# The Palaeogenomics of Arctic and Sub-Arctic Peoples: A Study of Population Genetics, Adaptations, and Pathogen Incidence

**Alison Margaret Semple Sutherland**

**St. Edmund's College**

**University of Cambridge**

**This thesis is submitted for the degree of Doctor of Philosophy**

**February 2023**

# Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text. It is not substantially the same as any work that has already been submitted before for any degree or other qualification except as declared in the preface and specified in the text. It does not exceed the prescribed word limit for the Degree Committee for the Faculty of Biology.

Cambridge, February 2023

Alison Margaret Semple Sutherland

# Summary

## The Palaeogenomics of Arctic and Sub-Arctic Peoples: A Study of Population Genetics, Adaptations, and Pathogen Incidence

Alison Margaret Semple Sutherland

The northeastern regions of Siberia and the North American Arctic are some of the last regions of the world to be inhabited by humans; there, sophisticated technologies were developed for hunting marine mammals. Continued migration waves of ancient Western Eurasians and ancient East Asians into northeastern Siberia led to the early formation of the “Palaeo-Siberians" in the Late Pleistocene and the “Neo-Siberians" in the Holocene, the latter being genetically continuous with present-day groups in the region. The North American Arctic was populated by two genetically distinct, archaeologically-defined cultural traditions of Neo-Siberian-related peoples: the Palaeo-Inuit (entering ~5.5 thousand years ago) and the Neo-Inuit (entering ~1 thousand years ago). Limited archaeological and palaeogenomic findings into prehistoric contacts and admixture between ancient Siberian and Arctic groups leave a knowledge gap in the demographic histories of these regions. This thesis comprises palaeogenomic, radiocarbon, and stable carbon and nitrogen isotope datasets of unmatched quality and scale, generated from 217 sets of human remains from northeastern Siberia and North America to investigate population histories, evidence of adaptations to the Arctic environment, and pathogen incidence.

Using allele-frequency based methods, the population genetics analyses in this thesis investigate three main research questions pertaining to Palaeo-Siberian, Palaeo-Inuit, and Neo-Inuit groups. Genetic similarity within and between these groups was determined, adding insight into ancient migrations and population interactions. Adaptations associated with fat metabolism and cold were examined in the ancient Arctic and sub-Arctic groups, at genetic *loci* that have been proposed to be under selection in present-day populations from the region. Ancient pathogens were identified from the sequencing data of the ancient individuals, expanding the catalogue of human pathogens in these regions over time.

The findings from this thesis elucidate the population histories of Arctic and sub-Arctic groups over time. Importantly, through continued community engagement and knowledge exchange with Indigenous peoples, this interdisciplinary project tells a more complete history of the peopling of the Siberian and North American Arctic.

*In dilectam memoriam aviarum mearum,*

Elizabeth Semple (*nata* Duffy) *et Doctoris* Eleanor Sutherland (*nata* Miller)

*Fons consilii et inspirationis semper erant*

## *Acknowledgments*

I would like to extend my appreciation and gratitude to a number of people:

To my supervisor, Prof. Eske Willerslev for the guidance, opportunities, and support; to my mentor, Dr. Maanasa Raghavan for the inspiration, kindness, patience, and pedagogy; to the postdocs, Dr. Hugh McColl and Dr. Victor Moreno-Mayar for the encouragement and endless help.

To my family for the limitless love and support.

To the Rothermere Foundation for this opportunity of a lifetime.

To the ancient individual whose remains are included in this work; to the Indigenous groups who supported this research, the Russian Association of Indigenous People Of the North, NIMA Corporation, Qanirtuuq Inc, Inuit Heritage Trust, Avataq Cultural Institute, Council of the Huron-Wendat Nation, Listuguj Mi'gmaq Council, Innu Nation, Maiwpukek First Nation, Nunatsiavut Government Research Advisory Committee, Qalipu First Nation, and the Greenland National Museum. With special thanks to Dr. Ripan Malhi for the guidance on interacting with Indigenous groups.

To all those who helped at the Lundbeck Foundation GeoGenetics Centre, Dr. Anna Razeto, Dr. Line Oslen, Pernille Vibeke Selmer Olsen, Judy Erichsen, Jesper Stenderup; to those who helped with laboratory advice and assistance Dr. Lasse Vinner, Dr. Charleen Gaunitz, Laerke Daniela Kjaersgaard Hansen, and Maria Madrona, who were also responsible for running the library preparation and library purification robots; to those at the Sequencing Centre, Dr. Mette Juul Jacobsen, Tina Blumensaadt Brand, and Julie Bitz-Thorsen, who were responsible for running the Illumnia sequencing machines and demultiplexing the sequencing data; to the Lundbeck Foundation Grant for funding the lab supplies.

To those who oversee the Lundbeck bioinformatics pipeline Dr. Abigail Ramsøe and Dr. Thorfinn Sand Korneliussen; for the help with the species identification Jie Chen and Dr. Yucheng Wang; for the assistance with the Y-chromosome haplogroup analysis from Thomaz Pinotti; for the assistance with the population genetics analyses, Dr. Victor Moreno-Mayar; for the advice on authentication of ancient pathogens from Dr. Frederik Valuer Seersholm; for the assistance with the pathogen pipeline and phylogenetic placement of the ancient variola virus from Dr. Martin Sikora; to the professors who advised me Dr. Ana Prohaska, Dr. Andrea Manica, Dr. Derek Smith, Dr. Morten Melgaard, and Dr. Terry Jones.

To Father Alban McCoy, Gordon Chesterman, and Dr. Michele Gemelos for the welfare support; for fidelity in the face of immorality, insidiousness, and injustice.

To my friends who supported and encouraged me, Dr. Aastha Dahal, Othman Al Bahri, Tharsika Vimala, Linnea Wesslén, Johannes Hoerning, Anna-Marie Robertson, Dr. Marianna Sykopetritou, Nicolas William Rhodes, Ella Davis, Raghav Rayasam, Dr. Riva Riley, Dominique Rain Bowden, and Robbie Whitefield. With special thanks to those who helped edit this thesis, Matthew Neville, Dr. Maria Novosolov, Saashi Bedford, Connie Elworthy, and Shaye Murray. And to Dr. Ahmad Khan, for standing by my side and believing in me.

# Table of Contents

# 1 Chapter 1 - Introduction

## *1.1 Outline of the Thesis*

This thesis is composed of three chapters of work pertaining to ancient Arctic and sub-Arctic peoples and this current introductory chapter, which outlines relevant terminology, population groups, and genomic concepts, while also providing background on aspects of this large interdisciplinary project. The second chapter describes the stepwise process of generating a dataset for applications in the field of palaeogenomics. The third chapter entails population genetic analyses performed on the newly generated Arctic and sub-Arctic genomes, which further elucidate the maternal and paternal lineages, population structure, genetic origins, and genetic admixture between these ancient groups. The third chapter continues with an examination of changes in allele frequencies in these populations, as they relate putatively to exposure to environmental pressures. The fourth chapter outlines the pathogens identified in the ancient tissues included in this project, with three ancient pathogens detailed as case studies.

## *1.2 A Note on Nomenclature*

Historically, the term ‘Eskimo’ was used to describe Indigenous individuals in the northeastern Siberia and North American Arctic. This name is thought to have originated from subArctic speakers of the Algonquin language family (1). Following the formation of the 1977 Inuit Circumpolar Council, which include the Yupik (Siberia), Inupiat and Yupik (Alaska), Inuit and Inuvialuit (Canada), and the Kalaallit (Greenland), leaders rejected the use of the term Eskimo as it is culturally insensitive and have put forward efforts to abolish it (1). While Palaeo-Eskimo and Neo-Eskimo are often used in archaeological/scientific literature as well as the Eskimo-Aleut language family in linguistic literature, the terms Palaeo-Inuit, Neo-Inuit, and Inuit-Aleut will be used where possible in this thesis. Palaeo-Inuit and Neo-Inuit are purely cultural labels as they are not genetically continuous populations (2).

Similarly, the term Native American is becoming outdated because of its vagueness and is therefore largely being replaced by the term Indigenous when referring to the first peoples of the Americas (1). There are many groups and individuals from Siberia and North America who still self-identify as Eskimo, Native American, AmerIndian, or Indian; however, for the purpose of this thesis, all ancient population groups will (where possible) be referred to by archaeological cultural affiliation, or as Indigenous (1,3).

## *1.3 Human History in Northeastern Siberia and the North American Arctic*

Humans have occupied regions of the Arctic in Siberia and North America for more than 40000 and 5500 years, respectively (2,4). The people of the tundra, that is, those who live above the treeline, have adapted to sustaining life in the harsh Arctic climate (**Figure 1.1**) (1,5). Arctic peoples developed sophisticated technologies for hunting marine mammals, namely nautical vessels and toggling harpoons (1,5,6). This was associated with a shift in dietary composition to that of high-fat, protein-rich subsistence (4). In addition to cultural and technological innovation, biological adaptations occurred in metabolic pathways enabling these populations to survive and thrive on a marine-mammal based diet, allowing for rapid migrations into the vast unpopulated regions of the North American Arctic (2,7–10).

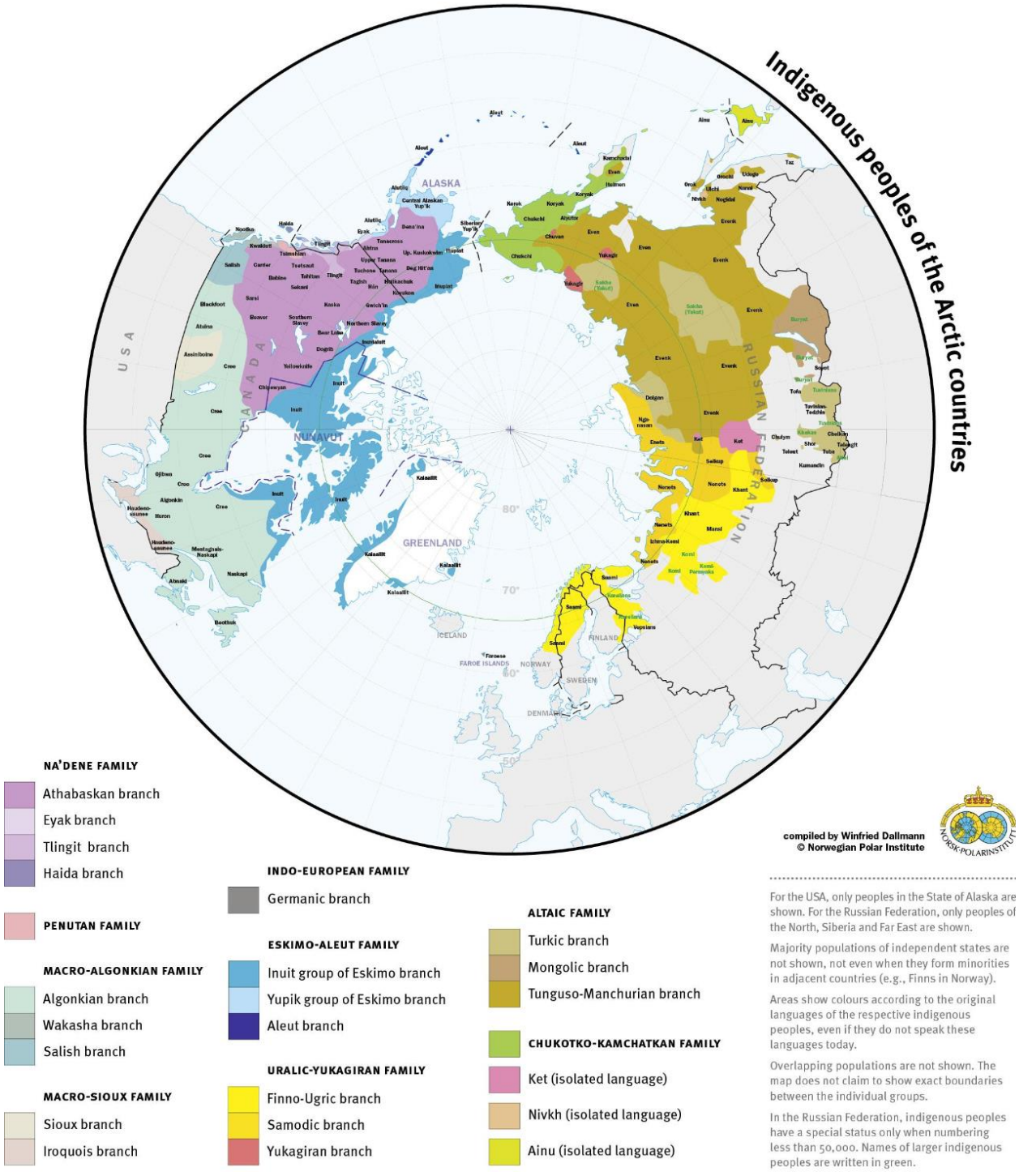



**Figure 1.1** - Map of the Indigenous peoples of the Arctic countries, detailing the language families spoken in regions of the Circumpolar North. Map from (11).

### 1.3.1 Present-Day Indigenous Population Groups of Siberia and North America

**Siberia**

Siberia refers to a vast region of central and eastern Russia that often has context-dependent boundaries. For the purpose of this thesis, Siberia refers to the Russian federal constituent entities of Yakutia (otherwise known as Republic of Sakha, shown in pink), the Magadan Oblast (shown in purple), the Kamchatka Territory (shown in dark green), and the Chukotka Autonomous Area (shown in light green), from which the ancient tissues included in this project originate (**Figure 1.2**) (12).


**Figure 1.2** - Map of Siberia, indicating the four federal constituent entities relevant in this work, Yakutia in pink, Magadan in purple, Kamchatka in dark green, and Chukotka in light green. Map adapted from (12).

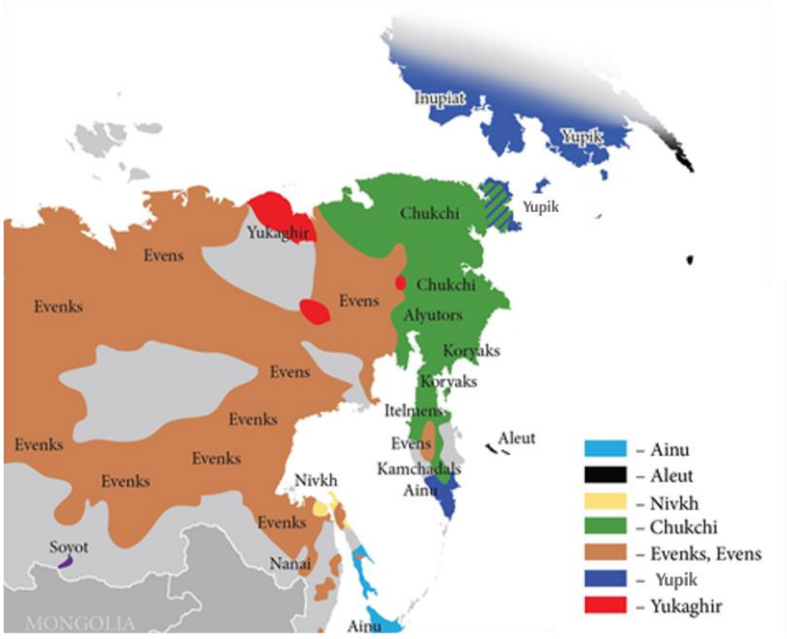



**Figure 1.3** - Map of the Indigenous regions of Siberian, depicted based on cultural groups. Most relevant to this project are the Chukchi, the Koryaks, the Itelmens, and the Yupik. Map adapted from (13).

The present-day Indigenous groups of these four regions include the Dolgany, Evenki, Even, Yukaghir, Kamchadals, Alyutors, Koryak, Itelmens, Aleut, Kereks, Chuvans, Chukchi, and Yupik (3). The Indigenous regions, outlined based on cultural group, are broadly depicted in **Figure 1.3** (13). Of particular relevance to this project are the Koryak, the Itelmens, the Chukchi, and the Yupik. The

Koryaks are a group of ~8700 individuals who inhabit coastal regions of Kamchatka and Magadan and are avid marine hunters (5). The Itelmens are a group of ~3200 individuals who occupy vast regions of Magadan, Kamchatka, and Chukotka, who are freshwater fishers as well as marine mammal hunters (5). The Chukchi are a group of ~15800 individuals, who predominantly live in Chukotka (as well as adjacent territories) and fall into two broad groupings: those that live inland and engage in reindeer herding, and those that live on the coast and hunt sea animals (5). The Yupik are a group of ~1800 individuals who live along the coast of Chukotka and the Arctic coast of North America, and are hunters of walrus, seal, and whale (5).

## Alaska

Alaska is occupied by Indigenous peoples that are broadly categorised into five groups: the Inuipiaq in the north, the Yupik and Cupik of the southwest, the Unangax and Alutiiq of the southern archipelago, the Athabascans in south-central interior, and the Eyak, Haida, Tsimshian, and Tlingit down the western coast of North America (**Figure 1.4**) (14). The Indigenous group most relevant to this work is the Yupik. The Yupik are a group of ~34000 individuals who are skilled marine mammal hunters and freshwater fishers (14).


**Figure 1.4** - Map of the Indigenous regions of Alaska, outlined based on population groups. Most relevant to this project at the Yupik. Map from (14).

## Canada

Canada's Indigenous peoples include the First Nations, Métis, and Inuit (1). The First Nations constitute Indigenous groups below the Arctic, and the Métis are admixed cultural group of early settlers and various first peoples groups (1). The Inuit are the people of the Arctic region

of Canada, known as Inuit Nunangat, comprising ~70500 individuals. Nunangat can be subdivided into four regions where different dialects of Inuktitut are spoken: Inuvialuit (spanning the northern coast of the Yukon and Northwest Territories), Nunavut (an autonomously governed Inuit territory), Nunavik (regions of northern Quebec, otherwise known as Avataq), and Nunatsiavut (northern coastline of Labrador) (**Figure 1.5**). While there are regional dialects of Inuktitut in Nunangat, they share the same tradition of sophisticated marine hunting (1).


**Figure 1.5** - Map of the regions of Inuit Nunangat (northern Canada): Inuvialuit, Nunavut, Nunavik, and Nunatsiavut. Map from (1).

**Greenland**

The coastal regions of the nation of Greenland, known to the Indigenous peoples as Kalaallit Nunaat, is inhabited by ~51000 Inuit (15). The Inuit in Greenland, who traditionally have been avid hunters of sea animals (namely seals) are broadly categorised into three groupings: the Kalaallit of the western coast, the Tunumiit of the east coast, and the Inughuit of the northern/ polar regions (depicted in blue **Figure 1.6**) (15).


**Figure 1.6** - Map of Kalaallit Nunaat (Greenland), indicating the regions inhabited by the three Inuit subgroupings: the Kalaallit, the Tunumiit, and the Inughuit. Map adapted from (11).

### 1.3.2 Population Genetics in Ancient Northeastern Siberia and the North American Arctic

As a field, population genetics examines genetic variation within and between population groups. Population genetics has been mainly concerned with the genetic diversity and population structure in present-day populations; however, the recent advent and rapid development of ancient DNA sequencing techniques has helped advance the understanding of past population demography such as admixture events, population replacements, and population bottlenecks (16,17). Novel population genomics methods and an increasing number of ancient Palaeogenomics datasets have enabled the reconstruction of past population sizes (18), population structure (19), genetic origins (20,21), split times (22), and genetic admixture between populations (23,24).

Population genetic studies of ancient northeastern Siberian and northwestern North American individuals have elucidated aspects of population origins, admixture, and migrations. The first known migration into this geographical region was by ancient western Eurasian Hunter-Gatherers and took place prior to 30 thousand years ago (kya) (25,26). These groups are known as the "Ancient North Siberians" (ANS), represented by the Yana individual from north-central

Sahka (Yakutia), and the Mal'ta and Afontova Gora individuals from south-central Irkutsk (25,26). The second large population migration was an East Asian-related group who admixed with ANS, giving rise to the ancestors of most present-day Indigenous Americans, which took place between 25 and 10 kya (25).

The ancestral population of the Indigenous peoples of the Americas diverged into three main genetic lineages: Ancient Beringians (AB), North Native Americans (NNA) and South Native Americans (SNA) (22,27,28). This demographic model was constructed following the sequencing of an AB genome from a ~11.5 kya child from central Alaska, which provided supportive evidence for an ~8000 year period of isolation in Beringia prior to entering the Americas (27,29). Furthermore, persistent gene flow (where genetic material from one population group enters another) from an East Asian-related population into northeastern Siberia during the Late Pleistocene and early Holocene led to a metapopulation termed "Palaeo-Siberians", represented by the 9.8 kya Kolyma individual from north-eastern Yakutia (25).

The third migration wave into northeastern Siberia took place after 10 kya, where another East Asian-related group migrated eastward, and became the "Neo-Siberians". These "Neo-Siberians" largely replaced the "Palaeo-Siberians", and share genetic continuity with present-day peoples in Siberia (25). An additional gene flow event took place into NNA, after ~11.5 kya, from a population who is genetically similar to the ANS and present-day Koryaks (25).

The North American Arctic has a distinct occupational history compared to the rest of the Americas. This region was populated much later by two distinct archaeologically-defined cultural traditions from Siberia: the Palaeo-Inuit who migrated into the North American Arctic ~5.5kya (represented by the ~4kya ancient Saqqaq individual from western Greenland (30)), and the Neo-Inuit who arrived in the North American Arctic in a later wave ~1 kya. Over a ~4500 year period, the Palaeo-Inuit migrated and occupied most of the coastal regions of the North American Arctic, and began to disappear from the archaeological record ~1000 years ago (31). These peoples have been archaeologically classified into the Saqqaq (in Greenland) and Pre-Dorset and Dorset Cultures (in Canada). The Palaeo-Inuit, represented by the ~4 kya Saqqaq individual from Greenland, can be modelled as a mix of Palaeo-Siberians and additional 20% East Asian ancestry

(25), the latter represented by ~8 kya genomes from Devil's Gate (32). The Neo-Inuit began to enter the North American Arctic ~1 kya and are reported as deriving genetic ancestry from NNA and a lineage closely related to the Saqqaq (2,27). This group later transitioned into present-day Inuit, to whom they are genetically and culturally ancestral (2). The archaeological record is inconclusive regarding contact between late Palaeo-Inuit and early Neo-Inuit groups (6,33).

There are still several unanswered questions relating to the palaeogenomics of northeastern Siberian and northern North American groups, which will benefit from higher depth genomes. Establishing migrations and genetic admixture between these ancient Arctic populations can contribute to refining their population histories and population structures, as well as resolving the cultural versus biological overlaps (2).

### 1.3.3 Adaptation in Northeastern Siberia and the North American Arctic

Over time and space, humans have undergone various adaptations to their environments. The field of palaeogenomics, when complemented by existing evidence from disciplines such as archaeology and anthropology, lends insights into genomic changes that may be associated with environmental pressures (34,35). These genomic changes are quantified by differences in allele frequencies over time, within or between population groups. This process can be due to selection where changes in allele frequency occurs through traits being selected for or against in the process of evolution, or can be due to genetic drift where allele frequency changes occur by chance events. These two processes can be difficult to discern, especially when a group has a small effective population size, such as in Arctic populations, where drift becomes stronger in driving allele frequency trajectories; therefore, additional considerations, such as variants in linkage blocks or in genes/pathways associated with a certain trait, must be taken into account before inferring that a change in allele frequency is the product of selection (7,36).

Arctic populations have adapted to regional environmental factors such as cold, diet, and exposure to novel infectious diseases through long-term survival in the harsh Arctic climate, where resources can be scarce and population sizes are small (7,36). Many studies have examined selection in Arctic populations by focusing primarily on adaptations to cold and changes in dietary

composition. Present-day Siberian and Canadian Inuit genomes show one of the strongest selective sweeps to have occurred in humans, where selected variants in the *CPT1A* gene led to advantageous adaptation to metabolising high-fat diets as well as resilience to cold climates (9,37). Similarly, present-day Siberian populations show evidence for positive selection for genes involved in brown adipose tissue production (*PLA2G2A*, *PLIN1*, and *ANGPTL8*) (10). Additionally, the genomes of present-day Greenlandic Inuit show the strongest signal for selection to be in the *FADS2* gene, a known metabolism gene that codes for a protein that assists in regulating the level of polyunsaturated fatty acids in the blood, which can affect an individual's height and weight (8).

Signals of selection in present-day genomes can be confounded by selective pressures that occurred in the distant past; however, palaeogenomics makes it possible to examine changes in allele frequencies over time and detect when a selected allele may have reached fixation. Candidate genes, such as those known to have substantial differences in allele frequencies, can be examined in ancient populations to estimate the timing, origin, and magnitude of selective pressures.

### 1.3.4 Ancient Pathogen Identification in Eurasia and the Americas

Infectious diseases have affected humans throughout history, but accurately identifying the true pathogenic agent is difficult to discern from skeletal pathology and historical records alone (38). Palaeogenomics can provide direct evidence from past epidemics through identifying pathogen genomes which have been obtained as a by-product of generating high quality human genomes (39). Teeth have been found to have high post-mortem DNA preservation levels (40,41) and, because of the high levels of circulation that reaches the oral cavity, the layer at the bottom of the tooth root (the cementum) is targeted in palaeogenomics studies as the best source of ancient pathogen genetic material (42).

Many investigations into ancient pathogens focus on bacteria, viruses, and protozoa (39,43). Because of the overall low incidence of pathogen recovery–due to factors such as low disease incidence, poor DNA preservation, or varying levels of the pathogen in the individual at

the time of death–can lead to low ancient pathogen recovery; therefore, several ancient pathogens studies have focused on bacteria that cause chronic infection, such as *Mycobacterium leprae* or *Mycobacterium tuberculosis* (44). Other ancient pathogen studies pertain to major human pandemics, such as the causative agent of the Black Death, *Yersinia pestis,* or the 1918 Spanish Influenza, A/H1N1 (45,46).

Ancient Indigenous populations from Siberia and the Americas are reported to have experienced low levels of transmissible illnesses through time due to populations having remained small and low density (47,48). However, genomic sequencing of ancient individuals from northern Eurasia and the Americas have identified several ancient pathogens, including bacterial species such as *Yersinia pestis* and *Mycobacterium tuberculosis*, as well as viral species such as parvovirus B19, variola virus, and hepatitis B virus (49–53). *Yersinia pestis* was identified in ancient Altai individuals dating to ~4800-4700 years before present (BP), showing evidence of an epidemic where fleas were not the zoonotic agent (50,51). *Mycobacterium tuberculosis* was found in three pre-Columbian individuals from southern Peru dating to ~1000 BP; interestingly, this strain is genetically closer to those found to infect seals and sea lions than to the known human-adapted forms and is therefore the suspected mode of transmission (52). A time transect of the parvovirus B19 virus was identified in three ancient individuals from Lake Baikal, Siberia, spanning four millennia (~5500 - ~1500 BP) (54). The variola virus was isolated from a ~300 year old mummy from the Sakha region of Yakutia, Siberia, and has been found in other regions of Eurasia (55,56). The hepatitis B virus (HBV), was identified in an individual from ~2000 BP in Mongolia (49), and in a 400-year-old mummified child from Korea, exemplifying its vast geographic distribution (53).

European colonists carried with them diseases such as measles, influenza, pertussis, smallpox, and intestinal infections–leading to significant mortality amongst Indigenous groups who were immunologically naive (57,58). To this day, Indigenous populations in Siberia and the Americas suffer from outbreaks of diseases such as *Mycobacterium tuberculosis* (at 10 times the national rate), meningitis, and sexually transmitted infections (58–63). Additionally, infections acquired from traditionally-prepared stored foods are common in these groups, such as *Clostridium botulinum* (found in fermented meat) or *Trichinella* (found in hunted game meats) (58,64,65), as

well as parasites acquired from Arctic fish species, namely *Diphyllobothrium* and *Corynosoma* (58).

Very few ancient human pathogens have been recovered from Siberia or the North American Arctic (39), leading to a poor understanding of the evolution of human pathogens and infections, which may impact human health in this region. High-quality pathogen genomes aid in reconstructing the evolutionary trajectories of pathogen strains that can lend insights into epidemiological factors such as virulence and regional adaptation. Additionally, consistencies between pathogen profiles of different Arctic and sub-Arctic populations could supplement analyses where population contacts are detected or even debated, particularly when minimal or no admixture is detected (66).

# 2 Chapter 2 – Palaeogenomic Dataset Generation

## *2.1 Introduction*

### 2.1.1 Cultural Affiliations of Ancient Tissues

All relevant information for the tissues included in this project is presented below and will be referred to throughout this thesis. The site and cultural group information is presented from west to east, and from oldest to youngest, where there are two tissues from the same region. Each tissue is presented with the registration number for the Lundbeck Foundation GeoGenetics Centre (CGG), the tissue type, the site name, the location, and the cultural affiliation as assigned by archaeological collaborators. Tissue information is additionally tabulated in Supplementary **Table S1-S8**.

#### Siberia

The majority of the Siberian tissues processed in this project were contributed by Dr. Pavel Grebenyuk in association with the North-East Interdisciplinary Scientific Research Institute, Far East branch of the Russian Academy of Sciences. Dr. Grebenyuk contributed 27 bones, three locks of hair, one petrous, and 21 teeth. Additionally, Dr. Vladimir Pitulko from the Palaeolithic Department, Institute for the History of Material Culture, Russian Academy of Sciences, contributed five bones, one tooth, and one lock of hair from the Zhokhovskaya site on the New Siberian Islands and Dr. Kirill Dneprovsky from the State Museum of Oriental Art in Moscow, Russia contributed five locks of hair from the Paipelghak site in Chukotka (**Figure 2.1**).

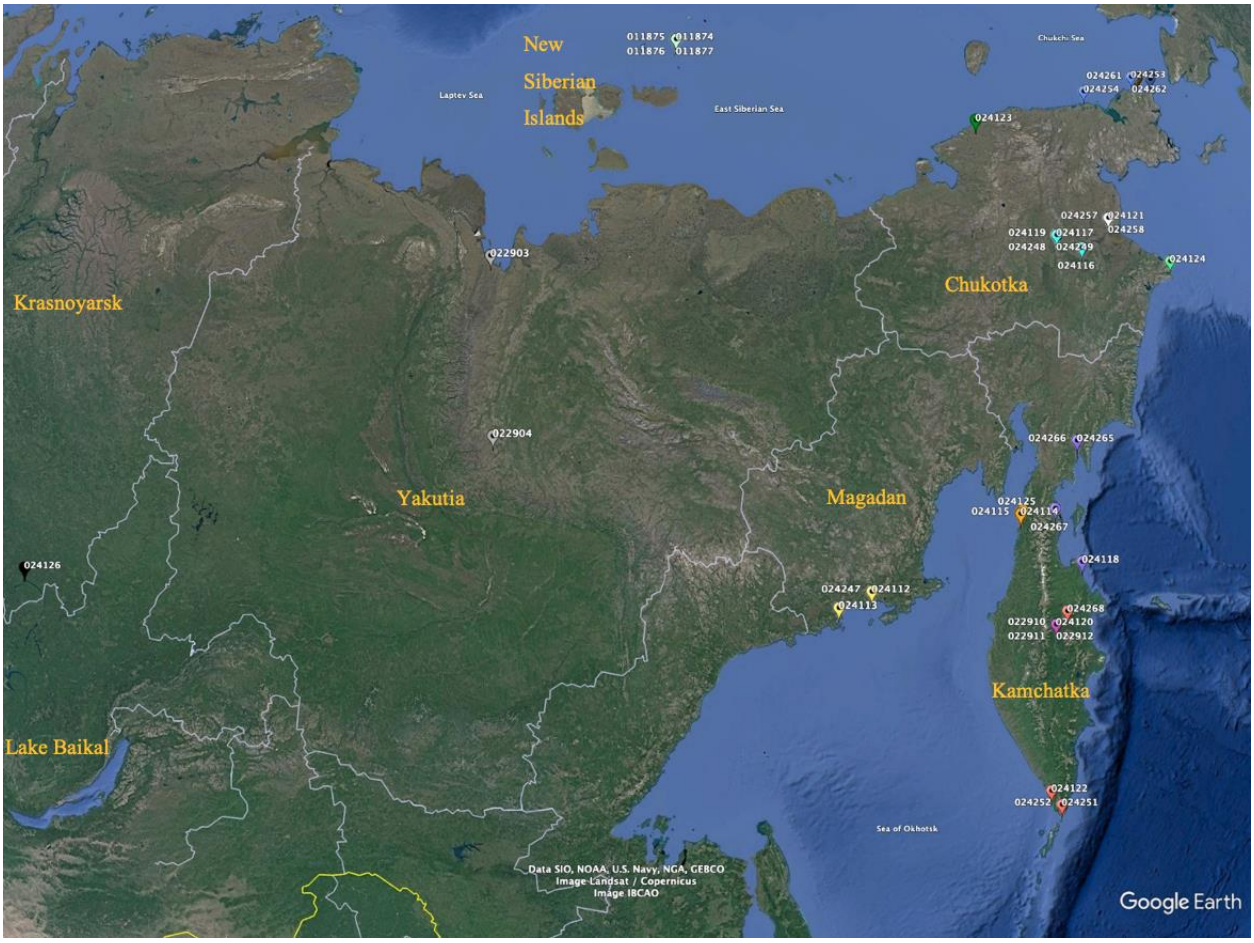



**Figure 2.1** - Map of eastern Russia, indicating the locations where the ancient Siberian tissues included in the project originated (67).

## Early Neolithic Culture

A single tooth from the Kutarey River Mouth site, located on the Angara River in the Krasnoyarsk Territory ~1000 km north of Lake Baikal, is a representative of the early Neolithic population in central Siberia (black pin, 024126, **Figure 2.1**). Archaeological evidence shows similarity in blade points between this site and the upper Kolyma early Holocene complex seen in Chukotka, Yakutia, and Kamchatka from ~8800–6000 BP (68).

### Surface Findings from Yakutia

Two bones from northern Yakutia were found at two different sites with minimal archaeological context surrounding them. One bone was found on the bank of the Omoloy River, 25 km above the mouth, near Timirdyakh-Khaya (light grey pin, 022903, **Figure 2.1**). The other bone was found on the bank of the Lena River near the Peschanaya Gora, about 120 km north of the city of Yakutsk (grey pin, 022904, **Figure 2.1**).

### Zhokhov Site

Five bones, one tooth, and one lock of hair, originating from the Zhokhovskaya site on the New Siberian Islands, were included in this project as representatives of the Mesolithic Sumnagin Cultural Complex (light aqua pins, 011874-80, **Figure 2.1** + **Figure 2.2**) (69). This site represents one of the most northern sites of ancient human occupation in the archaeological record, dating to ~8000 BP (69). Though these individuals lived at the coast, they had not adopted a lifestyle of marine mammal hunting and instead subsisted primarily on reindeer, polar bear, and fish (70). This site also holds the oldest sled dog technology on record (71).


**Figure 2.2** - Map of the New Siberian Islands, indicating the location where the ancient Zhokhov tissues included in the project originated (67).

### Tokarev Culture

Five bones from two sites were contributed as representatives of the Tokarev Culture, who inhabited the northern coasts of the Sea of Okhotsk ~2800-1500 BP (yellow pins, 024112-3, 024247, and 024263-4, **Figure 2.3**) (68,72). The Tokarev Culture is known for their marine based


**Figure 2.3** - Map of Magadan and the Kamchatka Peninsula, indicating where the ancient tissues from this region originated (67).

lifestyle in particular their seal hunting (most notably the toggling harpoon points) and fishing technologies (73). It has been suggested that a Tokarev-like tradition may have influenced the coastal cultures of the North Pacific, Bering Sea, and North American Arctic (74).

### Old Itelmen Culture

Two teeth and two bones from three sites from central and southern regions of the Kamchatka Peninsula were contributed as representatives of the Old Itelman Culture (red pins, 024122, 024251-2, 024268, **Figure 2.3**). The Old Itelman Culture is thought to have descended from the Tarya Culture (see below), with the transition taking place ~1000 BP (68,75,76). Assemblages from the earlier Old Itelmen sites include the toggling harpoon head, ski fragments, and items made of whalebone, indicating their subsistence activities included fishing, as well as land and sea animal hunting (68).

### Tevi Culture

Two locks of hair and one tooth originating from two sites on the northwest region of the Kamchatka Peninsula were contributed as representatives of the Tevi Culture (orange pins, 024114-5, 024125, **Figure 2.3**). The Tevi tradition, which is thought to have originated ~1000 BP, has archaeological evidence similar to Neo-Inuit and Aleutian cultural sites (68). It has been hypothesised that the Tevi Culture is the northwest counterpart to the Old Itelman and Old Koryak Cultures (75,77).

### Late Ushki Culture

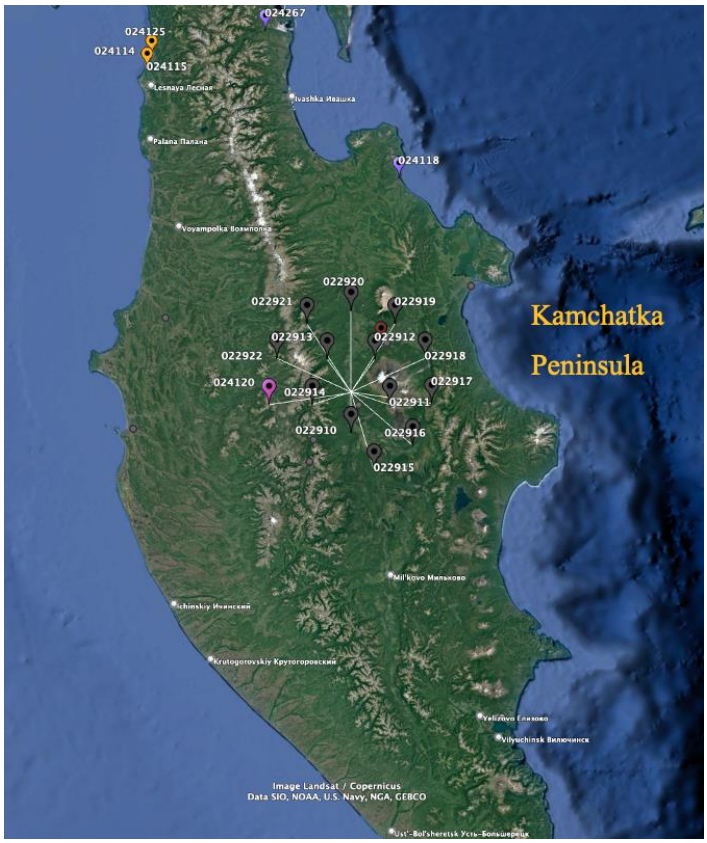



**Figure 2.4** - Map of the Kamchatka Peninsula, indicating where the ancient tissues from this region originated (67).

Thirteen small bags of fragmented bones, including the fragments of the crown of a single milk tooth, from the Ushki site located in the centre of the Kamchatka peninsula, were included in this project as representatives of the Late Ushki Culture of the Upper Palaeolithic period (dark grey pins, 022910-022922, **Figure 2.4**). The Ushki site consists of seven layers spanning four time periods, the Upper Palaeolithic, the Mesolithic and Neolithic, as well as the Early Metal Epoch (76,78). The Ushki site plays a pivotal role in the research of ancient northeastern Siberia. The complex of the Ushki sites contains both evidence of a non-microblade tradition (the Early Ushki Culture), and the Beringian tradition (the Late Ushki Culture). The Late Ushki cultural layer holds evidence of Beringian-like traditions, based on which it has been suggested to be the source of a late Pleistocene migration of *Palaeo-Inuit*-Aleutian people towards North America (79).There is archaeological evidence of overlap between cultural groups at the Ushki site and Mesolithic complexes in Chukotka, as well as links with Athabascan ancestors in North America (80).

### Tarya Culture

A single tooth from layer two of the Ushki site was contributed as a representative of the Tarya Culture (pink pin, 024120, **Figure 2.4**). The territory of the Tarya spreads from central to southern Kamchatka. Based on archaeological evidence, the Tarya people subsisted on fishing and

seal rookery hunting, which was discerned by the bone-made barbed harpoon heads found at Tarya sites (68). The presence of labret lip adornments suggests a proximity to the *Palaeo-Inuit*-Aleut Cultures (68). The evidence of seal hunting and consistency in embellishments suggest that Tarya Culture influenced the development of the Old Itelman and Old Koryak Cultures (75,76,81).

### Old Koryak Culture

Four teeth from three sites in the northeast of the Kamchatka Peninsula were included in this project as representatives of the Old Koryak Culture (purple pins, 024118, 024265-7, **Figure 2.3**). The Old Koryak Culture appeared ~1000 BP and is thought to have been formed by the Tokarev Culture. The early Old Koryak sites include bone-made toggling harpoons similar to those used by the Palaeo-Inuit populations in the North American Arctic, as well as evidence that their dwelling structures were built from whalebones (82). Archaeological evidence of the later stages of the Old Koryak Culture is dominated by seal hunting technologies, with the reduction in lithic tools and an increase in bone harpoon heads, with some containing slots for insertion of iron points (68).

### Pegtymel Complex

One tooth from the Pegtymel site located in Northern Chukotka was contributed to represent the Pegtymel Complex (dark green pin, 024123, **Figure 2.5**). This site is one of three Pegtymel Neolithic sites associated with the Pegtymel petroglyphs from the Bronze

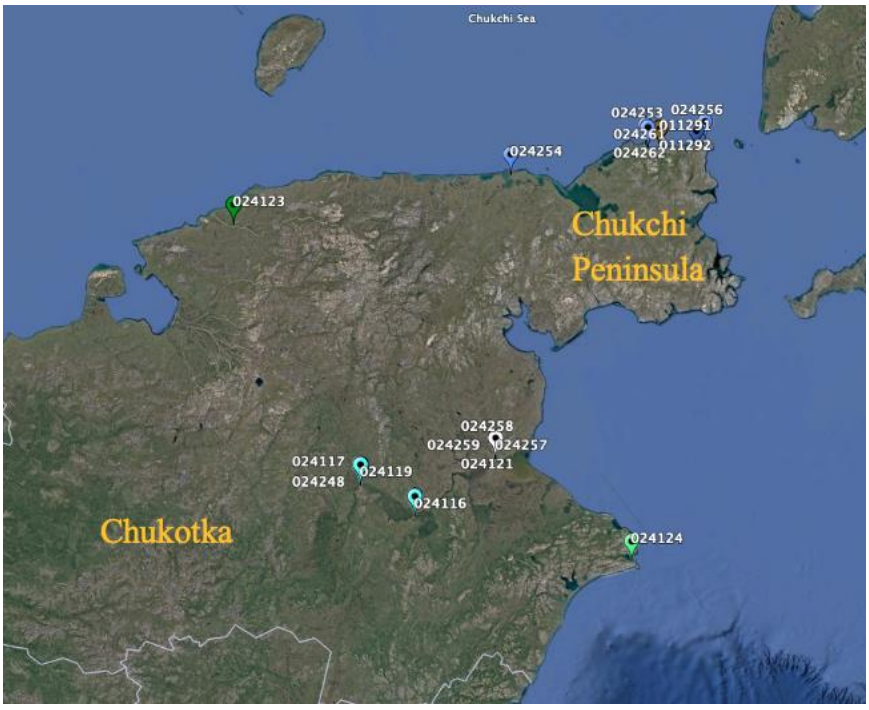



**Figure 2.5** - Map of Chukotka, including the Chukchi Peninsula, indicating where the ancient tissues from this region originated (67).

Age (83). The Pegtymel Complex is thought to have influenced the Palaeo-Inuit and the ancestors of the present-day Chukchi (76).

### Ust-Belaya Culture

Three teeth and three bones from two sites in southeastern Chukotka were included in this project as representatives of the Ust-Belaya Culture (aqua pins, 024116-7, 024119, 024248-50, **Figure 2.5**). The Ust-Belaya Culture is an inland group of seasonal hunters of reindeer and avid fishers, evidenced by the stone tools and bone inventory, which includes toggling harpoon points (76). It has been suggested that the Ust-Belaya Culture is ancestral to the present-day Chukchi population (84,85).

### Kanchalan Culture

Three teeth and two bones from the Sed'moi Prichal site in southeastern Chukotka, were contributed as representatives of the Kanchalan Culture (white pins, 024121, 024257-60, **Figure 2.5**). The coastal Kanchalan cultural sites on the Chukchi peninsula date from ~1300–500 BP and have evidence of reindeer and sea animal hunting (76,86). Sites associated with the Kanchalan Culture are characterised by a wealth of split bones of deer and sea mammals (68). Archaeologists consider these sites to have belonged to people of the tundra, who went only sporadically to the sea and began to settle along the coast, adapting to primarily hunting marine mammals (76).

### Lakhtin Culture

One lock of hair from the Gavriil Bay-II site, located on the southwest region of the Chukchi peninsula, was included in this project as a representative of the Lakhtin Culture (bright green pin, 024124, **Figure 2.5**). The Lakhtin Culture spans from ~2500-300 BP and has been suggested that its origins relate to the *Palaeo-Inuit*-Aleutian and Proto-Itelmen population groups (87).

### Old Bering Sea Culture

Four teeth, one bone, and one petrous from five sites in Chukotka were contributed as representatives of the Old Bering Sea Culture (dark blue pins, 024253-56, 024261-62, **Figure 2.6**). The Old Bering Sea Culture, as well as other Neo-Inuit Cultures (*e.g.*, Birnirk and Thule), began developing in the Bering Sea region ~3000 BP. It is thought that the Old Bering Sea Culture, which spanned the northeastern territory of Chukotka, developed through exchanges between Palaeo-Inuit and Neo-Inuit-like groups, as well as the Ust-Belaya Culture (4,76,88,89). The eastern region of the Chukchi Peninsula was a crucial centre for technological developments in marine hunting, in particular, the harpoon complex (68).


**Figure 2.6** - Map of the Chukchi Peninsula, indicating where the ancient tissues from this region originated (67).

### Birnirk Culture

Five locks of hair from the Paipelghak site were included in this project as representatives of the Birnirk Culture (brown pins, 011291-011295, **Figure 2.6**). This site on the shore of the Chukchi Sea has evidence of marine hunting, with an abundance of remains of whales, toggle harpoons, and ulus (an all-purpose knife, primarily used by women, for skinning and cleaning animals). The site dates to ~1000-700 BP, but the hair locks appear to be much older (~1500 BP), so are suspected to be the hair of the ancestors of the residence (90).

## Alaska

The Alaskan hair was contributed to this project by Dr. Kate Britton and Dr. Rick Knecht at the University of Aberdeen, in conjunction with Qanirtuuq Inc., an Indigenous-led group who are the official owners of the Nunalleq archaeological site.

### Yupik Culture

Eighty-four locks of hair from the Nunalleq site, close to Quinhagak, Alaska, were contributed as representatives of the Yupik Culture (purple pin, 011311-12, 011316-17, 011319-25, 011327, 01131-32, 01135-37, 011339, 011341-43, 023501-02, 023504, 024724-64, 024766-84, **Figure 2.7**). This permafrost-preserved site, dating as early as ~650 BP, represents a local phase of the Thule Culture on the south-eastern coast of the Bering Sea (91). Discrete clumps of hair were found buried in the floor of ancient settlements and are suspected to have been placed there in a non-mortuary context (92). The distribution of these buried locks of hair is of particular interest as it has been speculated that sex or social hierarchy may have dictated the placement of burials (92).


**Figure 2.7** - Map of Alaska, indicating where the ancient locks of hair from the Nunalleq site originated (67).

## Canada

The four teeth and five bones that originate from northern Canada were contributed to this project by the Canadian Museum of Natural History (CMH) in conjunction with the Inuit Heritage Trust and the Avataq Cultural Institute. The eight teeth that originate from Newfoundland and Labrador, Canada, were contributed to this project by Dr. Vaughan Grimes at Memorial University Archaeology Department in conjunction with The Rooms Museum in St. John's, Newfoundland.

### Pre-Dorset Culture

One bone from the Rocky Point site, on Devon Island in Nunavut, Canada, was included in this project as a representative of the Pre-Dorset Culture (black pin, 011435, **Figure 2.8**). The bone originated from an infant or foetal skeleton which was found buried in the floor of a tent ring (93,94). The site was anticipated to date between 3000 and 2800 BP, however this child skeleton has been found to be much older, dating to ~4210 BP (2,93,94). This site has evidence of marine mammal hunting technology, including various open-socketed harpoon heads, as well as an abundance of seal teeth and bones (93,94).


**Figure 2.8** - Map of the Canadian High Arctic, indicating where the ancient tissues from this region originated (67).

### Middle-Dorset Culture – Newfoundland


**Figure 2.9** - Map of the Great Northern Peninsula of Newfoundland, indicating where the ancient tissues from this region originated (67).

Seven teeth from two sites on the island of Newfoundland, on the east coast of Canada, were contributed as representatives of the Mid-Dorset Culture (bright blue pins, 011229-31, 011445-48, **Figure 2.9**). Archaeological evidence for a Dorset occupation on the island dates to ~5500 BP. Archaeological evidence for the presence of Dorset Culture on the island of Newfoundland is unique when compared to more northern regions of Canada (95). The Dorset sites of the Great Northern Peninsula, Newfoundland, are the most southerly Dorset sites ever found (96). At each of the three sites, toggling and barbed harpoons made of antler and bone were found alongside the Mid-Dorset skeletal remains, which were often interned in rock shelters and surrounded with ochre (96–98).

### Middle-Dorset Culture – Canadian High Arctic

One tooth and one bone from the Alarnek site, on the Melville peninsula in Nunavut, Canada, were included in this project as representatives of the Middle-Dorset Culture (yellow pins, 011436, 011442, **Figure 2.8**). The site is estimated to be ~1800 BP and has evidence of human occupation spanning centuries (99). This site shows evidence of marine mammal hunting including walrus tusks, seal bones, and various harpoon heads (99). One of the most infamous Dorset effigies

was found at the Alarnek site, a polar bear made of ivory with a small wooden compartment on the underside of the throat, which contained ochre (100). This carving as well as the harpoon heads indicate the high quality of craftsmanship of the Dorset Culture (100).

Two teeth from the Tayara site, in Sugluk, Avataq (otherwise referred to as Quebec), Canada, were contributed as representatives of the Mid-Dorset Culture (yellow pins, 011423, 011433, **Figure 2.8**). The Tayara site has been dated to ~2300 BP, although the stratigraphy indicates that the site may have been active during multiple phases of the Dorset Culture (101). This site was influential in suggesting that there was cultural continuity between Pre-Dorset and Dorset Cultures (102). An abundance of harpoon heads was found at this site demonstrating some regionally specific marine hunting technology (*i.e.*, the Tayara pointed harpoon head) (101).

### Middle-Late Dorset Culture

One bone from the Saatut site, on Baffin Island in Nunavut, Canada, was contributed as a representative of the Middle-Late Dorset Culture (purple pin, 011434, **Figure 2.8**). This site is estimated to be from ~2000-1200 BP and has at least two predominant occupational episodes (103). There have been discussions about whether this site represents Middle or Late Dorset Culture (102). The archaeological findings show evidence of hunting marine mammals, as was apparent by the numerous harpoon heads found, most notably the Nanook wasp waist type harpoon head (103).

### Late Dorset Culture

Two bones from the Angekok site, on Mansel Island in Nunavut, Canada, were included in this project as representatives of the Late Dorset Culture (dark blue pins, 011414-15, **Figure 2.8**). This site contained houses of more than one period of Dorset settlement (104). These skeletal remains were noted to have been found in a midden deposit, and unfortunately there is no description of other archaeological finds from the site (104).

### Huron-Wendat Nation

Three teeth from two sites in southern Ontario, Canada, were contributed as representatives of the Huron-Wendat Nation (orange pins, 022559-61, **Figure 2.10**). These individuals were originally mislabelled as Athabascans and were therefore included in this project. The Huron-Wendat are a non-Arctic First Nations group from the Eastern Woodlands who lived in an agricultural-based society with hunting and freshwater fishing practices (105). The Huron-Wendat Nation lived in villages with numerous long house structures. The sets of remains included in this project were found in middens or ossuaries and were thought to date to ~600 BP (105–107).


**Figure 2.10** - Map of eastern Canada (also known as the Eastern Woodlands), indicating where the ancient tissues from this region originated (67).

### Mi’gmaq Nation

One tooth from a site on Old Mission Point in New Brunswick, Canada, was included in this project as a representative of the Mi'gmaq Culture (red pin, 011444, **Figure 2.10**). This individual was included in the project to examine the relationship between the Dorset and neighbouring sub-Arctic populations. The site where this tooth originated is in the prehistoric village of Tjigog, which was a summer locale for the Mi'gmaq (108). The Mi'gmaq hunted both terrestrial and marine animals evidenced by the toggling harpoon head and beaver pelts found at this burial (109).

## Greenland

The 50 Greenlandic sets of remains included in this project were contributed by Dr. Marie Louise Schjellerup Jørkov and Dr. Niels Lynnerup in conjunction with the Greenland National Museum.

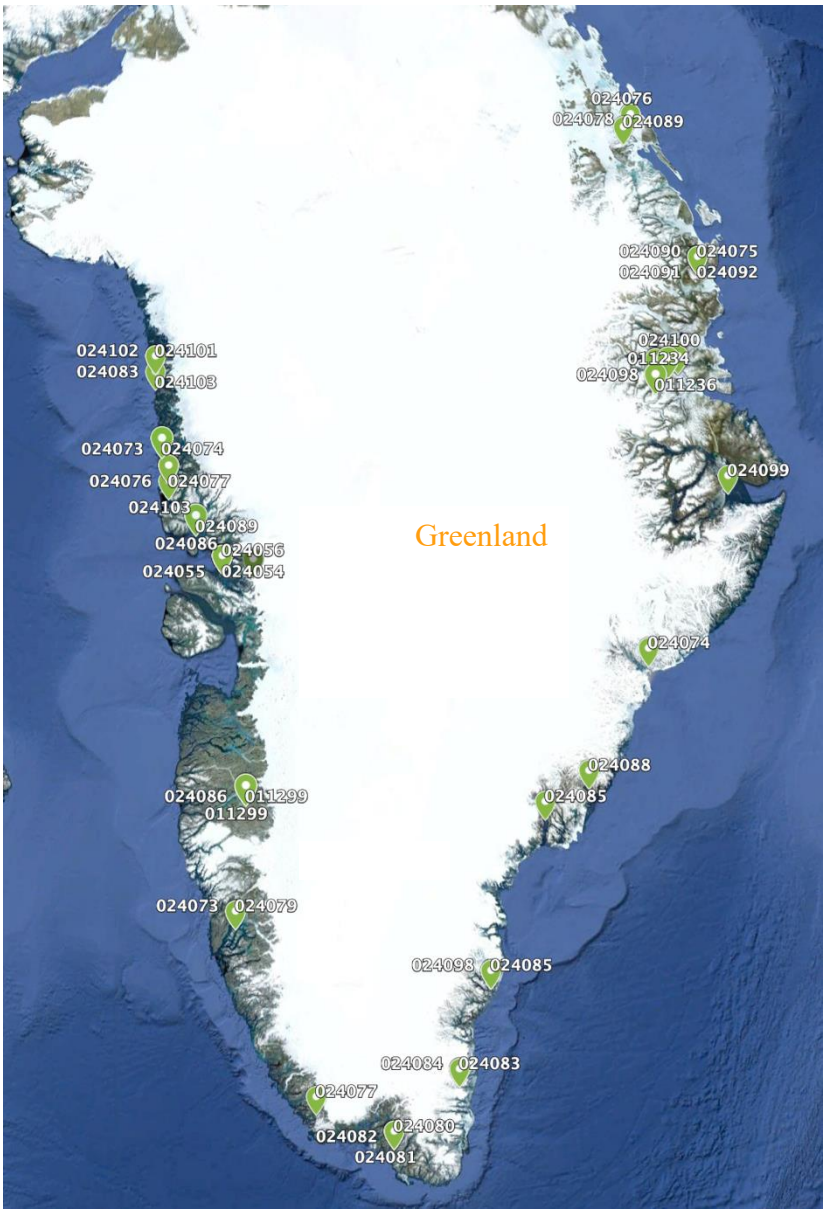



**Figure 2.11** - Map of Greenland, indicating where the ancient tissues from this region originated (67).

### Thule Culture

Two bones and 48 teeth originating from 16 sites along coastal Greenland, were contributed as representatives of the Thule Culture (light green pins, 011234, 011236, 011299, 024054-86, 024088-95, 024098-103, **Figure 2.11** + **Figure 2.12**). The Thule Culture is suggested to have arrived in Greenland ~700 BP, where they may have overlapped with late Palaeo-Inuit peoples in north and northwest Greenland (110,111). This temporal and geographic overlap is shown in the archaeological evidence; however, it is yet to be determined whether there was gene flow between the Palaeo- and Neo-Inuit groups (2,7,110). The minor population stratification observed in the present-day Inuit of Greenland suggests differing levels of gene flow from ancient Neo-Inuit (namely Thule) groups (7).

These tissues were designated as belonging to the Thule Culture based on the architectural remains of winter houses (112). Various ethnographic reports detail the dominance of seal hunting practices at these sites (113). Archaeological evidence suggests these sites were inhabited ~300 BP (112,113).


**Figure 2.12** - Map of western Greenland, indicating where the ancient tissues from the Uummannaq site originated (67).

## 2.1.2 Overview of Ethics in the Field of Palaeogenomics

As Western nations expanded across the world during the Colonial Era, personal and institutional osteological collections expanded (114). Over time, various biomolecular analyses have been developed which led to the destruction of these tissues, leading anthropologists, sociologists, and philosophers of science to describe the newly-emerging phenomenon as bio-colonialism (115). Commencing with simple techniques such as anthropometric measurements, progressing to techniques such as post-mortem tissue histology, and further to more destructive analysis, such as radiocarbon dating and isotopic analysis, the quantity and diversity within ancient remains in osteological collections began to diminish (116).

In the late 1990s and early 2000s, existing cloning and Sanger sequencing methods were applied to ancient mitochondrial DNA sequencing, leading to an increase in the rate of tissue destruction. The publication of the first ancient genome by Rasmussen in 2010 (30) began what

has been referred to as the ‘bone rush’, where scientists, predominantly in the West, accessed the colonially-collected remains and destroyed them for genomic analysis (116). In 2014, it was discovered that the highest endogenous DNA percentage in human skeletons is found in the internal auditory meatus of the petrous bone, leading to more targeted sampling for ancient DNA (117).

Professor of Native Studies, Dr. Laurelyn Whitt, and Professor and Canada Research Chair of Indigenous Peoples, Technoscience, and Society, Prof. Kim Tallbear, stress that archaeologists and geneticists in the field of ancient DNA should consider the bio-colonial aspects of the field, become educated on the cultural practices of the population groups they are working with, and have the influence of that knowledge direct their behaviours towards tissues, relationships with Indigenous groups, as well as project design and execution (115,118).

### 2.1.3 Purpose

The purpose of generating the dataset described in this chapter was to address gaps in the existing palaeogenomics knowledge base. To address these gaps, focus was set on three main goals. First, was to increase the data quality of the previously investigated tissues in Raghavan *et al*., 2014 (2). This was made possible by the immense technological progress in the field of ancient DNA sequencing, as well as the substantial number of resources available for this project (namely immense sequencing capacity). Second, was to increase the number of ancient genomes available from both the North American Arctic and potential source populations from Siberia. With new tissues becoming available through expanded collaborations, the temporal span and geographic scope was significantly broadened. Third, was to strengthen the existing relationships with the Indigenous groups associated with this work. This was addressed through increased communication, community engagement, and knowledge sharing.

### 2.1.4 Outline of Chapter

This chapter outlines the approach used for processing ancient tissues for the highest possible genomic quality. It begins by detailing the ethical considerations for work in the field of palaeogenomics. It goes on to describe the methodology for ancient DNA extraction, genomic library preparation and sequencing, radiocarbon dating, and isotopic analyses, as well as the subsequent bioinformatic pipeline used to authenticate the DNA as ancient, remove contamination, and assess which genomes are appropriate for inclusion in the final dataset. Finally, the novelty, scale, and potential of this dataset are discussed, as well as the proposed results dissemination and knowledge sharing efforts with Indigenous communities.

## *2.2 Methods*

### 2.2.1 Establishing Ethical Guidelines for Working with Ancient Remains

Principles for working with ethically sensitive remains were established at the Centre for GeoGenetics (now known as Lundbeck Foundation GeoGenetics Centre), in conjunction with Dr. Fernando Racimo, Dr. Mikkel Winther Pedersen, and Dr. Hannes Schroeder. This effort was done in order to increase transparency of current methodologies to archaeologists, academic and non-academic institutions, and communities who speak on behalf of ancestral remains. These principles were also established to provide a reference point for scientists working with ancient remains within the Centre. Though the process is extremely complex (as detailed in the flow chart of **Appendix 1**), both within the lab and within the social structures that exist outside the lab, four main objectives were detailed. These objectives are described as they would occur in a palaeogenomics project; however, it is represented in **Figure 2.13** without any numbering because it is cyclical, and no aspect is of greater importance than another.

The first aim is to establish a working relationship based on honesty and respect with those who are culturally or historically linked to the remains. To uphold this standard throughout the years that it often takes for palaeogenomic projects to conclude, a memorandum of understanding

is outlined, agreed upon by both parties (*i.e.*, the community representatives and the Centre), and signed.

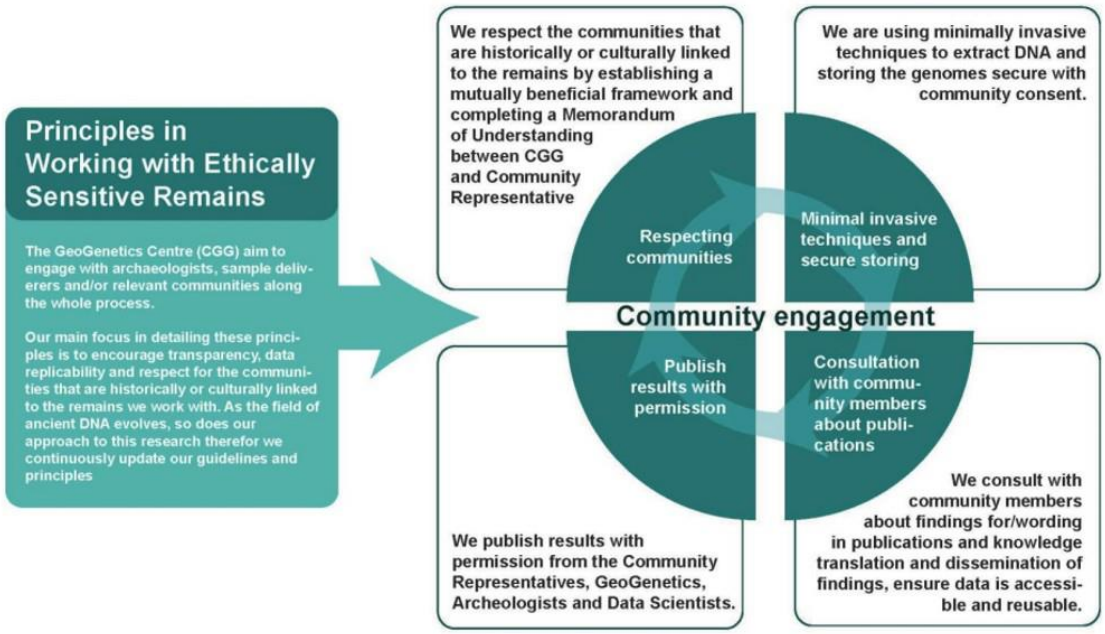



**Figure 2.13** - The CGG's principles for working with ethically sensitive remains. These objectives are described as they would occur in a palaeogenomics project; however, it is represented as cyclical because no aspect is of greater importance than another.

The second aim is for the tissues themselves to be treated with respect. This respect has two main themes, primarily that tissues need to be handled in a way that shows respect for the individual and the cultural group it represents, and secondly that the tissue needs to be processed in a way that is least wasteful and has the highest chance of contributing to science. When respect is shown in both these ways, it is likely that the processing will be done in a way that descendant and Indigenous communities are content with.

The third aim is to work alongside communities and discuss the potential analysis in which the genomic data could be included and ensure that the community agrees with the research questions being addressed. It is expected that scientists do their due diligence with regards to ensuring that the data is only used for analyses which have been approved by the communities. Following that, when the results become available, it is important that results dissemination is done through a means in which the public can appreciate the findings and not be dismissive of any other

lines of evidence that may be relevant to that cultural group, such as oral histories. Scientists should be receptive to new information which will enable knowledge sharing and move away from previous paternalistic approaches of science in the West.

The fourth aim is to consult with the communities regarding the expectations of data accessibility after publishing. Additionally, it is important to consider the potential socioeconomic detriments to Indigenous communities that could arise because of publishing genetic findings or the ways they may be framed (119). Reflecting on potential consequences and presenting findings in a clear and explicit manner that do not lend themselves to misinterpretations may help to decrease harm to descendant and present-day Indigenous communities.

### 2.2.2 Permissions and Community Engagement

Although existing permissions were in place for many tissues in this project both from and since the Raghavan *et al*., 2014 study (2), updated permissions were sought to confirm that the community members who speak on behalf of these ancestral remains continued to be in support of this work.

For the tissues of Siberian origin included in the project, permissions were pursued from RAIPON (Russian Association of Indigenous People Of the North). Permissions through RAIPON were granted at the regional level (**Appendix 2.1**) and are currently being pursued at the national level.

For the Alaskan hair locks, original permissions were granted in 2011 for tissue destruction, namely isotopic and ancient DNA analyses, from the NIMA Corporation (**Appendix 2.2**), and following documentation detailing the request for expanded analysis, updated permissions were granted in 2020 by Qanirtuuq Inc (**Appendix 2.3**). Community outreach is done by Dr. Rick Knecht, who often resides close to, and is the main archaeologist at, the Nunalleq site. A list of potential research questions was provided for the community to indicate what interests them the most, which helped guide the processing of the hair and the number of genomic libraries produced.

Many other community-based efforts surrounding this site take place and are detailed at nunalleq.wordpress.com/.

For the tissues from the High Arctic in Canada, permissions for continued analysis (that is, following the publication of Raghavan *et al*., 2014 (2)) were granted in 2018 and 2019 by the Inuit Heritage Trust (IHT) and the Avataq Cultural Institute (ACI), respectively (**Appendix 2.4**). A request for expanded analyses was distributed to the IHT and ACI through the Canadian Museum of History for which approvals were granted in 2022, after which direct contact between the geneticists and community-led organisations was permitted.

For the tissues from Southern Ontario, original permissions were granted for the work of Scheib *et al*., 2016 (107). A request for expanded analysis, sent via Dr. Ripan Malhi, was approved by the Council of the Huron-Wendat Nation (*i.e.*, Louis Lesage and Ron Williamson) in 2021 (**Appendix 2.5**). A public lecture-based research visit to descendant communities will take place both prior to publication and after conclusion of this project.

For the tooth from New Brunswick, original permissions for bioarchaeological assessment, sent via Memorial University of Newfoundland, were approved by the Listuguj Mi'gmaq community in 2011 (2). A request for expanded analysis was sent to the current Mi'gmaq Council in 2022, and their response is pending (**Appendix 2.6**).

For the ancient remains from the island of Newfoundland, permissions were sought from the Indigenous groups of Newfoundland and Labrador (Innu Nation, Miawpukek First Nation, Nunatsiavut Government Research Advisory Committee, Qalipu First Nation), through The Rooms Museum, and were granted in the summer of 2022 (**Appendix 2.7**). A public lecture was given at the National Indigenous Peoples Day at Memorial University in St. John's, Newfoundland, in 2018. Following that, community outreach was done via *Let's Talk Science Canada,* throughout Inuit communities along the coast of Labrador.

For the ancient remains from Greenland, permissions were granted through Dr. Marie Louise Schjellerup Jørkov at the Department of Forensic Medicine at the University of

Copenhagen in conjunction with the Greenland National Museum (**Appendix 2.8**). Results dissemination will take place throughout various communities in Greenland prior to publications.

### 2.2.3 Tissue Handling

All tissues contributed to this project were assigned CGG numbers when they were registered in the Centre's database. Once registered, tissues were processed for DNA extraction, with a sub quantity of the tissues sent for radiocarbon dating and isotopic analyses, where possible.

### 2.2.4 Radiocarbon Dating

Radiocarbon dating (C14 dating) (120) was performed on tissues with no previously published date, at the Chrono Centre at Queen's University Belfast and the W. M. Keck Carbon Cycle Accelerator Mass Spectrometer Facility. Successful radiocarbon dates were calibrated using IntCal20 on May 6$^{th}$, 2021 (121).

Various tissues did not undergo radiocarbon analyses due to factors such as; low preservation quality (minimal or no organic fraction remaining) resulting in radiocarbon dating techniques failing, the radiocarbon date being highly influenced by the marine reservoir effect (MRE, where the consumption of older, marine carbon impacts the C14 measurements, see Future Directions, Marine Reservoir Correction), or the tissues being too recent and therefore would result in large error due to wiggles in the C14 calibration curve (122,123). Undated tissues include the bone fragments from the Ushki site (low preservation), the Greenlandic Thule teeth and bones (suspected high MRE and wiggles), and the Alaskan Yupik hair (suspected high MRE and wiggles). In cases where no radiocarbon date was available, the archaeological context or indirect C14 dates are provided.

There are challenges presented when no C14 date is available for a set of ancient human remains. These challenges pertain to the use of relative dating methods, such as contextual seriation (using archeological context) or stratigraphy (using indirect C14 dates from proximal specimens) (124). In the case of contextual seriation, where changes in typology in the studied

locale are assumed to mimic chronological transitions, it is important to consider that the contextual information that accompanies ancient human remains from outside of Europe is provided through a 'colonial lens' (by a non-Indigenous person) and therefore can lead to misinterpretation and mis-categorisation. Seriation in Arctic cultures can be particularly difficult as there is not a consensus on cultural continuity between groups as well as affiliations of certain artifacts (2,102). Similar limitations are seen in the case of stratigraphy, where soil layers are used to construct a sequential categorisation of events at a site, with lower levels interpreted as older and top layers interpreted as younger (124). However, using the C14 date of a specimen from a similar soil layer may be not reliable, particularly in the context of the Arctic. Regions of the High Arctic have differential geological phenomena due to the intense cold, such as glaciers expanding and receding during period of cooling and warming or permafrost reducing erosion and deposition, leading to less informative stratigraphy (125). It is therefore important to consider limitations to relative dating methods and the consequences when interpreting estimated ages of ancient humans remains, especially those from the Arctic.

### 2.2.5 Bulk Isotope Analysis and Dietary Reconstructions

Bulk isotopic analyses can provide great insight into the diet of ancient populations. The carbon isotope ratio (the proportion of $C^{13}$ and $C^{12}$ isotopes in an organism [$\delta^{13}C$]) is indicative of the proportion of $C_4$ to $C_3$ plants consumed by that individual. In the majority of Arctic terrestrial ecosystems most plants use the $C_3$ photosynthetic pathway, where the average $\delta^{13}C$ is -27‰, with plants from more moist environments having a $\delta^{13}C$ of and as low as -35‰ and plants from drier environments having a $\delta^{13}C$ as high as -21‰ (126). In contrast, the $\delta^{13}C$ of marine phytoplankton range from -22‰ to -18‰ (127). Another indicator of dietary composition is stable nitrogen isotopes (the proportion of $N^{15}$ and $N^{14}$ isotopes in an organism [$\delta^{15}N$]), which can infer the plant-to-meat ratio in the diet. The $\delta^{15}N$ ratio increases based on trophic level being consumed, with +3‰ $\delta^{15}N$ increase per trophic level. Terrestrial ecosystems have few trophic levels, whereas marine ecosystems have many more trophic levels resulting in increased $\delta^{15}N$ from marine based diets (128).

Bulk carbon and nitrogen isotopic analyses were performed concurrently at the radiocarbon dating labs, and some undated tissues have previously published bulk isotope values (92,129); however, for consistency purposes, the Alaskan Stable Isotope Facility bulk dataset was used in this thesis (which was run as an internal calibration metric for the Compound Specific Isotopic Analyses – see Future Directions, Compound Specific Isotope Analysis).

For reconstructing the proportional contribution of dietary components, a three-end member (terrestrial, freshwater, and marine diets), two-isotope (carbon and nitrogen) mixing model was created as outlined in Appendix 3 (130–132). In some cases, the tissue-to-diet fractionation factor required adjustments following the methodology of Halffman *et al*., 2020 (130). This metric will inform the need for, and extent of, correction for the marine reservoir effect, see Future Directions, Marine Reservoir Correction.

### 2.2.6 DNA Extraction and Double-Stranded Library Preparation

The preliminary double-stranded libraries built from the extracted DNA were first sequenced to low coverage as part of a screening process. This was done in order to determine which tissues had high endogenous DNA and could therefore be sequenced to high coverage (*i.e.*, greater than 16x coverage), such that they would act as representatives for a cultural group or time period. All libraries were assessed for sequencing capacity, and each was sequenced to saturation to obtain the maximum amount of genomic information for each ancient individual.

Tissues with low endogenous DNA were sequenced to low-depth coverage (~0.1x) to generate a more robust panel for population genetic analyses (outlined in Chapter 3), as well as to generate more pathogen sequences (outlined in Chapter 4). Tissues with high endogenous DNA content were selected and underwent additional DNA extraction, library preparation, and USER treatment to increase molecular diversity, reduce terminal damage, and allow for deep sequencing of the genome to reach high depth.

To monitor contamination from the reagents and cross-contamination between ancient genomes, blanks were processed with each batch of pre-treatment, DNA extracts, genomic libraries, and library quantification.

## Pre-treatment

### Hair

Hair locks were weighed out in ~100 mg subsamples, where possible, and rinsed with water to remove large debris. The hair was then incubated with either 5vol% bleach or 5vol% sodium hypochlorite solution for 5 minutes, followed by another rinse with water to remove the bleach or sodium hypochlorite (133).

### Tooth and Bone

A circular saw blade was used to remove the outer surface of the bone or tooth in order to reduce contamination from environmental and modern human sources. The bone or tooth was then divided into <100 mg quantities, where possible. For teeth, the tooth root was removed and split in half to clean the inside and the cementum was preferentially selected when available (41). For the petrous bone the inner auditory meatus was preferentially selected (117).

## DNA Extraction

Prior to DNA extraction, a predigest incubation of 30 minutes at 37ºC was performed to remove exterior contamination (41).

### Hair

The pre-digestion and digestion buffers used for hair were composed of 10 mM NaCl, 2.5 mM EDTA (pH 8), 10 mM Tris-HCl (pH 8), 2% SDS, 5 mM CaCl2, 10% proteinase K, and 40

mM DTT (30). The pre-digestion buffer was removed, and fresh digestion buffer was used for 24-96 hours of incubation on a rotator at 55ºC. Additional proteinase K and DTT (1 M) was added for increased digestion when required (30).

**Tooth and bone**

The pre-digestion and digestion buffers used for tooth and bone were composed of EDTA 0.5 M (pH 8), 10 mM TE buffer, 0.5% N-lauryl sarcosine, and 10% Proteinase K (41). The pre-digestion buffer was removed, and fresh digestion buffer was used for 24-96 hours of incubation on a rotator at 37ºC - 55ºC, depending on the preservation level and tissue. Additional proteinase K was added for increased digestion when required (41).

Once tissue digestion was complete for hair, teeth, and bones, DNA was extracted from the digestion buffer using a binding buffer composed of 2 M GuHCl, 70% 2-propanol, and 0.05% Tween 20 (134). The binding buffer was added in a 20x ratio to the digestion buffer and the combined solution was spun through a Roche large volume silica column (Roche, Germany). The filter of the spin column was then washed twice with PE Buffer (Qiagen, Germany) and spun dry. Then, a 65 ul EBT buffer (10 mM Tris-HCl, 0.05% Tween 20) was added to the column for a 5-minute incubation at room temperature (Qiagen, Germany). The filter was then spun, and the eluate was reloaded to the filter and re-incubated for 5 minutes at room temperature before being spun down and transferred to a 1.5 ml Eppendorf LoBind tube for storage at -20ºC.

The ~65 ul eluate was then subdivided into three aliquots to build one non-USER-treated library for screening, and two USER-treated libraries for high-depth sequencing. The extracts from libraries built for high-depth sequencing underwent USER enzyme treatment for 3 hours at 37ºC to decrease the post-mortem ancient DNA deamination pattern and increase mappability of ancient reads to the reference genome (135).

## Double-Stranded Library Preparation

A NEB End-repair module E6050 master mix (NEBNext end repair buffer E6052, NEBNext end repair enzyme mix E6051) was combined with one-third of the USER or non-USER treated extract and incubated for 20 minutes at 12°C and then 15 minutes at 37°C. The end-repaired extract underwent a Qiagen MiniElute cleanup (binding buffer 10×vol PB, 1/25 NaOAc 3M pH=5, wash buffer PE, elution buffer EBT) and the EBT was left to incubate for 5-15 minutes at room temperature before being spun through the column.

Adapters were then ligated to the end-repaired strands using NEB quick ligation module E6056 (Quick ligation 5x buffer E6058, DNA adapters (25 uM), Quick T4 Ligase E6057) through a ~12-hour incubation at 20ºC. The adapter-ligated fragments underwent a Qiagen MiniElute cleanup. The EBT was left to incubate for 5-15 minutes at room temperature, before being spun through the column.

A fill-in reaction (10x ThermoPol Reaction Buffer, Bst DNA polymerase large fragment M0275, dNTP (stock=2.5 mM)) was then performed, incubating for 20 minutes at 65°C, then 20 minutes at 80°C for enzyme inactivation. In some cases, the Biomek i5 Automated Workstation was used for library preparation.

## Library Quantification, Amplification, Purification, and Re-quantification

For screening libraries, quantitative real time PCR was performed on the LightCycler 480 Instrument (Roche) to estimate the number of index PCR (iPCR) cycles for library amplification. For high-depth libraries, no quantitative PCR was performed and instead libraries underwent a set number of in-house pre-calibrated iPCR cycles for amplification (16 cycles).

The qPCR master mix (Roche LC480 Master Mix, P7 and P5 primers) was added to 1 ul of each library and ran for 10 minutes at 95º followed by quantitative detection for 30 cycles of 30 seconds at 95ºC, 30 seconds at 55ºC, 30 seconds at 72ºC, and was finished off with a melting curve detection by a single cycle of 30 seconds at 95ºC, 30 seconds at 55ºC, and 30 seconds at 95ºC.

Indexes provided by Illumina Innovative Technologies™ were then tagged to DNA fragments, which enabled multiplexing during the sequencing runs. The indexed libraries then underwent PCR amplification (iPCR) using the Kapa Hifi Uracil + ReadyMix Kit (Roche), where the number of iPCR cycles required was estimated from the qPCR results. For USER-treated libraries, a set number of 16 iPCR cycles was performed. Amplified libraries then underwent purification with SPRI magnetic beads, were washed twice with 80% ethanol, and eluted in EBT. For high-depth libraries, the CyBio FeliX Extraction Set was used to purify libraries.

In order to pool the purified libraries proportionally for sequencing, the libraries were quantified using a DNA High Sensitivity chip following the manufacturer's protocol and ran on the Agilent BioAnalyzer. USER-treated libraries were quantified using the 5300 Fragment Analyzer System (Agilent) with the HS NGS Fragment Kit 500 (Agilent), in order to pool them proportionally for sequencing.

### 2.2.7 Shotgun Sequencing

For screening libraries, pools of 12-60 libraries were prepared for sequencing. The screening libraries were sequenced on the Illumina HiSeq 4000 on one lane with single end reads at a read length of 80 base pairs (bp). For USER-treated libraries, pools of 6-90 libraries were prepared for sequencing. Those libraries were sequenced on the Illumina NovaSeq on either a single XP lane (2.5 billion reads) or over four lanes of a S4 full flow cell (10 billion reads) with paired end reads at a read length of 150 bp.

The base-calling of the sequencing data was done using the Illumina software CASAVA (v.1.8.2) and the libraries were de-multiplexed, while being constrained to an exact match of each of the two eight-nucleotide indexes that were ligated during library preparation.

For the most part, libraries only underwent one round of sequencing; however, based on the sequencing saturation metric, some libraries underwent deep sequencing to maximise the amount of genomic information gained per library.

### 2.2.8 Preliminary Mapping

Sequencing runs were demultiplexed and underwent quality control by the GeoGenetics Sequencing Core in Copenhagen, Denmark. The sequencing data was then uploaded to the Illumina BaseSpace Sequencing Hub application and preliminary sequencing statistics were generated (*e.g.*, the number of reads from each library, the proportion of the XP lane or S4 flow cell that was taken by that library, *etc.*). The preliminary mapping is composed of five main steps; the first two steps are done in parallel on the library level and then merged in step three.

**Adapter Removal**

Adapters were trimmed from the reads of double-stranded libraries, overlapping reads were collapsed, and reads below 30 base pairs were discarded. This was done using AdapterRemoval v2.3.2 (136), with the following parameters: --minlength 30, --collapse-conservatively, –adapter1 = 5'-AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC-3' and --adapter2 = 5'-AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3'.

**Mapping**

Paired-end and single-end fastq files were aligned separately to the hg38 human reference genome. Burrows-Wheeler Aligner (BWA aln v0.7.17, with parameter -l 32) was used with seeding enabled (137,138). Though seeding is not the standard in ancient DNA, because it results in 1% fewer reads mapped (due to the presence of ancient DNA damage at read ends), it was chosen for its increased processing speed in this preliminary mapping round. Following alignment, sam files were generated with samse for single-end and sampe for paired-end data, and samtools sort (v1.11.0) was used to sort the reads in the bam files (138,139).

#### Merging

Paired and single-end data for each library and each lane were then merged into a single bam file using samtools merge (v1.11.0) (139).

#### Mark Duplicates

Reads that are optical duplicates (an artefact of Illumina sequencing) of each other are flagged or 'marked' using Picard MarkDuplicates (v2.25.0) with an optical duplicate pixel distance of 12000 (140).

#### Cram Conversion

All bams were then converted to cram format, using samtools view (v1.11.0), to reduce the storage footprint (139).

### 2.2.9 Bioinformatics Pipeline

#### Data transfer to CGG servers

The ancient genomes were then transferred to the CGG servers and were processed through an in-house bioinformatics pipeline. Quality control was run to ensure complete transfer, namely file counts and samtools quickcheck (BaseSpaceCLI 1.5.1 -- built on 2021-12-16 at 14:53).

#### Filtering

Unmapped reads were removed (samtools view 1.10, parameter -F 4) and the statistics from the adapter removal '.settings' file were generated using Picard AddCommentsToBam 2.18.26-SNAPSHOT (140).

#### Removing PCR duplicates

Decluster, a tool in the ANGSD repository (parameters= -p 12000, -w, -T {reference genome}, -q 30), was used to determine the amount of PCR and cluster duplicates in each library, using single-end reads with a mapping quality of 30 and above (141).

## Complexity Projections

An in-house R script was used to estimate the saturation level of the library to measure the potential for further sequencing. This script uses the total number of reads sequenced and the superduper output to estimate the clonality at different projected depths (0.1x, 0.5x, 0.7x, 1x, 2x, 4x), as well as the number of reads required to achieve that depth. No output was given when the projected clonality was above 75%.

## Damage

To authenticate reads as ancient, the post-mortem deamination DNA degradation pattern was quantified using mapDamage2.0 (142).

## Contamination

Three different methods were used to estimate contamination levels in each library. ContaMix was run to calculate the proportion of non-endogenous mitochondrial reads to the mitochondrial consensus genome (143). For this, an in-house perl script was run that constructs two alternate versions of the endogenous mitochondrial genome. The first approach (Contamix Approximate MAP [Mapping Authenticity Parameter]) used sites with ≥1x coverage, where each basecall was assigned based on a 50% consensus of reads covering that site. The second approach (Contamix Precise MAP) used sites with ≥5x coverage, where each basecall was assigned based on a 70% consensus of reads covering that site. Both the approximate and precise methods of Contamix are reported with the lower and upper confidence intervals. Secondly, ANGSD was run to approximate nuclear contamination by calculating the percentage of X-chromosome homozygosity in males (141,144). Finally, CrosscheckFingerprints (CCF) was run to verify if genomic data originates from the same individual (140). This tool identifies variable regions of the genome to establish a fingerprint for the data and compares that fingerprint to other data to discern if it originated from the same individual or another individual in the dataset (140). All contamination estimates used reads filtered by a base quality of ≥20 and mapping quality of ≥30.

### Endogenous Percentage

The endogenous DNA percentage was calculated by the number of reads after cluster duplicates were removed over the total number of reads after trimming (*i.e.*, the reads after the adapter sequence were removed, overlapping reads were collapsed, and reads below 30 base pairs were discarded).

### Depth of Coverage

The depth of coverage for the mitochondrial and nuclear genome of each library was determined using samtools depth (139). The mean depth of coverage for each library was used to calculate the overall depth of coverage for each genome.

### Genetic Sex

The Skoglund method was used to assign genetic sex to the ancient genomes by computing the number of reads mapped to the Y-chromosome divided by the total number of reads mapped to the sex chromosomes, in a ratio termed R_y (145). Conservatively, the assignment cut-offs are based on the extreme values, including the confidence interval, in ancient XY (R_y = 0.077) and XX (R_y = 0.022) tissues observed by Skoglund *et al*., 2013 (145).

## 2.2.10 Species Identification

Some tissues were suspected to have originated from non-human animals. This was based on either the tissues appearing to be non-human (*i.e.*, extremely light weight bones or exceptionally coarse strands of hair), bulk isotope results falling outside the range of a human diet, or various bioinformatics tools giving suspicious results (*e.g.*, no mitochondrial haplogroup assignment despite adequate coverage).

Preliminarily, a subset of reads from genomes suspected to be of non-human origin were run through the computer program called MEGAN (Metagenome Analyzer) (146), where a species assignment was suggested. When additional sequencing data became available, an in-house taxa classification pipeline was used for more precise species identification. After quality control filtering, the initial mapping in this pipeline was performed against the mitochondrial reference

panel from the NCBI RefSeq mitochondrial database (147), which contains 13705 different taxa. Those reads were then consolidated and mapped in multiple alignments to the last common ancestor of all the hit taxa. These mapped reads were counted and the taxa with the most hits was assigned as the species.

### 2.2.11 Kinship

Kinship between individuals from the same geographic region (n=11, Central Siberia, Zhokhov Island, Magadan Kamchatka, Chukotka, Alaska, Nunangat, Eastern Woodlands, Newfoundland, Greenland Northwest, Greenland South, Greenland Northeast) was estimated using READ (Relationship Estimation from Ancient DNA) (148). This software was designed for pseudo-haploid data. The input genomes were divided into non-overlapping windows, and for each pair of genomes the non-matching alleles were calculated (148). These variable sites were then normalised for intra-population diversity, and relatedness was estimated. In some instances where effective population size is low, the estimated consanguinity coefficient is high, or there is a suspected founder effect, the related proportion between two individuals can be overestimated (148). Results are normalised based on the input data and therefore, when a larger number of genomes was input, the pooled results may differ and become more precise (148).

## *2.3 Results*

### 2.3.1 Tissues to Genomes

In total, 217 ancient tissues were processed for DNA extraction and genomic sequencing: 64 from Siberia, 84 from Alaska, 19 from Canada, and 50 from Greenland (detailed in **Table S1**). Of these tissues, 39 were bones, 93 were hair, one was petrous, and 84 were teeth – all of which varied in preservation quality.

The number of libraries for each ancient individual depended on the preservation of the tissue, the amount of resulting DNA extract, the uniqueness (*e.g.*, geographically or temporally), and the time and resources available. This resulted in a range of 1 to 44 libraries made per ancient individual. Including extraction and library blanks, a total of 1,696 genomic libraries were built in this thesis.

### 2.3.2 Radiocarbon Dating and Bulk Isotopes

Eighty-four tissues were sent for C14 dating, 70 of which resulted in radiocarbon dates ranging from 185 to 10550 BP. After calibration, these dates then range from 185 to 12552 Cal BP (calibrated years before present) (**Table S2**).

For tissues with adequate collagen, the $\delta^{13}C$ values ranged from -10.8‰ to -21.6‰ and the $\delta^{15}N$ values ranged from 10‰ to 23.4‰, excluding non-human tissues (detailed in **Table S3**). The results of the bulk isotope analyses are plotted in **Figure 2.14**, where the $\delta^{13}C$ levels are plotted on the x-axis and the $\delta^{15}N$ levels are plotted on the y-axis, and non-human tissues are crossed with an 'x'.

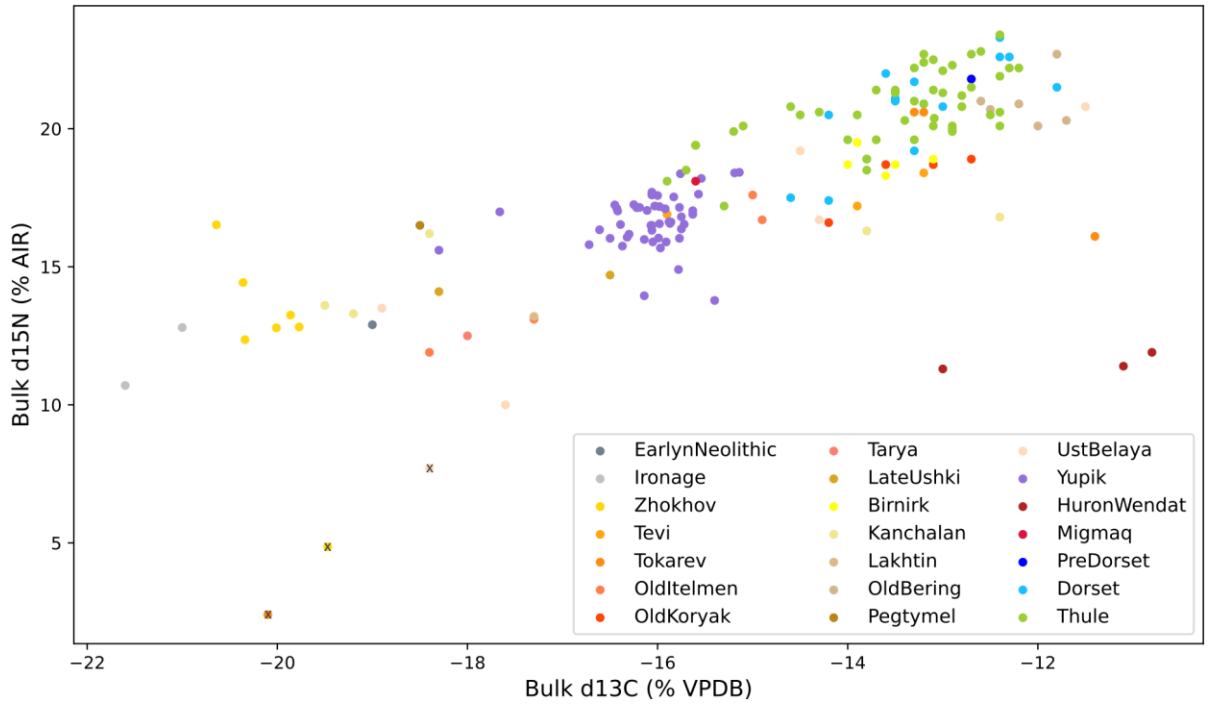



**Figure 2.14** - Scatter plot of the bulk carbon and nitrogen isotope results for the ancient individuals included in this project, where the colour is representative of the culture, and non-human tissues are crossed with an 'x'.

