Historical and Phylogeographic Influences on Mitochondrial DNA Diversity in Norwegians

Dana Kristjansson

Thesis for the degree of Philosophiae Doctor (PhD)
University of Bergen, Norway
2023

UNIVERSITY OF BERGEN

Historical and Phylogeographic Influences on Mitochondrial DNA Diversity in Norwegians

Dana Kristjansson

Thesis for the degree of Philosophiae Doctor (PhD)
at the University of Bergen

Date of defense: 13.01.2023

Year: 2023

Title: Historical and Phylogeographic Influences on Mitochondrial DNA Diversity in Norwegians

Name: Dana Kristjansson

Print: Skipnes Kommunikasjon / University of Bergen

Errata for
"Historical and Phylogeographic Influences on Mitochondrial DNA Diversity in Norwegians"

Dana Kristjansson

Thesis for the degree philosophiae doctor (PhD)

at the University of Bergen

9 Dec 2022 D. Kristjansson

(date and sign. of candidate)

12.12.22

(date and sign. of faculty)

Errata

Page 21 Missing word: fourth line from bottom, missing "were" after "values of 2%"

Page 41 Correction: In Figure 6, the Viking Age appears to begin at the beginning of the Iron Age. – correct to towards the end of the Iron Age.

Page 50 Correction: "and 14-10 cal kya" – corrected to "in 14-10 cal kya"

Contents

List of abbreviations

ADP or ATP: Adenosine Diphosphate (or Triphosphate for ATP)

AMOVA: Analysis of Molecular Variance

BIC: Bayesian Information Criterion

BCE: Before Common Era

bp: Base pair

BP: Before Present

BSP: Bayesian Skyline Plot

CE: Common Era

CeFH: Centre for Fertility and Health

DNA: Deoxyribonucleic acid

$FADH_2$: Flavin Adenine Dinucleotide (hydroquinone form)

hierBAPS: Hierarchical Bayesian Analytical Population Structure

HVR1/HVR2: Hypervariable region 1 or 2 of the mtDNA genome

mtDNA: Mitochondrial DNA

kya: Kilo years ago (kiloyears)

LGM: Last glacial maximum

MRCA: Most recent common ancestor

MDS: Multidimensional scaling

MCMC: Markov chain Monte Carlo

ML: Maximum Likelihood

NADH: Nicotinamide Adenine Dinucleotide Hydrogen

N_e: Effective population size

NIPH: Norwegian Institute of Public Health

NORBIS: Norwegian Research School in Bioinformatics, Biostatistics, and Systems Biology

np: Nucleotide position

OXPHOS: Oxidative phosphorylation

PCA: Principal Component Analysis

ROS: Reactive oxygen species

RSRS: Reconstructed Sapiens Reference Sequence

rRNA: Ribosomal RNA

tRNA: Transfer RNA

Scientific environment

The research work described in this PhD thesis was carried out at the Department of Genetics and Bioinformatics (until Jan 2022) and the Centre for Fertility and Health (CeFH), both of which are located at the Norwegian Institute of Public Health (NIPH) in Oslo, Norway. In addition, I held focused research meetings remotely with Prof. Theodore G. Schurr at the University of Pennsylvania (USA) on a weekly basis in the first year, and thereafter biweekly. In addition to participating in internal presentations and events at NIPH, I attended some meetings and gave a one-hour lecture at the Department of Global Public Health and Primary Care at the Faculty of Medicine, University of Bergen (Norway). Over the course of this PhD work, I was also a member of the Norwegian Research School in Bioinformatics, Biostatistics, and Systems Biology (NORBIS), and presented my research at two national annual meetings. In addition, I presented parts of this work at international research conferences including the American Society of Human Genetics and the European Society of Human Genetics annual meetings.

Supervisors:

Senior Scientist Astanand Jugessur

Department of Genetics and Bioinformatics (until Jan 2022) and the Center of Fertility and Health, NIPH, Oslo, Norway; Department of Global Public Health and Primary Care at the Faculty of Medicine, University of Bergen, Bergen, Norway

Professor Theodore G. Schurr

Department of Anthropology and Laboratory of Molecular Anthropology, University of Pennsylvania, Philadelphia, USA; Consulting Curator, Penn Museum of Archaeology and Anthropology

Senior Scientist Jon Bohlin

Center of Fertility and Health and Department of Method Development and Analytics, NIPH; Department of Veterinary Medicine, Norwegian University of Life Sciences, Oslo, Norway

Funding:

This research was funded by the Research Council of Norway, Grant/Award #:287961 to Astanand Jugessur and Theodore G. Schurr.

Acknowledgements

I would like to thank the Norwegian Institute of Public Health (NIPH), for support of this research, and the University of Bergen for the opportunity to complete this project as part of the PhD degree. I am also grateful to have received funding for this project from the Research Council of Norway, without which the results of this project would not have been possible. I thank the National Research School in Bioinformatics, Biostatistics, and Systems Biology (NORBIS) for offering several bioinformatics courses and the opportunity for skills development in data science.

I would like to express sincere gratitude to my PhD mentors Senior Scientist Astanand (Anil) Jugessur, Professor Theodore G. Schurr, and Senior Scientist Jon Bohlin for your unwavering support through thick and thin. This project would not have been possible without your steadfast beliefs in my scientific abilities. It is a great pleasure to work with a team of truly dedicated scientists and brilliant minds.

Furthermore, I am grateful for the privilege of working at CeFH under the leadership of Dr. Siri E. Håberg and Professor Per M. Magnus. I am also grateful to the Department of Genetics and Bioinformatics (HDGB) at the NIPH, Dr. Øyvind Helgeland who served as departmental head during my duration, as well as Kishan Kumar Chudasama and Ragnhild Valen. Thank you to my dear colleagues both at HDGB and CeFH for warmly welcoming me. Thank you to both departments for providing a friendly and open scientific research environment in which new ideas are encouraged and participation is welcomed. This PhD was all the more enjoyable thanks to your genuine interest in these findings and your intriguing questions during presentations.

I also would like to offer special thanks to Professor Håkon K. Gjessing, Dr. Jennifer Harris, Professor Rolv Terje Lie, Dr. Robert Lyle, Professor Vegard F. Skirbekk, Dr. Kåre Bævre, and Truc Trung Nguyen for showing an interest in my work and/or providing creative and useful feedback on my ideas and presentations.

Lastly, this project was carried through with the love and support of friends and family, both near and far.

Abstract in English

Mitochondrial DNA (mtDNA) has been an invaluable source of genetic information for understanding the long-term migration history, genetic diversity, and demographic structure of human populations. While well known for its Viking past, Norway's prehistoric and post-Viking population history and the influences that have shaped its mtDNA diversity are not fully understood. An understanding of the mitochondrial DNA lineages in Norway can further provide insight into medical conditions associated with the mtDNA variants present in those lineages. Thus, the overarching goal of this project was to analyze the mtDNA diversity of Norwegian populations, using phylogenetic (bioinformatic) methods, and examine the matrilineal ancestry of some lineages in relation to other Eurasian populations.

In this thesis, I explore the influences that may have led to the expansion of the most prevalent mtDNA lineages in Norwegians and estimate their arrival in the Scandinavian Peninsula in the context of known archaeological evidence and historical events. First, I describe mtDNA diversity in Norway in detail in terms of maternal lineages present in the country and their distribution in various counties and larger regions. Subsequently, I combine bioinformatic, phylogenetic approaches with time estimates for haplogroups J and U5 in Norway to elucidate their origins and later dispersal into Scandinavia. These are the second and third most frequent haplogroups in Scandinavia. Haplogroup H had already been extensively analysed and has a similar place of origin to haplogroup J, which has been less studied in Scandinavia.

Norwegian mtDNAs belonging to predominantly eight West Eurasian haplogroups were modestly differentiated by dialectal regions. Haplogroup U5 was found to be the earliest major haplogroup in Norway, with the earliest known carrier dating back to at least 9 kya. Haplogroup U5 is currently the second most frequent haplogroup in the country, with a higher frequency of this lineage occurring in the northern dialectal region of the country. My analysis revealed that the spread of subhaplogroup U5b was skewed from west-to-east while U5a was dispersed in both west-to-east and east-to-west directions. By contrast, haplogroup J entered

Scandinavia approximately six kya, with the youngest major subbranch, J1c, becoming the predominant J subhaplogroup in the region.

The detailed analysis of mtDNA variation conducted in this thesis will serve as the basis for further cross-disciplinary work linking public health, anthropology, and mitochondrial genetics.

Abstract in Norwegian

Mitokondrielt DNA (mtDNA) har vært en viktig kilde til genetisk informasjon for å forstå migrasjonshistorie, genetisk mangfold og demografisk struktur på lang sikt. Til tross for at Norge er kjent for sin Viking-fortid finnes det langt mindre kunnskap om befolkningen og mtDNA mangfold under andre tidsperioder. En økt forståelse for mitokondrielt DNA varianter i Norge kan bidra med mer innsikt om relevante mitokondrie relaterte sykdommer. Det overordnede formålet med dette prosjektet var derfor å analysere mtDNA diversiteten i den norske befolkning ved bruk av fylogenetiske (bioinfomatiske) metoder samt undersøke varianter av matrilineære aner og deres relasjon til andre Eurasiske befolkninger.

I denne avhandlingen undersøker jeg hvilke faktorer som kan ha forårsaket spredningen av nordmenns mest utbredte mtDNA varianter. Jeg estimerer også når de først kan ha dukket opp på den skandinaviske halvøy i samsvar med arkeologiske funn og historiske begivenheter. Jeg starter med å først beskrive maternelle mtDNA varianter som finnes i den norske befolkningen og deres fordeling i fylker og landsdeler. Videre belyser jeg deres opphav og senere spredning i Skandinavia ved å kombinere fylogenetiske og bioinformatiske metoder med tidsestimater for haplogruppene U5 og J. Disse er de henholdsvis nest- og tredje mest hyppige haplogruppene i Skandinavia. Omfattende analyser av den vanligste haplogruppen H finnes i rikt omfang samtidig som den har liknende opphav som halogruppe J som er lite studert i Skandinavia.

Norske mtDNA varianter tilhørende åtte vest-Eurasiske haplogrupper kunne delvis grupperes etter dialekter. Jeg fant at den nest mest hyppige haplogruppen U5 var den første store haplogruppen i Norge (ca. 9000 år siden) og at den er mest vanlig i Nord-Norge. Videre så jeg en viss grad av forskyvning av subhaplogruppe U5b fra vest til øst mens subhaplogruppe U5a var jevnt fordelt fra både øst til vest og vice versa. Haplogruppe J derimot ser ut til å ha først kommet til Skandinavia for omtrent 6000 år siden der den yngste subhaplogruppen J1c har blitt den mest vanlige.

Denne avhandlingens analyse av mtDNA variasjon i Norges befolkning vil forhåpentligvis kunne være et fundament for videre tverrfaglig samarbeid mellom folkehelse, antropologi og mitokondriell genetikk.

List of articles

Paper I

Kristjansson D, Bohlin J, Jugessur A, Schurr TG. (2021). “Matrilineal diversity and population history of Norwegians.” *American Journal of Physical Anthropology*. Sep;176(1):120-133.

Paper II

Kristjansson D, Bohlin J, Nguyen TT, Jugessur A, Schurr TG. (2022) “Evolution and dispersal of mitochondrial DNA haplogroup U5 in Northern Europe: Insights from an unsupervised learning approach to phylogeography.” *BMC Genomics.* Dec;23(1):1-25.

Paper III

Kristjansson D, Bohlin J, Schurr TG, Jugessur A. (2022) “Phylogeographic History of Mitochondrial Haplogroup J in Scandinavia.” *American Journal of Biological Anthropology* (under review)

1. Introduction

1.1 The mtDNA sequence

Mitochondria are cytoplasmic organelles in eukaryotic cells that generate adenosine triphosphate (ATP) for use in several cellular metabolic pathways (**Figure 1**). The origin of cellular mitochondria is thought to be linked to serial endosymbiotic events in which a free-living proto-mitochondrion was engulfed by a proto-eukaryote (1). It is hypothesized that this endosymbiosis was supported due to the exchange of metabolizable substrates and protection from the extracellular environment for the mitochondrion, and respiration-derived ATP from the mitochondrion for the cell. Molecular phylogenies based on mitochondrial genes have confirmed that mitochondria have a bacterial origin (2) and are distantly related to *Rickettsia prowazekii* (3,4).

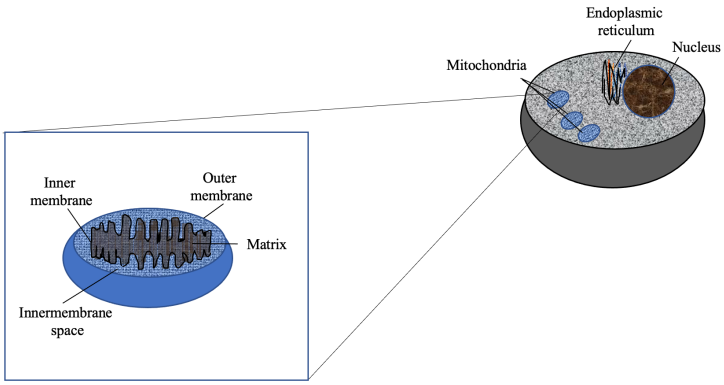

Figure 1. A simplified representation of a mitochondrion within the cell.

Located in the mitochondria, mitochondrial DNA (mtDNA) has also been found to be maternally inherited by both male and female offspring (5). The sequence and organization of human mtDNA was first described in 1981 by Anderson and colleagues (6,7). The mtDNA is a haploid and circular molecule consisting of 16,569 bp (**Figure 2**) (6). Human mtDNAs consist of one outer guanine-rich heavy (H)

strand and an inner cytosine-rich light strand (6). The heavy (H) strand encodes 12 proteins of the oxidative phosphorylation (OXPHOS) system, two ribosomal RNAs (12S and 16S), and 14 tRNAs (6,7). The light (L) strand encodes one subunit and 8 tRNAs (6,7).

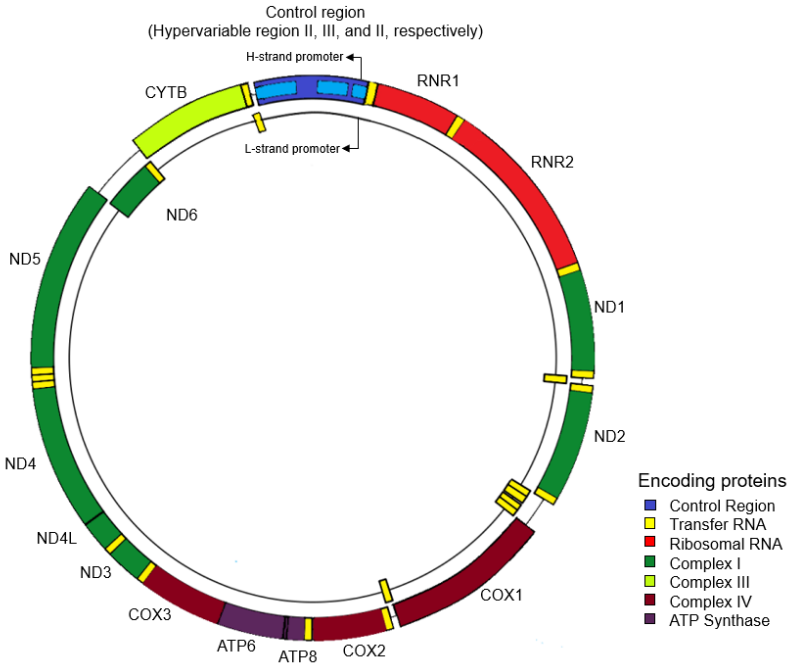

Figure 2. The organization of the human mitochondrial genome. The multicoloured, thicker circle represents the heavy (H) strand and the thin black line represents the light (L) strand. Shown here are the gene names in the circular mtDNA. The labels and gene names are based on the annotated human mtDNA sequences from van Oven *et al.* (2009) at Phylotree.org (8).

About 93% of the mitogenome consists of coding regions. They includes 22 transfer RNAs, 13 messenger RNAs, two ribosomal RNAs, cytochrome c oxidase subunits I, II and III, ATPase subunit 6, cytochrome b, and eight other genes that encode proteins involved in the OXPHOS process (6). The remaining proteins needed for mitochondrial functions are transcribed from genes located in the nucleus. They were transferred from the mitochondrial genome over evolutionary time, and are

translated in the cytosol before being transported across the mitochondrial membrane (9).

Approximately 7% of the mitogenome consists of the control region (labeled in dark- and light-blue in **Figure 2**), which initiates mtDNA replication and contains promotors for gene regulation (6). Three hypervariable regions (HVRs) are located within the control region, including HVR1 (np 16024-16383 bp; n = 400 bp), HVR2 (np 57-372 bp; n = 316 bp), and HVR3 (np 438-574; n = 137 bp) (7). The HVR sites are the most variable sites of those present in the mitogenome and are particularly useful for determining the haplogroup status of a particular mtDNA.

The majority of variants in mtDNA occur in the non-coding region (10). This is expected because the control region has a higher mutation rate than the coding region. In addition, the coding region is under greater functional constraint. An accumulation of mutations in the two first nucleotides of the codons in coding sequences would result in non-synonymous amino acid changes, thereby potentially threatening the survival and reproduction of a mtDNA lineage. However, mutations in the third codon nucleotide might accumulate as synonymous mutations without a change in the amino acid which it encodes (*aka* codon redundancy or degeneracy). Another reason that the control region has many variants is because it has a higher mutation rate than the coding region.

1.2 Mitochondrial function

Mitochondria generate approximately 90% of cellular ATP (11), an energy-carrying molecule that drives most cellular processes. The mitochondrion utilizes OXPHOS to combine the breakdown products of glucose with oxygen in the inner mitochondrial membrane through five protein Complexes (I through V) (12). NADH and $FADH_2$ produced from the Krebs cycle donate electrons to Complexes I and II, respectively (**Figure 3**). As the electrons pass through the electron transport chain (Complexes I through IV), they pump protons (H+) out of the mitochondrial inner

membrane into the intermembrane space (13). Complex IV receives an electron from cytochrome c, which donates an electron to molecular oxygen to form water (13). By the end of the electron transport chain, the accumulation of hydrogen ions in the inner membrane generates a mitochondrial membrane potential through the enzyme ATP synthase (13). This proton gradient enables a phosphate group to bind to adenosine diphosphate (ADP) to synthesize ATP (13).

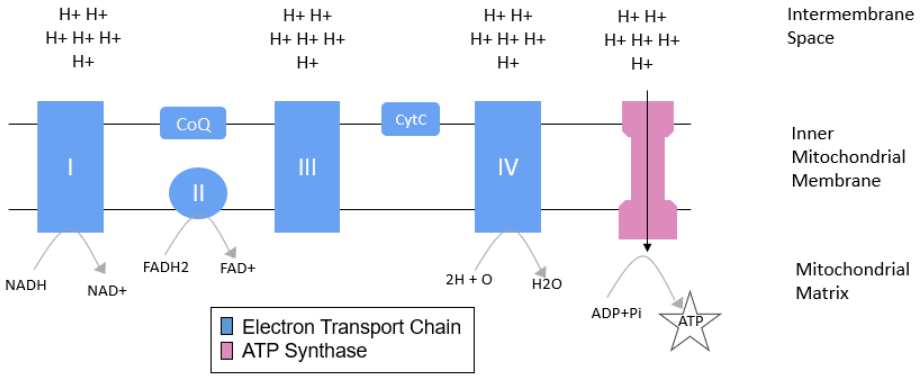

Figure 3. Representation of oxidative phosphorylation in the mitochondrion. CoQ: coenzyme Q; CytC: cytochrome c.

In addition to energy production, mitochondria also store calcium ions until they are needed for calcium signaling between the cytosol and the mitochondrial matrix. Cellular calcium ion signaling is an essential cellular process that regulates the activity of transporters, enzymes and proteins involved in muscle contraction, motility, synaptic transmission, membrane excitability, apoptosis, and gene transcription (14–17).

Mitochondria are further involved in several different cellular processes that may also influence overall cellular health. While many mtDNA mutations demarcate the evolutionary history of human matrilines, some may cause disease through the impairment of the OXPHOS system. Certain of these mitochondrial diseases are systemic, affecting vital tissues and organs that require a high amount of energy, such as the brain, muscles, and heart (18). Other mtDNA mutations may also be involved

in chronic diseases such as cancer, with several studies showing both germline and somatic mtDNA mutations in solid tumors and leukemias (19–25). However, the direct association of mtDNA with cancer is still under investigation.

At least one-third of the oxidation energy generated through OXPHOS is released as heat through a process called ‘proton leak’ instead of ATP synthesis (26). The large amount of heat for thermoregulation is mediated by mitochondrial uncoupling proteins which also control the production of reactive oxygen species (ROS) and fuel metabolism in adipose cells (27). In about 0.1–2% of electrons passing through the chain[1], oxygen is prematurely and incompletely reduced to the superoxide radical (·O2) that is the precursor of most ROS (28). These ROS can lead to progressive damage to mtDNA and the protein membrane structure of the mitochondrion (29–31).

1.3 Features of the mtDNA

The mtDNA has been an invaluable source of genetic information for understanding the long-term migration history of human populations. Key features of mtDNA that make it especially suitable for tracking lineages at the population level and over many generations are its clonal inheritance pattern, high reliable mutation rate, and low heteroplasmy.

1.3.1 Clonal (maternal) inheritance

With the exception of a few noted cases (32,33), mtDNA are inherited through the female germline (34). While the number of mtDNAs in a mature oocyte may exceed 100,000 (35), a sperm cell contains only about 700 to 1200 copies of mtDNA in the sperm flagellum (36). The paternal mitochondria, which normally do not penetrate the zona pellucida during fertilization, can still be eliminated by selective destruction, inactivation, or dilution by the vast surplus of oocyte mitochondria

[1] This number derives from studies in isolated mitochondria, though the exact rate in live organisms is yet to be fully agreed upon (28)

(37,38). Even if leakage of paternal mtDNA is rare, random replication probably reduces this contribution to zero in most individuals (38).

Since mtDNA are clonally inherited, the whole genome behaves as a single, non-recombining locus with all sites sharing a common genealogy (39). Whether or not mtDNA undergoes recombination has been disputed for several years (40–43). Although there has been some tentative evidence of exceptions for mtDNA recombination (44–46), later analyses have either found no evidence of recombination or that recombination is not sufficiently frequent to assume recombination in the vast majority of sequences (43,47–49). Moreover, after a reanalysis of evidence for recombination in two previously published studies (45,46) two separate articles cited possible errors in the data interpretation or errors in the sequences upon which the original interpretations were made (50,51).

The fact that mtDNA is clonally inherited from mother to all offspring considerably simplifies its representation within a population. This allows researchers to trace the maternal genetic connections of individuals and track population movements through time by following the dissemination or accumulation of specific maternal lineages, in addition to exploring the relatedness of populations from far-flung regions. Importantly, the more limited number of possible mtDNA lineages compared to longer, recombined nuclear DNA also simplifies the analysis of variation within populations.

1.3.2 Reliable mutation rate

Due to an absence of histones, the increased exposure to ROS by-products of OXPHOS and a lack of repair system, the overall mutation rate of mtDNA is greater than that of nuclear DNA (52–54). Humans have a genome-wide mutation rate ranging from 0.5×10^{-9} bp^{-1} $year^{-1}$ compared to a value of 2.0×10^{-7} bp^{-1} $year^{-1}$ in HVRI and HVRII of the mtDNA (55). However, the incidence of mitochondrial disease is low and estimated to be approximately 1 out of 5000 live births (56,57). This high mutation rate is also inconsistent with the observations of purifying selection (58), adaptive changes (59), codon bias in mitochondrial genes (60), and the low rate of mitochondrial mutations between generations that have been studied (61,62). Possible explanations for the low rate of mutations transmitted between

generations are that the female germline facilitates selection against pathogenic mitochondrial mutations through a number of possible mechanisms, although these mechanisms are unclear and continue to be debated (63).

The types of evolutionary change that animal mtDNA undergoes are relatively simple, being mainly base substitutions and length mutations (34). The mtDNA mutation rate across the mitochondrial genome varies. Synonymous sites and the mtDNA HVRs evolve faster than nonsynonymous sites, the central domain between HVRI and HVRII, tRNA, and rRNA genes (64). In 2009, Soares *et al.* (10) estimated the substitution rate for the mitogenome to be approximately 1.67×10^{-8} bp^{-1} $year^{-1}$. The lowest substitution rate was 6.91×10^{-9} for tRNA and the highest was 2.29×10^{-7} for HVR-II (10). These rates can be utilized to define mutation models in phylogenetic analyses and/or coalescent age estimates of maternal lineages (see Methods).

1.3.3 Low heteroplasmy

A cell may have several to many thousands of copies of an mtDNA. The number of mtDNA depends on the cell type and the energy demands of the specific tissue type (65). For example, a mature oocyte contains 100,000 mtDNA (35), cardiac muscle cells contain approximately 6,970 mtDNA (66), and erythrocytes contain no detectable mtDNAs (67). Erythrocytes do not require mitochondria-generated energy in order to shuttle oxygen to the tissues. In the reticulocyte stage, the precursor to the erythrocyte, the mitochondria undergoes autophagy in order to make room for oxygen-carrying hemoglobin (68).

Although various cell types of the same individual may share only one type of mtDNA, a condition known as homoplasmy, a mutation may arise in an individual which can lead to more than one mtDNA type being in a single cell or individual, a condition known as heteroplasmy. It is now generally believed that all individuals present some level of heteroplasmy, even though this level might be below the limits of detection by DNA sequence analysis (69,70). Heteroplasmy can be observed at different levels, and multiple mtDNA types may occur within a single mitochondrion, a single cell or between different cell types (71–73).

The level of heteroplasmy appears to be correlated with the extent of organ involvement, with post-mitotic tissues that no longer rapidly divide, such as skeletal muscle, favouring the accumulation of mtDNA mutations (73,74). By contrast, the regenerative ability of tissues like hemocytes or bone marrow may eliminate mutated mtDNA molecules (73). Very low-level variance in heteroplasmy (0.2–2% heteroplasmy) was found in peripheral blood cells and skeletal tissue of healthy individuals, but values of 2% only found in skeletal tissue (75). This is an important number because peripheral blood and bone marrow are the tissues from which mtDNAs most typically used in phylogenetic and population analyses are derived.

1.4 Applications of mtDNA analysis

Due to its unique pattern of inheritance, mtDNA is used for a wide range of scientific studies where tracing the deep ancestral roots of maternal lineages plays a key role in understanding population movement, health, and evolution. These fields include evolutionary biology (76), molecular anthropology (77), forensics (78), and medical studies that query mtDNA sequences for specific diseases or clinical outcomes (79). In several of these fields of study, mtDNA variants are typically discussed in terms of a specific group of alleles (haplotype) that are inherited together in a direct maternal lineage (80).

Individual haplotypes can be grouped into haplogroups based on their mutational characteristics (**Figure 4)**. A haplogroup represents a group of similar haplotypes that share a combination of ancestral polymorphisms inherited together, such as U5. Haplogroups were originally defined in the late 1980s and 1990s by grouping mtDNAs exhibiting similar patterns of restriction fragment length polymorphisms (RFLPs) in studies of mtDNA types from diverse populations (81,82). The first haplogroups were defined in aboriginal Siberian and Native American populations, and were labelled as A, B, C, and D (82,83). The list of haplogroups has since expanded to include all letters of the alphabet based on the time of their discovery and the populations in which they were defined, and without regard to the order of their actual evolutionary relationships.

The mtDNA haplogroups defined by restriction fragment length polymorphisms were then first primarily correlated with HVR polymorphisms (84–86). This led to the establishment of bioinformatic databases that assigned haplogroups solely based on control region sequences (87–89). Later, haplogroups and their specific variants from their entire mitogenome sequence were organized in a comprehensive phylogenetic tree called Phylotree (8).

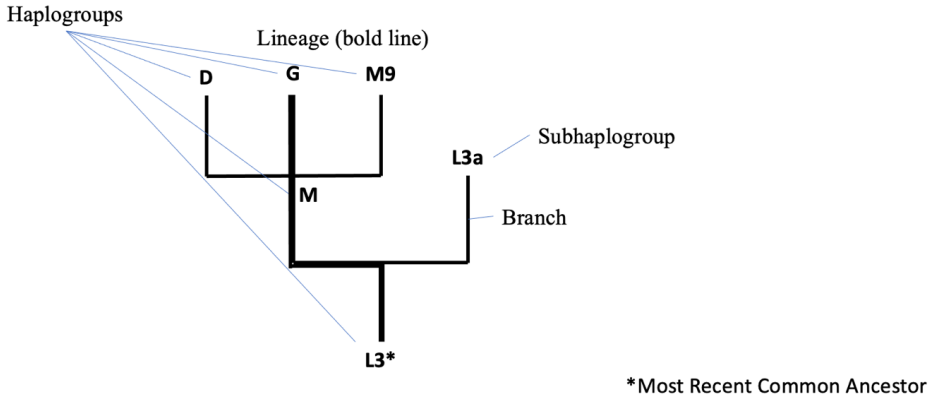

Figure 4. An example of a phylogenetic tree based on mtDNA sequences. The haplogroup nomenclature is based on the annotated human mtDNA sequences appearing in Phylotree (8).

This thesis utilizes specific definitions when discussing the details of mtDNA phylogenies. A *haplogroup* defines a major mtDNA lineage that has deep origins in human populations and may encompass a range of subbranches (subhaplogroups). A *subhaplogroup* is a branch of a haplogroup containing a subset of the sequences defined by the parent haplogroup but defined by its own set of mutation. For example, haplogroup U gives rise to multiple subbranches such as U5, which then split into U5a and U5b, with the latter subhaplogroup subsequently evolving to produce much smaller sublineages such as U5b1b1b. On a more general level, a *lineage* is a maternal line of descent often referred to in population studies, and a *branch* is a part of the phylogenetic tree that extends from a root or major trunk.

1.5 Human phylogeny

Phylogenetics is defined as the study of evolutionary relationships between or among species evolving from a common ancestral root, in which the splitting of different lineages and estimations of their rates of evolution can be determined (90). From a mitochondrial genetics perspective, the matrilineal most recent common ancestor (MRCA) of humans, the so-called "Mitochondrial Eve," is postulated to have lived in Africa about 100,000-300,000 years ago (91–93). Since the human dispersal within and outside of eastern Africa, numerous new mtDNA lineages have arisen and expanded over time in human populations. Because of the uniparental inheritance, lack of recombination, and reliable mutation rate, mutations arise in mtDNAs in a cumulative manner, making it is possible to track the evolution and dispersal of these maternal lineages over long periods of time.

Human mtDNA haplogroups can be broadly divided into African, West Eurasian, and East Eurasian types. This division is based on the origin of modern humans in Africa, and the conclusion that the initial split between western and eastern Eurasian haplogroups occurred somewhere between the modern-day Iran and the Indus Valley (94). Most major haplogroups (**Table 1**) arose approximately 136.3 to 20.9 kya and have defined different human populations (95). Haplogroup L mtDNAs arose in sub-Saharan Africa (~136.3–67.3 kya). Two haplogroups, M and N, arose from African haplogroup L3 approximately 58.9 to 49.6 kya as humans migrated out of eastern Africa (95). Haplogroup M lineages expanded into East Eurasia, eventually giving rise to haplogroups C, D, E, G, Q, and Z. Haplogroup N gave rise A, S, N, O, and Y, which stayed in East Eurasia (~53.5–24.2 kya), while later haplogroup N branches W, X, and I emerged in West Eurasia (~31.7–21.9 kya) (8,95). Haplogroup N also gave rise to haplogroup R, which is the root of the numerous West European haplogroups including HV, JT, U, and K (96). In East Asia, haplogroup R gave rise to haplogroups B, F, P, and various R sublineages in populations within this region (97,98).

Table 1. Major haplogroups and their approximate ages by regional origin.

Regional origin	Haplogroups	Approximate age[a] (kya)
Africa	L (L0, L1, L2, L4, L5, L6)	136.3 ± 11.8
	L3	67.3 ± 4.4
East Eurasia	A	24.2 ± 4.9
	B	49.5 ± 6.6
	C	23.9 ± 4.8
	D	38.4 ± 4.7
	E	23.7 ± 6.9
	F	42.9 ± 5.6
	G	31.6 ± 5.2
	M	49.6 ± 1.8
	N	58.9 ± 2.4
	O	52.1 ± 6.4
	P	54.8 ± 2.5
	Q	37.5 ± 5.6
	S	53.5 ± 5.5
	Y	24.6 ± 7.1
West Eurasia	HV	21.9 ± 2.8
	H	12.8 ± 0.8
	V	9.7 ± 1.4
	I	20.9 ± 3.6
	JT	46.9 ± 6.5
	R	56.5 ± 2.1
	U	46.5 ± 3.3
	K	26.7 ± 4.3
	W	17.6 ± 3.4
	X	31.7 ± 11.7
	Z	21.7 ± 8.4

a. Approximate ages of major haplogroups can be found in Behar *et al.* (95).

Since the early 2000s, numerous studies have characterized mtDNA diversity in global populations, leading to the identification of new subhaplogroups in the human mtDNA phylogeny. Haplogroup H, which comprises a large majority of the mtDNAs in Europe, includes numerous subhaplogroups that are found at a higher frequency in some populations (99). For example, subhaplogroup H2 is found with highest frequency in Eastern Europe (99), whereas H1 is found more frequently in Western Europe and Northern Africa (100). Thus, the use of phylogenetic methods can aid in tracking human migrations and characterizing mtDNA diversity in populations.

1.5.1 Methodological considerations for phylogeography

Identifying haplogroups and their subhaplogroups has become a cornerstone of phylogenetic studies of human mtDNA diversity and an important building block of biogeographic studies. Phylogenetic trees can aid in the visualization of existing genetic variation in human populations. In particular, phylogenetic analysis can help to elucidate the relationship between two or more populations by investigating their shared mtDNA lineages and exploring the differential effects of mutation, genetic drift, and gene flow on their genetic diversity over time. The ability to construct an accurate human phylogenetic tree is therefore important for elucidating migration patterns that can be evaluated in the context of archaeological, sociocultural, and linguistic evidence.

The use of median-joining networks was proposed as a method for the phylogenetic visualization of human mtDNA sequences based on point mutation changes (101). Median-joining networks have the advantage of computational speed and simplicity in the interpretation of genetic distances between sequences, although they, too, are not without limitations. The major shortcomings of median-joining networks include their reliance on distance-based *phenetics* [2] and their lack of a root from which to make historical inferences (102). This challenge is compounded when handling several sequences with small genetic distances between individuals (102).

By contrast, a maximum likelihood (ML)-based phylogeny can be rooted at a predetermined ancestral sequence. It can also take into account character transformations using different evolutionary models that can be validated using bootstrapping methods or bootstrap approximations (102–104). While ML is often employed to investigate the evolutionary relationship of non-human species, its use in human mtDNA analyses has been limited due to the tediousness of assigning and labelling each mitogenome sequence to a haplogroup defined in Phylotree. In addition, the similarity of the sequences in large human populations typically studied in these analyses can often result in unintelligible, dense, and complex trees. Consequently, the genetic relationships between groups of similar mitochondrial

[2] Overall similarity, without consideration of character transformation.

sequences become difficult to disentangle and use for making broader evolutionary inferences.

Two of the scientific articles in this thesis (**Paper II** and **Paper III**) attempted to create a ML phylogenetic tree with an inferred root. ML was used to cluster individual mtDNA sequences into haplogroups, some of which are potentially geographically specific. The approach of identifying large groups of similar haplotypes and tagging the geographic location of each sequence enables an investigation of the historical connection between these geographic regions and how their interactions may have contributed to the eventual distribution of maternal lineages within Norway.

