks when speciation with gene flow is ongoing, however, is problematic as F_{ST} is dependent on

genetic diversity (π), and other mechanisms can create similar F_{ST} profiles such as genetic drift, global adaptation, or simply reduced genetic diversity due to background selection (Cruickshank and Hahn 2014; Burri 2017; Ravinet et al. 2017; Booker et al. 2021). Joint genome scans including F_{ST}, d_{xy} and π facilitate a better depiction of the processes of selection and ecological speciation, particularly the use of d_{xy} which is not influenced by current levels of diversity (Cruickshank and Hahn 2014; Ravinet et al. 2017; Campbell et al. 2018; Shang et al. 2021). Shang et al. (2021) recently conceptualised the expected patterns F_{ST}, d_{xy} and π under four main modes of selection: divergence with gene flow, allopatric selection (i.e., positive selection post species split), balancing selection and background selection. Using pairs of *Populus* species along the speciation continuum, they investigated the genomic landscape to test clear predictions on the behaviour of F_{ST}, d_{xy} and π. Including correlations of π and F_{ST} to recombination rate (ρ) can help place a species on the continuum, with for example the correlation between F_{ST} and ρ is expected to become stronger with increasing divergence (Burri 2017; Shang et al. 2021). This approach ultimately allows for disentangling selection patterns from a neutral background based on those statistics, and helps decipher the evolutionary history of species.

White-tailed (*Odocoileus virginianus*; WTD) and mule deer (*O. hemionus*; MD) are abundant in North America with similar morphology, activity patterns and life-history traits, but they differ in several ecological aspects (Douzery and Randi 1997; Gilbert et al. 2006; Pitra et al. 2004; Brunjes et al. 2006; Berry et al. 2019). Mule deer favour food availability whereas WTD prioritise security and thermoregulation (Whittaker and Lindzey 2004); consequently, WTD prefer habitats at lower altitude and with denser visual cover than MD that favour open areas at higher elevation (Anthony and Smith

1977a; Brunjes et al. 2006). White-tailed deer and MD also differ in their sociality and associated predator response. Mule deer live in large cohesive groups including both sexes, and, as a group, adopt an aggressive behaviour in presence of predators, whereas WTD live in smaller female-biased groups and flee in response to predators (Lingle 2001; Lingle 2003). Both species represent a high economic value in North America as hunting-related activities generate billions of dollars annually (Cambronne 2013), and both species are an important cultural component of Indigenous communities (Adams and Hamilton 2011; Peres and Altman 2018).

Despite a species divergence date estimated at ~3.13 mya (Wright et al. 2022), WTD and MD hybridise in areas of sympatry with estimated hybridisation rates ranging from 1 to 19% depending on the region (Carr and Hughes 1993; Hornbeck and Mahoney 2000; Combe et al. 2021; Russell et al. 2021a). Based on divergent species distributions (Figure 2.1A) and likely separate refugia during glacial events (Greenslade 1998), both species likely spent considerable time in allopatry. Recent hybrid zone analyses mainly found backcrossed individuals rather than F1 hybrids (Combe et al. 2021; Russell et al. 2021a), and an interspecies F_{ST} up to 0.4 (Combe et al. 2021). It has also been shown that gene flow appears restricted but bidirectional, and there are signs of introgression at both mitochondrial and nuclear levels (Carr et al. 1986; Cronin 1991; Derr 1991; Carr and Hughes 1993; Cathey et al. 1998; Bradley et al. 2003; Russell et al. 2019), such that some MD acquired WTD mitochondrial DNA around ~1.32 mya (Wright et al. 2022). This pattern would suggest ancestral hybridisation and gene flow has taken place. Given their clear behavioural differences, and their documented hybridisation, white-tailed and mule deer do not present the reproductive isolation required by Mayr's biological species concept (Mayr 1942). Here, we hypothesised that WTD and MD have evolved via

ecological speciation, and more specifically a speciation with gene flow scenario. To test this hypothesis, we quantified genome-wide introgression and past admixture events and performed genome scans to look for signatures of four different divergence scenarios. We expected to find genetic signs of divergent selection, such as speciation islands, consistent with patterns of divergence with gene flow (DwGF), and a stronger F_{ST} - ρ correlation between species compared to within. We also predicted higher rates of admixture and historical introgression in areas of sympatry.

Materials and Methods

Sampling & Sequencing

We obtained tissue samples from harvested deer collected across the range of WTD & MD including areas of sympatry and allopatry (Figure 2.1A, Table 2.1). These areas were determined from NatureServe records and adjusted with IUCN range data (IUCN 2015a; IUCN 2015b; NatureServe 2021a; NatureServe 2021b). Specifically, the state of Washington and the province of British Columbia, for which the coasts are largely MD allopatric whereas the eastern parts are sympatric, were divided in two by the 120th meridian West. We extracted DNA from tissue using the Qiagen DNeasy Blood and Tissue Kit following manufacturer's instructions and checked the samples concentration using Invitrogen Qubit assays. WGS libraries were generated at The Centre for Applied Genomics in Toronto, Canada and sequenced to an average of 4x coverage on an Illumina HiSeqX.

Data processing

Raw reads quality was examined using FastQC (v0.11.9; Andrews S., 2010). We trimmed the reads using Trimmomatic (v0.36; Bolger et al., 2014), and aligned them to the WTD genome (GCA_014726795.1) with bwa-mem (v0.7.17; Li & Durbin, 2009). We

sorted the reads using SAMtools sort (v1.10; Li et al., 2009), identified duplicates reads with Picard MarkDuplicates (v2.23.2; Broad Institute, GitHub Repository., 2019) and removed them using Sambamba view (v0.7.0; Tarasov et al., 2015). We performed a local re-alignment using GATK RealignerTargetCreator and IndelRealigner (v4.1.7.0; McKenna et al., 2010). For further quality checks, we used Sambamba flagstat, mosdepth (v0.3.1; Pedersen and Quinlan 2018) and MultiQC (v1.10; Ewels et al., 2016). For the MSMC analysis, some samples were sequenced to a higher depth on multiple lanes. For those, before the duplicate removal step, the read groups were specified and both sequencing files merged using Picard AddOrReplaceReadGroups and MergeSamFiles. The rest of the pipeline was performed as described above with the addition of a final step: to be able to carry out other analyses with a coverage consistent across samples, we reduced the final coverage to the average coverage of all other samples (4x) using Picard DownsampleSam.

We created three datasets for our different analyses. The first contained the variants among all deer individuals (hereafter called DeerSNP dataset). We produced it using ANGSD (v0.918; Korneliussen et al., 2014) and estimated genotype likelihoods following the GATK model (-gl 2) and called genotypes (-doGeno 4) with the following filtering: SNPs with a minimum p-value of $1e^{-6}$, a minimum base quality of 20 and Minimum mapQ quality of 20 (-SNP_pval 1e-6, -minMapQ 20 & -minQ 20). The second dataset additionally included invariant sites (referred to as DeerALL), it was produced following the same method as DeerSNP except for the filtering on the SNP p-value (-SNP_pval 1e-6) which was removed. Our third dataset was designed for an ABBA-BABA analysis which requires samples from an additional species as outgroup. It comprises the 28 deer individuals and one caribou sample, all mapped to the caribou genome

(GCA_014898785.1; afterwards called DeerCAR dataset). For this dataset, we first sorted the deer bam files by read name and converted them to fastq using SAMtools sort and fastq respectively. We mapped these fastq to the caribou reference genome using bwa-mem. We ran ANGSD using the same procedure as for DeerSNP dataset, with the addition of a caribou genome sequence in bam format, obtained for this analysis from Dedato et al. (2022).

Population structure analyses and cross-species coalescent rate

To infer admixture proportions and compute a PCA on allele frequencies based on genotype likelihood data we used PCAngsd (Meisner and Albrechtsen 2018) and NGSadmix (Skotte et al. 2013) on the DeerSNP dataset. NGSadmix was run with K = 1 up to K =7 and the best K value was determined with CLUMPAK (Kopelman et al. 2015) following the Evanno method (Evanno et al. 2005) with 7 replicates for each K value. We investigated the correlation between the distribution on each PC and each individual's range in R using the package "stats" with a linear model including the species and either latitude or longitude, and a Pearson's correlation (e.g. lm(PC ~ Species + Latitude)).

For two samples with high coverage (Ov_ON6 at 18x - and Oh_WA1 at 22x) we implemented the multiple sequentially Markovian coalescent (MSMC) to infer the cross-coalescent rate. To do this we first generated a 35-mer mappability mask file using the SNPable pipeline as is required for MSMC (Schiffels and Wang 2020). We restricted analyses to scaffolds >500 Kbp (Gower et al. 2018) and phased samples using whatsHap 1.3 (Martin et al., 2016) prior to generating the MSMC input files. We ran MSMC2 using the time segment pattern 1×2+25×1+1×2+1×3 and estimated the cross-coalescence rate (CCR) across populations by comparing the first haplotype from each sample in a pairwise fashion using the -P flag and skipping any sites with ambiguous phasing. To

quantify the migration rate over time m(t) and the cumulative migration probability M(t) – which is related to the CCR (Schiffels and Durbin 2014) -– we implemented the IM model to the MSMC output (Wang et al. 2020). We assumed a generation time of 2 years (Demarais et al. 2000; Deyoung et al. 2003) and mutation rate of 1.33×10-8 mutations/site/generation.

Genome-wide ancestral introgression

To evaluate the extent of historical introgression between sympatric MD and WTD, we used the ABBA-BABA test, which allows for the detection of introgression between three populations P1, P2 & P3 and an outgroup (Green et al. 2010; Durand et al. 2011). We ran the ABBA-BABA analysis implemented in ANGSD on the DeerCAR dataset between individuals (-doAbbababa 1) and between populations (-doAbbababa2 1; Soraggi et al. 2018), using the caribou as outgroup. The D-statistic, standard error and Z-score were computed with ANGSD's accompanying R scripts: jackKnife.R and estAvgError.R. To further our understanding of the admixture events and build a maximum likelihood tree of our system, we used treemix (v1.13; Pickrell & Pritchard, 2012) on both DeerCAR and DeerSNP datasets. We constructed the maximum likelihood trees with migration events ranging from 0 to 5, either WTD_allopatry or Caribou as root, and accounted for linkage using 1000 SNPs per block. Tree and residuals visualisations were performed with associated R script plotting_funcs.R and the variance explained by each migration event was computed with the get_f() function.

In-windows historical introgression

The D statistic is sensitive to genomic variation (π), and should not be used to determine introgression on a small scale (Martin et al., 2015). The f_d statistic, proposed by Martin et al. (2015), is less dependent on diversity than D and therefore allows for the

detection of potentially introgressed regions of the genome. We used the python script ABBABABAwindows.py (Martin, 2021) to estimate in D and f_d in 50Kbp windows to detect potentially introgressed loci in our DeerCAR dataset on two comparisons: 1) P1 = WTD allopatry, P2 = WTD sympatry, P3 = MD sympatry and 2) P1 = MD allopatry, P2 = MD sympatry, P3 = WTD sympatry, both with the caribou as outgroup. We then identified potentially introgressed windows as those having a f_d value higher than the 97.5% quantile. Their position was used in BEDTools intersect (v2.30.0; Quinlan and Hall 2010) with the caribou annotation file to identify the genes present in putatively introgressed windows.

Genome scans

To detect islands of divergence between WTD & MD, we used the python script popgenWindows.py (Martin, 2021) to estimate individual heterozygosity, F_{ST}, d_{xy} and π in 50Kbp windows (note we filtered out scaffolds with less than 2 windows). Since the accurate computation of d_{xy} and π require a dataset including the invariant sites, we performed this analysis on the DeerALL dataset. To detect the outlier loci, we based our approach on the four selection scenarios developed in (Shang et al. 2021): I) Divergence with gene flow that presents as high F_{ST} and d_{xy} but low π; II) allopatric selection that shows high F_{ST}, low π and stable d_{xy}; III) background selection that presents as high F_{ST} but low π and d_{xy}; and IV) balancing selection that displays as low F_{ST} but high π and d_{xy}. Thresholds were set as the upper or lower 5% for a high or low criteria, and between 45 and 55% for a stable condition. Outliers were flagged as belonging to one of the four scenarios when they met the specific criteria shown in Table 2.2.

We identified genes present in outlier regions by comparing the coordinates of the outlier windows and the WTD annotation file in BEDTools intersect. We extracted

the gene names from those regions as well as those from their associated scaffolds. The gene list from outlier windows was uploaded to ShinyGO for enrichment analysis with the scaffold's gene list as background information and no species specification. We downloaded the enrichment analysis results for the molecular function, biological process, and cellular component GOs and analysed the results in R. We also uploaded the gene list to UniProt's Retrieve/ID mapping tool (Pundir et al. 2016) and downloaded the information for visual inspection.

To estimate the position of WTD & MD along the speciation continuum through the correlation between recombination rate (ρ), F_{ST}, and π, we computed ρ in windows using FastEPRR (v.2.0; Gao et al., 2016). Here, we used our four categories as a proxy for a speciation continuum with six comparisons: WTDa - WTDs and MDa - MDs representing no speciation, WTDs - MDs, WTDs - MDa and WTDa - MDs as intermediate stages and WTDa - MDa as full differentiation. We first phased the vcf using SHAPEIT (Delaneau et al. 2012) and split the resulting file into scaffolds. We then ran FastEPRR separately on our four populations to obtain the ρ estimation in 50kbp windows (winLength = 50000) on the 15 longest scaffolds of the WTD genome, with 100 repetitions. To be able to correlate ρ with F_{ST} and π, we re-ran the genome scan analysis on the population level. FastEPRR outputs ρ in $4N_e r$, to convert this value to ρ in cM/Mb, we computed N_e per population using the relationship $\pi = 4N_e\mu$. We used μ= 1.33×10-8 and the genome scans result to average π across windows in each population. After filtering to keep only windows that were present across all comparisons (12,500 windows), we generated Pearson's correlations between I) ρ in both populations (ρ_1 vs ρ_2), II) average ρ between populations and F_{ST} ($\rho_{1\text{-}2}$ vs F_{ST}), and III) average ρ and average π between populations ($\rho_{1\text{-}2}$ vs $\pi_{1\text{-}2}$) following Shang et al. (2021). The resulting

correlation coefficients were compared between all comparisons to infer trends over divergence time. All results were analysed in R version 4.1.0 "Camp Pontanezen" (R Core Team 2021) .

Results

We sequenced 28 individuals to an average coverage of approximately 4x. We called 103,970,889 SNPs in the DeerSNP dataset, which represented 4% of the genome. In the PCA based on allele frequencies, PC1 accounted for 59% of the variation and separated both species (Figure 2.1B). PC2 explained an additional 8% of the variation and showed a spread of MD individuals that is consistent with their longitudinal distribution (Pearson correlation = 0.8; *p-value* = 0.00057). Latitude was not important in shaping PC1 or PC2 variation (Pearson correlation *p-value* > 0.05). Admixture analysis showed a well-defined genetic clustering between species (Evanno's method: best K = 2, Figure 2.1C, Figure S2.1). The WTD individual Ov_MX2 was partially admixed in both analyses, with NGSAdmix assigning it 92.5% WTD and 7.44% MD ancestry (Figure 2.1C). This individuals' admixed pattern disappeared in the admixture analysis with higher K values (K = 3 to 7, Figure S2.1), though additional within species clusters are observed.

The MSMC2-IM model showed no support for migration between species once they split, with the CCR suggesting the speciation event for *Odocoileus* took place just under 1 mya (Figure 2.2).

To detect historical gene flow, we performed an ABBA-BABA test based on genotype likelihoods between all combinations of our 28 individuals in a first analysis and between all four populations in a second, with the caribou as outgroup (DeerCAR dataset, 72,438,766 SNPs). We computed 9828 ABBA-BABA individual topologies in ANGSD, of which 30 presented signatures of introgression. We found six potential inter-

species introgression events with a D-statistic ranging from 0.011 to -0.013, two of which are significant (Z-score > |3| (Barlow et al. 2021; Kirch et al. 2021), Figure 2.3A). D-statistic in intra-species comparisons range from 0.25 to 0.916 in ABBA and from -0.046 to -0.914 in BABA, all intra-species comparison show a significant Z-score (Figure 2.3A). In the comparisons between the four populations, three ABBA-BABA topologies presented signatures of introgression on a total of 12. The single inter-species comparison presents a D-statistic of -0.07 (Z-score = -7.66) for an introgression between sympatric MD and WTD. The other two comparisons show a reciprocal introgression between MD in allopatry and sympatry of 0.9 and 0.89 (Z-score = 1342.83 & 1376.69; Figure S2.3). The treemix analyses showed a topology with no migration that explained over 99% of the variation (Figure 2.3B). We further analysed historical introgression through the measure of f_d across 50kb non-overlapping windows in two different introgression scenarios: 1) WTDa, WTDs, MDs, Caribou and 2) MDa, MDs, WTDs, Caribou (Figure S2.4). We found 1243 outlier windows with a f_d higher than the 97.5% percentile in the first scenario and 1248 in the second, each representing ~2.45% of the genome with 68 windows overlapping between comparisons (Figure S2.4).

Measures of relative genomic divergence between WTD and MD were elevated (F_{ST} = 0.26) and absolute divergence d_{xy} was 0.011; genetic diversity (π) was higher in WTD than in MD (WTD = 0.008, MD = 0.004, Mann-Whitney test: $p < 2.2e-16$). We identified a total of 1183 windows presenting patterns consistent with one of the four selection scenarios (Figure 2.4, Figure S2.5). Of those, 1016 suggest a pattern of balancing selection, they were distributed across 236 scaffolds and represent 1.99% of the genome. Those windows contained 121 genes, some of which showed an enrichment in ontologies associated with the sense of smell and chemical stimuli, including three

categories presenting an enrichment above 15-fold (Figure S2.6). We also detected GOs related to the MHC and immunity, these include three categories with an enrichment above 24-fold (Figure S2.6). We identified patterns of background selection in 165 windows across 58 scaffolds, representing 0.32% of the genome. These windows harboured 208 genes identified with a UniProt ID and for which we found enrichment in GOs related to epigenetic factors such as "Unmethylated CpG binding" (25-fold) or "Histone demethylation" (15-fold) (Figure S2.7). These windows were either isolated or clustered together into putative islands of divergence (Figure S2.5). We found 2 windows under allopatric selection, each window containing one gene: *ACAP2*, a GTPase activating protein, and *PCDHB4* potentially involved in cell-binding (UniProt 2022a; UniProt 2022b). When we applied a more liberal cut-offs for $F_{ST,}$ $d_{xy,}$ and π we still observed no evidence for divergence with gene flow (Table S2.1). The correlations between π, F_{ST} and ρ are highly significant in all six comparisons (Figure S2.8). As predicted correlations coefficients for F_{ST} - ρ became stronger with divergent comparisons (Figure 2.5). Likewise, the π - ρ correlation coefficients were constant over divergence time, while the recombination landscape (ρ - ρ) differed between species (Figure 2.5).

Discussion

We investigated the speciation history of the *Odocoileus* genus through introgression analyses and genome scans. Our results suggest speciation between WTD & MD took place with negligible historical gene flow, despite contemporary and historical evidence of interbreeding (Stubblefield et al. 1986; Cronin et al. 1988; Cronin 1991; Derr 1991; Hughes and Carr 1993; Cathey et al. 1998; Hornbeck and Mahoney 2000; Combe et al. 2021; Russell et al. 2021a). Consistent with other estimates of divergence

(Douzery and Randi 1997; Combe et al. 2021), WTD and MD appear to have split ~1 mya (Figure 2.2). The main pattern of selection, however, appears to be that of balancing selection, with patterns of divergence with gene flow notably absent. Moreover, the scarcity of allopatric selection (i.e., positive selection post divergence) suggests isolation and drift primarily underlies the species differentiation, with WTD and MD likely only recently coming into secondary contact. The genome-wide scans combined with the absence of introgression suggests that WTD & MD are far along the speciation continuum (i.e. Feder 2012), despite contemporary hybridisation. Some signals of selection might have been lost by recombination, elevated F_{ST}, and time, but the paucity of DwGF and AS signals even at liberal cut-offs suggest speciation was driven primarily by drift.

Speciation with negligible historical introgression

Contemporary hybridisation in wild WTD & MD appears highly variable: rates between 1 and 19% have been observed, which can be explained by the region and degree of overlap, but also by the methodologies used that vary in resolution (Stubblefield et al. 1986; Cronin et al. 1988; Cronin 1991; Derr 1991; Hughes and Carr 1993; Cathey et al. 1998; Hornbeck and Mahoney 2000; Combe et al. 2021; Russell et al. 2021a). Moreover, a proposed historical mtDNA introgression would suggest ancestral hybridisation occurred (Wright et al. 2022). We therefore expected to find signs of introgression in our samples (e.g. Combe et al. 2021) assuming that historic matings would have left some detectable signal as is seen in humans (e.g. Green et al., 2010) and other species (Liu et al., 2021; Paijmans et al., 2021; Poelstra et al., 2014). Surprisingly, we observed negligible levels of ancestral gene flow and introgression. While Combe et al. (2021) did find evidence of introgression between *Odocoileus* species; we note, this

inference was based off a small number of topologies and included recent hybrids, both of which could skew D values. We did find signs for some contemporary admixture in one sample (Figure 2.1C), but supplementary analysis with an additional WTD from Mexico showed that the admixture pattern likely reflects population structure rather than hybridisation since it is detected in both individuals (Figure S2.2).

Our sampling surely is limited for the detection of recent hybridisation, as previous studies suggest gene flow is rather restricted (Cronin et al. 1988; Cronin 1991; Derr 1991; Hughes and Carr 1993; Russell et al. 2021a). For example, Russell et al., (2021) estimated a 1% hybridisation rate in Alberta by sampling 987 individuals in a range overlap of approximately 230,000 km^2. While signatures of f_d identified potentially introgressed windows in ~1% of the genome, this metric can be a sign of incomplete lineage sorting (ILS) rather than introgression, particularly with no signs of genome-wide introgression as evidenced in the D-statistic (Durand et al. 2011). The independent treemix and MSMC analyses do not support historical gene flow, thus collectively, we consider the f_d signal to be reflective of ILS which would be expected given the high genome-wide diversity and a relatively recent species-split (Cutter, 2013; Combe et al. 2021).

The absence of signatures of ancestral gene flow suggest that WTD & MD evolved in allopatry and have recently come in secondary contact. The dynamic history of the North American continent, notably the glacial cycles of the Quaternary, shaped the evolution and distribution of many taxa (Avise et al. 1998; Hewitt 2000; Shafer et al. 2010). The use of different refugia for prolonged periods of time during the glaciation events has increased the differentiation between populations in several species, including deer (Latch et al. 2014; Dussex et al. 2020; Kinoshita et al. 2020; Colella et al.

2021; Ito et al. 2021). Previous studies show that MD persisted in several refugia during the glacial cycles of the Pleistocene, increasing the intraspecific divergence (Latch et al. 2009; Latch et al. 2014; Wright et al. 2022). This divergence is highlighted here in the PCA which shows two distinct MD clusters, potentially segregating MD, and its black-tailed deer subspecies (Figure 2.1B). Environmental shifts following the LGM also seem to have impacted the divergence between subspecies of WTD & MD in Florida and Mexico for example (Ellsworth et al. 1994; Alminas et al. 2021). More recently, a form of allopatry was mediated through overharvest and habitat destruction of WTD where populations greatly decreased, even extirpated, in some regions (Deyoung et al. 2003; McDonald et al. 2004; Budd et al. 2018; Chafin et al. 2021). Here, the core of the current sympatric zone was greatly affected, notably Colorado, Montana, Idaho, Nebraska and Wyoming where WTD was almost extirpated before restocking efforts took place starting in the mid 40's (McDonald et al. 2004). Historical introgression could further be hindered by selection against hybrids, either through sexual selection, intrinsic or ecological incompatibilities (Rundle and Nosil 2005; Rundle and Nosil 2005; Nosil 2012). Given the genome-wide level of differentiation, we hypothesise that Dobzhansky-Muller incompatibles (DMIs) are in play, noting that they often occur early and rapidly (Schumer et al. 2015)

Little is known about hybrid fitness in the *Odocoileus* genus. Assessment of a single F1 male suggested it was fertile with no overt genetic defects (Derr et al. 1991). In three hybrids the spermatozoid phenotype showed a gradient of fertility with one sterile individual, one subfertile or infertile, and one potentially fertile but did not reproduce (Wishart et al. 1988); sperm phenotypic variation has been implicated in speciation (Albrechtová et al. 2012) and DMIs (Bhattacharyya et al. 2013). Sexual

selection can also be a powerful driver of selection against hybrids (Servedio 2004). In WTD, while females might favour males with larger antlers (Morina et al. 2018), sexual selection is poorly understood and most studies focus on male breeding success (DeYoung et al. 2006; DeYoung et al. 2009; Jones et al. 2011; Newbolt et al. 2017). Morphologically, however, WTD & MD are similar enough where the identification of their hybrids is often challenging with the metatarsal gland and their escape gaits being the only characters consistently distinguishing hybrids (e.g Lingle, 1992; Wishart, 1980). Thus, we propose a testable hypothesis that ecological incompatibilities combined with DMI maintain the species boundary. More research is also needed on WTD/MD hybrids to quantify their fitness and genomic regions underlying any selected hybrid phenotypes.

Mixed signatures of selection underly speciation in North American deer

Given ongoing hybridisation between WTD & MD and a history of recurrent glacial cycles, we expected to find deer genomes showing patterns of divergence with gene flow. The absence of introgression and the nonexistence of DwGF in our genome scans, even with more liberal thresholds, is surprising (Table S2.1). The further absence of allopatric selection suggests a diminished role of selection in deer speciation, lending support to drift-induced model more consistent with Mayr's (1942) view. While the number of windows in allopatric selection increased with more liberal thresholds (Table S2.1), they are never the dominant pattern. We suggest this absence could be a result of species being far along on the speciation continuum such that patterns of AS (specifically high F_{ST}) are no longer detectable or meaningful because they were reduced with time and recombination, which is consistent with late stages of ecological speciation (Feder 2012). Recombination rates, demography and selection can still

influence the patterns investigated here. We inferred ρ estimates that are consistent with those in other mammals (Jensen-Seaman et al. 2004), their correlations with divergence would suggest *Odocoileus* are well defined species and reflects a shift in recombination landscape between WTD & MD (Figure 2.5; Burri, 2017; Shang et al., 2021).

Signatures of balancing selection (BLS) were widespread with windows containing genes consistent with deer biology (olfactory receptors) and previous literature (i.e., MHC genes). The major histocompatibility complex (MHC) is involved in pathogen recognition in vertebrate species (Bernatchez and Landry 2003; Piertney and Oliver 2006). Polymorphism in this complex is directly linked to disease susceptibility, the most diverse are the most resistant, and there is evidence that MHC diversity is maintained by BLS (Bernatchez and Landry 2003; Piertney and Oliver 2006; Aguilar et al. 2004; Pierini and Lenz 2018; Zhang et al. 2018). The same is true for the results on sensory perception and particularly the olfactory receptors genes, both of which are under BLS (Liu et al., 2021). Each allele in the olfactory receptors genes is expressed by a single sensory neuron (Degl'Innocenti and D'Errico 2017), it is therefore plausible that diversity in these genes would be maintained by BLS. Olfactory receptors in WTD and MD, and more broadly deer, is critical to both predator detection and rutting behaviour (Ditchkoff 2011). Anatomically, WTD has an architecture optimised for the delivery of sensory stimuli to the receptors in the nose (Ranslow et al. 2014). Chamaillé-Jammes et al. (2014) showed that MD inspect and respond differently to different predators' olfactory cues, and both human and mammalian predators utilise wind direction and other mechanism to minimise scent detection by deer (Zagata and Haugen 1974; Cherry and Barton 2017). In other ungulates, olfactory reception is associated with

sexual activity (Cann et al. 2019), maternal behaviour (Keller and Lévy 2012; Blank and Yang 2017), territory choice (Deutsch and Nefdt 1992), predator response (Kuijper et al. 2014; Wikenros et al. 2015) and foraging (Hirata and Kusatake 2021). Low differentiation (F_{ST}) but high diversity (π) of olfactory receptors in deer is consistent with their biology and behaviour and would be expected to be under strong selection pressures.

Conclusion

Our results suggests that WTD & MD do not conform to a speciation with gene flow scenario despite evidence of contemporary hybridisation. We propose they were spatially separated during the Quaternary glaciation cycles where genome-wide differentiation accrued via drift. This is evidenced by the majority of the genome (> 97%) not matching a selection scenario. Increased sampling and model-based demographic assessment should help clarify the role of glaciers and secondary contact in North American deer. The near absence of patterns of allopatry and ancestral gene flow suggest that white-tailed and mule deer are far along the speciation continuum, the absence of introgression signs could suggest DMIs and selection against hybrids which would contribute to the reinforcement of reproductive isolation. Future studies should focus on assaying hybrid phenotypes and vigour.

Acknowledgements

Camille Kessler was supported by an International Graduate Scholarship. Eric Wootton was supported by three Natural Sciences and Engineering Research Council (NSERC) Undergraduate Student Research Awards. This work was supported by NSERC Discovery Grant (Grant Number: RGPIN-2017- 03934); ComputeCanada Resources for Research Groups (Grant Number: RRG gme-665-ab); Canadian Foundation for Innovation: John R. Evans Leaders Fund and the Ontario Early Researcher Award (Grant

Number: #36905). We thank Catherine Cullingham, Anh Dao, Orrin Duvuvuei, Brad Fulk, Steve Griffin, Levi Jaster, Lee Jeffers, Emily Latch, the NRDPFC, Charles Ruth, David Walter, Geoff Williams, Kevin White, Mark Wong, Kiana Young and Liana Zanette, for providing samples. We are also grateful to Andrew Foote, Jose Alberto Lòpez Alemàn and Emily Latch for their comments on the manuscript. Trent University is located on the traditional territory of the Michi Saagiig Anishnaabeg. Our sampling also covers most of North America, a stolen land which remains a home to many First Peoples. We are grateful to have had the opportunity to work on this land, would like to show our respect to the First Peoples, and thank them for their care, stewardship, and teachings. Miigwetch.

Figures and tables

Figure 2.1: Deer individuals in the study. (a) Sampling locations, colours represent the different groups: WTD allopatry (WTDa, blue), WTD sympatry (WTDs, violet), MD sympatry (MDs, green), MD allopatry (MDa, orange). Areas of sympatry are coloured Areas of sympatry are coloured in dark grey and areas of allopatry in light grey, from the IUCN range data for both species (IUCN, 2015a; IUCN, 2015b). (b) Principal component analysis performed in PCAngsd. (c) Admixture proportions for K = 2 calculated in NGSAdmix.

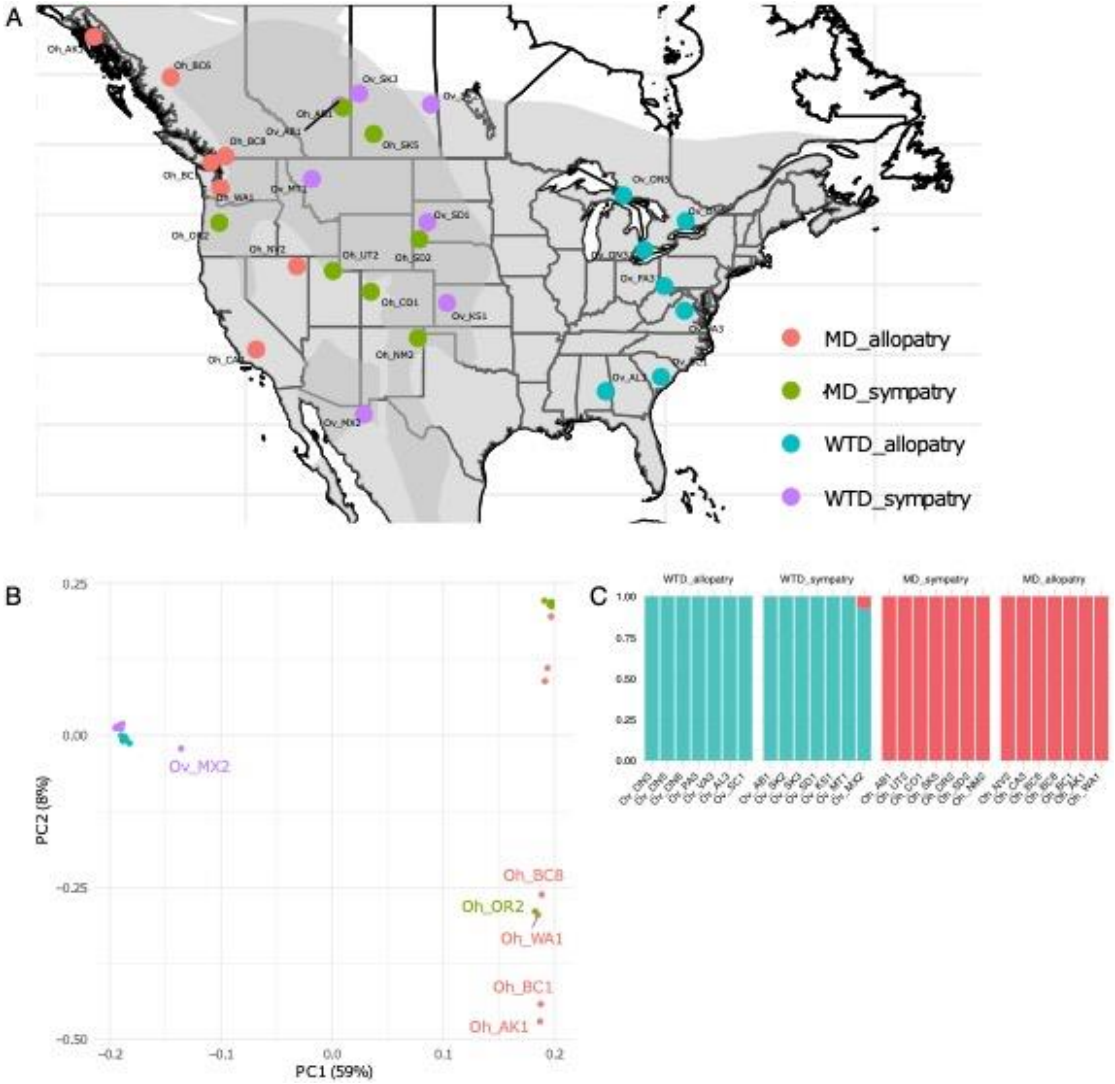


Figure 2.2: MSMC-IM analysis of white-tailed deer (Ov) and mule deer (Oh) from Ontario (ON) and Washington (WA). (a) Migration rates over time where dashed line indicates the time point where 50% of ancestry has merged. Blue shading indicates 99% percentile of the cumulative migration probability. (b) Cumulative migration probabilities M(t). Dashed lines indicate the relative cross-coalescence rate obtained from MSMC2.

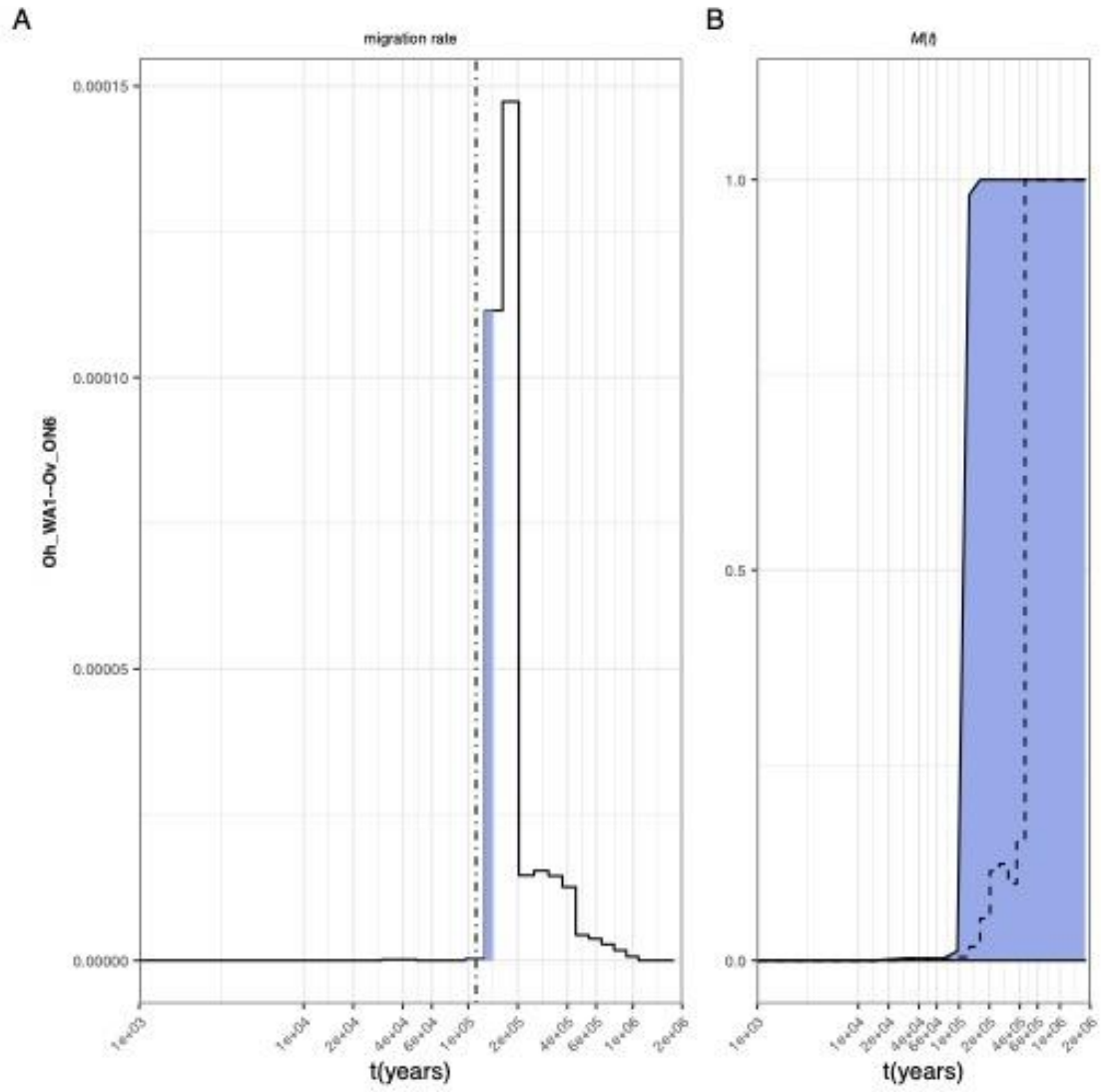


Figure 2.3: Introgression analyses. (a) ABBA-BABA analysis between individuals, individuals involved in gene flow are coloured depending on the species (WTD in blue, MD in orange), the colour gradient represents Z score, excess of ABBA depicted as points, excess of ABBA depicted as points, excess of BABA shown as triangles, significant tests highlighted with an *(Z-score > |3|). (b) Maximum likelihood tree inferred by Treemix for 0 migration (right) and the variance explained by each migration event (left).

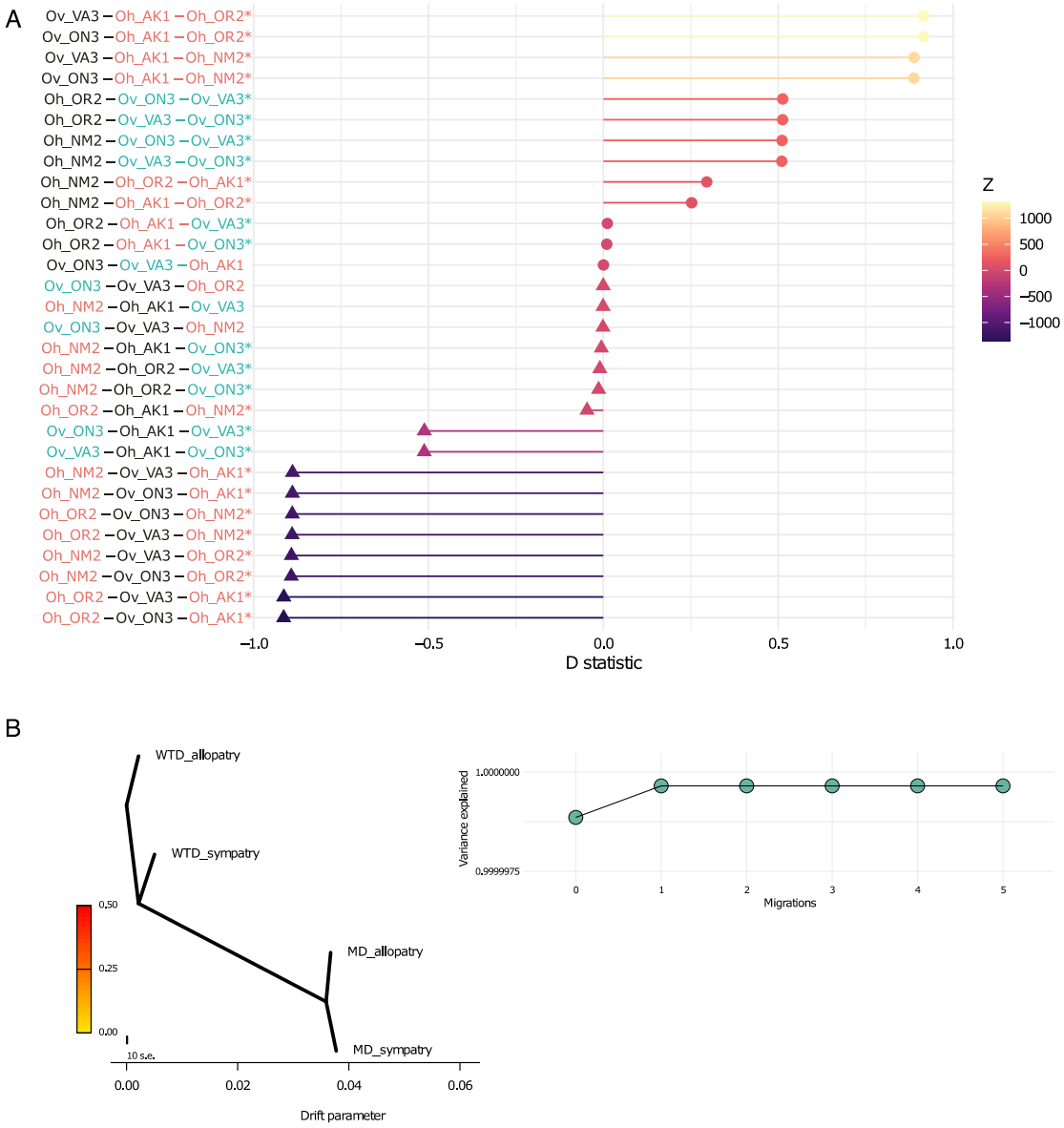


Figure 2.4: Distribution of windows presenting pattern of selection across the genome for d_{xy} (top), F_{ST} (middle) and π (bottom). Windows following a pattern of selection are coloured according to its corresponding evolutionary scenario, grey windows do not exhibit pattern of selection. The continuous and dashed line represent the mean and median respectively.

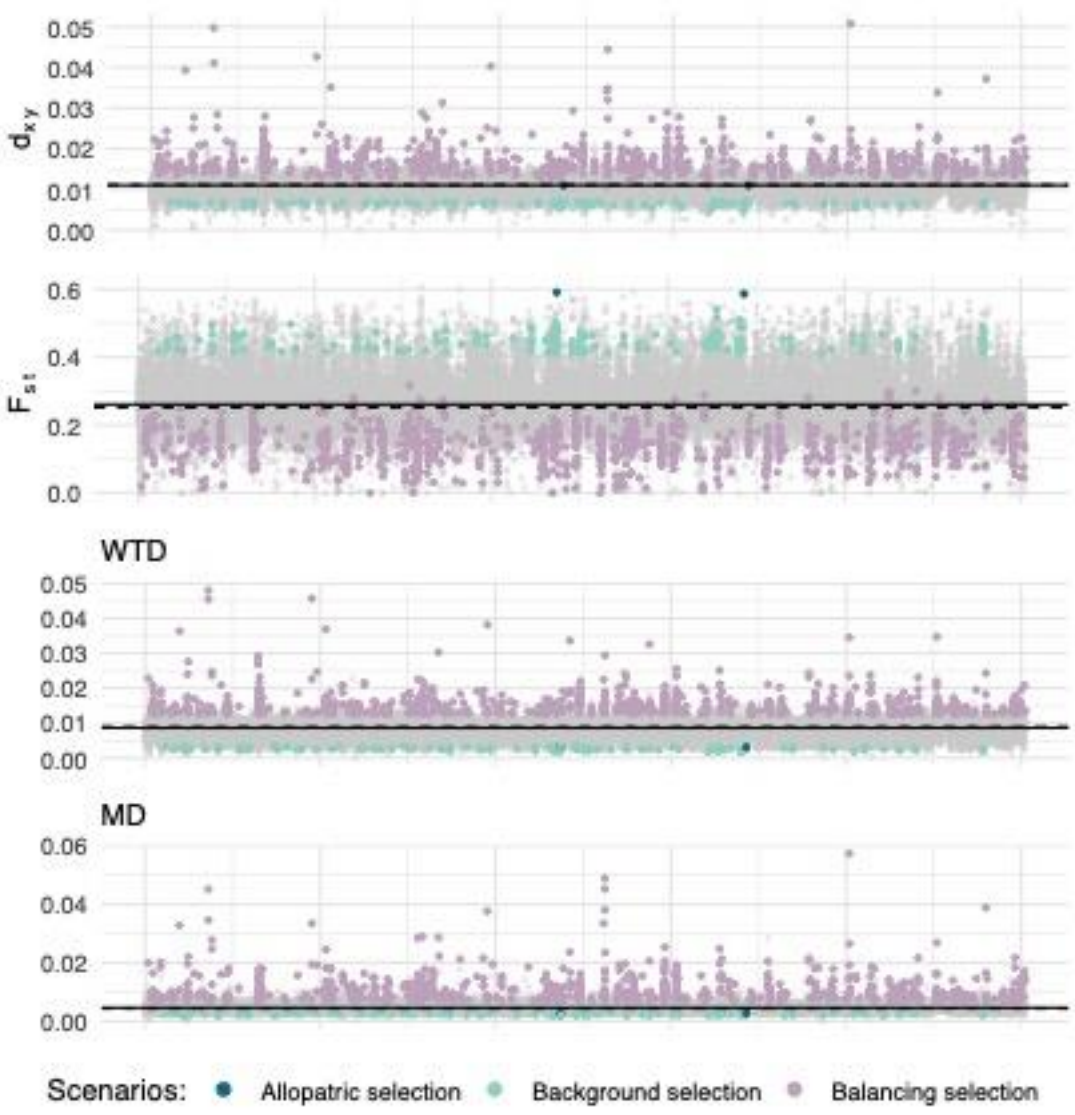


Figure 2.5: Correlation coefficients over divergence time for π - ρ (dark blue) F_{ST} – ρ (dark green) and ρ - ρ (yellow), trend lines represent a linear model. The speciation continuum is represented by six comparisons of our four groups).

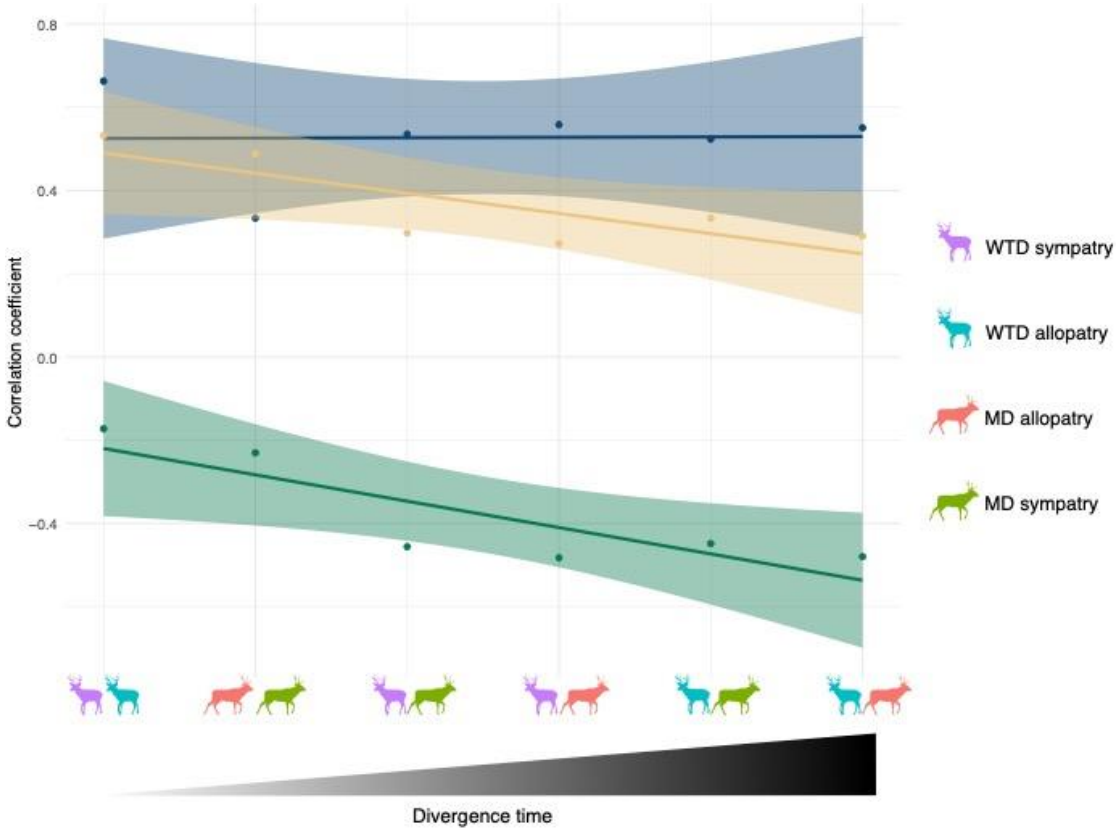


Table 2.1: Deer sample information

Genome_ID	Species	Location	Sex	Group
Ov_ON3	*O_virginianus*	Ontario	F	WTD_allopatry
Ov_ON5	*O_virginianus*	Ontario	M	WTD_allopatry
Ov_ON6	*O_virginianus*	Ontario	M	WTD_allopatry
Ov_PA3	*O_virginianus*	Pennsylvania	F	WTD_allopatry
Ov_VA3	*O_virginianus*	Virginia	F	WTD_allopatry
Ov_AL3	*O_virginianus*	Alabama	Unk	WTD_allopatry
Ov_SC1	*O_virginianus*	South Carolina	F	WTD_allopatry
Ov_AB1	*O_virginianus*	Alberta	F	WTD_sympatry
Ov_SK2	*O_virginianus*	Saskatchewan	M	WTD_sympatry
Ov_SK3	*O_virginianus*	Saskatchewan	M	WTD_sympatry
Ov_SD1	*O_virginianus*	South Dakota	M	WTD_sympatry
Ov_KS1	*O_virginianus*	Kansas	F	WTD_sympatry
Ov_MT1	*O_virginianus*	Montana	M	WTD_sympatry
Ov_MX2	*O_virginianus*	Mexico	M	WTD_sympatry
Oh_NV2	*O_hemionus*	Nevada	M	MD_allopatry
Oh_CA3	*O_hemionus*	California	Unk	MD_allopatry
Oh_BC6	*O_hemionus*	British Columbia	M	MD_allopatry
Oh_BC8	*O_hemionus*	British Columbia	M	MD_allopatry
Oh_BC1	*O_hemionus*	British Columbia	F	MD_allopatry
Oh_AK1	*O_hemionus*	Alaska	Unk	MD_allopatry
Oh_WA1	*O_hemionus*	Washington	M	MD_allopatry
Oh_AB1	*O_hemionus*	Alberta	F	MD_sympatry
Oh_UT2	*O_hemionus*	Utah	M	MD_sympatry
Oh_CO1	*O_hemionus*	Colorado	F	MD_sympatry
Oh_SK5	*O_hemionus*	Saskatchewan	Unk	MD_sympatry
Oh_OR2	*O_hemionus*	Oregon	Unk	MD_sympatry
Oh_SD2	*O_hemionus*	South Dakota	F	MD_sympatry
Oh_NM2	*O_hemionus*	New Mexico	M	MD_sympatry

Table 2.2: Four different evolutionary scenario and their thresholds for F_{ST}, π and d_{xy} percentiles applied in the genome scans analysis

Scenario	F_{st}	π	D_{xy}
Divergence with gene flow (DwGF)	> 0.95	< 0.05	> 0.95
Allopatric selection (AS)	> 0.95	< 0.05	Between 0.45 & 0.55
Background selection (BGS)	> 0.95	< 0.05	< 0.05
Balancing selection (BLS)	< 0.05	> 0.95	> 0.95

CHAPTER 3 : GENOMIC ANALYSES CAPTURE THE HUMAN-INDUCED DEMOGRAPHIC COLLAPSE AND RECOVERY IN A WIDE-RANGING CERVID

Authors:

Camille Kessler and Aaron B. A. Shafer

A version of this chapter was published in open access in *Molecular Biology and Evolution*.

Contributions:

CK and ABAS conceived the study, ABAS collected the samples, CK performed the molecular laboratory work and the bioinformatic analyses with contribution from ABAS for MSMC2, CK and ABAS wrote the manuscript.

Complete citation:

Kessler, C., & Shafer, A. B. A. (2024). Genomic Analyses Capture the Human-Induced Demographic Collapse and Recovery in a Wide-Ranging Cervid. *Molecular Biology and Evolution*, 41(3), msae038. https://doi.org/10.1093/molbev/msae038

Abstract

The glacial cycles of the Quaternary heavily impacted species through successions of population contractions and expansions. Similarly, populations have been intensely shaped by human pressures such as unregulated hunting and land use changes. White-tailed and mule deer survived in different refugia through the Last Glacial Maximum, and their populations were severely reduced after the European colonisation. Here, we analysed 73 re-sequenced deer genomes from across their North American range to understand the consequences of climatic and anthropogenic pressures on deer demographic and adaptive history. We found strong signals of climate-induced vicariance and demographic decline; notably, MSMC2 recovers a severe decline in mainland white-tailed deer effective population size (N_e) at the end of the Last Glacial Maximum. We found robust evidence for colonial overharvest in the form of a recent and dramatic drop in N_e in all analysed populations. Historical census size and restocking data show a clear parallel to historical N_e estimates, and temporal N_e/N_c ratio shows patterns of conservation concern for mule deer. Signatures of selection highlight genes related to temperature, including a cold receptor previously highlighted in woolly mammoth. We also detected immune-genes that we surmise reflect the changing land-use patterns in North America. Our study provides a detailed picture of anthropogenic and climatic-induced decline in deer diversity, and clues to understanding the conservation concerns of mule deer and the successful demographic recovery of white-tailed deer.

Keywords: *Odocoileus*, deer, last glacial maximum, island population, human impact

Introduction

The glacial cycles of the Quaternary with movement of ice sheets, corresponding sea level fluctuations, and global environmental changes, heavily impacted worldwide biota (Batchelor et al. 2019; Allen et al. 2020; Gowan et al. 2021). In the northern hemisphere, species adapted to the temperate climate were forced to adapt in cold periods, move or retreat to refugia, with populations decreasing or going extinct in the process; warming episodes allowed for population expansion, colonisations, and adaptive divergence (Pamilo and Savolainen 2004; Lorenzen et al. 2011; Martchenko and Shafer 2023). The peopling of North America ~15,000 years ago (kya; Willerslev and Meltzer 2021) followed by human expansion was also a driver of regional ecological shifts (Gajewski et al. 2019; Fulton and Yansa 2021; Commerford et al. 2022), although there is no consensus on whether this was involved in large mammal declines across the continent (Stuart 2015; Meltzer 2020; Stewart et al. 2021).

The more recent Anthropocene has dramatically shaped demographic trajectories of wildlife through farming, hunting and logging (Ceballos et al. 2015; IPCC 2022). Human pressures increased during the colonial era when settlers profoundly altered the land, leading to declines in endemic populations (Woinarski et al. 2015; Lindo et al. 2016; Tavares et al. 2019; Brain and Prosser 2022; Stevens et al. 2022). Indeed, colonial impact has been associated with the extinction of several species such as Australian rodents, and the blue antelope (*Hippotragus leucophaeus;* Roycroft et al. 2021; Hempel et al. 2022). Beyond the demographic impacts, European populations might also have altered selection pressures, with recent examples including brown bears (*Ursus arctos*) and moose (*Alces alces*; Allendorf and Hard 2009; Kvalnes et al. 2016; Van De Walle et al. 2018).

Severe population contractions, or bottlenecks, increase genetic drift and inbreeding risk, and while highly deleterious mutations can be purged, mildly deleterious mutations tend to accumulate (Frankham 2005; Charlesworth 2009) as seen in alpine ibex (*Capra ibex*) and woolly mammoth (*Mammuthus primigenius*; Palkopoulou et al. 2015; Grossen et al. 2020). Population size fluctuations also leave discernible genome-wide patterns: a declining population will exhibit a positive Tajima's *D* and a relatively flat site frequency spectrum (SFS), whereas an expanding population will present a negative Tajima's *D* and a SFS skewed towards low-frequency or rare alleles (Hahn 2018). Depending on the extent of the contraction, both in magnitude and duration, populations might also present with a low effective population size (N_e), particularly if the bottleneck was strong and the recovery long (Martchenko and Shafer 2023). In contrast, a population that remained steady over time, or one that recovered quickly from a bottleneck, will usually exhibit a larger N_e (Charlesworth 2009). Genomic data can be used to model changes in historical and, due to recent analytical advances, contemporary trends of N_e (Li and Durbin 2011; Schiffels and Durbin 2014; Santiago et al. 2020; Schiffels and Wang 2020), and when compared to census size (N_c), both historical and contemporary estimates of N_e/N_c reflect adaptive potential and extinction risk (Palstra and Ruzzante 2008; Wilder et al. 2023).

The species of the North American genus *Odocoileus* (deer) were heavily impacted by the glacial cycles of the Pleistocene and European colonisation. Divergence estimates suggest the speciation between white-tailed (*O. virginianus*; WTD) and mule deer (*O. hemionus*; MD) took place between 500 kya and 4.3 million years ago (mya; Baccus et al. 1983; Kessler et al. 2023), the genus became among the most successful Cervids during glacial cycles (Hewitt 2011). Both species survived in different

refugia during the Last Glacial Maximum (LGM, 26 - 19 kya (Clark et al. 2009; Ellsworth et al. 1994; Greenslade 1998; Latch et al. 2009; Latch et al. 2014) and share mitochondrial haplotypes due to incomplete lineage sorting (Klicka et al. 2023). While WTD & MD hybridise (eg. Carr and Hughes 1993; Russell et al. 2021), the hybridisation appears to be a result of secondary contact and did not impact speciation (Kessler et al. 2023).

Deer abundance is heavily monetised through hunting-related activities in North America (Cambronne 2013), and both species have been, and continue to be, important to indigenous communities as a source of food, clothing, and culture (Adams and Hamilton 2011; Peres and Altman 2018). In the late Holocene, WTD were likely the most abundant large-mammal prey in eastern North America, and while some Indigenous Peoples depended on them for subsistence, harvest appeared to minimally impact WTD numbers (Wolverton et al. 2008; Weitzel 2021). In contrast, over-hunting and forestry practices post European-colonisation resulted in heavily depleted and even extirpated WTD populations, leading to strict hunting regulations followed by restocking efforts across the USA in the 20th century (McDonald et al. 2004). Stocking efforts have left discernible admixture patterns in WTD genome (Chafin et al. 2021). Mule deer abundance is, and was, lower than WTD; their populations also declined post-colonisation and plunged in the early 1900's (Jensen et al. 2023). Stocking efforts were less pervasive for MD, and the species stabilised due to effective management strategies (Gruell 1986; Clements and Young 1997; Gill 1999; Bergman et al. 2015; Jensen et al. 2023). Previous historical demographic analysis for both species have generated both contrasting and coarse demographic reconstructions (Combe et al. 2021; Lamb et al. 2021).

Here, we aim to understand the consequences of climatic and anthropogenic pressures on WTD & MD populations through explicit demographic modelling and selection analyses. We hypothesise that the population genetic histories of North American deer were heavily impacted by the glacial cycles of the Pleistocene and by European colonisation, not the initial peopling of North America. Specifically, we predict that the LGM had a strong negative impact in the form of demographic contraction and population subdivision, and that the impact of European colonisation, notably overhunting and change in land use, would leave detectable signatures on the genome in a neutral and adaptive context. To test these predictions, we used whole genome sequence data of MD and WTD from across their North American range to compute summary statistics, reconstruct and model demographic histories, and quantify genomic variation linked to environmental factors and selective forces.

Materials and Methods

Sampling & Sequencing

We sequenced the genomes of 45 deer collected across North America (Figure 3.1, Table S3.1) and added WGS data from 28 published deer genomes (PRJNA830519; Kessler et al., 2023). The complete dataset comprised 20 MD and 53 WTD samples, including four WTD individuals from the Florida Keys subspecies (*O. v. clavium*). We extracted DNA using the Qiagen DNeasy Blood and Tissue Kit following manufacturer's instructions and checked concentration using Invitrogen Qubit assays. DNA was sent to The Centre for Applied Genomics in Toronto, Canada, for library preparation following the Illumina TruSeq PCR-free DNA Library Prep, and sequencing to generate 150bp paired end reads. All samples were sequenced to an average of 4x coverage on an Illumina HiSeqX, except for four samples which were re-sequenced to achieve higher

coverage (>15x; Table S3.1, Figure 3.1A). This 4x coverage was targeted as we generated genotype likelihoods that are designed for lower coverage data (Meisner and Albrechtsen 2018).

We obtained estimated census data for WTD and MD in the USA for the last 400 years from Webb (2018; complemented by Deer Friendly). This census data was generated using a process of environmental scanning that aggregated data from different sources such as state agencies, harvest records and historical sources for the most ancient timepoints. We also digitised restocking information for WTD in the USA from McDonald et al. (2004).

Data processing

We analysed and trimmed raw reads with FastQC (v0.11.9; Andrews 2010) and Trimmomatic (v0.36; Bolger et al. 2014), respectively. Using bwa-mem and default settings (v0.7.17; Li and Durbin 2009), we aligned raw reads to the WTD genome (GCA_014726795.1) and sorted the reads using SAMtools sort (v1.10; Li et al. 2009). We identified and removed duplicates using Picard MarkDuplicates (v2.23.2; Broad Institute, GitHub Repository. 2019) and Sambamba view (v0.7.0; Tarasov et al. 2015). We used GATK RealignerTargetCreator and IndelRealigner (v4.1.7.0; McKenna et al. 2010) to carry out a local re-alignment and used Sambamba flagstat, mosdepth (v0.3.1; Pedersen and Quinlan 2018) and MultiQC (v1.10; Ewels et al. 2016) for quality checks.

For samples sequenced on multiple lanes, we specified the read groups before removing duplicates and merged the sequencing files using Picard AddOrReplaceReadGroups and MergeSamFiles. To have a consistent coverage among samples for non-SMC demographic analyses, we down-sampled data to the average coverage of all other samples (4x) using Picard DownsampleSam. We called the

genotypes (-doGeno 4) and estimated genotype likelihood using the GATK model (-gl 2) in ANGSD (v0.918; Korneliussen et al. 2014). We retained SNPs with a minimum base quality and a minimum mapQ quality of 20 (-minQ 20 & minMapQ 20) and a minimum p-value of 1e-6 (-SNP_pval 1e-6).

Population subdivisions and summary statistics

We used PCAngsd (v1.02; Meisner and Albrechtsen 2018) to investigate population clustering between and within species, this program fits best our data as it is designed explicitly for low-coverage data and genotype likelihoods. We explored the relationship between the spread on the PCA and the geographic distribution of our samples using a Pearson's correlation test in R (v4.2.0; R Core Team 2021). We further generated an individual genetic distance matrix using the gene.dist() function from the ape R package (Paradis and Schliep 2019), with the dataset produced for the δaδi analysis (see below). Finally, to help decipher population structure in the MD samples, we further ran NGSadmix (Skotte et al. 2013), using K = 1 : 10 and 10 replicates, we identified the best K value using CLUMPAK (Evanno et al. 2005; Kopelman et al. 2015). We introduce the following convention to refer to populations supported by PCAngsd and NGSadmix and used in subsequent analyses: mainland WTD = WTD_{ML}, Florida Keys WTD = WTD_{KEY}, MD mule deer subspecies = MD_{MD}, and MD black-tailed deer subspecies = MD_{BTD}.

We computed the genome-wide Tajima's *D* for all populations following the ANGSD pipeline (http://popgen.dk/angsd/index.php/Thetas,Tajima,Neutrality_tests). Briefly, we produced a folded SFS with angsd -doSaf 1 and realSFS, computed theta with realSFS saf2theta followed by Tajima's *D* using thetaStat. Further, we converted the hardcalls generated in ANGSD (.geno file) into a .vcf file using a python script

(genoToVCF.py; Martin 2021) and retained only biallelic sites (--max-alleles 2) in VCFtools (v0.1.16; Danecek et al. 2011). Using VCFtools we computed weighted F_{ST} between the four populations, and nucleotide diversity (π) per population, all processed in windows of 50kbp. We used PLINK (v1.90b6.21; Chang et al. 2015) to estimate the inbreeding coefficient based on runs of homozygosity (F_{ROH}) on the 160 scaffolds representing the N90 of the WTD's genome assembly, using the following non-default criteria: 100kb window size containing a minimum of 50 SNPs, allowing up to three heterozygote sites and 10 missing calls per called ROH. We computed F_{ROH} per individual using the sum of ROH sizes divided by the sum of N90 scaffolds length.

Historical and recent demographic inference

We inferred the historical demography of WTD_{KEY}, WTD_{ML}, MD_{MD} and MD_{BTD} by implementing the multiple sequentially Markovian coalescent (MSMC2) model (Schiffels and Durbin 2014) that generates estimates of N_e over time. Here, we used four high-coverage samples, one per population (Figure 3.1A, Table S3.1), and first generated a 35-mer mappability mask file using SNPable, as required for MSMC2 with several samples (Schiffels and Wang 2020). Because of the small sample size of high coverage genomes, we phased the data with WhatsHap (v1.3; Patterson et al. 2015; Martin et al. 2016) and filtered out scaffolds shorter than 500 Kbp. We ran MSMC2, including 100 bootstraps, using 1×2+25×1+1×2+1×3 as time segment pattern. To estimate the cross-coalescence rate (CCR) across populations, we carried out a pairwise comparison of the first haplotype from each sample using the -P flag, skipping any site with ambiguous phasing. Finally, to test and account for ancestral migration, we implemented MSMC-IM (Wang et al. 2020), assuming a generation time of 2 years (Deyoung et al. 2003) and a mutation rate of 1.23×10^{-8} mutations/site/generation

(average from Table S16 in Chen et al. 2019). We complemented this inference with a stairway plot analysis (v2.1.2; Liu and Fu 2015; Liu and Fu 2020) to reconstruct historical population size changes. Using all samples in each population, we generated a folded site frequency spectra in easySFS (v0.0.1; Gutenkunst et al. 2009) and used the resulting SFS as input to stairway plot, along with the same mutation rate and generation time as used for MSMC2.

We used GONE (Santiago et al. 2020) to compute demographic history in the recent past from patterns of linkage disequilibrium (LD) for the four populations. We restricted this analysis to the 160 scaffolds representing the N90 of the WTD assembly. We ran GONE with 500 repetitions on 500 generations with 30,000 SNPs per chromosome, a minor allele frequency pruning of 0.01, a hc value of 0.04, a recombination rate of 0.5 for WTD populations and 0.35 for the MD populations (Kessler et al. 2023). We compared the census size (N_c) point estimates from Webb (2018) to the temporal estimates of N_e resulting from GONE to obtain a contemporary N_e/N_c ratio over time.

We further filtered the dataset in VCFtools to: 1) account for LD by removing sites that were within 10kbp of each other (--thin 10000); and 2) remove all missing data (--max-missing 1). To model explicitly the demographic history of WTD and MD, we used δaδi (v 2.1.1; Gutenkunst et al. 2009) and dadi_pipeline (v3.1.7; Portik et al. 2017) and performed two different demographic analyses with different outgroups: 1) WTD_{KEY}, WTD_{ML} and all MD, and 2) MD_{MD}, MD_{BTD} and WTD_{ML}. We selected relevant demographic models involving migration and asynchronous splitting (Table S3.2), optimised and fitted the models with four repetitions and default settings, the best model was selected based on the log-likelihood and its consistency across runs. This model was

additionally run with 100 bootstrapped frequency spectrums to quantify uncertainty. We obtained ancestral N_e/N_c ratio over time for both species using the ancestral N_e (NuA) from the δaδi analyses.

Adaptive divergence and selection scans

We used a redundancy analysis (RDA) to detect genotype-environment association and loci under local adaptation across individuals, RDA is a multivariate regression method that allows to identify loci associated with a variety of predictors (Legendre and Legendre 2012). In our case, the predictors were 19 bioclimatic variables (Table S3.3) which we collected at a resolution of 2.5' from WorldClim2 using the R package raster (v.3.6-3; Hijmans 2022; Fick and Hijmans 2017) and extracted the values corresponding to our samples' coordinates. We performed three analyses: one inter-specific with all individuals, and two intra-specific for WTD & MD using 19 bioclimatic variables, latitude, longitude, and species as predictors. As highly correlated variables are problematic in RDAs, we removed those that showed a Pearson correlation > 0.7 in each analysis (Dormann et al. 2013). We carried out the RDA with the R package vegan (v.2.6-4; Oksanen et al. 2022) based on the datasets created for demographic modelling. In addition to full models including non-correlated predictors, we investigated conditional models with the condition set as either the bioclimatic variables or longitude.

To detect regions under selection in the four populations, we computed the integrated haplotype score (iHS) on the scaffolds representing the N90 of the WTD reference genome. This statistic is based on extended haplotype decay to detect selective sweeps by comparing the state of decay between ancestral and derived allele (Sabeti et al. 2002; Voight et al. 2006). The state of decay for all alleles should be similar

in neutral regions and iHS should equal zero in the absence of selection. In the case of a selective sweep and as one allele will rise in frequency, haplotype decay will be slower and iHS values will become either positive or negative depending if the sweep takes place on the ancestral or derived allele, respectively (Voight et al. 2006). As this analysis included all samples, we started by phasing our data in SHAPEIT (v2.r904; O'Connell et al. 2014), then calculated and normalised iHS in vcflib (v1.0.3; Garrison et al. 2022) with the functions iHS and normalize-iHS and default settings. We identified peaks in the most extreme 1% iHS values in R using the function findPeaks (Grinberg 2019) and extracted gene IDs within 25Kbp upstream and downstream. The genes with available entry IDs were uploaded to Uniprot (Pundir et al. 2016) for ID mapping where we retrieved the gene names and functions. The gene list was manually curated, and each gene was assigned to one or more of the following categories based on its function: reproduction, immunity, physiology, metabolism, sensory perception, development, and other. We analysed all results and produced all the graphics in R.