**Table 1** summarises the average proportion of dietary contributions of the ancient individuals grouped by their cultural affinities. The dietary contributions were established using the three-end member, two-isotope mixing model as outlined in **Appendix 3**. The dietary profile is designated as marine where the marine contribution exceeds 80% (highlighted in blue **Table 1**). Palaeodiets established for non-human tissues plot far below the model, as would be expected from a non-human diet, where $\delta^{13}C$ is lower than -18‰ and the $\delta^{15}N$ is lower than 10‰.

**Table 1** - Summary of the average proportion of dietary contributions of the ancient individuals, grouped by cultural affiliations. Dietary profiles were established based on the average proportional contributions, designated as marine where the marine contribution exceeds 80%.

| Cultural Grouping | Number of Individuals | Average Marine Proportion | Average Terrestrial Proportion | Average Freshwater Proportion | Dietary Profile |
|---|---|---|---|---|---|
| Birnirk | 5 | 100 | 0 | 0 | Marine |
| Dorset (all) | 14 | 100 | 0 | 0 | Marine |
| Early Neolithic | 1 | 42 | 38 | 20 | Mixed |
| Huron-Wendat | 3 | $C_4$ | $C_4$ | $C_4$ | Agricultural |
| Kanchalan | 5 | 68.6 | 14.4 | 17 | Mixed |
| Lakhtin | 1 | 59 | 38 | 3 | Mixed Marine Terrestrial |
| Late Ushki | 1 | 54 | 27 | 19 | Mixed |
| Mi'gmaq | 1 | 90 | 0 | 10 | Marine |
| Old Bering Sea | 6 | 100 | 0 | 0 | Marine |
| Old Itelman | 4 | 75.25 | 22.25 | 2.5 | Mixed Marine/ Terrestrial |
| Old Koryak | 4 | 100 | 0 | 0 | Marine |
| Pegtymel Complex | 1 | 55 | 4 | 41 | Mixed Marine/ Freshwater |
| Pre-Dorset | 1 | 100 | 0 | 0 | Marine |
| Tarya | 1 | 50 | 44 | 6 | Mixed Marine/ Terrestrial |
| Tevi | 3 | 96 | 1 | 3 | Marine |
| Thule | 48 | 99.23 | 0.21 | 0.56 | Marine |
| Tokarev | 3 | 100 | 0 | 0 | Marine |
| Ust-Belaya | 5 | 76.2 | 19.4 | 4.4 | Mixed Marine Terrestrial |
| Yakutia Surface Findings | 2 | 16 | 46 | 38 | Mixed |
| Yupik | 50 | 84.22 | 6.44 | 9.34 | Marine |
| Zhokhov | 6 | 34.83 | 28.33 | 36.84 | Mixed |

### 2.3.3 Contamination

The resulting palaeogenomics dataset contained 1,608 genomic libraries, excluding extraction and library build blanks. A total of 148 libraries were removed due to contamination, 111 of which were flagged by either the Contamix, ANGSD, or CCF tools and 37 of which were too low coverage for the tools to run and were therefore labelled as inconclusive (140,141,143). An additional 305 libraries were labelled as contaminated, whether from flags from contamination tools or the analysis was inconclusive; however, those libraries remained in the project with the individual genome being tagged as contaminated for reasons explained in the Discussion section, Removing Contamination. After contamination removal there were 1,459 libraries, resulting in 173 clean genomes, 42 genomes flagged as contaminated, but were brought forward in some analyses, and two genomes which were removed from future analyses due to having a known source of contamination (detailed in **Table S4** + **Table S5**).

### 2.3.4 Damage

Ancient DNA damage, which was calculated on merged libraries without contamination (n=173), ranged from 0.007 to 0.332 (detailed in **Table S6**). It was observed that libraries with low amounts of DNA damage generally fell into two broad categories; low tissue preservation and DNA extracted from ancient hair (expanded upon in Discussion section, Low Tissue Preservation, Low Tissue Input).

### 2.3.5 Endogenous Percentage and Genomic Coverage

After contaminated libraries were removed, the endogenous DNA percentage was calculated as an average over all libraries from the same individual. The endogenous DNA percentages in clean genomes ranged from 0.007% to 63.488% (detailed in **Table S6**). After contaminated libraries were removed and the libraries of identical individuals were merged (see Results section, Kinship), the nuclear depth of coverage ranged from $5.4\times10^{-5}$x to 56.7x, and the mitochondrial depth of coverage ranged from 0.02x to 6469.1x (detailed in **Table S6**).

### 2.3.6 Genetic Sex Assignments

Genetic sex was assigned to 146 of the ancient individuals in this project, 87 of whom are assigned female (XX), and 59 of whom were assigned male (XY), shown in **Table S7**. A total of 57 genomes were not confidently assigned a genetic sex; 18 genomes were consistent with XY but not with XX, and 39 genomes were not assigned a genetic sex. Conservatively, the assignment cut-offs are based on the extreme values, including the confidence interval, observed in ancient XY (R_y = 0.077) and XX (R_y = 0.022) genomes (145).

### 2.3.7 Species Identification

Based on tissue morphology and the bulk carbon and nitrogen results, four tissues were investigated for species identification using the MEGAN tool (demarcated by * in **Table 2**) (146). Two additional genomes were investigated for species identification using the in-house pipeline. One was due to extremely high mitochondrial coverage with a very low probability in the haplogroup assignment (022903), and the other was due to having low endogenous percentage despite similar tissue preservation quality to other remains from the same site with high endogenous percentage (011442).

**Table 2** - Bulk carbon and nitrogen results for all tissues suspected to be of non-human origin.

| ID | Tissue | Bulk $\delta^{13}C$ (‰ VPDB) | Bulk $\delta^{15}N$ (‰ AIR) |
|---|---|---|---|
| 024263* | Bone | -19.4 | 2.5 |
| 011875* | Bone | -19.3 | 5 |
| 024119* | Tooth | -18.4 | 7.7 |
| 024124* | Hair | -17.3 | 13.2 |
| 022903 | Bone | -20.9 | 12.8 |
| 011442 | Bone | -14 | 20.6 |

Results from MEGAN are presented in **Appendix 4**, where the top hit for 024263 was *Ovis canadensis canadensis*, for 011875 was *Bos mutus*, and for 024119 was suggested to be *Ursus arctos horribilis* (with low level of certainty). Genome 024124, generated from a very coarse lock of hair, had inadequate reads for the tool to run.

**Table 3** - Results from the in-house species identification pipeline, showing the reads from six ancient genomes mapping to different taxa.

| Genome | Reads after QC | *Leontopithecus* | *Mazama* | *Neotragus* | *Dipus* | *Antechinus* | *Gorilla* | *Lutra* | *Rangifer* | *Muntiacus* | *Homo* | *Ursus* | *Ovis* | *Mustelidae* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **011875** | 20891454 | 0 | 30 | 5 | 9 | 9 | 0 | 33 | 2428 | 7 | 0 | 0 | 0 | 0 |
| **011442** | 45392014 | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 0 | 0 | 10 | 0 | 0 | 0 |
| **022903** | 92636204 | 125 | 0 | 0 | 20 | 75 | 15 | 32 | 0 | 0 | 3924 | 0 | 0 | 0 |
| **024119** | 84983542 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 34 | 0 | 0 |
| **024124** | 8702042 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 |
| **024263** | 5784378 | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 924 | 0 |

For the six genomes that were suspected to be of non-human origin, one library from each was sent through the in-house species identification pipeline, and the results are outlined in **Table**

**3**. The pipeline identified 011875 to be in the *Rangifer* genus, 024263 to be in the *Ovis* genus, and 022903 to be in the *Homo* genus. Genomes 011442, 024119, and 024124 did not have enough reads mapped to a reference genome to suggest a species assignment.

### 2.3.8 Kinship

The results of the kinship analyses are shown in **Table 4**. Within the 11 geographic areas, identical individuals (which could be identical twins, but due to there being no archaeological evidence to suggest as much, they will be considered identical individuals for the purpose of this thesis) were identified in five of the geographic areas. First-degree relatives were only identified at two geographic regions; that is, 13 pairs of primary relatives were found in the Alaskan subgrouping and two were found in the Northwest Greenland subgrouping. Similarly, second-degree relatives were only found in two geographic regions, 33 pairs were identified in the Alaskan subgrouping and one pair was identified in the Northeast Greenland subgrouping. As anticipated, given the nature of the READ tool, discrepancies in the output of the geographically divided and merged dataset runs were observed (148). As shown in **Table 4**, the number of first- and second-degree relatives differs in the geographically divided and merged dataset analyses, which is described further in Discussion section, Kinship).

**Table 4** - Results of the kinship analyses, run on the geographically divided and merged datasets.

| Geographic Area | Identical Individuals | Pairs of first-degree relatives | Pairs of second-degree relatives |
|---|---|---|---|
| Alaska | 7 | 13 | 33 |
| Central Siberia | 0 | 0 | 0 |
| Chukotka | 1 | 0 | 0 |
| Eastern Woodlands | 0 | 0 | 0 |
| Greenland Northeast | 0 | 0 | 1 |
| Greenland Northwest | 0 | 2 | 0 |
| Greenland South | 0 | 0 | 0 |
| Magadan Kamchatka | 1 | 0 | 0 |
| Newfoundland | 1 | 0 | 0 |
| Nunangat | 1 | 0 | 0 |
| Zhokhov | 0 | 0 | 0 |
| **Merged – All** | **11** | **48** | **139** |

## *2.4 Discussion*

### 2.4.1 Ethical Considerations for Working with Ancient Remains and Descendant Communities

The principles for working with ethically sensitive remains and engagement with Indigenous communities were observed through minimal tissue destruction, providing feedback and context to relevant communities, and through strengthening working relationships and knowledge sharing.

Minimal destruction of remains for a maximum return of information was pursued by initially processing a small quantity of each tissue and bringing it through the entire laboratory workflow before returning to the original tissue for further processing. This was done to ensure that only tissues with high endogenous percentage underwent additional destruction. Maximum amount of information from each library was obtained using shotgun sequencing (versus capture) and sequencing each library to saturation.

When requested, communities were provided with information on the current standard in the field pertaining to data accessibility following publication. This document detailed the current status of ancient genomes from similar cultural groups, the current status of present-day genomes from the same cultural group, and what options are now becoming available as the field evolves.

In efforts to strengthen the working relationship with the descendent and Indigenous Arctic communities, the results from this work will be brought to the Arctic, through community engagement and knowledge dissemination efforts. This will allow the voices and oral histories of Indigenous peoples to contribute to the publications, telling a more complete history of the peopling of the North.

### 2.4.2 Tissue to Genomic Analyses

#### Palaeodiets

The majority of individuals in the bulk carbon and nitrogen results scatter plot (**Figure 2.14**) plot outside (higher than) the mean marine diet end members, indicating that specific species consumed by these ancient individuals had a relatively higher $\delta^{15}N$ value than those used in the model (*e.g.*, marine fish including cod or wolf fish) or were consuming higher trophic level animals (*e.g.*, polar bears). Further information is available in the Alaska Stable Isotope Facility (ASIF) report, **Appendix 3**. These findings are consistent with the mean and ranges of the previous end members used by Raghavan *et al*., 2014 (2). Other ancient individuals have substantially different dietary profiles, namely the individuals from the Huron Wendat Nation, because of the $C_4$ plant and freshwater fish components to their diet.

#### Removing Contamination

A total of 453 libraries were labelled as contaminated, whether caused by flags from contamination tools or the analysis was inconclusive due to low depth of coverage; however, 305 those libraries remained in the project with the individual genome being tagged as contaminated. Although these individuals were not included in the analyses for this thesis, they will be included in analyses for publications to retain as many ancient individuals in the dataset as possible, with an analytical approach being taken in downstream analyses (149). Once genomes have been phased and imputed (which in some cases can reduce the contamination-associated signals), contaminated genomes will then be scrutinised in analyses such as MDS and *D*-statistics for effects of the contamination and, when appropriate, exclusions will be made.

#### Minimal Ancient DNA Damage Patterns

Libraries with low amounts of DNA damage generally fall into two broad categories. The first category comprises libraries that were generated from tissues with low preservation and therefore had low endogenous DNA percentage. This resulted in the sequenced reads being likely

from modern or environmental contaminating sources. The second category comprises libraries generated from locks of hair. The DNA extracted from hair, which is released from apoptotic skin cells and preserved in the keratin structure, does not usually contain substantial post-mortem degradation patterns, which is used to authenticate ancient DNA (150). Some libraries showed lower than expected DNA damage patterns given their C14 age, even with high endogenous percentage and low contamination flagged in the analyses. This is likely due to the tissue originating from the Arctic and being frozen for centuries or millennia, which maximises preservation (151).

## Genetic Sex Estimations

Of the 18 genomes that were consistent with XY but not consistent with XX, eight passed the recommended threshold of 100000 sequenced reads required to confidently assign genetic sex (**Table S7**) (145). All of these eight individuals had R_y values of more than 0.07, with the lower end of the 95% confidence interval (CI) not reaching the 0.077 threshold; however, three did exceed a lower CI of 0.0745, and given the number of reads assigned (*i.e.*, more than 19 million) an XY assignment may be justified. Of the 39 genomes that were not assigned a sex, 30 passed the recommended threshold of 100000 sequenced reads required to confidently assign sex (145), many of which had R_y values above 0.07; however, the upper limit of the 95% CI did not reach 0.077.

One individual, a lock of hair from the Nunalleq site in Quinhagak, Alaska (024727), seems to be an anomaly with reads assigned to the X and Y chromosomes (**Table S7**). This individual had more than 300 million sequenced and mapped reads, more than 10 million of which mapped to sex chromosomes, and almost 300000 mapping to the Y chromosome. This resulted in a R_y value of 0.0275, which suggests an intermediate ratio of X to Y mapped reads. Based on these metrics, it could be suggested that this individual is of XXY phenotype; however, further investigation is necessary (152,153).

## Species Identification

Genome 011875 was suggested to be *Bos mutus* by the MEGAN tool (146); this seemed unlikely as that species is not known to have existed in the region of the New Siberian Islands. In contrast when 011875 was suggested by the in-house pipeline to be *Rangifer*, it seemed far more likely based on the archaeological evidence and cultural affiliation to hunting reindeer. It is likely that data quality and/or reference database and assignment algorithm play a role in the observed discrepancy.

For genome 024119, few reads mapped to the *Ursus* genus, which is in line with the imprecise result from the MEGAN tool (146). The coarse hair, genome 024124, failed to be assigned a species with both approaches, this could be due to difficulties with DNA extraction from tissue with high melanin content, which often leads to downstream inhibition of amplification (154). Genome 011442 was simply a case of differential preservation levels of two human tissues from the same site. Genome 022903, from Yakutia and dated to the Iron Age, was investigated due to its high level of mitochondrial coverage and very low probability for the assigned mitochondrial haplogroup (50%, H2a2a) (155). 022903 was indeed identified as *Homo sapiens* and is likely to have been of Western Eurasian origin and the Haplogrep probability score is low because H2a2a is the default assignment when no haplogroup-defining variants are present (155).

## Kinship

The kinship analyses for both the geographically divided and merged dataset resulted in 11 pairs of identical individuals (with a low probability of them being identical twins). The agreement between the two methods allows for more confidence in the certainty of these individuals being identical. The Alaskan genomes (n=62) had the highest proportion of identical individuals, first-degree relatives, and second-degree relatives. The high number of identical individuals was anticipated as two locks of hair could easily have originated from the same person. The high level of relatedness at the Nunalleq site was also expected due to the high number of tissues analysed from the single site during a single time period and may indeed represent familial units buried together (92).

As was expected based on the nature of the READ tool, both the number of first-degree and second-degree pairs of relatives increased substantially when run on the merged dataset (148). This is because of the low genomic diversity in Arctic populations (7,9), resulting in longer runs of homozygosity and increased allele sharing between individuals, which can result in two individuals appearing more related to one another than their true kinship when compared to a larger dataset. In contrast to results from the geographically divided analysis, where 15 pairs of first-degree relatives were identified, 48 pairs of first-degree relatives were identified in the merged dataset. Similarly, while 34 pairs of second-degree relatives were identified in the geographically divided analysis, the merged dataset analysis identified 139 pairs of second-degree relatives (**Table 4**).

## Limitations

### Low Tissue Preservation, Low Tissue Input

A major limitation to the generation of this palaeogenomics dataset was the preservation level and the amount of each tissue that was available to be included in this project.

Preservation levels for tissues from cultural groups such as Late Ushki, Kanchalan, Tevi, and Tokarev from Siberia were very low, resulting in very few (in some cases no) genomes of adequate coverage. The Late Ushki site had 12 bags of heavily fragmented bones and one heavily fragmented milk tooth. Of these tissues, all were deemed contaminated, mostly due to contamination analysis failing because of low coverage (ranging from 0-0.13x, with the 0.13x genome being generated from the milk tooth fragments). The Ushki site is of great importance to research into ancient populations that migrated through Siberia into the Americas. The earlier layers at the Ushki site suggest that site may have been inhabited by the ancestors of present-day Indigenous groups in the Americas, and the later layers suggest that site may have been inhabited by the ancestors of present-day Indigenous groups in the Siberian and North American Arctic (79). Had the preservation levels of these tissues been better, more research questions pertaining to the

occupational history at this site, and consequences of the peopling of the Americas, could have been addressed. Similarly, low preservation levels in tissues associated with the Kanchalan, Tokarev, and Tevi Cultures resulted in minimal representatives from these groups being included in downstream analyses.

In some instances, there was very minimal tissue available for DNA extraction, often resulting in very low coverage genomes. This was the case for the Ushki milk tooth, which weighed only 38mg and had such thin fragments that its surface could not be cleaned. Similarly, some of the locks of hair were very low input. For example, many of the Nunalleq hair locks were as low as 6 mg of input and the locks of hair from the Tevi Culture were 9mg and 47mg (though it is still not clear whether the 47mg hair lock is of human origin). None of these low input hair locks produced genomes of adequate coverage.

### Gaps in Spatiotemporal Ascertainment of Ancient Genomes

As with many palaeogenomics studies, the dataset for this project was generated from tissues contributed to the study and approved for genomic sequencing by the representatives who speak on behalf of the ancient individuals. Despite being the largest palaeogenomics study of the Siberian and North American Arctic undertaken thus far, various spatiotemporal gaps in sampling exist in this Arctic and Sub-Arctic dataset. This limitation could be addressed in the future by obtaining tissues from two broad time periods and geographic regions. The first gap in the dataset are genomes from ancient Siberian individuals dating prior to 6-5 kya, which would enable researchers to address questions regarding the Palaeo-Inuit migration out of the Far East and into the North American Arctic. Some of these questions have remained unanswered owing to somewhat contradictory results from previous studies attempting to understand the genetic history of the Palaeo-Inuit and their relationship to ancient and present-day Siberians (2,25,156,157). The second gap in the dataset to address would be additional genomes from Alaska and Greenland representing the Palaeo-Inuit Culture, as well as additional genomes from Canada representing the Neo-Inuit cultural groups.

## Future Directions

### Compound Specific Isotope Analysis

There are some limitations to bulk isotopic analyses being used as a proxy to dietary composition, which can be addressed using Compound Specific Isotope Analysis (CSIA). CSIA examines the carbon and nitrogen atoms that make up the amino acids in the protein, lending additional insight to factors such as the proportional contribution of different food groups in the diet, as well as indicating what specific species were being consumed by the ancient individual (158). CSIA have been performed on most tissues at the ASIF, except for those ancient individuals where no tissue remained, or no permissions were in place. The CSIA results will be used to complement the palaeo-dietary profiles of the individuals included in this project.

### Marine Reservoir Correction

The Marine Reservoir Effect is a biological phenomenon where inaccuracies in radiocarbon dating results can be driven by the consumption of food sources from a marine ecosystem (123). Carbon degradation processes differ between atmospheric environments and marine environments. There are far fewer photosynthetic processes that occur in the deep sea resulting in carbon degradation to occur at a slower rate, causing the carbon to be older and the radiocarbon age to be inaccurate (123) .

Because many of the ancient individuals in this project had a diet rich in marine mammals, many of the radiocarbon ages will require Marine Reservoir Correction (MRC). To inform the need for, and extent of, MRC on the radiocarbon dates of each tissue, a correction factor will be calculated based on a combination of 1) the proportion of marine contribution in the individual's diet, and 2) the marine reservoir value ($\Delta R$) of that geographic region.

The first value required, the proportion of marine contribution in the diet, has been estimated using a three-end member (terrestrial, freshwater, and marine diets), two-isotope (carbon

and nitrogen) mixing model, outlined in **Appendix 3** (130–132). The results of which are outlined in section, Radiocarbon Dating and Bulk Isotopes. The second value required for MRC, the ΔR value, will be estimated for each of the following ten geographic regions or 'GeoGroups' using the average calculated on calib.org (159). GeoGroups shown in **Table 5**, were delineated based on known regions of differing marine reservoir values (160).

**Table 5** - Designated GeoGroups for correcting for the marine reservoir effect, based on known regions of differing values of marine reservoir effect.

| | |
|---|---|
| GeoGroup1 | Zhokhov/ New Siberian Islands |
| GeoGroup2 | Chukotka |
| GeoGroup3a | Western Kamchatka and Magadan |
| GeoGroup3b | Eastern Kamchatka, Commander Islands |
| GeoGroup4 | Western North America |
| GeoGroup5 | Eastern Canada |
| GeoGroup6a | High Arctic Canada |
| GeoGroup6b | Northeastern Canada |
| GeoGroup6c | Western Greenland |
| GeoGroup7 | Eastern Greenland |

### Constructing Ancient Pedigrees and Relatedness Over Time

In the instances where genomes were sequenced from multiple individuals at the same site and time frame, there is an opportunity for constructing ancient pedigrees and understanding kinship patterns in these ancient cultural groups (161). This will be done for four sites in this project: 1) the Nunalleq site in Alaska (n=84), 2) the Uummannaq site in Western Greenland (n=19), 3) the Dødemandsbugten site in Eastern Greenland (n=7), and 4) the Zhokhov site on the New Siberian Islands (n=6). Pedigrees of any close kin will be constructed using both uniparental markers and autosomal signals, comparing them to look for potential matrilocal versus patrilocal signals, as well as archaeological contexts (148,162). Investigation into kinship will also be extended to help shed light on whether the 'identical individuals' identified could in fact be identical twins (*i.e.*, 011229 and 011446 from Newfoundland, 011423 and 011433 from Nunangat, and 024247 and 024264 from Kamchatka). Additionally, using a method that leverages runs of homozygosity, including low coverage samples (163), parental relatedness over time in the Arctic will be evaluated which can provide insights into past social structures, mating patterns, and population sizes.

### Returning Results and Incorporating Oral histories

The results from this work will be returned to Indigenous groups and existing oral histories incorporated into the study. Findings from this research will be expressed as a different knowledge form to Indigenous knowledge, with no expectation for both to be concordant or to force reconciliation between the two. This approach may help to strengthen relations between palaeogeneticists and Indigenous groups, by bringing their knowledge to the forefront of population histories and shifting the common perception of genetic results being the main line of evidence for insights into the past (164).

## *2.5 Conclusion*

The palaeogenomic dataset generated for this project is more than adequate for addressing gaps in the existing ancient DNA knowledge base. Previously analysed tissues, as well as new tissues were processed with refined wet lab techniques, improved bioinformatic tools, and substantially increased sequencing capacity. This led to a dataset of unmatched quality and scale in Arctic palaeogenomics. This dataset, when accompanied by the improved working relationship with Indigenous communities, will enable comprehensive investigations into the occupational history of Siberia and the North American Arctic.

# 3 Chapter 3 – Population Genetics and Adaptation in Ancient Siberia and North American Arctic

## *3.1 Introduction*

### 3.1.1 Population Structure of Indigenous groups in Northeastern Siberia, the Americas, and the North American Arctic

Population genetics examines the genetic variation within and between population groups. Population genetics of present-day humans predominantly focuses on existing genetic diversity, from which past population demography can be difficult to discern because of compounding factors, *e.g.*, recent admixture events, population replacements, population bottlenecks. With the advent and rapid advancements in ancient DNA sequencing, research questions pertaining to ancient human history can be more thoroughly investigated (16,17). Together with novel population genomics methods, increasing volumes of ancient human genetic datasets have facilitated the reconstruction of historical population sizes, population structure (19), genetic origins (20,21), split times (22), and genetic admixture between populations (23,24).

Recent genome-wide examinations of ancient populations in northeastern Siberia and northwestern North America have provided insights on population origins, admixture, and migrations. The first known migration wave into this region was prior to 30 kya, by a group of "Ancient North Siberians" (ANS) who were related to Western Eurasian Hunter-Gatherers (represented by the Yana individual from north-central Yakutia, and the Mal'ta and Afontova Gora individuals from south-central Siberia) (25,26). Between 25 and 10 kya, there was a second migration of an East Asian-related population that admixed with ANS, giving rise to the ancestors of present-day Indigenous Americans. This ancestral Indigenous American population started splitting from East Asians ~36 kya, with persistent gene flow until ~25 kya, and ultimately diverged into the three main Indigenous American genetic lineages: Ancient Beringians (AB), North Native Americans (NNA) and South Native Americans (SNA) (22,27,28). The demographic

model inference, aided by an AB genome, provided evidence for the Beringian 'standstill model', with an ~8000 year period of isolation in Beringia prior to entering the Americas (27,29). Furthermore, continued East Asian-related gene flow into Northeastern Siberia during the Late Pleistocene and Early Holocene resulted in the metapopulation termed "Palaeo-Siberians" (represented by the Kolyma individual from north-eastern Yakutia) (25).

The third large migration into northeastern Siberia was during the Holocene, where another group of East Asian-related people moved eastward, becoming the "Neo-Siberians", who largely replaced the "Palaeo Siberians" and show genetic continuity with present-day groups in far eastern Siberia (**Figure 3.1**) (25). After ~11.5 kya there was a gene flow event from a northeastern Siberian population, genetically related to the Palaeo-Siberian Yana and present-day Koryaks, into NNA (25). During the last ~4 kya, there is also evidence for genetic admixture between NNA and the Siberian ancestors of present-day Inuit (25,27).

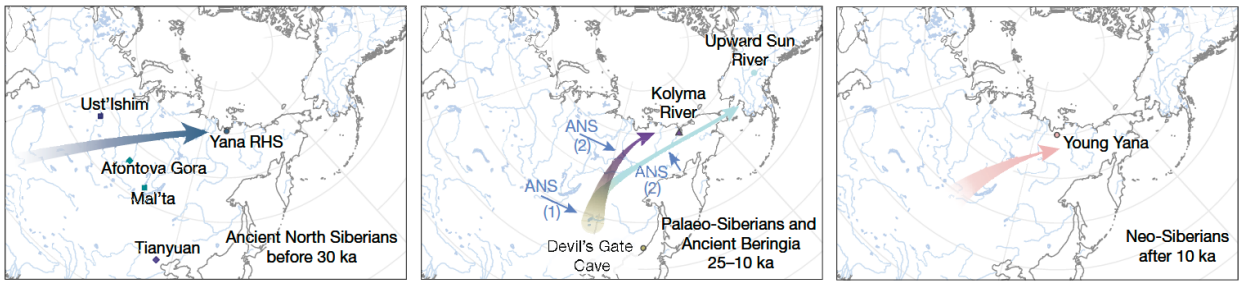



**Figure 3.1** - Suggested migration patterns of genetically distinct populations in the eastern Siberian and North American Arctic over time. The migration of the Ancient North Siberians before 30 kya is shown with a dark blue arrow, the migration of the Palaeo-Siberians and Ancient Beringians between 25-10 kya are shown with purple and light blue arrows, respectively, as well as the migration of the Neo-Siberians ~10 kya shown with a pink arrow (figure from Sikora *et al*., 2019, p. 5 (25)).

The North American Arctic has a distinct occupational history compared to the rest of the Americas and was populated much later by two distinct archaeologically defined cultural traditions from Siberia: the Palaeo-Inuit and the Neo-Inuit (**Figure 3.2**). Based on archaeological evidence, the Palaeo-Inuit migrated into the North American Arctic ~5.5 kya and disappeared from the archaeological record ~1 kya. In the span of ~4.5 kya, the Palaeo-Inuit tradition has been archaeologically classified into the Saqqaq (in Greenland) and Pre-Dorset and the Dorset Cultures (in Canada) (31). The Palaeo-Inuit show genetic continuity for ~4000 years and, therefore, may represent a single source population manifesting as different cultural units, perhaps in response to

changing climate and dietary ranges (2). Furthermore, the Palaeo-Inuit are genetically distinct from the Neo-Inuit and other Native American groups, including Athabascan speakers (2,156). The Neo-Inuit, who admixed with NNA, arrived in the North American Arctic in a later wave ~1 kya and transitioned into present-day Inuit, to whom they are genetically and culturally ancestral (2).

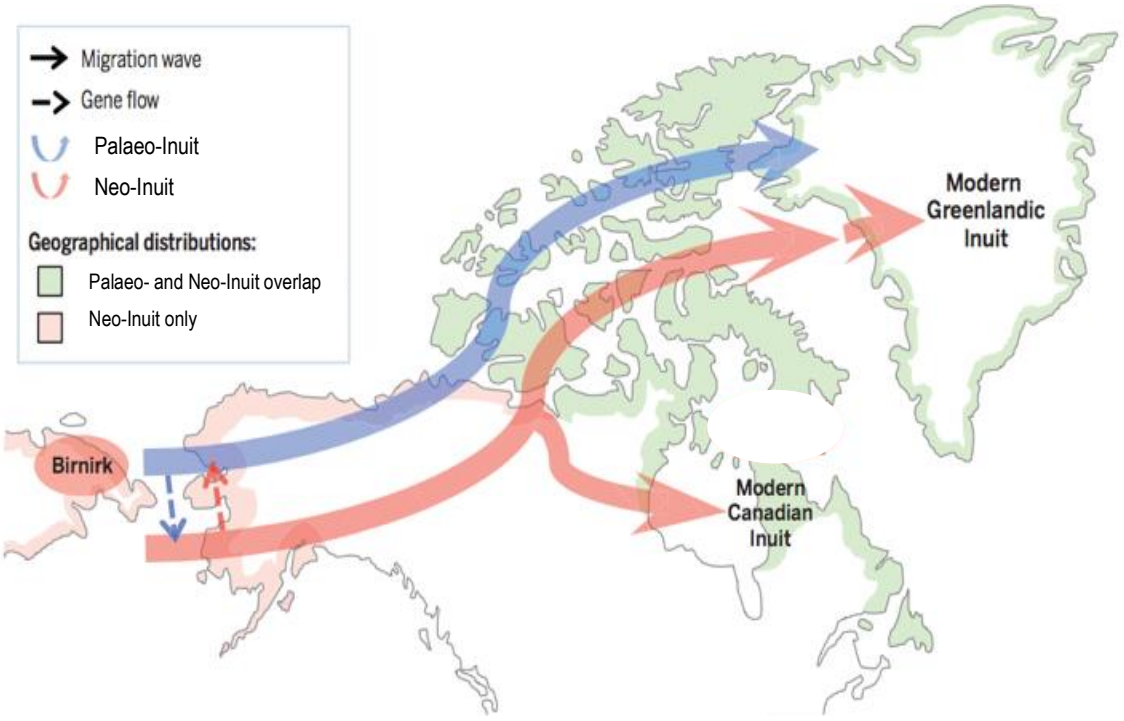



**Figure 3.2** - Migrations of genetically distinct populations into the North American Artic, with the Palaeo-Inuit migration represented by the blue arrow and the Neo-Inuit migration represented by the red arrow (figure adapted from Raghavan *et al*., 2014, p.1 (2)).

Establishing migrations and genetic admixture between ancient Arctic populations can contribute to refining their population histories and population structures, as well as resolving the cultural versus biological overlaps (2). There are currently limited archaeological findings that provide insight into prehistoric contacts and admixture between late Palaeo-Inuit and Neo-Inuit. Early radiocarbon dating from eastern regions of the Arctic indicates the Palaeo-Inuit disappeared more than 200 years prior to the Neo-Inuit entering that geographical region (6,33). In contrast, more recent radiocarbon dating suggests an overlap of between 50 and 200 years of the Palaeo- and Neo-Inuit, with geographical coexistence in western regions of the Arctic (6). This has led to an ongoing debate surrounding phenomena such as the rapid spread of the Neo-Inuit and near-concurrent cultural collapse of the Palaeo-Inuit across the Arctic.

The potential overlap between Palaeo- and Neo Inuit is also difficult to discern using published ancient genomes. Because of the reported ancient gene flow between these groups, subsequent gene flow events cannot be easily identified, especially with existing low coverage genomic data (*i.e.,* 0.001x-0.3x) (2). While the low depth genomes from these cultures have previously indicated potential genetic admixture between the Palaeo-Inuit and Neo-Inuit, studies were limited in performing quantitative assessment of the level, timing, and source of the admixture events (2). The only high-depth reference genome for Palaeo-Inuit has been the Saqqaq (Greenlandic Palaeo-Inuit population) genome, however it has relatively high error rates (30). Hence, there are still several unanswered questions relating to the population history of these ancient Arctic cultures that will benefit from higher depth genomes.

### 3.1.2 Natural Selection in Arctic and Sub-Arctic Populations

Over time and space, humans have undergone various adaptations to their environment. Using genomic data, coupled with demographic models and existing evidence from fields such as archaeology and anthropology, adaptations due to environmental pressures can be examined through population genetics methods (34,35). One way that selection can be investigated is by looking for *loci* with marked differences in allele frequencies between populations or over time. It must be considered, however, that such differences in allele frequencies could be also due to genetic drift (allele frequency changes occurring by chance). Separating true selective pressures from genetic drift can be difficult, particularly when a group has a small effective population size, making it more susceptible to drift. Therefore, additional considerations, such as if the highly differentiated genetic variants occur in linkage blocks or in genes associated with a certain trait, must be taken into account before it can be speculated that a particular adaptation is the product of selection.

Based on the long-term survival and ability of populations to thrive in the harsh Arctic climate, where resources can be scarce and population sizes were small, regions of the genome could have been selected for through adaptations to regional environment factors such as cold,

diet, and exposure to novel infectious diseases (7,36). Investigating selection using only present-day genomes is very complex, particularly with selective pressures that happened in the distant past. Through the advent and advancement of sequencing ancient human genomes, it became possible to examine selection events in real time and to establish when a selected allele may have reached fixation. Even if the derived allele has not yet reached fixation, examining allele frequency changes through time may help to narrow down the period in which the selective pressure was prevalent. Because many Arctic populations were founded by small groups of individuals, and sustained a small population size over time, the caveat surrounding the relative impacts of drift versus selection must be considered and investigated using other genomic features such as haplotype blocks or genic and functional context.

Many studies have been conducted on genetic sites under selection in Arctic genomes, which are often found to contribute to adaptation to the extreme environment, such as exposure to cold, shifts in dietary composition, changes in stature, *etc*. When examining genomes of present-day Siberian populations, one of the strongest selective sweeps on the human genomes was identified in the *CPT1A* gene, which codes for a protein that functions in both metabolising high-fat diets as well as resilience to cold climates (9,37). Another investigation into present-day Siberian populations showed evidence for positive selection for genes involved in brown adipose tissue production. Specifically, these genes were involved in lipid metabolism (*PLA2G2A*, *PLIN1*, and *ANGPTL8*) and result in an increase in basal metabolic rate and low serum lipid levels (10). Similarly, a study conducted on the Greenlandic Inuit found the strongest signal for selection to be in the *FADS2* gene (8). The *FADS2* gene codes for a protein that assists in regulating the level of polyunsaturated fatty acids in the blood, which can affect an individual's height and weight (8). Candidate genes such as these can be examined in ancient populations to estimate the timing, origin, and magnitude of selective pressures.

### 3.1.3 Purpose

The primary aim of this chapter is to conduct population genetics analyses, using population-wide low coverage genomes and select higher coverage genomes from ancient Siberian, Alaskan, Canadian, and Greenlandic populations. Population genetics tools will be used

to investigate source populations, population structure, as well as timing and magnitude of admixture between groups in the eastern Siberian and North American Arctic. This work aims to further refine and expand upon demographic models in the palaeogenomics knowledge base from low-coverage genomes in publications such as Raghavan *et al.,* 2014 (2).

A second aim of this chapter is to assess the effects of environmental pressures on the genomes of Arctic and sub-Arctic individuals, with emphasis on the impact of shifts in subsistence strategies from a terrestrial to marine-based diet. Palaeo-dietary profiles, which capture the lifestyle of these ancient groups based on stable carbon and nitrogen isotopic analyses (outlined in Chapter 2, section Radiocarbon Dating and Bulk Isotopes), will add context to the patterns of changes in allele frequencies identified in the genetic data and help characterise the impact of natural selection over time in the Arctic.

### 3.1.4 Research Questions

This chapter addresses three main research questions related to population structure and demographic history that are now feasible to investigate because of technological and methodological advances, as well as increased access to, and approvals for processing of ancient tissues. The first area of emphasis is on the genetic characterization of the Zhokhov individuals from the New Siberian Islands, who are geographically isolated and archaeologically unique due to their coastal Arctic lifestyle without any evidence of hunting marine mammals. The second area of emphasis is the Palaeo-Inuit populations in the Canadian Arctic and how they relate to the contemporary Saqqaq population (represented by the individual originally sequenced in Rasmussen *et al*., 2010 (30)) as well as to Neo-Inuit and present-day Inuit populations. The third area of emphasis will be on the formation of the Neo-Inuit populations and their relation to each other.

This chapter also examines the effect of environmental pressures, especially transitioning from terrestrial to marine subsistence strategies, on the distribution of variants pertaining to these adaptations in the genomes of populations inhabiting the Arctic through time.

### 3.1.5 Outline of Chapter

This chapter outlines the population genetics analyses used to investigate the genetic origins, relationships, and genetic admixture between ancient Arctic populations as well as with sub-Arctic populations. Additionally, the genomes of these ancient individuals are examined for putative evidence of selection to the Arctic landscape. The results from these analyses are then described and interpreted alongside existing evidence from previous palaeogenomics studies. The findings from this chapter are then discussed, the limitations are summarised, and the future directions for this work are outlined.

## *3.2 Methods*

To infer the maternal and paternal lineages, uniparental markers were investigated for individuals where coverage of the mitochondrial DNA and Y chromosome was adequate, respectively. On a genome-wide level, the low-coverage (screening) and high-coverage (selected as representative of population group) genomes were investigated for demographic features such as population history as well as admixture levels and timings. This was done using tools such as MDS (165), outgroup $f_3$ and *D*-Statistics (166), TreeMix (167), qpGraph (166), DATES (168). In addition, an in-house allele frequency-based approach was used for examining sites under putative selective pressure through space and time.

### 3.2.1 Mitochondrial Haplogroup

SNPs of the mitochondrial (mt) genome were identified using bcftools mpileup and were then run through Haplogrep for mt haplogroup assignment (155). A mitochondrial haplogroup was assigned and given a probability score. The majority of the haplogroup inferences were based on mitochondrial genomes with an average depth of coverage >5X. A subset of the haplogroup assignments were based on mt genomes with an average depth of coverage <5X to maximise information (details in **Table S8**). Despite the lower depth in these cases, these haplogroups were reported in the results section given the rarity of the haplogroups assigned to these individuals and their consistency within individuals of similar cultural groups.

### 3.2.2 Y-Chromosome Haplogroups

Y-chromosome haplogroups were called for 116 ancient individuals based on the genetic sex assignments shown in **Table S7**. This included individuals assigned XY (n=59), those 'Consistent with XY but not XX' (n=18), and those 'Not Assigned' (n=39). Y-chromosome reads were extracted using samtools view and the coverage was calculated using samtools depth with a reference bed file that correctly delineates the callable regions of the Y-chromosome for ancient DNA analyses (*i.e.*, the 10 Mb short-read mappable region) (169,170). Individuals with less than

0.01X coverage were then excluded. Bcftools mpileup and call functions were used to call genotypes on the Y-chromosome (169). Haplogroups were called using an in-house pipeline that determined the root-to-tip paths for haplogroups, the number of supported variants observed, and matched variants to the ISOGG database (171).

### 3.2.3 Reference Datasets

The reference dataset used for many of the population genomic analyses was compiled from various sources, as described in Moreno-Mayar *et al*., 2018 (Supplement Materials, p. 20 (22)). This dataset is a combination of genome-wide sequencing data from relevant palaeogenomics publications and SNP array data that was selected for Moreno-Mayar *et al*., 2018 (22). The resulting reference dataset is composed of 199208 genotyped SNPs from 2537 present-day individuals belonging to 167 ethnic groups from around the world, which were enriched for Native Americans. For present-day Native American individuals, Non-Native American ancestry tracts were masked to reduce compounding signals of recent admixture. Sites not included in the 1000 Genomes Project's strict-accessible regions were filtered out.

The 182 ancient human genomes generated in the project, which were found to have no identical individual or primary relative in the dataset and no evidence of contamination, were merged with the reference dataset for use in downstream analyses. Individuals were then filtered out based on having more than 95% missingness in the data (n=50), as shown in **Figure 3.3**. In this thesis, pseudo-haploid calls, where one random allele is selected at each site, were generated and used for all presented analyses.

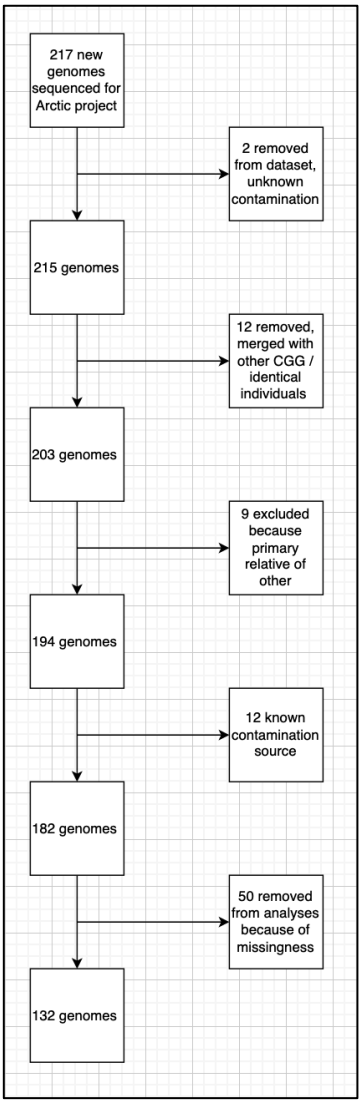



**Figure 3.3** - Flow chart depicting the exclusion of genomes from the population genomic and adaptation analyses.

### 3.2.4 Multidimensional Scaling

Multidimensional Scaling (MDS) is a data visualisation tool that uses the pairwise distance between data points to infer their similarity (172). MDS plots present each datapoint in a lower-dimensional representation to reduce the variability in the data, resulting in a scatter plot depicting clustering and stratification in the dataset (173). In the context of population genetics, MDS is used to visualise genetic relatedness between individuals, where distance between two individuals on the MDS plot is proportional to the allele-sharing distance between them (165).

In this thesis, MDS transformations are used to explore the genetic relationship between the ancient Arctic individuals and other ancient and present-day individuals. MDS transformations are computed on different subsets of individuals in the reference dataset. Beginning with a reference panel subset that includes world-wide representative populations, the MDS helps to inform the general positioning of the newly sequenced genomes. Subsequent MDS plots are generated with only a subset of the reference populations. Those that are not genetically similar are removed to focus on getting a better resolution on the genetic affinities of the ancient Arctic and sub-Arctic individuals. This allows for a closer examination of populations from specific geographic regions or archaeological groupings. Patterns observed on MDS plots inform research questions and hypotheses that are investigated using other methods discussed below.

### 3.2.5 Outgroup $f_3$-Statistics

Outgroup $f_3$-statistic can be used to investigate the genetic affinity between two populations when compared to an outgroup population. Here, the outgroup $f_3$-statistic was used to compare ancient Arctic populations to other ancient and present-day populations. Outgroup $f_3$-statistic in the form $f_3(\text{Pop}_{\text{out}}; \text{Pop}_A, \text{Pop}_B)$ was estimated for each individual and for each pool. Under the assumption of no admixture between the ingroups (*i.e.*, $\text{Pop}_A$ and $\text{Pop}_B$) and the outgroup (*i.e.*, $\text{Pop}_{\text{out}}$), the outgroup $f_3$-statistic indicates the shared drift between $\text{Pop}_A$ and $\text{Pop}_B$ after the divergence of the ancestral population from $\text{Pop}_{\text{out}}$. The Yoruba was used as an African outgroup in all the presented tests, based on the assumption that no Arctic or sub-Arctic population is likely to have a higher proportion of African ancestry than another. The outgroup $f_3$-statistic was

computed as described in Patterson *et al.,* 2012 (166), and the standard error was calculated for each $f_3$-statistic with a weighted block jackknife approach using 5Mb windows to account for linkage disequilibrium (26,166).

The outgroup $f_3$-statistic was computed for all pairs of individuals in this project independently, and then again on designated groups or pools of individuals with a similar genomic ancestry profile. This was done based on results from MDS plots, individually run outgroup $f_3$-statistic, and matching cultural affiliation from the archaeological record.

### 3.2.6 *D*-Statistics

*D*-statistics can be used to test for treeness between groups in a four-population unrooted tree of the form $D(Pop_{out}, Pop_{test}; Pop_A, Pop_B)$. The method compares all SNPs between the two ingroup populations ($Pop_A$ and $Pop_B$) and determines whether any of those populations share significantly more alleles with the third population ($Pop_{test}$) than the other does. When there is asymmetrical allele sharing between $Pop_{test}$ and $Pop_A$ or $Pop_B$ (*i.e.*, when the *D*-statistic deviates significantly from 0), the null hypothesis of treeness can be rejected. In most cases, the interpretation of the alternative hypothesis is that $Pop_A$ and $Pop_B$ do not form a clade to the exclusion of the other two populations or that there was admixture between one of the ingroups and $Pop_{test}$ or $Pop_{out}$ (174,175). Since the method is focused on the internal branches of the tree, the independent demographic histories of each population (*e.g.*, population bottlenecks) do not impact the outcome. To test for statistical significance, a leave-one-out weighted block jackknife approach (with 5Mb blocks), was used to estimate a standard error and compute a Z-score (166,174,175). If the Z-score was greater or equal to |3.3| (p-value ~0.001), the null hypothesis was rejected (166).

Herein, the *D*-statistic was applied to various postulated four-population sets that include Arctic and sub-Arctic groups and select worldwide populations. Trees were formulated based on independent ancient Arctic individuals acting as cultural representatives or based on 'pools' designated by cultural affiliations and similar groupings on a MDS plots. The reference

populations included in the figures generated from the *D*-statistics analysis are the same as the outgroup $f_3$-statistic (outlined above).

### 3.2.7 TreeMix

TreeMix is a tool that models population splits and admixture events in numerous populations concurrently by displaying these patterns in a bifurcating tree that is best fitting to the demographic history. This tool generates maximum likelihood, fixed-root trees that are constructed by genetic distances, which are based on the allele frequencies in each population. The tool then optimises the tree to better fit the data by adding in admixture events (167).

TreeMix was run on the ancient Arctic individuals whose genomes were sequenced to high coverage, acting as cultural representatives. This was done to support admixture graph results and to produce a tree that relates selected populations. SNPs were grouped into 5Mb blocks, to correct for linkage disequilibrium, and admixture graphs with zero to three migration edges (admixture events) were fitted (22). This was done by enabling a final rearrangement of the tree, using the '-global' parameter, while the built-in sample size correction procedure was disabled, using the '-noss' parameter (22). Because of the low number of populations and the low number of individuals per population, all replicates for each number of migration edges yielded the same likelihood. Therefore, trees of a single replicate are shown in the results section.

The output of the TreeMix tool is: 1) a maximum likelihood, fixed-root, tree of all populations and individuals being investigated, and 2) a matrix of pairwise residuals of the populations or individuals. The fixed-root tree has the drift parameter on the x-axis, which suggests the branch lengths estimate based on heterozygous alleles at each site. However, the genomes in this project are pseudo-haploid, while the reference populations were genotyped resulting in the external branch lengths to be disproportionately long. This is due to the fact that the tool is not able to assess the heterozygosity in pseudo-haploid individuals. The colour of the migration edges is proportional to the magnitude of the estimated admixture event, where larger amounts of gene flow are represented by warmer colours (oranges and reds). After fitting the tree, population pairwise residuals were summarised in a heatmap. When adding migration edges, the preferred

configuration was chosen such that the population pairwise residuals were minimised, thus addressing the populations with the largest number of unresolved genetic relationships.

### 3.2.8 *f*-Statistics-Based Admixture Graph

qpGraph is a tool for fitting arbitrarily complex admixture graphs relating different individuals or populations using genetic data (166). This tool generates tree-like models, based on patterns of allele frequency correlations between populations, where admixture events can be incorporated into leaves, which represent sampled individuals or populations (166). Models generated by qpGraph can be validated by the comparison of the observed $f_4$-statistics and the $f_4$-statistics that are predicted by the fitted model (166). Quantitatively testing whether a model fits the data better than a simpler, nested model is one major advantage that differentiates qpGraph from other tools for admixture graph fitting, such as TreeMix (166).

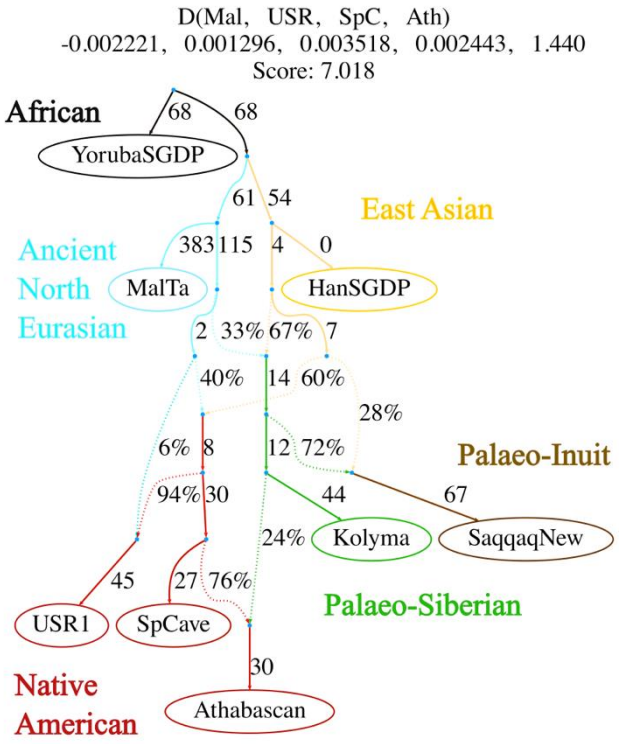



**Figure 3.4** - The 'Starting Graph', which was a reasonable fit for the key populations (composed of one or more individuals), from previous publications. The different colours broadly correspond to genetic ancestry components; black for African, blue for Ancient North Eurasians, orange for East Asians, green for Palaeo-Siberians, red for Native Americans, and brown for Palaeo-Inuit.

The qpGraph tool was used to explore the phylogenetic placement of ancient Arctic genomes on a tree and the genetic relationships between population groups. **Figure 3.4** depicts the 'Starting Graph', which was a reasonable fit for a small number of key populations (composed of one or more individuals), based on findings from previous palaeogenomic publications (2,22,25–27,30). The key populations are shown in different colours that broadly correspond to genetic ancestry components; black for African, blue for Ancient North Eurasians, orange for East Asians, green for Palaeo-Siberians, red for Native Americans, and

brown for Palaeo-Inuit. Using this Starting Graph, the Arctic genome of interest was grafted onto the model, on all possible branches and model testing was performed following the methodology in Moreno-Mayar *et al.,* 2018 (22).

Using the admixturegraph R package, the test population is added as a 'non-admixed' or as an 'admixed' leaf, otherwise known as a terminal node (176). For grafting leaves in a non-admixed model, the terminal nodes are branches of either internal or external edges of the Starting Graph. For grafting leaves in an admixed model, the test population is extended off any possible pair of edges in the starting graph. Then, qpGraph was used to optimise the admixture proportions and branch lengths as well as to calculate the fit score (log-likelihood), which corresponds to the likelihood of the data under the fitted model. A higher score indicates a worse likelihood. Therefore, for model testing, only models with the lowest fit scores are considered (166).

In each admixture graph result, as there is in the Starting Graph, there are three rows of text which indicate the worst $f_4$-statistics residual and the four-population set leading to this residual, the values of the $f_4$-statistics (both expected and observed), the value of the residual, the standard error, the residual Z-score, and the overall model fit score. Solid lines connecting the leaves illustrate direct linkages with the values to the right indicating optimised drift parameters. Dashed lines connecting the leaves illustrate admixed lineages, with percentages to the right indicating the admixture proportions.

The top five non-admixed and admixed admixture graphs are presented for each test population. The graphs in each panel were sorted from left to right according to the order of model fit score. In each case, fit scores are used to test if the non-admixed model can be rejected in favour of the admixed model. When a fit score (log-likelihood) difference greater than 3 (which corresponds to a p-value ~0.05) was observed, then the non-admixed model was rejected. Since the population history of northeastern Siberians is considerably complex and an admixture graph representation is likely an oversimplification, in some of the results below, a more conservative significance threshold of 4.6 (which corresponds to a p value ~0.01) was also considered for model testing (22,177).

### 3.2.9 DATES

DATES (Distribution of Ancestry Tract of Evolutionary Signals), is a tool that assesses the admixture linkage disequilibrium in an admixed genome, based on the difference in allele frequencies between two ancestral populations (168,178). This tool can estimate the time since an admixture event, measured in generation times of between 25 and 30 years (179,180), from a single diploid genome and two ancestral populations. DATES models the genotypes in an admixed genome as a combination of the two ancestral genomes, computing the likelihood of the observed genotype based on the allele frequencies and the estimated ancestry proportions of the two ancestral genomes, while considering recombination. As admixture linkage disequilibrium in a population decreases as a function of the time since admixture, this metric can be used to estimate the time in which the two source populations gave rise to the admixed population (168,178). An exponential function is fitted to the ancestry covariance decay curve from which the number of generation times can be estimated, where a steeper curve indicates an older admixture event, and a flatter curve represents a more recent admixture event.

DATES was used, with the default parameters and a generation time of 29 years (178,181), to approximate the date of the admixture event that led to the formation of six Neo-Inuit populations: the 'Old' Old Bering Sea Culture (OBSC1) from Chukotka (n=1), 'Young' Old Bering Sea Culture (OBSC3) from Chukotka (n=3), the Thule from northeast Greenland (n=14), the Thule from northwest Greenland (n=26), the Thule from south Greenland (n=8), and the ancient Yupik from Alaska (n=73). To examine the occurrence of these admixture events, representatives of the ancient Saqqaq Culture and Athabascans were used as proxies of ancestral source populations. Because only a single Saqqaq ancient genome and four present-day Athabascan genomes have been published, homogeneous source populations (based on previous admixture graph results) were pooled as following:

1. Saqqaq-like source population = one ancient Mid-Dorset genome from Nunavut, Canada (this project) + one Pre-Dorset genome from Nunavut, Canada (this project) + one Saqqaq genome from Qeqertarsuaq, Greenland (published, (30))

2. Athabascan-like source population = two present-day Athabascans (Dakelh, of Burns Lake First Nation) genomes from central British Columbia, Canada (published, (182)) + two present-day Chipewyan (Dene, of Fond du Lac Denesuline First Nation) genomes from northern Saskatchewan, Canada (published, (183)).

### 3.2.10 Allele Frequency Changes due to Environmental Pressures

An in-house computational tool was developed to explore allele frequency trajectories with the rationale that sites under putative selective pressure can be followed through time and space to visualise the occurrence and rate of selection. This tool depicts the frequency of an allele within a population, which in some cases is represented by only a single or few ancient individual(s). Sites of the genome known to be under selection for an adaptive trait can be examined together to discern the timeframe during which the adaptation may have occurred. When the genomes of many ancient individuals are included, observing changes in allele frequencies can indicate which population groups may have undergone an adaptive transition. When these sites are investigated on a larger scale, this tool may help to elucidate the causative variant. This approach can be used on data which has not been genotyped or phased and therefore can be used as a preliminary analysis before embarking on more detailed selection scans and analyses over time, such as the dynamic maps as described in Muktupavela et al., 2022 (184).

This tool was run on sites known to be under selection in Arctic genomes (predominantly from present-day genomic data), which were selected from the literature (8–10,34,37,185). As the causal SNPs among these selected sites are unknown, multiple *loci* were included in this analysis. In total, 372 SNPs were short-listed for inclusion in the analysis. These 372 SNPs were compared by combining diploid genotype calls for reference outgroup populations (that is, Africa and Oceania), and pseudo-haploid calls from previously published ancient genomes and the new ancient genomes from this project. The two possible alleles are reported as A1 and A2. The populations used as reference for each iteration of the analysis, whether reference genomes or ancient genomes, are those populations with non-missing data at that specific site.

Of the initial 372 sites selected for this analysis, only 246 were ultimately analysed. SNPs that were excluded were done so for one of the following reasons: the SNP site was not present in the strict mappability mask from the 1000 Genomes Project (186), the SNP site was not present in the diploid whole genome dataset (22), the SNP site was lost during lift over from the hg19 to hg38 human reference genomes, or the SNP site fell on an ALT contig instead of the primary assembly (these sites are often not reliable, especially in ancient genomic data).

One environmental pressure that was investigated in-depth was the transition from hunting terrestrial animals to hunting marine animals, since subsistence strategies and access to prey species are significant factors that dictate survival in the harsh Arctic landscape. To examine the genomic selection associated with this change in dietary composition, bulk nitrogen and carbon isotopic profiles of the ancient individuals were measured (and the results outlined in Chapter 2, section Radiocarbon Dating and Bulk Isotopes). Based on the ratio of these stable isotopes, the marine component of these palaeodiets was deduced and considered in reconstructing allele frequency-based trajectories.

## *3.3 Results*

### 3.3.1 Mitochondrial Haplogroups

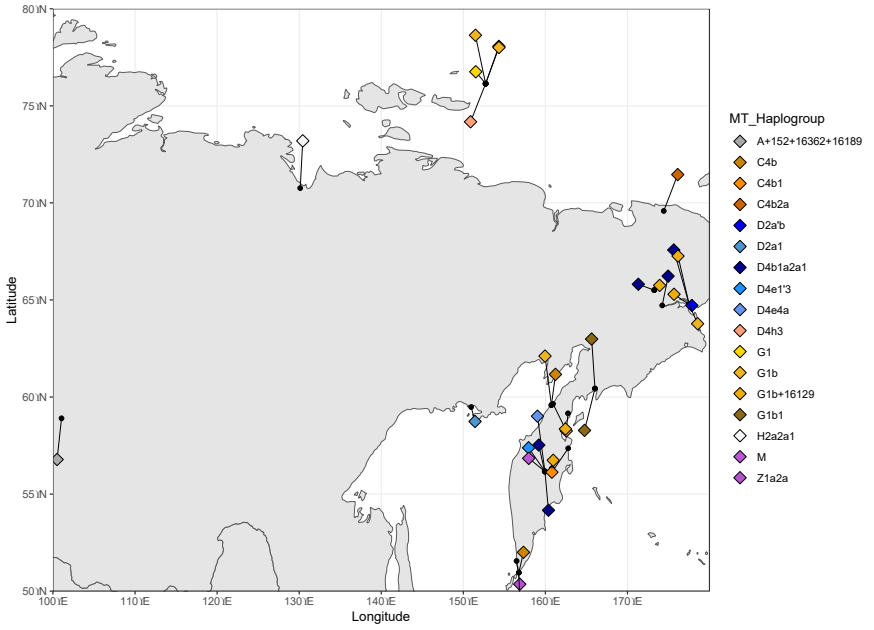



**Figure 3.5** - Map of eastern Siberia, indicating the mitochondrial haplogroups of the ancient individuals from the region.

In total, 172 ancient individuals had more than 5x coverage on their mitochondrial genome. With few exceptions, haplogroups were assigned with a probability of over 0.77, indicating high support. Haplogroups vary among cultural groups, with some trends seen in related populations. Siberian populations consistently display haplogroups G (G1, G1b, G1b-16129, G1b1), C (C4b, C4b1, C4b2a) and D (D2a'b, D2a1, D4b1a2a1, D4e1'3, D4e4a, and D4h3 in **Figure 3.5**), while the Palaeo-Inuit all have haplogroup D (D2a, D2a1 in **Figure 3.7**) and the Neo-Inuit populations predominantly display haplogroup A (A2a, A2a1, A2a3, and A2b1 in **Figure**

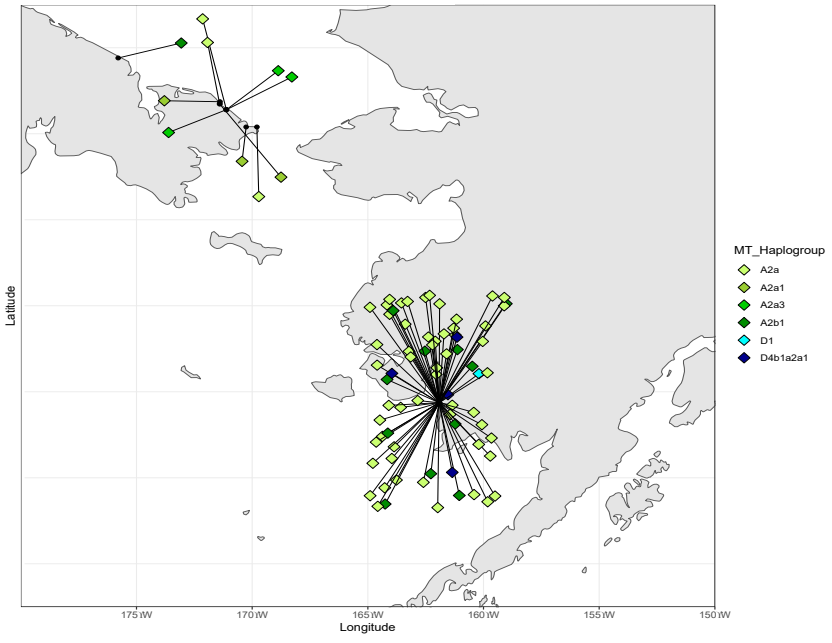



**Figure 3.6** - Map of the Bering Strait (including the Chukchi Peninsula and western Alaska), indicating the mitochondrial haplogroups of the ancient individuals from the region.

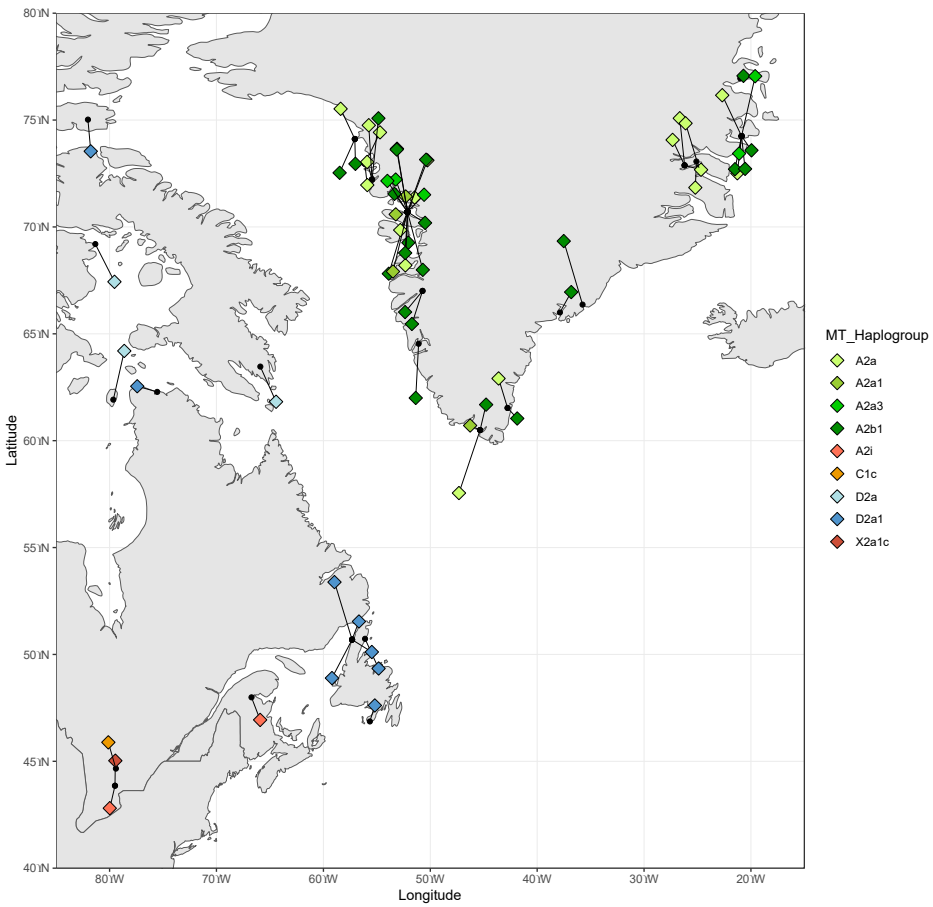



**Figure 3.7** - Map of eastern Canada and Greenland, indicating the mitochondrial haplogroups of the ancient individuals from the region.

**3.6**) and, in some cases, haplogroup D (D1, D4b1a2a1, **Figure 3.6** and **Figure 3.7**). Further specification of haplogroups is shown in **Table S8**. These results are contextualised in the Discussion, section Uniparental Markers.

## 3.3.2 Y-Chromosome Haplogroups

Of the 116 ancient individuals included in the Y haplogroup analysis, 84 had more than 0.01X coverage on the Y-chromosome (ranging from 0.0105x to 30.857x). Of these 84, 79 individuals were assigned to haplogroups Q, N, C, and R.

Haplogroup Q was identified in individuals originating from Central Siberia to Greenland: the Early Neolithic from north of Lake Baikal (Q1b-YP4010), the Zhokhov Islands (Q1b-YP4010) shown **Figure 3.8**, in the Old Bering Sea (Q1a-B143 and Q1b-M3 (Y4303)) and Birnirk Culture of the Chukchi Peninsula (Q1a-B143), the Alaskan Yupik (Q1, Q1a-B143, Q1b, Q1b-L53, Q1b-

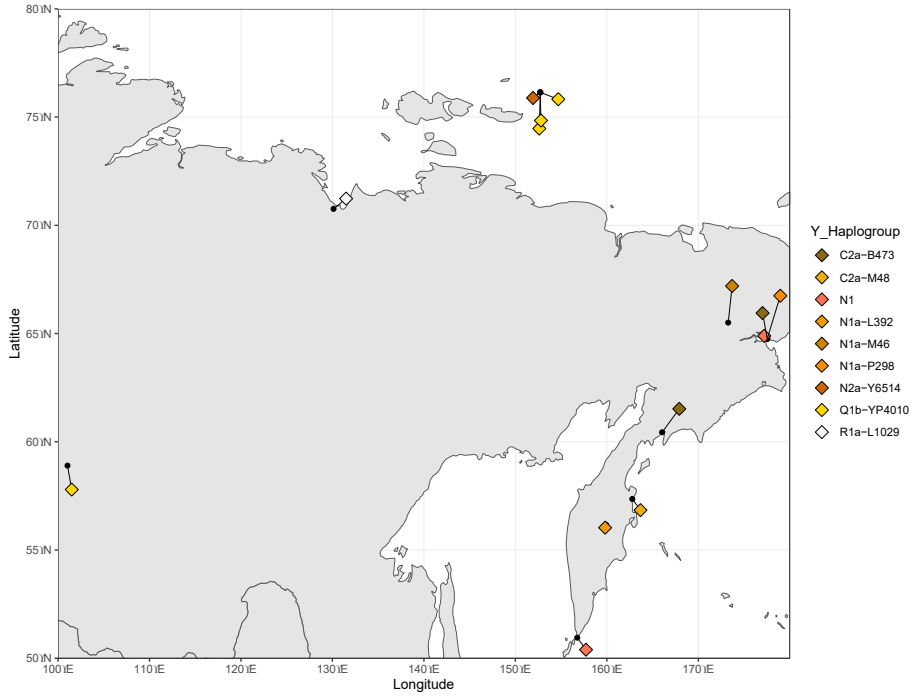



**Figure 3.8** - Map of eastern Siberia, indicating the Y-chromosome haplogroups of the ancient individuals from the region.

The N1 lineage was identified in individuals from the Old Itelmen (N1) and Tarya Culture (N1a-L392) from the Kamchatka Peninsula, the Kanchalan Culture (N1 and N1a-P298) shown in **Figure 3.8**, and Ust-Belaya Culture (N1a-M46) from Chukotka, and the Yupik from Alaska (N1a-L392) shown in M3 (Y4303)) shown in **Figure 3.9**, the Pre-Dorset (Q1a-B143) and Dorset (Q1a-B143) from northern Canada, the First Nations groups for the Canadian Eastern Woodlands (Q1b-M3 (M848), Q1b-M3 (Y4303)), and the Greenlandic Thule (Q1a-B143, Q1b-L53, Q1b-M3 (Y4303)), shown in **Figure 3.10**.

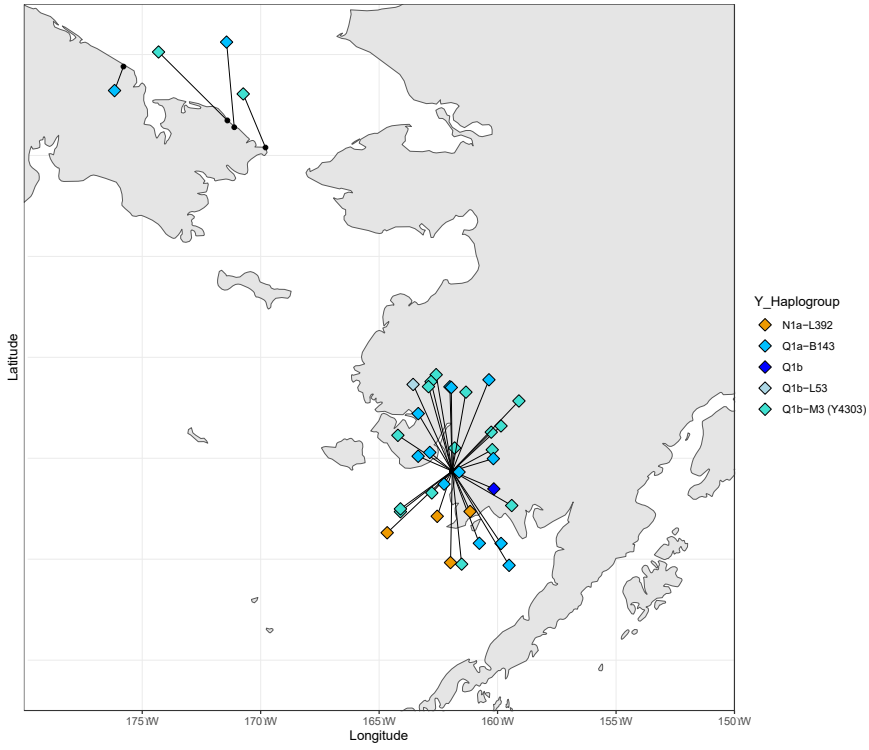



**Figure 3.9** - Map of the Bering Strait (including the Chukchi Peninsula and western Alaska), indicating the Y-chromosome haplogroups of the ancient individuals from the region.

**Figure 3.9**. A single N2a lineage was identified in one of the Zhokhov individuals (N2a-Y6514) shown in **Figure 3.8**.

The C2a lineage was identified in individuals from the Old Koryak Culture from Kamchatka (C2a-B473 and C2a-M48) shown in **Figure 3.8**, and the Kanchalan Culture from Chukotka (C2a-B473) shown in **Figure 3.9**.

A single R haplogroup was found in the individual from Yakutia (R1a-L1029) shown in **Figure 3.8**.

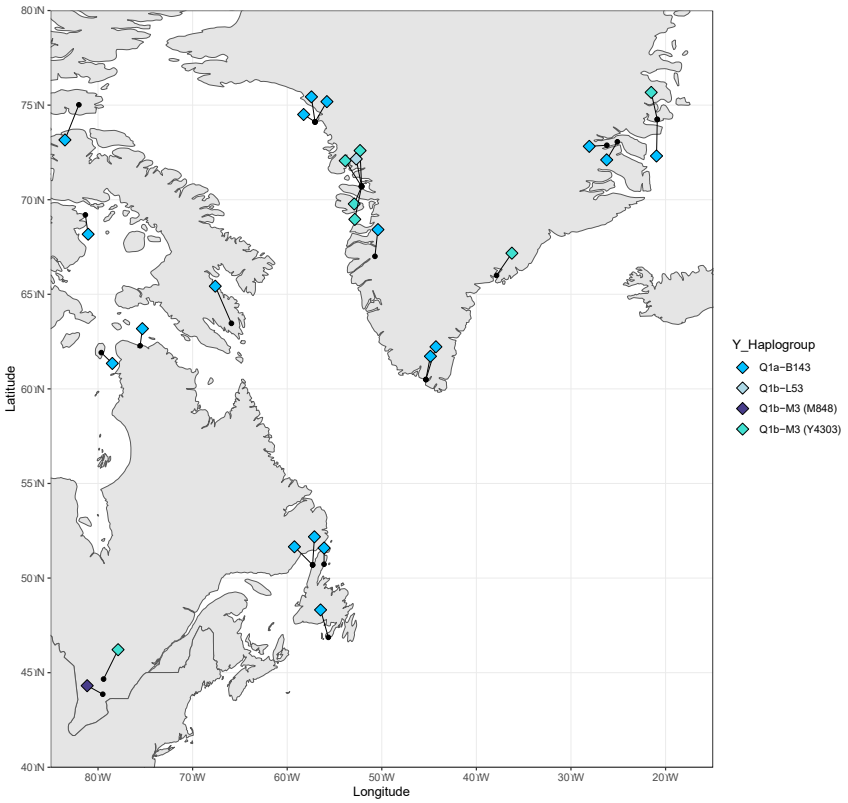



**Figure 3.10** - Map of eastern Canada and Greenland, indicating the Y-chromosome haplogroups of the ancient individuals from the region.

### 3.3.3 Multidimensional Scaling

The 132 ancient human genomes generated in this project that were retained after filtering for contamination and missingness are plotted in **Figure 3.11**, together with the reference data consisting of non-African populations. In this worldwide MDS, major continental groupings are separated from each other. In particular, individuals with European, East Asian, and Native American genetic ancestries are differentiated on the first two MDS dimensions. Moreover, admixed Oceanian individuals are placed along an Oceanian-Europe cline, South Asian individuals placed along a southeast Asian-West Eurasian cline, western and central Siberians (including Uralic, Yukaghir, and Altaic speakers) placed along a European-East Asian cline, and northeastern Siberians (including Chukotko-Kamchatkan speakers) placed along an East Asian-

Native American cline. These clinal patterns reflect shared genetic ancestries and prehistoric/historic genetic admixture.

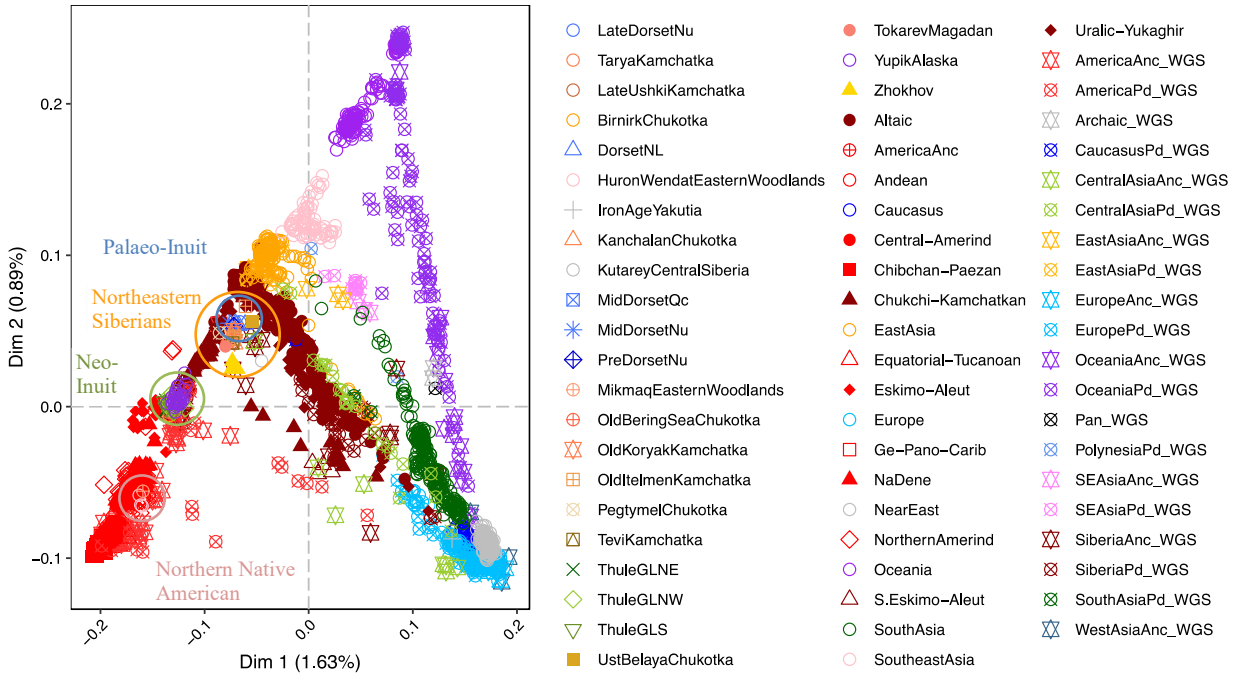



**Figure 3.11** – MDS plot showing the 132 ancient human genomes generated in this project, plotted with the non-African reference data. From this project, the northeastern Siberian are indicated within the orange circle, the Palaeo-Inuit within the blue circle, the Neo-Inuit within the green circle, and the Northern Native Americans within the light red circle.

Ancient individuals from Central Siberia (Kutarey) and the New Siberian Islanders (Zhokhov) fall in between the Altaic + Uralic – Yukaghir and Chukchi-Kamchatkan clines. Ancient individuals from northeastern Siberia from Magadan (Tokarev), Kamchatka (Tevi, Tarya, Old Itelmen, and Old Koryak), Chukotka (Ust-Belaya, Pegtymel, Kanchalan), and Palaeo-Inuit populations from northeastern Canada (Pre-Dorset, Dorset) were placed on the East Asian side of the northeastern Siberian cline, falling close to ancient Siberian individuals such as, Kolyma and Zhokhov. In contrast, the Neo-Inuit populations from Chukotka (Birnirk, Old Bering Sea), Alaska (Yupik), and Greenland (Thule), fall at the opposite end of the Chukchi-Kamchatkan cline, close to speakers of *Inuit*-Aleut languages. The Northern Native American (NNA) populations from the Eastern Woodlands (Huron-Wendat and Mi'gmaq) fall together with other NNA.

The ancient Arctic individuals fall roughly into four locations of the MDS plot. The Zhokhov individuals, together with the Kolyma individual, fall on the Siberian end of the Chukotko-Kamchatkan cline but are slightly pulled towards Native Americans. Other ancient northeastern Siberians fall predominantly on the Siberian end of the Chukotko-Kamchatkan cline. Palaeo-Inuit populations fall adjacent to the Chukchi-Kamchatkan cline. Ancient Neo-Inuit individuals fall between the Siberian and North American speakers of the *Inuit*-Aleut language family, and the populations from the Eastern Woodlands fall within the Native American cluster (**Figure 3.12**).

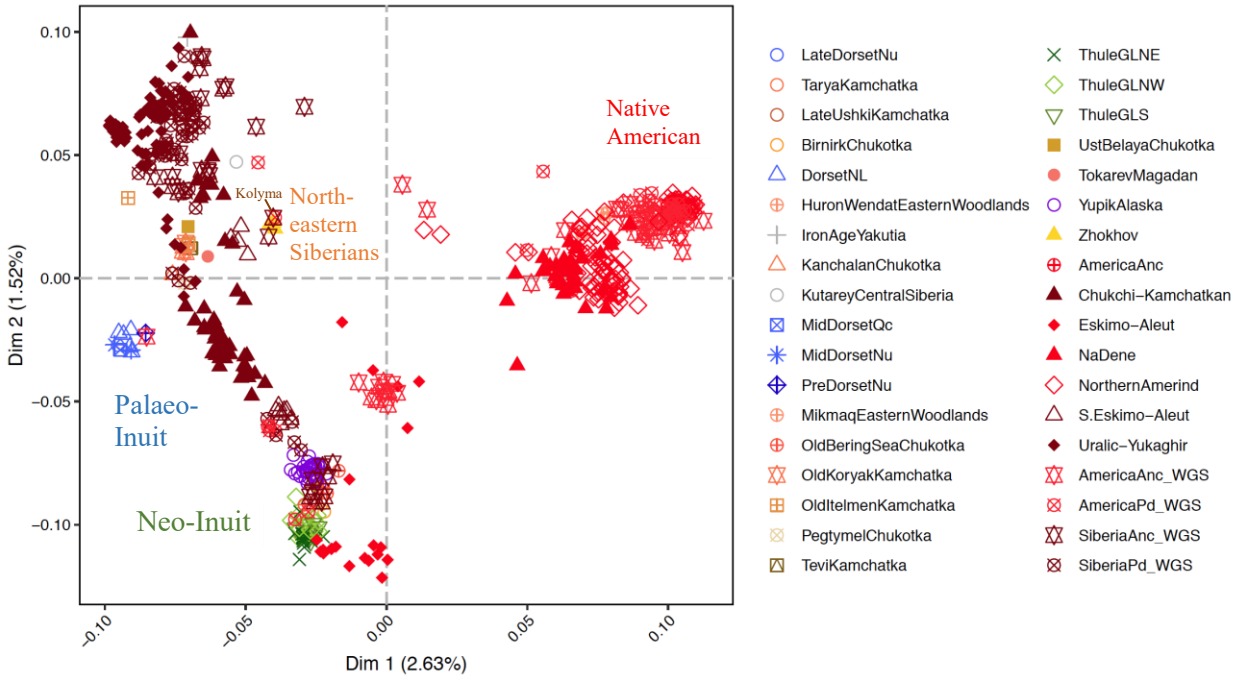



**Figure 3.12 –** MDS plot showing the 132 ancient Arctic and sub-Arctic individuals plotted with the reference data, where only the central and eastern Siberian as well as Native American reference populations are included. In this plot, Chukotko-Kamchatkan speakers fall in an intermediate position between western and central Siberians and Neo-Inuit individuals, whereas Native Americans are differentiated from Siberians on the first MDS dimension.

To further examine the broad genetic affinities of ancient northeastern Siberian individuals, all Neo-Inuit, Siberian, and Indigenous American individuals, as well as the Native American reference populations were excluded in **Figure 3.13**.

In **Figure 3.13**, eastern and central Siberians are differentiated from Chukotko-Kamchatkan speakers on the first dimension of the MDS. Whereas most ancient northeastern Siberians from this study were placed adjacent to Chukotko-Kamchatkan speakers, the Zhokhov individuals' group with Kolyma closer to the latter, thus suggesting that this group is genetically more similar to present-day and ancient Chukchi-Kamchatkans than to ancient and present-day Uralic and Yukaghir-speakers.

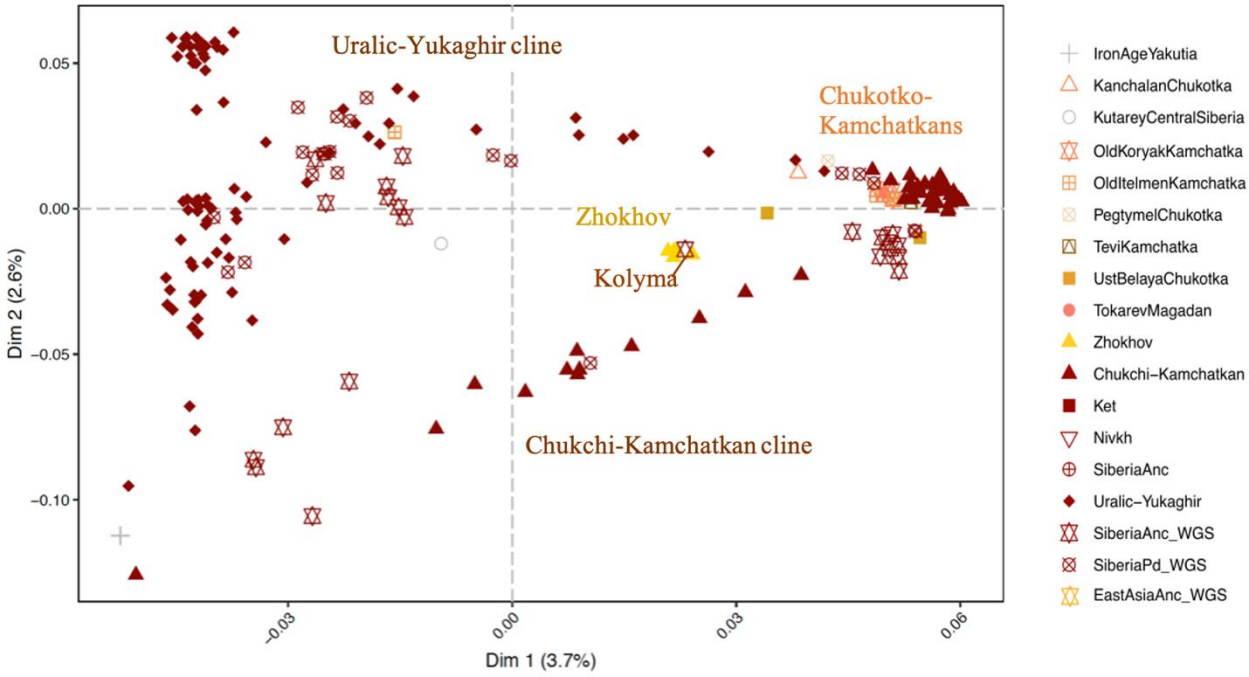



**Figure 3.13** - MDS plot showing the ancient Siberian genomes generated in this project, plotted with the central and eastern Siberian reference data. The Chukotko-Kamchatkan and Zhokhov groupings are visible, with the Zhokhov group falling with the ancient Kolyma individual.

To examine the Palaeo- and Neo-Inuit population groups, all populations besides speakers of the *Inuit*-Aleut and Na Dene languages were excluded, and the results are shown in **Figure 3.14**. In this transformation, Inuit, Native American, and Palaeo-Inuit are separated by the first two MDS dimensions. North American Inuit speakers are placed along a cline that extends from Greenlandic Inuit to ancient Siberians. More importantly, this plot exemplifies the difference between Palaeo-Inuit and non-Palaeo-Inuit populations, with the Palaeo-Inuit being genetically more differentiated from the other populations on the MDS. The Pre-Dorset individual falls next to the Saqqaq individual, which is consistent with their spatiotemporal archaeological context and with a

previous study (2). The Dorset, from various regions of Eastern Canada and the High Arctic, all group together as shown previously (2), very distantly from others on the plot, with the Pre-Dorset and Saqqaq individuals falling intermediately. The Dorset from Newfoundland, who are archaeologically and geographically distinct from other Dorset populations, plot separately from the other Dorset populations.

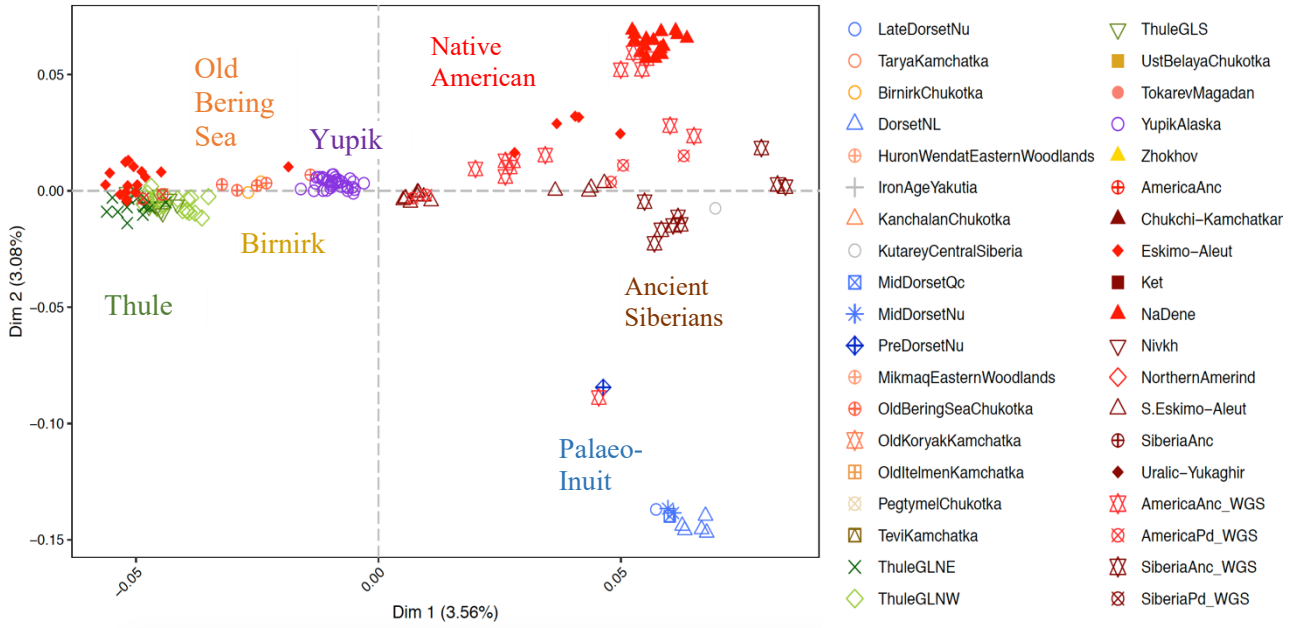



**Figure 3.14** - MDS plot showing the ancient Palaeo-Inuit and Neo-Inuit genomes generated in this project, plotted with some Siberian and Native American reference data. The Palaeo-Inuit grouping, as well as the Neo-Inuit subgroupings (Thule, Old Bering Sea, Birnirk, and Yupik) are visible.

The Neo-Inuit groups fall along the Inuit-Aleut cline, with the Alaskan Yupik and the Greenlandic Thule forming separate groupings with the Siberian Neo-Inuit populations (*i.e.*, the Old Bering Sea and Birnirk from Chukotka) plotting intermediately. Interestingly, the Old Bering Sea individuals form a time-dependant gradient as they span between the Alaskan Yupik and the Greenlandic Thule, with the oldest Old Bering Sea individual (~2230 BP) plotting closest to the Yupik and the youngest Old Bering Sea individual (~1470 BP) plotting with the Thule. Similarly, the Birnirk individuals plot consistently with this time gradient, with the ~1816 BP individual plotting closer to the Alaskan Yupik and the ~1726 BP individual plotting closer to the Greenlandic Thule. Also worth noting is the population stratification within the Greenlandic Thule, where the

western Greenland Thule are closer to the Chukotka Neo-Inuit populations than Thule from east Greenland.