1.6 A brief history of Norway

Since this thesis focuses on understanding the deep ancestry of mtDNA lineages in Norway, I will provide a backdrop for understanding some of the notable events in Norwegian topography and history that pertain to the population history of Norwegians.

Situated in the northwest of Europe, Norway is unique in its vast length and distance from the geographic starting point of human migrations from eastern Africa. Norway is also the longest country in Scandinavia, stretching 1,752 km in length, a distance equivalent to the length of the Netherlands to the center of Italy (105). Compared to southern Europe, Norway's climate had been relatively uninhabitable until the large Fennoscandian ice sheet shrouding the land began to melt during the last glacial maximum (LGM) approximately 23 kya (106,107). Topographically, Norway's fjords divide large portions of the coastline areas and mountainous terrain, and potentially limited gene flow between subpopulations living there. The highest mountain ranges begin at the latitude of Stavanger in the southwest and extend up to Trondheim in central Norway. This range includes the Jotunheimen Mountains, home to Norway's highest mountain called Galdhøpiggen at 2,469 m (108), which divides

the southern and most populated portion of the country into eastern and western halves.

Human migrations into Norway have influenced the genetic make-up of the Norwegian population over thousands of years. The earliest evidence of human settlement, which dates to 10,000 years ago, stems from the hunting tools and burial sites of hunter-gatherers found from the southeastern Oslo Fjord up to Finnmark in the far north (109,110). These early settlers exhibited a predominance of mtDNA haplogroup U5 (109). Neolithic farmers arrived in Scandinavia approximately six kya (111). For millennia, genetically distinct Neolithic farmers and local Scandinavian hunter-gatherers existed largely separately (112), although evidence for genetic admixture suggested that Scandinavian hunter-gatherers had already begun to be assimilated into agricultural communities approximately five kya (113).

After agriculture became the predominant subsistence strategy in Scandinavia (114), competition for resources coupled with increased social demands for personal wealth played a key role in motivating expansions during the Viking Age (793–1066 CE) (115,116). In addition to population expansion from Viking travels, the population in Norway was subjected to demographic and potentially adaptive pressures such as the estimated loss of 40–50% of the population during the Black Death (after 1349 CE in Scandinavia) (117,118). In addition to these historical events, contacts with the local Saami populations as well as traveling and trading populations likely contributed to present-day Norwegian mtDNA diversity (5).

Early modern inhabitants were influenced by few internal migrations before the 1750s (119,120). The regional distribution of Norwegians is reflected in the distinct spoken dialects in particular locations, which may also be correlated to differences in the distribution of maternal genetic lineages there. These dialects vary widely in grammar, syntax, tone, and pronunciation (121), and reflect the partitioning of genetic variation across vast regions within Norway over time. The major dialects include Eastern Norwegian (*østnorsk*), Western Norwegian (*vestnorsk*), Trøndelag in central Norway (*trøndersk*), and Northern Norwegian (*nordnorsk*) (121,122).

The work included in this thesis is the result of a multifaceted investigation into maternal genetic lineages that are present in Norway. **Paper I** presents a large-scale analysis of the maternal genetic structure of the Norwegian population within specific regions of the country. **Paper II** and **Paper III** present the results of in-depth explorations of two major haplogroups found within Norway, the timeline of their origins, and the manner in which they have arrived in Scandinavia within a global context. Since other populations also share these same mtDNA haplogroups, studying their phylogenies will also contribute to a better understanding of genetic diversity in populations outside Norway.

2. Aims of the thesis

The overall aim of this thesis is to use bioinformatic tools to explore the diversity of mtDNA haplogroups in Norwegian populations, particularly with respect to other Eurasian populations, and to better understand the forces that shape mtDNA diversity in Norway. First, I describe mtDNA diversity in Norway in detail and assess its distribution in various counties and larger regions within the country. I then combine bioinformatic approaches with phylogenetic tools and geographical data to study two of the most common haplogroups found in Norway, including their estimated ages and the events that led to their dispersal in Scandinavia. Since haplogroup H had already been studied extensively by others (99,123–126), this thesis focuses on the second and third most frequent haplogroups in Scandinavia, namely, U5 and J. Haplogroup U5 is likely to have been the earliest maternal lineage to have arrived in Norway, while haplogroup J arrived later with agricultural populations. The thesis combines the same techniques used in the field of molecular anthropology (i.e., mtDNA estimates of ancestry) with methods employed in public health training at NIPH for the analysis of these sequences at a population level.

These studies have the following objectives:

Paper I: To describe regional mtDNA diversity in Norway and to understand the influence of dialectal regions and internal historical migrations on the mtDNA haplogroup distribution found in the data.

Paper II: To combine hierBAPS analysis of haplogroup U5 mitogenome sequences with maximum likelihood (ML) phylogenetics to make inferences about the evolution and dispersal of this major maternal lineage in Scandinavia.

Paper III: To investigate haplogroup J variation, considering the genetic and archeological evidence, and provide further insight into the demographic and evolutionary history of haplogroup J frequency within Scandinavia.

3. Methods

3.1 mtDNA data

The human mtDNA sequence data used in this study were collected from various public sources, including previously published literature, GenBank, the European Nucleotide Archive, and various projects that are part of FamilyTreeDNA.[3] **Table 2** lists each of the studies from which the data were acquired, the region of the mtDNA genome sequenced, the sample sizes for each study population, and the characteristics of the populations being studied.

Table 2. Published mtDNA sequences from Modern Norwegians used in this study

	Sample Sources	mtDNA Region Analyzed	N	Population Characteristics
Paper I	• Krzewińska M *et al.,* 2014 • Helgason A *et al.,* 2001 • Opdal *et al.,* 1998 • Passarino *et al.,* 2002 • FamilyTreeDNA via GenBank	16024–16383	1,174	Norwegians
	• The Norway DNA Project	16024–16383	1,597	Norwegians with Norwegian ancestors
Paper II	• GenBank • European Nucleotide Archive	1–16569	873	Haplogroup U5 sequences worldwide
Paper III	• GenBank	1–16569	2,345	Haplogroup J sequences worldwide
	• European Nucleotide Archive	1–16569	55	Viking Age graves

[3] https://www.familytreedna.com/

For **Paper I**, I analysed a total of 1,174 partial mtDNA sequences (HVRI) from modern-day Norwegians obtained from various published sources on Norwegians genetic diversity (127–130). This study also included mtDNA sequences from the GenBank database (131) belonging to persons of Norwegian ethnicity and origin as of June 1, 2020. Data for the location or residence within Norway were available for 64% (n = 755) of these individuals.

In addition, HVRI sequences from 1,597 non-related individuals with matrilineal Norwegian ancestors were incorporated into this analysis to contextualize the human migration changes within Norway since the early modern ancestors of the 17th to 20th centuries. The mtDNA sequences were retrieved on June 1, 2020, from The Norway DNA Project, a subproject of FamilyTreeDNA containing mtDNA sequences submitted by consenting members for public display (132). This project is open to individuals with a Norwegian background, Norwegian ancestry, or residency in Norway. Only data from individuals with traced Norwegian maternal ancestry, but not necessarily residing in Norway, were used for this project.

Location information was confirmed using The National Archives of Norway from individuals who had submitted their mtDNA data and who had also indicated the local region of their earliest known matrilineal ancestor (87%; n = 1,396) (133). These archives contain annotated information derived from parish records, census records, and village books on Norwegians living from the 17th to the early 20th century. For these reasons, the Norway DNA Project data were used as a proxy for ancestral Norwegian populations which are henceforth called ‘Ancestors’.

For **Paper II**, haplogroup U5 mitogenome sequences were retrieved from the European Nucleotide Archive and GenBank (n = 873) (accessed on 31 May 2021) using a search for “whole mtDNA” and “haplogroup U5.” For **Papers II and III**, Nordic populations were separated into Saami, Scandinavia (Norway, Denmark, and Sweden), and Finland. Finland was kept separate from Scandinavia due to its geographic isolation from the Scandinavian Peninsula and its linguistic distinctiveness from the other Nordic languages. Specific information about the

ethnicity or original location of the individuals represented by these sequences was available for 855 (97.8%) of the total dataset.

Haplogroup J mitogenome sequences (**Paper III**) were retrieved from GenBank (n = 2,245) (accessed on 15 December 2021) using accession numbers found in MTree (134). Since the geographic focus of this study was Scandinavia, we also included haplogroup J data from Viking Age graves (hereafter called the 'Viking Burial Sites') available in the European Nucleotide Archive (n = 55). Specific information about the ethnicity or original location of the individuals was available for 2094 (91.0%) sequences in the total dataset. A total of 2,153 of the sequences (93.7%) derived from modern-day populations, while 147 sequences (6.4%) came from past populations.

3.2 Haplogroup identification

We used the Haplogrep software, version 2.1.21 (135), for **Papers I–III** to assign a haplogroup to each mitogenome sequence based on its mutational signature. Haplogrep estimates haplogroup classifications based on pre-calculated phylogenetic weights that correspond to the occurrence of a polymorphism per position in Phylotree Build 17 (8). These weights reflect the mutational stability of a variant. Mutations were identified relative to the Reconstructed Sapiens Reference Sequence (RSRS) (95), which allows for the naming and mapping of human mtDNA haplogroups from a hypothetical ancestral root.

3.3 Phylogenetic analysis

3.3.1 Maximum-likelihood phylogenic tree

Mitogenome sequences were aligned using the MAFFT software, v.7 (136). ML phylogenies for mitogenome sequences for **Papers I–II** were conducted using the software IQ-tree, v.1.6.12 (103). The phylogeny was reconstructed using the best fitting nucleotide substitution model inferred by jModelTest (137,138) based on the Bayesian Information Criterion (BIC). The approximate likelihood ratio test (aLRT)

was used to achieve branch support (139) based on resampling the estimated log-likelihood method with an effective collection scheme of candidate trees (137). I also conducted 10,000 ultrafast bootstrap replicates using the UFBoot algorithm (104). UFBoot overcomes the computational burden required by the standard nonparametric bootstrap, and can be interpreted as providing an unbiased bootstrap support with 95% support, which corresponds to a 95% probability that a clade is correctly assigned (140).

3.3.2 Partitioning mtDNA sequences using hierBAPS

A number of the U5 sequences for **Paper II** belonged to the same subhaplogroups as a result of sharing a common ancestral lineage, although their specific haplotypes may have differed. The hierBAPS algorithm was used to identify clusters of closely linked sequences (141). This algorithm groups DNA sequences into clusters in a hierarchical manner (142). The hierBAPS algorithm assumes that each individual sequence is drawn from one of several distinct genetic subpopulations, with each cluster having its own set of allele frequencies. For **Paper II**, the hierBAPS clusters were projected onto an independently derived ML-based phylogenetic tree to compare how accurately the hierBAPS algorithm clustered the individual mtDNA sequences.

To apply hierBAPS to mtDNA sequences, we utilized an R software implementation of the algorithm, RhierBAPS, that is available on the Comprehensive R Archive Network (142). Briefly, the hierBAPS algorithm attempts to maximize the posterior probability of an allocation of a sequence over other possible allocations, assigning each individual sequence to specific clusters. After the number of clusters (K) is specified, the algorithm partitions the sequences of the dataset into as many groupings as possible (up to K_{max} clusters). The initial number of K clusters can be chosen based on the number of subpopulations expected and can be increased on each re-run of the algorithm. The algorithm is typically re-run until the number of clusters ceases to increase.

The clusters were refined into levels of low to high resolution of cluster

specificity. We conducted three different cluster-level combinations: *Level 1*: 4 groups, *Level 2*: 11 groups, and *Level 3*: 24 groups. To distinguish Phylotree labels from hierBAPS groups for the demonstrative purposes of this study, alphabetical letters or Roman numerals were used to represent hierBAPS labels. It is important to note that the hierBAPS group labels are generated in an arbitrary order. In this thesis, the term *subclade* will be used to refer to a cluster of related haplotypes associated with a hierBAPS grouping in **Paper II**.

3.3.3 Definitions of geographic regions

Geographic regions were defined for each region based on the defining features of the study dataset (**Table 3**). **Paper I** compared the distribution of mtDNA sequences within Norway. Norway was divided into four main regions based on dialectal differences between them (121,122): North, West, Central, and Southeast. The mtDNA sequences among modern-day Norwegians were localized largely to cities, reflecting the current locations of most Norwegians, while mtDNA sequences for the Ancestors[4] were spread throughout towns, villages, and cities within different counties.

Papers II and III each focused on a single haplogroup, and areas were divided based on geographical areas with a detailed focus on Northern Europe. Nordic populations were separated into Saami, Scandinavia (Norway, Denmark, and Sweden), and Finland categories. As mentioned before, Finland was kept separate from Scandinavia in this analysis due to its geographic isolation from the Scandinavian Peninsula and its linguistic distinctiveness from the other Nordic languages.

[4] Refers to data from Norway DNA Project linked to ancestral data, where haplogroups were used as a proxy for ancestral Norwegian populations (defined in section 3.1).

Table 3. Definitions of geographic regions utilized for **Papers I through III**

	Defining features	Category name	Geographic areas or peoples represented
Paper I	Norwegian dialectal regions	Southeast (Eastern Norwegain dialect)	Hedmark, Oppland, Buskerud, Akershus and Oslo, Telemark, Vestfold, Østfold counties
		West (Western Norwegian dialect)	Agder, Rogaland, Hordaland (including Bergen), Sogn and Fjordane (including Førde), Møre and Romsdal counties
		Central (Trøndelag Norwegian dialect)	South-Trøndelag, North-Trøndelag (including Trondheim) counties
		North (Northern Norwegian dialect)	Nordland, Troms, Finnmark counties
Paper II	Global haplogroup U5 sequences	Africa	Burkina Faso, Berber, Fulbe, and Fulani peoples
		Western Europe	Ireland, Germany, United Kingdom
		Southern Europe	France, Italy, Spain, Sardinia
		Scandinavia	Denmark, Norway, Sweden
		Finland	Finland
		Saami	Saami from Scandinavia and Finland
		Central Europe	Czech Republic, Hungary (Roma), Poland, Serbia, Slovenia, Slovakia
		Eastern Europe	Baltic, Belarus, Caucasus, Russia
		Asia	India, Iran
Paper III	Global haplogroup J sequences	Sub-Saharan Africa	Uganda, Tanzania
		Northern Africa	Morocco, Tunisia, Algeria, and Egypt
		Arabian Peninsula	Yemen, Saudi Arabia, Kuwait, United Arab Emirates
		Near East	Lebanon, Syria, Iraq, Iran, Israel, Palestinian territories
		Southern Caucasus	Armenia, Azerbaijan, Georgia
		Central Asia	Tajikistan, Kyrgyzstan, Kazakhstan
		Southern Asia	Afghanistan, Pakistan, India
		Eastern Asia	Uyghurs, Mongolia, Oroqen
		Siberia	Evenks, Buryat, Yakut, Khanty, Mansi
		Eastern Europe	Russia west of Ural Mountains, Bulgaria, Ukraine, Belarus, Moldova
		Western Balkans	Croatia, Bosnia and Herzegovina, Albania, Serbia
		Central Europe	Czechia, Hungary, Romania, Slovakia, Poland
		Baltic	Estonia, Latvia, Lithuania
		Finland	Finland
		Scandinavia	Norway, Sweden, Denmark
		Mediterranean	Italy including Sardinia, France, Greece, Cyprus, Turkey
		Iberian Peninsula	Spain, Portugal
		Western Europe	Netherlands, Germany, Switzerland, Hutterite populations, Austria
		British Isles	United Kingdom, Ireland

3.3.4 Haplogroup age estimates

The TempEst software, v.1.5.1 (143), was used to perform a regression of root-to-tip genetic distances against year of sampling. This step was taken to confirm that there was sufficient temporal structure to estimate divergence times. Divergence times with 95% confidence intervals were estimated using the Least Squares Dating IQ-tree plugin (144). To calibrate the ages, we used a root age based on the reported 177 ± 11 kya age estimation for the RSRS sequence reported by Behar and colleagues (95), as well as radiocarbon dating for ancient samples for **Papers II and III**. In **Paper III**, long branches that could cause biased date estimates, such as those in the Viking Burial Site sequences, which collectively contain a high proportion of modern DNA contamination in specific regions (145), were excluded from the dating analysis.

3.4 Comparative data

Since the GenBank data for haplogroup U5 were not representative of the frequency of this maternal lineage in each region, with notably few sequences among the Saami population (n = 10), a literature search was conducted. The search was based on population-based descriptive studies reporting the frequency of U5 mtDNAs within various global populations that provided the geographical prevalence of this lineage, regardless of being based on mitogenome data or only control region sequence data. The U5 frequency from each specific region was tabulated. For more specific information about the major U5 subhaplogroups, we obtained data from 6,488 individuals in the public database on the U5 mtDNA Project available from FamilyTreeDNA (146). The overall frequencies of U5 mtDNAs were plotted on a geographic heat map using the graphical package ggplot2.3 implemented in the statistical programming language R, version 3.6.3 (The R Foundation) (147).

3.5 Statistical analyses

3.5.1 Analysis of mtDNA sequence variation

In order to evaluate the genetic differentiation of the sequences in the study populations for **Papers I and III**, an analysis of descriptive statistics of the mtDNA sequence data, including an analysis of molecular variance (AMOVA), was carried out using the Arlequin software, v.3.5.2.2 (148), and R v.3.6.3 (149). The statistical significance of fixation indices (F*st*) and their respective p-values were estimated by permutation analysis (10,000 permutations) assuming a Tamura-Nei (1993) model with a gamma distribution of 0.26. F*st* values were plotted on multidimensional scaling plots (MDS). Comparisons of haplogroup frequencies between geographic or dialectic regions were conducted using a Chi-square or Fisher's exact test, where appropriate. The DnaSP software, version 5.10.01 (150), was used to calculate the basic parameters of genetic diversity, including the number of haplotypes and Tajima's D. Tajima's D was calculated to evaluate neutrality under the null hypothesis that all mutations are selectively neutral if the studied populations evolve with a constant effective population size (151).

3.5.2 Bayesian Skyline Plot

To estimate population growth over time in the Norwegian population (**Paper I**), we generated a Bayesian skyline plot (BSP) using the BEAST software, version 1.10.4 (152). This is a cross-platform program for Bayesian analysis of molecular sequences, with gamma distributed rates (152–154). Each Markov chain Monte Carlo (MCMC) sample was based on a run of 10 million generations sampled every 1000 steps, allowing for a burn-in of 1%. The convergence of MCMC was assessed and visualized using Tracer v.1.7.1 (155). The mutation rate of 1.64×10^{-7} (10) was used to convert substitution rates into years and coalescent intensities into effective population sizes (154).

3.6 Ethical considerations

The work described in this thesis is based on open-access and publicly available datasets. The respective studies from which these data derive have undergone standard protocols to obtain informed consent from participants, clearance, and approval from the respective ethics committees for sample collection and analysis, as outlined in the associated publications.

4. Results of the studies

4.1 Study I: Matrilineal diversity of Norwegians

Among 1,174 modern-day Norwegians, the HVRI sequences comprised 268 distinct haplotypes. Overall haplotype diversity was high at 0.938 on a scale of 0 to 1, and similar to populations of Spain and Germany (both 0.938) (156,157), but slightly lower than in populations of the Middle East (mostly Saudi Arabia) (0.996) and Buryats (0.991) (156,158,159). The overall mean number of nucleotide differences was 3.18, with the Central and West dialectal regions having the highest values (3.52 and 3.16, respectively). Each region also had Tajima's D values that were consistent with population expansions, with a range of -2.41 (Southeast; $p < 0.01$) to -1.18 (North, not significant). The negative Tajima's D implies a bottleneck followed by a population expansion (151). The values estimated for Norwegians were of a similar value (in the order of -1 to -2) as several other populations, such as subpopulations of the South Caucasus and Italy (160,161).

The mtDNA haplogroups found in Norway could be broadly categorized into eight West Eurasian (N, HV, JT, I, U, K, N1, X, W), five East Eurasian (A, F, G, N11, Z), and one African (L2) haplogroups (**Table 4**). A detailed distribution of haplogroups and their proportions can also be found in Table 1 of **Paper I**.

Comparing these results from Norway to previous results from Denmark (162) and Sweden (163–165), haplogroup H is found at 47–50% in all three Scandinavian countries. This is followed by haplogroup U (13–17.5%) and haplogroup J (10–12%). Finland had higher proportions of haplogroup U (24–27.5%) and lower frequencies of haplogroup J (4.5–5.6%) (86,166–172). Haplogroup U in Norwegians overall (17.4%) was found to be lower than the countries of Eastern Europe (17.9–20.1%) but higher than the British Isles, Spain, Portugal, France, and Germany (9.5–14.8%) (173). Haplogroup J was found throughout most of Europe at about 7–15%, with the higher frequencies of this range found in the British Isles (173,174).

Table 4. Distribution of major mtDNA haplogroups in Norway

Regional origin	Haplogroup	n	%
West Eurasia	HV	512	43.6
	H	40	3.4
	V	24	2.1
	I	30	2.6
	JT	2	0.2
	J	149	13.0
	T	99	8.4
	K	62	5.3
	N, N1	6	0.5
	U	204	17.4
	W	19	1.6
	X	9	0.8
East Eurasia	A	1	0.1
	F	1	0.1
	G	1	0.1
	N11	6	0.5
	Z	8	0.7
Africa	L2	1	0.1

Note: These frequencies are based on data for 1,174 individuals.

Interestingly, the mutational signature of a U5a1 haplotype (16129G, 16187C, 16189T, 16192T, 16223C, 16230A, 16256T, 16270T, 16278C, and 16311T relative to RSRS) appearing among modern-day Norwegians was also found in the mtDNA of a Mesolithic individual analysed by Gunther and colleagues (109). Three J1b mtDNA partial sequences found in Viking Age burials in Hedmark, Oppland, (southwest Norway) and Nordland (northern Norway) (127) were found among modern-day Norwegians in this study. This suggests that these lineages had been prevalent in Norway since at least these time periods. The high prevalence of haplogroups U5 and J1b in the modern population in Norway also suggests some level of lineage continuity in the region for many years.

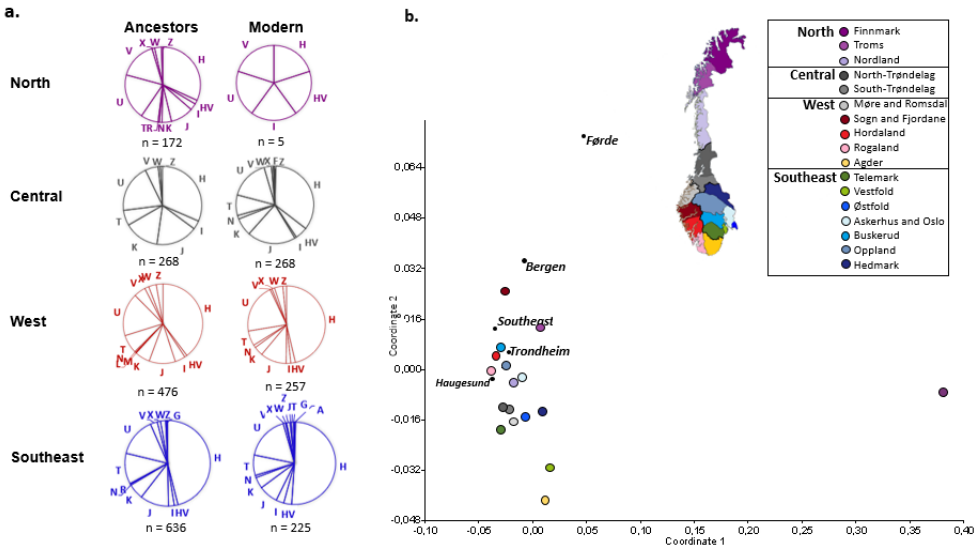

Figure 5. The mtDNA diversity by region in Norway. (a) The historical county map of Norway is colour-coded. Locations in black italics represent areas from which mtDNA data have been obtained among contemporary individuals. The distribution of haplogroups by larger dialectic region are coded as follows: violet: North; grey: Central; red: West; blue: Southeast. (b) A MDS plot of inter-population pairwise F*st* values calculated from mtDNA HVRI.

The haplogroup distribution by dialectal region within Norway is displayed in **Figure 5a**. Among the modern-day population, haplogroup H was most frequent in the Southeast dialectal region (50%) and least frequent in the Central dialectal region (35%) ($\chi^2 = 14.06$; $p < 0.001$). Haplogroup J appeared at the highest frequency in the Central dialectal region (20%) and the lowest in the Southeast dialectal region (8%) ($\chi^2 = 15.74$; $p < 0.001$).

Since there are few data for modern-day population living in the North, the distribution differences between the Ancestor populations in different regions can be compared. The frequency of haplogroup H in the Ancestor population was higher than in the Southeast dialectal region and lower in the North region ($\chi^2 = 11.31$, $p < 0.001$). Although the difference in the frequency of haplogroup U between the North and Southeast region was not statistically significant ($\chi^2 = 2.6$, $p = 0.11$), haplogroup

U5 appeared at high frequency in the North (for which 92% of the U mtDNAs belong to U5) compared to the Southeast region ($\chi^2 = 7.35$, $p = 0.007$).

The MDS plot of population differentiation of the Ancestral and Modern Norwegian mtDNA are displayed in **Figure 5b**. The overall level of population differentiation (F*st*) between different subpopulations was 1.68%, $p < 0.001$. The genetic differentiation by large dialectal regions independently was 0.41%, $p < 0.001$; by subpopulation, the differentiation was 0.81%, $p < 0.001$. The subpopulations with the largest genetic differentiation were in Ancestral Finnmark, Ancestral Agder, and Modern Førde.

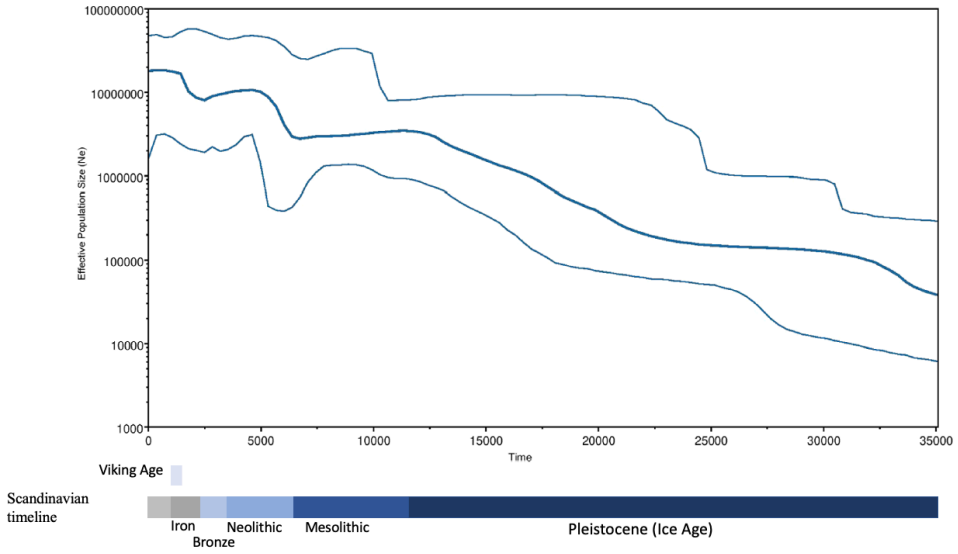

Figure 6. A Bayesian Skyline Plot (BSP) of Norwegian mtDNA sequences. A mutation rate of 1.64 $x10^{-7}$ was used to convert substitution rates into years (x-axis) and coalescent intensities into effective population sizes (y-axis). The Scandinavian timeline was added for ease of visualization of key Scandinavian epochs based on work by Josefsson *et al.* (175).

The BSP analysis of the mtDNA sequences in Norway juxtaposed with the archaeological timeline of Scandinavia as defined by Josefsson *et al*. (175) is displayed in **Figure 6**. According to the BSP, the effective population size (N_e) of the mtDNA lineages increased from approximately 35 kya and then stabilized around 12

kya as the inland ice sheets began melting at the end of the LGM (106). This early period of population growth likely corresponds to the initial settlement of Europe. The population size increased again around six kya and when a Neolithic agricultural lifestyle was dispersed into Norway (176). The population further increased from 2.5 to 1.6 kya (or around 300 to 400 CE) during the Late Bronze Age.

4.2 Study II: Evolution and diversification of haplogroup U5

Given that it is one of the most frequent haplogroups and likely the earliest to arrive in Norway, we explored the genetic structure of haplogroup U5 by employing an unsupervised learning approach for phylogenetic clustering. The most detailed analysis using the hierBAPS algorithm clustered 874 mitogenome sequences into 24 separate groups. RSRS was designated as a separate group (group VIII, not shown), while the other 23 groups corresponded to the specific haplogroup U5 subclades as shown in **Table 5**. After excluding the RSRS sequence, each of the 23 hierBAPS groups were found to share a set of polymorphisms that enabled the hierBAPS algorithm to generate specific clusters for them that also corresponded to several representative subclades.

All hierBAPS groups and the specific set of polymorphisms shared among them were mutually exclusive, i.e., no haplogroups were defined by a set of polymorphisms that was common to two different hierBAPS clusters. Additionally, the hierBAPS algorithm was able to accurately cluster all of the sequences belonging to a specific subclade when compared to Haplogrep2 subhaplogroup assignment.

Table 5. hierBAPS groups found for U5 and their representative subclade(s).

hierBAPS Groups	Subclade(s)	N	%
I	U5a2, U5a2b, U5a2c, U5a2d	90	10.3
II	U5a2e	11	1.3
III	U5a, U5a1	128	14.6
IV	U5b1, U5b1a, U5b1d, U5b1f, U5b1i, U5b3	44	5
V	U5b1 + 16189C! + 16192C!, U5b1b, U5b1c	72	8.2
VI	U5a1a2	26	3
VII	U5a1h	7	0.8
IX	U5a1d	18	2.1
X	U5a1c	28	3.2
XI	U5a1a1	89	10.2
XII	U5b2a	25	2.9
XIII	U5b2, U5b2c	11	1.3
XIV	U5b2a2	29	3.3
XV	U5b2b	19	2.2
XVI	U5b2b1	10	1.1
XVII	U5b1b1+16192C!	81	9.3
XVIII	U5b1b, U5b1b1	39	4.5
XIX	U5b2a1a +16311T!	32	3.7
XX	U5b2a1a2	5	0.6
XXI	U5a2a	70	8
XXII	U5a2a2a	8	0.9
XXIII	U5b1e	25	2.9
XXIV	U5b1e1 (+ T8337C)	6	0.7

The subclades represented by each of the 23 hierBAPS groups were mapped onto a ML phylogeny to determine how well they cohered with phylogenetic branches using the maximum-likelihood model (**Figure 7**). Although not every human U5 haplogroup sequence can be shown, the broad view of the phylogenetic tree, its subclades, and the geographic location of the samples provided insight into the way that U5 subhaplogroups were regionally related.

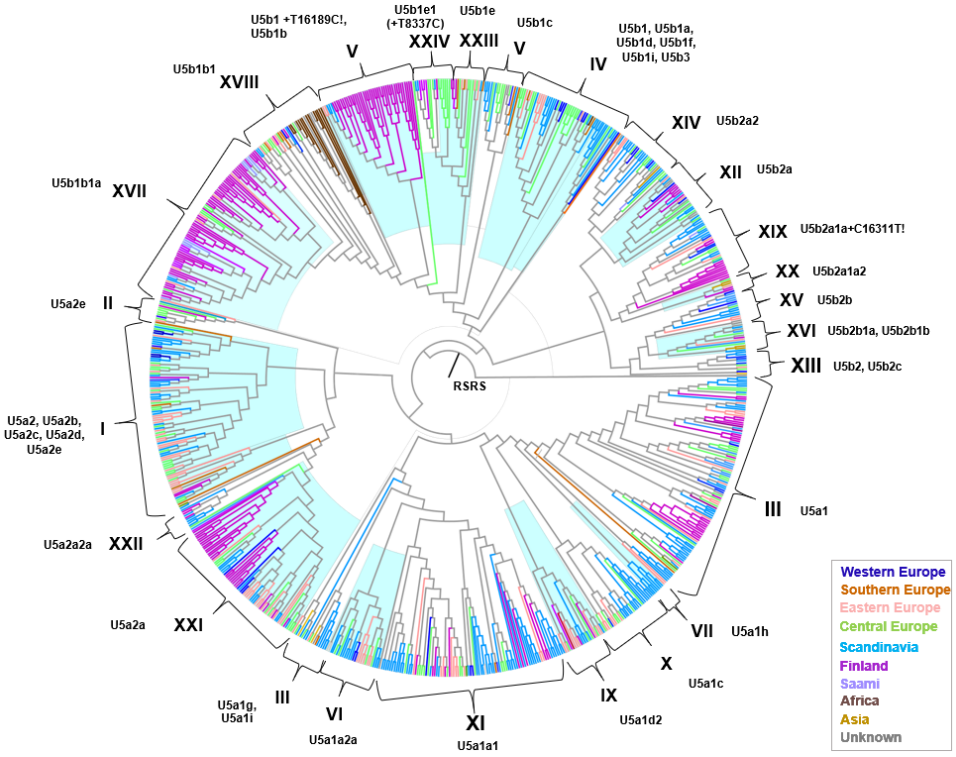

Figure 7. The hierBAPS groups (roman numerals) added to the ML-derived phylogenetic tree of haplogroup U5 sequences. The hierBAPS grouped sequences are separated by alternating light blue and white watermarks.

Several sequences in subclade XVII (subhaplogroup U5b1b1a) are found in Finns, Saami, Poles, Belarussians, and Yakuts of eastern Russia, although the vast majority of these mtDNAs appeared in the Saami and Finns. While a few U5b1ba and U5b1b1a1 haplotypes in the Saami and Finns were similar, a closer look at the haplotypes revealed that the Saami had U5b1b1a3 mtDNA sequences with a A16335G mutation that Finnish populations lacked. This finding suggested that a unique haplotype may have originated among the Saami.

To obtain further information about the dispersal of U5 mtDNAs, we aggregated sequence data for this haplogroup from published sources (**Paper II** Table S7 and Supplemental Material 1) and projected them onto a map of Africa, Asia, and

Europe (**Figure 8**). Haplogroup U5 was found most frequently (between 40–64.8%) among the Saami populations of Norway, Sweden, Finland, and the Kola Peninsula. Areas with the second highest frequencies included areas with Uralic speakers, mostly Finns (23.1% in higher latitudes of Finland to 15.6% in the southernmost part of the country) (86,168,170,172), and then Estonians, Karelians (16.0%) (168), Mordovians (15.9%) (168), and Russians from the Pskov Oblast (19.2%) (177), the latter region having long barrow burials pointing to early Finnic tribe settlements in the 9th–10th century (178). North-dwelling Norwegians (19.0%) (128–130,132) and Swedes (16.6%) (168,170) also had high frequencies of U5 mtDNAs. Overall, a higher proportion of haplogroup U5 mtDNAs was found among the Finns compared to Scandinavians (167).