Results

Population subdivisions and summary statistics

We called 144,634,190 SNPs across species from 73 individuals at a coverage of 4x on average. The PCA on allele frequencies of all individuals presents a clear species delimitation along PC1 which explained 46% of the variation, whereas PC2 separated the WTD_{KEY} individuals (Figure 3.1B). WTD-specific PCA showed the delimitation between WTD_{ML} and WTD_{KEY} individuals (PC1); PC2 correlated with longitude (Pearson correlation = -0.88, p-value = < 2.2e-16) and accounted for 7% of the variation (Figure S3.1A). In the MD PCA, PC1 correlated to longitude (Pearson correlation = 0.83, p-value = 5.96e-06), and PC2 was associated with latitude (Pearson correlation = 0.47, p-value =

0.037, Figure S3.1B). The MD PCA showed one clear cluster of 13 samples; we clarified the population identity of the seven remaining individuals using NGSadmix that assigned two admixed individuals to the first cluster (MD_{MD}) and identified a second population of five individuals (MD_{BTD}; Figure S3.1C). Individual genetic distances separated MD and WTD_{KEY} individuals from WTD (Figure S3.1D). The individual genetic distance matrix mirrored nucleotide diversity estimates which were highest in WTD_{ML} and similar in the other three populations (Table 3.1, Figure S3.1D).

Weighted F_{ST} measures were high in between species comparisons (0.36 - 0.58); intra-specific comparisons ranged from 0.12 to 0.17 (Table S3.4). Tajima's *D* values were negative only for the WTD_{ML} population (-0.76), the highest being 1.08 for MD_{BTD} (Table 3.1, Figure S3.2A). F_{ROH} values were lowest in the MD populations (<= 0.3%), whereas WTD_{KEY} presented the highest F_{ROH} value (1.4%; Table 3.1; Figure S3.2B).

Demographic inferences

We inferred the demography of four high-coverage samples, one per population, using MSMC2. The trajectories of the four samples split at approximately 600 kya where the MD_{MD} population start declining whereas WTD populations kept increasing (Figure 3.2A). Both MD population trajectories were similar and showed a general steady decline. WTD_{KEY} appear to decline ~50 kya, whereas WTD_{ML} drastically increased to peak at an N_e of >1 million individuals, followed by a clear crash at the end of the LGM (Figure 3.2A). Species split times inferred by the cross-coalescence rate of the MSMC-IM model is approximately 500 kya (Figure S3.3A), the within-species split times are estimated at ~ 25 kya between WTD_{ML} and WTD_{KEY}, and ~70 kya for MD_{MD} and MD_{BTD} (Figure S3.3B). Effective population size over time as inferred by the stairway plot analysis showed similar historic trajectories where all populations have declines

surrounding the LGM (Figure S3.4); however, both MD_{BTD} and WTD_{ML} presented a large recovery post-LGM followed by a recent crash; this is consistent with model overfitting in a stairway plot that produces spurious and complex patterns (Lapierre et al. 2017).

The contemporary N_e of WTD_{KEY} appears very low and stable for the last ~400 years, reaching its lowest at ~350 individuals (Figure 3.2B). Temporally, the WTD_{ML} population decreased with variable intensity that ended with a drastic decline to a N_e of ~2,000 near the end of the 19th century. Following this severe drop, the WTD_{ML} population rapidly rebounded to reach a current N_e in the millions (Figure S3.2B); this clearly correlates with census data and restocking efforts (Figure 3.2C). Both MD populations exhibited a slow but steady increase until approximately 300 years ago, after which both trajectories showed a gradual decline ending with a severe drop without recovery (Figure 3.2B). MD N_e patterns are more decoupled from the census data as N_c here showed a crash in both MD_{MD} and MD_{BTD} followed by a small recovery (Figure 3.2C). As such, the contemporary N_e/N_c ratio is higher and more variable for MD (Figure 3.3; mean = 0.49, var = 0.3 vs WTD mean = 0.18 & var = 0.01) with most values from the last century below 0.1. Both species have a low ancestral N_e/N_c ratio with little variability, the averages are of 0.06 and 0.01 for WTD and MD, respectively. White tailed deer's historical ratio (var = 0.03) is more variable than MD (var = 0.0001) and the contemporary WTD (Figure 3.3).

Finally, we modelled the demographic history of our four populations using the SFS. For our first comparison, we analysed the demographic history of the trio WTD_{KEY} – WTD_{ML} – MD, the best model supports adjacent secondary contact between populations, though migration rates were low ($m < 1e^{-5}$; Figure 3.4A, Table S3.5). This model suggests a MD/WTD split ~280 kya followed by a MD decline and a WTD

expansion. After the split between WTD_{KEY} and WTD_{ML}, both populations declined with notably a demographic crash in WTD_{KEY}; all of these predate the peopling of North America. In our second trio, we investigated the relationship between MD_{MD} – MD_{BTD} – WTD_{ML}. Here, the split WTD/MD is seen ~390 kya, the best model suggests a secondary contact between WTD_{ML} – MD_{MD} that predates the split between the two MD populations, ~13 kya, but all migration rates are extremely low ($m <= 4e^{-6}$). This model also suggests population expansion for WTD_{ML} after the split from MD, but a decline for both MD_{BTD} and MD_{MD} before and after the peopling of North America (Figure 3.4B, Table S3.5).

Adaptive divergence and selection

We conducted three redundancy analyses to identify genotypes linked to environmental variables and loci under local adaptation. The RDA with all individuals explained 16.5% of the variation, most of which was driven by the species variable (Table 3.2, Figure S3.7A). The full model for the WTD explained 6.32% of the variation and included five significant predictors - BIO1 (annual mean temperature), BIO2 (mean diurnal range), BIO8 (mean temperature of wettest quarter), BIO15 (precipitation seasonality) and longitude (Table 3.2, Figure S3.7B). There were 12,206 outlier SNPs, predominantly associated with annual mean temperature (Table 3.2 & S3.3, Figure S3.7C). In the RDA on MD, we found only BIO2 and BIO16 (precipitation of wettest quarter) as significant predictors, and 7434 outlier SNPs, mostly associated with BIO16 (Table 3.2). Conditional models of bioclimatic variable explained more variation than those of longitude in both species (Table 3.2).

We found a total of 443 genes within 25 Kbp upstream and downstream of an iHS peaks, including 249 with an existing Entry ID (Table S3.6). There were 121 genes around iHS peaks in the WTD_{ML} population, 45 of which we could assign an explicit

category including 18 genes assigned to immunity, of which two related to the major histocompatibility complex (MHC): *HLA-DQA1* and *HLA-DQB*. Additionally, we found two genes linked to sensory perception: one odorant receptor (*OR5D18*) and one cold receptor (*TRPM8*, Table S3.6). We identified 23 genes surrounding iHS peaks in the WTD_{KEY}, including one MHC gene: *HLA-DRB4*. For the MD_{MD} population, we could explicitly categorise 34 genes on a total of 95 neighbouring iHS peaks. They include four MHC genes: *HLA-DQA1*, *HLA-DQB1*, *HLA-DRB3* and *HLA-DRB4,* another gene related to the MHC: *HIVEP2*, three genes described as olfactory or odorant receptors (*OR51I1*, *OR5D18*, *TAAR5*), and one cold receptor (*GRIK2*). Finally, we identified 26 genes adjacent to iHS peaks in the MD_{BTD} population, including two related to immunity and three linked to reproduction

Discussion

The complex climatic cycles of the Pleistocene, concluding with the LGM, generally forced species to move from a wide range into smaller, often isolated refugia (Soltis et al. 2006; Shafer et al. 2010). The peopling of North America and the colonisation by Europeans several millennia later represented further changes and pressures to many populations inhabiting the continent (McDonald et al. 2004; Lindo et al. 2016; Smith et al. 2021). Here, we reconstructed the demographic history and signals of selection of WTD & MD to improve our understanding of climatic and anthropogenic pressures on these two North American species. Through extensive genomic analysis we observed a remarkable overlap between climate, recent human drivers, and N_e, and provide a temporal N_e/N_c assessment supporting the demographic concerns facing MD (Bergman et al. 2015; Webb 2018). Further, our selection scans that reflect climate and

disease pressures provide testable hypotheses for the impact of European colonisation.

Strong historical climatic impact on deer populations

The glacial cycles of the Quaternary forced populations to move or adapt, resulting in population divergence (Colella et al. 2021; Ito et al. 2021), decline (Ersmark et al. 2019; Baca 2020; Dussex et al. 2020; Meiri et al. 2020) and even replacement (Loog et al. 2020). Our MSMC2 results suggest a clear impact of climate on WTD during the late Quaternary as we observed an intense N_e drop for WTD_{ML} as well as a population split time that coincides with the LGM (Figure 3.2A, Figure S3.3B), supporting WTD populations being separated into different refugia during this period (Ellsworth et al. 1994; Moscarella et al. 2003; Combe et al. 2021; Wright et al. 2022). Results from the δaδi analysis do not pinpoint the LGM, but rather suggests population declines and subdivision approximately 50 kya, which coincides with large-scale climatic changes of the mid-Wisconsinian, 30 - 60 kya (Figure 3.4A; Fréchette and De Vernal 2013; Kerr et al. 2021). While we present here the best model for our data, both δaδi comparisons still present with relatively large residuals (Figure 3.4) suggesting the true story might be more complicated than depicted here. Additional modelling with different approaches might help render a more precise history. The Key deer showed a steady decline that predated the LGM, with a clear separation from WTD_{ML} ~50 kya (Figure 3.2A, Figure 3.4A, Table S3.5). All analyses suggest both populations started diverging before the isolation of the Florida Keys from the continent ~8,000 years ago (Ellsworth et al. 1994; Villanova et al. 2017). This isolation led to a clear loss of genomic diversity (Table 3.1, Figure S3.2) and accumulation of deleterious alleles in the endangered Key deer (Cars et al. 2023;

U.S. Fish & Wildlife Service 1967). Overall, the collective demographic methods used here pinpoint climate as a driver of subdivision and historical declines in WTD.

The LGM has also been implicated in the MD_{MD} and MD_{BTD} (*O. h. columbianus*) subdivision, as the two subspecies were separated in different refugia during that time (Latch et al. 2009; Latch et al. 2014; Wright et al. 2022). Our results are equivocal regarding the split time: the MSMC2 analysis suggest a steady decline of both populations and a MD/BTD split estimated at approximately 70 kya by the CCR (Figure 3.2A, Figure S3.2B); δaδi analysis, in contrast, places the split after the LGM and concomitant with a decline of ~30% in both populations (Figure 3.4, Table S3.5). Nevertheless, both analyses support changes pre-peopling of North America and are consistent pre-split: the N_e of the common ancestor of MD_{MD} and MD_{BTD} appeared low but stable through time (Figure 3.2A). As MD are more prone to population fluctuations (Forrester and Wittmer 2013), and N_e is most strongly influenced by long periods of low N_c (Peart et al. 2020), the overall stable historical N_e is consistent with population genetic theory.

Hahn (2018) suggested to focus on big picture, not point estimates, for demographic inferences, and while SMC reconstructions differ in magnitude based on data treatment, overall time and shape remains constant (Schiffels and Wang 2020). While the general historical patterns of demographic change observed here support previous assessments (Lamb et al. 2021), including the LGM impact in WTD, they are in contradiction with Combe et al. (2021). Combe et al's (2021) stairway plot analysis of ~25k SNPs suggests a single population contraction ~100 kya followed by an expansion for WTD, and population plateau at high number for over 200 kya before a drastic decline for MD. These times are inflated due to their generation time of five years versus

two used here (see also Deyoung et al. 2003). Simulations, have shown that a dataset of this size is cannot reliably resolve changes in N_e (Shafer et al. 2015), thus our explicit modelling and genome-wide data set has significantly more power and precision with respect to historical demographic inference. Stairway plots are also prone to over-fitting and erroneous reconstructions (Lapierre et al. 2017); for example, the explosive LGM recovery suggested by our WTD_{ML} stairway plot (Fig S3.4) and Combe et al. (2021) is at odds with our all other results and patterns seen in comparable species (e.g. Dussex et al. 2020; Taylor et al. 2021).

None of the deer populations analysed here showed any signal of recovery after the LGM and ice sheets recession. This is particularly striking given the massive loss of large mammals on the continent (Elias and Schreve 2007; Stuart 2015; Meltzer 2020), presumably opening new niches, reducing competition and predation. One explanation could be that the primary cause of population declines in deer were the environmental shifts caused by the deglaciation dynamics (Mottl et al. 2021; Hanberry 2023) and the changes in ecosystem function after the megafaunal extinctions (Malhi et al. 2016; Malhi et al. 2022). While the decline of mainland WTD overlapped to some degree with human presence (Bennett et al. 2021), early human populations were not large enough to drive the dramatic decline suggested by MSMC2. Further, given the trajectories of the three other populations analysed here, for which the declines start earlier than the peopling of North America, and those of comparable species (e.g. Dussex et al. 2020; Taylor et al. 2021), it strongly suggests that Pleistocene deer declines were not driven by human impact. The lack of full recovery is consistent with population dynamics of other temperate ungulate species (De Jong et al. 2020; Dussex et al. 2020; Taylor et al. 2021; Robin et al. 2022). Nevertheless, the limits of detection of MSMC2 and

GONE analyses (10 kya – 0.4 kya) do leave a gap in the Holocene for which it is difficult to reconstruct demographic history; however, δaδi should pick up a large Holocene change (e.g. Dedato et al. 2022), or the pattern seen in the stairway plot analysis, were they to have actually occurred.

Hybridisation between WTD and MD in areas of sympatry is well documented (e.g. Cronin 1991; Derr 1991; Carr and Hughes 1993), but it was suggested that introgression had no impact on species divergence and rather that gene flow is a result of secondary contact (Kessler et al. 2023). Likewise, Klicka et al. (2023) suggested incomplete lineage sorting of mitochondria, not historical hybridisation, explains shared haplotypes between species. This is consistent with our analyses as both δaδi comparisons suggest a secondary contact with low migration rates (Figure 3.4, Table S3.5), and while some ancient migration is detected, it is extremely low (Figure 3.4B, Table S3.5, Figure S3.3C). Our results therefore support the finding of negligible ancient gene flow between species. In fact, our analyses suggest a species split time under 500 kya (Figure 3.4, Table S3.5, Figure S3.3), which are more recent than previous molecular clock evaluations ranging between 750 kya and 4.3 mya (Baccus et al. 1983; Douzery and Randi 1997; Combe et al. 2021; Wright et al. 2022). All of these estimates were based either on a restricted number of nuclear and mitochondrial loci, presenting lower power of analysis than our whole genome assessment.

Human-induced collapse, varying selection pressures on white-tailed deer

Driven by overharvest and change in land use for agriculture and logging, the recent decline and near extirpation of WTD has been thoroughly documented in the USA (McDonald et al. 2004). The reconstructed demographic trajectory of the mainland WTD population mirror these historical sources and census size estimates (Figure 3.2B-

C); specifically, we captured the gradual decline followed by dramatic collapse at the end of the 19th century. This depletion led local authorities to implement protective laws and managers to supplement deer populations with translocations (see McDonald et al. 2004 for an extensive summary). Those measures were responsible for the quick rebound we observed for the WTD_{ML} population (Figure 3.2B) as analyses of nearly extirpated populations suggest the WTD bottleneck had minimal impact on diversity thanks to the rapid population expansion following translocations (Deyoung et al. 2003; Budd et al. 2018). Together with the stable relationship of N_e/N_c over time we observed (Figure 3.3), these results suggest that WTD are highly resilient; however, despite the recovery and high diversity, assessment of deleterious alleles also showed a high genetic load for WTD (Cars et al. 2023; Wootton et al. 2023).

Given the low N_e and stocking efforts of WTD, it is probable that any sweep that took place before the demographic collapse was either lost or fixed, with the latter generally having reduced power in haplotype scans (Szpiech 2021). In contrast to ungulates experiencing reduced human intervention (e.g. Martchenko and Shafer 2023), longitude had a minimal impact in our models which is consistent with stocking efforts largely erasing geographic signals. Combined with evidence showing rapid contemporary evolutionary responses (Zamorano et al. 2023), we suggest that the outliers and sweeps detected in our analyses could be the result of selection during European colonisation that coincided with large changes to the landscape. In mainland WTD, we found a variety of immunity-related genes potentially under selection which would suggest selection pressures from pathogens (Table S3.6), some of which might have arisen recently from livestock (Campbell and VerCauteren 2011). The presence of selection at immune genes in WTD scans is consistent with the development of

agriculture and large land use changes that have transpired in North America over the past two centuries. Nevertheless, as immune genes are often detected in genome scans and as hyper-variable regions of the genome are susceptible to misalignments (Manel et al. 2016), these results should be taken with caution. Recent selection at immune loci, and indeed the colonial impact, could be investigated by using historical samples: here, any sweep signal detected in the present analysis not seen in historical or ancient DNA would suggest a colonial selection event.

We also found selection patterns at the gene *TRPM8* which is involved in sensitivity to cold temperature (Table S3.6). Mutations identified in woolly mammoth's *TRPM8* were suggested to be linked to the species' adaptation to cold climates (Lynch et al. 2015), purifying selection connected to this gene was suggested to cause temperature sensitivity loss in mammoths (Chigurapati et al. 2018). Moreover, in the RDA between species, as well as in the WTD RDA, most outlier SNPs are associated with mean annual temperature (BIO1) rather than species or longitude meaning that temperature is a major driver of differentiation. WTD are exposed to a wide range of temperatures on the continent, but also rapidly expanding northward (Dawe and Boutin 2016). As climate change clearly impacted historical and contemporary WTD ranges (Dawe and Boutin 2016), outliers and selection on genes like *TRPM8* are expected, and likely partly responsible for the success of WTD on the continent. Only few genes were identified as under selection for MD_{BTD} and WTD_{KEY}, given the fewer number of samples from those populations the power of the selection scans analysis was clearly reduced (Klassmann and Gautier 2022).

Delayed collapse but similar selection pressures on mule deer

Human Settlements on the MD historic range were estimated at less than 20 in 1800, and while harvest likely depressed populations in the surrounding areas, the rest of the range probably saw minimal impact (Jensen et al. 2023). Over time and with the increased settlements in western North America, MD populations declined to reach their lowest at the end of the 19th century (Gill 1999; Bergman et al. 2015; Jensen et al. 2023). This is consistent with our analysis which shows a sharp decrease in the MD_{MD} population during that period, though the MD_{BTD} population decline appeared later (Figure 3.2B). These population reductions also appeared to happen after the WTD collapse, possibly reflecting the later western wave of settlers and more recent land use changes. As a migratory species, MD might respond more to human-induced habitat change, for example, the limitation in winter range habitat appears to be driving MD decline in Colorado (Gill 1999; Bergman et al. 2015). Collectively the delayed decline, the decoupling of N_e from N_c (Figure 3.3), the positive Tajima's *D* and the more muted temporal response of N_e are consistent with the species ecology and human intervention history. Despite this collapse and lack of recovery, both MD subspecies exhibit relatively high genetic diversity and low F_{ROH} (Table 3.1, Figure S3.2B), all promising signals for a species that is often described as declining. Assessment of MD genetic load should help further inform the management challenges facing this species.

In MD_{MD}, we found 14 immunity-related genes under selection (Table S3.6). Diseases affecting MD are often the same that infect WTD, including the risk of spilling over to and from livestock (Campbell and VerCauteren 2011). Chronic wasting disease is of particular concern for MD as it is considered as a major factor in the species decline in many regions (DeVivo et al. 2017), with the genomic and immune response clearly

polygenic (Seabury et al. 2020). Regarding sensory perception, we found selection patterns at the cold receptor *GRIK2*. Cold adaptation might have been a determinant factor for MD as they favour higher elevation habitats (Anthony and Smith 1977b; Brunjes et al. 2006), and are more tolerant to cold climates (Mautz et al. 1985). Supporting this was outliers associated with mean diurnal range (BIO2), which is linked to temperature (Table S3.6). Most outlier SNPs were associated with precipitation of the wettest quarter (BIO16, Table 3.2), likely reflective of MD_{BTD} inhabitation of temperate rainforests (Heffelfinger and Krausman 2023).

Variable N_e/N_c ratio through time

The N_e/N_c ratio is informative on the genomic health and demographic histories of species, which is why it has often been used to assess populations of management and conservation concern (Frankham 1995; Palstra and Ruzzante 2008; Ferchaud et al. 2016). Indeed, a population exhibiting a low contemporary N_e/N_c ratio is more subject to rapid loss in genetic diversity than populations with a higher ratio (Frankham 1995; Palstra and Ruzzante 2008). While, this ratio is highly variable between taxa, the generally accepted median of concern in wild populations is 0.1 (Frankham 1995; Waples et al. 2013). Our unique integrated approach allowed us to measure this ratio over time, with the seven most recent values for MD all below 0.1 (Figure 3.3). The larger fluctuations in MD's N_c in the last 100 years appear to have led to a decoupling of N_e and N_c and reflect the species higher demographic stochasticity. This recent reduced ratio could mean that the species is at risk of faster loss of genetic diversity and echoes other calls of concern (Gill 1999; Bergman et al. 2015). Interestingly, historical N_e/N_c ratio presents low variability for both species and no immediate red flag (Wilder et al. 2023); this ratio is clearly driven by the high N_c of both species, and focused sampling of

isolated populations might reveal a different pattern as seen in caribou (M.N. Dedato et al. 2022). Denser sampling of MD might also help identify key corridors facilitating gene flow (e.g. Fusco et al. 2023).

Conclusion

We observed strong signals of climate- and human-induced demographic declines with a remarkable overlap between both census and biogeographic data and historical inferences coming from multiple methods. The dramatic drop in N_e in the recent past of all deer populations studied here is robust evidence for a colonial impact through unregulated hunting and changes in land use. Even though WTD & MD populations were heavily depleted, both species showed signs of recovery, particularly white-tailed deer from restocking efforts. While diversity and demographic trends appear positive, there is also a high genetic load in WTD (Wootton et al. 2023), and a low contemporary N_e/N_c ratio in MD. Selection surrounding a wide variety of genes, including some of biological importance in WTD & MD such as genes related to immunity and temperature appear evident. Historical sampling of *Odocoileus* will prove informative for disentangling human-induced selection and genetic load in these wide-ranging cervids.

Acknowledgements

Camille Kessler was supported by an International Graduate Scholarship, an Ontario Graduate Scholarship, and a French American Charitable Trust Scholarship. This work was supported by NSERC Discovery Grant (Grant Number: RGPIN-2017-03934); ComputeCanada Resources for Research Groups (Grant Number: RRG gme-665-ab); Canadian Foundation for Innovation: John R. Evans Leaders Fund and the Ontario Early Researcher Award (Grant Number: #36905). For providing samples, we thank Alan

Cain, Steeve Côté, Anh Dao, Marco Festa-Bianchet, Brad Fulk, Eric Hoffman, Levi Jaster, Daniel Koelsch, Emily Latch, Brent Patterson, Joe Nocera, Don Stewart, Russell Easy, Jonathan Shaw, David Walter and Jon Wheeler. We also thank Jose Alberto López-Alemán for commenting on the text and three anonymous reviewers.

We sampled deer from across North America, a stolen land which remains a home to many First Peoples and Trent University is located on the territory of the Michi Saagiig Anishnaabeg. As settlers, we are grateful to have had the opportunity to live and work on this land, and to benefit from it. We would like to show our respect to the First Peoples and thank them for their care, stewardship, and teachings. Miigwetch.

Figures and tables

Figure 3.1: Sample information. (A) Sampling locations and population assignment, triangles represent high-coverage samples, shaded areas represent each species' IUCN range with areas of sympatry in a daker shade (B) PCA of all individuals, based on allele frequencies. Colour coded by assigned population.

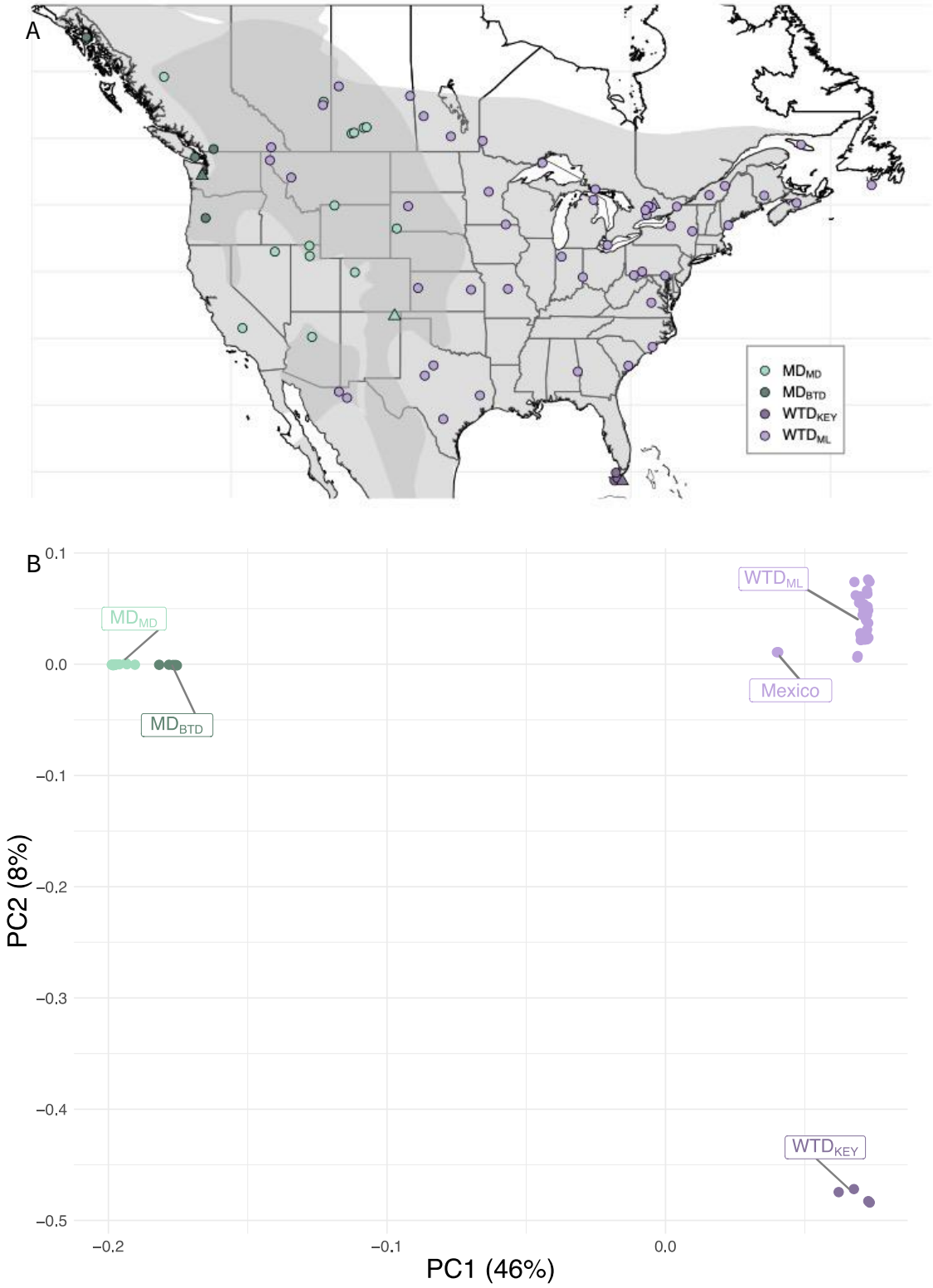


Figure 3.2: Demographic inference. (A) Changes in population size of four high coverage samples with bootstraps as estimated by MSMC2, shaded areas symbolise full glacial periods (Batchelor et al., 2019). (B) Recent changes in Ne computed by GONE on all samples with bootstraps. (C) Estimated WTD & MD population in the US from Webb (2018) with WTD restocking period for 30 US states (grey rectangles) and corresponding state's year with lowest WTD population (black points) from McDonald et al. (2004).

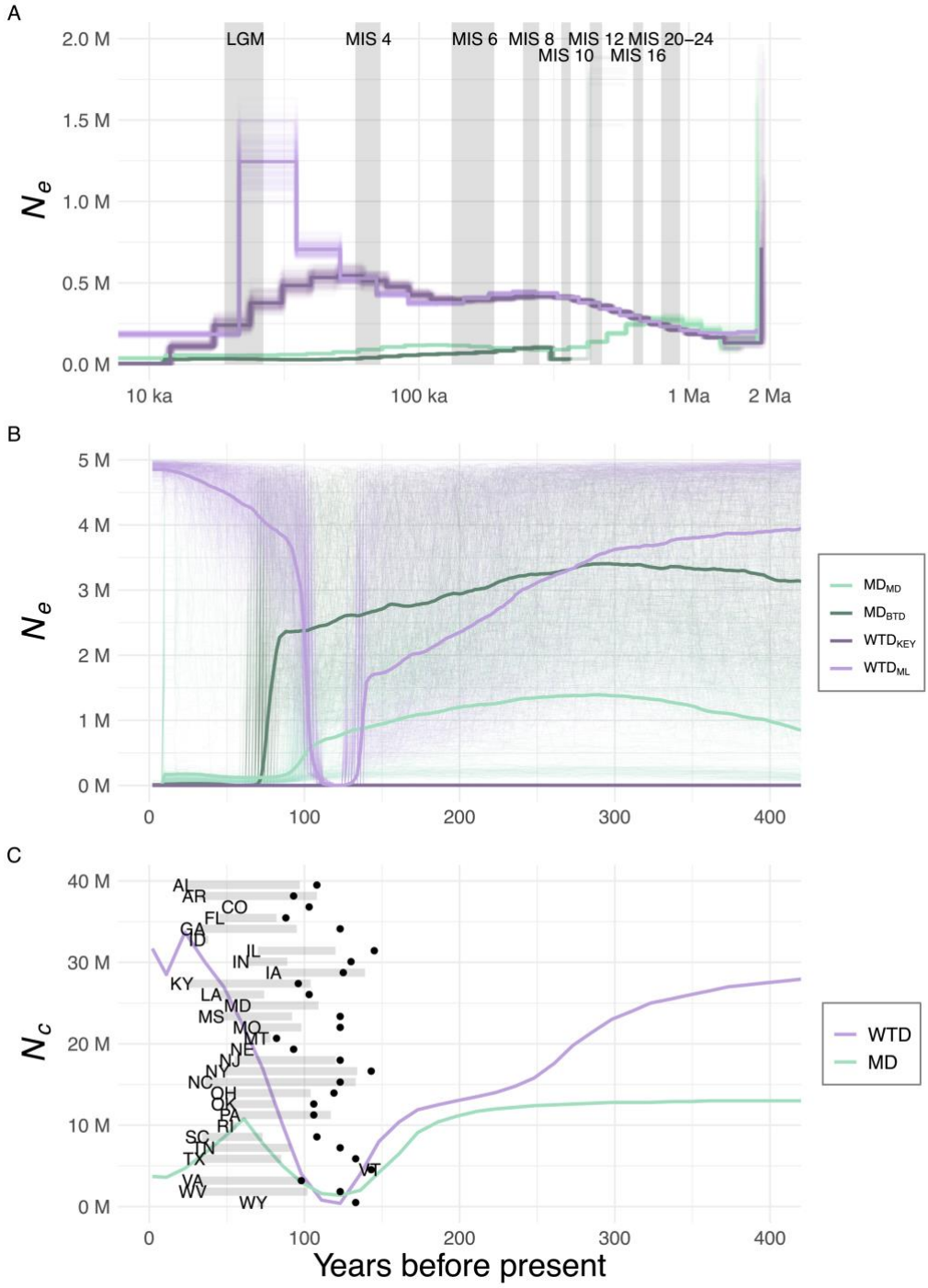


Figure 3.3: Contemporary and historical N_e /N_c ratio over time, point colour indicates time before present, N_e of the historical ratio taken from NuA values in the δaδi analysis.

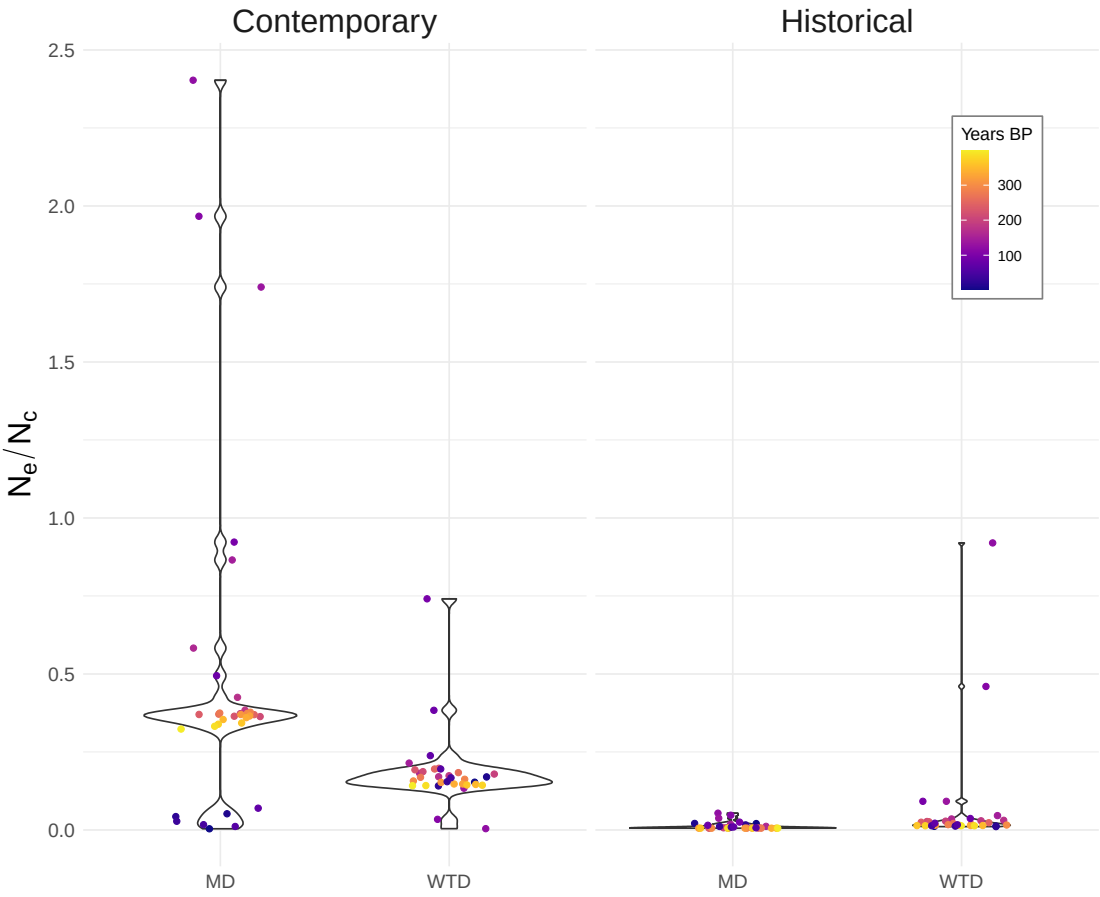


Figure 3.4: Three populations models in δaδi with best model schema on the left and 2D spectra on the right. (A) MD – WTD_{ML} – WTD_{KEY} comparisons, with MD in green, WTD_{ML} in light and WTDKEY in dark purple. (B) WTD_{ML} – MD_{MD} – MD_{BTD} comparison with WTD_{ML} in light purple, MD_{MD} in light green and MD_{BTD} in dark green. Population sizes (nu) and migration rates (m) estimates to scale, time is not to scale

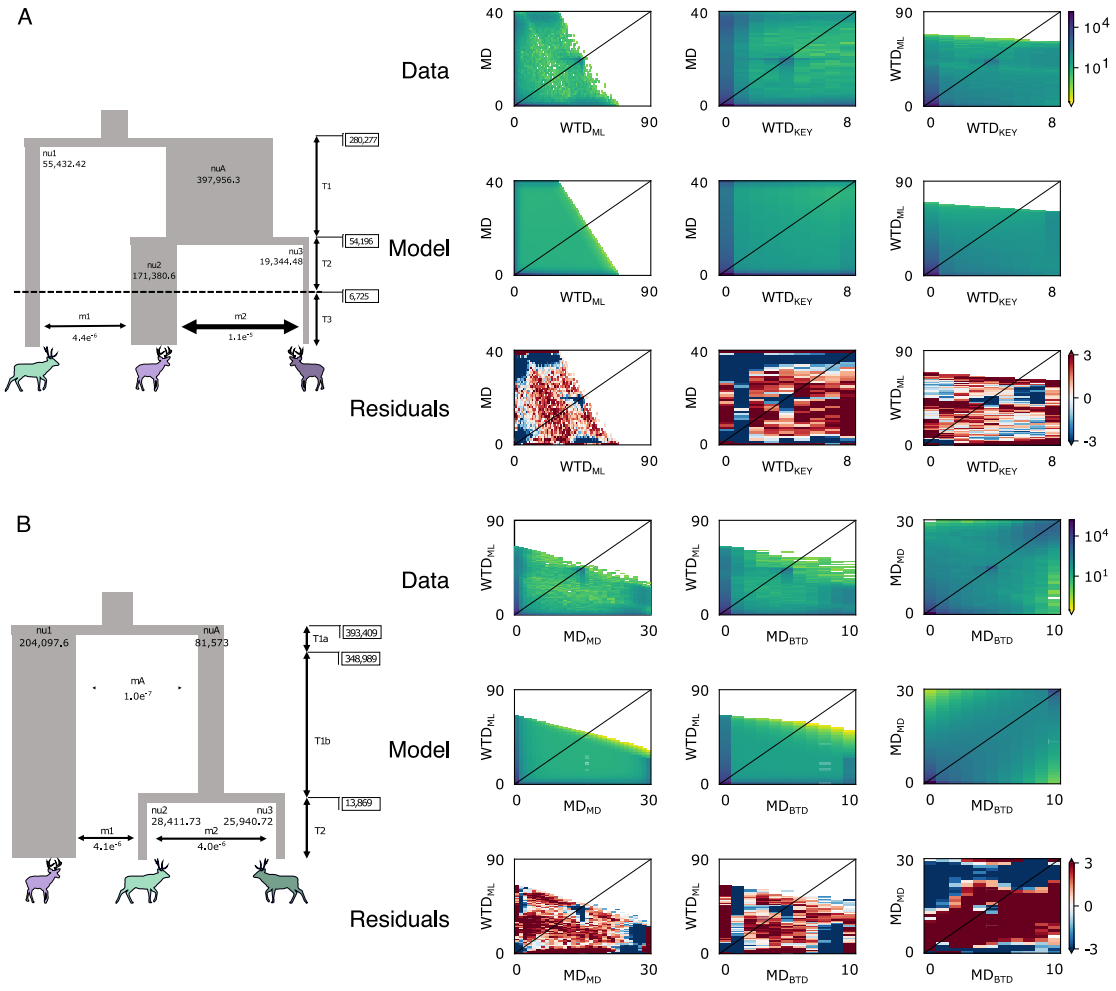


Table 3.1.: Tajima's D, nucleotide diversity and F_{ROH} for our four populations, F_{ROH} as percentage of the N90 scaffolds.

	Tajima's D	Nucleotide diversity (π)	F_{ROH}
WTD_{ML}	-0.76	0.0068	0.89
WTD_{KEY}	0.70	0.0038	1.38
MD_{MD}	0.81	0.0036	0.30
MD_{BTD}	1.08	0.0037	0.28

Table 3.2: Anova of the RDA for full and conditional models as well as number of outlier SNPs for each predictor in the full model. Significant p-values are highlighted in bold.

		Adjusted R^2	bio1	bio2	bio8	bio12	bio14	bio15	bio16	bio18	Longitude	Species
Interspecies												
	P-value		**0.004**	**0.001**	**0.012**	**0.009**	-	**0.001**	-	-	**0.001**	**0.001**
~ BIOs + Long + SP	Variance explained	16.55%	2.58%	3.12%	1.95%	1.95%	-	3.54%	-	-	3.55%	7.96%
	number of outlier SNPs		5448	768	760	2491		47			2820	2101
MD												
	P-value		0.093	**0.001**	-	-	0.116	0.355	**0.012**	0.374	0.219	-
~ BIOs + Long	Variance explained	8.82%	5.67%	9.82%	-	-	5.48%	4.82%	6.72%	4.83%	5.04%	-
	number of outlier SNPs		99	540	-	-	690	1	4452	0	1652	-
	P-value		0.119	**0.032**	-	-	0.167	0.336	**0.003**	0.476	-	-
~ BIOs + Condition(Long)	Variance explained	4.11%	5.96%	6.23%	-	-	5.84%	5.53%	7.29%	5.35%	-	-
	P-value		-	-	-	-	-	-	-	-	0.445	-
~ Condition(BIOs) + Long	Variance explained	0.35%	-	-	-	-	-	-	-	-	8.05%	-
WTD												
	P-value		**0.001**	**0.001**	**0.034**	-	-	**0.001**	-	-	**0.001**	-
~ BIOs + Long	Variance explained	6.32%	3.37%	3.89%	2.11%	-	-	2.99%	-	-	2.96%	-
	number of outlier SNPs		5625	4487	1025	-	-	209	-	-	777	-
	P-value		**0.001**	**0.001**	**0.038**	-	-	**0.001**	-	-	-	-
~ BIOs + Condition(Long)	Variance explained	4.64%	3.50%	3.57%	2.18%	-	-	2.94%	-	-	-	-
	P-value		-	-	-	-	-	-	-	-	**0.001**	-
~ Condition(BIOs) + Long	Variance explained	1.26%	-	-	-	-	-	-	-	-	3.38%	-

CHAPTER 4 : ANCIENT DNA OF THE TORONTO SUBWAY DEER ADDS TO THE EXTINCTION LIST OF ICE AGE MEGAFAUNA

Authors:

Camille Kessler, Oliver Haddrath, Burton K. Lim, and Aaron B. A. Shafer

A version of this chapter is under revision for *Current Biology*.

Contributions:

CK and ABAS conceived the study, OH and BKL provided the sample, CK performed the molecular laboratory work with contribution from OH, CK performed bioinformatic analyses, CK wrote the manuscript with general contribution from ABAS and contribution on taxonomy from BKL. All authors commented on the text.

Summary

The late Pleistocene was a time of global megafaunal extinctions that were particularly severe in North America. The continent lost many mammalian genera and species, but the validity of several of these taxa remains contentious. This ambiguity is best highlighted in North American Cervidae as the legitimacy of two of the four deer genera that went extinct is debated. A controversial taxon, *Torontoceros hypogaeus*, is represented by a single antler-bearing partial-skull specimen (ROMM75974) discovered in 1976 during the construction of the Toronto subway in Ontario, Canada. The species was described based on its unique antler morphology, but its near absence in the fossil record and the variable nature of antler morphology leads to uncertainty concerning the systematic relationships with other cervids. Here, we used whole genome sequencing of ancient DNA to clarify the taxonomic placement of ROMM75974 and the evolutionary history of Northern American deer. We performed mitochondrial and whole genome analyses with related cervids and showed that the specimen ROMM75974 has a close affinity, but relatively high divergence from the *Odocoileus* sister species white-tailed (*O. virginianus*) and mule deer (*O. hemionus*). With its unique antler morphology and divergent genome, *Torontoceros* should be recognised as a subgenus, and the species *Odocoileus (Torontoceros) hypogaeus* be included in the list of extinct North American taxa. Based on antler morphology, *O. (T.) hypogaeus* was likely adapted to open landscape; we hypothesise that its likely preferred habitat changed rapidly at the end of the Pleistocene, highlighting the role of climate change in the extinction of megafauna biodiversity at the end of the ice age.

Introduction

The late Pleistocene was characterised by an extinction event that particularly affected large mammals (Martin and Steadman 1999; Stuart 2015). The causes of these extinctions have been attributed to human impact, climate change, or a combination of the two (Barnosky et al. 2004; Koch and Barnosky 2006; Lorenzen et al. 2011; Broughton and Weitzel 2018; Meltzer 2020; Stewart et al. 2021; Bergman et al. 2023), although other factors have been suggested (Firestone et al. 2007; Dembitzer et al. 2022). This extinction event was severe in North America which lost between 33 and 37 large mammal genera (Martin and Steadman 1999; Barnosky et al. 2004; Koch and Barnosky 2006; Elias and Schreve 2007; Grayson 2007; Stuart 2015; Meltzer 2020; Stewart et al. 2021), and while some genera are consistently included in the extinction list (e.g., *Mammut*, *Equus*), others are not (e.g., *Holmesina*, *Capromeryx*). Four cervids (deer) disappeared at this time: *Bretzia, Cervalces, Torontoceros* and *Navahoceros,* which, together with the extant *Odocoileus* and *Rangifer*, represent the only genera in the deer family to have inhabited North America during the glacial cycles of the Pleistocene (Webb 2000; Boeskorov 2005; Heffelfinger 2011; Weckworth et al. 2012). The extant wapiti (*Cervus canadensis*) and moose (*Alces alces*) originated in Eurasia and colonised North America through Beringia after the Last Glacial Maximum (LGM; Meiri et al. 2014; Hu et al. 2019; Dussex et al. 2020; Meiri et al. 2020; Croitor 2022; Mackiewicz et al. 2022b).

Cervids first appeared in the North American fossil record approximately 5 million years ago (mya) and consisted of three medium-sized genera: *Eocoileus*, *Bretzia*, and *Odocoileus* (Webb 2000; Gustafson 2015). *Cervalces*, *Rangifer* and *Navahoceros* (American mountain deer) emerged later (Kurten 1975; Webb 2000), although the

taxonomic legitimacy of the latter is uncertain (Morejohn et al. 2004; Heffelfinger and Latch 2023). *Torontoceros hypogaeus* was described in 1982 (Churcher and Peterson 1982) and is represented by a single specimen (ROMM75974) dated at 11.3 thousand years ago (kya). The specimen consists of a partial cranium with incomplete antler beams, found in Toronto (Ontario, Canada) during excavation work for a metro line and was popularly dubbed the Subway deer. Similar in size to white-tailed deer (*Odocoileus virginianus*) and caribou (*Rangifer tarandus*), *T. hypogaeus* has a distinctive antler morphology that does not match any other American Cervidae (Churcher and Peterson 1982; Figure 4.1, Figure S4.1).

The genus *Torontoceros* was proposed primarily based on antler morphology, a highly variable and complex trait for which the use in systematics has been discouraged (Webb 2000). Two studies considered *T. hypogaeus* as a member of the genus *Rangifer* (Spiess et al. 1985; Croitor 2022), either because the morphology resembled that of some North American caribou populations (Spiess et al. 1985), or that it could reflect adaptation to open landscape (Croitor 2022). The ambiguity around *T. hypogaeus* taxonomy is heightened by the rarity of this genus in the fossil record. Perhaps due to this uncertainty, *T. hypogaeus* has not been consistently included in analyses of the Pleistocene megafaunal extinction (Table S4.1). Here, we sought to clarify the taxonomic affinity of the Subway deer (*T. hypogaeus*) using ancient DNA sequencing. We carried out phylogenetic analyses of mitochondrial genes and whole genome sequencing data to resolve the taxonomic position of this enigmatic fossil within Cervidae, and to further clarify deer evolutionary history and extinction events in North America during the Pleistocene.

Results and discussion

Late Pleistocene extinction studies usually focus on megafaunal mammals at the genus-level because fossil identification to the species-level is difficult. Genera-level inferences have several shortcomings; they quantitatively minimise the biodiversity loss by not accounting for species number (Graham and Lundelius 1984; i.e. up to five species in extinct North American genera (Meltzer 2020)); they also do not include extinct species whose genera survived (e.g., *Panthera spelaea* (Stanton et al. 2020)). A genus is not necessarily representative of functional diversity as species within the same genus can have distinctive functional traits (e.g., *Ursus arctos* & *Ursus maritimus*) and their loss can have drastically different impacts on the ecosystem. The advent of ancient DNA analyses allows for reconstructing the late Pleistocene extinction event with more resolution, and to fully quantify the loss of biodiversity that took place at the species level.

Taxonomic resolution of Subway deer

We successfully recovered 12 mitochondrial genes from ROMM75974 (Table S4.2). BLAST best hits consistently designated the *Odocoileus* genus, and the highly conserved *COX1* and *CYTB* genes presented an 88.28% and 91.56% similarity to mule deer (*O. hemionus)*, respectively (Table S.4.3). The placement of ROMM75974 appears nested within the *Odocoileus* genus, as a sister group to the white-tailed (*O. virginianus*) – mule deer (*O. hemionus*) clade (Figure 4.2, Figure 4.3, Figure S4.2), though some phylogenies show it fully integrated within this clade (Figure S4.3). The lower support in the mitochondrial phylogenies can be explained by the recognised phylogenetic discordance in white-tailed and mule deer, probably caused by incomplete lineage sorting (Klicka et al. 2023). Interestingly, our mitochondrial phylogenies showing

Torontoceros as nested within extant *Odocoileus* (Figure S4.3) would suggest that those species were likely part of an early radiation event from the same deer common ancestor.

The nuclear genome phylogenies showed complete support in the position of ROMM75974 as sister to the white-tailed-mule deer clade; divergence analyses present the same results and further support this finding (Figure 4.4, Figure S4.4, Figure S4.5). Likewise, d_{XY} shows a lower sequence divergence from white-tailed (0.01) and mule deer (0.009) than other Cervidae (Table S4.4). We also identified the molecular sex of ROMM75974 as male in datasets mapped to caribou (X/A = 0.668) and cattle (X/A = 0.624).

The rarity of *T. hypogaeus* in the fossil record and its description based on antler morphology has contributed to the uncertainty around the taxon's designation as a monotypic genus, with some studies assigning the specimen to *Rangifer* (Spiess et al. 1985; Croitor 2022). The phylogenetic proximity of ROMM75974 to *Odocoileus* in all our analyses contradicts the assignment of *T. hypogaeus* to its own genus or that of *Rangifer*. Instead, our results suggest ROMM75974 belongs to the genus *Odocoileus*, but as a sister taxon to the clade of white-tailed and mule deer and with *O. pandora* as more distant relative. Given this important divergence of *O. pandora* from other *Odocoileus* (Figure 4.2) and as the taxonomic transfer of *Mazama pandora* to *Odocoileus* (Gutiérrez et al. 2017) has been considered controversial, the taxonomic position of this species requires further analysis but is suggestive of either a distinct genus or subgenus. The *CYTB* sequence divergence of ROMM75974 with white-tailed and mule deer (Table S.4.3) is consistent with a distinct species (Bradley and Baker 2001), and the genomic divergence and phylogenies support this designation (Figure 4.3, Figure S4.2, Table

S4.4). Given the unique brow tine morphology, we consider *Torontoceros* as a subgenus and recognise ROMM75974 as *Odocoileus (Torontoceros) hypogaeus* (Supplemental data). The species identification of a museum specimen based on genomic data is increasingly common practice (Grewe et al. 2021; Raxworthy and Smith 2021; Plaxton et al. 2023), and arguably more reliable than morphological assignment, particularly based on antler structure (Webb 2000). As fossil identification to the species level of this genus is difficult (Jacobson 2003; Heffelfinger 2011), DNA assessment of additional *Odocoileus* museum specimens might identify other representatives of this new species and give a better idea of genetic variation prior to extinction (e.g. Rogers and Slatkin 2017; Sharko et al. 2021).

Adding to the Ice Age extinction list of North America

The late Pleistocene extinction event was global and affected many species. While megafaunal mammals received most of the attention, particularly the emblematic Proboscidea (*Mammut*, *Mammuthus*), smaller mammals, birds and trees were also impacted (Grayson 1977; Jackson and Weng 1999; Grayson 2007). North America saw over 30 mammal genera disappear from the continent (Elias and Schreve 2007; Stuart 2015; Broughton and Weitzel 2018; Meltzer 2020) and our results support the addition of *O. (T.) hypogaeus* to the North American Late Pleistocene extinction list.

The late Pleistocene landscape of southern Ontario, the location of current day Toronto, and more broadly the North American Great Lakes region, was similar to boreal woodland, dominated by spruce and sedge (Yu and Wright 2001; Shuman 2002). The area was home to a variety of extant fauna such as caribou and grizzly bears (*Ursus arctos–horribilis*; Peterson 1965; Long and Yahnke 2011; Lemke 2015) whose ranges have significantly shifted, but also the extinct mammoth (*Mammuthus sp.*; McAndrews

and Jackson 1988; Harington et al. 2012; Yansa and Adams 2012; Stewart et al. 2021), mastodon (*Mammut americamun*; McAndrews and Jackson 1988; Yansa and Adams 2012) and stag-moose (*C. scotti*; McDonald 1989; Long and Yahnke 2011; Glotzhober and McDonald 2015). One explanation for these range shifts and extinctions is the rapid vegetal transition from an open boreal woodland habitat to a more closed pine forest approximately 10 kya (Yu and Wright 2001; Shuman 2002; Long and Yahnke 2011; Yansa and Adams 2012; Allen et al. 2020; Stewart et al. 2021). This sudden vegetation replacement would have strongly impacted *O. (T.) hypogaeus*, as Croitor (2022) suggested its antler morphology reflected an adaptation to open landscapes. Collectively, recognising *Odocoileus (Torontoceros) hypogaeus* as a distinct but extinct species from the North American Great Lakes region adds weight to climate change being a major driver of megafaunal extinctions.

Material and methods

Ancient DNA extraction and sequencing

The specimen ROMM75974, described as the new genus and species *Torontoceros hypogaeus* by Churcher and Peterson (Churcher and Peterson 1982), is deposited at the Royal Ontario Museum and consists of an incomplete cranium with partial antlers and bone shards (Figure 4.1, Figure S4.1). We sampled a bone fragment from the posterior of the antler so as not to impact the integrity of the sample when exhibited. We performed aDNA extraction and library preparation in a dedicated laboratory at the Royal Ontario Museum. We irradiated the bone fragment with UV light for 11 minutes on all sides in a Stratagene UV Stratalinker 2400 before pulverizing it in liquid nitrogen with a mortar and pestle. We extracted DNA from approximately 60 mg of bone powder following the silica column protocol from Dehasque et al. (2022), and

prepared double-stranded sequencing libraries with USER enzyme treatment following Meyer and Kircher (2010) and including one extraction and one library prep negative control. Finally, we used unique barcodes to double-index and amplify the sequencing libraries as described in Díez-del-Molino et al. (2023), with four independent PCR amplifications under the following protocol: 95°C for 2 min followed by 12 cycles of 95°C for 15 s, 60°C for 30 s and 68 °C for 1 min. We sent the amplified libraries to The Centre for Applied Genetics in Toronto, Canada, for paired-end sequencing of 150 bp reads on an Illumina Novaseq 6000 with SP flowcells.

For comparison with modern data, we collected 17 published mitochondrial genomes of Cervidae, and whole genome resequencing data from six deer species (Table S4.5, Table S4.6, Figure S4.6), including all Northern American representatives (USA & Canada). Resequencing data was not available for all species investigated with mitochondrial genes.

Data processing

Mitochondrial genes

For the ancient sample of ROMM75974, we trimmed and merged paired end raw reads in AdapterRemoval (v2.3.2; Schubert et al. 2016) setting the minimum read length to 25 (--minlength 25) and mismatch rate to 3 (--mm 3; Plaxton et al. 2023). We converted the merged fastq file to a fasta format and filtered out reads belonging to microbial contaminants with Kraken2 (v2.1.3; Wood et al. 2019) using the viral, archaeal, bacterial and fungal databases; 10.13% of our data was classified as belonging to these databases (Table S4.7). We used MitoGeneExtractor (v1.9.5; Slater and Birney 2005; Brasseur et al. 2023) with default settings to extract the 13 protein-coding mitochondrial genes from trimmed and filtered reads (Table S4.2). As the phylogenetic

relationship of *T. hypogaeus* is uncertain and to avoid reference bias, we chose to use cattle (*Bos taurus*) as reference because the family Bovidae is an outgroup to Cervidae. We used BLAST (Altschul et al. 1990) for a first species identification step with each assembled protein-coding gene sequence.

Resequencing data

For ROMM75974 resequencing data preparation, we trimmed and merged the reads in SeqPrep (v1.2; John 2011) with default settings and minor source code edits (Palkopoulou et al. 2015). The above BLAST searches on the mitochondrial data narrowed the taxonomic affinity; accordingly, we mapped the merged reads to two reference genomes for different analyses: (1) to produce a phylogeny including six extant Cervidae species, we mapped the data to the cattle reference (GCF_002263795.2); and (2) to perform a PCA of the most closely related species (*Odocoileus* and *Rangifer*), we mapped to the caribou reference (GCA_949782905.1). For this, we used bwa aln (v0.7.17; Li and Durbin 2009) and parameters optimised for aDNA data: a maximum of two gaps (-o 2), allowing more substitution (-n 0.01), and deactivating seed (-l 16500; Schubert et al. 2012; van der Valk et al. 2022). We filtered mapped reads to a minimum length of 30 bp and a minimum mapping quality of 25 in SAMtools (v1.17; Heng Li et al. 2009). We identified and removed PCR duplicates with the python script samremovedup.py, which takes into account the read start and end coordinates and read length (Palkopoulou et al. 2018). We used BamTools stats (v2.5.1; Barnett et al. 2011) throughout the mapping process for quality checks and computed the final coverage in ANGSD (v0.939; Korneliussen et al. 2014b; Table S4.8). Finally, we assessed damage patterns in mapDamage (v2.2.1; Jónsson et al. 2013) and PMDtools (v2; Skoglund et al. 2014), the latter using the platypus option that allows damage

identification at CpG sites which remain unaffected by USER enzyme treatment (Figure S4.7).

For the modern cervid data (Table S4.6), we downsampled the data to a maximum of 100 million random read pairs using seqtk sample (v1.3; Li 2018) to even out the modern data, making sure the seed was set to the same number for each read pair file. We trimmed the downsampled reads in Trimmomatic (v0.36; Bolger et al. 2014b). Using bwa-mem (v0.7.17; Li and Durbin 2009), we mapped the data to the cattle reference genome for the Cervidae phylogeny, and *Rangifer* and *Odocoileus* data to caribou for the PCA. We used SAMtools (v1.10; Heng Li et al. 2009) to sort the aligned reads, then used Picard MarkDuplicates (v2.23.2; Broad Institute 2019) and Sambamba view (v0.7.0; Tarasov et al. 2015b) to identify and filter out duplicated reads. Finally, we used GATK RealignerTargetCreator and IndelRealigner (v4.1.7.0; McKenna et al. 2010b) to carry out a local realignment.

Divergence from Cervidae

Mitochondrial Phylogenies

We used MEGA (v11.0.13; Stecher et al. 2020) and the ClustalW algorithm with default settings to align *COX1* and *CYTB* sequences, and the collated partial mitochondrial genome of *T. hypogaeus* to that of 17 other Cervidae species with cattle as outgroup (Table S4.5), removing all gaps, ambiguous sites, and singletons from the alignment. We used IQ-TREE (v1.6.12; Nguyen et al. 2015; Kalyaanamoorthy et al. 2017; Hoang et al. 2018) to perform phylogenetic analyses, selected the best substitution model with ModelFinder (-m MFP; Kalyaanamoorthy et al. 2017), obtained branch support from the SH-like approximate likelihood ratio test (-alrt 1000) and used

ultrafast bootstrap (-bb 1000) optimised with a hill-climbing nearest neighbour interchange search (-bnni).

To date the divergence of ROMM75974 from other Cervidae species, we ran BEAST (v2.7.6; Bouckaert et al. 2014) to construct a calibrated *CYTB* phylogeny using the alignment produced by ClustalW but removing the outgroup. We applied a GTR+I+G nucleotide substitution model as selected by ModelFinder, an optimised relaxed clock model and a calibrated Yule tree prior with default settings. We further used three calibration points following a normal distribution prior: Cervinae at 10.9 ± 1 mya (Zurano et al. 2019), Muntiacinae at 8 ± 1 mya (Dong et al. 2004) and the American species at 5 ± 1 mya (Webb 2000; Gilbert et al. 2006). We sampled every 1,000 generations in three independent runs of 50 million generations and discarded the first 10% as burn-in. We assessed convergence in Tracer (v1.7.2; Rambaut et al. 2018), and generated a maximum clade credibility tree with median node heights using TreeAnnotator (v2.7.6; Bouckaert et al. 2014) before visualisation.

Whole genome phylogeny and allelic distance

Using the data mapped to cattle, we created an initial dataset in ANGSD including variant and invariant sites but no missing data to estimate nucleotide diversity (π) and sequence divergence (d_{XY}). We used a custom script (popgenWindows.py; Martin 2021) to compute π and d_{XY} in sliding windows of 500 bp, allowing a minimum of 50 sites per window (-w 500 -s 500 -m 50). We called genotypes and genotype likelihoods in ANGSD removing all missing data (-minInd 37) and filtering for a maximum p-value of 1 e−6 (-SNP_pval 1e-6). We further applied different filters using vcftools and created two datasets: (1) minimal LD filter dataset, where we allowed a maximum of one site per 100bp (--thin 100; 11,862 sites); (2) minimal LD and depth

filter dataset, based on the previous dataset with the addition of a genotype minimum depth of three (--minDP 3; 8,551 sites). We created the latter to increase the confidence in the ancient data, and phylogenetic and allelic distance analyses were performed on both datasets.

We used PLINK (v1.90b6.21; Chang et al. 2015) to compute allelic distance (--distance) between all individuals and plotted the resulting matrix in heatmaply (Galili et al. 2018). We inferred a time-calibrated Cervidae phylogeny based on SNP data using the multi-species coalescent model in SNAPP (v 1.6.1; Bryant et al. 2012; Stange et al. 2018), setting two calibration points following a normal distribution, one on the *Cervus* node (μ = 2.3, σ= 0.5; calibration based on Figure 4.2 and Hu et al. 2019; Mackiewicz et al. 2022b) and one on the Odocoileini node (μ = 5.8, σ= 0.5; Gilbert et al. 2006), and a starting tree modifying the allelic distance tree produced above into a species tree. We used a custom script (snapp_prep.rb; Stange et al. 2018) with default settings to generate two XML inputs for each dataset: one with all the filtered data and one that included transversions only as they should represent true mutations as opposed to transitions which can be the product of DNA damage (--transversions). We ran SNAPP in three independent runs of 500,000 MCMC iterations, excluding the first 10% as burn-in. As above, we used Tracer to check for convergence and generated a maximum clade credibility tree with median node heights before visualisation.

Divergence from *Odocoileus* & *Rangifer*

Because two previous studies considered ROMM75974 as a caribou (Spiess et al. 1985; Croitor 2022) and to obtain more resolution on the taxonomic affinity of the specimen, we compared the genomic sequences to all *Odocoileus* and *Rangifer* individuals mapped to caribou. We produced a beagle file in ANGSD, filtering for SNPs

(-SNP_pval $1e^{-6}$) and removing all sites with missing data to keep only the sites covered in our ancient sample (-minInd 24). We then used PCAngsd (v1.02; Meisner and Albrechtsen 2018) to compute a PCA based on allele frequencies of genotype likelihoods. The results were plotted in R (v 4.3.1; R Core Team 2021). Finally, because caribou females are antler-bearing, which is unique in Cervidae, identifying the molecular sex of ROMM75974 could inform on phylogenetic affinity. Females have two copies of autosomes and X chromosomes; their X/A ratio is expected to be close to one, whereas it should be approximately 0.5 in males who only have one X copy (Bro-Jørgensen et al. 2021). We used SAMtools idxstats to retrieve the number of reads mapped to each chromosome of the caribou and cattle reference, using chromosomes one and two, respectively, as they are closest in size to their X chromosome.

Acknowledgements

The specimen ROMM75974 was discovered on the ancestral land of different First Nations including the Anishinaabeg, the Haudenosaunee, and the Wendat. Historically, the region of Toronto was a busy trade route which connected many First Nations of the Great Lakes, it was a place of culture, important for fishing and hunting, and generally highly significant to the people who lived there for thousands of years. The land on which we work belongs to many First Nations, it is their ancestral home and unceded territory. Together, Trent University and the ROM are located on the territory of the Michi Saagiig Anishnaabeg, the Wendat, the Haudenosaunee Confederacy, and the Anishinaabeg Nation, including the Mississaugas of the Credit First Nation. As settlers, we are grateful to have had the opportunity to live and work on this land, and to benefit from it. We would like to show our respect to the First Peoples

and thank them for their care, stewardship, and teachings. Miigwetch, tiawenhk, niá:wen.

Camille Kessler was supported by an International Graduate Scholarship, an Ontario Graduate Scholarship, and a French American Charitable Trust Scholarship. This work was supported by NSERC Discovery Grant (Grant Number: RGPIN-2017-03934); ComputeCanada Resources for Research Groups (Grant Number: RRG gme-665-ab); Canadian Foundation for Innovation: John R. Evans Leaders Fund and the Ontario Early Researcher Award (Grant Number: #36905).

We would like to thank Marianne Dehasque for their precious help and sharing of laboratory methods and protocols, Marie-Laurence Cossette for their help in the lab, Jochen Wolf and Jose Alberto Lopez-Aleman for helpful comments, and Paul Eekhoff for photography of the antlers.

Data availability

Raw reads for ROMM75974 were deposited on the NCBI Sequence Read Archive (Accession number PRJNA1136003). All scripts are available on GitLab: https://gitlab.com/WiDGeT_TrentU/graduate_theses/-/tree/master/Kessler/CH_03

Figures and tables

Figure 4.1: Specimen ROMM75974 of *Torontoceros hypogaeus* from the Royal Ontario Museum, Toronto. A) Photograph from anterior aspect, credit: Paul Eekhoff B) Artistic rendering, credit: Sherri Owen.

Figure 4.2: Calibrated *CYTB* phylogeny of Cervidae produced by BEAST, nodes are labelled with the estimated divergence time in the past, colour represents posterior distribution support (white ≤ 70%, grey 71 - 89%, black ≥ 90%), square nodes show calibration points and error bars represent the 95% highest posterior density (HPD) intervals, x-axis timeline in units of millions of years ago (mya). Tree visualisation performed in FigTree and edited in Inkscape.

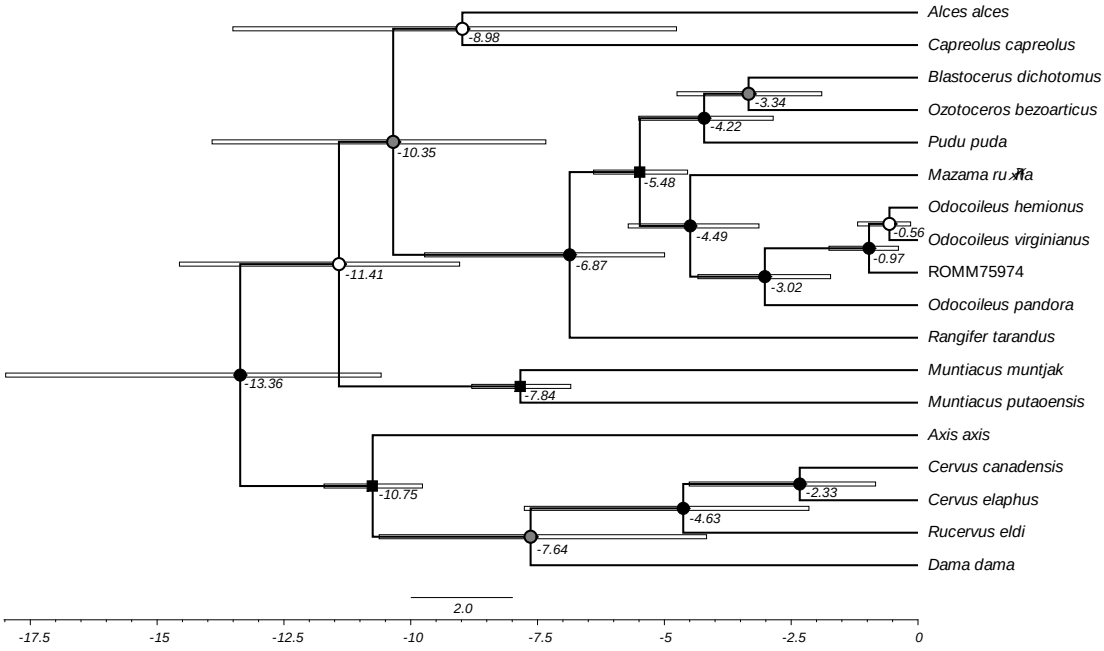


Figure 4.3: Calibrated whole genome phylogeny of Cervidae produced by SNAPP based on the minimal LD filter dataset (11862 sites). Nodes are labelled with the estimated divergence time in the past, colour represents posterior distribution support (white ≤ 70%, grey 71 - 89%, black ≥ 90%), square nodes show calibration points and error bars represent the 95% HPD intervals, x-axis timeline in units of mya. Tree visualisation performed in FigTree and edited in Inkscape.

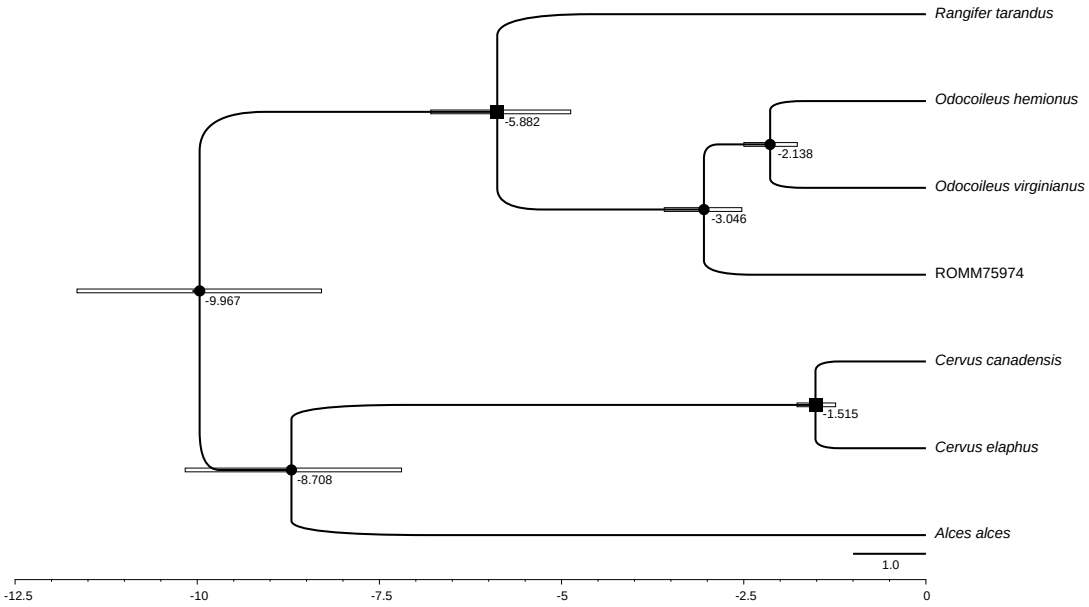
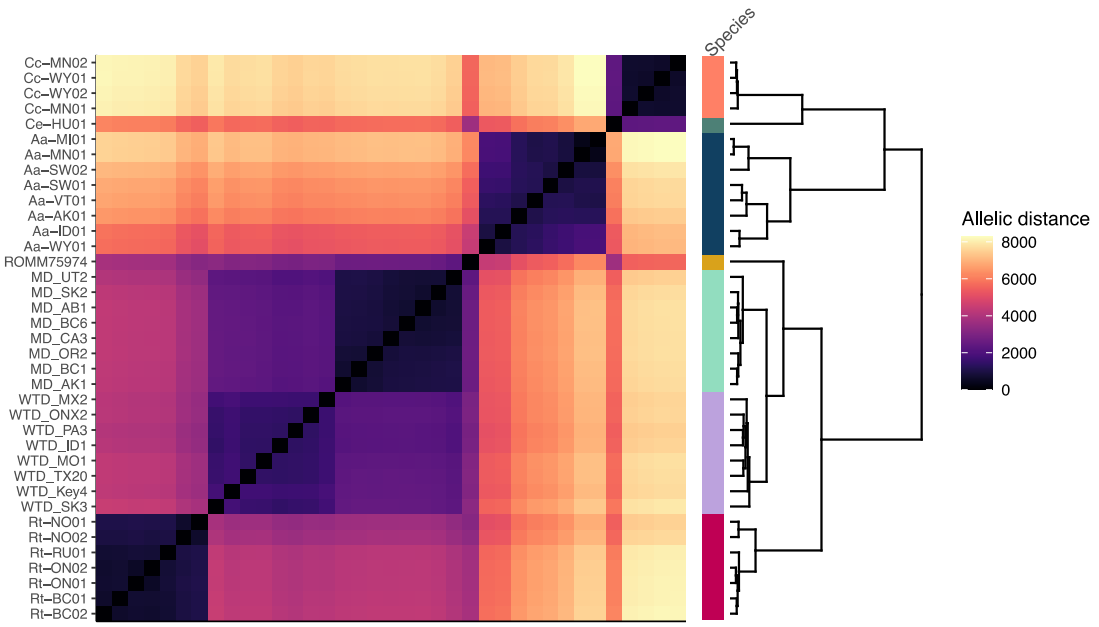


Figure 4.4: Allelic distance matrix heatmap and clustering based on minimal LD filter dataset (11862 sites) of Cervidae, side colour represents the species as abbreviated in sample name: Aa = *Alces alces*, Cc = *Cervus canadensis*, Ce = *Cervus elaphus*, Oh = *Odocoileus hemionus*, Ov = *Odocoileus virginianus*, Rt = *Rangifer tarandus*.