### 3.3.4 $f_3$ & *D*-Statistics

Based on archaeological labels, MDS findings, and the individually run outgroup $f_3$-statistics, some individuals were grouped for further outgroup $f_3$- and *D*-statistical analyses to increase the analytical power. The groupings that deviate from the cultural labels are the Old Bering Sea Culture split into three, the Old Itelman Culture split into two, and the Ust-Belaya Culture split into two. These groupings are consistent with the C14 ages of the individuals and are numbered from oldest to youngest (*i.e.*, OldBeringSeaCulture1 is older than OldBeringSeaCulture3).

**Figure 3.15** - **Figure 3.35** which were generated for the outgroup $f_3$- and *D*-statistical analyses only contain a subset of the reference panel, where two populations were selected from each broad geographic region. The excluded populations were removed as they are not the focus of this project (*i.e.*, Africa, Europe, Near East, South Asia, Southeast Asia, Central Asia). Meanwhile, selected populations from East Asia, Siberia, and the Americas of known relevance were included.

## Zhokhov Outgroup $f_3$-Statistic

The outgroup $f_3$-statistic for the four pooled Zhokhov individuals was performed to confirm their genetic affinity to Kolyma, as suggested by their consistent proximity on the MDS plots. This result shown in **Figure 3.15** indicate that, among the populations in the reference dataset, the Zhokhov group shares the most genetic drift with Kolyma. Following Kolyma, the Zhokhov population is genetically most similar to many of the ancient populations in Magadan and the Kamchatka Peninsula, Indigenous groups in Northern North America (speakers of Na Dene languages), and the Neo-Inuit and Palaeo-Inuit groups (speakers of *Inuit*-Aleut languages).


**Figure 3.15** - The outgroup $f_3$-statistic for the Zhokhov individuals showing that they share the most genetic drift with the ancient Kolyma individual, when compared to the reference dataset.

## Palaeo-Inuit Outgroup $f_3$-Statistic

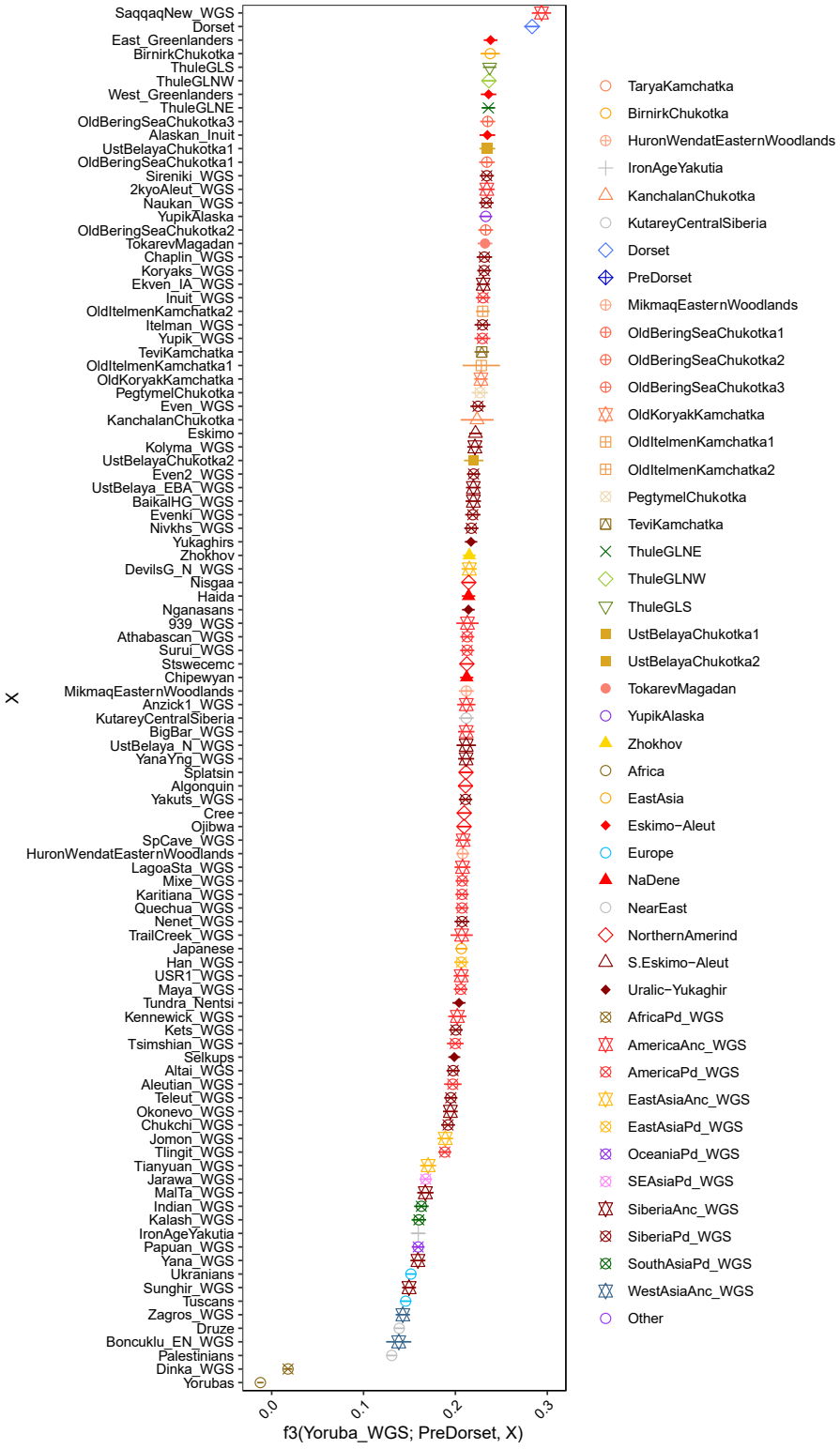



**Figure 3.16** - The outgroup $f_3$-statistic for the Pre-Dorset individual showing that they share the most genetic drift with the ancient Saqqaq individual from Greenland and the pooled Dorset individuals from Nunangat and Newfoundland, when compared to the reference dataset.

MDS results suggested that Palaeo-Inuit (Saqqaq, Dorset, Pre-Dorset) individuals grouped together but were substantially drifted from other Arctic and sub-Arctic populations. Outgroup $f_3$-statistics were used to explore their genetic affinities to each other, relative to other world-wide populations. Both the single Pre-Dorset individual from Nunavut and the nine pooled Dorset individuals from Newfoundland and the Canadian High Arctic are genetically more similar to Saqqaq than to any other worldwide population included in this analysis (**Figure 3.16** and **Figure 3.17**, respectively). Closely following the Palaeo-

Inuit grouping, the Neo-Inuit populations of both Chukotka and North America have the most shared drift, followed by Palaeo-Siberian-related, ancient East Asian-related, and present-day Siberian individuals/groups, and then followed by the Indigenous populations of the Americas (both NNA and Southern Native American [SNA] groups).


**Figure 3.17** - The outgroup $f_3$-statistic for the pooled Dorset individuals showing that they share the most genetic drift with the ancient Saqqaq individual from Greenland and the Pre-Dorset individual from Nunavut, when compared to the reference dataset.

## Neo-Inuit Outgroup $f_3$ Statistic

Although Neo-Inuit individuals were placed close to each other in the MDS analyses, they formed a cline that roughly followed their age and sampling locations. Outgroup $f_3$-statistics were used to explore their genetic affinities to each other, and to other world-wide populations.


**Figure 3.18** - The outgroup $f_3$-statistic for the pooled Old Bering Sea 1 individuals showing that, when compared to the reference dataset, they are most genetically similar to present-day Greenlandic and Alaskan Inuit groups, other Old Bering Sea groups, and the ancient Greenlandic Thule.

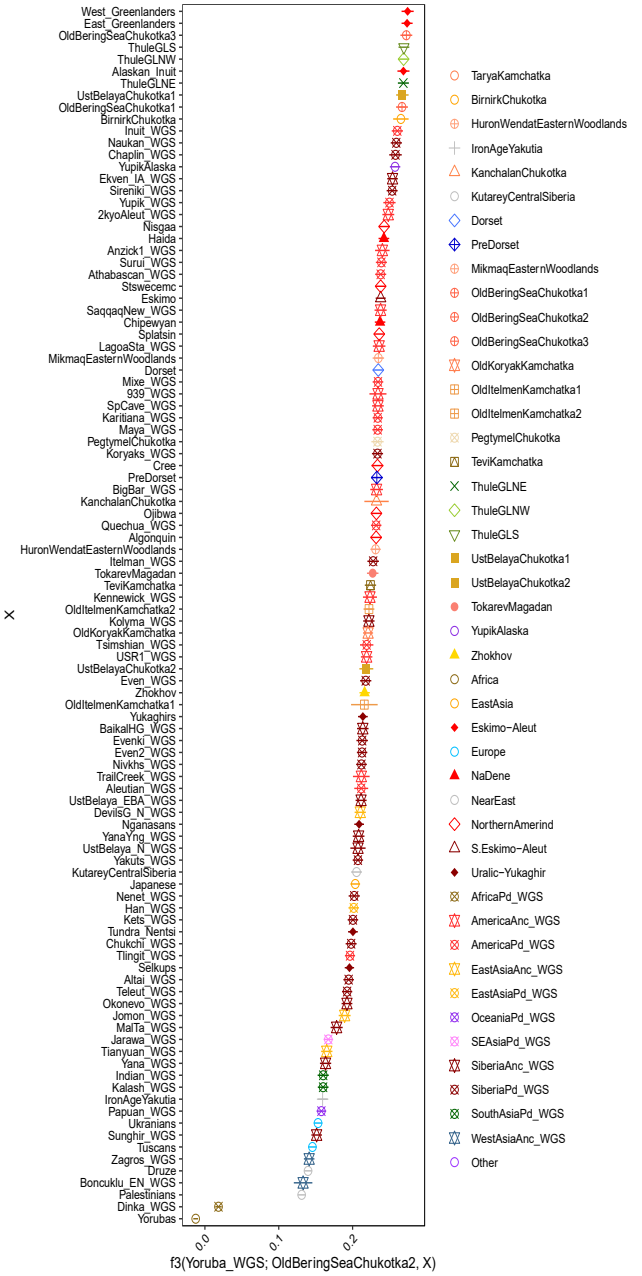



**Figure 3.19** - The outgroup $f_3$-statistic for the pooled Old Bering Sea 2 individuals showing that, when compared to the reference dataset, they are most genetically similar to present-day Greenlandic and Alaskan Inuit groups, other Old Bering Sea groups, and the ancient Greenlandic Thule.


**Figure 3.20** - The outgroup $f_3$-statistic for the pooled Old Bering Sea 3 individuals showing that, when compared to the reference dataset, they are most genetically similar to present-day Greenlandic and Alaskan Inuit groups, other Old Bering Sea groups, and the ancient Greenlandic Thule.

The outgroup $f_3$-statistic of the three Old Bering Sea pools shows most genetic similarity to present-day Greenlandic and Alaskan Inuit groups, other Old Bering Sea groups, and the ancient Greenlandic Thule (**Figure 3.18, Figure 3.19, Figure 3.20**). Following those, the genetically most similar are present-day Canadian Inuit, present-day Indigenous Siberian populations, and the ancient Alaskan Yupik. The NNA and SNA populations show a slightly stronger genetic affinity to the Old Bering Sea groups than the Palaeo-Inuit populations.

The outgroup $f_3$-statistic for the Birnirk population in Chukotka shows the most genetic similarity to the oldest and youngest Old Bering Sea groups followed by present-day Greenlandic and Alaskan Inuit groups, and the ancient Greenlandic Thule (**Figure 3.21**). Following those, the genetically most similar are the

ancient Alaskan Yupik, the present-day Canadian Inuit, and present-day Indigenous Siberian populations. The NNA and SNA populations show similar genetic affinity to the Birnirk as the

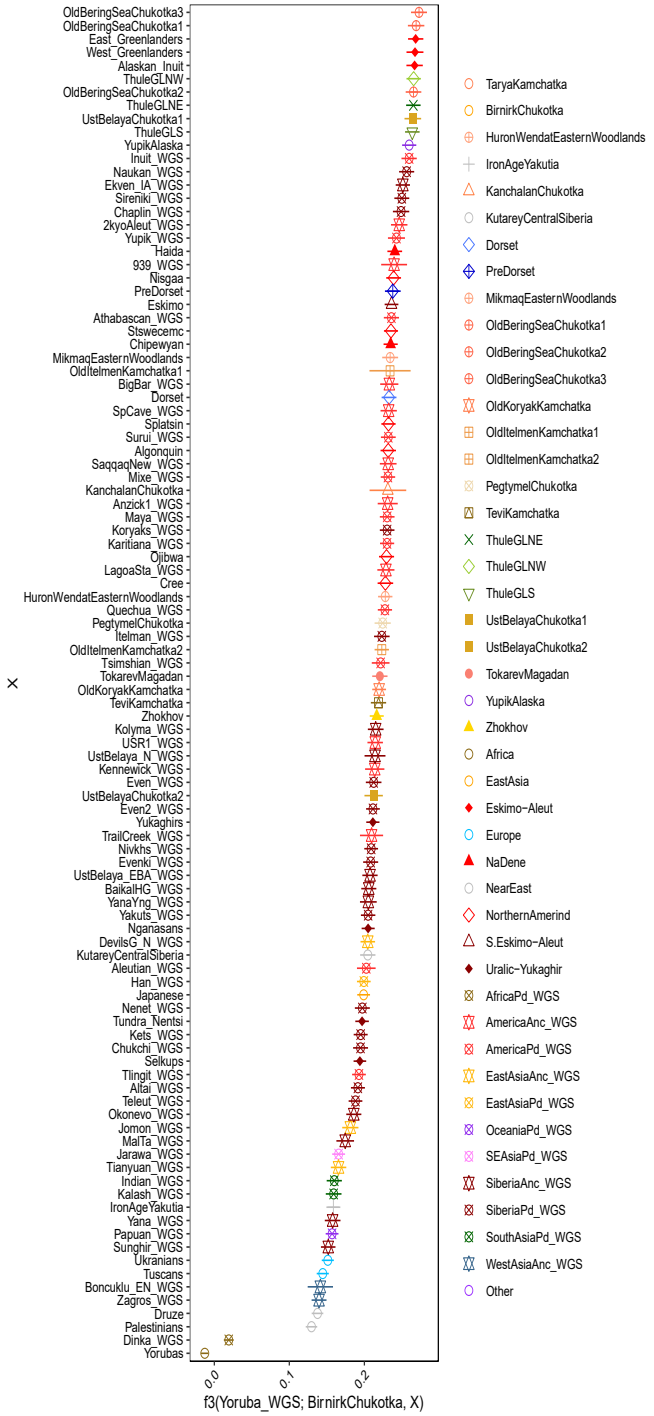



**Figure 3.21** - The outgroup $f_3$-statistic for the Birnirk population in Chukotka shows that, when compared to the reference dataset, they are the most genetically similar to the oldest and youngest Old Bering Sea groups followed by present-day Greenlandic and Alaskan Inuit groups, and the ancient Greenlandic Thule.

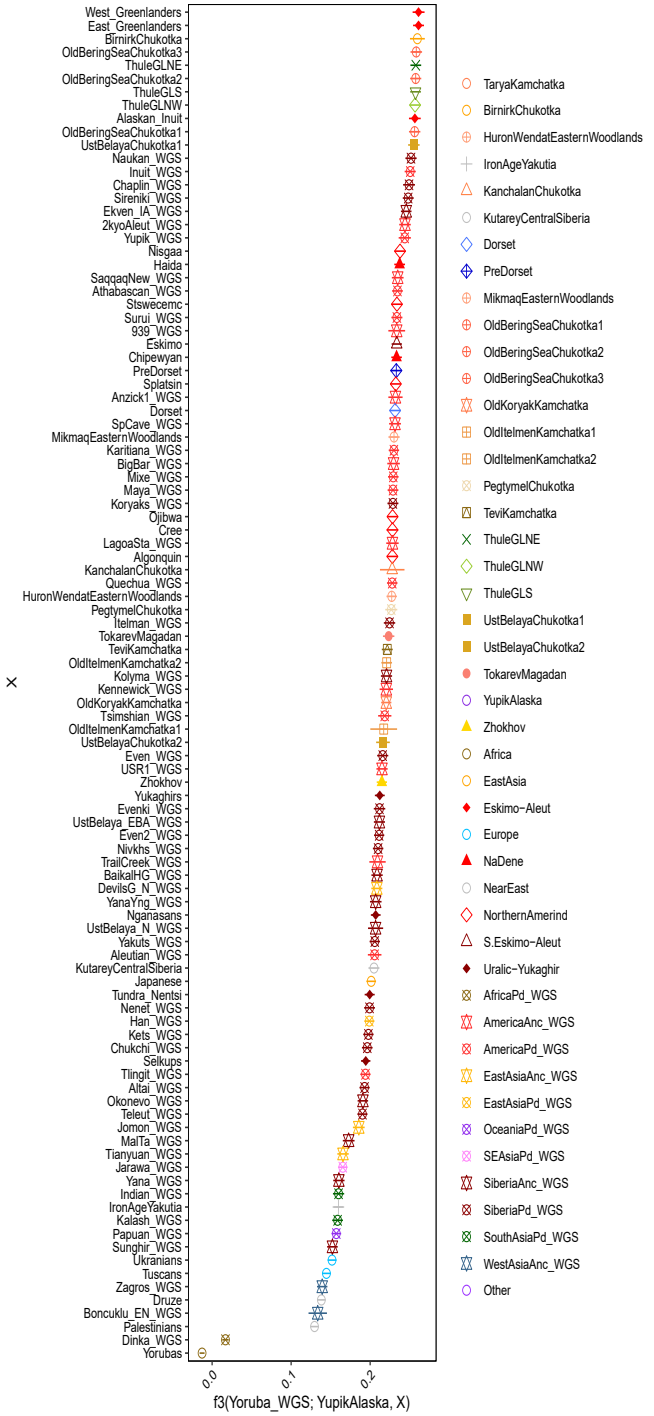



**Figure 3.22** - The outgroup $f_3$-statistic for the Alaskan Yupik shows that, when compared to the reference dataset, they are the most genetically similar to present-day Greenlandic Inuit groups, the Birnirk and Old Bering Sea Populations from Chukotka, and the ancient Greenlandic Thule.

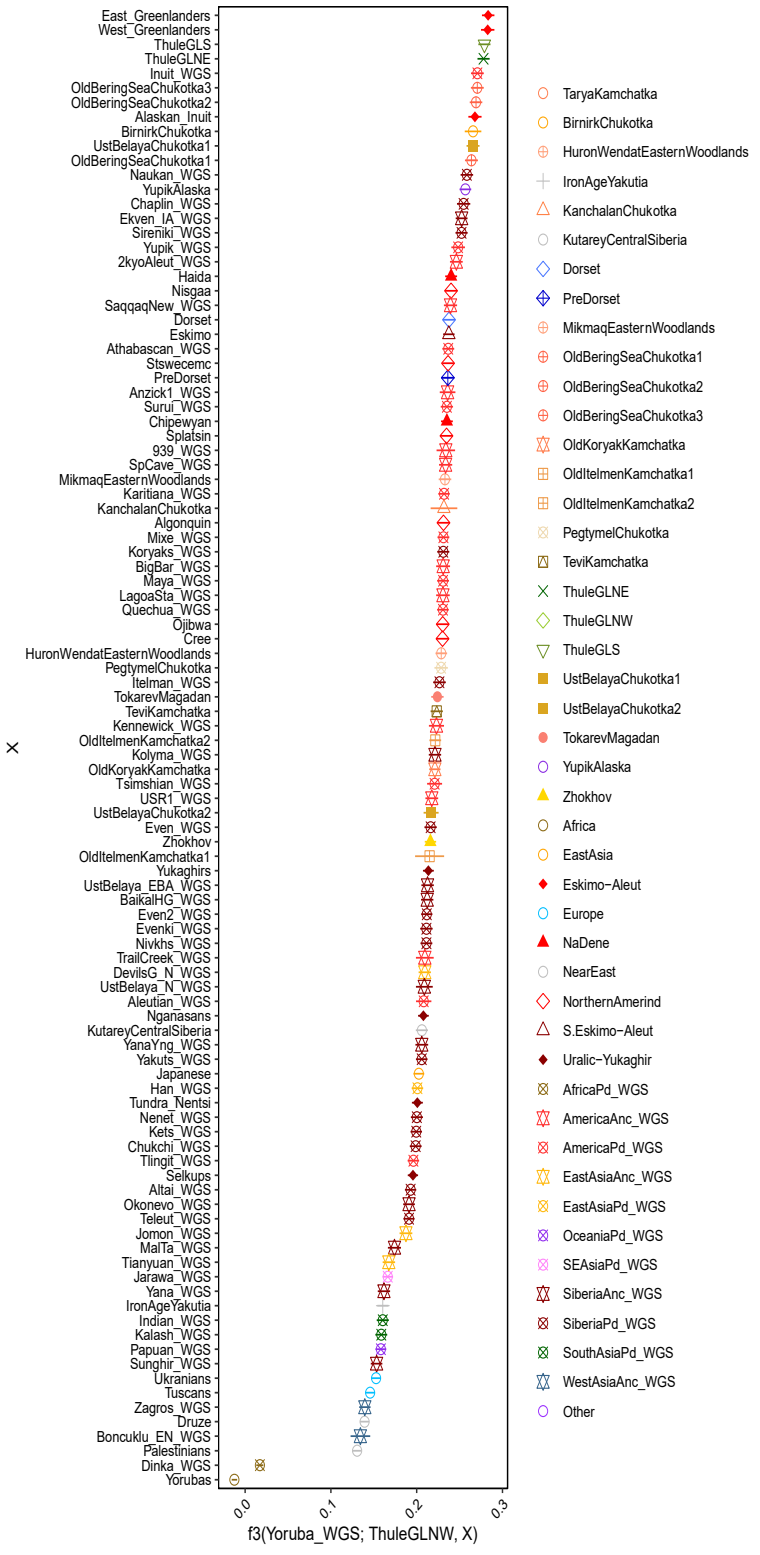



**Figure 3.23** - The outgroup $f_3$-statistics for the pooled western Greenlandic Thule shows that, when compared to the reference dataset, they are the most genetically similar to the other ancient Greenlandic Thule (from the south and east) as well as to present-day Greenlandic Inuit groups.

Palaeo-Inuit populations.

The outgroup $f_3$-statistic for the Alaskan Yupik shows the most genetic similarity to present-day Greenlandic Inuit groups, the Birnirk and Old Bering Sea Populations from Chukotka, and the ancient Greenlandic Thule (**Figure 3.22**). Following those, the genetically most similar are the ancient and present-day Indigenous Siberian populations and the present-day Canadian Inuit. The NNA populations show similar genetic affinity to the Yupik as the Palaeo-Inuit populations.

The outgroup $f_3$-statistics for the three pools of Greenlandic Thule show the most genetic similarity to each other and to present-day Greenlandic Inuit groups (**Figure 3.23, Figure 3.24, Figure 3.25**). Following that, the strongest genetic similarity is to present-day Canadian Inuit and the younger Old Bering Sea populations. There is also genetic affinity to present-day Indigenous Siberian populations and the

Birnirk from Chukotka and the Alaskan Yupik. The Palaeo-Inuit populations seem to each have equal genetic similarity to the ancient Greenlandic Thule groups as the NNA populations.


**Figure 3.24** - The outgroup $f_3$-statistics for the pooled southern Greenlandic Thule shows that, when compared to the reference dataset, they are the most genetically similar to the other ancient Greenlandic Thule (from the east and west) as well as to present-day Greenlandic Inuit groups.


**Figure 3.25** - The outgroup $f_3$-statistics for the pooled eastern Greenlandic Thule shows that, when compared to the reference dataset, they are the most genetically similar to the other ancient Greenlandic Thule (from the south and west) as well as to present-day Greenlandic Inuit groups.

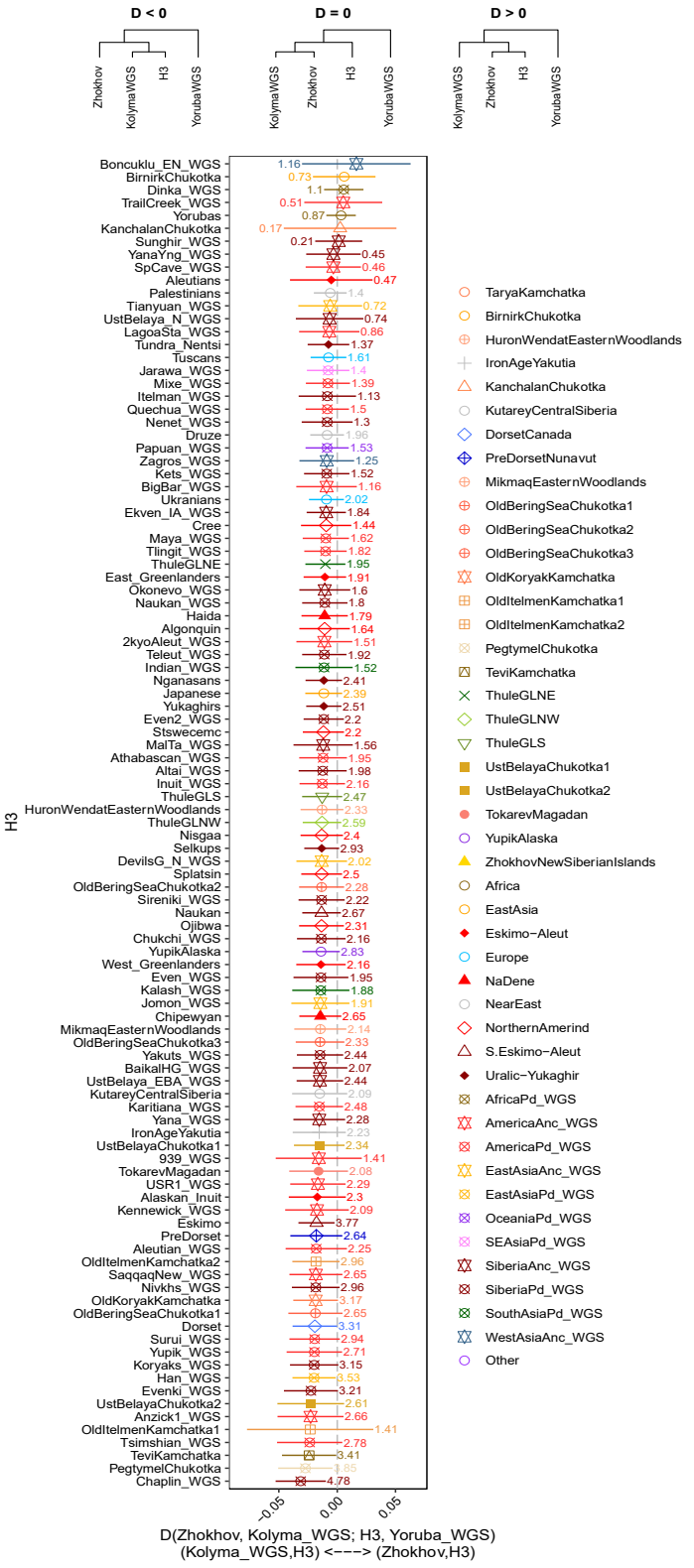



**Figure 3.26** - Results of a *D*-statistic of the form *D*(Zhokhov, Kolyma; H3, Yoruba) supporting a Kolyma+Zhokhov clade, where most of the populations in the reference panel do not have *D*-statistic values that differed significantly from 0 (|*Z*|<3).

## Kolyma Zhokhov *D*-Statistic

The Kolyma individual and the Zhokhov individuals consistently clustered together on MDS plots and showed strong genetic affinity in the outgroup $f_3$-statistics; therefore, a *D*-statistic of the form *D*(Zhokhov, Kolyma; H3, Yoruba) was performed to formally test if Zhokhov and Kolyma form a clade to the exclusion of all world-wide populations (H3) (**Figure 3.26**). For the majority of the populations in the reference panel, *D*-statistic values were not significantly different from 0 (|*Z*|<3), in support of a Kolyma+Zhokhov clade. Nevertheless, a number of Arctic and northeastern Siberian populations shared significantly more alleles with the Kolyma individual than with the Zhokhov group. Those groups are the present-day and ancient Koryaks, the Dorset, the Tevi, the Han Chinese, as well as the Siberian Yupik, the Pegtymel, and the Chaplin from the Chukchi Peninsula.

To confirm whether these populations form a clade with the Kolyma individual, *D*-statistics of the form *D*(H1, Kolyma; Zhokhov, Yoruba)

were computed. This resulted in *D*-statistics ranging between -0.078 and -0.146, with the Z-score ranging between -11.57 and -24.19 (**Table 6**). These results suggest that the Kolyma and Zhokhov individuals indeed form a clade with each other. Yet, it is likely that Kolyma is more closely related to the Palaeo-Siberian population that contributed to the ancestry of northeastern Siberian and American Arctic populations.

**Table 6** - *D*-statistics of the form *D*(H1, Kolyma; Zhokhov, Yoruba) were computed and confirmed that none of the populations listed in H1 form a clade with Kolyma to the exclusion of Zhokhov.

| H1pop | H2pop | H3pop | H4pop | D | SE | Z | nSNPs |
|---|---|---|---|---|---|---|---|
| Han | Kolyma | Zhokhov | Yoruba | -0.146 | 0.006 | -24.19 | 56280 |
| Eskimo | Kolyma | Zhokhov | Yoruba | -0.123 | 0.005 | -22.89 | 92818 |
| Evenki | Kolyma | Zhokhov | Yoruba | -0.123 | 0.007 | -17.93 | 35135 |
| Dorset | Kolyma | Zhokhov | Yoruba | -0.103 | 0.006 | -17.87 | 74041 |
| Chaplin | Kolyma | Zhokhov | Yoruba | -0.107 | 0.007 | -15.87 | 37103 |
| Old Koryak Kamchatka | Kolyma | Zhokhov | Yoruba | -0.086 | 0.006 | -14.15 | 64083 |
| Pegtymel Chukotka | Kolyma | Zhokhov | Yoruba | -0.103 | 0.007 | -14.05 | 28368 |
| Koryaks | Kolyma | Zhokhov | Yoruba | -0.089 | 0.007 | -13.38 | 49507 |
| Tevi Kamchatka | Kolyma | Zhokhov | Yoruba | -0.078 | 0.007 | -11.57 | 37006 |

## Palaeo-Inuit *D*-Statistic

The Saqqaq, Pre-Dorset, and Dorset individuals consistently clustered together (and drifted from other groups) on MDS plots and showed strong genetic affinity to each other in the outgroup $f_3$-statistics; therefore, a *D*-statistic was performed to formally test if the Palaeo-Inuit form a clade with each other to the exclusion of all worldwide populations. In every case, the *D*-Statistic results (**Figure 3.27, Figure 3.28, Figure 3.29**) show that a Palaeo-Inuit clade that excludes all other populations in the reference dataset cannot be rejected.

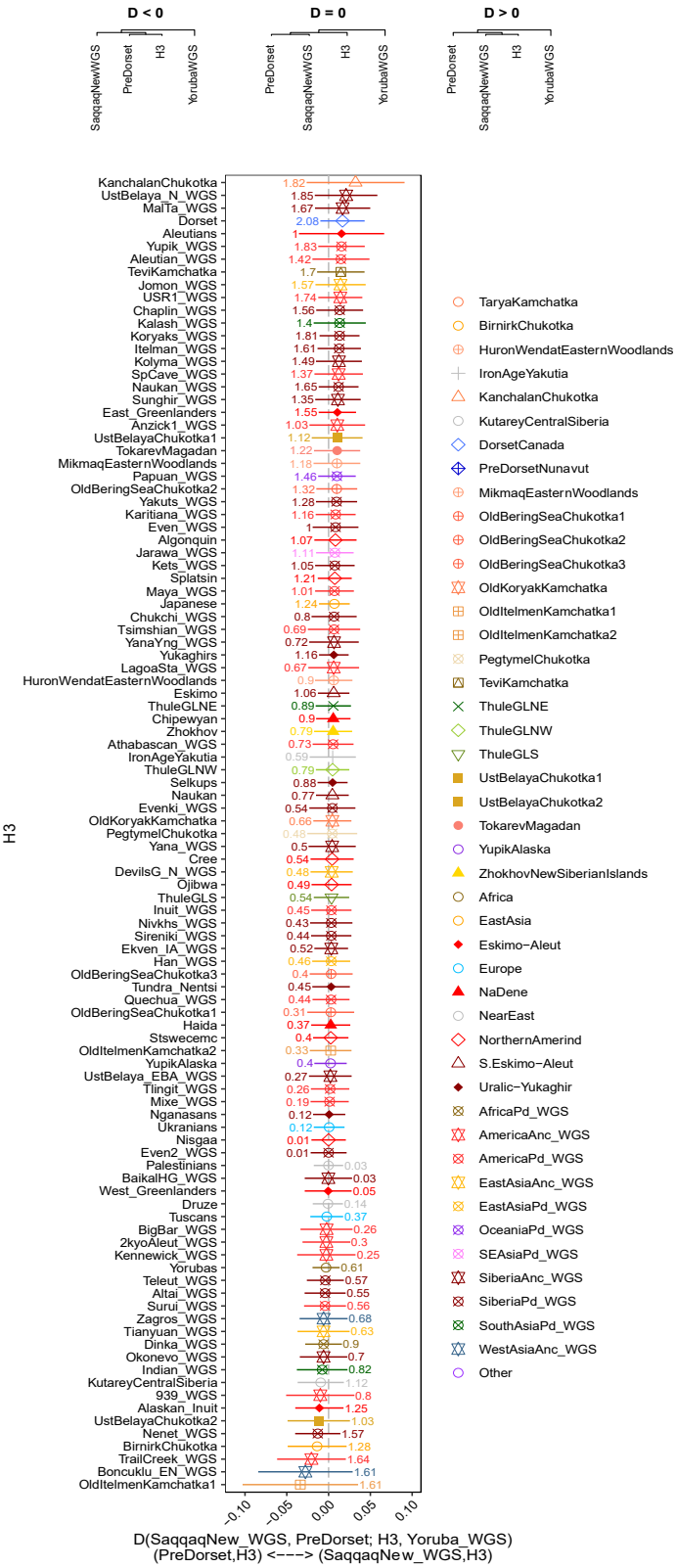



**Figure 3.27** - Results of a *D*-statistic of the form *D*(Saqqaq, Pre-Dorset; H3, Yoruba) supporting a Palaeo-Inuit clade, where no population in the reference panel have a *D*-statistic value that differed significantly from 0 (|Z|<3).

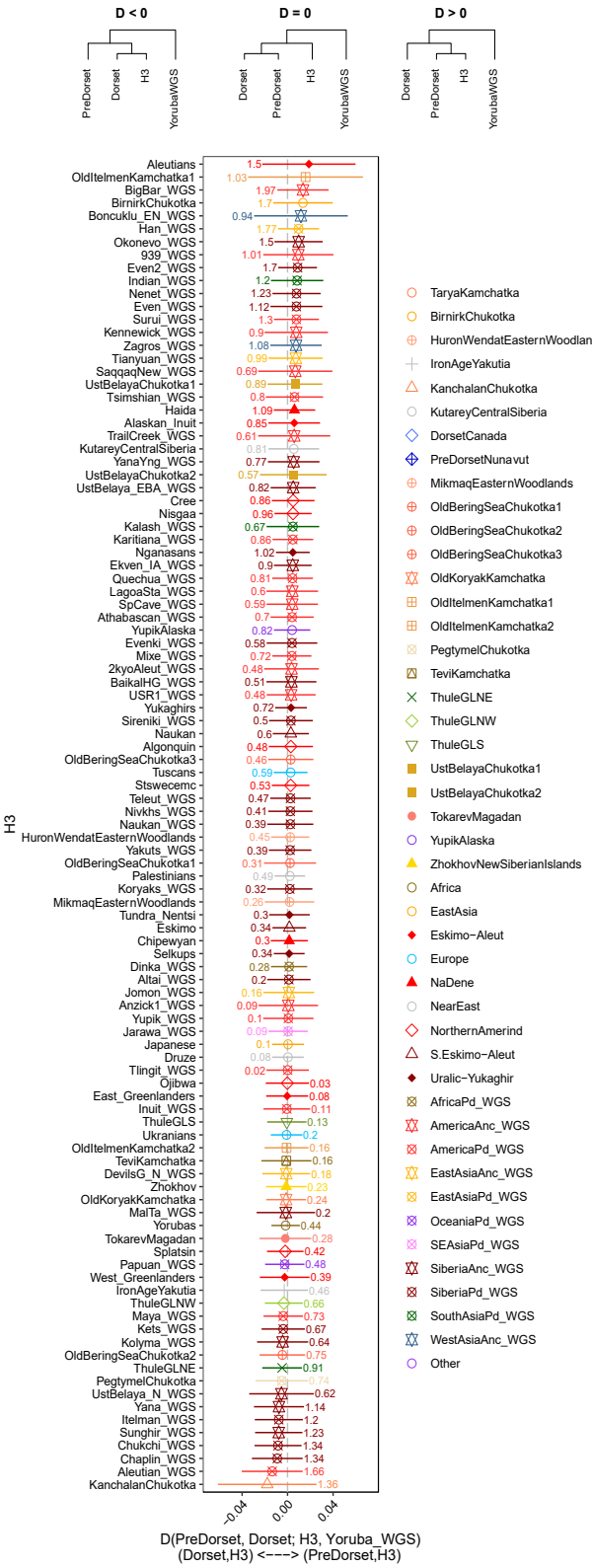



**Figure 3.28** - Results of a *D*-statistic of the form *D*(Pre-Dorset, Dorset; H3, Yoruba) supporting a Palaeo-Inuit clade, where no population in the reference panel have a *D*-statistic value that differed significantly from 0 (|*Z*|<3).

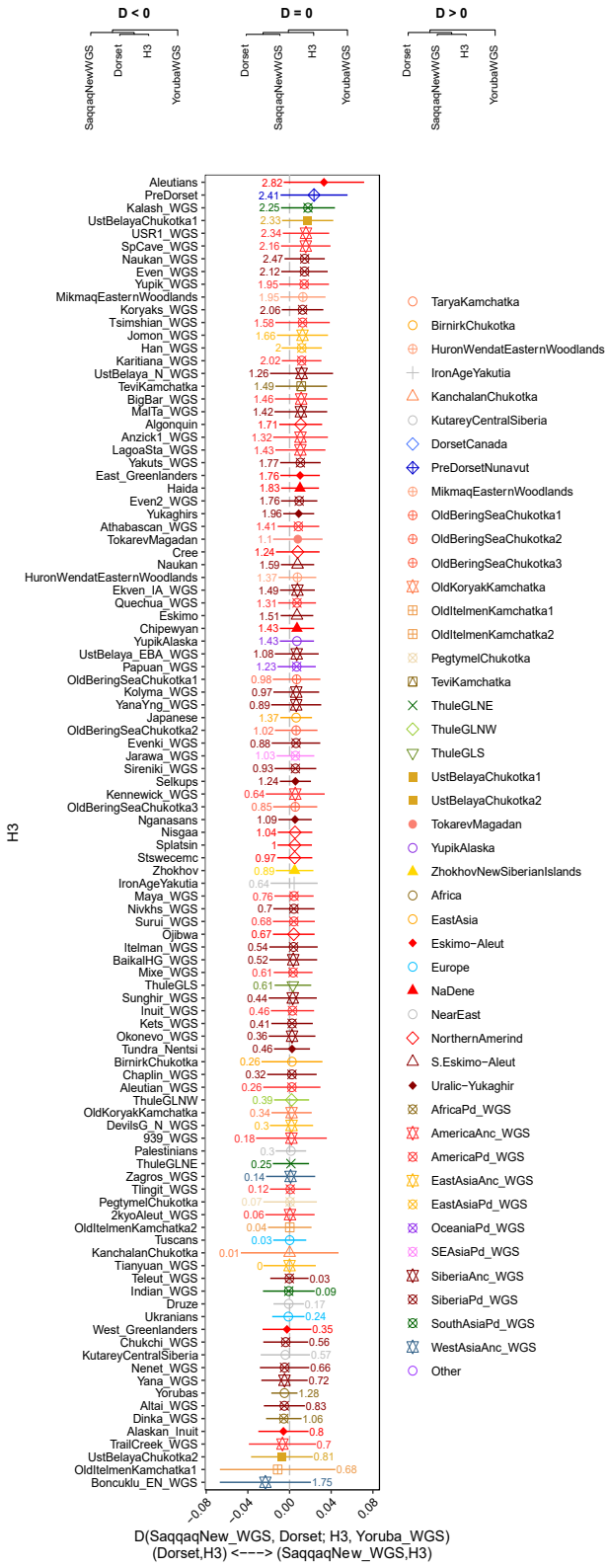



**Figure 3.29** - Results of a *D*-statistic of the form *D*(Saqqaq, Dorset; H3, Yoruba) supporting a Palaeo-Inuit clade, where no population in the reference panel have a *D*-statistic value that differed significantly from 0 (|*Z*|<3).

## Neo-Inuit *D*-Statistic

The Neo-Inuit individuals consistently clustered together on MDS plots, close to present-day Inuit groups. These findings were supported by genetic affinity between the Neo-Inuit group seen in the outgroup $f_3$-statistics results, as well as the differing levels of genetic similarity between Neo-Inuit and the Palaeo-Inuit and NNA populations.

Importantly, previous findings have suggested that Inuit populations derive from an admixture event between a populations related to Palaeo-Inuit and NNA ancestry (2,27). A *D*-statistic of the form *D*(Neo-Inuit, Athabascan; H3, Yoruba) was performed to further assess the context of the formation of the four Neo-Inuit groups. Here, Athabascans are a representative of North Native Americans and H3 represents the rest of the populations in the reference dataset. As expected, the *D*-statistic results show that most Native American populations have greater genetic

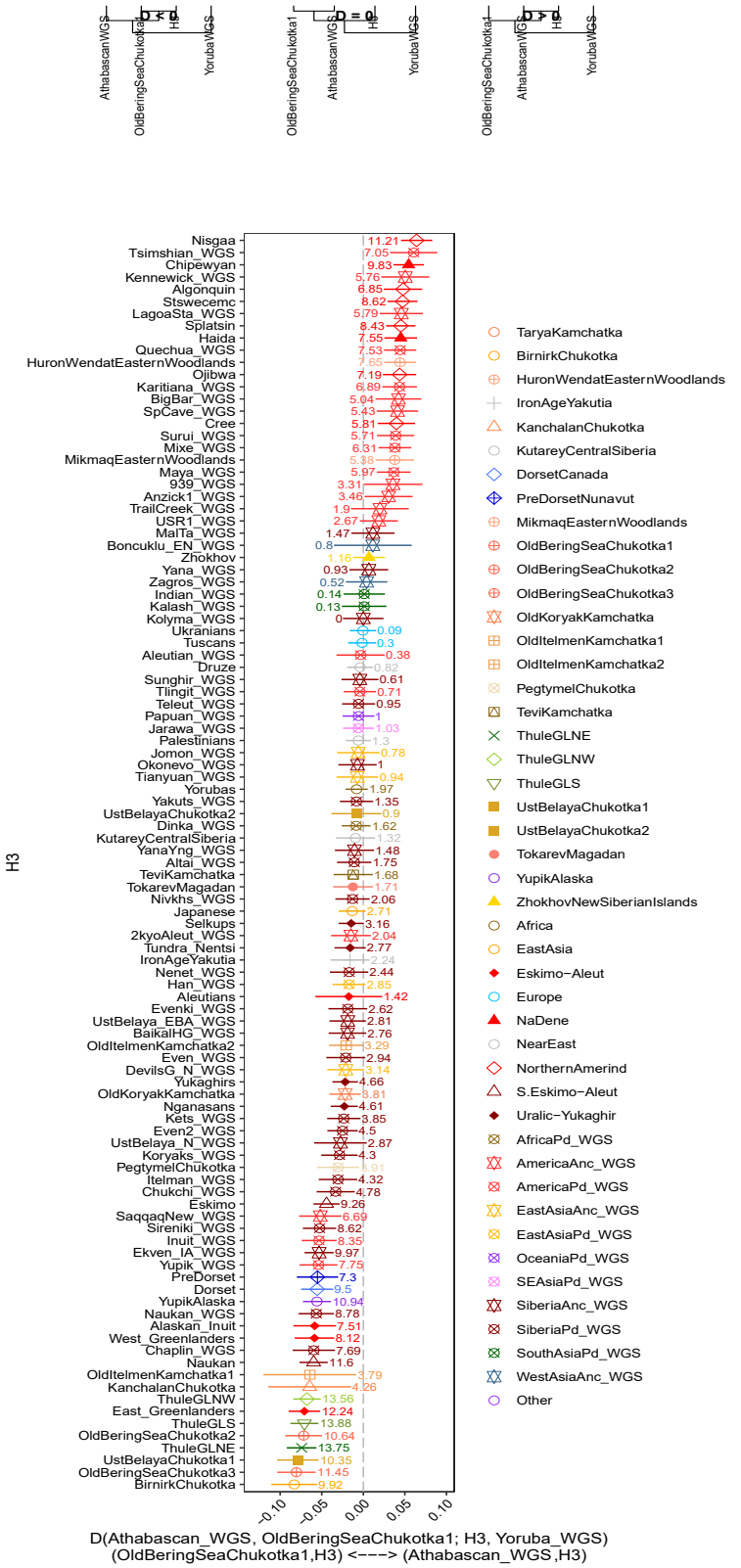



**Figure 3.30** - A *D*-statistic of the form *D*(OldBeringSeaChukotka1, Athabascan; H3, Yoruba) supporting the formation of the Neo-Inuit groups from an admixture event between a northeastern Siberian population and Northern Native Americans.

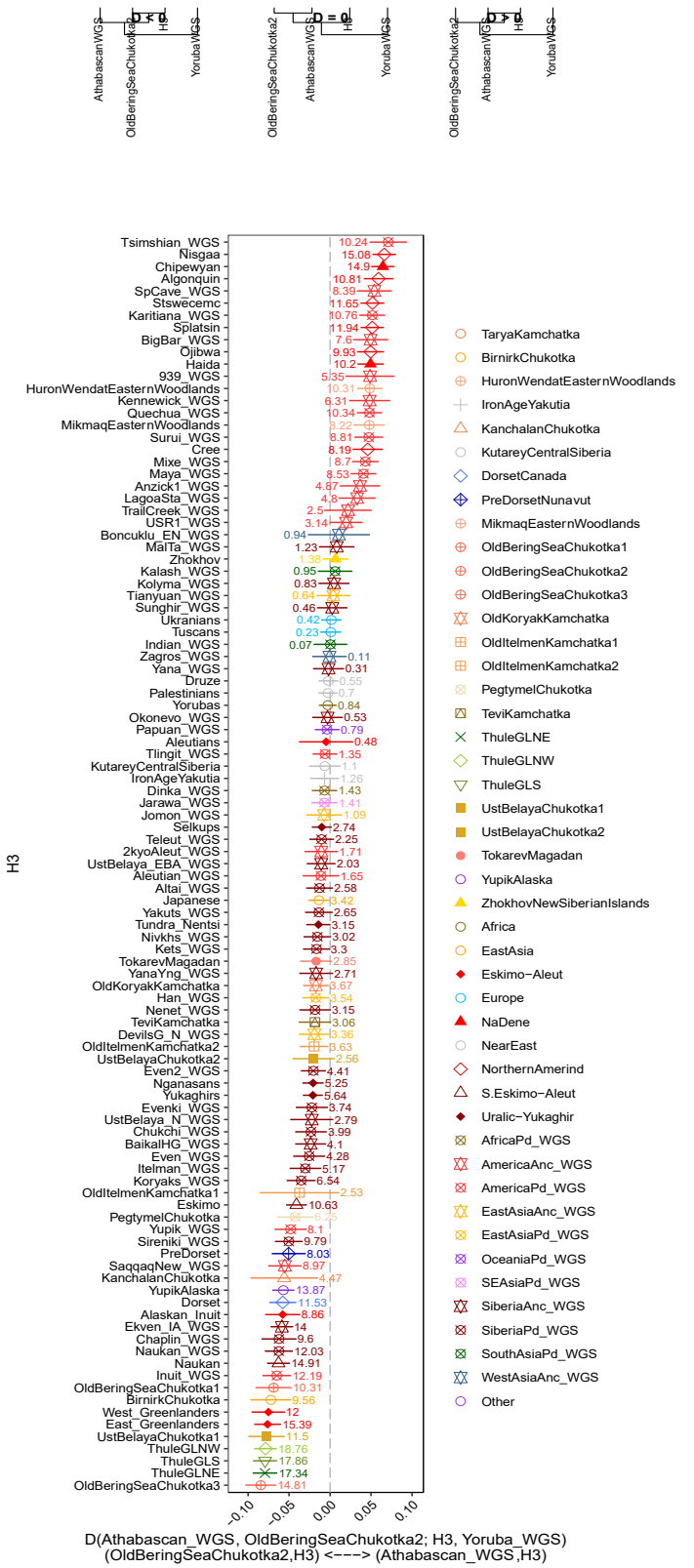



**Figure 3.31** - A *D*-statistic of the form *D*(OldBeringSeaChukotka2, Athabascan; H3, Yoruba) supporting the formation of the Neo-Inuit groups from an admixture event between a northeastern Siberian population and Northern Native Americans.

affinity to the Athabascans than to the Neo-Inuit. Furthermore, these results show that the Neo-Inuit share significantly more alleles with present-day Indigenous Siberians, Inuit, Neo-Inuit, and Palaeo-Inuit groups than Athabascans do with these latter populations. These results support that Neo-Inuit derive from an admixture event between a northeastern Siberian population and Northern Native Americans. Unfortunately, this analysis does not provide enough resolution to identify the population that is most closely related to the northeastern Siberian source (**Figure 3.30 - Figure 3.37**).

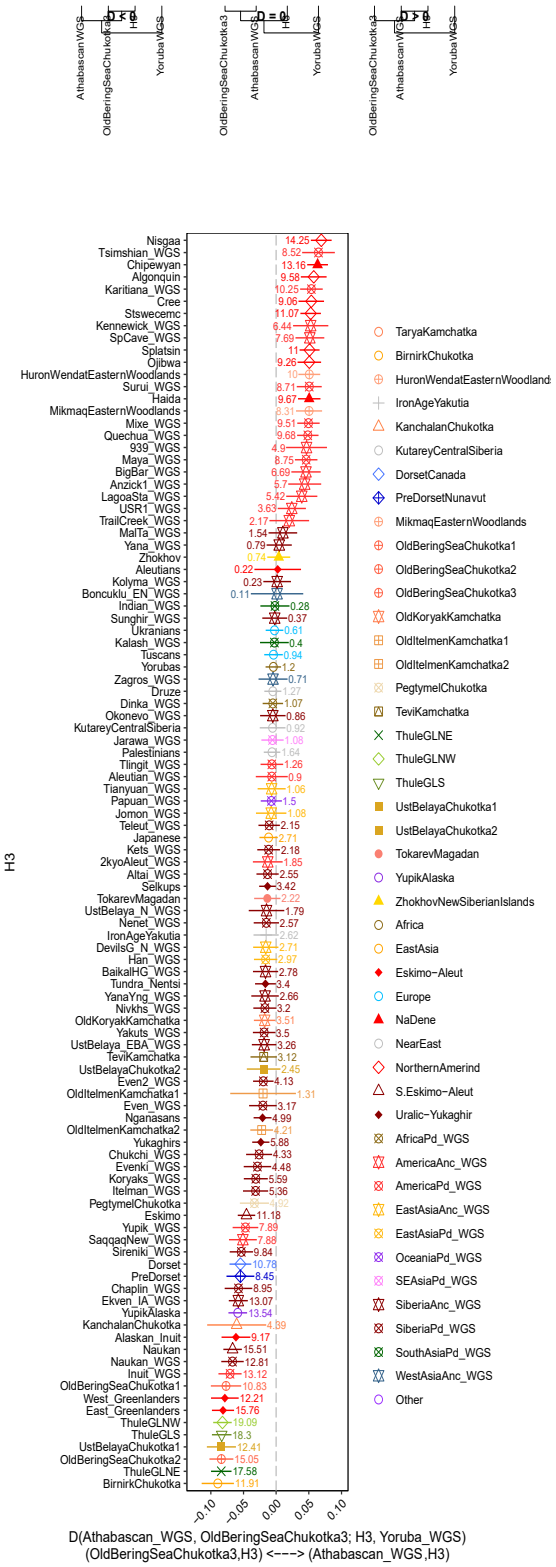



**Figure 3.32** - A *D*-statistic of the form *D*(OldBeringSeaChukotka3, Athabascan; H3, Yoruba) supporting the formation of the Neo-Inuit groups from an admixture event between a northeastern Siberian population and Northern Native Americans.

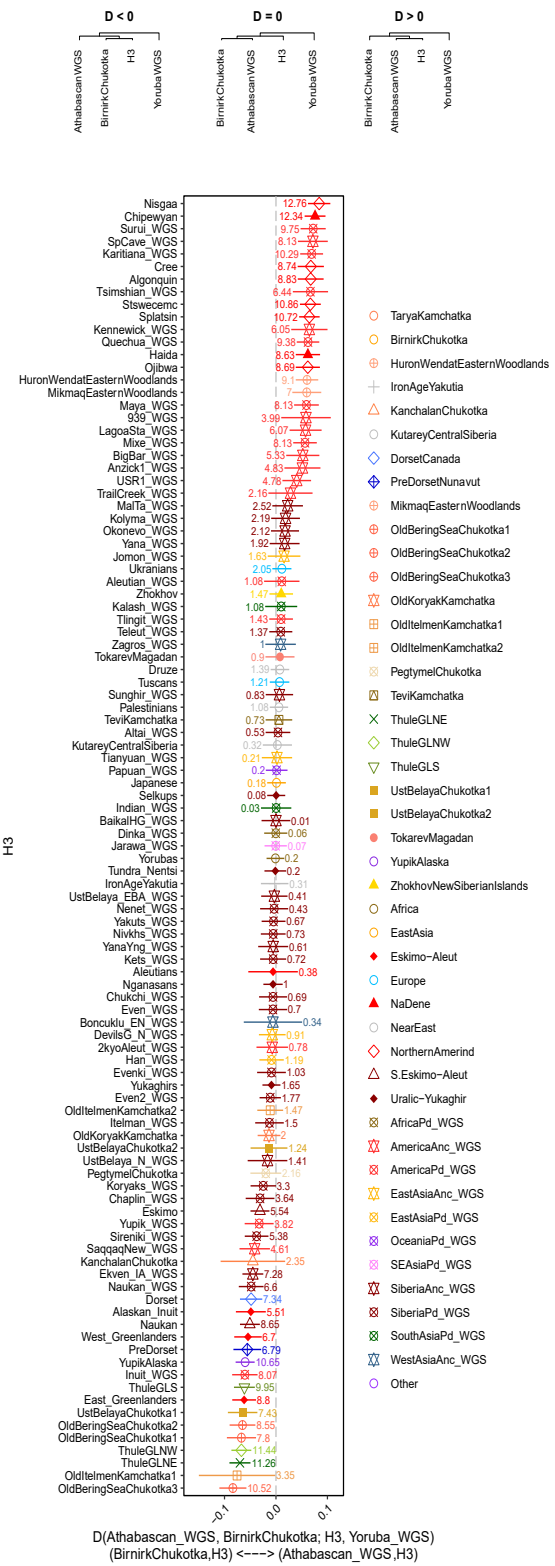



**Figure 3.33** - A *D*-statistic of the form *D*(Birnirk, Athabascan; H3, Yoruba) supporting the formation of the Neo-Inuit groups from an admixture event between a northeastern Siberian population and Northern Native Americans.

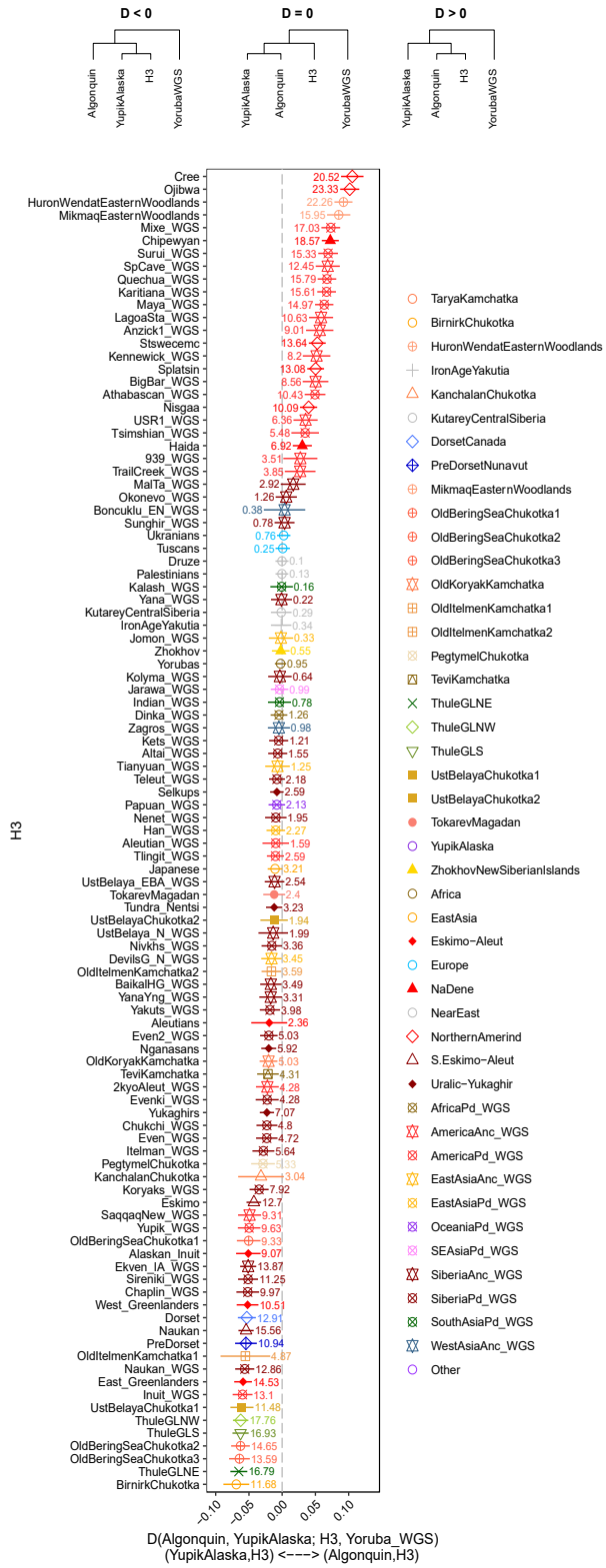



**Figure 3.34** - A *D*-statistic of the form *D*(YupikAlaska, Athabascan; H3, Yoruba) supporting the formation of the Neo-Inuit groups from an admixture event between a northeastern Siberian population and Northern Native Americans.

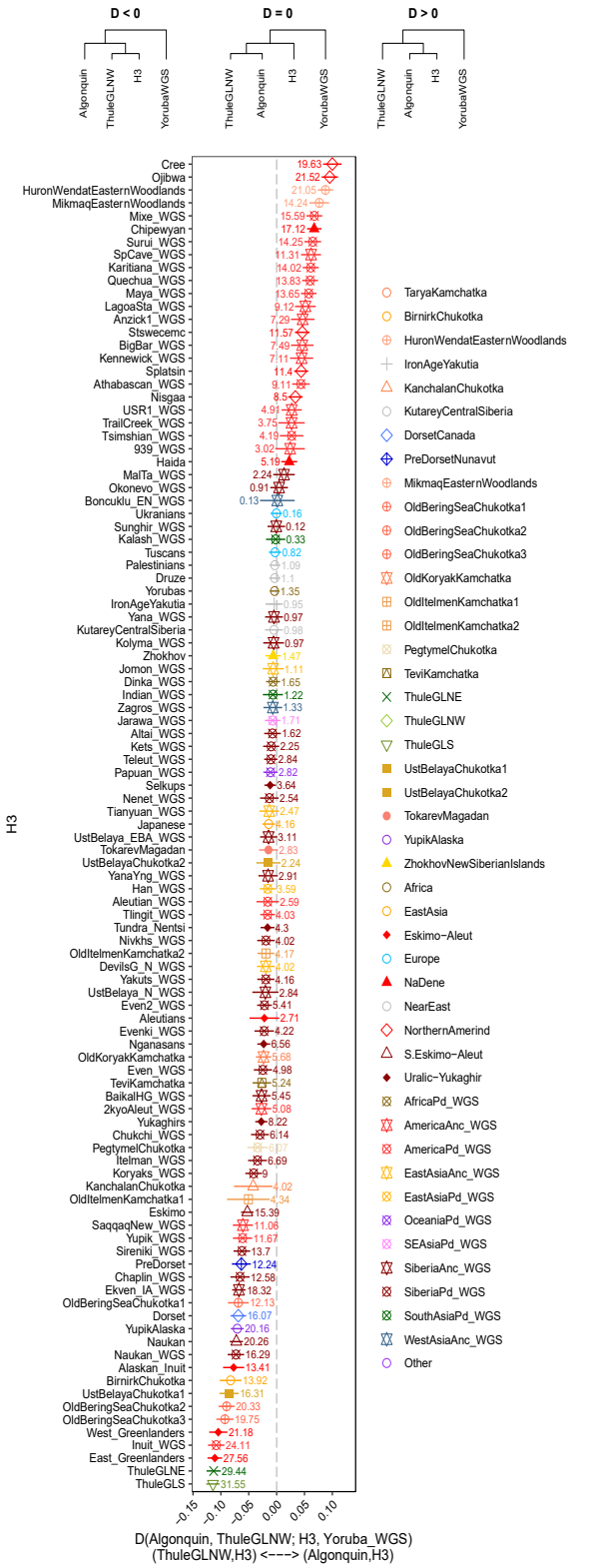



**Figure 3.35** - A *D*-statistic of the form *D*(ThuleGLNW, Athabascan; H3, Yoruba) supporting the formation of the Neo-Inuit groups from an admixture event between a northeastern Siberian population and Northern Native Americans.

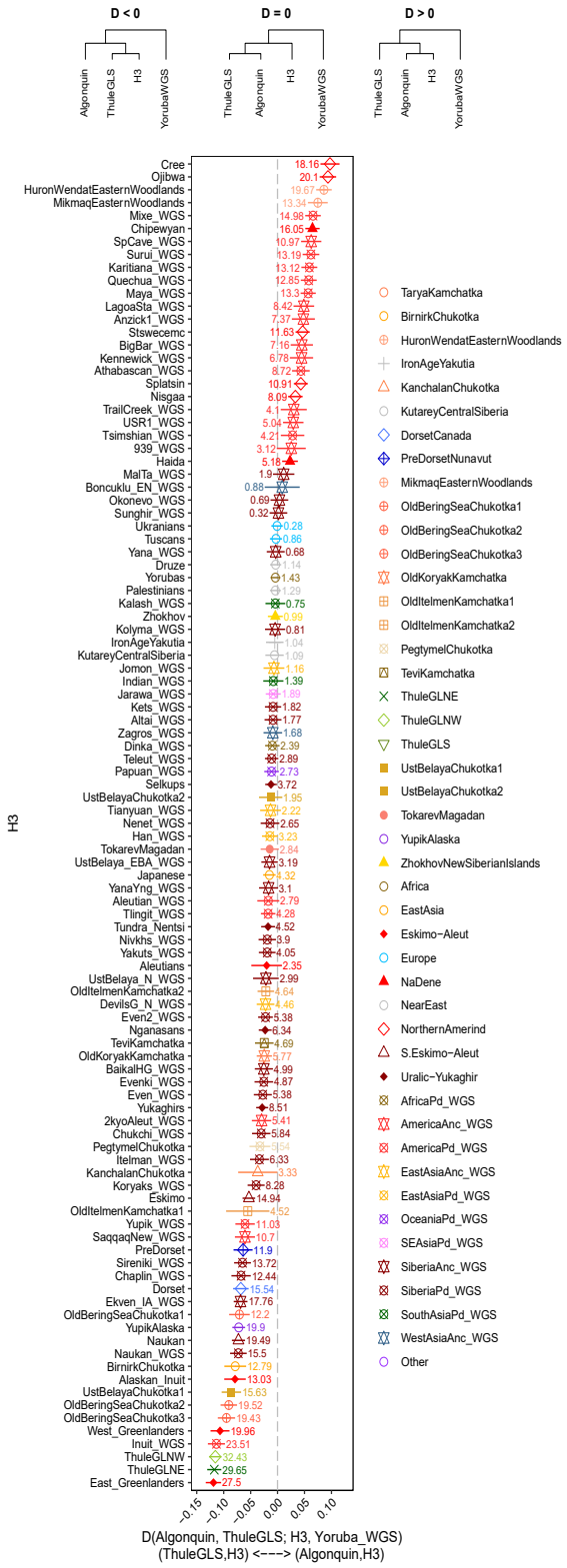



**Figure 3.36** - A *D*-statistic of the form *D*(ThuleGLS, Athabascan; H3, Yoruba) supporting the formation of the Neo-Inuit groups from an admixture event between a northeastern Siberian population and Northern Native Americans.

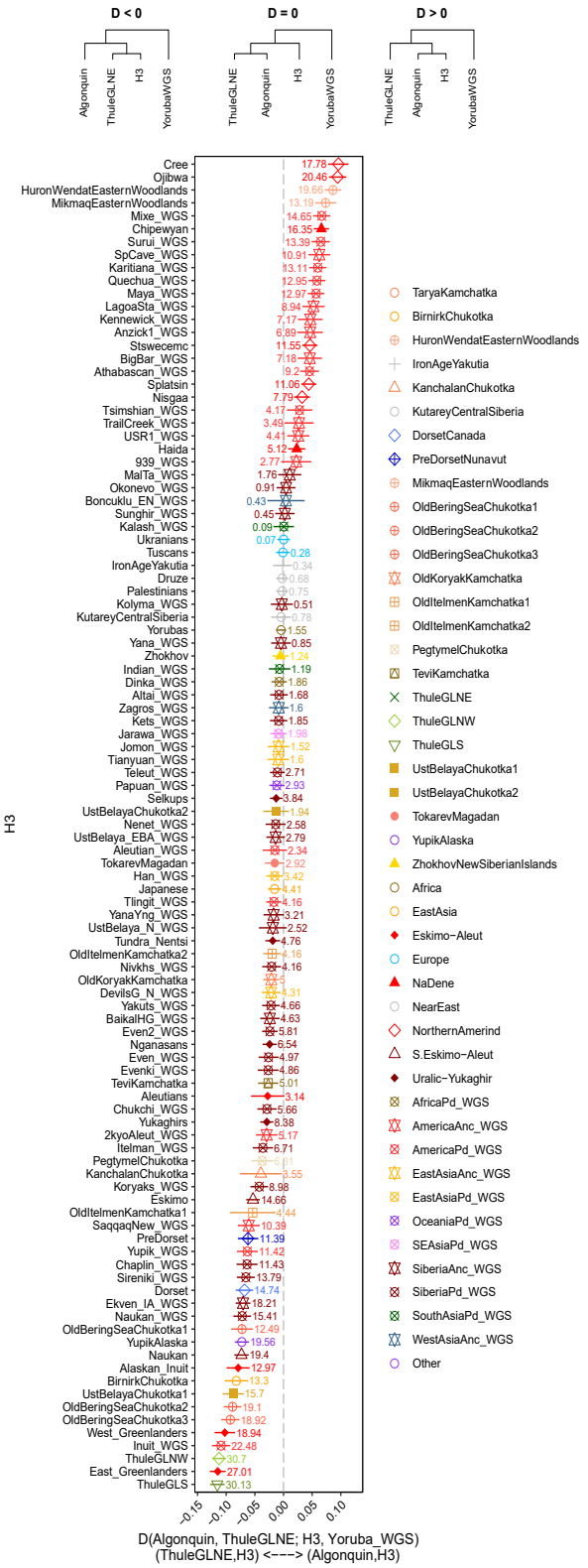



**Figure 3.37** - A *D*-statistic of the form *D*(ThuleGLNE, Athabascan; H3, Yoruba) supporting the formation of the Neo-Inuit groups from an admixture event between a northeastern Siberian population and Northern Native Americans.

### 3.3.5 Tree Mix

#### No Admixture

TreeMix was run on high-coverage ancient individuals to investigate the phylogeny of, and admixture between, selected populations. This analysis was run to investigate the clades and groupings established in previous analyses (MDS, outgroup-$f_3$, and *D*-statistics), which are the Kolyma+Zhokhov clade, the Palaeo-Inuit clade, and the formation of the Neo-Inuit groups through admixture between the Palaeo-Inuit and Northern Native Americans.

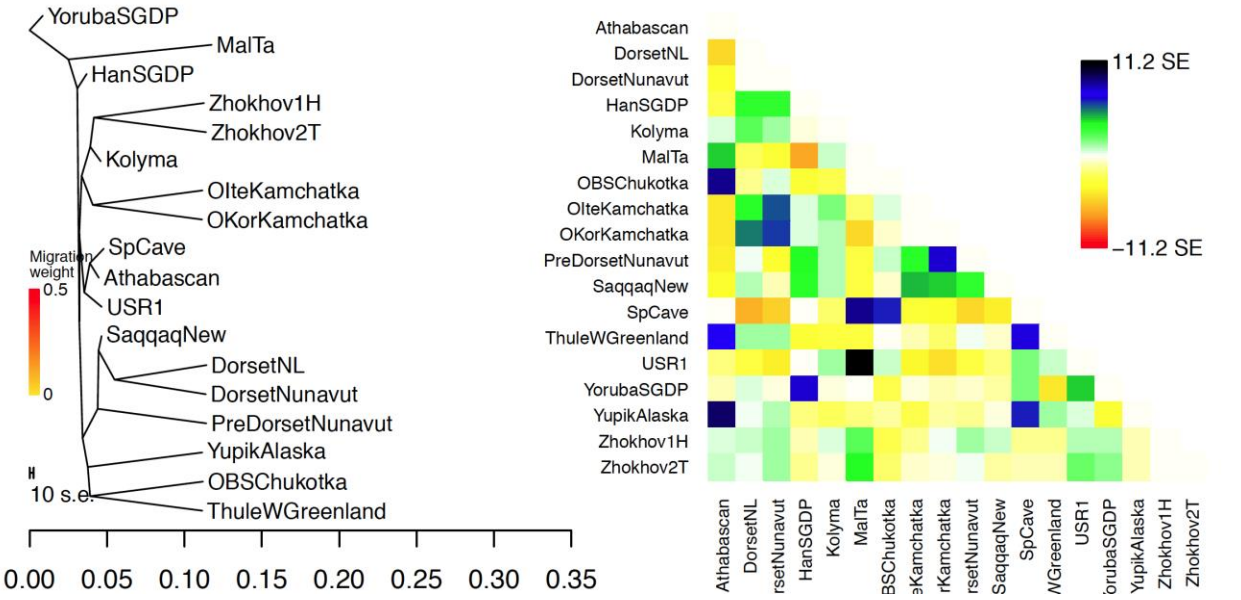



**Figure 3.38** – Phylogeny, with no admixture events, generated by TreeMix based on high-coverage ancient individuals acting as cultural representatives, where three main clades are shown: the "Kolyma+Zhokhov" clade, the "Native American" clade, and the "Palaeo-Inuit+Neo-Inuit" clade.

The tree generated by TreeMix in **Figure 3.38** categorises the ancient populations into roughly three clades: 1) the Kolyma+Zhokhov clade, which includes a subclade of the Kamchatkan populations, the Old Itelmen and Old Koryaks, 2) the ancient Native American clade, with USR1 as an outgroup to Spirit Cave (representing the SNA branch), and the Athabascans (representing the NNA branch), and 3) the Palaeo- and Neo-Inuit clade with the Palaeo-Inuit groups, Saqqaq and Pre-Dorset, as separate branches and Dorset cultural groups form a subclade, whereas with the

Neo-Inuit group, the Alaskan Yupik fall on a separate branch, and the Greenlandic Thule and Chukotka Old Bering Sea form a subclade.

In this tree, there is a large residual between the Neo-Inuit populations and both the Athabascans (NNA) and Spirit Cave (SNA) populations. This is likely due to the NNA component in the Neo-Inuit not being well represented. Similarly, this tree has a large residual between the Kamchatkan Old Itelmen and Old Koryak populations and the Palaeo-Inuit populations. This is likely due to shared ancestry between Chukotko-Kamchatkan groups and Palaeo-Inuits that remained unmodeled with a bifurcating tree.

## Single Admixture Event

The first admixture event added into this model is between the Athabascans and the Neo-Inuit populations **Figure 3.39**. Given the findings from previous publications (2) and previous analyses (MDS, admixture graphs, outgroup $f_3$-statistics, *D*-statistics), this is an expected first migration edge towards a better fitting model. The admixture event fitted by the model presented in the tree fixed the worst residuals from the non-admixed tree, leaving poor residuals between the Yoruba and Han, and USR1 and Mal'ta.

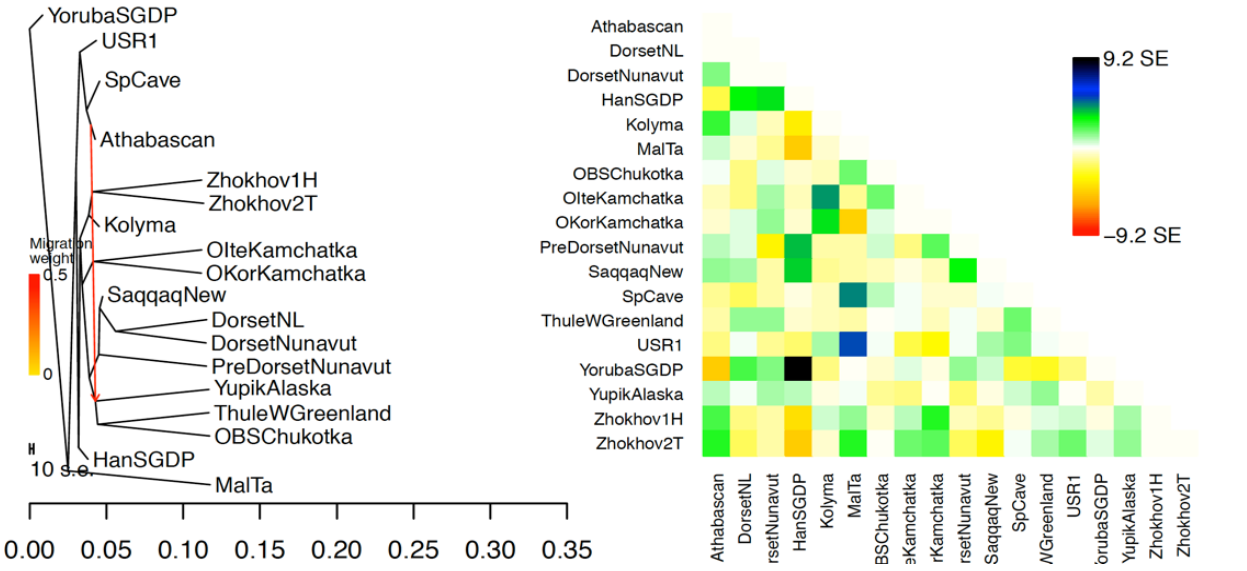



**Figure 3.39** - Phylogeny, with a single admixture event, generated by TreeMix based on high-coverage ancient individuals acting as cultural representatives, where three main clades are shown: the "Kolyma+Zhokhov" clade, the "Native American" clade, and the "Palaeo-Inuit+Neo-Inuit" clade.

## Two Admixture Events

For a second migration edge, TreeMix modelled admixture between the Yoruba and Han, shown in **Figure 3.40**. Although this admixture edge was not expected, it is likely that this edge is inferred due to the unbalanced sampling of different ancestries. In this case, the only East Asian population is the Han Chinese, representing one of the major ancestry components in northeastern Siberians and Indigenous Americans. However, the Han Chinese are still an outgroup to those populations. The inferred admixture edge between the Yoruba and the Han is thus attempting to model these two features of the data. This migration edge addresses the worst residuals in the Single Admixture Event model, but still has poor residuals between the Kolyma+Zhokhov clade and the Kamchatkan populations. This unresolved relationship is in agreement with *D*-statistics results suggesting that the Kolyma individual is more closely related to the population that contributed ancestry to Chukotko-Kamchatkan populations than the Zhokhov population is.

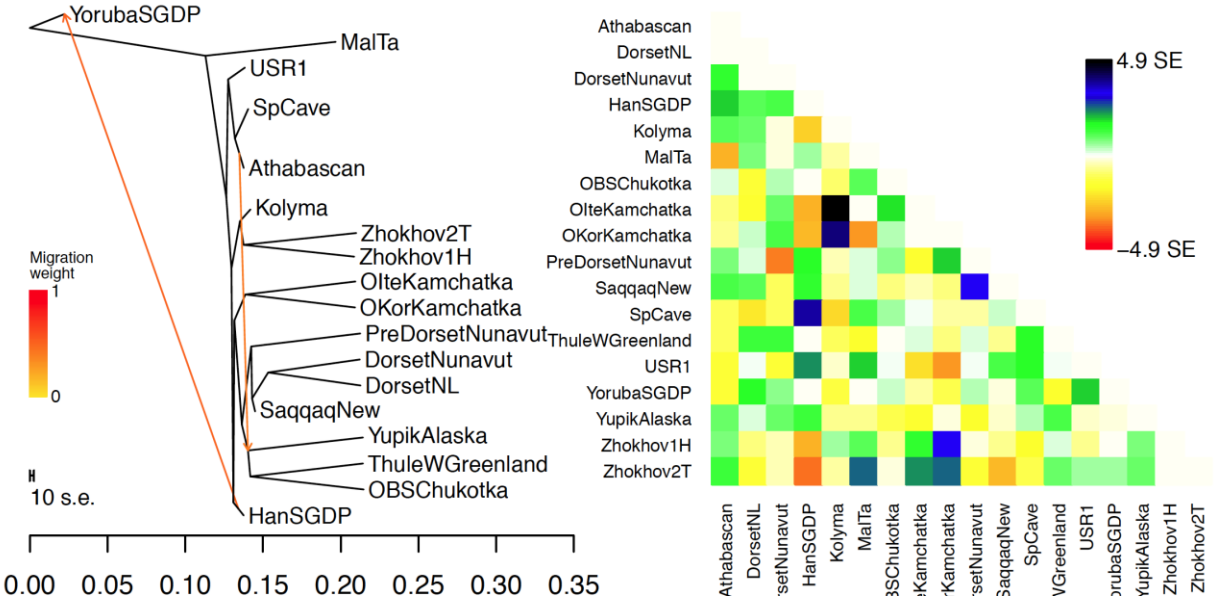



**Figure 3.40** - Phylogeny, with two admixture events, generated by TreeMix based on high-coverage ancient individuals acting as cultural representatives, where three main clades are shown: the "Kolyma+Zhokhov" clade, the "Native American" clade, and the "Palaeo-Inuit+Neo-Inuit" clade.

## Three Admixture Events

The third admixture edge added to the tree is from an Ancient North Eurasian population (a population basal to Mal'ta) into the Kolyma+Zhokhov clade, shown in **Figure 3.41**. This admixture event results in a very good residual between the Dorset and the Pre-Dorset, with a poor residual between Saqqaq and Pre-Dorset. This admixture edge is likely attempting to model the differing ANS ancestry proportions between the more ancient Palaeo-Siberians, *i.e.*, Kolyma and Zhokhov and the more recent Chukotko-Kamchatkan populations.

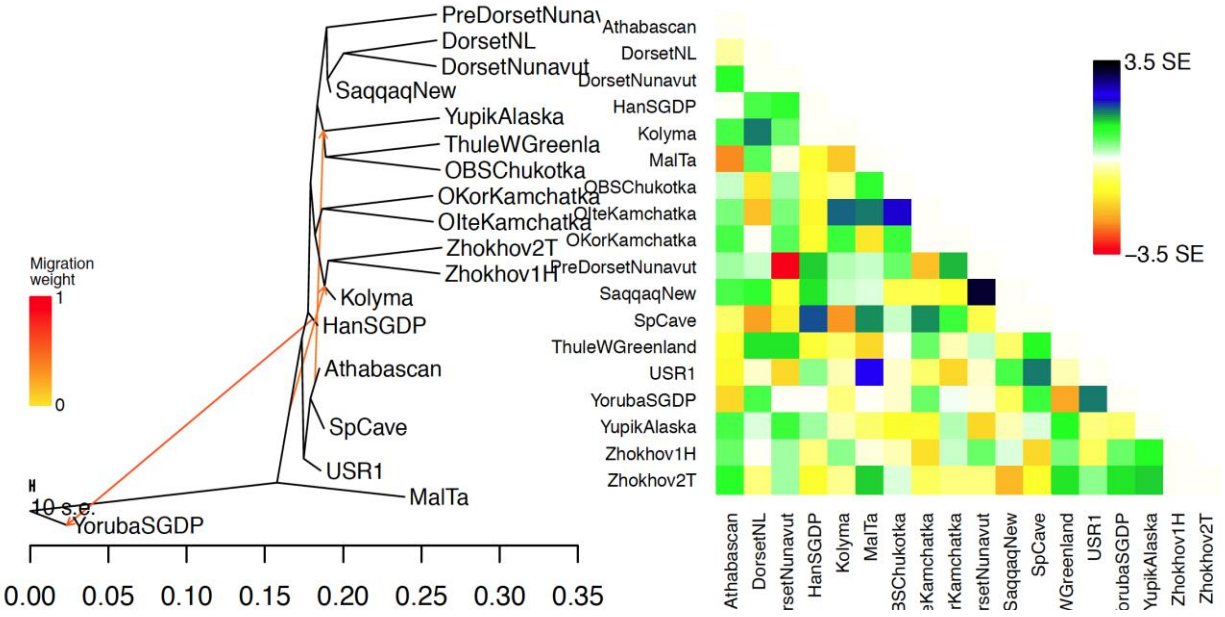



**Figure 3.41** - Phylogeny, with three admixture events, generated by TreeMix based on high-coverage ancient individuals acting as cultural representatives, where three main clades are shown: the "Kolyma+Zhokhov" clade, the "Native American" clade, and the "Palaeo-Inuit+Neo-Inuit" clade.

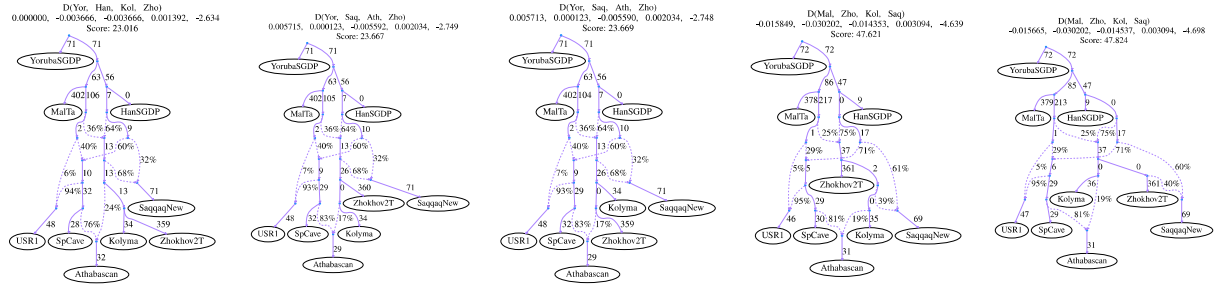
### *3.3.6* Admixture Graph

**Figure 3.42 - Figure 3.46** depict admixture graph results from high-depth cultural representatives of different groups added to the 'Starting Graph' using qpGraph. The top panel in the figure shows the five non-admixed extensions of the seed graph with the best fit scores, and the bottom panel shows the five possible admixed extensions of the seed graph with the best fit scores for each cultural representative. There are three rows of text which indicate the worst *D*-statistic residual and the four-population set leading to such a residual; the values of the *D*-statistic (both expected and observed), the value of the residual and the standard error, as well as the residual Z-score; and the overall model fit score. The graphs in each panel are sorted from left to right in order of model fit score. Solid lines connecting the leaves illustrate direct linkages with the values to the right indicating optimised drift parameters. Dashed lines connecting the leaves illustrate admixed lineages, with percentages to the right indicating the admixture proportions.

## Zhokhov

The difference in fit scores between the non-admixed and admixed models, for grafting the Zhokhov genome into the Starting Graph, is 12.027 (**Figure 3.42**). This difference in fit score is much greater than the conservative threshold of 4.6 that is needed for rejecting the non-admixed model. This indicates that the Zhokhov do not form a clean clade with Kolyma. In the admixed model, the Zhokhov receive 9% of an Ancient North Eurasian component which makes the model a much better fit. The worst residual *D*-statistic in the best fitting non-admixed model includes the feature of this graph under investigation (*i.e.*, the potential Kolyma+Zhokhov clade), which is more support for rejecting the non-admixed model. Although it is not a clean clade, the Zhokhov and Kolyma share substantial ancestry and this finding is supported by the MDS grouping, outgroup $f_3$-statistic, *D*-statistic, and TreeMix phylogeny.

No Admixture




With Admixture

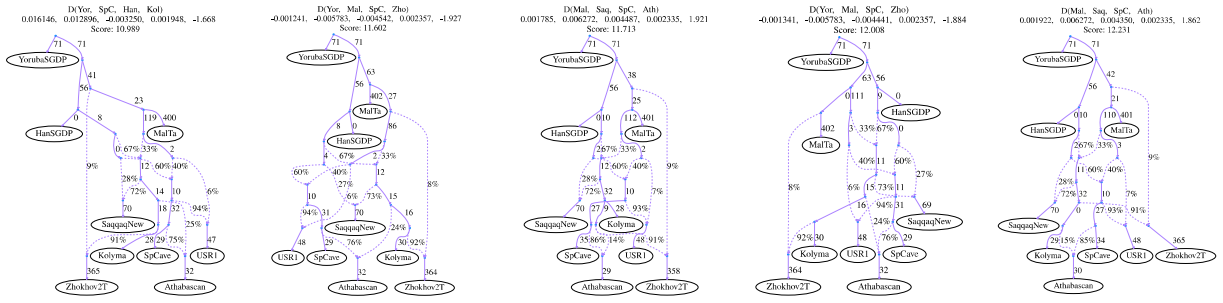
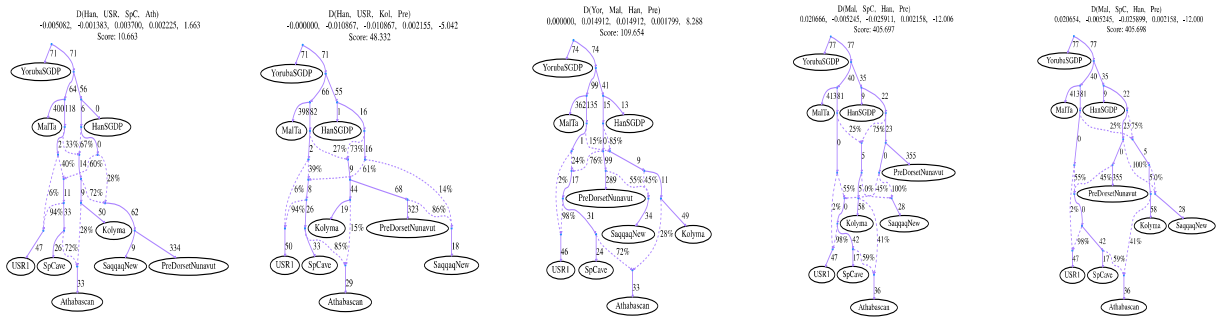



**Figure 3.42** - The admixture graph result of the high-depth Zhokhov cultural representative grafted on to the Starting Graph, where the difference in fit scores between the non-admixed and admixed models is 12.027, indicating that the Zhokhov do not form a clean clade with Kolyma. An enlarged version of this figure is available in Appendix 5, **Figure S5.1**.

## Pre-Dorset Nunavut

The difference in fit scores between the non-admixed and admixed models for grafting the Pre-Dorset individual into the Starting Graph, is 1.828 (**Figure 3.43**). This difference in fit scores does not exceed the threshold of 3, which is needed for rejecting the non-admixed model. This suggests that the Pre-Dorset form a clean clade with Saqqaq. The Pre-Dorset Saqqaq clade is also supported by the MDS grouping, where the Pre-Dorset and Saqqaq individuals always plotted together, as well as the outgroup $f_3$-statistic, $D$-statistic, and the grouping in the TreeMix model.

No admixture



With admixture

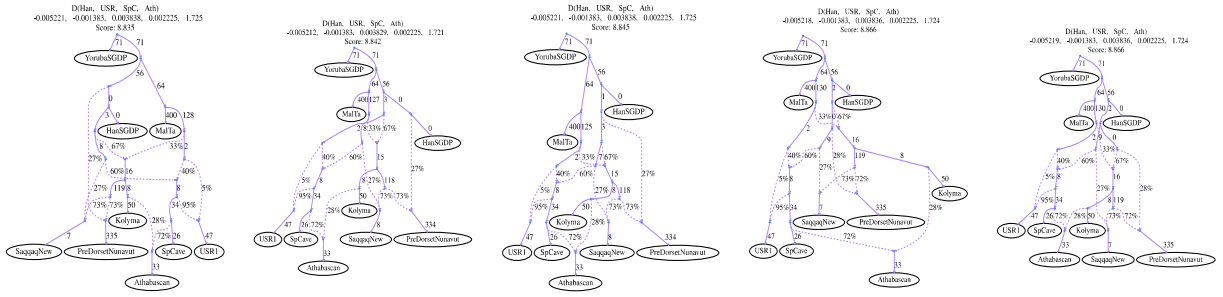
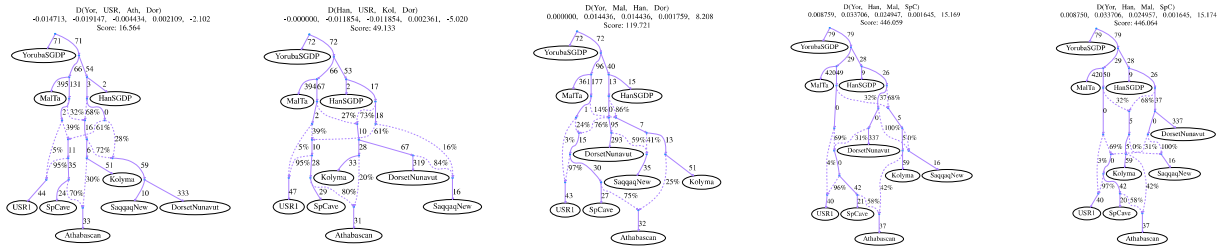


**Figure 3.43** - The admixture graph result of the high-depth Pre-Dorset cultural representative grafted on to the Starting Graph, where the difference in fit scores between the non-admixed and admixed models is 1.828, indicating that the Pre-Dorset form a clean clade with Saqqaq. An enlarged version of this figure is available in Appendix 5, **Figure S5.2**.