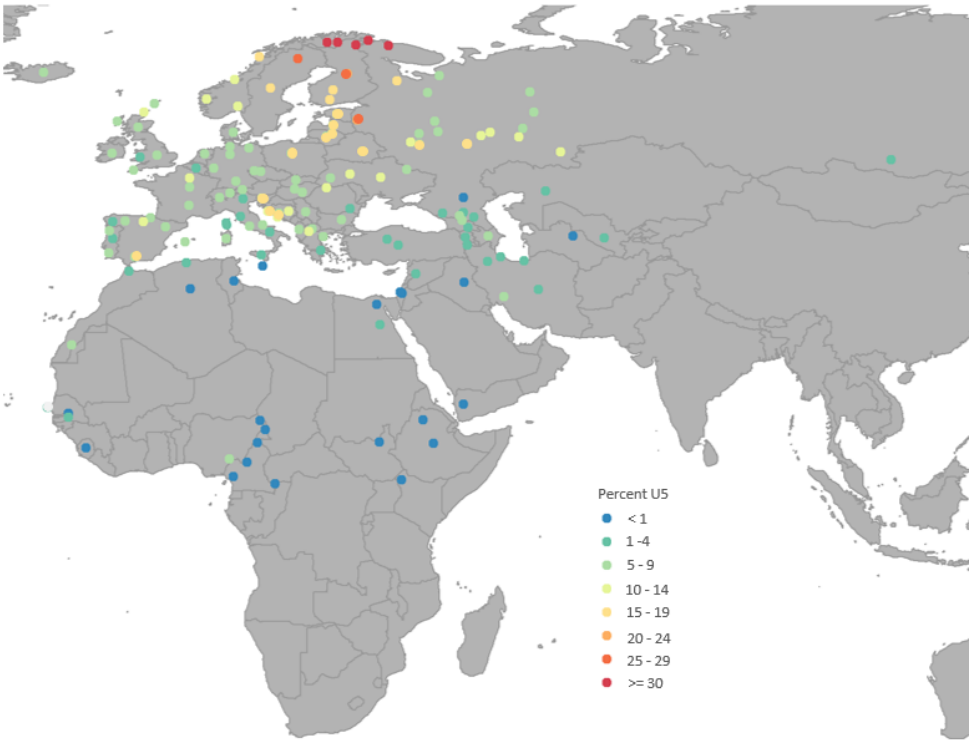

Figure 8. The frequency of haplogroup U5 mtDNAs by world area based on the literature (for percentage data and sources, see Table S7 in **Paper II**).

We found subhaplogroup U5b1 in nine out of 10 Saami sequences, which is not surprising, considering that other studies have found that the 40–65% of Saami U5 sequences belonged to this subhaplogroup (165,179,180). Even so, we observed a single Saami sequence in subclade XXI (U5a2a) which appears to have a separate evolutionary origin from those from the younger subclade XVII (U5b1b1a). It is therefore possible that U5 mtDNAs in the Saami have two distinct sources, the first being Southern Europe via the Franco-Cantabrian refuge (U5b1), and the other from Finland and/or Central Europe (subhaplogroup U5a2).

4.3 Study III: Phylogeographic history of haplogroup J in Scandinavia

Modern populations comprised 93.6% (n = 2,153) of the 2300 haplogroup J sequences, while the rest (6.4%; n = 147) came from past populations dating from the Neolithic to the Early Modern period. Each of the 3-digit J subhaplogroups (J1b, J1c, J1d, J2a, J2b) were found within both Europe and Asia, although there were notable regional differences (**Figure 9a**). J1c comprised over half of the J mtDNAs found within Europe, while it comprised less than 20% of these mtDNAs in the Near East and Arabian Peninsula ($\chi^2 = 38.0$; $p < 0.001$). Conversely, subhaplogroup J1b comprised at least 30% of the J mtDNAs in continental Asia but occurred at a very low frequency in Europe ($\chi^2 = 278.3$; $p < 0.001$). J1d was also more common in Asia compared to Europe ($\chi^2 = 195.5$; $p < 0.001$).

The Viking Burial Sites contained significantly higher frequency of J1b [21.8% to 11.4%, respectively; ($\chi^2 = 4.3$; $p = 0.04$)] and J2b [12.7% to 3.6%, respectively; ($\chi^2 = 7.0$; $p = 0.008$)] mtDNAs compared to modern Scandinavians. Over 70% of the modern Scandinavian population had subhaplogroup J1c. The larger proportions of J1d, J1b, and J2b among the Viking Burial Sites suggested a greater diversity of J subhaplogroups among these past populations compared to modern Scandinavians. However, when the Viking Burial Site sequences were compared with sequences in the MDS plot (**Figure 9b**), the Viking sequences were most similar to those from other European populations. When these Viking Burial Site haplogroups

were compared to data from a study of haplogroup determination of mtDNA control region sequences in a Viking Age population in Norway conducted by Krzewińska and colleagues (181), two J1d sequences and three J1b sequences were found.

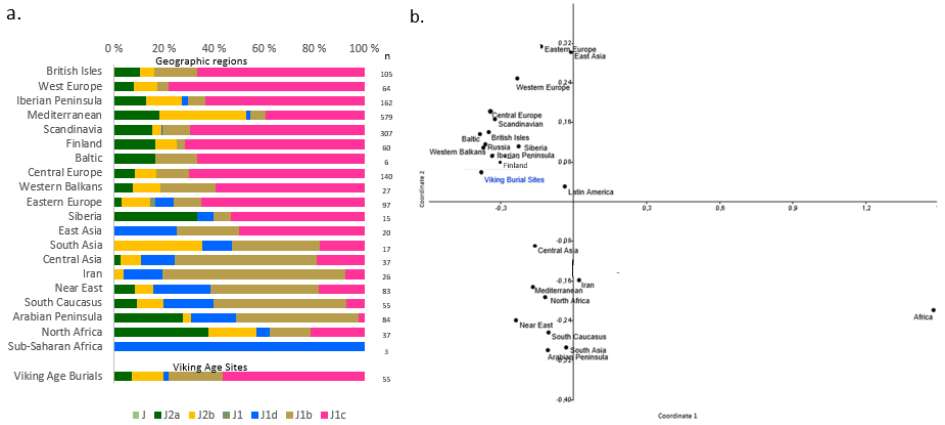

Figure 9. Population differences within haplogroup J. a.) Distribution of 3-digit haplogroup J subhaplogroups within major populations studied in **Paper III**. Haplogroup J is represented by one sequence in Mediterranean Europe. b.) MDS plot of F*st* values to infer the level of haplogroup J control region sequence similarity by geographical regions. Viking Burial Site is highlighted in blue font.

The earliest haplogroup J mtDNAs found in Scandinavia were likely introduced several thousand years after their origin in Western Asia, as a consequence of the northern migrations of Neolithic farmers. Subhaplogroup J2a1 [9.3 (95% CI: 6.3–16.9)] mtDNAs have been identified in several populations of the Near East, Mediterranean, and the Iberian Peninsula, with later derived branches appearing in the British Isles, Denmark, and Sweden. Some subhaplogroups may have also been the result of Viking Age trade and expansions. Subhaplogroup J2b1a, which has been dated to 9.1 kya (95% CI: 7.3–13.6), was found in a Viking Age burial in Sweden and is distantly related to J2b1a sequences identified in the UK and Spain. Likewise, Viking burials in Norway and the Orkney and Faroe Islands have J1b1a1a [3.8 kya (95% CI: 1.8–7.2)] mtDNAs, with these derived types arising from earlier branches that appear in England and Ireland.

The earliest haplogroup J (subhaplogroup J1c5) [8.1 kya (95% CI: 5.3–12.7)] mtDNA reported in Scandinavia was found in human remains from mainland Sweden

that were dated to 5.0–4.9 kya (Scandinavian Neolithic Age). Archaeologically, this individual also belonged to the Funnel Beaker Culture (182), of which settlements were likely to have occurred approximately 6.0–5.5 kya (183). Subhaplogroup J1c5 derives from J1c mtDNAs found in Slovenia, Hungary, and Italy [16.4 kya (95% CI: 11.9–23.4)]. Another early Scandinavian sequence belonging to haplogroup J1c2c1 was found in a 4th century CE burial in Nordland, Norway [4.1 kya (95% CI: 2.5–6.4)] (145). Both the J1c5 and J1c2c1 sequences can be linked to a sister branch in Scotland (subhaplogroup J1c3) [7.1 kya (95% CI: 4.7–11.4)] (**Figure 10**).

Most Scandinavian branches are younger than branches in the Mediterranean countries and the British Isles, suggesting that subhaplogroup expansion occurred relatively later in Scandinavia than in these other regions. The youngest branches of the subhaplogroup J1 — J1c3g [6.4 kya (95% CI: 4.4–9.5)] and J1c3c [4.6 kya (95% CI: 2.7–7.3)] — are shared by populations from Denmark and Germany. Subhaplogroups J1c4 [5.1 kya (95% CI: 3.5–8.3)] and J1c5 [8.1 kya (95% CI: 5.3–12.7)] are also shared between Italians and Danes. Collectively, these findings point to the Mediterranean, Germany, and British Isles as additional sources for the haplogroup J mtDNAs that were brought into Scandinavia.

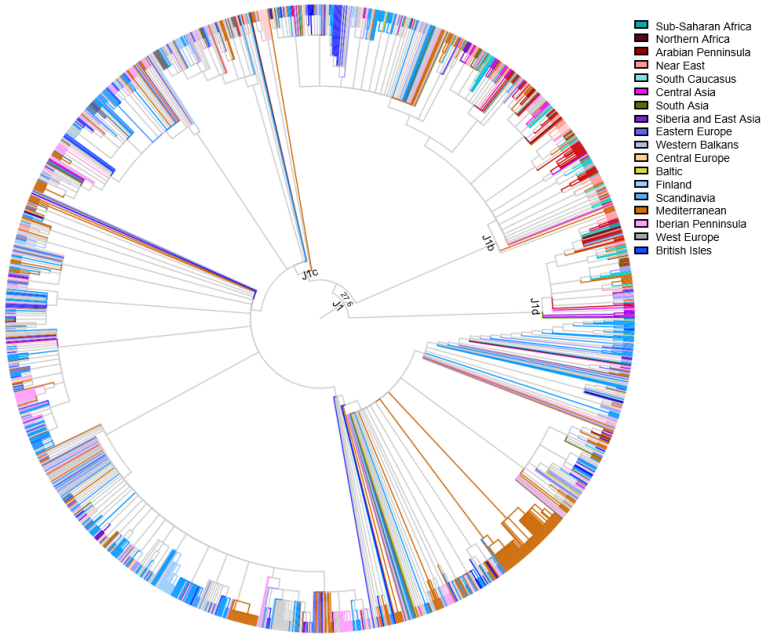

Figure 10. Phylogeographical tree of haplogroup J1.

5. Discussion

5.1 Summary of main findings

The overall aim of this thesis was to apply bioinformatic tools to analyze and explore the mtDNA haplogroups of Norwegian populations, particularly with respect to other Eurasian populations, and to understand the forces shaping mtDNA diversity. **Paper I** laid the groundwork for describing the diversity of mtDNA haplogroups among Norwegians and within specific regions of the country. Haplogroup U5 was the earliest major haplogroup identified in Norway and is currently the second most frequent haplogroup found in the country, occurring at its highest frequency in the northern dialectal region. **Paper II** draws inferences about the evolution of haplogroup U5 and its dispersal into Northern Europe by utilizing a combined phylogenetic approach, ML and hierBAPS, that dated and traced the geographic origin of each subhaplogroup. The phylogeographic analysis produced a visualization of the dispersal of U5 towards Scandinavia from both eastern and southwestern directions. **Paper III** explored the third most frequent mtDNA lineage in Scandinavia, haplogroup J, which appeared at the highest frequency within the Central and West dialectal regions of Norway. Haplogroup J entered Scandinavia approximately six kya, and its youngest branch, subhaplogroup J1c, subsequently became the predominant sublineage within the region.

5.2 Interpretation of the main findings

Norwegian maternal lineages belong to predominantly eight West Eurasian haplogroups. The varied proportions of mtDNA haplogroups within various regions of Norway are modestly influenced by dialectal region, with haplogroups J found at the highest frequency in the Central dialectal region and haplogroup U (mostly U5) found at the highest frequency in the North dialectal region (184). The high frequency of U5 haplogroups mtDNA among Norwegians suggested some level of genetic continuity over several millennia that was not observed in most other western European countries (185,186) (**Paper I**).

As the earliest identified haplogroup in Norway (dated ~9.4 kya) (109), a phylogeny of worldwide U5 mitogenomes was used to track the dispersal of this maternal lineage from an evolutionary perspective. Utilizing an unsupervised learning method, I generated an immediate and accurate identification of numerous closely related mtDNA sequences. When combined with a ML phylogenetic tree reconstruction, broad evolutionary conclusions could be made on the U5 subclades.

The spread of the U5 mtDNAs in Northern Europe was skewed from west-to-east as indicated by the distribution of U5b, although some U5a sublineages in Northern Europe appear to have been dispersed in both west-to-east and east-to-west directions. The deglaciation of the Norwegian shelf between the local LGM and 14-10 cal kya (187) allowed the west-to-east expansion of human groups.

Haplogroup U5 appears to have spread from a co-existing refuge in the so-called “periglacial zone” located in Ukraine and the West Siberian Plain (188). As ice retreated from the eastern portion of the Fennoscandian Ice Sheet, large ice-dammed lakes formed and separated the Baltic countries and northern Russia from Scandinavia (107), thereby preventing early human migrations there. The Baltic Ice Lake persisted until approximately 11,620 ± 100 cal BP when it began dissipating, with this happening before U5 lineages were brought to Scandinavia (189).

In addition to haplogroup U5, haplogroup J represents one of the lineages brought to Norway during a period of further population expansion in Scandinavia, as shown in **Paper I**. Haplogroup J likely arrived in Scandinavia with the Funnel Beaker Culture around six kya when Scandinavia began its transition from a hunter-gatherer to agricultural cultures. While the majority of the haplogroup J sequences found in Scandinavia belong to subhaplogroup J1c, some older subhaplogroups, such as J2a, J2b, and J1b, and J1d, also appear among modern-day Scandinavian and Viking Age individuals (**Paper III**). Differences in the distribution of haplogroup J lineages may reflect the incorporation of non-Norse individuals during Viking Age expansions as well as post-Viking historical migrations bringing individuals who were gradually assimilated into Scandinavian populations.

5.2.1 Periods of early lineage expansions supported by archeological and historical evidence

Although it was the earliest haplogroup found in Norway, haplogroup U5, in addition to haplogroups R, U, and U2 were also found throughout continental Europe prior to the LGM (45–25 kya) (190). Haplogroup U5 entered western Europe around 55 to 30 kya (10,95), and expanded into that region before the end of the LGM over 20 kya (10,84,86,191). At around this time, approximately 23 kya, the northwestern inflow of the Gulf Stream began to melt the Fennoscandian ice sheet covering Scandinavia and the coastline of Norway, resulting in that region becoming slowly inhabited by a new wealth of marine life, tree species, and other wildlife (106,107,192,193). The earliest evidence of human settlement, which dates to approximately 10 kya, comes from archaeological sites found in the Oslo Fjord through the northernmost region of Finnmark (109,110,194,195). By this time during the Holocene, haplogroup U5 became the predominant lineage found in European hunter-gatherers (190), with several early branches found within Scandinavia (**Paper II**).

A second influx of haplogroups from the Near East expanded at the end of the LGM concomitant with the introduction of agriculture and animal husbandry into Scandinavia (196). It is not until this influx accompanying later dispersals during the Neolithic period (approximately 11–6.5 kya) (197) that there is genetic evidence showing that the predominant U5 subhaplogroups had been diluted in Europe (198). The Neolithic agriculturalists of central Europe carried mainly haplogroup N1a, but also H, HV, J, K, T, V, and U3 mtDNAs (199). Archeological evidence further supports the idea that agricultural subsistence was brought to Scandinavia with northward migrating European farmers, and that indigenous hunter-gatherers were gradually assimilated into the farming cultures (200). Farming arrived in Scandinavia with the Funnel Beaker culture by 4000 BCE (201). Corded Ware/Single Grave communities later appeared on the Jutland Peninsula at approximately 2850 BCE, while the late Funnel Beaker culture continued for several hundred years in the eastern parts of southern Scandinavia together with the Pitted Ware Culture (201). Thus, several prehistoric western and central European cultures led to the emergence

of farming in Scandinavia, bringing with them considerable maternal lineage diversity. Ancient DNA evidence further suggests that the transition from hunting-foraging to farming started in southern Scandinavia, with diverse mtDNA haplogroups gradually replacing the predominant haplogroup U mtDNAs (175).

Early Neolithic sites in southern Scandinavia show the synchronous introduction of bread wheat and naked barley cultivars within a 300 year period (4000–3700 cal BCE) across the entire region of southern Scandinavia (202,203). Bones of domesticated cattle were found throughout southern Scandinavia during this same time period (202), consistent with evidence of cereal cultivation and fodder for cattle. Evidence of flint tool mining characteristic of the Michelsberg Culture (4400–3500 cal BCE) found in the modern-day Belgium and western Germany also appeared in Scandinavia around the same time (202).

The expansion of the Norwegian haplogroups (**Paper I**) likely occurred shortly before the Viking Age. However, later historical processes over the past millennium likely reshaped the distribution of haplogroup J mtDNAs in Scandinavia. Our analysis suggested that J1b mtDNAs are less common among Scandinavians today but may once have been frequent among inhabitants of the Late Iron Age to the Viking Age (145,181) (**Paper III**). Population changes following the Viking Age travels in the 10th and 11th centuries CE may have further shaped the distribution of J subhaplogroups. As carriers of maternal lineages, women have been cited as important agents in the Viking expansions as settlers both within and out of Scandinavia's borders (181,204). Based on archeological evidence found in burial site artefacts, the Vikings often incorporated non-Scandinavian individuals into their populations (127), and established several satellite communities outside of Scandinavia that persisted until the end of the Viking Age in the 11th century CE (205). Graves found in England, continental Europe, and eastern areas of Viking expansions (206–208) contained traditional Viking burial gifts (209–211), and genetic analysis suggests that several of these foreign graves contained individuals with Norse ancestry (145). Within Scandinavia, genetic evidence suggested that individuals who assimilated into Viking culture were also given Viking burials, as the

mtDNA haplogroups present in these individuals were uncommon among individuals with Norse ancestry (181). Over time, haplogroup J1c became the predominant J subhaplogroup in Scandinavia.

A comparison of the data from Ancestor and Modern populations showed that the frequencies of these haplogroups in the current Norwegian population within different dialectal regions have changed modestly over time, with slightly higher frequencies of U and K but slightly lower frequencies of H and J occurring today **(Paper I)**. The diversity of haplogroups among Modern Norwegians underwent expansion. All the haplogroups present in the Modern Norwegian population were also present in the Ancestor Norwegian population as expected, except for haplogroups N11 and F2a, which are present at low frequency in the Modern population.

A small number of mtDNAs in Norwegian populations belonged to lineages that did not originate in West Eurasia. Among these were mtDNAs belonging to East Eurasian haplogroups, which were likely contributed by people recently immigrating to Norway from regions in which these maternal lineages are common, e.g., Southeast Asia. By contrast, G2a1 and Z1a, which are both haplogroups of East Eurasian origin, have been present among Norwegian Ancestors since at least the 1600s and likely entered the population during recent prehistory. They are also present in Uralic-speaking aboriginal Siberians who genetically influenced ancestors of the Saami population (212,213). Haplogroup G2a is also present at low frequencies among populations of Central and Eastern Europe, while haplogroup Z is present at about 4 to 7% among Saami populations (165,212,214).

In addition, haplogroup L2, the only maternal lineage of African origin, found both in our study and that of Passarino *et al.* (129), appeared in one participant in The Norwegian DNA Project (not included in this analysis) and is associated with an individual from the Dominican Republic who arrived in Norway in 1860. A haplogroup L2 mtDNA also appears in the Hordaland Ancestor population in an individual with a Norwegian first name and surname. It is therefore probable that the

L2 mtDNA was introduced into Norway by an early modern immigrant and is now included among the maternal lineages found among Norwegians.

5.2.2 Migration routes and the importance of maritime travels

Despite its long length (1,752 km) (105), the entire Norwegian coastline was quickly settled by pioneering groups (215) who employed watercraft to swiftly move along the long coastline for hunting and fishing (216). The coastal-based location of the settlements, with each settling area opening up into a body of water, was confirmed based on two archaeological proxies: summed radiocarbon probability distributions and site count data of shoreline dated sites (195). These data also showed that coastal areas of Norway were of central importance for early human settlement. The abundance of resources inferred by the archaeological data suggests that marine resources may have been enough to sustain a stable population size during the Middle and Late Scandinavian Mesolithic periods, as well as the first part of the Neolithic period (195).

Our phylogeographic analysis showed that haplogroup U5 was an ancient maternal lineage and diversified quite extensively since its origin (**Paper II**). Interestingly, Neolithic maternal lineages such as haplogroup J did not extend as successfully far north, where U5 comprises over 50% of maternal lineages among the modern-day Saami. Among Finns and Scandinavians, U5 continues to be the second-most frequent haplogroup after H (168,179,184). Although the warmer temperatures during the period from eight to 4.8 kya allowed for human settlement (217–219), it is likely that the lack of arable land in Scandinavia above the Arctic circle limited agricultural opportunities, and hence the northward dispersal of haplogroup J-carrying Scandinavian farmers. In support of this interpretation, there is no evidence of Neolithic pottery above the Arctic circle in Norway and the few reported finds of cereal crops have been discounted (220,221).

The topography of Norway is different compared to the relatively uniform landscapes of Denmark and Sweden and has likely influenced human movement there. Mountainous terrain runs through in the innermost parts of the country, and the valleys and fjords running through the coastal areas separate large portions of the

population from each other. The Jotunheimen Mountain range, which includes the highest mountain in Norway called Galdhøpiggen at 2,469 m (108), divides the most populated portion of Norway into eastern and western halves in the southern and widest part of the elongated country. The matrilineal differences between otherwise closely located regions (e.g., the Ancestor West Norway versus Ancestor Southeast Norway showed significantly differences, $p < 0.001$) were likely shaped by the geographic barrier of tall mountain ranges that prevented frequent close contact between the two regions (**Paper I**).

Despite these geographical barriers, Norwegians maintained remarkable contact across long distances through maritime travel. The importance of marine vessels is reflected in their appearance in petroglyphs found all over Norway (222–224). While recent archaeological evidence indicates that, by 6180–6680 cal BP, early settlers sometimes traveled across the Jotunheimen Mountains to connect to the Southeast or West (225), this travel was likely restricted to the winter months when bogs and streams were frozen over. During the Neolithic, the noted rapid expansion of farming culture throughout the large geographic area of the Scandinavian peninsula (226) suggests that marine vessels were employed by these early Scandinavian farmers. Within Norway, haplogroup J mtDNAs are more frequent in Central and Western Norway, specifically Trondheim and the former Norwegian capital of Bergen (184), which have historically been major regions of trade and maritime commerce. Ships were also clearly crucial for the Norse expansion during the Viking Age (227). Moreover, the mid-19th century "golden age" of sailing vessels has been recorded as being a catalyst for major timber and fishing-based economic activities within Norway that led to eventual economic prosperity (228,229).

Maritime mastery during the Viking Age (750–1050 CE) had a major influence on Norway's gene pool and has been described in terms of both genomic admixture and changes in mtDNA haplogroup frequencies (145,181). This shift likely resulted from greater gene flow from the British Isles into modern Norway compared with Denmark and Sweden, which, in turn, was due to more frequent maritime routes of the western-dwelling Norse to the British Isles (145). For haplogroup J, the phylogenetic tree from this study showed that most Scandinavian branches are

younger than those in the Mediterranean countries, Western Europe, and British Isles, suggesting that these subhaplogroups arose relatively later in Scandinavia than in these other regions (**Paper III**).

Maritime travel influenced Norwegian society and culture long after the Viking expansions. The petty kingdoms of Norway were politically unified as a single entity by the 11th century (230,231), relatively earlier than other countries of similar size in Europe. Due to its climate, the economic dependence on natural goods in Norway varied by region and required timely ship-based mechanisms to distribute fresh cargo throughout the rest of the country. The processing and packaging of fish for export also took place in or near coastal towns along routes of great distance. Extending the length of the western coastline, the Norwegian fisheries industry was dispersed in scattered longitudinal rural settlements rather than more nucleated fishing communities as seen in other European countries such as Great Britain, Germany, France, and Denmark (228,232,233). By 1850, nearly all Norwegian towns west of Lindesnes in Adger County, Norway's southernmost point, were engaged in the processing and exportation of fish and/or shipping and shipbuilding (228). The majority of the country's 3% arable land in the Hedmark and Østfold Counties in the southeast region of Norway was dedicated to growing crops on large farms (234). With sparser forests, the central and northern parts of Norway have traditionally relied more on fishing (234).

Mass migrations started within Norway in the 1750–1780s (119,120) and boomed in the mid-1800s when Norwegians no longer needed official permissions to relocate (120,235). A large proportion of migrants were young couples looking for better economic prospects and opportunities to expand their families. The most comprehensive and most widely mentioned wave of internal migrants was that from southeastern Norway and southern Trøndelag to the Målselv and Bardu Valleys southeast of Tromsø (119). Between 1750–1801, as rural inland farming areas in the southeast became more overpopulated, people emigrated from the inland districts to the Oslo Fjord for better opportunities in agriculture and timber trade (236,237). Similarly, people moved from the inland southeastern, western, and central parts of

Norway to the northernmost coastlines for cod and herring fishing opportunities (238).

Some Norwegians also emigrated after famine and poverty spread in the 19th century. However, the majority of those who could not afford the ship fare and who did not want to enter into indentured servitude relocated within the country itself (237). Maritime routes remained the primary mode of trade and transportation until the use of motorized vehicles became popular in Norway in the early 20th century (239) and railway networks expanded, with the Bergen line (*Bergensbanen*) in the West reaching the inland Southeast in 1909 (228).

5.2.3 Regional differences within Norway

The detailed geographical analysis (**Paper I**) indicated that Norwegians from Finnmark, and to a lesser extent, Agder and Førde, appear to have become moderately differentiated from the general Norwegian population. The observed level of dissimilarity between these regions is similar to what had been reported for Y-chromosome (paternal) lineages in Norwegians (240). Evidence of genetic substructuring based on an AMOVA differentiation for the Y chromosome was reported for regions of Finnmark, Sogn and Fjordane (where Førde is situated) and Agder (240).

Northern Norwegians were found to have a large proportion of haplogroup U5, which is also found at high frequencies among the Saami. Norwegians have been living in the northernmost parts of the country since at least the 14th century (241). The Norwegian population was scattered along the coastal areas, and the region was also occupied by the Saami, an indigenous people in Norway that had been seasonally nomadic until the late 19th century (242). Interestingly, Krzewińska *et al.* (181) detected U5b1b mtDNAs in two individuals from the Late Iron Age who had received a Norse burial, suggesting that some Saami individuals may have been assimilated or accepted into Norse society as early as the 10th century CE. A separate analysis of the overall genetic structure of Norwegians, which was published within 30 days to **Paper I**, found that Saami ancestry was enriched in northern Norway

(243). Further northern settlements were encouraged in the 1780s by the appointed bailiff in order to take advantage of fishing opportunities there (119). These northward migrations continued until the 1830s, when the introduction of potato farming and smallpox vaccination made southeastern and western Norway more hospitable and prosperous (244,245).

The hierBAPS-ML phylogeny of haplogroup U5 showed that populations from Finland, Scandinavia, North Africa, and Central and Eastern Europe shared U5 subclades (**Paper II**). The diversification of haplogroup U5 in Southern and Western Europe indicated a west-to-east migration into Scandinavia. However, U5b1b may have spread with Uralic-speaking groups eastwards (165), as Saami-specific lineages found in Norway were also present in Uralic-speaking populations of Eastern Europe and Siberia (165). The results of our analysis supports the view that the divergence of U5b1b likely occurred via a scenario in which one subhaplogroup (U5b1b1) became prominent among African populations after ancient hunter-gatherers crossed the Strait of Gibraltar (246). Over a few thousand years, U5b1b1a, which dated back to 4.1 kya in our study, became more frequent farther north in Scandinavia with the spread of U5b1b1 (165,167). Our phylogenetic tree further showed that both lineages were distantly related to the younger subhaplogroups U5b1c and U5b1e1 in Central and Eastern Europe. This finding confirms that the early migrations of individuals with U5b1 mtDNAs likely occurred from west-to-east rather than the opposite direction.

A different pattern was found for the western part of Norway. While the haplogroup composition of Ancestor populations from Sogn and Fjordane was similar to those of other counties in the West, modern Førde has become more heterogeneous over time (**Paper I**). The frequency of haplogroup J mtDNA was higher in Central and West Norway, which have historically been major regions of commerce. This is supported by historical documentation of later immigration events during and following the medieval period, particularly from Great Britain, the Netherlands, and the influx of Hanseatic League settlers, all of which had a major influence on Scandinavian culture, religion, and economic trade (247–250). Before the 17th

century, Bergen was the international commercial city in Norway, and goods coming to and from Trondheim were channeled through Bergen (251). At this time, foreigners from the United Kingdom, Germany, and the Netherlands were legally able to freely trade with Norwegian peasants, but after the mid-17th century, official policy afforded privileges to Norwegian merchants (248,252,253). Foreign merchants circumvented this policy by establishing permanent residency in Trondheim (252,253). Our study found some evidence of J1c sequences from Germany, the British Isles, and Mediterranean Europe, connected to Scandinavian sequences collected from modern populations.

5.3 Methodological considerations

5.3.1 Sampling bias

This study utilized datasets that are open-access and publicly available, which may be subject to sampling bias. The majority of the data (78%) from **Paper I** was selected from published sources that specifically investigated the demography of mtDNA haplogroups in Norway (127,129,130). The criteria for this selection included sampling from several areas of Norway (127), ensuring national representation through questions about regional maternal ancestry of participants (129), and acquiring data from a non-biased source with respect to mtDNA haplogroups, such as male Norwegian military enrolees (129). Data from Opdal *et al.* (128) included Norwegians from the Oslo University Hospital catchment area whose mtDNAs were being studied for the number of polymorphisms related to sudden infant death syndrome. Since these were the only data available from the Oslo region, and no correlation between mtDNA haplogroups and disease was shown, we included these data in our analysis.

Paper I and **Paper II** utilized some data from FamilyTreeDNA, a commercial testing company based in the United States but globally available and utilized. Since members must pay a fee (currently 159 USD) to have their mtDNAs analyzed, these data are likely to be skewed towards individuals from higher income countries. However, **Paper I** focused only on ethnic Norwegians in Norway, a country with one

of the lowest levels of economic inequalities in the world (254), and it is unlikely that participation in FamilyTreeDNA would have biased the demographic profile of participants towards a particular maternal ancestry or region within Norway. However, we found in **Paper II**, which focused on continental European populations in which haplogroup U5 mtDNAs are primarily found, that there were too few Saami sequences (n = 10) in the dataset. To obtain more representative distributions of U5 subhaplogroups, we utilized control region mtDNA data from FamilyTreeDNA's U5 Project, which featured only one lineage. These data allowed us to expand our sample size and determine that U5b1 was indeed the predominant haplogroup in the northernmost region of Scandinavia among the Saami.

The public database containing mtDNA sequences, GenBank, was also used to obtain sequences for **Papers I–III.** The majority of the data (>99%) for **Paper II**, and **Paper III** from GenBank were selected with respect to the haplogroup being studied (haplogroup U5 or J). A larger proportion of global sequences were obtained for haplogroup J, where the regions with the largest proportion of haplogroup J, the British Isles, and the Near East, were adequately represented for the purpose of evolutionary inferences to Scandinavia (n = 109 sequences each) (**Paper III**). The frequency of haplogroup J is uniformly distributed at around 7–15% (174) across the European and Asian countries where it is predominantly found.

There were a few obstacles when gathering whole mitogenome U5 data. Firstly, a paucity of whole mitogenomes was found for the Saami, likely not reflecting previous studies that have found that U5 comprised 50% of Saami mtDNAs (168,179). Secondly, there was a limited number of studies on the overall distribution of haplogroup U5 in the world. Therefore, I sought to map out the representation of the proportion of U5 sequences in each country (**Paper II)**. Since the descriptive studies included in the literature were conducted with demographic details in mind and contained a larger amount of pooled data overall, they would be a closer representation of the actual global frequencies of U5 mtDNAs than our limited sample of 873 sequences. Accordingly, we aggregated data from mtDNA population-based studies in the literature and calculated the frequency of haplogroup U5 out of a total of 56,617 sequences. These estimates were then used to map the distribution of

haplogroup U5 in Norwegian populations, including the Saami. The Saami have roughly 50% haplogroup U5 mtDNAs, which is in close approximation to their frequency in other Scandinavian populations.

5.3.2 Limitations of partial (control region) sequences

Although whole mitogenome sequences would have been ideal for the analysis of matrilineal diversity in Norway (**Paper I**), we were limited in our analysis by using published control region sequences. While control regions enable an accurate estimation of haplogroups, this limited our confidence in predicting the detailed subhaplogroup for all sequences as well as in automating of the process of phylogenetic tree reconstruction. For these reasons, we focused our analyses on major haplogroups.

The non-coding region carries a large number of haplogroup-defining mutations. The substitution rate of the non-coding region was estimated to be nearly eight times higher than the coding region. Using 2,196 mitogenomes, Soares and colleagues (10) calculated the average number of substitutions per nucleotide position to 0.47 for the coding region and 3.74 for the control (non-coding) region. The algorithmic program, Haplogrep2, was used to predict the haplogroups based on these control region sequences. Haplogrep2 computes the haplogroup classifications on pre-calculated phylogenetic weights that correspond to the occurrence per position in Phylotree, Build 17 (8), which, in turn, reflects the mutational stability of a variant. The use of Haplogrep2 thus automated predictions of major haplogroups for these partial (control region) sequences.

In the case of partial (control region) sequences, creating a phylogenetic tree using a single, algorithmic program is not straight-forward. There is no single algorithm that considers back mutations, non-continuous mutations[5], and hotspots that are common in the mtDNA control region. This is also true when using mitogenome sequences as well. However, having more nucleotide positions in the input sequences can improve the identification the relationships between more detailed subhaplogroups that share the same or similar mutations. In other words,

[5] mutations diagnostic for one haplogroup, but not all daughter haplogroups

more detail can be gleaned from the relationships between mitogenomes with more sequence information. Additional resources and bioinformatic tools apart from phylogenetic programs (particularly Phylotree and Haplogrep2) are needed in order to examine particular mutations of interest, cross-check the phylogenetic tree arrangement, and add haplogroup labels.

Although a considerably smaller sample size was used for **Paper II** than **Paper I** (n = 873 vs. 1,174 for **Paper I**), **Paper II** utilizes mitogenome sequences to create a phylogenetic tree using a combination of ML phylogenetics and hierBAPS. Supplementation with hierBAPS in **Paper II** allowed us to assign haplogroup statuses to groups of individual sequences that varied slightly from each other in an ML tree that was also less cumbersome and error prone.

5.3.3 Generalizability

This project demonstrated that a ML phylogenetic visualization (**Paper II** and **Paper III**) can be utilized as a starting point for investigating the divergence of mtDNA haplogroups in evolutionary and geographical terms. The hierBAPS algorithm, which grouped large clusters of ancestrally derived sequences in **Paper II**, facilitated making broad inferences about their evolution and presented a more refined visual organization that may not have been evident based on detailed haplogroup labeling alone.

The phylogenetic trees discussed in this project were generated through ML estimation and included numerous branches of haplogroup U5 and J, each constituted by over 800 mitogenome sequences. Although these trees are not considered to be an exhaustive representation of every known branch of haplogroup U5 or J, many of the key subhaplogroups were represented. A major reason why haplogroups U5 and J were the focus of our study was because these haplogroups occurred at high frequency in Scandinavia. Thus, they had a higher likelihood of being represented in public databases. Their high frequencies also suggested that these haplogroups likely had a phylogenetic diversity worth exploring. It is also worth noting that adding more mitogenome sequences would result in a fundamentally similar phylogenetic structure and dating of the constituent subbranches via estimations conferred by the ML models utilized by IQ-tree.