CHAPTER 5 : ANCIENT AND MODERN PERSPECTIVES OF GENOMIC DIVERSITY IN NORTH AMERICAN CERVIDS

Authors:

Camille Kessler, Oliver Haddrath, Christopher N. Jass, Burton K. Lim, Aaron B. A. Shafer

This chapter is in preparation for submission.

Contributions:

CK and ABAS conceived the study, CNJ & BKL provided the sample, CK performed the molecular laboratory work with contribution from OH, CK performed bioinformatic analyses, CK wrote the manuscript with general contribution from ABAS. All authors will review and provide comments on the submitted manuscript.

Summary

Wild populations in North America have experienced many changes over the last two million years, with the glacial cycles of the Pleistocene, the peopling waves, the megafaunal mammal extinction event, the European colonisation and the current global climate crisis. These changes caused demographic fluctuations, population divergence, range shifts and even extinctions. Canada and the USA harbour five extant cervids: caribou (*Rangifer tarandus*), moose (*Alces alces*), mule deer (*Odocoileus hemionus*), wapiti (or elk, *Cervus canadensis*) and white-tailed deer (*Odocoileus virginianus*). All five species have experienced changes in population size, due to climatic oscillations, dispersal into the continent or human activities after European colonisation, all of which likely impacted their genomic diversity. Here, we used chronological sampling of ancient, historical and modern individuals of these five cervids to retrace changes in genetic diversity through time. Our results suggest rapid expansion of moose and wapiti following their arrival on the continent ~15 thousand years ago, with a continued expansion through the Holocene and recent loss of genetic diversity exemplified by modern samples. Wapiti was the only species presenting variant composition change over time which could indicate adaptation to novel environment. Our results suggest several colonisation waves of caribou, with Pleistocene levels of heterozygosity similar to that of modern samples. The phylogenetic placement of our *Odocoileus* samples gives weight to incomplete lineage sorting as source of the recognised mito-nuclear discordance in those species. Overall, our results are consistent with known dispersal and demographic histories of our target species and exemplify the utility of chronological sampling to gain a better understanding of how external pressures on wild species drive genomic changes.

Introduction

Over the last two million years (mya), North America has experienced intense climatic oscillations that caused drastic local and global environmental changes (Shuman 2002; Allen et al. 2020; Marean et al. 2020; Mottl et al. 2021; Hanberry 2023). The peopling of the continent happened at the very end of these oscillations, around the time of the last glacial maximum (LGM; Nielsen et al., 2017; Willerslev & Meltzer, 2021) and was closely followed by a large megafaunal extinction event approximately 11,000 years ago (kya), noting human impact is not a clear driver of this loss (Broughton and Weitzel 2018; Meltzer 2020; Stewart et al. 2021; Bergman et al. 2023). This extinction event was likely triggered by the deglaciation dynamics which caused particularly fast environmental shifts in the region (Shuman 2002; Mottl et al. 2021; Benfield et al. 2023; Hanberry 2023), with the loss of biodiversity and ecosystem functions itself affecting vegetation communities in a positive feedback loop (Malhi et al. 2016; Malhi et al. 2022). The Holocene also entailed climatic changes of lower intensity (Shuman and Marsicek 2016; Thompson et al. 2022; Zhang et al. 2022), and since the 1500's, colonisation from European nations affected wild populations through logging and farming practices, for example (Lindo et al. 2016; Smith et al. 2021; Brain and Prosser 2022). Finally, the current global climate crisis driven by human activities since the industrial revolution have been strongly detrimental to wild communities worldwide (Ceballos et al. 2020; IPCC 2022; Ceballos and Ehrlich 2023). Altogether, this dynamic history has influenced North American fauna, causing demographic fluctuations, population divergence, range shifts and extinctions (Lorenzen et al. 2011; Taylor et al. 2021; Da Silva Coelho et al. 2023; Hold et al. 2024).

Northern America (Canada & USA) is home to five extant Cervidae species: caribou (or reindeer, *Rangifer tarandus*), moose (*Alces alces*), wapiti (or elk, *Cervus canadensis*) and white-tailed deer (*Odocoileus virginianus*) and mule deer (*Odocoileus hemionus*). Both *Odocoileus* species and caribou were present in the region through the glacial cycles of the Pleistocene, where cold periods caused declines and diversification in white-tailed and mule deer through separate refuges (Combe et al. 2021; Lamb et al. 2021; Kessler and Shafer 2024); rapid warming events appeared to lead to declines in caribou (Taylor et al. 2021). Moose and wapiti entered the continent after the last glacial maximum (Hundertmark et al. 2002; Meiri et al. 2020; Mackiewicz et al. 2022a), shortly before the Pleistocene extinction event where several Cervidae species went extinct, including the Toronto subway deer (*Odocoileus (Torontoceros) hypogaeus*; Churcher & Peterson, 1982; Kessler et al., In prep.; Meltzer, 2020). While Indigenous impact on these species was likely minor (Wolverton et al. 2008; Weitzel 2021; Groves et al. 2022), all extant species were impacted by human activities after the colonisation, to the point of regional extirpations for white-tailed deer and wapiti (Gill 1999; O'Gara and Dundas 2002; Schaefer 2003; McDonald et al. 2004; Wattles and DeStefano 2011; DeCesare et al. 2014; Jensen et al. 2023).

Undoubtedly, each of these events impacted the species' genetic diversity. The development of ancient and historical DNA analysis has allowed to sample populations before and after an event of interest, and to accurately describe the impact of those events on the genome (Dehasque et al. 2020; Orlando et al. 2021). For example, chronological sampling has shown a major loss of genetic diversity due to a strong bottleneck in alpine ibex (Robin et al. 2022), revealed the colonisation route to a remote archipelago in caribou (Hold et al. 2024), and shed light on the extinction processes of

the blue antelope (Hempel et al. 2024). Extinct Proboscidea (e.g. woolly mammoth) are probably the group that have been most scrutinised with ancient DNA, and exemplifies the versatility of this tool to study admixture, adaptation and selection (e.g. Díez-del-Molino et al., 2023; van der Valk et al., 2021, 2022). Notably, ancient DNA analysis of mammoths provided information on extinction dynamics and on processes leading to accumulation of deleterious mutations, high genetic load and ultimately genetic meltdown (Palkopoulou et al. 2015; Rogers and Slatkin 2017). Therefore, using a chronological sequence of genomic data, we can retrace the onset of selection, identify drivers of adaptation, and measure changes in genetic diversity through time.

Here, we used ancient, historical and contemporary DNA analysis to explore changes in genetic diversity over time for five extant North American Cervidae. The late Pleistocene was a period of demographic changes for North American caribou, with a general decline starting approximately 50 kya (Taylor et al. 2021). Their populations expanded after the LGM (Weckworth et al. 2012), but the species is currently declining throughout North America (Schaefer 2003; Festa-Bianchet et al. 2011; Schmelzer 2020). This decline, while increasing genetic load, appears not to be accompanied by loss of genetic diversity or purging (Solmundson et al. 2023; Taylor et al. 2024). Accordingly, we predict ancient and modern caribou to present similar levels of genetic diversity and deleterious mutations. White-tailed deer populations have fluctuated throughout the Pleistocene, reaching very high numbers but plummeting during the LGM (Kessler and Shafer 2024). White-tailed deer population size was also severely impacted during the colonial-era, but appropriate management and conservation allowed a fast recovery (Kessler and Shafer 2024). Given that the historical bottleneck was short and the high resilience of this species, we expect genetic diversity of white-tailed deer to remain

stable over time. In contrast, Mule deer population size appears to have remained low but stable over the last 300 kya, with a dramatic bottleneck without recovery about 150 years ago (Kessler and Shafer 2024). Stability at low effective population size can lead to accumulation of deleterious mutations and an increased genetic load, although highly deleterious mutations should be purged by strong bottlenecks (Grossen et al. 2020; Wootton et al. 2023). Therefore, while the genetic diversity of mule deer should be stable in ancient samples, we expect a recent loss of genetic diversity in modern individuals and no (to slight) increase in deleterious mutations over time. The dispersal of moose and wapiti into North America was relatively recent and while information is lacking for wapiti, it was likely accompanied with slight bottleneck followed by rapid expansion in moose (Meiri et al. 2020). We assume wapiti followed the same pattern, and therefore predict both species to show low diversification and short phylogenetic branches in North American populations. Further, we expect both species to have expanded during the Holocene, we should see an increase of rare alleles and genetic diversity during that time, but a loss in diversity in modern samples due to the bottlenecks both moose and wapiti suffered after the European colonisation, with the wapiti decline arguably stronger (O'Gara and Dundas 2002; Wattles and DeStefano 2011).

Material and methods

Sampling and laboratory procedures

We sampled 11 North American museum specimens of our species of interest (Table S5.1), and gathered contemporary resequencing data from 36 published genomes, including North American and European representatives when relevant (Table S5.2). Radiocarbon dates were unknown for all but one museum specimen

(OTh_ON01; Churcher & Peterson, 1982), all other dates were inferred from bone fragments that were pre-processed for collagen extraction and purification at the Trent Environmental Archaeology Lab and radiocarbon dating was performed by the W. M. Keck Carbon Cycle AMS facility at the University of California, Irvine. We calibrated the dates in OxCal 4.4 (Ramsey 2009) using the IntCal20 calibration curve (Reimer 2020) with a resolution of two and 100 MCMC chains. While all ancient *Cervus* samples are labelled as *C. elaphus*, we consider them as *C. canadensis* as this species was recognised as distinct from its European counterpart relatively recently (Kuwayama and Ozawa 2000; Pitra et al. 2004; Lorenzini and Garofalo 2015; Hu et al. 2019). Both species are represented in our contemporary samples.

We performed all aDNA laboratory procedures in a dedicated facility in the Royal Ontario Museum (Canada). Briefly, we first irradiated the bone fragments with UV, which we then powdered in liquid nitrogen using a mortar and pestle to obtain approximately 50mg of bone powder. We followed Dehasque et al. (2022) for DNA extraction using the silica column protocol and double-stranded library preparation with USER enzyme treatment, then we followed Díez-del-Molino et al. (2023) to double-index and amplify the sequencing libraries. We performed four independent PCR amplifications following this protocol: 95°C for 2 min, then 12 cycles of 95°C for 15 s, 60°C for 30s and 68 °C for 1min and included one extraction and one library prep negative control (ddH_2O) per batch. We sent the amplified libraries to The Centre for Applied Genetics in Toronto, Canada, for paired-end sequencing of 150 bp reads on an Illumina Novaseq 6000 with SP flowcells.

Data processing

For the ancient data, we trimmed and merged the raw reads in SeqPrep (v1.2; John 2011) with default settings and a minor source code modification (Palkopoulou et al. 2015), then mapped the data with bwa aln (v0.7.17; Li and Durbin 2009) and settings optimised for aDNA (-o 2, -n 0.01, -l 16500; Schubert et al. 2012; van der Valk et al. 2022). We mapped all samples to their respective mitochondrial reference, to their specific (or closely related) genome reference assembly, and all to the cattle reference (Table S5.3). Using SAMtools (v1.17; Li et al. 2009), we excluded reads shorter than 30bp and with a mapping quality lower than 25, we removed PCR duplicates with the python script samremovedup.py (Palkopoulou et al. 2018). Finally, we used BamTools stats (v2.5.1; Barnett et al. 2011) for quality checks, computed the final coverage in ANGSD (v0.939; Korneliussen et al. 2014), and identified damage patterns at CpG sites using the platypus option in PMDtools (v2; Skoglund et al. 2014). We sexed all samples using the ratio of reads covering the autosome/X chromosome for the species with an chromosome-level reference, and SATC (Nursyifa et al. 2022) for species without.

We downsampled the modern data to make it approximately equivalent between samples; we selected 100 million random reads pairs using seqtk sample (v1.3; Li 2018), setting the same seed number for each read pair file. Then, we used Trimmomatic (v0.36; Bolger et al. 2014) to trim the reads, bwa-mem (v0.7.17; Li and Durbin 2009) to align the data to the appropriate complete and mitochondrial reference genome (see above; Table S5.3), and SAMtools (v1.10; Li et al. 2009) to sort the aligned reads. We identified and filtered duplicated reads using Picard MarkDuplicate (v2.23.2; Broad Institute 2019) and Sambamba view (v0.7.0; Tarasov et al. 2015), and performed

local realignment in GATK RealignerTargetCreator and IndelRealigner (v4.1.7.0; McKenna et al. 2010).

Haplotype networks and phylogenetic analysis

We produced mitochondrial consensus sequence for each sample in ANGSD (v0.939; Korneliussen et al. 2014), selecting the base with highest effective depth and setting a minimum depth of 3 (-doFasta 3; -setMinDepth 3). We aligned the consensus sequences in MEGA (v11.0.13; Stecher et al. 2020) using the ClustalW algorithm with default settings both independently for each species or species groups, and including all samples. With the specific alignments, we inferred median joining networks in popART(v1.7; Bandelt et al. 1999; Leigh and Bryant 2015). With the alignment comprising all samples, we removed all gaps and singletons, and produced a phylogenetic tree using IQ-TREE (v1.6.12; Nguyen et al. 2015; Kalyaanamoorthy et al. 2017; Hoang et al. 2018) with best substitution model selection (-m MFP; Kalyaanamoorthy et al. 2017), SH-like approximate likelihood ratio test (-alrt 1000), ultrafast bootstrap (-bb 1000) and hill-climbing nearest neighbour interchange search (-bnni).

Population structure and genetic diversity

We generated principal component analyses using smart_pca of the smartsnp R package (v1.1.0: Herrando-Pérez et al. 2021), independently for each species group and all samples together with data mapped to cattle. Using ANGSD, we produced a complete dataset including all SNPs passing the following filters: biallelic sites (-skipTriallelic 1), “good” reads that are uniquely mapping (-remove_bads 1 -uniqueOnly 1), with a base quality above 30 and mapping quality above 25 (-minMapQ 25 -minQ 30) and a p-value below $1e^{-6}$ (-SNP_pval 1e-6). We created a variant-major additive

component file (.traw) in PLINK (v2.0; Chang et al. 2015) for input into smart_pca which projects the ancient samples position on a PCA space computed on modern samples while accounting for drift.

We used ANGSD to compute individual heterozygosity using the same quality filters as above and applying two levels of site filtering: a minimum depth of three and an additional filter removing transitions (-rmTrans 1). The analysis on modern samples was restricted to the sites recovered in ancient samples of the same species group.

Variant analysis

Using VCFtools (Danecek et al. 2011) and BCFtools (Danecek et al. 2021), we isolated each ancient individual from the complete datasets produced above, applying a filter for minimum depth of 3 and removing all missing data. We further divided the complete dataset per population of modern individual and restricting it to the sites overlapping with ancient individuals. We ran snpEff (Cingolani et al. 2012) with default settings on each subset of the data, and used the SnpSift extension to format the output and filter out C to T and G to A transitions which could be spurious patterns of ancient DNA. We analysed all results in R (R Core Team 2021).

Results and discussion

Free-ranging wildlife in North America have been evolving through glacial cycles, peopling waves, extinction events, European colonisation and the global climate crisis. Here, we compared ancient and historical DNA from five North American Cervidae species to modern data to retrace changes in genetic diversity though time to better understand the impact of such events. We successfully sequenced 11 museum samples at ultra-low coverage at the whole-genome level (0.04 – 0.9x), and at low to medium coverage for mitochondrial DNA (1 - 11x;), all presenting patterns of DNA

damage consistent with aDNA and presenting up to 60% endogenous DNA (Table S5.4; Table S5.5; Figure S5.1).

Rapid diversification into North America from Eurasian moose and wapiti source populations

Moose and Wapiti entered North America through Beringia, likely after the LGM (Meiri et al. 2014; Meiri et al. 2020; Mackiewicz et al. 2022a). Our results showed European samples as most diverged and support Eurasia as the source population for North American moose and wapiti. The spread of North American samples on PC1 is indicative of a genetic diversity that is lower than Swedish moose, and the short phylogenetic branches supports rapid diversification of this group (Figure 5.1A; Figure 5.2). These results agree with previous literature (Hundertmark et al. 2002; Meiri et al. 2020), and support that moose dispersal into North America was accompanied by a slight loss of diversity and followed by rapid expansion across the continent. The pattern of lower diversity in North American compared to Swedish moose is not recovered in our heterozygosity analysis which shows similar values across modern samples (Figure 5.3; Table S5.6). Our analyses reveal clusters of modern North American individuals that are consistent with subspecies designation (Figure 5.1; Figure 5.2) and previous analyses that have shown a relatively high divergence between but low differentiation within moose subspecies (Ferrante et al. 2022). Interestingly, the ancient moose which would currently belong to *A. a. americana* (Aa_ON02) or *A. a. andersoni* (Aa_AB*) did not group with the modern geographic representatives of these subspecies. While the Ontario sample groups with *A. a. americana* in the mitochondrial analyses, the ones from Alberta do not show any clear clustering pattern (Figure 5.1; Figure 5.2). The lack of structure is likely reflective of high mobility and expansion of

ancient populations, likely facilitated by habitat connectivity prior to the major changes in land use brought to the region since the colonisation by Europeans. A higher habitat fragmentation and the current demographic decline observed in the species (Wattles and DeStefano 2011; DeCesare et al. 2014; Timmermann and Rodgers 2015; Priadka et al. 2024) could explain the significantly lower heterozygosity observed in modern samples (Figure 5.3; Table S5.6). Indeed, the most recent museum sample (Aa_ON02; 296 – 0 calBP) shows the highest heterozygosity which suggests a loss of genetic diversity in the last 200 years, though we detected no difference in variant composition over time (Figure 5.4).

The high divergence of the contemporary *C. elaphus* sample in our analyses is consistent with the species designation of *C. elaphus* and *C. canadensis* and supports the assignment of the museum samples to the latter (Figure 5.1; Figure 5.5). Our mitochondrial analyses suggest a rapid diversification of wapiti into North America with very short phylogenetic branches (Figure 5.1) and short genetic distance between haplotypes (Figure 5.5A). These distances are slightly higher for ancient samples (Figure 5.5A), suggesting a higher ancestral diversity and a loss of diversity over time, consistent with our heterozygosity estimates which recovered pattern of higher diversity in two ancient samples (Figure 5.3; Table S5.6). Speller et al. (2014) analysed the D-loop in ancient, historical and modern Alberta wapiti and showed that loss of genetic diversity over time was dependent on subspecies, with that of *C. c nelsoni* remaining stable and *C. c. manitobensis* decreasing with time. A direct comparison is challenging as we are lacking subspecies designation for our museum samples, it is nevertheless possible that our samples represent both subspecies which would explain the similar levels of heterozygosity between modern and Ce_AB07 (Figure 5.3; Table S5.6). Whole-genome

analyses show strong modern population structure in the PCA, which accounts for slightly more variation than the inter-specific separation along PC2 (Figure 5.5B), this pattern is also partially recovered in the haplotype network (Figure 5.5A). Here, the ancient samples are tightly clustered between two modern North America populations along PC1 and with *C. canadensis* along PC2 (Figure 5.5B). The lack of spread in these samples could be caused either by lack of signal in the data or by population structure in a population that was not sampled with modern data. We surmise the position of Ce_AB07 (0.9x) at least, reflects population structure which would suggest a third population that was not sampled with modern individuals. We detected differences in variant composition change in the oldest wapiti sample (Ce_AB05; 12,058 – 12,612 cal BP) compared to all other samples (Figure 5.4). The higher intergenic region modifier mutations upon the arrival on North America, and the loss of those mutations through time could indicate adaptation to the novel environment or some degree of purging.

Dispersal waves from Beringia caribou

Caribou originated in Beringia and the North American lineage diverged approximately 70 kya, although Beringia remained a refugia during the glacial cycles of the Pleistocene (Webb 2000; Weckworth et al. 2012; Taylor et al. 2021). This refugial population was the source of several dispersal waves that introgressed with the populations that remained south of the ice sheets (Flagstad and Røed 2003; Weckworth et al. 2012; Taylor et al. 2021); our results are consistent with this complex history. Mitochondrial analyses suggest high divergence in our samples, with the pair Rt_AB03 & Rt-ON01 as most diverged (Figure 5.1; Figure 5.6A). This pattern supports several dispersal waves from Beringia into North America, with those two samples resulting from an earlier wave. North American samples from the same region show higher

haplotype distance than the Norwegian samples, suggesting higher diversity in North America (Figure 5.6A), consistent with the heterozygosity patterns (Table S5.6), PCA (Figure 5.6B), and existing literature (Kuhn et al. 2010; Hold et al. 2024; Taylor et al. 2024). The PCA shows three clear clusters of modern samples with similar variation explained by PC1 and PC2 (Figure 5.6B). The spread along PC2 could be reflective of dispersal age, with the Norwegian and Ontarian populations as part of an initial migration wave, whereas samples from Russia and British Colombia would result from a subsequent wave. Similar to wapiti, the ancient caribou exhibits a central position, which in this case would likely reflect the low number of variants used (Table S5.4). Overall, our results support several migration waves from Beringia caribou into North America and Eurasia, resulting in large genetic diversity in North America. Here again, variant composition remained stable through time (Figure 5.4).

Gradient of ancient *Odocoileus*

White-tailed and mule deer exhibited a recognised mito-nuclear phylogenetic discordance (e.g. Cronin, 1991; Hopken, 2015; Wright et al., 2022), caused by incomplete lineage sorting (ILS; (Klicka et al. 2023) as there is no evidence for ancestral hybridisation (Kessler et al. 2023; Kessler and Shafer 2024). Our mitochondrial analyses show no species clustering in the *Odocoileus* genus (Figure 5.1; Figure 5.7A), giving further weight to the ILS hypothesis. These results also show the clear divergence of two subspecies: black-tailed deer (*O. h. colombianus & sitkensis*; MD BC1, OR2 & AK1), and Florida Keys deer (*O. v. clavium*; WTD_Key4; Figure 5.1; Figure 5.7A). Of note, the shared haplotype between the historical Oi_AB05 (4 - 281 calBP) and a mule deer from Saskatchewan would suggest this ancient unidentified individual belongs to *O. hemionus,* noting many museum specimens are only identified to the genus. The high

divergence of the ancient Ov_ON01 in mitochondrial analyses is likely caused by missing data, as is its position on the PCA (Table S5.5, Figure 5.7B). Principal component analysis of whole genome data shows clear species delimitation along PC1 (Figure 5.7B). Here, the ancient samples are spread following a gradient between the two species, with Oi_AB05 and Oh_AB02 as ancestral mule deer individuals, OTh_ON01 as the recently recognised extinct *O. (Torontoceros) hypogaeus*, and Ov_ON01 as intermediate between white-tailed and mule deer (Figure 5.7B). The patterns of poor data quality for the latter, despite a relatively recent radiocarbon date (306 - 433 calBP), is unexpected but not unheard of (e.g. Plaxton et al., 2023). Museum samples can be subject to various treatments for specimen preservation that are destructive for DNA molecules, further information on this specimen and its history should indicate whether it was treated in any way. Heterozygosity analysis in white-tailed and mule deer shows patterns of genetic diversity increasing over time, with a drop in modern samples (Figure 5.3; Table S5.6), likely reflective of the demographic plumet those species underwent after the European colonisation. Here again, the loss of genetic diversity does not seem to have impacted genetic load as we found no change in variant composition (Figure 5.4), consistent with observations from Wootton et al., (2023).

Conclusion

Using ancient, historical and modern DNA analysis, we were able to retrace the changes in genetic diversity through time for five extant cervids. Our results are consistent with dispersal and demographic histories of our target species. Our analyses reveal patterns of expansion through the Holocene for moose, wapiti and white-tailed & mule deer, and loss of genetic diversity in the last 200 years, consistently with declines in population size concomitant to the large-scale establishment of European settlers.

We also found that levels of heterozygosity in Pleistocene caribou was similar to that of modern samples. Overall, our results exemplify the utility of chronological sampling to better understand external pressures on wild species and how they drive genomic changes. Here, we focussed on DNA data of ultra-low coverage in a limited number of samples, increasing both those aspects would allow to answer precise evolutionary history questions of selection, adaptation or hybridisation for example.

Figures and tables

Figure 5.1: Phylogenetic analysis of mitochondrial consensus sequences performed in IQtree; (A) Consensus tree, nodes coloured by bootstrap support: black ≥ 90%, grey = 70 – 89%, white < 70%; (B) Cladogram, nodes labelled with bootstrap values. Ancient samples identified with and *, colour column represents the species. Species assignment columns and sample names were edited in Inkscape.

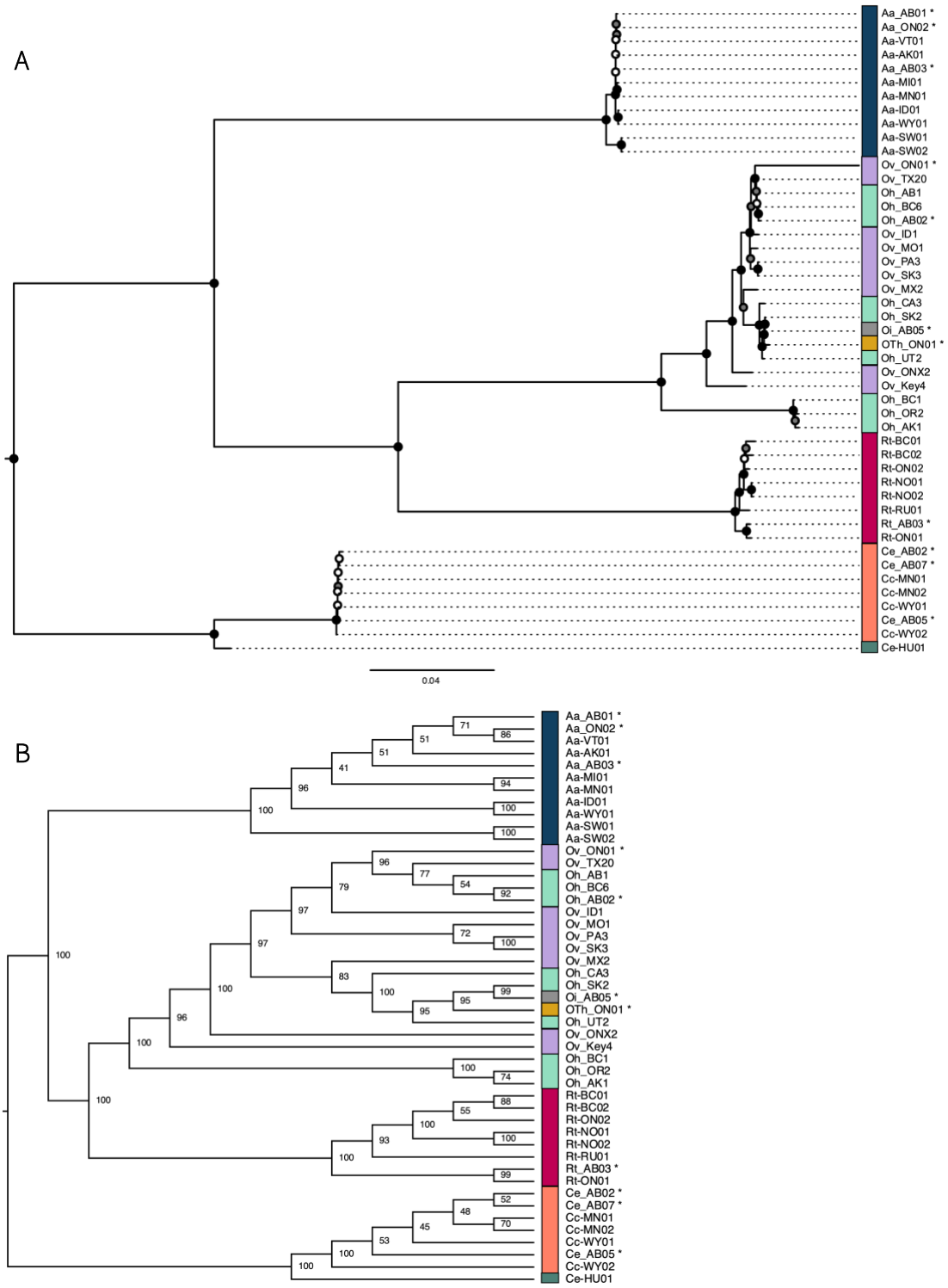


Figure 5.2: Moose (A) Haplotype network and (B) smartPCA; Eurasian samples in yellow, North American in purple and ancient in orange. Subspecies and colours edited in Inkscape.

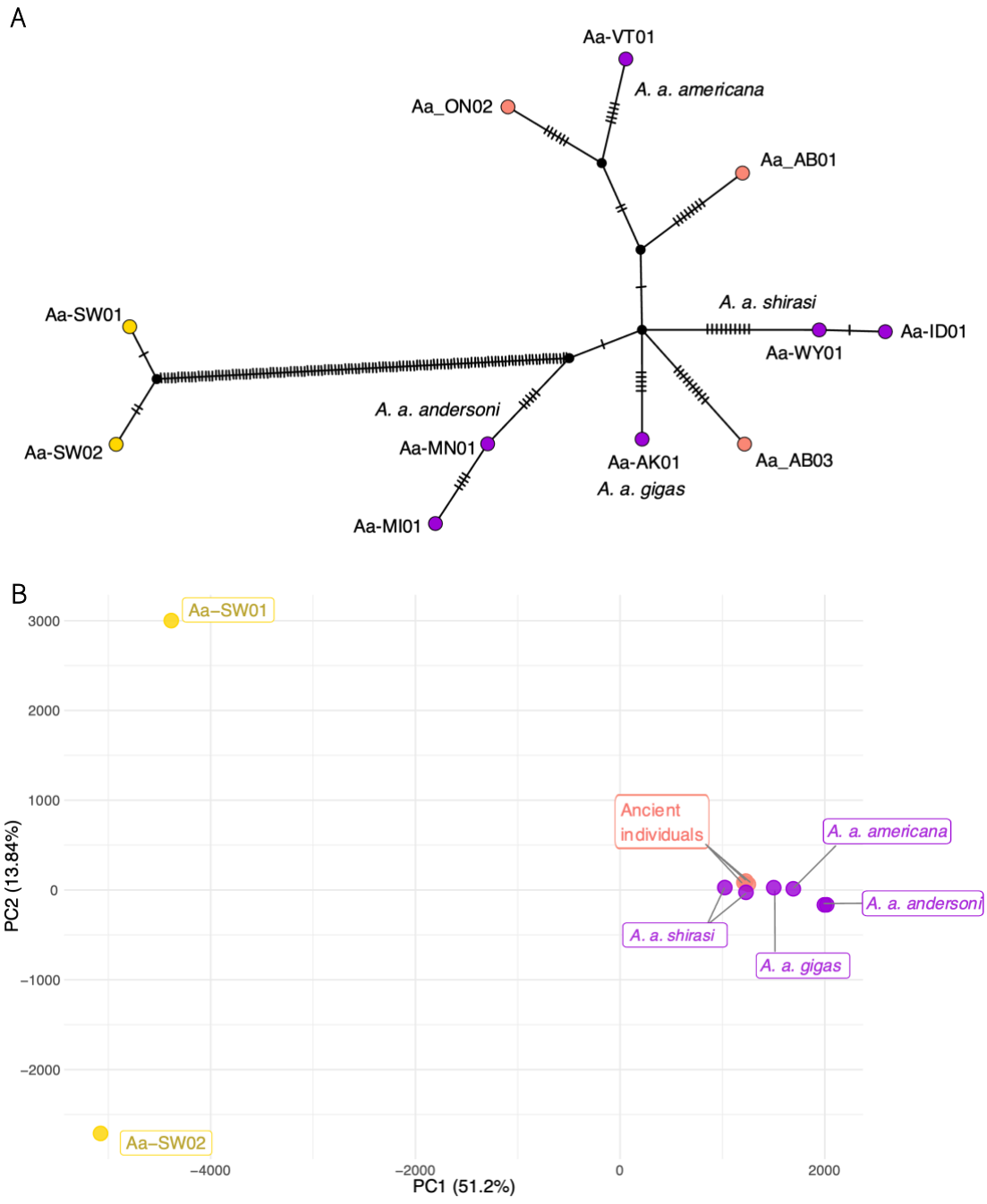


Figure 5.3: Heterozygosity computed in ANGSD, ancient samples as circles, modern samples as triangles, colour represents sample age, ** signifies a p-value < 0.01. Significance was aligned in Inkscape.

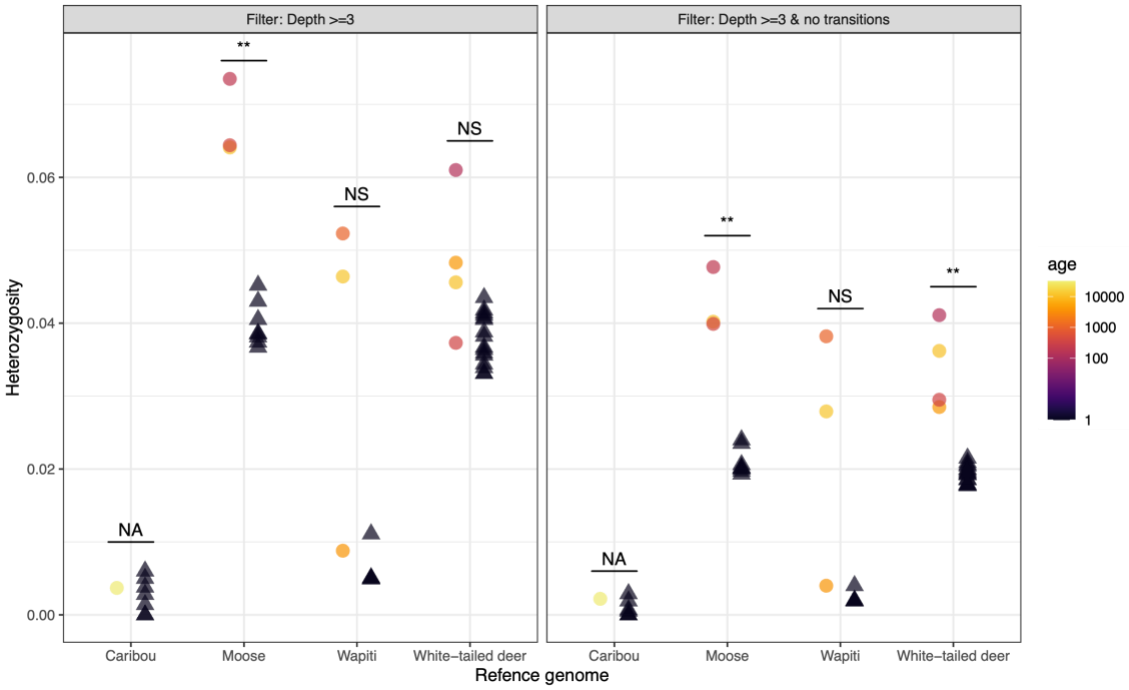


Figure 5.4: Proportion of variants effect, restricted to effects present >5% in one species, transparency denotes filtering: darker = depth ≥ 3, lighter = depth ≥ 3 and C → T & A → G transitions removed. Variant effect column was edited in Inkscape.

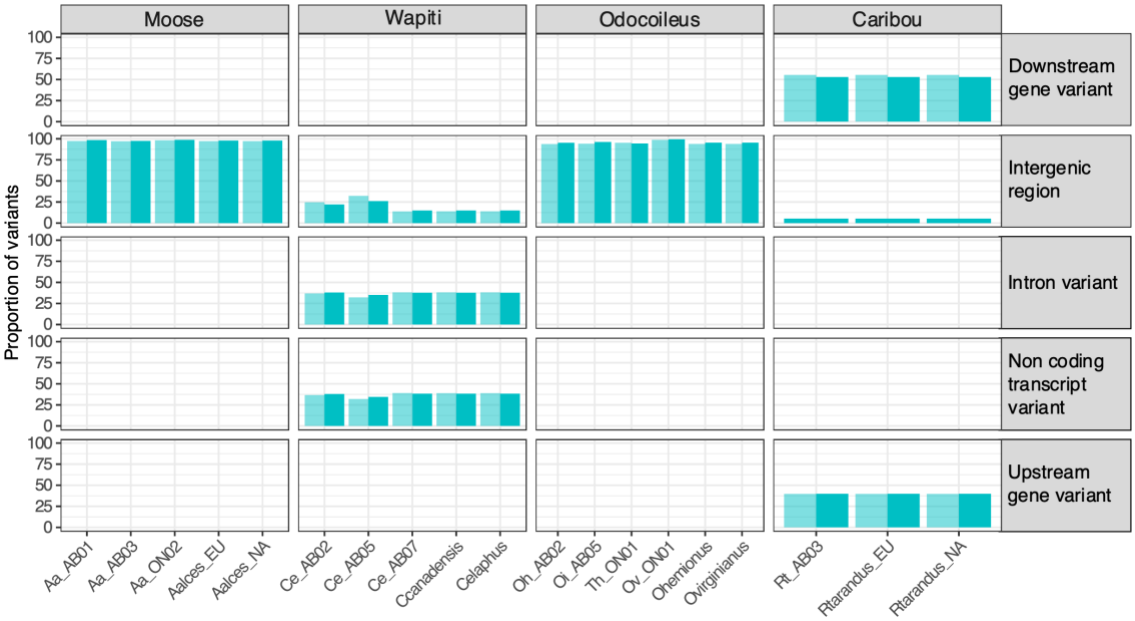


Figure 5.5: Wapiti (A) Haplotype network and (B) smartPCA; Eurasian samples in yellow, North American in purple and ancient in orange. Colours edited in Inkscape.

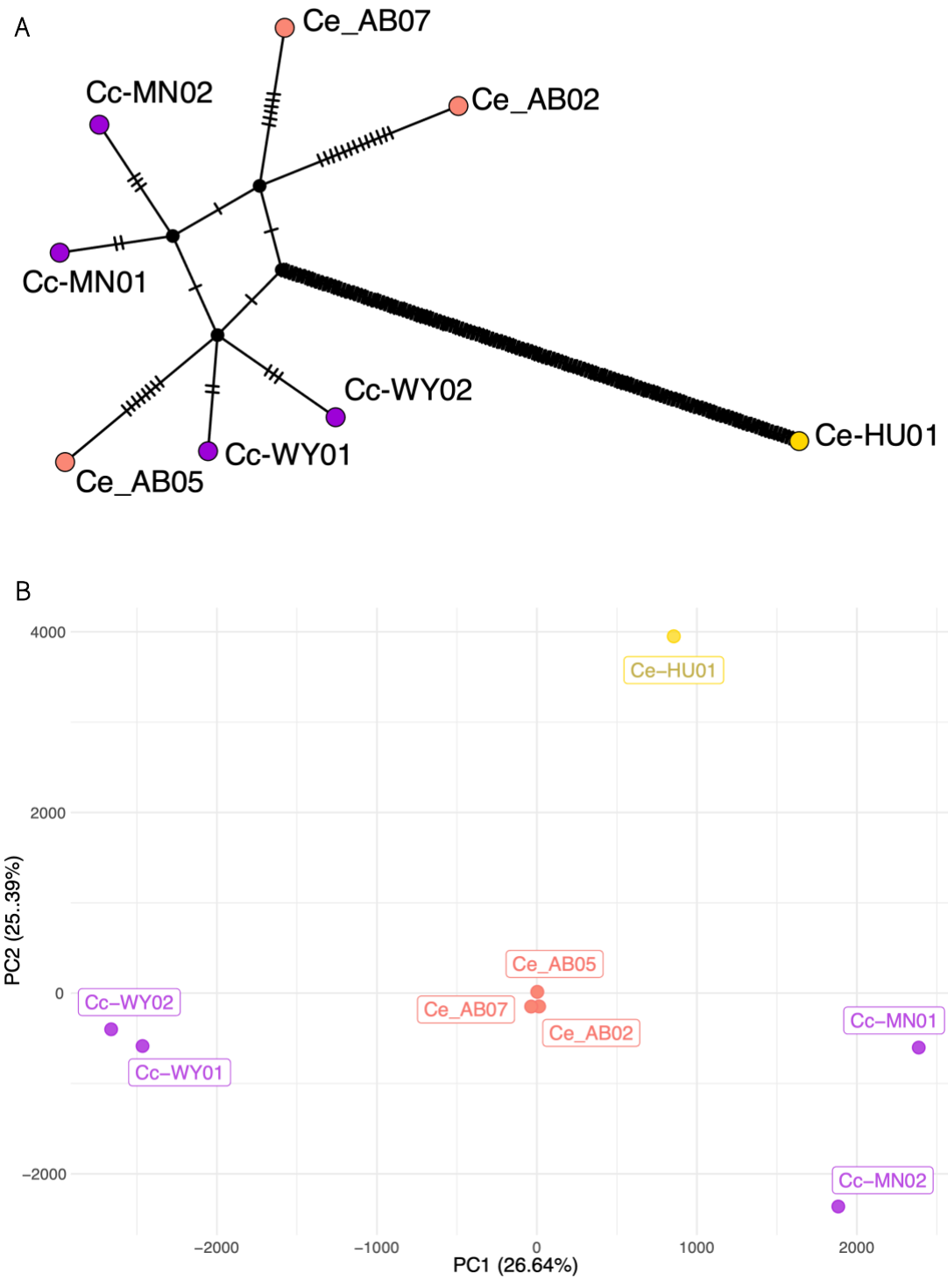
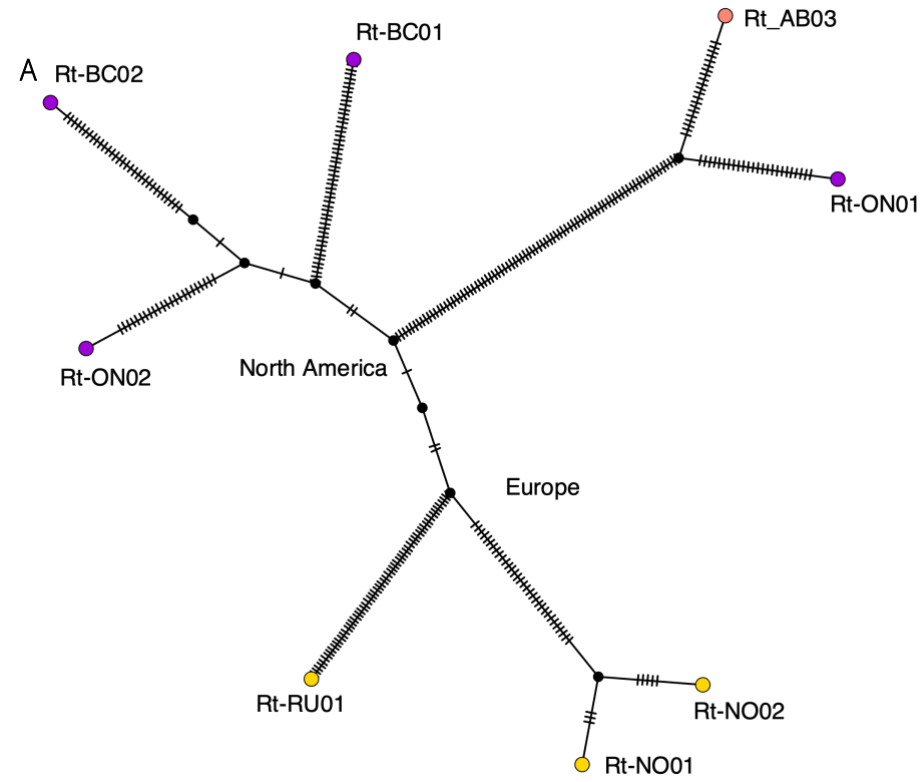


Figure 5.6: Caribou (A) Haplotype network and (B) smartPCA; Eurasian samples in yellow, North American in purple and ancient in orange. Populations and colours were edited in Inkscape.



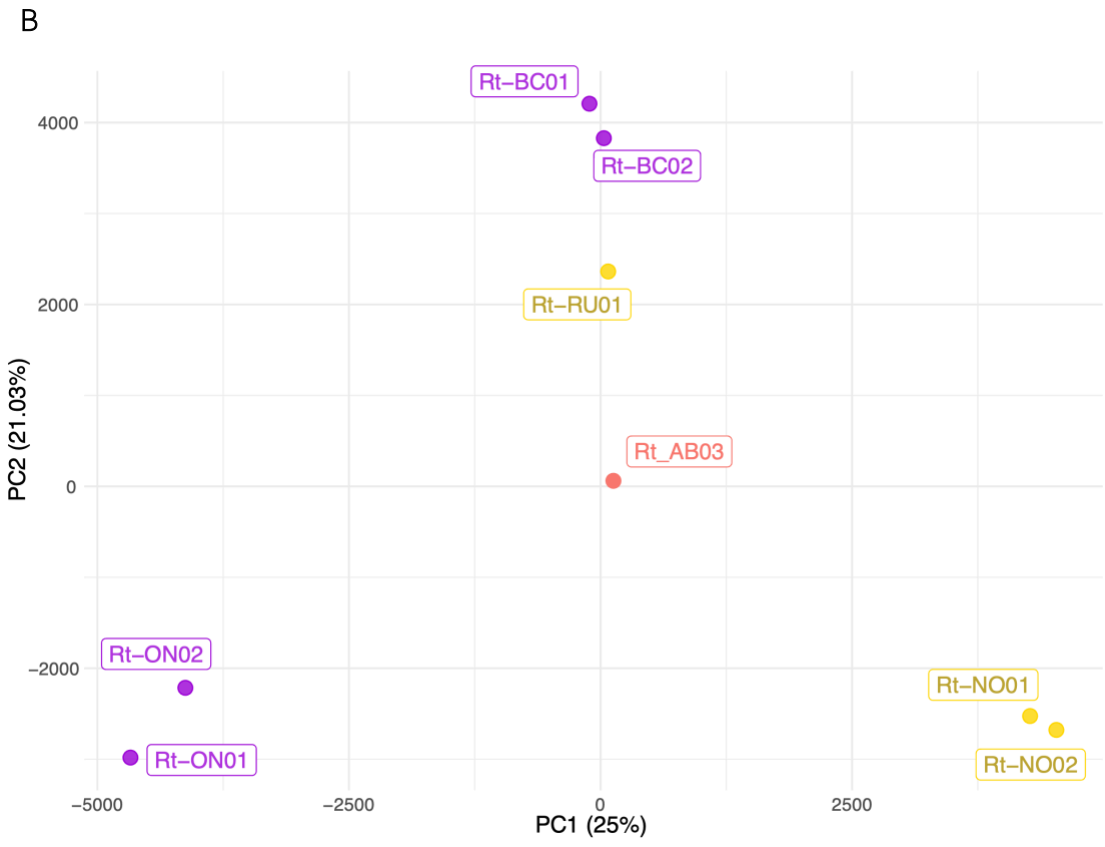
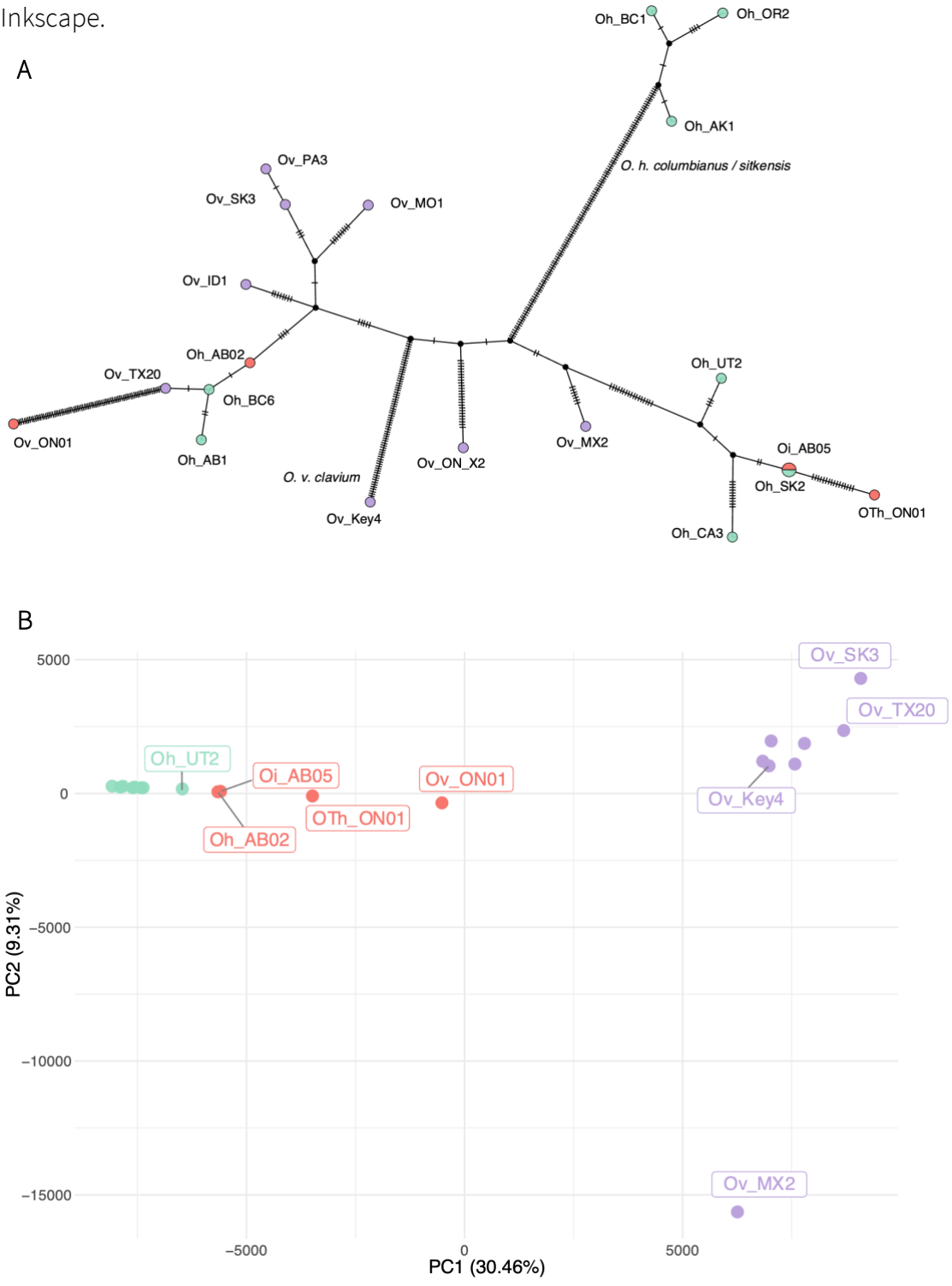


Figure 5.7: *Odocoileus* (A) Haplotype network and (B) smartPCA; Mule deer samples in green, White-tailed deer in purple and ancient in orange. Species and colours edited in Inkscape.



CHAPTER 6 : GENERAL DISCUSSION

Over the last 2 million years, varying climatic and anthropogenic pressures have been influencing North American vegetation and wildlife communities. North American populations of caribou, mule deer and white-tailed deer fluctuated during the glacial cycles of the Pleistocene, those cycles also impacted diversification in white-tailed and mule deer (Combe et al. 2021; Lamb et al. 2021; Taylor et al. 2021; Kessler and Shafer 2024). Moose and Wapiti, on the other hand, entered North America towards the end of those cycles, the bottleneck apparently not strong enough to largely impact genetic diversity (Dussex et al. 2020; Meiri et al. 2020; Ferrante et al. 2022). These dispersals were shortly followed by the extinction several Cervidae genera in the megafaunal extinction event, approximately 11 kya (e.g. Meltzer, 2020). While Indigenous impact on North American deer populations were likely minimal, each species was heavily impacted by the large scale settling of European colonisers in the 19th and 20th centuries (Gill 1999; O'Gara and Dundas 2002; Schaefer 2003; McDonald et al. 2004; DeCesare et al. 2014; Jensen et al. 2023; Rosenblatt et al. 2023). The central aim of my thesis was to understand how climatic and anthropogenic pressures affected Cervidae of Northern America. Using whole genome resequencing data of modern and ancient deer samples, I was able to glean novel insights on the evolutionary history of the most abundant Cervidae genus of northern America, and on changes in genetic diversity through time in this diverse family.

Detailed evolutionary history of *Odocoileus*

My results suggest that white-tailed and mule deer diverged between 280 kya and 2 mya (Kessler et al. 2021; Kessler and Shafer 2024), potentially during a radiation event that also gave rise to the Toronto subway deer (Chapter 4). These estimates

present a wide range and are more recent than previous estimates based on mitochondrial or restricted marker data (e.g. > 4mya in Baccus et al., 1983; > 3 mya in Wright et al., 2022). This wide range is likely due to the particulars of each analytical method which were shown to yield different results, with MSMC ranking among the best program to estimate divergence time from whole genome data (Zhou and Teo 2016). After their divergence, white-tailed populations expanded and fluctuated while mule deer populations declined and remained consistently low. I found that gene flow did not impact the speciation process, rather my results suggest that both species diverged due to drift, and likely remained in different refugia during glacial periods (Chapter 2). While some analyses of my third chapter do recover patterns of ancient gene flow between white-tailed and mule deer, migration rates are low over time and appear later in the speciation process, consistently with secondary contact and the results from Chapter 2. Most of the split times I recovered suggest the impact of major climatic pressures such as glacial periods or deglaciation dynamics, as seen in many species of the region (e.g. Colella et al., 2021; Da Silva Coelho et al., 2023; Taylor et al., 2021), however the inconsistencies between analyses make it hard to accurately pinpoint the timing of population diversification within both species (Chapter 3).

More recently, and additional to the climatic pressures, I found that effective population size (N_e) of white-tailed and mule deer plummeted, concomitantly with the establishment of European settlers on the continent (Chapter 3). The tendency of European settlers to overharvest and change the landscape has been implicated in the declines of many species worldwide (McDonald et al. 2004; Roycroft et al. 2021; Hempel et al. 2024). The demographic patterns I recovered further suggest that the conservation measures taken for white-tailed and mule deer had different outcomes. While white-

tailed deer effective population size shows a rapid recovery, that of mule deer does not (Chapter 3). This absence, coupled to high contemporary N_e/N_c values, are patterns of conservation concern for mule deer.

Altogether, my results picture a detailed evolutionary history of white-tailed and mule deer, even though some major timing uncertainties remain. Both species diverged due to drift during the glacial cycles of the Pleistocene which had diversifying and demographic impact on white-tailed and mule deer. Much later, human impact following European colonisation also affected those species population size; conservation and management measures had drastically different results for white-tailed and mule deer.

Museomics as tool to understand Late Pleistocene extinctions

DNA analysis of ancient or historical museum samples is becoming commonplace and informs on sample origin, systematics and extinction dynamics (e.g. Grewe et al. 2021; Plaxton et al. 2023; Hempel et al. 2024). There is great potential for this technology to increase our understanding of the biodiversity loss that took place in North America at the Pleistocene-Holocene transition. One aspect for which the use of aDNA will be primordial is taxonomy which is often unclear and debated. This was the case for the Toronto subway deer (initially *Torontoceros hypogaeus*) and with reason as my genomic assessment of ROMM75974 suggests it belonged to a distinct species of the *Odocoileus* genus: *O. (Torontoceros) hypogaeus* (Chapter 4). This species is the only known *Odocoileus* to have disappeared during the Late Pleistocene extinction event. This event caused the disappearance of over 30 megafaunal mammal genera in North America and its causes are still highly debated, although they are likely mixed (e.g. Broughton & Weitzel, 2018; Meltzer, 2020; Smith et al., 2022). The extinction of the

Toronto subway deer potentially supports the role of climate, although a complete understanding of the causes of this extinction event will only be possible by telling the story of each individual species, mammal or bird, large or small, extinct or extant.

Using ancient DNA to clarify evolutionary histories

Assessment of past levels of diversity and differentiation generally relies on analytical approaches (i.e. Chapter 3). Recently, the popularisation of aDNA analysis allowed to infer changes in diversity by sampling ancient populations, to reproduce a chronology of genomes with each sample from a specific time point representing the diversity of the species at that time (Dehasque et al. 2020; Orlando et al. 2021). Using this tool to sample populations before and after large scale events of interest, such as the LGM or the colonisation, allows to precisely understand the impact of those events on the genomic landscape of species. Here we used ancient DNA analyses to understand changes in genetic diversity through time in five extant Cervidae species (Chapter 5). I found patterns suggestive of rapid and continued expansion of moose and wapiti since their arrival on North America ~15 kya, with a recent loss of genetic diversity in modern moose samples. My results also suggest adaptation to novel environment in Wapiti which presents change of variant composition over time. Caribou genetic diversity seems to have remained stable over time, but I found signs of several colonisation waves from Beringia. Finally, analysis of *Odocoileus* samples support the hypothesis of incomplete lineage sorting as source of the recognised mito-nuclear discordance between white-tailed and mule deer. Overall, this work highlights exemplify the utility of aDNA to gain a better understanding of wild species' evolutionary history.

References

Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, Boughman J, Brelsford A, Buerkle CA, Buggs R, et al. 2013. Hybridization and Speciation. *J. Evol. Biol.* 26:229–246.

Adams KP, Hamilton RJ. 2011. Management History. In: Hewitt DG, editor. Biology and management of white-tailed deer. Boca Raton: CRC Press. p. 355–378.

Adavoudi R, Pilot M. 2021. Consequences of Hybridization in Mammals: A Systematic Review. *Genes* 13:50.

Aguilar A, Roemer G, Debenham S, Binns M, Garcelon D, Wayne RK. 2004. High MHC Diversity Maintained by Balancing Selection in an otherwise Genetically Monomorphic Mammal. *Proc. Natl. Acad. Sci. U.S.A.* 101:3490–3494.

Albrechtová J, Albrecht T, Baird SJE, Macholán M, Rudolfsen G, Munclinger P, Tucker PK, Piálek J. 2012. Sperm-related phenotypes implicated in both maintenance and breakdown of a natural species barrier in the house mouse. *Proc. R. Soc. B.* 279:4803–4810.

Allen JRM, Forrest M, Hickler T, Singarayer JS, Valdes PJ, Huntley B. 2020. Global vegetation patterns of the past 140,000 years. *J Biogeogr* 47:2073–2090.

Allendorf FW, Hard JJ. 2009. Human-induced evolution caused by unnatural selection through harvest of wild animals. *Proc. Natl. Acad. Sci. U.S.A.* 106:9987–9994.

Allendorf FW, Hohenlohe PA, Luikart G. 2010. Genomics and the future of conservation genetics. *Nat Rev Genet* 11:697–709.

Alminas OSV, Heffelfinger JR, Statham MJ, Latch EK. 2021. Phylogeography of Cedros and Tiburón Island Mule Deer in North America's Desert Southwest.Murphy WJ, editor. *Journal of Heredity* 112:260–275.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215:403–410.

Andrews. 2010. FastQC: a Quality Control Tool for High Throughput Sequence Data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Anthony RG, Smith NS. 1977a. Ecological Relationships between Mule Deer and White-Tailed Deer in Southeastern Arizona. *Ecological Monographs* 47:255–277.

Anthony RG, Smith NS. 1977b. Ecological Relationships between Mule Deer and White-Tailed Deer in Southeastern Arizona. *Ecological Monographs* 47:255–277.

Avise JC, Walker D, Johns GC. 1998. Speciation Durations and Pleistocene Effects on Vertebrate Phylogeography. *Proc. R. Soc. Lond. B* 265:1707–1712.

Baca M. 2020. Diverse responses of common vole (Microtus arvalis) populations to Late Glacial and Early Holocene climate changes - Evidence from ancient DNA. *Quaternary Science Reviews*:13.

Baccus R, Ryman N, Smith MH, Reuterwall C, Cameron D. 1983. Genetic Variability and Differentiation of Large Grazing Mammals. *Journal of Mammalogy* 64:109–120.

Bandelt HJ, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37–48.

Barlow A, Paijmans JLA, Alberti F, Gasparyan B, Bar-Oz G, Pinhasi R, Foronova I, Puzachenko AY, Pacher M, Dalén L, et al. 2021. Middle Pleistocene genome calibrates a revised evolutionary history of extinct cave bears. *Current Biology* 31:1771-1779.e7.

Barnett DW, Garrison EK, Quinlan AR, Strömberg MP, Marth GT. 2011. BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* 27:1691–1692.

Barnosky AD, Koch PL, Feranec RS, Wing SL, Shabel AB. 2004. Assessing the Causes of Late Pleistocene Extinctions on the Continents. *Science* 306:70–75.

Barratt CD, Bwong BA, Jehle R, Liedtke HC, Nagel P, Onstein RE, Portik DM, Streicher JW, Loader SP. 2018. Vanishing refuge? Testing the forest refuge hypothesis in coastal East Africa using genome-wide sequence data for seven amphibians. *Mol Ecol* 27:4289–4308.

Batchelor CL, Margold M, Krapp M, Murton DK, Dalton AS, Gibbard PL, Stokes CR, Murton JB, Manica A. 2019. The configuration of Northern Hemisphere ice sheets through the Quaternary. *Nat Commun* 10:3713.

Benfield AJ, Ivory SJ, Hodelka BN, Zimmerman SRH, McGlue MM. 2023. Terrestrial ecosystem transformations in response to rapid climate change during the last deglaciation around Mono Lake, California, USA. *Quat. res.* 113:87–104.

Bennett MR, Bustos D, Pigati JS, Springer KB, Urban TM, Holliday VT, Reynolds SC, Budka M, Honke JS, Hudson AM, et al. 2021. Evidence of humans in North America during the Last Glacial Maximum. *Science* 373:1528–1531.

Bergman EJ, Doherty PF, White GC, Holland AA. 2015. Density dependence in mule deer: a review of evidence. *Wildlife Biology* 21:18–29.

Bergman J, Pedersen RØ, Lundgren EJ, Lemoine RT, Monsarrat S, Pearce EA, Schierup MH, Svenning J-C. 2023. Worldwide Late Pleistocene and Early Holocene population declines in extant megafauna are associated with Homo sapiens expansion rather than climate change. *Nat Commun* 14:7679.

Bernatchez L, Landry C. 2003. MHC Studies in Nonmodel Vertebrates: What Have we Learned about Natural Selection in 15 Years? *Journal of Evolutionary Biology* 16:363–377.

Berry SL, Shipley LA, Long RA, Loggers C. 2019. Differences in Dietary Niche and Foraging Behavior of Sympatric Mule and White-Tailed Deer. *Ecosphere* 10:e02815.

Bhattacharyya T, Gregorova S, Mihola O, Anger M, Sebestova J, Denny P, Simecek P, Forejt J. 2013. Mechanistic basis of infertility of mouse intersubspecific hybrids. *Proc. Natl. Acad. Sci. U.S.A.* 110:E468–E477.

Blank DA, Yang W. 2017. Mother-young Recognition in Goitered Gazelle during Hiding Period. *Behavioural Processes* 142:21–28.

Boeskorov GG. 2005. A review of the systematics of Pliocene and Pleistocene moose, part 1. *Cranium* 22:26–55.

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

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

Booker TR, Yeaman S, Whitlock MC. 2021. Global Adaptation Complicates the Interpretation of Genome Scans for Local Adaptation. *Evolution Letters* 5:4–15.

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis.Prlic A, editor. *PLoS Comput Biol* 10:e1003537.

Bradley RD, Baker RJ. 2001. A Test of the Genetic Species Concept: Cytochrome-B Sequences and Mammals. *Journal of Mammalogy* 82:960–973.

Bradley RD, Bryant FC, Bradley LC, Haynie ML, Baker RJ. 2003. IMPLICATIONS OF HYBRIDIZATION BETWEEN WHITE-TAILED DEER AND MULE DEER. *The Southwestern Naturalist* 48:7.

Brain RA, Prosser RS. 2022. Human induced fish declines in North America, how do agricultural pesticides compare to other drivers? *Environ Sci Pollut Res* 29:66010–66040.

Brasseur MV, Astrin JJ, Geiger MF, Mayer C. 2023. MitoGeneExtractor: Efficient extraction of mitochondrial genes from next-generation sequencing libraries. *Methods Ecol Evol* 14:1017–1024.

Broad Institute. 2019. Picard Toolkit.

Broad Institute, GitHub Repository. 2019. "Picard Toolkit." Available from: http://broadinstitute.github.io/picard/

Bro-Jørgensen MH, Keighley X, Ahlgren H, Scharff-Olsen CH, Rosing-Asvid A, Dietz R, Ferguson SH, Gotfredsen AB, Jordan P, Glykou A, et al. 2021. Genomic sex identification of ancient pinnipeds using the dog genome. *Journal of Archaeological Science* 127:105321.

Broughton JM, Weitzel EM. 2018. Population reconstructions for humans and megafauna suggest mixed causes for North American Pleistocene extinctions. *Nat Commun* 9:5441.

Brunjes KJ, Ballard WB, Humphrey MH, Harwell F, Mcintyre NE, Krausman PR, Wallace MC. 2006. Habitat Use by Sympatric Mule and White-Tailed Deer in Texas. *Journal of Wildlife Management* 70:1351–1359.

Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A. 2012. Inferring Species Trees Directly from Biallelic Genetic Markers: Bypassing Gene Trees in a Full Coalescent Analysis. *Molecular Biology and Evolution* 29:1917–1932.

Budd K, Berkman LK, Anderson M, Koppelman J, Eggert LS. 2018. Genetic Structure and Recovery of White-tailed Deer in Missouri. *Jour. Wild. Mgmt.* 82:1598–1607.

Burri R. 2017. Interpreting Differentiation Landscapes in the Light of Long-term Linked Selection. *Evolution Letters* 1:118–131.

Cambronne A. 2013. Deerland: America's Hunt for Ecological Balance and the Essence of Wildness. Lyons Press

Campbell CR, Poelstra JW, Yoder AD. 2018. What is Speciation Genomics? The Roles of Ecology, Gene Flow, and Genomic Architecture in the Formation of Species. *Biological Journal of the Linnean Society* 124:561–583.

Campbell TA, VerCauteren KC. 2011. Diseases and Parasites. In: Biology and management of white-tailed deer. Boca Raton: CRC Press. p. 220–240.

Cann P, Chabi M, Delsart A, Le Danvic C, Saliou J-M, Chasles M, Keller M, Nagnan-Le Meillour P. 2019. The Olfactory Secretome Varies According to Season in Female Sheep and Goat. *BMC Genomics* 20:794.

Carr SM, Ballinger SW, Derr JN, Blankenship LH, Bickham JW. 1986. Mitochondrial DNA Analysis of Hybridization between Sympatric White-tailed Deer and Mule Deer in West Texas. *Proceedings of the National Academy of Sciences* 83:9576–9580.

Carr SM, Hughes GA. 1993. Direction of Introgressive Hybridization between Species of North American Deer (Odocoileus) as Inferred from Mitochondrial-Cytochrome-b Sequences. *Journal of Mammalogy* 74:331–342.

Cars BS, Kessler C, Hoffman EA, Côté SD, Koelsch D, Shafer ABA. 2023. Island demographics and trait associations in white-tailed deer. bioRxiv Available from: http://biorxiv.org/lookup/doi/10.1101/2023.08.01.551454

Cathey JC, Bickham JW, Patton JC. 1998. Introgressive Hybridization and Nonconcordant Evolutionary History of Maternal and Paternal Lineages in North American Deer. *Evolution* 52:1224–1229.

Ceballos G, Ehrlich PR. 2023. Mutilation of the tree of life via mass extinction of animal genera. *Proc. Natl. Acad. Sci. U.S.A.* 120:e2306987120.

Ceballos G, Ehrlich PR, Barnosky AD, García A, Pringle RM, Palmer TM. 2015. Accelerated modern human–induced species losses: Entering the sixth mass extinction. *Sci. Adv.* 1:e1400253.

Ceballos G, Ehrlich PR, Raven PH. 2020. Vertebrates on the brink as indicators of biological annihilation and the sixth mass extinction. *Proc. Natl. Acad. Sci. U.S.A.* 117:13596–13602.

Chafin TK, Zbinden ZD, Douglas MR, Martin BT, Middaugh CR, Gray MC, Ballard JR, Douglas ME. 2021. Spatial Population Genetics in Heavily Managed Species: Separating Patterns of Historical Translocation from Contemporary Gene Flow in White-tailed Deer. *Evol Appl* 14:1673–1689.

Chamaillé-Jammes S, Malcuit H, Le Saout S, Martin J-L. 2014. Innate Threat-sensitive Foraging: Black-tailed Deer Remain More Fearful of Wolf than of the less Dangerous Black Bear even after 100 years of Wolf Absence. *Oecologia* 174:1151–1158.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaSci* 4:7.

Charlesworth B. 2009. Effective population size and patterns of molecular evolution and variation. *g e n e ti c s*:11.

Cheddadi R, de Beaulieu J-L, Jouzel J, Andrieu-Ponel V, Laurent J-M, Reille M, Raynaud D, Bar-Hen A. 2005. Similarity of vegetation dynamics during interglacial periods. *Proceedings of the National Academy of Sciences* 102:13939–13943.

Chen L, Qiu Q, Jiang Y, Wang K, Lin Z, Li Z, Bibi F, Yang Y, Wang J, Nie W, et al. 2019. Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. *Science* 364:eaav6202.

Cherry MJ, Barton BT. 2017. Effects of Wind on Predator-Prey Interactions. *Food Webs* 13:92–97.

Chigurapati S, Sulak M, Miller W, Lynch VJ. 2018. Relaxed constraint and thermal desensitization of the cold-sensing ion channel TRPM8 in mammoths. Available from: http://biorxiv.org/lookup/doi/10.1101/397356

Churcher CS, Peterson RL. 1982. Chronologic and Environmental Implications of a New Genus of Fossil Deer from Late Wisconsin Deposits at Toronto, Canada. *Quat. res.* 18:184–195.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* 6:80–92.

Clark PU, Dyke AS, Shakun JD, Carlson AE, Clark J, Wohlfarth B, Mitrovica JX, Hostetler SW, McCabe AM. 2009. The Last Glacial Maximum. *Science* 325:710–714.