## Dorset Nunavut

The difference in fit scores between the non-admixed and admixed models, for grafting the Nunavut Dorset individual into the Starting Graph, is 3.746 (**Figure 3.44**). This difference in fit scores exceeds the threshold of 3, which is needed for rejecting the non-admixed model but does not exceed the more conservative threshold of 4.6. This suggests that the Dorset from Nunavut forms a clade with Saqqaq. The Nunavut Dorset+Saqqaq clade is also supported by the MDS plot, where the Dorset shift with the Pre-Dorset+Saqqaq grouping, as well as the outgroup $f_3$-statistic, the *D*-statistic, and the TreeMix phylogeny.

No admixture



With admixture

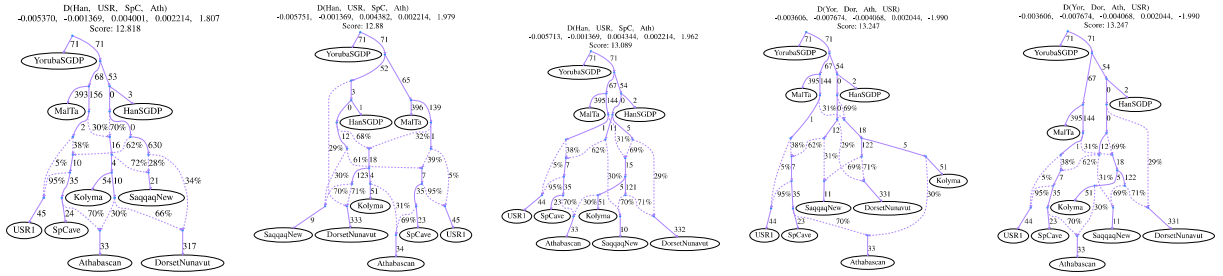
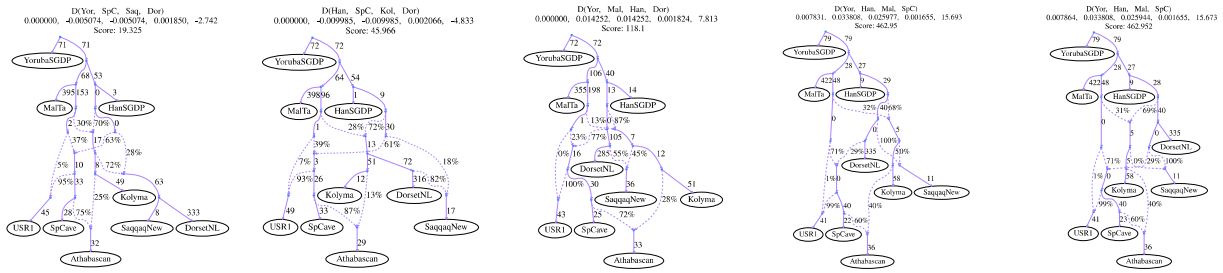


**Figure 3.44** - The admixture graph result of the high-depth Nunavut Dorset cultural representative grafted on to the Starting Graph, where the difference in fit scores between the non-admixed and admixed models is 3.746, indicating that the Nunavut Dorset form a clade with Saqqaq. An enlarged version of this figure is available in Appendix 5, **Figure S5.3**.

## Dorset Newfoundland

The difference in fit scores between the non-admixed and admixed models, for grafting the Newfoundland Dorset individual into the Starting Graph is 10.132 (**Figure 3.45**). This difference in fit score exceeds the conservative threshold of 4.6 that is needed for rejecting the more simplified (non-admixed) model. This suggests that the Newfoundland Dorset do not form a clean clade with Saqqaq. The Newfoundland Dorset and Saqqaq not forming a clade is likely due to a slight basal East Asian ancestry difference that is not well modelled in this admixture graph. This finding is not aligned with evidence for these population groups from other analyses such as the TreeMix phylogeny, outgroup $f_3$-statistic, the $D$-statistic, and the MDS groupings. However, there is a slight separation of the Newfoundland Dorset from the Nunavut Dorset on the MDS, which is perhaps due to this basal East Asian component, causing the Newfoundland Dorset to not join the other Dorset and Pre-Dorset populations in forming a clade with Saqqaq.

No admixture



With admixture

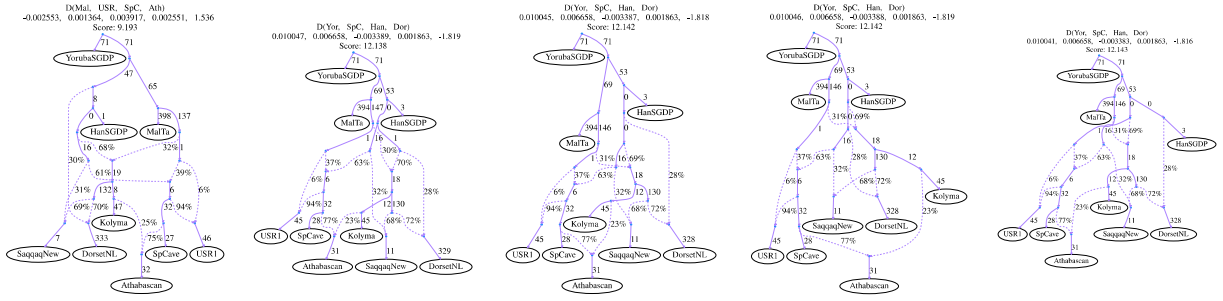
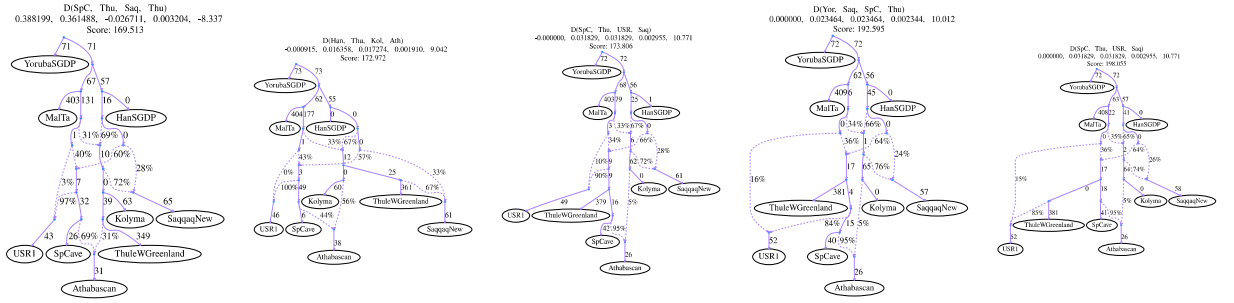


**Figure 3.45** - The admixture graph result of the high-depth Newfoundland Dorset cultural representative grafted on to the Starting Graph, where the difference in fit scores between the non-admixed and admixed models is 10.132, indicating that the Newfoundland Dorset do not form a clade with Saqqaq. An enlarged version of this figure is available in Appendix 5, **Figure S5.4**.

## Thule West Greenland

The difference in fit scores between the non-admixed and admixed models, for grafting the Greenlandic Thule individual into the Starting Graph is 162.443 (**Figure 3.46**). This difference in fit score far exceeds the conservative threshold of 4.6 that is needed for rejecting the non-admixed model. This suggests that the Greenlandic Thule do not form a clade with Athabascans, which is due to the known admixture with a Saqqaq-like population that is captured in the admixed model with the best likelihood. This finding is supported by the first admixture event modelled in TreeMix.

No admixture




With admixture

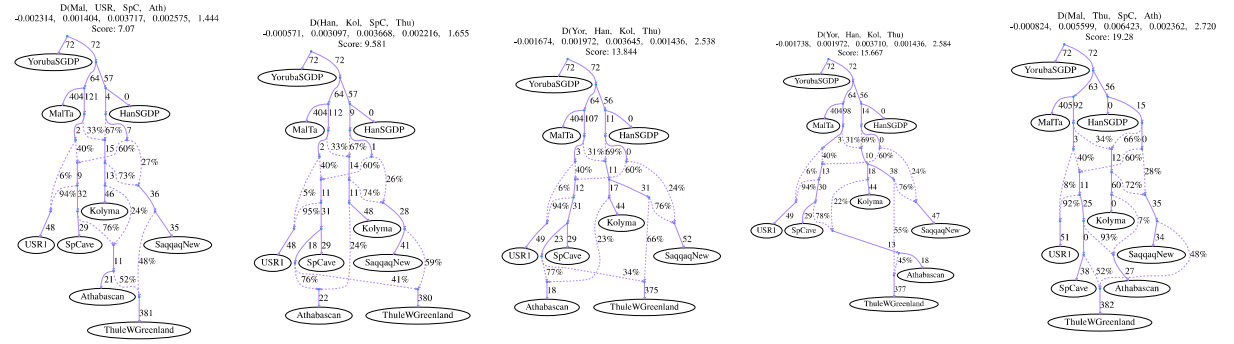
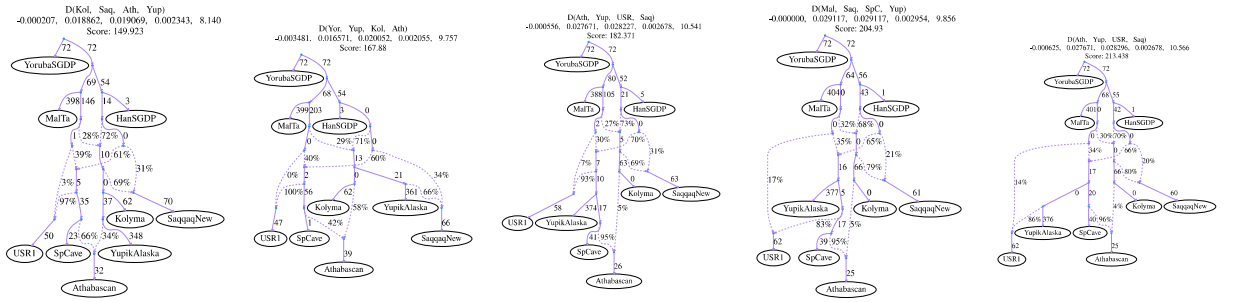



**Figure 3.46** - The admixture graph result of the high-depth Greenlandic Thule cultural representative grafted on to the Starting Graph, where the difference in fit scores between the non-admixed and admixed models is 162.443, indicating that the Greenlandic Thule do not form a clade with Athabascans. An enlarged version of this figure is available in Appendix 5, **Figure S5.5**.

## Yupik Alaska

The difference in fit scores between the non-admixed and admixed models, for grafting the Alaskan Yupik individual into the Starting Graph, is 141.819 (**Figure 3.47**). This difference in fit scores far exceeds the conservative threshold of 4.6 which is needed for rejecting the non-admixed model. This suggests that the Alaskan Yupik do not form a clade with Athabascans due to the known admixture with a Saqqaq-like population that is retrieved in the admixed model with the best likelihood. This finding is similar to the other Neo-Inuit population, the Greenlandic Thule, who also are not modelled well without admixture from a Saqqaq-like population.

No admixture



With admixture

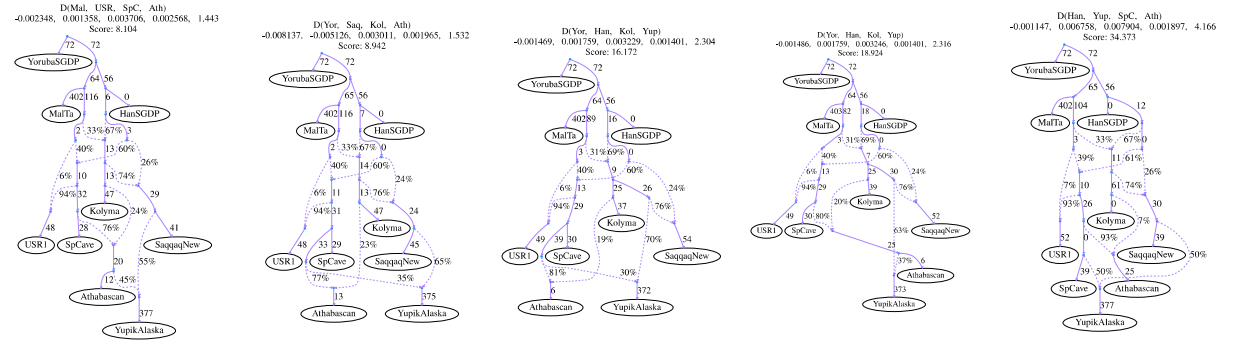


**Figure 3.47** - The admixture graph result of the high-depth Alaskan Yupik cultural representative grafted on to the Starting Graph, where the difference in fit scores between the non-admixed and admixed models is 141.819, indicating that the Alaskan Yupik do not form a clade with Athabascans. An enlarged version of this figure is available in Appendix 5, **Figure S5.6**.

### 3.3.7 DATES

DATES was used to investigate the timing of the formation of the Neo-Inuit gene pool. **Figure 3.48** depicts the weighted covariance decay curve for three Neo-Inuit populations, divided into six subgroupings, that were modelled as a mixture between a Saqqaq-related and an Athabascan-related source population. The estimated time (in number of generations) since the admixture event that led to each of the Neo-Inuit populations is presented in **Figure 3.49**, with the lower and upper bounds of the 95% confidence intervals included. The number of generations since the admixture event were then multiplied by the average generation time (29 years, (180)) and added to the age of that tissue, or average age of the tissues, where more than one genome was assessed (**Table 7**).

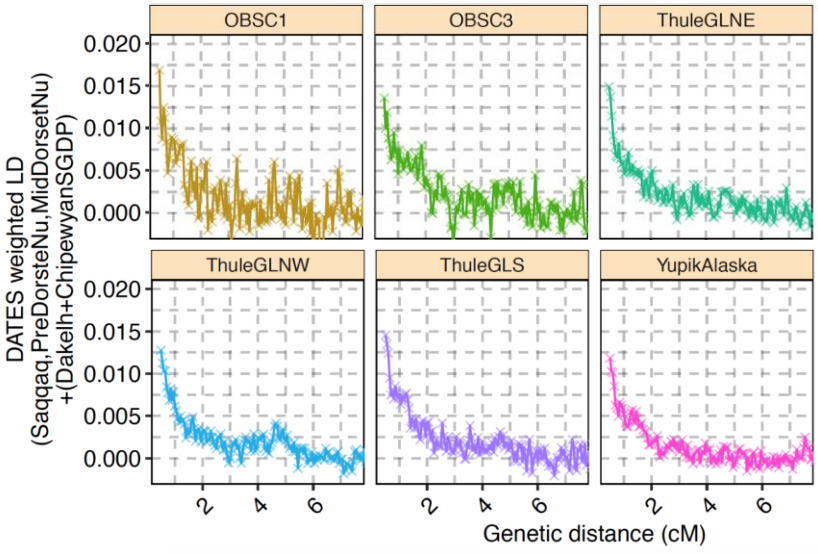



**Figure 3.48** - The weighted covariance decay curves for three Neo-Inuit populations, divided into six subgroupings, that were modelled as a mixture between a Saqqaq-related and an Athabascan-related source population.

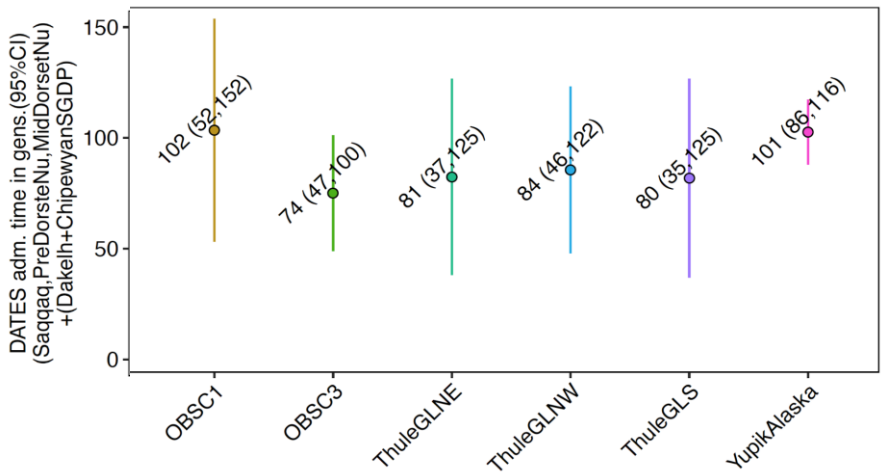



**Figure 3.49** - The estimated number of generations (with the lower and upper 95% confidence intervals) since the admixture event that led to the formation of each of the Neo-Inuit populations.

The older individuals from the Old Bering Sea Culture from Chukotka (OBSC1), who have a C14 age of 2230 BP, were estimated to derive from an admixture event that took place 102 generations ago (95% CI = 52-152), which

equates to ~5179 years ago. Whereas the three younger Old Bering Sea individuals from Chukotka (OBSC3), who's average C14 age is ~1457 BP, were formed from an admixture event that took place 74 generations ago (95% CI = 47-100), which equates to ~3603 years ago. Though the formation of the Thule populations from Greenland were analysed separately based on geographic region (*i.e.,* Greenland north-west [GLNW], Greenland south [GLS], and Greenland north-east [GLNE]), they have been found to have no geographic divide, and therefore the admixture dating results can be considered together. The 48 Greenlandic Thule, who are estimated to be ~300 years old, were formed from an admixture event that took place 82 generations ago (95% CI = 35-125), which equates to ~2668 years ago. The 73 Alaskan Yupik, who are estimated to be ~650 years old, were formed from an admixture event that took place 101 generations ago (95% CI = 86-116), which equates to ~3579 years ago.

**Table 7** - The time, presented both in generation time and years before present, since the admixture event took place that led to the formation of various Neo-Inuit population groups.

| Population Group | Generations, as Calculated by DATES | Lower 95% CI | Upper 95% CI | Average Age of Tissues (BP) | Age of Admixture event, 29y Gen Time | Lower 95% CI | Upper 95% CI |
|---|---|---|---|---|---|---|---|
| OBSC1 | 102 | 52 | 152 | 2221 | **5179** | 3729 | 6629 |
| OBSC3 | 74 | 47 | 100 | 1457 | **3603** | 2820 | 4357 |
| ThuleGLNE | 81 | 37 | 125 | 300 | **2649** | 1373 | 3925 |
| ThuleGLNW | 84 | 46 | 122 | 300 | **2736** | 1634 | 3838 |
| ThuleGLS | 80 | 35 | 125 | 300 | **2620** | 1315 | 3925 |
| YupikAlaska | 101 | 86 | 116 | 650 | **3579** | 3144 | 4014 |

### 3.3.8 Allele Frequency Trajectory

The allele frequency spatio-temporal trajectories of SNPs that are segregating or have reached fixation in Arctic and sub-Arctic populations are depicted in the figures below. The trajectories are accompanied by descriptions of the variant, including functional relevance where possible. Based on the function of the gene in which the variant is present, as well as the substitution type and position, the potential resultant adaptation is considered. Reference populations included in the figures below include African (present-day), European (present-day),

Andaman (present-day Indigenous group from the archipelago between India and Indonesia), and East Asian (present-day). Individuals only appear on figures when the SNP is present in their genome (some low depth individuals may not have adequate coverage at each site and are, therefore, excluded).

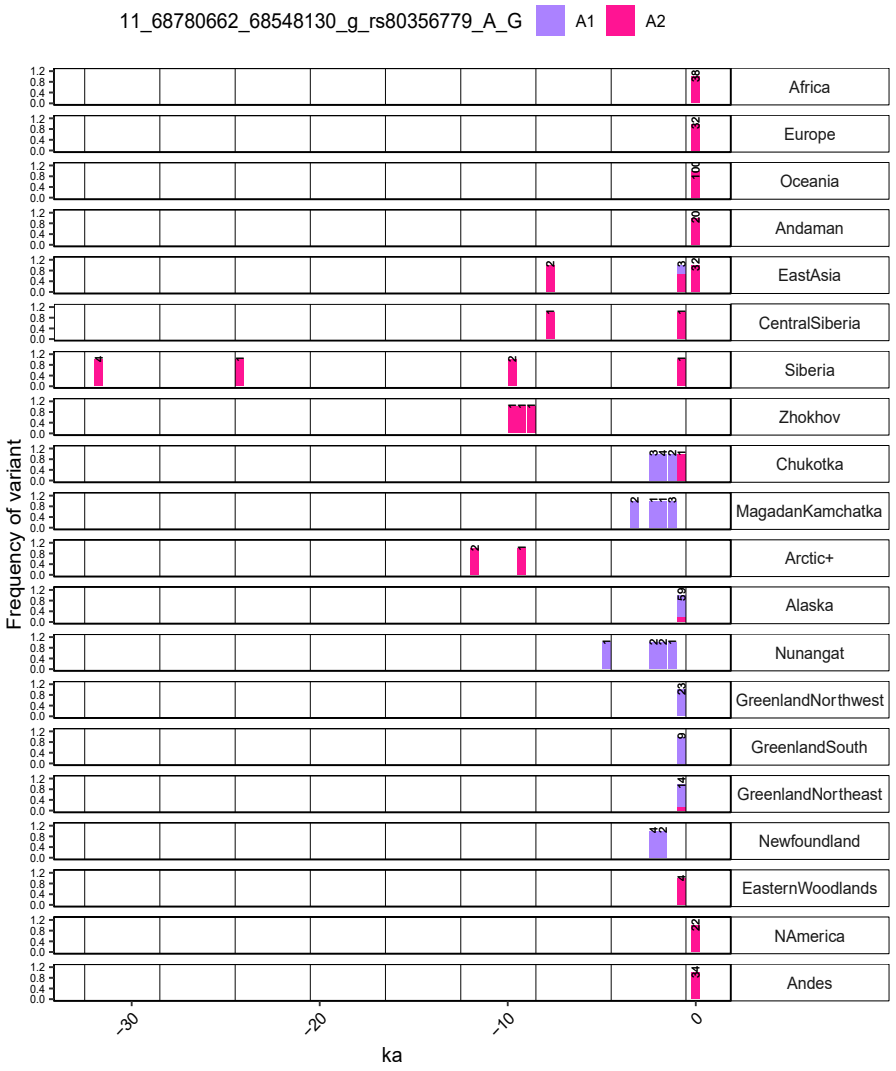



**Figure 3.50** - Visualisation of the segregating site, rs80356779 located in the *CPT1A* gene, where the derived allele (A, shown in purple) is present in most Arctic populations.

The segregating site in **Figure 3.50** shows the derived allele (A, purple) present in most Arctic populations in the last ~5000 years. The SNP at rs80356779 is a nonsynonymous G>A transition (Pro479Leu), found in the *CPT1A* gene, which is known to regulate oxidation of long-chain fatty-acids in the mitochondria (37)). The derived allele at this site is often correlated to hypoketotic hypoglycaemia that can result in mortality in infants. This allele is very common in present-day Canadian and Greenlandic Inuit, northeastern Siberians, and is also at elevated frequencies in the Aleutians, Nivkhs, and Athabascans as well as being present in the ancient Saqqaq individual (9,37). This allele contributes to one of the strongest selective sweeps on the human genome, thought to have taken place between 6-23 kya (37). This selective pressure is thought to be an adaptation to either living in a cold environment or eating a high-fat diet (37). This is supported

by the bulk isotope findings presented in **Figure 2.14**, where almost all individuals with the derived allele have diets of predominantly marine mammals (*i.e.*, the Neo-Inuit, Palaeo-Inuit, and Siberian groups, as outlined in Radiocarbon Dating and Bulk Isotopes).

**Figure 3.51** shows a site where one of the alleles (G, pink) has raised to high frequency exclusively in populations in the Americas, as early as the Late Pleistocene. The SNP at rs12577276 is an intronic variant in the *FADS2* gene, that codes for a fatty acid desaturase, which regulates the levels of polyunsaturated fatty acids (PUFA) in plasma (8). The derived allele (G, pink), at this site has been found to be associated with multiple metabolic phenotypes, as well as anthropometric traits, such as height and weight (8). Moreover, the derived allele is almost absent in Magadan and Kamchatkan populations, but is frequent in American Arctic populations who carry some Native American-related genetic ancestry. These results suggest that the selection of this allele likely started acting on the ancestral Native American population.

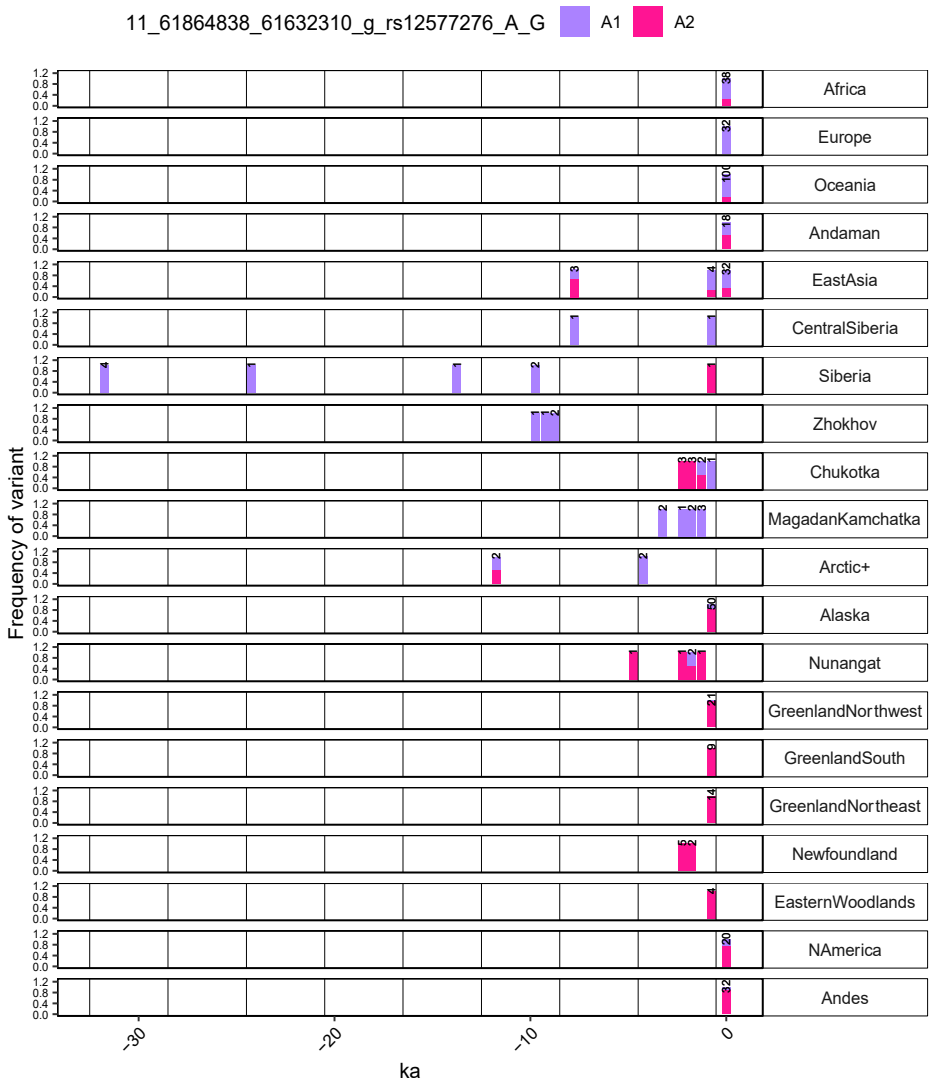



**Figure 3.51** - Visualisation of the segregating site rs12577276, located in the *FADS2* gene, where the derived allele (G, shown in pink) rises to high frequency in populations in the Americas, as early as the Late Pleistocene.

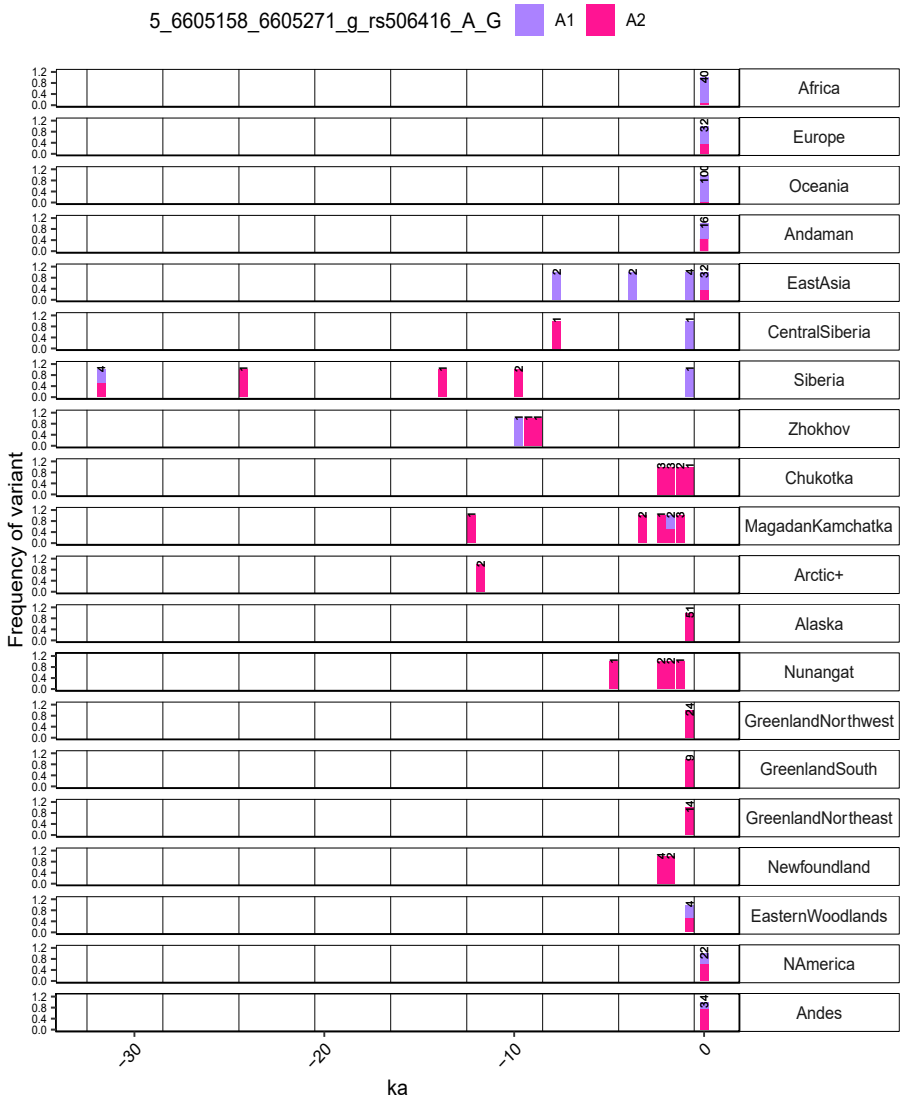



**Figure 3.52** - Visualisation of the segregating site rs506416, located in the *NSUN2* gene, where the derived allele (G, shown in pink) is present in Ancient North Eurasians and almost all Arctic and Indigenous populations in the Americas, but not present in East Asians.

The segregating site in **Figure 3.52** depicts a derived allele (G, pink) that is present in Ancient North Eurasians and almost all Arctic and Indigenous populations in the Americas, but not present in East Asians. The SNP at rs506416 is an intronic variant found in the *NSUN2* gene, which is a methyltransferase involved in the formation of 5-methylcytosine (187) . This gene has been found to play a role in cognitive function (188); however there is no known functional relevance of this particular variant and therefore no adaptation is considered. The geographic and temporal distribution of the derived allele (G, pink) suggest that it likely rose to high frequencies in Ancient North Siberians during the Pleistocene, and that it retained its high frequency in the related ancient Palaeo-Siberian, Native American, Palaeo-Inuit, and Inuit populations.

The segregating site in **Figure 3.53** shows the derived allele (G, purple) present in many Arctic populations but is not present in Ancient North Eurasians, Palaeo-Siberians or Native Americans. The SNP at rs1984651 is an intronic variant found in the pseudogene *GTF2IRD1P1*. No functional relevance for this SNP is known; however, a proximal SNP in the same gene was found to reduce central corneal thickness, a phenotype that is common in people from cold and dry climates (189,190).

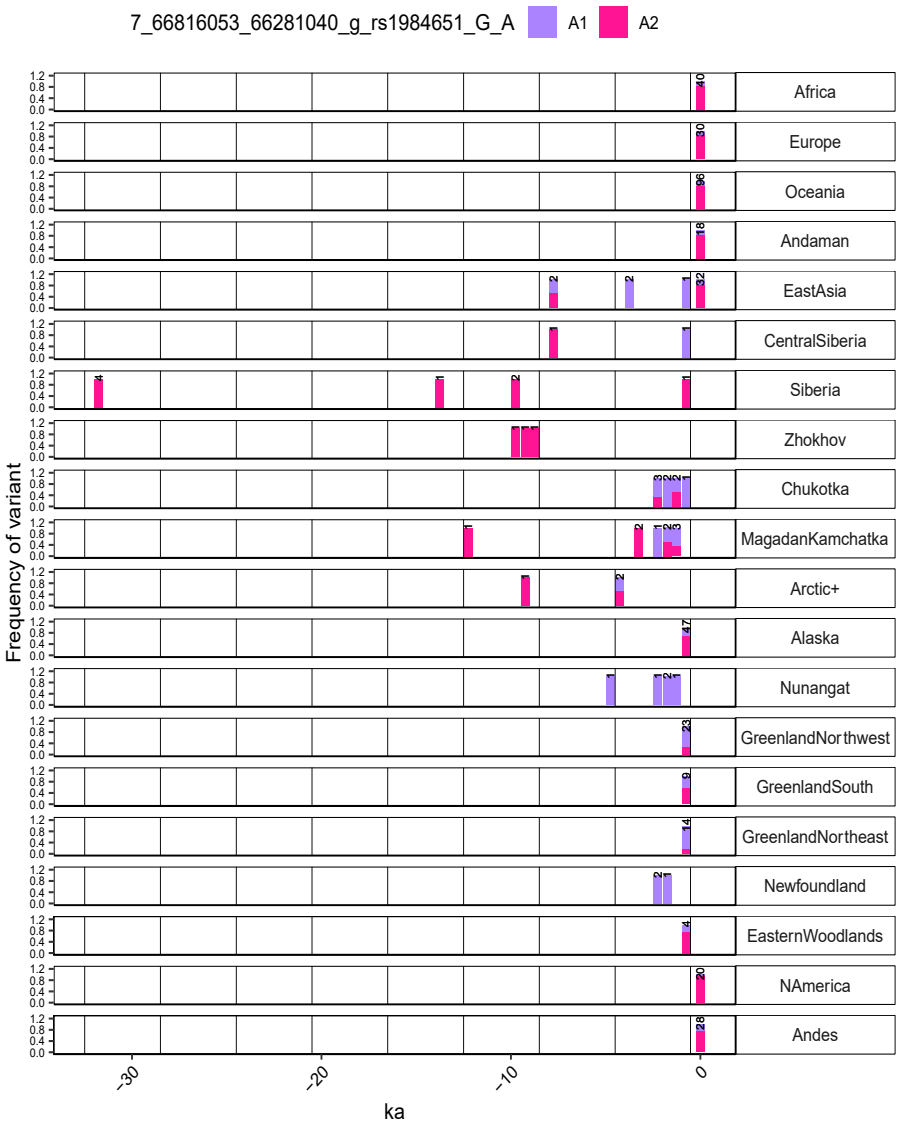



**Figure 3.53** - Visualisation of the segregating site rs1984651, located in the *GTF2IRD1P1* gene, where the derived allele (G, shown in purple) is present in many Arctic populations but is not present in Ancient North Eurasians, Palaeo-Siberians, or Native Americans.

The spatio-temporal distribution of this derived allele suggests that its frequency rose in Arctic populations in Siberia and the Americas, but only during the Holocene, since more ancient Siberians carry the ancestral allele (A, pink).

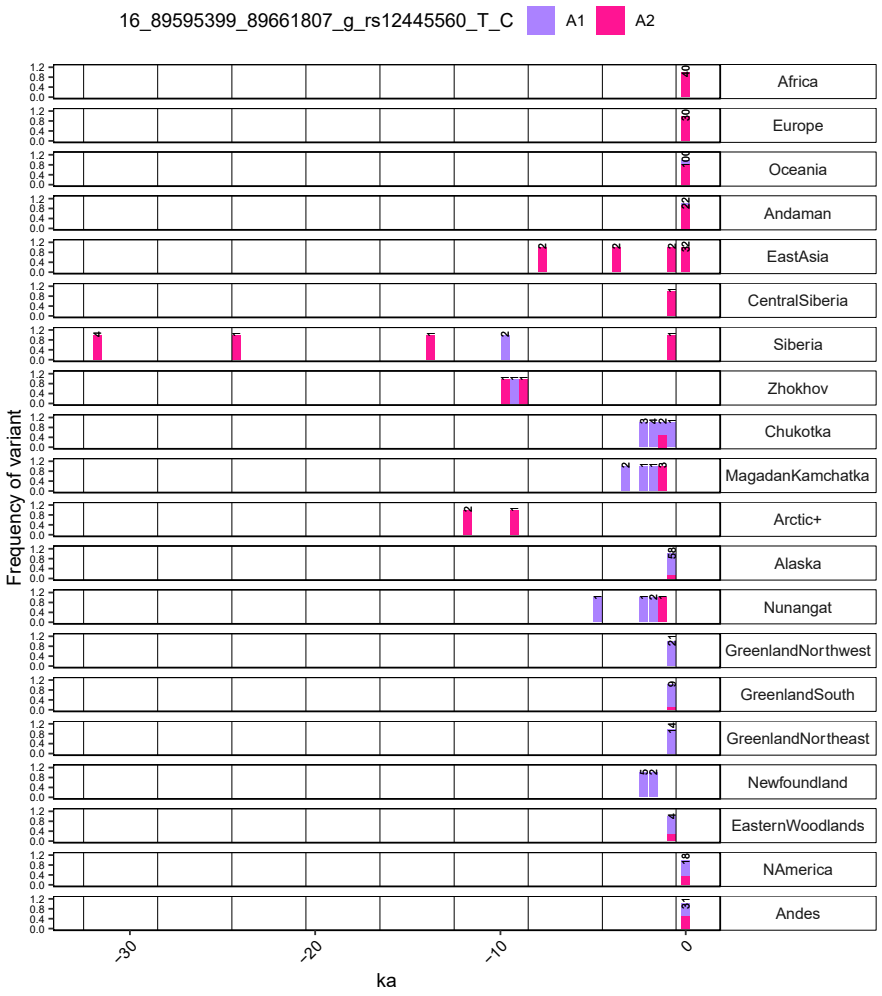



**Figure 3.54** - Visualisation of the segregating site rs12445560, located in the *CPNE7* gene, where the derived allele (T, shown in purple) is present in Arctic populations and Indigenous populations in the Americas.

**Figure 3.54** shows a site segregating in Arctic populations and Indigenous populations in the Americas. The derived allele (T, purple) is found in some Palaeo-Siberians but is not found in East Asians or Ancient North Eurasians. The SNP at rs12445560 is a synonymous substitution in an alternative splicing variant in the *CPNE7* gene, that codes for a copine protein, involved in cell adhesion and cell migration. Copines function in lipid metabolism pathways and have been suggested to be associated with heart disease and obesity (9). The observation that one of the alleles is not present in any of the putative source populations that contributed to the formation of northeast Siberians and Native Americans, *i.e.*, East Asians and Ancient North Siberians, suggests the frequency of the alternative allele rose after the Last Glacial Maximum, once East Asian-related populations started moving northwards.

## *3.4 Discussion*

### 3.4.1 Uniparental Markers

#### Mitochondrial Haplogroups

There is congruency in the haplogroup assignments in the Palaeo-Inuit groups in northern Canada, which is consistent with previous publications (*i.e.*, D2a, D2a1) (2,191,192). The only other individual with a haplogroup from the D2 lineage in the dataset belongs to the Tokarev Culture. The Tokarev Culture has been noted to have similar archaeological artefacts to Palaeo-Inuit groups (68).

Consistency in haplogroup assignments was also observed amongst the Neo-Inuit populations, where most individuals had haplogroups A2a, A2a1, A2a3, or A2b1; however, one instance of D1 and four instances of D4b1a2a1 are observed in the Alaskan Yupik (2,29,193). The D4b1a2a1 haplogroup was also observed in the Ust-Belaya Culture, the Kanchalan Culture, and in two layers of the Ushki site, one representing the Tarya Culture and one representing the Late Ushki Culture (though there were not adequate damage patterns for the Late Ushki site). The matrilineal discontinuity between the Palaeo- and Neo-Inuit was taken as early evidence to suggest that the latter constituted a new migration wave into the American Arctic, distinct from the resident Palaeo-Inuit (2,191,192).

Various individuals from eastern Siberia shared similar haplogroups. For example, haplogroups C4b and C4b1 were identified in the Old Itelman Culture, the Tevi Culture, and the Old Koryak Culture. Similarly, there is sharing of the G1b and G1b+16129 haplogroup across the vast region of eastern Siberia as it was identified in the Zhokhov, the Old Itelman Culture, the Tevi Culture, the Old Koryak Culture, the Ust-Belaya Culture, and the Kanchalan Culture. Interestingly two Zhokhov individuals were found to have the haplogroup D4h3, which is most commonly found in present-day Thai people (194).

The H2a2a1 haplogroup is the default haplogroup assignment when no distinctive variants are found; however, this haplogroup is also present in some Eurasian populations (195). This was noted because the Yakutia individual 022903 was assigned H2a2a1 despite the 481X coverage on the mitochondrial genome. This prompted an investigation into species identification, which proved to be human, and it was later determined that this individual does in fact have haplogroup H2a2a1 despite the low haplogroup probability assignment and, therefore, this individual's maternal lineage is likely of West Eurasian origin. This is also supported by this individual's Y-chromosome haplogroup assignment of R1a-L1029, as well as the MDS results, where 022903 (grey plus symbol) falls with populations from Europe and the Caucasus **Figure 3.11**.

## Y-Chromosome Haplogroups

Haplogroup Q branches are common in Indigenous populations of the Americas and Asia (196). The Q1b-YP4010 haplogroup is an extremely rare lineage that is widely distributed in Eurasia (previous findings in Poland and near Lake Baikal) (197), as well as having a sub-branch that is found in Native Americans, most notably in the 1800 Lovelock individual from Nevada, USA (22). The Q1a-B143 haplogroup is characteristic and almost exclusively found in Palaeo-Siberians and Palaeo-Inuit groups, and presently is found in Neo-Inuit populations on both sides of the Bering Sea (2,25), which is consistent with it being found in the Palaeo-Inuit and Neo-Inuit population groups in this project. The Q1b-M3 (M848) lineage is the most common Native American Y-chromosome haplogroup, being found in 85-90% of all Native Americans (198,199), including one of the Huron-Wendat individuals. The other Huron-Wendat individual has a sister branch to Q1b-M3 (M848), the Q1b-M3 (Y4303) haplogroup, that is more common in northern regions of the Americas (198), which is consistent with it also being found in Neo-Inuit populations from Chukotka to Greenland. The presence of Q1b-M3 in the Old Bering Sea indicates genetic backflow to Eurasia from North America or shared origins of OBS and Neo-Inuit, in line with previous studies (2,25).

The N1a haplogroup is a common circumpolar lineage, found in the region spanning from Sweden to Chukotka (200). The extremely rare N2a-Y6514 haplogroup was identified in one of the Zhokhov individuals. This haplogroup is found in very few present-day people in different

parts of Russia. This extremely rare haplogroup finding is aligned with the autosomal lineage of the Zhokhov, which does not contribute to present-day populations.

The C2 haplogroup is most commonly found in central Asia and eastern Siberia and, to a lesser extent, in some Indigenous populations in North America (198,201,202). One of the derived lineages of C2a-B473, defined by SNPs P39/Z30536, is found in present-day Athabascans (198,201). Other individuals with C2a-B473 have been found outside of the Americas, such as the ancient individual UKY from south of Lake Baikal (203). The C2a-M48 is a rare lineage found amongst eastern Siberians and some Japanese populations, which is consistent with being found across the Sea of Okhotsk in an ancient individual on the Kamchatka Peninsula (202).

The R1a-L1029 lineage is common in Eastern Europe and has been found as far east as Lake Baikal (204), which is consistent with the geographic and genetic placement of individual 022903.

Overall, analysis of the uniparental markers recapitulates known genetic relationships between Arctic populations in the Americas and Siberia, both in prehistory and present-day.

### 3.4.2 Broad Genetic Groupings

#### Palaeo-Siberians

The $f_3$- and *D*-statistic results give resolution to the different lineages of Kolyma-related ancestry. Most notable is the high level of shared drift between the Kolyma individual and the Zhokhov group. This shows genetic continuity in northern Yakutia, including the New Siberian Islands, for more than 2000 years. This is a logical finding as Kolyma and Zhokhov are contemporary in a time when the mammoth steppe was changing into a tundra landscape and the Holocene Climate Optimum was approaching, offering a refuge for humans spread over a vast area (205).

Outside of the Kolyma+Zhokhov clade, there are different groups with varying proportions of Palaeo-Siberian related ancestry that are discernible through the outgroup $f_3$-statistic analysis. Palaeo-Siberian lineages dispersed and intermixed with successive waves of ancient East Asian-related populations resulting in the southward groups from Magadan and Kamchatka (both present-day and ancient Itelmen and Koryaks), the eastward groups in Chukotka, and the North American Arctic (the Palaeo- and Neo-Inuit). Whereas the Kolyma and Zhokhov individuals were consistently found to be closely related to each other, *D*-statistics results suggest that the Kolyma individual was more closely related to the Palaeo-Siberian population that contributed to the genomic ancestry of later groups including Chukotko-Kamchatkan, Palaeo-Inuit, and Inuit groups. Potentially, the isolation of the Zhokhov group in the New Siberian Islands could have been the cause for the observed asymmetrical allele sharing pattern.

## Palaeo-Inuit

The amount of shared drift (outgroup $f_3$-statistics) between individuals in the Palaeo-Inuit clade and their consistent grouping in the MDS plots separate them from other northeast Siberian and Native American populations. This result suggests that these groups were isolated from other populations, while retaining broad genetic continuity for at least 2500 years (based on the dates from the individuals in this study). This is in keeping with the genetic continuity previously suggested to have spanned as long as 4000 years (2). Importantly, we observe that Palaeo-Inuit groups lack the NNA admixture signal that characterises Neo-Inuit and Inuit groups. Admixture graph fitting results support these findings and show that Palaeo-Inuit groups were closely related to Palaeo-Siberians but carried additional East Asian-related ancestry that was likely introduced into northeastern Siberia before the Palaeo-Inuit ancestors moved into the American Arctic.

## Neo-Inuit

There is a general trend in the outgroup $f_3$-statistics for the Neo-Inuit groups to be genetically most similar to each other and other Inuit groups, followed by the Palaeo-Inuit and Native American populations that gave rise to the Inuit and Neo-Inuit groups, and followed by the Palaeo-Siberian and other Siberian populations related to the putative Siberian source.

The Neo-Inuit groups showed varying levels of high genetic similarity with one another, that are mostly in keeping with the genetic cline seen on the MDS plots. The cline shows the Neo-Inuit populations spreading away from clusters of other Arctic and Indigenous groups, with the Yupik being more proximal, followed by the Old Bering Sea and Birnirk groups, and the eastern Greenland Thule being the most distal to Palaeo-Inuit and Siberian-related clusters. Though the MDS showed the Yupik being most similar to the Siberian Inuit, the outgroup-$f_3$ indicates that they share the most genetic drift with present-day Greenlandic Inuit. A similar trend is observable with the Birnirk and the OBSC1 group. Further investigation, using *D*-statistic of different conformations, can help to further resolve these relationships.

The Neo-Inuit groups and the present-day Inuit have very similar genetic ancestry profiles. These results suggest that the ancient Inuit-related population that were sampled do not represent the Siberian source population that subsequently admixed with NNA to give rise to present-day Inuit. Instead, they are the direct genetic ancestors of the present-day Inuit, which is in keeping with the cultural continuity in the archaeological and oral history records (31,102,206–208).

The outgroup $f_3$-statistics showed genetic similarity between the Neo-Inuit groups and the Palaeo-Siberians and other Siberian populations. All of the Neo-Inuit groups share more genetic drift with Kolyma than with the Zhokhov grouping, other than the Birnirk who are closer to the Zhokhov group. This is a rare instance of Zhokhov contributing more ancestry to a group than Kolyma.

The *D*-statistic results for investigating the formation of the Neo-Inuit groups shows that the Dorset and Pre-Dorset are closer to all of the Neo-Inuit groups than Saqqaq, with the Alaskan

Yupik having the highest genetic affinity to the Dorset/Pre-Dorset. From this, it can be inferred that the Dorset/Pre-Dorset are more likely the population that contributes ancestry to the Neo-Inuit, when compared to Saqqaq. This is a logical finding as the archaeological record shows that Dorset groups existed late enough in time to contribute to the formation of the Neo-Inuit (6,33,209,209–211); whereas the Saqqaq population had already reached their demise.

### 3.4.3 Timing Admixture in Neo-Inuit groups

The dates of the admixture events that led to the formation of the various Inuit populations are all relatively similar, ranging from 74 to 102 generations ago. These date estimates are in line with the hypothesis that Neo-Inuit populations formed in a similar time frame across the vast geographical region. It is interesting to see the difference between the date of the admixture event that led to the formation of OBSC1 and OBSC3, which are the upper and lower limits of the formation of Inuit populations. This result suggests that there are 32 generations between OldBeringSeaCulture1 and OldBeringSeaCulture3, which contextualises the population structure in Old Bering Sea Culture seen in the presented analyses in this thesis and when comparing differences in allele frequencies within this population group over time.

Though the Greenlandic Thule were divided by geographic region for this analysis, the result in generation times since admixture is extremely similar across the groups with a slightly older age estimated for individuals from the northwest of Greenland. The Alaskan Yupik had an average time since admixture event very similar to OBSC3. Since the OBSC3 individuals lived approximately 1000 years before the Alaskan Yupik, this finding could suggest that the OBSC3 and Yupik are different branches of a population formed from the same admixture event.

Both the Greenlandic Thule and the Alaskan Yupik tissues included in this project did not undergo radiocarbon dating and therefore the indirect dates assigned for the purpose of this thesis should be considered with reservation. Tissues from both population groups were not dated because of the large impact of the Marine Reservoir Effect - and in the case of the Greenlandic Thule, where the anticipated age was too recent and therefore impacted by wiggles in the

radiocarbon calibration curve. For both sets of tissues relative dating methods, both seriation and stratigraphy, were used to estimate the age of the group. These limitations should be considered when interpreting the results from DATES, where the indirect dates were used to estimate the time since admixture.

OBSC2 was not included in the DATES output because, although only marginally, this genome was not at high enough resolution to determine the timing of the admixture event, resulting in a confidence interval that was twice as large as the estimate. This low resolution is likely due to the genomes from this project being pseudo-haploid, and this pool being composed of a single low-coverage individual (less than 16x). With imputation and genotype calling, it is likely that the confidence interval will decrease, providing a more accurate timing of the admixture event.

Overall, the findings from this thesis are in keeping with other lines of evidence, such as archaeological records, regarding the origins of the Neo-Inuit and their spread eastwards into the North American Arctic. The archaeological record suggests that the Neo-Inuit ancestral population formed in Eastern Siberia/Beringia (6,102,209,212–215). The presented results provide support for this because the oldest admixture dates are for Neo-Inuit populations in Siberia, followed by Beringia, western North America, and lastly for Neo-Inuit populations in eastern North America (Greenland). This deduction is based on the distance required to travel between eastern Beringia and the eastern Arctic.

### 3.4.4 Allele Frequency Distributions

The in-house tool created for depicting segregating sites in genomes of different populations through time and space was useful in identifying both when allele frequencies were changing as well as visualising in which lineage(s) putative selection signals may have arisen. The data from this project, though not genotyped or phased, was adequate for investigating sites that are likely under selection in Arctic and sub-Arctic populations. However, the current data quality should be considered when interpreting the results. Another important consideration is, as with output from GWAS or selection scans, the identified associations are not necessarily causal.

Though some of the variants may have been selected for over time (or hitchhiked due to proximity to the causal *loci*), others may be changing in allele frequency solely based on genetic drift.

### 3.4.5 Limitations

There are limitations to the dataset and analyses in this chapter. The limitations on data quality are outlined in Chapter 2, section Limitations, as well as artefacts such as high error rate in ancient DNA data due to deamination, short reads sequencing, and short read mapping (216). The genomes in this dataset are a combination of UDG-treated and non-UDG-treated reads that may still carry signals of damage, impacting downstream inferences, particularly for genomes at lower depth of coverage. Moreover, the data was analysed in pseudo-haploid format and has not been genotyped, imputed, or phased. Pseudo-haploid calls invariably underestimate the actual heterozygosity and have been previously shown to suffer from reference bias (217,218). In future, genomes will undergo imputation and calling of diploid genotypes for inclusion in other analyses, such as ChromoPainter and fineSTRUCTURE (219).

There are also limitations to the dataset that are not technical, but instead biological in nature. Many populations being studied in this thesis are heavily isolated and have become genetically homogenous overtime. This results in reduced genetic variation between individuals from the same or similar population groups, making it difficult to resolve fine-scale differences. Another demographic factor that can confound results is the numerous admixture events with incoming migration waves that impacted the structure of these populations (*e.g.*, East Asian migrations into the northeast of Siberia, Siberian migrations into the Americas, North Native American back migrations into Siberia) (2,25–28). Another limitation is in the sparsity of the archaeological record in Arctic and sub-Arctic regions. The population size of these groups was always small and, accounting for variable burial practices such as leaving the dead in the open or on sea-ice, few sets of remains were left behind and exist today (7,212,220).

There are some limitations to the reference dataset that should also be considered. The reference dataset is composed of less than 200k SNP sites, most of which are transitions, and most

of which were not ascertained in Indigenous Siberians or American genomes. Though this reference dataset is adequate for the purpose of this thesis, as it suffices to identify genetic phenomena that occurred on a large scale, it is unlikely to contribute to disentangling more in-depth research questions.

There are also limitations to the population genetic analyses applied to answer the research questions in this chapter. The largest limitation to the tools is the input bias, where only populations represented in the reference data or project data can be examined for genetic similarity. For example, some ancient groups remain ghost populations as there are no ancient genomes sequenced from them for representation. Therefore, any descendant populations with the ghost ancestry component will be difficult to resolve in population genetic analyses. Similarly, if some populations are underrepresented either in the reference data or the project data, it will be difficult to reach the level of granularity required to draw conclusions about the genetics of that group.

### 3.4.6 Future Directions

#### Location of admixture of the Neo-Inuit

Though it was possible to estimate the timings of the admixture events that led to the formation of the Neo-Inuit groups, it was not possible to infer the locations of these admixture events. Discerning the locations may become possible as the quality of the data increases, through genotype calling, imputation and phasing, and/or when more genomic data from ancient Arctic and sub-Arctic peoples becomes available. An expanded time transect of the Alaskan Yupik, or the Greenlandic Thule would be beneficial in addressing this research question, as well as additional Neo-Inuit genomes from Canada and Eastern Siberia from yet unsampled cultures such as Punuk (eastern Siberia and Alaska) and Ipiutak (Alaska).

## Selection Scans

Performing selection scans on this dataset would enable further investigations into the adaptations that these populations underwent over time. Selection scans will become possible when the dataset undergoes genotype calling and imputation, which is currently being optimised in the group for ancient genomics datasets mapped to the human reference genome, hg38. Results from these selection scans will then be considered with known historical events, such as changes in climate or population contacts (between Indigenous groups in Siberia or the Americas, or with European colonists). These population contacts cannot always be discerned through admixture and, therefore, pathogen profiles (of the diseases that an individual contracted or was exposed to in life) can be compared to infer potential population overlaps and contact (66), as well as examining the pathogen load as a selection driver.

## Oral histories

A review of the traditional knowledge systems will be conducted, while considering the findings from this project, to assess the relationship between oral and genetic histories from Arctic and sub-Arctic population groups. Through community engagement and knowledge exchange these interdisciplinary overlaps will assist in telling a more complete story of lessons of resilience in harsh Arctic climates and the peopling of the North.

## *3.5 Conclusion*

The population genetic analyses conducted in this chapter adequately answered the research questions pertaining to the genetic relationship between the Kolyma individual and the Zhokhov group, the genetic similarity amongst the Palaeo-Inuit individuals, and the formation of three Neo-Inuit groups from across the Arctic. Both uniparental markers and autosomal data were investigated to resolve the shared drift and admixture between groups. Sites that have been proposed to be under selection in Arctic and sub-Arctic populations were visualised to explore when and in which lineages these changes in allele frequencies may have risen. The findings from this chapter will further elucidate the population genetic histories of these groups once dataset quality is increased through imputation and a connection to Indigenous oral histories is made.

# 4 Chapter 4 - Pathogen Incidence in Ancient Arctic and Sub-Arctic Populations

## *4.1 Introduction*

### 4.1.1 Overview of Ancient Pathogen Genomics

Although infectious diseases have long afflicted humans, skeletal pathology and historical records often do not suffice in accurately categorising the true pathogenic agent (38). Prior to the advent of, and recent advancements in ancient DNA extraction and sequencing technologies, direct evidence from past epidemics was not possible. With the development and optimization of laboratory and bioinformatic techniques, pathogen sequences can now be obtained as a by-product of sequencing human genomes (39). Some tissues, such as teeth, have been found to have higher DNA preservation levels (40,41) and, because of the high levels of circulation that reach the oral cavity, the layer at the bottom of the tooth root (the cementum) has been found to be the best source of ancient pathogen genetic material in humans (42).

Many investigations into ancient pathogens have focused on bacteria, viruses, and protozoa (39,43). Because of the overall low incidence of pathogen recovery (due to factors such as low disease incidence in a population, poor DNA preservation, and varying levels of pathogen in the individual at the time of death), many studies of ancient pathogens have focused on bacteria that cause chronic infection, such as *Mycobacterium leprae* or *Mycobacterium tuberculosis* (44). The first complete human pathogenic bacterial genome was of *Yersinia pestis*, published in 2011 (45). Following that, many more studies have been conducted into ancient bacteria that cause illness in humans, including *Helicobacter pylori* (*H. pylori*), *Brucella melitensis*, *Treponema pallidum* (syphilis), and *Salmonella enterica* (221–224). The first ancient virus to be extracted was five segments of the A/H1N1 influenza genome that caused the 1918 pandemic, isolated from lung tissue that was paraffin-embedded and formalin-fixed (45,46). Since then, many studies have been conducted on ancient viruses such as human papillomavirus, human parvovirus, human T-cell

leukaemia virus, and variola virus, with the oldest being a hepatitis B virus extracted from a 7000-year-old human tooth from Germany (49,55,56,225–229). The field has expanded to include protozoan parasites such as the causal agent of malaria, *Plasmodium falciparum* (230). However, as is common with large eukaryotic genomes assembled through extracts from culture, the malarial reference genome has human contamination, resulting in human reads mapping and inauthentic findings (231).

### 4.1.2 Pathogens in Northern Eurasia and the Americas

#### Historical & Modern Pathogens

Ancient Indigenous populations from Siberia and the Americas have experienced low levels of transmissible illnesses through time as populations remained small and widely dispersed (47,48). However, after contact with early colonists, the pathogen load in these groups rose substantially, with the introduction of diseases such as measles, influenza, pertussis, smallpox, and intestinal infections (57,58). These pathogenic agents are thought to be particularly lethal amongst Indigenous groups as they did not have the same coevolutionary history with European diseases (57). The mortality rate amongst Indigenous groups led to a sharp decline in pre-colonial pathogen incidence and therefore many remain undocumented (232). To this day, Indigenous populations in Siberia and the Americas suffer from outbreaks of tuberculosis (at 10 times the national rate), meningitis, and sexually transmitted infections (58–63).

Some pathogens remain at higher incidence specifically in present-day Arctic populations because of cultural practices. Infections from stored foods are common, such as *Clostridium botulinum* (found in fermented meat that has been traditionally prepared) or *Trichinella* (found in hunted game meats) (58,64,65). Similarly, several types of parasites such as *Diphyllobothrium* and *Corynosoma* are very common in the Arctic fish species and can infect consumers with tapeworms (58).

## Ancient Pathogens

Studies of ancient individuals from northern Eurasia and the Americas have identified various ancient pathogens, including bacterial species such as *Yersinia pestis* and *Mycobacterium tuberculosis*, as well as viral species such as parvovirus, variola virus, and hepatitis B virus (49–53).

Bacterial pathogens *Yersinia pestis* and *Mycobacterium tuberculosis* have been sequenced from ancient Siberian and Indigenous American remains, respectively (50–52). *Yersinia pestis*, the causative agent of the bubonic, pneumonic, and septicaemic plague has been identified in ancient Altai individuals dating to 4800-4700 BP, where there was evidence of an epidemic at that time (50). One of these individual, RISE 509, from the Afanasievo culture, was found to have a strain of *Y. pestis* (in this instance causing bubonic plague) that was basal to all other known strains, in which a promoter mutation suggested that fleas were not the zoonotic agent (51). *Mycobacterium tuberculosis* complex, commonly known as TB, causes deterioration of lung tissue when in active disease state (233). Genomic and osteological evidence of TB was found in three pre-Columbian individuals from southern Peru dating to ~1000 BP (52). This study contrasted with previous theories of the introduction of TB to the Americas by European colonists and showed that the ancient Peruvian strain is more analogous to the TB adapted to seals and sea lions than to the known human-adapted forms (52).

Viral pathogens Parvovirus B19, variola virus, and the hepatitis B virus have been identified in ancient remains from Siberia and East Asia (54–56). Three ancient individuals from the Lake Baikal region were identified as having the parvovirus B19 virus (otherwise known as B19V), dating to ~5500, ~3500, and ~1500 BP (54). This time transect indicates that the Parvovirus was present in that region for more than 4000 years (54). When analysed with B19V genomes from other Eurasian individuals, this study suggested a long-term association of the disease with human populations (54). Infection by the variola virus, colloquially known as smallpox, causes intense fever and vesicular rash and has led to hundreds of millions of deaths throughout human history (234). An ancient variola virus was isolated from a ~300 year old mummy from the Sakha region of Yakutia, Siberia, where evidence for pulmonary haemorrhaging

(another symptom of variola infections) was present in the lungs (55,56). The hepatitis B virus (HBV), which causes chronic liver infection, was identified in an individual from ~2000 BP in Mongolia (49). The study found that the geographical distribution of ancient viral genotypes are not consistent with present-day geographical distributions, which is plausibly due to redistribution through large human migrations in the Bronze and Iron Ages (49). HBV was also identified in a 400-year-old mummified child from Korea, exemplifying its vast geographic distribution (53).

### 4.1.3 Purpose

There are very few ancient human pathogens recovered from Siberia or the North American Arctic; therefore, the purpose of this chapter is to sequence and catalogue the genomes of pathogens across Arctic populations, which have been obtained as a by-product of generating high quality human genomes. Based on this information, the aim is to characterise the ancient pathogen load in the Arctic in order to, eventually (beyond the scope of this thesis), explore potentially drastic shifts in pathogen species during cultural or climatic transitions, as well as proposed or known periods of population contact. With the pathogen incidence information generated in this thesis, future studies can undertake deeper sequencing of the pathogen-positive DNA extracts and retrieve pathogen genomes at higher coverage. High quality pathogen genomes will aid in reconstructing their evolutionary histories, and possibly lend insights on epidemiological factors such as virulence and regional adaptation. Additionally, consistencies between pathogen profiles of different Arctic and sub-Arctic populations could supplement analyses in which population contacts are detected or even debated (66), particularly when minimal or no admixture is apparent.

### 4.1.4 Outline of Chapter

This chapter outlines the methodology for identifying pathogens in DNA extracted from ancient Arctic and sub-Arctic humans and cataloguing the pathogen load (presence/absence and species assignment) for use in future works. This begins by detailing the approach to ascertaining and authenticating ancient pathogen reads. The chapter then outlines the pathogens found in the

ancient Arctic tissues and categorises the incidence of disease through space and time. Three pathogens are then detailed as case studies, in which the authentication is described, and the relevance of the finding is discussed. Finally, the future directions for this research are outlined.

## *4.2 Methods*

The non-USER treated libraries from the 217 ancient tissues included in this project, sequenced as outlined in Chapter 2, section Shotgun Sequencing, were screened for the presence of pathogens. Using AdapterRemoval, reads were trimmed, collapsed and filtered to those at least 30 bp long. Reads retained were then input to an in-house, high-throughput pathogen identification pipeline, with a database composed of microbial genomes from species that have been associated with humans.

### K-mer-based Taxonomic Classification and Comparative Mapping

The KrakenUniq tool was used for metagenomic classification of ancient DNA reads (235). KrakenUniq first shreds the DNA reads into sequence substrings, referred to as k-mers, of 31 base pairs. K-mers are then classified taxonomically by comparison to the in-house curated reference database. The **KrakenkmerRank** is determined based on the number of unique k-mers that are classified to each genus. This metric can inform authentication, as an increased number of unique k-mers matching indicates an increased portion of the reference genome being covered. Based on the KrakenUniq results, a decision is made regarding which species to analyse using traditional mapping. For species represented by more than 50 kmers in a sample, all reads that were classified to the same species are then comparatively mapped, using minimap2, in parallel to all different reference genomes in that genus (236). A bam file is created for each of these species' assignments, from which summary statistics are generated. The PCR and optical duplicate reads in this bam file are then marked and removed and only reads with a mapping quality of 30 are retained.

### Authentication Metrics

The ancient **DNA damage** pattern on retained pathogen reads was calculated using MetaDamage (237). When establishing the threshold for exclusion it is important to consider the

geographic region from which the human remains originated, as well as to compare the damage on the pathogen reads to the damage pattern from the corresponding human reads.

The next step in authenticating ancient pathogen findings is assessing the distribution of reads mapping to the reference genome, or evenness of coverage. If there is not an equal distribution of reads over the genome, indicating that the reads are stacked in conserved regions, then it is expected to be a false positive. To determine how evenly reads are distributed across the reference genome, the depth of coverage is calculated (139,169). Based on this metric, there is a theoretical expectation of coverage across the genome, which, when divided by the observed coverage, indicates how evenly the mapped reads are distributed, termed the coveragePRatio. When the coveragePRatio is greater than 0.5 it indicates even coverage, where reads are relatively evenly distributed across the genome. When the coveragePRatio is less than 0.5 it indicates that reads may be stacked, where there is an uneven distribution of coverage.

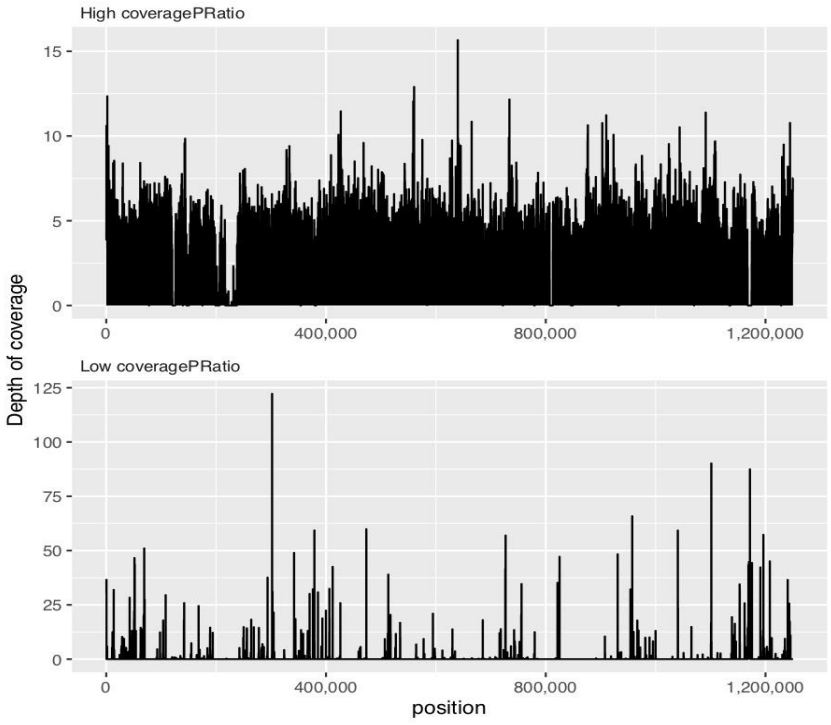



**Figure 4.1** - Bar plot indicating the number of reads mapping to an ancient human pathogen, where both findings have approximately the same number of reads mapping (note the difference in scale on the y-axis) and the same average coverage; however, there is clear read stacking and unequal coverage in the bottom panel.

**Figure 4.1** is a depiction of the reads mapping to an ancient human pathogen, where both findings have approximately the same number of reads mapping and the same average coverage (note the difference in scale on the y-axis). The reads in the top panel are equally distributed, with a more even breadth of coverage, whereas in the bottom panel, reads are stacked in particular regions, resulting in the genome having unequal coverage.

The similarity to the reference genome should also be assessed when authenticating ancient pathogens, as it is important in discerning between true findings and matches to similar species. This is calculated through the **edit distance (**also referred to as the Levenshtein distance), which is the number of changes required to convert one string into another (238), indicating in this case the difference between the read and the reference genome. The lower the edit distance, the more likely it is that the finding is authentic.

Another central metric is the **average nucleotide identity** (ANI) which is the number of mismatches, normalised against the length of the read (239). The ANI provides the average similarity of the set of reads to the reference genome provided; therefore, this value should be as close to one as possible, with low ANI being less than 0.97 and high ANI being greater than 0.97. The aniRank100 is a complementary metric restricted to hits within a genus that have at least 100 reads mapped.

### 4.2.2 Low Complexity Mask

The low complexity regions of pathogen genomes in the reference database were masked using SDUST (240). Masking involves replacing low complexity genomic regions with Ns, resulting in most analysis being focused on the middle regions of pathogen genomes. The mappability mask can affect the coveragePRatio because k-mers cannot align to the regions on either side of the mappability mask, so it is difficult to achieve 100% coverage. To correct for this, the **coveragePRatioCorr** metric normalises the number of aligned k-mers to the mappable regions of the pathogen genome.

### 4.2.3 Species Exclusion

Many microbial species exist in nature that are not necessarily pathogenic to humans. On this basis, findings were excluded based on two considerations, 1) if there is an analogous species that is present in soil (*e.g.*, *Clostridium* and *Klebsiella*) (241,242) or 2) if they are common

commensals that cause no infection (including constituents of the oral microbiome (*e.g.*, *Streptococcus* and *Staphylococcus*) (39,243), or are species of unknown clinical relevance (*e.g.*, *Yersinia intermedia*) (244). For example, *Streptococcus* is a known species in the human oral microbiome, so when these species are identified they are often authentic; however, these commensals are typically non-pathogenic, and it is difficult to determine the source (243). Similarly, *Staphylococcus* findings are often authentic, but it is difficult to link the species to a particular infection (39). *Yersinia intermedia* is not a pathogenic strain of *Yersinia,* but may act as an opportunistic pathogen (244).

### 4.2.4 Phylogenetic Trees

Where ancient pathogens of interest were identified and authenticated, and where previous ancient DNA studies have been conducted on that species, the newly sequenced genomes were integrated into the existing phylogeny (56). In this case, all sequencing data from the ancient individual where that pathogen hit was found were remapped to all reference genomes of the species identified, using bowtie2 (245). A consensus sequence was then generated from the mapped bam file, where the allele was called when there was 80% consensus of reads at that site (139,169). Because there is often low coverage of ancient genomes, there was no threshold for the number of reads covering one site. Alignment of the consensus sequence with the reference sequences was done using MAFFT with the –add flag, where the new sequence is added to the existing alignments rather than realigning all sequences (246). Grappa was used to place the newly identified pathogen onto a reference phylogenetic tree using –graft, which converts the weighted placement to an attachment point along that branch (247). The tree was generated using an in-house R-Script based on ggtree v 3.6.2 (248).

## *4.3 Results*

### 4.3.1 Catalogue of Ancient Pathogens

A summary of relevant pathogen hits is available in **Table S9**. This table is composed of pathogens that passed the authentication metrics, are hits from microbial genomes (rather than lab-generated contigs), and species known to infect humans. In total, 25 ancient microbial species were found in 34 individuals: 15 hits from Siberia (originating from six sites), 12 hits from Alaska (originating from a single site), eight hits from Canada (originating from 12 sites), and seven hits from Greenland (originating from six sites). The most pathogens identified were from the Nunalleq site in Alaska, where five of the 12 authenticated findings were assigned to the human polyomavirus 5. Four of the individuals from the Zhokhov site were found to have *Yersinia intermedia*, which was the only pathogen found for the New Siberian Islands. The individual with the most pathogens identified was 022561, with four different findings (*Haemophilus aegyptius*, *Haemophilus influenzae*, *Pseudomonas stutzeri*, *Yersinia enterocolitica*). Further data and analyses are required to confirm whether this is a true case of co-infections.

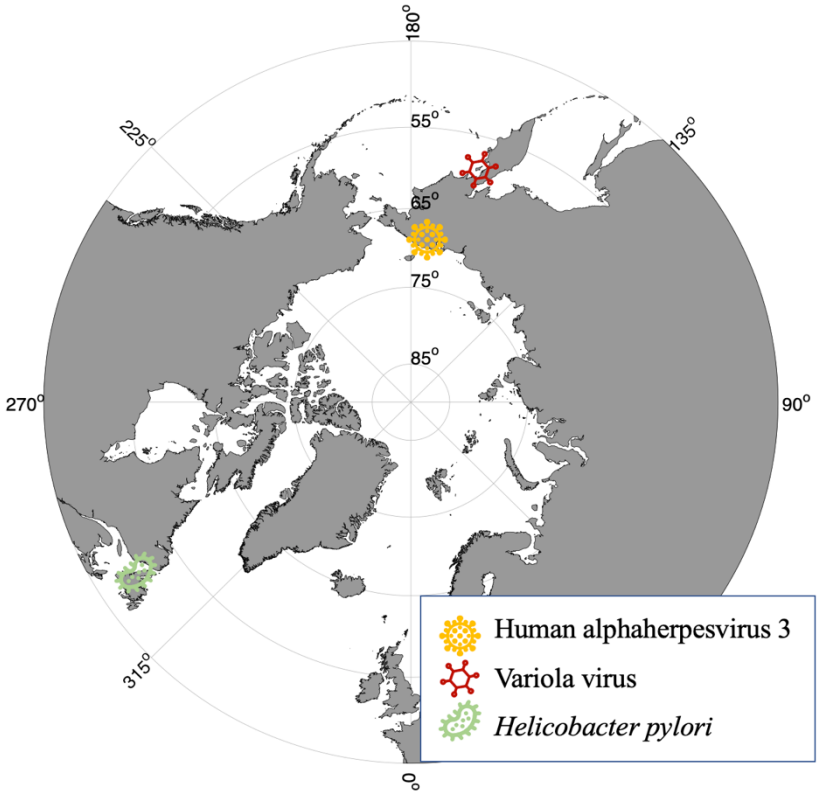



**Figure 4.2** - Map of the locations of origin of the ancient pathogens (variola virus from Kamchatka, the human alphaherpesvirus 3 from Chukotka, and *H. pylori* from Newfoundland) that are detailed as case studies in this chapter.

Of the ancient pathogens identified, the variola virus, human alphaherpesvirus 3, and *H. pylori* were selected for

further analyses and discussion (**Figure 4.2**) (249). The pathogens in **Table 8** are restricted to species that have; been identified as causing infections in humans, have either a KrakenkmerRank or an aniRank of 1, have more than 50 reads mapping, have a coveragePRatioCorr of ≥0.9, and have an ANI of over 97%. Figures below were generated on genomes where more than 100 reads were found.

**Table 8** - List of pathogens that are detailed as case studies. This selection was restricted to species that have been identified as causing infections in humans, have either a KrakenkmerRank or an aniRank of 1, have more than 50 reads mapping, have a coveragePRatioCorr of ≥0.9, and have an ANI of over 97%.

| ID | Taxa Species | nReads | dam5p | ANI | Edit Dist Avg | Human dam5p | Culture | Age of Tissue |
|---|---|---|---|---|---|---|---|---|
| 024266 | Variola virus | 104 | 0.077 | 0.993 | 0.356 | 0.086 | Old Koryak | 1095 BP |
| 024261 | Variola virus | 57 | 0.000 | 0.990 | 0.544 | 0.083 | Old Bering | 1580 BP |
| 011447 | *Helicobacter pylori* | 196 | 0.286 | 0.982 | 0.908 | 0.183 | Dorset | 1935 BP |
| 011448 | *Helicobacter pylori* | 78 | 0.063 | 0.984 | 0.846 | 0.094 | Dorset | 2030 BP |
| 024098 | *Helicobacter pylori* | 67 | 0.250 | 0.982 | 0.910 | 0.078 | Thule | ~300 yo |
| 024123 | Human alphaherpesvirus 3 | 194 | 0.146 | 0.980 | 1.180 | 0.115 | Bronze age | 939 BP |

## Variola Virus

The variola virus was identified in a tooth from an Old Koryak individual, 024266, from the coast of the Kamchatka Peninsula dated to 1095 BP (**Figure 4.3**). The variola viral genome has 104 reads mapping from 024266, with an ANI of 0.99 and an average edit distance of 0.356. The average ancient DNA damage pattern at the 5-prime end of the variola viral reads was 7.8%, which is comparable to the human reads showing 8.6% terminal damage.

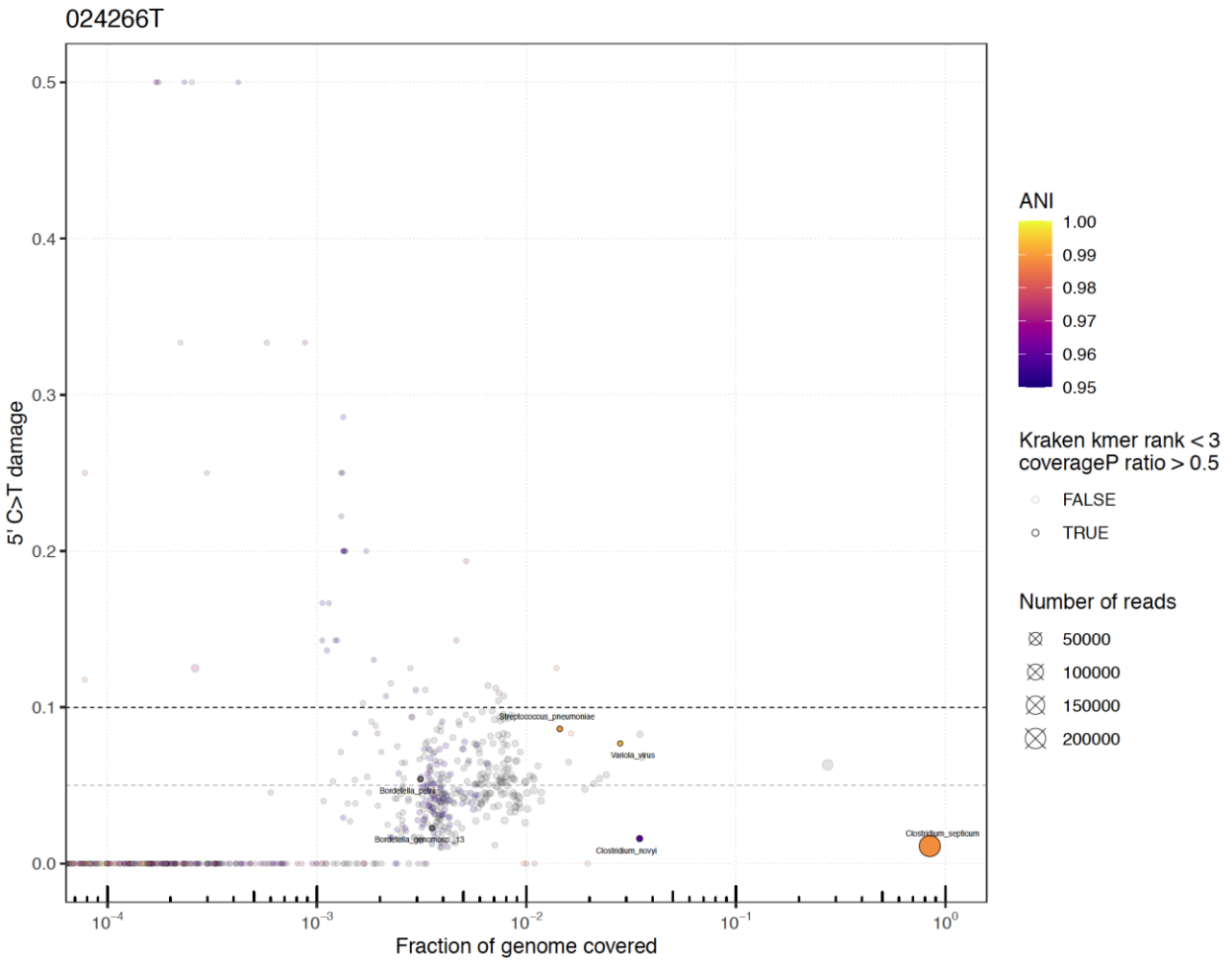



**Figure 4.3** - Plot depicting all the microbial species (including the variola virus) identified from 024266, a tooth from the Kamchatka Peninsula. Opaque circles indicate a coveragePRatio of >0.5, with the top hits from each identified species accompanied by a label. This circle's size is proportional to the number of reads mapping and the circle's colour indicates the ANI. The dotted line indicates 10% 5' C>T ancient damage pattern.

**Figure 4.3** - Plot depicting all the microbial species (including the variola virus) identified from 024266. Circles that are opaque have a KrakenkmerRank of less than three and a coveragePRatio of greater than 0.5, with labels accompanying the top hits from each species (labelling up to two). The size of the circle is proportional to the number of reads mapped to that species and the colour indicates the ANI, where yellower colours are more likely to be authentic and bluer colours are less likely. Circles crossed with an 'x' are hits that are considered authentic based on the following four parameters: a coveragePRatio of >0.8, a KrakenKmerRank of 1, an aniRank100 of 1, and damage of >10%. The variola virus hit was not crossed with an 'x' because it did not pass the 10% damage, but because this is an Arctic tissue and there is lower than expected damage on the human reads as well, this finding was still authenticated.

The variola virus was also found in a tooth of an Old Bering Sea individual, 024261, from the Chukchi Peninsula dated to 1580 BP. This variola virus genome has 57 reads mapping from 024261, an ANI of 0.99, and an average edit distance of 0.554; however, there was no damage pattern on the reads, making it difficult to authenticate. This contrasts with the human reads from this individual showing 8.3% terminal damage.

## Phylogeny

To further investigate the variola virus finding, all sequencing data (both non-USER- and USER-treated) from the ancient individual 024266 was remapped to the variola viral genome. This resulted in 7316 reads mapping and an increase to 1.6x coverage of the variola genome, with a coveragePRatio of 0.918, the ANI of 0.995, and an average editDistance of 0.201 (**Table 9**).

**Table 9** - Results from the in-depth analysis of 024266, the tooth from which the variola virus was identified, where all data (both non-USER and USER-treated) was run through the pathogen pipeline. The table shows the species name, number of reads mapping, the average coverage of the genome, the coveragePRatio, an ANI, and the average editDistance.

| ID | taxNameSpecies | nReads | coverageAvg | coveragePRatio | ani | editDistAvg |
|---|---|---|---|---|---|---|
| 024266 | variola virus | 7316 | 1.581 | 0.918 | 0.995 | 0.201 |

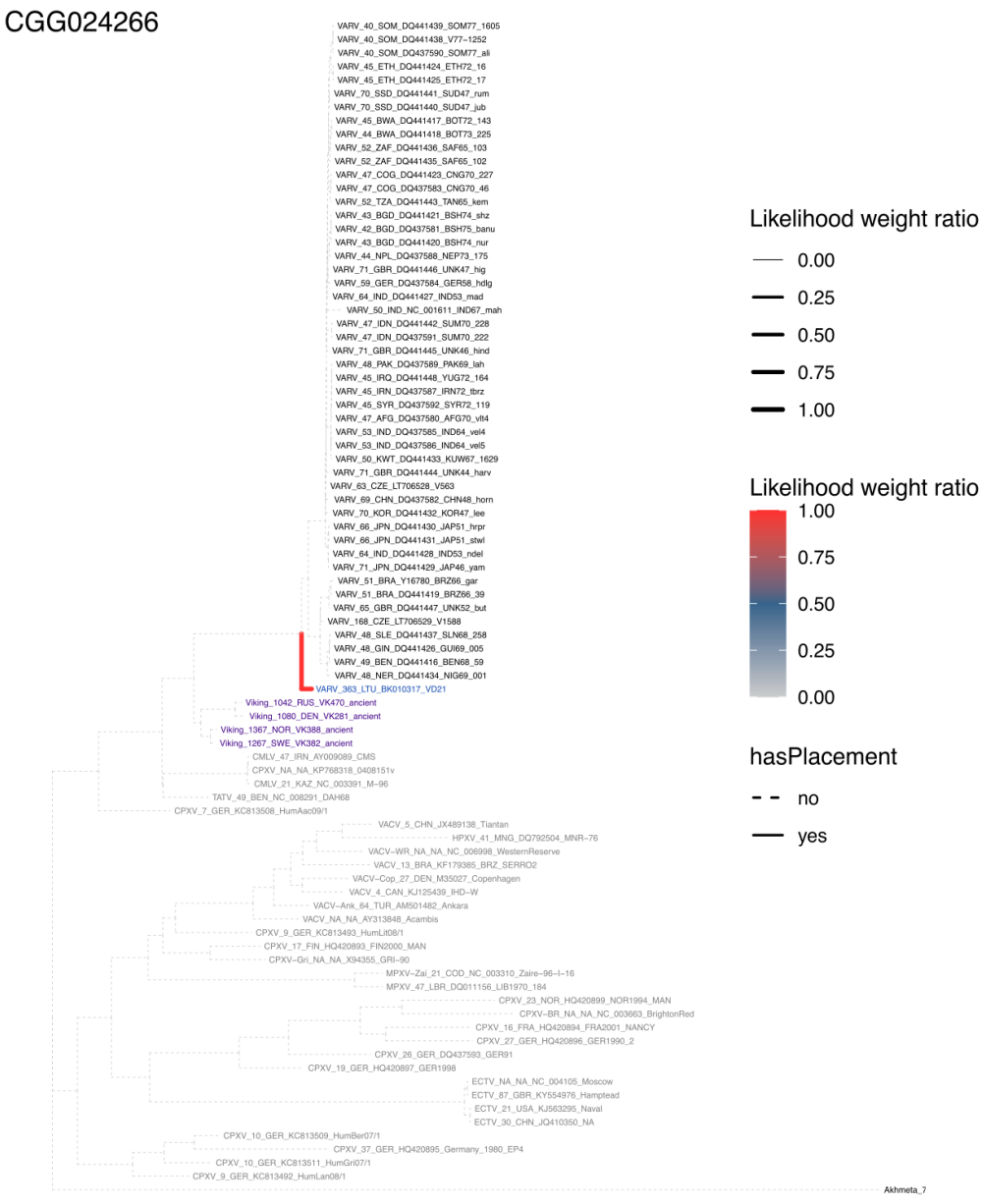



**Figure 4.4** - Phylogenetic tree of the orthopoxvirus, from the Mühlemann *et al*., 2020 study, showing the likelihood weight ratio (depicted in red) of the placement of the variola virus identified in 024266. The orthopoxvirus tree consists of historical variola strains written in black text, the ancient Lithuanian mummy variola strain in blue text, ancient Viking variola strains written in purple text, and non-human orthopoxviruses written in grey text.

The higher depth variola virus genome from 024266 was placed on an existing phylogenetic tree from the Mühlemann *et al*., 2020 study (56). **Figure 4.4** shows the orthopoxvirus

tree, with historical variola strains written in black text, ancient Viking variola strains written in purple text, a single variola strain identified in a Lithuanian mummy (~400 BP (228)) written in blue text, and non-human orthopoxviruses written in grey. The likelihood weight ratio of the placement of the 024266 variola strain is depicted, with both the colour and the thickness of the line indicating a likelihood of close to one. This indicates that the variola virus isolated from 024266 (written in red text) forms a clade with the variola strain from the Lithuanian mummy, as shown in **Figure 4.5**.


**Figure 4.5** - Phylogenetic tree of the orthopoxvirus, from the Mühlemann *et al*., 2020 study, showing the variola virus isolated from 024266 (written in red text) forms a clade with the variola strain from the Lithuanian mummy (written in blue text). The orthopoxvirus tree consists of historical variola strains written in black text, the ancient Lithuanian mummy variola strain in blue text, ancient Viking variola strains written in purple text, and non-human orthopoxviruses written in grey text.

## Helicobacter pylori

The *H. pylori* bacterial genome was identified in a tooth from a Dorset individual, 011447, from coastal Newfoundland dated to 1935 BP (**Figure 4.6**). The *H. pylori* bacterial genome had 196 reads mapping from 011447, with an ANI of 0.982 and an average edit distance of 0.908. The average ancient DNA damage pattern at the 5-prime end of the *H. pylori* reads was 28.6%, which is higher than the human reads of 18.3%.

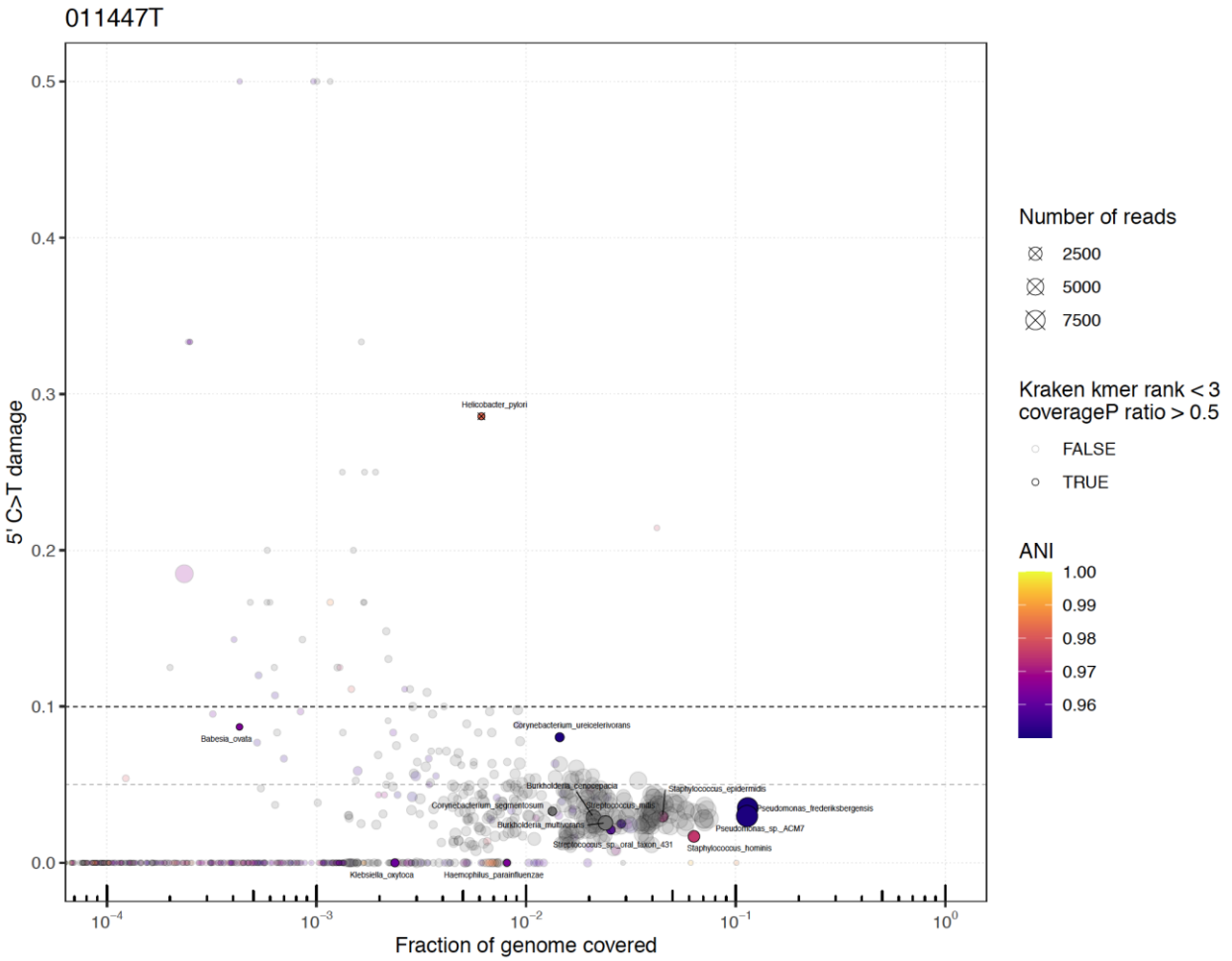



**Figure 4.6** - Plot depicting all the microbial species (including *Helicobacter pylori*) identified from 011447, a tooth from northern Newfoundland. Opaque circles indicate a coveragePRatio of >0.5, with the top hits from each identified species accompanied by a label. This circle's size is proportional to the number of reads mapping and the circle's colour indicates the ANI. The dotted line indicates 10% 5' C>T ancient damage pattern. The *H. pylori* finding is crossed with an 'x' because it was authenticated by the four parameters: coveragePRatio of >0.8, a KrakenKmerRank of 1, an aniRank100 of 1, and damage of >10%.