A major challenge of this project was achieving an adequate representation of the subbranches within a haplogroup and producing a legible phylogenetic tree, particularly with regards to the geographic origin of the samples. Despite not having access to every known U5 or J haplogroup sequence, this study presents findings that bear on the phylogenetic history of haplogroup U5 and J, the phylogeographic history of these maternal lineages in Europe, and the cultural, historical, and geological factors that shaped their distributions in contemporary Scandinavian populations.

5.4 Public health implications

While the characterization of mtDNA diversity in human populations has become an essential component of molecular anthropological research, the clinical significance of sequence variation in mtDNA haplogroups has been less studied. Furthermore, the extent to which the predominance of certain haplogroups, particularly in the region of Scandinavia, is related to survival or adaptation to external conditions remains to be fully elucidated.

Although it is well known that particular mtDNA variants can lead to mitochondrial disease, substantially less is known about the association of specific types of mtDNA with phenotypes of complex origins, such as Parkinson's disease or longevity (255). A recent study using data from the UK Biobank found that particular mtDNA variants were associated with diseases that were also reported by previous studies (79). One study found that Russians with haplogroup J, and specifically those carrying the A10398G mutation, were more likely to have multiple sclerosis (256). A different study showed that the T8655C mutation in the *ATP6* gene in a haplogroup L subclade among Africans was associated with type II diabetes (257). In addition, subhaplogroup J2b, which occurs predominantly among individuals in Mediterranean Europe and South Asia (as identified in **Paper III)**, was associated with iron deficiency anaemia caused by iron deficiency (79). Subhaplogroup U5a was further associated with polyuria, joint disorders, and lesions of the plantar nerve (79).

Several other studies have linked specific mtDNA haplogroups to clinical diseases. A case-control study of 406 patients and 183 healthy controls found a

favourable statistical association between haplogroup U5 and the risk of cardiovascular infarction, but a higher risk of low ventricular ejection fraction (<40%) (258). Other studies have found biological mechanisms that supported higher sperm motility among patients with U5 mtDNAs (259). Yet, another study confirmed an association between sperm motility and both the OXPHOS system and mtDNA haplogroup (260), although this finding was not supported by later research (261). Finally, a study has suggested that haplogroup U mtDNAs occurred at high frequency among patients with elevated risk for occipital brain infarct (262), with a related study suggested that the association was due to a high frequency of haplogroup U5 (263).

These findings represent correlations and may not take into account confounders or interactions with alleles in the nuclear DNA (264). A more comprehensive understanding of mtDNA mutations with regards to their role in disease and interactions with other genes, as well as the mechanisms by which mutations influence phenotypes, will promote the understanding of the contribution of particular mtDNA haplogroups to human health.

It is uncertain whether adaptation is an essential factor at play in the regional prevalence of certain haplogroups found in Scandinavia. While there are differences in the frequency of haplogroup U5 in Europe, this pattern may reflect genetic drift. This view is supported by studies of nuclear DNA SNPs and microsatellite markers, which show a high degree of linkage disequilibrium, and thereby less genetic diversity, in the Saami population compared to Scandinavians (265–268). Coupled with geographic isolation and historical segregation policies (269), genetic drift likely contributed to the relatively high proportion of haplogroup U5 mtDNAs in far northern Saami populations.

By contrast, the relatively uniform distribution of haplogroup J in Scandinavia might be due to an adaptive advantage conferred by the mutational features of this maternal lineage. Haplogroup J mtDNAs share polymorphisms that have been predicted to increase an individual's basal metabolic rate, possibly as a result of a survival advantage in colder climates (270). Certain amino acid substitutions were also found to be highly conserved in certain subbranches of haplogroup J, where they

also represent lineages from colder Europe compared to the warmer climate origins of haplogroup J (260). Other studies suggest that several of the highly conserved haplogroup J polymorphisms are correlated with an increase of energy deficiency and longevity. More specifically, the T14798C substitution associated with subhaplogroup J1 may affect an inner CoQ-binding site that could reduce proton pumping and coupling efficiency (271,272), providing more heat release in colder environments (260).

In **Paper III**, we found that subhaplogroup J1b was prevalent among the Viking Age burials compared to modern-day Scandinavians. Krzewińska and colleagues (181) noted the same trend in three Viking Age individuals. By contrast, J1b was associated with a case of plantar nerve lesions (79). We found no evidence for an association between J1b mtDNAs and decreased survival in Scandinavia. However, certain J1b mutations were found to be associated with a decreased likelihood of Parkinson's disease (273).

Numerous studies have explored the possible influence of haplogroup J on longevity (260,274–277). If longevity is indeed affected by haplogroup J, then its high frequency could be a contributing factor to the overall longer life spans observed in Scandinavian countries (278). However, studies examining the relationship between haplogroup J and longevity have yielded mixed results. Studies of populations from Finland (274), Italy (275,277), and Spain (276) found a significantly higher frequency of haplogroup J among healthy individuals over 90 years of age than in younger controls. These proportional differences were not found for haplogroup J among individuals in Southern Italy (279), Ashkenazi Jews (280) or Uyghurs (281). A similar study of Irish individuals also did not find an overall correlation, although, when specific haplogroup J branches were considered, subhaplogroup J2 occurred more frequently among individuals over 80 years of age (282). Similarly, J2 was also found to be overrepresented among older individuals in Northern Spain (276), while another study found no evidence of such associations between haplogroup J and Spanish centenarians (283).

Studies of haplogroup J mtDNAs and life expectancy are limited by features that are typical of correlative studies in which a causal relationship cannot be

established. Additionally, some countries have a limited population of older adults for comparison, such as Tunisia in which the average lifespan is 70 years, thus limiting the conclusions that can be reached about mtDNA variation and longevity (284). Most importantly, variation in lifespan depends on interactions between multiple genetic loci, epigenetics, and manifold environmental exposures throughout the life-course that modulate susceptibility to age-related diseases (285,286).

A combination of particular mtDNA alleles, rather than single alleles, showed a stronger association with the ageing process (287). As an example, it was shown that the combination of three mtDNA polymorphisms — C150T, T489C, A10398G — is associated with longer life expectancy in both Finland and Japan (287). While these alleles were associated with a specific haplogroup, namely haplogroup J2, none of the individual mutations occur exclusively in haplogroup J2. Furthermore, not all haplogroup J2 mtDNAs contain this same combination of variants. In this regard, it should be noted that none of the J subhaplogroups in this project contained this combination of variants.

5.5 Future perspectives

It is well-established that haplogroups and subhaplogroups arose in specific geographic regions and their lineages can be tracked across multiple generations. Within specific regions, it becomes possible to explore the relationship between mutational changes within mtDNA haplogroups and the complex traits or phenotypes that are also associated with these regions. The connections between haplogroup and disease prevalence establishes the groundwork for researchers to develop projects at the crossroads of public health, anthropology, and mitochondrial genetics.

Whether the presence of mtDNA lineages in certain regions of the world are merely markers of human migrations or also reflections of biological advantage(s) conferred to human hosts remains unclear. Several studies have speculated that the continuity of particular mtDNA lineages in northern latitudes may due a protective advantage from cold weather, but supportive evidence for this claim has been

inconsistent (260,270–272). Additional studies in larger populations are necessary to demonstrate the reproducibility of results and determine the directionality of the effects or associations. These studies can be further strengthened by investigating the function of haplogroup-specific polymorphisms on cellular metabolism in experimental studies that use *in vitro* models.

Whether or not mtDNA haplogroups affect human health remains an open question. The associations found in retrospective studies draw correlative conclusions which beg further investigation. Prospective study designs are needed to test the strength of the causal relationship between a specific haplogroup and a complex trait. Selection of complex, clinically relevant traits found in a particular population along with haplogroups predominant in that region can be a starting point for further investigation in epidemiological studies.

For haplogroup-specific public health studies, the investigation should include haplogroup determination for mitogenomes found within the population, including subpopulations, and a detailed outline of unique variants found within highly prevalent haplogroups. For larger, population-based studies, the study design should account for the prevalence of the specific haplogroups to be studied in order to have enough sequences available to represent haplotype variation. Since statistical power may be limited for haplogroups occurring less frequently in a population, cohort studies monitoring incident cases of a common disease may be necessary to investigate the relationship between a less frequent haplogroup and a disease. Furthermore, future studies investigating complex chronic diseases must take into account life-course risk factors associated with the conditions studied (for example, tobacco smoking and alcohol drinking in association with cancer). As with genome-wide association studies, mitochondrial genome-wide association studies may also be conducted to reveal the associations between specific, heritable variants and clinical disease, as some have been already completed (288,289).

Through many millennia, the population of Norway has changed in its diversity and substructure. These environmentally and historically driven changes may have resulted in the loss of mtDNA lineages that have yet to be defined, but would be of great interest to population geneticists, anthropologists, and

demographers alike. While there have been some attempts to compare the distribution of mtDNA lineages within populations across time, e.g., pre- and post- Black Death in London and Denmark (290), these comparisons have largely focused on single-digit major haplogroups (e.g., haplogroup J). With the availability of high-throughput sequencing and advancements in bioinformatic tools, researchers can now rapidly identify subhaplogroups within these major haplogroups. These tools further enable researchers to conduct a more detailed examination of subhaplogroup lineage continuity and identify temporal shifts in subhaplogroup frequency within a haplogroup before and after an event.

6. Conclusions

Combined with archeological, linguistic, and sociocultural evidence, our genetic analyses reveal that Norwegian maternal ancestry has been shaped by major evolutionary, cultural and demographic events over many generations. This project has been, in part, an attempt to chronicle the expansion of some of the less studied haplogroups within Norway and other Scandinavian countries. It has also demonstrated the efficacy of a new bioinformatic approach for investigating patterns of matrilineal diversity and the phylogenetic structure of specific haplogroups.

During the Mesolithic period, haplogroup U5 was the predominant maternal lineage in Scandinavia. The combined hierBAPS-ML based phylogeny analysis demonstrated that haplogroup U5 was brought into Norway from both eastern and western directions, providing insights into the phylogeographic partitioning of haplotypes within this lineage. Haplogroup J was brought to Scandinavia with the dispersal of Neolithic farmers from West Asia and Europe around six kya, with J1c eventually becoming the most common sublineage there.

During the Viking Age marked by extensive historical interactions with foreign regions, maternal lineage diversity in Norway increased, with additional sublineages of haplogroup J entering at this time. Following the Viking Age, continued external migrations as well as internal migration policies influenced the pattern of mtDNA diversity in Norway. Our analysis further revealed the importance of geographic regions as boundaries to gene flow among a people deeply influenced by Norway's topography and maritime travel around the country.

Bibliography

1. Margulis L. On the origin of mitosing cells. J Theor Biol. 1967;14(3).

2. Gray MW, Cedergren R, Abel Y, Sankoff D. On the evolutionary origin of the plant mitochondrion and its genome. Proc Natl Acad Sci. 1989;86(7).

3. Gray MW. Rickettsia, typhus and the mitochondrial connection. Vol. 396, Nature. 1998.

4. Andersson SGE, Zomorodipour A, Andersson JO, Sicheritz-Pontén T, Alsmark UCM, Podowski RM, et al. The genome sequence of Rickettsia prowazekii and the origin of mitochondria. Nature. 1998;396(6707).

5. Giles RE, Blanc H, Cann HM, Wallace aD. C. Maternal inheritance of human mitochondrial DNA. Proc Natl Acad Sci U S A. 1980;77(11 I).

6. Anderson S, Bankier AT, Barrell BG, De Bruijn MHL, Coulson AR, Drouin J, et al. Sequence and organization of the human mitochondrial genome. Nature. 1981;290(5806).

7. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the cambridge reference sequence for human mitochondrial DNA [5]. Vol. 23, Nature Genetics. 1999.

8. van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat [Internet]. 2009 [cited 2021 Jan 11];30(2):E386-94. Available from: https://pubmed.ncbi.nlm.nih.gov/18853457/

9. Calvo SE, Mootha VK. The mitochondrial proteome and human disease. Vol. 11, Annual Review of Genomics and Human Genetics. 2010.

10. Soares P, Ermini L, Thomson N, Mormina M, Rito T, Röhl A, et al. Correcting

for purifying selection: an improved human mitochondrial molecular clock. Am J Hum Genet [Internet]. 2009 Jun 12 [cited 2021 Jan 11];84(6):740–59. Available from: https://pubmed.ncbi.nlm.nih.gov/19500773/

11. Wallace DC. Mitochondrial DNA in aging and disease. Sci Am. 1997;277(2).

12. Hatefi Y. The mitochondrial electron transport and oxidative phosphorylation system. Annu Rev Biochem. 1985;54(1).

13. Zapico SC, Ubelaker DH. mtDNA mutations and their role in aging, diseases and forensic sciences. Vol. 4, Aging and Disease. 2013.

14. Rossi A, Pizzo P, Filadi R. Calcium, mitochondria and cell metabolism: A functional triangle in bioenergetics. Vol. 1866, Biochimica et Biophysica Acta - Molecular Cell Research. 2019.

15. Patron M, Granatiero V, Espino J, Rizzuto R, De Stefani D. MICU3 is a tissue-specific enhancer of mitochondrial calcium uptake. Cell Death Differ. 2019;26(1).

16. Srinivasan S, Guha M, Kashina A, Avadhani NG, Biophys B, Author A. Mitochondrial Dysfunction and Mitochondrial Dynamics-The Cancer Connection Graphical abstract HHS Public Access Author manuscript. Biochim Biophys Acta. 2017;1858(8).

17. Glancy B, Willis WT, Chess DJ, Balaban RS. Effect of calcium on the oxidative phosphorylation cascade in skeletal muscle mitochondria. Biochemistry. 2013;52(16).

18. Greaves LC, Reeve AK, Taylor RW, Turnbull DM. Mitochondrial DNA and disease. J Pathol. 2012;226(2):274–86.

19. Fliss MS, Usadel H, Caballero OL, Wu L, Buta MR, Eleff SM, et al. Facile detection of mitochondrial DNA mutations in tumors and bodily fluids. Science (80-). 2000;287(5460).

20. Jerónimo C, Nomoto S, Caballero OL, Usadel H, Henrique R, Varzim G, et al. Mitochondrial mutations in early stage prostate cancer and bodily fluids. Oncogene. 2001;20(37).

21. Triska P, Kaneva K, Merkurjev D, Sohail N, Falk MJ, Triche TJ, et al. Landscape of germline and somatic mitochondrial DNA mutations in pediatric malignancies. Cancer Res. 2019;79(7).

22. Lund M, Melbye M, Diaz LJ, Duno M, Wohlfahrt J, Vissing J. Mitochondrial dysfunction and risk of cancer. Br J Cancer. 2015;112(6).

23. Petros JA, Baumann AK, Ruiz-Pesini E, Amin MB, Sun CQ, Hall J, et al. MtDNA mutations increase tumorigenicity in prostate cancer. Proc Natl Acad Sci U S A. 2005;102(3).

24. Park JS, Sharma LK, Li H, Xiang R, Holstein D, Wu J, et al. A heteroplasmic, not homoplasmic, mitochondrial DNA mutation promotes tumorigenesis via alteration in reactive oxygen species generation and apoptosis. Hum Mol Genet. 2009;18(9).

25. He I, Luo L, Proctor SJ, Middleton PG, Blakely EL, Taylor RW, et al. Somatic mitochondrial DNA mutations in adult-onset leukaemia. Leukemia. 2003;17(12).

26. Brand MD, Chien LF, Ainscow EK, Rolfe DFS, Porter RK. The causes and functions of mitochondrial proton leak. Vol. 1187, BBA - Bioenergetics. 1994.

27. Lettieri-Barbato D. Redox control of non-shivering thermogenesis. Vol. 25, Molecular Metabolism. 2019.

28. Armani C, Landini Jr. L, Leone A. Molecular and Biochemical Changes of the Cardiovascular System due to Smoking Exposure. Curr Pharm Des. 2009;15(10).

29. Ballinger SW, Van Houten B, Conklin CA, Jin GF, Godley BF. Hydrogen

peroxide causes significant mitochondrial DNA damage in human RPE cells. Exp Eye Res. 1999;68(6).

30. Yakes FM, Van Houten B. Mitochondrial DNA damage is more extensive and persists longer than nuclear DNA damage in human cells following oxidative stress. Proc Natl Acad Sci U S A. 1997;94(2).

31. Suematsu N, Tsutsui H, Wen J, Kang D, Ikeuchi M, Ide T, et al. Oxidative stress mediates tumor necrosis factor-α-induced mitochondrial DNA damage and dysfunction in cardiac myocytes. Circulation. 2003;107(10).

32. Luo S, Valencia CA, Zhang J, Lee NC, Slone J, Gui B, et al. Biparental inheritance of mitochondrial DNA in humans. Proc Natl Acad Sci U S A. 2018;115(51).

33. Schwartz M, Vissing J. Paternal Inheritance of Mitochondrial DNA. N Engl J Med. 2002;347(8).

34. Wison AC, Cann RL, Carr SM, George M, Gyllensten UB, Helm-Bychowski KM, et al. Mitochondrial DNA and two perspectives on evolutionary genetics. Biol J Linn Soc. 1985;26(4).

35. Pikó L, Taylor KD. Amounts of mitochondrial DNA and abundance of some mitochondrial gene transcripts in early mouse embryos. Dev Biol. 1987;123(2).

36. Díez-Sánchez C, Ruiz-Pesini E, Lapeña AC, Montoya J, Pérez-Martos A, Enríquez JA, et al. Mitochondrial DNA content of human spermatozoa. Biol Reprod. 2003;68(1).

37. Cummins JM, Wakayama T, Yanagimachi R. Fate of microinjected sperm components in the mouse oocyte and embryo. Zygote. 1997;5(4).

38. Birky CW. Uniparental inheritance of mitochondrial and chloroplast genes: Mechanisms and evolution. Vol. 92, Proceedings of the National Academy of Sciences of the United States of America. 1995.

39. Galtier N, Nabholz B, GlÉmin S, Hurst GDD. Mitochondrial DNA as a marker of molecular diversity: A reappraisal. Vol. 18, Molecular Ecology. 2009.

40. Hagelberg E. Recombination or mutation rate heterogeneity? Implications for mitochondrial eve. Trends Genet. 2003;19(2).

41. Eyre-Walker A. Do mitochondria recombine in humans? Philos Trans R Soc B Biol Sci. 2000;355(1403).

42. Saville BJ, Kohli Y, Anderson JB. mtDNA recombination in a natural population. Proc Natl Acad Sci U S A. 1998;95(3).

43. Elson JL, Andrews RM, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Analysis of European mtDNAs for recombination. Am J Hum Genet. 2001;68(1).

44. Hagelberg E, Goldman N, Lió P, Whelan S, Schiefenhövel W, Clegg JB, et al. Evidence for mitochondrial DNA recombination in a human population of island Melanesia. Proc R Soc B Biol Sci. 1999;266(1418).

45. Eyre-Walker A, Smith NH, Maynard Smith J. How clonal are human mitochondria? Proc R Soc B Biol Sci. 1999;266(1418).

46. Awadalla P, Eyre-Walker A, Smith JM. Linkage disequilibrium and recombination in hominid mitochondrial DNA. Science (80-). 1999;286(5449).

47. Ingman M, Kaessmannē H, Paabo S, Gyllensten U. Mitochondrial genome variation and the origin of modern humans. Nature. 1999;408:708–13.

48. Finnilä S, Lehtonen MS, Majamaa K. Phylogenetic network for european mtDNA. Am J Hum Genet. 2001;68(6).

49. Herrnstadt C, Elson JL, Fahy E, Preston G, Turnbull DM, Anderson C, et al. Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. Am J Hum Genet. 2002;70(5).

50. Kivisild T, Villems R. Questioning evidence for recombination in human mitochondrial DNA. Science (80-). 2000;288(5473).

51. Piganeau G, Eyre-Walker A. A reanalysis of the indirect evidence for recombination in human mitochondrial DNA. Heredity (Edinb). 2004;92(4).

52. Wallace DC, Ye J, Neckelmann SN, Singh G, Webster KA, Greenberg BD. Sequence analysis of cDNAs for the human and bovine ATP synthase β subunit: mitochondrial DNA genes sustain seventeen times more mutations. Curr Genet. 1987;12(2).

53. Bogenhagen DF. Mitochondrial DNA nucleoid structure. Vol. 1819, Biochimica et Biophysica Acta - Gene Regulatory Mechanisms. 2012.

54. Radzvilavicius AL, Hadjivasiliou Z, Pomiankowski A, Lane N. Selection for Mitochondrial Quality Drives Evolution of the Germline. PLoS Biol. 2016;14(12).

55. Scally A, Durbin R, Shriver MD, Mei R, Parra EJ, Sonpar V, et al. Revising the human mutation rate: implications for understanding human evolutionLarge-scale SNP analysis reveals clustered and continuous patterns of human genetic variation. Nat Rev Genet. 2012;13(10).

56. Sanderson S, Green A, Preece MA, Burton H. The incidence of inherited metabolic disorders in the West Midlands, UK. Arch Dis Child. 2006;91(11).

57. Schaefer AM, Taylor RW, Turnbull DM, Chinnery PF. The epidemiology of mitochondrial disorders - Past, present and future. In: Biochimica et Biophysica Acta - Bioenergetics. 2004.

58. da Fonseca RR, Johnson WE, O'Brien SJ, Ramos MJ, Antunes A. The adaptive evolution of the mammalian mitochondrial genome. BMC Genomics. 2008;9.

59. James JE, Piganeau G, Eyre-Walker A. The rate of adaptive evolution in animal mitochondria. Vol. 25, Molecular Ecology. 2016.

60. Yang Z, Nielsen R. Mutation-selection models of codon substitution and their use to estimate selective strengths on codon usage. Mol Biol Evol. 2008;25(3).

61. Stewart JB, Freyer C, Elson JL, Wredenberg A, Cansu Z, Trifunovic A, et al. Strong purifying selection in transmission of mammalian mitochondrial DNA. PLoS Biol. 2008;6(1).

62. Hill JH, Chen Z, Xu H. Selective propagation of functional mitochondrial DNA during oogenesis restricts the transmission of a deleterious mitochondrial variant. Nat Genet. 2014;46(4).

63. Burr SP, Pezet M, Chinnery PF. Mitochondrial DNA Heteroplasmy and Purifying Selection in the Mammalian Female Germ Line. Vol. 60, Development Growth and Differentiation. 2018.

64. Pesole G, Gissi C, De Chirico A, Saccone C. Nucleotide substitution rate of mammalian mitochondrial genomes. J Mol Evol. 1999;48(4).

65. Naue J, Hörer S, Sänger T, Strobl C, Hatzer-Grubwieser P, Parson W, et al. Evidence for frequent and tissue-specific sequence heteroplasmy in human mitochondrial DNA. Mitochondrion. 2015;20(1).

66. Miller FJ, Rosenfeldt FL, Zhang C, Linnane AW, Nagley P. Precise determination of mitochondrial DNA copy number in human skeletal and cardiac muscle by a PCR-based assay: lack of change of copy number with age. Nucleic Acids Res. 2003;31(11).

67. Shuster RC, Rubenstein AJ, Wallace DC. Mitochondrial DNA in anucleate human blood cells. Biochem Biophys Res Commun. 1988;155(3).

68. Liang R, Menon V, Qiu J, Arif T, Renuse S, Lin M, et al. Mitochondrial localization and moderated activity are key to murine erythroid enucleation. Blood Adv. 2021;5(10).

69. Bendall KE, Macaulay VA, Baker JR, Sykes BC. Heteroplasmic point

mutations in the human mtDNA control region. Am J Hum Genet. 1996;59(6).

70. Tully LA, Parsons TJ, Steighner RJ, Holland MM, Marino MA, Prenger VL. A sensitive denaturing gradient-gel electrophoresis assay reveals a high frequency of heteroplasmy in hypervariable region 1 of the human mtDNA control region. Am J Hum Genet. 2000;67(2).

71. Tagliabracci A, Turchi C. mtDNA exploitation in forensics. In: The Human Mitochondrial Genome: From Basic Biology to Disease. 2020.

72. He Y, Wu J, Dressman DC, Iacobuzio-Donahue C, Markowitz SD, Velculescu VE, et al. Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. Nature. 2010;464(7288).

73. Krjutškov K, Koltšina M, Grand K, Võsa U, Sauk M, Tõnisson N, et al. Tissue-specific mitochondrial heteroplasmy at position 16,093 within the same individual. Curr Genet. 2014;60(1).

74. Avital G, Buchshtav M, Zhidkov I, Tuval J, Dadon S, Rubin E, et al. Mitochondrial DNA heteroplasmy in diabetes and normal adults: Role of acquired and inherited mutational patterns in twins. Hum Mol Genet. 2012;21(19).

75. Payne BAI, Wilson IJ, Yu-Wai-Man P, Coxhead J, Deehan D, Horvath R, et al. Universal heteroplasmy of human mitochondrial DNA. Hum Mol Genet. 2013;22(2).

76. DeSalle R, Schierwater B, Hadrys H. MtDNA: The small workhorse of evolutionary studies. Vol. 22, Frontiers in Bioscience - Landmark. 2017.

77. Schurr TG. The peopling of the New World: Perspectives from molecular anthropology. Annu Rev Anthropol. 2004;33.

78. Budowle B, Allard MW, Wilson MR, Chakraborty R. Forensics and Mitochondrial DNA: Applications, Debates, and Foundations. Vol. 4, Annual

Review of Genomics and Human Genetics. 2003.

79. Yonova-Doing E, Calabrese C, Gomez-Duran A, Schon K, Wei W, Karthikeyan S, et al. An atlas of mitochondrial DNA genotype–phenotype associations in the UK Biobank. Nat Genet. 2021;53(7).

80. Cox CB, Healey IN, Moore PD. Biogeography: An Ecological and Evolutionary Approach. Syst Bot. 1977;2(3):106.

81. Butler J. Mitochondrial DNA analysis. In: Advanced Topics in Forensic DNA Typing: Methodology. San Diego, CA: Elsevier Academic Press:; 2012. p. 405–56.

82. Torroni A, Sukernik RI, Schurr TG, Starikovskaya YB, Cabell MF, Crawford MH, et al. mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. Am J Hum Genet. 1993;53(3).

83. Schurr TG, Ballinger SW, Gan YY, Hodge JA, Merriwether DA, Lawrence DN, et al. Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. Am J Hum Genet. 1990;46(3).

84. Richards MB, Macaulay VA, Bandelt H-J, Sykes BC. Phylogeography of mitochondrial DNA in western Europe. Ann Hum Genet. 1998;62(3):241–60.

85. Torroni A, Lott MT, Cabell MF, Chen YS, Lavergne L, Wallace DC. mtDNA and the origin of Caucasians: Identification of ancient Caucasian- specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. Am J Hum Genet. 1994;55(4).

86. Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, et al. Tracing european founder lineages in the near eastern mtDNA pool. Am J Hum Genet. 2000;67(5):1251–76.

87. Young HY, Song I, Ha E, Cho SB, Yang WI, Shin KJ. mtDNAmanager: A

Web-based tool for the management and quality analysis of mitochondrial DNA control-region sequences. BMC Bioinformatics. 2008;9.

88. Behar DM, Rosset S, Blue-Smith J, Balanovsky O, Tzur S, Comas D, et al. The genographic project public participation mitochondrial DNA database. PLoS Genet. 2007;3(6).

89. Prieto L, Zimmermann B, Goios A, Rodriguez-Monge A, Paneto GG, Alves C, et al. The GHEP-EMPOP collaboration on mtDNA population data - A new resource for forensic casework. Forensic Sci Int Genet. 2011;5(2).

90. Gaffney ES. An Introduction to the Logic of Phylogeny Reconstruction. In: Phylogenetic Analysis and Paleontology. 2019.

91. RL Cann, M Stoneking and AW. Mitochondrial DNA and human evolution. Nature. 1987;325(6099):31–6.

92. Maca-Meyer N, González AM, Larruga JM, Flores C, Cabrera VM. Major genomic mitochondrial lineages delineate early human expansions. BMC Genet. 2001;2(1):1–8.

93. Stoneking M, Sherry ST, Redd AJ, Vigilant L. New approaches to dating suggest a recent age for the human mtDNA ancestor. Philos Trans R Soc Lond B Biol Sci. 1992;337(1280).

94. Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, et al. Most of the extant mtDNA boundaries in South and Southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans (BMC Genetics). BMC Genet. 2004;5(1):1–25.

95. 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 [Internet]. 2012 Apr 6 [cited 2021 Jan 11];90(4):675–84. Available from: /pmc/articles/PMC3322232/?report=abstract

96. Torroni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, et al. Classification of european mtDNAs from an analysis of three European populations. Genetics. 1996;144(4):1835–50.

97. Hill C, Soares P, Mormina M, Macaulay V, Meehan W, Blackburn J, et al. Phylogeography and ethnogenesis of aboriginal Southeast Asians. Mol Biol Evol. 2006;23(12).

98. Friedlaender J, Schurr T, Gentz F, Koki G, Friedlaender F, Horvat G, et al. Expanding Southwest Pacific mitochondrial haplogroups P and Q. Mol Biol Evol. 2005;22(6).

99. Roostalu U, Kutuev I, Loogväli EL, Metspalu E, Tambets K, Reidla M, et al. Origin and expansion of haplogroup H, the dominant human mitochondrial DNA lineage in west Eurasia: The Near Eastern and Caucasian perspective. Mol Biol Evol. 2007;24(2).

100. Hernández CL, Dugoujon JM, Novelletto A, Rodríguez JN, Cuesta P, Calderón R. The distribution of mitochondrial DNA haplogroup H in southern Iberia indicates ancient human genetic exchanges along the western edge of the Mediterranean. BMC Genet. 2017;18(1).

101. Bandelt HJ, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol [Internet]. 1999 [cited 2021 Jan 11];16(1):37–48. Available from: https://pubmed.ncbi.nlm.nih.gov/10331250/

102. Kong S, Sánchez-Pacheco SJ, Murphy RW. On the use of median-joining networks in evolutionary biology. Cladistics. 2016;32(6):691–9.

103. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol [Internet]. 2015 Jan 1 [cited 2021 Jan 11];32(1):268–74. Available from: https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msu300

104. Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018;35(2):518–22.

105. Hervik A, Tretvik T, Øvstedal L. Norway: Crossing Fjords and Mountains. In Springer, Dordrecht; 1993 [cited 2021 Jan 11]. p. 349–65. Available from: https://link.springer.com/chapter/10.1007/978-94-015-8118-9_20

106. Glørstad H, Gundersen J, Kvalø F, Nymoen P, Simpson D, Skar B. Norway: submerged stone age from a norwegian perspective. In: Coastal Research Library [Internet]. Springer; 2020 [cited 2021 Jan 11]. p. 125–40. Available from: https://doi.org/10.1007/978-3-030-37367-2_6

107. Stroeven AP, Hättestrand C, Kleman J, Heyman J, Fabel D, Fredin O, et al. Deglaciation of Fennoscandia. Quat Sci Rev. 2016 Sep 1;147:91–121.

108. Vistad OI, Wold LC, Daugstad K, Haukeland JV. Mimisbrunnr Climate Park – A network for heritage learning, tourism development, and climate consciousness. J Herit Tour [Internet]. 2016 Jan 1 [cited 2021 Jan 10];11(1):43–57. Available from: https://www.tandfonline.com/doi/abs/10.1080/1743873X.2015.1082570

109. Günther T, Malmström H, Svensson EM, Omrak A, Sánchez-Quinto F, Kılınç GM, et al. Population genomics of Mesolithic Scandinavia: Investigating early postglacial migration routes and high-latitude adaptation. Barton N, editor. PLOS Biol [Internet]. 2018 Jan 9 [cited 2021 Jan 11];16(1):e2003703. Available from: https://dx.plos.org/10.1371/journal.pbio.2003703

110. Bang-Andersen S. Southwest Norway at the Pleistocene/Holocene Transition: Landscape Development, Colonization, Site Types, Settlement Patterns. Nor Archaeol Rev [Internet]. 2003 [cited 2021 Jan 11];36(1):5–25. Available from: https://www.tandfonline.com/doi/abs/10.1080/00293650307293

111. Malmer MP. The Neolithic of south Sweden: the Funnel Beaker culture (TRB),

the Pitted Ware culture (GRK) and the Battle Axe culture (STR). In: Royal Academy of Letters & Antiquities. Stockholm, Sweden: Akademibokhandelsgruppen AB; 2002. p. 5–284.

112. Malmström H, Linderholm A, Skoglund P, Storå J, Sjödin P, Gilbert MTP, et al. Ancient mitochondrial DNA from the northern fringe of the Neolithic farming expansion in Europe sheds light on the dispersion process. Philos Trans R Soc B Biol Sci [Internet]. 2015 Jan 19 [cited 2021 Feb 11];370(1660):20130373. Available from: http://dx.doi.org/10.1098/rstb.2013.0373

113. Skoglund P, Malmström H, Omrak A, Raghavan M, Valdiosera C, Günther T, et al. Genomic diversity and admixture differs for stone-age Scandinavian foragers and farmers. Science (80-) [Internet]. 2014 May 16 [cited 2021 Feb 11];344(6185):747–50. Available from: https://science.sciencemag.org/content/344/6185/747

114. Skoglund P, Malmström H, Raghavan M, Storå J, Hall P, Willerslev E, et al. Origins and genetic legacy of neolithic farmers and hunter-gatherers in Europe. Science (80-) [Internet]. 2012 Apr 27 [cited 2021 Jan 25];336(6080):466–9. Available from: https://science.sciencemag.org/content/336/6080/466

115. Bell-Fialkoff A. The Vikings. In: The Role of Migration in the History of the Eurasian Steppe. New York: Palgrave Macmillan; 2000. p. 151–79.

116. Baug I, Skre D, Heldal T, Jansen ØJ. The Beginning of the Viking Age in the West. J Marit Archaeol. 2019;14(1).

117. Berglund B, Briksson K, Holm I, Karlsson H, Karlsson J, Pettersson S, et al. The Historical Archaeology of the Medieval Crisis in Scandinavia. Curr Swedish Archaeol. 2021;17(1).

118. Myrdal J. [Digerdöden, pestvågor och ödeläggelse.]. Stockholm: Sällskapet Runica et Mediævalia; 2003. 166–189 p.

119. Thorvaldsen G. Internal migration in 19th and 20th century Norway. An overview 1865 to 1960. . In: Nominative Data in Demographic Research in the East and the West: monograph [Internet]. Издательство Уральского университета. Publishing house of the Ural University.; 2019 [cited 2021 Jan 12]. p. 166–84. Available from: https://www.norgeshistorie.no/industrialisering-og-demokrati/kommunikasjon-

120. Svalestuen AA. Om den Regionale Spreiinga av Norsk Utvandring før 1865. In: Engen A, editor. Utvandringa-Det Store Opbrotet. Oslo: Det Norske Samlaget; 1978. p. 77.

121. Skjekkeland M. Dialektar i Noreg–Tradisjon og Fornying [Internet]. Kristiansand: Høyskoleforlaget.; 2005 [cited 2021 Jan 11]. Available from: https://scholar.google.com/scholar?hl=no&as_sdt=0%2C5&q=Skjekkeland%2C+M.+%282005%29.+Dialektar+i+Noreg–Tradisjon+og+Fornying.+Kristiansand%3A+Høyskoleforlaget.&btnG=

122. Venås K, Skjekkeland M. dialekter i Norge – inndeling – Store norske leksikon [Internet]. 2020 [cited 2021 Jan 12]. Available from: https://snl.no/dialekter_i_Norge_-_inndeling

123. Brotherton P, Haak W, Templeton J, Brandt G, Soubrier J, Jane Adler C, et al. Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. Nat Commun. 2013;4.