Clements D, Young A. 1997. A viewpoint: Rangeland health and mule deer habitat. *JOURNAL OF RANGE MANAGEMENT*.

Colella JP, Lan T, Talbot SL, Lindqvist C, Cook JA. 2021. Whole-genome Resequencing Reveals Persistence of Forest-Associated Mammals in Late Pleistocene Refugia along North America's North Pacific Coast. *J. Biogeogr.* 48:1153–1169.

Combe FJ, Jaster L, Ricketts A, Haukos D, Hope AG. 2021. Population Genomics of Free-ranging Great Plains White-tailed and Mule deer Reflects a Long History of Inter-specific Hybridization. *Evolutionary Applications* 15:111–132.

Commerford JL, Gittens G, Gainforth S, Wilson JJ, Bird BW. 2022. Differences in forest composition following two periods of settlement by pre-Columbian Native Americans. *Veget Hist Archaeobot* 31:467–480.

Croitor R. 2022. Paleobiogeography of Crown Deer. *Earth* 3:1138–1160.

Cronin MA. 1991. Mitochondrial and Nuclear Genetic Relationships of Deer (Odocoileus spp.) in Western North America. *Can. J. Zool.* 69:1270–1279.

Cronin MA, Vyse ER, Cameron DG. 1988. Genetic Relationships between Mule Deer and White-Tailed Deer in Montana. *The Journal of Wildlife Management* 52:320.

Cruickshank TE, Hahn MW. 2014. Reanalysis Suggests that Genomic Islands of Speciation are Due to Reduced Diversity, not Reduced Gene Flow. *Mol Ecol* 23:3133–3157.

Cutter AD. 2013. Integrating Phylogenetics, Phylogeography and Population Genetics through Genomes and Evolutionary Theory. *Molecular Phylogenetics and Evolution* 69:1172–1185.

Da Silva Coelho FA, Gill S, Tomlin CM, Papavassiliou M, Farley SD, Cook JA, Sonsthagen SA, Sage GK, Heaton TH, Talbot SL, et al. 2023. Ancient bears provide insights into Pleistocene ice age refugia in Southeast Alaska. *Molecular Ecology*:mec.16960.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27:2156–2158.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience* 10:giab008.

Darwin C 1809-1882. 1859. On the Origin of Species by Means of Natural Selection, or Preservation of Favoured Races in the Struggle for Life. London : John Murray, 1859 Available from: https://search.library.wisc.edu/catalog/9934839413602122

Dawe KL, Boutin S. 2016. Climate change is the primary driver of white-tailed deer (Odocoileus virginianus) range expansion at the northern extent of its range; land use is secondary. *Ecol Evol* 6:6435–6451.

De Jong MJ, Li Z, Qin Y, Quéméré E, Baker K, Wang W, Hoelzel AR. 2020. Demography and adaptation promoting evolutionary transitions in a mammalian genus that diversified during the Pleistocene. *Molecular Ecology* 29:2777–2792.

DeCesare NJ, Smucker TD, Garrott RA, Gude JA. 2014. Moose Status and Management in Montana. *ALCES VOL.* 50.

Dedato M, Robert C, Taillon J, Shafer A, Cote S. 2022. Demographic History and Conservation Genomics of Caribou (Rangifer tarandus) in Québec. Preprints Available from: https://www.authorea.com/users/407174/articles/552124-demographic-history-and-conservation-genomics-of-caribou-rangifer-tarandus-in-qu%C3%A9bec?commit=d3fe88936f0714b1711f9dba33bc149835a4aba3

Dedato MN, Robert C, Taillon J, Shafer ABA, Côté SD. 2022. Demographic history and conservation genomics of caribou (Rangifer tarandus) in Québec. *Evolutionary Applications* 15:2043–2053.

Deer Friendly. Deer Population. *Deer Friendly* [Internet]. Available from: https://www.deerfriendly.com/decline-of-deer-populations

Degl'Innocenti A, D'Errico A. 2017. Regulatory Features for Odorant Receptor Genes in the Mouse Genome. *Frontiers in Genetics* 8:8.

Dehasque M, Ávila-Arcos MC, Díez-del-Molino D, Fumagalli M, Guschanski K, Lorenzen ED, Malaspinas A, Marques-Bonet T, Martin MD, Murray GGR, et al. 2020. Inference of natural selection from ancient DNA. *Evolution Letters* 4:94–108.

Dehasque M, Pečnerová P, Kempe Lagerholm V, Ersmark E, Danilov GK, Mortensen P, Vartanyan S, Dalén L. 2022. Development and Optimization of a Silica Column-Based Extraction Protocol for Ancient DNA. *Genes* 13:687.

Delaneau O, Marchini J, Zagury J-F. 2012. A linear complexity phasing method for thousands of genomes. *Nat Methods* 9:179–181.

Demarais S, Miller KV, Jacobson HA. 2000. White-tailed deer. In: Demarais S, Krausman PR, editors. Ecology and management of large mammals in North America. Upper Saddle River, NJ: Prentice Hall.

Dembitzer J, Castiglione S, Raia P, Meiri S. 2022. Small brains predisposed Late Quaternary mammals to extinction. *Sci Rep* 12:3453.

Derr JN. 1991. Genetic Interactions between White-Tailed and Mule Deer in the Southwestern United States. *The Journal of Wildlife Management* 55:228.

Derr JN, Hale DW, Ellsworth DL, Bickham JW. 1991. Fertility in an F1 Male Hybrid of White-tailed Deer (Odocoileus virginianus) x Mule deer (O. hemionus). :7.

Deutsch JC, Nefdt RJC. 1992. Olfactory Cues Influence Female Choice in two Lek-breeding Antelopes. *Nature* 356:596–598.

DeVivo MT, Edmunds DR, Kauffman MJ, Schumaker BA, Binfet J, Kreeger TJ, Richards BJ, Schätzl HM, Cornish TE. 2017. Endemic chronic wasting disease causes mule deer population decline in Wyoming.Zabel MD, editor. *PLoS ONE* 12:e0186512.

Deyoung RW, Demarais S, Honeycutt RL, Rooney AP, Gonzales RA, Gee KL. 2003. Genetic Consequences of White-tailed Deer (Odocoileus virginianus) Restoration in Mississippi. *Molecular Ecology* 12:3237–3252.

Díez-del-Molino D, Dehasque M, Chacón-Duque JC, Pečnerová P, Tikhonov A, Protopopov A, Plotnikov V, Kanellidou F, Nikolskiy P, Mortensen P, et al. 2023. Genomics of adaptive evolution in the woolly mammoth. *Current Biology* 33:1753-1764.e4.

Ditchkoff SS. 2011. Anatomy and Physiology. In: Hewitt DG, editor. Biology and management of white-tailed deer. Boca Raton: CRC Press. Available from: http://site.ebrary.com/id/11007093

Dong W, Pan Y, Liu J. 2004. The earliest Muntiacus (Artiodactyla, Mammalia) from the Late Miocene of Yuanmou, southwestern China. *Comptes Rendus Palevol* 3:379–386.

Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, Marquéz JRG, Gruber B, Lafourcade B, Leitão PJ, et al. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36:27–46.

Douzery E, Randi E. 1997. The Mitochondrial Control Region of Cervidae: Evolutionary Patterns and Phylogenetic Content. *Molecular Biology and Evolution* 14:1154–1166.

Durand EY, Patterson N, Reich D, Slatkin M. 2011. Testing for Ancient Admixture between Closely Related Populations. *Molecular Biology and Evolution* 28:2239–2252.

Dussex N, Alberti F, Heino MT, Olsen R-A, van der Valk T, Ryman N, Laikre L, Ahlgren H, Askeyev IV, Askeyev OV, et al. 2020. Moose Genomes Reveal Past Glacial Demography and the Origin of Modern Lineages. *BMC Genomics* 21:854.

Elias S, Schreve D. 2007. Late Pleistocene Megafaunal Extinctions. In: Encyclopedia of Quaternary Science. 2nd ed. Encyclopedia of Quaternary Science. Amsterdam: Elsevier. p. 700–712.

Ellsworth DL, Honeycutt RL, Silvy NJ, Bickham JW, Klimstra WD. 1994. Historical Biogeography and Contemporary Patterns Of Mitochondrial DNA Variation In White-Tailed Deer From The Southeastern United States. *Evolution* 48:122–136.

Ersmark E, Baryshnikov G, Higham T, Argant A, Castaños P, Döppes D, Gasparik M, Germonpré M, Lidén K, Lipecki G, et al. 2019. Genetic turnovers and northern survival during the last glacial maximum in European brown bears. *Ecol Evol* 9:5891–5905.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the Number of Clusters of Individuals Using the Software Structure: a Simulation Study. *Mol Ecol* 14:2611–2620.

Ewels P, Magnusson M, Lundin S, Käller M. 2016. MultiQC: Summarize Analysis Results for Multiple Tools and Samples in a Single Report. *Bioinformatics* 32:3047–3048.

Feder JL. 2012. The Genomics of Speciation-With-Gene-Flow. 28:9.

Ferchaud A-L, Perrier C, April J, Hernandez C, Dionne M, Bernatchez L. 2016. Making sense of the relationships between Ne, Nb and Nc towards defining conservation thresholds in Atlantic salmon (Salmo salar). *Heredity* 117:268–278.

Ferrante JA, Smith CH, Thompson LM, Hunter ME. 2022. Genome-wide SNP analysis of three moose subspecies at the southern range limit in the contiguous United States. *Conserv Genet* 23:109–121.

Festa-Bianchet M, Ray JC, Boutin S, Côté SD, Gunn A. 2011. Conservation of caribou (Rangifer tarandus) in Canada: an uncertain future. *Can. J. Zool.* 89:419–434.

Fick SE, Hijmans RJ. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37:4302–4315.

Firestone RB, West A, Kennett JP, Becker L, Bunch TE, Revay ZS, Schultz PH, Belgya T, Kennett DJ, Erlandson JM, et al. 2007. Evidence for an extraterrestrial impact 12,900 years ago that contributed to the megafaunal extinctions and the Younger Dryas cooling. *Proc. Natl. Acad. Sci. U.S.A.* 104:16016–16021.

Firneno TJ, O'Neill JR, Portik DM, Emery AH, Townsend JH, Fujita MK. 2020. Finding complexity in complexes: Assessing the causes of mitonuclear discordance in a problematic species complex of Mesoamerican toads. *Mol Ecol* 29:3543–3559.

Flagstad Ø, Røed KH. 2003. REFUGIAL ORIGINS OF REINDEER (RANGIFER TARANDUS L.) INFERRED FROM MITOCHONDRIAL DNA SEQUENCES. *Evolution* 57:658–670.

Forrester TD, Wittmer HU. 2013. A review of the population dynamics of mule deer and black-tailed deer Odocoileus hemionus in North America: Population dynamics of mule deer and black-tailed deer. *Mammal Review* 43:292–308.

Frankham R. 1995. Effective population size/adult population size ratios in wildlife: a review. *Genet. Res.* 66:95–107.

Frankham R. 2005. Genetics and extinction. *Biological Conservation* 126:131–140.

Fréchette B, De Vernal A. 2013. Evidence for large-amplitude biome and climate changes in Atlantic Canada during the last interglacial and mid-Wisconsinan periods. *Quat. res.* 79:242–255.

Fulton AE, Yansa CH. 2021. Onset of the Paleoanthropocene in the Lower Great Lakes Region of North America: An Archaeological and Paleoecological Synthesis. *Annals of the American Association of Geographers* 111:771–783.

Fusco NA, Cosentino BJ, Gibbs JP, Allen ML, Blumenfeld AJ, Boettner GH, Carlen EJ, Collins M, Dennison C, DiGiacopo D, et al. 2023. Population genomic structure of a widespread, urban-dwelling mammal: The eastern grey squirrel (Sciurus carolinensis). *Molecular Ecology*:mec.17230.

Gajewski K, Kriesche B, Chaput MA, Kulik R, Schmidt V. 2019. Human–vegetation interactions during the Holocene in North America. *Veget Hist Archaeobot* 28:635–647.

Galili T, O'Callaghan A, Sidi J, Sievert C. 2018. heatmaply: an R package for creating interactive cluster heatmaps for online publishing.Wren J, editor. *Bioinformatics* 34:1600–1602.

Gao F, Ming C, Hu W, Li H. 2016. New Software for the Fast Estimation of Population Recombination Rates (FastEPRR) in the Genomic Era. *G3 Genes|Genomes|Genetics* 6:1563–1571.

Garrison E, Kronenberg ZN, Dawson ET, Pedersen BS, Prins P. 2022. A spectrum of free software tools for processing the VCF variant call format: vcflib, bio-vcf, cyvcf2, hts-nim and slivar.Schneidman-Duhovny D, editor. *PLoS Comput Biol* 18:e1009123.

Gilbert C, Ropiquet A, Hassanin A. 2006. Mitochondrial and Nuclear Phylogenies of Cervidae (Mammalia, Ruminantia): Systematics, Morphology, and Biogeography. *Molecular Phylogenetics and Evolution* 40:101–117.

Gill RB. 1999. Declining Mule Deer Populations in Colorado: Reasons and Responses (a report to the Colorado Legislature, November 1999). 58: Colorado Division of Wildlife

Glotzhober RC, McDonald HG. 2015. Partial Skeleton of an Elk-moose, Cervalces scotti, from Chippewa Lake, Medina County, Ohio. *PaleoAmerica* 1:332–342.

Gowan EJ, Zhang X, Khosravi S, Rovere A, Stocchi P, Hughes ALC, Gyllencreutz R, Mangerud J, Svendsen J-I, Lohmann G. 2021. A new global ice sheet reconstruction for the past 80 000 years. *Nat Commun* 12:1199.

Gower G, Tuke J, Rohrlach A, Soubrier J, Llamas B, Bean N, Cooper A. 2018. Population Size History from Short Genomic Scaffolds: How Short is Too Short? Genetics Available from: http://biorxiv.org/lookup/doi/10.1101/382036

Graham RW, Lundelius EL. 1984. Coevolutionary Disequilibrium and Pleistocene Extinctions. In: Martin PS, Klein RG, editors. Quaternary Extinctions: A Prehistoric Revolution. University of Arizona Press. p. 223–249. Available from: http://www.jstor.org/stable/10.2307/j.ctv264f91j

Grayson DK. 1977. Pleistocene Avifaunas and the Overkill Hypothesis. *Science* 195:691–693.

Grayson DK. 2007. Deciphering North American Pleistocene Extinctions. *Journal of Anthropological Research* 63:185–213.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MHY, et al. 2010. A Draft Sequence of the Neandertal Genome. *Science* 328:710–722.

Greenslade AD. 1998. Inter- and intraspecific phylogeography of North American Odocoileus Deer based on mitochondrial DNA sequences. Available from: https://research.library.mun.ca/1190/

Grewe F, Kronforst MR, Pierce NE, Moreau CS. 2021. Museum genomics reveals the Xerces blue butterfly (Glaucopsyche xerces) was a distinct species driven to extinction. *Biol. Lett.* 17:20210123.

Grinberg S. 2019. findPeaks. *GitHub* [Internet]. Available from: https://github.com/stas-g/findPeaks/blob/master/find_peaks.R

Grossen C, Guillaume F, Keller LF, Croll D. 2020. Purging of highly deleterious mutations through severe bottlenecks in Alpine ibex. *Nat Commun* 11:1001.

Groves P, Mann DH, Kunz ML. 2022. Prehistoric perspectives can help interpret the present: 14 000 years of moose (*Alces alces*) in the Western Arctic. *Can. J. Zool.* 100:732–746.

Gruell GE. 1986. Post-1900 mule deer irruptions in the Intermountain West: Principal cause and influences. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station Available from: https://www.fs.usda.gov/treesearch/pubs/37804

Gustafson EP. 2015. An early Pliocene North American deer: Bretzia pseudalces, its osteology, biology, and place in cervid history. Museum of Natural History, University of Oregon Eugene, OR, USA

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data.McVean G, editor. *PLoS Genet* 5:e1000695.

Gutiérrez EE, Helgen KM, McDonough MM, Bauer F, Hawkins MTR, Escobedo-Morales LA, Patterson BD, Maldonado JE. 2017. A gene-tree test of the traditional taxonomy of American deer: the importance of voucher specimens, geographic data, and dense sampling. *ZK* 697:87–131.

Hahn MW. 2018. Molecular population genetics. New York : Sunderland, MA: Oxford University Press ; Sinauer Associates

Hanberry BB. 2023. Shifting potential tree species distributions from the Last Glacial Maximum to the Mid-Holocene in North America, with a correlation assessment. *J Quaternary Science* 38:829–839.

Harington CR, Mol D, Van Der Plicht J. 2012. The Muirkirk Mammoth: A Late Pleistocene woolly mammoth (Mammuthus primigenius) skeleton from southern Ontario, Canada. *Quaternary International* 255:106–113.

Heffelfinger J, Krausman PR eds. 2023. Ecology and management of black-tailed and mule deer of North America. First edition. Boca Raton, FL: CRC Press

Heffelfinger JR. 2011. Taxonomy, Evolutionary History, and Distribution. In: Hewitt DG, editor. Biology and management of white-tailed deer. Boca Raton: CRC Press. Available from: http://site.ebrary.com/id/11007093

Heffelfinger JR, Latch EK. 2023. Origin, classification, and distribution. In: Ecology and management of black-tailed and mule deer of north america. CRC Press. p. 3–24.

Hempel E, Bibi F, Faith JT, Koepfli K-P, Klittich AM, Duchêne DA, Brink JS, Kalthoff DC, Dalén L, Hofreiter M, et al. 2022. Blue Turns to Gray: Paleogenomic Insights into the Evolutionary

History and Extinction of the Blue Antelope (Hippotragus leucophaeus). *Molecular Biology and Evolution* 39:msac241.

Hempel E, Faith JT, Preick M, Jager D de, Barish S, Hartmann S, Grau JH, Moodley Y, Gedman G, Pirovich KM, et al. 2024. Colonial-driven extinction of the blue antelope despite genomic adaptation to low population size. *Current Biology* [Internet] 0. Available from: https://www.cell.com/current-biology/abstract/S0960-9822(24)00391-9

Herrando-Pérez S, Tobler R, Huber CD. 2021. smartsnp, an r package for fast multivariate analyses of big genomic data. *Methods in Ecology and Evolution* 12:2084–2093.

Hewitt DG. 2011. Biology and management of white-tailed deer. Boca Raton: CRC Press

Hewitt G. 2000. The Genetic Legacy of the Quaternary Ice Ages. *Nature* 405:907–913.

Hijmans RJ. 2022. raster: Geographic Data Analysis and Modeling. Available from: https://CRAN.R-project.org/package=raster

Hirata M, Kusatake N. 2021. Relative Importance of Senses in Forage Discrimination by Cattle Depends on the Sensory Contrast between the Discrimination Targets: a Preliminary Study. *Anim Cogn* 24:99–106.

Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution* 35:518–522.

Hold K, Lord E, Brealey JC, Le Moullec M, Bieker VC, Ellegaard MR, Rasmussen JA, Kellner FL, Guschanski K, Yannic G, et al. 2024. Ancient reindeer mitogenomes reveal island-hopping colonisation of the Arctic archipelagos. *Sci Rep* 14:4143.

Hopken MW. 2015. Molecular assessment of translocation and management of an endangered subspecies of white-tailed deer (Odocoileus virginianus). *Conserv Genet*:13.

Hornbeck GE, Mahoney JM. 2000. Introgressive Hybridization of Mule Deer and White-Tailed Deer in Southwestern Alberta. :5.

Hu P, Shao Y, Xu J, Wang T, Li Y, Liu H, Rong M, Su W, Chen B, Cui S, et al. 2019. Genome-wide study on genetic diversity and phylogeny of five species in the genus Cervus. *BMC Genomics* 20:384.

Hughes GA, Carr SM. 1993. Reciprocal Hybridization between White-tailed Deer (Odocoileus virginianus) and Mule Deer (O.hemionus) in Western Canada: Evidence from Serum Albumin and mtDNA Sequences. *Can. J. Zool.* 71:524–530.

Hundertmark KJ, Shields GF, Udina IG, Bowyer RT, Danilkin AA, Schwartz CC. 2002. Mitochondrial Phylogeography of Moose (Alces alces): Late Pleistocene Divergence and Population Expansion. *Molecular Phylogenetics and Evolution* 22:375–387.

Iacolina L, Corlatti L, Buzan E, Safner T, Šprem N. 2019. Hybridisation in European Ungulates: an Overview of the Current Status, Causes, and Consequences. *Mam Rev* 49:45–59.

IPCC. 2022. Climate change 2022: Impacts, adaptation and vulnerability. :3–33.

Ito T, Hayakawa T, Suzuki–Hashido N, Hamada Y, Kurihara Y, Hanya G, Kaneko A, Natsume T, Aisu S, Honda T, et al. 2021. Phylogeographic History of Japanese Macaques. *J Biogeogr*:jbi.14087.

IUCN. 2015a. Odocoileus virginianus: Gallina, S. and Lopez Arevalo, H.: The IUCN Red List of Threatened Species 2016: e.T42394A22162580. Available from: http://www.iucnredlist.org/details/42394/0

IUCN. 2015b. Odocoileus hemionus: Sanchez Rojas, G. and Gallina Tessaro, S.: The IUCN Red List of Threatened Species 2016: e.T42393A22162113. Available from: http://www.iucnredlist.org/details/42393/0

Jackson ST, Weng C. 1999. Late Quaternary extinction of a tree species in eastern North America. *Proc. Natl. Acad. Sci. U.S.A.* 96:13847–13852.

Jacobson JA. 2003. Identification of Mule Deer *(Odocoileus Hemionus)* and White-Tailed Deer *(Odocoileus Virginianus)* Postcranial Remains as a Means of Determining Human Subsistence Strategies. *Plains Anthropologist* 48:287–297.

Jensen WF, Bleich VC, Whittaker DG. 2023. Historical Trends in Black-tailed Deer, Mule Deer, and Their Habitats. In: Heffelfinger J, Krausman PR, Cox A, editors. Ecology and management of black-tailed and mule deer of North America. First edition. Boca Raton, FL: CRC Press. p. 25–42.

Jensen-Seaman MI, Furey TS, Payseur BA, Lu Y, Roskin KM, Chen C-F, Thomas MA, Haussler D, Jacob HJ. 2004. Comparative Recombination Rates in the Rat, Mouse, and Human Genomes. *Genome Res.* 14:528–538.

John. 2011. SeqPrep: tool for stripping adaptors and/or merging paired reads with overlap into single reads. Available from: https://githubcom/ jstjohn/SeqPrep.

Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L. 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29:1682–1684.

Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589.

Keller M, Lévy F. 2012. The Main but not the Accessory Olfactory System is Involved in the Processing of Socially Relevant Chemosignals in Ungulates. *Front. Neuroanat.* [Internet] 6. Available from: http://journal.frontiersin.org/article/10.3389/fnana.2012.00039/abstract

Kerr PJ, Tassier-Surine SA, Kilgore SM, Bettis EA, Dorale JA, Cramer BD. 2021. Timing, provenance, and implications of two MIS 3 advances of the Laurentide Ice Sheet into the Upper Mississippi River Basin, USA. *Quaternary Science Reviews* 261:106926.

Kessler C, Brambilla A, Waldvogel D, Camenisch G, Biebach I, Leigh DM, Grossen C, Croll D. 2021. A robust sequencing assay of a thousand amplicons for the high-throughput population monitoring of Alpine ibex immunogenetics. *Mol Ecol Resour*:1755-0998.13452.

Kessler C, Haddrath O, Lim BK, Shafer ABA. In prep. Ancient DNA of the Toronto Subway Deer Adds to the Extinction List of Ice Age Megafauna.

Kessler C, Shafer ABA. 2024. Genomic analyses capture the human-induced demographic collapse and recovery in a wide-ranging cervid.Meyer D, editor. *Molecular Biology and Evolution* 41:msae038.

Kessler C, Wootton E, Shafer ABA. 2023. Speciation without gene-flow in hybridizing deer. *Molecular Ecology* 32:1117–1132.

Kinoshita E, Kosintsev PA, Abramov AV, Solovyev VA, Saveljev AP, Nishita Y, Masuda R. 2020. Holocene Changes in the Distributions of Asian and European Badgers (Carnivora: Mustelidae: Meles) Inferred from Ancient DNA Analysis. *Biological Journal of the Linnean Society* 129:594–602.

Kirch M, Romundset A, Gilbert MTP, Jones FC, Foote AD. 2021. Ancient and modern stickleback genomes reveal the demographic constraints on adaptation. *Current Biology*:S0960982221002682.

Klassmann A, Gautier M. 2022. Detecting selection using extended haplotype homozygosity (EHH)-based statistics in unphased or unpolarized data.Rashid MAR, editor. *PLoS ONE* 17:e0262024.

Klicka LB, Najar N, Vázquez-Miranda H, Zink R. 2023. Relationships among North American deer based on mitochondrial DNA and ultraconserved elements, with comments on mito-nuclear discordance. In Review Available from: https://www.researchsquare.com/article/rs-3006994/v1

Koch PL, Barnosky AD. 2006. Late Quaternary Extinctions: State of the Debate. *Annu. Rev. Ecol. Evol. Syst.* 37:215–250.

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. 2015. Clumpak: a Program for Identifying Clustering Modes and Packaging Population Structure Inferences Across K. *Mol Ecol Resour* 15:1179–1191.

Korneliussen TS, Albrechtsen A, Nielsen R. 2014a. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* 15:356.

Korneliussen TS, Albrechtsen A, Nielsen R. 2014b. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* 15:356.

Kuhn TS, Mcfarlane KA, Groves P, Mooers AØ, Shapiro B. 2010. Modern and ancient DNA reveal recent partial replacement of caribou in the southwest Yukon. *Molecular Ecology* 19:1312–1323.

Kuijper DPJ, Verwijmeren M, Churski M, Zbyryt A, Schmidt K, Jędrzejewska B, Smit C. 2014. What Cues Do Ungulates Use to Assess Predation Risk in Dense Temperate Forests?Sueur C, editor. *PLoS ONE* 9:e84607.

Kurten B. 1975. A New Pleistocene Genus of American Mountain Deer. *Journal of Mammalogy* 56:507–508.

Kuwayama R, Ozawa T. 2000. Phylogenetic Relationships among European Red Deer, Wapiti, and Sika Deer Inferred from Mitochondrial DNA Sequences. *Molecular Phylogenetics and Evolution* 15:115–123.

Kvalnes T, Saether B-E, Haanes H, Røed KH, Engen S, Solberg EJ. 2016. Harvest-induced phenotypic selection in an island population of moose, Alces alces. *Evolution* 70:1486–1500.

Lamb S, Taylor AM, Hughes TA, McMillan BR, Larsen RT, Khan R, Weisz D, Dudchenko O, Aiden EL, Edelman NB, et al. 2021. De novo chromosome-length assembly of the mule deer (Odocoileus hemionus) genome. *Gigabyte* 2021:1–13.

Lapierre M, Lambert A, Achaz G. 2017. Accuracy of Demographic Inferences from the Site Frequency Spectrum: The Case of the Yoruba Population. *Genetics* 206:439–449.

Latch EK, Heffelfinger JR, Fike JA, RHODES Jr OE. 2009. Species-wide Phylogeography of North American Mule Deer (Odocoileus hemionus): Cryptic Glacial Refugia and Postglacial Recolonization. *Molecular Ecology* 18:1730–1745.

Latch EK, Reding DM, Heffelfinger JR, Alcalá-Galván CH, Rhodes OE. 2014. Range-wide Analysis of Genetic Structure in a Widespread, Highly Mobile Species (Odocoileus hemionus) Reveals the Importance of Historical Biogeography. *Mol Ecol* 23:3171–3190.

Lavretsky P, DaCosta JM, Sorenson MD, McCracken KG, Peters JL. 2019. ddRAD-seq Data Reveal Significant Genome-Wide Population Structure and divergent Genomic Regions that Distinguish the Mallard and Close Relatives in North America. *Mol Ecol* 28:2594–2609.

Legendre P, Legendre L. 2012. Numerical ecology. Third English edition. Amsterdam: Elsevier

Leigh JW, Bryant D. 2015. popart: full-feature software for haplotype network construction. *Methods in Ecology and Evolution* 6:1110–1116.

Lemke AK. 2015. Great Lakes Rangifer and Paleoindians: Archaeological and Paleontological Caribou Remains from Michigan. *PaleoAmerica* 1:276–283.

Li H. 2018. Seqtk. Available from: https://github.com/lh3/seqtk

Li H, Durbin R. 2009. Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform. *Bioinformatics* 25:1754–1760.

Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* 475:493–496.

Li H., Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.

Li Heng, 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* 25:2078–2079.

Lindo J, Huerta-Sánchez E, Nakagome S, Rasmussen M, Petzelt B, Mitchell J, Cybulski JS, Willerslev E, DeGiorgio M, Malhi RS. 2016. A time transect of exomes from a Native American population before and after European contact. *Nat Commun* 7:13175.

Lingle S. 1992. Escape Gaits of White-Tailed Deer, Mule Deer and Their Hybrids: Gaits Observed and Patterns of Limb Coordination. *Behaviour* 122:153–181.

Lingle S. 2001. Anti-Predator Strategies and Grouping Patterns in White-Tailed Deer and Mule Deer. *Ethology* 107:295–314.

Lingle S. 2003. Group Composition and Cohesion in Sympatric White-tailed Deer and Mule deer. *Can. J. Zool.* 81:1119–1130.

Liu Q, Mishra M, Saxena AS, Wu H, Qiu Y, Zhang X, You X, Ding S, Miyamoto MM. 2021. Balancing Selection Maintains Ancient Polymorphisms at Conserved Enhancers for the Olfactory Receptor Genes of a Chinese Marine Fish. *Mol Ecol* 30:4023–4038.

Liu S, Westbury MV, Dussex N, Mitchell KJ, Sinding M-HS, Heintzman PD, Duchêne DA, Kapp JD, von Seth J, Heiniger H, et al. 2021. Ancient and modern genomes unravel the evolutionary history of the rhinoceros family. *Cell* 184:4874-4885.e16.

Liu X, Fu Y-X. 2015. Exploring population size changes using SNP frequency spectra. *Nat Genet* 47:555–559.

Liu X, Fu Y-X. 2020. Stairway Plot 2: demographic history inference with folded SNP frequency spectra. *Genome Biol* 21:280.

Long CA, Yahnke CJ. 2011. End of the Pleistocene: elk-moose (Cervalces) and caribou (Rangifer) in Wisconsin. *Journal of Mammalogy* 92:1127–1135.

Loog L, Thalmann O, Sinding MS, Schuenemann VJ, Perri A, Germonpré M, Bocherens H, Witt KE, Samaniego Castruita JA, Velasco MS, et al. 2020. Ancient DNA suggests modern wolves trace their origin to a Late Pleistocene expansion from Beringia. *Mol Ecol* 29:1596–1610.

Lorenzen ED, Nogués-Bravo D, Orlando L, Weinstock J, Binladen J, Marske KA, Ugan A, Borregaard MK, Gilbert MTP, Nielsen R, et al. 2011. Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature* 479:359–364.

Lorenzini R, Garofalo L. 2015. Insights into the evolutionary history of *Cervus* (Cervidae, tribe Cervini) based on Bayesian analysis of mitochondrial marker sequences, with first indications for a new species. *J Zoolog Syst Evol Res* 53:340–349.

Lynch VJ, Bedoya-Reina OC, Ratan A, Sulak M, Drautz-Moses DI, Perry GH, Miller W, Schuster SC. 2015. Elephantid Genomes Reveal the Molecular Bases of Woolly Mammoth Adaptations to the Arctic. *Cell Reports* 12:217–228.

Mackiewicz P, Matosiuk M, Świsłocka M, Zachos FE, Hajji GM, Saveljev AP, Seryodkin IV, Farahvash T, Rezaei HR, Torshizi RV, et al. 2022a. Phylogeny and evolution of the genus Cervus (Cervidae, Mammalia) as revealed by complete mitochondrial genomes. *Sci Rep* 12:16381.

Mackiewicz P, Matosiuk M, Świsłocka M, Zachos FE, Hajji GM, Saveljev AP, Seryodkin IV, Farahvash T, Rezaei HR, Torshizi RV, et al. 2022b. Phylogeny and evolution of the genus Cervus (Cervidae, Mammalia) as revealed by complete mitochondrial genomes. *Sci Rep* 12:16381.

Malhi Y, Doughty CE, Galetti M, Smith FA, Svenning J-C, Terborgh JW. 2016. Megafauna and ecosystem function from the Pleistocene to the Anthropocene. *Proc. Natl. Acad. Sci. U.S.A.* 113:838–846.

Malhi Y, Lander T, Le Roux E, Stevens N, Macias-Fauria M, Wedding L, Girardin C, Kristensen JÅ, Sandom CJ, Evans TD, et al. 2022. The role of large wild animals in climate change mitigation and adaptation. *Current Biology* 32:R181–R196.

Mallet J. 2005. Hybridization as an Invasion of the Genome. *Trends in Ecology & Evolution* 20:229–237.

Manel S, Perrier C, Pratlong M, Abi-Rached L, Paganini J, Pontarotti P, Aurelle D. 2016. Genomic resources and their influence on the detection of the signal of positive selection in genome scans. *Molecular Ecology* 25:170–184.

Marean CW, Cowling RM, Franklin J. 2020. The Palaeo-Agulhas Plain: Temporal and spatial variation in an extraordinary extinct ecosystem of the Pleistocene of the Cape Floristic Region. *Quaternary Science Reviews* 235:106161.

Marques DA, Lucek K, Sousa VC, Excoffier L, Seehausen O. 2019. Admixture Between Old Lineages Facilitated Contemporary Ecological Speciation in Lake Constance Stickleback. *Nat Commun* 10:4240.

Martchenko D, Shafer ABA. 2023. Contrasting whole-genome and reduced representation sequencing for population demographic and adaptive inference: an alpine mammal case study. *Heredity* 131:273–281.

Martin M, Patterson M, Garg S, O Fischer S, Pisanti N, Klau GW, Schöenhuth A, Marschall T. 2016. WhatsHap: Fast and Accurate Read-based Phasing. Bioinformatics Available from: http://biorxiv.org/lookup/doi/10.1101/085050

Martin PS, Steadman DW. 1999. Prehistoric Extinctions on Islands and Continents. In: MacPhee RDE, editor. Extinctions in Near Time. Boston, MA: Springer US. p. 17–55. Available from: http://link.springer.com/10.1007/978-1-4757-5202-1_2

Martin S. 2021. genomics_general. *GitHub* [Internet]. Available from: https://github.com/simonhmartin/genomics_general

Martin S, Davey JW, Jiggins CD. 2015. Evaluating the Use of ABBA–BABA Statistics to Locate Introgressed Loci. *Molecular Biology and Evolution* 32:244–257.

Mautz W, Pekins P, Warren J. 1985. Cold temperature effects on metabolic rate of white-tailed, mule, and black-tailed deer in winter coat. *Biology of deer production. Bulletin* 22:453–457.

Mayr E. 1942. Systematics and the Origin of Species.

McAndrews JH, Jackson LJ. 1988. Age and environment of Late Pleistocene mastodont and mammoth in southern Ontario. v. 33:161–172. Available from: https://ehrafarchaeology.yale.edu/document?id=w130-050

McDonald HG. 1989. New Records of the Elk-moose Cervalces scotti from Ohio. *American Midland Naturalist* 122:349.

McDonald JS, Miller KV, Association QDM, Resources DBWS of F. 2004. A History of White-tailed Deer Restocking in the United States, 1878 to 2004. Quality Deer Management Assoc.

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

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

Meiri M, Lister A, Kosintsev P, Zazula G, Barnes I. 2020. Population dynamics and range shifts of moose (Alces alces) during the Late Quaternary. *J Biogeogr* 47:2223–2234.

Meiri M, Lister AM, Collins MJ, Tuross N, Goebel T, Blockley S, Zazula GD, Van Doorn N, Dale Guthrie R, Boeskorov GG, et al. 2014. Faunal record identifies Bering isthmus conditions as constraint to end-Pleistocene migration to the New World. *Proc. R. Soc. B.* 281:20132167.

Meisner J, Albrechtsen A. 2018. Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data. *Genetics* 210:719–731.

Meltzer DJ. 2020. Overkill, glacial history, and the extinction of North America's Ice Age megafauna. *Proc. Natl. Acad. Sci. U.S.A.* 117:28555–28563.

Meyer M, Kircher M. 2010. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harb Protoc* 2010:pdb.prot5448.

Momigliano P, Jokinen H, Fraimout A, Florin A-B, Norkko A, Merilä J. 2017. Extraordinarily Rapid Speciation in a Marine Fish. :6.

Morejohn GV, Dailey DC, Sierra College Natural History Museum. 2004. The identity and postcranial osteology of Odocoileus lucasi (Hay) 1927 : a plio-Pleistocene deer from California and Idaho. Rocklin, Calif.: Sierra College Natural History Museum Rocklin, Calif.

Morina DL, Demarais S, Strickland BK, Larson JE. 2018. While Males Fight, Females Choose: Male Phenotypic Quality Informs Female Mate Choice in Mammals. *Animal Behaviour* 138:69–74.

Moscarella RA, Aguilera M, Escalante AA. 2003. Phylogeography, Population Structure, and Implications for Conservation of White-tailed Deer (Odocoileus Virginianus) in Venezuela. *Journal of Mammalogy* 84:1300–1315.

Mottl O, Flantua SGA, Bhatta KP, Felde VA, Giesecke T, Goring S, Grimm EC, Haberle S, Hooghiemstra H, Ivory S, et al. 2021. Global acceleration in rates of vegetation change over the past 18,000 years. *Science* 372:860–864.

NatureServe. 2021a. NatureServe Explorer. *NatureServe Mule Deer* [Internet]. Available from: https://explorer.natureserve.org/Taxon/ELEMENT_GLOBAL.2.101365/Odocoileus_hemionus

NatureServe. 2021b. NatureServe Explorer [web application] WTD. *NatureServe, Arlington, Virginia* [Internet]. Available from: https://explorer.natureserve.org/Taxon/ELEMENT_GLOBAL.2.104777/Odocoileus_virginianus

Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* 32:268–274.

Nielsen R, Akey JM, Jakobsson M, Pritchard JK, Tishkoff S, Willerslev E. 2017. Tracing the peopling of the world through genomics. *Nature* 541:302–310.

Nosil P. 2012. Ecological Speciation. Oxford ; New York: Oxford University Press

Nursyifa C, Brüniche-Olsen A, Garcia-Erill G, Heller R, Albrechtsen A. 2022. Joint identification of sex and sex-linked scaffolds in non-model organisms using low depth sequencing data. *Molecular Ecology Resources* 22:458–467.

O'Connell J, Gurdasani D, Delaneau O, Pirastu N, Ulivi S, Cocca M, Traglia M, Huang J, Huffman JE, Rudan I, et al. 2014. A General Approach for Haplotype Phasing across the Full Spectrum of Relatedness.Gibson G, editor. *PLoS Genet* 10:e1004234.

O'Gara B, Dundas R. 2002. Distribution: past and present. In: Toweill DE, Thomas JW, Metz DP, editors. North American elk: ecology and management. 1st ed. Washington, D.C: Smithsonian Institution Press.

Oksanen J, Simpson GL, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Solymos P, Stevens MHH, Szoecs E, et al. 2022. vegan: Community Ecology Package. Available from: https://CRAN.R-project.org/package=vegan

Orlando L, Allaby R, Skoglund P, Der Sarkissian C, Stockhammer PW, Ávila-Arcos MC, Fu Q, Krause J, Willerslev E, Stone AC, et al. 2021. Ancient DNA analysis. *Nat Rev Methods Primers* 1:14.

Paijmans JLA, Barlow A, Becker MS, Cahill JA, Fickel J, Förster DWG, Gries K, Hartmann S, Havmøller RW, Henneberger K, et al. 2021. African and Asian leopards are highly differentiated at the genomic level. *Current Biology*:S0960982221004577.

Palkopoulou E, Lipson M, Mallick S, Nielsen S, Rohland N, Baleka S, Karpinski E, Ivancevic AM, To T-H, Kortschak RD, et al. 2018. A comprehensive genomic history of extinct and living elephants. *Proc Natl Acad Sci USA* 115:E2566–E2574.

Palkopoulou E, Mallick S, Skoglund P, Enk J, Rohland N, Li H, Omrak A, Vartanyan S, Poinar H, Götherström A, et al. 2015. Complete Genomes Reveal Signatures of Demographic and Genetic Declines in the Woolly Mammoth. *Current Biology* 25:1395–1400.

Palstra FP, Ruzzante DE. 2008. Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild population persistence? *Molecular Ecology* 17:3428–3447.

Pamilo P, Savolainen O. 2004. Post-Glacial Colonization, Drift, Local Selection and Conservation Value of Populations: A Northern Perspective. *Hereditas* 130:229–238.

Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35:526–528.

Patterson M, Marschall T, Pisanti N, Van Iersel L, Stougie L, Klau GW, Schönhuth A. 2015. WhatsHap : Weighted Haplotype Assembly for Future-Generation Sequencing Reads. *Journal of Computational Biology* 22:498–509.

Peart CR, Tusso S, Pophaly SD, Botero-Castro F, Wu C-C, Aurioles-Gamboa D, Baird AB, Bickham JW, Forcada J, Galimberti F, et al. 2020. Determinants of genetic variation across eco-evolutionary scales in pinnipeds. *Nat Ecol Evol* 4:1095–1104.

Pedersen BS, Quinlan AR. 2018. Mosdepth: Quick Coverage Calculation for Genomes and Exomes.Hancock J, editor. *Bioinformatics* 34:867–868.

Peres TM, Altman H. 2018. The Magic of Improbable Appendages: Deer Antler Objects in the Archaeological Record of the American South. *Journal of Archaeological Science: Reports* 20:888–895.

Pérez-González J, Carranza J, Anaya G, Broggini C, Vedel G, De La Peña E, Membrillo A. 2023. Comparative Analysis of Microsatellite and SNP Markers for Genetic Management of Red Deer. *Animals* 13:3374.

Peterson RL. 1965. A Well-preserved Grizzly Bear Skull recovered from a Late Glacial Deposit near Lake Simcoe, Ontario. *Nature* 208:1233–1234.

Pickrell JK, Pritchard JK. 2012. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data.Tang H, editor. *PLoS Genet* 8:e1002967.

Pierini F, Lenz TL. 2018. Divergent Allele Advantage at Human MHC Genes: Signatures of Past and Ongoing Selection.Wilke C, editor. *Molecular Biology and Evolution* 35:2145–2158.

Piertney SB, Oliver MK. 2006. The Evolutionary Ecology of the Major Histocompatibility Complex. *Heredity* 96:7–21.

Pitra C, Fickel J, Meijaard E, Groves C. 2004. Evolution and Phylogeny of Old World Deer. *Molecular Phylogenetics and Evolution* 33:880–895.

Plaxton L, Hempel E, Marsh WA, Portela Miguez R, Waurick I, Kitchener AC, Hofreiter M, Lister AM, Zachos FE, Brace S. 2023. Assessing the identity of rare historical museum specimens of the extinct blue antelope (Hippotragus leucophaeus) using an ancient DNA approach. *Mamm Biol* [Internet]. Available from: https://link.springer.com/10.1007/s42991-023-00373-4

Poelstra JW, Vijay N, Bossu CM, Lantz H, Ryll B, Muller I, Baglione V, Unneberg P, Wikelski M, Grabherr MG, et al. 2014. The Genomic Landscape Underlying Phenotypic Integrity in the Face of Gene Flow in Crows. *Science* 344:1410–1414.

Portik DM, Leaché AD, Rivera D, Barej MF, Burger M, Hirschfeld M, Rödel M, Blackburn DC, Fujita MK. 2017. Evaluating mechanisms of diversification in a Guineo-Congolian tropical forest frog using demographic model selection. *Mol Ecol* 26:5245–5263.

Priadka P, Moses B, Kozmik C, Kell S, Popp JN. 2024. Weaving Indigenous and Western knowledge systems to discern drivers of mooz (moose) population decline. *People and Nature* [Internet] n/a. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/pan3.10706

Pundir S, Martin MJ, O'Donovan C, The UniProt Consortium. 2016. UniProt Tools. *Current Protocols in Bioinformatics* [Internet] 53. Available from: https://onlinelibrary.wiley.com/doi/10.1002/0471250953.bi0129s53

Quinlan AR, Hall IM. 2010. BEDTools: a Flexible Suite of Utilities for Comparing Genomic Features. *Bioinformatics* 26:841–842.

R Core Team. 2021. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing Available from: https://www.R-project.org/

Ragavan P, Zhou R, Ng WL, Rana TS, Mageswaran T, Mohan PM, Saxena A. 2017. Natural Hybridization in Mangroves – an Overview. *Botanical Journal of the Linnean Society* 185:208–224.

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7.Susko E, editor. *Systematic Biology* 67:901–904.

Ramsey CB. 2009. Bayesian Analysis of Radiocarbon Dates. *Radiocarbon* 51:337–360.

Ranslow AN, Richter JP, Neuberger T, Van Valkenburgh B, Rumple CR, Quigley AP, Pang B, Krane MH, Craven BA. 2014. Reconstruction and Morphometric Analysis of the Nasal Airway of the White-Tailed Deer (Odocoileus virginianus) and Implications Regarding Respiratory and Olfactory Airflow: White-Tailed Deer Nasal Anatomy and Function. *Anat. Rec.* 297:2138–2147.

Ravinet M, Faria R, Butlin RK, Galindo J, Bierne N, Rafajlović M, Noor MAF, Mehlig B, Westram AM. 2017. Interpreting the Genomic Landscape of Speciation: a Road Map for Finding Barriers to Gene Flow. *J. Evol. Biol.* 30:1450–1477.

Raxworthy CJ, Smith BT. 2021. Mining museums for historical DNA: advances and challenges in museomics. *Trends in Ecology & Evolution* 36:1049–1060.

Reimer PJ. 2020. Composition and consequences of the IntCal20 radiocarbon calibration curve. *Quaternary Research* 96:22–27.

Robin M, Ferrari G, Akgül G, Münger X, Von Seth J, Schuenemann VJ, Dalén L, Grossen C. 2022. Ancient mitochondrial and modern whole genomes unravel massive genetic diversity loss during near extinction of Alpine ibex. *Molecular Ecology* 31:3548–3565.

Rogers RL, Slatkin M. 2017. Excess of genomic defects in a woolly mammoth on Wrangel island.Barsh GS, editor. *PLoS Genet* 13:e1006601.

Rosenblatt E, Gieder K, Donovan T, Murdoch J, Smith TPL, Heaton MP, Kalbfleisch TS, Murdoch BM, Bhattarai S, Pacht E, et al. 2023. Genetic diversity and connectivity of moose (Alces americanus americanus) in eastern North America. *Conserv Genet* 24:235–248.

Roycroft E, MacDonald AJ, Moritz C, Moussalli A, Portela Miguez R, Rowe KC. 2021. Museum genomics reveals the rapid decline and extinction of Australian rodents since European settlement. *Proc Natl Acad Sci USA* 118:e2021390118.

Rundle HD, Nosil P. 2005. Ecological Speciation. *Ecology Letters* 8:336–352.

Russell T, Cullingham C, Ball M, Pybus M, Coltman D. 2021a. Extent and Direction of Introgressive Hybridization of Mule and White-tailed Deer in Western Canada. *Evol Appl*:eva.13250.

Russell T, Cullingham C, Ball M, Pybus M, Coltman D. 2021b. Extent and direction of introgressive hybridization of mule and white-tailed deer in western Canada. *Evol Appl* 14:1914–1925.

Russell T, Cullingham C, Kommadath A, Stothard P, Herbst A, Coltman D. 2019. Development of a Novel Mule Deer Genomic Assembly and Species-Diagnostic SNP Panel for Assessing Introgression in Mule Deer, White-Tailed Deer, and Their Interspecific Hybrids. *G3*:g3.200838.2018.

Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, et al. 2002. Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419:832–837.

Santiago E, Novo I, Pardiñas AF, Saura M, Wang J, Caballero A. 2020. Recent Demographic History Inferred by High-Resolution Analysis of Linkage Disequilibrium. *Molecular Biology and Evolution* 37:3642–3653.

Schaefer JA. 2003. Long-Term Range Recession and the Persistence of Caribou in the Taiga. *Conservation Biology* 17:1435–1439.

Schiffels S, Durbin R. 2014. Inferring Human Population Size and Separation History from Multiple Genome Sequences. *Nat Genet* 46:919–925.

Schiffels S, Wang K. 2020. MSMC and MSMC2: The Multiple Sequentially Markovian Coalescent. In: Dutheil JY, editor. Statistical Population Genomics. Vol. 2090. Methods in Molecular Biology. New York, NY: Springer US. p. 147–166.

Schmelzer I. 2020. Boreal caribou survival in a warming climate, Labrador, Canada 1996-2014. *Global Ecology and Conservation*:25.

Schorger AW. 1955. The passenger pigeon, its natural history and extinction. University of Wisconsin Press Available from: https://books.google.ch/books?id=6TXxAAAAMAAJ

Schubert M, Ginolhac A, Lindgreen S, Thompson JF, AL-Rasheid KA, Willerslev E, Krogh A, Orlando L. 2012. Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics* 13:178.

Schubert M, Lindgreen S, Orlando L. 2016. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res Notes* 9:88.

Schumer M, Cui R, Rosenthal GG, Andolfatto P. 2015. Reproductive Isolation of Hybrid Populations Driven by Genetic Incompatibilities.Payseur BA, editor. *PLoS Genet* 11:e1005041.

Seabury CM, Oldeschulte DL, Bhattarai EK, Legare D, Ferro PJ, Metz RP, Johnson CD, Lockwood MA, Nichols TA. 2020. Accurate Genomic Predictions for Chronic Wasting Disease in U.S. White-Tailed Deer. *G3 Genes|Genomes|Genetics* 10:1433–1441.

Servedio MR. 2004. The Evolution of Premating Isolation: Local Adaptation and Natural and Sexual Selection Against Hybrids. *Evolution* 58:913–924.

Shafer ABA, Cullingham CI, Côté SD, Coltman DW. 2010. Of Glaciers and Refugia: a Decade of Study Sheds New Light on the Phylogeography of Northwestern North America. *Molecular Ecology* 19:4589–4621.

Shafer ABA, Gattepaille LM, Stewart REA, Wolf JBW. 2015. Demographic inferences using short-read genomic data in an approximate Bayesian computation framework: in silico evaluation of power, biases and proof of concept in Atlantic walrus. *Mol Ecol* 24:328–345.

Shafer ABA, Wolf JBW. 2013. Widespread Evidence for Incipient Ecological Speciation: a Meta-analysis of Isolation-by-ecology.Wiens J, editor. *Ecol Lett* 16:940–950.

Shang H, Rendón-Anaya M, Paun O, Field DL, Hess J, Vogl C, Liu J, Ingvarsson PK, Lexer C, Leroy T. 2021. Conserved Genomic Landscapes of Differentiation Across Populus Speciation Continuum. Evolutionary Biology Available from: http://biorxiv.org/lookup/doi/10.1101/2021.08.26.457771

Sharko FS, Boulygina ES, Tsygankova SV, Slobodova NV, Alekseev DA, Krasivskaya AA, Rastorguev SM, Tikhonov AN, Nedoluzhko AV. 2021. Steller's sea cow genome suggests

this species began going extinct before the arrival of Paleolithic humans. *Nat Commun* 12:2215.

Shuman B. 2002. The anatomy of a climatic oscillation: vegetation change in eastern North America during the Younger Dryas chronozone. *Quaternary Science Reviews* 21:1777–1791.

Shuman B, Marsicek J. 2016. The structure of Holocene climate change in mid-latitude North America. *Quaternary Science Reviews* 141:38–51.

Skoglund P, Northoff BH, Shunkov MV, Derevianko AP, Pääbo S, Krause J, Jakobsson M. 2014. Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc. Natl. Acad. Sci. U.S.A.* 111:2229–2234.

Skotte L, Korneliussen TS, Albrechtsen A. 2013. Estimating Individual Admixture Proportions from Next Generation Sequencing Data. *Genetics* 195:693–702.

Slater G, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6:31.

Smith BT, Gehara M, Harvey MG. 2021. The demography of extinction in eastern North American birds. *Proc. R. Soc. B.* 288:20201945.

Smith FA, Elliott Smith EA, Villaseñor A, Tomé CP, Lyons SK, Newsome SD. 2022. Late Pleistocene megafauna extinction leads to missing pieces of ecological space in a North American mammal community. *Proc. Natl. Acad. Sci. U.S.A.* 119:e2115015119.

Smith-Jones C. 2022. A guide to the deer of the world. Stackpole Books Available from: https://books.google.ch/books?id=RtrczgEACAAJ

Solmundson K, Bowman J, Manseau M, Taylor RS, Keobouasone S, Wilson PJ. 2023. Genomic population structure and inbreeding history of Lake Superior caribou. *Ecology and Evolution* 13:e10278.

Soltis DE, Morris AB, McLACHLAN JS, Manos PS, Soltis PS. 2006. Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology* 15:4261–4293.

Soraggi S, Wiuf C, Albrechtsen A. 2018. Powerful Inference with the D-statistic on Low-coverage Whole-genome Data. *G3: Genes, Genomes, Genetics* 8.

Speller CF, Kooyman B, Rodrigues AT, Langemann EG, Jobin RM, Yang DY. 2014. Assessing prehistoric genetic structure and diversity of North American elk (Cervus elaphus) populations in Alberta, Canada. *Can. J. Zool.* 92:285–298.

Spiess AE, Curran ML, Grimes JR. 1985. Caribou (Rangifer Tarandus L.) Bones from New England Paleoindian Sites. *North American Archaeologist* 6:145–159.

Stange M, Sánchez-Villagra MR, Salzburger W, Matschiner M. 2018. Bayesian Divergence-Time Estimation with Genome-Wide Single-Nucleotide Polymorphism Data of Sea Catfishes (Ariidae) Supports Miocene Closure of the Panamanian Isthmus.Hahn M, editor. *Systematic Biology* 67:681–699.

Stankowski S, Ravinet M. 2021. Defining the Speciation Continuum. *Evolution* 75:1256–1273.

Stanton DWG, Alberti F, Plotnikov V, Androsov S, Grigoriev S, Fedorov S, Kosintsev P, Nagel D, Vartanyan S, Barnes I, et al. 2020. Early Pleistocene origin and extensive intra-species diversity of the extinct cave lion. *Sci Rep* 10:12621.

Stecher G, Tamura K, Kumar S. 2020. Molecular Evolutionary Genetics Analysis (MEGA) for macOS.Russo C, editor. *Molecular Biology and Evolution* 37:1237–1239.

Stevens N, Bond W, Feurdean A, Lehmann CER. 2022. Grassy Ecosystems in the Anthropocene. *Annu. Rev. Environ. Resour.* 47:261–289.

Stewart M, Carleton WC, Groucutt HS. 2021. Climate change, not human population growth, correlates with Late Quaternary megafauna declines in North America. *Nat Commun* 12:965.

Stuart AJ. 2015. Late Quaternary megafaunal extinctions on the continents: a short review. *Geol. J.* 50:338–363.

Stubblefield SS, Warren RJ, Murphy BR. 1986. Hybridization of Free-Ranging White-Tailed and Mule Deer in Texas. *The Journal of Wildlife Management* 50:688.

Szpiech ZA. 2021. selscan 2.0: scanning for sweeps in unphased data. Evolutionary Biology Available from: http://biorxiv.org/lookup/doi/10.1101/2021.10.22.465497

Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. 2015a. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* 31:2032–2034.

Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. 2015b. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* 31:2032–2034.

Tavares GM, Reales G, Bortolini MC, Fagundes NJR. 2019. Measuring the impact of European colonization on Native American populations in Southern Brazil and Uruguay: Evidence from mtDNA. *Am J Hum Biol* 31:e23243.

Taylor RS, Manseau M, Keobouasone S, Liu P, Mastromonaco G, Solmundson K, Kelly A, Larter NC, Gamberg M, Schwantje H, et al. 2024. High genetic load without purging in caribou, a diverse species at risk. *Current Biology*:S0960982224001441.

Taylor RS, Manseau M, Klütsch CFC, Polfus JL, Steedman A, Hervieux D, Kelly A, Larter NC, Gamberg M, Schwantje H, et al. 2021. Population dynamics of caribou shaped by glacial cycles before the last glacial maximum. *Mol Ecol*:mec.16166.

Taylor SA, Larson EL. 2019. Insights from Genomes into the Evolutionary Importance and Prevalence of Hybridization in Nature. *Nat Ecol Evol* 3:170–177.

Thompson AJ, Zhu J, Poulsen CJ, Tierney JE, Skinner CB. 2022. Northern Hemisphere vegetation change drives a Holocene thermal maximum. *Sci. Adv.* 8:eabj6535.

Timmermann HR, Rodgers AR. 2015. The Status and Management of Moose in North America – Circa 2015. *ALCES VOL.* 53:22.

UniProt. 2022a. Q9Y5E5 · PCDB4_HUMAN. *UniProt beta* [Internet]. Available from: https://beta.uniprot.org/uniprotkb/Q9Y5E5/entry

UniProt. 2022b. Q15057 · ACAP2_HUMAN. *UniProt beta* [Internet]. Available from: https://beta.uniprot.org/uniprotkb/Q15057/entry

U.S. Fish & Wildlife Service. 1967. Endangered Species List. Available from: https://www.govinfo.gov/content/pkg/FR-1967-03-11/pdf/FR-1967-03-11.pdf#page=1

van der Valk T, Dehasque M, Chacón-Duque JC, Oskolkov N, Vartanyan S, Heintzman PD, Pečnerová P, Díez-del-Molino D, Dalén L. 2022. Evolutionary consequences of genomic deletions and insertions in the woolly mammoth genome. *iScience* 25:104826.

van der Valk T, Pečnerová P, Díez-del-Molino D, Bergström A, Oppenheimer J, Hartmann S, Xenikoudakis G, Thomas JA, Dehasque M, Sağlıcan E, et al. 2021. Million-year-old DNA sheds light on the genomic history of mammoths. *Nature* 591:265–269.

Van De Walle J, Pigeon G, Zedrosser A, Swenson JE, Pelletier F. 2018. Hunting regulation favors slow life histories in a large carnivore. *Nat Commun* 9:1100.

Villanova VL, Hughes PT, Hoffman EA. 2017. Combining genetic structure and demographic analyses to estimate persistence in endangered Key deer (Odocoileus virginianus clavium). *Conserv Genet* 18:1061–1076.

Voight BF, Kudaravalli S, Wen X, Pritchard JK. 2006. A Map of Recent Positive Selection in the Human Genome.Hurst L, editor. *PLoS Biol* 4:e72.

Wang K, Mathieson I, O'Connell J, Schiffels S. 2020. Tracking Human Population Structure through Time from Whole Genome Sequences.Schierup MH, editor. *PLoS Genet* 16:e1008552.

Waples RS, Luikart G, Faulkner JR, Tallmon DA. 2013. Simple life-history traits explain key effective population size ratios across diverse taxa. *Proc. R. Soc. B.* 280:20131339.

Wattles D, DeStefano S. 2011. Status and management of moose in the northeastern United States. *Alces* 47:53–68.

Webb DS. 2000. Evolutionary History of New World Cervidae. In: Vrba ES, Schaller GB, Yale Institute for Biospheric Studies, Wildlife Conservation Society (New York, N.Y.), editors. Antelopes, deer, and relatives: fossil record, behavioral ecology, systematics, and conservation. New Haven: Yale University Press.

Webb GK. 2018. Searching the internet to estimate deer population trends in the U.S., California, and Connecticut. *IIS* 19:163–173.

Weckworth BV, Musiani M, McDEVITT AD, Hebblewhite M, Mariani S. 2012. Reconstruction of caribou evolutionary history in Western North America and its implications for conservation: CARIBOU GENETICS IN WESTERN NORTH AMERICA. *Molecular Ecology* 21:3610–3624.

Weitzel EM. 2021. Investigating overhunting of white-tailed deer (Odocoileus virginianus) in the late Holocene Middle Tennessee River Valley. *Southeastern Archaeology* 40:1–19.

Whittaker DG, Lindzey FG. 2004. Habitat Use Patterns of Sympatric Deer Species on Rocky Mountain Arsenal, Colorado. *Wildlife Society Bulletin* 32:1114–1123.

Wikenros C, Kuijper DPJ, Behnke R, Schmidt K. 2015. Behavioural Responses of Ungulates to Indirect Cues of an Ambush Predator. *Behav* 152:1019–1040.

Wilder AP, Supple MA, Subramanian A, Mudide A, Swofford R, Serres-Armero A, Steiner C, Koepfli K-P, Genereux DP, Karlsson EK, et al. 2023. The contribution of historical processes to contemporary extinction risk in placental mammals. *Science* 380:eabn5856.

Willerslev E, Meltzer DJ. 2021. Peopling of the Americas as inferred from ancient genomics. *Nature* 594:356–364.

Wishart WD. 1980. Hybrids of White-Tailed and Mule Deer in Alberta. *Journal of Mammalogy* 61:716–719.

Wishart WD, Hrudka F, Schmutz SM, Flood PF. 1988. Observations on Spermatogenesis, Sperm Phenotype, and Fertility in White-tailed × Mule Deer Hybrids and a Yak × Cow Hybrid. *Can. J. Zool.* 66:1664–1671.

Woinarski JCZ, Burbidge AA, Harrison PL. 2015. Ongoing unraveling of a continental fauna: Decline and extinction of Australian mammals since European settlement. *Proc. Natl. Acad. Sci. U.S.A.* 112:4531–4540.

Wolverton S, Nagaoka L, Densmore J, Fullerton B. 2008. White-tailed deer harvest pressure & within-bone nutrient exploitation during the mid- to late Holocene in southeast Texas. *Before Farming* 2008:1–23.

Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. *Genome Biol* 20:257.

Wootton E, Robert C, Taillon J, Côté SD, Shafer ABA. 2023. Genomic health is dependent on long-term population demographic history. *Molecular Ecology* 32:1943–1954.

Wright EA, Roberts EK, Platt RN, Bayouth JV, Conway WC, Bradley RD. 2022. Mitochondrial capture and subsequent genetic divergence generates a novel haplogroup: evidence from ancient and ongoing hybridization in mule and white-tailed deer. *Journal of Mammalogy* 103:723–736.

Yansa CH, Adams KM. 2012. Mastodons and Mammoths in the Great Lakes Region, USA and Canada: New Insights into their Diets as they Neared Extinction. *Geography Compass* 6:175–188.

Yu Z, Wright HE. 2001. Response of interior North America to abrupt climate oscillations in the North Atlantic region during the last deglaciation. *Earth-Science Reviews* 52:333–369.

Zagata MD, Haugen AO. 1974. Influence of Light and Weather on Observability of Iowa Deer. *The Journal of Wildlife Management* 38:220.

Zamorano LS, Gompert Z, Fronhofer EA, Feder JL, Nosil P. 2023. A stabilizing eco-evolutionary feedback loop in the wild. *Current Biology* 33:3272-3278.e3.

Zhang P, Huang K, Zhang B, Dunn DW, Chen D, Li F, Qi X, Guo S, Li B. 2018. High Polymorphism in MHC-DRB Genes in Golden Snub-nosed Monkeys Reveals Balancing Selection in Small, Isolated Populations. *BMC Evol Biol* 18:29.

Zhang W, Wu H, Cheng J, Geng J, Li Q, Sun Y, Yu Y, Lu H, Guo Z. 2022. Holocene seasonal temperature evolution and spatial variability over the Northern Hemisphere landmass. *Nat Commun* 13:5334.

Zhou J, Teo Y-Y. 2016. Estimating time to the most recent common ancestor (TMRCA): comparison and application of eight methods. *Eur J Hum Genet* 24:1195–1201.

Zurano JP, Magalhães FM, Asato AE, Silva G, Bidau CJ, Mesquita DO, Costa GC. 2019. Cetartiodactyla: Updating a time-calibrated molecular phylogeny. *Molecular Phylogenetics and Evolution* 133:256–262.

Appendix I: Supplementary material for Chapter 2

Figures

Figure S2.1: Evanno plot for K = 2 through to K = 6 (top). Estimates of admixture proportions from NGSAdmix for different values of K, from K = 2 through K = 7 (bottom).

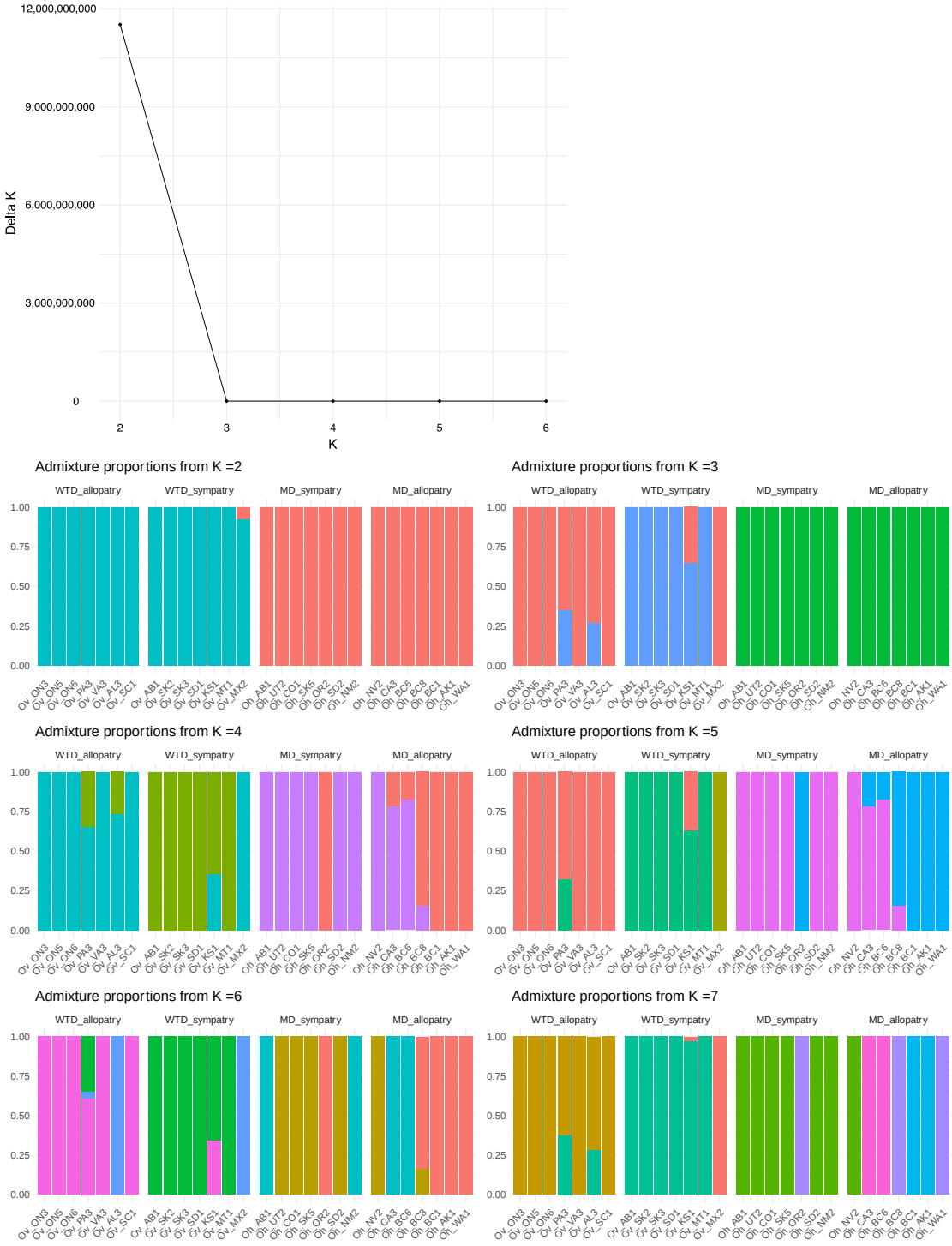


Figure S2.2: PCA of 73 Odocoileus individuals from unpublished data. WTD coloured un blue, MD in orange.

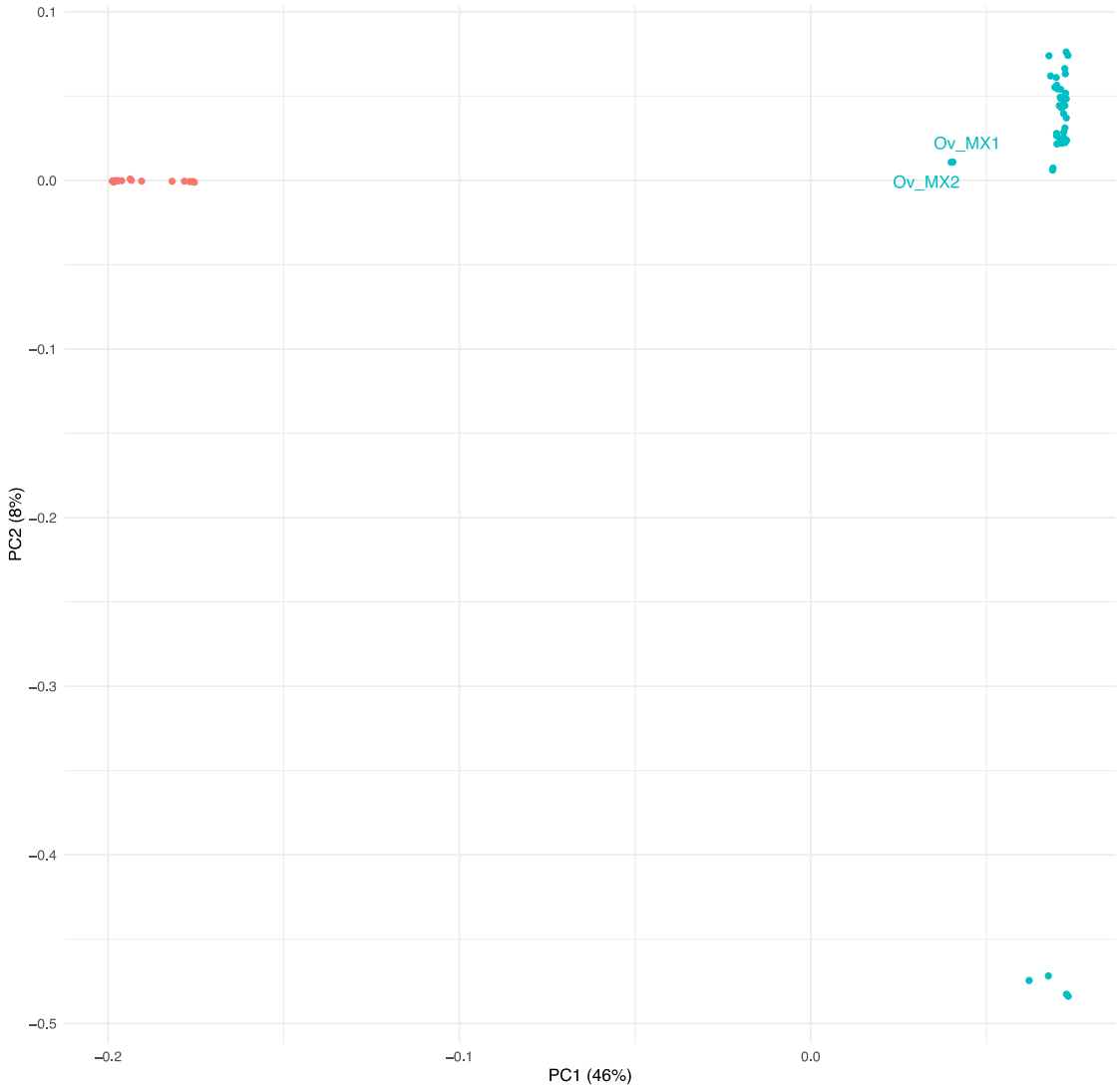


Figure S2.3: ABBA-BABA analysis between populations, colour gradient represents Z-score, excess of ABBA depicted as points, excess of BABA shown as triangles.