**Figure 4.6** shows all the microbial species identified from 011447. Circles that are opaque have a KrakenkmerRank of less than three and a coveragePRatio of greater than 0.5, with labels accompanying the top hits from each species (labelling up to two). The size of the circle is proportional to the number of reads mapped to that species and the colour indicates the ANI, where yellower colours are more likely to be authentic and bluer colours are less likely. Circles crossed with an 'x' are hits that are considered authentic based on the following four parameters: a coveragePRatio of >0.8, a KrakenKmerRank of 1, an aniRank100 of 1, and damage of >10%. The *H. pylori* finding was authenticated by all four parameters.

*H. pylori* was also identified in two other individuals, 011448 and 024098. Individual 011448 is also a Dorset, from a site approximately 120 km away from 011447 in northern Newfoundland. The 78 *H. pylori* reads isolated from 011448 had an ANI of 0.984 and an average edit distance of 0.846. Individual 024098 is from Suessland in northeast Greenland, where no other *H. pylori* was detected. The 67 *H. pylori* reads isolated from 024098 had an ANI of 0.982 and an average edit distance of 0.910.

## Human alphaherpesvirus 3

The human alphaherpesvirus 3 was identified from individual 024123, from the Pegtymel Complex from northern Chukotka, dated to 939 BP (**Figure 4.7**). The Human alphaherpesvirus 3 genome has 194 reads mapping that were isolated from 024123, with an ANI of 0.98 and an average edit distance of 0.115. The average ancient DNA damage pattern at the 5-prime end of the viral reads was 14.6%, which is comparable to the human reads with 11.5%. No other human alphaherpesvirus 3 was isolated from the ancient individuals included in this project.

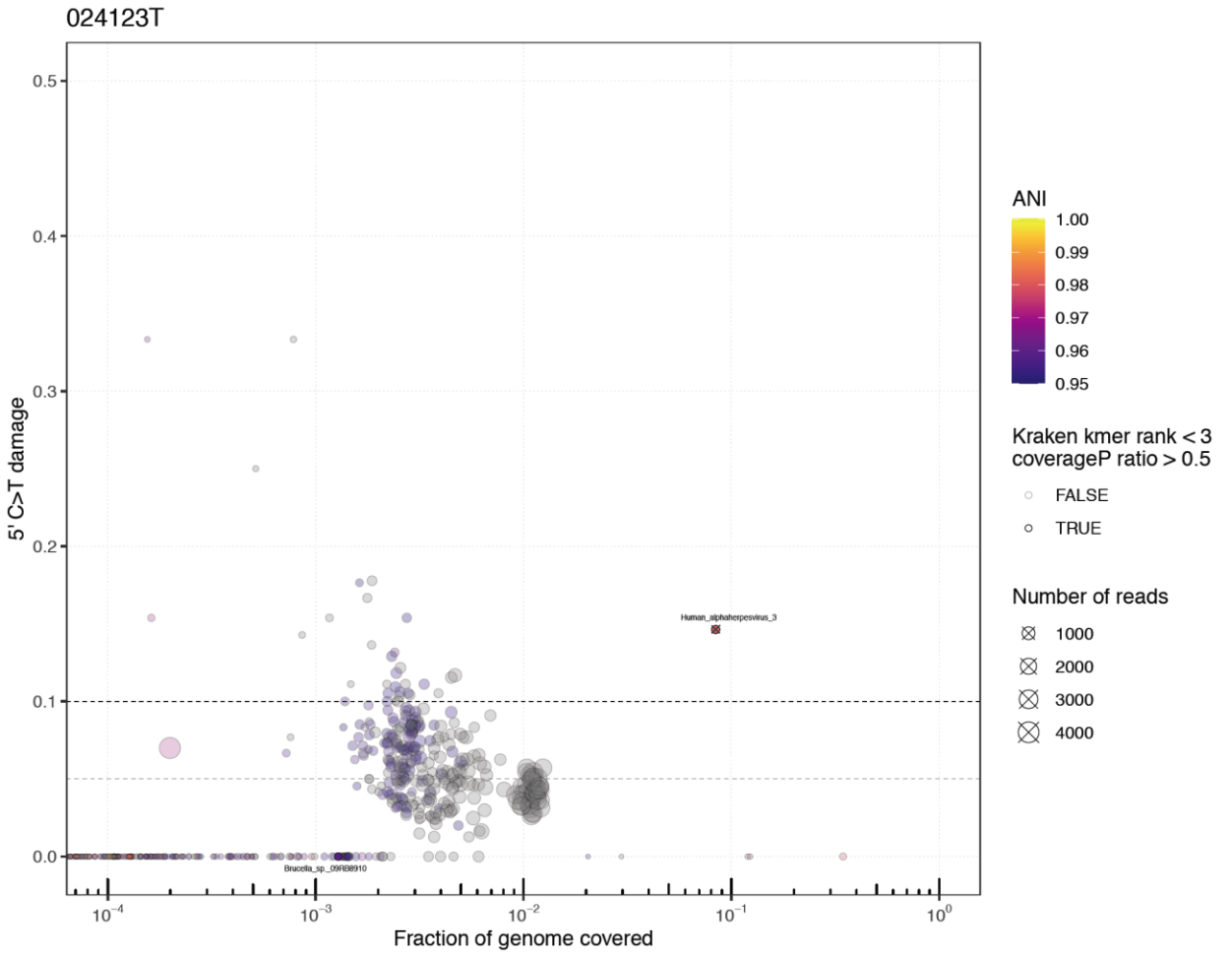



**Figure 4.7** - Plot depicting all the microbial species (including human alphaherpesvirus 3) identified from 024123, a tooth from northern Chukotka. Opaque circles indicate a coveragePRatio of >0.5, with the top hits from each identified species accompanied by a label. This circle's size is proportional to the number of reads mapping and the circle's colour indicates the ANI. The dotted line indicates 10% 5' C>T ancient damage pattern. The human alphaherpesvirus 3 finding is crossed with an 'x' because it was authenticated by the four parameters: coveragePRatio of >0.8, a KrakenkmerRank of 1, an aniRank100 of 1, and damage of >10%.

**Figure 4.7** shows all the microbial species identified from 024123. Circles that are opaque have a KrakenkmerRank of less than three and a coveragePRatio of greater than 0.5, with labels accompanying the top hits from each species (labelling up to two). The size of the circle is proportional to the number of reads mapped to that species and the colour indicates the ani, where yellower colours are more likely to be authentic and bluer colours are less likely. Circles crossed with an 'x' are hits that are considered authentic based on the following four parameters: a coveragePRatio of >0.8, a KrakenKmerRank of 1, an aniRank100 of 1, and damage of >10%. The human alphaherpesvirus 3 hit was authenticated by all four parameters.

## *4.4 Discussion*

### 4.4.1 Ancient Pathogens Identified

#### Variola virus

Smallpox is a communicable disease caused by the variola virus that results in an abrupt onset of symptoms including fever, headache, and back pain (250). The infection then causes a rash to develop in the mouth and oropharynx, followed closely by a vesicular rash on the face which spreads to the extremities and the remainder of the body. Smallpox infections have caused between 300 million and 500 million deaths in the 20$^{th}$ century and afflicted widespread mortality for centuries before (234). In 1796, Edward Jenner developed the smallpox vaccine that led to the eradication of the disease by 1980 (251).

The variola virus, and other orthopoxviruses are in the genus of poxviridae, which have double stranded DNA genomes that are large and linear, normally ranging between ~186000 and ~228000 nucleotides in length (252). The only known host for the variola virus are humans but many analogous viruses exist in other species, such as camelpox virus, taterapox virus, cowpox virus, monkeypox virus, and vaccinia virus (253,254).

The variola virus isolated from 024266 has an unexpected placement on the orthopoxvirus phylogenetic tree, as 024266 differs both geographically and temporally from the Lithuanian mummy. The clade where the variola strain from the Lithuanian mummy falls has a median root age of 372-429 years into the past (Supplement Materials, p. 108 (56)). This root age post-dates the C14 age of the individual 024266, which is 1095 BP, (or 994 Cal BP); however, there are a few factors that may mitigate this discrepancy in time. Firstly, this median root age was estimated based on a tree that did not include the variola strain from 024266, once included there would likely be a change in the date of this node. Secondly, the C14 age of 024266 will change after correction for the marine reservoir effect, making it younger (details of marine reservoir effect are outlined in Chapter 2, Future Directions Section, Marine Reservoir Correction). Lastly, it could be

that this discrepancy in timings is because of a potential change in the substitution rate (which is ~5.5 substitution sites per year) on the branch of the tree leading to the 024266 and the Lithuanian mummy strain. Similarly, this placement could be impacted because it is falling in between well calibrated genomes within the broader clade such as the historical smallpox genomes (black text in **Figure 4.4**), and the ancient Viking genomes (purple text in **Figure 4.4**) (56).

Variola genomes obtained in this project will assist in resolving the virus' evolutionary history. While beyond the scope of this thesis, this data may also lend insight to the transmission of the virus between ancient human populations.

## Helicobacter pylori

*H. pylori* bacterial infections, which are transmitted through bodily fluids, affect nearly one half of present-day humans (255). Infection leads to upper gastrointestinal tract ulcers as well as an elevated risk of developing gastric carcinoma (255).

*H. pylori* has a circular DNA genome that has an unusually high mutation and recombination rate, making it a good candidate for reconstructing human migrations (**Figure 4.8**) (256). The *H. pylori* infection has been reported to have coexisted with humans since the migrations out of Africa, originally through present-day sequences, based on the global population structure of human (genetic distance of microsatellites, uniparental markers, and autosomal *loci* between present-day human populations) and *H. pylori* (sequences from the housekeeping gene in the same population) (256,257). The coevolutionary history of humans and *H. pylori* also lends support for the proposed timings and directionality of migrations into Siberia, across the Bering Strait, and into the Americas (258). Strains of *H. pylori* isolated from present-day populations across Siberia have shown the persistence of human occupation of the Far East during the Pleistocene, as well as the recolonisation during the Holocene (258). Evidence of a *H. pylori* strain known to be carried by Indigenous people in the Americas was also found in northern Eurasian

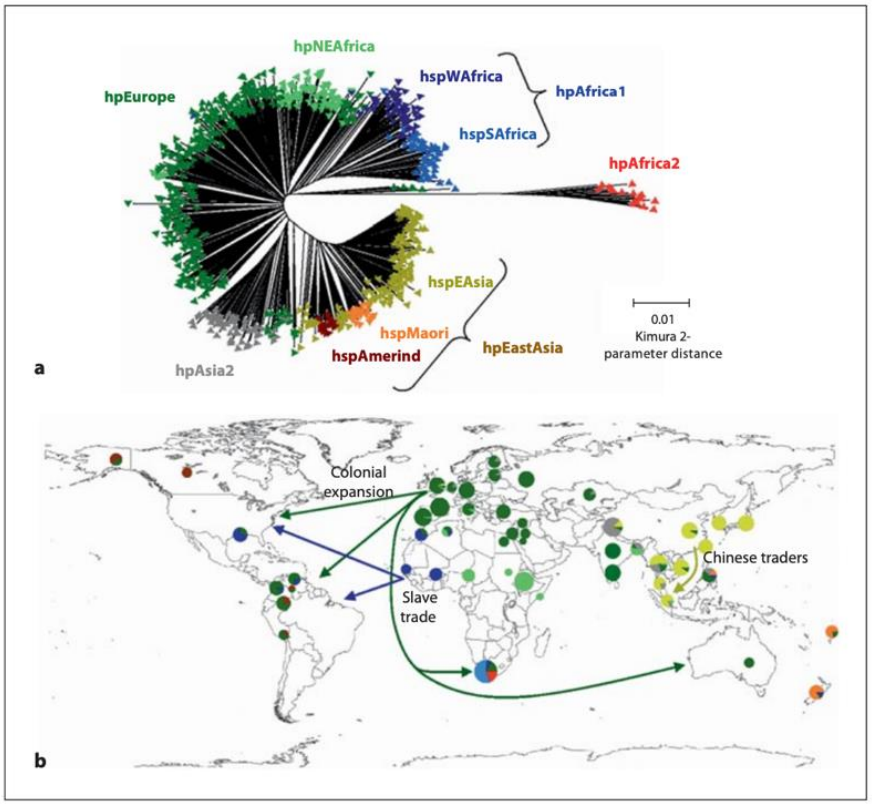



**Figure 4.8** – Map of the global distribution of *H. pylori* strains, which parallel large human migrations (Figure from Moodley et al., 2009 p. 3 (250)).

populations, and investigations into this strain supported a back migration from the Americas (258).

*H. pylori* has been found to circulate in the bloodstream; however, serum levels are not as high as other pathogenic bacteria (such as *Yersinia pestis* or *Mycobacterium tuberculosis*), resulting in no other documented recovery of *H. pylori* genomes from ancient bones or teeth (39,52,221,259). Sequences coding for biomolecules associated with *H. pylori* infections have also been isolated from stool, gastric tissues, and stomach tissue from ancient individuals from Chile, Mexico, Alaska, and South Korea; however, the only full *H. pylori* genome recovered was from a 5,300-year-old Copper Age mummy from the Italian Alps (221,260–263). The 5,300-year-old *H. Pylori* genome was an unadmixed strain of Asian origin, which differs from the modern strains in that region, that are hybrids with an African strain that arrived in Europe in the last few thousand years (221).

Additional strains for *H. pylori*, such as those identified in this research, will complement the existing ancient and present-day insights into human and *H. pylori* dispersal out of Eurasia and into the Americas.

### Human alphaherpesvirus 3

Human alphaherpesvirus 3, otherwise known as the varicella zoster virus (VZV), can manifest as chickenpox (varicella) or shingles (zoster) (264). VZV is one of nine herpes viruses known to infect humans and has a double stranded DNA genome of ~125000 bps (264,265). Infections from the VZV begin by impacting the upper respiratory epithelium before reaching the lymphatic system, where they are transported to the bloodstream resulting in a vesicular rash (264). Simultaneously, the virus enters the cell bodies of sensory nerves and establishes a latent infection which, when reactivated, can lead to zoster skin lesions (264). Contracting VZV is unlikely to result in mortality and the incidence of infections has been prevalent throughout human history; however, it is difficult to quantify the diseases impact on the human population through time as VZV was only distinguished as a separate disease to the variola virus in the 1700's (266).

Though extensive research has been conducted into the evolution and phylogeography of VZV (267), this is the first reported human alphaherpesvirus 3 isolated from ancient human remains.

## 4.4.2 Limitations

### Ascertainment Bias

There are ascertainment biases to consider with ancient pathogens studies, including the analyses described in this thesis, which fall within two broad categories - biological and analytical. Biologically, considerations for ascertainment bias should focus on elements such as: which tissues are archaeologically recovered (influenced by population density in that geographic region through time), the incidence of the pathogens in ancient populations, the preservation of the tissues of infected individuals over time (pathogens are most frequently isolated from teeth and not all skeletal remains include teeth), if the infectious agent can be preserved and isolated from the tissues being worked on (such as isolating pathogens from ancient hair), as well as if a species in soil is genetically similar enough to a species that is infectious to humans that the finding is discarded. Analytically, consideration for ascertainment bias should pertain to whether extraction

methods are optimised for microbial recovery, how exhaustive the reference databases are, and whether extinct species have modern counterparts that are genomically similar enough to be identified.

#### Tool Constraints

While computational efficiency is an advantage of the KrakenUniq software, a major limitation of this tool exists when two pathogen genomes share high levels of nucleotide identity (*e.g.*, *Yersinia pestis* and *Yersinia pseudotuberculosis*, which share ~97% average nucleotide identity in 75% of genes) (268). In such cases, the k-mers will not uniquely align to one species and will instead assign those reads at a higher taxonomic level, limiting the analytical resolution. Another disadvantage of KrakenUniq is that it does not include an authentication step, which is instead conducted in the downstream analysis.

### 4.4.3 Future Directions

#### In-House Pipeline Improvements

There are upcoming improvements to the in-house pipeline that will allow for more precise identification of authentic ancient pathogen hits. These changes include incorporating an entropy metric for more accurately assessing the distribution of reads mapping across the genome and a Bayesian estimation of ancient DNA damage patterns.

Another assessment of read distribution is to measure the entropy. The entropy of read distribution can be determined by considering the start position of each read, which is where the first base pair of the read lands on the reference genome. The entropy is calculated by the fraction of reads that have the first base pair in each 100 bp region (bins) of the reference genome. Even coverage of the genome results in high entropy and uneven coverage of the genome results in low entropy.

The ancient DNA damage pattern on pathogen reads will be calculated using the metaDMG tool, developed specially for metagenomic and environmental DNA datasets (269). This tool includes a feature which allows a model of ancient DNA damage to be fitted to the observed data, providing a Bayesian estimate to measure the statistical significance of the level of damage in the data.

## Increasing Genomic Coverage, Resolving Phylogenies, and Dating Divergence Times

To increase the genomic coverage of the pathogens identified in this project, tissues, DNA extracts, genomic libraries, and existing USER-treated libraries will be reprocessed for additional sequencing or pathogen capture (270). Increased genomic coverage will allow for greater certainty in the variant calls, resulting in more accurate placement on a phylogenetic tree. Additional genomic resolution would help to better estimate divergence times, substitution rates, and pathogen adaptation using bioinformatic tools such as BEAST and SnpEff (271,272).

## Pipeline for Pathogen Identification from Hair

This project contains the largest dataset of ancient genomes extracted from human hair (n = 93) reported in the ancient genomics literature thus far. This thesis is focused mainly on pathogens extracted from teeth, as this is where the highest incidence of pathogen hits are recovered, but in future we will shift the focus to pathogens that can be carried in hair, such as the human polyomavirus, which is known to be carried by head lice and is common in human hair follicles. Additional considerations will also be implemented, such as including additional pathogenic species that are known to adhere to hair, fur, or fabrics.

## Applications of Findings

The pathogen genomes, which will increase in coverage after additional sequencing, will aid in reconstructing the evolutionary trajectories of different pathogenic species, lending insights into adaptations that may impact virulence or transmissibility and resistance. Other applications of

these findings extend to using intersections in pathogen load between populations as a proxy for population contact, which could supplement analyses where contact is proposed with limited support of genetic evidence. For example, findings of pathogens unique to incoming Neo-Inuit groups in resident Palaeo-Inuit groups after the archaeologically known date of the arrival of the former in the North American Arctic, may signal some level of direct contact between these groups (2,6,33,66).

Another future direction of this research is to complement the pathogen catalogue with selection scans on the human host genome (as outlined in Chapter 3, Selection Scans), which could indicate correlations between the incidence of specific pathogens in a population and selection for immunological resistance. Previous studies have shown that European diseases would have exerted novel selective pressures on the immune genes of Indigenous groups in the Americas, which would have left signatures on their genomes through selection for advantageous variants (*e.g.,* those conferring immunity to new diseases) or through the purging/reduction in frequency of disadvantageous/neutral variants (*e.g*., those that were beneficial before contact but not against the onslaught of novel pathogens). One study showed that genomes from the north-west of North America show the strongest selection signal at the *HLA-DQA1* locus (gene involved in antigen presentation in the adaptive immune system), with alleles that were close to fixation in the past but underwent a frequency decrease of 64% in comparative present-day populations (273). This selection is attributed to European colonisation changing the endogenous environment and ensuing selective pressures, negative in this case, on existing genetic traits and variation (273).

## *4.5 Conclusion*

An in-house, high-throughput pathogen pipeline was used to screen genomic libraries built from DNA extracted from 217 ancient human tissues for various human pathogens. The high depth sequencing runs outlined in Chapter 2 of this thesis were entered into the pipeline and the output was sorted and examined for authenticity. The authentic pathogens identified were then catalogued in a brief overview of human pathogens in the Arctic over time. Three pathogens (variola virus, *Helicobacter Pylori*, and human alphaherpesvirus 3) were further investigated - and in the case of the variola virus, placed onto a phylogenetic tree. These results demonstrate the contributions of this dataset to the growing database of ancient pathogens and show that the authenticated findings can be used for future work that aims to correlate the major changes in pathogen species composition in Arctic populations with known events such as migrations, population contact, climate change, and more.

# 5 Chapter 5 - Conclusion

Generating this ancient DNA dataset addresses gaps in the existing palaeogenomics knowledge base. This was done by increasing the data quality of the previously investigated individuals, by increasing the number of ancient genomes available from both the North American Arctic and potential source populations from Siberia, and by strengthening existing relationships with Indigenous groups.

Population genetic analyses were conducted, using population-wide low coverage and select higher coverage genomes from ancient Arctic and sub-Arctic groups to investigate population origins, population structure, and the timing and magnitude of admixture. Three main research questions pertaining to the demographic history of specific groups were addressed in this work, 1) the genetic characterization of the geographically isolated and archaeologically unique Zhokhov individuals from the New Siberian Islands; 2) examining the Palaeo-Inuit populations and how they relate to the contemporary Saqqaq, past Neo-Inuit, and present-day Inuit populations; 3) investigating the formation of the Neo-Inuit populations and their relation to each other.

Multiple lines of population genetics analyses found the geographically isolated and archaeologically unique Zhokhov individuals from the New Siberian Islands to be most similar to the 9.8 kya ancient Kolyma individual from northern Yakutia (25). This genetic grouping, termed the Palaeo-Siberians, indicates at least 2000 years of genetic continuity in that region. This finding is in keeping with the impending Holocene Climate Optimum, when the mammoth steppe was evolving into a tundra landscape offering refuge for the hunter-gather groups of Palaeo-Siberians (205). While the Zhokhov individuals remained geographically isolated, leading to their ultimate demise, Kolyma-related ancestry contributed to successive population groups in northeastern Siberia and the North American Arctic while admixing with incoming groups with ancient East Asian-related ancestry. These findings are consistent with those describe a previous work, where some of these ancient Zhokhov individuals had their genomes sequenced to very low coverage (274).

Investigating the Palaeo-Inuit genetic affinities, that is the Canadian Pre-Dorset and Dorset individuals as well as the ancient Saqqaq individual from Greenland, suggests that they form a clade to the exclusion of all other groups from the Americas and Siberia included in the population genetics analyses. These results are consistent with broader genetic continuity in Paleo-Inuit populations of the North American Arctic for at least 2500 years, which was proposed (upwards of 4000 years) in the Raghavan et al., 2014 publication (2). Exploratory analyses of the Palaeo-Inuit population indicates that they did not admix with NNA populations. The Palaeo-Inuit were also found to have genetic similarity to the Palaeo-Siberians, but possess an additional ancient East-Asian-related lineage that was likely obtained from an admixture event that occurred in northeastern Siberia prior to the Palaeo-Inuit expansion into the North American Arctic. Though there is a high level of shared drift between the Palaeo-Inuit and the succeeding Neo-Inuit populations, as well as present-day Inuit populations, these successive populations differ from the former by possessing a substantial genetic component related to NNA.

Examining genetic data from the Neo-Inuit populations, that is the Siberian Birnirk and Old Bering Sea groups, the Alaskan Yupik, and the Greenlandic Thule, suggests that they are most closely related to each other and other present-day Inuit groups, followed by the Palaeo-Inuit and Native American population groups. This finding is in keeping with the results of the Neo-Inuit (low depth genomes) and present-day Inuit (high depth genomes) included in Raghavan et al., 2014, where the Neo-Inuit are found to be the direct genetic ancestor of the present-day Inuit (2). Oral histories and the archeological record are also consistent with this line of evidence (31,102,206–208). Furthermore, population genetic analyses showed differing proportions of relatedness of the Neo-Inuit groups to other Arctic population groups, where the Alaskan Yupik were found to be most similar to non-Neo-Inuit Arctic groups, followed by the Siberian Old Bering Sea and Birnirk groups, and the Greenlandic Thule being most drifted. The Canadian Dorset and Pre-Dorset showed more genetic similarity to the Neo-Inuit groups than to the Greenlandic Saqqaq, suggesting that the Dorset/Pre-Dorset are more likely to have contributed to the formation of the Neo-Inuit when compared to the Saqqaq. This result is supported by the spatio-temporal distribution of the ancient groups, where the Dorset groups survived long enough to possibly have contributed to the formation of the Neo-Inuit groups (estimated to have taken place between 74-102 generations ago in the DATES analysis), whereas the Saqqaq+Pre-Dorset had already

undergone demise as a cultural entity (30). Further analyses indicate that most Neo-Inuit groups share more genetic drift with Kolyma than with the Zhokhov individuals, with the rare exception of the Birnirk who have more Zhokhov-related ancestry.

The genomes of these ancient individuals were also investigated for the putative effects of environmental pressures, especially transitioning from terrestrial to marine subsistence strategies, on the distribution of genetic variants pertaining to adaptations to survive in the harsh cold climate of the Arctic landscape. Changes in allele frequency in the genomes of ancient and present-day Arctic populations (and some sub-Arctic populations) were observed in genes related to fat metabolism and fat distribution. These findings support previous reports of sites that segregate in Arctic populations and serve as a proof of principle for the newly developed allele frequency-based approach to visualising changes in SNPs over time (8–10,37,185).

These population genetic and adaptation findings will undergo further investigation once the ancient genomes generated in this work undergo imputation for increased data quality. This increased data quality will allow for improved resolution of the admixture events that took place that resulted in the formation of the Neo-Inuit groups. Similarly, the imputed dataset will also lend itself to selection scans, which may resolve the timing and magnitude of the changes in allele frequency as well as potentially identifying additional variants under selection in these Arctic and sub-Arctic populations.

Ancient pathogens were identified as a by-product of generating high quality human genomes and broadened the current ancient pathogen catalogue from Siberia and the North American Arctic. This catalogue will be expanded by running the USER-treated data through the pathogen pipeline, as well as sending the libraries with positive hits for additional sequencing or pathogen capture. Increasing pathogen coverage will resolve the evolutionary trajectories of, and can complement established changes in, species composition found in Arctic and sub-Arctic populations and may provide support for known events such as migrations, contact between populations, and changes in climate.

To expand the potential of the existing dataset, additional supporting evidence will be collected, which are currently beyond the scope of this thesis. This entails the compound specific isotopic analyses to further elucidate the dietary components of these ancient individuals, correcting the calibrated C14 dates of these individuals to better establish the temporal distribution of these genomes, determining relatedness between individuals from the same site or region to establish kinship practices, and communicating with Indigenous groups on these findings to gain a more complete understanding of the history of these ancient Arctic and sub-Arctic groups.

## *Bibliography*


1. Parrott, Z. Indigenous Peoples in Canada [Internet]. Historica Canada; 2023. (The Canadian Encyclopedia). Available from: https://www.thecanadianencyclopedia.ca/en/article/aboriginal-people

2. Raghavan M, DeGiorgio M, Albrechtsen A, Moltke I, Skoglund P, Korneliussen TS, et al. The genetic prehistory of the New World Arctic. Science. 2014 Aug 29;345(6200):1255832.

3. Russian Association of Indigenous Peoples of the North (RAIPON). Association of Indigenous Peoples of the North, Siberia and the Far East [Internet]. 2021. Available from: https://raipon.info/narody/

4. Friesen TM, Mason OK, editors. The Oxford handbook of the prehistoric Arctic. New York: Oxford University Press; 2016. 988 p. (The Oxford handbooks series).

5. Tishkov, V. A., Zhuravsky, A. V. Peoples of Russia: Atlas of Cultures and Religions. Pritvorov, A.P., editor. Moscow; 2008.

6. Friesen TM, Arnold CD. The Timing of the Thule Migration: New Dates from the Western Canadian Arctic. Am Antiq. 2008 Jul;73(3):527–38.

7. Moltke I, Fumagalli M, Korneliussen TS, Crawford JE, Bjerregaard P, Jørgensen ME, et al. Uncovering the Genetic History of the Present-Day Greenlandic Population. Am J Hum Genet. 2015 Jan;96(1):54–69.

8. Fumagalli M, Moltke I, Grarup N, Racimo F, Bjerregaard P, Jørgensen ME, et al. Greenlandic Inuit show genetic signatures of diet and climate adaptation. Science. 2015 Sep 18;349(6254):1343–7.

9. Zhou S, Xie P, Quoibion A, Ambalavanan A, Dionne-Laporte A, Spiegelman D, et al. Genetic architecture and adaptations of Nunavik Inuit. Proc Natl Acad Sci. 2019 Aug 6;116(32):16012–7.

10. Hallmark B, Karafet TM, Hsieh P, Osipova LP, Watkins JC, Hammer MF. Genomic Evidence of Local Adaptation to Climate and Diet in Indigenous Siberians. Hernandez R, editor. Mol Biol Evol. 2019 Feb 1;36(2):315–27.

11. Norsk Polarinstitutt. Indigenous Peoples of the Arctic Countries [Internet]. Norwegian Polar Institute; 2007 [cited 2023 Feb 15]. Available from: https://www.npolar.no/ansipra/image/Arctic04E.jpg

12. Romanova NA, Zhirnov AB, Yust NA, Fucheng X. Influence of forest growth conditions on the density of wood in the Amur region. Cent Eur For J. 2019 Mar 1;65(1):41–50.

13. Kardash OV, Girchenko EA. Resettlement and Adaptation of Asian Tribes on the Territory of Northern Eurasia. Adv Appl Sociol. 2018;08(06):486–93.

14. Alaska Native Heritage Center. Traditional territories of Alaska Native Cultures [Internet]. Anchorage, AK: Alaska Native Heritage Center; 2023 [cited 2023 Feb 21]. Available from: https://www.alaskanative.net/cultural-knowledge/

15. Rasmussen, Rasmus Ole. Encyclopedia Britannica [Internet]. 2022 [cited 2023 Feb 2]. (Greenland). Available from: https://www.britannica.com/place/Greenland

16. Skoglund P, Mathieson I. Ancient Genomics of Modern Humans: The First Decade. Annu Rev Genomics Hum Genet. 2018 Aug 31;19(1):381–404.

17. Orlando L, Allaby R, Skoglund P, Der Sarkissian C, Stockhammer PW, Ávila-Arcos MC, et al. Ancient DNA analysis. Nat Rev Methods Primer. 2021 Feb 11;1(1):14.

18. Speidel L, Cassidy L, Davies RW, Hellenthal G, Skoglund P, Myers SR. Inferring Population Histories for Ancient Genomes Using Genome-Wide Genealogies. Harris K, editor. Mol Biol Evol. 2021 Aug 23;38(9):3497–511.

19. Hajdinjak M, Mafessoni F, Skov L, Vernot B, Hübner A, Fu Q, et al. Initial Upper Palaeolithic humans in Europe had recent Neanderthal ancestry. Nature. 2021 Apr 8;592(7853):253–7.

20. Rasmussen M, Sikora M, Albrechtsen A, Korneliussen TS, Moreno-Mayar JV, Poznik GD, et al. The ancestry and affiliations of Kennewick Man. Nature. 2015 Jul;523(7561):455–8.

21. Malaspinas AS, Lao O, Schroeder H, Rasmussen M, Raghavan M, Moltke I, et al. Two ancient human genomes reveal Polynesian ancestry among the indigenous Botocudos of Brazil. Curr Biol. 2014 Nov;24(21):R1035–7.

22. Moreno-Mayar JV, Vinner L, de Barros Damgaard P, de la Fuente C, Chan J, Spence JP, et al. Early human dispersals within the Americas. Science. 2018 Dec 7;362(6419):eaav2621.

23. Allentoft ME, Sikora M, Sjögren KG, Rasmussen S, Rasmussen M, Stenderup J, et al. Population genomics of Bronze Age Eurasia. Nature. 2015 Jun 11;522(7555):167–72.

24. Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, et al. Massive migration from the steppe was a source for Indo-European languages in Europe. Nature. 2015 Jun 11;522(7555):207–11.

25. Sikora M, Pitulko VV, Sousa VC, Allentoft ME, Vinner L, Rasmussen S, et al. The population history of northeastern Siberia since the Pleistocene. Nature. 2019 Jun;570(7760):182–8.

26. Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I, et al. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature. 2014 Jan;505(7481):87–91.

27. Moreno-Mayar JV, Potter BA, Vinner L, Steinrücken M, Rasmussen S, Terhorst J, et al. Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. Nature. 2018 Jan;553(7687):203–7.

28. Willerslev E, Meltzer DJ. Peopling of the Americas as inferred from ancient genomics. Nature. 2021 Jun 17;594(7863):356–64.

29. Tamm E, Kivisild T, Reidla M, Metspalu M, Smith DG, Mulligan CJ, et al. Beringian Standstill and Spread of Native American Founders. Carter D, editor. PLoS ONE. 2007 Sep 5;2(9):e829.

30. Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, Moltke I, et al. Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature. 2010 Feb;463(7282):757–62.

31. Avataq Cultural Institute. Discovering Archaeology. 2015. Arctic Chronology. Available from: http://www.avataq.qc.ca/en/Institute/Departments/Archaeology/Discovering-Archaeology/Arctic-Chronology

32. Siska V, Jones ER, Jeon S, Bhak Y, Kim HM, Cho YS, et al. Genome-wide data from two early Neolithic East Asian individuals dating to 7700 years ago. Sci Adv. 2017 Feb 3;3(2):e1601877.

33. Appelt M, editor. Identities and cultural contacts in the Arctic: proceedings from a Conference at the Danish National Museum, Copenhagen, November 30 to December 2, 1999. Copenhagen: The Danish National Museum & Danish Polar Center; 2000. 267 p. (Danish Polar Center Publication).

34. Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, et al. Genome-wide patterns of selection in 230 ancient Eurasians. Nature. 2015 Dec 24;528(7583):499–503.

35. Fumagalli M, Sironi M, Pozzoli U, Ferrer-Admettla A, Pattini L, Nielsen R. Signatures of Environmental Genetic Adaptation Pinpoint Pathogens as the Main Selective Pressure through Human Evolution. Akey JM, editor. PLoS Genet. 2011 Nov 3;7(11):e1002355.

36. Desjardins SPA, Jordan PD. Arctic Archaeology and Climate Change. Annu Rev Anthropol. 2019 Oct 21;48(1):279–96.

37. Clemente FJ, Cardona A, Inchley CE, Peter BM, Jacobs G, Pagani L, et al. A Selective Sweep on a Deleterious Mutation in CPT1A in Arctic Populations. Am J Hum Genet. 2014 Nov;95(5):584–9.

38. Raoult D, Drancourt M, editors. Paleomicrobiology: Past Human Infections [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2008 [cited 2023 Feb 9]. Available from: http://link.springer.com/10.1007/978-3-540-75855-6

39. Spyrou MA, Bos KI, Herbig A, Krause J. Ancient pathogen genomics as an emerging tool for infectious disease research. Nat Rev Genet. 2019 Jun;20(6):323–40.

40. Hansen HB, Damgaard PB, Margaryan A, Stenderup J, Lynnerup N, Willerslev E, et al. Comparing Ancient DNA Preservation in Petrous Bone and Tooth Cementum. Caramelli D, editor. PLOS ONE. 2017 Jan 27;12(1):e0170940.

41. Damgaard PB, Margaryan A, Schroeder H, Orlando L, Willerslev E, Allentoft ME. Improving access to endogenous DNA in ancient bones and teeth. Sci Rep. 2015 Jun 17;5(1):11184.

42. Margaryan A, Hansen HB, Rasmussen S, Sikora M, Moiseyev V, Khoklov A, et al. Ancient pathogen DNA in human teeth and petrous bones. Ecol Evol. 2018 Mar;8(6):3534–42.

43. Warinner C, Herbig A, Mann A, Fellows Yates JA, Weiß CL, Burbano HA, et al. A Robust Framework for Microbial Archaeology. Annu Rev Genomics Hum Genet. 2017 Aug 31;18(1):321–56.

44. Marciniak S, Poinar HN. Ancient Pathogens Through Human History: A Paleogenomic Perspective. In: Lindqvist C, Rajora OP, editors. Paleogenomics [Internet]. Cham: Springer International Publishing; 2018 [cited 2023 Feb 9]. p. 115–38. (Population Genomics). Available from: http://link.springer.com/10.1007/13836_2018_52

45. Bos KI, Schuenemann VJ, Golding GB, Burbano HA, Waglechner N, Coombes BK, et al. A draft genome of Yersinia pestis from victims of the Black Death. Nature. 2011 Oct 27;478(7370):506–10.

46. Taubenberger JK, Reid AH, Krafft AE, Bijwaard KE, Fanning TG. Initial Genetic Characterization of the 1918 "Spanish" Influenza Virus. Science. 1997 Mar 21;275(5307):1793–6.

47. Darling MI, Donoghue HD. Insights from paleomicrobiology into the indigenous peoples of pre-colonial America - A Review. Mem Inst Oswaldo Cruz. 2014 Apr;109(2):131–9.

48. Podmaskin, V. V. Medico-demographic consequences of ethnocultural contacts of the indigenous peoples of the Far East of Russia. Bull Far East Branch Russ Acad Sci. 2012;(2):162.

49. Mühlemann B, Jones TC, Damgaard P de B, Allentoft ME, Shevnina I, Logvin A, et al. Ancient hepatitis B viruses from the Bronze Age to the Medieval period. Nature. 2018 May 17;557(7705):418–23.

50. Rascovan N, Sjögren KG, Kristiansen K, Nielsen R, Willerslev E, Desnues C, et al. Emergence and Spread of Basal Lineages of Yersinia pestis during the Neolithic Decline. Cell. 2019 Jan;176(1–2):295-305.e10.

51. Rasmussen S, Allentoft ME, Nielsen K, Orlando L, Sikora M, Sjögren KG, et al. Early Divergent Strains of Yersinia pestis in Eurasia 5,000 Years Ago. Cell. 2015 Oct;163(3):571–82.

52. Bos KI, Harkins KM, Herbig A, Coscolla M, Weber N, Comas I, et al. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. Nature. 2014 Oct;514(7523):494–7.

53. Kahila Bar-Gal G, Kim MJ, Klein A, Shin DH, Oh CS, Kim JW, et al. Tracing hepatitis B virus to the 16th century in a Korean mummy. Hepatology. 2012 Nov;56(5):1671–80.

54. Mühlemann B, Margaryan A, Damgaard P de B, Allentoft ME, Vinner L, Hansen AJ, et al. Ancient human parvovirus B19 in Eurasia reveals its long-term association with humans. Proc Natl Acad Sci. 2018 Jul 17;115(29):7557–62.

55. Biagini P, Thèves C, Balaresque P, Géraut A, Cannet C, Keyser C, et al. Variola Virus in a 300-Year-Old Siberian Mummy. N Engl J Med. 2012 Nov 22;367(21):2057–9.

56. Mühlemann B, Vinner L, Margaryan A, Wilhelmson H, de la Fuente Castro C, Allentoft ME, et al. Diverse variola virus (smallpox) strains were widespread in northern Europe in the Viking Age. Science. 2020 Jul 24;369(6502):eaaw8977.

57. Collen EJ, Johar AS, Teixeira JC, Llamas B. The immunogenetic impact of European colonization in the Americas. Front Genet. 2022 Aug 5;13:918227.

58. Dudarev AlexeyA, Chupakhin ValeryS, Øyvind Odland J. Health and society in Chukotka: an overview. Int J Circumpolar Health. 2013 Jan 31;72(1):20469.

59. Grygier PS. A Long Way from Home: The Tuberculosis Epidemic Among the Inuit [Internet]. McGill-Queen's University Press; 1997. (McGill-Queen'S/Hannah Institute Studies in the History of Medicine, Health and Society, 2). Available from: https://books.google.co.uk/books?id=Nc1hTqNUq58C

60. Bjerregaard P. Infectious Diseases in Greenlanders of Upernavik. Scand J Prim Health Care. 1985 Jan;3(3):163–9.

61. Petersen GM, Silimperi DR, Chiu CY, Ward JI. Effects of Age, Breast Feeding, and Household Structure on Haemophilus influenzae Type b Disease Risk and Antibody Acquisition in Alaskan Eskimos. Am J Epidemiol. 1991 Nov 15;134(10):1212–21.

62. Wainwright RB, Bulkow LR, Parkinson AJ, Zanis C, McMahon BJ. Protection Provided by Hepatitis B Vaccine in a Yupik Eskimo Population--Results of a 10-Year Study. J Infect Dis. 1997 Mar 1;175(3):674–7.

63. Toomey KE. Unrecognized High Prevalence of Chlamydia trachomatis Cervical Infection in an Isolated Alaskan Eskimo Population. JAMA J Am Med Assoc. 1987 Jul 3;258(1):53.

64. Shaffer N, Wainwright RB, Middaugh JP, Tauxe RV. Botulism among Alaska Natives. The role of changing food preparation and consumption practices. West J Med. 1990 Oct;153(4):390–3.

65. Proulx J, MacLean JD, Gyorkos TW, Leclair D, Richter A, Serhir B, et al. Novel Prevention Program for Trichinellosis in Inuit Communities. Clin Infect Dis. 2002 Jun;34(11):1508–14.

66. Guzmán-Solís AA, Villa-Islas V, Bravo-López MJ, Sandoval-Velasco M, Wesp JK, Gómez-Valdés JA, et al. Ancient viral genomes reveal introduction of human pathogenic viruses into Mexico during the transatlantic slave trade. eLife. 2021 Aug 5;10:e68612.

67. Google Maps. Google Earth Pro [Internet]. Keyhole, Inc.; 2015 [cited 2023 Feb 15]. (Image IBCAO). Available from: https://earth.google.com/web/

68. Grebenyuk PS, Fedorchenko AYu, Dyakonov VM, Lebedintsev AI, Malyarchuk BA. Ancient Cultures and Migrations in Northeastern Siberia. In: Bocharnikov VN, Steblyanskaya AN, editors. Humans in the Siberian Landscapes [Internet]. Cham: Springer International Publishing; 2022 [cited 2023 Feb 8]. p. 89–133. (Springer Geography). Available from: https://link.springer.com/10.1007/978-3-030-90061-8_4

69. Pitul'ko V. Terminal Pleistocene—Early Holocene occupation in northeast Asia and the Zhokhov assemblage. Quat Sci Rev. 2001 Jan;20(1–3):267–75.

70. Pitulko VV. Geoarchaeology, Age and Chronology of the Zhokhov Site. Institute of the History of Material Culture, Russian Academy of Sciences, Peter the Great Museum of Anthropology and Ethnography (Kunstkamera), Russian Academy of Sciences, editors. Vestn St Petersburg Univ Hist. 2022;67(4):1253–95.

71. Pitulko VV, Kasparov AK. Archaeological dogs from the Early Holocene Zhokhov site in the Eastern Siberian Arctic. J Archaeol Sci Rep. 2017 Jun;13:491–515.

72. Lebedint͡sev AI. Early Maritime cultures of North-Western Priokhotye/Drevnie primorskie kul′tury Severo-Zapadnogo Priokhot′i͡a. Leningrad: "Nauka," Leningradskoe otd-nie; 1990. 258 p.

73. Lebedintsev AI, Bland RL, Gokhman A, Workman W, Workman K. Maritime Cultures of the North Coast of the Sea of Okhotsk. Arct Anthropol. 1998;35(1):296–320.

74. Arutiunov, S. A. Problems of ethnic history in the Bering Sea : the Ekven Cemetery [Internet]. National Park Service, Shared Beringian Heritage Program; 1975. Available from: https://search.library.wisc.edu/catalog/9910062441002121

75. Ponomarenko, A. K. Ancient Culture of the Itelmen of Eastern Kamchatka/Drevni͡a͡ia Kul′tura Itel′menov Vostochnoĭ Kamchatki. Moscow: Nauka; 1985.

76. Dikov, N. N. Early Cultures of Northeastern Asia. /Drevnie kul′tury Severo-Vostochnoĭ Azii. Vol. 331. U.S. Dept. of the Interior, National Park Service, Shared Beringian Heritage Program, Anchorage, Alaska; 2004.

77. Ptashinskiy, A. V. Culture of marine hunters in the Northwestern of Kamchatka. In:Lebedintsev AI (ed) Researches on archaeology of the North of the Far East. Magadan: NEISRI FEB RAS Press; 1999.

78. Dikov, N. N. Archaeological sites of Kamchatka, Chukotka, and the Upper Kolyma [Internet]. Anchorage, Alaska : U.S. Dept. of the Interior, National Park Service, Shared Beringian Heritage Program, 2003.; 2003. Available from: https://search.library.wisc.edu/catalog/9910376807202121

79. Dikov, N. N. Paleolithic of Kamchatka and Chukotka in in connection with the problems of initial peopling of America/Paleolit Kamchatki i Chukotki v sviazi s problemoi pervonachal'nogo zaseleniia Ameriki. Ross Akad Nauk Dal'nevostochnoe Otd Sev-Vostochnyi Kompleks Nauchno-Issledovatel'skii Inst Magadan Russ. 1993;

80. Goebel T, Waters MR, Dikova M. The Archaeology of Ushki Lake, Kamchatka, and the Pleistocene Peopling of the Americas. Science. 2003 Jul 25;301(5632):501–5.

81. Dikova, T. M. Archaeology of southern Kamchatka in connection with the Ainu occupation problem./Arkheologiia iuzhnoĭ Kamchatki v sviazi s Problemoĭ rasseleniia Aĭnov. Moscow: Nauka; 1983.

82. Vasil′evskiĭ, R. S. Origin and ancient culture of the Koryaks./ Proiskhozhdenie i drevniaia kul′tura Koriakov. Vol. 252. Novosibirsk: Nauka; 1971.

83. Bland RL. Another Look at the Pegtymel' Petroglyphs. Arct Anthropol. 2010;47(2):22–31.

84. Burykin, A. A. The role of the toponymic study for archaeological and historical-ethnographic research. Dikov Readings (Dikovckie qteni). Lebedintsev, A. I., editor. Magadan: NEISRI FEB RAS Press; 2001.

85. Burykin, A. A. Migration from east to west and the problems of the ethno-cultural history of the peoples of the coast of the Arctic Ocean. Chairkina, N. M., editor. Khanty-Mansiysk, Yekaterinburg; 2002. (Northern Archaeological Congress: abstracts.).

86. Dikov, N. N. The history of Chukotka from ancient times to the present day./ Istoriia Chukotki s drevneĭshikh vremen do nashikh dneĭ. Moscow: Mysl′; 1989. 492 p.

87. Dikov, N. N. The main contours of the ethnic history of Northeast Siberia in the light of archaeological data. Derevyanko, A. P., editor. Novosibirsk: SB AS USSR Press; 1973. (Problems of ethnogenesis of the peoples of Siberia and the Far East.).

88. Arutyunov, S. A., Sergeev, D. A. Ancient cultures of Asian Eskimos (Uelen cemetery)/Древние культуры азиатских эскимосов: (Узленский могильнык). Moscow: Nauka,; 1969. 206 p.

89. Mason OK, Rasic JT. Walrusing, whaling and the origins of the Old Bering Sea culture. World Archaeol. 2019 May 27;51(3):454–83.

90. Dneprovsky, K. A Late Birnirk House at Paipelghak in Northern Chukotka: A Preliminary Report based on the Excavations from 2002-2004. Alsk J Anthropol. 2006;4:34–53.

91. Britton K, Knecht R, Nehlich O, Hillerdal C, Davis RS, Richards MP. Maritime adaptations and dietary variation in prehistoric Western Alaska: Stable isotope analysis of permafrost-preserved human hair: Diet in Prehistoric Western Alaska. Am J Phys Anthropol. 2013 Jul;151(3):448–61.

92. Britton K, McManus-Fry E, Nehlich O, Richards M, Ledger PM, Knecht R. Stable carbon, nitrogen and sulphur isotope analysis of permafrost preserved human hair from rescue excavations (2009, 2010) at the precontact site of Nunalleq, Alaska. J Archaeol Sci Rep. 2018 Feb;17:950–63.

93. Helmer, J. W., Kennedy, B. V. Early Palaeo-Eskimo Skeletal Remains from North Devon Island, High Arctic Canada. Can J Archaeol J Can Archéologie. 1986;10:127–43.

94. Helmer JW. The Palaeo-Eskimo Prehistory of the North Devon Lowlands. ARCTIC. 1991 Jan 1;44(4):301–17.

95. Renouf >M. A. P. Prehistory of Newfoundlandhunter-gatherers: Extinctions or adaptations? World Archaeol. 1999 Feb;30(3):403–20.

96. Brown, S. Archaeological Investigations at Crow Head Cave and the Gargamelle Rockshelter in the Port aux Choix National Historical Park, Newfoundland. Halifax, Nova Scotia: Parks Canada; 1988.

97. Renouf MAP, editor. The cultural landscapes of Port Au Choix: precontact hunter-gatherers of northwestern Newfoundland. New York: Springer; 2011. 321 p. (Interdisciplinary contributions to archaeology).

98. Harp, E. J., Hughes, D. R. Five Prehistoric Burials from Port Aux Choix, Newfoundland. Polar Notes. 1968;8:1–47.

99. Meldgaard, J. Origin and evolution of Eskimo culture in the eastern Arctic. Can Geogr J. 1960;60:64–75.

100. Lynnerup N, Meldgaard J, Jakobsen J, Appelt M, Koch A, Frøhlich B. Human Dorset Remains from Igloolik, Canada. ARCTIC. 2003 Jan 1;56(4):349–58.

101. Taylor, Jr. W. E. The Arnapik and Tyara Sites. An Archaeological Study of Dorset Culture Origins. VanStone JW, editor. ARCTIC. 1971;24(4):311–2.

102. Grønnow, B. Dynamics of northern societies: proceedings of the SILA/NABO Conference on Arctic and North Atlantic Archaeology, Copenhagen, May 10th-14th, 2004. Arneborg J, Greenland Research Centre (Nationalmuseet : Denmark), North Atlantic Biocultural Organisation, Nationalmuseet (Denmark), editors. Copenhagen: National Museum of Denmark; 2006. 415 p. (Publications from the National Museum. Studies in archaeology & history).

103. Mary-Rousselière, G. Nunguvik et Saatut : site paléoeskimaux de Navy Board Inlet, île de Baffin. Hull, Quebec: Canadian Museum of Civilization; 2002. (Archaeological Survey of Canada).

104. Taylor, Jr. W. E. Archaeological Work in Ungava and Mansel Island. Arct Circ. 1959;11:66–7.

105. Williamson, R. The Archaeological history of the Wendat to AD 1651: An overview. Ont Archaeol. 2014;(94):3–64.

106. Dewar G, Ginter JK, Shook BAS, Ferris N, Henderson H. A bioarchaeological study of a Western Basin tradition cemetery on the Detroit River. J Archaeol Sci. 2010 Sep;37(9):2245–54.

107. Scheib CL, Li H, Desai T, Link V, Kendall C, Dewar G, et al. Ancient human parallel lineages within North America contributed to a coastal expansion. Science. 2018 Jun;360(6392):1024–7.

108. Turnbull, C. J. Old Mission Point 1973: Report for an Archaeological Survey of Canada Salvage Contract. Ottawa; 1974. (Archaeological Survey of Canada).

109. Garlie, T. N. An Ethnohistorical and Archaeological Review regarding Aboriginal Mortuary Remains reported from Nova Scotia and New Brunswick and the Potential for Future Research. Unpublished Honours Essay, Memorial University; 1992.

110. Gulløv HC. The Nature of Contact between Native Greenlanders and Norse. J N Atl. 2008 Jul;1:16–24.

111. Gulløv, H. C., Appelt, M. Tunit, Norsemen and Inuit in Thirteenth-Century Northwest Greenland: Dorset between the Devil and the Deep Sea. In: The Northern World, AD 900-1400. Salt Lake City, Utah: University of Utah Press; 2009. p. 300–20.

112. Mathiassen T. Eskimo Migrations in Greenland. Geogr Rev. 1935 Jul;25(3):408.

113. Holtved, E. Archaeological investigations in the Thule district. København: C.A. Reitzel; 1944. (Meddelelser om Grønland).

114. Fletcher A, Antoine D, Hill JD. Regarding the dead: human remains in the British Museum. London: British museum; 2014. (The British Museum Research Publication).

115. Whitt L. Science, Colonialism, and Indigenous Peoples: The Cultural Politics of Law and Knowledge [Internet]. 1st ed. Cambridge University Press; 2009 [cited 2023 Feb 9]. Available from: https://www.cambridge.org/core/product/identifier/9780511760068/type/book

116. Fox K, Hawks J. Use ancient remains more wisely. Nature. 2019 Aug 29;572(7771):581–3.

117. Gamba C, Jones ER, Teasdale MD, McLaughlin RL, Gonzalez-Fortes G, Mattiangeli V, et al. Genome flux and stasis in a five millennium transect of European prehistory. Nat Commun. 2014 Oct 21;5(1):5257.

118. TallBear K. Native American DNA: Tribal Belonging and the False Promise of Genetic Science [Internet]. University of Minnesota Press; 2013 [cited 2023 Feb 9]. Available from: https://academic.oup.com/minnesota-scholarship-online/book/19819

119. Garrison NA, Hudson M, Ballantyne LL, Garba I, Martinez A, Taualii M, et al. Genomic Research Through an Indigenous Lens: Understanding the Expectations. Annu Rev Genomics Hum Genet. 2019 Aug 31;20(1):495–517.

120. Libby WF, Anderson EC, Arnold JR. Age Determination by Radiocarbon Content: World-Wide Assay of Natural Radiocarbon. Science. 1949 Mar 4;109(2827):227–8.

121. Reimer PJ, Austin WEN, Bard E, Bayliss A, Blackwell PG, Bronk Ramsey C, et al. The IntCal20 Northern Hemisphere Radiocarbon Age Calibration Curve (0–55 cal kBP). Radiocarbon. 2020 Aug;62(4):725–57.

122. Bronk Ramsey C, van der Plicht J, Weninger B. 'Wiggle Matching' Radiocarbon Dates. Radiocarbon. 2001;43(2A):381–9.

123. Ascough P, Cook G, Dugmore A. Methodological approaches to determining the marine radiocarbon reservoir effect. Prog Phys Geogr Earth Environ. 2005 Dec;29(4):532–47.

124. Michels JW. Dating Methods. Annu Rev Anthropol. 1972 Oct;1(1):113–26.

125. Alexanderson H, Backman J, Cronin TM, Funder S, Ingólfsson Ó, Jakobsson M, et al. An Arctic perspective on dating Mid-Late Pleistocene environmental history. Quat Sci Rev. 2014 May;92:9–31.

126. Wooller MJ, Zazula GD, Edwards M, Froese DG, Boone RD, Parker C, et al. Stable Carbon Isotope Compositions of Eastern Beringian Grasses and Sedges: Investigating Their Potential as Paleoenvironmental Indicators. Arct Antarct Alp Res. 2007 May;39(2):318–31.

127. Lee SA, Kim TH, Kim G. Tracing terrestrial versus marine sources of dissolved organic carbon in a coastal bay using stable carbon isotopes. Biogeosciences. 2020 Jan 15;17(1):135–44.

128. Schoeninger MJ, DeNiro MJ. Nitrogen and carbon isotopic composition of bone collagen from marine and terrestrial animals. Geochim Cosmochim Acta. 1984 Apr;48(4):625–39.

129. Nelson DE, Lynnerup N, Arneborg J. A First Isotopic Dietary Study of the Greenlandic Thule Culture. J N Atl. 2012 Oct;301:51–64.

130. Halffman CM, Potter BA, McKinney HJ, Tsutaya T, Finney BP, Kemp BM, et al. Ancient Beringian paleodiets revealed through multiproxy stable isotope analyses. Sci Adv. 2020 Sep 4;6(36):eabc1968.

131. Choy K, Potter BA, McKinney HJ, Reuther JD, Wang SW, Wooller MJ. Chemical profiling of ancient hearths reveals recurrent salmon use in Ice Age Beringia. Proc Natl Acad Sci. 2016 Aug 30;113(35):9757–62.

132. Sponheimer M, Robinson T, Ayliffe L, Roeder B, Hammer J, Passey B, et al. Nitrogen isotopes in mammalian herbivores: hair ?15N values from a controlled feeding study. Int J Osteoarchaeol. 2003 Jan;13(1–2):80–7.

133. Gilbert MTP, Tomsho LP, Rendulic S, Packard M, Drautz DI, Sher A, et al. Whole-Genome Shotgun Sequencing of Mitochondria from Ancient Hair Shafts. Science. 2007 Sep 28;317(5846):1927–30.

134. Glocke I, Meyer M. Extending the spectrum of DNA sequences retrieved from ancient bones and teeth. Genome Res. 2017 Jul;27(7):1230–7.

135. Rohland N, Harney E, Mallick S, Nordenfelt S, Reich D. Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. Philos Trans R Soc B Biol Sci. 2015 Jan 19;370(1660):20130624.

136. Schubert M, Lindgreen S, Orlando L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res Notes. 2016 Dec;9(1):88.

137. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009 Jul 15;25(14):1754–60.

138. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013 [cited 2023 Feb 10]; Available from: https://arxiv.org/abs/1303.3997

139. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. GigaScience. 2021 Jan 29;10(2):giab008.

140. Picard Toolkit [Internet]. Broad Institute; 2019. (Broad Institute, GitHub repository). Available from: https://broadinstitute.github.io/picard/

141. Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics. 2014 Dec;15(1):356.

142. Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics. 2013 Jul;29(13):1682–4.

143. Fu Q, Mittnik A, Johnson PLF, Bos K, Lari M, Bollongino R, et al. A Revised Timescale for Human Evolution Based on Ancient Mitochondrial Genomes. Curr Biol. 2013 Apr;23(7):553–9.

144. Rasmussen M, Guo X, Wang Y, Lohmueller KE, Rasmussen S, Albrechtsen A, et al. An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia. Science. 2011 Oct 7;334(6052):94–8.

145. Skoglund P, Storå J, Götherström A, Jakobsson M. Accurate sex identification of ancient human remains using DNA shotgun sequencing. J Archaeol Sci. 2013 Dec;40(12):4477–82.

146. Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. Genome Res. 2007 Mar;17(3):377–86.

147. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 2016 Jan 4;44(D1):D733–45.

148. Monroy Kuhn JM, Jakobsson M, Günther T. Estimating genetic kin relationships in prehistoric populations. Calafell F, editor. PLOS ONE. 2018 Apr 23;13(4):e0195491.

149. Allentoft ME, Sikora M, Refoyo-Martínez A, Irving-Pease EK, Fischer A, Barrie W, et al. Population Genomics of Stone Age Eurasia [Internet]. Evolutionary Biology; 2022 May [cited 2023 Feb 7]. Available from: http://biorxiv.org/lookup/doi/10.1101/2022.05.04.490594

150. Wang K, Bleasdale M, Le Moyne C, Freund C, Krause J, Boivin N, et al. 4000-year-old hair from the Middle Nile highlights unusual ancient DNA degradation pattern and a potential source of early eastern Africa pastoralists. Sci Rep. 2022 Dec 3;12(1):20939.

151. Dabney J, Meyer M, Paabo S. Ancient DNA Damage. Cold Spring Harb Perspect Biol. 2013 Jul 1;5(7):a012567–a012567.

152. Moilanen U, Kirkinen T, Saari NJ, Rohrlach AB, Krause J, Onkamo P, et al. A Woman with a Sword? – Weapon Grave at Suontaka Vesitorninmäki, Finland. Eur J Archaeol. 2022 Feb;25(1):42–60.

153. O'Sullivan N, Posth C, Coia V, Schuenemann VJ, Price TD, Wahl J, et al. Ancient genome-wide analyses infer kinship structure in an Early Medieval Alemannic graveyard. Sci Adv. 2018 Sep 7;4(9):eaao1262.

154. Eckhart L, Bach J, Ban J, Tschachler E. Melanin Binds Reversibly to Thermostable DNA Polymerase and Inhibits Its Activity. Biochem Biophys Res Commun. 2000 May;271(3):726–30.

155. Weissensteiner H, Pacher D, Kloss-Brandstätter A, Forer L, Specht G, Bandelt HJ, et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic Acids Res. 2016 Jul 8;44(W1):W58–63.

156. Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, et al. Reconstructing Native American population history. Nature. 2012 Aug 16;488(7411):370–4.

157. Flegontov P, Altınışık NE, Changmai P, Rohland N, Mallick S, Adamski N, et al. Palaeo-Eskimo genetic ancestry and the peopling of Chukotka and North America. Nature. 2019 Jun;570(7760):236–40.

158. Webb EC, Honch NV, Dunn PJH, Linderholm A, Eriksson G, Lidén K, et al. Compound-specific amino acid isotopic proxies for distinguishing between terrestrial and aquatic resource consumption. Archaeol Anthropol Sci. 2018 Feb;10(1):1–18.

159. Stuiver, M., Reimer, P.J. CALIB rev. 8. 1993. 215–230 p. (Radiocarbon; vol. 35).

160. Pieńkowski AJ, Coulthard RD, Furze MFA. Revised marine reservoir offset ( ΔR ) values for molluscs and marine mammals from Arctic North America. Boreas. 2022 Nov 3;bor.12606.

161. Mittnik A, Massy K, Knipper C, Wittenborn F, Friedrich R, Pfrengle S, et al. Kinship-based social inequality in Bronze Age Europe. Science. 2019 Nov 8;366(6466):731–4.

162. Lipatov M, Sanjeev K, Patro R, Veeramah K. Maximum Likelihood Estimation of Biological Relatedness from Low Coverage Sequencing Data [Internet]. Genetics; 2015 Jul [cited 2023 Feb 10]. Available from: http://biorxiv.org/lookup/doi/10.1101/023374

163. Ringbauer H, Novembre J, Steinrücken M. Parental relatedness through time revealed by runs of homozygosity in ancient DNA. Nat Commun. 2021 Sep 14;12(1):5425.

164. TallBear K. Genomic articulations of indigeneity. Soc Stud Sci. 2013 Aug;43(4):509–33.

165. Malaspinas AS, Tange O, Moreno-Mayar JV, Rasmussen M, DeGiorgio M, Wang Y, et al. *bammds:* a tool for assessing the ancestry of low-depth whole-genome data using multidimensional scaling (MDS). Bioinformatics. 2014 Oct 15;30(20):2962–4.

166. Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, et al. Ancient Admixture in Human History. Genetics. 2012 Nov 1;192(3):1065–93.

167. Pickrell JK, Pritchard JK. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. Tang H, editor. PLoS Genet. 2012 Nov 15;8(11):e1002967.

168. Chintalapati M, Patterson N, Moorjani P. The spatiotemporal patterns of major human admixture events during the European Holocene. eLife. 2022 May 30;11:e77625.

169. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078–9.

170. Poznik GD, Henn BM, Yee MC, Sliwerska E, Euskirchen GM, Lin AA, et al. Sequencing Y Chromosomes Resolves Discrepancy in Time to Common Ancestor of Males Versus Females. Science. 2013 Aug 2;341(6145):562–5.

171. International Society of Genetic Genealogy. Y-DNA Haplogroup Tree, Version: 15.73 [Internet]. 2020. Available from: http://www.isogg.org/tree/

172. Mead A. Review of the Development of Multidimensional Scaling Methods. The Statistician. 1992;41(1):27.

173. Borg I, Groenen PJF. Modern multidimensional scaling: theory and applications. 2nd ed. New York: Springer; 2005. (Springer Series in Statistics).

174. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, et al. A Draft Sequence of the Neandertal Genome. Science. 2010 May 7;328(5979):710–22.

175. Durand EY, Patterson N, Reich D, Slatkin M. Testing for Ancient Admixture between Closely Related Populations. Mol Biol Evol. 2011 Aug;28(8):2239–52.

176. Leppälä K, Nielsen SV, Mailund T. admixturegraph: an R package for admixture graph manipulation and fitting. Wren J, editor. Bioinformatics. 2017 Jun 1;33(11):1738–40.

177. Lipson M, Reich D. A working model of the deep relationships of diverse modern human genetic lineages outside of Africa. Mol Biol Evol. 2017 Jan 10;msw293.

178. Narasimhan VM, Patterson N, Moorjani P, Rohland N, Bernardos R, Mallick S, et al. The formation of human populations in South and Central Asia. Science. 2019 Sep 6;365(6457):eaat7487.

179. Tremblay M, Vézina H. New Estimates of Intergenerational Time Intervals for the Calculation of Age and Origins of Mutations. Am J Hum Genet. 2000 Feb;66(2):651–8.

180. Sun JX, Helgason A, Masson G, Ebenesersdóttir SS, Li H, Mallick S, et al. A direct characterization of human mutation based on microsatellites. Nat Genet. 2012 Oct;44(10):1161–5.

181. Chintalapati M, Patterson N, Moorjani P. The spatiotemporal patterns of major human admixture events during the European Holocene. eLife. 2022 May 30;11:e77625.

182. Raghavan M, Steinrücken M, Harris K, Schiffels S, Rasmussen S, DeGiorgio M, et al. Genomic evidence for the Pleistocene and recent population history of Native Americans. Science. 2015 Aug 21;349(6250):aab3884.

183. Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature. 2016 Oct 13;538(7624):201–6.

184. Muktupavela RA, Petr M, Ségurel L, Korneliussen T, Novembre J, Racimo F. Modeling the spatiotemporal spread of beneficial alleles using ancient genomes. eLife. 2022 Dec 20;11:e73767.

185. G. Amorim CE, Nunes K, Meyer D, Comas D, Bortolini MC, Salzano FM, et al. Genetic signature of natural selection in first Americans. Proc Natl Acad Sci. 2017 Feb 28;114(9):2195–9.

186. The 1000 Genomes Project Consortium, Corresponding authors, Auton A, Abecasis GR, Steering committee, Altshuler DM, et al. A global reference for human genetic variation. Nature. 2015 Oct 1;526(7571):68–74.

187. Brzezicha B, Schmidt M, Makałowska I, Jarmołowski A, Pieńkowska J, Szweykowska-Kulińska Z. Identification of human tRNA:m5C methyltransferase catalysing intron-dependent m5C formation in the first position of the anticodon of the ${\hbox{ pre-tRNA }}_{\left(\hbox{ CAA }\right)}^{\hbox{ Leu }}$. Nucleic Acids Res. 2006 Nov;34(20):6034–43.

188. Abbasi-Moheb L, Mertel S, Gonsior M, Nouri-Vahid L, Kahrizi K, Cirak S, et al. Mutations in NSUN2 Cause Autosomal- Recessive Intellectual Disability. Am J Hum Genet. 2012 May;90(5):847–55.

189. Bonnemaijer PWM, Leeuwen EM van, Iglesias AI, Gharahkhani P, Vitart V, Khawaja AP, et al. Multi-trait genome-wide association study identifies new loci associated with optic disc parameters. Commun Biol. 2019 Nov 27;2(1):435.

190. Ma J, Zhao L, Yang Y, Yun D, Yu-Wai-Man P, Zhu Y, et al. Associations Between Regional Environment and Cornea-Related Morphology of the Eye in Young Adults: A Large-Scale Multicenter Cross-Sectional Study. Investig Opthalmology Vis Sci. 2021 Feb 23;62(2):35.

191. Hayes MG, Coltrain JB, O'Rourke DH. Molecular archaeology of the Dorset, Thule, and Sadlermiut: Ancestor-descendant relationships in eastern North American arctic prehistory. Contrib Study Dorset Paleo-Eskimos. 2005;167:11–32.

192. Gilbert MTP, Kivisild T, Grønnow B, Andersen PK, Metspalu E, Reidla M, et al. Paleo-Eskimo mtDNA Genome Reveals Matrilineal Discontinuity in Greenland. Science. 2008 Jun 27;320(5884):1787–9.

193. Tackney J, Jensen AM, Kisielinski C, O'Rourke DH. Molecular analysis of an ancient Thule population at Nuvuk, Point Barrow, Alaska. Am J Phys Anthropol. 2019 Feb;168(2):303–17.

194. Derenko M, Malyarchuk B, Grzybowski T, Denisova G, Rogalla U, Perkova M, et al. Origin and Post-Glacial Dispersal of Mitochondrial DNA Haplogroups C and D in Northern Asia. Gilbert MTP, editor. PLoS ONE. 2010 Dec 21;5(12):e15214.

195. Behar DM, van Oven M, Rosset S, Metspalu M, Loogväli EL, Silva NM, et al. A "Copernican" Reassessment of the Human Mitochondrial DNA Tree from its Root. Am J Hum Genet. 2012 Apr;90(4):675–84.

196. The 1000 Genomes Project Consortium, Poznik GD, Xue Y, Mendez FL, Willems TF, Massaia A, et al. Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. Nat Genet. 2016 Jun;48(6):593–9.

197. Kılınç GM, Kashuba N, Koptekin D, Bergfeldt N, Dönertaş HM, Rodríguez-Varela R, et al. Human population dynamics and *Yersinia pestis* in ancient northeast Asia. Sci Adv. 2021 Jan 6;7(2):eabc4587.

198. Pinotti T, Bergström A, Geppert M, Bawn M, Ohasi D, Shi W, et al. Y Chromosome Sequences Reveal a Short Beringian Standstill, Rapid Expansion, and early Population structure of Native American Founders. Curr Biol. 2019 Jan;29(1):149-157.e3.

199. Grugni V, Raveane A, Ongaro L, Battaglia V, Trombetta B, Colombo G, et al. Analysis of the human Y-chromosome haplogroup Q characterizes ancient population movements in Eurasia and the Americas. BMC Biol. 2019 Dec;17(1):3.

200. Ilumäe AM, Reidla M, Chukhryaeva M, Järve M, Post H, Karmin M, et al. Human Y Chromosome Haplogroup N: A Non-trivial Time-Resolved Phylogeography that Cuts across Language Families. Am J Hum Genet. 2016 Jul;99(1):163–73.

201. Dulik MC, Owings AC, Gaieski JB, Vilar MG, Andre A, Lennie C, et al. Y-chromosome analysis reveals genetic divergence and new founding native lineages in Athapaskan- and Eskimoan-speaking populations. Proc Natl Acad Sci. 2012 May 29;109(22):8471–6.

202. Sun J, Ma P, Cheng H, Wang C, Li Y, Cui Y, et al. Post-last glacial maximum expansion of Y-chromosome haplogroup C2A-L1373 in northern Asia and its implications for the origin of Native Americans. Am J Phys Anthropol. 2021 Feb;174(2):363–74.

203. Yu H, Spyrou MA, Karapetian M, Shnaider S, Radzevičiūtė R, Nägele K, et al. Paleolithic to Bronze Age Siberians Reveal Connections with First Americans and across Eurasia. Cell. 2020 Jun;181(6):1232-1245.e20.

204. de Barros Damgaard P, Martiniano R, Kamm J, Moreno-Mayar JV, Kroonen G, Peyrot M, et al. The first horse herders and the impact of early Bronze Age steppe expansions into Asia. Science. 2018 Jun 29;360(6396):eaar7711.

205. Makeyev VM, Ponomareva DP, Pitulko VV, Chernova GM, Solovyeva DV. Vegetation and Climate of the New Siberian Islands for the Past 15,000 Years. Arct Antarct Alp Res. 2003 Feb;35(1):56–66.

206. Collins, Henry Bascom. Excavations at Thule Culture Sites Near Resolute Bay, Cornwallis Island, N.W.T. 1951. Report No.: Bulletin. (123).

207. Jordan, R. H. Arctic. Damas D, editor. Washington: Smithsonian Institution; 1984. 829 p. (Handbook of North American Indians; vol. 5).

208. Meldgaard, J. Danish-Netherlands symposium on developments in Greenlandic Arctic culture. Kylstra H. P., editor. Groningen: University of Groningen; 1976.

209. Maschner HDG, Mason OK, McGhee R, editors. The Northern world, AD 900-1400. Salt Lake City: University of Utah Press; 2009. 349 p.

210. Friesen TM. Contemporaneity of Dorset and Thule Cultures in the North American Arctic: New Radiocarbon Dates from Victoria Island, Nunavut. Curr Anthropol. 2004 Dec;45(5):685–91.

211. Morrison DA, Pilon JL, Canadian Museum of Civilization, editors. Threads of Arctic prehistory: papers in honour of William E. Taylor, Jr. Hull, Quebec: Canadian Museum of Civilization; 1994. 422 p. (Mercury series).

212. McGhee, R. Ancient people of the Arctic. Savelle JM, editor. Polar Rec. 1998 Jan;34(188):70–70.

213. Appelt M, Dansk Polarcenter, editors. Late Dorset in High Arctic Greenland: final report on the Gateway to Greenland project. Copenhagen: Danish National Museum [u.a.]; 1999. 83 p. (Danish Polar Center publication).

214. Park RW. The Dorset-Thule Succession in Arctic North America: Assessing Claims for Culture Contact. Am Antiq. 1993 Apr;58(2):203–34.

215. Fitzhugh, W. W. Encyclopedia of archaeology. Pearsall DM, editor. Oxford: Academic Press/Elsevier; 2008.

216. Briggs AW, Stenzel U, Johnson PLF, Green RE, Kelso J, Prüfer K, et al. Patterns of damage in genomic DNA sequences from a Neandertal. Proc Natl Acad Sci. 2007 Sep 11;104(37):14616–21.

217. Martiniano R, Cassidy LM, Ó'Maoldúin R, McLaughlin R, Silva NM, Manco L, et al. The population genomics of archaeological transition in west Iberia: Investigation of ancient substructure using imputation and haplotype-based methods. Di Rienzo A, editor. PLOS Genet. 2017 Jul 27;13(7):e1006852.

218. Günther T, Nettelblad C. The presence and impact of reference bias on population genomic studies of prehistoric human populations. Di Rienzo A, editor. PLOS Genet. 2019 Jul 26;15(7):e1008302.

219. Lawson DJ, Hellenthal G, Myers S, Falush D. Inference of Population Structure using Dense Haplotype Data. Copenhaver GP, editor. PLoS Genet. 2012 Jan 26;8(1):e1002453.

220. Maxwell MS. Prehistory of the eastern Arctic. Orlando [Fla.]: Academic Press; 1985. 327 p. (New World archaeological record).

221. Maixner F, Krause-Kyora B, Turaev D, Herbig A, Hoopmann MR, Hallows JL, et al. The 5300-year-old *Helicobacter pylori* genome of the Iceman. Science. 2016 Jan 8;351(6269):162–5.

222. Kay GL, Sergeant MJ, Giuffra V, Bandiera P, Milanese M, Bramanti B, et al. Recovery of a Medieval Brucella melitensis Genome Using Shotgun Metagenomics. Keim PS, editor. mBio. 2014 Aug 29;5(4):e01337-14.

223. Schuenemann VJ, Kumar Lankapalli A, Barquera R, Nelson EA, Iraíz Hernández D, Acuña Alonzo V, et al. Historic Treponema pallidum genomes from Colonial Mexico retrieved from archaeological remains. Norris SJ, editor. PLoS Negl Trop Dis. 2018 Jun 21;12(6):e0006447.

224. Vågene ÅJ, Herbig A, Campana MG, Robles García NM, Warinner C, Sabin S, et al. Salmonella enterica genomes from victims of a major sixteenth-century epidemic in Mexico. Nat Ecol Evol. 2018 Jan 15;2(3):520–8.

225. Fornaciari G, Zavaglia K, Giusti L, Vultaggio C, Ciranni R. Human papillomavirus in a 16th century mummy. The Lancet. 2003 Oct;362(9390):1160.

226. Li HC, Fujiyoshi T, Lou H, Yashiki S, Sonoda S, Cartier L, et al. The presence of ancient human T-cell lymphotropic virus type I provirus DNA in an Andean mummy. Nat Med. 1999 Dec;5(12):1428–32.

227. Pajer P, Dresler J, Kabíckova H, Písa L, Aganov P, Fucik K, et al. Characterization of Two Historic Smallpox Specimens from a Czech Museum. Viruses. 2017 Jul 27;9(8):200.

228. Duggan AT, Perdomo MF, Piombino-Mascali D, Marciniak S, Poinar D, Emery MV, et al. 17th Century Variola Virus Reveals the Recent History of Smallpox. Curr Biol. 2016 Dec;26(24):3407–12.

229. Krause-Kyora B, Susat J, Key FM, Kühnert D, Bosse E, Immel A, et al. Neolithic and medieval virus genomes reveal complex evolution of hepatitis B. eLife. 2018 May 10;7:e36666.

230. Marciniak S, Prowse TL, Herring DA, Klunk J, Kuch M, Duggan AT, et al. Plasmodium falciparum malaria in 1 st –2 nd century CE southern Italy. Curr Biol. 2016 Dec;26(23):R1220–2.

231. Oyola SO, Gu Y, Manske M, Otto TD, O'Brien J, Alcock D, et al. Efficient Depletion of Host DNA Contamination in Malaria Clinical Sequencing. J Clin Microbiol. 2013 Mar;51(3):745–51.

232. Martin DL. Health conditions before Columbus: paleopathology of native North Americans. West J Med. 2002 Jan 1;176(1):65–8.

233. Torrelles JB, Schlesinger LS. Integrating Lung Physiology, Immunology, and Tuberculosis. Trends Microbiol. 2017 Aug;25(8):688–97.

234. Hopkins DR. Princes and peasants: smallpox in history. Chicago: University of Chicago Press; 1983. 380 p.

235. Breitwieser FP, Baker DN, Salzberg SL. KrakenUniq: confident and fast metagenomics classification using unique k-mer counts. Genome Biol. 2018 Nov 16;19(1):198.

236. Li H. Minimap2: pairwise alignment for nucleotide sequences. Birol I, editor. Bioinformatics. 2018 Sep 15;34(18):3094–100.

237. Everett R, Cribdon B. MetaDamage tool: Examining post-mortem damage in sedaDNA on a metagenomic scale. Front Ecol Evol. 2023 Jan 24;10:888421.

238. Levenshtein VI. Binary Codes Capable of Correcting Deletions, Insertions and Reversals. Sov Phys Dokl. 1966 Feb 1;10:707.

239. Yoon SH, Ha S min, Lim J, Kwon S, Chun J. A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie Van Leeuwenhoek. 2017 Oct;110(10):1281–6.

240. Morgulis A, Gertz EM, Schäffer AA, Agarwala R. A Fast and Symmetric DUST Implementation to Mask Low-Complexity DNA Sequences. J Comput Biol. 2006 Jun;13(5):1028–40.

241. Bagley ST. Habitat Association of *Klebsiella* Species. Infect Control. 1985 Feb;6(2):52–8.

242. Glassing A, Dowd SE, Galandiuk S, Davis B, Chiodini RJ. Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. Gut Pathog. 2016 Dec;8(1):24.

243. Arriola LA, Cooper A, Weyrich LS. Palaeomicrobiology: Application of Ancient DNA Sequencing to Better Understand Bacterial Genome Evolution and Adaptation. Front Ecol Evol. 2020 Jun 16;8:40.

244. Batt CA, Tortorello ML, editors. Encyclopedia of food microbiology. 2. ed. Amsterdam: AP, Academic Press/Elsevier; 2014.

245. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012 Apr;9(4):357–9.

246. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol. 2013 Apr 1;30(4):772–80.

247. Czech L, Barbera P, Stamatakis A. Genesis and Gappa: processing, analyzing and visualizing phylogenetic (placement) data. Schwartz R, editor. Bioinformatics. 2020 May 1;36(10):3263–5.

248. Yu G. Using ggtree to Visualize Data on Tree-Like Structures. Curr Protoc Bioinforma [Internet]. 2020 Mar [cited 2023 Feb 10];69(1). Available from: https://onlinelibrary.wiley.com/doi/10.1002/cpbi.96

249. Vihtakari, M. _ggOceanMaps: Plot Data on Oceanographic Maps using 'ggplot2'_. [Internet]. GitHub; 2022. Available from: https://mikkovihtakari.github.io/ggOceanMaps/

250. Breman JG, Henderson DA. Diagnosis and Management of Smallpox. N Engl J Med. 2002 Apr 25;346(17):1300–8.

251. Jenner E. On the Origin of the Vaccine Inoculation. Med Phys J. 1801 Jun;5(28):505–8.

252. Hendrickson RC, Wang C, Hatcher EL, Lefkowitz EJ. Orthopoxvirus Genome Evolution: The Role of Gene Loss. Viruses. 2010 Sep 15;2(9):1933–67.

253. Parker S, Crump R, Hartzler H, Buller R. Evaluation of Taterapox Virus in Small Animals. Viruses. 2017 Aug 1;9(8):203.

254. Baxby D. Jenner's smallpox vaccine: the riddle of vaccinia virus and its origin. London: Heinemann Educational Books; 1981. 214 p.

255. Correa P, Piazuelo MB. Natural history of Helicobacter pylori infection. Dig Liver Dis. 2008 Jul;40(7):490–6.

256. Moodley Y, Linz B. *Helicobacter pylori* Sequences Reflect Past Human Migrations. In: de Reuse H, Bereswill S, editors. Genome Dynamics [Internet]. S. Karger AG; 2009 [cited 2023 Feb 10]. p. 62–74. Available from: https://www.karger.com/Article/FullText/235763

257. Falush D, Wirth T, Linz B, Pritchard JK, Stephens M, Kidd M, et al. Traces of Human Migrations in *Helicobacter pylori* Populations. Science. 2003 Mar 7;299(5612):1582–5.

258. Moodley Y, Brunelli A, Ghirotto S, Klyubin A, Maady AS, Tyne W, et al. *Helicobacter pylori* 's historical journey through Siberia and the Americas. Proc Natl Acad Sci. 2021 Jun 22;118(25):e2015523118.

259. Aspholm M, Olfat FO, Nordén J, Sondén B, Lundberg C, Sjöström R, et al. SabA Is the H. pylori Hemagglutinin and Is Polymorphic in Binding to Sialylated Glycans. Hultgren SJ, editor. PLoS Pathog. 2006 Oct 27;2(10):e110.

260. Correa P, Willis D, Allison MJ, Gerszten E. Helicobacter pylori in pre-Columbian mummies. Gastroenterology. 1998 Apr;114:A956.

261. Castillo-Rojas G, Cerbón MA, López-Vidal Y. Presence of Helicobacter pylori in a Mexican Pre-Columbian Mummy. BMC Microbiol. 2008 Dec;8(1):119.

262. Swanston T, Haakensen M, Deneer H, Walker EG. The Characterization of Helicobacter pylori DNA Associated with Ancient Human Remains Recovered from a Canadian Glacier. Ahmed N, editor. PLoS ONE. 2011 Feb 16;6(2):e16864.

263. Shin DH, Oh CS, Hong JH, Lee H, Lee SD, Lee E. Helicobacter pylori DNA obtained from the stomach specimens of two 17th century Korean mummies. Anthropol Anz. 2018 Feb 1;75(1):75–87.

264. Zerboni L, Sen N, Oliver SL, Arvin AM. Molecular mechanisms of varicella zoster virus pathogenesis. Nat Rev Microbiol. 2014 Mar;12(3):197–210.

265. Hayat MA, Grose, C. Introduction to Autophagy. In: Autophagy: Cancer, Other Pathologies, Inflammation, Immunity, Infection, and Aging [Internet]. Elsevier; 2015 [cited 2023 Feb 10]. p. 1–51. Available from: https://linkinghub.elsevier.com/retrieve/pii/B9780128010327000010

266. Etymologia: Varicella Zoster Virus. Emerg Infect Dis. 2015 Apr;21(4):698–698.

267. Grose C. Pangaea and the Out-of-Africa Model of Varicella-Zoster Virus Evolution and Phylogeography. J Virol. 2012 Sep 15;86(18):9558–65.

268. Beauregard A, Smith E, Petrone B, Singh N, Karch C, McDonough K, et al. Identification and characterization of small RNAs in Yersinia pestis . RNA Biol. 2013 Mar;10(3):397–405.

269. Michelsen C, Pedersen MW, Fernandez-Guerra A, Zhao L, Petersen TC, Korneliussen TS. metaDMG – A Fast and Accurate Ancient DNA Damage Toolkit for Metagenomic Data [Internet]. Bioinformatics; 2022 Dec [cited 2023 Feb 10]. Available from: http://biorxiv.org/lookup/doi/10.1101/2022.12.06.519264

270. Bos KI, Jäger G, Schuenemann VJ, Vågene ÅJ, Spyrou MA, Herbig A, et al. Parallel detection of ancient pathogens via array-based DNA capture. Philos Trans R Soc B Biol Sci. 2015 Jan 19;370(1660):20130375.

271. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w $^{1118}$ ; iso-2; iso-3. Fly (Austin). 2012 Apr;6(2):80–92.

272. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. Pertea M, editor. PLOS Comput Biol. 2019 Apr 8;15(4):e1006650.

273. Lindo J, Huerta-Sánchez E, Nakagome S, Rasmussen M, Petzelt B, Mitchell J, et al. A time transect of exomes from a Native American population before and after European contact. Nat Commun. 2016 Dec;7(1):13175.

274. Ilardo MA. Man and the Sea: Genetics in Maritime Populations. Natural History Museum of Denmark, Faculty of Science, University of Copenhagen; 2018.