124. Loogväli EL, Roostalu U, Malyarchuk BA, Derenko M V., Kivisild T, Metspalu E, et al. Disuniting uniformity: A pied cladistic canvas of mtDNA haplogroup H in Eurasia. Mol Biol Evol. 2004;21(11).

125. Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, Scozzari R, et al. The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. Am J Hum Genet. 2004;75(5).

126. Ochir-Goryaeva MA, Kornienko I V., Faleeva TG, Aramova OY, Makhotkin MA, Kekeev EA, et al. Ancestry and identity in Bronze Age Catacomb culture burials: A meta-tale of graves, skeletons, and DNA. J Archaeol Sci Reports. 2021;37.

127. Krzewińska M. Human origins and migrations in Norway inferred from ancient and modern DNA analysis. Published PhD thesis. [Oslo]: University of Oslo; 2014.

128. Opdal SHS, Rognum TOT, Vege Å, Stave AKA, Dupuy BMB, Egeland T. Increased number of substitutions in the D-loop of mitochondrial DNA in the sudden infant death syndrome. Acta Paediatr Int J Paediatr [Internet]. 1998 Jan 2 [cited 2021 Jan 11];87(10):1039–44. Available from: http://doi.wiley.com/10.1111/j.1651-2227.1998.tb01410.x

129. Passarino G, Cavalleri GL, Lin AA, Cavalli-Sforza LL, Børresen-Dale AL, Underhill PA. Different genetic components in the Norwegian population revealed by the analysis of mtDNA and Y chromosome polymorphisms. Eur J Hum Genet [Internet]. 1998 Aug 13 [cited 2021 Jan 11];10(9):521–9. Available from: http://lgb.unige.ch/

130. Helgason A, Hickey E, Goodacre S, Bosnes V, Stefánsson K, Ward R, et al. mtDNA and the Islands of the North Atlantic: Estimating the proportions of Norse and Gaelic ancestry. Am J Hum Genet. 2001 Mar 1;68(3):723–37.

131. Sayers EW, Cavanaugh M, Clark K, Ostell J, Pruitt KD, Karsch-Mizrachi I. GenBank. Nucleic Acids Res. 2019;47(D1).

132. The Norway DNA Project Group. FamilyTreeDNA - The Norway DNA - Norge Project [Internet]. FamilyTreeDNA. 2014 [cited 2020 Jun 11]. Available from: https://www.familytreedna.com/group-join.aspx?Group=Norway

133. National Archives of Norway. Arkivverket [Internet]. [cited 2020 Jun 1]. Available from: https://www.arkivverket.no/

134. YFull. MTree 1.02.16171 [Internet]. 2021 [cited 2021 Dec 4]. Available from: https://www.yfull.com/mtree/J/

135. 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 [Internet]. 2016 Jul 8 [cited 2021 Jan 11];44(W1):W58–63. Available from: https://pubmed.ncbi.nlm.nih.gov/27084951/

136. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol. 2013 Apr;30(4):772–80.

137. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 2003;52(5):696–704.

138. Darriba D, Taboada GL, Doallo R, Posada D. JModelTest 2: More models, new heuristics and parallel computing. Nat Methods. 2012;9(8):772.

139. Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst Biol [Internet]. 2006 Aug 1 [cited 2021 Jun 24];55(4):539–52. Available from: https://academic.oup.com/sysbio/article/55/4/539/1675125

140. Minh BQ, Nguyen MAT, Von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol [Internet]. 2013 May 1 [cited 2021 Jun 24];30(5):1188–95. Available from: https://academic.oup.com/mbe/article/30/5/1188/997508

141. Corander J, Marttinen P, Sirén J, Tang J. Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. BMC Bioinformatics. 2008;9.

142. Tonkin-Hill G, Lees JA, Bentley SD, Frost SDW, Corander J. RhierBAPs: An R implementation of the population clustering algorithm hierbaps. Wellcome

Open Res [Internet]. 2018 [cited 2021 Jun 21];3(93). Available from: /pmc/articles/PMC6178908/

143. Rambaut A, Lam TT, Carvalho LM, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol. 2016;2(1).

144. To TH, Jung M, Lycett S, Gascuel O. Fast Dating Using Least-Squares Criteria and Algorithms. Syst Biol. 2016;65(1):82–97.

145. Margaryan A, Lawson DJ, Sikora M, Racimo F, Rasmussen S, Moltke I, et al. Population genomics of the Viking world. Nature [Internet]. 2020 Sep 17 [cited 2021 Feb 1];585(7825):390–6. Available from: https://www.nature.com/articles/s41586-020-2688-8

146. FamilyTreeDNA. FamilyTreeDNA - The U5 Project. FamilyTreeDNA. 2021.

147. RStudio Team. RStudio: Integrated Development for R. [Internet]. Boston, MA: RStudio, PBC; 2020 [cited 2021 Oct 25]. Available from: http://www.rstudio.com/.

148. Excoffier L, Lischer HEL. Arlequin suite version 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour. 2010 May;10(3):564–7.

149. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2018.

150. Librado P, Rozas J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009 Jun;25(11):1451–2.

151. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989;123(3).

152. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10.

Virus Evol [Internet]. 2018 Jan 1 [cited 2021 Jan 12];4(1):vey016. Available from: https://academic.oup.com/ve/article/doi/10.1093/ve/vey016/5035211

153. Soares P, Alshamali F, Pereira JB, Fernandes V, Silva NM, Afonso C, et al. The expansion of mtDNA haplogroup L3 within and out of Africa [Internet]. Vol. 29, Molecular Biology and Evolution. Mol Biol Evol; 2012 [cited 2021 Jan 12]. p. 915–27. Available from: https://pubmed.ncbi.nlm.nih.gov/22096215/

154. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol [Internet]. 2005 May 1 [cited 2021 Jan 12];22(5):1185–92. Available from: https://pubmed.ncbi.nlm.nih.gov/15703244/

155. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst Biol [Internet]. 2018 [cited 2021 Jan 12];67(5):901–4. Available from: /pmc/articles/PMC6101584/?report=abstract

156. Picornell A, Gómez-Barbeito L, Tomàs C, Castro JA, Ramon MM. Mitochondrial DNA HVRI variation in Balearic populations. Am J Phys Anthropol. 2005;128(1).

157. Pfeiffer H, Brinkmann B, Hühne J, Rolf B, Morris AA, Steighner R, et al. Expanding the forensic German mitochondrial DNa control region database: Genetic diversity as a function of sample size and microgeography. Int J Legal Med. 1999;112(5).

158. Derenko M V., Grzybowski T, Malyarchuk BA, Dambueva IK, Denisova GA, Czarny J, et al. Diversity of mitochondrial DNA lineages in South Siberia. Ann Hum Genet. 2003;67(5).

159. Di Rienzo A, Wilson AC. Branching pattern in the evolutionary tree for human mitochondrial DNA. Proc Natl Acad Sci U S A. 1991;88(5).

160. Schönberg A, Theunert C, Li M, Stoneking M, Nasidze I. High-throughput sequencing of complete human mtDNA genomes from the Caucasus and West Asia: High diversity and demographic inferences. Eur J Hum Genet. 2011;19(9).

161. Messina F, Scorrano G, Labarga CM, Rolfo MF, Rickards O. Mitochondrial DNA variation in an isolated area of Central Italy. Ann Hum Biol. 2010;37(3).

162. Bybjerg-Grauholm J, Hagen CM, Gonçalves VF, Bækvad-Hansen M, Hansen CS, Hedley PL, et al. Complex spatio-temporal distribution and genomic ancestry of mitochondrial DNA haplogroups in 24,216 danes. PLoS One. 2018;13(12):e0208829.

163. Tillmar AO, Coble MD, Wallerström T, Holmlund G. Homogeneity in mitochondrial DNA control region sequences in Swedish subpopulations. Int J Legal Med. 2010;124(2).

164. Sajantila A, Salem AH, Savolainen P, Bauer K, Gierig C, Pääbo S. Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population. Proc Natl Acad Sci U S A. 1996;93(21):12035–9.

165. Tambets K, Rootsi S, Kivisild T, Help H, Serk P, Loogväli EL, et al. The western and eastern roots of the Saami - the story of genetic “outliers” told by mitochondrial DNA and Y chromosomes. Am J Hum Genet. 2004 Apr 1;74(4):661–82.

166. Hedman M, Brandstätter A, Pimenoff V, Sistonen P, Palo JU, Parson W, et al. Finnish mitochondrial DNA HVS-I and HVS-II population data. Forensic Sci Int. 2007;172(2–3).

167. Meinilä M, Finnilä S, Majamaa K. Evidence for mtDNA admixture between the Finns and the Saami. Hum Hered. 2001;52(3):160–70.

168. Sajantila A, Lahermo P, Anttinen T, Lukka M, Sistonen P, Savontaus ML, et al. Genes and languages in Europe: An analysis of mitochondrial lineages.

Genome Res. 1995 Aug 1;5(1):42–52.

169. Richards M, Corte-Real H, Forster P, Macaulay V, Wilkinson-Herbots H, Demaine A, et al. Paleolithic and neolithic lineages in the European mitochondrial gene pool. Am J Hum Genet. 1996;59(1):185.

170. Kittles RA, Bergen AW, Urbanek M, Virkkunen M, Linnoila M, Goldman D, et al. Autosomal, mitochondrial, and Y chromosome DNA variation in Finland: Evidence for a male-specific bottleneck. Am J Phys Anthropol. 1999;108(4):381–99.

171. Pult I, Sajantila A, Simanainen J, Georgiev O, Schaffner W, Paabo S. Mitochondrial DNA sequences from Switzerland reveal striking homogeneity of European populations. Biol Chem Hoppe Seyler. 1994;375(12).

172. Lahermo P, Sajantila A, Sistonen P, Lukka M, Aula P, Peltonen L, et al. The genetic relationship between the Finns and the Finnish Saami (Lapps): Analysis of nuclear DNA and mtDNA. Am J Hum Genet [Internet]. 1996 [cited 2021 Feb 11];58(6):1309–22. Available from: /pmc/articles/PMC1915079/?report=abstract

173. Richard C, Pennarun E, Kivisild T, Tambets K, Tolk HV, Metspalu E, et al. An mtDNA perspective of French genetic variation. Ann Hum Biol [Internet]. 2007 Jan [cited 2021 Feb 9];34(1):68–79. Available from: https://www.tandfonline.com/doi/abs/10.1080/03014460601076098

174. Simoni L, Calafell F, Pettener D, Bertranpetit J, Barbujani G. Geographic patterns of mtDNA diversity in Europe. Am J Hum Genet. 2000;66(1):262–78.

175. Josefsson T, Ramqvist PH, Hörnberg G. The history of early cereal cultivation in northernmost Fennoscandia as indicated by palynological research. Veg Hist Archaeobot. 2014;23(6).

176. Hjelle KL, Hufthammer AK, Bergsvik KA. Hesitant hunters: A review of the introduction of agriculture in western Norway. Environ Archaeol [Internet].

2006 [cited 2021 Jan 11];11(2):147–70. Available from: https://www.tandfonline.com/doi/abs/10.1179/174963106x123188

177. Malyarchuk B, Derenko M, Grzybowski T, Lunkina A, Czarny J, Rychkov S, et al. Differentiation of mitochondrial DNA and Y chromosomes in Russian populations. Hum Biol. 2004;76(6):877–900.

178. Tvauri A. Migrants or Natives? The Research History of Long Barrows in Russia and Estonia in the 5th–10th Centuries. 32nd ed. Nuorluoto J, editor. Vol. 32, Slavica Helsingiensia. Helsinki: University of Helsinki; 2007. 247–285 p.

179. Dupuy BM, Olaisen B. mtDNA sequences in the Norwegian Saami and main populations. In: Carracedo A., Brinkmann B. BW, editor. Advances in Forensic Haemogenetics [Internet]. 1st ed. Berlin, Heidelberg: Springer; 1996 [cited 2021 Feb 4]. p. 23–5. Available from: https://link.springer.com/chapter/10.1007/978-3-642-80029-0_6

180. Delghandi M, Utsi E, Krauss S. Saami mitochondrial DNA reveals deep maternal lineage clusters. Hum Hered. 1998;48(2).

181. Krzewińska M, Bjørnstad G, Skoglund P, Olason PI, Bill J, Götherström A, et al. Mitochondrial DNA variation in the Viking age population of Norway. Philos Trans R Soc B Biol Sci [Internet]. 2015 Jan 19 [cited 2021 Jan 11];370(1660):20130384. Available from: https://royalsocietypublishing.org/doi/10.1098/rstb.2013.0384

182. Malmström H, Günther T, Svensson EM, Juras A, Fraser M, Munters AR, et al. The genomic ancestry of the Scandinavian Battle Axe Culture people and their relation to the broader Corded Ware horizon. Proc R Soc B Biol Sci. 2019;286(1912):20191528.

183. Gron KJ, Montgomery J, Nielsen PO, Nowell GM, Peterkin JL, Sørensen L, et al. Strontium isotope evidence of early Funnel Beaker Culture movement of

cattle. J Archaeol Sci Reports. 2016;6.

184. Kristjansson D, Bohlin J, Jugessur A, Schurr TG. Matrilineal diversity and population history of Norwegians. Am J Phys Anthropol [Internet]. 2021;176(1):120–33. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/ajpa.24345

185. Bramanti B, Thomas MG, Haak W, Unterlaender M, Jores P, Tambets K, et al. Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. Science (80-). 2009;326(5949):137–41.

186. Röhl A, Brinkmann B, Forster L, Forster P. An annotated mtDNA database. Int J Legal Med. 2001;115(1).

187. Hughes ALC, Gyllencreutz R, Lohne ØS, Mangerud J, Svendsen JI. The last Eurasian ice sheets - a chronological database and time-slice reconstruction, DATED-1. Boreas. 2016;45(1).

188. Dolukhanov PM. Modern Humans' Expansion in Eurasia: One Flew East. Open Anthropol J. 2008;1(1):26–32.

189. Stroeven AP, Heyman J, Fabel D, Björck S, Caffee MW, Fredin O, et al. A new Scandinavian reference 10Be production rate. Quat Geochronol. 2015;29.

190. Posth C, Renaud G, Mittnik A, Drucker DG, Rougier H, Cupillard C, et al. Pleistocene mitochondrial genomes suggest a single major dispersal of non-africans and a late glacial population turnover in Europe. Curr Biol [Internet]. 2016 Mar 21 [cited 2021 Jan 29];26(6):827–33. Available from: http://dx.doi.org/10.1016/j.cub.2016.01.037

191. Malyarchuk B, Derenko M, Grzybowski T, Perkova M, Rogalla U, Vanecek T, et al. The peopling of Europe from the mitochondrial haplogroup U5 perspective. PLoS One [Internet]. 2010 [cited 2021 Jul 1];5(4):10285. Available from: /pmc/articles/PMC2858207/

192. Aarnes I, Kühl N, Birks HH. Quantitative climate reconstruction from late-glacial and early Holocene plant macrofossils in western Norway using the probability density function approach. Rev Palaeobot Palynol. 2012;170.

193. Eide W, Birks HH, Bigelow NH, Peglar SM, Birks HJB. Holocene forest development along the Setesdal valley, southern Norway, reconstructed from macrofossil and pollen evidence. Veg Hist Archaeobot. 2006;15(2).

194. Breivik HM. Palaeo-oceanographic development and human adaptive strategies in the Pleistocene–Holocene transition: A study from the Norwegian coast. Holocene. 2014;24(11).

195. Solheim S, Persson P. Early and mid-Holocene coastal settlement and demography in southeastern Norway: Comparing distribution of radiocarbon dates and shoreline-dated sites, 8500–2000 cal. BCE. J Archaeol Sci Reports. 2018;19.

196. Pala M, Olivieri A, Achilli A, Accetturo M, Metspalu E, Reidla M, et al. Mitochondrial DNA signals of late glacial recolonization of europe from near eastern refugia. Am J Hum Genet. 2012;90(5).

197. Diamond J, Bellwood P. Farmers and their languages: The first expansions. Vol. 300, Science. 2003.

198. Brandt G, Haak W, Adler CJ, Roth C, Szécsényi-Nagy A, Karimnia S, et al. Ancient DNA reveals key stages in the formation of Central European mitochondrial genetic diversity. Science (80-). 2013;342(6155):257–61.

199. Haak W, Forster P, Bramanti B, … SM-, 2005 U. Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. science.sciencemag.org [Internet]. 2005 [cited 2021 Jul 2];310(5750):1016–8. Available from: https://science.sciencemag.org/content/310/5750/1016.abstract

200. Fischer A. Food for Feasting: An evaluation of explanations of the neolithisation of Denmark and southern Sweden. In: Fischer A, Kristiansen K,

editors. The neolithisation of Denmark: 150 years of debate. Sheffield: JR. Collis Publications; 2002. p. 343–93.

201. Iversen R, Kroonen G. Talking neolithic: Linguistic and archaeological perspectives on how Indo-European was implemented in Southern Scandinavia. Am J Archaeol. 2017;121(4):511–25.

202. Sørensen L, Karg S. The expansion of agrarian societies towards the north - new evidence for agriculture during the Mesolithic/Neolithic transition in Southern Scandinavia. J Archaeol Sci. 2014;51:98–114.

203. Zohary D, Hopf M, Weiss E. Plant remains at representative archaeological sites. 4th editio. Zohary D, Hopf M, Weiss E, editors. Domestication of Plants in the Old World: The origin and spread of cultivated plants in West Asia, Europe and the Nile Valley. New York: Oxford University Press; 2012. 189–90 p.

204. Goodacre S, Helgason A, Nicholson J, Southam L, Ferguson L, Hickey E, et al. Genetic evidence for a family-based Scandinavian settlement of Shetland and Orkney during the Viking periods. Heredity (Edinb). 2005;95(2):129–35.

205. Downham C. Viking Ethnicities: A Historiographic Overview. Hist Compass. 2012;10(1):1–2.

206. Price N. The Viking World. In: Brink S, Price N, editors. The Viking World. 1st ed. New York: Routledge; 2008. p. 257–71.

207. Halsall G. The Viking presence in England? The burial evidence reconsidered. In: Hadley D, Richards J, editors. Cultures in Contact Scandinavian Settlement in England in the Ninth and Tenth Centuries. First edit. Turnhout: Brepols; 2000. p. 259–76.

208. Androshchuk F. [En man i Osebergsgraven?]. Fornvannen. 2005;100:115–28.

209. Lund J. Connectedness with things. Animated objects of Viking Age

Scandinavia and early medieval Europe. Archaeol Dialogues. 2017;24(1):89–108.

210. Lund J. Thresholds and Passages: The Meanings of Bridges and Crossings in the Viking Age and Early Middle Ages. Viking Mediev Scand. 2005;1:109–35.

211. Lund J. Banks, Borders and Bodies of Water in a Viking Age Mentality. J Wetl Archaeol. 2008;8(1):53–72.

212. Ingman M, Gyllensten U. A recent genetic link between Sami and the Volga-Ural region of Russia. Eur J Hum Genet [Internet]. 2007 Jan 20 [cited 2021 Feb 13];15(1):115–20. Available from: www.fluxus-engineering.com

213. Tambets K, Yunusbayev B, Hudjashov G, Ilumäe AM, Rootsi S, Honkola T, et al. Genes reveal traces of common recent demographic history for most of the Uralic-speaking populations. Genome Biol. 2018;19(1):1–20.

214. Mielnik-Sikorska M, Daca P, Malyarchuk B, Derenko M, Skonieczna K, Perkova M, et al. The History of Slavs Inferred from Complete Mitochondrial Genome Sequences. Pereira LMSM, editor. PLoS One [Internet]. 2013 Jan 14 [cited 2021 Feb 13];8(1):e54360. Available from: https://dx.plos.org/10.1371/journal.pone.0054360

215. Bang-Andersen S. Colonizing contrasting landscapes.: The pioneer coast settlement and inland utilization in southern Norway 10,000-9500 years before present. Oxford J Archaeol [Internet]. 2012 May 1 [cited 2021 Feb 1];31(2):103–20. Available from: http://doi.wiley.com/10.1111/j.1468-0092.2012.00381.x

216. Helskog K. Boats and meaning: A study of change and continuity in the Alta fjord, arctic Norway, from 4200 to 500 years B.C. J Anthropol Archaeol. 1985 Sep 1;4(3):177–205.

217. Bjerck HB. Settlements and Seafaring: Reflections on the Integration of Boats and Settlements Among Marine Foragers in Early Mesolithic Norway and the

Yámana of Tierra del Fuego. J Isl Coast Archaeol. 2017;12(2).

218. Seppä H, Bjune AE, Telford RJ, Birks HJB, Veski S. Last nine-thousand years of temperature variability in Northern Europe. Clim Past. 2009;5(3).

219. Kullman L. Further Details on Holocene Treeline, Glacier/Ice Patch and Climate History in Swedish Lapland. Int J Res Geogr. 2017;3(4):61–9.

220. Price TD. The introduction of farming in northern Europe. In: Price TD, editor. Europe's first farmers. First. Cambridge, UK: Cambridge University Press; 2000. p. 260–300.

221. Prescott C. Was there really a Neolithic in Norway? Antiquity. 1996;70(267).

222. Ballard C, Bradley R, Myhre LN, Wilson M. The ship as symbol in the prehistory of Scandinavia and southeast Asia. World Archaeol [Internet]. 2004 [cited 2021 Jan 12];35(3):385–403. Available from: https://www.tandfonline.com/doi/abs/10.1080/0043824042000185784

223. Bjerck H. On the outer fringe of the human world: pheno- menological perspectives on anthropomorphic cave paintings in Norway. In: Bergsvik KA, Skeates R, editors. Caves in Context: the cultural significance of caves and rockshelters in Europe. Oxford and Oxbow: Oxbow Books; 2012. p. 48–64.

224. Ljunge M. Capturing images: knowledge, ownership and the materiality of cave art. In: Klevnäs A, Hedenstierna-Jonson C, editors. Own or be owned: Archaeological approaches to the concept of possession. Stockholm: Stockholm University; 2015. p. 131–40.

225. Pilø LH, Barrett JH, Eiken T, Finstad E, Grønning S, Post-Melbye JR, et al. Interpreting archaeological site-formation processes at a mountain ice patch: A case study from Langfonne, Norway. The Holocene [Internet]. 2020 Nov 25 [cited 2021 Jan 12];095968362097277. Available from: http://journals.sagepub.com/doi/10.1177/0959683620972775

226. Lindqvist C, Possnert G. The subsistence economy and diet at Jakobs/Ajvide and Stora Förvar, Eksta Parish and other prehistoric dwelling and burial sites on Gotland in long-term perspective. In: Burenhult G, editor. Remote Sensing, Theses and Papers in North-European Archaeology. vol 1. Hasselholm: Department of Archaeology, Stockholm University; 1997. p. 29–90.

227. Østmo E. When the Norsemen learned to row: a technological innovation for shipping in the early Iron Age. Viking Nor Archaeol Yearb. 2003;66:7–29.

228. Haaland A, Svihus Å. Coastal and maritime Norway. [Internet]. Art Council Norway, The Norwegian Coastal Administration, The Directorate for Cultural Heritage and Directory of Fisheries; 2011. p. 12. Available from: https://ra.brage.unit.no/ra-xmlui/bitstream/handle/11250/176922/Fortellinger_kystNorge_Kyst_og_havlandet_eng.pdf?sequence=1

229. Fischer LR, Nordvik HW. From namsos to halden: Myths and realities in the history of norwegian seamen's wages, 1850-1914. Scand Econ Hist Rev [Internet]. 1987 Jan 1;35(1):41–64. Available from: https://www.tandfonline.com/action/journalInformation?journalCode=sehr20

230. Bagge S. Nationalism in Norway in the middle ages. Scand J Hist [Internet]. 1995 [cited 2021 Feb 4];20(1):1–18. Available from: https://www.tandfonline.com/doi/abs/10.1080/03468759508579290

231. Larsen K. A History of Norway. 2nd ed. Princeton: Princeton University Press; 1950. 80 p.

232. Kleppe JI. Chapter 10: Desolate landscapes or shifting landscapes? Late glacial/early post-glacial settlement of northernmost Norway in the light of new data from Eastern Finnmark. In book: Lateglacial and Postglacial Pioneers in Northern Europe. BAR Intern. Riede F, Tallavaara M, editors. Archaeopress; 2014. 121–145 p.

233. Wickler S, Narmo LE. Tracing the Development of Fishing Settlement From the Iron Age to the Modern Period in Northern Norway: A Case Study From Borgvær in the Lofoten Islands. J Isl Coast Archaeol [Internet]. 2014 [cited 2021 Jan 12];9(1):72–87. Available from: https://www.tandfonline.com/doi/abs/10.1080/15564894.2013.810678

234. Knutsen H. Norwegian agriculture: status and trends 2019. 6(8); 2020.

235. Pryser T. På flyttefot: Innanlands vandring på 1800-talet. In: Gjerdåker B, editor. Oslo: Det Norske Samlaget; 1981. p. 59–69.

236. Dyrvik S. Historical demography in Norway 1660-1801: A short survey. Scand Econ Hist Rev [Internet]. 1972 Jan [cited 2021 Feb 3];20(1):27–44. Available from: http://www.tandfonline.com/doi/abs/10.1080/03585522.1972.10407709

237. Sølvi S. Folkevekst og flytting: en historisk-demografisk studie i 1700-årenes Øst-Norge [Internet]. Oslo: Universitetsforlaget; 1979. 93 p. Available from: http://urn.nb.no/URN:NBN:no-nb_digibok_2009062901084

238. Niemi E, Myhre JE, Kjeldstadli K. I nasjonalstatens tid 1814-1940. In: Kjeldstadli K, editor. Norwegian Immigration History. 2nd ed. Oslo; 2003.

239. Jentoft S, Finstad BP. Building fisheries institutions through collective action in Norway. Marit Stud [Internet]. 2018 Apr 1 [cited 2021 Feb 1];17(1):13–25. Available from: https://doi.org/10.1007/s40152-018-0088-6

240. Dupuy BM, Stenersen M, Lu TT, Olaisen B. Geographical heterogeneity of Y-chromosomal lineages in Norway. Forensic Sci Int. 2006 Dec 1;164(1):10–9.

241. Opsahl E. Avaldsnes' Position in Norway in the 14th Century. In: Skre D, editor. Rulership in 1st to 14th century Scandinavia Royal graves and sites at Avaldsnes and beyond. Berlin/Boston: De Gryuter; 2020. p. 517–29.

242. Ahren M. Indigenous Peoples' Culture, Customs, and Traditions and Customary Law - The Saami People's Perspective. Ariz J Int Comp Law

[Internet]. 2004 [cited 2021 Feb 4];21. Available from: https://heinonline.org/HOL/Page?handle=hein.journals/ajicl21&id=81&div=&collection=

243. Mattingsdal M, Ebenesersdóttir SS, Moore KHS, Andreassen OA, Hansen TF, Werge T, et al. The genetic structure of Norway. Eur J Hum Genet. 2021;29(11).

244. Gjerde J. From peasants to farmers: The migration from Balestrand, Norway, to the Upper Middle West. . Cambridge: Cambridge University Press; 1989. 7–15 p.

245. Tryland M. Kopper og koppevirus – 200 år siden første vaksinasjon i Norge. Tidsskr Den Nor legeforening [Internet]. 2001 Dec 10 [cited 2021 Feb 15]; Available from: https://tidsskriftet.no/2001/12/medisinsk-historie/kopper-og-koppevirus-200-ar-siden-forste-vaksinasjon-i-norge

246. Achilli A, Rengo C, Battaglia V, Pala M, Olivieri A, Fornarino S, et al. Saami and Berbers - An unexpected mitochondrial DNA link. Am J Hum Genet [Internet]. 2005 [cited 2021 Jul 1];76(5):883–6. Available from: https://pubmed.ncbi.nlm.nih.gov/15791543/

247. Burkhardt M. The German Hanse and Bergen -new perspectives on an old subject. Scand Econ Hist Rev. 2010;58(1):60–79.

248. Helle K. Bergen's role in the medieval North Atlantic trade. AmS-Skrifter. 2019;(27):43–51.

249. Kent HSK. The Anglo-Norwegian Timber Trade in the Eighteenth Century. Econ Hist Rev. 1955;8(1):62–74.

250. Lloyd TH. England and the German Hanse, 1157-1611: a study of their trade and commercial diplomacy. Paperback. Cambridge, UK: Cambridge University Press; 1991. 40–53 p.

251. Gardiner M, Mehler N. Introduction: German trade in the North Atlantic. AmS-Skrifter. 2020;(27).

252. Bull I. Integration of Immigrating Merchants in Trondheim in the Seventeenth and Eighteenth Century. In: Paper to the 7th European Urban History Association Conference. : Athens-Piraeu; 2004. p. 1–6.

253. Bull I. Merchant households and their networks in eighteenth-century Trondheim. Contin Chang. 2002;17(2).

254. World Bank. Gini index: All countries and economies [Internet]. The World Bank Group. 2021 [cited 2022 Jun 10]. Available from: https://data.worldbank.org/indicator/SI.POV.GINI?most_recent_value_desc=true

255. Samuels DC, Carothers AD, Horton R, Chinnery PF. The power to detect disease associations with mitochondrial DNA haplogroups. Am J Hum Genet. 2006;78(4).

256. Kozin MS, Kulakova OG, Kiselev IS, Balanovsky OP, Boyko AN, Favorova OO. Variants of mitochondrial genome and risk of multiple sclerosis development in Russians. Acta Naturae. 2018;10(4).

257. Preste R, Vitale O, Clima R, Gasparre G, Attimonelli M. Hmtvar: A new resource for human mitochondrial variations and pathogenicity data. Nucleic Acids Res. 2019;47(D1).

258. Golubenko M V., Salakhov RR, Makeeva OA, Goncharova IA, Kashtalap V V., Barbarash OL, et al. Association of mitochondrial DNA polymorphism with myocardial infarction and prognostic signs for atherosclerosis. Mol Biol. 2015;49(6).

259. Montiel-Sosa F, Ruiz-Pesini E, Enríquez JA, Marcuello A, Díez-Sánchez C, Montoya J, et al. Differences of sperm motility in mitochondrial DNA haplogroup U sublineages. Gene. 2006;368(1–2).

260. Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC. Effects of Purifying and Adaptive Selection on Regional Variation in Human mtDNA. Science (80-) [Internet]. 2004 Jan 9 [cited 2021 Jan 12];303(5655):223–6. Available from: https://pubmed.ncbi.nlm.nih.gov/14716012/

261. Pereira L, Gonçalves J, Franco-Duarte R, Silva J, Rocha T, Arnold C, et al. No evidence for an mtDNA Role in sperm motility: Data from complete sequencing of asthenozoospermic males. Mol Biol Evol. 2007;24(3).

262. Majamaa K, Turkka J, Kärppä M, Winqvist S, Hassinen IE. The common MELAS mutation A3243G in mitochondrial DNA among young patients with an occipital brain infarct. Neurology. 1997;49(5).

263. Finnilä S, Hassinen IE, Ala-Kokko L, Majamaa K. Phylogenetic network of the mtDNA haplogroup U in northern Finland based on sequence analysis of the complete coding region by conformation-sensitive gel electrophoresis. Am J Hum Genet. 2000;66(3).

264. Healy TM, Burton RS. Strong selective effects of mitochondrial DNA on the nuclear genome. Proc Natl Acad Sci U S A. 2020;117(12).

265. Laan M, Pääbo S. Demographic history and linkage disequilibrium in human populations. Nat Genet. 1997;17(4).

266. Kaessmann H, Zöllner S, Gustafsson AC, Wiebe V, Laan M, Lundeberg J, et al. Extensive linkage disequilibrium in small human populations in Eurasia. Am J Hum Genet. 2002;70(3).

267. Ross AB, Johansson Å, Ingman M, Gyllensten U. Lifestyle, genetics, and disease in Sami [Internet]. Vol. 47, Croatian Medical Journal. Medicinska Naklada; 2006 [cited 2021 Jul 5]. p. 553–65. Available from: www.cmj.hr

268. Johansson Å, Vavruch-Nilsson V, Edin-Liljegren A, Sjölander P, Gyllensten U. Linkage disequilibrium between microsatellite markers in the Swedish Sami relative to a worldwide selection of populations. Hum Genet. 2005;116(1–2).

269. Eidemiller KY, Geht AB, Samylovskaya EA, Kulik S V., Krylova EA. History and prospects of the Saami issue. In: IOP Conference Series: Earth and Environmental Science. 2020. p. 012004.

270. Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, et al. Natural selection shaped regional mtDNA variation in humans. Proc Natl Acad Sci U S A. 2003;100(1):171–6.

271. Di Rago JP, Colson AM. Molecular basis for resistance to antimycin and diuron, Q-cycle inhibitors acting at the Q(i) site in the mitochondrial ubiquinol-cytochrome c reductase in Saccharomyces cerevisiae. J Biol Chem. 1988;263(25):12564–70.

272. Fisher N, Rich PR. A motif for quinone binding sites in respiratory and photosynthetic systems. J Mol Biol. 2000;296(4):1153–62.

273. Hudson G, Nalls M, Evans JR, Breen DP, Winder-Rhodes S, Morrison KE, et al. Two-stage association study and meta-analysis of mitochondrial DNA variants in Parkinson disease. Neurology. 2013;80(22).

274. Niemi AK, Hervonen A, Hurme M, Karhunen PJ, Jylhä M, Majamaa K. Mitochondrial DNA polymorphisms associated with longevity in a Finnish population. Hum Genet. 2003;112(1).

275. De Benedictis G, Rose G, Carrieri G, De Luca M, Falcone E, Passarino G, et al. Mitochondrial DNA inherited variants are associated with successful aging and longevity in humans. FASEB J. 1999;13(12).

276. Domínguez-Garrido E, Martínez-Redondo D, Martín-Ruiz C, Gómez-Durán A, Ruiz-Pesini E, Madero P, et al. Association of mitochondrial haplogroup J and mtDNA oxidative damage in two different North Spain elderly populations. Biogerontology. 2009;10(4).

277. Rose G, Passarino G, Carrieri G, Altomare K, Greco V, Bertolini S, et al. Paradoxes in longevity: Sequence analysis of mtDNA haplogroup J in

centenarians. Eur J Hum Genet. 2001;9(9).

278. Nordic Burden of Disease Collaborators. Life expectancy and disease burden in the Nordic countries: results from the Global Burden of Diseases, Injuries, and Risk Factors Study 2017. Lancet Public Heal. 2019;4(12).

279. Dato S, Passarino G, Rose G, Altomare K, Bellizzi D, Mari V, et al. Association of the mitochondrial DNA haplogroup J with longevity is population specific. Eur J Hum Genet. 2004;12(12).

280. Shlush LI, Atzmon G, Weisshof R, Behar D, Yudkovsky G, Barzilai N, et al. Ashkenazi Jewish centenarians do not demonstrate enrichment in mitochondrial haplogroup J. PLoS One. 2008;3(10).

281. Ren WH, Li XH, Zhang HG, Deng FM, Liao WQ, Pang Y, et al. Mitochondrial DNA haplogroups in a Chinese Uygur population and their potential association with longevity. Clin Exp Pharmacol Physiol. 2008;35(12).