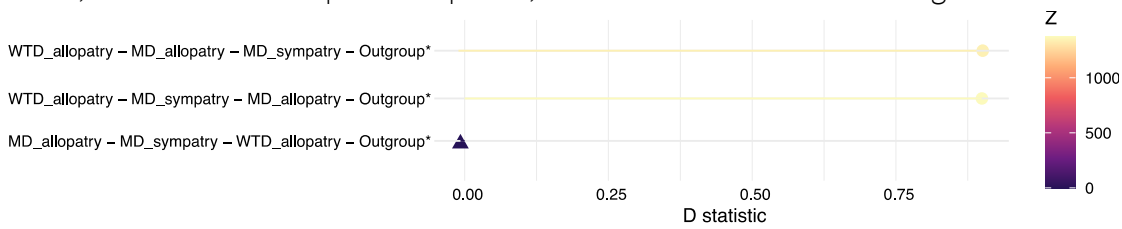


Figure S2.4: In windows ABBA-BABA outliers of fd for two comparisons

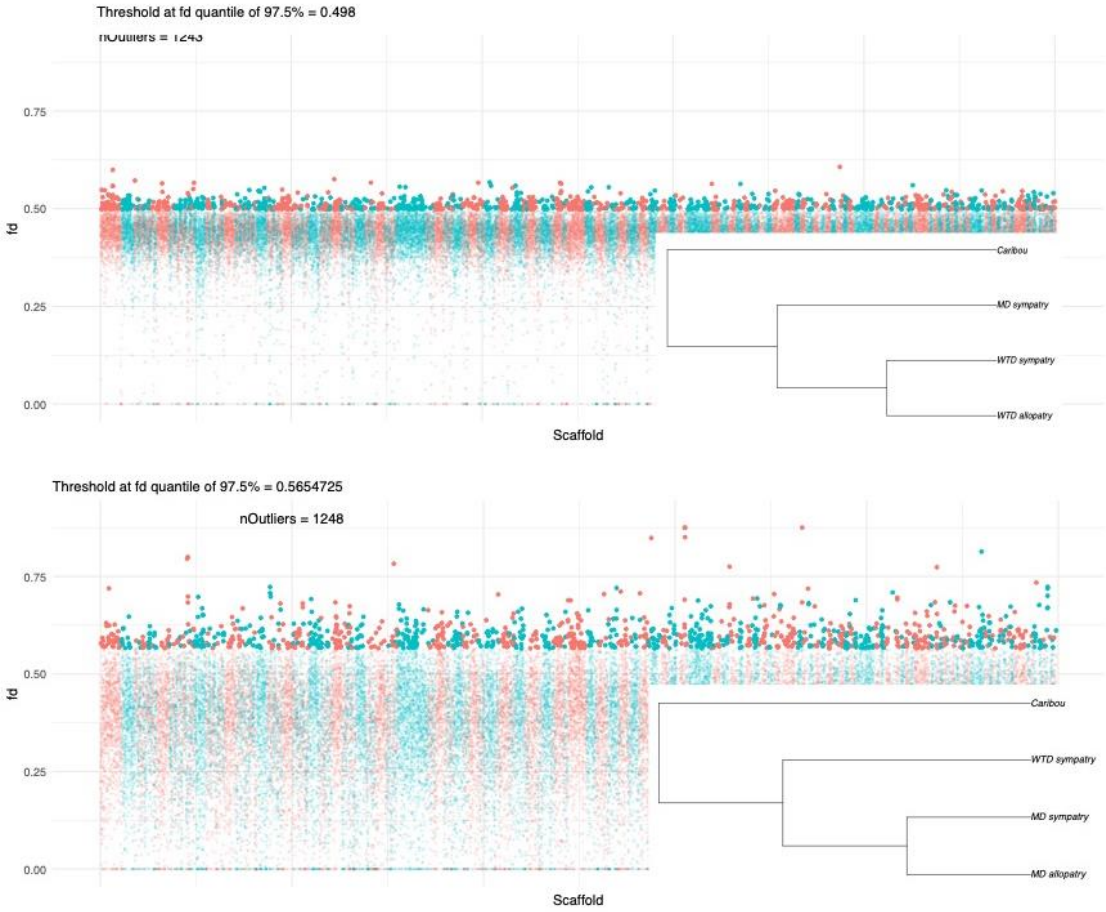


Figure S2.5: Count of outlier windows per scaffolds (top) and example of outliers in balancing selection across scaffold 1816 (right) and scaffold 393 (left; Outlier in darker shades)

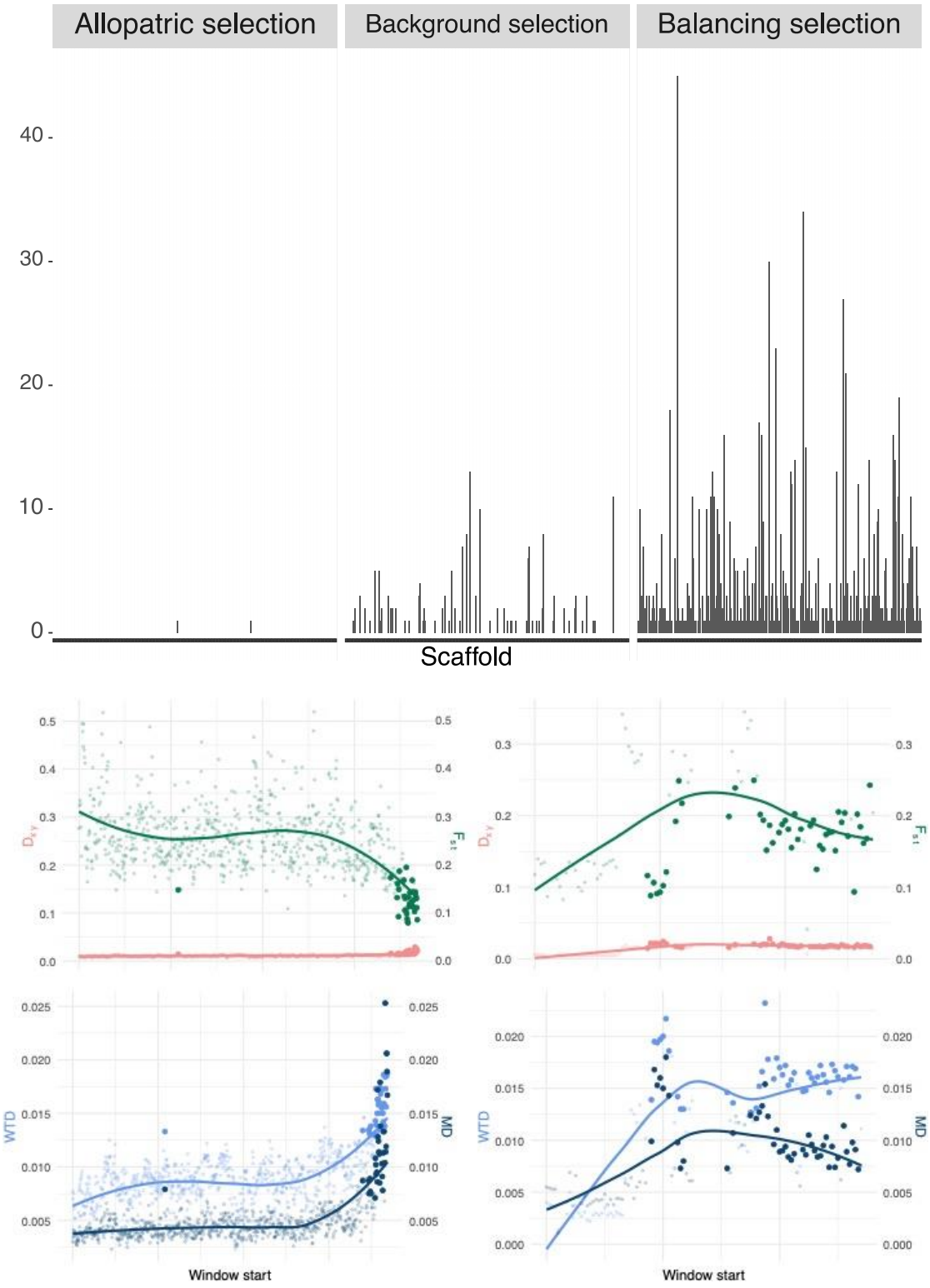


Figure S2.6: Gene enrichment analysis for windows in balancing selection for three GO categories: Biological Process (top), Cellular Component (middle) and Molecular Function (bottom)

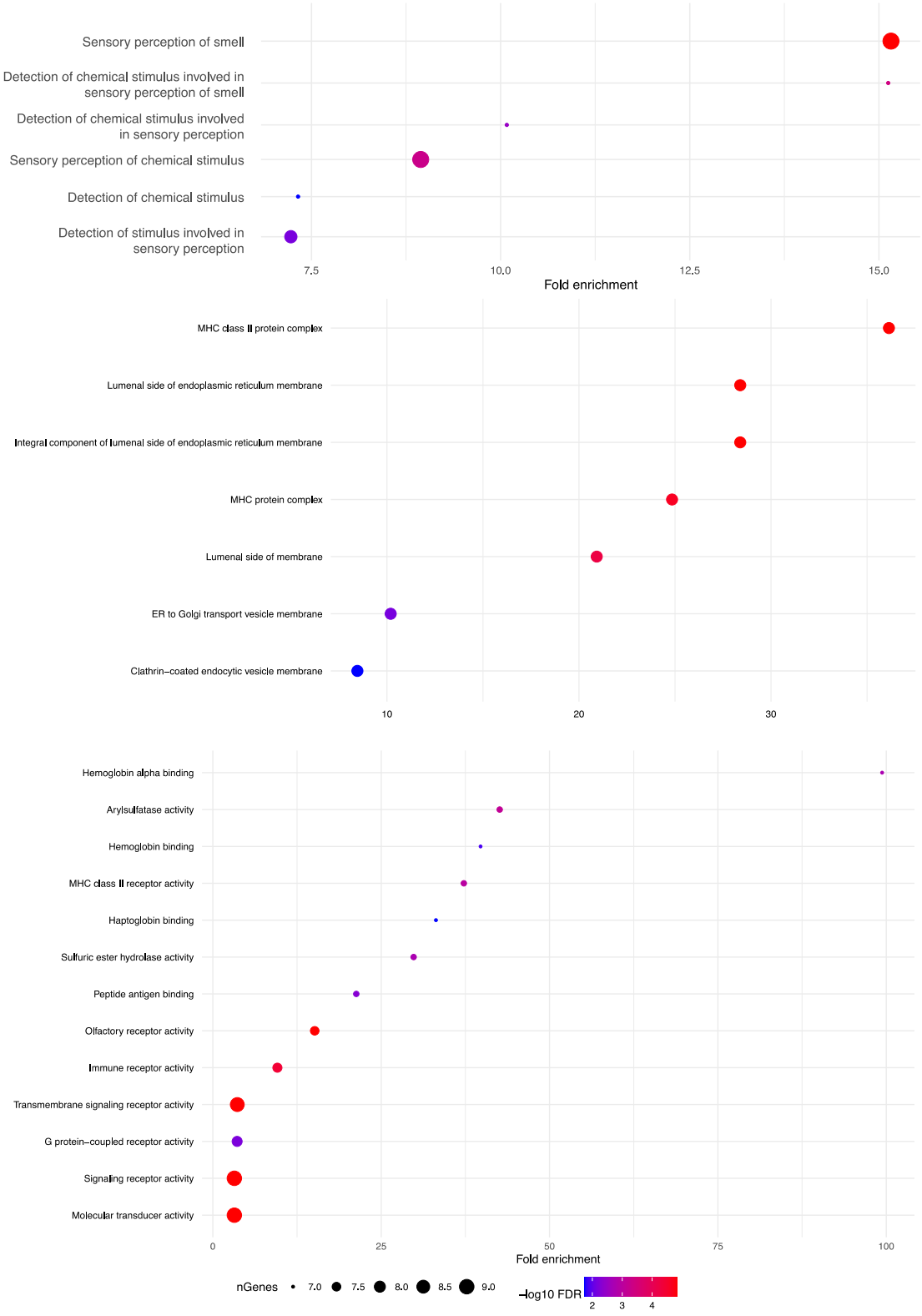
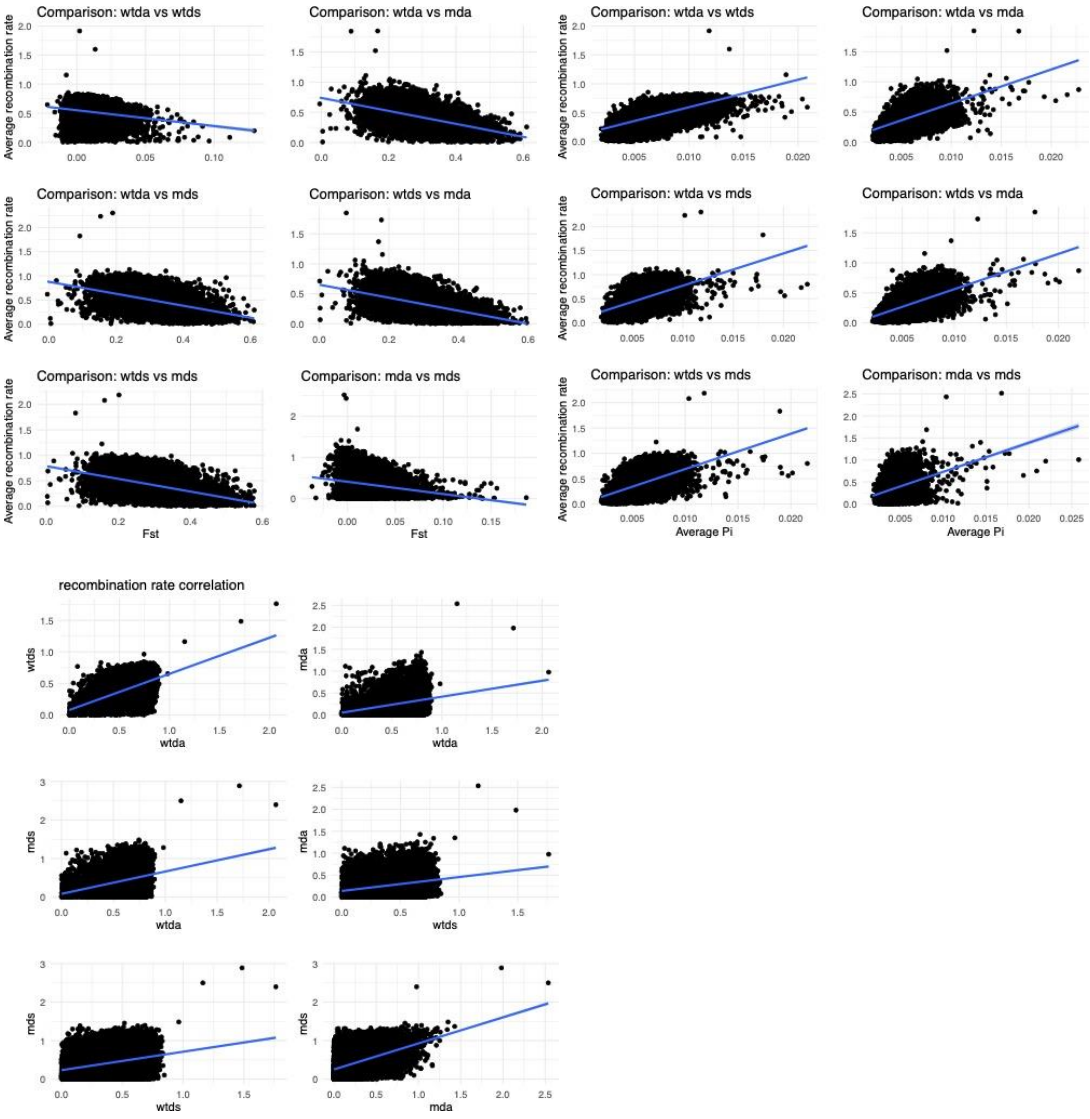


Figure S2.7: Gene enrichment analysis for windows in background selection for three GO categories: Biological Process (top), Cellular Component (middle) and Molecular Function (bottom)

Figure S2.8: Correlations between r1-2 vs FST (top-left), r1-2 vs p1-2 (top-right) and r1 vs r2 (bottom-left) for all six comparisons



Tables

Table S2.1: Number of windows in the evolutionary scenarios using different thresholds, (†) indicates which thresholds we used for the analyses.

Thresholds	DwGF	AS	BGS	BLS	Total
†Lower = 5% Upper = 95% Stable = Between 45 & 55%	0	2	165	1016	1183
Lower = 10% Upper = 90% Stable = Between 35 & 65%	0	35	961	2314	3310
Lower = 15% Upper = 85% Stable = Between 30 & 70%	0	153	2007	3658	5818
Lower = 20% Upper = 80% Stable = Between 20 & 80%	2	739	2977	5207	8925

Table S2.2: Additional sample information, approximate sampling coordinates.

Genome_ID	Location	Latitude	Longitude
Ov_ON3	Ontario	42.4800	-82.0000
Ov_ON5	Ontario	46.3600	-84.0000
Ov_ON6	Ontario	44.5300	-78.0500
Ov_PA3	Pennsylvania	39.8944	-80.0870
Ov_VA3	Virginia	38.1300	-78.1600
Ov_AL3	Alabama	32.3823	-85.6858
Ov_SC1	South Carolina	33.4148	-80.4105
Ov_AB1	Alberta	52.7800	-111.0148
Ov_SK2	Saskatchewan	52.8600	-102.3600
Ov_SK3	Saskatchewan	53.6330	-109.2000
Ov_SD1	South Dakota	44.4415	-102.6855
Ov_KS1	Kansas	38.6900	-100.8158
Ov_MT1	Montana	47.5275	-113.7110
Ov_MX2	Mexico	30.7295	-108.7348
Oh_NV2	Nevada	41.3036	-115.1377
Oh_CA3	California	35.3733	-119.0187
Oh_BC6	British Columbia	54.7824	-127.1686
Oh_BC8	British Columbia	49.1579	-121.9515
Oh_BC1	British Columbia	48.6955	-123.3223
Oh_AK1	Alaska	57.6853	-134.4895
Oh_WA1	Washington	46.8806	-122.4092
Oh_AB1	Alberta	52.6251	-110.7518
Oh_UT2	Utah	40.9733	-111.6903
Oh_CO1	Colorado	39.4853	-108.0932
Oh_SK5	Saskatchewan	50.7442	-107.8104
Oh_OR2	Oregon	44.4100	-122.5340
Oh_SD2	South Dakota	43.2240	-103.4512
Oh_NM2	New Mexico	36.1846	-103.5916

Appendix II: Supplementary material for Chapter 3

Figures

Figure S3.1: Population structure analyses. (A) PCA of WTD individuals. (B) PCA of MD individuals. (C) Admixture proportions of the MD individuals. (D) Genetic distance in the form of number of divergent loci between individuals, out of 249,380 snps.

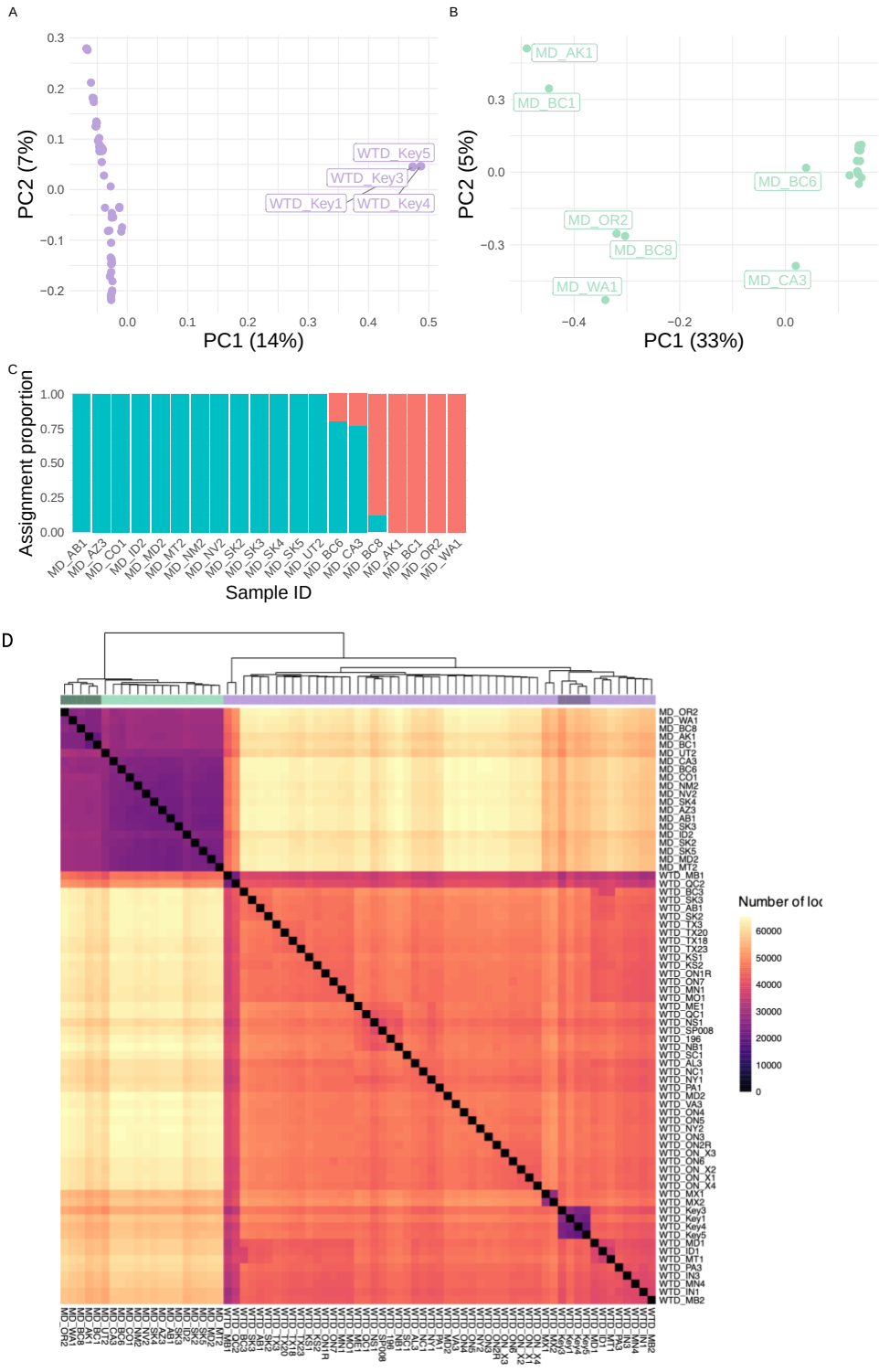


Figure S3.2: Summary statistics, (A) Genome-wide Tajima's D distribution per population, with average across scaffolds (B) F_{ROH} distribution per individual per population. The y-axis F_{ROH} are expressed as percentage of the genome.

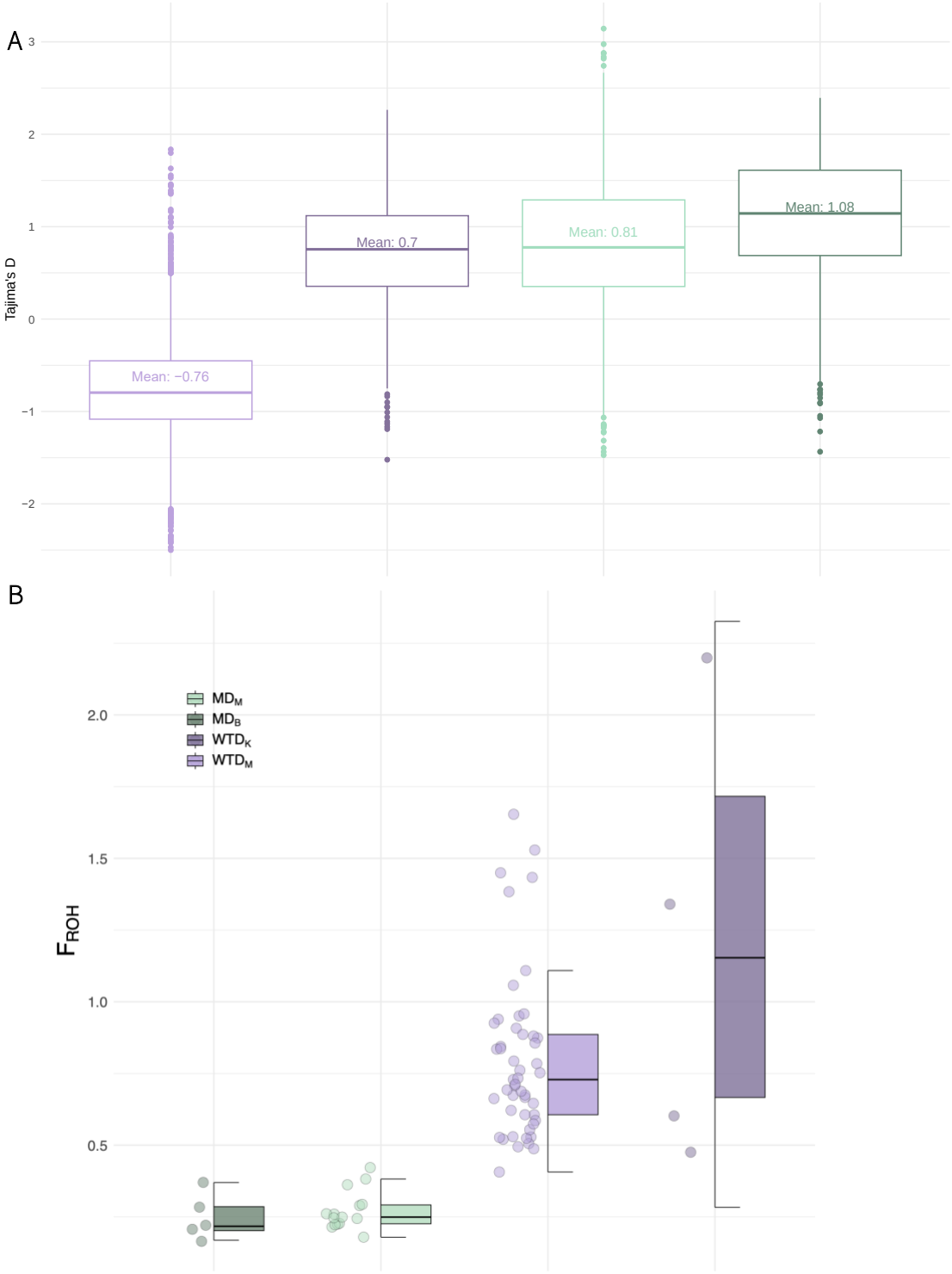


Figure S3.3: MSMC2 cross coalescent rates between species (A) and within species (B), migration rate (C). CCR of one means that populations are not differentiated, we used CCR = 0.5 as population differentiation.

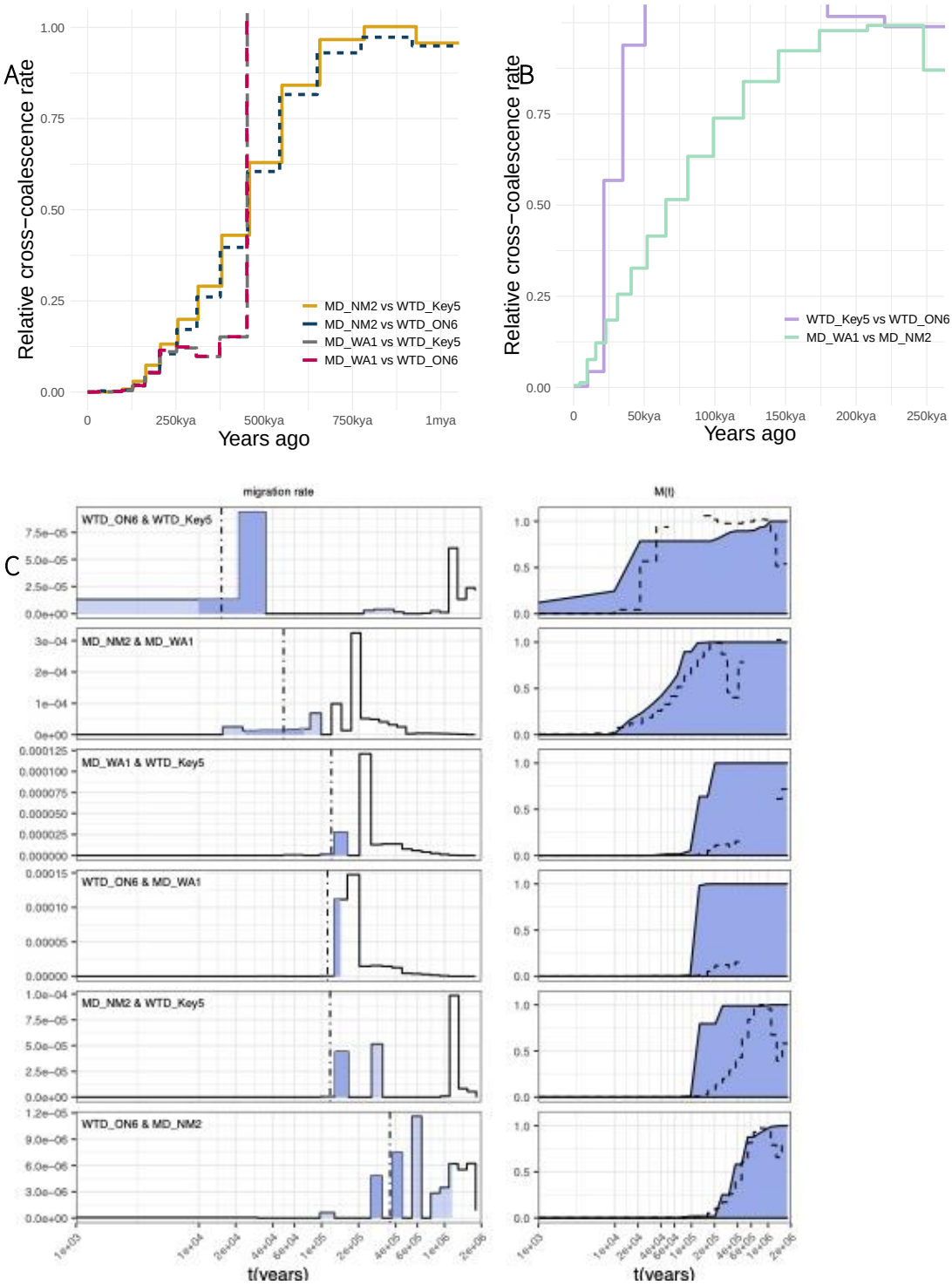


Figure S3.4: Stairway plot analysis for each population, grey rectangle represents the LGM; dashed lines represent 95% confidence interval.

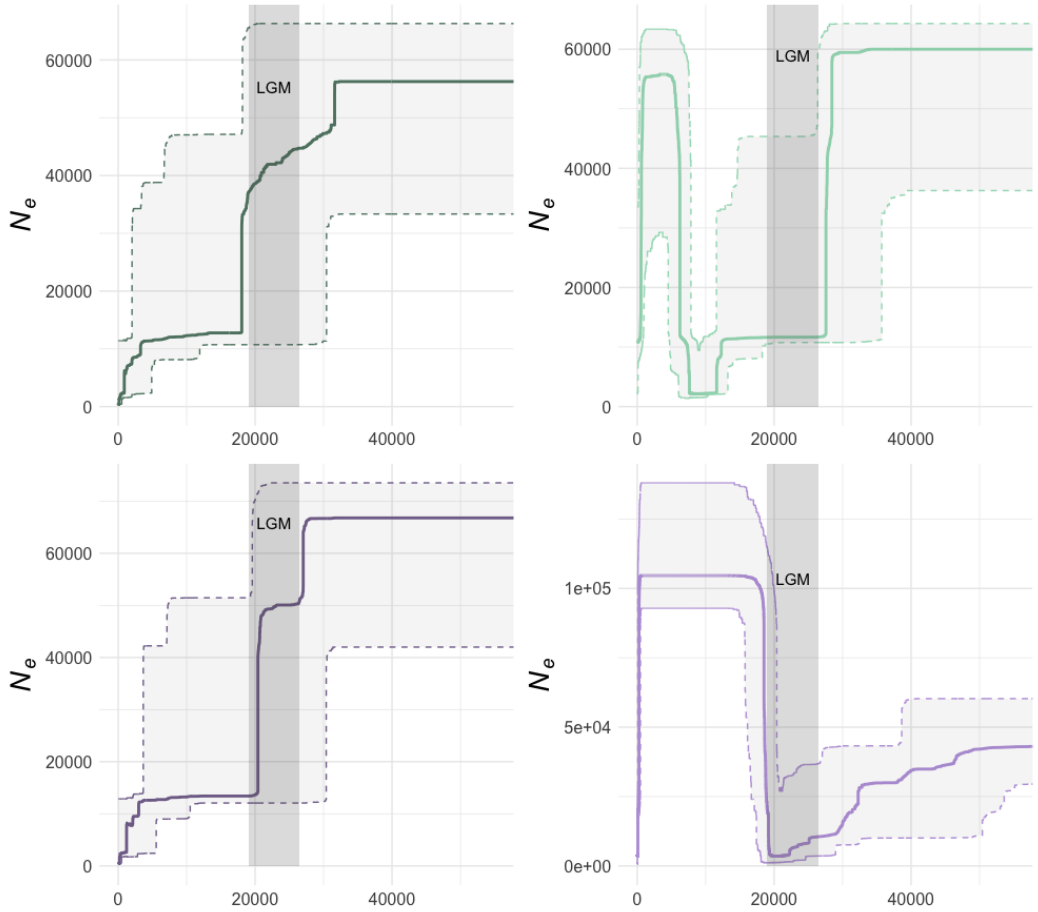


Figure S3.5: GONE uncertainty for the last 200 generations (400 years), y-axis in log10 scale, x-axis in time bins of 10 generations, error bars represent 95% CI, colour represents population.

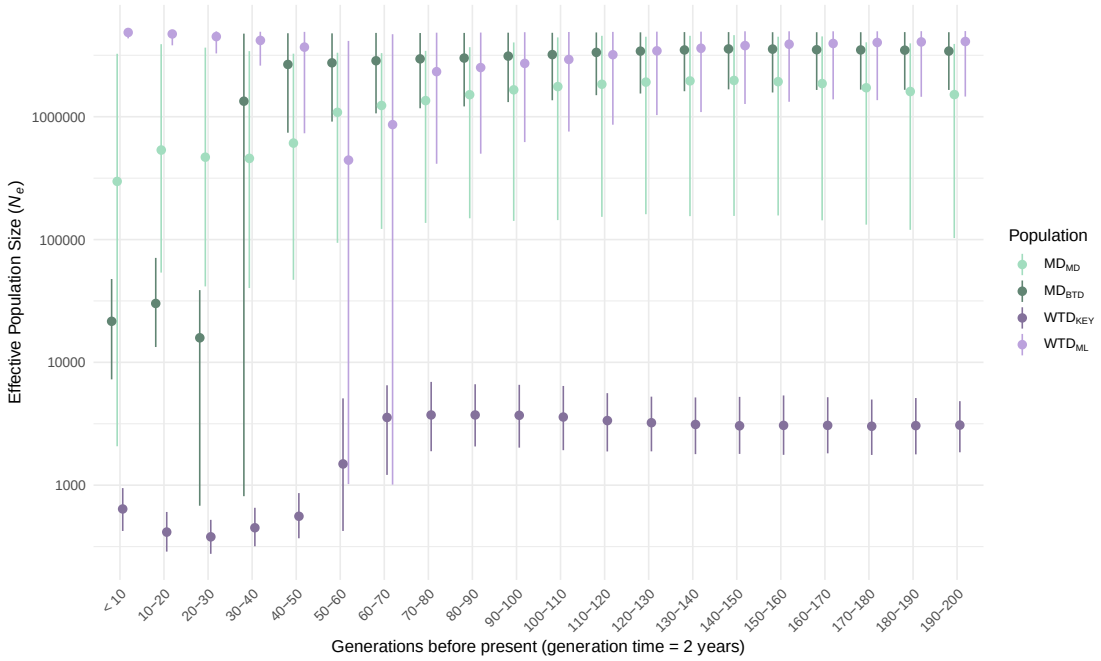


Figure S3.6: δaδi parameters uncertainty: Densities of parameters and best parameter value for the WTD_{KEY}, WTD_{ML} and all MD comparison (A), and the MD_{MD}, MD_{BTD} and WTD_{ML} comparison (B).

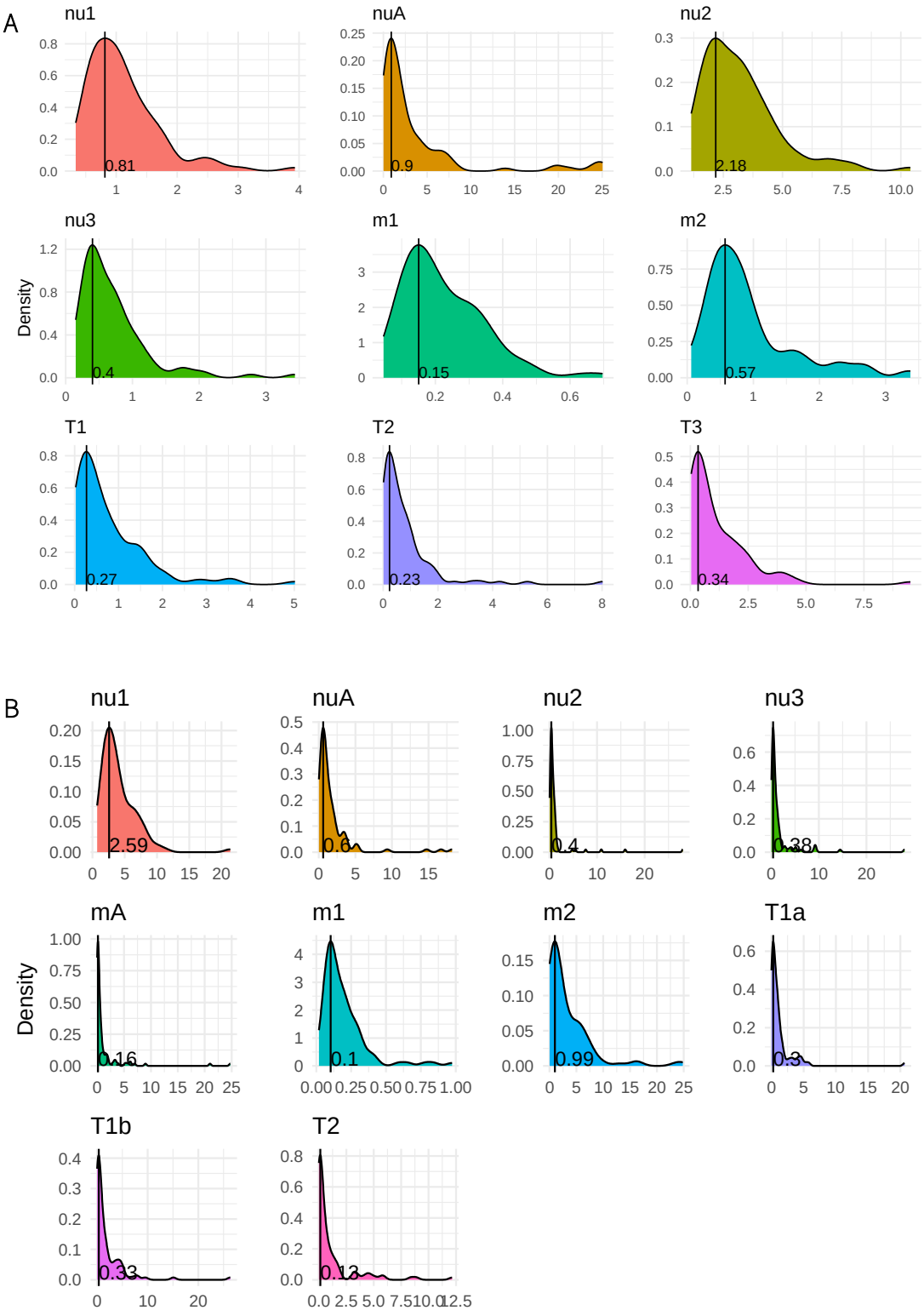
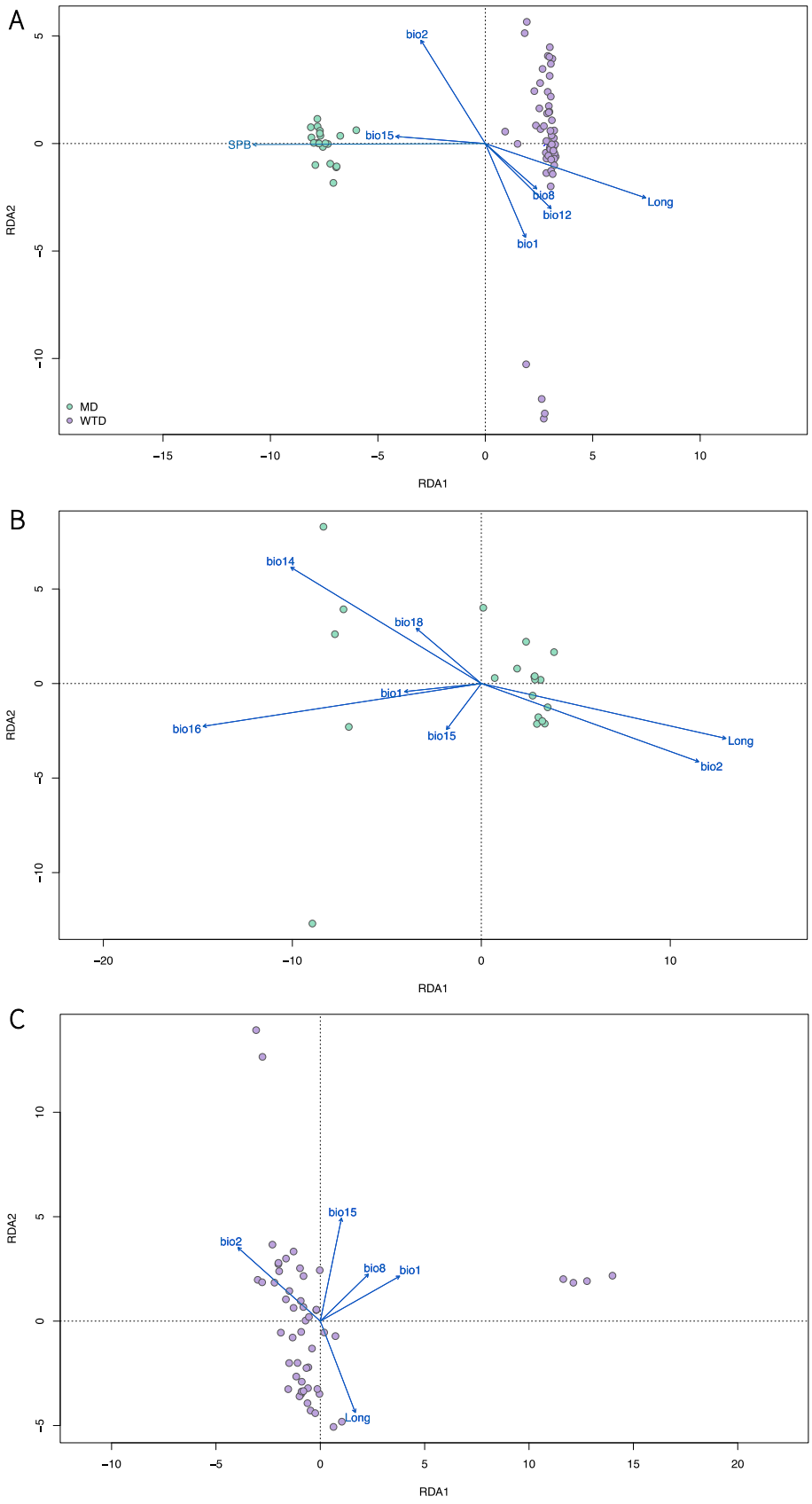


Figure S3.7: RDA between species (A), within MD (B) and within WTD (C).



Tables

Table S3.1: Sample information; The four re-sequenced samples are highlighted in bold, assigned population taken from the PCA and NGSAdmix analyses with corresponding colour code, sample coordinates were not taken at sampling site but estimated from the approximate sampling location.

Sample ID	Species	Location	Assigned population	Sex	Latitude (approx)	Longitude (approx)
MD_BC1	*O_hemionus*	British Columbia	MD_{BTD}	F	48.69	-123.32
MD_WA1	**O_hemionus**	**Washington**	**MD_{BTD}**	**M**	**46.88**	**-122.40**
MD_AB1	O_hemionus	Alberta	MD_{MD}	F	52.62	-110.75
MD_AK1	*O_hemionus*	Alaska	MD_{BTD}	NA	57.68	-134.48
MD_AZ3	*O_hemionus*	Arizona	MD_{MD}	NA	35.19	-111.65
MD_BC6	*O_hemionus*	British Columbia	MD_{MD}	M	54.78	-127.16
MD_BC8	*O_hemionus*	British Columbia	MD_{BTD}	M	49.15	-121.95
MD_CA3	*O_hemionus*	California	MD_{MD}	NA	35.37	-119.01
MD_CO1	*O_hemionus*	Colorado	MD_{MD}	F	39.48	-108.09
MD_ID2	*O_hemionus*	Idaho	MD_{MD}	F	42.05	-111.90
MD_MT2	*O_hemionus*	Montana	MD_{MD}	M	45.42	-109.73
MD_NM2	**O_hemionus**	**New Mexico**	**MD_{MD}**	**M**	**36.18**	**-103.59**
MD_NV2	*O_hemionus*	Nevada	MD_{MD}	M	41.30	-115.13
MD_OR2	*O_hemionus*	Oregon	MD_{BTD}	NA	44.41	-122.53
MD_SD2	*O_hemionus*	South Dakota	MD_{MD}	F	43.22	-103.45
MD_SK2	*O_hemionus*	Saskatchewan	MD_{MD}	NA	51.02	-106.43
MD_SK3	*O_hemionus*	Saskatchewan	MD_{MD}	NA	51.02	-106.43
MD_SK4	*O_hemionus*	Saskatchewan	MD_{MD}	NA	50.74	-107.81
MD_SK5	*O_hemionus*	Saskatchewan	MD_{MD}	NA	50.74	-107.81
MD_UT2	*O_hemionus*	Utah	MD_{MD}	M	40.97	-111.69
WTD_AB1	*O_virginianus*	Alberta	WTD_{ML}	F	52.78	-111.01
WTD_AL3	*O_virginianus*	Alabama	WTD_{ML}	NA	32.38	-85.68
WTD_BC3	*O_virginianus*	British Columbia	WTD_{ML}	M	49.51	-115.76
WTD_IN1	*O_virginianus*	Indiana	WTD_{ML}	NA	41.25	-87.22
WTD_IN3	*O_virginianus*	Indiana	WTD_{ML}	NA	39.25	-85.04
WTD_Key1	*O_virginianus*	Florida	WTD_{KEY}	NA	24.69	-81.32
WTD_Key3	*O_virginianus*	Florida	WTD_{KEY}	NA	24.69	-81.32
WTD_Key4	*O_virginianus*	Florida	WTD_{KEY}	NA	24.69	-81.32
WTD_Key5	**O_virginianus**	**Florida**	**WTD_{KEY}**	**NA**	**24.69**	**-81.32**
WTD_KS1	*O_virginianus*	Kansas	WTD_{ML}	F	38.69	-100.81
WTD_KS2	*O_virginianus*	Kansas	WTD_{ML}	M	38.43	-96.18
WTD_MB1	*O_virginianus*	Manitoba	WTD_{ML}	M	51.52	-100.39
WTD_MB2	*O_virginianus*	Manitoba	WTD_{ML}	M	50.61	-97.56
WTD_MD2	*O_virginianus*	Maryland	WTD_{ML}	F	39.62	-78.61
WTD_ME1	*O_virginianus*	Maine	WTD_{ML}	F	43.66	-70.63
WTD_MN1	*O_virginianus*	Minnesota	WTD_{ML}	M	46.35	-94.20
WTD_MN4	*O_virginianus*	Minnesota	WTD_{ML}	M	43.67	-92.08

WTD_MO1	*O_virginianus*	Missouri	WTD_{ML}	M	38.38	-91.90
WTD_MT1	*O_virginianus*	Montana	WTD_{ML}	M	47.52	-113.71
WTD_MX1	*O_virginianus*	Mexico	WTD_{ML}	M	30.72	-108.73
WTD_MX2	*O_virginianus*	Mexico	WTD_{ML}	M	30.72	-108.73
WTD_NB1	*O_virginianus*	New Brunswick	WTD_{ML}	M	45.68	-66.69
WTD_NC1	*O_virginianus*	North Carolina	WTD_{ML}	M	34.65	-77.47
WTD_NS1	*O_virginianus*	Nova Scotia	WTD_{ML}	M	45.05	-63.16
WTD_NY1	*O_virginianus*	New York	WTD_{ML}	F	42.64	-73.74
WTD_NY2	*O_virginianus*	New York	WTD_{ML}	F	43.02	-76.17
WTD_ON1R	*O_virginianus*	Ontario	WTD_{ML}	M	48.24	-89.12
WTD_ON2R	*O_virginianus*	Ontario	WTD_{ML}	M	45.12	-75
WTD_ON3	*O_virginianus*	Ontario	WTD_{ML}	F	42.48	-82
WTD_ON4	*O_virginianus*	Ontario	WTD_{ML}	M	44.92	-83.4
WTD_ON5	*O_virginianus*	Ontario	WTD_{ML}	M	46.36	-84
WTD_ON6	**O_virginianus**	**Ontario**	**WTD_{ML}**	**M**	**44.53**	**-78.05**
WTD_ON7	*O_virginianus*	Ontario	WTD_{ML}	M	49.76	-94.48
WTD_ONX1	*O_virginianus*	Ontario	WTD_{ML}	F	44.58	-78.07
WTD_ONX2	*O_virginianus*	Ontario	WTD_{ML}	F	44.58	-78.07
WTD_ONX3	*O_virginianus*	Ontario	WTD_{ML}	F	44.58	-78.25
WTD_ONX4	*O_virginianus*	Ontario	WTD_{ML}	M	44.62	-78.13
WTD_PA1	*O_virginianus*	Pennsylvania	WTD_{ML}	M	39.96	-76.72
WTD_PA3	*O_virginianus*	Pennsylvania	WTD_{ML}	F	39.89	-80.08
WTD_QC1	*O_virginianus*	Quebec	WTD_{ML}	M	46.9	-70.43
WTD_QC2	*O_virginianus*	Quebec	WTD_{ML}	F	45.40	-71.89
WTD_SC1	*O_virginianus*	South Carolina	WTD_{ML}	F	33.41	-80.41
WTD_SD1	*O_virginianus*	South Dakota	WTD_{ML}	M	44.44	-102.68
WTD_SK2	*O_virginianus*	Saskatchewan	WTD_{ML}	M	52.86	-102.36
WTD_SK3	*O_virginianus*	Saskatchewan	WTD_{ML}	M	53.63	-109.2
WTD_SP008	*O_virginianus*	St-Pierre et Miquelon	WTD_{ML}	F	46.86	-56.24
WTD_TX18	*O_virginianus*	Texas	WTD_{ML}	F	30.71	-94.81
WTD_TX20	*O_virginianus*	Texas	WTD_{ML}	F	33.18	-99.27
WTD_TX23	*O_virginianus*	Texas	WTD_{ML}	M	31.78	-101.07
WTD_TX3	*O_virginianus*	Texas	WTD_{ML}	F	29.29	-99.01
WTD_VA3	*O_virginianus*	Virginia	WTD_{ML}	F	38.13	-78.16
WTD_196	*O_virginianus*	Anticosti Island	WTD_{ML}	NA	49.47	-63.09
WTD_ID1	*O_virginianus*	Idaho	WTD_{ML}	F	48.68	-116.36

Table S3.2: List of 3D models run in dadi_pipeline (Portik et al. 2017) for two comparisons

Comparison	Model	Reference
WTD_{KEY} - WTK_{ML} - MD	split_nomig	Portik et al., 2017
	split_symmig_adjacent	
	refugia_adj_1	
	refugia_adj_2	
	refugia_adj_3	
	ancmig_adj_1	
	ancmig_adj_2	
	ancmig_adj_3	
	ancmig_2_size	Barratt et al., 2018
	split_sym_mig_adjacent_var1	Firneno et al., 2020
	split_sym_mig_adjacent_var2	
MD_{BTD} - MD_{MD} - WTD_{ML}	split_nomig	Portik et al., 2017
	split_symmig_all	
	split_symmig_adjacent	
	refugia_adj_1	
	refugia_adj_2	
	refugia_adj_3	
	refugia_adj_2_var_sym	Firneno et al., 2020
	refugia_adj_3_var_sym	
	refugia_adj_2_var_uni	
	refugia_adj_3_var_uni	
	split_sym_mig_adjacent_var1	
	split_uni_mig_adjacent_var1	

Table S3.3: Bioclimatic variables from WorldClim2

Code	Bioclimatic variable
BIO1	Annual Mean Temperature
BIO2	Mean Diurnal Range (Mean of monthly (max temp - min temp))
BIO3	Isothermality (BIO2/BIO7) (×100)
BIO4	Temperature Seasonality (standard deviation ×100)
BIO5	Max Temperature of Warmest Month
BIO6	Min Temperature of Coldest Month
BIO7	Temperature Annual Range (BIO5-BIO6)
BIO8	Mean Temperature of Wettest Quarter
BIO9	Mean Temperature of Driest Quarter
BIO10	Mean Temperature of Warmest Quarter
BIO11	Mean Temperature of Coldest Quarter
BIO12	Annual Precipitation
BIO13	Precipitation of Wettest Month
BIO14	Precipitation of Driest Month
BIO15	Precipitation Seasonality (Coefficient of Variation)
BIO16	Precipitation of Wettest Quarter
BIO17	Precipitation of Driest Quarter
BIO18	Precipitation of Warmest Quarter
BIO19	Precipitation of Coldest Quarter

Table S3.4: Weighted F_{ST} values between populations

	WTD_{ML}	WTD_{KEY}	MD_{MD}
WTD_{KEY}	0.17		
MD_{MD}	0.4	0.58	
MD_{BTD}	0.36	0.56	0.12

Table S3.5: Best model parameter values from δaδi per comparison and corresponding 95% CI computed on 100 bootstraps. Converted parameters values found by solving for $\Theta = 4N_{ref}\mu L$.

		nu				m			T				
	Theta	**nu1**	**nu2**	**nu3**	**nuA**	**m1**	**m2**	**mA**	**T1**	**T1a**	**T1b**	**T2**	**T3**
WTD_{KEY} & WTD_{ML} & MD													
Parameters	21'150.17	0.56	1.73	0.20	4.02	0.86	2.15	-	0.57	-	-	0.12	0.02
2.5% CI		0.44	1.34	0.18	0.06	0.07	0.11	-	0.03	-	-	0.02	0.06
97.5% CI		2.61	7.47	2.04	24.23	0.49	2.70	-	3.27	-	-	3.95	4.10
Converted parameters		55'432	171'381	19'344	397'956	4.4E-06	1.1E-05	-	226'081	-	-	47'471	6'725
WTD_{ML} & MD_{BTD} & MD_{MD}													
Parameters	20'887.17	2.09	0.29	0.27	0.84	0.79	0.79	0.02	-	0.11	0.86	0.04	-
2.5% CI		1.19	0.15	0.10	0.08	0.05	0.10	0.03	-	0.03	0.05	0.02	-
97.5% CI		9.82	9.34	9.28	12.18	0.72	16.21	17.37	-	5.07	8.83	7.19	-
Converted parameters		204'098	28'412	25'941	81'573	4.1E-06	4.0E-06	1.0E-07	-	44'420	335'120	13'869	-

Table S3.6: 249 genes within 25kbp of iHS peaks.

Entry	Gene Names	Function (uniprot.org)	Category	Pop
O14497	*ARID1A BAF250 BAF250A C1orf4 OSA1 SMARCF1*	FUNCTION: Involved in transcriptional activation and repression of select genes by chromatin remodeling (alteration of DNA-nucleosome topology). Component of SWI/SNF chromatin remodeling complexes that carry out key enzymatic activities, changing chromatin structure by altering DNA-histone contacts within a nucleosome in an ATP-dependent manner. Binds DNA non-specifically. Belongs to the neural progenitors-specific chromatin remodeling complex (npBAF complex) and the neuron-specific chromatin remodeling complex (nBAF complex). During neural development a switch from a stem/progenitor to a postmitotic chromatin remodeling mechanism occurs as neurons exit the cell cycle and become committed to their adult state. The transition from proliferating neural stem/progenitor cells to postmitotic neurons requires a switch in subunit composition of the npBAF and nBAF complexes. As neural progenitors exit mitosis and differentiate into neurons, npBAF complexes which contain ACTL6A/BAF53A and PHF10/BAF45A, are exchanged for homologous alternative ACTL6B/BAF53B and DPF1/BAF45B or DPF3/BAF45C subunits in neuron-specific complexes (nBAF). The npBAF complex is essential for the self-renewal/proliferative capacity of the multipotent neural stem cells. The nBAF complex along with CREST plays a role regulating the activity of genes essential for dendrite growth (By similarity). .	*Development*	WTDML
O60229	*KALRN DUET DUO HAPIP TRAD*	FUNCTION: Promotes the exchange of GDP by GTP. Activates specific Rho GTPase family members, thereby inducing various signaling mechanisms that regulate neuronal shape, growth, and plasticity, through their effects on the actin cytoskeleton. Induces lamellipodia independent of its GEF activity. .	*Development*	WTDML
P12272	*PTHLH PTHRP*	FUNCTION: Neuroendocrine peptide which is a critical regulator of cellular and organ growth, development, migration, differentiation and survival and of epithelial calcium ion transport. Regulates endochondral bone development and epithelial-mesenchymal interactions during the formation of the mammary glands and teeth. Required for skeletal homeostasis. Promotes mammary mesenchyme differentiation and bud outgrowth by modulating mesenchymal cell responsiveness to BMPs. Up-regulates BMPR1A expression in the mammary mesenchyme and	*Development*	WTDML

		this increases the sensitivity of these cells to BMPs and allows them to respond to BMP4 in a paracrine and/or autocrine fashion. BMP4 signaling in the mesenchyme, in turn, triggers epithelial outgrowth and augments MSX2 expression, which causes the mammary mesenchyme to inhibit hair follicle formation within the nipple sheath (By similarity). Promotes colon cancer cell migration and invasion in an integrin alpha-6/beta-1-dependent manner through activation of Rac1. .; FUNCTION: Osteostatin is a potent inhibitor of osteoclastic bone resorption. .		
Q19T08	*ECSCR ECSM2*	FUNCTION: Regulates endothelial chemotaxis and tube formation. Has a role in angiogenesis and apoptosis via modulation of the actin cytoskeleton and facilitation of proteasomal degradation of the apoptosis inhibitors BIRC3/IAP1 and BIRC2/IAP2. .	*Development*	WTDML
Q4G0U5	*CFAP221 PCDP1*	FUNCTION: May play a role in cilium morphogenesis. .	*Development*	WTDML
Q5HYA8	*TMEM67 MKS3*	FUNCTION: Required for ciliary structure and function. Part of the tectonic-like complex which is required for tissue-specific ciliogenesis and may regulate ciliary membrane composition (By similarity). Involved in centrosome migration to the apical cell surface during early ciliogenesis. Involved in the regulation of cilia length and appropriate number through the control of centrosome duplication. Is a key regulator of stereociliary bundle orientation (By similarity). Required for epithelial cell branching morphology. Essential for endoplasmic reticulum-associated degradation (ERAD) of surfactant protein C (SFTPC). Involved in the negative regulation of canonical Wnt signaling, and activation of the non-canonical cascade stimulated by WNT5A . In non-canonical Wnt signaling, it may act as ROR2 coreceptor (By similarity). .	*Development*	MDBTD
Q5VUG0	*SFMBT2 KIAA1617*	FUNCTION: Transcriptional repressor of HOXB13 gene. .	*Development*	WTDKEY
Q68J44	*DUSP29 DUPD1 DUSP27*	FUNCTION: Dual specificity phosphatase able to dephosphorylate phosphotyrosine, phosphoserine and phosphothreonine residues within the same substrate, with a preference for phosphotyrosine as a substrate . Involved in the modulation of intracellular signaling cascades. In skeletal muscle regulates systemic glucose homeostasis by activating, AMPK, an energy sensor protein kinase (By similarity). Affects MAP kinase signaling though modulation of the MAPK1/2 cascade in skeletal muscle promoting	*Development*	WTDKEY

		muscle cell differentiation, development and atrophy (By similarity). .		
Q6PI78	*TMEM65*	FUNCTION: May play an important role in cardiac development and function. May regulate cardiac conduction and the function of the gap junction protein GJA1. May contribute to the stability and proper localization of GJA1 to cardiac intercalated disk thereby regulating gap junction communication (By similarity). May also play a role in the regulation of mitochondrial respiration and mitochondrial DNA copy number maintenance . .	*Development*	WTDKEY
Q7L9L4	*MOB1B* *MOB4A* *MOBKL1A*	FUNCTION: Activator of LATS1/2 in the Hippo signaling pathway which plays a pivotal role in organ size control and tumor suppression by restricting proliferation and promoting apoptosis. The core of this pathway is composed of a kinase cascade wherein STK3/MST2 and STK4/MST1, in complex with its regulatory protein SAV1, phosphorylates and activates LATS1/2 in complex with its regulatory protein MOB1, which in turn phosphorylates and inactivates YAP1 oncoprotein and WWTR1/TAZ. Phosphorylation of YAP1 by LATS1/2 inhibits its translocation into the nucleus to regulate cellular genes important for cell proliferation, cell death, and cell migration. Stimulates the kinase activity of STK38L. .	*Development*	WTDKEY
Q8IZT6	*ASPM* *MCPH5*	FUNCTION: Involved in mitotic spindle regulation and coordination of mitotic processes. The function in regulating microtubule dynamics at spindle poles including spindle orientation, astral microtubule density and poleward microtubule flux seems to depend on the association with the katanin complex formed by KATNA1 and KATNB1. Enhances the microtubule lattice severing activity of KATNA1 by recruiting the katanin complex to microtubules. Can block microtubule minus-end growth and reversely this function can be enhanced by the katanin complex . May have a preferential role in regulating neurogenesis. .	*Development*	MDMD
Q8WVF2	*UCMA* *C10orf49*	FUNCTION: May be involved in the negative control of osteogenic differentiation of osteochondrogenic precursor cells in peripheral zones of fetal cartilage and at the cartilage-bone interface. .	*Development*	WTDML
Q8WYB5	*KAT6B* *KIAA0383* *MORF* *MOZ2* *MYST4*	FUNCTION: Histone acetyltransferase which may be involved in both positive and negative regulation of transcription. Required for RUNX2-dependent transcriptional activation. May be involved in cerebral cortex development. Component of the MOZ/MORF complex which has a histone H3 acetyltransferase activity. .	*Development*	WTDKEY

Q9H799	*CPLANE1* *C5orf42* *JBTS17*	FUNCTION: Involved in ciliogenesis . Involved in the establishment of cell polarity required for directional cell migration. Proposed to act in association with the CPLANE (ciliogenesis and planar polarity effectors) complex. Involved in recruitment of peripheral IFT-A proteins to basal bodies (By similarity). .	*Development*	WTDML
Q9UH90	*FBXO40* *FBX40* *KIAA1195*	FUNCTION: Probable substrate-recognition component of the SCF (SKP1-CUL1-F-box protein)-type E3 ubiquitin ligase complex that may function in myogenesis. .	*Development*	MDMD
Q9UMX5	*NENF* *CIR2* *SPUF*	FUNCTION: Acts as a neurotrophic factor in postnatal mature neurons enhancing neuronal survival . Promotes cell proliferation and neurogenesis in undifferentiated neural progenitor cells at the embryonic stage and inhibits differentiation of astrocytes (By similarity). Its neurotrophic activity is exerted via MAPK1/ERK2, MAPK3/ERK1 and AKT1/AKT pathways (By similarity). Neurotrophic activity is enhanced by binding to heme (By similarity). Acts also as an anorexigenic neurotrophic factor that contributes to energy balance (By similarity). .	*Development*	MDMD
P98164	*LRP2*	FUNCTION: Multiligand endocytic receptor (By similarity). Acts together with CUBN to mediate endocytosis of high-density lipoproteins (By similarity). Mediates receptor-mediated uptake of polybasic drugs such as aprotinin, aminoglycosides and polymyxin B (By similarity). In the kidney, mediates the tubular uptake and clearance of leptin (By similarity). Also mediates transport of leptin across the blood-brain barrier through endocytosis at the choroid plexus epithelium (By similarity). Endocytosis of leptin in neuronal cells is required for hypothalamic leptin signaling and leptin-mediated regulation of feeding and body weight (By similarity). Mediates endocytosis and subsequent lysosomal degradation of CST3 in kidney proximal tubule cells (By similarity). Mediates renal uptake of 25-hydroxyvitamin D3 in complex with the vitamin D3 transporter GC/DBP (By similarity). Mediates renal uptake of metallothionein-bound heavy metals . Together with CUBN, mediates renal reabsorption of myoglobin (By similarity). Mediates renal uptake and subsequent lysosomal degradation of APOM (By similarity). Plays a role in kidney selenium homeostasis by mediating renal endocytosis of selenoprotein SEPP1 (By similarity). Mediates renal uptake of the antiapoptotic protein BIRC5/survivin which may be important for functional integrity of the kidney . Mediates renal uptake of matrix metalloproteinase MMP2 in complex with metalloproteinase inhibitor TIMP1 (By similarity).	*Development, physiology*	MDMD

		Mediates endocytosis of Sonic hedgehog protein N-product (ShhN), the active product of SHH (By similarity). Also mediates ShhN transcytosis (By similarity). In the embryonic neuroepithelium, mediates endocytic uptake and degradation of BMP4, is required for correct SHH localization in the ventral neural tube and plays a role in patterning of the ventral telencephalon (By similarity). Required at the onset of neurulation to sequester SHH on the apical surface of neuroepithelial cells of the rostral diencephalon ventral midline and to control PTCH1-dependent uptake and intracellular trafficking of SHH (By similarity). During neurulation, required in neuroepithelial cells for uptake of folate bound to the folate receptor FOLR1 which is necessary for neural tube closure (By similarity). In the adult brain, negatively regulates BMP signaling in the subependymal zone which enables neurogenesis to proceed (By similarity). In astrocytes, mediates endocytosis of ALB which is required for the synthesis of the neurotrophic factor oleic acid (By similarity). Involved in neurite branching (By similarity). During optic nerve development, required for SHH-mediated migration and proliferation of oligodendrocyte precursor cells (By similarity). Mediates endocytic uptake and clearance of SHH in the retinal margin which protects retinal progenitor cells from mitogenic stimuli and keeps them quiescent (By similarity). Plays a role in reproductive organ development by mediating uptake in reproductive tissues of androgen and estrogen bound to the sex hormone binding protein SHBG (By similarity). Mediates endocytosis of angiotensin-2 (By similarity). Also mediates endocytosis of angiotensis 1-7 (By similarity). Binds to the complex composed of beta-amyloid protein 40 and CLU/APOJ and mediates its endocytosis and lysosomal degradation (By similarity). Required for embryonic heart development (By similarity). Required for normal hearing, possibly through interaction with estrogen in the inner ear (By similarity). .		
Q13113	*PDZK1IP1 MAP17*	FUNCTION: May play an important role in tumor biology.	*Development, physiology*	WTDML
Q9BSE2	*TMEM79 MATT*	FUNCTION: Contributes to the epidermal integrity and skin barrier function. Plays a role in the lamellar granule (LG) secretory system and in the stratum corneum (SC) epithelial cell formation (By similarity). .	*Development, physiology*	MDMD
O95760	*IL33 C9orf26*	FUNCTION: Cytokine that binds to and signals through the IL1RL1/ST2 receptor which in turn activates NF-kappa-B and MAPK signaling pathways in target cells . Involved in the	*Immunity*	WTDML