# Supplemental Tables

## Table S1 – Tissue Metadata

List of all tissues included in this project (n=217), sorted by the ID assigned at the Centre for GeoGenetics (CGG) and accompanied by the alias from the museum, cultural centre, or laboratory they were previously housed, the tissue type, the location, province or state, and region from which they originated, as well as the archaeologically assigned cultural affiliation.

| ID | Alias | Tissue | Location | Province/ State | Region | Culture |
|---|---|---|---|---|---|---|
| 011229 | MARC480 | Tooth | Port au Choix | Newfoundland | Canada | Dorset |
| 011230 | MARC481 | Tooth | Port au Choix | Newfoundland | Canada | Dorset |
| 011231 | MARC482 | Tooth | Port au Choix | Newfoundland | Canada | Dorset |
| 011234 | KAL-0043x01 | Bone | Suessland | Nordøstkysten | Greenland | Thule |
| 011236 | KAL-0041x01 | Bone | Suessland | Suessland | Greenland | Thule |
| 011291 | A-3 N179 | Hair | Paipelghak | Chukchi Peninsula | Siberia | Birnirk |
| 011292 | A-2 N79 | Hair | Paipelghak | Chukchi Peninsula | Siberia | Birnirk |
| 011293 | A/3 N177 | Hair | Paipelghak | Chukchi Peninsula | Siberia | Birnirk |
| 011294 | A/3 N181 | Hair | Paipelghak | Chukchi Peninsula | Siberia | Birnirk |
| 011295 | A-2 N83 | Hair | Paipelghak | Chukchi Peninsula | Siberia | Birnirk |
| 011299 | KAL-0048x01 | Tooth | Skærgaardshalvæ | Kangerlussuaq | Greenland | Thule |
| 011311 | B-003-1 | Hair | Nunalleq | Alaska | United States | Yupik |
| 011312 | B-248-1 | Hair | Nunalleq | Alaska | United States | Yupik |
| 011313 | B-248-2 | Hair | Nunalleq | Alaska | United States | Yupik |
| 011316 | B-248-3 | Hair | Nunalleq | Alaska | United States | Yupik |
| 011317 | B-248-4 | Hair | Nunalleq | Alaska | United States | Yupik |
| 011319 | B-248-6 | Hair | Nunalleq | Alaska | United States | Yupik |
| 011320 | B-248-7 | Hair | Nunalleq | Alaska | United States | Yupik |
| 011321 | B-248-9 | Hair | Nunalleq | Alaska | United States | Yupik |
| 011322 | B-248-10 | Hair | Nunalleq | Alaska | United States | Yupik |
| 011323 | B-248-12 | Hair | Nunalleq | Alaska | United States | Yupik |
| 011324 | B-248-15 | Hair | Nunalleq | Alaska | United States | Yupik |
| 011325 | B-248-16 | Hair | Nunalleq | Alaska | United States | Yupik |
| 011327 | B-248-19 | Hair | Nunalleq | Alaska | United States | Yupik |
| 011332 | B-248-26 | Hair | Nunalleq | Alaska | United States | Yupik |
| 011335 | B-248-30 | Hair | Nunalleq | Alaska | United States | Yupik |
| 011336 | B-248-31 | Hair | Nunalleq | Alaska | United States | Yupik |
| 011337 | B-248-32 | Hair | Nunalleq | Alaska | United States | Yupik |
| 011339 | B-248-36 | Hair | Nunalleq | Alaska | United States | Yupik |
| 011341 | B-248-39 | Hair | Nunalleq | Alaska | United States | Yupik |
| 011342 | B-248-40 | Hair | Nunalleq | Alaska | United States | Yupik |
| 011343 | B-248-44 | Hair | Nunalleq | Alaska | United States | Yupik |
| 011414 | XIV-C:340 | Bone | Angekok, Mansel Island | Nunavut | Canada | Late Dorset |
| 011415 | XIV-C:340 | Bone | Angekok, Mansel Island | Nunavut | Canada | Late Dorset |
| 011423 | XIV-C:3 or XIV-C:4 | Tooth | Sugluk, Avataq | Quebec | Canada | Middle Dorset |
| 011433 | XIV-C:3 or XIV-C:4 | Tooth | Sugluk, Avataq | Quebec | Canada | Middle Dorset |
| 011434 | XIV-H:126 | Bone | Saatut, Baffin Island | Nunavut | Canada | Middle-Late Dorset |
| 011435 | XIV-H:168 | Bone | North Devon Island | Nunavut | Canada | Pre-Dorset |
| 011436 | Alarnek-3, NhHd-1 | Tooth | Melville Peninsula | Nunavut | Canada | Middle Dorset |
| 011442 | Alarnek-1 | Bone | Melville Peninsula | Nunavut | Canada | Middle Dorset |
| 011444 | MARC1492 | Tooth | Old Mission Point | New Brunswick | Canada | Mi'gmaq |
| 011445 | MARC1493 | Tooth | Port au Choix | Newfoundland | Canada | Dorset |
| 011446 | MARC1490 | Tooth | Port au Choix | Newfoundland | Canada | Dorset |
| 011447 | MARC1489 | Tooth | Eastern Point | Newfoundland | Canada | Dorset |
| 011448 | MARC1491 | Tooth | Englee | Newfoundland | Canada | Dorset |
| 011874 | SR-8474 | Hair | Zhokhov | New Siberian Islands | Siberia | Zhokhovo |
| 011875 | SR-8475 | Bone | Zhokhov | New Siberian Islands | Siberia | Zhokhovo |
| 011876 | SR-8476 | Tooth | Zhokhov | New Siberian Islands | Siberia | Zhokhovo |
| 011877 | SR-8477 | Bone | Zhokhov | New Siberian Islands | Siberia | Zhokhovo |
| 011878 | SR-8478 | Bone | Zhokhov | New Siberian Islands | Siberia | Zhokhovo |
| 011879 | SR-8479 | Bone | Zhokhov | New Siberian Islands | Siberia | Zhokhovo |
| 011880 | SR-8480 | Bone | Zhokhov | New Siberian Islands | Siberia | Zhokhovo |

| ID | Alias | Tissue | Location | Province/ State | Region | Culture |
|---|---|---|---|---|---|---|
| 022559 | ALGV-231 | Tooth | Vaughan | Ontario | Canada | Huron-Wendat |
| 022560 | 11SP-83 | Tooth | Orillia | Ontario | Canada | Huron-Wendat |
| 022561 | 05SP-46 | Tooth | Vaughan | Ontario | Canada | Huron-Wendat |
| 022903 | Y-1 | Bone | Iron age | Northern Yakutia | Siberia | Yakut |
| 022904 | Y-2 | Bone | Iron age | Central Yakutia | Siberia | Yakut |
| 022910 | U-1 | Bone | Ushki-I, layer VI | Kamchatka | Siberia | Late Ushki |
| 022911 | U-2 | Bone | Ushki-I, layer VI | Kamchatka | Siberia | Late Ushki |
| 022912 | U-3 | Bone | Ushki-I, layer VI | Kamchatka | Siberia | Late Ushki |
| 022913 | U-4 | Bone | Ushki-I, layer VI | Kamchatka | Siberia | Late Ushki |
| 022914 | U-5 | Bone | Ushki-I, layer VI | Kamchatka | Siberia | Late Ushki |
| 022915 | U-6 | Bone | Ushki-I, layer VI | Kamchatka | Siberia | Late Ushki |
| 022916 | U-7a | Tooth | Ushki-I, layer VI | Kamchatka | Siberia | Late Ushki |
| 022917 | U-7b | Bone | Ushki-I, layer VI | Kamchatka | Siberia | Late Ushki |
| 022918 | U-8 | Bone | Ushki-I, layer VII | Kamchatka | Siberia | Late Ushki |
| 022919 | U-9 | Bone | Ushki-I, layer VII | Kamchatka | Siberia | Late Ushki |
| 022920 | U-10 | Bone | Ushki-I, layer VII | Kamchatka | Siberia | Late Ushki |
| 022921 | U-11 | Bone | Ushki-I, layer VII | Kamchatka | Siberia | Late Ushki |
| 022922 | U-12 | Bone | Ushki-V, layer VII | Kamchatka | Siberia | Late Ushki |
| 023501 | B-248-18 | Hair | Nunalleq | Alaska | United States | Yupik |
| 023502 | B-248-25 | Hair | Nunalleq | Alaska | United States | Yupik |
| 023504 | B-248-33 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024054 | KAL-0172x01 | Tooth | Uummannaq | Uummannaq | Greenland | Thule |
| 024055 | KAL 0166x01 | Tooth | Uummannaq | Uummannaq | Greenland | Thule |
| 024056 | KAL-0173x01 | Tooth | Uummannaq | Uummannaq | Greenland | Thule |
| 024057 | KAL-0175x01 | Tooth | Uummannaq | Uummannaq | Greenland | Thule |
| 024058 | KAL 0256x01 | Tooth | Umanak | Umanak | Greenland | Thule |
| 024059 | KAL 0169x01 | Tooth | Uummannaq | Uummannaq | Greenland | Thule |
| 024060 | KAL-0231x01 | Tooth | Uummannaq | Uummannaq | Greenland | Thule |
| 024061 | KAL-0232x01 | Tooth | Uummannaq | Uummannaq | Greenland | Thule |
| 024062 | KAL-0248x01 | Tooth | Uummannaq | Uummannaq | Greenland | Thule |
| 024063 | KAL 0247x01 | Tooth | Uummannaq | Uummannaq | Greenland | Thule |
| 024064 | KAL 0602x01 | Tooth | Uummannaq | Uummannaq | Greenland | Thule |
| 024065 | KAL 0165x01 | Tooth | Uummannaq | Uummannaq | Greenland | Thule |
| 024066 | KAL 0283x01 | Tooth | Uummannaq | Uummannaq | Greenland | Thule |
| 024067 | KAL 0607x01 | Tooth | Uummannaq | Uummannaq | Greenland | Thule |
| 024068 | KAL-0245x01 | Tooth | Uummannaq | Uummannaq | Greenland | Thule |
| 024069 | KAL 0246x01 | Tooth | Uummannaq | Uummannaq | Greenland | Thule |
| 024070 | KAL 0244x01 | Tooth | Uummannaq | Uummannaq | Greenland | Thule |
| 024071 | KAL 0238x01 | Tooth | Uummannaq | Uummannaq | Greenland | Thule |
| 024072 | KAL 0242x01 | Tooth | Uummannaq | Uummannaq | Greenland | Thule |
| 024073 | KAL 0001x02 | Tooth | Tununngassoq | Inussuk | Greenland | Thule |
| 024074 | KAL 0001x01 | Tooth | Tununngassoq | Inussuk | Greenland | Thule |
| 024075 | KAL-0060x01 | Tooth | Dødemandsbugten | Dødemandsbugten | Greenland | Thule |
| 024076 | KAL 1549x01 | Tooth | Illutalik | Anap Nunaa | Greenland | Thule |
| 024077 | KAL 1547x01 | Tooth | Illutalik | Anap Nunaa | Greenland | Thule |
| 024078 | KAL-1548x01 | Tooth | Illutalik | Anap Nunaa | Greenland | Thule |
| 024079 | KAL 1440x01 | Tooth | Qoornoq | Nuup Kangarlua | Greenland | Thule |
| 024080 | KAL-1467x01 | Tooth | Uunartoq | Uunartoq Fjord | Greenland | Thule |
| 024081 | KAL-1461x01 | Tooth | Uunartoq | Uunartoq Fjord | Greenland | Thule |
| 024082 | KAL 1460x01 | Tooth | Uunartoq | Uunartoq Fjord | Greenland | Thule |
| 024083 | KAL 1422x01 | Tooth | Ruinnæsset | Akorninap Kangerlua | Greenland | Thule |
| 024084 | KAL 1425x01 | Tooth | Ruinnæsset | Akorninap Kangerlua | Greenland | Thule |
| 024085 | KAL 0122x01 | Tooth | Sukertii | Sermilik | Greenland | Thule |
| 024086 | KAL-1410x01 | Tooth | Skærgaardshalvæ | Kangerlussuaq | Greenland | Thule |
| 024088 | KAL 0126x01 | Tooth | Kangertittivatsiaq | Saqqarmiut | Greenland | Thule |
| 024089 | KAL 0077x01 | Tooth | Rypefjeld | Mørke Fjord | Greenland | Thule |
| 024090 | KAL 0058x01 | Tooth | Dødemandsbugten | Clavering Ø | Greenland | Thule |
| 024091 | KAL 0030x01 | Tooth | Dødemandsbugten | Clavering Ø | Greenland | Thule |
| 024092 | KAL 0059x01 | Tooth | Dødemandsbugten | Clavering Ø | Greenland | Thule |
| 024093 | KAL 0055x01 | Tooth | Dødemandsbugten | Dødemandsbugten | Greenland | Thule |
| 024094 | KAL 0061x01 | Tooth | Dødemandsbugten | Dødemandsbugten | Greenland | Thule |
| 024095 | KAL 0723x01 | Tooth | Dødemandsbugten | Clavering Ø | Greenland | Thule |
| 024098 | KAL 0065x01 | Tooth | Suessland | Suessland | Greenland | Thule |
| 024099 | KAL-0044x01 | Tooth | Suessland | Suessland | Greenland | Thule |
| 024100 | KAL-0067x01 | Tooth | Kap Harry | Ella Ø | Greenland | Thule |

| ID | Alias | Tissue | Location | Province/ State | Region | Culture |
|---|---|---|---|---|---|---|
| 024101 | KAL-0158x01 | Tooth | Nuussuaq omr. | Nuugaaq | Greenland | Thule |
| 024102 | KAL 0159x01 | Tooth | Nuussuaq omr. | Nuugaaq | Greenland | Thule |
| 024103 | KAL 0160x01 | Tooth | Nuussuaq omr. | Nuugaaq | Greenland | Thule |
| 024112 | Mag 5a | Bone | Olskaya | Northern Priokhotye | Siberia | Tokarev |
| 024113 | Mag 2 | Bone | Tokareva Site | Northern Priokhotye | Siberia | Tokarev |
| 024114 | Mag 30K | Hair | Galgan-1 | Northwest Kamchatka | Siberia | Tevi |
| 024115 | Mag 31K | Tooth | Galgan-1 | Northwest Kamchatka | Siberia | Tevi |
| 024116 | Mag 23 | Tooth | Chikaevo | Southeastern Chukotka | Siberia | Ust-Belaya |
| 024117 | Mag 6 | Tooth | Ust-Belaya cemetery | Southeastern Chukotka | Siberia | Ust-Belaya |
| 024118 | Mag 36K | Tooth | Ozernaya | Northeast Kamchatka | Siberia | Old Koryak |
| 024119 | Mag 8 | Tooth | Ust-Belaya cemetery | Southeastern Chukotka | Siberia | Ust-Belaya |
| 024120 | Mag 25 | Tooth | Ushki-I, layer II | Central Kamchatka | Siberia | Tarya |
| 024121 | Mag 17 | Tooth | Sed'moi Prichal | Southeastern Chukotka | Siberia | Kanchalan |
| 024122 | Mag 22 | Tooth | Yavino-1 | South Kamchatka | Siberia | Old Itelmen |
| 024123 | Mag 24 | Tooth | Pegtymel | Northern Chukotka | Siberia | Bronze age |
| 024124 | Mag 9 | Hair | Gavriil Bay-II | Northwest Bering Sea | Siberia | Lakhtin |
| 024125 | Mag 37K | Hair | Tevi | Northwest Kamchatka | Siberia | Tevi |
| 024126 | Mag 0 | Tooth | Kutarey | Cis-Baikal | Siberia | Early neolithic |
| 024247 | Mag 1a | Bone | Olskaya | Northern Priokhotye | Siberia | Tokarev |
| 024248 | Mag 3a | Bone | Ust-Belaya cemetery | Southeastern Chukotka | Siberia | Ust-Belaya |
| 024249 | Mag 4 | Bone | Ust-Belaya cemetery | Southeastern Chukotka | Siberia | Ust-Belaya |
| 024250 | Mag 7a | Bone | Ust-Belaya cemetery | Southeastern Chukotka | Siberia | Ust-Belaya |
| 024251 | Mag 10a | Bone | Lopatka-I | South Kamchatka | Siberia | Old Itelmen |
| 024252 | Mag 11a | Bone | Lopatka-I | South Kamchatka | Siberia | Old Itelmen |
| 024253 | Mag 12a | Tooth | Seshan cemetery | Chukchi Peninsula | Siberia | Old Bering |
| 024254 | Mag 13a | Tooth | Vankarem | Northern Chukotka | Siberia | Old Bering |
| 024255 | Mag 14a | Bone | Inchoun cemetery | Chukchi Peninsula | Siberia | Old Bering |
| 024256 | Mag 15 | Petrous | Uelen cemetery | Chukchi Peninsula | Siberia | Old Bering |
| 024257 | Mag 16 | Tooth | Sed'moi Prichal | Southeastern Chukotka | Siberia | Kanchalan |
| 024258 | Mag 17 | Tooth | Sed'moi Prichal | Southeastern Chukotka | Siberia | Kanchalan |
| 024259 | Mag 18 | Bone | Sed'moi Prichal | Southeastern Chukotka | Siberia | Kanchalan |
| 024260 | Mag 19 | Bone | Sed'moi Prichal | Southeastern Chukotka | Siberia | Kanchalan |
| 024261 | Mag 20 | Tooth | Chegitun cemetery | Chukchi Peninsula | Siberia | Old Bering |
| 024262 | Mag 21 | Tooth | Chegitun cemetery | Chukchi Peninsula | Siberia | Old Bering |
| 024263 | Mag 28 | Bone | Olskaya | Northern Priokhotye | Siberia | Tokarev |
| 024264 | Mag 29 | Bone | Olskaya | Northern Priokhotye | Siberia | Tokarev |
| 024265 | Mag 32K | Tooth | Tilichiki | Northwest Kamchatka | Siberia | Old Koryak |
| 024266 | Mag 33K | Tooth | Tilichiki | Northwest Kamchatka | Siberia | Old Koryak |
| 024267 | Mag 34K | Tooth | Karaga | Northeast Kamchatka | Siberia | Old Koryak |
| 024268 | Mag 35K | Tooth | Klyuchi | Central Kamchatka | Siberia | Old Itelmen |
| 024724 | S-1207 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024725 | S-1070 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024726 | S-1382 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024727 | S-1457 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024728 | B-248-50 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024729 | B-248-48 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024730 | B-248-46 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024731 | B-248-49 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024732 | B-248-53 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024733 | B-248-22 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024734 | B-248-29 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024735 | B-248-21 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024736 | B-248-23 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024737 | B-248-47 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024738 | B-248-45 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024739 | B-248-38 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024740 | B-248-28 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024741 | B-248-27 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024742 | B-248-42 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024743 | B-248-51 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024744 | B-248-54 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024745 | B-248-14 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024746 | B-248-17 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024747 | S-1053 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024748 | S-1295 | Hair | Nunalleq | Alaska | United States | Yupik |

| ID | Alias | Tissue | Location | Province/ State | Region | Culture |
|---|---|---|---|---|---|---|
| 024749 | S-1350 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024750 | S-1460 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024751 | S-1470 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024752 | S-1473 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024753 | S-1559 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024754 | S-1565 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024755 | S-1569 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024756 | S-1592 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024757 | S-14160 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024758 | S-14214 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024759 | S-14337 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024760 | S-14373 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024761 | S-14392 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024762 | S-14403 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024763 | S-14404 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024764 | S-15051 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024766 | S-15126 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024767 | S-15151 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024768 | S-15165 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024769 | S-15166 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024770 | S-15167 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024771 | S-15169 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024772 | S-15171 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024773 | S-15184 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024774 | S-15260 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024775 | S-15262 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024776 | S-15326 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024777 | S-15332 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024778 | S-15383 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024779 | S-15400 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024780 | S-15408 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024781 | S-15411 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024782 | S-15412 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024783 | B-248-5 | Hair | Nunalleq | Alaska | United States | Yupik |
| 024784 | B-248-20 | Hair | Nunalleq | Alaska | United States | Yupik |

## Table S2 – Tissue Dating

List of all tissues included in this project (n=217), sorted by the ID assigned at the CGG and accompanied by the radiocarbon dating ID provided by the labs where the C14 dating took place, the radiocarbon date in years before present (BP), the uncertainty in years of the C14 date, the calibrated C14 date using IntCal20, the bulk carbon and nitrogen isotopes (provided either by the C14 dating lab or from published works). For those tissues that were not radiocarbon dated, an approximate date (based on the average of dates from that site or archaeological approximation), are shown in column 5.

| ID | Dating ID | Radiocarbon Date (BP) | ± | Calibrated Date, IntCal20 (BP) | Bulk d13C (‰ VPDB) | Bulk d15N (‰ AIR) |
|---|---|---|---|---|---|---|
| 011229 | UBA-41565 | 1993 | 26 | 1928 | -12.4 | 23.3 |
| 011230 | UBA-41566 | 1923 | 27 | 1840 | -12.3 | 22.6 |
| 011231 | UBA-41567 | 2027 | 26 | 1963 | -11.8 | 21.5 |
| 011234 | | No direct date | | ~300 | -15.9 | 18.1 |
| 011236 | | No direct date | | ~300 | -15.3 | 17.2 |
| 011291 | UBA-41390 | 1680 | 24 | 1567 | -13.9 | 19.5 |
| 011292 | UBA-41391 | 1712 | 24 | 1596 | -13.2 | 18.6 |
| 011293 | UBA-41392 | 1726 | 23 | 1611 | -12.8 | 19.3 |
| 011294 | UBA-41393 | 1816 | 29 | 1715 | -14 | 18.7 |
| 011295 | UBA-41394 | 1716 | 24 | 1600 | -13.5 | 18.7 |
| 011299 | | No direct date | | ~300 | -13.4 | 19 |
| 011311 | | No direct date | | ~650 (site date) | | |
| 011312 | | No direct date | | ~650 (site date) | | |
| 011313 | | No direct date | | ~650 (site date) | | |
| 011316 | | No direct date | | ~650 (site date) | | |
| 011317 | | No direct date | | ~650 (site date) | | |
| 011319 | | No direct date | | ~650 (site date) | | |
| 011320 | | No direct date | | ~650 (site date) | -16.4 | 14.9 |
| 011321 | | No direct date | | ~650 (site date) | | |
| 011322 | | No direct date | | ~650 (site date) | | |
| 011323 | | No direct date | | ~650 (site date) | | |
| 011324 | | No direct date | | ~650 (site date) | | |
| 011325 | | No direct date | | ~650 (site date) | | |
| 011327 | | No direct date | | ~650 (site date) | | |
| 011332 | | No direct date | | ~650 (site date) | | |
| 011335 | | No direct date | | ~650 (site date) | | |
| 011336 | | No direct date | | ~650 (site date) | | |
| 011337 | | No direct date | | ~650 (site date) | | |
| 011339 | | No direct date | | ~650 (site date) | | |
| 011341 | | No direct date | | ~650 (site date) | | |
| 011342 | | No direct date | | ~650 (site date) | | |
| 011343 | | No direct date | | ~650 (site date) | | |
| 011414 | AAR-17518 | 1295 | 25 | 1226 | -13.7 | 19.7 |
| 011415 | AAR-17518 | 1295 | 25 | 1226 | -13.7 | 19.7 |
| 011423 | AAR-17521 | 2193 | 28 | 2233 | -12.6 | 20.1 |
| 011433 | AAR-17521 | 2193 | 28 | 2233 | -12.6 | 20.1 |
| 011434 | AAR-17520 | 1915 | 30 | 1827 | -13.1 | 23.1 |
| 011435 | AAR-17520 | 4205 | 15 | 4741 | -12.6 | 21.7 |
| 011436 | AAR-17516 | 1765 | 26 | 1651 | -14.4 | 20.1 |
| 011442 | UCIAMS-236956 | 2920 | 15 | 3063 | -14 | 20.6 |
| 011444 | UCIAMS107246 | 740 | 20 | 675 | -15.6 | 18.1 |
| 011445 | UBA-41564 | 1987 | 28 | 1920 | -12.4 | 22.6 |
| 011446 | UCIAMS-125914 | 1905 | 15 | 1806 | -13.3 | 21.7 |
| 011447 | UCIAMS-125902 | 1935 | 15 | 1857 | -13 | 20.8 |
| 011448 | UCIAMS-125907 | 2030 | 15 | 1964 | -13.2 | 20.9 |
| 011874 | UCIAMS-236948 | 8710 | 20 | 9634 | -22.7 | 14.5 |
| 011875 | UCIAMS-236949 | 7985 | 20 | 8874 | -19.3 | 5 |
| 011876 | UCIAMS-236950 | 8220 | 20 | 9190 | -19.1 | 13.9 |
| 011877 | UCIAMS-236951 | 7865 | 20 | 8630 | -20 | 13.8 |
| 011878 | UCIAMS-236952 | 8070 | 20 | 9003 | -19.7 | 13.2 |

| ID | Dating ID | Radiocarbon Date (BP) | ± | Calibrated Date, IntCal20 (BP) | Bulk d13C (‰ VPDB) | Bulk d15N (‰ AIR) |
|---|---|---|---|---|---|---|
| 011879 | UCIAMS-236953 | 8085 | 20 | 9010 | -19.7 | 12.8 |
| 011880 | UCIAMS-236954 | 7860 | 20 | 8626 | -20 | 12.6 |
| 022559 | UCIAMS-233726 | 485 | 15 | 519 | -11.2 | 11 |
| 022560 | UCIAMS-233727 | 720 | 15 | 669 | -13.4 | 11.3 |
| 022561 | UCIAMS-233728 | 465 | 15 | 512 | -11.3 | 11.3 |
| 022903 | UCIAMS-225304 | 565 | 20 | 598 | -20.9 | 12.8 |
| 022904 | UCIAMS-225305 | 185 | 20 | 185 | -21.3 | 11.2 |
| 022910 | | Failed dating | | ~10500 (site average) | | |
| 022911 | | Failed dating | | ~10500 (site average) | | |
| 022912 | | Failed dating | | ~10500 (site average) | | |
| 022913 | | Failed dating | | ~10500 (site average) | | |
| 022914 | | Failed dating | | ~10500 (site average) | | |
| 022915 | | Failed dating | | ~10500 (site average) | | |
| 022916 | | Failed dating | | ~10500 (site average) | | |
| 022917 | UCIAMS-225306 | 10550 | 30 | 12552 | -16.5 | 14.7 |
| 022918 | | Failed dating | | ~10500 (site average) | | |
| 022919 | | Failed dating | | ~10500 (site average) | | |
| 022920 | | Failed dating | | ~10500 (site average) | | |
| 022921 | | Failed dating | | ~10500 (site average) | | |
| 022922 | | Failed dating | | ~10500 (site average) | | |
| 023501 | | No direct date | | ~650 (site date) | -15.7 | 13.6 |
| 023502 | | No direct date | | ~650 (site date) | -15.2 | 17.6 |
| 023504 | | No direct date | | ~650 (site date) | -16.8 | 17 |
| 024054 | | No direct date | | ~300 | | |
| 024055 | | No direct date | | ~300 | | |
| 024056 | | No direct date | | ~300 | | |
| 024057 | | No direct date | | ~300 | | |
| 024058 | | No direct date | | ~300 | | |
| 024059 | | No direct date | | ~300 | | |
| 024060 | | No direct date | | ~300 | | |
| 024061 | | No direct date | | ~300 | | |
| 024062 | | No direct date | | ~300 | | |
| 024063 | | No direct date | | ~300 | | |
| 024064 | | No direct date | | ~300 | | |
| 024065 | | No direct date | | ~300 | | |
| 024066 | | No direct date | | ~300 | | |
| 024067 | | No direct date | | ~300 | | |
| 024068 | | No direct date | | ~300 | | |
| 024069 | | No direct date | | ~300 | | |
| 024070 | | No direct date | | ~300 | | |
| 024071 | | No direct date | | ~300 | | |
| 024072 | | No direct date | | ~300 | | |
| 024073 | | No direct date | | ~300 | -13 | 21.4 |
| 024074 | | No direct date | | ~300 | -13.1 | 21.3 |
| 024075 | | No direct date | | ~300 | -15.3 | 17.2 |
| 024076 | | No direct date | | ~300 | -13.5 | 21.8 |
| 024077 | | No direct date | | ~300 | -12.7 | 21.4 |
| 024078 | | No direct date | | ~300 | -13.4 | 21.6 |
| 024079 | | No direct date | | ~300 | -12.8 | 19.4 |
| 024080 | | No direct date | | ~300 | -13.2 | 19 |
| 024081 | | No direct date | | ~300 | -13.3 | 19.7 |
| 024082 | | No direct date | | ~300 | -13.4 | 19.6 |
| 024083 | | No direct date | | ~300 | -13.8 | 19.3 |
| 024084 | | No direct date | | ~300 | -13.3 | 19.4 |
| 024085 | | No direct date | | ~300 | -13.7 | 18.8 |
| 024086 | | No direct date | | ~300 | -13.7 | 20.3 |
| 024088 | | No direct date | | ~300 | | |
| 024089 | | No direct date | | ~300 | -15.5 | 18.7 |
| 024090 | | No direct date | | ~300 | -14.9 | 19.3 |
| 024091 | | No direct date | | ~300 | -15.5 | 19.3 |
| 024092 | | No direct date | | ~300 | -14.8 | 19.6 |
| 024093 | | No direct date | | ~300 | -16.1 | 17.3 |
| 024094 | | No direct date | | ~300 | -15.3 | 17.1 |
| 024095 | | No direct date | | ~300 | -15.6 | 16.8 |

| ID | Dating ID | Radiocarbon Date (BP) | ± | Calibrated Date, IntCal20 (BP) | Bulk d13C (‰ VPDB) | Bulk d15N (‰ AIR) |
|---|---|---|---|---|---|---|
| 024098 | | No direct date | | ~300 | -15.7 | 17.4 |
| 024099 | | No direct date | | ~300 | -16.1 | 16.8 |
| 024100 | | No direct date | | ~300 | -15.6 | 18 |
| 024101 | | No direct date | | ~300 | -12.9 | 20.9 |
| 024102 | | No direct date | | ~300 | -13 | 21 |
| 024103 | | No direct date | | ~300 | -13.5 | 19.7 |
| 024112 | UBA-41381 | 2756 | 29 | 2839 | -11.4 | 16.7 |
| 024113 | | Failed dating | | ~2500 | | |
| 024114 | UCIAMS-233695 | 1940 | 15 | 1862 | -15.9 | 16.9 |
| 024115 | UCIAMS-233722 | 1955 | 15 | 1879 | -13.2 | 18.4 |
| 024116 | UBA-41382 | 2160 | 25 | 2148 | -11.5 | 21.3 |
| 024117 | UBA-41384 | 3099 | 28 | 3302 | -18.9 | 13.8 |
| 024118 | UBA-41385 | 1442 | 26 | 1330 | -14.2 | 16.7 |
| 024119 | UCIAMS-233723 | 1915 | 15 | 1831 | -18.4 | 7.7 |
| 024120 | UBA-41387 | 2928 | 29 | 3078 | -17.9 | 12.5 |
| 024121 | UCIAMS-233724 | 565 | 15 | 598 | -19.2 | 13.3 |
| 024122 | UCIAMS-233725 | 1685 | 15 | 1565 | -14.9 | 16.7 |
| 024123 | UBA-41388 | 939 | 30 | 848 | -18.6 | 17.1 |
| 024124 | UBA-41395 | 2339 | 23 | 2348 | -17.3 | 13.2 |
| 024125 | UCIAMS-233696 | 2130 | 15 | 2100 | -13.9 | 17.2 |
| 024126 | UBA-41389 | 6703 | 38 | 7573 | -19.1 | 13 |
| 024247 | UCIAMS-225310 | 3005 | 20 | 3194 | -13 | 20.3 |
| 024248 | UCIAMS-225311 | 2520 | 20 | 2586 | -17.5 | 11 |
| 024249 | UCIAMS-225312/3 | 4478 | 20 | 5183 | -18 | 13.4 |
| 024250 | UCIAMS-225314 | 1025 | 20 | 938 | -14.5 | 17.3 |
| 024251 | UCIAMS-225315 | 2650 | 20 | 2757 | -18.5 | 12.1 |
| 024252 | UCIAMS-225316 | 980 | 20 | 859 | -14.4 | 17.6 |
| 024253 | UCIAMS-225317 | 1945 | 20 | 1869 | -11.5 | 21.3 |
| 024254 | UCIAMS-225318 | 2230 | 20 | 2221 | -12.8 | 20.5 |
| 024255 | UCIAMS-225319/20 | 1710 | 20 | 1592 | -12.2 | 19.9 |
| 024256 | UCIAMS-225321/2 | 1670 | 20 | 1556 | -12.4 | 20.6 |
| 024257 | UCIAMS-225324 | 785 | 20 | 701 | -13.9 | 14.6 |
| 024258 | UCIAMS-225325 | 580 | 20 | 604 | -18.8 | 13.7 |
| 024259 | UCIAMS-225326/7 | 943 | 20 | 849 | -13.8 | 16.4 |
| 024260 | UCIAMS-225328 | 870 | 20 | 760 | -18.6 | 16 |
| 024261 | UCIAMS-225333 | 1580 | 20 | 1465 | -11.8 | 20.1 |
| 024262 | UCIAMS-225334 | 1470 | 15 | 1349 | -12.5 | 20.5 |
| 024263 | UCIAMS-225335 | 1910 | 20 | 1818 | -19.4 | 2.5 |
| 024264 | UCIAMS-225336 | 3065 | 20 | 3287 | -13.1 | 20.2 |
| 024265 | UCIAMS-225337 | 1115 | 20 | 1009 | -12.7 | 18.3 |
| 024266 | UCIAMS-225338 | 1095 | 20 | 994 | -13.1 | 18.3 |
| 024267 | UCIAMS-225339 | 1185 | 20 | 1106 | -13.1 | 18.4 |
| 024268 | UCIAMS-225340 | 1540 | 20 | 1402 | -15 | 17.8 |
| 024724 | | No direct date | | ~650 (site date) | -16.9 | 16.4 |
| 024725 | | No direct date | | ~650 (site date) | -15.6 | 17.4 |
| 024726 | | No direct date | | ~650 (site date) | -16 | 15.9 |
| 024727 | | No direct date | | ~650 (site date) | | |
| 024728 | | No direct date | | ~650 (site date) | -15.5 | 16.2 |
| 024729 | | No direct date | | ~650 (site date) | -15.5 | 16.2 |
| 024730 | | No direct date | | ~650 (site date) | -15.9 | 16.7 |
| 024731 | | No direct date | | ~650 (site date) | -15.4 | 16.1 |
| 024732 | | No direct date | | ~650 (site date) | -15.5 | 16.7 |
| 024733 | | No direct date | | ~650 (site date) | -15.4 | 16 |
| 024734 | | No direct date | | ~650 (site date) | -15.8 | 17 |
| 024735 | | No direct date | | ~650 (site date) | -15.9 | 16.2 |
| 024736 | | No direct date | | ~650 (site date) | -15.5 | 16.7 |
| 024737 | | No direct date | | ~650 (site date) | -15.8 | 15.7 |
| 024738 | | No direct date | | ~650 (site date) | -15.7 | 11.9 |
| 024739 | | No direct date | | ~650 (site date) | | |
| 024740 | | No direct date | | ~650 (site date) | -15.1 | 16.3 |
| 024741 | | No direct date | | ~650 (site date) | -15.4 | 14.5 |
| 024742 | | No direct date | | ~650 (site date) | -15.7 | 17.4 |
| 024743 | | No direct date | | ~650 (site date) | -15.5 | 17.2 |
| 024744 | | No direct date | | ~650 (site date) | -16.1 | 17.4 |

| ID | Dating ID | Radiocarbon Date (BP) | ± | Calibrated Date, IntCal20 (BP) | Bulk d13C (‰ VPDB) | Bulk d15N (‰ AIR) |
|---|---|---|---|---|---|---|
| 024745 | | No direct date | | ~650 (site date) | -17.1 | 15.3 |
| 024746 | | No direct date | | ~650 (site date) | -15.7 | 18.2 |
| 024747 | | No direct date | | ~650 (site date) | -15.3 | 17.5 |
| 024748 | | No direct date | | ~650 (site date) | -15.4 | 18 |
| 024749 | | No direct date | | ~650 (site date) | -16.5 | 16.3 |
| 024750 | | No direct date | | ~650 (site date) | | |
| 024751 | | No direct date | | ~650 (site date) | | |
| 024752 | | No direct date | | ~650 (site date) | | |
| 024753 | | No direct date | | ~650 (site date) | | |
| 024754 | | No direct date | | ~650 (site date) | | |
| 024755 | | No direct date | | ~650 (site date) | | |
| 024756 | | No direct date | | ~650 (site date) | | |
| 024757 | | No direct date | | ~650 (site date) | -16.3 | 16.7 |
| 024758 | | No direct date | | ~650 (site date) | -15.7 | 16.6 |
| 024759 | | No direct date | | ~650 (site date) | -15.1 | 17.2 |
| 024760 | | No direct date | | ~650 (site date) | -15.9 | 17 |
| 024761 | | No direct date | | ~650 (site date) | -16 | 17 |
| 024762 | | No direct date | | ~650 (site date) | -16 | 15.8 |
| 024763 | | No direct date | | ~650 (site date) | -15.7 | 16.7 |
| 024764 | | No direct date | | ~650 (site date) | | |
| 024766 | | No direct date | | ~650 (site date) | -15.6 | 16.4 |
| 024767 | | No direct date | | ~650 (site date) | | |
| 024768 | | No direct date | | ~650 (site date) | -15.9 | 17.6 |
| 024769 | | No direct date | | ~650 (site date) | -14.6 | 18.4 |
| 024770 | | No direct date | | ~650 (site date) | -16 | 17.4 |
| 024771 | | No direct date | | ~650 (site date) | | |
| 024772 | | No direct date | | ~650 (site date) | | |
| 024773 | | No direct date | | ~650 (site date) | -16.3 | 16.5 |
| 024774 | | No direct date | | ~650 (site date) | -15.7 | 16.6 |
| 024775 | | No direct date | | ~650 (site date) | -15.8 | 16.5 |
| 024776 | | No direct date | | ~650 (site date) | -15.7 | 16.6 |
| 024777 | | No direct date | | ~650 (site date) | -15.5 | 16.2 |
| 024778 | | No direct date | | ~650 (site date) | | |
| 024779 | | No direct date | | ~650 (site date) | -15.2 | 17 |
| 024780 | | No direct date | | ~650 (site date) | -15.9 | 17.3 |
| 024781 | | No direct date | | ~650 (site date) | | |
| 024782 | | No direct date | | ~650 (site date) | -15.2 | 17.8 |
| 024783 | | No direct date | | ~650 (site date) | -15.5 | 17.1 |
| 024784 | | No direct date | | ~650 (site date) | -15.7 | 17.9 |

## Table S3 – Bulk Isotope Dietary Information

List of all tissues included in this project (n=217), sorted by the ID assigned at the CGG and accompanied by the bulk carbon and nitrogen isotope results from the Alaska Stable Isotope Facility (ASIF), the percentage of marine dietary proportion, the percentage of terrestrial dietary proportion, the percentage of freshwater dietary proportion, and the estimated palaeodietary profile. Individuals with over 80% marine contribution were assigned a marine based diet, which is relevant to many aspects of this project. Profiles labelled as $C_4$ are assigned to individuals with a diet based on $C_4$ plants (*i.e.,* those from the Huron-Wendat Nation). Many of the samples that were not analysed at ASIF were still assigned dietary profiles based on the stable isotope results from the other facilities where they were C14 dated and concurrently measured for stable isotopes.

| ID | Bulk $\delta^{13}C$ (ASIF) | Bulk $\delta^{15}N$ (ASIF) | % Marine | % Terrestrial | % Freshwater | Dietary Profile |
|---|---|---|---|---|---|---|
| 011229 | Not sent to ASIF | Not sent to ASIF | 100 | 0 | 0 | marine |
| 011230 | Not sent to ASIF | Not sent to ASIF | 100 | 0 | 0 | marine |
| 011231 | Not sent to ASIF | Not sent to ASIF | 100 | 0 | 0 | marine |
| 011234 | Not sent to ASIF | Not sent to ASIF | 90 | 0 | 10 | marine |
| 011236 | Not sent to ASIF | Not sent to ASIF | 90 | 0 | 10 | marine |
| 011291 | -13.9 | 19.5 | 100 | 0 | 0 | marine |
| 011292 | -13.6 | 18.3 | 100 | 0 | 0 | marine |
| 011293 | -13.1 | 18.9 | 100 | 0 | 0 | marine |
| 011294 | -14 | 18.7 | 100 | 0 | 0 | marine |
| 011295 | -13.5 | 18.7 | 100 | 0 | 0 | marine |
| 011299 | -13.09 | 20.38 | 100 | 0 | 0 | marine |
| 011311 | -15.54 | 18.2 | 90 | 0 | 10 | marine |
| 011312 | -16.14 | 13.95 | 70 | 30 | 0 | mixed marine+terrestrial |
| 011313 | Not sent to ASIF | Not sent to ASIF | | | | |
| 011316 | Not sent to ASIF | Not sent to ASIF | | | | |
| 011317 | Not sent to ASIF | Not sent to ASIF | | | | |
| 011319 | -16.32 | 16.07 | 82 | 10 | 8 | marine |
| 011320 | -16.5 | 16.03 | 78 | 12 | 10 | mixed marine+terrestrial |
| 011321 | Not sent to ASIF | Not sent to ASIF | | | | |
| 011322 | -15.99 | 16.04 | 83 | 12 | 5 | marine |
| 011323 | -16.03 | 17.2 | 88 | 0 | 12 | marine |
| 011324 | -16.05 | 15.9 | 82 | 13 | 5 | marine |
| 011325 | -15.75 | 16.81 | 89 | 4 | 7 | marine |
| 011327 | -16.06 | 16.32 | 84 | 8 | 8 | marine |
| 011332 | -15.98 | 16.56 | 86 | 6 | 8 | marine |
| 011335 | -15.75 | 16.37 | 87 | 9 | 4 | marine |
| 011336 | -16.45 | 17.24 | 84 | 0 | 16 | marine |
| 011337 | -15.97 | 15.68 | 82 | 15 | 3 | marine |
| 011339 | -15.78 | 14.9 | 79 | 21 | 0 | mixed marine+terrestrial |
| 011341 | Not sent to ASIF | Not sent to ASIF | | | | |
| 011342 | -16.14 | 15.99 | 82 | 12 | 6 | marine |
| 011343 | -15.83 | 17.53 | 90 | 0 | 10 | marine |
| 011414 | -14.2 | 17.4 | 100 | 0 | 0 | marine |
| 011415 | -14.6 | 17.5 | 100 | 0 | 0 | marine |
| 011423 | -13.8 | 18.9 | 100 | 0 | 0 | marine |
| 011433 | -13.3 | 19.2 | 100 | 0 | 0 | marine |
| 011434 | -13.6 | 22 | 100 | 0 | 0 | marine |
| 011435 | -12.7 | 21.8 | 100 | 0 | 0 | marine |
| 011436 | -13.5 | 21.1 | 100 | 0 | 0 | marine |
| 011442 | -14.2 | 20.5 | 100 | 0 | 0 | marine |
| 011444 | Not sent to ASIF | Not sent to ASIF | 90 | 0 | 10 | marine |
| 011445 | Not sent to ASIF | Not sent to ASIF | 100 | 0 | 0 | marine |
| 011446 | Not sent to ASIF | Not sent to ASIF | 100 | 0 | 0 | marine |
| 011447 | Not sent to ASIF | Not sent to ASIF | 100 | 0 | 0 | marine |
| 011448 | Not sent to ASIF | Not sent to ASIF | 100 | 0 | 0 | marine |
| 011874 | -20.34 | 12.36 | 27 | 42 | 31 | mixed |
| 011875 | -19.47 | 4.86 | 0 | 100 | 0 | terrestrial |

| ID | Bulk $\delta^{13}C$ (ASIF) | Bulk $\delta^{15}N$ (ASIF) | % Marine | % Terrestrial | % Freshwater | Dietary Profile |
|---|---|---|---|---|---|---|
| 011876 | -20.64 | 16.52 | 42 | 0 | 58 | mixed marine+freshwater |
| 011877 | -20.36 | 14.43 | 37 | 20 | 43 | mixed |
| 011878 | -19.86 | 13.25 | 36 | 33 | 31 | mixed |
| 011879 | -19.77 | 12.82 | 35 | 37 | 28 | mixed |
| 011880 | -20.01 | 12.79 | 32 | 38 | 30 | mixed |
| 022559 | -11.1 | 11.4 | $C_4$ | $C_4$ | $C_4$ | $C_4$ |
| 022560 | -13 | 11.3 | $C_4$ | $C_4$ | $C_4$ | $C_4$ |
| 022561 | -10.8 | 11.9 | $C_4$ | $C_4$ | $C_4$ | $C_4$ |
| 022903 | -21 | 12.8 | 23 | 36 | 41 | mixed |
| 022904 | -21.6 | 10.7 | 9 | 56 | 35 | mixed |
| 022910 | -18.3 | 14.1 | 54 | 27 | 19 | mixed |
| 022911 | No collagen | No collagen | No Collagen | No Collagen | No Collagen | |
| 022912 | No collagen | No collagen | No Collagen | No Collagen | No Collagen | |
| 022913 | No collagen | No collagen | No Collagen | No Collagen | No Collagen | |
| 022914 | No collagen | No collagen | No Collagen | No Collagen | No Collagen | |
| 022915 | No collagen | No collagen | No Collagen | No Collagen | No Collagen | |
| 022916 | No collagen | No collagen | No Collagen | No Collagen | No Collagen | |
| 022917 | No collagen | No collagen | No Collagen | No Collagen | No Collagen | |
| 022918 | No collagen | No collagen | No Collagen | No Collagen | No Collagen | |
| 022919 | No collagen | No collagen | No Collagen | No Collagen | No Collagen | |
| 022920 | No collagen | No collagen | No Collagen | No Collagen | No Collagen | |
| 022921 | No collagen | No collagen | No Collagen | No Collagen | No Collagen | |
| 022922 | No collagen | No collagen | No Collagen | No Collagen | No Collagen | |
| 023501 | Not sent to ASIF | Not sent to ASIF | | | | |
| 023502 | Not sent to ASIF | Not sent to ASIF | | | | |
| 023504 | -17.66 | 16.99 | 72 | 0 | 28 | mixed marine+freshwater |
| 024054 | -13.1 | 21.4 | 83 | 10 | 7 | marine |
| 024055 | -13.2 | 22.4 | 100 | 0 | 0 | marine |
| 024056 | -13.3 | 22.2 | 100 | 0 | 0 | marine |
| 024057 | -13.5 | 21.4 | 100 | 0 | 0 | marine |
| 024058 | -13.5 | 21.3 | 100 | 0 | 0 | marine |
| 024059 | No collagen | 0 | No Collagen | No Collagen | No Collagen | |
| 024060 | -12.4 | 20.1 | 100 | 0 | 0 | marine |
| 024061 | -12.3 | 22.2 | 100 | 0 | 0 | marine |
| 024062 | -12.4 | 21.9 | 100 | 0 | 0 | marine |
| 024063 | -13.2 | 20.9 | 100 | 0 | 0 | marine |
| 024064 | -13 | 22.1 | 100 | 0 | 0 | marine |
| 024065 | -13.3 | 21 | 100 | 0 | 0 | marine |
| 024066 | -12.4 | 23.4 | 100 | 0 | 0 | marine |
| 024067 | -12.6 | 22.8 | 100 | 0 | 0 | marine |
| 024068 | -13.7 | 21.4 | 100 | 0 | 0 | marine |
| 024069 | -12.4 | 20.6 | 100 | 0 | 0 | marine |
| 024070 | -13.2 | 22.7 | 100 | 0 | 0 | marine |
| 024071 | -13.4 | 20.3 | 100 | 0 | 0 | marine |
| 024072 | No collagen | 0 | No Collagen | No Collagen | No Collagen | |
| 024073 | -13 | 21.3 | 100 | 0 | 0 | marine |
| 024074 | -12.8 | 21.2 | 100 | 0 | 0 | marine |
| 024075 | -15.7 | 18.5 | 100 | 0 | 0 | marine |
| 024076 | -12.7 | 22.7 | 100 | 0 | 0 | marine |
| 024077 | -13.1 | 22.5 | 100 | 0 | 0 | marine |
| 024078 | -12.9 | 22.3 | 100 | 0 | 0 | marine |
| 024079 | -12.9 | 20 | 100 | 0 | 0 | marine |
| 024080 | -12.9 | 19.9 | 100 | 0 | 0 | marine |
| 024081 | -12.5 | 20.5 | 100 | 0 | 0 | marine |
| 024082 | -13.8 | 18.9 | 100 | 0 | 0 | marine |
| 024083 | -12.9 | 20.1 | 100 | 0 | 0 | marine |
| 024084 | -13.1 | 20.1 | 100 | 0 | 0 | marine |
| 024085 | -14 | 19.6 | 100 | 0 | 0 | marine |
| 024086 | -13.9 | 20.5 | 100 | 0 | 0 | marine |
| 024088 | -13.8 | 18.5 | 100 | 0 | 0 | marine |
| 024089 | -13.3 | 19.6 | 100 | 0 | 0 | marine |
| 024090 | -14.3 | 20.6 | 100 | 0 | 0 | marine |
| 024091 | -14.5 | 20.5 | 100 | 0 | 0 | marine |
| 024092 | -14.6 | 20.8 | 100 | 0 | 0 | marine |

| ID | Bulk $\delta^{13}C$ (ASIF) | Bulk $\delta^{15}N$ (ASIF) | % Marine | % Terrestrial | % Freshwater | Dietary Profile |
|---|---|---|---|---|---|---|
| 024093 | -15.1 | 20.1 | 100 | 0 | 0 | marine |
| 024094 | -15.6 | 19.4 | 100 | 0 | 0 | marine |
| 024095 | -15.2 | 19.9 | 100 | 0 | 0 | marine |
| 024098 | -15.6 | 19.4 | 100 | 0 | 0 | marine |
| 024099 | -15.6 | 19.4 | 100 | 0 | 0 | marine |
| 024100 | -13.7 | 19.6 | 100 | 0 | 0 | marine |
| 024101 | -12.7 | 21.5 | 100 | 0 | 0 | marine |
| 024102 | -12.2 | 22.2 | 100 | 0 | 0 | marine |
| 024103 | -12.8 | 20.8 | 100 | 0 | 0 | marine |
| 024112 | -11.4 | 16.1 | 100 | 0 | 0 | marine |
| 024113 | No collagen | 0 | No Collagen | No Collagen | No Collagen | |
| 024114 | Not sent to ASIF | Not sent to ASIF | 88 | 3 | 9 | marine |
| 024115 | Not sent to ASIF | Not sent to ASIF | 100 | 0 | 0 | marine |
| 024116 | -11.4 | 20.3 | 100 | 0 | 0 | marine |
| 024117 | -18.9 | 13.5 | 46 | 32 | 22 | mixed |
| 024118 | -14.2 | 16.6 | 100 | 0 | 0 | marine |
| 024119 | Not sent to ASIF | Not sent to ASIF | 15 | 85 | 0 | mixed marine+terrestrial |
| 024120 | -18 | 12.5 | 50 | 44 | 6 | mixed marine+terrestrial |
| 024121 | Not sent to ASIF | Not sent to ASIF | 41 | 35 | 24 | mixed |
| 024122 | Not sent to ASIF | Not sent to ASIF | 100 | 0 | 0 | marine |
| 024123 | -18.5 | 16.5 | 55 | 4 | 41 | mixed |
| 024124 | -17.3 | 13.2 | 59 | 38 | 3 | mixed marine+terrestrial |
| 024125 | Not sent to ASIF | Not sent to ASIF | 100 | 0 | 0 | marine |
| 024126 | -19 | 12.9 | 42 | 38 | 20 | mixed |
| 024247 | -13.2 | 20.6 | 100 | 0 | 0 | marine |
| 024248 | -17.6 | 10 | 35 | 65 | 0 | mixed marine+terrestrial |
| 024249 | -14.5 | 19.2 | 100 | 0 | 0 | marine |
| 024250 | -14.3 | 16.7 | 100 | 0 | 0 | marine |
| 024251 | -18.4 | 11.9 | 43 | 49 | 8 | mixed marine+terrestrial |
| 024252 | -17.3 | 13.1 | 58 | 40 | 2 | mixed marine+terrestrial |
| 024253 | -11.8 | 22.7 | 100 | 0 | 0 | marine |
| 024254 | -12.6 | 21 | 100 | 0 | 0 | marine |
| 024255 | -12 | 20.1 | 100 | 0 | 0 | marine |
| 024256 | -12.2 | 20.9 | 100 | 0 | 0 | marine |
| 024257 | -12.4 | 16.8 | 100 | 0 | 0 | marine |
| 024258 | -19.5 | 13.6 | 40 | 31 | 29 | mixed |
| 024259 | -13.8 | 16.3 | 100 | 0 | 0 | marine |
| 024260 | -18.4 | 16.2 | 62 | 6 | 32 | mixed |
| 024261 | -11.7 | 20.3 | 100 | 0 | 0 | marine |
| 024262 | -12.5 | 20.7 | 100 | 0 | 0 | marine |
| 024263 | -20.1 | 2.4 | 0 | 100 | 0 | terrestrial |
| 024264 | -13.3 | 20.6 | 100 | 0 | 0 | marine |
| 024265 | -13.6 | 18.7 | 100 | 0 | 0 | marine |
| 024266 | -12.7 | 18.9 | 100 | 0 | 0 | marine |
| 024267 | -13.1 | 18.7 | 100 | 0 | 0 | marine |
| 024268 | -15 | 17.6 | 100 | 0 | 0 | marine |
| 024724 | -15.19 | 18.4 | 100 | 0 | 0 | marine |
| 024725 | -15.92 | 17.1 | 100 | 0 | 0 | marine |
| 024726 | -16.06 | 17.63 | 86 | 0 | 14 | marine |
| 024727 | -16 | 17.58 | 90 | 0 | 10 | marine |
| 024728 | -15.91 | 15.9 | 83 | 14 | 3 | marine |
| 024729 | Not sent to ASIF | Not sent to ASIF | | | | |
| 024730 | -15.72 | 16.54 | 88 | 7 | 5 | marine |
| 024731 | -15.77 | 16.03 | 85 | 12 | 3 | marine |
| 024732 | -15.87 | 16.57 | 87 | 6 | 7 | marine |
| 024733 | Not sent to ASIF | Not sent to ASIF | | | | |
| 024734 | Not sent to ASIF | Not sent to ASIF | | | | |
| 024735 | Not sent to ASIF | Not sent to ASIF | | | | |
| 024736 | Not sent to ASIF | Not sent to ASIF | | | | |
| 024737 | -16.37 | 15.75 | 78 | 14 | 8 | mixed marine+terrestrial |
| 024738 | Not sent to ASIF | Not sent to ASIF | | | | |
| 024739 | Not sent to ASIF | Not sent to ASIF | | | | |
| 024740 | Not sent to ASIF | Not sent to ASIF | | | | |
| 024741 | -15.4 | 13.78 | 50 | 30 | 20 | mixed |

| ID | Bulk $\delta^{13}C$ (ASIF) | Bulk $\delta^{15}N$ (ASIF) | % Marine | % Terrestrial | % Freshwater | Dietary Profile |
|---|---|---|---|---|---|---|
| 024742 | Not sent to ASIF | Not sent to ASIF | | | | |
| 024743 | Not sent to ASIF | Not sent to ASIF | | | | |
| 024744 | -16.42 | 17.02 | 83 | 7 | 10 | marine |
| 024745 | Not sent to ASIF | Not sent to ASIF | | | | |
| 024746 | Not sent to ASIF | Not sent to ASIF | | | | |
| 024747 | -16.11 | 17.04 | 87 | 1 | 12 | marine |
| 024748 | -16.06 | 17.6 | 85 | 0 | 15 | marine |
| 024749 | -15.77 | 17.15 | 90 | 0 | 10 | marine |
| 024750 | Not sent to ASIF | Not sent to ASIF | | | | |
| 024751 | Not sent to ASIF | Not sent to ASIF | | | | |
| 024752 | Not sent to ASIF | Not sent to ASIF | | | | |
| 024753 | -15.14 | 18.42 | 100 | 0 | 0 | marine |
| 024754 | Not sent to ASIF | Not sent to ASIF | | | | |
| 024755 | -18.3 | 15.6 | 60 | 12 | 28 | mixed |
| 024756 | Not sent to ASIF | Not sent to ASIF | | | | |
| 024757 | Not sent to ASIF | Not sent to ASIF | | | | |
| 024758 | -16.07 | 16.5 | 85 | 6 | 9 | marine |
| 024759 | -15.63 | 17.03 | 90 | 3 | 7 | marine |
| 024760 | Not sent to ASIF | Not sent to ASIF | | | | |
| 024761 | -16.43 | 17.12 | 84 | 0 | 16 | marine |
| 024762 | -16.3 | 16.18 | 81 | 10 | 9 | marine |
| 024763 | Not sent to ASIF | Not sent to ASIF | | | | |
| 024764 | -16.72 | 15.8 | 75 | 13 | 12 | mixed |
| 024766 | -15.88 | 16.63 | 87 | 6 | 7 | marine |
| 024767 | -16.39 | 16.53 | 82 | 6 | 12 | marine |
| 024768 | Not sent to ASIF | Not sent to ASIF | | | | |
| 024769 | Not sent to ASIF | Not sent to ASIF | | | | |
| 024770 | -16.06 | 17.7 | 85 | 0 | 15 | marine |
| 024771 | Not sent to ASIF | Not sent to ASIF | | | | |
| 024772 | -16.06 | 16.5 | 85 | 6 | 9 | marine |
| 024773 | -16.61 | 16.34 | 80 | 7 | 13 | marine |
| 024774 | -15.76 | 18.37 | 90 | 0 | 10 | marine |
| 024775 | Not sent to ASIF | Not sent to ASIF | | | | |
| 024776 | -15.86 | 16.62 | 87 | 6 | 7 | marine |
| 024777 | Not sent to ASIF | Not sent to ASIF | | | | |
| 024778 | -15.63 | 16.9 | 90 | 4 | 6 | marine |
| 024779 | -15.98 | 17.18 | 88 | 0 | 12 | marine |
| 024780 | -16.25 | 17.25 | 86 | 0 | 14 | marine |
| 024781 | Not sent to ASIF | Not sent to ASIF | | | | |
| 024782 | -15.57 | 17.63 | 90 | 0 | 10 | marine |
| 024783 | -16.22 | 17.14 | 86 | 0 | 14 | marine |
| 024784 | Not sent to ASIF | Not sent to ASIF | | | | |

**Table S4 – Autosomal Contamination Estimates**

List of all tissues included in this project (n=217), sorted by the ID assigned at the CGG and accompanied by the reported metrics from the autosomal contamination estimation tool Contamix. Contamix Approximate MAP [Mapping Authenticity Parameter] used sites with ≥1x coverage, with basecalls assigned by a 50% consensus of reads covering that site. A second approach called Contamix Precise MAP was also employed that used sites with ≥5x coverage, with basecalls assigned by a 70% consensus of reads covering that site. Both the approximate and precise methods of Contamix are reported with the lower (LO) and upper (HI) confidence intervals. Contamination estimates were generated from the contamination-removed merged genomes. Where no result is shown, genomes were either merged (where identical individuals were identified, indicated in column 7 of Table S6), or the genomic coverage was too low for the tools to accurately estimate contamination. Where contamination estimates are lower than 0.95 in either the Approximate or Precise method, they were flagged as contaminated in column 6 of Table S5, (where the results from the mitochondrial contamination ANGSD are shown).

| ID | Contamix Approx MAP | Contamix Approx LO | Contamix Approx HI | Contamix Precise MAP | Contamix Precise LO | Contamix Precise HI |
|---|---|---|---|---|---|---|
| 011229 | 0.9949 | 0.9879 | 0.9988 | 0.9950 | 0.9882 | 0.9988 |
| 011230 | 0.9920 | 0.9878 | 0.9948 | 0.9921 | 0.9878 | 0.9948 |
| 011231 | 0.9960 | 0.9919 | 0.9980 | 0.9958 | 0.9917 | 0.9980 |
| 011234 | 0.9998 | 0.9955 | 1.0000 | 0.9998 | 0.9959 | 1.0000 |
| 011236 | 0.9998 | 0.9965 | 1.0000 | 0.9998 | 0.9966 | 1.0000 |
| 011291 | 0.9998 | 0.9964 | 1.0000 | 0.9998 | 0.9962 | 1.0000 |
| 011292 | 0.9871 | 0.9784 | 0.9947 | 0.9877 | 0.9783 | 0.9945 |
| 011293 | 0.9999 | 0.9982 | 1.0000 | 0.9999 | 0.9983 | 1.0000 |
| 011294 | 0.9999 | 0.9984 | 1.0000 | 0.9999 | 0.9983 | 1.0000 |
| 011295 | | | | | | |
| 011299 | 0.9562 | 0.9375 | 0.9716 | 0.9985 | 0.9930 | 0.9998 |
| 011311 | 0.9433 | 0.1702 | 0.9884 | | | |
| 011312 | 0.9595 | 0.3111 | 0.9925 | | | |
| 011313 | 0.9998 | 0.9949 | 1.0000 | 0.9998 | 0.9952 | 1.0000 |
| 011316 | 0.9999 | 0.9967 | 1.0000 | 0.9999 | 0.9966 | 1.0000 |
| 011317 | | | | | | |
| 011319 | 0.9999 | 0.9972 | 1.0000 | 0.9999 | 0.9972 | 1.0000 |
| 011320 | 0.9999 | 0.9973 | 1.0000 | 0.9999 | 0.9973 | 1.0000 |
| 011321 | | | | | | |
| 011322 | 0.9998 | 0.9955 | 1.0000 | 0.9999 | 0.9975 | 1.0000 |
| 011323 | | | | | | |
| 011324 | 0.9999 | 0.9969 | 1.0000 | 0.9999 | 0.9969 | 1.0000 |
| 011325 | 0.9994 | 0.9859 | 0.9999 | 0.9993 | 0.9849 | 0.9999 |
| 011327 | | | | | | |
| 011332 | 0.9998 | 0.9965 | 1.0000 | 0.9998 | 0.9965 | 1.0000 |
| 011335 | | | | | | |

| ID | Contamix Approx MAP | Contamix Approx LO | Contamix Approx HI | Contamix Precise MAP | Contamix Precise LO | Contamix Precise HI |
|---|---|---|---|---|---|---|
| 011336 | 0.9997 | 0.9941 | 0.9999 | 0.9997 | 0.9941 | 0.9999 |
| 011337 | 0.9999 | 0.9979 | 1.0000 | 0.9999 | 0.9979 | 1.0000 |
| 011339 | 0.9998 | 0.9966 | 1.0000 | 0.9999 | 0.9977 | 1.0000 |
| 011341 | 0.9983 | 0.9922 | 0.9999 | 0.9998 | 0.9952 | 1.0000 |
| 011342 | | | | | | |
| 011343 | 0.9996 | 0.9940 | 0.9999 | 0.9999 | 0.9973 | 1.0000 |
| 011414 | 0.9997 | 0.9944 | 0.9999 | 0.9998 | 0.9961 | 1.0000 |
| 011415 | | | | | | |
| 011423 | 0.9878 | 0.9803 | 0.9928 | 0.9879 | 0.9799 | 0.9926 |
| 011433 | | | | | | |
| 011434 | 0.9958 | 0.9915 | 0.9978 | 0.9955 | 0.9915 | 0.9978 |
| 011435 | 0.9991 | 0.9969 | 0.9997 | 0.9991 | 0.9969 | 0.9997 |
| 011436 | 0.9673 | 0.9533 | 0.9769 | 0.9961 | 0.9919 | 0.9989 |
| 011442 | 0.7896 | 0.6525 | 0.8826 | 0.5991 | 0.1398 | 0.8769 |
| 011444 | 0.9955 | 0.9904 | 0.9986 | 0.9997 | 0.9963 | 1.0000 |
| 011445 | 0.9998 | 0.9970 | 1.0000 | 0.9998 | 0.9971 | 1.0000 |
| 011446 | | | | | | |
| 011447 | 0.9995 | 0.9961 | 0.9999 | 0.9994 | 0.9960 | 0.9999 |
| 011448 | 0.9998 | 0.9970 | 1.0000 | 0.9999 | 0.9981 | 1.0000 |
| 011874 | 0.9790 | 0.9711 | 0.9852 | 0.9916 | 0.9872 | 0.9950 |
| 011875 | 0.9843 | 0.9434 | 0.9959 | 0.9985 | 0.9656 | 0.9998 |
| 011876 | 0.9972 | 0.9955 | 0.9984 | 0.9973 | 0.9957 | 0.9985 |
| 011877 | 0.9925 | 0.9830 | 0.9975 | 0.9928 | 0.9827 | 0.9975 |
| 011878 | 0.9899 | 0.7874 | 0.9983 | | | |
| 011879 | 0.9230 | 0.7586 | 0.9855 | 0.8376 | 0.1894 | 0.9807 |
| 011880 | 0.9963 | 0.9894 | 0.9996 | 0.9996 | 0.9932 | 0.9999 |
| 022559 | 0.9997 | 0.9940 | 0.9999 | 0.9997 | 0.9955 | 0.9999 |
| 022560 | 0.9690 | 0.9530 | 0.9790 | 0.9872 | 0.9780 | 0.9936 |
| 022561 | 0.9989 | 0.9951 | 0.9998 | 0.9987 | 0.9948 | 0.9998 |
| 022903 | 0.9997 | 0.9946 | 1.0000 | 0.9997 | 0.9947 | 1.0000 |
| 022904 | 0.9906 | 0.7974 | 0.9984 | | | |
| 022910 | 0.9559 | 0.2943 | 0.9915 | | | |
| 022911 | 0.9618 | 0.3285 | 0.9933 | | | |
| 022912 | 0.9594 | 0.2850 | 0.9930 | | | |
| 022913 | 0.9696 | 0.4292 | 0.9949 | | | |
| 022914 | 0.9852 | 0.6857 | 0.9975 | | | |
| 022915 | 0.9852 | 0.9707 | 0.9933 | 0.9851 | 0.9702 | 0.9931 |
| 022916 | 0.9282 | 0.8971 | 0.9501 | 0.9253 | 0.8942 | 0.9484 |
| 022917 | 0.6081 | 0.5429 | 0.6685 | 0.7604 | 0.7112 | 0.7964 |
| 022918 | 0.9867 | 0.7254 | 0.9979 | | | |
| 022919 | 0.9755 | 0.5025 | 0.9957 | | | |
| 022920 | 0.9848 | 0.6898 | 0.9977 | | | |

| ID | Contamix Approx MAP | Contamix Approx LO | Contamix Approx HI | Contamix Precise MAP | Contamix Precise LO | Contamix Precise HI |
|---|---|---|---|---|---|---|
| 022921 | 0.9691 | 0.4563 | 0.9951 | | | |
| 022922 | 0.9782 | 0.8471 | 0.9971 | 0.9737 | 0.5309 | 0.9957 |
| 023501 | 0.9959 | 0.9097 | 0.9994 | 0.9825 | 0.6617 | 0.9972 |
| 023502 | 0.9998 | 0.9962 | 1.0000 | 0.9999 | 0.9978 | 1.0000 |
| 023504 | 0.9998 | 0.9967 | 1.0000 | 0.9998 | 0.9966 | 1.0000 |
| 024054 | 0.9708 | 0.9633 | 0.9775 | 0.9714 | 0.9637 | 0.9779 |
| 024055 | 0.9991 | 0.9814 | 0.9999 | 0.9992 | 0.9821 | 0.9999 |
| 024056 | 0.9955 | 0.9884 | 0.9987 | 0.9954 | 0.9880 | 0.9986 |
| 024057 | 0.9985 | 0.9946 | 0.9998 | 0.9982 | 0.9943 | 0.9997 |
| 024058 | 0.9991 | 0.9959 | 0.9999 | 0.9990 | 0.9958 | 0.9999 |
| 024059 | 0.9993 | 0.9850 | 0.9999 | 0.9994 | 0.9862 | 0.9999 |
| 024060 | 0.9989 | 0.9950 | 0.9997 | 0.9988 | 0.9951 | 0.9997 |
| 024061 | 0.9998 | 0.9964 | 1.0000 | 0.9998 | 0.9965 | 1.0000 |
| 024062 | 0.9971 | 0.9915 | 0.9991 | 0.9970 | 0.9914 | 0.9990 |
| 024063 | 0.9997 | 0.9949 | 0.9999 | 0.9998 | 0.9960 | 1.0000 |
| 024064 | 0.9976 | 0.9923 | 0.9994 | 0.9994 | 0.9940 | 0.9999 |
| 024065 | 0.9990 | 0.9947 | 0.9999 | 0.9991 | 0.9947 | 0.9999 |
| 024066 | 0.9999 | 0.9974 | 1.0000 | 0.9999 | 0.9974 | 1.0000 |
| 024067 | 0.9997 | 0.9957 | 1.0000 | 0.9997 | 0.9957 | 0.9999 |
| 024068 | 0.9992 | 0.9956 | 0.9999 | 0.9993 | 0.9956 | 0.9999 |
| 024069 | 0.9997 | 0.9945 | 0.9999 | 0.9998 | 0.9963 | 1.0000 |
| 024070 | 0.9999 | 0.9977 | 1.0000 | 0.9999 | 0.9977 | 1.0000 |
| 024071 | 0.9993 | 0.9963 | 0.9998 | 0.9992 | 0.9963 | 0.9998 |
| 024072 | 0.9693 | 0.9523 | 0.9778 | 0.9667 | 0.9526 | 0.9775 |
| 024073 | 0.9956 | 0.9868 | 0.9990 | 0.9999 | 0.9971 | 1.0000 |
| 024074 | 0.9857 | 0.9753 | 0.9923 | 0.9871 | 0.9768 | 0.9934 |
| 024075 | 0.9999 | 0.9970 | 1.0000 | 0.9999 | 0.9970 | 1.0000 |
| 024076 | 0.9998 | 0.9967 | 1.0000 | 0.9998 | 0.9967 | 1.0000 |
| 024077 | 0.9994 | 0.9966 | 0.9999 | 0.9994 | 0.9965 | 0.9999 |
| 024078 | 0.9999 | 0.9967 | 1.0000 | 0.9998 | 0.9968 | 1.0000 |
| 024079 | 0.9990 | 0.9949 | 0.9998 | 0.9988 | 0.9949 | 0.9998 |
| 024080 | 0.9536 | 0.9421 | 0.9630 | 0.9535 | 0.9417 | 0.9630 |
| 024081 | 0.9987 | 0.9950 | 0.9997 | 0.9987 | 0.9950 | 0.9997 |
| 024082 | 0.9996 | 0.9927 | 0.9999 | 0.9997 | 0.9938 | 1.0000 |
| 024083 | 0.9999 | 0.9968 | 1.0000 | 0.9999 | 0.9970 | 1.0000 |
| 024084 | 0.9565 | 0.9315 | 0.9735 | 0.9996 | 0.9942 | 0.9999 |
| 024085 | 0.9985 | 0.9941 | 0.9998 | 0.9985 | 0.9942 | 0.9998 |
| 024086 | 0.9999 | 0.9976 | 1.0000 | 0.9999 | 0.9975 | 1.0000 |
| 024088 | 0.9996 | 0.9944 | 0.9999 | 0.9997 | 0.9950 | 1.0000 |
| 024089 | 0.9994 | 0.9925 | 0.9999 | 0.9995 | 0.9926 | 0.9999 |
| 024090 | 0.9711 | 0.9507 | 0.9829 | 0.9996 | 0.9953 | 0.9999 |
| 024091 | 0.9998 | 0.9961 | 1.0000 | 0.9998 | 0.9962 | 1.0000 |

| ID | Contamix Approx MAP | Contamix Approx LO | Contamix Approx HI | Contamix Precise MAP | Contamix Precise LO | Contamix Precise HI |
|---|---|---|---|---|---|---|
| 024092 | 0.9999 | 0.9973 | 1.0000 | 0.9999 | 0.9973 | 1.0000 |
| 024093 | 0.9979 | 0.9923 | 0.9997 | 0.9990 | 0.9940 | 0.9999 |
| 024094 | 0.9999 | 0.9979 | 1.0000 | 0.9999 | 0.9979 | 1.0000 |
| 024095 | 0.9998 | 0.9968 | 1.0000 | 0.9998 | 0.9970 | 1.0000 |
| 024098 | 0.9980 | 0.9930 | 0.9998 | 0.9982 | 0.9929 | 0.9998 |
| 024099 | 0.9999 | 0.9976 | 1.0000 | 0.9999 | 0.9976 | 1.0000 |
| 024100 | 0.9990 | 0.9950 | 0.9998 | 0.9989 | 0.9950 | 0.9999 |
| 024101 | 0.9994 | 0.9955 | 0.9999 | 0.9995 | 0.9956 | 0.9999 |
| 024102 | 0.9999 | 0.9971 | 1.0000 | 0.9999 | 0.9971 | 1.0000 |
| 024103 | 0.9985 | 0.9952 | 0.9996 | 0.9985 | 0.9951 | 0.9996 |
| 024112 | 0.9808 | 0.6228 | 0.9969 | | | |
| 024113 | 0.9691 | 0.4092 | 0.9944 | | | |
| 024114 | 0.9907 | 0.8953 | 0.9987 | 0.9133 | 0.6122 | 0.9855 |
| 024115 | 0.9924 | 0.9855 | 0.9969 | 0.9928 | 0.9853 | 0.9970 |
| 024116 | 0.9965 | 0.9931 | 0.9985 | 0.9966 | 0.9932 | 0.9986 |
| 024117 | 0.9976 | 0.9674 | 0.9996 | 0.9980 | 0.9623 | 0.9997 |
| 024118 | 0.9918 | 0.9791 | 0.9975 | 0.9940 | 0.9816 | 0.9980 |
| 024119 | 0.9236 | 0.0558 | 0.9832 | | | |
| 024120 | 0.9976 | 0.9686 | 0.9996 | 0.9961 | 0.9639 | 0.9995 |
| 024121 | 0.9976 | 0.9921 | 0.9997 | 0.9978 | 0.9920 | 0.9998 |
| 024122 | 0.9900 | 0.9842 | 0.9941 | 0.9947 | 0.9895 | 0.9973 |
| 024123 | 0.9919 | 0.9849 | 0.9967 | 0.9914 | 0.9843 | 0.9961 |
| 024124 | 0.9455 | 0.1784 | 0.9895 | | | |
| 024125 | 0.9984 | 0.9639 | 0.9998 | 0.9975 | 0.9476 | 0.9996 |
| 024126 | 0.9957 | 0.9885 | 0.9984 | 0.9958 | 0.9882 | 0.9983 |
| 024247 | 0.9946 | 0.9892 | 0.9974 | 0.9945 | 0.9893 | 0.9974 |
| 024248 | 0.9911 | 0.8089 | 0.9986 | | | |
| 024249 | 0.9831 | 0.6329 | 0.9971 | | | |
| 024250 | 0.9998 | 0.9961 | 1.0000 | 0.9998 | 0.9960 | 1.0000 |
| 024251 | 0.9977 | 0.9859 | 0.9998 | 0.9992 | 0.9879 | 0.9999 |
| 024252 | 0.9924 | 0.8475 | 0.9987 | | | |
| 024253 | 0.9999 | 0.9974 | 1.0000 | 0.9999 | 0.9972 | 1.0000 |
| 024254 | 0.9973 | 0.9910 | 0.9996 | 0.9993 | 0.9934 | 0.9999 |
| 024255 | 0.9854 | 0.9485 | 0.9962 | 0.9845 | 0.9444 | 0.9962 |
| 024256 | 0.9981 | 0.9939 | 0.9998 | 0.9982 | 0.9938 | 0.9997 |
| 024257 | 0.9969 | 0.9919 | 0.9991 | 0.9976 | 0.9930 | 0.9995 |
| 024258 | 0.9997 | 0.9931 | 0.9999 | 0.9996 | 0.9929 | 0.9999 |
| 024259 | 0.9767 | 0.9316 | 0.9933 | 0.9731 | 0.9200 | 0.9922 |
| 024260 | 0.9881 | 0.9679 | 0.9959 | 0.9884 | 0.9690 | 0.9959 |
| 024261 | 0.9999 | 0.9970 | 1.0000 | 0.9998 | 0.9971 | 1.0000 |
| 024262 | 0.9998 | 0.9962 | 1.0000 | 0.9998 | 0.9963 | 1.0000 |
| 024263 | 0.9962 | 0.9192 | 0.9994 | 0.9954 | 0.8973 | 0.9992 |

| ID | Contamix Approx MAP | Contamix Approx LO | Contamix Approx HI | Contamix Precise MAP | Contamix Precise LO | Contamix Precise HI |
|---|---|---|---|---|---|---|
| 024264 | | | | | | |
| 024265 | 0.9997 | 0.9955 | 0.9999 | 0.9997 | 0.9963 | 1.0000 |
| 024266 | 0.9998 | 0.9963 | 1.0000 | 0.9914 | 0.9852 | 0.9954 |
| 024267 | 0.9675 | 0.9435 | 0.9803 | 0.9954 | 0.9911 | 0.9982 |
| 024268 | 0.9896 | 0.9817 | 0.9960 | 0.9917 | 0.9833 | 0.9967 |
| 024724 | 0.9992 | 0.9847 | 0.9999 | 0.9993 | 0.9851 | 0.9999 |
| 024725 | 0.9651 | 0.9306 | 0.9825 | 0.9575 | 0.9189 | 0.9797 |
| 024726 | 0.9998 | 0.9958 | 1.0000 | 0.9998 | 0.9960 | 1.0000 |
| 024727 | 0.9835 | 0.9681 | 0.9932 | 0.9995 | 0.9912 | 0.9999 |
| 024728 | 0.9997 | 0.9932 | 0.9999 | 0.9997 | 0.9942 | 0.9999 |
| 024729 | 0.9998 | 0.9954 | 1.0000 | 0.9998 | 0.9953 | 1.0000 |
| 024730 | 0.9996 | 0.9931 | 0.9999 | 0.9998 | 0.9956 | 1.0000 |
| 024731 | 0.9764 | 0.9594 | 0.9873 | 0.9998 | 0.9960 | 1.0000 |
| 024732 | 0.9998 | 0.9941 | 1.0000 | 0.9998 | 0.9940 | 1.0000 |
| 024733 | 0.9998 | 0.9967 | 1.0000 | 0.9998 | 0.9964 | 1.0000 |
| 024734 | 0.9998 | 0.9962 | 1.0000 | 0.9999 | 0.9965 | 1.0000 |
| 024735 | 0.9998 | 0.9962 | 1.0000 | 0.9998 | 0.9961 | 1.0000 |
| 024736 | 0.9999 | 0.9969 | 1.0000 | 0.9999 | 0.9978 | 1.0000 |
| 024737 | 0.9998 | 0.9965 | 1.0000 | 0.9999 | 0.9976 | 1.0000 |
| 024738 | 0.9295 | 0.0314 | 0.9801 | | | |
| 024739 | 0.9999 | 0.9973 | 1.0000 | 0.9999 | 0.9973 | 1.0000 |
| 024740 | | | | | | |
| 024741 | 0.9452 | 0.1767 | 0.9891 | | | |
| 024742 | 0.9998 | 0.9959 | 1.0000 | 0.9998 | 0.9959 | 1.0000 |
| 024743 | 0.9999 | 0.9970 | 1.0000 | 0.9999 | 0.9970 | 1.0000 |
| 024744 | 0.9998 | 0.9964 | 1.0000 | 0.9998 | 0.9964 | 1.0000 |
| 024745 | 0.9998 | 0.9966 | 1.0000 | 0.9999 | 0.9964 | 1.0000 |
| 024746 | 0.9997 | 0.9951 | 1.0000 | 0.9997 | 0.9953 | 1.0000 |
| 024747 | 0.9998 | 0.9965 | 1.0000 | 0.9998 | 0.9965 | 1.0000 |
| 024748 | 0.9996 | 0.9918 | 0.9999 | 0.9996 | 0.9917 | 0.9999 |
| 024749 | 0.9999 | 0.9967 | 1.0000 | 0.9998 | 0.9966 | 1.0000 |
| 024750 | 0.9999 | 0.9974 | 1.0000 | 0.9999 | 0.9974 | 1.0000 |
| 024751 | 0.9530 | 0.2039 | 0.9900 | | | |
| 024752 | 0.9999 | 0.9969 | 1.0000 | 0.9999 | 0.9969 | 1.0000 |
| 024753 | | | | | | |
| 024754 | | | | | | |
| 024755 | 0.9990 | 0.9940 | 0.9999 | 0.9997 | 0.9957 | 0.9999 |
| 024756 | 0.9997 | 0.9944 | 1.0000 | 0.9997 | 0.9948 | 1.0000 |
| 024757 | 0.9998 | 0.9961 | 1.0000 | 0.9998 | 0.9961 | 1.0000 |
| 024758 | 0.9542 | 0.9187 | 0.9765 | 0.9990 | 0.9861 | 0.9998 |
| 024759 | 0.9001 | 0.8775 | 0.9178 | 0.8987 | 0.8759 | 0.9166 |
| 024760 | 0.9988 | 0.9934 | 0.9999 | 0.9986 | 0.9934 | 0.9999 |

| ID | Contamix Approx MAP | Contamix Approx LO | Contamix Approx HI | Contamix Precise MAP | Contamix Precise LO | Contamix Precise HI |
|---|---|---|---|---|---|---|
| 024761 | 0.9781 | 0.9625 | 0.9881 | 0.9856 | 0.9745 | 0.9926 |
| 024762 | 0.9971 | 0.9622 | 0.9995 | 0.9989 | 0.9761 | 0.9998 |
| 024763 | 0.9995 | 0.9885 | 0.9999 | 0.9995 | 0.9895 | 0.9999 |
| 024764 | 0.9999 | 0.9974 | 1.0000 | 0.9999 | 0.9975 | 1.0000 |
| 024766 | 0.9999 | 0.9968 | 1.0000 | 0.9999 | 0.9967 | 1.0000 |
| 024767 | 0.9987 | 0.9812 | 0.9998 | 0.9998 | 0.9947 | 1.0000 |
| 024768 | 0.9997 | 0.9941 | 1.0000 | 0.9998 | 0.9943 | 1.0000 |
| 024769 | 0.9769 | 0.9620 | 0.9869 | 0.9959 | 0.9864 | 0.9991 |
| 024770 | 0.9998 | 0.9965 | 1.0000 | 0.9998 | 0.9965 | 1.0000 |
| 024771 | 0.9998 | 0.9966 | 1.0000 | 0.9999 | 0.9970 | 1.0000 |
| 024772 | 0.9841 | 0.9738 | 0.9925 | 0.9854 | 0.9763 | 0.9996 |
| 024773 | 0.9964 | 0.9867 | 0.9996 | 0.9996 | 0.9931 | 0.9999 |
| 024774 | 0.9999 | 0.9970 | 1.0000 | 0.9999 | 0.9978 | 1.0000 |
| 024775 | 0.9999 | 0.9975 | 1.0000 | 0.9999 | 0.9976 | 1.0000 |
| 024776 | 0.9998 | 0.9965 | 1.0000 | 0.9998 | 0.9965 | 1.0000 |
| 024777 | 0.9844 | 0.9299 | 0.9978 | 0.9900 | 0.9258 | 0.9983 |
| 024778 | 0.9997 | 0.9940 | 1.0000 | 0.9997 | 0.9934 | 1.0000 |
| 024779 | 0.9998 | 0.9962 | 1.0000 | 0.9998 | 0.9961 | 1.0000 |
| 024780 | 0.9999 | 0.9973 | 1.0000 | 0.9999 | 0.9972 | 1.0000 |
| 024781 | 0.9999 | 0.9971 | 1.0000 | 0.9999 | 0.9971 | 1.0000 |
| 024782 | 0.9871 | 0.9687 | 0.9967 | 0.9976 | 0.9898 | 0.9997 |
| 024783 | 0.9998 | 0.9964 | 1.0000 | 0.9998 | 0.9968 | 1.0000 |
| 024784 | 0.9998 | 0.9964 | 1.0000 | 0.9999 | 0.9975 | 1.0000 |

**Table S5 – Mitochondrial Contamination Estimates**

List of all tissues included in this project (n=217), sorted by the ID assigned at the CGG and accompanied by the genetic sex assignment (from Table S7), the reported metrics from the mitochondrial contamination estimation tool ANGSD (which estimates nuclear contamination by calculating X-chromosome homozygosity, where males should be below 0.05 and females above 0.1), the number of SNPs included in the estimate, the number of SNP sites flanking the SNPs included in the estimate, and the contamination flag from any contamination estimate. Contamination estimates were generated from the contamination-removed merged genomes, where possible. Otherwise, the contamination estimates were generated on contaminated genomes, where unknown contamination has occurred, or the genomic coverage was too low for the tools for accurate contamination estimations. Contamination estimates begin in Table S4.

| ID | Genetic Sex Assignment | ANGSD X Contamination Estimate | ANGSD X nSNPs | ANGSD X nSNPs with flanking | Flags from either Contamix or ANGSD-X |
|---|---|---|---|---|---|
| 011229 | XY | 0.0165 | 57551 | 517959 | |
| 011230 | XY | 0.2128 | 16314 | 146826 | contaminated |
| 011231 | XX | 0.1117 | 28191 | 253719 | |
| 011234 | XY | 0.0036 | 43103 | 387927 | |
| 011236 | XY | 0.0041 | 44170 | 397530 | |
| 011291 | Not Assigned | 0.0000 | 4 | 36 | |
| 011292 | Not Assigned | 0.0000 | 8 | 72 | |
| 011293 | consistent with XY but not XX | 0.1181 | 493 | 4437 | |
| 011294 | Not Assigned | 0.0050 | 1820 | 16380 | |
| 011295 | | | | | |
| 011299 | XX | 0.2956 | 57753 | 519777 | |
| 011311 | consistent with XY but not XX | | | | contaminated |
| 011312 | consistent with XY but not XX | | | | |
| 011313 | consistent with XY but not XX | 0.0000 | 1 | 9 | |
| 011316 | XX | 0.1926 | 19497 | 175473 | |
| 011317 | | | | | known contamination |
| 011319 | Not Assigned | 0.0173 | 331 | 2979 | |
| 011320 | XX | 0.1835 | 13874 | 124866 | |
| 011321 | | | | | |
| 011322 | XX | 0.1772 | 20986 | 188874 | |
| 011323 | | | | | |
| 011324 | Not Assigned | 0.0083 | 501 | 4509 | |
| 011325 | consistent with XY but not XX | | 0 | 0 | |
| 011327 | | | | | known contamination |
| 011332 | Not Assigned | 0.0094 | 623 | 5607 | |
| 011335 | | | | | |
| 011336 | XX | 0.1688 | 97 | 873 | |
| 011337 | Not Assigned | 0.0092 | 673 | 6057 | |
| 011339 | XX | 0.1910 | 34951 | 314559 | |
| 011341 | Not Assigned | -0.0046 | 161 | 1449 | |
| 011342 | | | | | |
| 011343 | Not Assigned | 0.0046 | 8971 | 80739 | |
| 011414 | XY | 0.0188 | 23375 | 210375 | |
| 011415 | | | | | |
| 011423 | XY | 0.0070 | 57511 | 517599 | |
| 011433 | | | | | |
| 011434 | XY | 0.0027 | 57756 | 519804 | |
| 011435 | XY | 0.0054 | 57680 | 519120 | |
| 011436 | XX | 0.1961 | 57777 | 519993 | |
| 011442 | XY | 0.0770 | 11 | 99 | contaminated |

| ID | Genetic Sex Assignment | ANGSD X Contamination Estimate | ANGSD X nSNPs | ANGSD X nSNPs with flanking | Flags from either Contamix or ANGSD-X |
|---|---|---|---|---|---|
| 011444 | XX | 0.2931 | 57657 | 518913 | |
| 011445 | XX | 0.1396 | 30076 | 270684 | |
| 011446 | | | | | |
| 011447 | XY | 0.0289 | 11319 | 101871 | |
| 011448 | XY | 0.0041 | 57757 | 519813 | |
| 011874 | XY | 0.0121 | 49523 | 445707 | |
| 011875 | XX | -0.0215 | 44 | 396 | |
| 011876 | XY | 0.0098 | 57191 | 514719 | |
| 011877 | XY | 0.0118 | 807 | 7263 | |
| 011878 | consistent with XY but not XX | | 0 | 0 | contaminated |
| 011879 | Not Assigned | 0.0000 | 2 | 18 | contaminated |
| 011880 | XY | 0.0090 | 6172 | 55548 | |
| 022559 | XY | 0.0124 | 44904 | 404136 | |
| 022560 | XY | 0.0124 | 21116 | 190044 | |
| 022561 | XX | 0.2927 | 57810 | 520290 | |
| 022903 | XY | 0.0089 | 25523 | 229707 | |
| 022904 | XX | | 0 | 0 | contaminated |
| 022910 | consistent with XY but not XX | | 0 | 0 | contaminated |
| 022911 | Not Assigned | | 0 | 0 | contaminated |
| 022912 | Not Assigned | 0.0000 | 1 | 9 | contaminated |
| 022913 | Not Assigned | | 0 | 0 | contaminated |
| 022914 | Not Assigned | | 0 | 0 | contaminated |
| 022915 | XX | 0.0710 | 19 | 171 | contaminated |
| 022916 | XX | 0.1416 | 595 | 5355 | contaminated |
| 022917 | Not Assigned | -0.0124 | 15 | 135 | contaminated |
| 022918 | XX | | 0 | 0 | contaminated |
| 022919 | Not Assigned | | 0 | 0 | contaminated |
| 022920 | Not Assigned | | 0 | 0 | contaminated |
| 022921 | Not Assigned | | 0 | 0 | contaminated |
| 022922 | Not Assigned | | 0 | 0 | contaminated |
| 023501 | consistent with XY but not XX | | 0 | 0 | |
| 023502 | XX | 0.1614 | 7631 | 68679 | |
| 023504 | XX | 0.1878 | 18975 | 170775 | |
| 024054 | XY | 0.0051 | 54821 | 493389 | |
| 024055 | Not Assigned | | 0 | 0 | |
| 024056 | XX | 0.1922 | 13096 | 117864 | |
| 024057 | XX | 0.2404 | 57202 | 514818 | |
| 024058 | XX | 0.3115 | 57809 | 520281 | |
| 024059 | XY | | 0 | 0 | |
| 024060 | XX | 0.2160 | 55455 | 499095 | |
| 024061 | XX | 0.1839 | 29118 | 262062 | |
| 024062 | XY | 0.0032 | 29865 | 268785 | |
| 024063 | XX | 0.1433 | 46973 | 422757 | |
| 024064 | XX | 0.1558 | 32017 | 288153 | |
| 024065 | XX | 0.2522 | 53277 | 479493 | |
| 024066 | XY | 0.0036 | 50950 | 458550 | |
| 024067 | XY | 0.0053 | 19148 | 172332 | |
| 024068 | XX | 0.1885 | 56905 | 512145 | |
| 024069 | XX | 0.1835 | 57779 | 520011 | |
| 024070 | XX | 0.2541 | 57728 | 519552 | |
| 024071 | XX | 0.2821 | 57727 | 519543 | |
| 024072 | Not Assigned | 0.1750 | 3238 | 29142 | contaminated |
| 024073 | XX | 0.2487 | 57681 | 519129 | |
| 024074 | XX | 0.2284 | 40580 | 365220 | |
| 024075 | XX | 0.2180 | 57676 | 519084 | |
| 024076 | XX | 0.2148 | 54690 | 492210 | |

| ID | Genetic Sex Assignment | ANGSD X Contamination Estimate | ANGSD X nSNPs | ANGSD X nSNPs with flanking | Flags from either Contamix or ANGSD-X |
|---|---|---|---|---|---|
| 024077 | XX | 0.2788 | 57698 | 519282 | |
| 024078 | XX | 0.0819 | 34667 | 312003 | |
| 024079 | XX | 0.1673 | 56286 | 506574 | |
| 024080 | XX | 0.3431 | 57786 | 520074 | contaminated |
| 024081 | XY | 0.0026 | 57708 | 519372 | |
| 024082 | XY | 0.0101 | 28870 | 259830 | |
| 024083 | XX | 0.2214 | 56171 | 505539 | |
| 024084 | XX | 0.3140 | 57746 | 519714 | |
| 024085 | XY | 0.0030 | 57718 | 519462 | |
| 024086 | XY | 0.0032 | 56142 | 505278 | |
| 024088 | XX | 0.1938 | 56445 | 508005 | |
| 024089 | XX | 0.2306 | 57394 | 516546 | |
| 024090 | XX | 0.2407 | 57772 | 519948 | |
| 024091 | XX | 0.1793 | 55041 | 495369 | |
| 024092 | XX | 0.2251 | 57694 | 519246 | |
| 024093 | XY | 0.0034 | 57605 | 518445 | |
| 024094 | XY | 0.0031 | 57745 | 519705 | |
| 024095 | XX | 0.2971 | 57804 | 520236 | |
| 024098 | XX | 0.3076 | 57785 | 520065 | |
| 024099 | XX | 0.3133 | 57767 | 519903 | |
| 024100 | XX | 0.2772 | 57761 | 519849 | |
| 024101 | XY | 0.0155 | 1419 | 12771 | |
| 024102 | XY | 0.0077 | 27228 | 245052 | |
| 024103 | XY | 0.0020 | 57731 | 519579 | |
| 024112 | XX | | 0 | 0 | contaminated |
| 024113 | Not Assigned | 0.0000 | 3 | 27 | contaminated |
| 024114 | consistent with XY but not XX | | 0 | 0 | contaminated |
| 024115 | XX | 0.2855 | 57723 | 519507 | |
| 024116 | XX | 0.1831 | 56811 | 511299 | |
| 024117 | XY | 0.0502 | 22 | 198 | contaminated |
| 024118 | XY | -0.0021 | 76 | 684 | |
| 024119 | Not Assigned | | 0 | 0 | contaminated |
| 024120 | XY | 0.0467 | 17 | 153 | contaminated |
| 024121 | XX | 0.1830 | 4 | 36 | |
| 024122 | XX | 0.2766 | 57700 | 519300 | |
| 024123 | XX | 0.1392 | 21741 | 195669 | |
| 024124 | consistent with XY but not XX | | 0 | 0 | contaminated |
| 024125 | Not Assigned | | 0 | 0 | contaminated |
| 024126 | XY | 0.0070 | 21981 | 197829 | |
| 024247 | XX | 0.2162 | 49348 | 444132 | |
| 024248 | consistent with XY but not XX | | 0 | 0 | contaminated |
| 024249 | XX | | 0 | 0 | contaminated |
| 024250 | XX | 0.0803 | 6425 | 57825 | |
| 024251 | XY | -0.0034 | 86 | 774 | |
| 024252 | XX | 0.0000 | 1 | 9 | contaminated |
| 024253 | XX | 0.2663 | 56824 | 511416 | |
| 024254 | XY | 0.0045 | 49535 | 445815 | |
| 024255 | XX | 0.1305 | 47 | 423 | contaminated |
| 024256 | XY | 0.0050 | 57695 | 519255 | |
| 024257 | XY | 0.0000 | 6 | 54 | |
| 024258 | XX | | 0 | 0 | |
| 024259 | XY | -0.0090 | 19 | 171 | contaminated |
| 024260 | XY | 0.0213 | 104 | 936 | |
| 024261 | XY | 0.0039 | 57676 | 519084 | |
| 024262 | XX | 0.1872 | 31265 | 281385 | |
| 024263 | XX | 0.0000 | 2 | 18 | |

| ID | Genetic Sex Assignment | ANGSD X Contamination Estimate | ANGSD X nSNPs | ANGSD X nSNPs with flanking | Flags from either Contamix or ANGSD-X |
|---|---|---|---|---|---|
| 024264 | | | | | |
| 024265 | XY | 0.0026 | 57606 | 518454 | |
| 024266 | XX | 0.2862 | 57685 | 519165 | |
| 024267 | XX | 0.2980 | 57766 | 519894 | |
| 024268 | XX | 0.1745 | 29500 | 265500 | |
| 024724 | XX | 0.3194 | 56615 | 509535 | |
| 024725 | consistent with XY but not XX | nan | | | |
| 024726 | Not Assigned | 0.2474 | 33448 | 301032 | |
| 024727 | XX | 0.2261 | 49130 | 442170 | |
| 024728 | consistent with XY but not XX | -0.0166 | 10 | 90 | |
| 024729 | Not Assigned | 0.0000 | 1 | 9 | |
| 024730 | Not Assigned | 0.0030 | 4459 | 40131 | |
| 024731 | XY | 0.0024 | 30267 | 272403 | |
| 024732 | Not Assigned | 0.0000 | 1 | 9 | |
| 024733 | Not Assigned | 0.1020 | 67 | 603 | contaminated |
| 024734 | XX | 0.1574 | 250 | 2250 | |
| 024735 | XY | 0.0279 | 39 | 351 | |
| 024736 | XX | 0.1728 | 41822 | 376398 | |
| 024737 | Not Assigned | 0.0041 | 5972 | 53748 | |
| 024738 | consistent with XY but not XX | | 0 | 0 | contaminated |
| 024739 | XX | 0.1494 | 4062 | 36558 | |
| 024740 | | | | | |
| 024741 | consistent with XY but not XX | | 0 | 0 | contaminated |
| 024742 | Not Assigned | -0.0333 | 13 | 117 | |
| 024743 | Not Assigned | 0.0032 | 2462 | 22158 | |
| 024744 | Not Assigned | 0.1738 | 634 | 5706 | contaminated |
| 024745 | XX | 0.1766 | 607 | 5463 | |
| 024746 | XX | 0.1915 | 1035 | 9315 | |
| 024747 | XX | 0.1992 | 26784 | 241056 | |
| 024748 | XY | 0.0028 | 4288 | 38592 | |
| 024749 | XX | 0.1773 | 3877 | 34893 | |
| 024750 | XX | 0.2183 | 46737 | 420633 | |
| 024751 | Not Assigned | | 0 | 0 | |
| 024752 | XX | 0.1603 | 9477 | 85293 | |
| 024753 | | | | | |
| 024754 | | | | | |
| 024755 | consistent with XY but not XX | 0.0182 | 3337 | 30033 | |
| 024756 | XY | 0.0066 | 1563 | 14067 | |
| 024757 | XY | 0.0050 | 5497 | 49473 | |
| 024758 | Not Assigned | 0.0104 | 4806 | 43254 | |
| 024759 | XY | 0.0113 | 850 | 7650 | contaminated |
| 024760 | XY | 0.1382 | 5958 | 53622 | contaminated |
| 024761 | Not Assigned | 0.0084 | 4462 | 40158 | |
| 024762 | XY | 0.1828 | 10158 | 91422 | contaminated |
| 024763 | Not Assigned | 0.0004 | 1387 | 12483 | |
| 024764 | XX | 0.2088 | 24659 | 221931 | |
| 024766 | XX | 0.2441 | 57390 | 516510 | |
| 024767 | Not Assigned | 0.0016 | 2597 | 23373 | |
| 024768 | XY | 0.0029 | 42580 | 383220 | |
| 024769 | consistent with XY but not XX | 0.0155 | 3366 | 30294 | |
| 024770 | XY | 0.0127 | 7883 | 70947 | |
| 024771 | XX | 0.1699 | 9738 | 87642 | |
| 024772 | XY | 0.0021 | 11977 | 107793 | |
| 024773 | XY | 0.0261 | 8131 | 73179 | |
| 024774 | XX | 0.0520 | 28592 | 257328 | |
| 024775 | XX | 0.1781 | 16200 | 145800 | |

| ID | Genetic Sex Assignment | ANGSD X Contamination Estimate | ANGSD X nSNPs | ANGSD X nSNPs with flanking | Flags from either Contamix or ANGSD-X |
|---|---|---|---|---|---|
| 024776 | XY | 0.0016 | 8416 | 75744 | |
| 024777 | XY | -0.0106 | 77 | 693 | contaminated |
| 024778 | consistent with XY but not XX | 0.0000 | 2 | 18 | |
| 024779 | XX | 0.2215 | 42079 | 378711 | |
| 024780 | XY | 0.0046 | 6668 | 60012 | |
| 024781 | XX | 0.1436 | 344 | 3096 | |
| 024782 | Not Assigned | 0.0182 | 610 | 5490 | |
| 024783 | XX | 0.1764 | 19761 | 177849 | |
| 024784 | XX | 0.1557 | 3226 | 29034 | |