282. Ross OA, McCormack R, Curran MD, Alistair Duguid R, Barnett YA, Maeve Rea I, et al. Mitochondrial DNA polymorphism: Its role in longevity of the Irish population. Exp Gerontol. 2001;36(7).

283. Pinós T, Nogales-Gadea G, Ruiz JR, Rodríguez-Romo G, Santiago-Dorrego C, Fiuza-Luces C, et al. Are mitochondrial haplogroups associated with extreme longevity? A study on a Spanish cohort. Age (Omaha). 2012;34(1).

284. Costa MD, Cherni L, Fernandes V, Freitas F, Ammar el Gaaied A Ben, Pereira L. Data from complete mtDNA sequencing of Tunisian centenarians: Testing haplogroup association and the “golden mean” to longevity. Mech Ageing Dev. 2009;130(4).

285. Capri M, Salvioli S, Sevini F, Valensin S, Celani L, Monti D, et al. The genetics of human longevity. In: Annals of the New York Academy of Sciences. 2006.

286. Perls T, Terry D. Genetics of exceptional longevity. In: Experimental Gerontology. 2003.

287. Niemi AK, Moilanen JS, Tanaka M, Hervonen A, Hurme M, Lehtimäki T, et al. A combination of three common inherited mitochondrial DNA polymorphisms promotes longevity in Finnish and Japanese subjects. Eur J Hum Genet. 2005;13(2).

288. Børte S, Zwart JA, Skogholt AH, Gabrielsen ME, Thomas LF, Fritsche LG, et al. Mitochondrial genome-wide association study of migraine - the HUNT Study. Cephalalgia. 2020;40(6).

289. Miller B, Torres M, Jiang X, McKean-Cowdin R, Nousome D, Kim SJ, et al. A mitochondrial genome-wide association study of cataract in a latino population. Transl Vis Sci Technol. 2020;9(6).

290. Klunk J, Duggan AT, Redfern R, Gamble J, Boldsen JL, Golding GB, et al. Genetic resiliency and the Black Death: No apparent loss of mitogenomic diversity due to the Black Death in medieval London and Denmark. Am J Phys Anthropol. 2019;169(2):240–52.

Paper I

Received: 16 February 2021 | Revised: 19 May 2021 | Accepted: 23 May 2021

DOI: 10.1002/ajpa.24345

RESEARCH ARTICLE

American Journal of PHYSICAL ANTHROPOLOGY | WILEY

Matrilineal diversity and population history of Norwegians

Dana Kristjansson[1,2,3] | Jon Bohlin[2,4] | Astanand Jugessur[1,2,3] | Theodore G. Schurr[5]

[1]Department of Genetics and Bioinformatics, Norwegian Institute of Public Health, Oslo, Norway

[2]Center of Fertility and Health, Norwegian Institute of Public Health, Oslo, Norway

[3]Department of Global Public Health and Primary Care, Faculty of Medicine, University of Bergen, Bergen, Norway

[4]Department of Infection Epidemiology and Modelling, Norwegian Institute of Public Health, Oslo, Norway

[5]Department of Anthropology, University of Pennsylvania, Philadelphia, Pennsylvania, USA

Correspondence
Dana Kristjansson, The Norwegian Institute of Public Health, P.O. Box 222 Skøyen, 0213 Oslo, Norway, Norway.
Email: dana.kristjansson@fhi.no

Funding information
Norges Forskningsråd, Grant/Award Number: 287961

Abstract

Background: While well known for its Viking past, Norway's population history and the influences that have shaped its genetic diversity are less well understood. This is particularly true with respect to its demography, migration patterns, and dialectal regions, despite there being curated historical records for the past several centuries. In this study, we undertook an analysis of mitochondrial DNA (mtDNA) diversity within the country to elaborate this history from a matrilineal genetic perspective.

Methods: We aggregated 1174 partial modern Norwegian mtDNA sequences from the published literature and subjected them to detailed statistical and phylogenetic analysis by dialectal regions and localities. We further contextualized the matrilineal ancestry of modern Norwegians with data from Mesolithic, Iron Age, and historic period populations.

Results: Modern Norwegian mtDNAs fell into eight West Eurasian (N, HV, JT, I, U, K, X, W), five East Eurasian (A, F, G, N11, Z), and one African (L2) haplogroups. Pairwise analysis of molecular variance (AMOVA) estimates for all Norwegians indicated they were differentiated from each other at 1.68% ($p < 0.001$). Norwegians within the same dialectal region also showed genetic similarities to each other, although differences between subpopulations within dialectal regions were also observed. In addition, certain mtDNA lineages in modern Norwegians were also found among prehistoric and historic period populations, suggesting some level of genetic continuity over hundreds to many thousands of years.

Conclusions: This analysis of mtDNA diversity provides a detailed picture of the genetic variation within Norway in light of its topography, settlement history, and historical migrations over the past several centuries.

KEYWORDS
haplogroup, haplotype, language, mtDNA, Norway

1 | INTRODUCTION

Norway lies on the western edge of the Scandinavian Peninsula, with the majority of the country surrounded by water. Norway is also the longest country in Scandinavia, stretching 1752 km in length, a distance equivalent to the length of the Netherlands to the center of Italy (Hervik et al., 1993). Taking into consideration fjords and islands, the coastline is ~83,000 km in total length (Grabbe et al., 2009).

Topographically, Norway consists of largely mountainous terrain in the innermost parts of most of the country that is broken up into valleys and fjords running through the coastal areas. The highest mountain ranges begin at the horizontal level of Stavanger in the southwest and extend up to Trondheim in central Norway. This stretch includes the Jotunheimen mountain range, home to Norway's highest mountain called Galdhøpiggen at 2469 m (Vistad et al., 2016), which divides the southern and most populated portion of the country into eastern and western halves.

After the Fennoscandian ice sheet covering Norway began to melt during the Last Glacial Maximum ~23,000 years ago, Norway's coastline slowly became inhabited by rich forests and wildlife (Glørstad et al., 2020; Stroeven et al., 2016). Much of the earliest evidence of human settlement, which dates to 10,000 years ago, comes from hunting tools and burial sites found from the southeastern Oslo Fjord all of the way up to Finnmark in the far north (Bang-Andersen, 2003; Günther et al., 2018). These findings suggest that the entire coastline was quickly settled by pioneering groups (Bang-Andersen, 2012), who employed watercraft to move along the long coastline for hunting and fishing (Helskog, 1985).

Based on recent ancient DNA work, Norway was initially settled by hunter-gatherers groups migrating into the country from the both southwest and northeast (Günther et al., 2018). These groups were later partially incorporated into expanding farming cultures (Malmström et al., 2015; Skoglund et al., 2012). A major genetic shift in Norway's population structure occurred after the Viking Age (750–1050 ACE), and has been described in terms of both genomic admixture (Margaryan et al., 2020) and changes in mtDNA haplogroup frequencies (Krzewińska et al., 2015). This shift likely resulted from greater gene flow from the British Isles into modern-day Norway compared with Denmark and Sweden, which, in turn, was due to more frequent maritime routes of the western-dwelling Norse to the British Isles (Margaryan et al., 2020).

These maritime routes continued to shape and influence Norwegian society and culture long after the Viking expansions. Due to its climate, the economic dependence on natural goods varies by region in Norway and requires timely ship-based mechanisms to distribute fresh cargo throughout the rest of the country. The processing and packaging of fish for export also took place either in or near coastal towns along routes of great distance. Extending the length of the western coastline, the Norwegian fisheries industry was dispersed in scattered longitudinal rural settlements rather than more nucleated fishing communities as seen in European countries such as Great Britain, Germany, France, and Denmark (Haaland & Svihus, 2011; Kleppe, 2014; Wickler & Narmo, 2014). By 1850, nearly all Norwegian towns west of Lindesnes in Adger County, Norway's southernmost point, were engaged in the processing and exportation of fish and/or shipping and shipbuilding (Haaland & Svihus, 2011). The majority of the country's 3% arable land in Hedmark and Østfold Counties in the southeast region of Norway was dedicated to growing crops on large farms (Knutsen, 2019). With sparser forests, the central and northern parts of Norway have traditionally relied more on fishing (Knutsen, 2019).

Mass migrations started within Norway around the 1750–1780s (Svalestuen, 1978; Thorvaldsen, 2019) and boomed in the mid-1800s when Norwegians no longer needed official permissions to relocate (Pryser, 1981; Svalestuen, 1978). A large proportion of migrants were young couples looking for better economic opportunities and to expand their families. The most comprehensive and most widely mentioned wave of internal migrants was that from southeastern Norway and southern Trøndelag to the Målselv and Bardu valleys southeast of Tromsø (Thorvaldsen, 2019). Later, between 1750 and 1801, as rural inland farming populations in the southeast became more overpopulated, people moved from the inland districts to the Oslo Fjord for better opportunities in agriculture and timber trade (Dyrvik, 1972; Sølvi, 1979). Similarly, people moved from the inland southeast, west, and central Norway to the northernmost coastlines for cod and herring fishing opportunities (Niemi et al., 2003). Some Norwegians also emigrated after famine and poverty spread in the 19th century. However, the majority could not afford the ship fare and did not want to enter into indentured servitude relocated within the country itself (Sølvi, 1979).

The regional distribution of Norwegians is further reflected in the distinct spoken dialects in those locations, which may have been influenced by maritime routes and internal migration patterns. These dialects vary substantially from each other in terms of grammar, syntax, tone, and pronunciation (Skjekkeland, 2005). The major dialects include Eastern Norwegian (*østnorsk*), Western Norwegain (*vestnorsk*) and Trøndelag in central Norway (*trøndersk*), and Northern Norwegian (*nordnorsk*) (Skjekkeland, 2005; Venås & Skjekkeland, 2020). During Danish rule (1357–1814), a colloquial Norwegian-like Danish was spoken widely by Norwegians who lived in cities and towns. By the first half of the16th century, the Old Norwegian written language had fallen out of favor and, by the 18th century, Danish grammar, pronunciation, and vocabulary were preferred or insisted upon over Norwegian in education and theater (Haugen, 1959). During the less restrictive union between Sweden and Norway (1814–1905), the rise of a unique Norwegian identity led to the return of Norwegian dialects and lexicon back into acceptance (Derry, 2012; Haugen, 1959). As a result of this language history, several elements of the four regional Norwegian dialects have persisted, with several of their distinctive features being preserved (Skjekkeland, 2005).

Mitochondrial DNA (mtDNA) has been a frequently studied source of genetic information for understanding long-term population shifts, diversity, and demographic structure. It is a maternally inherited, nonrecombining DNA molecule at the HVS I region that evolves at a clock-like rate (Soares et al., 2009), making its invaluable for reconstructing populations dynamics based on matrilineal genetic diversity in human populations. Thus, we investigated matrilineal variation in Norway by dialect regions, we attempted to characterize regional mitochondrial DNA (mtDNA) diversity in Norway to determine the influence of dialectal region, subpopulation, and topography on its distribution in the country. We further explored the impact of internal historical migrations on the present-day gene pool.

2 | MATERIALS AND METHODS

2.1 | Mitochondrial DNA data

Data for a total of 1174 partial mtDNA sequences from modern Norwegians were gathered from various published sources, including Krzewińska, 2014; Opdal et al., 1998; Passarino et al., 2002; Helgason et al., 2001. This study also included mtDNA sequences from the GenBank database as of June 1, 2020, that belonged to persons of Norwegian ethnicity and origin. Specific information about the collection location or residence within Norway was available for 64% of the total data set ($n = 755$). Information about these samples, including their GenBank accession numbers and the specific tables within the published sources from which the sequences were obtained, are provided in Table S1.

In addition, partial mtDNA sequences from 1597 non-related individuals with matrilineal Norwegian ancestors were incorporated into this analysis to contextualize the human migration changes within Norway since the early modern ancestors of the 17th to 20th centuries. The mtDNA sequences were obtained on June 1, 2020, from The Norway DNA Project (http://www.norwaydna.no), which is a subproject of the Family Tree DNA database (http://www.familytreedna.com) containing mtDNA sequences submitted by consenting members (The Norway DNA Project Group, 2014). Membership is open to individuals who have a Norwegian background, Norwegian ancestry, or who live in Norway. Location information was derived from individuals submitting their mtDNA data who had also indicated the local region of their earliest known matrilineal ancestor (87%; $n = 1396$), using The Norwegian National Archives (https://www.arkivverket.no/). These archives contain annotated information on Norwegians living from the 17th to the early 20th century that derives from parish records, census records, and village books. For these reasons, this Family Tree DNA data set was used as a proxy for ancestral Norwegian populations, and for this reason its constituent members will be called the "Ancestors."

2.2 | Geographic and dialectic regions of Norway

MtDNA sequences among Modern Norwegians were localized largely to cities, which reflects the current locations of most Norwegians, while mtDNA sequence data among Ancestors were spread throughout towns, villages, and cities within different counties. For comparative purposes, the specific locations were separated into the following geographical regions based on the major dialects of Norway: (1) The *Southeast (Eastern Norwegian dialect) region included Ancestors born or from the settlements of* Hedmark, Oppland, Buskerud, Akershus and Oslo, Telemark, Vestfold, and Østfold Counties. For Modern Norwegians, it compassed the Southeast area of Norway, which largely compasses the capital of Oslo, as well as the above-mentioned neighboring southeastern counties. (2) The *West (Western Norwegian dialect) region included Ancestors born in or from the settlements of* Agder, Rogaland, Hordaland, Sogn and Fjordane (including Førde), and Møre and Romsdal Counties. For Modern Norwegians, it compassed Haugesund (within Rogaland county), Bergen (within Hordaland county), and Førde (within Sogn and Fjordane county). (3) The *Central (Trøndelag Norwegian dialect) region included Ancestors born in or from the settlements of* South-Trøndelag and North-Trøndelag Counties. For Modern Norwegians, this region compassed Trondheim, where samples were taken at St. Olav's Hospital and the population consists of individuals residing in North-Trøndelag County. (4) *The North (Northern Norwegian dialect) region included Ancestors born in or from the settlements of* Nordland, Troms, and Finnmark Counties. For Modern Norwegians, it compassed the same counties.

2.3 | Ethics statement

This study is based on open-access and publicly available data sets. The respective studies from which these data derive have gone through standard protocols to obtain informed consent from participants, clearance, and approval from the respective ethics committees for sample collection and analysis, as outlined in the associated publications.

2.4 | Phylogenetic and statistical analysis

All mtDNA sequences were delimited to reads between nucleotide positions (np) 16,024 to 16,383, that is, the first hypervariable region segment (HVS1) of the mtDNA control region (CR). The HVSI mtDNA sequences were then aligned against the Reconstructed Sapiens Reference Sequence (RSRS), which allows for naming and mapping of the mtDNA haplogroups from an ancestral base (Behar et al., 2012) using MAFFT version 7 (Katoh & Standley, 2013).

Haplogroup classifications of the HSVI sequences were made with HaploGrep2 (Weissensteiner et al., 2016), which uses PhyloTree Build 17 as its reference (van Oven & Kayser, 2009). Haplogrep2 computes the haplogroup classifications on pre-calculated phylogenetic weights that correspond to the occurrence per position in Phylotree Build 17 (http://www.phylotree.org), which, in turn, reflects the mutational stability of a variant.

Haplogrep2 classifications were verified using a maximum likelihood (ML) phylogeny for the unique haplogroups using IQ-TREE 1.6.12 software (http://www.iqtree.org) (Nguyen et al., 2015). The phylogeny was constructed under the general time-reversible nucleotide substitution model with a proportion of invariant sites (TPM3u + F + I + G4) which was inferred in jModelTest (https://github.com/ddarriba/jmodeltest2) as the best fitting model. This tree was then updated to take into account back mutations, (point mutations that revert to the ancestral state) and noncontinuous mutations in accordance with PhyloTree, Build 17 (van Oven & Kayser, 2009). Accordingly, we ignored hot-spot mutations such as cytosine (C) insertions/deletions at position 16,193, and expansions of cytosomes that affect the number of adenosines at positions 16,182 and 16,183, because these common polymorphisms are also excluded from Phylotree, are

not diagnostic for a particular haplogroup assignment, and may lead to inaccuracies in algorithmic predictions of phylogenetic organization.

DnaSP 5.10.01 (Librado & Rozas, 2009) was used to calculate the basic parameters of genetic diversity. The analysis of molecular variance (AMOVA) was carried out using Arlequin 3.5.2.2 (Excoffier & Lischer, 2010) and R (R Core Team, 2018). The statistical significance of fixation indices (F*st*) and their respective p-values was estimated by permutation analysis (10,000 permutations) assuming a Tamura-Nei (1993) model with a gamma distribution of 0.26. Comparisons of haplogroup frequencies between geographic or dialectic regions were conducted using a Chi-square or Fisher's exact test, where appropriate.

The structuring of mtDNA sequence diversity by geography was assessed through the analysis of F*st* values with multidimensional scaling (MDS), using the PAST v.2.17b software (Hammer et al., 2001). All analyses were conducted by location listed in the public data sets. Locations were combined into the four main dialect regions of Norway—North, Central, West, and Southeast (Venås & Skjekkeland, 2020)—for both the Ancestor and Modern Norwegian data sets, as outlined above.

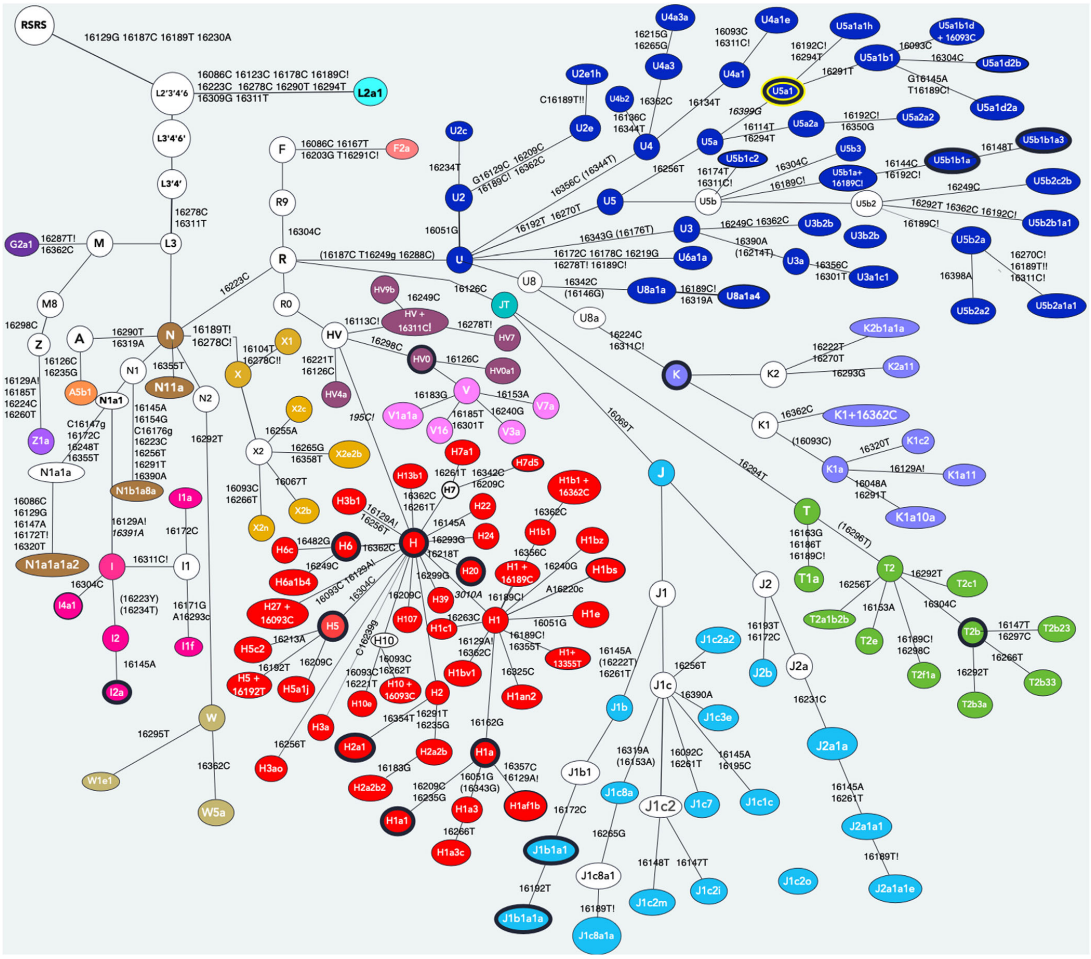

FIGURE 1 Phylogenetic network of maternal lineages represented by mitochondrial DNA. (mtDNA) haplogroups among present-day Norwegians. The white circles indicate haplogroups not found among the Norwegian data. The thick black borders around haplogroups indicate the representation among late Iron Age inhabitants of Norway (Krzewińska et al., 2015). The yellow circle indicates haplogroup representation from Mesolithic inhabitants of Norway, as sequenced by (Günther et al., 2018). Back mutations are indicated with an exclamation mark (!) and two exclamation marks (!!) indicate a double back mutation. Noncontinuous mutations that do not follow through all subsequent haplogroups are indicated in parentheses. Mutations outside of the np16,024–16,383 region that support some regions of the tree are indicated in italics

To estimate population growth over time, we generated a Bayesian skyline plot (BSP) using BEAST 1.10.4, cross-platform program for Bayesian analysis of molecular sequences, (Suchard et al., 2018) with gamma distributed rates (Drummond et al., 2005; Soares et al., 2012; Suchard et al., 2018). Briefly, each Markov chain Monte Carlo (MCMC) sample was based on a run of 10 million generations sampled every 1000 steps, of which the first 1% was discarded to allow for burn-in. The mutation rate of 1.64×10^{-7} (Soares et al., 2009) was used to convert substitution rates into years (x-axis) and coalescent intensities into effective population sizes (y-axis) (Drummond et al., 2005). BEAST outputs were visualized with the Tracer v.1.7.1 program (Rambaut et al., 2018). The BSP analysis was restricted to the database for contemporary Norwegians because the information for the Ancestors' sequences from Family Tree DNA did not specify the individuals' modern familial residence (whether inside or outside Norway).

3 | RESULTS

3.1 | Distribution of Norwegian maternal lineages.

In this study, a total of 1174 Modern Norwegian HVS1 sequences were analyzed. Among them, 151 unique haplogroups were identified based on the Phylotree Build 17 nomenclature (Figure 1). These haplogroups fell into eight West Eurasian (N, HV, JT, I, U, K, X, W), five East Eurasian (A, F, G, N11, Z), and one African (L2) major haplogroups. Overall, 98.4% ($n = 1155$) of Norwegians had a mtDNA that belonged to a West Eurasian maternal lineage.

As can be seen in the haplogroup network displayed in Figure 1, a large number of Modern Norwegian mtDNAs belonged to haplogroups H, J, and U, similar to what has been seen for other northern and western European populations, including the studies on Norwegian mtDNA diversity used in this paper (Helgason et al., 2001; Krzewińska, 2014; Lembring et al., 2013; Li et al., 2014; Passarino et al., 2002). We also noted that the mutational signature of the U5a1 haplotype (16129G, 16187C, 16189T, 16192T, 16223C, 16230A, 16256T, 16270T, 16278C, and 16311T) appearing in this network was also found in the mtDNA of a Mesolithic individual analyzed by (Günther et al., 2018), while sequence motifs for haplogroups HV0, H5, H6, H20, H, H1a, H1a1, H2a1, I2a, J, J1b1a1, J1b1a1a, T2b, U5b1b1a3, and U5b1b1a in contemporary Norwegians were also detected in Iron Age individuals analyzed by Krzewinska and coworkers (Krzewińska et al., 2015).

These results were not unexpected. Recent analysis of the genomic makeup of Viking Age Scandinavians (793–1066 CE) reflected gene flow from other European populations, as well as primarily genetic ancestry from populations preceding the Iron Age (500 BCE–800 CE) (Margaryan et al., 2020). While further analysis is needed to determine whether the observed similarity of the polymorphisms extends further to include whole mitogenomes, these findings suggested that certain haplogroups have been present in the region for many thousands of years.

Of the Modern Norwegian samples, 64% were localized to specific dialectal regions of Norway, whereas 36% could not be localized to any specific region. Nevertheless, the proportion of major haplogroups among those with no locations was similar to the

TABLE 1 Distribution of major mtDNA haplogroups by major dialectal region in Norway among Norwegians

	Southeast $n = 225$		West $n = 257$		Central $n = 268$		North $n = 5$		No location[a] $n = 419$		Total $n = 1174$	
Hg	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
A	1	0.44	0	0	0	0	0	0	0	0	1	0.09
F	0	0	0	0	1	0.37	0	0	0	0	1	0.09
G	1	0.44	0	0	0	0	0	0	0	0	1	0.09
H	113	50.22	119	45.92	94	35.07	1	20	185	44.15	512	43.6
HV	7	3.1	5	1.95	14	5.22	1	20	13	3.11	40	3.42
I	7	3.11	5	1.95	2	0.74	1	20	15	3.58	30	2.57
J	17	7.55	28	10.89	54	20.16	0	0	50	11.93	149	12.69
JT	2	0.89	0	0	0	0	0	0	0	0	2	0.17
K	10	4.43	10	3.9	20	7.46	0	0	22	5.25	62	5.3
L	0	0	0	0	0	0	0	0	1	0.24	1	0.09
N	2	0.88	3	1.17	4	1.5	0	0	3	0.72	12	1.03
T	19	8.44	16	6.23	23	8.57	0	0	41	9.78	99	8.43
U	39	17.33	53	20.63	44	16.78	1	20	67	16	204	17.41
V	2	0.88	2	0.78	8	2.99	1	20	11	2.63	24	2.06
W	3	1.33	9	3.5	1	0.37	0	0	6	1.43	19	1.63
X	1	0.44	3	1.17	2	0.74	0	0	3	0.72	9	0.79
Z	1	0.44	4	1.56	1	0.37	0	0	2	0.48	8	0.68

Note: All mtDNA sequences were analyzed between np 16,024 and 16,383.
[a]Individuals for whom a geographic place of origin was not listed.

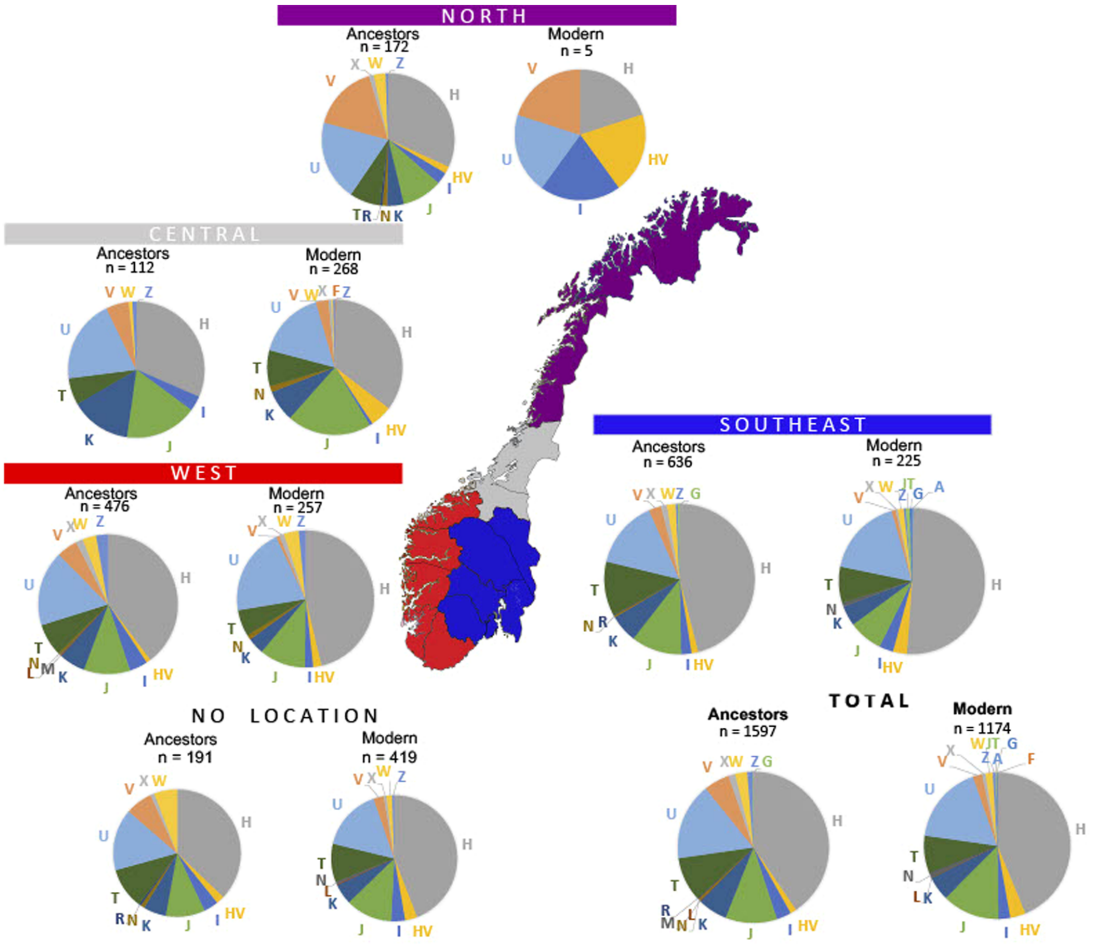

FIGURE 2 Comparison of mtDNA haplogroup distribution among ancestral populations of Norway (17th–20th century) versus the modern population by dialect region. The map of Norway is modified from a public domain map found at: https://en.wikipedia.org/wiki/Counties_of_Norway#/media/File:Nye_fylker_-_regjeringen.no.svg

proportions for Norwegians, overall (Fisher's $p = 0.20$). As a result, the individuals with no location were deemed to be representative of Norway as a whole. Accordingly, the mtDNA gene pool of Norwegians is defined, predominantly, by haplogroups H (44%), U (17%), J (13%), T (8%), K (5%), I (3%), HV (3%), V (2%), and W (2%), while all other haplogroups were present at 1% or less. These frequencies are similar to what was reported in the studies from which these data were obtained (Helgason et al., 2001; Krzewińska, 2014; Passarino et al., 2002).

Haplogroup distribution varied by dialectal region. Haplogroup H was most frequent in the Southeast dialectal region (50%) and least frequent in the Central dialectal region (35%) ($\chi^2 = 14.06$; $p < 0.001$) (Table 1). Haplogroup J appeared at the highest frequency in the Central dialectal region (20%) and the lowest in the Southeast dialectal region (8%) ($\chi^2 = 15.74$; $p < 0.001$). In the Central dialectal region, the haplogroup T occurred at the highest frequency (9%) and the lowest in the West dialectal region (6%), although this difference was not statistically significant ($\chi^2 = 0.33$; $p = 0.56$). Since only five individuals derived from the North dialectal region, the small sample size did not allow for any effective estimate of haplogroup frequencies there. Additional haplogroup detail for these mtDNA sequences are provided in Table S2.

A comparison of the distributional differences among the Modern population compared with the Ancestral population is displayed in Figure 2. Most noticeably, the frequency of haplogroup H in the Ancestral population was higher than in the Southeast dialectal region and lower in the North region ($\chi^2 = 11.31$, $p < 0.001$). Although the difference in the frequency of haplogroup U between the North and

Southeast region was not statistically significant ($\chi^2 = 2.6$, $p = 0.11$), haplogroup U5 appeared at high frequency in the North (for which 92% of the U mtDNAs belong to U5) compared to the Southeast ($\chi^2 = 7.35$, $p = 0.007$) region. The North was distinct from the other regions in the terms of the higher frequency of haplogroup V ($\chi^2 = 11.31$, $p < 0.001$), although haplogroups I and HV occurred at similar frequencies in the Ancestral North compared to the other regions (both $p > 0.05$).

3.2 | Genetic diversity and differentiation.

Overall, the level of genetic variation among Norwegian mtDNA was similar to that seen in other European populations, with about 50% being belonging to haplogroup H and the majority of the rest to other sequences consisting of other West Eurasian lineages described (Côrte-Real et al., 1996; Li et al., 2014; Richard et al., 2007). Summary statistics with the parameters of genetic diversity for Norwegians sorted by dialectal region are shown in Table 2. The HVSI sequences in Norwegians comprised 268 distinct haplotypes defined by 131 variable sites being identified among 1174 individuals. The mean number of nucleotide differences was 3.18, with the Central and West dialectal regions having the highest values (3.52 and 3.16, respectively). Similar values of haplotype and nucleotide diversity among Norwegians were noted in all dialect regions, except the North, for which only five individuals were identified. Each region also had Tajima's D that was consistent with a recent population expansions.

TABLE 2 Diversity indices for Norwegian subpopulations based on mtDNA HVS1 sequences

Population	Samples (*n*)	Haplotypes (h)	Variable sites (s)	Haplotype diversity (hd)	Nucleotide diversity (Pi)	Mean # of nucleotide differences (k)	Tajima D's	*p*-value for Tajima's D
Southeast	225	106	81	0.9271	0.00958	2.971	−2.41	$p < 0.01$
West	257	86	69	0.9192	0.01019	3.159	−2.09	$p < 0.05$
Central	268	81	69	0.9334	0.01134	3.517	−2.05	$p < 0.05$
North	5	4	4	0.9	0.00519	1.6	−1.18	N.S.
No Location	419	142	94	0.9207	0.01004	3.082	−2.27	$p < 0.01$
Total	1174	270	131	0.938	0.01026	3.182	−2.35	$p < 0.001$

Note: "N.S." indicated "not significant."