	IL1F11 NFHEV	maturation of Th2 cells inducing the secretion of T-helper type 2-associated cytokines. Also involved in activation of mast cells, basophils, eosinophils and natural killer cells. Acts as a chemoattractant for Th2 cells, and may function as an 'alarmin', that amplifies immune responses during tissue injury . Induces rapid UCP2-dependent mitochondrial rewiring that attenuates the generation of reactive oxygen species and preserves the integrity of Krebs cycle required for persistent production of itaconate and subsequent GATA3-dependent differentiation of inflammation-resolving alternatively activated macrophages. .; FUNCTION: In quiescent endothelia the uncleaved form is constitutively and abundantly expressed, and acts as a chromatin-associated nuclear factor with transcriptional repressor properties, it may sequester nuclear NF-kappaB/RELA, lowering expression of its targets . This form is rapidely lost upon angiogenic or pro-inflammatory activation		
P01909	*HLA-DQA1*	FUNCTION: Binds peptides derived from antigens that access the endocytic route of antigen presenting cells (APC) and presents them on the cell surface for recognition by the CD4 T-cells. The peptide binding cleft accommodates peptides of 10-30 residues. The peptides presented by MHC class II molecules are generated mostly by degradation of proteins that access the endocytic route, where they are processed by lysosomal proteases and other hydrolases. Exogenous antigens that have been endocytosed by the APC are thus readily available for presentation via MHC II molecules, and for this reason this antigen presentation pathway is usually referred to as exogenous. As membrane proteins on their way to degradation in lysosomes as part of their normal turn-over are also contained in the endosomal/lysosomal compartments, exogenous antigens must compete with those derived from endogenous components. Autophagy is also a source of endogenous peptides, autophagosomes constitutively fuse with MHC class II loading compartments. In addition to APCs, other cells of the gastrointestinal tract, such as epithelial cells, express MHC class II molecules and CD74 and act as APCs, which is an unusual trait of the GI tract. To produce a MHC class II molecule that presents an antigen, three MHC class II molecules (heterodimers of an alpha and a beta chain) associate with a CD74 trimer in the ER to form a heterononamer. Soon after the entry of this complex into the endosomal/lysosomal system where antigen processing occurs, CD74 undergoes a sequential degradation by various	*Immunity*	WTDML MDMD

		proteases, including CTSS and CTSL, leaving a small fragment termed CLIP (class-II-associated invariant chain peptide). The removal of CLIP is facilitated by HLA-DM via direct binding to the alpha-beta-CLIP complex so that CLIP is released. HLA-DM stabilizes MHC class II molecules until primary high affinity antigenic peptides are bound. The MHC II molecule bound to a peptide is then transported to the cell membrane surface. In B-cells, the interaction between HLA-DM and MHC class II molecules is regulated by HLA-DO. Primary dendritic cells (DCs) also to express HLA-DO. Lysosomal microenvironment has been implicated in the regulation of antigen loading into MHC II molecules, increased acidification produces increased proteolysis and efficient peptide loading.		
P01920	*HLA-DQB1 HLA-DQB*	FUNCTION: Binds peptides derived from antigens that access the endocytic route of antigen presenting cells (APC) and presents them on the cell surface for recognition by the CD4 T-cells. The peptide binding cleft accommodates peptides of 10-30 residues. The peptides presented by MHC class II molecules are generated mostly by degradation of proteins that access the endocytic route, where they are processed by lysosomal proteases and other hydrolases. Exogenous antigens that have been endocytosed by the APC are thus readily available for presentation via MHC II molecules, and for this reason this antigen presentation pathway is usually referred to as exogenous. As membrane proteins on their way to degradation in lysosomes as part of their normal turn-over are also contained in the endosomal/lysosomal compartments, exogenous antigens must compete with those derived from endogenous components. Autophagy is also a source of endogenous peptides, autophagosomes constitutively fuse with MHC class II loading compartments. In addition to APCs, other cells of the gastrointestinal tract, such as epithelial cells, express MHC class II molecules and CD74 and act as APCs, which is an unusual trait of the GI tract. To produce a MHC class II molecule that presents an antigen, three MHC class II molecules (heterodimers of an alpha and a beta chain) associate with a CD74 trimer in the ER to form a heterononamer. Soon after the entry of this complex into the endosomal/lysosomal system where antigen processing occurs, CD74 undergoes a sequential degradation by various proteases, including CTSS and CTSL, leaving a small fragment termed CLIP (class-II-associated invariant chain peptide). The removal of CLIP is facilitated by HLA-DM via direct binding to the	*Immunity*	WTDML MDMD

		alpha-beta-CLIP complex so that CLIP is released. HLA-DM stabilizes MHC class II molecules until primary high affinity antigenic peptides are bound. The MHC II molecule bound to a peptide is then transported to the cell membrane surface. In B-cells, the interaction between HLA-DM and MHC class II molecules is regulated by HLA-DO. Primary dendritic cells (DCs) also to express HLA-DO. Lysosomal microenvironment has been implicated in the regulation of antigen loading into MHC II molecules, increased acidification produces increased proteolysis and efficient peptide loading.		
P05107	*ITGB2* *CD18* *MFI7*	FUNCTION: Integrin ITGAL/ITGB2 is a receptor for ICAM1, ICAM2, ICAM3 and ICAM4. Integrin ITGAL/ITGB2 is also a receptor for the secreted form of ubiquitin-like protein ISG15; the interaction is mediated by ITGAL . Integrins ITGAM/ITGB2 and ITGAX/ITGB2 are receptors for the iC3b fragment of the third complement component and for fibrinogen. Integrin ITGAX/ITGB2 recognizes the sequence G-P-R in fibrinogen alpha-chain. Integrin ITGAM/ITGB2 recognizes P1 and P2 peptides of fibrinogen gamma chain. Integrin ITGAM/ITGB2 is also a receptor for factor X. Integrin ITGAD/ITGB2 is a receptor for ICAM3 and VCAM1. Contributes to natural killer cell cytotoxicity . Involved in leukocyte adhesion and transmigration of leukocytes including T-cells and neutrophils . Triggers neutrophil transmigration during lung injury through PTK2B/PYK2-mediated activation . Integrin ITGAL/ITGB2 in association with ICAM3, contributes to apoptotic neutrophil phagocytosis by macrophages . In association with alpha subunit ITGAM/CD11b, required for CD177-PRTN3-mediated activation of TNF primed neutrophils . .	*Immunity*	WTDML
P13762	*HLA-DRB4*	FUNCTION: Binds peptides derived from antigens that access the endocytic route of antigen presenting cells (APC) and presents them on the cell surface for recognition by the CD4 T-cells. The peptide binding cleft accommodates peptides of 10-30 residues. The peptides presented by MHC class II molecules are generated mostly by degradation of proteins that access the endocytic route, where they are processed by lysosomal proteases and other hydrolases. Exogenous antigens that have been endocytosed by the APC are thus readily available for presentation via MHC II molecules, and for this reason this antigen presentation pathway is usually referred to as exogenous. As membrane proteins on their way to degradation in lysosomes as part of their normal turn-over are also contained in the	*Immunity*	WTDKEY MDMD

		endosomal/lysosomal compartments, exogenous antigens must compete with those derived from endogenous components. Autophagy is also a source of endogenous peptides, autophagosomes constitutively fuse with MHC class II loading compartments. In addition to APCs, other cells of the gastrointestinal tract, such as epithelial cells, express MHC class II molecules and CD74 and act as APCs, which is an unusual trait of the GI tract. To produce a MHC class II molecule that presents an antigen, three MHC class II molecules (heterodimers of an alpha and a beta chain) associate with a CD74 trimer in the ER to form a heterononamer. Soon after the entry of this complex into the endosomal/lysosomal system where antigen processing occurs, CD74 undergoes a sequential degradation by various proteases, including CTSS and CTSL, leaving a small fragment termed CLIP (class-II-associated invariant chain peptide). The removal of CLIP is facilitated by HLA-DM via direct binding to the alpha-beta-CLIP complex so that CLIP is released. HLA-DM stabilizes MHC class II molecules until primary high affinity antigenic peptides are bound. The MHC II molecule bound to a peptide is then transported to the cell membrane surface. In B-cells, the interaction between HLA-DM and MHC class II molecules is regulated by HLA-DO. Primary dendritic cells (DCs) also to express HLA-DO. Lysosomal microenvironment has been implicated in the regulation of antigen loading into MHC II molecules, increased acidification produces increased proteolysis and efficient peptide loading.		
P14317	*HCLS1 HS1*	FUNCTION: Substrate of the antigen receptor-coupled tyrosine kinase. Plays a role in antigen receptor signaling for both clonal expansion and deletion in lymphoid cells. May also be involved in the regulation of gene expression.	*Immunity*	MDMD
P20701	*ITGAL CD11A*	FUNCTION: Integrin ITGAL/ITGB2 is a receptor for ICAM1, ICAM2, ICAM3 and ICAM4. Integrin ITGAL/ITGB2 is a receptor for F11R . Integrin ITGAL/ITGB2 is a receptor for the secreted form of ubiquitin-like protein ISG15; the interaction is mediated by ITGAL . Involved in a variety of immune phenomena including leukocyte-endothelial cell interaction, cytotoxic T-cell mediated killing, and antibody dependent killing by granulocytes and monocytes. Contributes to natural killer cell cytotoxicity . Involved in leukocyte adhesion and transmigration of leukocytes including T-cells and neutrophils . Required for generation of common lymphoid progenitor cells in bone marrow, indicating a role in lymphopoiesis (By similarity). Integrin	*Immunity*	WTDML

		ITGAL/ITGB2 in association with ICAM3, contributes to apoptotic neutrophil phagocytosis by macrophages . .		
P26022	*PTX3* *TNFAIP5* *TSG14*	FUNCTION: Plays a role in the regulation of innate resistance to pathogens, inflammatory reactions, possibly clearance of self-components and female fertility. .	*Immunity*	WTDKEY
P31629	*HIVEP2*	FUNCTION: This protein specifically binds to the DNA sequence 5'-GGGACTTTCC-3' which is found in the enhancer elements of numerous viral promoters such as those of SV40, CMV, or HIV1. In addition, related sequences are found in the enhancer elements of a number of cellular promoters, including those of the class I MHC, interleukin-2 receptor, somatostatin receptor II, and interferon-beta genes. It may act in T-cell activation.	*Immunity*	MDMD
P31994	*FCGR2B* *CD32* *FCG2* *IGFR2*	FUNCTION: Receptor for the Fc region of complexed or aggregated immunoglobulins gamma. Low affinity receptor. Involved in a variety of effector and regulatory functions such as phagocytosis of immune complexes and modulation of antibody production by B-cells. Binding to this receptor results in down-modulation of previous state of cell activation triggered via antigen receptors on B-cells (BCR), T-cells (TCR) or via another Fc receptor. Isoform IIB1 fails to mediate endocytosis or phagocytosis. Isoform IIB2 does not trigger phagocytosis.	*Immunity*	MDMD
P32456	*GBP2*	FUNCTION: Interferon (IFN)-inducible GTPase that plays important roles in innate immunity against a diverse range of bacterial, viral and protozoan pathogens . Hydrolyzes GTP to GMP in 2 consecutive cleavage reactions, but the major reaction product is GDP . Following infection, recruited to the pathogen-containing vacuoles or vacuole-escaped bacteria and acts as a positive regulator of inflammasome assembly by promoting the release of inflammasome ligands from bacteria (By similarity). Acts by promoting lysis of pathogen-containing vacuoles, releasing pathogens into the cytosol (By similarity). Following pathogen release in the cytosol, promotes recruitment of proteins that mediate bacterial cytolysis: this liberates ligands that are detected by inflammasomes, such as lipopolysaccharide (LPS) that activates the non-canonical CASP4/CASP11 inflammasome or double-stranded DNA (dsDNA) that activates the AIM2 inflammasome (By similarity). Confers protection to the protozoan pathogen Toxoplasma gondii (By similarity). Independently of its GTPase activity, acts as an inhibitor of various viruses infectivity, such as HIV-1, Zika and	*Immunity*	WTDML

		influenza A viruses, by inhibiting FURIN-mediated maturation of viral envelope proteins . .		
P40305	*IFI27*	FUNCTION: Probable adapter protein involved in different biological processes . Part of the signaling pathways that lead to apoptosis . Involved in type-I interferon-induced apoptosis characterized by a rapid and robust release of cytochrome C from the mitochondria and activation of BAX and caspases 2, 3, 6, 8 and 9 . Also functions in TNFSF10-induced apoptosis . May also have a function in the nucleus, where it may be involved in the interferon-induced negative regulation of the transcriptional activity of NR4A1, NR4A2 and NR4A3 through the enhancement of XPO1-mediated nuclear export of these nuclear receptors . May thereby play a role in the vascular response to injury (By similarity). In the innate immune response, has an antiviral activity towards hepatitis C virus/HCV . May prevent the replication of the virus by recruiting both the hepatitis C virus non-structural protein 5A/NS5A and the ubiquitination machinery via SKP2, promoting the ubiquitin-mediated proteasomal degradation of NS5A . Promotes also virus-induced pyroptosis by activating CASP3 in the mitochondria after 'Lys-6'-linked ubiquitination by TRIM21 . .	*Immunity*	WTDML
P40933	*IL15*	FUNCTION: Cytokine that plays a major role in the development of inflammatory and protective immune responses to microbial invaders and parasites by modulating immune cells of both the innate and adaptive immune systems . Stimulates the proliferation of natural killer cells, T-cells and B-cells and promotes the secretion of several cytokines . In monocytes, induces the production of IL8 and monocyte chemotactic protein 1/CCL2, two chemokines that attract neutrophils and monocytes respectively to sites of infection . Unlike most cytokines, which are secreted in soluble form, IL15 is expressed in association with its high affinity IL15RA on the surface of IL15-producing cells and delivers signals to target cells that express IL2RB and IL2RG receptor subunits . Binding to its receptor triggers the phosphorylation of JAK1 and JAK3 and the recruitment and subsequent phosphorylation of signal transducer and activator of transcription-3/STAT3 and STAT5 . In mast cells, induces the rapid tyrosine phosphorylation of STAT6 and thereby controls mast cell survival and release of cytokines such as IL4 (By similarity). .	*Immunity*	MDMD
P79483	*HLA-DRB3*	FUNCTION: A beta chain of antigen-presenting major histocompatibility complex class II (MHCII) molecule. In complex with the alpha chain HLA-DRA, displays antigenic peptides on professional	*Immunity*	MDMD

antigen presenting cells (APCs) for recognition by alpha-beta T cell receptor (TCR) on HLA-DRB3-restricted CD4-positive T cells. This guides antigen-specific T-helper effector functions, both antibody-mediated immune response and macrophage activation, to ultimately eliminate the infectious agents and transformed cells. Typically presents extracellular peptide antigens of 10 to 30 amino acids that arise from proteolysis of endocytosed antigens in lysosomes . In the tumor microenvironment, presents antigenic peptides that are primarily generated in tumor-resident APCs likely via phagocytosis of apoptotic tumor cells or macropinocytosis of secreted tumor proteins (By similarity). Presents peptides derived from intracellular proteins that are trapped in autolysosomes after macroautophagy, a mechanism especially relevant for T cell selection in the thymus and central immune tolerance (By similarity). The selection of the immunodominant epitopes follows two processing modes: 'bind first, cut/trim later' for pathogen-derived antigenic peptides and 'cut first, bind later' for autoantigens/self-peptides. The anchor residue at position 1 of the peptide N-terminus, usually a large hydrophobic residue, is essential for high affinity interaction with MHCII molecules (By similarity). .; FUNCTION: ALLELE DRB3*01:01: Exclusively presents several immunogenic epitopes derived from C. tetani neurotoxin tetX, playing a significant role in immune recognition and long-term protection . Presents viral epitopes derived from HHV-6B U11, TRX2/U56 and U85 antigens to polyfunctional CD4-positive T cells with cytotoxic activity implicated in control of HHV-6B infection . .; FUNCTION: ALLELE DRB3*02:02 Exclusively presents several immunogenic epitopes derived from C. tetani neurotoxin tetX, playing a significant role in immune recognition and long-term protection . Upon EBV infection, presents to CD4-positive T cells latent antigen EBNA2 (PRSPTVFYNIPPMPLPPSQL) and lytic antigen BZLF1 (LTAYHVSTAPTGSWF) peptides, driving oligoclonal expansion and selection of virus-specific memory T cell subsets with cytotoxic potential to directly eliminate virus-infected B cells . Presents viral epitopes derived from HHV-6B U11, gB/U39 and gH/U48 antigens to polyfunctional CD4-positive T cells with cytotoxic activity implicated in control of HHV-6B infection . Plays a minor role in CD4-positive T cell immune response against Dengue virus by presenting conserved peptides from capsid and non-structural NS3 proteins . Displays peptides

		derived from IAV matrix protein M, implying a role in protection against IAV infection . In the context of tumor immunesurveillance, may present to T-helper 1 cells an immunogenic epitope derived from tumor-associated antigen WT1 (KRYFKLSHLQMHSRKH), likely providing for effective antitumor immunity in a wide range of solid and hematological malignancies . Presents to Vbeta2-positive T-helper 1 cells specifically an immunodominant peptide derived from tumor antigen CTAG1A/NY-ESO-1(PGVLLKEFTVSGNILTIRLTAADHR) and confers protective memory response . In metastatic epithelial tumors, presents to intratumoral CD4-positive T cells a TP53 neoantigen (HYNYMCNSSCMGSMNRRPILTIITL) carrying G245S hotspot driver mutation and may mediate tumor regression . .; FUNCTION: ALLELE DRB3*03:01: Presents a series of conserved peptides derived from the M. tuberculosis PPE family of proteins, in particular PPE29 and PPE33, known to be highly immunogenic . Presents immunogenic epitopes derived from C. tetani neurotoxin tetX, playing a role in immune recognition and long-term protection . Displays immunodominant viral peptides from HCV non-structural protein NS2, as part of a broad range T-helper response to resolve infection . .		
Q01804	*OTUD4* *HIN-1* *KIAA1046*	FUNCTION: Deubiquitinase which hydrolyzes the isopeptide bond between the ubiquitin C-terminus and the lysine epsilon-amino group of the target protein . May negatively regulate inflammatory and pathogen recognition signaling in innate immune response. Upon phosphorylation at Ser-202 and Ser-204 residues, via IL-1 receptor and Toll-like receptor signaling pathway, specifically deubiquitinates 'Lys-63'-polyubiquitinated MYD88 adapter protein triggering down-regulation of NF-kappa-B-dependent transcription of inflammatory mediators . Independently of the catalytic activity, acts as a scaffold for alternative deubiquitinases to assemble specific deubiquitinase-substrate complexes. Associates with USP7 and USP9X deubiquitinases to stabilize alkylation repair enzyme ALKBH3, thereby promoting the repair of alkylated DNA lesions . .	*Immunity*	MDMD
Q14684	*RRP1B* *KIAA0179*	FUNCTION: Positively regulates DNA damage-induced apoptosis by acting as a transcriptional coactivator of proapoptotic target genes of the transcriptional activator E2F1 . Likely to play a role in ribosome biogenesis by targeting serine/threonine protein phosphatase PP1 to the nucleolus . Involved in regulation of mRNA splicing (By similarity). Inhibits SIPA1 GTPase	*Immunity*	MDMD

		activity (By similarity). Involved in regulating expression of extracellular matrix genes (By similarity). Associates with chromatin and may play a role in modulating chromatin structure . .; FUNCTION: (Microbial infection) Following influenza A virus (IAV) infection, promotes viral mRNA transcription by facilitating the binding of IAV RNA-directed RNA polymerase to capped mRNA. .		
Q6P9F5	*TRIM40 RNF35*	FUNCTION: E3 ubiquitin-protein ligase that plays a role in the limitation of the innate immune response . Mediates inhibition of the RLR signaling pathway by ubiquitinating RIGI and IFIH1 receptors, leading to their proteasomal degradation . Promotes also the neddylation of IKBKG/NEMO, stabilizing NFKBIA, and thereby inhibiting of NF-kappa-B nuclear translocation and activation . .	*Immunity*	MDMD
Q6UX01	*LMBR1L KIAA1174 LIMR UNQ458/ PRO783*	FUNCTION: Plays an essential role in lymphocyte development by negatively regulating the canonical Wnt signaling pathway (By similarity). In association with UBAC2 and E3 ubiquitin-protein ligase AMFR, promotes the ubiquitin-mediated degradation of CTNNB1 and Wnt receptors FZD6 and LRP6 (By similarity). LMBR1L stabilizes the beta-catenin destruction complex that is required for regulating CTNNB1 levels (By similarity). Acts as a LCN1 receptor and can mediate its endocytosis . .	*Immunity*	WTDML
Q7L513	*FCRLA FCRL FCRL1 FCRLM1 FCRX FREB UNQ291/ PRO329*	FUNCTION: May be implicated in B-cell differentiation and lymphomagenesis. .	*Immunity*	MDMD
Q86WV6	*STING1 ERIS MITA STING TMEM173*	FUNCTION: Facilitator of innate immune signaling that acts as a sensor of cytosolic DNA from bacteria and viruses and promotes the production of type I interferon (IFN-alpha and IFN-beta) . Innate immune response is triggered in response to non-CpG double-stranded DNA from viruses and bacteria delivered to the cytoplasm . Acts by binding cyclic dinucleotides: recognizes and binds cyclic di-GMP (c-di-GMP), a second messenger produced by bacteria, and cyclic GMP-AMP (cGAMP), a messenger produced by CGAS in response to DNA virus in the cytosol . Upon binding of c-di-GMP or cGAMP, STING1 oligomerizes, translocates from the endoplasmic reticulum and is phosphorylated by TBK1 on the pLxIS motif, leading to recruitment and subsequent activation of the transcription factor	*Immunity*	WTDML

		IRF3 to induce expression of type I interferon and exert a potent anti-viral state . In addition to promote the production of type I interferons, plays a direct role in autophagy . Following cGAMP-binding, STING1 buds from the endoplasmic reticulum into COPII vesicles, which then form the endoplasmic reticulum-Golgi intermediate compartment (ERGIC) . The ERGIC serves as the membrane source for WIPI2 recruitment and LC3 lipidation, leading to formation of autophagosomes that target cytosolic DNA or DNA viruses for degradation by the lysosome . The autophagy- and interferon-inducing activities can be uncoupled and autophagy induction is independent of TBK1 phosphorylation . Autophagy is also triggered upon infection by bacteria: following c-di-GMP-binding, which is produced by live Gram-positive bacteria, promotes reticulophagy (By similarity). Exhibits 2',3' phosphodiester linkage-specific ligand recognition: can bind both 2'-3' linked cGAMP (2'-3'-cGAMP) and 3'-3' linked cGAMP but is preferentially activated by 2'-3' linked cGAMP . The preference for 2'-3'-cGAMP, compared to other linkage isomers is probably due to the ligand itself, whichs adopts an organized free-ligand conformation that resembles the STING1-bound conformation and pays low energy costs in changing into the active conformation . May be involved in translocon function, the translocon possibly being able to influence the induction of type I interferons . May be involved in transduction of apoptotic signals via its association with the major histocompatibility complex class II (MHC-II) (By similarity). .; FUNCTION: (Microbial infection) Antiviral activity is antagonized by oncoproteins, such as papillomavirus (HPV) protein E7 and adenovirus early E1A protein . Such oncoproteins prevent the ability to sense cytosolic DNA . .		
Q8IXQ6	*PARP9 BAL BAL1*	FUNCTION: ADP-ribosyltransferase which, in association with E3 ligase DTX3L, plays a role in DNA damage repair and in immune responses including interferon-mediated antiviral defenses . Within the complex, enhances DTX3L E3 ligase activity which is further enhanced by PARP9 binding to poly(ADP-ribose) . In association with DTX3L and in presence of E1 and E2 enzymes, mediates NAD(+)-dependent mono-ADP-ribosylation of ubiquitin which prevents ubiquitin conjugation to substrates such as histones . During DNA repair, PARP1 recruits PARP9/BAL1-DTX3L complex to DNA damage sites via PARP9 binding to ribosylated PARP1 . Subsequent PARP1-dependent PARP9/BAL1-DTX3L-mediated	*Immunity*	WTDML

		ubiquitination promotes the rapid and specific recruitment of 53BP1/TP53BP1, UIMC1/RAP80, and BRCA1 to DNA damage sites . In response to DNA damage, PARP9-DTX3L complex is required for efficient non-homologous end joining (NHEJ); the complex function is negatively modulated by PARP9 activity . Dispensable for B-cell receptor (BCR) assembly through V(D)J recombination and class switch recombination (CSR) (By similarity). In macrophages, positively regulates pro-inflammatory cytokines production in response to IFNG stimulation by suppressing PARP14-mediated STAT1 ADP-ribosylation and thus promoting STAT1 phosphorylation . Also suppresses PARP14-mediated STAT6 ADP-ribosylation . .		
Q8NHX9	*TPCN2* *TPC2*	FUNCTION: Intracellular channel initially characterized as a non-selective Ca(2+)-permeable channel activated by NAADP (nicotinic acid adenine dinucleotide phosphate), it is also a highly-selective Na(+) channel activated directly by PI(3,5)P2 (phosphatidylinositol 3,5-bisphosphate) . Localizes to the lysosomal and late endosome membranes where it regulates organellar membrane excitability, membrane trafficking, and pH homeostasis. Is associated with a plethora of physiological processes, including mTOR-dependent nutrient sensing, skin pigmentation and autophagy . Ion selectivity is not fixed but rather agonist-dependent and under defined ionic conditions, can be readily activated by both NAADP and PI(3,5)P2 . As calcium channel, it increases the pH in the lysosomal lumen, as sodium channel, it promotes lysosomal exocytosis . Plays a crucial role in endolysosomal trafficking in the endolysosomal degradation pathway and is potentially involved in the homeostatic control of many macromolecules and cell metabolites (By similarity) . Also expressed in melanosomes of pigmented cells where mediates a Ca(2+) channel and/or PI(3,5)P2-activated melanosomal Na(+) channel to acidify pH and inhibit tyrosinase activity required for melanogenesis and pigmentation . Unlike the voltage-dependent TPCN1, TPCN2 is voltage independent and can be activated solely by PI(3,5)P2 binding. In contrast, PI(4,5)P2, PI(3,4)P2, PI(3)P and PI(5)P have no obvious effect on channel activation . .; FUNCTION: (Microbial infection) During Ebola virus (EBOV) infection, controls the movement of endosomes containing virus particles and is required by EBOV to escape from the endosomal network into the cell cytoplasm. .; FUNCTION: (Microbial infection) Required for cell entry of coronaviruses SARS-CoV	*Immunity*	WTDML

		and SARS-CoV-2, as well as human coronavirus EMC (HCoV-EMC), by endocytosis. .		
Q8TDB6	*DTX3L BBAP*	FUNCTION: E3 ubiquitin-protein ligase which, in association with ADP-ribosyltransferase PARP9, plays a role in DNA damage repair and in interferon-mediated antiviral responses . Monoubiquitinates several histones, including histone H2A, H2B, H3 and H4 . In response to DNA damage, mediates monoubiquitination of 'Lys-91' of histone H4 (H4K91ub1) . The exact role of H4K91ub1 in DNA damage response is still unclear but it may function as a licensing signal for additional histone H4 post-translational modifications such as H4 'Lys-20' methylation (H4K20me) . PARP1-dependent PARP9-DTX3L-mediated ubiquitination promotes the rapid and specific recruitment of 53BP1/TP53BP1, UIMC1/RAP80, and BRCA1 to DNA damage sites . By monoubiquitinating histone H2B H2BC9/H2BJ and thereby promoting chromatin remodeling, positively regulates STAT1-dependent interferon-stimulated gene transcription and thus STAT1-mediated control of viral replication . Independently of its catalytic activity, promotes the sorting of chemokine receptor CXCR4 from early endosome to lysosome following CXCL12 stimulation by reducing E3 ligase ITCH activity and thus ITCH-mediated ubiquitination of endosomal sorting complex required for transport ESCRT-0 components HGS and STAM . In addition, required for the recruitment of HGS and STAM to early endosomes . In association with PARP9, plays a role in antiviral responses by mediating 'Lys-48'-linked ubiquitination of encephalomyocarditis virus (EMCV) and human rhinovirus (HRV) C3 proteases and thus promoting their proteasomal-mediated degradation . .	*Immunity*	WTDML
Q8WWU7	*ITLN2 UNQ2789 /PRO7179*	FUNCTION: May play a role in the defense system against pathogens. .	*Immunity*	MDBTD
Q96BN8	*OTULIN FAM105B*	FUNCTION: Deubiquitinase that specifically removes linear ('Met-1'-linked) polyubiquitin chains to substrates and acts as a regulator of angiogenesis and innate immune response . Required during angiogenesis, craniofacial and neuronal development by regulating the canonical Wnt signaling together with the LUBAC complex . Acts as a negative regulator of NF-kappa-B by regulating the activity of the LUBAC complex . OTULIN function is mainly restricted to homeostasis of the LUBAC complex: acts by removing 'Met-1'-linked autoubiquitination of the LUBAC complex, thereby preventing inactivation of the LUBAC complex . Acts as a key negative	*Immunity*	WTDML

		regulator of inflammation by restricting spontaneous inflammation and maintaining immune homeostasis . In myeloid cell, required to prevent unwarranted secretion of cytokines leading to inflammation and autoimmunity by restricting linear polyubiquitin formation . Plays a role in innate immune response by restricting linear polyubiquitin formation on LUBAC complex in response to NOD2 stimulation, probably to limit NOD2-dependent pro-inflammatory signaling . .		
Q96LT7	*C9orf72 DENND9 DENNL72*	FUNCTION: Component of the C9orf72-SMCR8 complex, a complex that has guanine nucleotide exchange factor (GEF) activity and regulates autophagy . In the complex, C9orf72 and SMCR8 probably constitute the catalytic subunits that promote the exchange of GDP to GTP, converting inactive GDP-bound RAB8A and RAB39B into their active GTP-bound form, thereby promoting autophagosome maturation . The C9orf72-SMCR8 complex also acts as a regulator of autophagy initiation by interacting with the ULK1/ATG1 kinase complex and modulating its protein kinase activity . As part of the C9orf72-SMCR8 complex, stimulates RAB8A and RAB11A GTPase activity in vitro . Positively regulates initiation of autophagy by regulating the RAB1A-dependent trafficking of the ULK1/ATG1 kinase complex to the phagophore which leads to autophagosome formation . Acts as a regulator of mTORC1 signaling by promoting phosphorylation of mTORC1 substrates . Plays a role in endosomal trafficking . May be involved in regulating the maturation of phagosomes to lysosomes (By similarity). Promotes the lysosomal localization and lysosome-mediated degradation of CARM1 which leads to inhibition of starvation-induced lipid metabolism (By similarity). Regulates actin dynamics in motor neurons by inhibiting the GTP-binding activity of ARF6, leading to ARF6 inactivation . This reduces the activity of the LIMK1 and LIMK2 kinases which are responsible for phosphorylation and inactivation of cofilin, leading to CFL1/cofilin activation . Positively regulates axon extension and axon growth cone size in spinal motor neurons . Required for SMCR8 protein expression and localization at pre- and post-synaptic compartments in the forebrain, also regulates protein abundance of RAB3A and GRIA1/GLUR1 in post-synaptic compartments in the forebrain and hippocampus (By similarity). Plays a role within the hematopoietic system in restricting inflammation and the development of autoimmunity (By similarity). .; FUNCTION: [Isoform 1]: Regulates stress granule assembly in	*Immunity*	WTDML

		response to cellular stress. .; FUNCTION: [Isoform 2]: Does not play a role in regulation of stress granule assembly in response to cellular stress. .		
Q96PP9	*GBP4*	FUNCTION: Interferon (IFN)-inducible GTPase that plays important roles in innate immunity against a diverse range of bacterial, viral and protozoan pathogens (By similarity). Negatively regulates the antiviral response by inhibiting activation of IRF7 transcription factor (By similarity). .	*Immunity*	WTDML
Q9C030	*TRIM6 RNF89*	FUNCTION: E3 ubiquitin ligase that plays a crucial role in the activation of the IKBKE-dependent branch of the type I interferon signaling pathway . In concert with the ubiquitin-conjugating E2 enzyme UBE2K, synthesizes unanchored 'Lys-48'-linked polyubiquitin chains that promote the oligomerization and autophosphorylation of IKBKE leading to stimulation of an antiviral response . Ubiquitinates also MYC and inhibits its transcription activation activity, maintaining the pluripotency of embryonic stem cells (By similarity). Promotes the association of unanchored 'Lys-48'-polyubiquitin chains with DHX16 leading to enhanced RIGI-mediated innate antiviral immune response . .; FUNCTION: (Microbial infection) Ubiquitinates ebolavirus protein VP35 leading to enhanced viral transcriptase activity. .	*Immunity*	MDBTD
Q9UDY6	*TRIM10 RFB30 RNF9*	FUNCTION: E3 ligase that plays an essential role in the differentiation and survival of terminal erythroid cells. May directly bind to PTEN and promote its ubiquitination, resulting in its proteasomal degradation and activation of hypertrophic signaling (By similarity). In addition, plays a role in immune response regulation by repressing the phosphorylation of STAT1 and STAT2 in the interferon/JAK/STAT signaling pathway independent of its E3 ligase activity. Mechanistically, interacts with the intracellular domain of IFNAR1 and thereby inhibits the association between TYK2 and IFNAR1 . .	*Immunity*	MDMD
O60911	*CTSV CATL2 CTSL2 CTSU UNQ268/ PRO305*	FUNCTION: Cysteine protease. May have an important role in corneal physiology. .	*Immunity, physiology*	MDMD
Q8IW41	*MAPKAPK 5 PRAK*	FUNCTION: Tumor suppressor serine/threonine-protein kinase involved in mTORC1 signaling and post-transcriptional regulation. Phosphorylates FOXO3, ERK3/MAPK6, ERK4/MAPK4, HSP27/HSPB1, p53/TP53 and RHEB. Acts as a tumor suppressor by mediating Ras-induced senescence and phosphorylating p53/TP53. Involved in post-transcriptional regulation of MYC	*Immunity, physiology*	WTDML

		by mediating phosphorylation of FOXO3: phosphorylation of FOXO3 leads to promote nuclear localization of FOXO3, enabling expression of miR-34b and miR-34c, 2 post-transcriptional regulators of MYC that bind to the 3'UTR of MYC transcript and prevent MYC translation. Acts as a negative regulator of mTORC1 signaling by mediating phosphorylation and inhibition of RHEB. Part of the atypical MAPK signaling via its interaction with ERK3/MAPK6 or ERK4/MAPK4: the precise role of the complex formed with ERK3/MAPK6 or ERK4/MAPK4 is still unclear, but the complex follows a complex set of phosphorylation events: upon interaction with atypical MAPK (ERK3/MAPK6 or ERK4/MAPK4), ERK3/MAPK6 (or ERK4/MAPK4) is phosphorylated and then mediates phosphorylation and activation of MAPKAPK5, which in turn phosphorylates ERK3/MAPK6 (or ERK4/MAPK4). Mediates phosphorylation of HSP27/HSPB1 in response to PKA/PRKACA stimulation, inducing F-actin rearrangement. .		
Q9NQZ5	*STARD7 GTT1*	FUNCTION: May play a protective role in mucosal tissues by preventing exaggerated allergic responses. .	*Immunity, physiology*	WTDML
Q16553	*LY6E 9804 RIGE SCA2 TSA1*	FUNCTION: GPI-anchored cell surface protein that regulates T-lymphocytes proliferation, differentiation, and activation. Regulates the T-cell receptor (TCR) signaling by interacting with component CD3Z/CD247 at the plasma membrane, leading to CD3Z/CD247 phosphorylation modulation (By similarity). Restricts the entry of human coronaviruses, including SARS-CoV, MERS-CoV and SARS-CoV-2, by interfering with spike protein-mediated membrane fusion . Also plays an essential role in placenta formation by acting as the main receptor for syncytin-A (SynA). Therefore, participates in the normal fusion of syncytiotrophoblast layer I (SynT-I) and in the proper morphogenesis of both fetal and maternal vasculatures within the placenta. May also act as a modulator of nicotinic acetylcholine receptors (nAChRs) activity (By similarity). .; FUNCTION: (Microbial infection) Promotes entry, likely through an enhanced virus-cell fusion process, of various viruses including HIV-1, West Nile virus, dengue virus and Zika virus . In contrast, the paramyxovirus PIV5, which enters at the plasma membrane, does not require LY6E . Mechanistically, adopts a microtubule-like organization upon viral infection and enhances viral uncoating after endosomal escape . .	*Immunity, reproduction, development*	WTDML
P06858	*LPL LIPD*	FUNCTION: Key enzyme in triglyceride metabolism. Catalyzes the hydrolysis of triglycerides from circulating chylomicrons and	*Metabolism*	WTDKEY

		very low density lipoproteins (VLDL), and thereby plays an important role in lipid clearance from the blood stream, lipid utilization and storage . Although it has both phospholipase and triglyceride lipase activities it is primarily a triglyceride lipase with low but detectable phospholipase activity . Mediates margination of triglyceride-rich lipoprotein particles in capillaries . Recruited to its site of action on the luminal surface of vascular endothelium by binding to GPIHBP1 and cell surface heparan sulfate proteoglycans . .		
P51857	*AKR1D1* *SRD5B1*	FUNCTION: Catalyzes the stereospecific NADPH-dependent reduction of the C4-C5 double bond of bile acid intermediates and steroid hormones carrying a delta(4)-3-one structure to yield an A/B cis-ring junction. This cis-configuration is crucial for bile acid biosynthesis and plays important roles in steroid metabolism. Capable of reducing a broad range of delta-(4)-3-ketosteroids from C18 (such as, 17beta-hydroxyestr-4-en-3-one) to C27 (such as, 7alpha-hydroxycholest-4-en-3-one). .	*Metabolism*	WTDML
Q02318	*CYP27A1* *CYP27*	FUNCTION: Cytochrome P450 monooxygenase that catalyzes regio- and stereospecific hydroxylation of cholesterol and its derivatives. Hydroxylates (with R stereochemistry) the terminal methyl group of cholesterol side-chain in a three step reaction to yield at first a C26 alcohol, then a C26 aldehyde and finally a C26 acid . Regulates cholesterol homeostasis by catalyzing the conversion of excess cholesterol to bile acids via both the 'neutral' (classic) and the 'acid' (alternative) pathways . May also regulate cholesterol homeostasis via generation of active oxysterols, which act as ligands for NR1H2 and NR1H3 nuclear receptors, modulating the transcription of genes involved in lipid metabolism . Plays a role in cholestanol metabolism in the cerebellum. Similarly to cholesterol, hydroxylates cholestanol and may facilitate sterol diffusion through the blood-brain barrier to the systemic circulation for further degradation . Also hydroxylates retinal 7-ketocholesterol, a noxious oxysterol with pro-inflammatory and pro-apoptotic effects, and may play a role in its elimination from the retinal pigment epithelium . May play a redundant role in vitamin D biosynthesis. Catalyzes 25-hydroxylation of vitamin D3 that is required for its conversion to a functionally active form . .	*Metabolism*	WTDML
Q49AA0	*ZFP69* *ZNF642*	FUNCTION: Putative transcription factor that appears to regulate lipid metabolism. .	*Metabolism*	WTDML MDMD

Q8N118	*CYP4X1 UNQ1929 /PRO4404*	FUNCTION: A cytochrome P450 monooxygenase that selectively catalyzes the epoxidation of the last double bond of the arachidonoyl moiety of anandamide, potentially modulating endocannabinoid signaling. Has no hydroxylase activity toward various fatty acids, steroids and prostaglandins. Mechanistically, uses molecular oxygen inserting one oxygen atom into a substrate, and reducing the second into a water molecule, with two electrons provided by NADPH via cytochrome P450 reductase (CPR; NADPH-ferrihemoprotein reductase). .	*Metabolism*	WTDML
Q9BV23	*ABHD6*	FUNCTION: Lipase that preferentially hydrolysis medium-chain saturated monoacylglycerols including 2-arachidonoylglycerol . Through 2-arachidonoylglycerol degradation may regulate endocannabinoid signaling pathways (By similarity). Also has a lysophosphatidyl lipase activity with a preference for lysophosphatidylglycerol among other lysophospholipids (By similarity). Also able to degrade bis(monoacylglycero)phosphate (BMP) and constitutes the major enzyme for BMP catabolism . BMP, also known as lysobisphosphatidic acid, is enriched in late endosomes and lysosomes and plays a key role in the formation of intraluminal vesicles and in lipid sorting . .	*Metabolism*	WTDML
Q9H0X9	*OSBPL5 KIAA1534 OBPH1 ORP5*	FUNCTION: Lipid transporter involved in lipid countertransport between the endoplasmic reticulum and the plasma membrane: specifically exchanges phosphatidylserine with phosphatidylinositol 4-phosphate (PI4P), delivering phosphatidylserine to the plasma membrane in exchange for PI4P, which is degraded by the SAC1/SACM1L phosphatase in the endoplasmic reticulum. Binds phosphatidylserine and PI4P in a mutually exclusive manner . May cooperate with NPC1 to mediate the exit of cholesterol from endosomes/lysosomes . Binds 25-hydroxycholesterol and cholesterol . .	*Metabolism*	MDMD
O94933	*SLITRK3 KIAA0848*	FUNCTION: Suppresses neurite outgrowth. .	*Physiology*	WTDML
P21439	*ABCB4 MDR3 PGY3*	FUNCTION: [Isoform 1]: Energy-dependent phospholipid efflux translocator that acts as a positive regulator of biliary lipid secretion. Functions as a floppase that translocates specifically phosphatidylcholine (PC) from the inner to the outer leaflet of the canalicular membrane bilayer into the canaliculi of hepatocytes. Translocation of PC makes the biliary phospholipids available for extraction into the canaliculi lumen by bile salt mixed micelles and therefore protects the biliary tree from the	*Physiology*	MDMD

		detergent activity of bile salts . Plays a role in the recruitment of phosphatidylcholine (PC), phosphatidylethanolamine (PE) and sphingomyelin (SM) molecules to nonraft membranes and to further enrichment of SM and cholesterol in raft membranes in hepatocytes . Required for proper phospholipid bile formation (By similarity). Indirectly involved in cholesterol efflux activity from hepatocytes into the canalicular lumen in the presence of bile salts in an ATP-dependent manner . Promotes biliary phospholipid secretion as canaliculi-containing vesicles from the canalicular plasma membrane . In cooperation with ATP8B1, functions to protect hepatocytes from the deleterious detergent activity of bile salts . Does not confer multidrug resistance (By similarity). .		
P32238	*CCKAR CCKRA*	FUNCTION: Receptor for cholecystokinin. Mediates pancreatic growth and enzyme secretion, smooth muscle contraction of the gall bladder and stomach. Has a 1000-fold higher affinity for CCK rather than for gastrin. It modulates feeding and dopamine-induced behavior in the central and peripheral nervous system. This receptor mediates its action by association with G proteins that activate a phosphatidylinositol-calcium second messenger system.	*Physiology*	WTDML
Q14916	*SLC17A1 NPT1*	FUNCTION: Important for the resorption of phosphate by the kidney . May be involved in actively transporting phosphate into cells via Na(+) cotransport in the renal brush border membrane . Plays a role in urate transport in the kidney . .	*Physiology*	WTDML
Q8IUK8	*CBLN2 UNQ1892 /PRO4338*	FUNCTION: Acts as a synaptic organizer in specific subsets of neurons in the brain (By similarity). Essential for long-term maintenance but not establishment of excitatory synapses (By similarity). .	*Physiology*	MDBTD
Q8NB66	*UNC13C*	FUNCTION: May play a role in vesicle maturation during exocytosis as a target of the diacylglycerol second messenger pathway. May be involved in the regulation of synaptic transmission at parallel fiber - Purkinje cell synapses (By similarity). .	*Physiology*	MDMD
Q92954	*PRG4 MSF SZP*	FUNCTION: Plays a role in boundary lubrication within articulating joints. Prevents protein deposition onto cartilage from synovial fluid by controlling adhesion-dependent synovial growth and inhibiting the adhesion of synovial cells to the cartilage surface.; FUNCTION: Isoform F plays a role as a growth factor acting on the primitive cells of both hematopoietic and endothelial cell lineages.	*Physiology*	MDMD
Q9H195	*MUC3B*	FUNCTION: Major glycoprotein component of a variety of mucus gels. Thought to provide a	*Physiology*	WTDML

		protective, lubricating barrier against particles and infectious agents at mucosal surfaces (By similarity). .		
Q9UBS5	*GABBR1 GPRC3A*	FUNCTION: Component of a heterodimeric G-protein coupled receptor for GABA, formed by GABBR1 and GABBR2 . Within the heterodimeric GABA receptor, only GABBR1 seems to bind agonists, while GABBR2 mediates coupling to G proteins . Ligand binding causes a conformation change that triggers signaling via guanine nucleotide-binding proteins (G proteins) and modulates the activity of down-stream effectors, such as adenylate cyclase . Signaling inhibits adenylate cyclase, stimulates phospholipase A2, activates potassium channels, inactivates voltage-dependent calcium-channels and modulates inositol phospholipid hydrolysis . Calcium is required for high affinity binding to GABA (By similarity). Plays a critical role in the fine-tuning of inhibitory synaptic transmission . Pre-synaptic GABA receptor inhibits neurotransmitter release by down-regulating high-voltage activated calcium channels, whereas postsynaptic GABA receptor decreases neuronal excitability by activating a prominent inwardly rectifying potassium (Kir) conductance that underlies the late inhibitory postsynaptic potentials . Not only implicated in synaptic inhibition but also in hippocampal long-term potentiation, slow wave sleep, muscle relaxation and antinociception (Probable). Activated by (-)-baclofen, cgp27492 and blocked by phaclofen . .; FUNCTION: Isoform 1E may regulate the formation of functional GABBR1/GABBR2 heterodimers by competing for GABBR2 binding. This could explain the observation that certain small molecule ligands exhibit differential affinity for central versus peripheral sites.	*Physiology*	WTDML
Q9BT56	*SPX C12orf39*	FUNCTION: Plays a role as a central modulator of cardiovascular and renal function and nociception. Also plays a role in energy metabolism and storage. Inhibits adrenocortical cell proliferation with minor stimulation on corticosteroid release (By similarity). .; FUNCTION: [Spexin-1]: Acts as a ligand for galanin receptors GALR2 and GALR3 . Intracerebroventricular administration of the peptide induces an increase in arterial blood pressure, a decrease in both heart rate and renal excretion and delayed natriuresis. Intraventricular administration of the peptide induces antinociceptive activity. Also induces contraction of muscarinic-like stomach smooth muscles. Intraperitoneal administration of the peptide induces a reduction in food consumption and body weight. Inhibits long	*Physiology, metabolism*	MDMD

		chain fatty acid uptake into adipocytes (By similarity). .; FUNCTION: [Spexin-2]: Intracerebroventricular administration of the peptide induces a decrease in heart rate, but no change in arterial pressure, and an increase in urine flow rate. Intraventricular administration of the peptide induces antinociceptive activity (By similarity). .		
P07686	*HEXB* *HCC7*	FUNCTION: Hydrolyzes the non-reducing end N-acetyl-D-hexosamine and/or sulfated N-acetyl-D-hexosamine of glycoconjugates, such as the oligosaccharide moieties from proteins and neutral glycolipids, or from certain mucopolysaccharides . The isozyme B does not hydrolyze each of these substrates, however hydrolyzes efficiently neutral oligosaccharide . Only the isozyme A is responsible for the degradation of GM2 gangliosides in the presence of GM2A . During fertilization is responsible, at least in part, for the zona block to polyspermy. Present in the cortical granules of non-activated oocytes, is exocytosed during the cortical reaction in response to oocyte activation and inactivates the sperm galactosyltransferase-binding site, accounting for the block in sperm binding to the zona pellucida (By similarity). .	*Reproduction*	MDMD
P48378	*RFX2*	FUNCTION: Transcription factor that acts as a key regulator of spermatogenesis. Acts by regulating expression of genes required for the haploid phase during spermiogenesis, such as genes required for cilium assembly and function (By similarity). Recognizes and binds the X-box, a regulatory motif with DNA sequence 5'-GTNRCC(0-3N)RGYAAC-3' present on promoters . Probably activates transcription of the testis-specific histone gene H1-6 (By similarity). .	*Reproduction*	WTDML MDMD
Q07617	*SPAG1*	FUNCTION: May play a role in the cytoplasmic assembly of the ciliary dynein arms (By similarity). May play a role in fertilization. Binds GTP and has GTPase activity. .	*Reproduction*	WTDML
Q5BJF6	*ODF2*	FUNCTION: Seems to be a major component of sperm tail outer dense fibers (ODF). ODFs are filamentous structures located on the outside of the axoneme in the midpiece and principal piece of the mammalian sperm tail and may help to maintain the passive elastic structures and elastic recoil of the sperm tail. May have a modulating influence on sperm motility. Functions as a general scaffold protein that is specifically localized at the distal/subdistal appendages of mother centrioles. Component of the centrosome matrix required for the localization of PLK1 and NIN to the centrosomes. Required for the formation and/or maintenance of normal CETN1 assembly. .	*Reproduction*	MDMD

Q6GV28	*TMEM225 PMP22CD*	FUNCTION: Probably inhibits protein phosphatase 1 (PP1) in sperm via binding to catalytic subunit PPP1CC. .	*Reproduction*	WTDML
Q6HA08	*ASTL*	FUNCTION: Oocyte-specific oolemmal receptor involved in sperm and egg adhesion and fertilization. Plays a role in the polyspermy inhibition. Probably acts as a protease for the post-fertilization cleavage of ZP2. Cleaves the sperm-binding ZP2 at the surface of the zona pellucida after fertilization and cortical granule exocytosis, rendering the zona pellucida unable to support further sperm binding. .	*Reproduction*	WTDML
Q8N6M8	*IQCF1*	FUNCTION: Involved in sperm capacitation and acrosome reaction. .	*Reproduction*	MDBTD
Q9UBK7	*RABL2A*	FUNCTION: Plays an essential role in male fertility, sperm intra-flagellar transport, and tail assembly. Binds, in a GTP-regulated manner, to a specific set of effector proteins including key proteins involved in cilia development and function and delivers them into the growing sperm tail. .	*Reproduction*	WTDML
Q9UKJ8	*ADAM21*	FUNCTION: May be involved in sperm maturation and/or fertilization. May also be involved in epithelia functions associated with establishing and maintaining gradients of ions or nutrients.	*Reproduction*	WTDML
O43323	*DHH*	FUNCTION: [Desert hedgehog protein]: The C-terminal part of the desert hedgehog protein precursor displays an autoproteolysis and a cholesterol transferase activity (By similarity). Both activities result in the cleavage of the full-length protein into two parts (N-product and C-product) followed by the covalent attachment of a cholesterol moiety to the C-terminal of the newly generated N-product (By similarity). Both activities occur in the reticulum endoplasmic (By similarity). Functions in cell-cell mediated juxtacrine signaling . Promotes endothelium integrity . Binds to PTCH1 receptor, which functions in association with smoothened (SMO), to activate the transcription of target genes in endothelial cells . In Schwann cells, controls the development of the peripheral nerve sheath and the transition of mesenchymal cells to form the epithelium-like structure of the perineurial tube (By similarity). .; FUNCTION: [Desert hedgehog protein N-product]: The dually lipidated desert hedgehog protein N-product is essential for a variety of patterning events during development (By similarity). Binds to the patched (PTCH1) receptor, which functions in association with smoothened (SMO), to activate the transcription of target genes . Required for normal testis development and spermatogenesis, namely for the formation of adult-type Leydig cells and normal development of peritubular cells and seminiferous tubules (By similarity). Activates	*Reproduction, development*	WTDML

		primary cilia signaling on neighboring valve interstitial cells through a paracrine mechanism (By similarity). May induce motor neurons in lateral neural tube and may have a polarizing activity . Prevents the desert hedgehog protein precursor binding to PTCH1 . .		
O75752	*B3GALNT1 B3GALT3 UNQ531/PRO1074*	FUNCTION: Transfers N-acetylgalactosamine onto globotriaosylceramide . Plays a critical role in preimplantation stage embryonic development (By similarity). .	*Reproduction, development*	MDBTD
P0CW00	*TSPY8*	FUNCTION: May be involved in sperm differentiation and proliferation. .	*Reproduction, development*	MDMD
P83110	*HTRA3 PRSP*	FUNCTION: Serine protease that cleaves beta-casein/CSN2 as well as several extracellular matrix (ECM) proteoglycans such as decorin/DCN, biglycan/BGN and fibronectin/FN1. Inhibits signaling mediated by TGF-beta family proteins possibly indirectly by degradation of these ECM proteoglycans (By similarity). May act as a tumor suppressor. Negatively regulates, in vitro, trophoblast invasion during placental development and may be involved in the development of the placenta in vivo. May also have a role in ovarian development, granulosa cell differentiation and luteinization . .	*Reproduction, development*	WTDKEY
Q9NX45	*SOHLH2 TEB1*	FUNCTION: Transcription regulator of both male and female germline differentiation. Suppresses genes involved in spermatogonial stem cells maintenance, and induces genes important for spermatogonial differentiation. Coordinates oocyte differentiation without affecting meiosis I (By similarity). .	*Reproduction, development*	MDBTD
O14804	*TAAR5 PNR*	FUNCTION: Olfactory receptor specific for trimethylamine, a trace amine. Also activated at lower level by dimethylethylamine. Trimethylamine is a bacterial metabolite found in some animal odors, and to humans it is a repulsive odor associated with bad breath and spoiled food. This receptor is probably mediated by the G(s)-class of G-proteins which activate adenylate cyclase. .	*Sensory perception*	MDMD
Q13002	*GRIK2 GLUR6*	FUNCTION: Ionotropic glutamate receptor. L-glutamate acts as an excitatory neurotransmitter at many synapses in the central nervous system. Binding of the excitatory neurotransmitter L-glutamate induces a conformation change, leading to the opening of the cation channel, and thereby converts the chemical signal to an electrical impulse. The receptor then desensitizes rapidly and enters a transient inactive state, characterized by the presence of bound agonist . Modulates cell surface expression of NETO2 (By similarity). .; FUNCTION: Independent of its	*Sensory perception*	MDMD

		ionotropic glutamate receptor activity, acts as a thermoreceptor conferring sensitivity to cold temperatures . Functions in dorsal root ganglion neurons (By similarity). .		
Q8NGL1	*OR5D18*	FUNCTION: Odorant receptor. .	*Sensory perception*	WTDML MDMD
Q96RI8	*TAAR6 TA4 TAR4 TRAR4*	FUNCTION: Orphan receptor. Could be a receptor for trace amines. Trace amines are biogenic amines present in very low levels in mammalian tissues. Although some trace amines have clearly defined roles as neurotransmitters in invertebrates, the extent to which they function as true neurotransmitters in vertebrates has remained speculative. Trace amines are likely to be involved in a variety of physiological functions that have yet to be fully understood.	*Sensory perception*	MDMD
Q9H343	*OR51I1*	FUNCTION: Odorant receptor. .	*Sensory perception*	MDMD
Q7Z2W7	*TRPM8 LTRPC6 TRPP8*	FUNCTION: Receptor-activated non-selective cation channel involved in detection of sensations such as coolness, by being activated by cold temperature below 25 degrees Celsius. Activated by icilin, eucalyptol, menthol, cold and modulation of intracellular pH. Involved in menthol sensation. Permeable for monovalent cations sodium, potassium, and cesium and divalent cation calcium. Temperature sensing is tightly linked to voltage-dependent gating. Activated upon depolarization, changes in temperature resulting in graded shifts of its voltage-dependent activation curves. The chemical agonist menthol functions as a gating modifier, shifting activation curves towards physiological membrane potentials. Temperature sensitivity arises from a tenfold difference in the activation energies associated with voltage-dependent opening and closing. In prostate cancer cells, shows strong inward rectification and high calcium selectivity in contrast to its behavior in normal cells which is characterized by outward rectification and poor cationic selectivity. Plays a role in prostate cancer cell migration . Isoform 2 and isoform 3 negatively regulate menthol- and cold-induced channel activity by stabilizing the closed state of the channel. .	*Sensory perception, physiology*	WTDML
A6NP61	*ZAR1L*	FUNCTION: mRNA-binding protein required for maternal mRNA storage, translation and degradation during oocyte maturation (By similarity). Probably promotes formation of some phase-separated membraneless compartment that stores maternal mRNAs in oocytes: acts by undergoing liquid-liquid phase separation upon binding to maternal mRNAs (By similarity). Binds	*Other*	MDMD

		to the 3'-UTR of maternal mRNAs, inhibiting their translation (By similarity). .		
A7E2V4	*ZSWIM8 KIAA0913*	FUNCTION: Substrate recognition component of a SCF-like E3 ubiquitin-protein ligase complex that promotes target-directed microRNA degradation (TDMD), a process that mediates degradation of microRNAs (miRNAs) . The SCF-like E3 ubiquitin-protein ligase complex acts by catalyzing ubiquitination and subsequent degradation of AGO proteins (AGO1, AGO2, AGO3 and/or AGO4), thereby exposing miRNAs for degradation . Specifically recognizes and binds AGO proteins when they are engaged with a TDMD target . May also act as a regulator of axon guidance: specifically recognizes misfolded ROBO3 and promotes its ubiquitination and subsequent degradation . .	*Other*	MDMD
O00764	*PDXK C21orf124 C21orf97 PKH PNK PRED79*	FUNCTION: Catalyzes the phosphorylation of the dietary vitamin B6 vitamers pyridoxal (PL), pyridoxine (PN) and pyridoxamine (PM) to form pyridoxal 5'-phosphate (PLP), pyridoxine 5'-phosphate (PNP) and pyridoxamine 5'-phosphate (PMP), respectively (Probable). PLP is the active form of vitamin B6, and acts as a cofactor for over 140 different enzymatic reactions. .	*Other*	MDMD
O14686	*KMT2D ALR MLL2 MLL4*	FUNCTION: Histone methyltransferase that catalyzes methyl group transfer from S-adenosyl-L-methionine to the epsilon-amino group of 'Lys-4' of histone H3 (H3K4) . Part of chromatin remodeling machinery predominantly forms H3K4me1 methylation marks at active chromatin sites where transcription and DNA repair take place . Acts as a coactivator for estrogen receptor by being recruited by ESR1, thereby activating transcription . .	*Other*	WTDML
O14709	*ZNF197 ZKSCAN9 ZNF166*	FUNCTION: May be involved in transcriptional regulation.	*Other*	WTDML
O14798	*TNFRSF10C DCR1 LIT TRAILR3 TRID UNQ321/PRO366*	FUNCTION: Receptor for the cytotoxic ligand TRAIL. Lacks a cytoplasmic death domain and hence is not capable of inducing apoptosis. May protect cells against TRAIL mediated apoptosis by competing with TRAIL-R1 and R2 for binding to the ligand.	*Other*	MDMD
O15427	*SLC16A3 MCT3 MCT4*	FUNCTION: Proton-dependent transporter of monocarboxylates such as L-lactate and pyruvate . Plays a predominant role in L-lactate efflux from highly glycolytic cells (By similarity). .	*Other*	MDMD
O43150	*ASAP2 DDEF2 KIAA0400*	FUNCTION: Activates the small GTPases ARF1, ARF5 and ARF6. Regulates the formation of post-Golgi vesicles and modulates constitutive secretion. Modulates phagocytosis mediated by	*Other*	WTDML

		Fc gamma receptor and ARF6. Modulates PXN recruitment to focal contacts and cell migration. .		
O43818	*RRP9 RNU3IP2 U355K*	FUNCTION: Component of a nucleolar small nuclear ribonucleoprotein particle (snoRNP) thought to participate in the processing and modification of pre-ribosomal RNA (pre-rRNA) . Part of the small subunit (SSU) processome, first precursor of the small eukaryotic ribosomal subunit. During the assembly of the SSU processome in the nucleolus, many ribosome biogenesis factors, an RNA chaperone and ribosomal proteins associate with the nascent pre-rRNA and work in concert to generate RNA folding, modifications, rearrangements and cleavage as well as targeted degradation of pre-ribosomal RNA by the RNA exosome . .	*Other*	MDBTD
O60343	*TBC1D4 AS160 KIAA0603*	FUNCTION: May act as a GTPase-activating protein for RAB2A, RAB8A, RAB10 and RAB14. Isoform 2 promotes insulin-induced glucose transporter SLC2A4/GLUT4 translocation at the plasma membrane, thus increasing glucose uptake. .	*Other*	WTDML
O75564	*JRK JH8*	FUNCTION: May bind DNA. .	*Other*	MDMD
O75691	*UTP20 DRIM*	FUNCTION: Part of the small subunit (SSU) processome, first precursor of the small eukaryotic ribosomal subunit. During the assembly of the SSU processome in the nucleolus, many ribosome biogenesis factors, an RNA chaperone and ribosomal proteins associate with the nascent pre-rRNA and work in concert to generate RNA folding, modifications, rearrangements and cleavage as well as targeted degradation of pre-ribosomal RNA by the RNA exosome. Involved in 18S pre-rRNA processing. Associates with U3 snoRNA. .	*Other*	WTDML
O75818	*RPP40 RNASEP1*	FUNCTION: Component of ribonuclease P, a ribonucleoprotein complex that generates mature tRNA molecules by cleaving their 5'-ends . Also a component of the MRP ribonuclease complex, which cleaves pre-rRNA sequences . .	*Other*	WTDML
O95391	*SLU7*	FUNCTION: Required for pre-mRNA splicing as component of the spliceosome . Participates in the second catalytic step of pre-mRNA splicing, when the free hydroxyl group of exon I attacks the 3'-splice site to generate spliced mRNA and the excised lariat intron. Required for holding exon 1 properly in the spliceosome and for correct AG identification when more than one possible AG exists in 3'-splicing site region. May be involved in the activation of proximal AG. Probably also involved in alternative splicing regulation. .	*Other*	MDMD
P05060	*CHGB SCG1*	FUNCTION: Secretogranin-1 is a neuroendocrine secretory granule protein, which may be the precursor for other biologically active peptides.	*Other*	MDBTD

P05091	*ALDH2* *ALDM*	FUNCTION: Required for clearance of cellular formaldehyde, a cytotoxic and carcinogenic metabolite that induces DNA damage. .	*Other*	WTDML
P05160	*F13B*	FUNCTION: The B chain of factor XIII is not catalytically active, but is thought to stabilize the A subunits and regulate the rate of transglutaminase formation by thrombin. .	*Other*	MDMD
P07602	*PSAP* *GLBA* *SAP1*	FUNCTION: Saposin-A and saposin-C stimulate the hydrolysis of glucosylceramide by beta-glucosylceramidase (EC 3.2.1.45) and galactosylceramide by beta-galactosylceramidase (EC 3.2.1.46). Saposin-C apparently acts by combining with the enzyme and acidic lipid to form an activated complex, rather than by solubilizing the substrate.; FUNCTION: Saposin-B stimulates the hydrolysis of galacto-cerebroside sulfate by arylsulfatase A (EC 3.1.6.8), GM1 gangliosides by beta-galactosidase (EC 3.2.1.23) and globotriaosylceramide by alpha-galactosidase A (EC 3.2.1.22). Saposin-B forms a solubilizing complex with the substrates of the sphingolipid hydrolases.; FUNCTION: Saposin-D is a specific sphingomyelin phosphodiesterase activator (EC 3.1.4.12).; FUNCTION: [Prosaposin]: Behaves as a myelinotrophic and neurotrophic factor, these effects are mediated by its G-protein-coupled receptors, GPR37 and GPR37L1, undergoing ligand-mediated internalization followed by ERK phosphorylation signaling. .; FUNCTION: Saposins are specific low-molecular mass non-enzymic proteins, they participate in the lysosomal degradation of sphingolipids, which takes place by the sequential action of specific hydrolases.	*Other*	MDBTD
P08237	*PFKM* *PFKX*	FUNCTION: Catalyzes the phosphorylation of D-fructose 6-phosphate to fructose 1,6-bisphosphate by ATP, the first committing step of glycolysis.	*Other*	WTDML
P12270	*TPR*	FUNCTION: Component of the nuclear pore complex (NPC), a complex required for the trafficking across the nuclear envelope. Functions as a scaffolding element in the nuclear phase of the NPC essential for normal nucleocytoplasmic transport of proteins and mRNAs, plays a role in the establishment of nuclear-peripheral chromatin compartmentalization in interphase, and in the mitotic spindle checkpoint signaling during mitosis. Involved in the quality control and retention of unspliced mRNAs in the nucleus; in association with NUP153, regulates the nuclear export of unspliced mRNA species bearing constitutive transport element (CTE) in a NXF1- and KHDRBS1-independent manner. Negatively regulates both the association of CTE-containing mRNA with large polyribosomes and translation	*Other*	MDMD

		initiation. Does not play any role in Rev response element (RRE)-mediated export of unspliced mRNAs. Implicated in nuclear export of mRNAs transcribed from heat shock gene promoters; associates both with chromatin in the HSP70 promoter and with mRNAs transcribed from this promoter under stress-induced conditions. Modulates the nucleocytoplasmic transport of activated MAPK1/ERK2 and huntingtin/HTT and may serve as a docking site for the XPO1/CRM1-mediated nuclear export complex. According to some authors, plays a limited role in the regulation of nuclear protein export . Also plays a role as a structural and functional element of the perinuclear chromatin distribution; involved in the formation and/or maintenance of NPC-associated perinuclear heterochromatin exclusion zones (HEZs). Finally, acts as a spatial regulator of the spindle-assembly checkpoint (SAC) response ensuring a timely and effective recruitment of spindle checkpoint proteins like MAD1L1 and MAD2L1 to unattached kinetochore during the metaphase-anaphase transition before chromosome congression. Its N-terminus is involved in activation of oncogenic kinases. .		
P17516	*AKR1C4* *CHDR*	FUNCTION: Cytosolic aldo-keto reductase that catalyzes the NADH and NADPH-dependent reduction of ketosteroids to hydroxysteroids. Liver specific enzyme that acts as NAD(P)(H)-dependent 3-, 17- and 20-ketosteroid reductase on the steroid nucleus and side chain . Displays the ability to catalyze both oxidation and reduction in vitro, but most probably acts as a reductase in vivo since the oxidase activity measured in vitro is inhibited by physiological concentration of NADPH . Acts preferentially as a 3-alpha-hydroxysteroid dehydrogenase (HSD) with a subsidiary 3-beta-HSD activity . Catalyzes efficiently the transformation of the potent androgen 5-alpha-dihydrotestosterone (5alpha-DHT or 17beta-hydroxy-5alpha-androstan-3-one) into the less active form, 5-alpha-androstan-3-alpha,17-beta-diol (3-alpha-diol) . Catalyzes the reduction of estrone into 17beta-estradiol but with low efficiency . Metabolizes a broad spectrum of natural and synthetic therapeutic steroid and plays an important role in metabolism of androgens, estrogens, progestereone and conjugated steroids . Catalyzes the biotransformation of the pesticide chlordecone (kepone) to its corresponding alcohol leading to increased biliary excretion of the pesticide and concomitant reduction of its neurotoxicity since bile is the major excretory route . .	*Other*	WTDML

P18887	*XRCC1*	FUNCTION: Scaffold protein involved in DNA single-strand break repair by mediating the assembly of DNA break repair protein complexes . Negatively regulates ADP-ribosyltransferase activity of PARP1 during base-excision repair in order to prevent excessive PARP1 activity . Recognizes and binds poly-ADP-ribose chains: specifically binds auto-poly-ADP-ribosylated PARP1, limiting its activity . .	*Other*	MDMD
P20226	*TBP* *GTF2D1* *TF2D* *TFIID*	FUNCTION: The TFIID basal transcription factor complex plays a major role in the initiation of RNA polymerase II (Pol II)-dependent transcription . TFIID recognizes and binds promoters with or without a TATA box via its subunit TBP, a TATA-box-binding protein, and promotes assembly of the pre-initiation complex (PIC) . The TFIID complex consists of TBP and TBP-associated factors (TAFs), including TAF1, TAF2, TAF3, TAF4, TAF5, TAF6, TAF7, TAF8, TAF9, TAF10, TAF11, TAF12 and TAF13 . The TFIID complex structure can be divided into 3 modules TFIID-A, TFIID-B, and TFIID-C . TBP forms the TFIID-A module together with TAF3 and TAF5 . TBP is a general transcription factor that functions at the core of the TFIID complex . During assembly of the core PIC on the promoter, as part of TFIID, TBP binds to and also bends promoter DNA, irrespective of whether the promoter contains a TATA box . Component of a BRF2-containing transcription factor complex that regulates transcription mediated by RNA polymerase III . Component of the transcription factor SL1/TIF-IB complex, which is involved in the assembly of the PIC during RNA polymerase I-dependent transcription . The rate of PIC formation probably is primarily dependent on the rate of association of SL1 with the rDNA promoter . SL1 is involved in stabilization of nucleolar transcription factor 1/UBTF on rDNA . .	*Other*	WTDML
P24385	*CCND1* *BCL1* *PRAD1*	FUNCTION: Regulatory component of the cyclin D1-CDK4 (DC) complex that phosphorylates and inhibits members of the retinoblastoma (RB) protein family including RB1 and regulates the cell-cycle during G(1)/S transition . Phosphorylation of RB1 allows dissociation of the transcription factor E2F from the RB/E2F complex and the subsequent transcription of E2F target genes which are responsible for the progression through the G(1) phase . Hypophosphorylates RB1 in early G(1) phase . Cyclin D-CDK4 complexes are major integrators of various mitogenenic and antimitogenic signals . Also a substrate for SMAD3, phosphorylating SMAD3 in a cell-cycle-dependent manner and repressing its transcriptional activity . Component of the ternary complex, cyclin D1/CDK4/CDKN1B, required for nuclear	*Other*	WTDML

		translocation and activity of the cyclin D-CDK4 complex . Exhibits transcriptional corepressor activity with INSM1 on the NEUROD1 and INS promoters in a cell cycle-independent manner . .		
P28074	*PSMB5 LMPX MB1 X*	FUNCTION: Component of the 20S core proteasome complex involved in the proteolytic degradation of most intracellular proteins. This complex plays numerous essential roles within the cell by associating with different regulatory particles. Associated with two 19S regulatory particles, forms the 26S proteasome and thus participates in the ATP-dependent degradation of ubiquitinated proteins. The 26S proteasome plays a key role in the maintenance of protein homeostasis by removing misfolded or damaged proteins that could impair cellular functions, and by removing proteins whose functions are no longer required. Associated with the PA200 or PA28, the 20S proteasome mediates ubiquitin-independent protein degradation. This type of proteolysis is required in several pathways including spermatogenesis (20S-PA200 complex) or generation of a subset of MHC class I-presented antigenic peptides (20S-PA28 complex). Within the 20S core complex, PSMB5 displays a chymotrypsin-like activity. .	*Other*	MDMD
P30040	*ERP29 C12orf8 ERP28*	FUNCTION: Does not seem to be a disulfide isomerase. Plays an important role in the processing of secretory proteins within the endoplasmic reticulum (ER), possibly by participating in the folding of proteins in the ER.	*Other*	WTDML
P35052	*GPC1*	FUNCTION: Cell surface proteoglycan that bears heparan sulfate. Binds, via the heparan sulfate side chains, alpha-4 (V) collagen and participates in Schwann cell myelination (By similarity). May act as a catalyst in increasing the rate of conversion of prion protein PRPN(C) to PRNP(Sc) via associating (via the heparan sulfate side chains) with both forms of PRPN, targeting them to lipid rafts and facilitating their interaction. Required for proper skeletal muscle differentiation by sequestering FGF2 in lipid rafts preventing its binding to receptors (FGFRs) and inhibiting the FGF-mediated signaling. .	*Other*	MDMD
P35241	*RDX*	FUNCTION: Probably plays a crucial role in the binding of the barbed end of actin filaments to the plasma membrane.	*Other*	WTDML
P35813	*PPM1A PPPM1A*	FUNCTION: Enzyme with a broad specificity. Negatively regulates TGF-beta signaling through dephosphorylating SMAD2 and SMAD3, resulting in their dissociation from SMAD4, nuclear export of the SMADs and termination of the TGF-beta-mediated signaling. Dephosphorylates PRKAA1 and PRKAA2. Plays an important role in the termination of TNF-alpha-mediated NF-kappa-B	*Other*	WTDKEY

		activation through dephosphorylating and inactivating IKBKB/IKKB. .		
P40616	*ARL1*	FUNCTION: GTP-binding protein that recruits several effectors, such as golgins, arfaptins and Arf-GEFs to the trans-Golgi network, and modulates their functions at the Golgi complex . Plays thereby a role in a wide range of fundamental cellular processes, including cell polarity, innate immunity, or protein secretion mediated by arfaptins, which were shown to play a role in maintaining insulin secretion from pancreatic beta cells . .	*Other*	WTDML
P48595	*SERPINB10 PI10*	FUNCTION: Protease inhibitor that may play a role in the regulation of protease activities during hematopoiesis and apoptosis induced by TNF. May regulate protease activities in the cytoplasm and in the nucleus. .	*Other*	WTDML
P48730	*CSNK1D HCKID*	FUNCTION: Essential serine/threonine-protein kinase that regulates diverse cellular growth and survival processes including Wnt signaling, DNA repair and circadian rhythms. It can phosphorylate a large number of proteins. Casein kinases are operationally defined by their preferential utilization of acidic proteins such as caseins as substrates. Phosphorylates connexin-43/GJA1, MAP1A, SNAPIN, MAPT/TAU, TOP2A, DCK, HIF1A, EIF6, p53/TP53, DVL2, DVL3, ESR1, AIB1/NCOA3, DNMT1, PKD2, YAP1, PER1 and PER2. Central component of the circadian clock. In balance with PP1, determines the circadian period length through the regulation of the speed and rhythmicity of PER1 and PER2 phosphorylation. Controls PER1 and PER2 nuclear transport and degradation. YAP1 phosphorylation promotes its SCF(beta-TRCP) E3 ubiquitin ligase-mediated ubiquitination and subsequent degradation. DNMT1 phosphorylation reduces its DNA-binding activity. Phosphorylation of ESR1 and AIB1/NCOA3 stimulates their activity and coactivation. Phosphorylation of DVL2 and DVL3 regulates WNT3A signaling pathway that controls neurite outgrowth. Phosphorylates NEDD9/HEF1 (By similarity). EIF6 phosphorylation promotes its nuclear export. Triggers down-regulation of dopamine receptors in the forebrain. Activates DCK in vitro by phosphorylation. TOP2A phosphorylation favors DNA cleavable complex formation. May regulate the formation of the mitotic spindle apparatus in extravillous trophoblast. Modulates connexin-43/GJA1 gap junction assembly by phosphorylation. Probably involved in lymphocyte physiology. Regulates fast synaptic transmission mediated by glutamate. .	*Other*	MDMD