## Table S6 – Genomic Library Information

List of all tissues included in this project (n=217), sorted by the ID assigned at the CGG and accompanied by the number of genomic libraries constructed, the number of remaining libraries after filtering for contamination, the amount of post-mortem DNA damage (reported as the frequency of C to T transitions at the 5' end of each read), the average percentage of endogenous DNA across all libraries for that individual (contaminated libraries excluded), the autosomal depth of coverage (contaminated libraries excluded), and information about known sources of contamination (from lab errors), relatedness, and species identification.

| ID | Number of Libraries | Number of Libraries after Contamination Removed | Damage (5pCtoT Frequency) | Average Endogenous Percentage | Autosomal Depth | Known Contamination, Relatedness, and Species Information |
|---|---|---|---|---|---|---|
| 011229 | 19 | 19 | 0.20 | 9.42 | 30.36 | |
| 011230 | 20 | 19 | 0.12 | 9.03 | 2.04 | |
| 011231 | 15 | 15 | 0.15 | 3.07 | 1.63 | |
| 011234 | 12 | 9 | 0.05 | 12.26 | 5.66 | |
| 011236 | 14 | 11 | 0.09 | 11.14 | 5.87 | |
| 011291 | 15 | 11 | 0.02 | 0.48 | 0.02 | |
| 011292 | 15 | 15 | 0.03 | 0.55 | 0.03 | |
| 011293 | 15 | 10 | 0.03 | 5.28 | 0.25 | |
| 011294 | 25 | 20 | 0.02 | 3.03 | 0.51 | |
| 011295 | 16 | 12 | 0.03 | 2.62 | | Identical Individual to/merged with CGG011293 |
| 011299 | 13 | 13 | 0.07 | 38.23 | 35.92 | |
| 011311 | 3 | 3 | #N/A | #N/A | 0.00 | |
| 011312 | 1 | 1 | 0.01 | 0.02 | 0.00 | |
| 011313 | 1 | 1 | 0.03 | 3.67 | 0.01 | |
| 011316 | 2 | 2 | 0.02 | 37.75 | 1.34 | |
| 011317 | 1 | #N/A | #N/A | #N/A | | known contamination |
| 011319 | 2 | 2 | 0.04 | 11.96 | 0.21 | |
| 011320 | 3 | 3 | 0.03 | 23.93 | 1.12 | |
| 011321 | 2 | 2 | 0.03 | 9.60 | | Identical Individual to/merged with CGG011322 |
| 011322 | 2 | 2 | 0.03 | 13.31 | 1.41 | |
| 011323 | 2 | 2 | 0.02 | 21.91 | | Identical Individual to/merged with CGG024736 |
| 011324 | 2 | 2 | 0.03 | 4.85 | 0.27 | |
| 011325 | 1 | 1 | 0.02 | 0.04 | 0.00 | |
| 011327 | 1 | #N/A | #N/A | #N/A | | known contamination |
| 011332 | 2 | 2 | 0.02 | 3.98 | 0.31 | |
| 011335 | 2 | 2 | 0.02 | 20.23 | | Identical Individual to/merged with CGG023502 |
| 011336 | 2 | 2 | 0.03 | 7.55 | 0.05 | |
| 011337 | 2 | 2 | 0.03 | 2.90 | 0.34 | |
| 011339 | 2 | 2 | 0.02 | 32.55 | 2.25 | |
| 011341 | 2 | 2 | 0.03 | 1.71 | 0.15 | |
| 011342 | 2 | 2 | 0.02 | 11.33 | | Identical Individual to/merged with CGG024737 |
| 011343 | 2 | 2 | 0.03 | 25.44 | 1.51 | |
| 011414 | 21 | 9 | 0.16 | 8.01 | 2.73 | |
| 011415 | 6 | 6 | 0.17 | 1.08 | | Identical Individual to/merged with CGG011414 |
| 011423 | 37 | 30 | 0.09 | 12.77 | 18.93 | |
| 011433 | 10 | 9 | 0.23 | 1.34 | | Identical Individual to/merged with CGG011423 |
| 011434 | 8 | 8 | 0.04 | 63.49 | 33.58 | |
| 011435 | 32 | 29 | 0.06 | 21.58 | 26.83 | |
| 011436 | 44 | 22 | 0.12 | 12.29 | 34.10 | |
| 011442 | 15 | 15 | 0.04 | 0.16 | 0.05 | |
| 011444 | 20 | 20 | 0.09 | 47.10 | 26.68 | |
| 011445 | 12 | 12 | 0.17 | 11.57 | 1.70 | |
| 011446 | 25 | 21 | 0.18 | 8.28 | | Identical Individual to/merged with CGG011229 |
| 011447 | 10 | 7 | 0.18 | 9.83 | 1.62 | |
| 011448 | 18 | 17 | 0.09 | 23.85 | 44.15 | |
| 011874 | 13 | 13 | 0.06 | 18.13 | 16.43 | |
| 011875 | 13 | 13 | 0.06 | 0.15 | 0.02 | *Rangifer tarandus* |
| 011876 | 16 | 16 | 0.15 | 32.51 | 18.81 | |

| ID | Number of Libraries | Number of Libraries after Contamination Removed | Damage (5pCtoT Frequency) | Average Endogenous Percentage | Autosomal Depth | Known Contamination, Relatedness, and Species Information |
|---|---|---|---|---|---|---|
| 011877 | 14 | 11 | 0.13 | 0.86 | 0.35 | |
| 011878 | 12 | 12 | 0.09 | 0.03 | 0.00 | |
| 011879 | 9 | 9 | 0.11 | 0.14 | 0.02 | |
| 011880 | 17 | 15 | 0.20 | 1.03 | 1.13 | |
| 022559 | 17 | 15 | 0.16 | 5.25 | 7.18 | |
| 022560 | 16 | 14 | 0.16 | 8.37 | 3.47 | |
| 022561 | 21 | 20 | 0.15 | 32.71 | 40.29 | |
| 022903 | 12 | 12 | 0.10 | 9.96 | 3.06 | |
| 022904 | 6 | 6 | 0.12 | 0.03 | 0.00 | |
| 022910 | 6 | 6 | 0.01 | 0.01 | 0.00 | |
| 022911 | 6 | 6 | 0.01 | 0.04 | 0.00 | |
| 022912 | 6 | 6 | 0.01 | 0.06 | 0.00 | |
| 022913 | 6 | 6 | 0.01 | 0.03 | 0.00 | |
| 022914 | 6 | 6 | 0.01 | 0.06 | 0.00 | |
| 022915 | 7 | 6 | 0.01 | 0.52 | 0.02 | |
| 022916 | 4 | 4 | 0.01 | 0.52 | 0.13 | |
| 022917 | 34 | 32 | 0.03 | 0.07 | 0.04 | |
| 022918 | 6 | 5 | 0.05 | 0.01 | 0.00 | |
| 022919 | 6 | 5 | 0.02 | 0.10 | 0.00 | |
| 022920 | 6 | 6 | 0.01 | 0.25 | 0.01 | |
| 022921 | 6 | 6 | 0.01 | 0.13 | 0.00 | |
| 022922 | 6 | 6 | 0.01 | 0.11 | 0.00 | |
| 023501 | 2 | 1 | 0.02 | 0.05 | 0.00 | |
| 023502 | 3 | 3 | 0.03 | 13.01 | 0.79 | |
| 023504 | 4 | 4 | 0.02 | 24.88 | 1.58 | |
| 024054 | 5 | 5 | 0.07 | 25.71 | 9.91 | |
| 024055 | 2 | 2 | 0.12 | 0.63 | 0.02 | |
| 024056 | 6 | 6 | 0.22 | 20.51 | 0.92 | |
| 024057 | 5 | 5 | 0.06 | 49.74 | 7.14 | |
| 024058 | 9 | 9 | 0.05 | 60.06 | 41.52 | |
| 024059 | 2 | 2 | 0.13 | 0.48 | 0.01 | |
| 024060 | 4 | 4 | 0.09 | 57.44 | 5.03 | |
| 024061 | 5 | 5 | 0.11 | 13.25 | 1.73 | |
| 024062 | 5 | 5 | 0.09 | 54.18 | 3.47 | |
| 024063 | 5 | 5 | 0.08 | 19.48 | 3.09 | |
| 024064 | 5 | 5 | 0.07 | 7.38 | 1.97 | First degree relative of 024068 |
| 024065 | 4 | 3 | 0.07 | 34.79 | 4.18 | |
| 024066 | 5 | 5 | 0.09 | 33.74 | 7.30 | |
| 024067 | 5 | 5 | 0.09 | 7.13 | 2.41 | First degree relative of 024066 |
| 024068 | 5 | 5 | 0.06 | 47.16 | 10.37 | |
| 024069 | 11 | 11 | 0.05 | 61.54 | 42.12 | |
| 024070 | 4 | 4 | 0.04 | 52.98 | 14.63 | |
| 024071 | 5 | 5 | 0.05 | 50.58 | 20.01 | |
| 024072 | 9 | 9 | 0.13 | 1.09 | 0.41 | |
| 024073 | 9 | 9 | 0.09 | 26.32 | 12.01 | |
| 024074 | 4 | 4 | 0.12 | 15.26 | 2.54 | |
| 024075 | 5 | 5 | 0.08 | 46.47 | 10.39 | |
| 024076 | 5 | 4 | #N/A | 31.18 | 5.17 | |
| 024077 | 5 | 5 | 0.08 | 54.23 | 14.11 | |
| 024078 | 5 | 4 | #N/A | 32.91 | 1.94 | |
| 024079 | 5 | 5 | 0.06 | 10.16 | 5.99 | |
| 024080 | 6 | 6 | 0.05 | 49.76 | 38.91 | |
| 024081 | 5 | 5 | 0.05 | 51.81 | 24.61 | |
| 024082 | 5 | 5 | 0.10 | 14.94 | 3.39 | |
| 024083 | 5 | 5 | 0.10 | 39.10 | 5.61 | |
| 024084 | 11 | 11 | 0.09 | 23.98 | 27.85 | |
| 024085 | 10 | 10 | 0.07 | 53.31 | 38.05 | |
| 024086 | 5 | 5 | 0.07 | 15.50 | 11.08 | |
| 024088 | 5 | 5 | 0.07 | 16.89 | 6.19 | |
| 024089 | 9 | 9 | 0.11 | 24.61 | 8.21 | |
| 024090 | 5 | 2 | 0.07 | 48.73 | 18.15 | |
| 024091 | 5 | 5 | 0.10 | 24.10 | 5.06 | |

| ID | Number of Libraries | Number of Libraries after Contamination Removed | Damage (5pCtoT Frequency) | Average Endogenous Percentage | Autosomal Depth | Known Contamination, Relatedness, and Species Information |
|---|---|---|---|---|---|---|
| 024092 | 5 | 5 | 0.08 | 21.11 | 11.99 | |
| 024093 | 10 | 9 | 0.06 | 19.13 | 19.37 | |
| 024094 | 9 | 8 | 0.07 | 61.87 | 56.73 | |
| 024095 | 5 | 5 | 0.09 | 52.11 | 33.20 | |
| 024098 | 11 | 11 | 0.08 | 40.02 | 34.57 | |
| 024099 | 5 | 5 | 0.05 | 56.23 | 27.00 | |
| 024100 | 5 | 5 | 0.08 | 54.06 | 13.35 | |
| 024101 | 11 | 10 | 0.12 | 1.97 | 0.48 | |
| 024102 | 5 | 5 | 0.10 | 13.90 | 3.32 | |
| 024103 | 5 | 5 | 0.05 | 61.37 | 44.91 | |
| 024112 | 6 | 6 | 0.08 | 0.07 | 0.00 | |
| 024113 | 6 | 6 | 0.05 | 0.03 | 0.00 | |
| 024114 | 7 | 7 | 0.05 | 0.09 | 0.00 | |
| 024115 | 10 | 10 | 0.10 | 24.48 | 16.29 | |
| 024116 | 11 | 11 | 0.07 | 19.44 | 7.02 | |
| 024117 | 9 | 8 | 0.12 | 1.23 | 0.06 | |
| 024118 | 9 | 8 | 0.04 | 0.68 | 0.11 | |
| 024119 | 6 | 6 | 0.06 | 0.00 | 0.00 | *Ursus (*?) |
| 024120 | 12 | 12 | 0.10 | 0.65 | 0.05 | |
| 024121 | 6 | 4 | 0.15 | 0.15 | 0.01 | |
| 024122 | 18 | 17 | 0.10 | 22.37 | 23.37 | |
| 024123 | 10 | 9 | 0.12 | 10.15 | 1.85 | |
| 024124 | 14 | 14 | 0.04 | 0.56 | 0.00 | Unidentified species |
| 024125 | 14 | 14 | 0.05 | 3.86 | 0.00 | |
| 024126 | 23 | 19 | 0.33 | 4.65 | 2.66 | |
| 024247 | 15 | 11 | 0.17 | 10.36 | 3.50 | |
| 024248 | 6 | 6 | 0.27 | 0.14 | 0.00 | |
| 024249 | 6 | 6 | 0.15 | 0.04 | 0.00 | |
| 024250 | 8 | 4 | #N/A | 2.95 | 0.59 | |
| 024251 | 13 | 8 | 0.16 | 0.81 | 0.10 | |
| 024252 | 6 | 6 | 0.09 | 0.05 | 0.00 | |
| 024253 | 17 | 16 | 0.12 | 8.19 | 6.32 | |
| 024254 | 15 | 15 | 0.08 | 10.98 | 6.74 | |
| 024255 | 6 | 6 | 0.18 | 0.08 | 0.03 | |
| 024256 | 21 | 20 | 0.07 | 49.16 | 49.57 | |
| 024257 | 12 | 12 | 0.11 | 0.15 | 0.04 | |
| 024258 | 5 | 5 | 0.13 | 0.01 | 0.00 | |
| 024259 | 11 | 11 | 0.18 | 0.65 | 0.05 | |
| 024260 | 12 | 8 | 0.23 | 1.11 | 0.13 | |
| 024261 | 23 | 23 | 0.08 | 25.15 | 26.05 | |
| 024262 | 10 | 10 | 0.13 | 6.24 | 1.95 | |
| 024263 | 5 | 4 | 0.12 | 0.16 | 0.00 | *Ovis nivicola* |
| 024264 | 19 | 16 | 0.18 | 3.36 | | Identical Individual to/merged with CGG024247 |
| 024265 | 13 | 13 | 0.07 | 16.32 | 20.78 | |
| 024266 | 18 | 18 | 0.09 | 16.87 | 14.71 | |
| 024267 | 22 | 20 | 0.12 | 12.92 | 19.97 | |
| 024268 | 9 | 9 | 0.15 | 2.27 | 1.73 | |
| 024724 | 10 | 4 | #N/A | 30.30 | 19.52 | |
| 024725 | 1 | 1 | 0.02 | 6.79 | 0.00 | |
| 024726 | 7 | 6 | 0.02 | 21.82 | 4.04 | |
| 024727 | 2 | 2 | 0.02 | 38.80 | 5.96 | |
| 024728 | 1 | 1 | 0.03 | 1.88 | 0.03 | |
| 024729 | 1 | 1 | 0.03 | 5.43 | 0.01 | |
| 024730 | 2 | 2 | 0.02 | 26.32 | 1.01 | |
| 024731 | 2 | 2 | 0.02 | 39.20 | 5.06 | |
| 024732 | 1 | 1 | 0.03 | 1.49 | 0.01 | |
| 024733 | 2 | 2 | 0.03 | 4.27 | 0.08 | |
| 024734 | 2 | 2 | 0.03 | 34.61 | 0.08 | |
| 024735 | 2 | 1 | 0.03 | 18.26 | 0.07 | |
| 024736 | 2 | 2 | 0.02 | 33.57 | 3.96 | |
| 024737 | 2 | 2 | 0.02 | 26.09 | 1.14 | |
| 024738 | 1 | 1 | 0.02 | 0.03 | 0.00 | |

| ID | Number of Libraries | Number of Libraries after Contamination Removed | Damage (5pCtoT Frequency) | Average Endogenous Percentage | Autosomal Depth | Known Contamination, Relatedness, and Species Information |
|---|---|---|---|---|---|---|
| 024739 | 2 | 2 | 0.02 | 39.64 | 0.43 | First degree relative of 024731 and 024740 |
| 024740 | 2 | 2 | 0.03 | 31.59 | | Identical Individual to/merged with CGG024731 |
| 024741 | 1 | 1 | 0.02 | 0.01 | 0.00 | |
| 024742 | 1 | 1 | 0.03 | 26.00 | 0.03 | |
| 024743 | 2 | 2 | 0.02 | 26.93 | 0.67 | |
| 024744 | 2 | 2 | 0.03 | 4.77 | 0.21 | |
| 024745 | 2 | 2 | 0.03 | 8.32 | 0.14 | |
| 024746 | 3 | 3 | 0.03 | 10.58 | 0.18 | |
| 024747 | 2 | 2 | 0.03 | 34.42 | 1.95 | |
| 024748 | 2 | 2 | 0.01 | 21.74 | 1.05 | |
| 024749 | 2 | 2 | 0.02 | 6.51 | 0.45 | First degree relative of 011321 and 011322 |
| 024750 | 2 | 2 | 0.03 | 38.13 | 5.20 | |
| 024751 | 1 | 1 | 0.01 | 0.01 | 0.00 | |
| 024752 | 2 | 2 | 0.03 | 31.22 | 0.80 | |
| 024753 | 2 | 2 | 0.01 | 18.62 | | Identical Individual to/merged with CGG024750 |
| 024754 | 2 | 2 | 0.03 | 4.39 | | Identical Individual to/merged with CGG024755 |
| 024755 | 2 | 2 | 0.02 | 14.37 | 0.83 | |
| 024756 | 2 | 2 | 0.03 | 16.95 | 0.51 | First degree relative of 024754 and 024755 |
| 024757 | 2 | 2 | 0.03 | 18.95 | 1.30 | First degree relative of 024773 |
| 024758 | 2 | 2 | 0.03 | 30.01 | 1.11 | |
| 024759 | 1 | 1 | 0.02 | 38.19 | 0.39 | |
| 024760 | 2 | 2 | 0.03 | 43.08 | 1.21 | |
| 024761 | 2 | 2 | 0.03 | 14.63 | 1.05 | First degree relative of 024775 |
| 024762 | 2 | 2 | 0.01 | 15.07 | 1.83 | |
| 024763 | 2 | 2 | 0.02 | 39.06 | 0.51 | First degree relative of 011319 and 024767 |
| 024764 | 2 | 2 | 0.02 | 36.80 | 1.99 | |
| 024766 | 10 | 10 | 0.02 | 41.83 | 25.04 | |
| 024767 | 3 | 3 | 0.02 | 21.46 | 0.70 | |
| 024768 | 7 | 7 | 0.02 | 30.12 | 7.59 | |
| 024769 | 2 | 2 | 0.03 | 37.12 | 0.80 | |
| 024770 | 2 | 2 | 0.02 | 20.23 | 1.53 | |
| 024771 | 2 | 2 | 0.03 | 40.37 | 0.76 | |
| 024772 | 2 | 2 | 0.02 | 31.28 | 2.06 | |
| 024773 | 2 | 2 | 0.02 | 31.85 | 1.63 | |
| 024774 | 2 | 2 | 0.02 | 37.22 | 2.24 | |
| 024775 | 2 | 2 | 0.02 | 39.20 | 1.47 | |
| 024776 | 2 | 2 | 0.02 | 29.31 | 1.41 | First degree relative of 024779 |
| 024777 | 1 | 1 | 0.04 | 26.35 | 0.06 | |
| 024778 | 1 | 1 | 0.02 | 0.60 | 0.01 | |
| 024779 | 2 | 2 | 0.02 | 31.90 | 3.51 | |
| 024780 | 2 | 2 | 0.01 | 18.56 | 1.45 | |
| 024781 | 2 | 2 | 0.02 | 10.58 | 0.11 | |
| 024782 | 2 | 2 | 0.03 | 14.14 | 0.32 | |
| 024783 | 2 | 2 | 0.03 | 18.56 | 1.64 | |
| 024784 | 2 | 2 | 0.03 | 8.38 | 0.42 | |

**Table S7 – Genetic Sex Assignments**

List of all tissues included in this project (n=217), sorted by the ID assigned at the CGG and accompanied by the total number of mapping reads from the merged libraries, the combined number of reads mapping to the sex chromosomes, the number of reads mapping to the Y chromosome, the number of reads mapped to the Y chromosome divided by the total number of reads mapped to the sex chromosomes (termed R_y), the genetic sex assignment, and the 95% confidence interval for the sex assignment. See the last column of Table S6 for information on the IDs with blank entries (*e.g.* either non-human, or merged with another sample due to sample duplication).

| ID | Nseqs | NchrY+NchrX | NchrY | R_y | Genetic Sex Assignment | 95% CI |
|---|---|---|---|---|---|---|
| 011229 | 1908815726 | 53989889 | 4218872 | 0.0781 | XY | 0.0781-0.0782 |
| 011230 | 115460230 | 3399536 | 257294 | 0.0757 | XY | 0.0754-0.076 |
| 011231 | 73787067 | 3771804 | 9481 | 0.0025 | XX | 0.0025-0.0026 |
| 011234 | 247532286 | 7048541 | 590442 | 0.0838 | XY | 0.0836-0.084 |
| 011236 | 235527633 | 6787003 | 580132 | 0.0855 | XY | 0.0853-0.0857 |
| 011291 | 1163521 | 27946 | 1936 | 0.0693 | Not Assigned | 0.0663-0.0723 |
| 011292 | 2417115 | 55036 | 3992 | 0.0725 | Not Assigned | 0.0704-0.0747 |
| 011293 | 19875374 | 461591 | 34921 | 0.0757 | consistent with XY but not XX | 0.0749-0.0764 |
| 011294 | 39018131 | 962650 | 70894 | 0.0736 | Not Assigned | 0.0731-0.0742 |
| 011295 | | | | | | |
| 011299 | 1813341190 | 93181974 | 138728 | 0.0015 | XX | 0.0015-0.0015 |
| 011311 | 7531 | 360 | 25 | 0.0694 | consistent with XY but not XX | 0.0432-0.0957 |
| 011312 | 5823 | 224 | 8 | 0.0357 | consistent with XY but not XX | 0.0114-0.06 |
| 011313 | 537033 | 13974 | 1047 | 0.0749 | consistent with XY but not XX | 0.0706-0.0793 |
| 011316 | 104503015 | 5162941 | 9901 | 0.0019 | XX | 0.0019-0.002 |
| 011317 | | | | | | |
| 011319 | 17400312 | 444296 | 30621 | 0.0689 | Not Assigned | 0.0682-0.0697 |
| 011320 | 87767428 | 4143400 | 11014 | 0.0027 | XX | 0.0026-0.0027 |
| 011321 | | | | | | |
| 011322 | 116751301 | 5934855 | 10602 | 0.0018 | XX | 0.0018-0.0018 |
| 011323 | | | | | | |
| 011324 | 21213544 | 530777 | 35548 | 0.067 | Not Assigned | 0.0663-0.0676 |
| 011325 | 51418 | 1317 | 93 | 0.0706 | consistent with XY but not XX | 0.0568-0.0845 |
| 011327 | | | | | | |
| 011332 | 25629318 | 646781 | 44893 | 0.0694 | Not Assigned | 0.0688-0.07 |
| 011335 | | | | | | |
| 011336 | 4324076 | 210666 | 533 | 0.0025 | XX | 0.0023-0.0027 |
| 011337 | 28041526 | 704864 | 48588 | 0.0689 | Not Assigned | 0.0683-0.0695 |
| 011339 | 173681026 | 9072835 | 19024 | 0.0021 | XX | 0.0021-0.0021 |
| 011341 | 12060765 | 315466 | 22450 | 0.0712 | Not Assigned | 0.0703-0.0721 |
| 011342 | | | | | | |
| 011343 | 118928028 | 3233077 | 231460 | 0.0716 | Not Assigned | 0.0713-0.0719 |
| 011414 | 139149179 | 4008495 | 324937 | 0.0811 | XY | 0.0808-0.0813 |
| 011415 | | | | | | |
| 011423 | 979354194 | 28562849 | 2322082 | 0.0813 | XY | 0.0812-0.0814 |
| 011433 | | | | | | |
| 011434 | 1202998840 | 37560064 | 3303419 | 0.088 | XY | 0.0879-0.088 |
| 011435 | 1228341318 | 36178543 | 2974174 | 0.0822 | XY | 0.0821-0.0823 |
| 011436 | 1421251597 | 72492565 | 102806 | 0.0014 | XX | 0.0014-0.0014 |

| ID | Nseqs | NchrY+NchrX | NchrY | R_y | Genetic Sex Assignment | 95% CI |
|---|---|---|---|---|---|---|
| 011442 | 2393262 | 69033 | 5564 | 0.0806 | XY | 0.0786-0.0826 |
| 011444 | 1640823870 | 82335341 | 100434 | 0.0012 | XX | 0.0012-0.0012 |
| 011445 | 90569925 | 4709105 | 8666 | 0.0018 | XX | 0.0018-0.0019 |
| 011446 | | | | | | |
| 011447 | 89783167 | 2545673 | 203416 | 0.0799 | XY | 0.0796-0.0802 |
| 011448 | 1680676607 | 48544687 | 4147277 | 0.0854 | XY | 0.0854-0.0855 |
| 011874 | 1177155631 | 30989425 | 2382487 | 0.0769 | XY | 0.0768-0.077 |
| 011875 | 1668345 | 48940 | 715 | 0.0146 | XX | 0.0135-0.0157 |
| 011876 | 1101731273 | 32763066 | 2547458 | 0.0778 | XY | 0.0777-0.0778 |
| 011877 | 21648521 | 605843 | 47344 | 0.0781 | XY | 0.0775-0.0788 |
| 011878 | 338878 | 9594 | 674 | 0.0703 | consistent with XY but not XX | 0.0651-0.0754 |
| 011879 | 1128473 | 32532 | 2341 | 0.072 | Not Assigned | 0.0692-0.0748 |
| 011880 | 69282607 | 1973052 | 153329 | 0.0777 | XY | 0.0773-0.0781 |
| 022559 | 263451496 | 7506554 | 648633 | 0.0864 | XY | 0.0862-0.0866 |
| 022560 | 200108363 | 4993826 | 405393 | 0.0812 | XY | 0.0809-0.0814 |
| 022561 | 1588492061 | 81027721 | 126044 | 0.0016 | XX | 0.0015-0.0016 |
| 022903 | 151371894 | 4284548 | 350788 | 0.0819 | XY | 0.0816-0.0821 |
| 022904 | 196636 | 9986 | 28 | 0.0028 | XX | 0.0018-0.0038 |
| 022910 | 30889 | 1113 | 72 | 0.0647 | consistent with XY but not XX | 0.0502-0.0791 |
| 022911 | 66282 | 2585 | 92 | 0.0356 | Not Assigned | 0.0284-0.0427 |
| 022912 | 73502 | 2590 | 107 | 0.0413 | Not Assigned | 0.0336-0.049 |
| 022913 | 106290 | 4538 | 190 | 0.0419 | Not Assigned | 0.036-0.0477 |
| 022914 | 78516 | 2777 | 133 | 0.0479 | Not Assigned | 0.04-0.0558 |
| 022915 | 976212 | 47440 | 292 | 0.0062 | XX | 0.0055-0.0069 |
| 022916 | 4098325 | 198737 | 2501 | 0.0126 | XX | 0.0121-0.0131 |
| 022917 | 2464709 | 75834 | 4890 | 0.0645 | Not Assigned | 0.0627-0.0662 |
| 022918 | 141420 | 6525 | 25 | 0.0038 | XX | 0.0023-0.0053 |
| 022919 | 44214 | 1666 | 48 | 0.0288 | Not Assigned | 0.0208-0.0368 |
| 022920 | 194261 | 7609 | 232 | 0.0305 | Not Assigned | 0.0266-0.0344 |
| 022921 | 55932 | 2424 | 79 | 0.0326 | Not Assigned | 0.0255-0.0397 |
| 022922 | 123896 | 5168 | 162 | 0.0313 | Not Assigned | 0.0266-0.0361 |
| 023501 | 22352 | 927 | 10 | 0.0108 | consistent with XY but not XX | 0.0041-0.0174 |
| 023502 | 65415255 | 2872979 | 8040 | 0.0028 | XX | 0.0027-0.0029 |
| 023504 | 125472005 | 5785784 | 16374 | 0.0028 | XX | 0.0028-0.0029 |
| 024054 | 500647620 | 14375241 | 1185780 | 0.0825 | XY | 0.0823-0.0826 |
| 024055 | 1089950 | 32292 | 2309 | 0.0715 | Not Assigned | 0.0687-0.0743 |
| 024056 | 56634172 | 2810878 | 4134 | 0.0015 | XX | 0.0014-0.0015 |
| 024057 | 359952815 | 18508482 | 27667 | 0.0015 | XX | 0.0015-0.0015 |
| 024058 | 1718993401 | 88553052 | 120929 | 0.0014 | XX | 0.0014-0.0014 |
| 024059 | 781231 | 22135 | 1787 | 0.0807 | XY | 0.0771-0.0843 |
| 024060 | 257388093 | 13123550 | 19523 | 0.0015 | XX | 0.0015-0.0015 |
| 024061 | 104497434 | 5243096 | 8221 | 0.0016 | XX | 0.0015-0.0016 |
| 024062 | 159986437 | 4626904 | 380603 | 0.0823 | XY | 0.082-0.0825 |
| 024063 | 163233693 | 8417125 | 12342 | 0.0015 | XX | 0.0014-0.0015 |
| 024064 | 100356286 | 4988069 | 8268 | 0.0017 | XX | 0.0016-0.0017 |
| 024065 | 228916954 | 11682918 | 17112 | 0.0015 | XX | 0.0014-0.0015 |
| 024066 | 346747591 | 10109381 | 832648 | 0.0824 | XY | 0.0822-0.0825 |
| 024067 | 112830217 | 3228026 | 269081 | 0.0834 | XY | 0.0831-0.0837 |
| 024068 | 469978520 | 15205048 | 21594 | 0.0014 | XX | 0.0014-0.0014 |
| 024069 | 1964015581 | 100966352 | 121962 | 0.0012 | XX | 0.0012-0.0012 |
| 024070 | 596305342 | 30845280 | 43460 | 0.0014 | XX | 0.0014-0.0014 |
| 024071 | 927657247 | 47208908 | 69575 | 0.0015 | XX | 0.0015-0.0015 |

| ID | Nseqs | NchrY+NchrX | NchrY | R_y | Genetic Sex Assignment | 95% CI |
|---|---|---|---|---|---|---|
| 024072 | 19822262 | 816776 | 21645 | 0.0265 | Not Assigned | 0.0262-0.0268 |
| 024073 | 528090183 | 27253454 | 38904 | 0.0014 | XX | 0.0014-0.0014 |
| 024074 | 148637311 | 7609828 | 10871 | 0.0014 | XX | 0.0014-0.0015 |
| 024075 | 426245599 | 21839032 | 33559 | 0.0015 | XX | 0.0015-0.0016 |
| 024076 | 267594405 | 13563990 | 18804 | 0.0014 | XX | 0.0014-0.0014 |
| 024077 | 629958922 | 32206710 | 45112 | 0.0014 | XX | 0.0014-0.0014 |
| 024078 | 114970976 | 5878404 | 7566 | 0.0013 | XX | 0.0013-0.0013 |
| 024079 | 281611373 | 14315588 | 21855 | 0.0015 | XX | 0.0015-0.0015 |
| 024080 | 1668410605 | 85823711 | 113086 | 0.0013 | XX | 0.0013-0.0013 |
| 024081 | 984466029 | 29576755 | 2491394 | 0.0842 | XY | 0.0841-0.0843 |
| 024082 | 178888310 | 5173380 | 420704 | 0.0813 | XY | 0.0811-0.0816 |
| 024083 | 297659036 | 15164221 | 22896 | 0.0015 | XX | 0.0015-0.0015 |
| 024084 | 1321747865 | 67640823 | 97846 | 0.0014 | XX | 0.0014-0.0015 |
| 024085 | 1636746638 | 48800838 | 4090954 | 0.0838 | XY | 0.0838-0.0839 |
| 024086 | 478558666 | 14087113 | 1185229 | 0.0841 | XY | 0.084-0.0843 |
| 024088 | 298356122 | 15136417 | 22943 | 0.0015 | XX | 0.0015-0.0015 |
| 024089 | 458790770 | 23204526 | 34116 | 0.0015 | XX | 0.0015-0.0015 |
| 024090 | 700649137 | 36442571 | 50533 | 0.0014 | XX | 0.0014-0.0014 |
| 024091 | 228749636 | 11687818 | 20096 | 0.0017 | XX | 0.0017-0.0017 |
| 024092 | 479905622 | 24549961 | 39644 | 0.0016 | XX | 0.0016-0.0016 |
| 024093 | 786917175 | 23045932 | 1937278 | 0.0841 | XY | 0.0839-0.0842 |
| 024094 | 2456891369 | 74613032 | 6160997 | 0.0826 | XY | 0.0825-0.0826 |
| 024095 | 1318812004 | 67782145 | 102389 | 0.0015 | XX | 0.0015-0.0015 |
| 024098 | 1450689600 | 74794514 | 115216 | 0.0015 | XX | 0.0015-0.0015 |
| 024099 | 1182550667 | 59836457 | 79726 | 0.0013 | XX | 0.0013-0.0013 |
| 024100 | 510779098 | 26792326 | 39317 | 0.0015 | XX | 0.0015-0.0015 |
| 024101 | 26920467 | 767889 | 62440 | 0.0813 | XY | 0.0807-0.0819 |
| 024102 | 163596505 | 4738766 | 385995 | 0.0815 | XY | 0.0812-0.0817 |
| 024103 | 1843601750 | 56891765 | 4752180 | 0.0835 | XY | 0.0835-0.0836 |
| 024112 | 116393 | 5093 | 58 | 0.0114 | XX | 0.0085-0.0143 |
| 024113 | 68784 | 3235 | 79 | 0.0244 | Not Assigned | 0.0191-0.0297 |
| 024114 | 13899 | 379 | 29 | 0.0765 | consistent with XY but not XX | 0.0498-0.1033 |
| 024115 | 697303260 | 35844330 | 47051 | 0.0013 | XX | 0.0013-0.0013 |
| 024116 | 333506095 | 16951188 | 26313 | 0.0016 | XX | 0.0015-0.0016 |
| 024117 | 3680571 | 107368 | 8520 | 0.0794 | XY | 0.0777-0.081 |
| 024118 | 5887309 | 169066 | 13933 | 0.0824 | XY | 0.0811-0.0837 |
| 024119 | 35522 | 1576 | 38 | 0.0241 | Not Assigned | 0.0165-0.0317 |
| 024120 | 3448580 | 99924 | 7996 | 0.08 | XY | 0.0783-0.0817 |
| 024121 | 516101 | 23882 | 143 | 0.006 | XX | 0.005-0.007 |
| 024122 | 1231797426 | 62036707 | 101540 | 0.0016 | XX | 0.0016-0.0016 |
| 024123 | 91843632 | 4374226 | 14091 | 0.0032 | XX | 0.0032-0.0033 |
| 024124 | 20703 | 863 | 19 | 0.022 | consistent with XY but not XX | 0.0122-0.0318 |
| 024125 | 48255 | 1363 | 63 | 0.0462 | Not Assigned | 0.0351-0.0574 |
| 024126 | 126193598 | 3612996 | 301537 | 0.0835 | XY | 0.0832-0.0837 |
| 024247 | 172282714 | 8664732 | 13325 | 0.0015 | XX | 0.0015-0.0016 |
| 024248 | 201600 | 5812 | 426 | 0.0733 | consistent with XY but not XX | 0.0666-0.08 |
| 024249 | 89933 | 4478 | 14 | 0.0031 | XX | 0.0015-0.0048 |
| 024250 | 29201359 | 1440132 | 2548 | 0.0018 | XX | 0.0017-0.0018 |
| 024251 | 6455125 | 183922 | 14689 | 0.0799 | XY | 0.0786-0.0811 |
| 024252 | 155882 | 7888 | 38 | 0.0048 | XX | 0.0033-0.0063 |
| 024253 | 334718770 | 16999053 | 25543 | 0.0015 | XX | 0.0015-0.0015 |
| 024254 | 357699130 | 10427844 | 846429 | 0.0812 | XY | 0.081-0.0813 |

| ID | Nseqs | NchrY+NchrX | NchrY | R_y | Genetic Sex Assignment | 95% CI |
|---|---|---|---|---|---|---|
| 024255 | 2369834 | 118835 | 246 | 0.0021 | XX | 0.0018-0.0023 |
| 024256 | 2378295807 | 68319925 | 5639671 | 0.0825 | XY | 0.0825-0.0826 |
| 024257 | 2188086 | 59604 | 5336 | 0.0895 | XY | 0.0872-0.0918 |
| 024258 | 83886 | 3518 | 13 | 0.0037 | XX | 0.0017-0.0057 |
| 024259 | 3502346 | 96914 | 7583 | 0.0782 | XY | 0.0766-0.0799 |
| 024260 | 8478670 | 237768 | 18393 | 0.0774 | XY | 0.0763-0.0784 |
| 024261 | 1122578985 | 31814986 | 2692243 | 0.0846 | XY | 0.0845-0.0847 |
| 024262 | 82081233 | 4084944 | 6567 | 0.0016 | XX | 0.0016-0.0016 |
| 024263 | 206415 | 5405 | 66 | 0.0122 | XX | 0.0093-0.0151 |
| 024264 | | | | | | |
| 024265 | 820605893 | 23599646 | 1983973 | 0.0841 | XY | 0.084-0.0842 |
| 024266 | 695519685 | 35272142 | 48799 | 0.0014 | XX | 0.0014-0.0014 |
| 024267 | 795501701 | 40421402 | 55700 | 0.0014 | XX | 0.0014-0.0014 |
| 024268 | 78365971 | 4004626 | 7053 | 0.0018 | XX | 0.0017-0.0018 |
| 024724 | 1420630906 | 67199053 | 113974 | 0.0017 | XX | 0.0017-0.0017 |
| 024725 | 52807 | 1318 | 105 | 0.0797 | consistent with XY but not XX | 0.065-0.0943 |
| 024726 | 305504873 | 10830832 | 297713 | 0.0275 | Not Assigned | 0.0274-0.0276 |
| 024727 | 440532511 | 21286251 | 38496 | 0.0018 | XX | 0.0018-0.0018 |
| 024728 | 2014936 | 51431 | 3783 | 0.0736 | consistent with XY but not XX | 0.0713-0.0758 |
| 024729 | 969029 | 25205 | 1737 | 0.0689 | Not Assigned | 0.0658-0.072 |
| 024730 | 79891892 | 2011203 | 139259 | 0.0692 | Not Assigned | 0.0689-0.0696 |
| 024731 | 383253975 | 10118042 | 791297 | 0.0782 | XY | 0.078-0.0784 |
| 024732 | 577569 | 14686 | 1009 | 0.0687 | Not Assigned | 0.0646-0.0728 |
| 024733 | 6340863 | 167900 | 11908 | 0.0709 | Not Assigned | 0.0697-0.0722 |
| 024734 | 6987500 | 325974 | 665 | 0.002 | XX | 0.0019-0.0022 |
| 024735 | 5593878 | 145052 | 11200 | 0.0772 | XY | 0.0758-0.0786 |
| 024736 | 309936652 | 14727644 | 30959 | 0.0021 | XX | 0.0021-0.0021 |
| 024737 | 91945091 | 2427380 | 176624 | 0.0728 | Not Assigned | 0.0724-0.0731 |
| 024738 | 12239 | 501 | 12 | 0.024 | consistent with XY but not XX | 0.0106-0.0373 |
| 024739 | 34665251 | 1646094 | 3836 | 0.0023 | XX | 0.0023-0.0024 |
| 024740 | | | | | | |
| 024741 | 7923 | 209 | 19 | 0.0909 | consistent with XY but not XX | 0.0519-0.1299 |
| 024742 | 2688968 | 71951 | 4943 | 0.0687 | Not Assigned | 0.0669-0.0705 |
| 024743 | 51946522 | 1321917 | 93198 | 0.0705 | Not Assigned | 0.0701-0.0709 |
| 024744 | 16870734 | 599625 | 19722 | 0.0329 | Not Assigned | 0.0324-0.0333 |
| 024745 | 11330167 | 541409 | 797 | 0.0015 | XX | 0.0014-0.0016 |
| 024746 | 14851974 | 730230 | 1473 | 0.002 | XX | 0.0019-0.0021 |
| 024747 | 155377462 | 7476390 | 15960 | 0.0021 | XX | 0.0021-0.0022 |
| 024748 | 67282007 | 1695580 | 142027 | 0.0838 | XY | 0.0833-0.0842 |
| 024749 | 35296602 | 1640773 | 3814 | 0.0023 | XX | 0.0023-0.0024 |
| 024750 | 398756491 | 18727701 | 37826 | 0.002 | XX | 0.002-0.002 |
| 024751 | 2697 | 94 | 4 | 0.0426 | Not Assigned | 0.0017-0.0834 |
| 024752 | 63299623 | 2997823 | 8921 | 0.003 | XX | 0.0029-0.003 |
| 024753 | | | | | | |
| 024754 | | | | | | |
| 024755 | 65567590 | 1643831 | 123626 | 0.0752 | consistent with XY but not XX | 0.0748-0.0756 |
| 024756 | 40411089 | 1042587 | 81301 | 0.078 | XY | 0.0775-0.0785 |
| 024757 | 97788834 | 2387036 | 206226 | 0.0864 | XY | 0.086-0.0868 |
| 024758 | 89109850 | 2218981 | 164771 | 0.0743 | Not Assigned | 0.0739-0.0746 |
| 024759 | 27435738 | 684906 | 59211 | 0.0865 | XY | 0.0858-0.0871 |
| 024760 | 96937195 | 2432302 | 191298 | 0.0786 | XY | 0.0783-0.079 |
| 024761 | 77251427 | 1898882 | 140775 | 0.0741 | Not Assigned | 0.0738-0.0745 |

| ID | Nseqs | NchrY+NchrX | NchrY | R_y | Genetic Sex Assignment | 95% CI |
|---|---|---|---|---|---|---|
| 024762 | 140747326 | 3625630 | 273991 | 0.0756 | XY | 0.0753-0.0758 |
| 024763 | 40179246 | 934973 | 63678 | 0.0681 | Not Assigned | 0.0676-0.0686 |
| 024764 | 151036632 | 6898576 | 13051 | 0.0019 | XX | 0.0019-0.0019 |
| 024766 | 1692008967 | 80896608 | 146805 | 0.0018 | XX | 0.0018-0.0018 |
| 024767 | 53727644 | 1359411 | 94637 | 0.0696 | Not Assigned | 0.0692-0.07 |
| 024768 | 547685620 | 14739011 | 1169107 | 0.0793 | XY | 0.0792-0.0795 |
| 024769 | 60986595 | 1597824 | 119694 | 0.0749 | consistent with XY but not XX | 0.0745-0.0753 |
| 024770 | 117319970 | 2962214 | 239189 | 0.0807 | XY | 0.0804-0.0811 |
| 024771 | 62727516 | 3100865 | 5564 | 0.0018 | XX | 0.0017-0.0018 |
| 024772 | 147647499 | 3805197 | 313329 | 0.0823 | XY | 0.0821-0.0826 |
| 024773 | 123060916 | 3064819 | 243466 | 0.0794 | XY | 0.0791-0.0797 |
| 024774 | 179227461 | 8416050 | 15522 | 0.0018 | XX | 0.0018-0.0019 |
| 024775 | 118230804 | 5262225 | 11859 | 0.0023 | XX | 0.0022-0.0023 |
| 024776 | 112797846 | 3060948 | 251040 | 0.082 | XY | 0.0817-0.0823 |
| 024777 | 5431616 | 160640 | 13217 | 0.0823 | XY | 0.0809-0.0836 |
| 024778 | 444365 | 11415 | 887 | 0.0777 | consistent with XY but not XX | 0.0728-0.0826 |
| 024779 | 270076079 | 13304716 | 25900 | 0.0019 | XX | 0.0019-0.002 |
| 024780 | 107117385 | 2602781 | 225719 | 0.0867 | XY | 0.0864-0.0871 |
| 024781 | 8930055 | 432806 | 2713 | 0.0063 | XX | 0.006-0.0065 |
| 024782 | 25229705 | 609163 | 43029 | 0.0706 | Not Assigned | 0.07-0.0713 |
| 024783 | 137225663 | 6167315 | 13117 | 0.0021 | XX | 0.0021-0.0022 |
| 024784 | 33717320 | 1477308 | 3263 | 0.0022 | XX | 0.0021-0.0023 |

**Table S8 – Mitochondrial and Y-Haplogroup Assignments**

List of all tissues included in this project (n=217), sorted by the ID assigned at the CGG and accompanied by the depth of coverage on the mitochondrial genome, the mitochondrial haplogroup assignment, the mitochondrial haplogroup assignment probability, the depth of coverage of the Y chromosome, the Y-chromosome ISOGG haplogroup assignment, the major Y-haplogroup assignment, and the minor Y-haplogroup assignment.

| ID | MT Depth | MT Haplogroup | MT Probability | Y Depth | ISOGG Haplo | Haplo Major | Haplo Minor |
|---|---|---|---|---|---|---|---|
| 011229 | 1690.4 | D2a1 | 0.92 | 15.14 | Q1a1b | Q1a | Q1a-B143 |
| 011230 | 1682.5 | D2a1 | 0.97 | 1.02 | Q1a1b | Q1a | Q1a-B143 |
| 011231 | 841.7 | D2a1 | 0.90 | | | | |
| 011234 | 296.3 | A2a | 0.98 | 2.90 | Q1a1b | Q1a | Q1a-B143 |
| 011236 | 467.0 | A2a | 1.00 | 3.12 | Q1a1b | Q1a | Q1a-B143 |
| 011291 | 193.1 | A2a3 | 0.96 | | | | |
| 011292 | 429.2 | A2a3 | 0.96 | | | | |
| 011293 | 1098.9 | A2a3 | 0.98 | 0.10 | Q1a1b | Q1a | Q1a-B143 |
| 011294 | 1174.0 | A2a | 1.00 | | | | |
| 011295 | | | | | | | |
| 011299 | 1316.3 | A2b1 | 0.94 | | | | |
| 011311 | 0.0 | B4b1 | 0.51 | | | | |
| 011312 | 0.0 | H2a2a1 | 0.50 | | | | |
| 011313 | 61.6 | A2a | 0.91 | | | | |
| 011316 | 272.2 | A2a | 0.97 | | | | |
| 011317 | | | | | | | |
| 011319 | 185.3 | A2a | 0.93 | 0.09 | N1a1a1a1a | N1a | N1a-L392 |
| 011320 | 264.0 | A2a | 0.95 | | | | |
| 011321 | | | | | | | |
| 011322 | 510.5 | A2a | 0.95 | | | | |
| 011323 | | | | | | | |
| 011324 | 188.6 | A2a | 0.93 | 0.10 | Q1a1b | Q1a | Q1a-B143 |
| 011325 | 18.5 | A2a | 0.81 | | | | |
| 011327 | | | | | | | |
| 011332 | 164.5 | A2a | 0.95 | 0.12 | Q1b1a1a2 | Q1b | Q1b-M3 (Y4303) |
| 011335 | | | | | | | |
| 011336 | 150.2 | A2a | 0.93 | | | | |
| 011337 | 230.3 | A2b1 | 0.98 | 0.13 | Q1a1b | Q1a | Q1a-B143 |
| 011339 | 221.9 | A2a | 0.97 | | | | |
| 011341 | 155.2 | D1 | 0.94 | | | | |
| 011342 | | | | | | | |
| 011343 | 221.6 | A2b1 | 0.96 | 0.60 | Q1b1a1a2a~ | Q1b | Q1b-M3 (Y4303) |
| 011414 | 351.9 | D2a | 0.92 | 1.43 | Q1a1b | Q1a | Q1a-B143 |
| 011415 | | | | | | | |
| 011423 | 1374.6 | D2a1 | 0.92 | 9.86 | Q1a1b | Q1a | Q1a-B143 |
| 011433 | | | | | | | |
| 011434 | 505.3 | D2a | 0.94 | 19.62 | Q1a1b | Q1a | Q1a-B143 |
| 011435 | 1461.5 | D2a1 | 0.97 | 14.53 | Q1a1b | Q1a | Q1a-B143 |
| 011436 | 3426.1 | D2a | 0.91 | | | | |
| 011442 | 3.2 | H2a2a | 0.55 | 0.02 | Q1a1b | Q1a | Q1a-B143 |
| 011444 | 1210.2 | A2i | 0.92 | | | | |
| 011445 | 840.6 | D2a1 | 0.92 | | | | |

| ID | MT Depth | MT Haplogroup | MT Probability | Y Depth | ISOGG Haplo | Haplo Major | Haplo Minor |
|---|---|---|---|---|---|---|---|
| 011446 | | | | | | | |
| 011447 | 6418.0 | D2a1 | 0.93 | 0.84 | Q1a1b | Q1a | Q1a-B143 |
| 011448 | 3919.4 | D2a1 | 0.97 | 23.27 | Q1a1b | Q1a | Q1a-B143 |
| 011874 | 1474.0 | D4h3 | 0.90 | 6.86 | N2a~ | N2a | N2a-Y6514 |
| 011875 | 5.5 | H2a2a | 0.57 | | | | |
| 011876 | 2007.6 | G1b | 0.91 | 9.87 | Q1b1b~ | Q1b | Q1b-YP4010 |
| 011877 | 80.3 | G1b | 0.93 | 0.17 | Q1b1b~ | Q1b | Q1b-YP4010 |
| 011878 | 0.4 | H2a2a | 0.55 | | | | |
| 011879 | 2.3 | G1 | 0.56 | | | | |
| 011880 | 226.3 | D4h3 | 0.89 | 0.57 | Q1b1b~ | Q1b | Q1b-YP4010 |
| 022559 | 1820.4 | X2a1c | 0.95 | 3.70 | Q1b1a1a1 | Q1b | Q1b-M3 (M848) |
| 022560 | 921.2 | C1c | 0.91 | 1.58 | Q1b1a1a2b2a~ | Q1b | Q1b-M3 (Y4303) |
| 022561 | 1470.4 | A2i | 0.92 | | | | |
| 022903 | 481.0 | H2a2a1 | 0.50 | 1.56 | R1a1a1b1a1a1c1g1~ | R1a | R1a-L1029 |
| 022904 | 0.5 | H2a2a1 | 0.50 | | | | |
| 022910 | 0.1 | H2a2a1 | 0.50 | | | | |
| 022911 | 0.1 | H2a2a | 0.74 | | | | |
| 022912 | 0.3 | M30d | 0.52 | | | | |
| 022913 | 0.2 | H2a2a1 | 0.50 | | | | |
| 022914 | 0.3 | D4e1'3 | 0.55 | | | | |
| 022915 | 25.5 | D4b1a2a1 | 0.92 | | | | |
| 022916 | 28.5 | D4e4a | 0.82 | | | | |
| 022917 | 41.8 | M | 0.87 | | | | |
| 022918 | 0.4 | Z1 | 0.69 | | | | |
| 022919 | 0.1 | T2f1a | 0.54 | | | | |
| 022920 | 0.3 | J1c2a3 | 0.53 | | | | |
| 022921 | 0.1 | H2a2a1 | 0.50 | | | | |
| 022922 | 1.0 | H2a2a | 0.54 | | | | |
| 023501 | 1.1 | A2a | 0.71 | | | | |
| 023502 | 450.0 | A2a | 0.93 | | | | |
| 023504 | 309.6 | A2a | 0.96 | | | | |
| 024054 | 423.8 | A2a | 1.00 | 5.11 | Q1b1a1a2a~ | Q1b | Q1b-M3 (Y4303) |
| 024055 | 25.9 | A2a | 0.78 | | | | |
| 024056 | 436.6 | A2a1 | 0.96 | | | | |
| 024057 | 447.2 | A2b1 | 0.98 | | | | |
| 024058 | 794.7 | A2a1 | 0.93 | | | | |
| 024059 | 33.9 | A2b1 | 0.77 | | | | |
| 024060 | 318.6 | A2b1 | 0.96 | | | | |
| 024061 | 449.5 | A2b1 | 0.96 | | | | |
| 024062 | 171.0 | A2a | 0.96 | 1.79 | Q1b1a1a2a~ | Q1b | Q1b-M3 (Y4303) |
| 024063 | 435.5 | A2a3 | 0.91 | | | | |
| 024064 | 453.9 | A2b1 | 0.96 | | | | |
| 024065 | 208.6 | A2a3 | 0.93 | | | | |
| 024066 | 332.4 | A2b1 | 1.00 | 3.82 | Q1b1a1a2a~ | Q1b | Q1b-M3 (Y4303) |
| 024067 | 675.0 | A2b1 | 0.92 | 1.24 | Q1b1a1a2a~ | Q1b | Q1b-M3 (Y4303) |
| 024068 | 275.3 | A2b1 | 0.93 | | | | |
| 024069 | 1029.1 | A2b1 | 0.96 | | | | |
| 024070 | 520.5 | A2a3 | 0.99 | | | | |
| 024071 | 353.6 | A2a | 1.00 | | | | |
| 024072 | 520.5 | A2a1 | 0.93 | 0.09 | Q1b1 | Q1b | Q1b-L53 |
| 024073 | 846.5 | A2a | 0.96 | | | | |
| 024074 | 121.1 | A2a | 0.96 | | | | |

| ID | MT Depth | MT Haplogroup | MT Probability | Y Depth | ISOGG Haplo | Haplo Major | Haplo Minor |
|---|---|---|---|---|---|---|---|
| 024075 | 345.4 | A2a | 0.98 | | | | |
| 024076 | 413.5 | A2b1 | 1.00 | | | | |
| 024077 | 418.7 | A2a | 0.96 | | | | |
| 024078 | 171.1 | A2a | 0.96 | | | | |
| 024079 | 544.3 | A2b1 | 0.94 | | | | |
| 024080 | 663.4 | A2b1 | 0.98 | | | | |
| 024081 | 594.5 | A2a1 | 1.00 | 13.35 | Q1a1b | Q1a | Q1a-B143 |
| 024082 | 481.6 | A2a | 0.96 | 1.76 | Q1a1b | Q1a | Q1a-B143 |
| 024083 | 236.1 | A2a | 0.98 | | | | |
| 024084 | 1255.2 | A2b1 | 0.96 | | | | |
| 024085 | 917.4 | A2b1 | 0.96 | 20.61 | Q1b1a1a2a~ | Q1b | Q1b-M3 (Y4303) |
| 024086 | 431.5 | A2b1 | 1.00 | 5.91 | Q1a1b | Q1a | Q1a-B143 |
| 024088 | 610.3 | A2b1 | 0.96 | | | | |
| 024089 | 655.9 | A2b1 | 0.96 | | | | |
| 024090 | 268.8 | A2b1 | 0.99 | | | | |
| 024091 | 333.5 | A2a3 | 0.98 | | | | |
| 024092 | 576.3 | A2b1 | 0.96 | | | | |
| 024093 | 1254.9 | A2b1 | 0.96 | 10.18 | Q1b1a1a2a~ | Q1b | Q1b-M3 (Y4303) |
| 024094 | 760.8 | A2a | 1.00 | 30.86 | Q1a1b | Q1a | Q1a-B143 |
| 024095 | 694.4 | A2a3 | 0.96 | | | | |
| 024098 | 1376.1 | A2a | 0.96 | | | | |
| 024099 | 400.7 | A2a | 0.98 | | | | |
| 024100 | 432.7 | A2a | 0.98 | | | | |
| 024101 | 727.4 | A2b1 | 0.96 | 0.24 | Q1a1b | Q1a | Q1a-B143 |
| 024102 | 796.7 | A2a | 0.96 | 1.72 | Q1a1b | Q1a | Q1a-B143 |
| 024103 | 480.8 | A2b1 | 0.98 | 24.84 | Q1a1b | Q1a | Q1a-B143 |
| 024112 | 0.1 | H2a2a | 0.69 | | | | |
| 024113 | 0.2 | H2a2a1 | 0.50 | | | | |
| 024114 | 2.5 | G1b | 0.77 | | | | |
| 024115 | 590.4 | C4b | 0.93 | | | | |
| 024116 | 1157.7 | D4b1a2a1 | 0.96 | | | | |
| 024117 | 9.4 | D4b1a2a1 | 0.85 | 0.03 | N1a1 | N1a | N1a-M46 |
| 024118 | 57.9 | C4b1 | 0.92 | 0.06 | C2a1a2b1b | C2a | C2a-M48 |
| 024119 | 0.0 | H2a2a1 | 0.50 | | | | |
| 024120 | 9.5 | D4b1a2a1 | 0.78 | 0.03 | N1a1a1a1a | N1a | N1a-L392 |
| 024121 | 152.7 | G1b+16129 | 0.93 | | | | |
| 024122 | 729.7 | Z1a2a | 0.94 | | | | |
| 024123 | 611.1 | C4b2a | 0.95 | | | | |
| 024124 | 0.0 | JT | 0.54 | | | | |
| 024125 | 4.8 | G1b+16129 | 0.86 | | | | |
| 024126 | 451.7 | A+152+16362+16189 | 0.94 | 1.40 | Q1b1b~ | Q1b | Q1b-YP4010 |
| 024247 | 328.9 | D2a1 | 0.96 | | | | |
| 024248 | 0.3 | H2a2a | 0.54 | | | | |
| 024249 | 0.3 | H2a2a | 0.61 | | | | |
| 024250 | 156.5 | G1b+16129 | 0.98 | | | | |
| 024251 | 27.8 | C4b | 0.89 | 0.05 | N1 | N1 | N1 |
| 024252 | 0.7 | G1b+16129 | 0.67 | | | | |
| 024253 | 1076.7 | A2a1 | 0.94 | | | | |
| 024254 | 565.4 | A2b1 | 0.94 | 3.49 | Q1a1b | Q1a | Q1a-B143 |
| 024255 | 8.3 | A2a1 | 0.88 | | | | |
| 024256 | 1417.7 | A2a | 0.96 | 25.58 | Q1b1a1a2a~ | Q1b | Q1b-M3 (Y4303) |

| ID | MT Depth | MT Haplogroup | MT Probability | Y Depth | ISOGG Haplo | Haplo Major | Haplo Minor |
|---|---|---|---|---|---|---|---|
| 024257 | 431.1 | D4b1a2a1 | 0.95 | 0.02 | N1 | N1 | N1 |
| 024258 | 45.7 | G1b+16129 | 0.91 | | | | |
| 024259 | 8.1 | G1b | 0.92 | 0.03 | N1a1a1a1 | N1a | N1a-P298 |
| 024260 | 20.9 | D2a'b | 0.86 | 0.06 | C2a1a1b2~ | C2a | C2a-B473 |
| 024261 | 1644.7 | A2a | 0.94 | 13.49 | Q1b1a1a2a~ | Q1b | Q1b-M3 (Y4303) |
| 024262 | 623.2 | A2a1 | 0.96 | | | | |
| 024263 | 0.5 | U5a1d | 0.54 | | | | |
| 024264 | | | | | | | |
| 024265 | 885.3 | G1b1 | 0.95 | 10.72 | C2a1a1b1a2b~ | C2a | C2a-B473 |
| 024266 | 1739.3 | G1b1 | 0.95 | | | | |
| 024267 | 1449.1 | G1b+16129 | 0.95 | | | | |
| 024268 | 513.9 | G1b+16129 | 0.94 | | | | |
| 024724 | 2415.4 | A2a | 0.99 | | | | |
| 024725 | 8.3 | D4b1a2a1 | 0.94 | | | | |
| 024726 | 4950.6 | A2a | 0.99 | 0.79 | Q1b1a1a2a~ | Q1b | Q1b-M3 (Y4303) |
| 024727 | 333.9 | A2a | 0.93 | | | | |
| 024728 | 99.1 | A2a | 0.96 | | | | |
| 024729 | 89.8 | A2a | 0.96 | | | | |
| 024730 | 310.7 | A2a | 0.93 | 0.38 | Q1a1b | Q1a | Q1a-B143 |
| 024731 | 835.5 | A2a | 0.93 | 2.21 | Q1a1b | Q1a | Q1a-B143 |
| 024732 | 43.9 | A2a | 0.93 | | | | |
| 024733 | 166.7 | A2a | 0.96 | 0.03 | Q1b | Q1b | Q1b |
| 024734 | 90.8 | A2a | 0.95 | | | | |
| 024735 | 97.5 | A2a | 0.93 | 0.03 | Q1b1 | Q1b | Q1b-L53 |
| 024736 | 428.6 | A2a | 0.97 | | | | |
| 024737 | 436.5 | A2a | 0.95 | 0.46 | Q1a1b | Q1a | Q1a-B143 |
| 024738 | 0.0 | H2a2a | 1.00 | | | | |
| 024739 | 242.0 | A2a | 0.96 | | | | |
| 024740 | | | | | | | |
| 024741 | 0.1 | A2b1 | 0.54 | | | | |
| 024742 | 97.9 | A2a | 0.95 | 0.01 | Q1 | Q1 | Q1 |
| 024743 | 156.6 | A2a | 0.95 | 0.25 | Q1b1a1a2a~ | Q1b | Q1b-M3 (Y4303) |
| 024744 | 145.5 | A2a | 0.95 | 0.06 | Q1a1b | Q1a | Q1a-B143 |
| 024745 | 166.5 | A2a | 0.98 | | | | |
| 024746 | 172.3 | A2b1 | 0.96 | | | | |
| 024747 | 117.3 | A2a | 0.96 | | | | |
| 024748 | 395.1 | A2a | 0.99 | 0.44 | Q1b1a1a2 | Q1b | Q1b-M3 (Y4303) |
| 024749 | 302.6 | A2a | 0.99 | | | | |
| 024750 | 471.6 | A2a | 0.95 | | | | |
| 024751 | 0.0 | H76a | 0.52 | | | | |
| 024752 | 187.4 | A2a | 0.95 | | | | |
| 024753 | | | | | | | |
| 024754 | | | | | | | |
| 024755 | 427.7 | A2a | 0.93 | 0.34 | Q1b1a1a2 | Q1b | Q1b-M3 (Y4303) |
| 024756 | 136.6 | A2b1 | 0.96 | 0.22 | Q1b1a1a2 | Q1b | Q1b-M3 (Y4303) |
| 024757 | 205.9 | A2b1 | 0.98 | 0.54 | Q1b1a1a2 | Q1b | Q1b-M3 (Y4303) |
| 024758 | 525.3 | A2a | 0.95 | 0.44 | Q1a1b | Q1a | Q1a-B143 |
| 024759 | 90.4 | A2a | 0.85 | 0.17 | Q1b1a1a2 | Q1b | Q1b-M3 (Y4303) |
| 024760 | 243.4 | A2a | 0.93 | 0.52 | N1a1a1a1a | N1a | N1a-L392 |
| 024761 | 486.3 | D4b1a2a1 | 0.94 | 0.41 | Q1b1a1a2 | Q1b | Q1b-M3 (Y4303) |
| 024762 | 1003.9 | A2a | 0.99 | 0.76 | Q1b1a1a2 | Q1b | Q1b-M3 (Y4303) |
| 024763 | 552.5 | A2a | 0.95 | 0.18 | N1a1a1a1a | N1a | N1a-L392 |

| ID | MT Depth | MT Haplogroup | MT Probability | Y Depth | ISOGG Haplo | Haplo Major | Haplo Minor |
|---|---|---|---|---|---|---|---|
| 024764 | 186.5 | A2b1 | 0.98 | | | | |
| 024766 | 6469.1 | A2a | 0.97 | | | | |
| 024767 | 426.8 | A2a | 0.93 | 0.28 | N1a1a1a1a | N1a | N1a-L392 |
| 024768 | 5510.3 | A2a | 0.99 | 3.38 | Q1b1a1a2a~ | Q1b | Q1b-M3 (Y4303) |
| 024769 | 492.7 | D4b1a2a1 | 0.94 | 0.33 | Q1b1a1a2 | Q1b | Q1b-M3 (Y4303) |
| 024770 | 239.6 | A2b1 | 0.96 | 0.65 | Q1a1b | Q1a | Q1a-B143 |
| 024771 | 196.5 | A2b1 | 0.94 | | | | |
| 024772 | 179.7 | A2a | 0.93 | 0.91 | Q1b1a1a2 | Q1b | Q1b-M3 (Y4303) |
| 024773 | 320.4 | A2b1 | 0.96 | 0.68 | Q1b1a1a2 | Q1b | Q1b-M3 (Y4303) |
| 024774 | 180.2 | D4b1a2a1 | 0.96 | | | | |
| 024775 | 174.8 | A2a | 0.93 | | | | |
| 024776 | 135.6 | A2a | 0.93 | 0.66 | Q1a1b | Q1a | Q1a-B143 |
| 024777 | 4.7 | A2a | 0.70 | 0.03 | Q1a1b | Q1a | Q1a-B143 |
| 024778 | 58.9 | A2a | 0.91 | | | | |
| 024779 | 292.3 | A2a | 0.96 | | | | |
| 024780 | 184.0 | A2a | 0.99 | 0.62 | Q1a1b | Q1a | Q1a-B143 |
| 024781 | 141.2 | A2b1 | 0.96 | | | | |
| 024782 | 330.7 | A2a | 0.94 | 0.12 | Q1a1b | Q1a | Q1a-B143 |
| 024783 | 203.8 | A2a | 0.94 | | | | |
| 024784 | 194.8 | A2a | 0.96 | | | | |

## Table S9 – Identified Pathogens

List of relevant microbial species identified from the tissues included in this project, ordered by the CGG ID of the ancient individual. The location and region of the tissue are indicated, along with the taxa name, the number of reads mapping to those taxa, the damage pattern, the ani, and the average edit distance.

| ID | Location | Region | taxNameSpecies | nReads | dam5p | ani | Edit Dist Avg |
|---|---|---|---|---|---|---|---|
| 011229T | Port au Choix | Canada | Enterococcus_faecalis | 86 | 0.000 | 0.982 | 1.651 |
| 011332H | Nunalleq | United States | Treponema_succinifaciens | 389 | 0.038 | 0.984 | 1.139 |
| 011342H | Nunalleq | United States | Human_polyomavirus_5 | 151 | 0.000 | 0.990 | 0.834 |
| 011447T | Eastern Point | Canada | Helicobacter_pylori | 196 | 0.286 | 0.982 | 0.908 |
| 011448T | Englee | Canada | Helicobacter_pylori | 78 | 0.063 | 0.984 | 0.846 |
| 011448T | Englee | Canada | Human_mastadenovirus_D | 74 | 0.217 | 0.984 | 0.811 |
| 011874H | Zhokhov | Siberia | Yersinia_intermedia | 5478 | 0.031 | 0.983 | 1.495 |
| 011874H | Zhokhov | Siberia | Yersinia_sp._228 | 5554 | 0.030 | 0.983 | 1.479 |
| 011875B | Zhokhov | Siberia | Yersinia_intermedia | 22784 | 0.070 | 0.979 | 1.646 |
| 011875B | Zhokhov | Siberia | Yersinia_sp._228 | 23185 | 0.071 | 0.979 | 1.636 |
| 011877B | Zhokhov | Siberia | Yersinia_intermedia | 65 | 0.158 | 0.984 | 1.077 |
| 011878B | Zhokhov | Siberia | Yersinia_intermedia | 307 | 0.137 | 0.971 | 2.088 |
| 022561T | Vaughan | Canada | Haemophilus_aegyptius | 187 | 0.239 | 0.981 | 1.107 |
| 022561T | Vaughan | Canada | Haemophilus_influenzae | 182 | 0.217 | 0.979 | 1.231 |
| 022561T | Vaughan | Canada | Pseudomonas_stutzeri | 62807 | 0.158 | 0.972 | 2.132 |
| 022561T | Vaughan | Canada | Yersinia_enterocolitica | 2749 | 0.187 | 0.988 | 0.680 |
| 022912B | Ushki-I, layer VI | Siberia | Pseudomonas_aeruginosa | 387 | 0.008 | 0.971 | 2.791 |
| 022913B | Ushki-I, layer VI | Siberia | Enterococcus_cecorum | 52 | 0.000 | 0.977 | 2.558 |
| 022915B | Ushki-I, layer VI | Siberia | Pseudomonas_aeruginosa | 124 | 0.000 | 0.977 | 2.121 |
| 022916T | Ushki-I, layer VI | Siberia | Campylobacter_hominis | 105 | 0.000 | 0.974 | 2.381 |
| 022916T | Ushki-I, layer VI | Siberia | Pseudomonas_oleovorans | 4678 | 0.017 | 0.972 | 2.633 |
| 022920B | Ushki-I, layer VII | Siberia | Pseudomonas_fluorescens | 296933 | 0.020 | 0.984 | 1.212 |
| 024069T | Uummannaq | Greenland | Treponema_denticola | 91 | 0.111 | 0.970 | 1.912 |
| 024079T | Qoornoq | Greenland | Pseudomonas_psychrophila | 724 | 0.032 | 0.981 | 1.384 |
| 024082T | Uunartoq | Greenland | Haemophilus_influenzae | 202 | 0.063 | 0.977 | 1.361 |
| 024082T | Uunartoq | Greenland | Yersinia_intermedia | 1316 | 0.082 | 0.991 | 0.533 |
| 024084T | Ruinnæsset | Greenland | Listeria_monocytogenes | 69 | 0.000 | 0.974 | 2.029 |
| 024088T | Kangertittivatsiaq | Greenland | Yersinia_intermedia | 384 | 0.095 | 0.991 | 0.591 |
| 024098T | Suessland | Greenland | Helicobacter_pylori | 67 | 0.250 | 0.982 | 0.910 |
| 024113B | Tokareva Site | Siberia | Pseudomonas_sp._SXM-1 | 454 | 0.040 | 0.972 | 2.441 |
| 024123T | Pegtymel | Siberia | Human_alphaherpesvirus_3 | 194 | 0.146 | 0.980 | 1.180 |
| 024261T | Chegitun cemetery | Siberia | Variola_virus | 57 | 0.000 | 0.990 | 0.544 |
| 024266T | Tilichiki | Siberia | Variola_virus | 104 | 0.077 | 0.993 | 0.356 |
| 024267T | Karaga | Siberia | Yersinia_intermedia | 19002 | 0.050 | 0.983 | 1.209 |
| 024267T | Karaga | Siberia | Yersinia_sp_228 | 19123 | 0.050 | 0.983 | 1.199 |
| 024727H | Nunalleq | United States | [Brevibacterium]_frigoritolerans | 230 | 0.021 | 0.983 | 1.165 |
| 024733H | Nunalleq | United States | Enterococcus_cecorum | 581 | 0.000 | 0.977 | 2.000 |
| 024735H | Nunalleq | United States | Corynebacterium_variabile | 465 | 0.040 | 0.979 | 1.432 |
| 024747H | Nunalleq | United States | Human_polyomavirus_5 | 68 | 0.000 | 0.993 | 0.441 |
| 024747H | Nunalleq | United States | Rickettsia_felis | 75 | 0.000 | 0.976 | 1.587 |
| 024753H | Nunalleq | United States | Human_polyomavirus_5 | 84 | 0.000 | 0.991 | 0.643 |
| 024761H | Nunalleq | United States | Human_polyomavirus_5 | 101 | 0.000 | 0.991 | 0.723 |
| 024761H | Nunalleq | United States | Yersinia_intermedia | 677 | 0.041 | 0.980 | 1.383 |
| 024762H | Nunalleq | United States | Human_polyomavirus_5 | 4326 | 0.007 | 0.986 | 1.341 |
| 024781H | Nunalleq | United States | Listeria_grayi | 52 | 0.000 | 0.980 | 1.519 |

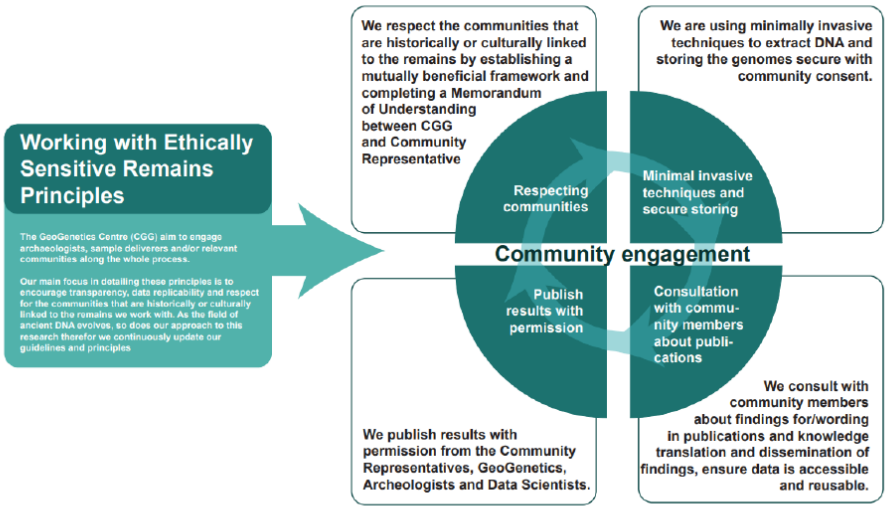
# Appendices

## *Appendix 1 - Policies on Working with Ethically Sensitive Remains*

The policies for working with ethically sensitive remains at the Centre for GeoGenetics. These parameters were outlined to increase transparency of current methodologies to archaeologists, to academic and non-academic institutions, and to communities who speak on behalf of ancestral remains.

UNIVERSITY OF COPENHAGEN
GLOBE INSTITUTE

### Policies on Working with Ethically Sensitive Remains




Please see a detailed model of CGGs approach working with ethically sensitive remains in the bottom of this document

#### Introduction

At the Globe Institute and the GeoGenetics Centre (CGG), we are heavily involved in many steps of the research data lifecycle, including collection and processing of samples, and bioinformatic analysis. We, therefore, believe it is important to set up clear and open principles of ethical data collection and management, to which we aim to adhere as part of the research process. Our main aim in detailing these principles is to encourage transparency, data replicability and respect for the communities that are historically or culturally linked to the remains we work with.

FIGURE

Below, we list our guiding principles for data collection and management in more detail.

1. **Ethical review process**

At GLOBE and CGG, we follow the ethical review guidelines by the Research Ethics Committee (REC) of the Faculty of Health and Medical Sciences at the University of Copenhagen and Horizon 2020. The REC reviews research proposals and/or publications involving human study subjects and/or data from human study subjects or archaeological remains. The

REC does not review projects with a medical focus. Where applicable projects with a medical focus are reviewed by the regional Health Research Ethics Committee.

## 2. Principles of engagement and support of Indigenous and First People's rights

We adhere to the CARE principles for indigenous data governance: https://www.gida-global.org/care

This means that we believe that:

- "Data ecosystems shall be designed and function in ways that enable Indigenous Peoples to derive benefit from the data."
- "Indigenous Peoples' rights and interests in Indigenous data must be recognised and their authority to control such data be empowered. Indigenous data governance enables Indigenous Peoples and governing bodies to determine how Indigenous Peoples, as well as Indigenous lands, territories, resources, knowledges and geographical indicators, are represented and identified within data."
- "Those working with Indigenous data have a responsibility to share how those data are used to support Indigenous Peoples' self-determination and collective benefit. Accountability requires meaningful and openly available evidence of these efforts and the benefits accruing to Indigenous Peoples."
- "Indigenous Peoples' rights and wellbeing should be the primary concern at all stages of the data life cycle and across the data ecosystem."

## 3. Principles of data collection for ancient samples

We adhere to the following principles of data collection:

- All incoming samples at GeoGenetics are covered by an agreement letter (Memorandum of Understanding) signed by the signing parties, generally the sample provider and the researcher. This details our respective obligations. This MoU is signed before any sample is provided. At the request of the sample provider, the leftover of the samples can be returned.
- We make an effort to keep careful records of previous agreements that have been made in the past at our institution, to ensure these adhere to our current data handling principles to the best of our ability. Documents are digitized and saved on our server at KU.
- Our database curator prepares, receives and organizes these agreement documents, so that every time a sample arrives, they know which agreement covers it. If there is no agreement yet, they have to prepare one.
- We respect wishes of local and descendant communities as to how ancestral remains should be handled, and work with the communities to ensure that cultural and religious ceremonies and rituals are respected, both before and/or after sampling.

## 4. Principles of sequencing and storage of ancient data

At GLOBE/CGG when we process an ancient sample, we store its genomic and associated contextual data at different levels:

- We understand ancient DNA extraction is a destructive sampling process. For this reason, we aim to maximize the information we obtain when performing this process, and generally carry out whole-genome shotgun sequencing on the DNA extracts, for this reason.

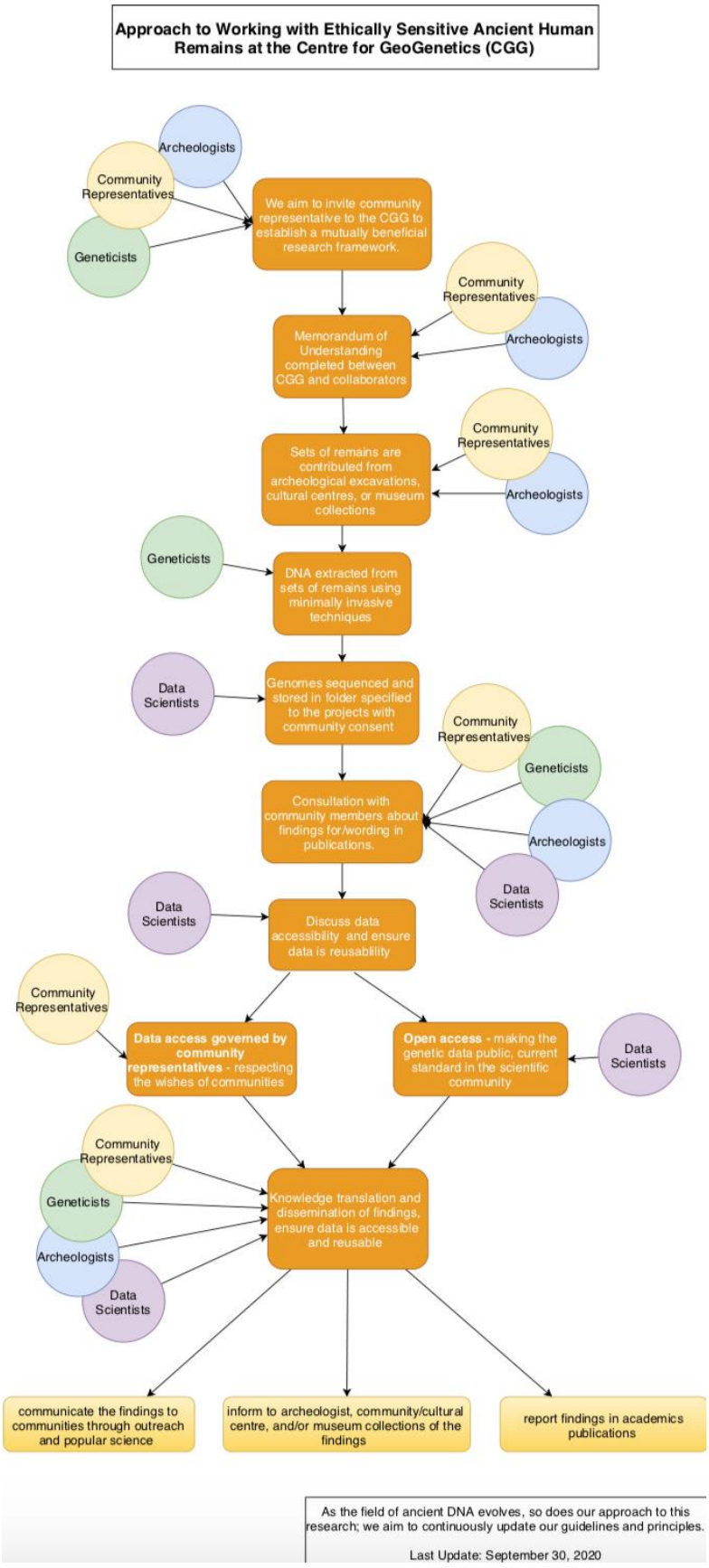
- We screen the microbial communities associated with the ancient samples, identifying pathogenic and non-pathogenic microorganisms. We believe these are important sources of information that may reflect the life and habits of the individuals whose genomes we are sequencing.
- In addition to genetic data, we aim to store as much contextual data and archaeological context information as possible, so as to maximize the insights we can obtain. This includes radiocarbon dating and isotopic analyses, wherever possible. Sometimes however, a sample may not lend itself for radiometric or isotopic analyses (e.g. due to poor preservation), so this cannot be guaranteed for every sample we collect.

Our workflow is designed to assure the FAIRness of the data produced at GLOBE/GCC. Making our data findable and accessible is a top priority, and it is reflected in our active role in the development and adoption of standards for reporting ancient genomic data developed by the Genomics Standards Consortium. We are leading the development of an extension of the MIxS checklist for reporting ancient sequences (MInAS) in collaboration with other members of the ancient DNA community. With this initiative we aim to provide a set of best practices for sharing and archiving any ancient sequence data, that will be crucial to cope with the ever-increasing number of ancient datasets produced and to facilitate the design of new sampling campaigns.

## 5. Principles of data storage and analysis of sensitive present-day human data

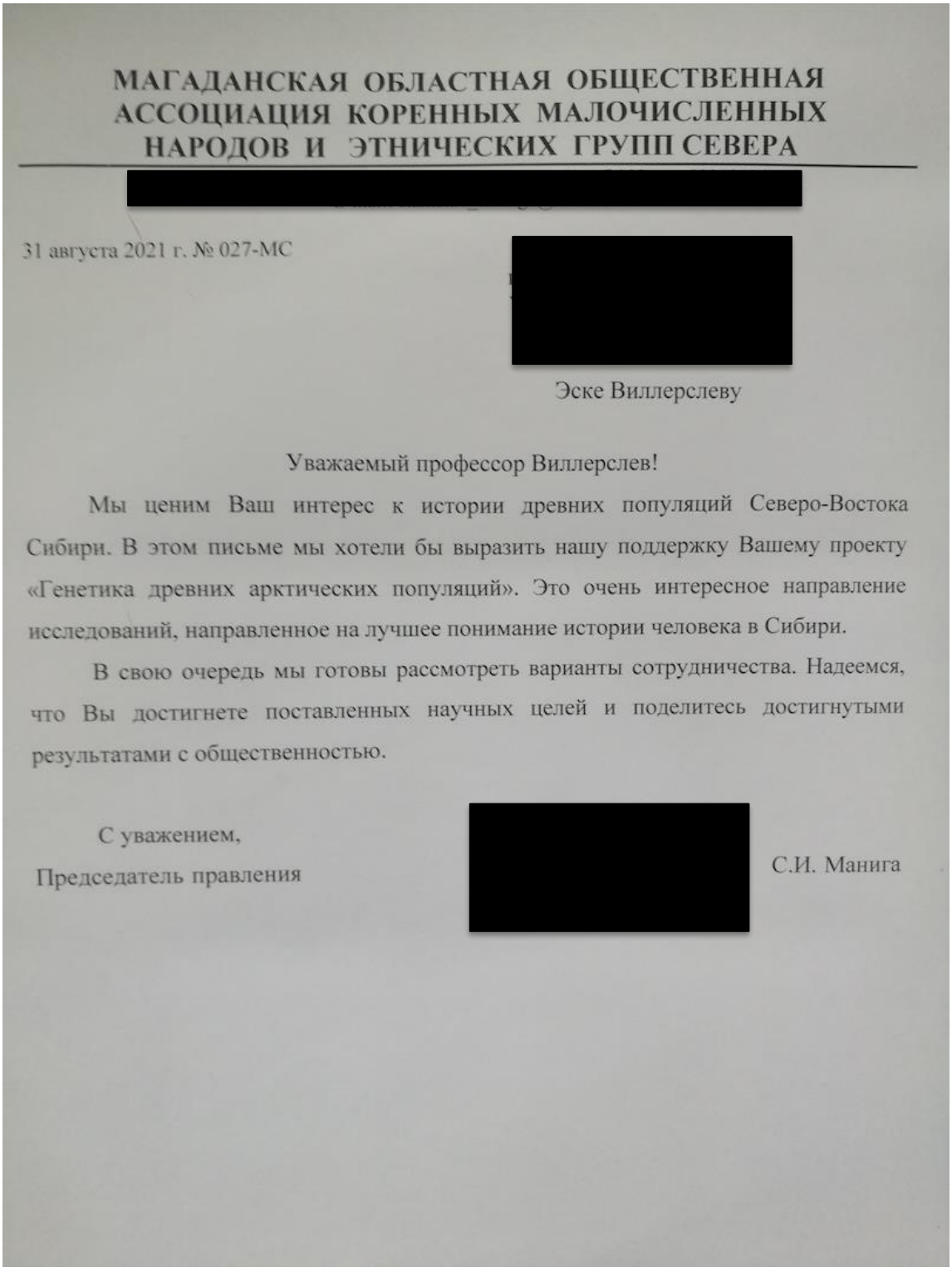
When working with sensitive present-day human data, we adhere to both the university review process detailed above and the principles and regulations established by the corresponding study that collected the data. This might include, among other things, the use of restricted computer servers (separate from the general-use servers, and accessible only to project-specific researchers) for storing and analyzing this type of data. We work closely with the University of Copenhagen's IT support (https://it.ku.dk/english/contact/faculty_it/) to ensure the necessary computational infrastructure is set up for each project.