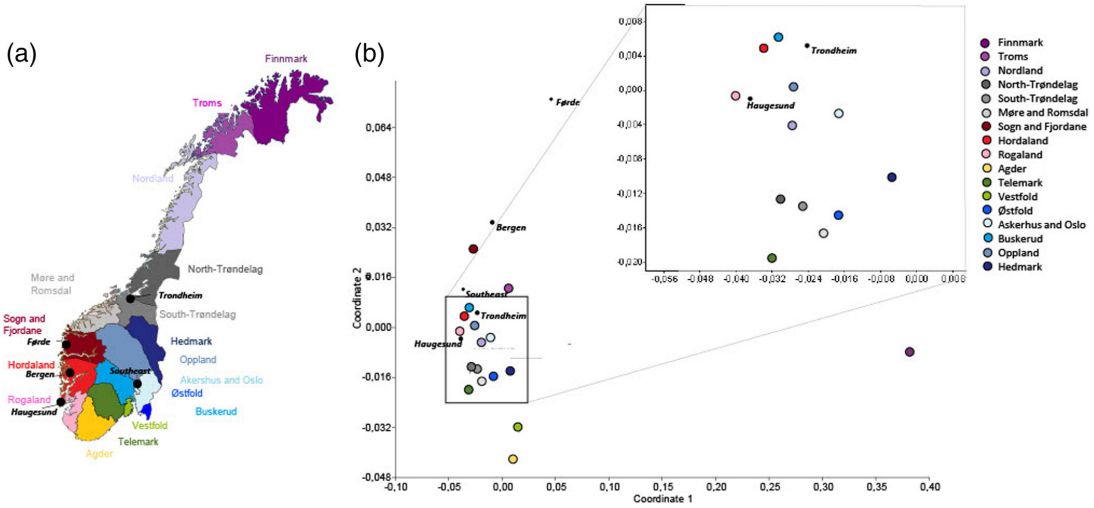

FIGURE 3 Multidimensional scaling plot of mtDNA diversity from specific locations in Norway (a) historical counties of Norway that represent the location of the ancestors' maternal lineages obtained from family tree DNA are in colored dots. Locations in black italics represent areas from which mtDNA data have been obtained among contemporary individuals. (b) A MDS plot of inter-population pairwise *Fst* values calculated from mtDNA HVSI sequences, with a magnified view of the center coordinates. Coordinates 1 and 2 for the modern populations are as follows: Møre and Romsdal (0.55, −0.12), Nordland (−0.13, 0.21), and Finnmark (0.46, 0.55). These points have not been included as the small sample sizes are not representative of mtDNA diversity in these regions and skews the overall plot to present these regions as false outliers

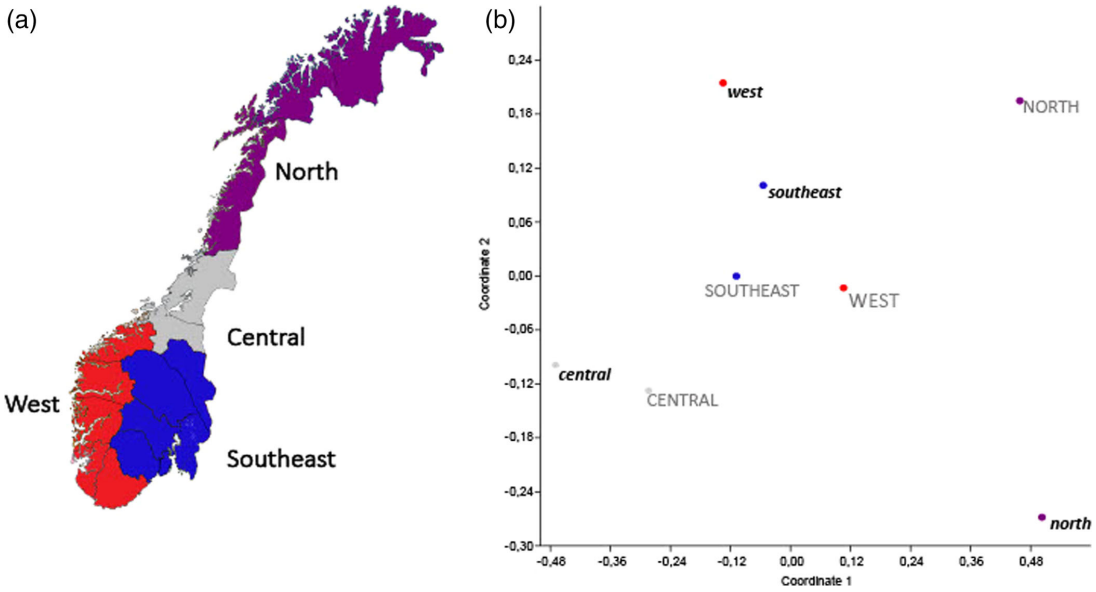

FIGURE 4 Influence of linguistic subdivisions on mtDNA diversity in Norway. (a) Black line divisions indicate the four general dialectal regions of Norway (north, central, southeast, and west). (b) MDS plot of inter-population pairwise *Fst* values calculated from mtDNA HVSI sequence data based on the four dialectal regions of Norway. mtDNA of descendants of historical regions (the ancestors) (capitalized gray font) and modern individuals and their current location in Norway (italicized black font)

3.3 | Inter-population relationships by locality and dialectal region

Using mtDNA data from Norwegians who had submitted information about the geographic location of their earliest matrilineal ancestor from The Norway DNA Project, we plotted these locations and those for modern Norwegians onto the county map of Norway (Figure 3a). Estimation of pairwise *Fst* values from HVSI sequences by specific Norwegian counties showed similarities between the locations for the Ancestral and Modern individuals (Table S3). When comparing the Fst values for Modern cities were compared to those of the Ancestral counties, the extent of differentiation between them was low: Bergen vs. Hordaland (F*st* = 0.0137; $p = 0.05$); Førde versus Sogn and Fjordane (F*st* = 0.0177, $p = 0.08$); and Haugesund vs. Rogaland (F*st* = 0.0050; $p = 0.95$). The exception was Askerhus and Oslo compared with the modern Southeast (F*st* = 0.0092; $p = 0.03$).

The Ancestral and the Modern Norwegian mtDNA data sets were subjected to AMOVA to determine the influence of geography and language on mtDNA diversity. The results pointed to generally low levels of population differentiation among Norwegians in different subpopulations (F*st* = 1.68%, $p < 0.001$) (Table S4). This analysis also showed that there was modest genetic differentiation by large dialectal regions (0.41%, $p < 0.001$) as well as subpopulation (0.81%; $p < 0.001$). Most of the differences came from Ancestral Finnmark, which had moderate genetic differentiation, followed by Ancestral Agder and Modern Førde (Table S3). In fact, the pattern of mtDNA diversity in Ancestral Finnmark differed significantly from that of all other Norwegian counties and Modern cities, with pairwise F*st* values ranging from 0.05 to 0.1 with *p*-values <0.001.

To visualize the genetic relationships between different Norwegian populations, an MDS plot was constructed using the pairwise F*st* values estimated for each of the Norwegian subpopulations (Figure 3b). The results showed that the majority of Norwegian counties were genetically similar, and that Norwegians with ancestry from the counties of the Southeast and Central regions had similar mtDNA backgrounds.

The Ancestral and Modern Norwegian populations were also partly differentiated to some degree based on the four major dialectal regions of the country (Figure 4a,b; Table S5). In general, the Ancestral populations tended to cluster closely toward the center of the plot, while there were greater levels of separation between Modern populations. Notably, the Southeast Modern and Ancestral populations had the lowest pairwise mtDNA distances relative to the other regions of Norway to each other (F*st* all 0.0050–0.0118) (Table S5). The Modern West population was distinct from the Ancestral West (F*st* = 0.0078; $p < 0.000$). The Modern West population was differentiated from the Ancestral Southeast (F*st* = 0.0119; $p < 0.001$), but had a lower F*st* value relative to the Modern Southeast (F*st* = 0.0052; $p = 0.04$). These findings suggest recent changes in the distribution of maternal lineages over the past few centuries that have made the Modern West less dissimilar to the Modern Southeast. By contrast, the Central Modern and Ancestral populations were not significantly different from each other (F*st* = 0.0064; $p = 0.13$). The North Ancestral population further differed from all other regional

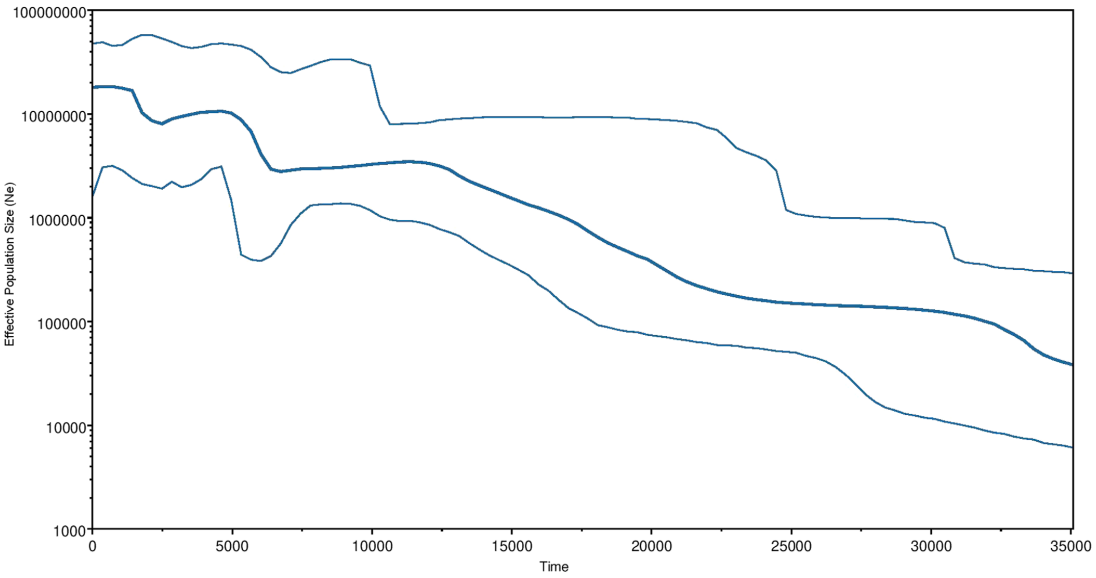

FIGURE 5 A Bayesian skyline plot (BSP) of Norwegian mtDNA sequences. A mutation rate of 1.64×10^{-7} was used to convert substitution rates into years (x-axis) and coalescent intensities into effective population sizes (y-axis)

populations ($p < 0.001$), except for the Modern North, from which it could not be differentiated due to its small sample size.

3.4 | Demography of Norwegian mtDNAs

To better understand the historical demography of the Norwegian population in terms of changes in the effective population size related to coalescent events, a demographic model with Bayesian analysis was conducted using the 1174 Modern samples (Figure 5). The x-axis shows the time from the present in units of thousands of years, and the y-axis is equal to Ne·μ, the product of the effective population size and the HVSI mutation rate calculated by Soares and colleagues (Soares et al., 2009). In this Bayesian skyline plot (BSP), the thick solid line represents the median posterior effective population size through time, while the thin lines show the 95% highest posterior density limits.

As seen in this BSP the effective population size (Ne) slowly grew from about 35 kya and then stabilized around 12 kya as the inland ice sheets began melting at the end of the Late Glacial Maximum (Glørstad et al., 2020). The earlier period of population growth likely corresponds to the initial settlement of the European continent by anatomically modern humans. The population size then increased again around 6 kya, which corresponds to the late Neolithic period when agriculture expanded into Norway (Hjelle et al., 2006). The population further increased from 2.5 to 1.6 to kya (or around 300 to 400 CE) during the Late Bronze Age, when the proto-Norse runic alphabet was established (Imer, 2011).

The BSP appears to slightly overestimate the effective population size based on this data set. However, this result is well within the uncertainty admitted by the 95% hypothetical posterior density limits (and may well be due to the stochastic error associated with this particular simulation). The growth projection was also similar to the BSP estimated from Danish mtDNA sequences using similar methodology (Li et al., 2014).

4 | DISCUSSION

We analyzed mtDNA HVS1 sequences from different Norwegian subpopulations delineated by dialectal region and geography to investigate the source of their matrilineal descent and to identify genetic differences that might be related to population movements. Overall, the Norwegian matrilineal gene pool is represented by a diverse set of mitochondrial lineages that belong primarily to eight West Eurasian haplogroups, which are distributed differently in various regional areas. Our findings indicate that Norwegian mtDNA diversity was modestly influenced by geography, and to lesser extent by language (dialect region).

While Norwegians share many maternal lineages in common, their subpopulations differed slightly from each other in regions linked to known maritime routes around the perimeter of the country rather than latitudinal land-traversing routes. The differences between otherwise closely located regions (e.g., the Ancestral West Norway versus Ancestral Southeast Norway, which were significantly different, $p < 0.001$) were likely shaped by the geographic barrier of tall mountain ranges that prevented frequent close contact between the two

regions. As a result, there are genetic similarities between several settlements located along the country's longitudinal coastline that would have been otherwise difficult to reach by land due to the numerous fjords that separate them.

The mountainous regions and fjords that topographically separate regions of Norway have led to maritime travel becoming an essential component of human movements within the country. Maritime activities have molded the development of Norwegian culture and society for many millennia. The importance of marine vessels is reflected in their appearance in early rock art found all over Norway (Ballard et al., 2004; Bjerck, 2012; Ljunge, 2015). While recent archaeological evidence indicates that, by 6180–6680 cal yr BP, people sometimes traveled through the Jotunheimen Mountains that separate the West from the Southeast (Pilø et al., 2020), this travel was likely restricted to the winter months when bogs and streams were frozen over. Ships were also clearly crucial for the Norse expansion during the Viking Age (Østmo, 2003). Moreover, the mid-19th century "golden age" of sailing vessels has been recorded as being a catalyst for major timber and fishing-based economic activities within in Norway that led to eventual economic prosperity (Fischer & Nordvik, 1987; Haaland & Svihus, 2011).

Given its geographical barriers, Norwegians maintained remarkable contact across long distances through maritime travel. The terms "Norway" and "Norwegian" had been in use since at least the 9th century by the Viking Age seafarer Ottar from Hålogaland (in the North), and the petty kingdoms of Norway had been politically unified as a single entity by the 11th century (Bagge, 1995; Larsen, 1950), relatively earlier than other countries of similar size in Europe. Maritime routes remained the primary mode of trade and transportation until the use of motorized vehicles became popular in Norway in the early 20th century (Jentoft & Finstad, 2018) and railway networks expanded, with the Bergen line (*Bergenbanen*) in the West reaching the inland Southeast in 1909 (Haaland & Svihus, 2011). More recent more routes of modern travel may have led to closer genetic similarities between the western city of Bergen and the southeastern city of Oslo in the modern data set.

The detailed locality analyses indicate that Norwegian subpopulations are genetically similar, while Norwegians from Finnmark, and to a lesser extent, Agder and modern Førde, appear to have become moderately differentiated. The observed level of dissimilarity between these regions is similar to what had been reported for Y-chromosome (paternal) lineages in Norwegians (Dupuy et al., 2006). Dupuy et al. (2006) reported regional variation in Y-chromosome variation for Finnmark (north), Sogn and Fjordane (west), and Agder (south), and indicated that these three areas had the highest degree of dissimilarity. While men were more often the migrants who traveled long distances in Norway (Thorvaldsen, 2019), parish registers indicate that many migrations also involved young couples who had given birth to children only after moving to a new locality (Svalestuen, 1978; Thorvaldsen, 2019). By 1920, it was as usual for women as men to be migrants within Norway (Thorvaldsen, 2019).

Some of the observed regional differences may be explained by specific internal migration patterns. First, the difference between Finnmark and other Norwegian regions may be due to its geographical isolation for many centuries. Norwegians have been living in the northernmost parts of the country since at least the 14th century (Opsahl, 2020). However, the population was scattered along coastal areas, and the region was also occupied by the Saami, an indigenous people in Norway that had been seasonally nomadic until the late 19th century (Ahren, 2004). Interestingly, Norwegian Saami populations contain a higher proportion of U5b1b mtDNAs than Norwegians (Dupuy & Olaisen, 1996; Tambets et al., 2004). However, Krzewinska et al. detected this haplogroup in two individuals from the Late Iron Age who had received a Norse burial, suggesting that some Saami individuals may have been assimilated or accepted into Norse society as early as the 10th century CE (Krzewińska et al., 2015).

It is worth noting that the F*st* value of dissimilarity among Norwegians of Ancestral Finnmark are lower than those of Norwegian Saami population. The F*st* values in the order of 0.2 (all $p < 0.05$ for Bergen, Oslo, Førde and Trøndelag) (Saami as reported by Krzewińska, 2014) vs. 0.1 (all $p < 0.001$ for the same regions) (Norwegian Finnmark in this study). No Norwegians in our study nor in the study by Krzewińska, 2014 had F*st* values above 0.2. Thus, while the northern Finnmark subpopulation was somewhat more genetically dissimilar to all other Norwegian subpopulations, the level of genetic dissimilarity is higher among the Norwegian Saami, who also live in the north but who have a distinct history and migration patterns (Tambets et al., 2004).

In the 1780s, a mass movement of settlers from the southern to the more northern regions of Norway was encouraged by the appointed bailiff in order to take advantage of fishing opportunities there (Thorvaldsen, 2019). Involving some 1000 individuals, this northward migration continued until the 1830s, when the introduction of potato farming and smallpox vaccination made southeastern and western Norway more hospitable and prosperous (Gjerde, 1989; Tryland, 2001).

A second and more recent migration is the movement of rural farming populations to city locations in the southeast. Our analysis of modern (primarily urban-located) Norwegians in Oslo and each of the surrounding farming areas, namely, Oppland, Hedmark, and Buskerud, show strong genetic similarities, and suggest that migration from the inland to the coastal southeast during industrialization contributed their genetic make-up. In addition, the economic opportunities within the capital of Oslo have attracted populations from the entire country, which reflected in the diversity of mtDNA haplogroups represented in the area and has continued to the present day. According to the National Statistics Bureau archives, Oslo has had a positive net migration rate (i.e., once Norwegians migrate to Oslo, they tend to settle there permanently), as well as the highest overall migration rate within the country since at least 1966. By contrast, those in northern Norway as well as the rural southeastern areas of Hedmark and Oppland Counties have had negative net migration rates (Longva, 2000). In the south of Norway, Agder County's dependence on shipbuilding and its strategically located southern fishing shores may have kept that part of the country more independent and isolated from the rest of the southeastern regional area (Gjerde, 1989) and more connected with

the fishing coasts of the southwest. Much of the Norwegian population today still lives in or near coastal cities, in particular Oslo, Bergen, and Trondheim (Statistics Norway: Statistics Central Bureau, 2020).

A third and different pattern was found for the western part of Norway. While the haplogroup composition of Ancestral populations from Sogn and Fjordane is like other counties of the West, modern Førde has become more heterogeneous over time. According to census records, 17.2 per 1000 individuals from the Sogn municipality emigrated to America yearly from 1856 to 1865, with a total of 6430 emigrants leaving Norway during this period (Gjerde, 1989; Svalestuen, 1978). By 1905, 40% of Sogn and Fjordane inlanders had left for the US, while 20%–30% left in the central parts of Sogn and Fjordane (Gjerde, 1989; Østrem, 2015). As population growth outpaced the available food supply, many Norwegians in the western part of the country were lured by the possibility of owning arable farmland in the northwestern United States that could not be offered in the western Norway (Gjerde, 1989). After Ireland, the second largest number of immigrants to the US and the majority came from Norway, in particular the north-central part of the West (Gjerde, 2007).

We found that the modern Førde population was not significantly different from Ancestral Sogn and Fjordane ($p = 0.08$). Nevertheless, the loss of some maternal lineages may have widened the difference between modern Førde and other Norwegian subpopulations over time. Indeed, the uniqueness of Sogn and Fjordane had been reported based on both Y chromosome variation (Dupuy et al., 2006) and the higher proportion of the Kell (K+) blood group (Kornstad, 1997). In addition, the uniqueness of modern-day Førde may be due to the natural boundaries of the region (Krzewińska, 2014), with the highest peak of Jotunheimen being located at this county's eastern border.

On a larger time scale, our phylogeographic analysis pointed to some founder maternal lineages that are still shared by Norwegians from the earliest inhabitants of the Mesolithic era 9000 BP and Late Iron Age (500 to 1050 CE) (Günther et al., 2018; Krzewińska et al., 2015). The frequency of these haplogroups among the current Norwegian population has changed modestly over time, with slightly higher frequencies of U and K but slightly lower frequencies of H and J (Krzewińska et al., 2015). Furthermore, the early infiltration of Orkney, English, Scottish, Irish, and other Scandinavian lineages through slave trading since the Viking expansion has been reported by Krzewińska et al. (2015), who also showed their continued genetic affinities among the samples from Haugesund, Bergen, Førde, and Trondheim (Krzewińska, 2014).

In more recent times, the diversity of haplogroups among ethnic Norwegians shows a modest expansion. All of the haplogroups present in the Modern Norwegian population were also present among the Ancestor Norwegian population as expected, except for haplogroups N11 and F2a, which are present at low frequency within the Modern population. Because these mtDNAs belong to East Eurasian haplogroups, their presence is likely due to recent immigration of people from regions in which these maternal lineages are more common, that is, South-East Asia. In addition, haplogroup L2, the only maternal lineage of African origin, found both our study and that of Passarino et al., (2002), also appears among one participant in The Norwegian DNA Project (not included in this analysis) and is associated with an individual from the Dominican Republic who arrived in Norway in 1860. A haplogroup L2 mtDNA also appears in the Hordaland Ancestor population in an individual with a Norwegian first name and surname. Thus, L2 mtDNA was introduced into Norway by early immigrants. Conversely, G2a1 and Z1a, which are both haplogroups of East Eurasian origin, have been present among Norwegian Ancestors since at least the 1600s and likely entered the population during recent prehistory. Haplogroup G2a is also present in similarly low frequencies among populations of Central and Eastern Europe, while haplogroup Z is present at about 4% to 7% among Saami populations (Ingman & Gyllensten, 2007; Mielnik-Sikorska et al., 2013; Tambets et al., 2004).

In conclusion, our study provides an extensive survey of mtDNA haplogroup distributions among Norwegians. It reveals the importance of geographic regions as boundaries of gene flow among a people deeply influenced by Norway's topography and maritime travel around the country. The study also serves as a comprehensive framework for understanding how the pattern of genetic variation in the Norwegian population has been shaped by major historical events over several generations. Further studies of Norwegian maternal lineages, specifically those focusing on mitogenome sequence variation, will provide a more comprehensive evolutionary portrait of Norwegian population history, demography, and migration.

ACKNOWLEDGMENTS

We would like to thank Maja Krzewinska for providing her published thesis, from which further mitochondrial DNA data among Norwegians could be abstracted.

CONFLICT OF INTEREST

All authors report no conflicts of interest.

AUTHOR CONTRIBUTIONS

Dana Kristjansson: Conceptualization; data curation; formal analysis; investigation; methodology; software; validation; visualization; writing-original draft; writing-review & editing. **Jon Bohlin:** Resources; writing-review & editing. **Astanand Jugessur:** Funding acquisition; project administration; writing-review & editing. **Theodore Schurr:** Conceptualization; funding acquisition; investigation; methodology; project administration; resources; supervision; writing-review & editing.

DATA AVAILABILITY STATEMENT

No new data was generated in this study. The data that support the findings of this study are available from public repositories and the published literature. The data obtained from Genbank can be found at https://www.ncbi.nlm.nih.gov/genbank/, reference numbers (AY026032.1-AY025708.1, EU684448.1, EU980593.1, FJ499472.1, FJ652065.1, GU815340.1, HQ153430.1, HQ660704.1, HQ676806.1, HQ698894.1, HQ711364.1, HQ917079.1, JF825889.1, JN603188.1, JQ735910.1, JQ763435.1, JQ898578.1, KC170990.1, KF057946.1, KF817593.1, KJ603459.1, KP136794.1, KP407173.1, KP733897.1, KP969064.1,

KT210950.1, KT886412.1, KU057167.1, KU873089.1. KX129707.1, KX980415.1, KY000078.1, KY115220.1, MF103670.1, MF103671.1, MF116363.1, MF116367.1, MF597726.1, MF693153.1, MG436774.1, MG687433.1, MH142589.1, MH550114.1, MH899455.1, MK434282.1, MK792836.1, MN318468.1, MN599048.1.) Data from the published literature can be found in the following published articles referenced within the bibliography of this article: Opdal SH et al (1998), Table 1, DOI: 10.1080/080352598750031347 and Passarino G et al (2002), Table 3, DOI: 10.1038/sj.ejhg.5200834. Data from Krzewińska, M. (2014). Tables S2-S11 are published in the doctoral thesis Human origins and migrations in Norway inferred from ancient and modern DNA analysis. Theses at University of Oslo are available from the university library upon request (Object ID: 71499613630002201). The FamilyTreeDNA data were derived from the following resource that is openly available: https://www.familytreedna.com/public/Norway?iframe=mtresults.

ORCID

Dana Kristjansson https://orcid.org/0000-0003-0687-2910

Theodore G. Schurr https://orcid.org/0000-0001-9323-9237

REFERENCES

Ahren, M. (2004). Indigenous Peoples' culture, customs, and traditions and customary law - the Saami People's perspective. *Arizona Journal of International and Comparative Law*, *21*, 63. https://heinonline.org/HOL/Page?handle=hein.journals/ajicl21&id=81&div=&collection

Bagge, S. (1995). Nationalism in Norway in the middle ages. *Scandinavian Journal of History*, *20*(1), 1–18. https://doi.org/10.1080/03468759508579290

Ballard, C., Bradley, R., Myhre, L. N., & Wilson, M. (2004). The ship as symbol in the prehistory of Scandinavia and Southeast Asia. *World Archaeology*, *35*(3), 385–403. https://doi.org/10.1080/0043824042000185784

Bang-Andersen, S. (2003). Southwest Norway at the Pleistocene/Holocene transition: Landscape development, colonization, site types, settlement patterns. *Norwegian Archaeological Review*, *36*(1), 5–25. https://doi.org/10.1080/00293650307293

Bang-Andersen, S. (2012). Colonizing contrasting landscapes.: The pioneer coast settlement and inland utilization in southern Norway 10,000-9500 years before present. *Oxford Journal of Archaeology*, *31*(2), 103–120. https://doi.org/10.1111/j.1468-0092.2012.00381.x

Behar, D. M., Van Oven, M., Rosset, S., Metspalu, M., Loogväli, E. L., Silva, N. M., Kivisild, T., Torroni, A., & Villems, R. (2012). A "copernican" reassessment of the human mitochondrial DNA tree from its root. *American Journal of Human Genetics*, *90*(4), 675–684. https://doi.org/10.1016/j.ajhg.2012.03.002

Bjerck, H. (2012). On the outer fringe of the human world: Phenomenological perspectives on anthropomorphic cave paintings in Norway. In K. A. Bergsvik & R. Skeates (Eds.), *Caves in Context: the cultural significance of caves and rockshelters in Europe (pp. 48–64)*. Oxbow Books.

Côrte-Real, H. B. S. M., Macaulay, V. A., Richards, M. B., Hariti, G., Issad, M. S., Cambon-Thomsen, A., Papiha, S., Bertranpetit, J., & Sykes, B. C. (1996). Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. *Annals of Human Genetics*, *60*(4), 331–350. https://doi.org/10.1111/j.1469-1809.1996.tb01196.x

Derry, T. K. (2012). *A history of Scandinavia: Norway, Sweden, Denmark, Finland and Iceland* (13th ed.). University of Minnesota Press.

Drummond, A. J., Rambaut, A., Shapiro, B., & Pybus, O. G. (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, *22*(5), 1185–1192. https://doi.org/10.1093/molbev/msi103

Dupuy, B. M., & Olaisen, B. (1996). *mtDNA sequences in the Norwegian Saami and main populations* 23–25. Springer, . https://doi.org/10.1007/978-3-642-80029-0_6

Dupuy, B. M., Stenersen, M., Lu, T. T., & Olaisen, B. (2006). Geographical heterogeneity of Y-chromosomal lineages in Norway. *Forensic Science International*, *164*(1), 10–19. https://doi.org/10.1016/j.forsciint.2005.11.009

Dyrvik, S. (1972). Historical demography in Norway 1660-1801: A short survey. *Scandinavian Economic History Review*, *20*(1), 27–44. https://doi.org/10.1080/03585522.1972.10407709

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and windows. *Molecular Ecology Resources*, *10*(3), 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Fischer, L. R., & Nordvik, H. W. (1987). From Namsos to Halden: Myths and realities in the history of norwegian seamen's wages, 1850-1914. *Scandinavian Economic History Review*, *35*(1), 41–64. https://doi.org/10.1080/03585522.1987.10408081

Gjerde, J. (1989). *From peasants to farmers: The migration from Balestrand, Norway, to the upper middle west*. Cambridge University Press.

Gjerde, J. (2007). Fiksjon, fakta og forskning: seminar om den tidlige utvandringa til Amerika. In O. Østrem (Ed.), *Echoes of freedom: The Norwegian encounter with America*. (48–59). Stavanger, Norway: University of Stavanger.

Glørstad, H., Gundersen, J., Kvalø, F., Nymoen, P., Simpson, D., & Skar, B. (2020). Norway: Submerged stone age from a norwegian perspective. In *Coastal research library* (Vol. 35, pp. 125–140). Springer. https://doi.org/10.1007/978-3-030-37367-2_6

Grabbe, M., Lalander, E., Lundin, S., & Leijon, M. (2009). A review of the tidal current energy resource in Norway. *Renewable and Sustainable Energy Reviews*, *13*, 1898–1909. https://doi.org/10.1016/j.rser.2009.01.026

Günther, T., Malmström, H., Svensson, E. M., Omrak, A., Sánchez-Quinto, F., Kılınç, G. M., Krzewińska, M., Eriksson, G., Fraser, M., Edlund, H., Munters, A. R., Coutinho, A., Simões, L. G., Vicente, M., Sjölander, A., Jansen Sellevold, B., Jørgensen, R., Claes, P., Shriver, M. D., … Jakobsson, M. (2018). Population genomics of Mesolithic Scandinavia: Investigating early postglacial migration routes and high-latitude adaptation. *PLoS Biology*, *16*(1), e2003703. https://doi.org/10.1371/journal.pbio.2003703

Haaland, A., & Svihus, Å. (2011). Coastal and maritime Norway. Art Council Norway, The Norwegian Coastal Administration, The Directorate for Cultural Heritage and Directory of Fisheries. https://ra.brage.unit.no/ra-xmlui/bitstream/handle/11250/176922/Fortellinger_kystNorge_Kyst_og_havlandet_eng.pdf?sequence=1

Hammer, D. A. T., Ryan, P. D., Hammer, Ø., & Harper, D. A. T. (2001). Past: paleontological statistics software package for education and data analysis. *Palaeontologia Electronica*, *4*(1), 1–9. http://palaeo-electronica.orghttp//palaeo-electronica.org/2001_1/past/issue1_01.htm

Haugen, E. (1959). Planning for a standard language in modern Norway. *Anthropological Linguistics*, *1*, 8–21.

Helgason, A., Hickey, E., Goodacre, S., Bosnes, V., Stefánsson, K., Ward, R., & Sykes, B. (2001). mtDNA and the islands of the North Atlantic: Estimating the proportions of Norse and Gaelic ancestry. *American Journal of Human Genetics*, *68*(3), 723–737. https://doi.org/10.1086/318785

Helskog, K. (1985). Boats and meaning: A study of change and continuity in the Alta fjord, arctic Norway, from 4200 to 500 years B.C. *Journal of Anthropological Archaeology*, *4*(3), 177–205. https://doi.org/10.1016/0278-4165(85)90002-9

Hervik, A., Tretvik, T., & Øvstedal, L. (1993). *Norway: Crossing Fjords and Mountains* (pp. 349–365). Springer. https://doi.org/10.1007/978-94-015-8118-9_20

Hjelle, K. L., Hufthammer, A. K., & Bergsvik, K. A. (2006). Hesitant hunters: A review of the introduction of agriculture in western Norway. *Environmental Archaeology*, *11*(2), 147–170. https://doi.org/10.1179/174963106x123188

Imer, L. (2011). The oldest runic monuments in the north. *NOWELE. North-Western European Language EvolutionNOWELE / North-Western European Language EvolutionNOWELE*, *62–63*(63), 169–212. https://doi.org/10.1075/nowele.62-63.04ime

Ingman, M., & Gyllensten, U. (2007). A recent genetic link between Sami and the Volga-Ural region of Russia. *European Journal of Human Genetics*, *15*(1), 115–120. https://doi.org/10.1038/sj.ejhg.5201712

Jentoft, S., & Finstad, B. P. (2018). Building fisheries institutions through collective action in Norway. *Maritime Studies*, *17*(1), 13–25. https://doi.org/10.1007/s40152-018-0088-6

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kleppe, J. I. (2014). Desolate landscapes or shifting landscapes? Late glacial/early post-glacial settlement of northernmost Norway in the light of new data from eastern Finnmark. In F. Riede & M. Tallavaara (Eds.), BAR Intern *Lateglacial and postglacial pioneers in northern Europe*. Archaeopress.

Knutsen, H. (2019). NIBIO POP 6(8)2020.

Kornstad, L. (1997). Frequency of the blood group antigen K and the A1A2BO groups in the Norwegian counties. *Gene Geography: A Computerized Bulletin on Human Gene Frequencies*, *11*(1), 37–46. https://europepmc.org/article/med/9615212

Krzewińska, M. (2014). Human origins and migrations in Norway inferred from ancient and modern DNA analysis. Ph.D., Museum of Cultural Heritage and Department of Biosciences, University of Oslo.

Krzewińska, M., Bjørnstad, G., Skoglund, P., Olason, P. I., Bill, J., Götherström, A., & Hagelberg, E. (2015). Mitochondrial DNA variation in the Viking age population of Norway. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *370*(1660), 20130384. https://doi.org/10.1098/rstb.2013.0384

Larsen, K. (1950). *A history of Norway* (2nd ed.). Princeton University Press.

Lembring, M., Van Oven, M., Montelius, M., & Allen, M. (2013). Mitochondrial DNA analysis of Swedish population samples. *International Journal of Legal Medicine*, *127*(6), 1097–1099. https://doi.org/10.1007/s00414-013-0908-6

Li, S., Besenbacher, S., Li, Y., Kristiansen, K., Grarup, N., Albrechtsen, A., Sparsø, T., Korneliussen, T., Hansen, T., Wang, J., Nielsen, R., Pedersen, O., Bolund, L., & Schierup, M. H. (2014). Variation and association to diabetes in 2000 full mtDNA sequences mined from an exome study in a Danish population. *European Journal of Human Genetics*, *22*(8), 1040–1045. https://doi.org/10.1038/ejhg.2013.282

Librado, P., & Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, *25*(11), 1451–1452. https://doi.org/10.1093/bioinformatics/btp187

Ljunge, M. (2015). Capturing images: Knowledge, ownership and the materiality of cave art. In A. Klevnäs & C. Hedenstierna-Jonson (Eds.), *Own or be owned: Archaeological approaches to the concept of possession (pp. 131–140)*. Stockholm University.

Longva, S. (2000). Population statistics 1998 with figures as of 1 January 1999.

Malmström, H., Linderholm, A., Skoglund, P., Storå, J., Sjödin, P., Gilbert, M. T. P., Holmlund, G., Willerslev, E., Jakobsson, M., Lidén, K., & Götherström, A. (2015). Ancient mitochondrial DNA from the northern fringe of the Neolithic farming expansion in Europe sheds light on the dispersion process. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *370*(1660), 20130373. https://doi.org/10.1098/rstb.2013.0373

Margaryan, A., Lawson, D. J., Sikora, M., Racimo, F., Rasmussen, S., Moltke, I., Cassidy, L. M., Jørsboe, E., Ingason, A., Pedersen, M. W., Korneliussen, T., Wilhelmson, H., Buś, M. M., de Barros Damgaard, P., Martiniano, R., Renaud, G., Bhérer, C., Moreno-Mayar, J. V., Fotakis, A. K., … Willerslev, E. (2020). Population genomics of the Viking world. *Nature*, *585*(7825), 390–396. https://doi.org/10.1038/s41586-020-2688-8

Mielnik-Sikorska, M., Daca, P., Malyarchuk, B., Derenko, M., Skonieczna, K., Perkova, M., Dobosz, T., & Grzybowski, T. (2013). The history of Slavs inferred from complete mitochondrial genome sequences. *PLoS One*, *8*(1), e54360. https://doi.org/10.1371/journal.pone.0054360

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Niemi, E., Myhre, J. E., & Kjeldstadli, K. (2003). I nasjonalstatens tid 1814-1940. In K. Kjeldstadli (Ed.), *Norwegian immigration history* (2nd ed.). Oslo, Norway: Pax.

Opdal, S. H., Rognum, T. O., Vege, Å., Stave, A. K., Dupuy, B. M., & Egeland, T. (1998). Increased number of substitutions in the D-loop of mitochondrial DNA in the sudden infant death syndrome. *Acta Paediatrica, International Journal of Paediatrics*, *87*(10), 1039–1044. https://doi.org/10.1080/080352598750031347

Opsahl, E. (2020). Avaldsnes' position in Norway in the 14th century. In D. Skre (Ed.), *Rulership in 1st to 14th century Scandinavia. Royal graves and sites at Avaldsnes and beyond* (pp. 517–529). De Gryuter.

Østmo, E. (2003). When the Norsemen learned to row: A technological innovation for shipping in the early iron age. *Viking: Norwegian Archaeological Yearbook*, *66*, 7–29.