P49368	*CCT3* *CCTG* *TRIC5*	FUNCTION: Component of the chaperonin-containing T-complex (TRiC), a molecular chaperone complex that assists the folding of proteins upon ATP hydrolysis . The TRiC complex mediates the folding of WRAP53/TCAB1, thereby regulating telomere maintenance . As part of the TRiC complex may play a role in the assembly of BBSome, a complex involved in ciliogenesis regulating transports vesicles to the cilia . The TRiC complex plays a role in the folding of actin and tubulin (Probable). .	*Other*	MDMD
P51587	*BRCA2* *FACD* *FANCD1*	FUNCTION: Involved in double-strand break repair and/or homologous recombination. Binds RAD51 and potentiates recombinational DNA repair by promoting assembly of RAD51 onto single-stranded DNA (ssDNA). Acts by targeting RAD51 to ssDNA over double-stranded DNA, enabling RAD51 to displace replication protein-A (RPA) from ssDNA and stabilizing RAD51-ssDNA filaments by blocking ATP hydrolysis. Part of a PALB2-scaffolded HR complex containing RAD51C and which is thought to play a role in DNA repair by HR. May participate in S phase checkpoint activation. Binds selectively to ssDNA, and to ssDNA in tailed duplexes and replication fork structures. May play a role in the extension step after strand invasion at replication-dependent DNA double-strand breaks; together with PALB2 is involved in both POLH localization at collapsed replication forks and DNA polymerization activity. In concert with NPM1, regulates centrosome duplication. Interacts with the TREX-2 complex (transcription and export complex 2) subunits PCID2 and SEM1, and is required to prevent R-loop-associated DNA damage and thus transcription-associated genomic instability. Silencing of BRCA2 promotes R-loop accumulation at actively transcribed genes in replicating and non-replicating cells, suggesting that BRCA2 mediates the control of R-loop associated genomic instability, independently of its known role in homologous recombination . .	*Other*	MDMD
P52849	*NDST2* *HSST2*	FUNCTION: Essential bifunctional enzyme that catalyzes both the N-deacetylation and the N-sulfation of glucosamine (GlcNAc) of the glycosaminoglycan in heparan sulfate. Modifies the GlcNAc-GlcA disaccharide repeating sugar backbone to make N-sulfated heparosan, a prerequisite substrate for later modifications in heparin biosynthesis. Plays a role in determining the extent and pattern of sulfation of heparan sulfate. Required for the exosomal release of SDCBP, CD63 and syndecan . .	*Other*	MDMD
P53992	*SEC24C* *KIAA0079*	FUNCTION: Component of the coat protein complex II (COPII) which promotes the formation	*Other*	MDMD

		of transport vesicles from the endoplasmic reticulum (ER). The coat has two main functions, the physical deformation of the endoplasmic reticulum membrane into vesicles and the selection of cargo molecules for their transport to the Golgi complex . Plays a central role in cargo selection within the COPII complex and together with SEC24D may have a different specificity compared to SEC24A and SEC24B . May more specifically package GPI-anchored proteins through the cargo receptor TMED10 . May also be specific for IxM motif-containing cargos like the SNAREs GOSR2 and STX5 . .		
P53999	*SUB1 PC4 RPO2TC1*	FUNCTION: General coactivator that functions cooperatively with TAFs and mediates functional interactions between upstream activators and the general transcriptional machinery. May be involved in stabilizing the multiprotein transcription complex. Binds single-stranded DNA. Also binds, in vitro, non-specifically to double-stranded DNA (ds DNA). .	*Other*	WTDKEY
P54619	*PRKAG1*	FUNCTION: AMP/ATP-binding subunit of AMP-activated protein kinase (AMPK), an energy sensor protein kinase that plays a key role in regulating cellular energy metabolism. In response to reduction of intracellular ATP levels, AMPK activates energy-producing pathways and inhibits energy-consuming processes: inhibits protein, carbohydrate and lipid biosynthesis, as well as cell growth and proliferation. AMPK acts via direct phosphorylation of metabolic enzymes, and by longer-term effects via phosphorylation of transcription regulators. Also acts as a regulator of cellular polarity by remodeling the actin cytoskeleton; probably by indirectly activating myosin. Gamma non-catalytic subunit mediates binding to AMP, ADP and ATP, leading to activate or inhibit AMPK: AMP-binding results in allosteric activation of alpha catalytic subunit (PRKAA1 or PRKAA2) both by inducing phosphorylation and preventing dephosphorylation of catalytic subunits. ADP also stimulates phosphorylation, without stimulating already phosphorylated catalytic subunit. ATP promotes dephosphorylation of catalytic subunit, rendering the AMPK enzyme inactive. .	*Other*	WTDML
P54840	*GYS2*	FUNCTION: Transfers the glycosyl residue from UDP-Glc to the non-reducing end of alpha-1,4-glucan.	*Other*	MDMD
P58397	*ADAMTS12 UNQ1918/PRO4389*	FUNCTION: Metalloprotease that may play a role in the degradation of COMP. Cleaves also alpha-2 macroglobulin and aggregan. Has anti-tumorigenic properties. .	*Other*	WTDKEY

P61221	*ABCE1 RLI RNASEL1 RNASELI RNS4I OK/SW-cl.40*	FUNCTION: Nucleoside-triphosphatase (NTPase) involved in ribosome recycling by mediating ribosome disassembly . Able to hydrolyze ATP, GTP, UTP and CTP . Splits ribosomes into free 60S subunits and tRNA- and mRNA-bound 40S subunits . Acts either after canonical termination facilitated by release factors (ETF1/eRF1) or after recognition of stalled and vacant ribosomes by mRNA surveillance factors (PELO/Pelota) . Involved in the No-Go Decay (NGD) pathway: recruited to stalled ribosomes by the Pelota-HBS1L complex, and drives the disassembly of stalled ribosomes, followed by degradation of damaged mRNAs as part of the NGD pathway . Also plays a role in quality control of translation of mitochondrial outer membrane-localized mRNA . As part of the PINK1-regulated signaling, ubiquitinated by CNOT4 upon mitochondria damage; this modification generates polyubiquitin signals that recruit autophagy receptors to the mitochondrial outer membrane and initiate mitophagy . RNASEL-specific protein inhibitor which antagonizes the binding of 2-5A (5'-phosphorylated 2',5'-linked oligoadenylates) to RNASEL . Negative regulator of the anti-viral effect of the interferon-regulated 2-5A/RNASEL pathway . .; FUNCTION: (Microbial infection) May act as a chaperone for post-translational events during HIV-1 capsid assembly. .; FUNCTION: (Microbial infection) Plays a role in the down-regulation of the 2-5A/RNASEL pathway during encephalomyocarditis virus (EMCV) and HIV-1 infections. .	*Other*	MDMD
P68363	*TUBA1B*	FUNCTION: Tubulin is the major constituent of microtubules, a cylinder consisting of laterally associated linear protofilaments composed of alpha- and beta-tubulin heterodimers . Microtubules grow by the addition of GTP-tubulin dimers to the microtubule end, where a stabilizing cap forms . Below the cap, tubulin dimers are in GDP-bound state, owing to GTPase activity of alpha-tubulin . .	*Other*	WTDML
P83436	*COG7 UNQ3082 /PRO10013*	FUNCTION: Required for normal Golgi function. .	*Other*	WTDML
P84074	*HPCA BDR2*	FUNCTION: Calcium-binding protein that may play a role in the regulation of voltage-dependent calcium channels . May also play a role in cyclic-nucleotide-mediated signaling through the regulation of adenylate and guanylate cyclases (By similarity). .	*Other*	WTDML

P86397	*HTD2*	FUNCTION: Mitochondrial 3-hydroxyacyl-thioester dehydratase, which may be involved in fatty acid biosynthesis. .	*Other*	WTDML
Q05923	*DUSP2 PAC1*	FUNCTION: Dephosphorylates both phosphorylated Thr and Tyr residues in MAPK1, and dephosphorylation of phosphotyrosine is slightly faster than that of phosphothreonine . Can dephosphorylate MAPK1 (By similarity). .	*Other*	WTDML
Q13398	*ZNF211*	FUNCTION: May be involved in transcriptional regulation.	*Other*	WTDML
Q13614	*MTMR2 KIAA1073*	FUNCTION: Phosphatase that acts on lipids with a phosphoinositol headgroup. Has phosphatase activity towards phosphatidylinositol 3-phosphate and phosphatidylinositol 3,5-bisphosphate . Binds phosphatidylinositol 4-phosphate, phosphatidylinositol 5-phosphate, phosphatidylinositol 3,5-bisphosphate and phosphatidylinositol 3,4,5-trisphosphate (By similarity). Stabilizes SBF2/MTMR13 at the membranes (By similarity). Specifically in peripheral nerves, stabilizes SBF2/MTMR13 protein (By similarity). .	*Other*	WTDML MDMD
Q14257	*RCN2 ERC55*	FUNCTION: Not known. Binds calcium.	*Other*	WTDML
Q14CX7	*NAA25 C12orf30 MDM20 NAP1*	FUNCTION: Non-catalytic subunit of the NatB complex which catalyzes acetylation of the N-terminal methionine residues of peptides beginning with Met-Asp, Met-Glu, Met-Asn and Met-Gln. May play a role in normal cell-cycle progression. .	*Other*	WTDML
Q15393	*SF3B3 KIAA0017 SAP130*	FUNCTION: Involved in pre-mRNA splicing as a component of the splicing factor SF3B complex, a constituent of the spliceosome . SF3B complex is required for 'A' complex assembly formed by the stable binding of U2 snRNP to the branchpoint sequence (BPS) in pre-mRNA. Sequence independent binding of SF3A/SF3B complex upstream of the branch site is essential, it may anchor U2 snRNP to the pre-mRNA . May also be involved in the assembly of the 'E' complex . As a component of the minor spliceosome, involved in the splicing of U12-type introns in pre-mRNAs (Probable). .	*Other*	WTDML
Q16342	*PDCD2 RP8 ZMYND7*	FUNCTION: May be a DNA-binding protein with a regulatory function. May play an important role in cell death and/or in regulation of cell proliferation.	*Other*	WTDML
Q16690	*DUSP5 VH3*	FUNCTION: Dual specificity protein phosphatase; active with phosphotyrosine, phosphoserine and phosphothreonine residues. The highest relative activity is toward ERK1. .	*Other*	MDMD
Q17RD7	*SYT16 STREP14*	FUNCTION: May be involved in the trafficking and exocytosis of secretory vesicles in non-neuronal tissues. Is Ca(2+)-independent.	*Other*	MDBTD

	SYT14L SYT14R			
Q495W5	*FUT11*	FUNCTION: [Isoform 1]: Has minor fucosyltransferase activity toward biantennary N-glycan acceptors. Does not fucosylate GlcNAc residue within type 2 lactosamine unit. .; FUNCTION: [Isoform 2]: Has fucosyltransferase activity toward biantennary N-glycan acceptors. Does not fucosylate GlcNAc residue within type 2 lactosamine unit. .	*Other*	MDMD
Q53GS7	*GLE1 GLE1L*	FUNCTION: Required for the export of mRNAs containing poly(A) tails from the nucleus into the cytoplasm. May be involved in the terminal step of the mRNA transport through the nuclear pore complex (NPC). .	*Other*	MDMD
Q5SGD2	*PPM1L PP2CE*	FUNCTION: Acts as a suppressor of the SAPK signaling pathways by associating with and dephosphorylating MAP3K7/TAK1 and MAP3K5, and by attenuating the association between MAP3K7/TAK1 and MAP2K4 or MAP2K6. .	*Other*	MDBTD
Q5SWA1	*PPP1R15 B*	FUNCTION: Maintains low levels of EIF2S1 phosphorylation in unstressed cells by promoting its dephosphorylation by PP1. .	*Other*	MDMD
Q5VVC0	*SPO16 C1orf146 SCRE*	FUNCTION: Plays a key role in reinforcing the integrity of the central element of the synaptonemal complex (SC) thereby stabilizing SC, ensuring progression of meiotic prophase I in male and female germ cells (By similarity). Promotes homologous recombination and crossing-over in meiotic prophase I via its association with SHOC1 (By similarity). Required for the localization of TEX11 and MSH4 to recombination intermediates (By similarity). .	*Other*	WTDML
Q641Q2	*WASHC2A FAM21A FAM21B*	FUNCTION: Acts at least in part as component of the WASH core complex whose assembly at the surface of endosomes inhibits WASH nucleation-promoting factor (NPF) activity in recruiting and activating the Arp2/3 complex to induce actin polymerization and is involved in the fission of tubules that serve as transport intermediates during endosome sorting. Mediates the recruitment of the WASH core complex to endosome membranes via binding to phospholipids and VPS35 of the retromer CSC. Mediates the recruitment of the F-actin-capping protein dimer to the WASH core complex probably promoting localized F-actin polymerization needed for vesicle scission. Via its C-terminus binds various phospholipids, most strongly phosphatidylinositol 4-phosphate (PtdIns-(4)P), phosphatidylinositol 5-phosphate (PtdIns-(5)P) and phosphatidylinositol 3,5-bisphosphate (PtdIns-(3,5)P2). Involved in the endosome-to-plasma membrane trafficking and recycling of SNX27-retromer-dependent cargo	*Other*	WTDML

		proteins, such as GLUT1. Required for the association of DNAJC13, ENTR1, ANKRD50 with retromer CSC subunit VPS35. Required for the endosomal recruitment of CCC complex subunits COMMD1 and CCDC93 as well as the retriever complex subunit VPS35L. .		
Q68DI1	*ZNF776*	FUNCTION: May be involved in transcriptional regulation. .	*Other*	WTDKEY WTDML
Q6GYQ0	*RALGAPA 1 GARNL1 KIAA0884 TULIP1*	FUNCTION: Catalytic subunit of the heterodimeric RalGAP1 complex which acts as a GTPase activator for the Ras-like small GTPases RALA and RALB. .	*Other*	WTDKEY
Q6SJ93	*FAM111B CANP*	FUNCTION: Serine protease. .	*Other*	WTDML
Q6UWI2	*PARM1 UNQ1879 /PRO4322*	FUNCTION: May regulate TLP1 expression and telomerase activity, thus enabling certain prostatic cells to resist apoptosis. .	*Other*	WTDML
Q6UWZ7	*ABRAXAS 1 ABRA1 CCDC98 FAM175A UNQ496/ PRO1013*	FUNCTION: Involved in DNA damage response and double-strand break (DSB) repair. Component of the BRCA1-A complex, acting as a central scaffold protein that assembles the various components of the complex and mediates the recruitment of BRCA1. The BRCA1-A complex specifically recognizes 'Lys-63'-linked ubiquitinated histones H2A and H2AX at DNA lesion sites, leading to target the BRCA1-BARD1 heterodimer to sites of DNA damage at DSBs. This complex also possesses deubiquitinase activity that specifically removes 'Lys-63'-linked ubiquitin on histones H2A and H2AX. .	*Other*	WTDKEY
Q6ZXV5	*TMTC3*	FUNCTION: Transfers mannosyl residues to the hydroxyl group of serine or threonine residues. The 4 members of the TMTC family are O-mannosyl-transferases dedicated primarily to the cadherin superfamily, each member seems to have a distinct role in decorating the cadherin domains with O-linked mannose glycans at specific regions. Also acts as O-mannosyl-transferase on other proteins such as PDIA3 . Involved in the positive regulation of proteasomal protein degradation in the endoplasmic reticulum (ER), and the control of ER stress response. .	*Other*	WTDKEY
Q75T13	*PGAP1 UNQ3024 /PRO9822*	FUNCTION: Involved in inositol deacylation of GPI-anchored proteins. GPI inositol deacylation may important for efficient transport of GPI-anchored proteins from the endoplasmic reticulum to the Golgi (By similarity). .	*Other*	MDMD
Q7L590	*MCM10 PRO2249*	FUNCTION: Acts as a replication initiation factor that brings together the MCM2-7 helicase and the DNA polymerase alpha/primase complex in order to initiate DNA replication. Additionally, plays a role in preventing DNA damage during replication. Key effector of the RBBP6 and ZBTB38-mediated regulation of DNA-replication and common fragile	*Other*	WTDML

		sites stability; acts as a direct target of transcriptional repression by ZBTB38 . .		
Q7Z419	*RNF144B IBRDC2 P53RFP*	FUNCTION: E3 ubiquitin-protein ligase which accepts ubiquitin from E2 ubiquitin-conjugating enzymes UBE2L3 and UBE2L6 in the form of a thioester and then directly transfers the ubiquitin to targeted substrates such as LCMT2, thereby promoting their degradation. Induces apoptosis via a p53/TP53-dependent but caspase-independent mechanism. However, its overexpression also produces a decrease of the ubiquitin-dependent stability of BAX, a pro-apoptotic protein, ultimately leading to protection of cell death; But, it is not an anti-apoptotic protein per se. .	*Other*	MDMD
Q7Z5R6	*APBB1IP PREL1 RARP1 RIAM*	FUNCTION: Appears to function in the signal transduction from Ras activation to actin cytoskeletal remodeling. Suppresses insulin-induced promoter activities through AP1 and SRE. Mediates Rap1-induced adhesion. .	*Other*	WTDKEY
Q86XI2	*NCAPG2 LUZP5*	FUNCTION: Regulatory subunit of the condensin-2 complex, a complex which establishes mitotic chromosome architecture and is involved in physical rigidity of the chromatid axis. .	*Other*	MDMD
Q8IUB5	*WFDC13 C20orf138 WAP13*	FUNCTION: Putative acid-stable proteinase inhibitor. .	*Other*	MDBTD
Q8IXH6	*TP53INP2 C20orf110 DOR PINH*	FUNCTION: Dual regulator of transcription and autophagy. Positively regulates autophagy and is required for autophagosome formation and processing. May act as a scaffold protein that recruits MAP1LC3A, GABARAP and GABARAPL2 and brings them to the autophagosome membrane by interacting with VMP1 where, in cooperation with the BECN1-PI3-kinase class III complex, they trigger autophagosome development. Acts as a transcriptional activator of THRA. .	*Other*	WTDML
Q8IXI2	*RHOT1 ARHT1*	FUNCTION: Mitochondrial GTPase involved in mitochondrial trafficking . Probably involved in control of anterograde transport of mitochondria and their subcellular distribution . Promotes mitochondrial fission during high calcium conditions . .	*Other*	WTDML
Q8IYS1	*PM20D2 ACY1L2*	FUNCTION: Catalyzes the peptide bond hydrolysis in dipeptides having basic amino acids lysine, ornithine or arginine at C-terminus. Postulated to function in a metabolite repair mechanism by eliminating alternate dipeptide by-products formed during carnosine synthesis. .	*Other*	MDMD
Q8N0Z2	*ABRA*	FUNCTION: Acts as an activator of serum response factor (SRF)-dependent transcription possibly by inducing nuclear translocation of MKL1 or MKL2	*Other*	WTDML

		and through a mechanism requiring Rho-actin signaling. .		
Q8N3J9	*ZNF664 ZFOC1 ZNF176*	FUNCTION: May be involved in transcriptional regulation.	*Other*	MDMD
Q8N3P4	*VPS8 KIAA0804*	FUNCTION: Plays a role in vesicle-mediated protein trafficking of the endocytic membrane transport pathway. Believed to act as a component of the putative CORVET endosomal tethering complexes which is proposed to be involved in the Rab5-to-Rab7 endosome conversion probably implicating MON1A/B, and via binding SNAREs and SNARE complexes to mediate tethering and docking events during SNARE-mediated membrane fusion. The CORVET complex is proposed to function as a Rab5 effector to mediate early endosome fusion probably in specific endosome subpopulations . Functions predominantly in APPL1-containing endosomes . .	*Other*	MDMD
Q8N465	*D2HGDH D2HGD*	FUNCTION: Catalyzes the oxidation of D-2-hydroxyglutarate (D-2-HG) to alpha-ketoglutarate . Also catalyzes the oxidation of other D-2-hydroxyacids, such as D-malate (D-MAL) and D-lactate (D-LAC) . Exhibits high activities towards D-2-HG and D-MAL but a very weak activity towards D-LAC . .	*Other*	WTDML
Q8N556	*AFAP1 AFAP*	FUNCTION: Can cross-link actin filaments into both network and bundle structures (By similarity). May modulate changes in actin filament integrity and induce lamellipodia formation. May function as an adapter molecule that links other proteins, such as SRC and PKC to the actin cytoskeleton. Seems to play a role in the development and progression of prostate adenocarcinoma by regulating cell-matrix adhesions and migration in the cancer cells. .	*Other*	MDMD
Q8N573	*OXR1 Nbla00307*	FUNCTION: May be involved in protection from oxidative damage. .	*Other*	WTDML
Q8NA72	*POC5 C5orf37*	FUNCTION: Essential for the assembly of the distal half of centrioles, required for centriole elongation. .	*Other*	WTDML MDMD
Q8TDG4	*HELQ HEL308*	FUNCTION: Single-stranded 3'-5' DNA helicase that plays a key role in homology-driven double-strand break (DSB) repair . Involved in different DSB repair mechanisms that are guided by annealing of extensive stretches of complementary bases at break ends, such as microhomology-mediated end-joining (MMEJ), single-strand annealing (SSA) or synthesis-dependent strand annealing (SDSA) . Possesses both DNA unwinding and annealing activities . Forms a complex with RAD51, stimulating HELQ	*Other*	WTDKEY

		DNA helicase activity and ability to unwing DNA . Efficiently unwinds substrates containing 3' overhangs or a D-loop . In contrast, interaction with the replication protein A (RPA/RP-A) complex inhibits DNA unwinding by HELQ but strongly stimulates DNA strand annealing . Triggers displacement of RPA from single-stranded DNA to facilitate annealing of complementary sequences		
Q8WWB7	*GLMP C1orf85 PSEC0030 UNQ2553 /PRO6182*	FUNCTION: Required to protect lysosomal transporter MFSD1 from lysosomal proteolysis and for MFSD1 lysosomal localization. .	*Other*	MDMD
Q8WWQ2	*HPSE2 HPA2*	FUNCTION: Binds heparin and heparan sulfate with high affinity, but lacks heparanase activity. Inhibits HPSE, possibly by competing for its substrates (in vitro). .	*Other*	WTDML
Q8WXF0	*SRSF12 SFRS13B SFRS19 SRRP35*	FUNCTION: Splicing factor that seems to antagonize SR proteins in pre-mRNA splicing regulation. .	*Other*	MDMD
Q8WYH8	*ING5*	FUNCTION: Component of the HBO1 complex, which specifically mediates acetylation of histone H3 at 'Lys-14' (H3K14ac) and, to a lower extent, acetylation of histone H4 . Component of the MOZ/MORF complex which has a histone H3 acetyltransferase activity . Through chromatin acetylation it may regulate DNA replication and may function as a transcriptional coactivator . Inhibits cell growth, induces a delay in S-phase progression and enhances Fas-induced apoptosis in an INCA1-dependent manner . .	*Other*	WTDML
Q969S9	*GFM2 EFG2 MSTP027*	FUNCTION: Mitochondrial GTPase that mediates the disassembly of ribosomes from messenger RNA at the termination of mitochondrial protein biosynthesis. Acts in collaboration with MRRF. GTP hydrolysis follows the ribosome disassembly and probably occurs on the ribosome large subunit. Not involved in the GTP-dependent ribosomal translocation step during translation elongation. .	*Other*	MDMD
Q96A04	*TSACC C1orf182*	FUNCTION: Co-chaperone that facilitates HSP-mediated activation of TSSK6. .	*Other*	MDMD
Q96A23	*CPNE4*	FUNCTION: Probable calcium-dependent phospholipid-binding protein that may play a role in calcium-mediated intracellular processes. .	*Other*	WTDML
Q96B42	*TMEM18*	FUNCTION: Transcription repressor. Sequence-specific ssDNA and dsDNA binding protein, with preference for GCT end CTG repeats. Cell migration modulator which enhances the glioma-specific migration ability of neural stem cells (NSC) and neural precursor cells (NPC). .	*Other*	WTDML MDMD

Q96BM0	*IFI27L1* *FAM14B*	FUNCTION: Plays a role in the apoptotic process and has a pro-apoptotic activity. .	*Other*	WTDML
Q96D46	*NMD3* *CGI-07*	FUNCTION: Acts as an adapter for the XPO1/CRM1-mediated export of the 60S ribosomal subunit. .	*Other*	MDBTD
Q96IF1	*AJUBA* *JUB*	FUNCTION: Adapter or scaffold protein which participates in the assembly of numerous protein complexes and is involved in several cellular processes such as cell fate determination, cytoskeletal organization, repression of gene transcription, mitosis, cell-cell adhesion, cell differentiation, proliferation and migration. Contributes to the linking and/or strengthening of epithelia cell-cell junctions in part by linking adhesive receptors to the actin cytoskeleton. May be involved in signal transduction from cell adhesion sites to the nucleus. Plays an important role in regulation of the kinase activity of AURKA for mitotic commitment. Also a component of the IL-1 signaling pathway modulating IL-1-induced NFKB1 activation by influencing the assembly and activity of the PRKCZ-SQSTM1-TRAF6 multiprotein signaling complex. Functions as an HDAC-dependent corepressor for a subset of GFI1 target genes. Acts as a transcriptional corepressor for SNAI1 and SNAI2/SLUG-dependent repression of E-cadherin transcription. Acts as a hypoxic regulator by bridging an association between the prolyl hydroxylases and VHL enabling efficient degradation of HIF1A. Positively regulates microRNA (miRNA)-mediated gene silencing. Negatively regulates the Hippo signaling pathway and antagonizes phosphorylation of YAP1. .	*Other*	MDMD
Q96IY4	*CPB2*	FUNCTION: Cleaves C-terminal arginine or lysine residues from biologically active peptides such as kinins or anaphylatoxins in the circulation thereby regulating their activities. Down-regulates fibrinolysis by removing C-terminal lysine residues from fibrin that has already been partially degraded by plasmin. .	*Other*	WTDML
Q96JM3	*CHAMP1* *C13orf8* *CAMP* *CHAMP* *KIAA1802* *ZNF828*	FUNCTION: Required for proper alignment of chromosomes at metaphase and their accurate segregation during mitosis. Involved in the maintenance of spindle microtubules attachment to the kinetochore during sister chromatid biorientation. May recruit CENPE and CENPF to the kinetochore. .	*Other*	WTDML
Q96NW7	*LRRC7* *KIAA1365* *LAP1*	FUNCTION: Required for normal synaptic spine architecture and function. Necessary for DISC1 and GRM5 localization to postsynaptic density complexes and for both N-methyl D-aspartate receptor-dependent and metabotropic glutamate receptor-dependent long term depression. .	*Other*	MDMD

Q96QG7	*MTMR9 C8orf9 MTMR8*	FUNCTION: Acts as an adapter for myotubularin-related phosphatases . Increases lipid phosphatase MTMR6 catalytic activity, specifically towards phosphatidylinositol 3,5-bisphosphate and MTMR6 binding affinity for phosphorylated phosphatidylinositols . Positively regulates lipid phosphatase MTMR7 catalytic activity (By similarity). Increases MTMR8 catalytic activity towards phosphatidylinositol 3-phosphate . The formation of the MTMR6-MTMR9 complex, stabilizes both MTMR6 and MTMR9 protein levels . Stabilizes MTMR8 protein levels . Plays a role in the late stages of macropinocytosis possibly by regulating MTMR6-mediated dephosphorylation of phosphatidylinositol 3-phosphate in membrane ruffles . Negatively regulates autophagy, in part via its association with MTMR8 . Negatively regulates DNA damage-induced apoptosis, in part via its association with MTMR6 . Does not bind mono-, di- and tri-phosphorylated phosphatidylinositols, phosphatidic acid and phosphatidylserine . .	*Other*	WTDML
Q99547	*MPHOSPH6 MPP6*	FUNCTION: RNA-binding protein that associates with the RNA exosome complex. Involved in the 3'-processing of the 7S pre-RNA to the mature 5.8S rRNA and play a role in recruiting the RNA exosome complex to pre-rRNA; this function may include C1D. .	*Other*	MDMD
Q9BQE5	*APOL2*	FUNCTION: May affect the movement of lipids in the cytoplasm or allow the binding of lipids to organelles.	*Other*	WTDKEY
Q9BTT4	*MED10 L6 TRG17 TRG20*	FUNCTION: Component of the Mediator complex, a coactivator involved in the regulated transcription of nearly all RNA polymerase II-dependent genes. Mediator functions as a bridge to convey information from gene-specific regulatory proteins to the basal RNA polymerase II transcription machinery. Mediator is recruited to promoters by direct interactions with regulatory proteins and serves as a scaffold for the assembly of a functional preinitiation complex with RNA polymerase II and the general transcription factors.	*Other*	WTDML
Q9BUT1	*BDH2 DHRS6 SDR15C1 UNQ6308 /PRO20933*	FUNCTION: NAD(H)-dependent dehydrogenase/reductase with a preference for cyclic substrates (By similarity). Catalyzes stereoselective conversion of 4-oxo-L-proline to cis-4-hydroxy-L-proline, likely a detoxification mechanism for ketoprolines . Mediates the formation of 2,5-dihydroxybenzoate (2,5-DHBA), a siderophore that chelates free cytoplasmic iron and associates with LCN2, thereby regulating iron transport and homeostasis while protecting cells against free radical-induced oxidative stress. The iron-siderophore complex is imported into	*Other*	WTDML

		mitochondria, providing an iron source for mitochondrial metabolic processes in particular heme synthesis (By similarity). May act as a 3-hydroxybutyrate dehydrogenase . .		
Q9BXR6	*CFHR5* *CFHL5* *FHR5*	FUNCTION: Involved in complement regulation. The dimerized forms have avidity for tissue-bound complement fragments and efficiently compete with the physiological complement inhibitor CFH. .	*Other*	MDMD
Q9BY50	*SEC11C* *SEC11L3* *SPC21* *SPCS4C*	FUNCTION: Catalytic component of the signal peptidase complex (SPC) which catalyzes the cleavage of N-terminal signal sequences from nascent proteins as they are translocated into the lumen of the endoplasmic reticulum . Specifically cleaves N-terminal signal peptides that contain a hydrophobic alpha-helix (h-region) shorter than 18-20 amino acids . .	*Other*	MDMD
Q9BZH6	*WDR11* *BRWD2* *KIAA1351* *WDR15*	FUNCTION: Involved in the Hedgehog (Hh) signaling pathway, is essential for normal ciliogenesis . Regulates the proteolytic processing of GLI3 and cooperates with the transcription factor EMX1 in the induction of downstream Hh pathway gene expression and gonadotropin-releasing hormone production . WDR11 complex facilitates the tethering of Adaptor protein-1 complex (AP-1)-derived vesicles. WDR11 complex acts together with TBC1D23 to facilitate the golgin-mediated capture of vesicles generated using AP-1 . .	*Other*	WTDKEY
Q9H6Z4	*RANBP3*	FUNCTION: Acts as a cofactor for XPO1/CRM1-mediated nuclear export, perhaps as export complex scaffolding protein. Bound to XPO1/CRM1, stabilizes the XPO1/CRM1-cargo interaction. In the absence of Ran-bound GTP prevents binding of XPO1/CRM1 to the nuclear pore complex. Binds to CHC1/RCC1 and increases the guanine nucleotide exchange activity of CHC1/RCC1. Recruits XPO1/CRM1 to CHC1/RCC1 in a Ran-dependent manner. Negative regulator of TGF-beta signaling through interaction with the R-SMAD proteins, SMAD2 and SMAD3, and mediating their nuclear export. .	*Other*	WTDML MDBTD
Q9H765	*ASB8* *PP14212*	FUNCTION: May be a substrate-recognition component of a SCF-like ECS (Elongin-Cullin-SOCS-box protein) E3 ubiquitin-protein ligase complex which mediates the ubiquitination and subsequent proteasomal degradation of target proteins. .	*Other*	WTDML
Q9H777	*ELAC1* *D29*	FUNCTION: Zinc phosphodiesterase, which displays some tRNA 3'-processing endonuclease activity . Specifically involved in tRNA repair: acts downstream of the ribosome-associated quality control (RQC) pathway by removing a 2',3'-cyclic phosphate from tRNAs following cleavage by ANKZF1 . tRNAs are then processed by TRNT1 . .	*Other*	WTDML

Q9H790	*EXO5 C1orf176 DEM1*	FUNCTION: Single-stranded DNA (ssDNA) bidirectional exonuclease involved in DNA repair. Probably involved in DNA repair following ultraviolet (UV) irradiation and interstrand cross-links (ICLs) damage. Has both 5'-3' and 3'-5' exonuclease activities with a strong preference for 5'-ends. Acts as a sliding exonuclease that loads at ssDNA ends and then slides along the ssDNA prior to cutting; however the sliding and the 3'-5' exonuclease activities are abolished upon binding to the replication protein A (RPA) complex that enforces 5'-directionality activity. .	*Other*	WTDML MDMD
Q9H813	*PACC1 C1orf75 TMEM206*	FUNCTION: Proton-activated chloride channel that mediates import of chloride ion in response to extracellular acidic pH . Involved in acidosis-induced cell death by mediating chloride influx and subsequent cell swelling . .	*Other*	MDMD
Q9H9E3	*COG4*	FUNCTION: Required for normal Golgi function . Plays a role in SNARE-pin assembly and Golgi-to-ER retrograde transport via its interaction with SCFD1 . .	*Other*	WTDML
Q9HC07	*TMEM165 TPARL*	FUNCTION: May function as a calcium/proton transporter involved in calcium and in lysosomal pH homeostasis. Therefore, it may play an indirect role in protein glycosylation. .	*Other*	WTDML
Q9NP91	*SLC6A20 SIT1 XT3 XTRP3*	FUNCTION: Mediates the Na(+)- and Cl(-)-dependent uptake of imino acids such as L-proline, N-methyl-L-proline and pipecolate as well as N-methylated amino acids . Also transports glycine, regulates proline and glycine homeostasis in the brain playing a role in the modulation of NMDAR currents . .	*Other*	MDMD
Q9NQV7	*PRDM9 PFM6*	FUNCTION: Histone methyltransferase that sequentially mono-, di-, and tri-methylates both 'Lys-4' (H3K4) and 'Lys-36' (H3K36) of histone H3 to produce respectively trimethylated 'Lys-4' (H3K4me3) and trimethylated 'Lys-36' (H3K36me3) histone H3 and plays a key role in meiotic prophase by determining hotspot localization thereby promoting meiotic recombination . Can also methylate all four core histones with H3 being the best substrate and the most highly modified . Is also able, on one hand, to mono and di-methylate H4K20 and on other hand to trimethylate H3K9 with the di-methylated H3K9 as the best substrate (By similarity). During meiotic prophase, binds specific DNA sequences through its zinc finger domains thereby determining hotspot localization where it promotes local H3K4me3 and H3K36me3 enrichment on the same nucleosomes through its histone methyltransferase activity . Thereby promotes double-stranded breaks (DSB) formation, at this subset of PRDM9-binding sites, that initiates meiotic recombination for the	*Other*	WTDML

		proper meiotic progression (By similarity). During meiotic progression hotspot-bound PRDM9 interacts with several complexes; in early leptonema binds CDYL and EHMT2 followed by EWSR1 and CXXC1 by the end of leptonema. EWSR1 joins PRDM9 with the chromosomal axis through REC8 (By similarity). In this way, controls the DSB repair pathway, pairing of homologous chromosomes and sex body formation (By similarity). Moreover plays a central role in the transcriptional activation of genes during early meiotic prophase thanks to H3K4me3 and H3K36me3 enrichment that represents a specific tag for epigenetic transcriptional activation (By similarity). In addition performs automethylation (By similarity). Acetylation and phosphorylation of histone H3 attenuate or prevent histone H3 methylation (By similarity). .		
Q9NR71	*ASAH2 HNAC1*	FUNCTION: Plasma membrane ceramidase that hydrolyzes sphingolipid ceramides into sphingosine and free fatty acids at neutral pH . Ceramides, sphingosine, and its phosphorylated form sphingosine-1-phosphate are bioactive lipids that mediate cellular signaling pathways regulating several biological processes including cell proliferation, apoptosis and differentiation . Also catalyzes the reverse reaction allowing the synthesis of ceramides from fatty acids and sphingosine . Together with sphingomyelinase, participates in the production of sphingosine and sphingosine-1-phosphate from the degradation of sphingomyelin, a sphingolipid enriched in the plasma membrane of cells . Also participates in the hydrolysis of ceramides from the extracellular milieu allowing the production of sphingosine-1-phosphate inside and outside cells (By similarity). This is the case for instance with the digestion of dietary sphingolipids in the intestinal tract (By similarity). .	*Other*	WTDML
Q9NTJ5	*SACM1L KIAA0851 SAC1*	FUNCTION: Phosphoinositide phosphatase which catalyzes the hydrolysis of phosphatidylinositol 4-phosphate (PtdIns(4)P) . Can also catalyze the hydrolysis of phosphatidylinositol 3-phosphate (PtdIns(3)P) and has low activity towards phosphatidylinositol-3,5-bisphosphate (PtdIns(3,5)P2) (By similarity). Shows a very robust PtdIns(4)P phosphatase activity when it binds PtdIns(4)P in a 'cis' configuration in the cellular environment, with much less activity seen when it binds PtdIns(4)P in 'trans' configuration . PtdIns(4)P phosphatase activity (when it binds PtdIns(4)P in 'trans' configuration) is enhanced in the presence of PLEKHA3 . .	*Other*	MDMD
Q9NUD9	*PIGV*	FUNCTION: Alpha-1,6-mannosyltransferase involved in glycosylphosphatidylinositol-anchor	*Other*	WTDML

		biosynthesis. Transfers the second mannose to the glycosylphosphatidylinositol during GPI precursor assembly. .		
Q9NV58	*RNF19A* *RNF19*	FUNCTION: E3 ubiquitin-protein ligase which accepts ubiquitin from E2 ubiquitin-conjugating enzymes UBE2L3 and UBE2L6 in the form of a thioester and then directly transfers the ubiquitin to targeted substrates, such as SNCAIP or CASR. Specifically ubiquitinates pathogenic SOD1 variants, which leads to their proteasomal degradation and to neuronal protection. .	*Other*	WTDML
Q9NXL6	*SIDT1*	FUNCTION: In vitro binds long double-stranded RNA (dsRNA) (500 and 700 base pairs), but not dsRNA shorter than 300 bp. Not involved in RNA autophagy, a process in which RNA is directly imported into lysosomes in an ATP-dependent manner, and degraded. .	*Other*	WTDML
Q9NYQ8	*FAT2* *CDHF8* *KIAA0811* *MEGF1*	FUNCTION: Involved in the regulation of cell migration . May be involved in mediating the organization of the parallel fibers of granule cells during cerebellar development (By similarity). .	*Other*	WTDML
Q9UGI9	*PRKAG3* *AMPKG3*	FUNCTION: AMP/ATP-binding subunit of AMP-activated protein kinase (AMPK), an energy sensor protein kinase that plays a key role in regulating cellular energy metabolism. In response to reduction of intracellular ATP levels, AMPK activates energy-producing pathways and inhibits energy-consuming processes: inhibits protein, carbohydrate and lipid biosynthesis, as well as cell growth and proliferation. AMPK acts via direct phosphorylation of metabolic enzymes, and by longer-term effects via phosphorylation of transcription regulators. AMPK also acts as a regulator of cellular polarity by remodeling the actin cytoskeleton; probably by indirectly activating myosin. The AMPK gamma3 subunit is a non-catalytic subunit with a regulatory role in muscle energy metabolism . It mediates binding to AMP, ADP and ATP, leading to AMPK activation or inhibition: AMP-binding results in allosteric activation of alpha catalytic subunit (PRKAA1 or PRKAA2) both by inducing phosphorylation and preventing dephosphorylation of catalytic subunits. ADP also stimulates phosphorylation, without stimulating already phosphorylated catalytic subunit. ATP promotes dephosphorylation of catalytic subunit, rendering the AMPK enzyme inactive. .	*Other*	WTDML
Q9UGP8	*SEC63* *SEC63L*	FUNCTION: Mediates cotranslational and post-translational transport of certain precursor polypeptides across endoplasmic reticulum (ER) . Proposed to play an auxiliary role in recognition of precursors with short and apolar signal peptides. May cooperate with SEC62 and	*Other*	WTDML

		HSPA5/BiP to facilitate targeting of small presecretory proteins into the SEC61 channel-forming translocon complex, triggering channel opening for polypeptide translocation to the ER lumen . Required for efficient PKD1/Polycystin-1 biogenesis and trafficking to the plasma membrane of the primary cilia (By similarity). .		
Q9UHC7	*MKRN1* *RNF61*	FUNCTION: E3 ubiquitin ligase catalyzing the covalent attachment of ubiquitin moieties onto substrate proteins. These substrates include FILIP1, p53/TP53, CDKN1A and TERT. Keeps cells alive by suppressing p53/TP53 under normal conditions, but stimulates apoptosis by repressing CDKN1A under stress conditions. Acts as a negative regulator of telomerase. Has negative and positive effects on RNA polymerase II-dependent transcription. .	*Other*	WTDML
Q9UHK6	*AMACR*	FUNCTION: Catalyzes the interconversion of (R)- and (S)-stereoisomers of alpha-methyl-branched-chain fatty acyl-CoA esters . Acts only on coenzyme A thioesters, not on free fatty acids, and accepts as substrates a wide range of alpha-methylacyl-CoAs, including pristanoyl-CoA, trihydroxycoprostanoyl-CoA (an intermediate in bile acid synthesis), and arylpropionic acids like the anti-inflammatory drug ibuprofen (2-(4-isobutylphenyl)propionic acid) but neither 3-methyl-branched nor linear-chain acyl-CoAs . .	*Other*	MDMD
Q9UJA5	*TRMT6* *KIAA1153* *TRM6* *CGI-09*	FUNCTION: Substrate-binding subunit of tRNA (adenine-N(1)-)-methyltransferase, which catalyzes the formation of N(1)-methyladenine at position 58 (m1A58) in initiator methionyl-tRNA . Together with the TRMT61A catalytic subunit, part of a mRNA N(1)-methyltransferase complex that mediates methylation of adenosine residues at the N(1) position of a small subset of mRNAs: N(1) methylation takes place in tRNA T-loop-like structures of mRNAs and is only present at low stoichiometries . .	*Other*	MDBTD
Q9UKN1	*MUC12* *MUC11*	FUNCTION: Involved in epithelial cell protection, adhesion modulation, and signaling. May be involved in epithelial cell growth regulation. Stimulated by both cytokine TNF-alpha and TGF-beta in intestinal epithelium. .	*Other*	WTDML
Q9UMX9	*SLC45A2* *AIM1* *MATP*	FUNCTION: Proton-associated glucose and sucrose transporter (By similarity). May be able to transport also fructose (By similarity). Expressed at a late melanosome maturation stage where functions as proton/glucose exporter which increase lumenal pH by decreasing glycolysis . Regulates melanogenesis by maintaining melanosome neutralization that is initially initiated by transient OCA2 and required for a proper function of the tyrosinase TYR . .	*Other*	MDMD

Q9Y2C5	*SLC17A4*	FUNCTION: Acts as a membrane potential-dependent organic anion transporter, the transport requires a low concentration of chloride ions . Mediates chloride-dependent transport of urate . Mediates sodium-independent high affinity transport of thyroid hormones including L-thyroxine (T4) and 3,3',5-triiodo-L-thyronine (T3) . Can actively transport inorganic phosphate into cells via Na(+) cotransport . .	*Other*	WTDML
Q9Y2W2	*WBP11* *NPWBP* *SIPP1* *SNP70*	FUNCTION: Activates pre-mRNA splicing. May inhibit PP1 phosphatase activity. .	*Other*	WTDML
Q9Y3E0	*GOLT1B* *GCT2* *GOT1A* *CGI-141* *HDCMA39P* *UNQ432/PRO793*	FUNCTION: May be involved in fusion of ER-derived transport vesicles with the Golgi complex.	*Other*	MDMD
Q9Y6F1	*PARP3* *ADPRT3* *ADPRTL3*	FUNCTION: Mono-ADP-ribosyltransferase that mediates mono-ADP-ribosylation of target proteins and plays a key role in the response to DNA damage . Mediates mono-ADP-ribosylation of glutamate, aspartate or lysine residues on target proteins . In contrast to PARP1 and PARP2, it is not able to mediate poly-ADP-ribosylation . Involved in DNA repair by mediating mono-ADP-ribosylation of a limited number of acceptor proteins involved in chromatin architecture and in DNA metabolism, such as histone H2B, XRCC5 and XRCC6 . ADP-ribosylation follows DNA damage and appears as an obligatory step in a detection/signaling pathway leading to the reparation of DNA strand breaks . Involved in single-strand break repair by catalyzing mono-ADP-ribosylation of histone H2B on 'Glu-2' (H2BE2ADPr) of nucleosomes containing nicked DNA . Cooperates with the XRCC5-XRCC6 (Ku80-Ku70) heterodimer to limit end-resection thereby promoting accurate NHEJ . Suppresses G-quadruplex (G4) structures in response to DNA damage . Associates with a number of DNA repair factors and is involved in the response to exogenous and endogenous DNA strand breaks . Together with APLF, promotes the retention of the LIG4-XRCC4 complex on chromatin and accelerate DNA ligation during non-homologous end-joining (NHEJ) . May link the DNA damage surveillance network to the mitotic fidelity checkpoint . Acts as a negative regulator of immunoglobulin class switch recombination,	*Other*	MDBTD

		probably by controlling the level of AICDA /AID on the chromatin (By similarity). In addition to proteins, also able to ADP-ribosylate DNA: mediates DNA mono-ADP-ribosylation of DNA strand break termini via covalent addition of a single ADP-ribose moiety to a 5'- or 3'-terminal phosphate residues in DNA containing multiple strand breaks . .		
Q9Y6R4	*MAP3K4* *KIAA0213* *MAPKKK4* *MEKK4* *MTK1*	FUNCTION: Component of a protein kinase signal transduction cascade. Activates the CSBP2, P38 and JNK MAPK pathways, but not the ERK pathway. Specifically phosphorylates and activates MAP2K4 and MAP2K6. .	*Other*	WTDML
A3QJZ7	*PRAMEF2* *7*			MDMD
A6NFZ4	*FAM24A*			MDBTD
A6NHY2	*ANKDD1B*			WTDML MDMD
H3BQB6	*STMND1*			WTDML
P08243	*ASNS* *TS11*			MDBTD
P15086	*CPB1 CPB* *PCPB*			MDBTD
P15088	*CPA3*			MDBTD
Q14094	*CCNI*			MDMD
Q15434	*RBMS2* *SCR3*			MDMD
Q68CR7	*LRRC66*			WTDML
Q7Z2Y8	*GVINP1* *GVIN1* *VLIG1*			MDBTD WTDKEY
Q8IWF9	*CCDC83* *HSD9*			MDMD
Q8IXL9	*IQCF2*			MDBTD
Q8IZP2	*ST13P4* *FAM10A4*			MDMD
Q8N7B9	*EFCAB3*			WTDML MDMD MDBTD
Q8N7C4	*TMEM217* *C6orf128*			MDMD
Q8N865	*C7orf31*			MDBTD
Q8NCL8	*TMEM116*			WTDML
Q8TBF8	*FAM81A*			WTDML
Q8TEQ0	*SNX29* *RUNDC2A*			WTDML
Q969K7	*TMEM54* *BCLP* *CAC1*			WTDML

Q96BP2	*CHCHD1* *C10orf34* *MRPS37*	MDMD
Q9H425	*C1orf198*	MDMD
Q9H972	*C14orf93*	MDMD
Q9P2S6	*ANKMY1* *TSAL1* *ZMYND13*	MDMD
Q9UKR5	*ERG28* *C14orf1* *AD-011* *HSPC288* *x0006*	WTDKEY
Q9Y3D5	*MRPS18C* *CGI-134*	WTDKEY
Q9Y6X4	*FAM169A* *KIAA0888*	MDBTD

Appendix III: Supplementary material for Chapter 4

Supplemental data

Taxonomy

Genus Odocoileus Rafinesque, 1832

Types Species: *Odocoileus speleus* Rafinesque, 1832 (= *Dama virginianus* Zimmermann, 1780).

Contents: Two subgenera, *Odocoileus* Rafinesque, 1832 and *Torontoceros* Churcher and Peterson, 1982.

Description: Medium-sized deer with antlers that are cylindrical and not flattened.

Subgenus Odocoileus Rafinesque, 1832

Types Species: *Odocoileus speleus* Rafinesque, 1832 (= *Dama virginianus* Zimmermann, 1780).

Contents: Two species, *hemionus* Rafinesque, 1817; and *virginianus* Zimmermann, 1780.

Description: The pelage is typically grayish brown with antlers that are upright and rugose with repeated forks and no brow tine (Churcher and Peterson, 1982).

Subgenus Torontoceros Churcher & Peterson, 1982

Type Species: *Torontoceros hypogaeus* Churcher and Peterson, 1982

Contents: One species, extinct.

Description: Known only by the holotype specimen, which is “a damaged braincase with main beams of antlers and bases of brow and bez tines” (Churcher and Peterson, 1982).

Figures

Figure S4.1: Specimen ROMM75974 from the Royal Ontario Museum, Toronto, additional pictures.

Figure S4.2: Calibrated whole genome phylogeny of Cervidae produced by SNAPP, nodes are labelled with the estimated divergence time in the past, colour represents posterior distribution support (white ≤ 70%, grey 71 - 89%, black ≥ 90%), square nodes show calibration points and blue bars represent the 95% HPD intervals, x-axis timeline in units of mya. Tree visualisation performed in FigTree and edited in Inkscape. (A) minimal LD filter dataset, transversions only (3670 sites), (B) minimal LD and depth filter dataset (8551 sites), (C) minimal LD and depth filter dataset, transversions only (2648 sites).

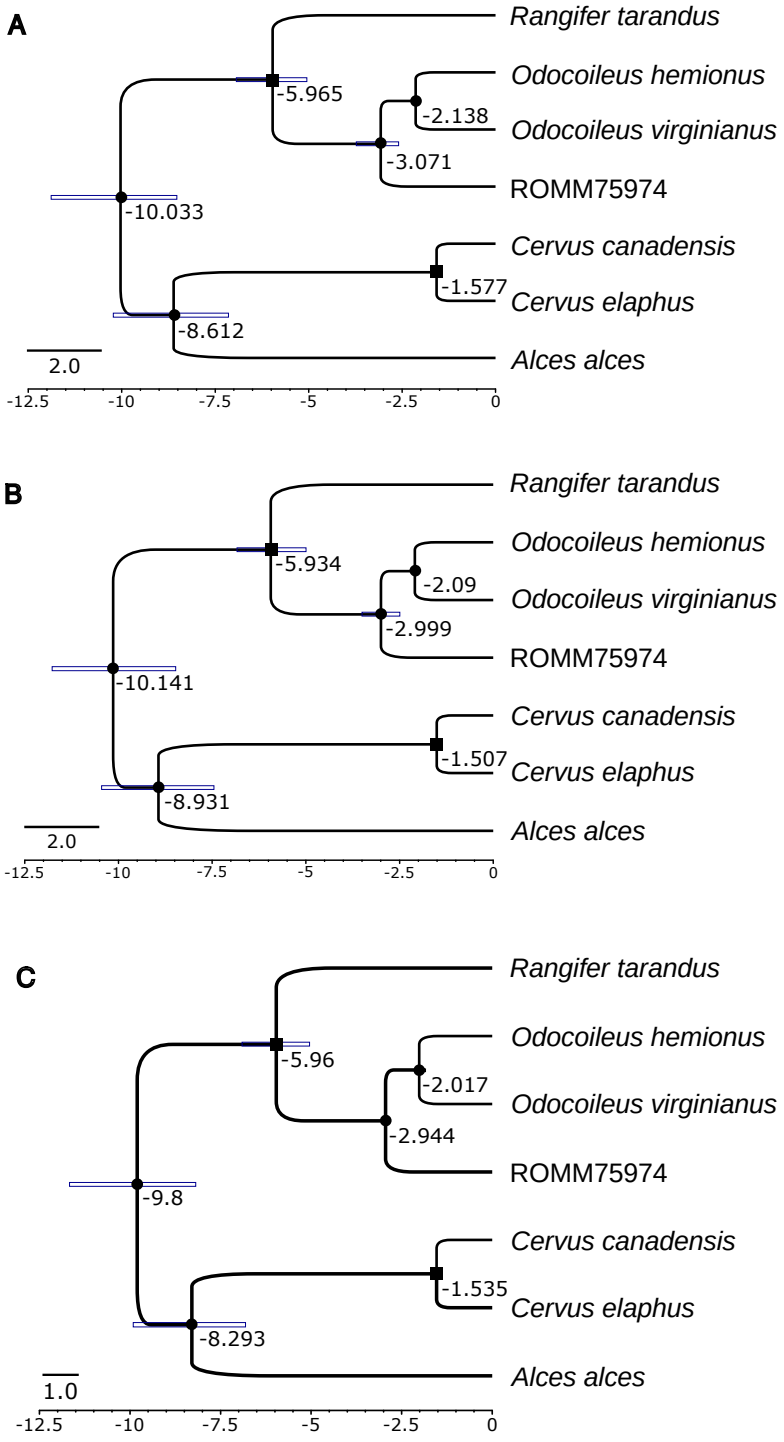


Figure S4.3: Cervidae phylogenies performed in IQtree of (A) *CYTB*, (B) *COX1* and (C) partial mitochondrial genome in the form of collated sequences of 12 genes. Includes ROMM75974, 17 other Cervidae species and cattle as outgroup, node colour represents bootstrap support (white ≤ 70%, grey 71 - 89%, black ≥ 90%).

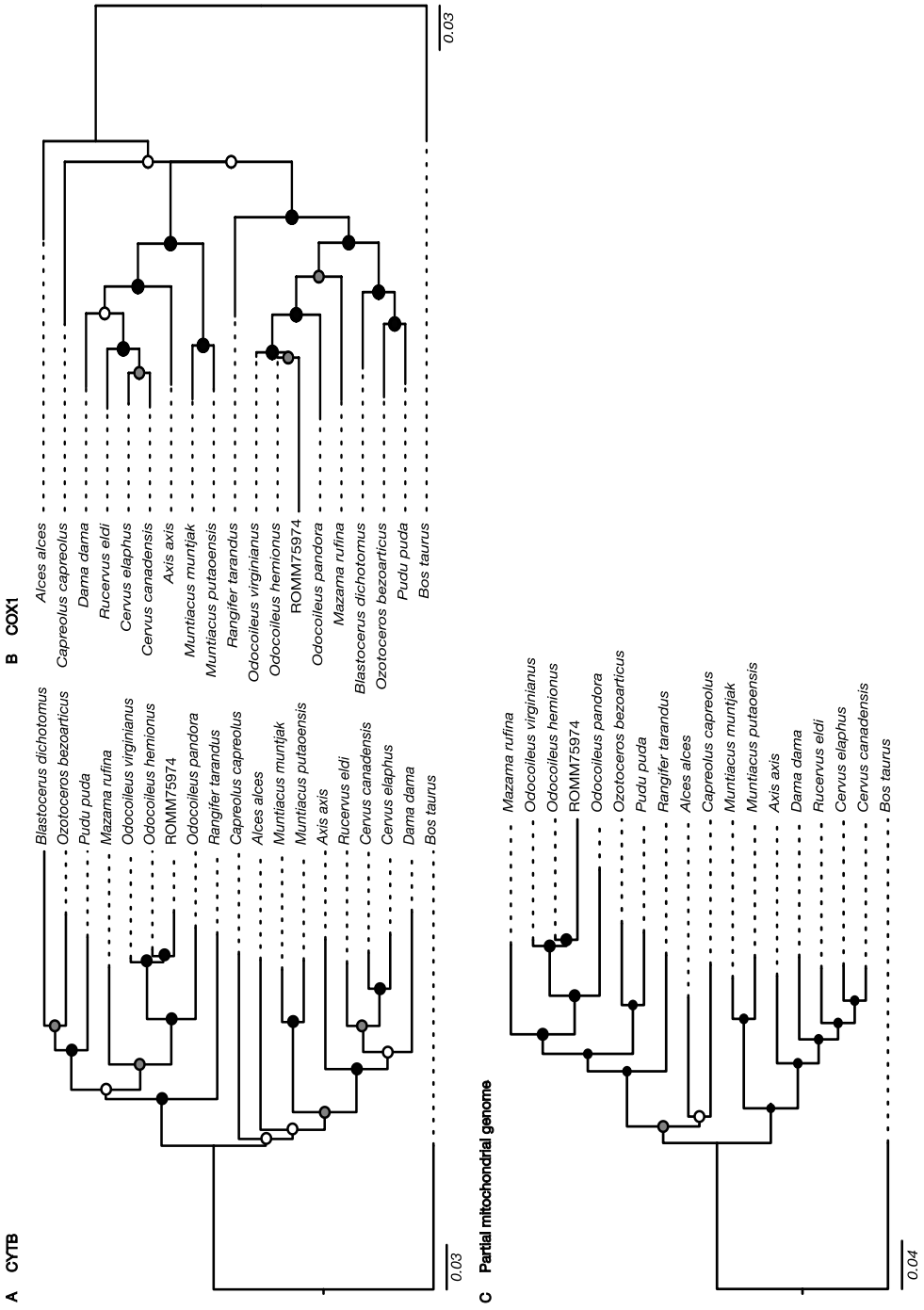


Figure S4.4: Allelic distance matrix heatmap and clustering based on minimal LD and depth filter dataset (8551 sites) of Cervidae, side colour represents the species as abbreviated in sample name: Aa = *Alces alces*, Cc = *Cervus canadensis*, Ce = *Cervus elaphus*, Oh = *Odocoileus hemionus*, Ov = *Odocoileus virginianus*, Rt = *Rangifer tarandus*.

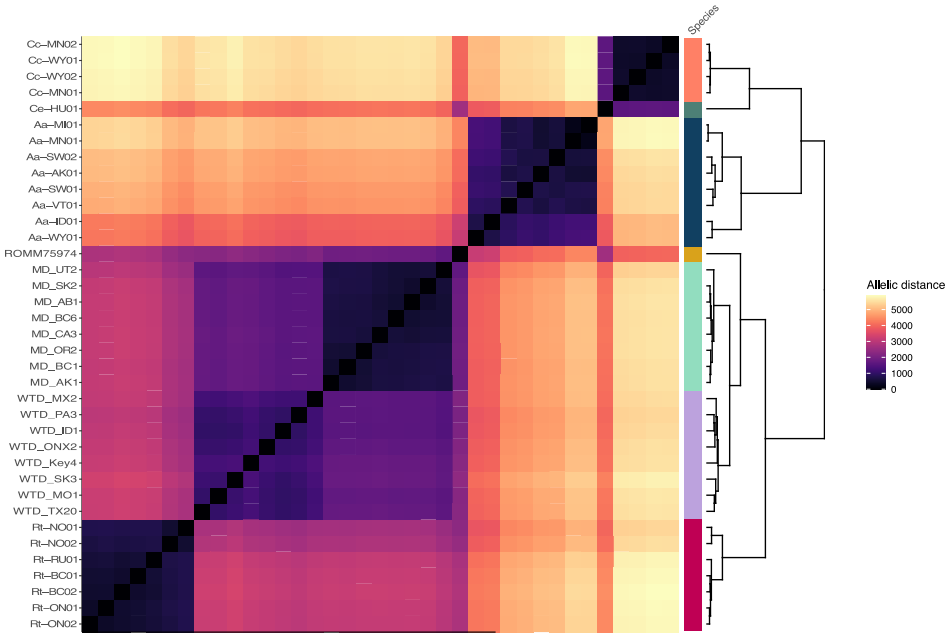


Figure S4.5: Principal component analysis based on allele frequencies of ROMM75974, caribou, white-tailed and mule deer samples, all mapped to caribou reference genome. Based on a total of 94520 sites without missing data.

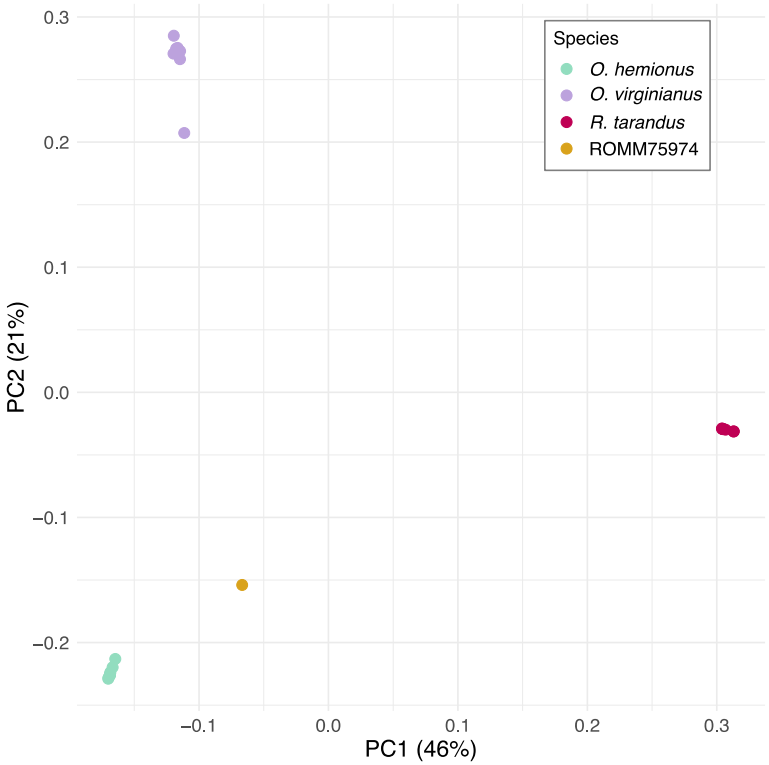


Figure S4.6: Approximate sampling locations, coloured by species.

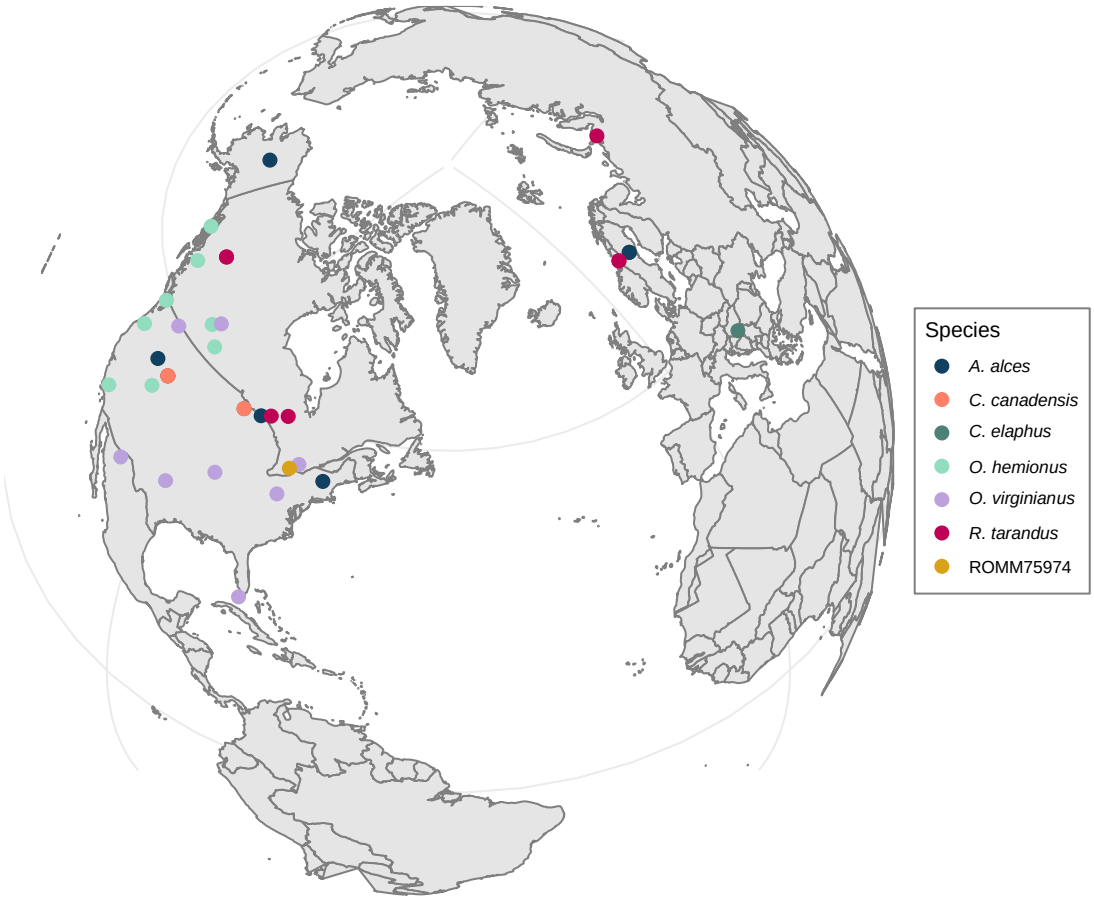
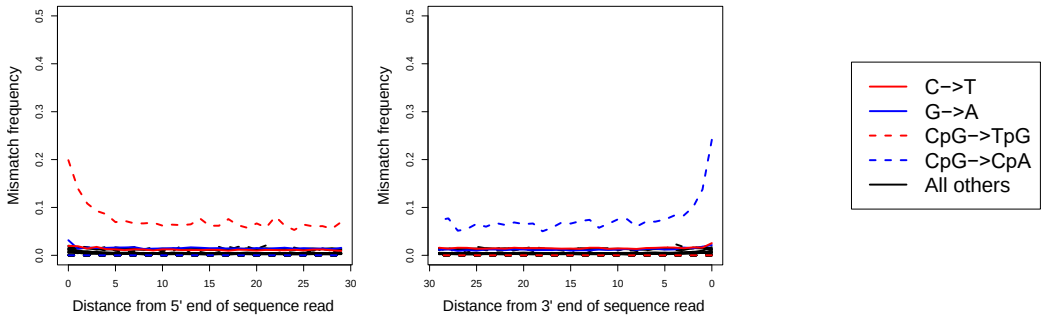


Figure S4.7: ROMM75974 damage patterns PMDtools when mapped to caribou. Deamination at CpG sites identified by PDMtools is unaffected by USER enzyme treatment.



Tables

Table S4.1: mini literature review of papers investigating extinction and including or excluding *Torontoceros hypogeaus*.

List of extinct NA mammals	Examples
Included	Barnosky, A.D., Koch, P.L., Feranec, R.S., Wing, S.L., and Shabel, A.B. (2004). Assessing the Causes of Late Pleistocene Extinctions on the Continents. Science 306, 70–75. 10.1126/science.1101476. Elias, S., and Schreve, D. (2007). Late Pleistocene Megafaunal Extinctions. In Encyclopedia of Quaternary Science. (Elsevier), pp. 3202–3217. 10.1016/B0-444-52747-8/00266-0. Jackson, L.J. (1988). Fossil Cervids and Fluted Point Hunters: A Review for Southern Ontario. Ontario Archaeology OA48, 15. Koch, P.L., and Barnosky, A.D. (2006). Late Quaternary Extinctions: State of the Debate. Annu. Rev. Ecol. Evol. Syst. 37, 215–250. 10.1146/annurev.ecolsys.34.011802.132415. Martin, P.S., and Steadman, D.W. (1999). Prehistoric Extinctions on Islands and Continents. In Extinctions in Near Time, R. D. E. MacPhee, ed. (Springer US), pp. 17–55. 10.1007/978-1-4757-5202-1_2. Meltzer, D. J., & Mead, J. I. (1985). Dating late Pleistocene extinctions: theoretical issues, analytical bias, and substantive results. Environments and Extinctions: Man in Late Glacial North America, 145-173. Spaulding, W.G. (1983). The Overkill Hypothesis as a Plausible Explanation for the Extinctions of Late Wisconsin Megafauna. Quat. res. 20, 110–112. 10.1016/0033-5894(83)90069-8. Faith, J.T., and Surovell, T.A. (2009). Synchronous extinction of North America's Pleistocene mammals. Proc. Natl. Acad. Sci. U.S.A. 106, 20641–20645. 10.1073/pnas.0908153106.
Excluded	Grayson, D.K. (2007). Deciphering North American Pleistocene Extinctions. Journal of Anthropological Research 63, 185–213. 10.3998/jar.0521004.0063.205. Grayson, D.K. (1991). Late Pleistocene mammalian extinctions in North America: Taxonomy, chronology, and explanations. J World Prehist 5, 193–231. 10.1007/BF00974990. Meltzer, D.J. (2020). Overkill, glacial history, and the extinction of North America's Ice Age megafauna. Proc. Natl. Acad. Sci. U.S.A. 117, 28555–28563. 10.1073/pnas.2015032117. Stewart, M., Carleton, W.C., and Groucutt, H.S. (2021). Climate change, not human population growth, correlates with Late Quaternary megafauna declines in North America. Nat Commun 12, 965. 10.1038/s41467-021-21201-8. Stuart, A.J. (2015). Late Quaternary megafaunal extinctions on the continents: a short review. Geol. J. 50, 338–363. 10.1002/gj.2633.

Table S4.2: MitoGeneExtractor mapping statistics for each mitochondrial gene, the gene was considered successfully extracted if mean coverage > 1.

Gene	Reference accession	Aligned reads	Alignment length	Mean coverage	Median coverage	Max coverage
ATP6	YP_209210.1	19	678	2.33	2	9
ATP8	YP_209209.1	8	198	3.38	4	6
COX1	YP_209207.1	164	1542	9.34	5	58
COX2	YP_209208.1	19	681	2.21	2	6
COX3	YP_209211.1	39	780	4.13	3	15
CYTB	YP_209217.1	42	1137	3	2	9
NADH1	YP_209205.1	231	954	23.19	7	156
NADH2	YP_209206.1	17	1041	1.34	1	4
NADH3	YP_209212.2	13	345	3.11	2	9
NADH4	YP_209214.1	81	1377	5.38	3	37
NADH4L	YP_209213.1	1	294	0.26	0	1
NADH5	YP_209215.1	124	1818	6.21	3	47
NADH6	YP_209216.1	7	525	1.18	1	3

Table S.4.3: 20 blast best hits for each of the 12 mitochondrial genes recovered in MitoGeneExtractor.