Additionally, when working with identifiable present-day human genetic and/or phenotypic data, we are bound by the EU General Data Protection Regulation (GDPR), which places an additional set of obligations to our Centre researchers, in order to be accountable for data protection. For more information, see: https://gdpr.eu/


Approach to Working with Ethically Sensitive Ancient Human Remains at the Centre for GeoGenetics (CGG)
Archeologists
Community Representatives
Geneticists
We aim to invite community representative to the CGG to establish a mutually beneficial research framework.
Community Representatives
Archeologists
Memorandum of Understanding completed between CGG and collaborators
Community Representatives
Archeologists
Sets of remains are contributed from archeological excavations, cultural centres, or museum collections
Geneticists
DNA extracted from sets of remains using minimally invasive techniques
Data Scientists
Genomes sequenced and stored in folder specified to the projects with community consent
Community Representatives
Geneticists
Archeologists
Data Scientists
Consultation with community members about findings for/wording in publications.
Data Scientists
Discuss data accessibility and ensure data is reusability
Community Representatives
Data access governed by community representatives - respecting the wishes of communities
Open access - making the genetic data public, current standard in the scientific community
Data Scientists
Community Representatives
Geneticists
Archeologists
Data Scientists
Knowledge translation and dissemination of findings, ensure data is accessible and reusable
communicate the findings to communities through outreach and popular science
inform to archeologist, community/cultural centre, and/or museum collections of the findings
report findings in academics publications
As the field of ancient DNA evolves, so does our approach to this research; we aim to continuously update our guidelines and principles.
Last Update: September 30, 2020

## *Appendix 2.1 - Permissions for work on ancient Russian tissues*

Ethical permissions granted through RAIPON (Russian Association of Indigenous People Of the North) for the tissues of Siberian origin included in this project.



МАГАДАНСКАЯ ОБЛАСТНАЯ ОБЩЕСТВЕННАЯ АССОЦИАЦИЯ КОРЕННЫХ МАЛОЧИСЛЕННЫХ НАРОДОВ И ЭТНИЧЕСКИХ ГРУПП СЕВЕРА

[redacted]

31 августа 2021 г. № 027-МС

[redacted]

Эске Виллерслеву

Уважаемый профессор Виллерслев!

Мы ценим Ваш интерес к истории древних популяций Северо-Востока Сибири. В этом письме мы хотели бы выразить нашу поддержку Вашему проекту «Генетика древних арктических популяций». Это очень интересное направление исследований, направленное на лучшее понимание истории человека в Сибири.

В свою очередь мы готовы рассмотреть варианты сотрудничества. Надеемся, что Вы достигнете поставленных научных целей и поделитесь достигнутыми результатами с общественностью.

С уважением,

Председатель правления [redacted] С.И. Манига

## *Appendix 2.2 - Permissions for work on ancient Alaskan tissues*

Ethical permissions granted from the NIMA Corporation (via collaborator Dr. Rick Knecht) for the hair included in this project from the Nunalleq Site, in Quinhagak, Alaska.

NIMA CORPORATION

[redacted]

13 September 2011

Dear Rick,

You have informed us that among the archaeological materials excavated on our property are cut pieces of human hair which were encountered on some of the prehistoric house floors. We understand that these scraps of hair are not from burials but the remains of hair trimming from living people, much like would be found on the floor of a modern barbershop. We understand that pieces of animal fur are also found on these house floors and along with the human hair isotope analysis can provide important information about the diet and subsistence practices of our ancestors. We also understand that DNA analysis may be possible on human hair and animal fur from house floors and that these may also provide clues about the origin of our ancestors and the animals they subsisted on.

We understand that strands of fur and hair may be destroyed in the analytical process, just as small quantities of charcoal and wood samples are destroyed to get C14 dating information.

We hereby grant you permission to conduct isotope and DNA analysis of discarded scraps of human hair and animal fur, with the understanding that we will receive copies of any and all papers on the subject for our review and approval before they are published.

Sincerely,

[redacted]

Wayne W. Don
Chairman
NIMA Corporation

NIMA Corporation Board of Directors 2011-2012
Chairman Wayne W. Don, Vice-Chairman Terry D. Don, Secretary Viola S. Smith,
Treasurer Abraham David, Member Brenda Schott, Member Sandra L. King. Member Clarence Kolerok

## *Appendix 2.3 - Permissions for work on ancient Alaskan tissues*

Ethical permissions granted from Qanirtuuq Inc., (via collaborator Dr. Rick Knecht) for the hair included in this project from the Nunalleq Site, in Quinhagak, Alaska.

QANIRTUUQ
INCORPORATED

[redacted]

July 16, 2020

Dr. Rick Knecht
Department of Archaeology
University of Aberdeen
[redacted]

Dear Rick,

On behalf of Qanirtuuq, Inc., I'd like to express our continued support for ongoing laboratory research on the Nunalleq Archaeological project being done by you and your colleagues in the UK, Denmark and beyond. As you know, Qanirtuuq, Inc. is the owner of the Nunalleq Archaeological site along with the collections and samples recovered from it. We have previously reviewed and approved collection and analysis of human hair, deciduous teeth and fingernails from the site, including DNA and stable isotope work and those permissions remain in effect.

Our corporation and community welcome new DNA and stable isotope research on the prehistoric materials from Nunalleq. We look forward to learning about what you and your colleagues discover about the food they ate, adaptations to cold, the way people were related, the population size, their social differences and more. Any new discoveries about the origins of Yup'ik people would be particularly exciting.

We look forward to learning the results of your research and ask that you and your colleagues continue to share share your draft papers with us before final publication.

Sincerely,

[redacted]

Warren Jones, CEO

## *Appendix 2.4 - Permissions for work on ancient Canadian tissues (Nunangat)*

Ethical permissions granted from the Avataq Cultural Institute and the Inuit Heritage Trust, for the tissues from Quebec, and Nunavut, Canada, respectively. The approvals were communicated to Dr. Maanasa Raghavan (co-Principal Investigator on this project) and curators at the Canadian Museum of History.



Subject **RE: Request to conduct destructive analysis**
From Stacey Girling-Chris [redacted]
To 'Maanasa Raghavan' [redacted]
Date 2019-11-20 08:46

Hi Maanasa,

No, thanks for the prodding, it's been hectic.

Janet confirmed that the next stage you proposed is fine and we do not need any further information at this point in time. Please proceed.

Cheers
Stacey

-----Original Message-----
From: Maanasa Raghavan [redacted]
Sent: November 18, 2019
To: Stacey Girling-Chri
Subject: RE: Request to conduct destructive analysis

Hi again Stacey,

Sorry to rush you on this, but we've started sending other samples from the broader Arctic project (except the CMH samples) for carbon, nitrogen, and strontium isotope analyses, so just want to ensure that the CMH samples do not get left behind in case there is community and museum consent. These analyses will not need any additional material; we will send over the left-over samples from the DNA analysis. Together, the DNA and isotope analyses will enable us to reconstruct high-resolution genetic, dietary, and migratory profiles of the analyzed individuals.

Please do let me know in case you need any further information on the newly proposed isotope studies.

Best,

Maanasa

On 2019-10-25 17:53, Maanasa Raghavan wrote:
> Hi Stacey,
>
> Hope you're well. I just wanted to circle back to this and confirm whether the approvals have also considered my last request of doing isotope analysis (in addition to the approved DNA work).
>
> Best wishes,
>
> Maanasa
>
> On 2019-09-03 08:57, Maanasa Raghavan wrote:
> > Hi Stacey,
> >
> > That is great news. Could you kindly confirm if the permission is for the DNA analysis of Alarnerk/NhHd-1 Individuals 1 and 3, or does it additionally include the isotope analysis proposed in my recent email?
> >
> > Thanks to you and Janet for working together with Mr. Beveridge and the IRT to make this happen. I will ensure that the samples are handled with the utmost respect and my team will share the final report with the CMH and IHT.

Best,

Maanasa

On 2019-08-26 13:25, Stacey Girling-Christie wrote:
Hello Maanasa,

I've spoken with Janet about the processing of XIV-H:164 and we have not received any news.

I'm meeting with Janet again this week.

Cheers,
Stacey

-----Original Message-----
From: Maanasa Raghavan [redacted]
Sent: August 8, 2019 3:
To: Stacey Girling-Chri
Subject: RE: Request to conduct destructive analysis

Hi Stacey,

Hope things are well with you. Any chance there has been an update with regards to the permits for processing XIV-H:164? I also wanted to clarify that this includes both individuals 1 and 3. The original email from you (below) mentions individual 1, but not 3.

Additionally, for all samples tested for DNA at our centre, we are also implementing an isotope pipeline, which will give us maximal information on other aspects of the ancient individual's lifetime. In
particular, we will be measuring bulk and compound-specific (single amino acid) carbon and nitrogen isotopes for dietary information, which will provide us with not just the nutritional status of these individuals but also aid in marine reservoir correction of the radiocarbon dates. Additionally, we will also be performing strontium isotope analysis to see if the individual was indigenous to the region or migrated over their lifetime.
Is this something that the museum, Avataq and the IHT might be additionally interested in? I'll note here that bulk carbon and nitrogen isotopes were already measured for some of these samples when we obtained radiocarbon dates for the 2015 Science publication.
I list all the samples from CMH and the pending analyses with regards to radiocarbon dates and isotopes, assuming all parties are happy to go ahead with these:

XIV-C:340 - compound-specific (single amino acid) carbon and nitrogen isotope, strontium isotope
XIV-C:3 or XIV-C:4 - radiocarbon dating, bulk and compound-specific (single amino acid) carbon and nitrogen isotope, strontium isotope.
We did obtain a C14 date and bulk C/N isotopes for a bone from this sample; but it would be safer to get these measurements directly from the material from which we've retrieved DNA (i.e. tooth)
XIV-H:126 - compound-specific (single amino acid) carbon and nitrogen isotope, strontium isotope
XIV-H:168 - compound-specific (single amino acid) carbon and nitrogen isotope, strontium isotope
XIV-H:164 Alarnek individual 3 (pending DNA approval) - compound-specific (single amino acid) carbon and nitrogen isotope,

Best,

Maanasa

On 2019-09-03 06:47, Stacey Girling-Christie wrote:

> Good morning Maanasa,
>
> We have received good news from William Beveridge at the IHT. You may proceed with the testing as requested.
>
> Please remind everyone involved with the project to use the utmost respect when handling the remains. You are already aware of this but I will repeat Mr. Beveridge's request asking you to send IHT a copy of your final report.
>
> Regards,
> Stacey
>
> -----Original Message-----
> From: Maanasa Raghavan [redacted]
> Sent: August 28, 2019 4
> To: Stacey Girling-Chri
> Subject: RE: Request to conduct destructive analysis
>
> Hi Stacey,
>
> That's great, thank you for the update.
>
> Noted re: both the individuals from Alarnerk/NhHd-1.
>
> Best,
>
> Maanasa
>
> On 2019-08-28 13:28, Stacey Girling-Christie wrote:
>
> > Good Afternoon Maanasa,
> >
> > I've just met with Janet. She is going to contact the Inuit Heritage Trust to see if they can provide her with an update.
> >
> > Thanks for bringing the following to my attention:
> > XIV-H:164 Alarnerk individual 3 (pending DNA approval) - compound-specific (single amino acid) carbon and nitrogen isotope, strontium isotope
> >
> > Please note, the request to IHT included
> > XIV-H:164 NhHd-1 Individual 1
> > and XIV-H:166 NhHd-1 Individual 3 (see attached e-mail)
> >
> > Cheers
> > Stacey
> >
> > -----Original Message-----
> > From: Maanasa Raghavan [redacted]
> > Sent: August 26, 2019 1
> > To: Stacey Girling-Chri
> > Subject: RE: Request to conduct destructive analysis
> >
> > Thank you, Stacey.

strontium isotope

If you wish, I'd be happy to call you to discuss the above in further detail. I look forward to hearing from you.

Best,

Maanasa

On 2019-03-08 16:40, Stacey Girling-Christie wrote:

Hello Maanasa,

Sorry, no we have heard nothing at all.

Cheers
Stacey

-----Original Message-----
From: Maanasa Raghavan [redacted]
Sent: February 20, 2019
To: Stacey Girling-Chri[redacted]
Subject: Re: Request to conduct destructive analysis

Dear Stacey,

Hope you're well. I am writing to see if there was any response from the IHT.

Best,

Maanasa

On 2018-10-30 13:33, Maanasa Raghavan wrote:

Dear Stacey and Janet,

This is brilliant news. Many thanks to you both and the communities/the Avataq Institute for supporting continued scientific research on the Nunavik material. I will keep you posted on the results and am happy to draft progress reports for the Avataq Institute, if need be.

Will also wait for news regarding individual XIV-H:164 from the IHT.

Best wishes,

Maanasa

On 2018-10-30 08:55, Stacey Girling-Christie wrote:

Good morning Maanasa,

I am happy to say that we received approval for you to conduct destructive analysis on the material from the Nunavik Marine Region.

This means all the material except Individual 1 XIV-H:164
NhHd-1
Alarnerk site.

We are waiting for a response from the Inuit Heritage Trust
for
XIV-H:164

Sincerely,

Stacey

Stacey Girling-Christie

Spécialiste de l'information sur les collections - Archéologie

Centre de Ressources

Partenariats d'affaires et Gestion de l'information (PAGI)

Musée canadien de l'histoire

Collections Information Specialist - Archaeology

_Resource Centre___

_Business Partnerships and Information Management (BPIM)_

_Canadian Museum of History_

## *Appendix 2.5 - Permissions for work on ancient Canadian tissues (Ontario)*

Ethical permissions granted from the Council of the Huron-Wendat Nation for the tissues from Southern Ontario, Canada. The request and approvals were communicated via Dr. Ripan Malhi.

**From:** **Malhi, Ripan S** [redacted]
**Subject:** Re: Request for Permissions on Expanded Analyses
**Date:** July 28, 2021 at 6:19 PM
**To:** Alison Sutherland
**Cc:** Maanasa Raghav [redacted] Eske Willerslev [redacted]

Hi Alison, Maanasa and Eske -

That was a quick meeting! Louis just wanted to make sure he knew what project this was as he is involved in numerous collaborations righ: now. He said there is new Council but we have the green light to continue with the project as you described it. I informed him that once there are results that Louis and others in the First Nation will be able to review them and we can possible have a Zoom meeting if they thought that would be beneficial. Also, prior to publication, we would share the draft of the manuscript with Louis and the Huron/Wendat for comments and edits.

We can also discuss data storage and use as biocultural labels now is taking on projects to label data submitted to Geome. See https://localcontexts.org/

Please let me know if you have any questions or concerns.

best,
Ripan

On Jul 21, 2021, at 8:52 AM, Alison Sutherland [redacted] wrote:

Hi Ripan,

Thanks so much for reaching out to Louis Lesage and Ron Williamson about these expanded analyses.

Yes, the plan is to do demographic modelling for population history and evolutionary analyses for environmental selective pressures on the genome. These selective pressures include (but are not limited to) - shifts in diet, disease exposure, and cold adaptations. For contextual information on these environmental pressures we ran isotopic analyses to generate dietary profiles for these individuals and we will be running these genomes through an in-house pipeline to identify pathogens these individuals were subjected to in life.

If any questions come up during your meeting with Louis next week you can call me [redacted] and I'll do my best to clarify.

Best,
Alison

On Jul 20, 2021, at 3:38 AM, Malhi, Ripan S <[redacted]> wrote:

Hi Alison, Maanasa and Eske -

a quick update that I am meeting with Louis Lesage next week to talk about the project and the samples from Huron-Wendat Ancestors I sent you. I'm assuming the plan is to do evolutionary analyses for population history and demographics along with a selection analysis? Are there other details I should know about before I meet with him?

thanks!
Ripan

Begin forwarded message:

**From:** Louis Lesage [redacted]
**Subject:** **RE: Request for Permissions on Expanded Analyses**
**Date:** July 16, 2021 at 7:26:27 AM CDT
**To:** "Malhi, Ripan S" [redacted]

Kwe Ripan,

I propose you July 27th at 10:00 or 28th at 1:00.

<image001.jpg>

**De :** Malhi, Ripan S <[redacted]>
**Envoyé :** 15 juillet 2021 15:27
**À :** Louis Lesage <[redacted]>
**Objet :** Re: Request for Permissions on Expanded Analyses

Yes, absolutely we can have a phone call. I'm available from July 26-29 that week. Please let me know what day/time would be best for you. best, ripan

On Jul 15, 2021, at 12:52 PM, Louis Lesage <[redacted]> wrote:

Kwe Ripan,

Thanks for the update.

Can we have a phone call about this. I think it could be better if we have a chat first to talk more in details about this.

Would the week of July 26 th could be an option? Next, week, I'm in Ontario to visit different archaeological work by our crew.

Önenh!

<image002.jpg>

---

**De :** Ron Williamson <[redacted]>
**Envoyé :** 13 juillet 2021 09:07
**À :** Malhi, Ripan S <[redacted]>; Louis Lesage <[redacted]>
**Cc :** Alison Sutherland <[redacted]>; Maanasa Raghavan <[redacted]>; Eske Willerslev <[redacted]>
**Objet :** RE: Request for Permissions on Expanded Analyses

Hi Ripan (and folks),
I would be fine with this but it is really up to Louis.
Good luck with your research,
Ron

**Ronald F. Williamson, PhD**
Founder • Senior Associate

ASI • Providing Archaeological & Cultural Heritage Services

<image003.gif>[redacted]

---

**From:** Malhi, Ripan S <[redacted]>
**Sent:** July 12, 2021 10:50 PM
**To:** louis.lesage <[redacted]>; Ron Williamson <[redacted]>
**Cc:** Alison Sutherland <[redacted]>; Maanasa Raghavan <[redacted]>; Eske Willerslev <[redacted]>
**Subject:** Fwd: Request for Permissions on Expanded Analyses

Kwe Louis and Ron,

Our collaborators on the samples from the ancestors you sent us are confirming that you and Huron-Wendat Nation are okay with selection analysis from the DNA results generated. Please read below and let us know.

my best,
Ripan

> Begin forwarded message:
>
> **From:** Alison Sutherland <[redacted]>
> **Subject: Request for Permissions on Expanded Analyses**
> **Date:** July 6, 2021 at 1:37:57 PM CDT
> **To:** "Malhi, Ripan S" <[redacted]>
> **Cc:** Eske Willerslev <[redacted]>, Maanasa Raghavan <[redacted]>
>
> Dear Ripan,
>
> We are writing to enquire about the ethical permissions regarding the exploratory genomic

We are writing to enquire about the ethical permissions regarding the exploratory genomic analyses on the skeletal remains of individuals from the Huron-Wendat Nation which have been approved for destructive analyses, DNA sequencing, and isotopic work at the Centre for GeoGenetics.

Results from both the sequenced genomes and the compound-specific isotope analyses are becoming available and there are very interesting patterns emerging. The sequenced genomes are high quality and are therefore promising for deep-sequencing analyses.

This project's original aims included demographic reconstructions, cataloging ancient pathogens, as well as assessing the **selective pressure on the genomes of ancient individuals because of past pathogen exposure**. Now that both genomic and isotopic data are available, this enables us to ask another important question: **has selection occurred because of changes in dietary composition**? The genetic impact of environmental pressures, such as large shift in diet, is likely to show signals of selection in metabolism-related genes.

Similarly, when these selection analyses are run, we are likely to observe **signals of selection for genes that are responsible for adaptation to cold and other environmental factors**. Understanding this genomic impact can help to indicate how the genomes of northern populations differ from others and can, ultimately, inform health care decision and potentially improve the health status of individuals in present day populations (the latter applications are beyond the scope of our project).

As this project expands, we want to be transparent about the potential research trajectory and ensure that these analyses are supported by the representatives who speak on behalf of the ancient individuals included in this research. It is very important to us that all parties are aware and consent to any investigations we perform within the scope of the collaboration.

Do you think that these expanded selection analyses would be of interest to the community representatives? For more details on these analyses please see the information attached. If you would like clarification on anything, please do not hesitate to get into contact with us.

Best,

Alison Sutherland, on behalf of Maanasa Raghavan and Eske Willerslev

Ripan Singh Malhi
Professor
Dept. of Anthropology, Dept. of Evolution, Ecology & Behavior (affiliate) & American Indian Studies Program (affiliate)
Carl R. Woese Institute for Genomic Biology
[redacted]

Executive Editor of Human Biology
Wayne State University Press, Detroit

Wayne State University Press, Detroit
http://digitalcommons.wayne.edu/humbiol/

Ripan Singh Malhi
Professor
Dept. of Anthropology, Dept. of Evolution, Ecology & Behavior (affiliate) & American Indian Studies Program (affiliate)
Carl R. Woese Institute for Genomic Biology
University of Illinois Urbana-Champaign
[redacted]

Executive Editor of Human Biology
Wayne State University Press, Detroit
http://digitalcommons.wayne.edu/humbiol/

Ripan Singh Malhi
Professor
Dept. of Anthropology, Dept. of Evolution, Ecology & Behavior (affiliate) & American Indian Studies Program (affiliate)
Carl R. Woese Institute for Genomic Biology
University of Illinois Urbana-Champaign
[redacted]

Executive Editor of Human Biology
Wayne State University Press, Detroit
http://digitalcommons.wayne.edu/humbiol/

## *Appendix 2.6 - Request for expanded analyses on ancient Canadian tissues (New Brunswick)*

The request for expanded analyses, sent to the current Mi'gmaq Council of Listuguj, New Brunswick, Canada for the ancient tooth that originated from their land that is included in this project.

Dear Chief Darcy Gray and the Listuguj Mi'gmaq community,

Our research group would like to request your feedback and approval for expanded analyses in our study on the genetics of ancient humans in North America.

We are enquiring about ethical permissions regarding the genomic analyses on the remains of a Mi'gmaq Ancestor from the prehistoric village of Tjigog, (representing the site of Old Mission Point, New Brunswick), which was previously approved for ancient DNA sequencing at the Centre for GeoGenetics in Copenhagen, Denmark.

Previously our group has worked, in collaboration with your community and the Archaeology Department at Memorial University, on one tooth from the ancestral individual referred to as MARC1492. The genome sequenced from this ~740-year-old individual, along with genomes sequenced from other ancestral individuals from other regions of the Eastern Woodlands, provided insight on the migration waves that populated North America and were published in Raghavan et al., 2015.

There have been major advancements in the field of ancient DNA research, DNA sequencing technology, and reconstruction of ancient dietary patterns since our initial study that allows us to gain a much better understanding of human genetic history. These advancements will enable us to run more detailed analyses than thought possible when the original approvals for this work were granted. We are, therefore, requesting to expand the analysis component of our work on the Mi'gmaq Ancestor from Tjigog, New Brunswick, <u>without requesting additional human tissues.</u>

The expanded analyses include:

1- **higher resolution demographic models** which help to determine population origins, genetic diversity over time, and past genetic interactions

2- **selection analyses** which identify regions of the genome that have changed due to environmental pressures (e.g., adaptation to cold, exposure to infectious disease, or shift in dietary composition)

3- **pathogen profiles** which help identify infectious agents like disease-causing bacteria and viruses (also called pathogens) that ancient individuals were exposed to in their lifetime

4- **improved dietary patterns** which help better understand possible connections between dietary choices, genetics, and health, within the context of cultural adaptations to their environment

As this project expands, we want to be transparent about the potential research trajectory and ensure that these analyses are supported by the representatives who speak on behalf of the Ancestors. We have attached a document with additional information on our research group, the field of ancient DNA (palaeogenomics), the ongoing ancient genetics project, and the analyses we

are hoping to perform if granted the permission to proceed. It is very important to us that all parties are aware and consent to any research we perform within the scope of the collaboration.

Do you think that these expanded selection analyses would be of interest to your community? Your support in this research is very important to us and we will be very happy to answer any questions you have regarding our work.

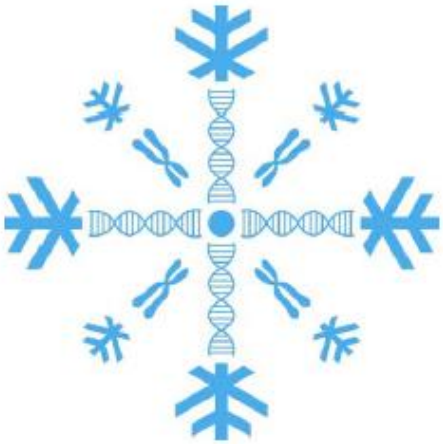
Additionally, if there are local researchers who are interested in contributing to this research, for example through sharing oral histories, providing the correct terminology, and/or learning more about becoming a genomics researcher themselves, please put them in touch as we would be very interested in working with them. Contributions to this research results in accreditation in the scientific publication.

Thank you very much for taking the time to read our request. We look forward to hearing from you.

Best,

Eske Willerslev, Maanasa Raghavan, and Alison Sutherland

**Additional Information for Chief Darcy Gray and the Listuguj Mi'gmaq Community**



**Sections**

1. Our Research Group
2. The Field of Palaeogenomics
3. The Genetics of Ancient Arctic Populations
4. Request for Permissions on Expanded Analyses

## Our Research Group

**Alison Sutherland, MSc**

Alison is a PhD student at the University of Cambridge and is the lead researcher on this project. Alison completed a bachelor's degree in Biological Sciences at the University of Guelph and a Masters in Medicine (Human Genetics) at Memorial University of Newfoundland. Alison was the 2017 winner of the Rothermere Foundation Fellowship Award enabling her to complete her PhD in the United Kingdom. With this award Alison hopes to contribute back to the knowledge base of Eastern Canada. https://gazette.mun.ca/student-life/opportunity-of-a-lifetime/

**Dr. Maanasa Raghavan**

Maanasa is an Assistant Professor at the University of Chicago whose research interests span questions and applications in multiple fields, including population genetics/genomics, anthropology, archaeology, and medical genetics. Maanasa completed her undergraduate in her hometown of Toronto, Ontario and went on to complete a Masters in Archeological Sciences at the University of Oxford. Maanasa went on to study Paleogenomics at the University of Copenhagen for her PhD where her geographical area of emphasis was the Americas. Maanasa is a leading expert on the genetics of ancient Arctic populations having published an article on the peopling of the North American Arctic in Nature in 2014. For further details on Maanasa's work please visit: https://hgen.uchicago.edu/program/faculty/maanasa-raghavan

**Prof. Eske Willerslev**

Eske is a world-renowned scientist focused primarily on the fields of ecology and evolution. Previous to entering academia, Eske was a fur-trapper in Siberia and led several expeditions to collect ethnographic material from the Yakut, Even, Chuckchi, Koryak, and Jakagir populations, as well as plant material and prehistoric material from mammoth, steppe bison, and wild horse for the Prehistoric Museum, Moesgaard. Eske began as an environmental microbiologist and then moved into the field of invertebrate systematics, and later into mammalian population genetics and community ecology.

Over recent years he has focused major parts of his research on understanding processes forming contemporary human genetic diversity, distribution, and disease load while in his role as Director of Centre of Excellence in GeoGenetics and Professor at the University of Copenhagen and Cambridge. Eske is known for his multidisciplinary approach and knowledge dissemination abilities. Eske is one of the few scientists that has succeeded in approaching and being accepted by several of the world's Indigenous peoples, which even led in 2014 to the honorary adoption into the Crow tribe, Montana (Apsaalooke), with name ChiitdeeXia'ssee (meaning Well Known Scout). For more information about Eske's work please visit: https://www.zoo.cam.ac.uk/directory/professor-eske-willerslev

## The Field of Palaeogenomics

The study of human populations is a long-standing field that involves different disciplines, from the natural sciences to the humanities. Genetics, the study of DNA and how DNA is transmitted across generations, allows us to understand the biology of an individual and their family. Population genetics, the study of DNA at a population level, allows us to understand the genetic differences within and between human groups. Understanding the differences and similarities between populations help us to get a better insight into human variation and health as well as human history. It enables us to use biological information to understand where present-day populations originate, trace past interactions and exchange of genetic material between populations, how present-day populations acquired beneficial, or adaptive, traits such as the ability to live in extremely cold environments, and similar questions of interest to geneticists, anthropologists and, importantly, communities themselves.

Palaeogenomics is the examination of ancient genomes at both the individual and population level through space and time. The first ancient human genome sequenced was of a ~4000-year-old seal hunter from Qeqertasussuk, Western Greenland, affiliated to the Saqqaq culture of Greenland, the earliest regional culture as per the archaeological record (Rasmussen et al., 2010). This effort was made possible by our research group at the Centre for GeoGenetics in

Copenhagen and was led by Prof Eske Willerslev, in close collaboration with the Greenlandic National Museum. This work provided several insights on the genetic relationship of the Saqqaq individual to present-day living populations in the region as well as aspects of this ancient individual's biological adaptations and health.

## The Genetics of Ancient North American Populations

Through the DNA analysis of ancient North American populations, we can address research questions about population migrations, population interactions, human adaptation, human evolution, and more. The goal of this project is therefore to have a better understanding of the human diversity and population history of the region. This non-profit research aims to gain a better understanding of human history in the North, specifically, the genetic impact of shifts in dietary composition and exposure to infectious diseases on individuals through time.

This project includes ancestral remains which have been contributed by several collaborators and communities from the circumpolar North. In addition to existing genomes such as the Saqqaq seal hunter mentioned above, this research includes many individuals from various regions of the Eastern Woodlands (who are from the same time period as the ancient individual from Old Mission Point, New Brunswick - more details below). These tissues were contributed to this project by the Huron Wendat Nation, who provide community and ethical oversight.

The one set of ancestral remains that we are consulting with you on has previously been approved for scientific analyses by the Listuguj Mi'gmaq community, in conjunction with the Archeology Department of Memorial University of Newfoundland. This tooth has been radiocarbon dated to ~740 years ago and had bulk carbon and nitrogen isotopic profiles evaluated, which indicated a diet with a mixed of marine and terrestrial protein sources.

**Table 1** - Information pertaining to the ancestral Mi'gmaq remains from New Brunswick

| Site ID | MUN Archeology Alias | Culture | Tissue | Location | Date in Years CE |
|---|---|---|---|---|---|
| C1Dq-1 | MARC1492 | Mi'gmaq | Tooth right mandibular first premolar (RPM1) | prehistoric village of Tjigog (representing the site of Old Mission Point, New Brunswick) | 1260 |

Previously, this tooth underwent DNA extraction at the Centre for GeoGenetics, Natural History Museum of Denmark. The genome was sequenced and published in Genomic evidence for the Pleistocene and recent population history of Native Americans, Raghavan et al., 2015. These genomes were sequenced in 2011 with older laboratory techniques that were widely available then and did not yield very much genetic material. Despite less data, the research group was able to estimate the timings of the first migrations into the Americas. This research was also able to estimate the timing of the split between Northern and Southern Native Americans to around 13,000. Because there have been significant advances in sequencing technology over the last ten years, we are motivated to re-examine these migrations, as well as other genetic phenomena, with more genomic data.

## Request for Permissions on Expanded Analyses

The desired expanded analyses for this research project include demographic modeling and selection analyses.

Demographic modeling helps to identify population origins, population diversity, and population size - as well as interactions with other groups in the region. In this instance, we would examine the overlap of the ancient Mi'gmaq with other neighbouring populations.

Selection analyses identify regions of the genome that have changed due to environmental pressures. Over time and space, humans have undergone various adaptations to their environment. These adaptations are observable on the molecular level when genomic data is compared between populations or between individuals within a population. These comparisons indicate where these differences (otherwise known as genetic variations) occur in the genome. When these variations occur in genes that have a function known to cause a certain trait, and that trait is related to the adaptation that occurred, it is concluded that this is associated with selective pressures.

Within this project, we will be predominately exploring the effects of two environmental pressures: dietary intake and infectious disease. Dietary profiles can be observed through compound specific isotopic analyses which are able to determine not only the major protein source (i.e., marine vs terrestrial) but also the specific species that an individual was consuming. For example, these analyses are detailed enough to discern between people eating ringed seals or bearded seals. Similarly, infectious disease profiles can be observed because when the genomes of ancient humans are sequenced, so are the genomes of the pathogens (vectors of infectious diseases e.g. bacteria and virus) they were exposed to in life. Previous studies based in different regions of the world have found ancient individuals who carried microbes responsible for the plague, hepatitis B, salmonella, and many more infectious diseases. Pathogen exposure can be used as a proxy to population interactions and in this instance, we will be investigating pathogen exchange between

Arctic (and sub-Arctic) ancient indigenous populations, as well as pathogen exchange after European colonization.

Below is an image from Fan et al., 2016, that depicts various human adaptations to their environment. As seen in this figure there have been numerous selection-based explorations on adaptations to cold, diet, and stature in the modern Greenlandic Inuit, informed primarily by previous studies on present-day populations (Fumagalli et al., 2015).

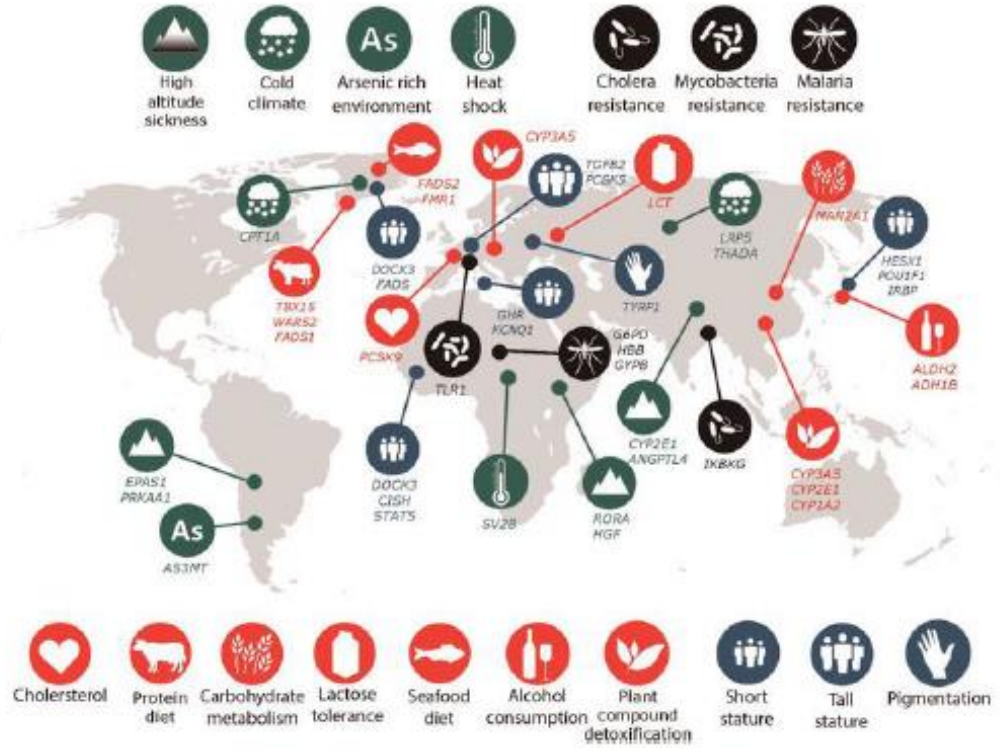



**Figure 1** - Fan et al. (2016), environmental adaptations are indicated by symbols representing either the physiological trait or environmental pressure. The gene under selection due to that adaptation is indicated below the symbols.

## References

**Rasmussen, M.,** Li, Y., Lindgreen, S., Skou Pedersen, J., Albrechtsen, A., Moltke, I., … Willerslev, E. (2010). Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature*, *463*. https://doi.org/10.1038/nature08835

**Raghavan, M.,** Degiorgio, M., Albrechtsen, A., Moltke, I., Korneliussen, T. S., Grønnow, B., … Willerslev, E. (2014). The Genetic Prehistory of the New World Arctic. *Science*, *345*(6200) https://doi.org/10.1126/science.1255832

**Fan S.,** Hansen ME., Lo Y., Tishkoff SA., (2016 ). Going global by adapting local: A review of recent human adaptation. Science. 354(6308):54-59, https://doi.org/10.1126/science.aaf5098

**Fumagalli M.,** Moltke I., Grarup N., Racimo F., … Nielsen R. (2015 ) Greenlandic Inuit show genetic signatures of diet and climate adaptation. Science. 349(6254):1343-7, https://doi.org/10.1126/science.aab2319

## *Appendix 2.7 - Permissions for work on ancient Canadian tissues (Newfoundland)*

Permissions granted, via curators at The Rooms Museum in St. John's, Newfoundland, from the Indigenous groups of Newfoundland and Labrador (the Innu Nation, the Miawpukek First Nation, the Nunatsiavut Government Research Advisory Committee, and the Qalipu First Nation).

### Permissions for Genetics Research on Ancient Remains from the Indigenous Groups of Newfoundland and Labrador

**1- Permission from Nunatsiavut**

Hi Kate,
my team and I have reviewed this application. We have no concerns with the research request. We would like to have a copy of the final report when it comes out.
One thing I would like to see in future is research requests such as this to go to the NGRAC at [redacted]. NGRAC is the Nunatsiavut Government Research Advisory Committee.
Thank you, nakummek.

Lena

---

**From:** Wolforth, Kate <[redacted]
**Sent:** May 2, 2022 9:34 AM
**To:** Lena Onalik <[redacted]>
**Subject:** Research request

**CAUTION: This email originated from outside of the organization. Do not click links or open attachments unless you recognize the sender and know the content is safe.**

Dear Lena,

In November 2021, Dr. Eske Willerslev, an internationally renowned researcher at Cambridge University, forwarded a letter to all Indigenous leaders about the genetic analysis project below. The Rooms and the Provincial Archeology Office are in support of this research as it does not require any transfer or extraction of further material and it presents a valuable opportunity to learn more about early Arctic peoples and their lives. The DNA available is ancient and has no genetic connection to living peoples so far. This research project has received support from other Indigenous groups and organizations in the circumpolar region in Canada and beyond.

We've advised Dr. Willeslev that we'd reach out to our own contacts within the Indigenous governments and organizations about this again, and that if no comments are received by him by June 1, we would consider all groups advised and the research may proceed.

Can you give me a call, Lena, if you'd like to discuss?
Thanks so much – I hope to speak with you soon.
Kate

**Kate Wolforth (she/her) BFA, MMSt.**
**Acting Director**
Museums and Galleries
The Rooms
[redacted]
therooms.ca

The Rooms acknowledges the province of Newfoundland and Labrador as the ancestral homelands of many diverse populations of Indigenous peoples who have contributed to 9,000 years of history including the Beothuk on the Island of Newfoundland. Today, this province is home to diverse populations of Indigenous and other people. We would also like to acknowledge with respect the diverse histories and cultures of the Mi'kmaq, Innu, and Inuit.

## 2- Permission from the Innu Nation

Hello Kate!

This looks like very interesting work and I think would be great to see what kind of information he can come up with.

Thanks for keeping me in the loop! Jodie

Sent from my iPhone

On May 2, 2022, at 6:26 AM, Wolforth, Kate [redacted] wrote:

Dear Jodie,

In November 2021, Dr. Eske Willerslev, an internationally renowned researcher at Cambridge University, forwarded a letter to all Indigenous leaders about the genetic analysis project below. The Rooms and the Provincial Archeology Office are in support of this research as it does not require any transfer or extraction of further material and it presents a valuable opportunity to learn more about early Arctic peoples and their lives. The DNA available is ancient and has no genetic connection to living peoples so far. This research project has received support from other Indigenous groups and organizations in the circumpolar region in Canada and beyond.

We've advised Dr. Willeslev that we'd reach out to our own contacts within the Indigenous governments and organizations about this again, and that if no comments are received by him by June 1, we would consider all groups advised and the research may proceed.

Can you give me a call, Jodie, if you'd like to discuss?
Thanks so much – I hope to speak with you soon.
Kate

**Kate Wolforth (she/her) BFA, MMSt.**
**Acting Director**
Museums and Galleries
The Rooms
[redacted]
therooms.ca

The Rooms acknowledges the province of Newfoundland and Labrador as the ancestral homelands of many diverse populations of Indigenous peoples who have contributed to 9,000 years of history including the Beothuk on the Island of Newfoundland. Today, this province is home to diverse populations of Indigenous and other people. We would also like to acknowledge with respect the diverse histories and cultures of the Mi'kmaq, Innu, and Inuit.

## 3- Permission from Miawpukek First Nation

Hi Kate,

I apologize for the delay in getting back to you. It seems that the new internet phone system the Band is installing is messing up our server and emails.

I have read the information provided on the research project of Dr. Willeslev regarding ancient DNA of Artic Peoples and believe it to be a worthy project. On behalf of Miawpukek First Nation we approve of the project and hope to be a part of the group that gets to read the final report.

Regards [redacted]

Colleen

**From:** Wolforth, Kate <[redacted]>
**Sent:** Tuesday, May 10, 2022 5:14 PM
**To:** [redacted]
**Subject:** Approval of research

Hi Colleen,

This is the email reminder you asked for regarding approval of the research Dr. Willeslev is doing into ancient DNA of Arctic Peoples. Could you confirm that this is fine to proceed with?

Many thanks!
Kate

**Kate Wolforth (she/her) BFA, MMSt.**
**Acting Director**
Museums and Galleries
The Rooms
[redacted]
therooms.ca

The Rooms acknowledges the province of Newfoundland and Labrador as the ancestral homelands of many diverse populations of Indigenous peoples who have contributed to 9,000 years of history including the Beothuk on the Island of Newfoundland. Today, this province is home to diverse populations of Indigenous and other people. We would also like to acknowledge with respect the diverse histories and cultures of the Mi'kmaq, Innu, and Inuit.

## 4- Permission from the Qalipu First Nation

Good afternoon Kate,

As project is using data already collected and no new sampling of human remains will be conducted, the Qalipu First Nation supports this project.

Take care

Keith Goulding
[redacted]



**From:** Wolforth, Kate <[redacted]
**Sent:** May 9, 2022 3:58 PM
**To:** Brendan Mitchell [redacted]
**Cc:** Paulette Brinston [redacted]
**Subject:** Research request
**Importance:** High

Dear Chief Mitchell,

In November 2021, Dr. Eske Willerslev, an internationally renowned researcher at Cambridge University, forwarded a letter to all Indigenous leaders in NL about the genetic analysis project outlined below. The Rooms and the Provincial Archeology Office are in support of this research as it does not require any transfer or extraction of further material and it presents a valuable opportunity to learn more about early Arctic peoples and their lives. The DNA available is ancient and has no genetic connection to living peoples so far. This research project has received support from other Indigenous groups and organizations in the circumpolar region in Canada and beyond. We advised Dr. Willerslev that we generally work with all groups in NL when consulting about peoples who have no known living relations.

We've advised Dr. Willeslev that we'd reach out to Indigenous governments and organizations that we work with and that if no comments were received by him by June 1, we would consider all groups advised and the research may proceed.

Thank you for your time and consideration of Dr. Willeslev's request. Please let me know if you or any of your members have any questions about the role of The Rooms, or about the research proposed.

Regards,

**Kate Wolforth (she/her) BFA, MMSt.**
**Acting Director**
Museums and Galleries
The Rooms
[redacted]
therooms.ca

The Rooms acknowledges the province of Newfoundland and Labrador as the ancestral homelands of many diverse populations of Indigenous peoples who have contributed to 9,000 years of history including the Beothuk on the Island of Newfoundland. Today, this province is home to diverse populations of Indigenous and other people. We would also like to acknowledge with respect the diverse histories and cultures of the Mi'kmaq, Innu, and Inuit.

## 5- Email to NunatuKavut

*** **Important Context**: at the time of this email exchange NunatuKavut (otherwise known as the Inuit-Metis or Labrador Metis) is was a unrecognised Inuit territory. "The Inuit Tapiriit Kanatami, which represents Inuit in Canada, reject the NunatuKavut community coucil's stance as a separate Iunit organization." https://www.cbc.ca/news/canada/newfoundland-labrador/itk-letter-federal-government-exclude-nunatukavut-1.6209086

And this was the email sent to NunatuKavut CC that we did not receive a reply to.

Best,
Kate

**From:** Wolforth, Kate
**Sent:** May 2, 2022 10:05 AM
**To:** Bryn Wood[redacted]
**Subject:** Research request

Dear Bryn,

In November 2021, Dr. Eske Willerslev, an internationally renowned researcher at Cambridge University, forwarded a letter to all Indigenous leaders about the genetic analysis project below. The Rooms and the Provincial Archeology Office are in support of this research as it does not require any transfer or extraction of further material and it presents a valuable opportunity to learn more about early Arctic peoples and their lives. The DNA available is ancient and has no genetic connection to living peoples so far. This research project has received support from other Indigenous groups and organizations in the circumpolar region in Canada and beyond.

We've advised Dr. Willeslev that we'd reach out to our own contacts within the Indigenous governments and organizations about this again, and that if no comments are received by him by June 1, we would consider all groups advised and the research may proceed.

Can you give me a call, Bryn, if you'd like to discuss?
Thanks so much – I hope to speak with you soon.
Kate

**Kate Wolforth (she/her) BFA, MMSt.**
**Acting Director**
Museums and Galleries
The Rooms
[redacted]
therooms.ca

The Rooms acknowledges the province of Newfoundland and Labrador as the ancestral homelands of many diverse populations of Indigenous peoples who have contributed to 9,000 years of history including the Beothuk on the Island of Newfoundland. Today, this province is home to diverse populations of Indigenous and other people. We would also like to acknowledge with respect the diverse histories and cultures of the Mi'kmaq, Innu, and Inuit.

## *Appendix 2.8 - Permissions for work on ancient Greenlandic tissues*

Permissions for work on the 50 ancient tissues from Greenland, granted by the Greenland National Museum, communicated through Dr. Niels Lynnerup.

**Fra:** Niels Lynnerup
**Sendt:** 7. april 2022 15:01
**Til:** Christian Koch Madsen [redacted]
**Emne:** VS: aDNA-Projekt

Kære Christian,

Hermed vedr Alison´s og Manasas projekt.
Mvh
Niels

**Niels Lynnerup**
Institutleder, professor

**Københavns Universitet**
Retsmedicinsk Institut
[redacted]

KØBENHAVNS UNIVERSITET

Sådan beskytter vi persondata

**Fra:** Christian Koch Madsen [redacted]
**Sendt:** 26. oktober 2018 19:49
**Til:** Niels Lynnerup [redacted]; Daniel Thorleifsen <[redacted]>
**Emne:** aDNA-Projekt

Kære Niels

NKA giver hermed sin tilladelse til sampling af humane rester fra Grønland opbevaret på Panum Instituttet under projektet: "*The Genetic and Immunological Impact of European Contact on Indigenous North Americans and Greenlandic Inuit*" (projektbeskrivelse vedhæftet).

--
Inussiarnersumik inuulluaqquakkit
Med venlig hilsen / Best regards

Christian Koch Madsen
Pisortap tullersortaa, katersugaasivilerisoq/ Deputy Director
Postdoctoral researcher

Nunatta Katersugaasivia Allagaateqarfialu/
Greenland National Museum & Archives
[redacted]

[redacted]

On 2018-10-27 08:47, Niels Lynnerup wrote:

t.o.

FRA: Christian Koch Madsen [redacted]
SENDT: 26. oktober 2018 19:49
TIL: Niels Lynnerup; Daniel Thorleifsen
EMNE: aDNA-Projekt

Kære Niels

NKA giver hermed sin tilladelse til sampling af humane rester fra Grønland opbevaret på Panum Instituttet under projektet: "_The Genetic and Immunological Impact of European Contact on

Indigenous
North Americans and Greenlandic Inuit_" (projektbeskrivelse vedhæftet).

--

Inussiarnersumik inuulluaqquakkit

Med venlig hilsen / Best regards

Christian Koch Madsen

Pisortap tullersortaa, katersugaasivilerisoq/ Deputy Director

Postdoctoral researcher

Nunatta Katersugaasivia Allagaateqarfialu/

Greenland National Museum & Archives

Links:
------
[1] http://www.natmus.gl

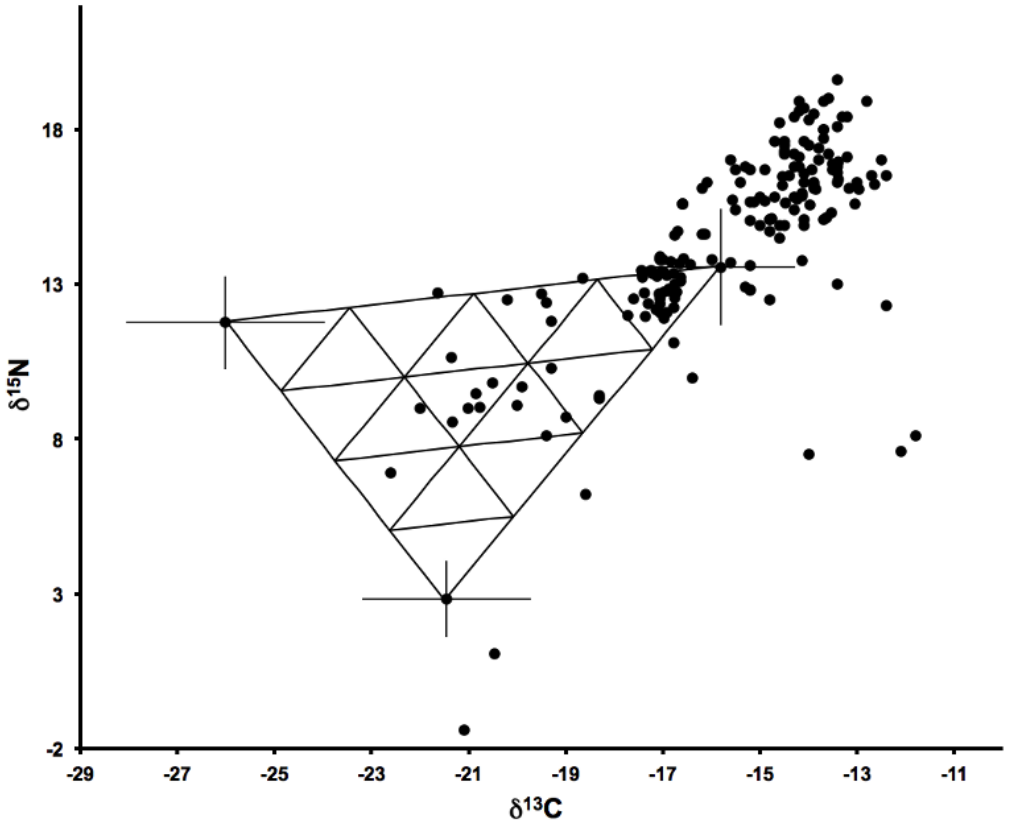
## *Appendix 3 - Report of Bulk Isotope and Dietary Contributions*

Report detailing the three-end member (terrestrial, freshwater, and marine diets), two-isotope (carbon and nitrogen) mixing model created for reconstructing the proportional contribution of dietary components of the ancient individuals included in this project. This report was prepared by Stormy Fields, Dr. Ben Barst, and Dr. Matthew Wooller from the Alaska Stable Isotope Facility.

Wooller, Barst and Fields ASIF initial bulk isotope report. Oct 2022

All of the samples run by ASIF that produced bulk isotope data also produced good atomic C:N ratios (~3.5), indicating good collagen preservation. Some samples did not produce adequate masses of collagen to produce bulk isotope data. There were a very few (4 out 129) samples where the ASIF isotope data relative to previous data produced by other labs were marginally different. Sample numbers 11874 (Hair) and 16048 had marginally different d13C data. Samples 24249 and 24252 marginally different d13C and d15N data. None of these samples that demonstrated an apparent difference between data sets were those previously analyzed by Britton et al. The mismatch in the hair sample could be because of differences in the methods used to prepare the hair for analysis. In ASIF we sampled the whole hair to give an average 'lifetime' value. Previous data on the hair sample might have been a section of the hair. ASIF's analyses along some of the hairs have shown significant seasonal variation, consistent with shifting seasonal diets towards and away from inclusion of marine resources. There were strong positive relationships between the ASIF data and previously run data (d13C = R2 0.9 and d15N = R2 0.9). In the cases of samples 24249, 16048, and 24252 the ASIF data were a closer match to similar samples from the same site and context compared to the prior data. In the case of the sample 11874 this sample was a hair and we might expect some mismatch (see above). Given all of these preliminary analyses we conclude that for further mixing model analyses using the bulk data we were justified to consistently use the ASIF data for internal consistency and because the ASIF data set ended up being more complete.

We used a simple three end member (diet - marine, terrestrial and freshwater), two isotope (carbon and nitrogen) mixing model to determine the proportional contribution of marine, terrestrial and marine resources to the diet of each individual (e.g. Choy et al.). Tissue to diet fractionation factor adjustments followed




previous research (Hoffman et al. 2020). The results are plotted in the figure above. Initial observations of the mixing model triangle are that a large mass of individuals fall higher than the mean marine diet end members. This indicates that these individuals may have been targeting specific species with relatively higher d15N values compared to the average values used in the model (e.g. some marine fish such as Cod and Wolf Fish) or higher trophic level pray (e.g. polar bears). The mean isotope values for the end members used here are also consistent with the mean and ranges of the previous end members used by Raghavan et al., 2014 (*The Genetic Prehistory of The New World Arctic*). This emphasizes the need for the compound specific isotope data to disentangle the influences of trophic level on elevating the bulk d15N values. For example an individual could feed on lower trophic shoreline marine resources like shellfish while another focused on including consumption of high trophic level food resources such as polar bear. Samples 22559, 22560 all 22561 all fall out of the mixing model triangle in a space that is different from being hyper-marine resource individuals. These are all Huron-Wendat in Canada about 600 years ago. It is likely that these individuals are including some C4 plant material in their diet along with freshwater fish. Some individuals clearly place towards more terrestrial and aquatic diet end members and therefore would be less likely to be

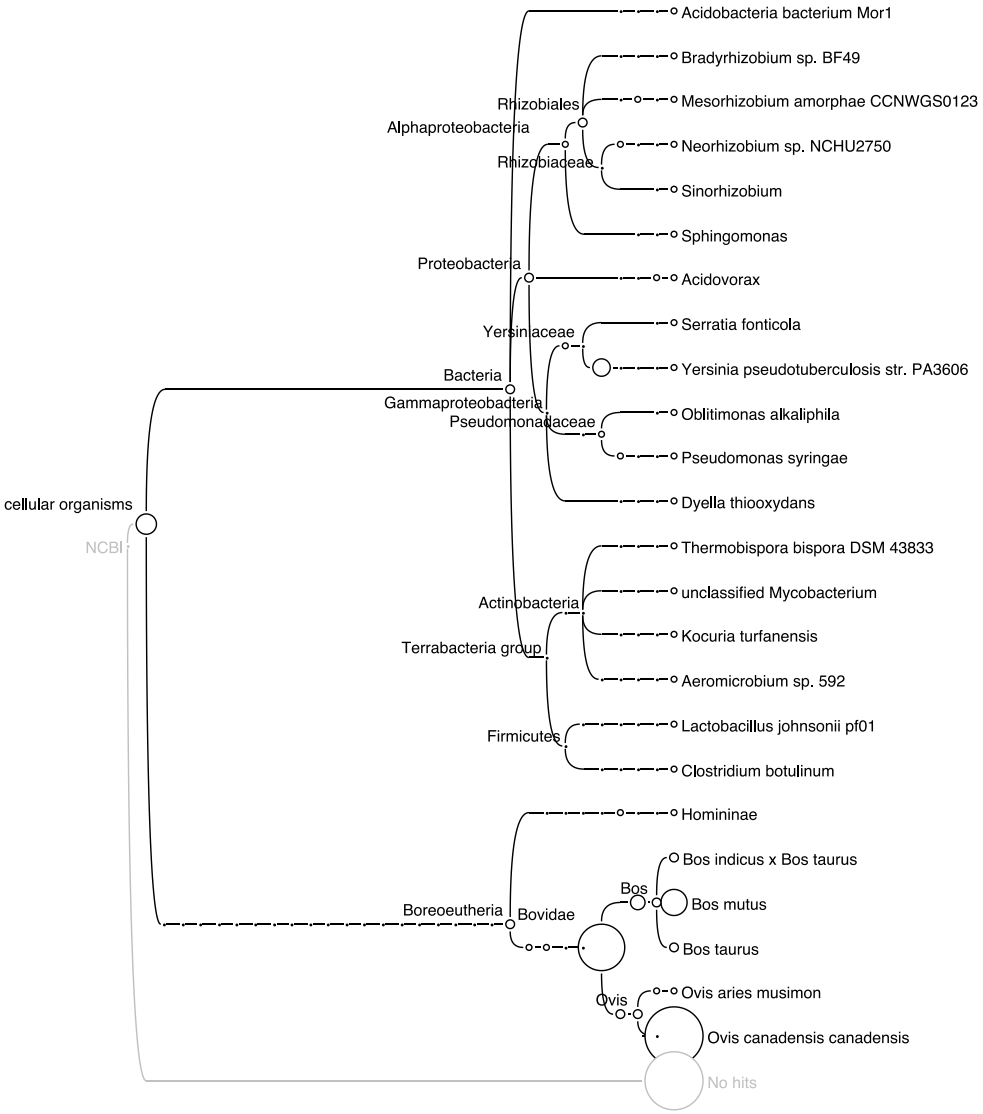
influenced by a marine reservoir influence on the radiocarbon ages. Broadly speaking the mixing model results here produce similar inferences to those produced in previous studies. Some anomalies in the database are highlighted in yellow in the first sheet, including a collection of 'orphaned' data and specimens highlighted in yellow at the base of sheet 1.

K. Choy#, B. A. Potter, H. J. McKinney#, J. D. Reuther, Shiway W. Wang*, and **M.J. Wooller**. (2016) Chemical profiling of ancient hearths reveals recurrent salmon use in Ice Age Beringia. PNAS 113 (35): 9757–9762

Halffman et al., et al, **M.J. Wooller**, (2020). Ancient Beringian paleodiets revealed through multiproxy stable isotope analyses. Science Advances. DOI: 10.1126/sciadv.abc1968

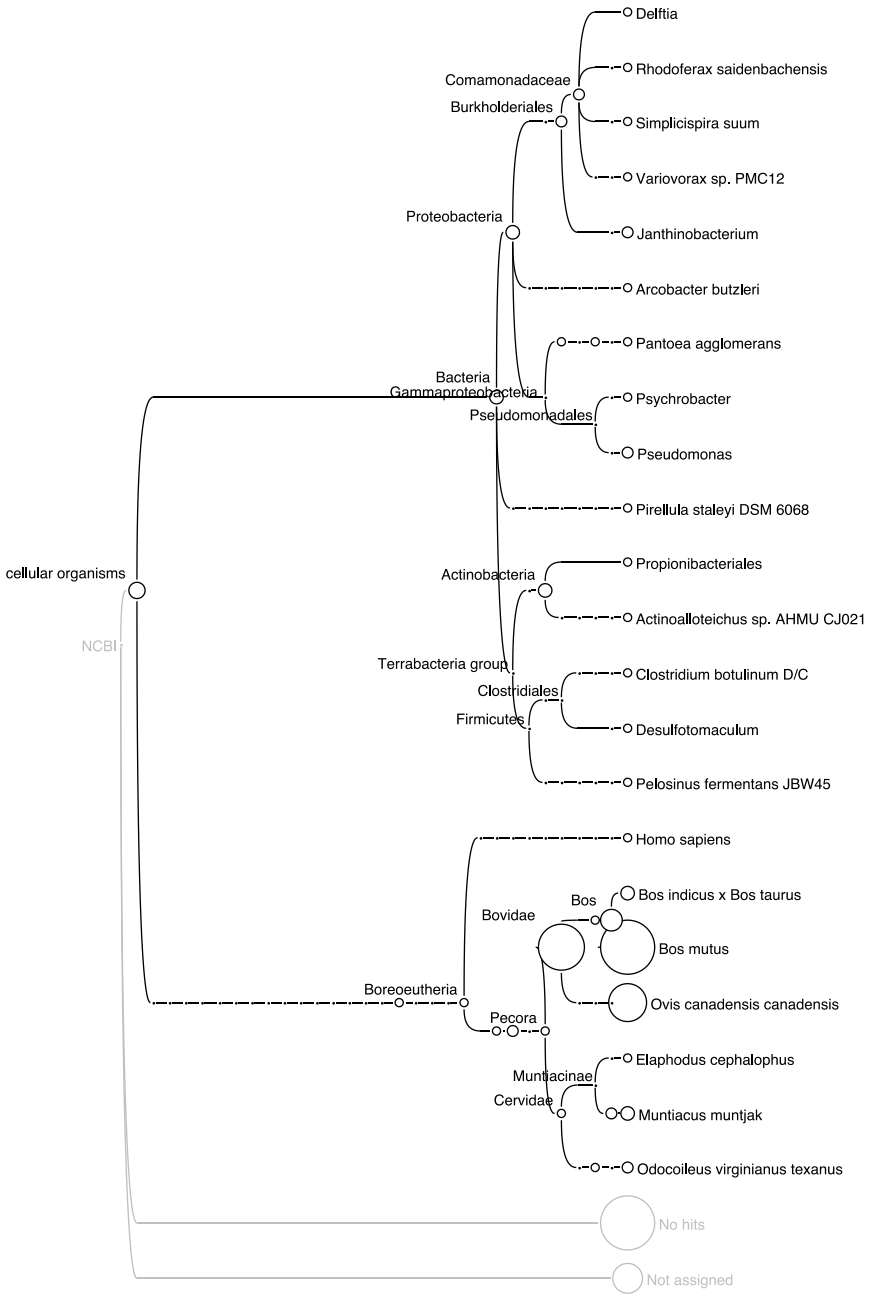
## *Appendix 4.1 - Preliminary investigation into species identification of 024263*

The MEGAN (Metagenome Analyzer) result for tissue 024263, where the most reads mapped to the *Ovis canadensis canadensis* species within the *Bovidae* family.



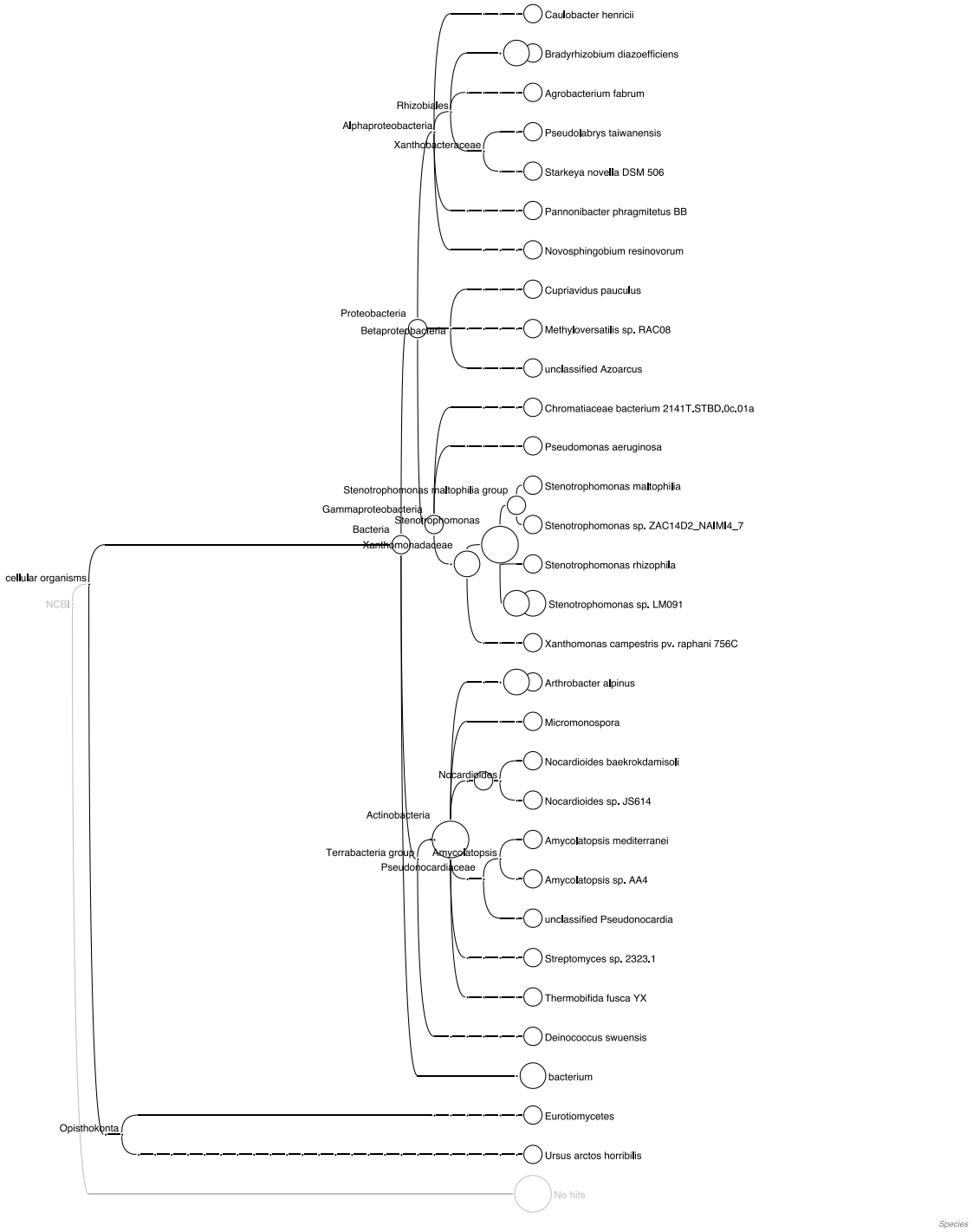
## *Appendix 4.2 - Preliminary investigation into species identification of 011875*

The MEGAN result for tissue 011875, where the most reads mapped to the *Bos mutus* species within the *Bovidae* family.



## *Appendix 4.3 - Preliminary investigation into species identification of 024119*

The MEGAN result for tissue 024119, where the only reads mapping to a mammalian genome was the *Ursus arctos horribilis* species.

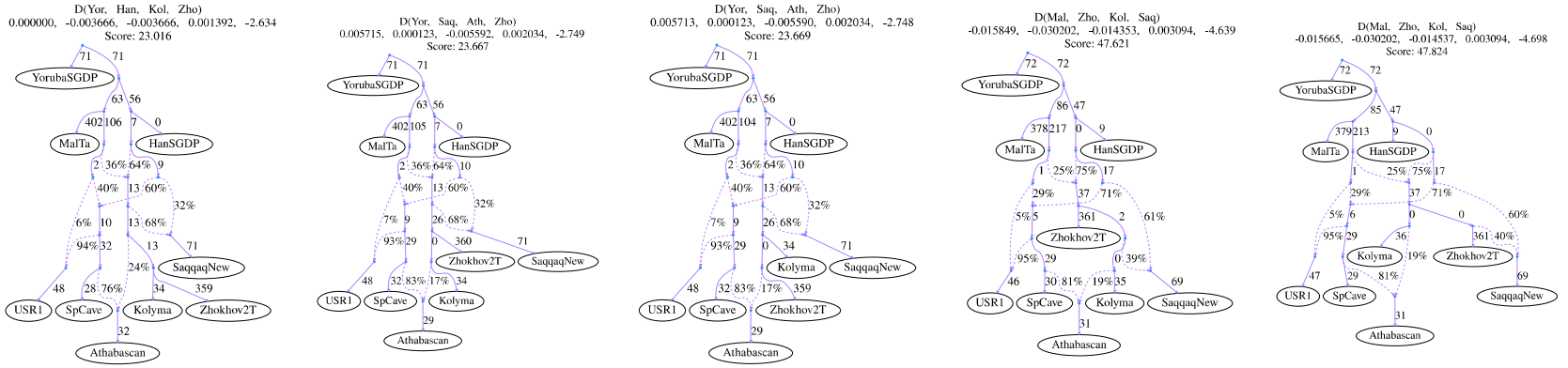


## *Appendix 5 - Enlarged figures of Admixture Graph results*

Admixture Graph **Figures S5.1**- **Figures S5.6** depict admixture graph results from high-depth cultural representatives of different groups added to the 'Starting Graph' using qpGraph. The top panel in the figure shows the five non-admixed extensions of the seed graph with the best fit scores, and the bottom panel shows the five possible admixed extensions of the seed graph with the best fit scores for each cultural representative. There are three rows of text which indicate the worst *D*-statistic residual and the four-population set leading to such a residual; the values of the *D*-statistic (both expected and observed), the value of the residual and the standard error, as well as the residual Z-score; and the overall model fit score. The graphs in each panel are sorted from left to right in order of model fit score. Solid lines connecting the leaves illustrate direct linkages with the values to the right indicating optimised drift parameters. Dashed lines connecting the leaves illustrate admixed lineages, with percentages to the right indicating the admixture proportions.

# Zhokhov Non-admixed and Admixed Models

## No Admixture




## With admixture

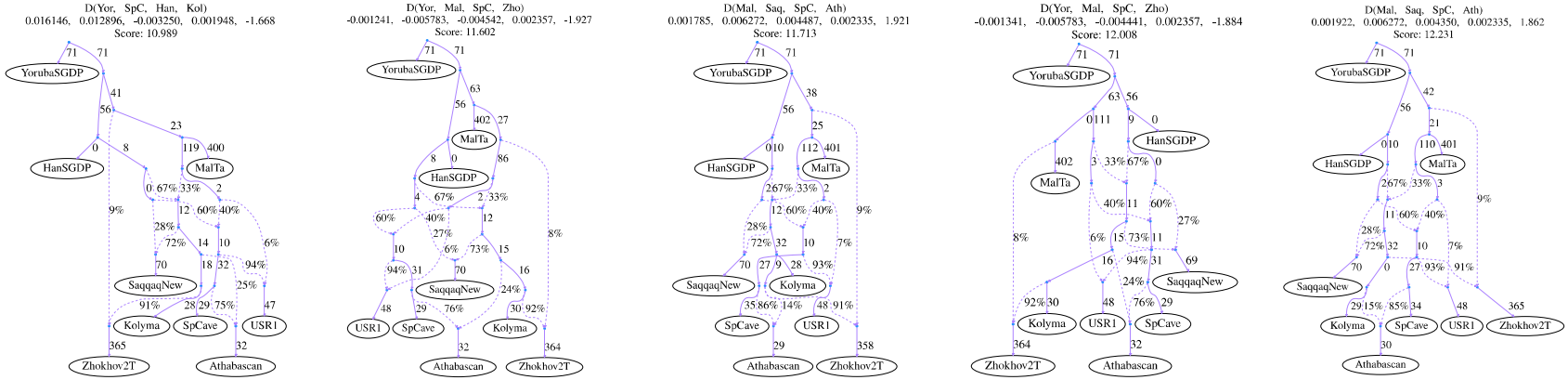
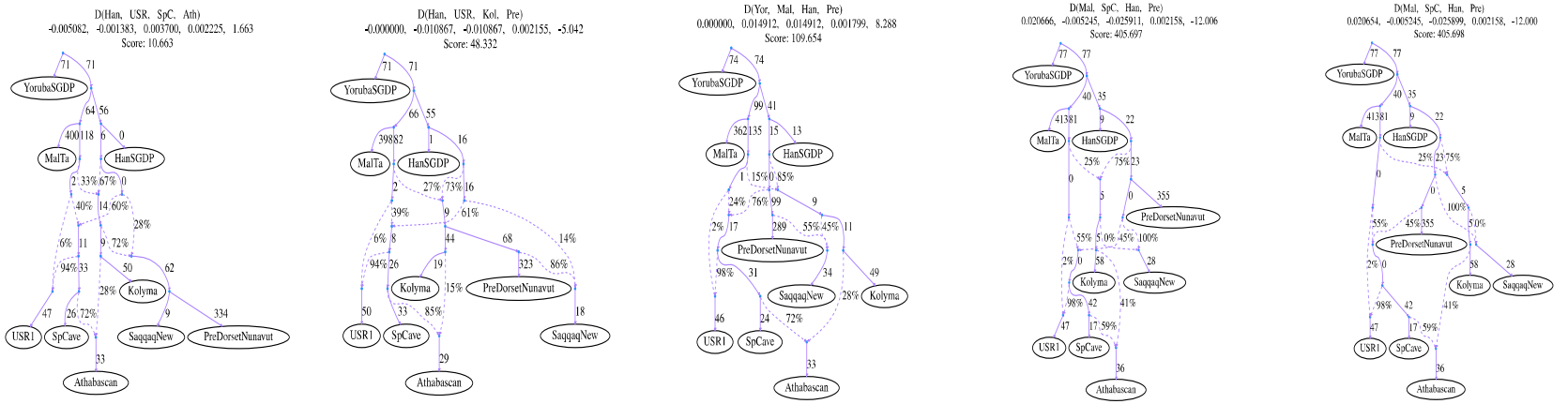



**Figure S5.1** – Enlarged version of **Figure 3.42,** the admixture graph results from the high-depth cultural representative of the Zhokhov.

# Pre-Dorset Nunavut Non-admixed and Admixed Models

No admixture




With admixture

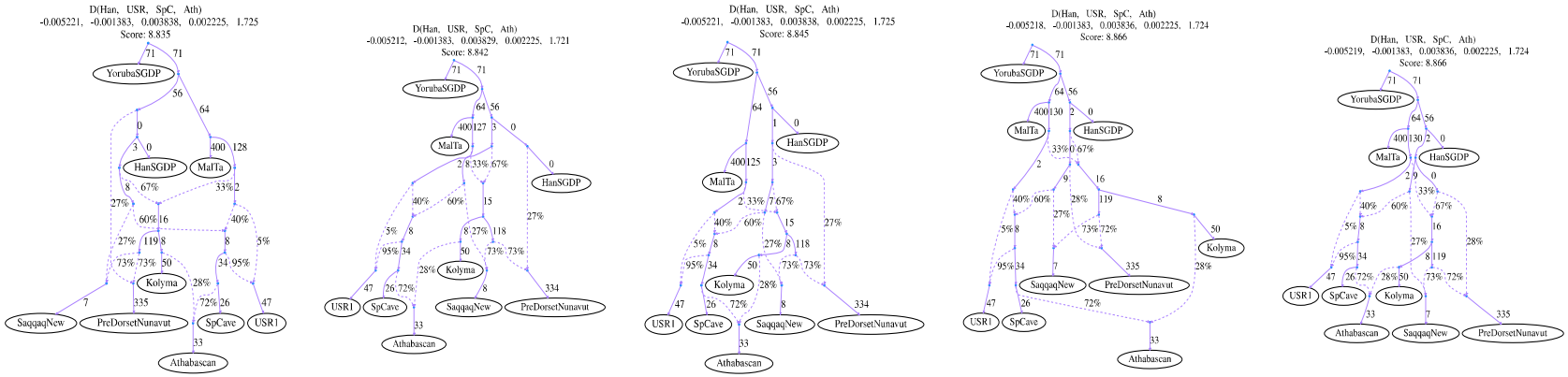
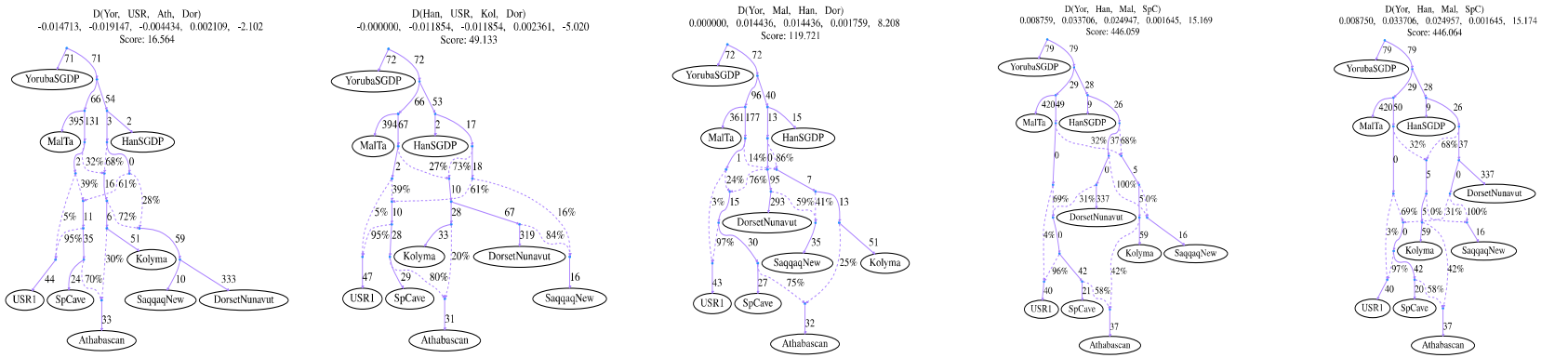



**Figure S5.2** - Enlarged version of **Figure 3.43**, the admixture graph results from the high-depth cultural representative of the Pre-Dorset.

## Dorset Nunavut Non-admixed and Admixed Models

No admixture




With admixture

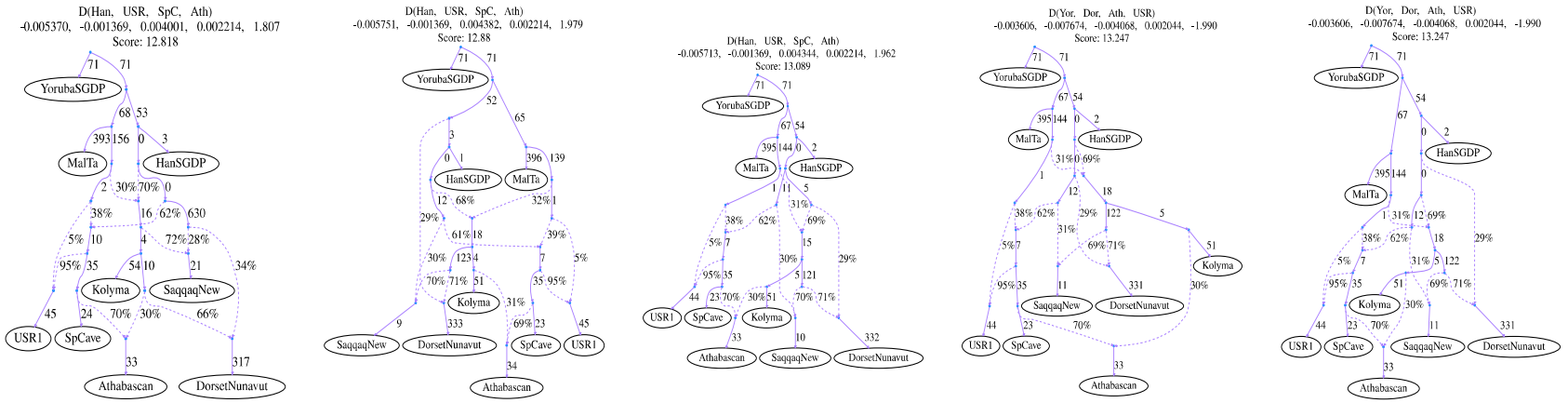
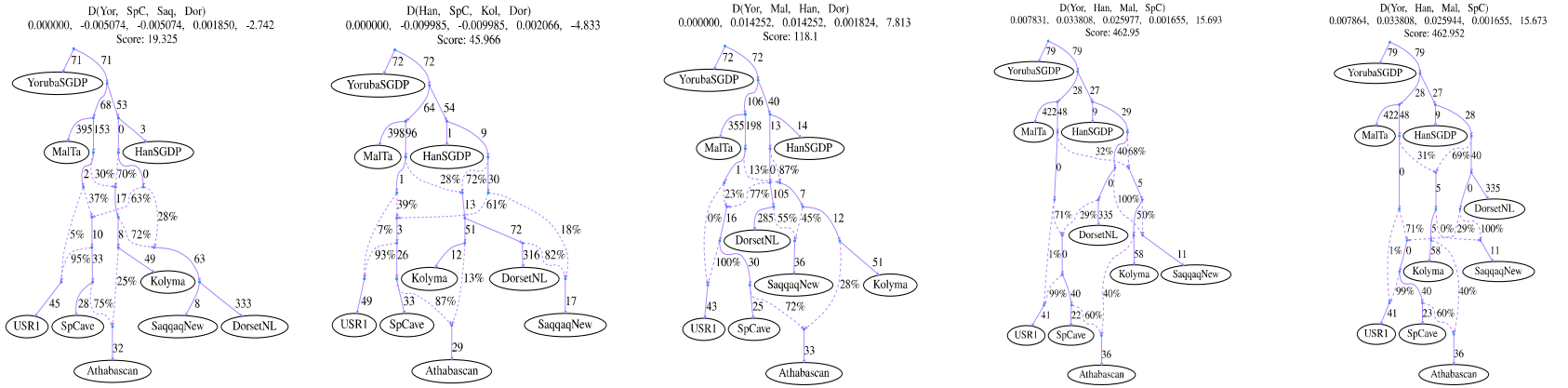



**Figure S5.3** - Enlarged version of **Figure 3.44**, the admixture graph results from the high-depth cultural representative of the Dorset from Nunavut.

# Dorset Newfoundland Non-admixed and Admixed Models

## No admixture




## With admixture

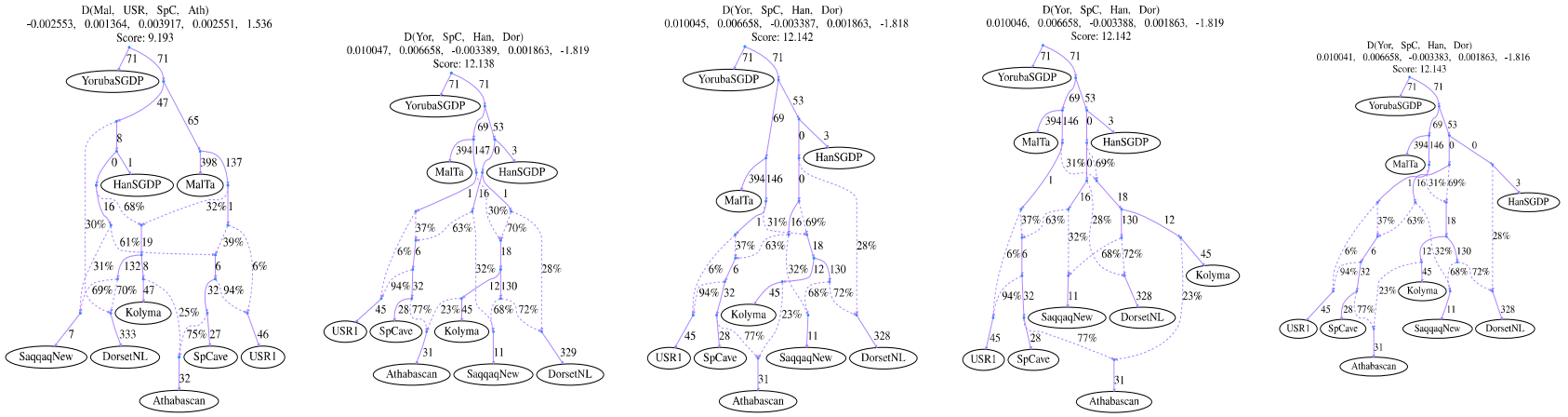
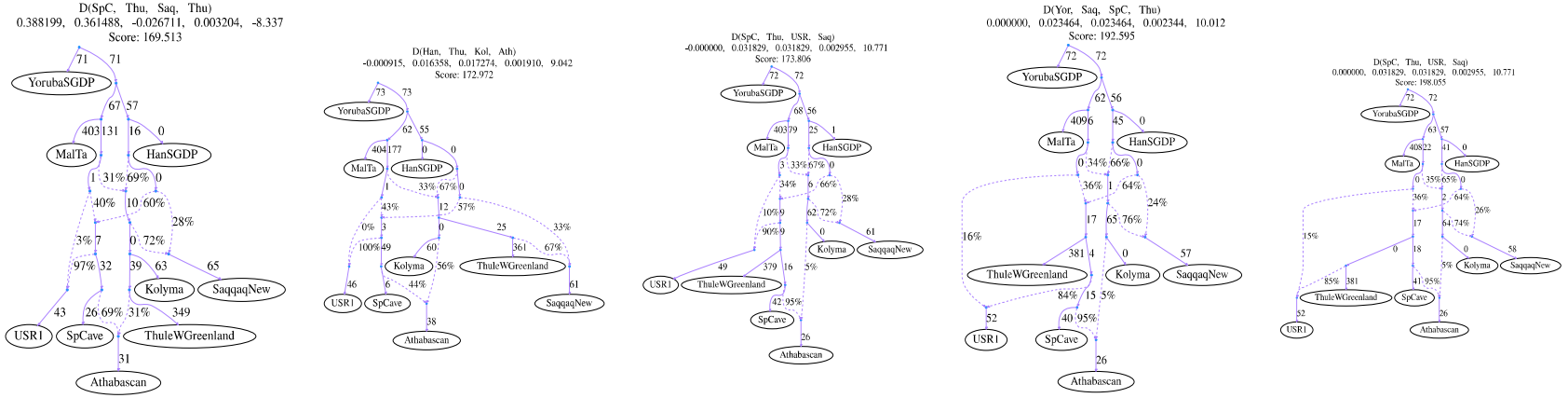



**Figure S5.4** - Enlarged version of **Figure 3.45**, the admixture graph results from the high-depth cultural representative of the Dorset from Newfoundland.

# Thule West Greenland Non-admixed and Admixed Models

## No admixture




## With admixture

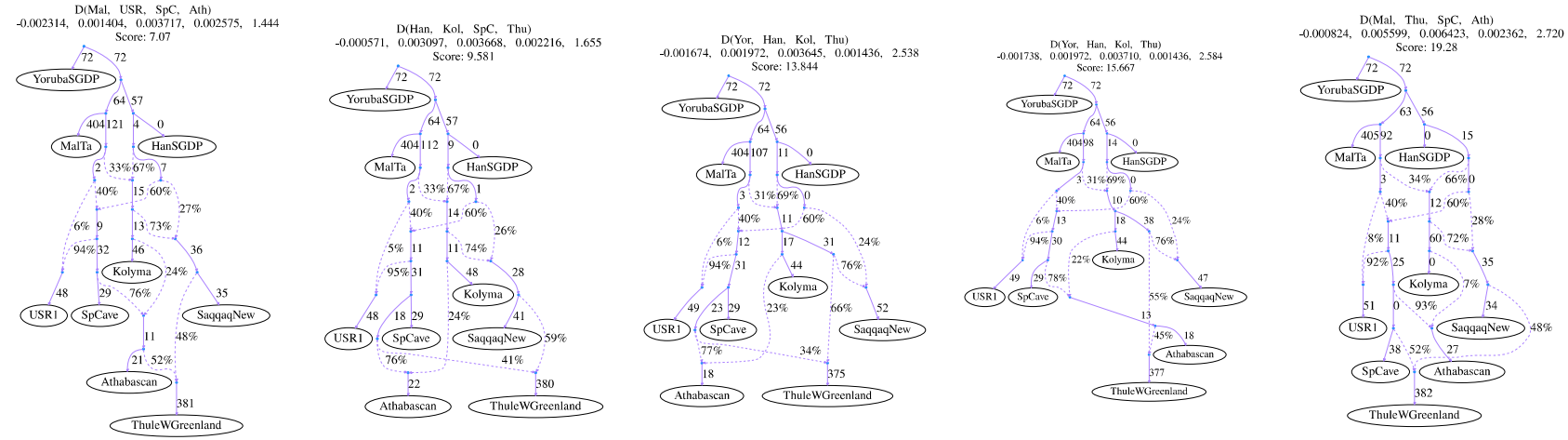
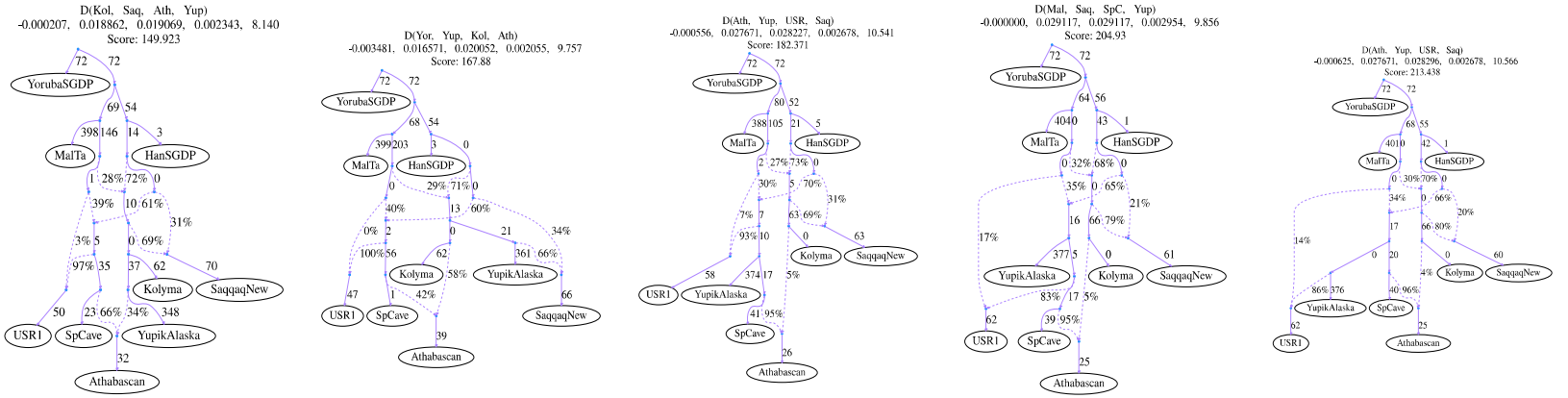



**Figure S5.5** - Enlarged version of **Figure 3.46**, the admixture graph results from the high-depth cultural representative of the Thule from west Greenland.

# Yupik Alaska Non-admixed and Admixed Models

## No admixture




## With admixture

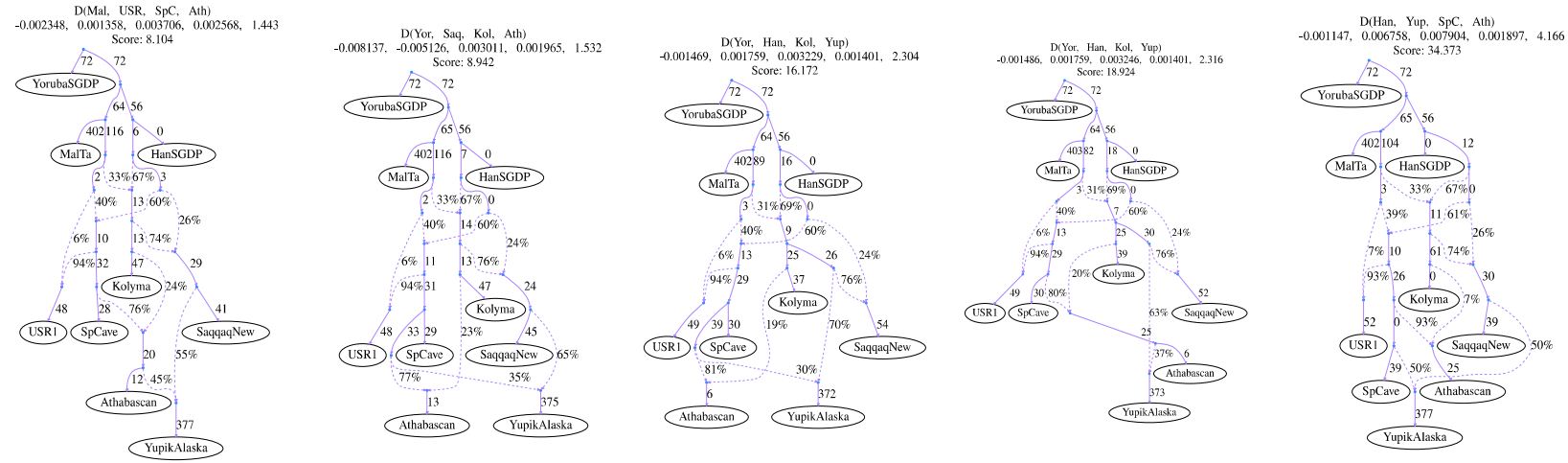



**Figure S5.6** - Enlarged version of **Figure 3.47**, the admixture graph results from the high-depth cultural representative of the Yupik from Alaska