ATP6								
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. ident	Acc. Len	Accession
Odocoileus virginianus isolate Ovmex31 mitochondrion, complete genome	Odocoileus virginianus	416	416	0.34	0	99.13	16477	KM612278.1
Odocoileus virginianus isolate Ovtol1 mitochondrion, complete genome	Odocoileus virginianus	416	416	0.34	0	99.13	16477	KM612276.1
Odocoileus virginianus isolate Ovcou3 mitochondrion, complete genome	Odocoileus virginianus	416	416	0.34	0	99.13	16479	KM612275.1
Odocoileus virginianus isolate Ovsin17 mitochondrion, complete genome	Odocoileus virginianus	416	416	0.34	0	99.13	16478	KM612274.1
Odocoileus virginianus isolate Ovtex439 mitochondrion, complete genome	Odocoileus virginianus	411	411	0.34	0	98.7	16477	KM612273.1
Odocoileus virginianus isolate Ovver13 mitochondrion, complete genome	Odocoileus virginianus	411	411	0.34	0	98.7	16478	KM612271.1
Odocoileus hemionus isolate T1766 mitochondrion, complete genome	Odocoileus hemionus	411	559	0.63	0	98.7	16482	NC_020729.1
Odocoileus virginianus mitochondrion, complete genome	Odocoileus virginianus	407	407	0.34	0	98.27	16477	NC_015247.1
Odocoileus virginianus isolate CYTO mitochondrion, complete genome	Odocoileus virginianus	399	399	0.34	0	97.84	16480	JN632672.1
Odocoileus virginianus isolate OvacpCM mitochondrion, complete genome	Odocoileus virginianus	394	394	0.34	0	97.4	16471	KM612279.1
Odocoileus virginianus isolate Ovoax2 mitochondrion, complete genome	Odocoileus virginianus	394	394	0.34	0	97.4	16470	KM612277.1
Odocoileus virginianus isolate Ovyuc61 mitochondrion, complete genome	Odocoileus virginianus	394	394	0.34	0	97.4	16469	KM612272.1
Odocoileus virginianus isolate T4887 mitochondrion, complete genome	Odocoileus virginianus	372	372	0.34	0	95.67	16465	JN632671.1
Mazama americana voucher NUPECCE T297 mitochondrion, complete genome	Mazama americana	355	355	0.34	0	94.37	16474	OQ731411.1

Odocoileus pandora voucher NUPECCE T365 topotype mitochondrion, complete genome	Odocoileus pandora	350	350	0.34	0	93.94	16474	OQ731410.1
Odocoileus pandora voucher NUPECCE MX06 mitochondrion, complete genome	Odocoileus pandora	350	350	0.34	0	93.94	16418	OQ731409.1
Odocoileus pandora voucher NUPECCE MX02 mitochondrion, complete genome	Odocoileus pandora	350	350	0.34	0	93.94	16362	OQ731408.1
Odocoileus pandora voucher UNAM:CNMA:26624 mitochondrion, partial genome	Odocoileus pandora	350	350	0.34	0	93.94	16422	OP764435.1
TPA: Odocoileus pandora mitochondrion, complete genome	Odocoileus pandora	350	350	0.34	0	93.94	16419	BK062825.1
Mazama americana voucher NUPECCE:T358 mitochondrion, complete genome	Mazama americana	350	350	0.34	0	93.94	16482	MZ350857.1
ATP8								
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. ident	Acc. Len	Accession
Odocoileus virginianus isolate Ovtol1 mitochondrion, complete genome	Odocoileus virginianus	350	350	0.95	0	100	16477	KM612276.1
Odocoileus hemionus isolate T1766 mitochondrion, complete genome	Odocoileus hemionus	350	350	0.95	0	100	16482	NC_020729.1
Odocoileus virginianus isolate Ovcou3 mitochondrion, complete genome	Odocoileus virginianus	344	344	0.95	0	99.47	16479	KM612275.1
Odocoileus virginianus isolate Ovsin17 mitochondrion, complete genome	Odocoileus virginianus	344	344	0.95	0	99.47	16478	KM612274.1
Odocoileus virginianus isolate Ovtex439 mitochondrion, complete genome	Odocoileus virginianus	344	344	0.95	0	99.47	16477	KM612273.1
Odocoileus virginianus isolate Ovver13 mitochondrion, complete genome	Odocoileus virginianus	344	344	0.95	0	99.47	16478	KM612271.1
Odocoileus virginianus mitochondrion, complete genome	Odocoileus virginianus	344	344	0.95	0	99.47	16477	NC_015247.1
Odocoileus virginianus isolate Ovmex31 mitochondrion, complete genome	Odocoileus virginianus	339	339	0.95	0	98.94	16477	KM612278.1
Odocoileus virginianus isolate Ovyuc61	Odocoileus virginianus	339	339	0.95	0	98.94	16469	KM612272.1

mitochondrion, complete genome								
Odocoileus virginianus isolate CYTO mitochondrion, complete genome	Odocoileus virginianus	339	339	0.95	0	98.94	16480	JN632672.1
Odocoileus virginianus isolate Ovoax2 mitochondrion, complete genome	Odocoileus virginianus	333	333	0.95	0	98.41	16470	KM612277.1
Odocoileus virginianus isolate OvacpCM mitochondrion, complete genome	Odocoileus virginianus	327	327	0.95	0	97.88	16471	KM612279.1
Odocoileus virginianus isolate T4887 mitochondrion, complete genome	Odocoileus virginianus	313	313	0.94	0	96.79	16465	JN632671.1
Ozotoceros bezoarticus isolate MRGOb2 mitochondrion, complete genome	Ozotoceros bezoarticus	305	305	0.95	0	95.77	16357	NC_020766.1
Ozotoceros bezoarticus voucher NUPECCE:FNMA40 mitochondrion, complete genome	Ozotoceros bezoarticus	300	300	0.95	0	95.24	16369	MZ350860.1
Mazama americana voucher NUPECCE:T358 mitochondrion, complete genome	Mazama americana	300	300	0.95	0	95.24	16482	MZ350857.1
Mazama americana isolate MAZ9472 mitochondrion, complete genome	Mazama americana	300	300	0.95	0	95.24	16478	NC_020719.1
Mazama americana voucher NUPECCE T297 mitochondrion, complete genome	Mazama americana	294	294	0.95	0	94.71	16474	OQ731411.1
Rangifer tarandus platyrhyncus genome assembly, organelle: mitochondrion	Rangifer tarandus platyrhyncus	294	294	0.95	0	94.71	16362	OX460346.1
Mazama temama voucher NUPECCE:T366 mitochondrion, complete genome	Mazama temama	294	294	0.95	0	94.71	16401	MZ350864.1
COX1								
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. ident	Acc. Len	Accession
Odocoileus hemionus isolate T1766 mitochondrion, complete genome	Odocoileus hemionus	1419	1603	0.77	0	88.28	16482	NC_020729.1
Odocoileus virginianus isolate Ovtol1 mitochondrion, complete genome	Odocoileus virginianus	1402	1581	0.77	0	88.01	16477	KM612276.1
Odocoileus virginianus isolate Ovtex439 mitochondrion, complete genome	Odocoileus virginianus	1393	1577	0.77	0	87.83	16477	KM612273.1

Odocoileus virginianus isolate Ovver13 mitochondrion, complete genome	Odocoileus virginianus	1393	1577	0.77	0	87.83	16478	KM612271.1
Odocoileus virginianus isolate Ovsin17 mitochondrion, complete genome	Odocoileus virginianus	1391	1575	0.77	0	87.83	16478	KM612274.1
Odocoileus virginianus isolate Ovmex31 mitochondrion, complete genome	Odocoileus virginianus	1387	1571	0.77	0	87.74	16477	KM612278.1
Odocoileus virginianus mitochondrion, complete genome	Odocoileus virginianus	1387	1568	0.77	0	87.74	16477	NC_015247.1
Odocoileus virginianus isolate CYTO mitochondrion, complete genome	Odocoileus virginianus	1376	1560	0.77	0	87.56	16480	JN632672.1
Odocoileus virginianus isolate Ovyuc61 mitochondrion, complete genome	Odocoileus virginianus	1371	1544	0.77	0	87.47	16469	KM612272.1
Odocoileus virginianus isolate Ovoax2 mitochondrion, complete genome	Odocoileus virginianus	1365	1538	0.77	0	87.38	16470	KM612277.1
Odocoileus virginianus isolate OvacpCM mitochondrion, complete genome	Odocoileus virginianus	1358	1531	0.77	0	87.28	16471	KM612279.1
Odocoileus virginianus isolate T4887 mitochondrion, complete genome	Odocoileus virginianus	1354	1527	0.77	0	87.19	16465	JN632671.1
Odocoileus virginianus isolate MRGOv14 mitochondrion, complete genome	Odocoileus virginianus	1192	1370	0.77	0	84.6	16483	JN632673.1
Mazama americana isolate MAZ9472 mitochondrion, complete genome	Mazama americana	1177	1344	0.77	0	84.29	16478	NC_020719.1
Mazama americana voucher NUPECCE:T358 mitochondrion, complete genome	Mazama americana	1166	1328	0.77	0	84.11	16482	MZ350857.1
Mazama americana voucher NUPECCE T297 mitochondrion, complete genome	Mazama americana	1160	1328	0.77	0	84.01	16474	OQ731411.1
Pudu mephistophiles isolate MRGPm2 mitochondrion, complete genome	Pudu mephistophiles	1072	1072	0.71	0	82.58	16426	NC_020739.1
Rangifer tarandus platyrhyncus genome assembly, chromosome: 5	Rangifer tarandus platyrhyncus	1070	1070	0.71	0	82.59	1E+08	OX596089.1
Mazama rufina isolate MRGMr4	Mazama rufina	1059	1059	0.71	0	82.4	16429	NC_020721.1

mitochondrion, complete genome								
Rangifer tarandus platyrhyncus genome assembly, chromosome: 22	Rangifer tarandus platyrhyncus	1018	1163	0.77	0	81.78	6E+07	OX596106.1
COX2								
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. ident	Acc. Len	Accession
Odocoileus virginianus isolate Ovmex31 mitochondrion, complete genome	Odocoileus virginianus	691	901	0.89	0	92.54	16477	KM612278.1
Odocoileus virginianus isolate Ovtol1 mitochondrion, complete genome	Odocoileus virginianus	691	901	0.89	0	92.54	16477	KM612276.1
Odocoileus hemionus isolate T1766 mitochondrion, complete genome	Odocoileus hemionus	691	907	0.89	0	92.54	16482	NC_020729.1
Odocoileus virginianus isolate Ovcou3 mitochondrion, complete genome	Odocoileus virginianus	686	896	0.89	0	92.32	16479	KM612275.1
Odocoileus virginianus isolate Ovsin17 mitochondrion, complete genome	Odocoileus virginianus	686	896	0.89	0	92.32	16478	KM612274.1
Odocoileus virginianus mitochondrion, complete genome	Odocoileus virginianus	676	886	0.89	0	91.89	16477	NC_015247.1
Odocoileus virginianus isolate Ovoax2 mitochondrion, complete genome	Odocoileus virginianus	675	879	0.89	0	91.89	16470	KM612277.1
Odocoileus virginianus isolate Ovtex439 mitochondrion, complete genome	Odocoileus virginianus	675	879	0.89	0	91.89	16477	KM612273.1
Odocoileus virginianus isolate Ovver13 mitochondrion, complete genome	Odocoileus virginianus	675	884	0.89	0	91.89	16478	KM612271.1
Odocoileus virginianus isolate OvacpCM mitochondrion, complete genome	Odocoileus virginianus	669	873	0.89	0	91.67	16471	KM612279.1
Odocoileus virginianus isolate CYTO mitochondrion, complete genome	Odocoileus virginianus	669	873	0.89	0	91.67	16480	JN632672.1
Odocoileus virginianus isolate T4887 mitochondrion, complete genome	Odocoileus virginianus	652	851	0.89	0	91.01	16465	JN632671.1
Odocoileus virginianus isolate Ovyuc61 mitochondrion, complete genome	Odocoileus virginianus	641	851	0.89	0	90.57	16469	KM612272.1
Odocoileus virginianus cytochrome c oxidase II	Odocoileus virginianus	641	840	0.89	0	90.57	684	U18816.1

gene, mitochondrial gene encoding mitochondrial protein, complete cds								
Mazama nana voucher NUPECCE T107 mitochondrion, complete genome	Mazama nana	625	801	0.89	0	89.91	16427	NC_065788.1
Mazama americana voucher NUPECCE T297 mitochondrion, complete genome	Mazama americana	619	812	0.88	0	89.85	16474	OQ731411.1
Mazama americana isolate MAZ9472 mitochondrion, complete genome	Mazama americana	619	812	0.88	0	89.85	16478	NC_020719.1
Mazama americana voucher NUPECCE:T358 mitochondrion, complete genome	Mazama americana	614	807	0.88	0	89.62	16482	MZ350857.1
Mazama temama voucher MZFC:4668 mitochondrion, complete genome	Mazama temama	606	606	0.66	0	89.38	16480	OP712670.1
Mazama temama voucher MZFC:4650 mitochondrion, complete genome	Mazama temama	606	606	66%	####	89.38	16479	OP712669.1
COX3								
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. ident	Acc. Len	Accession
Odocoileus virginianus isolate Ovtol1 mitochondrion, complete genome	Odocoileus virginianus	464	839	0.75	0	89.53	16477	KM612276.1
Odocoileus virginianus isolate Ovcou3 mitochondrion, complete genome	Odocoileus virginianus	464	845	0.75	0	89.53	16479	KM612275.1
Odocoileus virginianus isolate Ovsin17 mitochondrion, complete genome	Odocoileus virginianus	464	845	0.75	0	89.53	16478	KM612274.1
Odocoileus virginianus isolate Ovmex31 mitochondrion, complete genome	Odocoileus virginianus	459	834	0.75	0	89.26	16477	KM612278.1
Odocoileus virginianus isolate Ovtex439 mitochondrion, complete genome	Odocoileus virginianus	459	834	0.75	0	89.26	16477	KM612273.1
Odocoileus virginianus isolate Ovver13 mitochondrion, complete genome	Odocoileus virginianus	459	828	0.75	0	89.26	16478	KM612271.1
Odocoileus virginianus isolate CYTO mitochondrion, complete genome	Odocoileus virginianus	459	783	0.71	0	89.26	16480	JN632672.1
Odocoileus hemionus isolate T1766	Odocoileus hemionus	459	845	0.75	0	89.26	16482	NC_020729.1

mitochondrion, complete genome								
Odocoileus virginianus mitochondrion, complete genome	Odocoileus virginianus	459	834	0.75	0	89.26	16477	NC_015247.1
Mazama temama voucher NUPECCE T362 mitochondrion, complete genome	Mazama temama	436	683	0.71	0	88.15	16356	NC_065375.1
Mazama temama voucher MZFC:4650 mitochondrion, complete genome	Mazama temama	435	681	0.71	0	88.12	16479	OP712669.1
Mazama temama voucher MZFC:4668 mitochondrion, complete genome	Mazama temama	429	676	0.71	0	87.85	16480	OP712670.1
Mazama temama voucher NUPECCE:T366 mitochondrion, complete genome	Mazama temama	429	676	0.71	0	87.85	16401	MZ350864.1
Mazama americana isolate MRGMa40 mitochondrion, complete genome	Mazama americana	429	743	0.75	0	87.85	16473	JN632657.1
Odocoileus virginianus isolate OvacpCM mitochondrion, complete genome	Odocoileus virginianus	425	790	0.75	0	87.6	16471	KM612279.1
Mazama americana voucher NUPECCE:T358 mitochondrion, complete genome	Mazama americana	424	681	0.71	0	87.57	16482	MZ350857.1
Mazama americana voucher NUPECCE:T253 mitochondrion, complete genome	Mazama americana	424	749	0.75	0	87.57	16476	MZ350856.1
Odocoileus virginianus isolate MRGOv14 mitochondrion, complete genome	Odocoileus virginianus	424	749	0.75	0	87.57	16483	JN632673.1
Odocoileus virginianus isolate Ovoax2 mitochondrion, complete genome	Odocoileus virginianus	420	795	0.75	0	87.33	16470	KM612277.1
Odocoileus virginianus isolate Ovyuc61 mitochondrion, complete genome	Odocoileus virginianus	420	790	0.75	0	87.33	16469	KM612272.1
CYTB								
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. ident	Acc. Len	Accession
Odocoileus hemionus hemionus haplotype AB_728_24 cytochrome b (cytb) gene, partial cds; mitochondrial	Odocoileus hemionus hemionus	1009	1335	0.8	0	91.56	1028	FJ188725.1
Odocoileus virginianus x Odocoileus hemionus isolate TK24403 cytochrome b (cytb)	Odocoileus virginianus x Odocoileus hemionus	1003	1464	0.87	0	91.42	1140	ON924828.1

gene, complete cds; mitochondrial								
Odocoileus virginianus x Odocoileus hemionus isolate TK201631 cytochrome b (cytb) gene, complete cds; mitochondrial	Odocoileus virginianus x Odocoileus hemionus	1003	1470	0.87	0	91.42	1140	ON924827.1
Odocoileus hemionus isolate TK24425 cytochrome b (cytb) gene, complete cds; mitochondrial	Odocoileus hemionus	1003	1464	0.87	0	91.42	1140	ON924802.1
Odocoileus hemionus isolate TK211316 cytochrome b (cytb) gene, complete cds; mitochondrial	Odocoileus hemionus	1003	1464	0.87	0	91.42	1140	ON924801.1
Odocoileus hemionus isolate TK211305 cytochrome b (cytb) gene, complete cds; mitochondrial	Odocoileus hemionus	1003	1464	0.87	0	91.42	1140	ON924799.1
Odocoileus hemionus isolate TK24413 cytochrome b (cytb) gene, complete cds; mitochondrial	Odocoileus hemionus	1003	1464	0.87	0	91.42	1140	ON924796.1
Odocoileus hemionus isolate 101MDSF cytochrome b (cytb) gene, partial cds; and tRNA-Thr gene, partial sequence; mitochondrial	Odocoileus hemionus	1003	1413	0.85	0	91.42	1118	OK668115.1
Odocoileus hemionus voucher TTUM147715 cytochrome b (cytb) gene, complete cds; mitochondrial	Odocoileus hemionus	1003	1459	0.87	0	91.42	1140	MT846755.1
Odocoileus hemionus voucher TK26441 cytochrome b (cytb) gene, complete cds; mitochondrial	Odocoileus hemionus	1003	1459	0.87	0	91.42	1140	MT846749.1
Odocoileus hemionus isolate RF-2 cytochrome b (cytb) gene, complete cds; mitochondrial	Odocoileus hemionus	1003	1464	0.87	0	91.42	1140	HM222707.1
Odocoileus hemionus hemionus haplotype WY_SC_10 cytochrome b (cytb) gene, partial cds; mitochondrial	Odocoileus hemionus hemionus	1003	1329	0.8	0	91.42	1028	FJ188896.1
Odocoileus hemionus hemionus haplotype OR_NS_02 cytochrome b (cytb) gene, partial cds; mitochondrial	Odocoileus hemionus hemionus	1003	1324	0.8	0	91.42	1028	FJ188857.1
Odocoileus hemionus hemionus haplotype NM_RT_08 cytochrome	Odocoileus hemionus hemionus	1003	1313	0.8	0	91.42	1028	FJ188845.1

b (cytb) gene, partial cds; mitochondrial								
Odocoileus hemionus hemionus haplotype NM_RT_07 cytochrome b (cytb) gene, partial cds; mitochondrial	Odocoileus hemionus hemionus	1003	1318	0.8	0	91.42	1028	FJ188844.1
Odocoileus hemionus hemionus haplotype ID_SE_07 cytochrome b (cytb) gene, partial cds; mitochondrial	Odocoileus hemionus hemionus	1003	1324	0.8	0	91.42	1028	FJ188822.1
Odocoileus hemionus hemionus haplotype ID_BO_16 cytochrome b (cytb) gene, partial cds; mitochondrial	Odocoileus hemionus hemionus	1003	1307	0.8	0	91.42	1028	FJ188820.1
Odocoileus hemionus eremicus haplotype CA_IM_08 cytochrome b (cytb) gene, partial cds; mitochondrial	Odocoileus hemionus eremicus	1003	1324	0.8	0	91.42	1028	FJ188775.1
Odocoileus hemionus hemionus haplotype BC_CR_20 cytochrome b (cytb) gene, partial cds; mitochondrial	Odocoileus hemionus hemionus	1003	1324	0.8	0	91.42	1028	FJ188765.1
Odocoileus hemionus hemionus haplotype AZ_KB_16 cytochrome b (cytb) gene, partial cds; mitochondrial	Odocoileus hemionus hemionus	1003	1324	0.8	0	91.42	1028	FJ188748.1
NADH1								
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. ident	Acc. Len	Accession
Odocoileus virginianus isolate OvacpCM mitochondrion, complete genome	Odocoileus virginianus	145	370	0.24	0	94.44	16471	KM612279.1
Odocoileus virginianus isolate Ovmex31 mitochondrion, complete genome	Odocoileus virginianus	145	262	0.16	0	94.44	16477	KM612278.1
Odocoileus virginianus isolate Ovoax2 mitochondrion, complete genome	Odocoileus virginianus	145	370	0.24	0	94.44	16470	KM612277.1
Odocoileus virginianus isolate Ovtol1 mitochondrion, complete genome	Odocoileus virginianus	145	365	0.24	0	94.44	16477	KM612276.1
Odocoileus virginianus isolate Ovtex439 mitochondrion, complete genome	Odocoileus virginianus	145	365	0.24	0	94.44	16477	KM612273.1
Odocoileus virginianus isolate Ovyuc61 mitochondrion, complete genome	Odocoileus virginianus	145	370	0.24	0	94.44	16469	KM612272.1
Odocoileus virginianus isolate Ovver13	Odocoileus virginianus	145	365	0.24	0	94.44	16478	KM612271.1

mitochondrion, complete genome								
Odocoileus virginianus isolate CYTO mitochondrion, complete genome	Odocoileus virginianus	145	365	0.24	0	94.44	16480	JN632672.1
Odocoileus hemionus isolate T1766 mitochondrion, complete genome	Odocoileus hemionus	145	365	0.24	0	94.44	16482	NC_020729.1
Odocoileus virginianus mitochondrion, complete genome	Odocoileus virginianus	145	359	0.24	0	94.44	16477	NC_015247.1
Odocoileus virginianus isolate Ovcou3 mitochondrion, complete genome	Odocoileus virginianus	139	359	0.24	0	93.33	16479	KM612275.1
Odocoileus virginianus isolate Ovsin17 mitochondrion, complete genome	Odocoileus virginianus	139	359	0.24	0	93.33	16478	KM612274.1
Odocoileus virginianus isolate T4887 mitochondrion, complete genome	Odocoileus virginianus	139	363	0.24	0	93.33	16465	JN632671.1
Mazama americana voucher NUPECCE T297 mitochondrion, complete genome	Mazama americana	134	245	0.16	0	92.22	16474	OQ731411.1
Mazama americana voucher NUPECCE:T358 mitochondrion, complete genome	Mazama americana	134	245	0.16	0	92.22	16482	MZ350857.1
Odocoileus pandora voucher NUPECCE T365 topotype mitochondrion, complete genome	Odocoileus pandora	132	344	0.23	0	92.13	16474	OQ731410.1
Mazama americana isolate MAZ9472 mitochondrion, complete genome	Mazama americana	128	240	0.16	0	91.11	16478	NC_020719.1
Mazama americana isolate MRGMa40 mitochondrion, complete genome	Mazama americana	126	126	0.12	0	82.93	16473	JN632657.1
Odocoileus virginianus isolate MRGOv14 mitochondrion, complete genome	Odocoileus virginianus	122	326	0.24	0	90	16483	JN632673.1
Beatragus hunteri mitochondrion, complete genome	Beatragus hunteri	121	121	0.12	0	82.11	16359	NC_023542.1
NADH2								
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. ident	Acc. Len	Accession
Odocoileus hemionus isolate T1766 mitochondrion, complete genome	Odocoileus hemionus	641	801	0.67	0	86.71	16482	NC_020729.1
Odocoileus virginianus isolate Ovtol1	Odocoileus virginianus	632	792	0.67	0	86.32	16477	KM612276.1

mitochondrion, complete genome								
Odocoileus virginianus isolate Ovmex31 mitochondrion, complete genome	Odocoileus virginianus	625	796	0.67	0	86.1	16477	KM612278.1
Odocoileus virginianus isolate Ovtex439 mitochondrion, complete genome	Odocoileus virginianus	625	796	0.67	0	86.1	16477	KM612273.1
Odocoileus virginianus isolate Ovver13 mitochondrion, complete genome	Odocoileus virginianus	625	796	0.67	0	86.1	16478	KM612271.1
Odocoileus virginianus isolate CYTO mitochondrion, complete genome	Odocoileus virginianus	619	790	0.67	0	85.91	16480	JN632672.1
Odocoileus virginianus mitochondrion, complete genome	Odocoileus virginianus	619	785	0.67	0	85.91	16477	NC_015247.1
Odocoileus virginianus isolate Ovyuc61 mitochondrion, complete genome	Odocoileus virginianus	614	779	0.67	0	85.71	16469	KM612272.1
Odocoileus virginianus isolate Ovoax2 mitochondrion, complete genome	Odocoileus virginianus	603	763	0.67	0	85.33	16470	KM612277.1
Odocoileus virginianus isolate Ovcou3 mitochondrion, complete genome	Odocoileus virginianus	603	774	0.67	0	85.33	16479	KM612275.1
Odocoileus virginianus isolate Ovsin17 mitochondrion, complete genome	Odocoileus virginianus	603	774	0.67	0	85.33	16478	KM612274.1
Odocoileus virginianus isolate T4887 mitochondrion, complete genome	Odocoileus virginianus	597	763	0.67	0	85.14	16465	JN632671.1
Odocoileus virginianus isolate OvacpCM mitochondrion, complete genome	Odocoileus virginianus	592	746	0.67	0	84.94	16471	KM612279.1
Mazama nana voucher NUPECCE T107 mitochondrion, complete genome	Mazama nana	547	547	0.49	0	83.4	16427	NC_065788.1
Mazama temama voucher MZFC:4668 mitochondrion, complete genome	Mazama temama	542	703	0.67	0	83.2	16480	OP712670.1
Mazama americana isolate MRGMa40 mitochondrion, complete genome	Mazama americana	542	674	0.67	0	83.2	16473	JN632657.1
Mazama temama voucher MZFC:4650 mitochondrion, complete genome	Mazama temama	536	698	0.67	0	83.01	16479	OP712669.1

Mazama temama voucher NUPECCE:T366 mitochondrion, complete genome	Mazama temama	536	698	0.67	0	83.04	16401	MZ350864.1
Mazama temama voucher NUPECCE T362 mitochondrion, complete genome	Mazama temama	536	698	0.67	0	83.01	16356	NC_065375.1
Odocoileus virginianus isolate MRGOv14 mitochondrion, complete genome	Odocoileus virginianus	536	698	0.67	0	82.98	16483	JN632673.1
NADH3								
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. ident	Acc. Len	Accession
Odocoileus hemionus isolate T1766 mitochondrion, complete genome	Odocoileus hemionus	305	305	0.6	0	91.47	16482	NC_020729.1
Odocoileus virginianus isolate Ovtol1 mitochondrion, complete genome	Odocoileus virginianus	300	300	0.6	0	91	16477	KM612276.1
Odocoileus virginianus isolate Ovcou3 mitochondrion, complete genome	Odocoileus virginianus	300	300	0.6	0	91	16479	KM612275.1
Odocoileus virginianus isolate Ovsin17 mitochondrion, complete genome	Odocoileus virginianus	300	300	0.6	0	91	16478	KM612274.1
Odocoileus virginianus isolate T4887 mitochondrion, complete genome	Odocoileus virginianus	300	300	0.6	0	91	16465	JN632671.1
Odocoileus virginianus isolate OvacpCM mitochondrion, complete genome	Odocoileus virginianus	294	294	0.6	0	90.52	16471	KM612279.1
Odocoileus virginianus isolate Ovyuc61 mitochondrion, complete genome	Odocoileus virginianus	294	294	0.6	0	90.52	16469	KM612272.1
Odocoileus virginianus isolate CYTO mitochondrion, complete genome	Odocoileus virginianus	294	294	0.6	0	90.52	16480	JN632672.1
Odocoileus virginianus mitochondrion, complete genome	Odocoileus virginianus	291	291	0.6	0	90.05	16477	NC_015247.1
Odocoileus virginianus isolate Ovmex31 mitochondrion, complete genome	Odocoileus virginianus	289	289	0.6	0	90.05	16477	KM612278.1
Odocoileus virginianus isolate Ovoax2 mitochondrion, complete genome	Odocoileus virginianus	289	289	0.6	0	90.05	16470	KM612277.1
Odocoileus virginianus isolate Ovtex439 mitochondrion, complete genome	Odocoileus virginianus	289	289	0.6	0	90.05	16477	KM612273.1

Odocoileus virginianus isolate Ovver13 mitochondrion, complete genome	Odocoileus virginianus	289	289	0.6	0	90.05	16478	KM612271.1
Odocoileus virginianus isolate MRGOv14 mitochondrion, complete genome	Odocoileus virginianus	281	281	0.6	0	89.47	16483	JN632673.1
Mazama americana voucher NUPECCE T297 mitochondrion, complete genome	Mazama americana	272	272	0.6	0	88.63	16474	OQ731411.1
Mazama americana isolate MAZ9472 mitochondrion, complete genome	Mazama americana	272	272	0.6	0	88.63	16478	NC_020719.1
Mazama temama voucher MZFC:4668 mitochondrion, complete genome	Mazama temama	270	270	0.6	0	88.52	16480	OP712670.1
Mazama temama voucher NUPECCE:T366 mitochondrion, complete genome	Mazama temama	270	270	0.6	0	88.52	16401	MZ350864.1
Mazama temama voucher NUPECCE T362 mitochondrion, complete genome	Mazama temama	270	270	0.6	0	88.52	16356	NC_065375.1
Odocoileus pandora voucher NUPECCE T365 topotype mitochondrion, complete genome	Odocoileus pandora	267	267	0.6	0	88.15	16474	OQ731410.1
NADH4								
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. ident	Acc. Len	Accession
Odocoileus hemionus isolate T1766 mitochondrion, complete genome	Odocoileus hemionus	1090	1816	0.84	0	99.01	16482	NC_020729.1
Odocoileus virginianus isolate Ovmex31 mitochondrion, complete genome	Odocoileus virginianus	1074	1800	0.84	0	98.51	16477	KM612278.1
Odocoileus virginianus isolate Ovver13 mitochondrion, complete genome	Odocoileus virginianus	1074	1794	0.84	0	98.51	16478	KM612271.1
Odocoileus virginianus isolate Ovtex439 mitochondrion, complete genome	Odocoileus virginianus	1062	1783	0.84	0	98.18	16477	KM612273.1
Odocoileus virginianus mitochondrion, complete genome	Odocoileus virginianus	1062	1792	0.84	0	98.18	16477	NC_015247.1
Odocoileus virginianus isolate Ovcou3 mitochondrion, complete genome	Odocoileus virginianus	1057	1794	0.84	0	98.02	16479	KM612275.1
Odocoileus virginianus isolate Ovsin17	Odocoileus virginianus	1057	1789	0.84	0	98.02	16478	KM612274.1

mitochondrion, complete genome								
Odocoileus virginianus isolate Ovtol1 mitochondrion, complete genome	Odocoileus virginianus	1046	1777	0.84	0	97.69	16477	KM612276.1
Odocoileus virginianus isolate CYTO mitochondrion, complete genome	Odocoileus virginianus	1035	1761	0.84	0	97.36	16480	JN632672.1
Odocoileus virginianus isolate OvacpCM mitochondrion, complete genome	Odocoileus virginianus	1029	1750	0.84	0	97.19	16471	KM612279.1
Odocoileus virginianus isolate Ovoax2 mitochondrion, complete genome	Odocoileus virginianus	1024	1739	0.84	0	97.03	16470	KM612277.1
Odocoileus virginianus isolate T4887 mitochondrion, complete genome	Odocoileus virginianus	996	1711	0.84	0	96.2	16465	JN632671.1
Mazama americana voucher NUPECCE:T253 mitochondrion, complete genome	Mazama americana	913	1543	0.84	0	93.73	16476	MZ350856.1
Mazama americana voucher NUPECCE T297 mitochondrion, complete genome	Mazama americana	907	1550	0.84	0	93.56	16474	OQ731411.1
Odocoileus virginianus isolate Ovyuc61 mitochondrion, complete genome	Odocoileus virginianus	896	1578	0.84	0	93.23	16469	KM612272.1
Mazama bororo voucher NUPECCE T215 mitochondrion, complete genome	Mazama bororo	896	1521	0.84	0	93.23	16430	NC_065787.1
Mazama americana isolate MRGMa40 mitochondrion, complete genome	Mazama americana	896	1537	0.84	0	93.23	16473	JN632657.1
Mazama americana voucher NUPECCE:T358 mitochondrion, complete genome	Mazama americana	891	1517	0.84	0	93.07	16482	MZ350857.1
Mazama americana isolate MAZ9472 mitochondrion, complete genome	Mazama americana	891	1517	0.84	0	93.07	16478	NC_020719.1
Mazama temama voucher NUPECCE:T366 mitochondrion, complete genome	Mazama temama	885	1512	0.84	0	92.9	16401	MZ350864.1
NADH5								
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. ident	Acc. Len	Accession
Odocoileus virginianus isolate Ovcou3 mitochondrion, complete genome	Odocoileus virginianus	1020	1657	0.64	0	97.94	16479	KM612275.1

Odocoileus virginianus isolate Ovsin17 mitochondrion, complete genome	Odocoileus virginianus	1020	1657	0.64	0	97.94	16478	KM612274.1
Odocoileus virginianus isolate Ovtex439 mitochondrion, complete genome	Odocoileus virginianus	1020	1659	0.64	0	97.94	16477	KM612273.1
Odocoileus virginianus isolate Ovver13 mitochondrion, complete genome	Odocoileus virginianus	1020	1659	0.64	0	97.94	16478	KM612271.1
Odocoileus hemionus isolate T1766 mitochondrion, complete genome	Odocoileus hemionus	1020	1675	0.64	0	97.94	16482	NC_020729.1
Odocoileus virginianus isolate Ovmex31 mitochondrion, complete genome	Odocoileus virginianus	1009	1647	0.64	0	97.59	16477	KM612278.1
Odocoileus virginianus mitochondrion, complete genome	Odocoileus virginianus	1009	1647	0.64	0	97.59	16477	NC_015247.1
Odocoileus virginianus isolate Ovtol1 mitochondrion, complete genome	Odocoileus virginianus	1000	1633	0.64	0	97.25	16477	KM612276.1
Odocoileus virginianus isolate CYTO mitochondrion, complete genome	Odocoileus virginianus	970	1609	0.64	0	96.39	16480	JN632672.1
Odocoileus virginianus isolate OvacpCM mitochondrion, complete genome	Odocoileus virginianus	931	1542	0.64	0	95.19	16471	KM612279.1
Odocoileus virginianus isolate Ovyuc61 mitochondrion, complete genome	Odocoileus virginianus	931	1542	0.64	0	95.19	16469	KM612272.1
Odocoileus virginianus isolate Ovoax2 mitochondrion, complete genome	Odocoileus virginianus	928	1533	0.64	0	95.02	16470	KM612277.1
Odocoileus virginianus isolate T4887 mitochondrion, complete genome	Odocoileus virginianus	926	1526	0.64	0	95.02	16465	JN632671.1
Odocoileus pandora voucher NUPECCE T365 topotype mitochondrion, complete genome	Odocoileus pandora	837	1365	0.53	0	92.27	16474	OQ731410.1
Mazama americana voucher NUPECCE:T253 mitochondrion, complete genome	Mazama americana	835	1238	0.47	0	92.25	16476	MZ350856.1
TPA: Odocoileus pandora mitochondrion, complete genome	Odocoileus pandora	832	1359	0.53	0	92.1	16419	BK062825.1
Odocoileus pandora voucher NUPECCE MX06	Odocoileus pandora	826	1354	0.53	0	91.92	16418	OQ731409.1

mitochondrion, complete genome								
Odocoileus pandora voucher NUPECCE MX02 mitochondrion, complete genome	Odocoileus pandora	826	1354	0.53	0	91.92	16362	OQ731408.1
Mazama temama voucher MZFC:4650 mitochondrion, complete genome	Mazama temama	821	1102	0.43	0	91.75	16479	OP712669.1
Odocoileus virginianus isolate MRGOv14 mitochondrion, complete genome	Odocoileus virginianus	821	1099	0.42	0	91.75	16483	JN632673.1
NADH6								
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. ident	Acc. Len	Accession
Odocoileus hemionus isolate T1766 mitochondrion, complete genome	Odocoileus hemionus	636	636	0.76	0	94.75	16482	NC_020729.1
Odocoileus virginianus isolate Ovtol1 mitochondrion, complete genome	Odocoileus virginianus	625	625	0.76	0	94.25	16477	KM612276.1
Odocoileus virginianus isolate Ovcou3 mitochondrion, complete genome	Odocoileus virginianus	619	619	0.76	0	94	16479	KM612275.1
Odocoileus virginianus isolate Ovsin17 mitochondrion, complete genome	Odocoileus virginianus	619	619	0.76	0	94	16478	KM612274.1
Odocoileus virginianus isolate Ovmex31 mitochondrion, complete genome	Odocoileus virginianus	614	614	0.76	0	93.73	16477	KM612278.1
Odocoileus virginianus isolate CYTO mitochondrion, complete genome	Odocoileus virginianus	614	614	0.76	0	93.75	16480	JN632672.1
Odocoileus virginianus mitochondrion, complete genome	Odocoileus virginianus	614	614	0.76	0	93.75	16477	NC_015247.1
Odocoileus virginianus isolate Ovver13 mitochondrion, complete genome	Odocoileus virginianus	608	608	0.76	0	93.5	16478	KM612271.1
Odocoileus virginianus isolate Ovtex439 mitochondrion, complete genome	Odocoileus virginianus	603	603	0.76	0	93.25	16477	KM612273.1
Odocoileus virginianus isolate T4887 mitochondrion, complete genome	Odocoileus virginianus	603	603	0.76	0	93.25	16465	JN632671.1
Odocoileus virginianus isolate Ovoax2 mitochondrion, complete genome	Odocoileus virginianus	597	597	0.76	0	93	16470	KM612277.1
Odocoileus virginianus isolate OvacpCM	Odocoileus virginianus	592	592	0.76	0	92.75	16471	KM612279.1

mitochondrion, complete genome								
Mazama americana voucher NUPECCE:T358 mitochondrion, complete genome	Mazama americana	586	586	0.76	0	92.5	16482	MZ350857.1
Mazama americana isolate MAZ9472 mitochondrion, complete genome	Mazama americana	586	586	0.76	0	92.5	16478	NC_020719.1
Mazama americana voucher NUPECCE T297 mitochondrion, complete genome	Mazama americana	580	580	0.76	0	92.25	16474	OQ731411.1
Mazama temama voucher MZFC:4650 mitochondrion, complete genome	Mazama temama	580	580	0.76	0	92.23	16479	OP712669.1
Mazama temama voucher NUPECCE T362 mitochondrion, complete genome	Mazama temama	580	580	0.76	0	92.23	16356	NC_065375.1
Odocoileus virginianus isolate Ovyuc61 mitochondrion, complete genome	Odocoileus virginianus	575	575	0.76	0	92	16469	KM612272.1
Mazama temama voucher MZFC:4668 mitochondrion, complete genome	Mazama temama	569	569	0.76	0	91.73	16480	OP712670.1
Mazama temama voucher NUPECCE:T366 mitochondrion, complete genome	Mazama temama	569	569	0.76	0	91.73	16401	MZ350864.1

Table S4.4: Nucleotide diversity (π) and sequence divergence (d_{XY}) for data mapped to cattle.

		ROMM75974	*A. alces*	*C. canadensis*	*C. elaphus*	*O. hemionus*	*O. virginianus*	*R. tarandus*
π		0.0100	0.0063	0.0039	0.0136	0.0053	0.0081	0.0054
d_{XY}	ROMM75974	-	0.0170	0.0183	0.0144	0.0089	0.0106	0.0122
	A. alces	0.0170	-	0.0232	0.0193	0.0198	0.0200	0.0206
	C. canadensis	0.0183	0.0232	-	0.0088	0.0233	0.0236	0.0241
	C. elaphus	0.0144	0.0193	0.0088	-	0.0185	0.0188	0.0194
	O. hemionus	0.0089	0.0198	0.0233	0.0185	-	0.0092	0.0138
	O. virginianus	0.0106	0.0200	0.0236	0.0188	0.0092	-	0.0139
	R. tarandus	0.0122	0.0206	0.0241	0.0194	0.0138	0.0139	-

Table S4.5: Mitochondrial genomes accession summary: Information regarding the 17 Cervidae and the cattle mitochondrial genomes used in the phylogenetic analyses, with genomic range of the *COX1* and *CYTB* genes.

Description	Accession	Range (bp)	
		COX1	CYTB
Alces alces isolate Kazakhstan mitochondrion, complete genome	NC_020677.1	5329-6873	14156-15295
Axis axis isolate CYTO mitochondrion, complete genome	NC_020680.1	5329-6873	14155-15294
Blastocerus dichotomus isolate MRGBd8 mitochondrion, complete genome	NC_020682.1	5325-6869	14152-15291
Bos taurus mitochondrion, complete genome	NC_006853.1	5687-7231	14514-15653
Capreolus capreolus isolate CYTO mitochondrion, complete genome	NC_020684.1	5326-6870	14151-15290
Cervus canadensis isolate 19FC017 mitochondrion, complete genome	NC_050863.1	5331-6875	14158-15297
Cervus elaphus mitochondrion, complete genome	NC_007704.2	5336-6880	14162-15301
Dama dama isolate CYTO mitochondrion, complete genome	NC_020700.1	5331-6875	14158-15297
Mazama rufina isolate MRGMr4 mitochondrion, complete genome	NC_020721.1	5323-6867	14150-15292
Muntiacus muntjak mitochondrion, complete genome	NC_004563.1	5329-6873	14155-15294
Muntiacus putaoensis mitochondrion, complete genome	NC_036430.1	5331-6875	14156-15295
Odocoileus hemionus isolate T1766 mitochondrion, complete genome	NC_020729.1	5327-6871	14154-15293
Odocoileus pandora voucher NUPECCE T365 topotype mitochondrion, complete genome	OQ731410.1	5328-6872	14155-15297
Odocoileus virginianus mitochondrion, complete genome	NC_015247.1	380-1924	9207-10346
Ozotoceros bezoarticus isolate MRGOb2 mitochondrion, complete genome	NC_020766.1	5323-6867	14149-15288
Pudu puda isolate M92144 mitochondrion, complete genome	NC_020740.1	5323-6867	14150-15289
Rangifer tarandus mitochondrion, complete genome	NC_007703.1	5329-6873	14156-15297
Rucervus eldi mitochondrion, complete genome	NC_014701.1	5332-6876	14157-15296

Table S4.6: Whole genome data summary. SRA accessions and biosample information of the 37 modern samples used, coloured by species.

Species	Sample ID	Accession	Locality	Sex
Alces alces	Aa-AK01	SRR6079187	Alaska	Male
Alces alces	Aa-ID01	SRR6079177	Idaho	Male
Alces alces	Aa-MI01	SRR18899225	Michigan	NA
Alces alces	Aa-MN01	SRR18899222	Minnesota	NA
Alces alces	Aa-SW01	ERR12087977	Sweden	NA
Alces alces	Aa-SW02	ERR12087904	Sweden	NA
Alces alces	Aa-VT01	SRR6079181	Vermont	Male
Alces alces	Aa-WY01	SRR6079200	Wyoming	Female
Cervus canadensis	Cc-WY01	SRR12450513	Wyoming	Female
Cervus canadensis	Cc-MN01	SRR12450505	Minnesota	Male
Cervus canadensis	Cc-MN02	SRR12450502	Minnesota	Female
Cervus canadensis	Cc-WY02	SRR12450506	Wyoming	Female
Cervus elaphus	Ce-HU01	SRR956941	Hungary	Male
Rangifer tarandus	Rt-BC01	SRR27590283	British Columbia	NA
Rangifer tarandus	Rt-BC02	ERR11471728	British Columbia	NA
Rangifer tarandus	Rt-ON01	ERR11471702	Ontario	Male
Rangifer tarandus	Rt-ON02	SRR24951207	Ontario	NA
Rangifer tarandus	Rt-NO01	SRR24951210	Norway	NA
Rangifer tarandus	Rt-NO02	SRR15459420	Norway	NA
Rangifer tarandus	Rt-RU01	SRR15459424	Russia	NA
Odocoileus hemionus	Oh_AB1	SRS12707053	Alberta	Female
Odocoileus hemionus	Oh_AK1	SRS12707054	Alaska	NA
Odocoileus hemionus	Oh_BC1	SRS12707032	British Columbia	Female
Odocoileus hemionus	Oh_BC6	SRS12707055	British Columbia	Male
Odocoileus hemionus	Oh_CA3	SRS12707057	California	NA
Odocoileus hemionus	Oh_OR2	SRS12707035	Oregon	NA
Odocoileus hemionus	Oh_SK2	SRS16492111	Saskatchewan	NA
Odocoileus hemionus	Oh_UT2	SRS12707038	Utah	Male
Odocoileus virginianus	Ov_ID1	SRS16492129	Idaho	Female
Odocoileus virginianus	Ov_Key4	SRS16492093	Florida	NA
Odocoileus virginianus	Ov_MO1	SRS16492102	Missouri	Male
Odocoileus virginianus	Ov_MX2	SRS12707043	Mexico	Male
Odocoileus virginianus	Ov_ONX2	SRS16492116	Ontario	Female
Odocoileus virginianus	Ov_PA3	SRS12707047	Pennsylvania	Female
Odocoileus virginianus	Ov_SK3	SRS12707051	Saskatchewan	Male
Odocoileus virginianus	Ov_TX20	SRS16492126	Texas	Female
Odocoileus virginianus	Ov_TX20	SRS16492126	Texas	Female

Table S4.7: Kraken2 results, number of reads belonging to target taxonomic groups.

Percentage of reads	Number of reads	Taxon
89.84	53653069	Unclassified
10.04	5993203	Bacteria
0.05	31199	Archaea
0.02	13779	Fungi
0.02	8976	Viruses

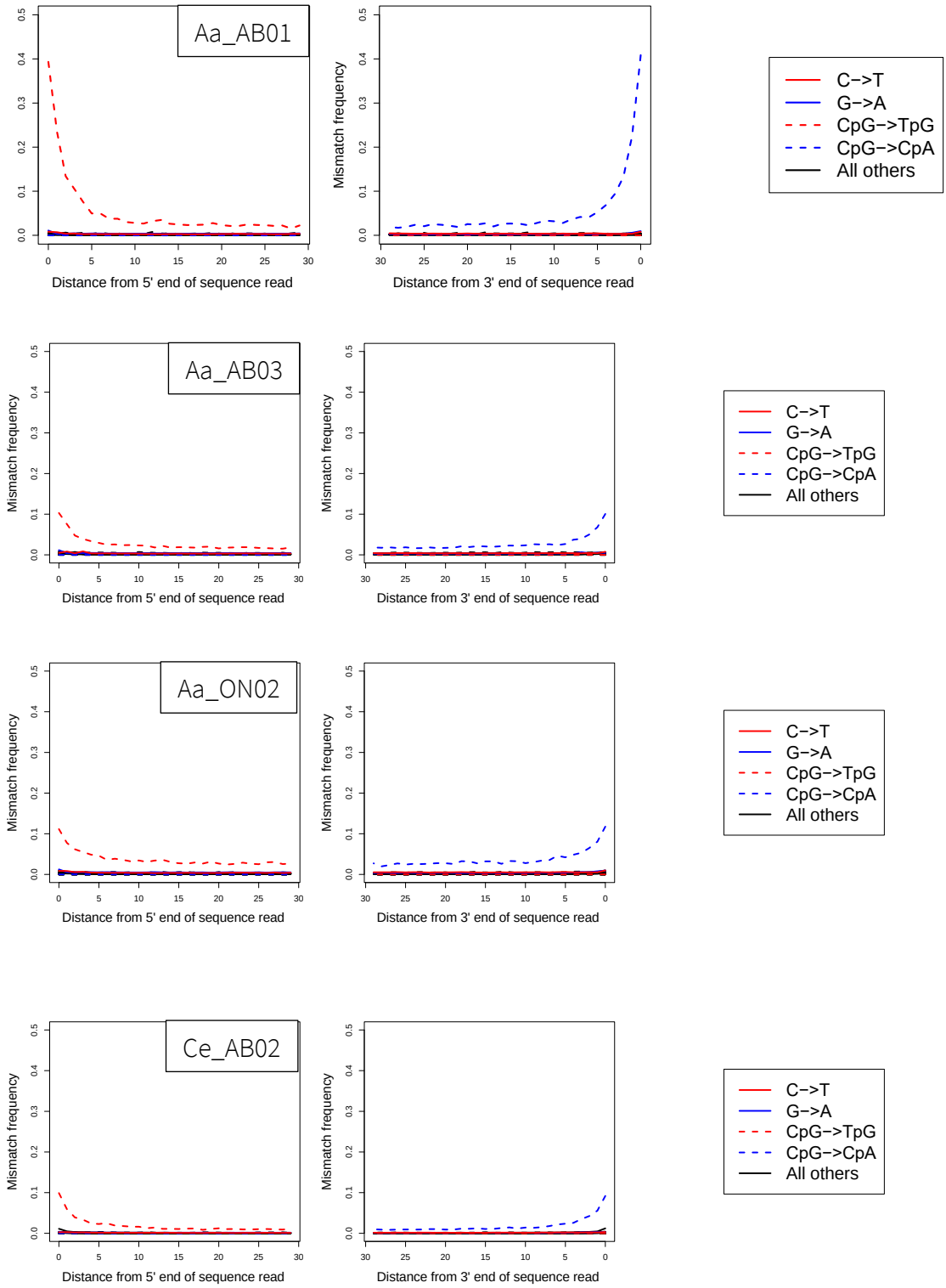
Table S4.8: Quality check for ROMM75974 whole genome data mapped to cattle and caribou. Final read number represents reads over 25 bp with a minimum quality of 30 and with duplicates removed.

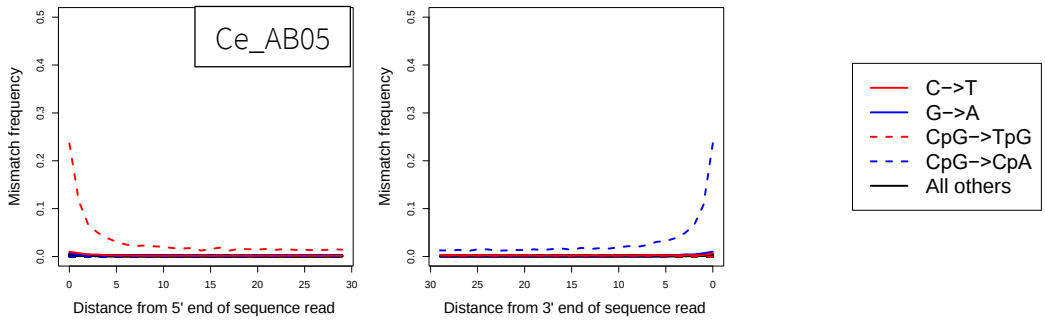
	Reference	Reads total	Reads mapped (#)	Reads mapped (%)	Final filtered reads (#)	Final average read length	ANGSD depth	ANGSD variant (#)
ROMM75794	Cattle (GCF_002263795.2)	26579834	61310	0.230664	39941	62	0.007	41794
	Caribou (GCA_949782905.1)	26579834	128407	0.483099	81022	70	0.237	138714

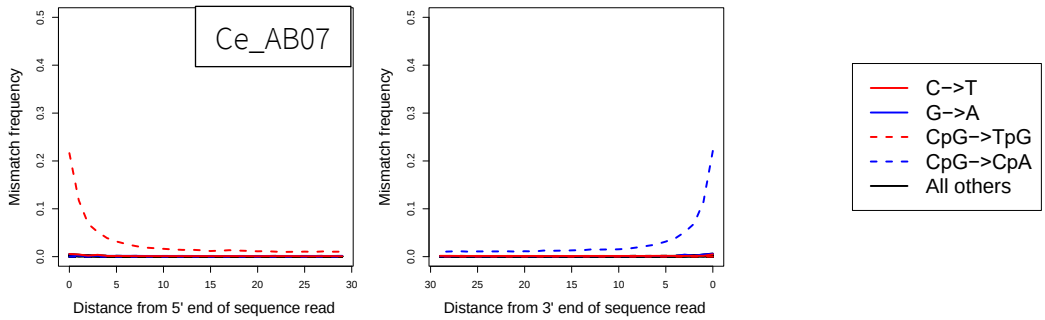
APPENDIX IV: SUPPLEMENTARY MATERIAL FOR CHAPTER 5

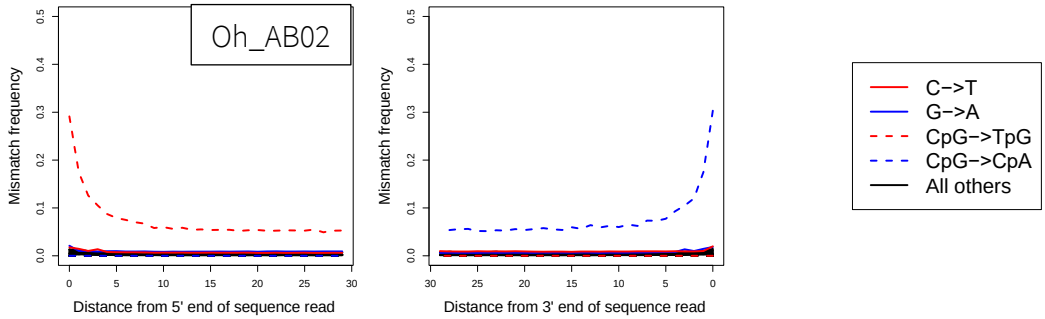
Figures

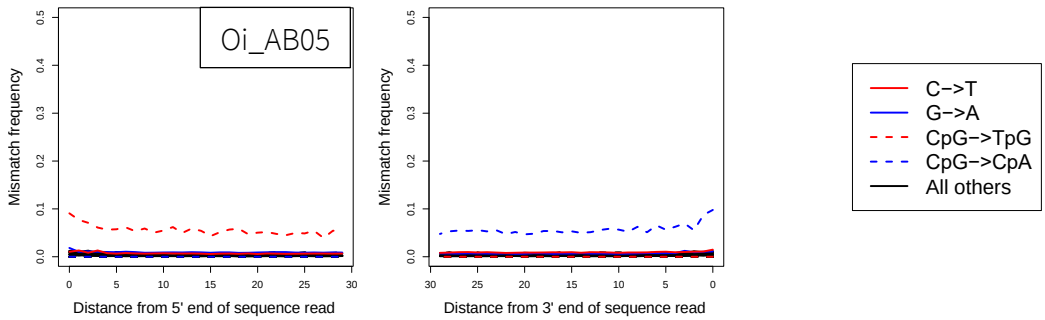
Figure S5.1: Damage patterns inferred in PMDtools in each ancient sample when mapped to their target species. Deamination at CpG sites identified by PDMtools is unaffected by USER enzyme treatment.

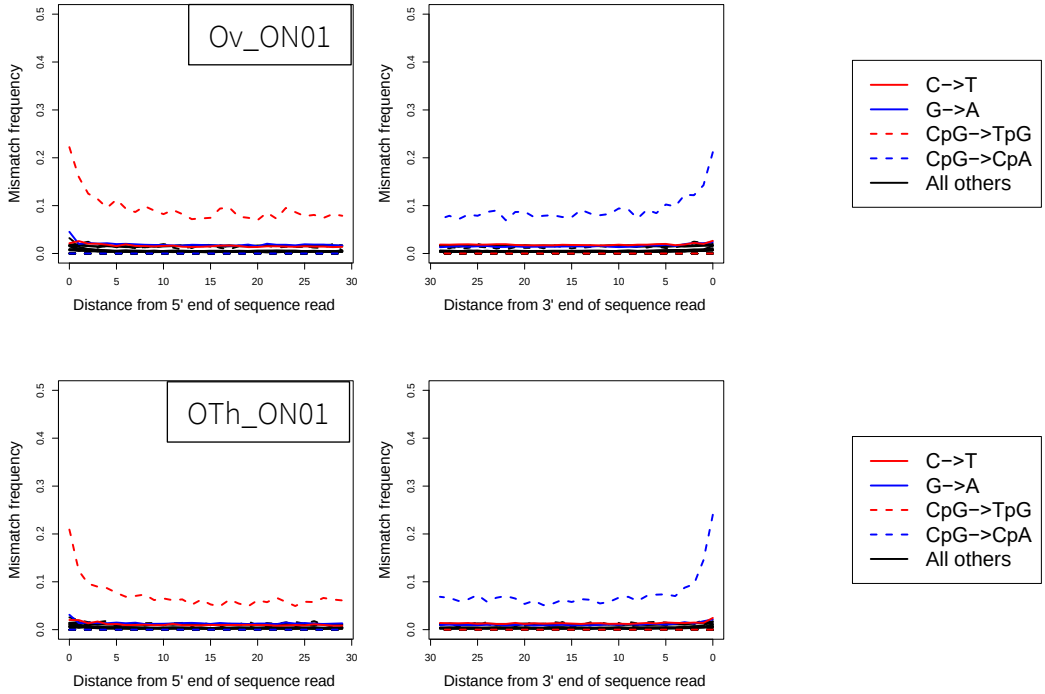



Ce_AB05
Mismatch frequency
Distance from 5' end of sequence read
Distance from 3' end of sequence read
C->T
G->A
CpG->TpG
CpG->CpA
All others


Ce_AB07
Mismatch frequency
Distance from 5' end of sequence read
Distance from 3' end of sequence read
C->T
G->A
CpG->TpG
CpG->CpA
All others


Oh_AB02
Mismatch frequency
Distance from 5' end of sequence read
Distance from 3' end of sequence read
C->T
G->A
CpG->TpG
CpG->CpA
All others


Oi_AB05
Mismatch frequency
Distance from 5' end of sequence read
Distance from 3' end of sequence read
C->T
G->A
CpG->TpG
CpG->CpA
All others


Ov_ON01
Mismatch frequency
Distance from 5' end of sequence read
Distance from 3' end of sequence read
C->T
G->A
CpG->TpG
CpG->CpA
All others

OTh_ON01
Mismatch frequency
Distance from 5' end of sequence read
Distance from 3' end of sequence read
C->T
G->A
CpG->TpG
CpG->CpA
All others

Tables

Table S5.1: Museum sample information. Short ID was used in paper and figures for clarity as it contains the species and broad location; specimen ID as used by the museums; provider abbreviation : ROM = Royal Ontario Museum and RAM = Royal Alberta Museum; Species as labelled in museum records; C14 age (BP) raw estimate from KCCAMS labs; Calibrated C14 age as computed in OxCal 4.4.

Short ID	Specimen ID	Sample source	Provider	Species	C14 age (BP)	Calibrated C14 age range BP (95.4%)	Mean calibrated C14 age BP
Aa_AB01	P19.385.6	Right mandible	RAM	*Alces alces*	9,980 ± 45	11,263 - 11,692	11477.5
Aa_AB03	P02.9.34	Left metacarpal	RAM	*Alces alces*	355 ± 15	319 - 482	400.5
Aa_ON02	ROMM33679	Upper jaw	ROM	*Alces alces*	205 ± 15	296 - out of range	NA
Ce_AB02	P02.9.67	Left tibia	RAM	*Cervus elaphus*	1,150 ± 15	975 - 1,175	1075
Ce_AB05	P85.10.1	Skull	RAM	*Cervus elaphus*	10,435 ± 50	12,058 - 12,612	12335
Ce_AB07	P18.236.4	Partial mandible	RAM	*Cervus elaphus*	3,770 ± 20	4,013 - 4,235	4124
Oh_AB02	P14.9.3	Right metatarsal	RAM	*Odocoileus hemionus*	3,905 ± 20	4,250 - 4,415	4332.5
Oi_AB05	P16.12.1	Right mandible	RAM	*Odocoileus sp. indet.*	155 ± 15	4 - 281	142.5
Ov_ON01	ROMM11048	Mandible	ROM	*Odocoileus virginianus*	305 ± 15	306 - 433	369.5
Rt_AB03	P94.1.297	Left metatarsal	RAM	*Rangifer tarandus*	27,330 ± 470	30,465 - 32,901	31683
OTh_ON01	ROMM75974	Antler	ROM	*Odocoileus (Torontoceros) hypogaeus*	11,315 ± 325		

Table S5.2: Modern sample information as retrieved from NCBI SRA metadata information. Sample ID designated for clarity.

Species	Sample ID	Accession	Locality	Sex
Alces alces ●	Aa-AK01	SRR6079187	Alaska	Male
Alces alces ●	Aa-ID01	SRR6079177	Idaho	Male
Alces alces ●	Aa-MI01	SRR18899225	Michigan	NA
Alces alces ●	Aa-MN01	SRR18899222	Minnesota	NA
Alces alces ●	Aa-SW01	ERR12087977	Sweden	NA
Alces alces ●	Aa-SW02	ERR12087904	Sweden	NA
Alces alces ●	Aa-VT01	SRR6079181	Vermont	Male
Alces alces ●	Aa-WY01	SRR6079200	Wyoming	Female
Cervus canadensis ●	Cc-WY01	SRR12450513	Wyoming	Female
Cervus canadensis ●	Cc-MN01	SRR12450505	Minnesota	Male
Cervus canadensis ●	Cc-MN02	SRR12450502	Minnesota	Female
Cervus canadensis ●	Cc-WY02	SRR12450506	Wyoming	Female
Cervus elaphus ●	Ce-HU01	SRR956941	Hungary	Male
Rangifer tarandus ●	Rt-BC01	SRR27590283	British Columbia	NA
Rangifer tarandus ●	Rt-BC02	ERR11471728	British Columbia	NA
Rangifer tarandus ●	Rt-ON01	ERR11471702	Ontario	Male
Rangifer tarandus ●	Rt-ON02	SRR24951207	Ontario	NA
Rangifer tarandus ●	Rt-NO01	SRR24951210	Norway	NA
Rangifer tarandus ●	Rt-NO02	SRR15459420	Norway	NA
Rangifer tarandus ●	Rt-RU01	SRR15459424	Russia	NA
Odocoileus hemionus ●	Oh_AB1	SRS12707053	Alberta	Female
Odocoileus hemionus ●	Oh_AK1	SRS12707054	Alaska	NA
Odocoileus hemionus ●	Oh_BC1	SRS12707032	British Columbia	Female
Odocoileus hemionus ●	Oh_BC6	SRS12707055	British Columbia	Male
Odocoileus hemionus ●	Oh_CA3	SRS12707057	California	NA
Odocoileus hemionus ●	Oh_OR2	SRS12707035	Oregon	NA
Odocoileus hemionus ●	Oh_SK2	SRS16492111	Saskatchewan	NA
Odocoileus hemionus ●	Oh_UT2	SRS12707038	Utah	Male
Odocoileus virginianus ●	Ov_ID1	SRS16492129	Idaho	Female
Odocoileus virginianus ●	Ov_Key4	SRS16492093	Florida	NA
Odocoileus virginianus ●	Ov_MO1	SRS16492102	Missouri	Male
Odocoileus virginianus ●	Ov_MX2	SRS12707043	Mexico	Male
Odocoileus virginianus ●	Ov_ONX2	SRS16492116	Ontario	Female
Odocoileus virginianus ●	Ov_PA3	SRS12707047	Pennsylvania	Female
Odocoileus virginianus ●	Ov_SK3	SRS12707051	Saskatchewan	Male
Odocoileus virginianus ●	Ov_TX20	SRS16492126	Texas	Female

Table S5.3: Reference genome information per species.

Reference	Samples mapped	Accession
Whole genome reference		
Alces alces	Moose	GCA_015832495.2
Bos taurus	All	GCF_002263795.2
Cervus canadensis	Wapiti and red deer	GCF_019320065.1
Rangifer tarandus	Caribou	GCA_949782905.1
Odocoileus virginianus	Moose, mule deer & white-tailed deer	GCA_014726795.1
Mitochondrial reference		
Alces alces isolate Kazakhstan mitochondrion, complete genome	Moose	NC_020677.1
Bos taurus mitochondrion, complete genome	All	AY526085.1
Cervus canadensis isolate 19FC017 mitochondrion, complete genome	Wapiti and red deer	NC_050863.1
Odocoileus virginianus mitochondrion, complete genome	Caribou	NC_015247.1
Rangifer tarandus platyrhyncus genome assembly, organelle: mitochondrion	Mule deer & white-tailed deer	OX460346.1

Table S5.4: Ancient samples nuclear genome QC; filtering QC refers to final read numbers using minQ > 25, Length >30, no duplicates; damage patterns visually inferred from plot.

	Mapping			Filtering		Coverage		Sex ID		Damage pattern
	Reference species	Mapped reads to self (#)	Mapped reads to self (%)	Fitlered reads (#)	Filtered read length (avg)	depth	# variant with depth >= 3	X/A ratio	Inferred sex	PMDtools
Aa_AB01	*Alces alces*	424790	11.5667	216876	58	0.1546	4467	0.519	Male	0.4
Aa_AB03	*Alces alces*	1283762	4.15461	556103	72	0.4878	12089	0.515	Male	0.1
Aa_ON02	*Alces alces*	713289	7.19809	298525	64	0.2338	7224	0.505	Male	0.1
Ce_AB02	*Cervus canadensis*	1205635	60.2798	603807	66	0.0963	2118	0.993	Female	0.1
Ce_AB05	*Cervus canadensis*	1382086	24.7406	486523	59	0.0696	1872	0.588	Male	0.25
Ce_AB07	*Cervus canadensis*	12311879	31.844	5516610	67	0.8911	25219	0.949	Female	0.2
Oh_AB02	*Odocoileus virginianus*	1502949	10.7382	828120	74	0.4225	23552	0.441	Male	0.3
Oi_AB05	*Odocoileus virginianus*	303780	10.3202	181249	84	0.1046	4118	0.456	Male	0.1
Ov_ON01	*Odocoileus virginianus*	176790	2.12721	96938	63	0.0421	1740	0.752	Female	0.2
Rt_AB03	*Rangifer tarandus*	105255	1.2914	56501	69	0.1623	162	0.846	Female	0.25
OTh_ON01	*Odocoileus virginianus*	134663	0.506636	82542	71	0.0401	1056	0.516	Male	0.2

Table S5.5: Ancient samples mitochondrial genome QC; filtering QC refers to final read numbers using minQ > 25, Length >30, no duplicates.

	Mapping		Filtering	Coverage	
	Reference species	Mapped reads to self (#)	Fitlered reads (#)	depth	%N in consensus sequence
Aa_AB01	*Alces alces*	1159	1055	4.5421	1.105
Aa_AB03	*Alces alces*	3429	2432	11.5671	0.257
Aa_ON02	*Alces alces*	1466	1074	4.7762	2.120
Ce_AB02	*Cervus canadensis*	228	216	1.5861	43.982
Ce_AB05	*Cervus canadensis*	1027	737	3.1776	5.980
Ce_AB07	*Cervus canadensis*	1719	1298	5.0664	1.270
Oh_AB02	*Odocoileus virginianus*	1310	1138	5.2301	1.606
Oi_AB05	*Odocoileus virginianus*	1312	1209	6.0828	1.041
Ov_ON01	*Odocoileus virginianus*	136	116	1.3198	65.717
Rt_AB03	*Rangifer tarandus*	1795	1304	6.0782	1.450
OTh_ON01	*Odocoileus virginianus*	252	237	1.7957	35.604

Table S5.6: Heterozygosity values per sample, filters comprise all sites recovered in ancient individuals, a minimum depth of 3, with and without removing transitions. Nsites indicates how many filtered sites were analysed per sample.

Sample	Species	Age (calBP)	heterozygosity in ANGSD			
			Depth >=3	nsites	Depth >=3, no transition	nsites
Aa_AB01	Moose	11,263 - 11,692	0.0641	43926	0.0402	30746
Aa_AB03	Moose	319 - 482	0.0644	184611	0.0399	131000
Aa_ON02	Moose	296 - out of range	0.0735	78401	0.0477	54694
Aa-AK01	Moose	Modern	0.0374	186831	0.0197	127453
Aa-ID01	Moose	Modern	0.0430	181350	0.0235	122686
Aa-MI01	Moose	Modern	0.0381	196304	0.0201	134634
Aa-MN01	Moose	Modern	0.0385	196743	0.0202	134956
Aa-SW01	Moose	Modern	0.0405	192609	0.0206	131566
Aa-SW02	Moose	Modern	0.0386	194731	0.0200	133035
Aa-VT01	Moose	Modern	0.0367	189847	0.0193	129724
Aa-WY01	Moose	Modern	0.0452	178171	0.0240	120477
Ce_AB02	Wapiti	975 - 1,175	0.0523	30993	0.0382	21377
Ce_AB05	Wapiti	12,058 - 12,612	0.0464	27665	0.0279	19757
Ce_AB07	Wapiti	4,013 - 4,235	0.0088	2150463	0.0040	1622741
Cc-MN01	Wapiti	Modern	0.0050	1983986	0.0019	1484773
Cc-MN02	Wapiti	Modern	0.0049	2037630	0.0019	1526754
Cc-WY01	Wapiti	Modern	0.0051	2038923	0.0020	1527576
Cc-WY02	Wapiti	Modern	0.0050	2035471	0.0020	1524555
Ce-HU01	Red deer	Modern	0.0111	424624	0.0040	317770
Oh_BC1	Mule deer	Modern	0.0367	134829	0.0193	98067
Oh_AB1	Mule deer	Modern	0.0331	141309	0.0177	102812
Oh_BC6	Mule deer	Modern	0.0339	137855	0.0179	100450
Oh_CA3	Mule deer	Modern	0.0331	139797	0.0178	101974
Oh_OR2	Mule deer	Modern	0.0356	141279	0.0186	102705
Oh_SK2	Mule deer	Modern	0.0344	138372	0.0185	100820
Oh_UT2	Mule deer	Modern	0.0364	130322	0.0192	94961
Oh_AK1	Mule deer	Modern	0.0382	134089	0.0198	97307
Oh_AB02	Mule deer	4,250 - 4,415	0.0483	146620	0.0285	109808
OTh_ON01	Toronto subway deer	11,315	0.0456	6126	0.0362	4183
Oi_AB05	Odocoileus indet.	4 - 281	0.0610	21333	0.0411	15436
Ov_ON01	White-tailed deer	306 - 433	0.0373	12143	0.0295	8506
Ov_ID1	White-tailed deer	Modern	0.0418	136079	0.0206	97294
Ov_Key4	White-tailed deer	Modern	0.0405	135856	0.0209	98248
Ov_MO1	White-tailed deer	Modern	0.0388	136763	0.0194	98367
Ov_MX2	White-tailed deer	Modern	0.0411	137339	0.0196	98622
Ov_ONX2	White-tailed deer	Modern	0.0435	135297	0.0215	96696
Ov_PA3	White-tailed deer	Modern	0.0407	132502	0.0204	95132

Ov_SK3	White-tailed deer	Modern	0.0359	142954	0.0178	103319
Ov_TX20	White-tailed deer	Modern	0.0414	136247	0.0200	97782
Rt_AB03	Caribou	30,465 - 32,901	0.0037	5116	0.0022	4290
Rt-BC01	Caribou	Modern	0.0028	3844	0.0007	3226
Rt-BC02	Caribou	Modern	0.0014	4233	0.0005	3563
Rt-NO01	Caribou	Modern	0.0000	3532	0.0000	2962
Rt-NO02	Caribou	Modern	0.0000	3325	0.0000	2834
Rt-ON01	Caribou	Modern	0.0050	3806	0.0029	3182
Rt-ON02	Caribou	Modern	0.0038	3743	0.0019	3141
Rt-RU01	Caribou	Modern	0.0060	4006	0.0000	3361