Østrem, N. O. (2015). *Suget frå Amerika - Norgeshistorie*. https://www.norgeshistorie.no/industrialisering-og-demokrati/1546-suget-fra-amerika.html

Passarino, G., Cavalleri, G. L., Lin, A. A., Cavalli-Sforza, L. L., Børresen-Dale, A. L., & Underhill, P. A. (2002). Different genetic components in the Norwegian population revealed by the analysis of mtDNA and Y chromosome polymorphisms. *European Journal of Human Genetics*, *10*(9), 521–529. https://doi.org/10.1038/sj.ejhg.5200834

Pilø, L. H., Barrett, J. H., Eiken, T., Finstad, E., Grønning, S., Post-Melbye, J. R., Nesje, A., Rosvold, J., Solli, B., & Ødegård, R. S. (2020). Interpreting archaeological site-formation processes at a mountain ice patch: A case study from Langfonne, Norway. *The Holocene*, *3*, 469–482. https://doi.org/10.1177/0959683620972775

Pryser, T. (1981). In B. Gjerdåker (Ed.), *På flyttefot: Innanlands vandring på 1800-talet* (pp. 59–69). Det Norske Samlaget.

R Core Team. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing.

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Systematic Biology*, *67*(5), 901–904. https://doi.org/10.1093/sysbio/syy032

Richard, C., Pennarun, E., Kivisild, T., Tambets, K., Tolk, H. V., Metspalu, E., Reidla, M., Chevalier, S., Giraudet, S., Lauc, L. B., Peričić, M., Rudan, P., Claustres, M., Journel, H., Dorval, I., Müller, C., Villems, R., Chaventré, A., & Moisan, J. P. (2007). An mtDNA perspective of French genetic variation. *Annals of Human Biology*, *34*(1), 68–79. https://doi.org/10.1080/03014460601076098

Skjekkeland, M. (2005). *Dialektar i Noreg–Tradisjon og Fornying*. Høyskoleforlaget. https://scholar.google.com/scholar?hl=no&as_sdt=0%2C5&q=Skjekkeland%2C+M.+%282005%29.+Dialektar+i+Noreg–Tradisjon+og+Fornying.+Kristiansand%3A+Høyskoleforlaget.&btnG

Skoglund, P., Malmström, H., Raghavan, M., Storå, J., Hall, P., Willerslev, E., Gilbert, M. T. P., Götherström, A., & Jakobsson, M. (2012). Origins and genetic legacy of neolithic farmers and hunter-gatherers in Europe. *Science*, *336*(6080), 466–469. https://doi.org/10.1126/science.1216304

Soares, P., Alshamali, F., Pereira, J. B., Fernandes, V., Silva, N. M., Afonso, C., Costa, M. D., Musilová, E., MacAulay, V., Richards, M. B., Černý, V., & Pereira, L. (2012). The expansion of mtDNA haplogroup L3 within and out of Africa. *Molecular Biology and Evolution*, *29*, 915–927. https://doi.org/10.1093/molbev/msr245

Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Röhl, A., Salas, A., Oppenheimer, S., Macaulay, V., & Richards, M. B. (2009). Correcting for purifying selection: An improved human mitochondrial molecular clock. *American Journal of Human Genetics*, *84*(6), 740–759. https://doi.org/10.1016/j.ajhg.2009.05.001

Sølvi, S. (1979). *Folkevekst og flytting: en historisk-demografisk studie i 1700-årenes Øst-Norge*. Universitetsforlaget. http://urn.nb.no/URN:NBN:no-nb_digibok_2009062901084

Statistics Norway: Statistics Central Bureau. (2020). *Facts about the Population*. https://www.ssb.no/en/befolkning/statistikker/beftett

Stroeven, A. P., Hättestrand, C., Kleman, J., Heyman, J., Fabel, D., Fredin, O., Goodfellow, B. W., Harbor, J. M., Jansen, J. D., Olsen, L., Caffee, M. W., Fink, D., Lundqvist, J., Rosqvist, G. C., Strömberg, B., & Jansson, K. N. (2016). Deglaciation of Fennoscandia. *Quaternary Science Reviews*, *147*, 91–121. https://doi.org/10.1016/j.quascirev.2015.09.016

Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., & Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution*, *4*(1), 1–5. https://doi.org/10.1093/ve/vey016

Svalestuen, A. A. (1978). Om den Regionale Spreiinga av Norsk Utvandring før 1865. In A. Engen (Ed.), *Utvandringa-Det Store Opbrotet (p. 77)*. Det Norske Samlaget.

Tambets, K., Rootsi, S., Kivisild, T., Help, H., Serk, P., Loogväli, E. L., Tolk, H. V., Reidla, M., Metspalu, E., Pliss, L., Balanovsky, O., Pshenichnov, A., Balanovska, E., Gubina, M., Zhadanov, S., Osipova, L., Damba, L., Voevoda, M., Kutuev, I., … Villems, R. (2004). The Western and eastern roots of the Saami - the story of genetic "outliers" told by mitochondrial DNA and Y chromosomes. *American Journal of Human Genetics*, *74*(4), 661–682. https://doi.org/10.1086/383203

Tamura, K., & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, *10*(3), 512–526. https://doi.org/10.1093/oxfordjournals.molbev.a040023

The Norway DNA Project Group. (2014). *FamilyTreeDNA - The Norway DNA - Norge Project*. https://www.familytreedna.com/group-join.aspx?Group=Norway

Thorvaldsen, G. (2019). Internal migration in 19th and 20th century Norway. An overview 1865 to 1960. In *Nominative Data in Demographic Research in the East and the West: Monograph* (pp. 166–184). Publishing house of the Ural University. https://doi.org/10.15826/B978-5-7996-2656-3.10

Tryland, M. (2001). Kopper og koppevirus – 200 år siden første vaksinasjon i Norge. *Tidsskrift for Den Norske Legeforening*. https://tidsskriftet.no/2001/12/medisinsk-historie/kopper-og-koppevirus-200-ar-siden-forste-vaksinasjon-i-norge

van Oven, M., & Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation*, *30*(2), E386–E394. https://doi.org/10.1002/humu.20921

Venås, K., & Skjekkeland, M. (2020). dialekter i Norge – inndeling – Store norske leksikon. https://snl.no/dialekter_i_Norge_-_inndeling

Vistad, O. I., Wold, L. C., Daugstad, K., & Haukeland, J. V. (2016). Mimisbrunnr climate park – A network for heritage learning, tourism development, and climate consciousness. *Journal of Heritage Tourism*, *11*(1), 43–57. https://doi.org/10.1080/1743873X.2015.1082570

Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H. J., Kronenberg, F., Salas, A., & Schönherr, S. (2016). HaploGrep 2: Mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Research*, *44*(W1), W58–W63. https://doi.org/10.1093/nar/gkw233

Wickler, S., & Narmo, L. E. (2014). Tracing the development of fishing settlement from the iron age to the modern period in northern Norway: A case study from Borgvær in the Lofoten Islands. *Journal of Island and Coastal Archaeology*, *9*(1), 72–87. https://doi.org/10.1080/15564894.2013.810678

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Kristjansson, D., Bohlin, J., Jugessur, A., & Schurr, T. G. (2021). Matrilineal diversity and population history of Norwegians. *American Journal of Physical Anthropology*, *176*(1), 120–133. https://doi.org/10.1002/ajpa.24345

Paper II

Kristjansson *et al. BMC Genomics* (2022) 23:354
https://doi.org/10.1186/s12864-022-08572-y

BMC Genomics

RESEARCH

Open Access

Evolution and dispersal of mitochondrial DNA haplogroup U5 in Northern Europe: insights from an unsupervised learning approach to phylogeography

Dana Kristjansson[1,2*], Jon Bohlin[1,3], Truc Trung Nguyen[4], Astanand Jugessur[1,2] and Theodore G. Schurr[5]

Abstract

Background: We combined an unsupervised learning methodology for analyzing mitogenome sequences with maximum likelihood (ML) phylogenetics to make detailed inferences about the evolution and diversification of mitochondrial DNA (mtDNA) haplogroup U5, which appears at high frequencies in northern Europe.

Methods: Haplogroup U5 mitogenome sequences were gathered from GenBank. The hierarchal Bayesian Analysis of Population Structure (hierBAPS) method was used to generate groups of sequences that were then projected onto a rooted maximum likelihood (ML) phylogenetic tree to visualize the pattern of clustering. The haplogroup statuses of the individual sequences were assessed using Haplogrep2.

Results: A total of 23 hierBAPS groups were identified, all of which corresponded to subclades defined in Phylotree, v.17. The hierBAPS groups projected onto the ML phylogeny accurately clustered all haplotypes belonging to a specific haplogroup in accordance with Haplogrep2. By incorporating the geographic source of each sequence and subclade age estimates into this framework, inferences about the diversification of U5 mtDNAs were made. Haplogroup U5 has been present in northern Europe since the Mesolithic, and spread in both eastern and western directions, undergoing significant diversification within Scandinavia. A review of historical and archeological evidence attests to some of the population interactions contributing to this pattern.

Conclusions: The hierBAPS algorithm accurately grouped mitogenome sequences into subclades in a phylogenetically robust manner. This analysis provided new insights into the phylogeographic structure of haplogroup U5 diversity in northern Europe, revealing a detailed perspective on the diversity of subclades in this region and their distribution in Scandinavian populations.

Keywords: Scandinavia, Migration, Phylogeny, Clade, Haplotype

Significance statement

We wanted to explore the genetic structure of haplogroup U5 in northern Europe by employing an unsupervised learning approach for phylogenetic clustering. We accurately identified groups of mitochondrial DNA (mtDNA) sequences that were mapped onto a phylogenetic tree in order to make historical inferences about human population history. Our results support previous

*Correspondence: dana.kristjansson@fhi.no

[1] Center for Fertility and Health, Norwegian Institute of Public Health, Oslo, Norway
Full list of author information is available at the end of the article

hypotheses that haplogroup U5 mtDNAs expanded throughout Northern, Southern and Central Europe with more recent expansions into Western Europe and Africa. The results further allow us to explain how U5 mtDNAs are now found with high frequency in Northern Europe, as well as delineate the origins of the specific U5 subhaplogroups found in that part of Europe. We have found that the integration of hierBAPS clustering with a maximum likelihood phylogenetic analysis delineates clusters of similar sequences at sufficient resolution to allow broader inferences about lineage diversification and population migration history to be made from them. To our knowledge, this combined approach has not yet been applied to studies focused on human mtDNA diversity

Introduction

Over the past three decades, mitochondrial DNA (mtDNA) variation has been used to trace human ancestry in population genetic studies. The mtDNA is particularly informative for evolutionary studies because it represents a non-recombining part of the human genome, is maternally inherited, and evolves at a clock-like rate [1]. For this reason, many tens of thousands of mitogenomes from different human populations have been sequenced in an effort to reconstruct the phylogeographic history of our species.

Since its first build was published in 2008, Phylotree has become one of the most comprehensive libraries of known global human mtDNA variation [2]. It provides a systematic haplogroup nomenclature based on signature polymorphisms observed in the published sequences entered in the database. Until recently, Phylotree has been continually updated with newly available mitogenome sequences, and currently incorporates data from 24,275 mitogenomes [2].

Despite it being a valuable resource, the nomenclature used in Phylotree to assign a haplogroup status to individual sequences remains tedious and prone to errors. This is especially the case when a haplogroup must be assigned to several sequences from a human population in which several branches of an ancestral haplogroup may have slightly varying mutations. Although algorithmic software that incorporates Phylotree nomenclature can aid in haplogroup identification [3–7], constructing a phylogenetic tree that is consistent with Phylotree haplogroup labeling still remains an iterative and slow process.

A maximum likelihood (ML) phylogeny based on single nucleotide polymorphism (SNP) calling can be referentially rooted at an ancestral sequence, and also take into account character transformations using different evolutionary models that can be validated using bootstrapping methods or bootstrap approximations [8–10]. While ML is often employed to understand the evolutionary relationship of non-human species, its use in human mtDNA analyses has been limited due to the tediousness of assigning each mitogenome sequence to a Phylotree haplogroup. In addition, the similarity of the sequences in large human populations typically studied in these analyses can often result in unintelligible, dense, and unorganized trees. As a consequence, the genetic relationships of groups of similar sequences become difficult to disentangle and categorize for broader, evolutionary inferences. Furthermore, since Phylotree was last updated back in February 2016, several haplogroups have been recently defined but not integrated into the current nomenclatural system [11–15]. Thus, a method that could quickly categorize new sequences at high resolution would be extremely useful for phylogenetic studies.

One such method of making these classifications is the hierarchical Bayesian Analysis of Population Structure (hierBAPS) algorithm. This algorithm identifies clusters of sequences based on the corresponding allele frequencies within that cluster [16]. It is especially useful for quickly grouping sequences from several individuals who have different haplotypes but share a common ancestral lineage. The grouping of large clusters of ancestrally derived sequences further allows broader inferences to be made about their evolution, and can lead to a more refined visual organization that may not be evident based on detailed haplogroup labeling alone.

The hierBAPS clustering has usually been conducted in studies of haploid DNA from microorganisms [17, 18]. In particular, it has been utilized for several years in combination with ML phylogenetics for studies of bacterial populations [16, 19, 20]. However, this combined methodology has yet to be applied to an evolutionary analysis of human mtDNAs.

Haplogroup U5 as a case study

Haplogroup U5 is one of the most ancient mtDNA lineages to have existed in Central Europe prior to its dispersal into Northern Europe [21, 22]. This haplogroup is thought to have evolved in the western steppe region [23] and then entered Europe around 30 to 55 kya [1, 24]. It appears to have expanded into Europe before the end of the Last Glacial Maximum (LGM) over 20 thousand years ago (kya) [1, 25–27], i.e., before the thick ice sheets covering most of northern continental Europe were in the final stages of dissipating away from the interior.

Today, the frequency of U5 varies between 5–12% in most European countries [28, 29]. Its frequency varies particularly widely within Northern Europe. Haplogroup U5 mtDNAs are present in northern Saami populations at over 50% [30–32], while their corresponding frequencies in the southern areas of the Scandinavian countries (Norway, Sweden, and Denmark) lie between 6–15%

[31, 33, 34]. These differing frequencies raise interesting questions about the phylogenetic structure of this major lineage and the timing of the dispersal of its subbranches within the European continent.

On this note, while both Saami and Finns speak Finno-Uralic languages, the two populations do not share a close genetic relationship based on nuclear DNA marker loci [35]. This pattern is also true to some extent based on mtDNA data. Apart from Scandinavia, U5b mtDNAs with the "Saami motif" (defined by the T16144C, T16189C, and C16270T control-region variants; Tambets et al. 2004) have been observed at significant frequencies in populations from the northwestern Pskov Oblast and the Republic of Karelia in Russia [31, 36]. This distribution points the emergence of U5b mtDNA in ancestral Saami (Uralic) groups, and their dispersal into surrounding Indo-European populations through admixture.

Based on this evidence, it is generally agreed that the Saami are genetically distinct from other European populations [32, 37, 38], although the source of U5 mtDNAs among these European populations is not entirely clear. Therefore, a broader analysis of the phylogeographic features of haplogroup U5 is necessary to fill this knowledge gap. The aim of this study is thus to combine hierBAPS analysis of haplogroup U5 mitogenome sequences with maximum likelihood (ML) phylogenetics to make inferences about the evolution and dispersal of this major maternal lineage in Northern Europe.

Materials and methods

Mitogenome sequences

Data for haplogroup U5 mitogenome sequences were retrieved from the European Nucleotide Archive and GenBank ($n=873$) (accessed on 31 May 2021) and the search was limited to "whole mtDNA" and "haplogroup U5". For the purposes of this study, we separated Nordic populations into Saami, Scandinavia (Norway, Denmark, and Sweden), and Finland categories. Finland was separated from Scandinavia in this analysis due to its geographic isolation from the Scandinavian Peninsula and its linguistic distinctiveness. Specific information about the ethnicity or original location of the individuals represented by these sequences was available for 855 (97.8%) of the total dataset. The accession numbers of the samples are provided in the data availability statement.

Phylogenetic analysis

Maximum-likelihood phylogeny

We constructed a ML phylogeny from the 873 U5 mitogenome sequences with the software IQ-tree 1.6.12 [9]. The phylogeny was constructed under the best fitting nucleotide substitution model inferred by jModelTest [39, 40], which was TIM3+F+R3 based on the Bayesian Information Criterion (BIC). Branch support was achieved by the approximate likelihood ratio test (aLRT) [41] based on resampling the estimated log-likelihood method with a simple but effective collection scheme of candidate trees [39]. This was accomplished by applying the UFBoot algorithm [10] for 10,000 replicates. UFBoot overcomes the computational burden required by the standard nonparametric bootstrap, and can be interpreted as providing an unbiased bootstrap support with 95% support which corresponds to a 95% probability that a clade is true [42].

Partitioning mtDNA sequences using hierBAPS

To identify clusters of closely linked sequences within the 873 U5 mitogenome sequences, we employed the hierBAPS algorithm [43]. This algorithm groups DNA sequences into clusters in a hierarchical manner, and can be used to project the grouped sequences onto an independently derived phylogenetic tree [19]. The hierBAPS algorithm assumes that each individual sequence is drawn from one of several distinct genetic subpopulations, with each cluster having its own set of allele frequencies.

To apply hierBAPS to mtDNA sequences, we utilized an R software implementation of algorithm, RhierBAPS, that is available on the Comprehensive R Archive Network [19]. Briefly, the hierBAPS algorithm attempts to maximize the posterior probability of an allocation of a sequence over other possible allocations, assigning each individual sequence to specific clusters. After the number of clusters (K) is specified, the algorithm partitions the sequences of the dataset into as many groupings as possible (up to K_{max} clusters). The initial number of K clusters can be chosen based on the number of subpopulations expected, and can be increased on each re-run of the algorithm. The algorithm is typically re-run until the number of clusters stops increasing.

The clusters were refined into levels of low to high resolution of cluster specificity. We conducted three different cluster-level combinations: *Level 1*: 4 groups, *Level 2*: 11 groups, and *Level 3*: 24 groups. To distinguish Phylotree labels from hierBAPS groups for the demonstrative purposes of this study, alphabetical letters or roman numerals were used to represent hierBAPS labels. It is important to note here that the hierBAPS group labels provided by the algorithm, denoted by roman numerals, are generated in arbitrary order.

We also explored hierBAPS clustering using only the coding regions of the mitogenome sequences. This step was carried out by extracting the coding regions of the sequences using the Harvesttools package [44]. We conducted four cluster-level combinations on these data: *Level 1*: 3 groups, *Level 2*: 6 groups, *Level 3*: 12 groups,

and *Level 4*: 18 groups. The highest resolution results for both the coding region only and the whole mitogenome sequences were then compared.

Haplogroup identification

We used Haplogrep, version 2.1.21 [5] to assign a haplogroup to each mitogenome sequence based on its mutational signature, independent of the hierBAPS grouping. Haplogrep computes these classifications on pre-calculated phylogenetic weights that correspond to the occurrence of a polymorphism per position in Phylotree Build 17 [2], which, in turn, reflects the mutational stability of a variant. Mutations were identified relative to the Reconstructed Sapiens Reference Sequence (RSRS) [24], which allows for the naming and mapping of human mtDNA haplogroups from an ancestral root.

To be clear about the outcome of this analysis, we have utilized the following definitions when discussing the details of the U5 phylogeny. First, a *haplogroup* is a group of similar haplotypes that share a combination of ancestral polymorphisms commonly inherited together, such as U5. Similarly, a *subhaplogroup* is a branch of a haplogroup containing a subset of the sequences defined by the parent haplogroup but defined by its own set of mutation, such as U5b or even more specifically U5b1b1b. By contrast, a *subclade* is a cluster of related haplotypes associated with a hierBAPS grouping. On a more general level, a *lineage* is a maternal line of descent often referred to in population studies, and a *branch* is a part of the phylogenetic tree that extends from a root or major trunk.

Haplogroup age estimates

A temporal framework for the divergence of haplogroup U5 branches was assessed with TempEst v.1.5.1 [45]. Age estimates with 95% confidence intervals were calculated using the Least Squares Dating IQ-tree plugin [46]. To calibrate the ages, we used a root age based on the reported 177 ± 11 kya age estimation for the RSRS sequence reported by Behar and colleagues [24], as well as radiocarbon dating for ancient samples bearing U5 mtDNAs [22, 47, 48].

Comparative data analysis

Due to the fact that the GenBank sequences were collected for specific research purposes, had a low sample size per region, and did not encompass all geographic locations, it was not possible to make conclusions about haplogroup prevalence based solely on these data. Thus, the GenBank sequences were only utilized in this study for the purposes of making conclusions about the groupings and evolutionary relationships between sequences from an ancestral inference point.

To understand the geographical prevalence of U5 based on more representative data, we conducted a search of studies reporting the frequency of U5 mtDNAs within various populations. The frequency from each specific region was then tabulated. For more specific information about the major subhaplogroups within U5, we obtained data from 6488 individuals from the public database on the U5 mtDNA Project available from FamilyTreeDNA [49]. The overall frequencies of U5 mtDNAs were plotted on a geographic heat map using the statistical programming language R, version 3.6.3 (The R Foundation), and its graphical package ggplot2.3 [50].

Results

Bayesian Analysis of hierBAPS Groups

The least detailed hierBAPS analysis (*Level 1*) identified four major clusters within haplogroup U5. These included A: U5a1; B: U5a2; C: U5b1 + U5b3; and D: U5b2. The most detailed hierBAPS analysis (*Level 3*) identified 24 groups. The 24-group analysis listed the RSRS separately as group VIII, while the other 23 groups corresponded to the specific subclades listed in Table 1. Excluding the RSRS sequence, each of the 23 hierBAPS groups shared a set of polymorphisms that enabled the hierBAPS algorithm to generate specific clusters for them (Table 2). About 32.5% ($n = 28$) of the group-defining polymorphisms occurred in the non-coding control region of the mitogenome sequence.

All hierBAPS groups and the specific set of polymorphisms shared among them were mutually exclusive, i.e., no haplogroups were defined by a set of polymorphisms that was common to two different hierBAPS clusters. Additionally, the hierBAPS algorithm was able to accurately cluster all sequences belonging to a specific subhaplogroup even though each member of a hierBAPS group did not contain all diagnostic polymorphisms for a haplogroup defined by Phylotree. For example, not all sequences clustering in subclade III, represented by subhaplogroup U5a1, contained the polymorphisms 14793G and 16256 T, which are diagnostic for this subhaplogroup according to Phylotree, build 17. However, all subclade III sequences contained a sufficient number of common polymorphisms unique to them such that they could be partitioned to this branch within the U5 phylogeny.

The hierBAPS analysis also revealed considerable substructure within subhaplogroup U5b. Subhaplogroup U5b3, which is present in less than 1% in most human populations [51], was placed in subclade IV along with several other U5b1 sequences. Despite having other differences between them, the haplotypes within subclade IV shared two specific control region mutations, 16233C and 16230A, which caused them to cluster together both in the ML phylogenetic tree and in subclade IV.

Table 1 HierBAPS groups and their representative subclade(s) based on the human mtDNA U5 haplogroup

hierBAPS Groups	*Broad haplogroup (4-digits)*	*Major Subclade(s) or Haplogroups*	*Specific Haplogroups Included*	*N*	*%*
I	U5a2	U5a2, U5a2b, U5a2c, U5a2d	U5a2, U5a2 + 16294 T, U5a2b, U5a2b1a, U5a2b1b, U5a2b1c, U5a2b1d, U5a2b2, U5a2b2a, U5a2b3, U5a2b3a, U5a2b3a1, U5a2b4, U5a2b4a, U5a2c, U5a2c1, U5a2c3a, U5a2c4, U5a2d, U5a2d1, U5a2d1a	90	10.3
II	U5a2	U5a2e	U5a2e	11	1.3
III	U5a1	U5a1, U5a1g, U5a1i	U5a1, U5a1b, U5a1b + 16362C, U5a1b1, U5a1b1a, U5a1b1b, U5a1b1b1, U5a1b1c, U5a1b1c1, U5a1b1c2, U5a1b1d + 16093C, U5a1b1d1, U5a1b1e, U5a1b1g, U5a1b1h, U5a1b2, U5a1b3, U5a1b3a, U5a1b3a1, U5a1b4, U5a1d, U5a1d1, U5a1e, U5a1f1a, U5a1f2, U5a1g, U5a1g1, U5a1i, U5a1i1 U5a1j	128	14.6
IV	U5b1 + U5b3	U5b1, U5b1a, U5b1d, U5b1f, U5b1i, U5b3	U5b1, U5b1a, U5b1d1a, U5b1d1b, U5b1d1c, U5b1d2, U5b1f, U5b1f1, U5b1f1a, U5b1i, U5b3, U5b3a1a, U5b3a2, U5b3b1, U5b3b2, U5b3e, U5b3h	44	5
V	U5b1 + U5b3	U5b1 + 16189C!, U5b1b, U5b1c	U5b1 + 16189C!, U5b1b, U5b1b2, U5b1b2a, U5b1b2b, U5b1c, U5b1c1a, U5b1c1a1, U5b1c2, U5b1c2a, U5b1c2b	72	8.2
VI	U5a1	U5a1a2	U5a1a2, U5a1a2a, U5a1a2a1, U5a1a2a1a, U5a1a2b1	26	3
VII	U5a1	U5a1h	U5a1h	7	0.8
	RSRS	*RSRS*	*RSRS*	1	*0.1*
IX	U5a1	U5a1d2	U5a1d2a, U5a1d2a1, U5a1d2b	18	2.1
X	U5a1	U5a1c	U5a1c	28	3.2
XI	U5a1	U5a1a1	U5a1a1, U5a1a1a, U5a1a1b, U5a1a1c, U5a1a1d, U5a1a1h, U5a1a1i	89	10.2
XII	U5b2	U5b2a	U5b2a, U5b2a1b, U5b2a3, U5b2a3a, U5b2a4, U5b2a4a, U5b2a5, U5b2a5a, U5b2a6	25	2.9
XIII	U5b2	U5b2, U5b2c	U5b2, U5b2c1, U5b2c2, U5b2c2b	11	1.3
XIV	U5b2	U5b2a2	U5b2a2, U5b2a2a1, U5b2a2b, U5b2a2b1, U5a2a2c	29	3.3
XV	U5b2	U5b2b	U5b2b, U5b2b2, U5b2b3a1a, U5b2b4, U5b2b4a, U5b2b5	19	2.2
XVI	U5b2	U5b2b1	U5b2b1a, U5b2b1a1, U5b2a1a2, U5b2b1b	10	1.1
XVII	U5b1 + U5b3	U5b1b1a	U5b1b1a, U5b1b1a1, U5b1b1a1a, U5b1b1a1a1, U5b1b1a1b, U5b1b1a2, U5b1b1a3	83	9.3
XVIII	U5b1 + U5b3	U5b1b1	U5b1b1, U5b1b1 + 152C!, U5b1b1b, U5b1b1d, U5b1b1e, U5b1b1f, U5b1b1g1, U5b1b1g1a	39	4.5
XIX	U5b2	U5b2a1a + 16311 T!	U5b2a1a + 16311 T!, U5b2a1a1, U5b2a1a1a, U5b2a1a1d	32	3.7
XX	U5b2	U5b2a1a2	U5b2a1a2	5	0.6
XXI	U5a2	U5a2a	U5a2a, U5a2a1, U5a2a1 + 152C!, U5a2a1a, U5a2a1b, U5a2a1b1, U5a2a1c, U5a2a1e	70	8
XXII	U5a2	U5a2a2a	U5a2a2a	8	0.9
XXIII	U5b1 + U5b3	U5b1e1	U5b1e1, U5b1e1a	25	2.9
XXIV	U5b1 + U5b3	U5b1e1 (+T8337C)	U5b1e1 (+T8337C)	6	0.7

By contrast, the sequences assigned to subclade V appeared in two places in the phylogenetic tree. One was situated between subclades IV and XXIII, and the other between subclades XXIV and XVIII. This subclade is also part of subhaplogroup U5b1, although all its constituent subhaplogroups (e.g., U5b1b, U5b1c) arose after the T16189C! mutational event. Subhaplogroup U5b1b1 was placed in subclade XVIII, while its daughter branches in U5b1b1a were clustered into subclade XVIII.

In addition, the hierBAPS algorithm grouped subhaplogroup U5b1e1 + T8337C (subclade XXIV) with its parent haplogroup U5b1e1 (subclade XXIII). This distinction was not previously noted in Phylotree (Build 17). Both subclades XXIII and XXIV contained a set of polymorphisms diagnostic for subhaplogroup U5b1e, with subclade XXIV sequences also having the T8337C polymorphism in the mtDNA $tRNA^{Lys}$ gene.

ML phylogenetic tree projection

The hierBAPS group results were projected onto an ML tree from lowest to highest number of clusters (Figure S1). The *Level 3*:24 group analysis provided the most

Table 2 Shared mtDNA polymorphisms per hierBAPS group[a]

		hierBAPS Group													
		I	II	III	IV	V	VI	VII		IX	X	XI	XII	XIII	XIV
		U5a2, U5a2b, U5a2c, U5a2d	U5a2e	U5a1,U5a1g, U5a1i	U5b1, U5b1a, U5b1d, U5b1f, U5b1i, U5b3	U5b1 + T16189C! U5b1b, U5b1c	U5a1a2	U5a1h	*RSRS (Ancestral)*	U5a1d	U5a1c	U5a1a1	U5b2a	U5b2, U5b2c	U5b2a2
Number of shared mitochondrial polymorphisms within each U5 hierBAP group (n)		19	21	16	18	19	29	45	*1*	20	28	21	27	33	35
Nucleotide Position	Location														
146	HVS-II		T					T	T		T		T	T	T
150	HVS-II			T				T	C				T	T	T
151	HVS-II		T						C						
152	HVS-II		C!	T			T	T	T						T
195	HVS-II		T				T	T	T						T
247	HVS-II	G	G	G			G	G	G		G	G	G	G	G
523	HVS-III							A	A					A	
524	HVS-III							C	C					C	
769	12S_rRNA	G	G	G	G	G	G	G	G	G	G	G		G	G
1303	12S_rRNA							A	G						
1700	16S_rRNA						C		T			C			
1721	16S_rRNA								C				T	T	T
2757	16S_rRNA								A						
3027	16S_rRNA								T	C					
3107	preserves historical genome annotation numbering						d	d	-						d
3192	16S_rRNA							T	C						
3197	16S_rRNA	C	C		C	C	C	C	T	C	C	C	C	C	C
3212	16S_rRNA								C						T
3552	ND1 (Ala—3rd position in codon)								T	C					
3591	ND1 (Leu—3rd position in codon)							A	G						

Table 2 (continued)

		hierBAPS Group													
		I	II	III	IV	V	VI	VII		IX	X	XI	XII	XIII	XIV
		U5a2, U5a2b, U5a2c, U5a2d	U5a2e	U5a1,U5a1g, U5a1i	U5b1, U5b1a, U5b1d, U5b1f, U5b1i, U5b3	U5b1 + T16189C! U5b1b, U5b1c	U5a1a2	U5a1h	*RSRS (Ancestral)*	U5a1d	U5a1c	U5a1a1	U5b2a	U5b2, U5b2c	U5b2a2
3768	ND1 (Leu—3rd position in codon)		G						A						
4592	ND2 (Ser—3rd position in codon)							C	T						
4732	ND2 (Asn—2nd position in codon)								A				G		G
5452	ND2 (Thr—2nd position in codon)								C						
5495	ND2 (Phe—3nd position in codon)								T			C			
5656	position between tRNA-Ala and tRNA-Asn					G			A						
7146	CO1 (Ala—1st position in codon)	A	A	A	A	A	A	A	A		A	A		A	A
7256	CO1 (Asn—3rd position in codon)	C	C	C	C	C	C	C	C	C	C	C	C	C	
7521	tRNA-Asp	G	G		G	G	G	G	G	G	G	G	G	G	G
7768	CO2 (Met—3rd position in codon)				G	G			A				G	G	G
7853	CO2 (Val—1st position in codon)								G						
8337	tRNA-Lys								T						
8701	ATP6 (Ala—1st position codon)	A	A	A		A	A	A	A	A	A	A	A	A	A

Table 2 (continued)

		hierBAPS Group													
		I	II	III	IV	V	VI	VII		IX	X	XI	XII	XIII	XIV
		U5a2, U5a2b, U5a2c, U5a2d	U5a2e	U5a1,U5a1g, U5a1i	U5b1, U5b1a, U5b1d, U5b1f, U5b1i, U5b3	U5b1 + T16189C! U5b1b, U5b1c	U5a1a2	U5a1h	*RSRS (Ancestral)*	U5a1d	U5a1c	U5a1a1	U5b2a	U5b2, U5b2c	U5b2a2
8705	ATP6 (Met—2nd position codon)								T						
9477	CO3 (Val—1st position codon)	A	A	A	A	A	A	A	G	A	A		A	A	A
9540	CO3 (Leu—1st position codon)	T	T	T	T	T	T	T	T	T	T	T	T	T	T
10,283	ND3 (Leu—3rd position codon)								A						
10,398	ND3 (Ala—1st position codon)	A	A	A	A	A	A	A	A	A	A		A	A	A
10,810	ND4 (Leu—3rd position codon)	T	T	T	T	T		T	T	T	T	T	T	T	T
10,873	ND4 (Pro—3rd position codon)	T	T		T	T	T	T	T	T	T	T	T	T	T
10,915	ND4 (Cys—3rd position codon)		T	T	T	T	T	T	C	T	T	T	T	T	T
10,927	ND4 (Phe—3rd position codon)								T						
11,296	ND4 (Leu—3rd position codon)							T	C						
11,653	ND4 (Val—3rd position codon)								A						
11,914	ND4 (Thr—3rd position codon)	G	G		G		G	G	G	G	G		G	G	G

Table 2 (continued)

		hierBAPS Group													
		I	II	III	IV	V	VI	VII		IX	X	XI	XII	XIII	XIV
		U5a2, U5a2b, U5a2c, U5a2d	U5a2e	U5a1,U5a1g, U5a1i	U5b1, U5b1a, U5b1d, U5b1f, U5b1i, U5b3	U5b1 + T16189C! U5b1b, U5b1c	U5a1a2	U5a1h	*RSRS (Ancestral)*	U5a1d	U5a1c	U5a1a1	U5b2a	U5b2, U5b2c	U5b2a2
11,938	ND4 (Leu—3rd position codon)							T	C						
12,308	tRNA-Leu		G	G		G	G	G	A	G	G	G	G	G	G
12,346	ND5 (His—1st position codon)						T		C						
12,372	ND5 (Leu—3rd position codon)	A	A	A		A	A	A	G	A	A	A	A	A	A
12,406	ND5 (Val—1st position codon)								G						
12,616	ND5 (Leu—1st position codon)								T						
12,618	ND5 (Leu—3rd position codon)							A	G						
12,634	ND5 (Ile—1st position codon)								A						
12,705	ND5 (Ile—3rd position codon)								C						
13,105	ND5 (Val—1st position codon)	A		A	A	A		A	A	A	A	A	A	A	A
13,145	ND5 (Ser—2nd position codon)								G						
13,276	ND5 (Val—2nd position codon)								A						
13,617	ND5 (Ile—3rd position codon)		C	C		C	C	C	T	C	C	C	C	C	C

Table 2 (continued)

		hierBAPS Group													
		I	II	III	IV	V	VI	VII		IX	X	XI	XII	XIII	XIV
		U5a2, U5a2b, U5a2c, U5a2d	U5a2e	U5a1,U5a1g, U5a1i	U5b1, U5b1a, U5b1d, U5b1f, U5b1i, U5b3	U5b1 + T16189C! U5b1b, U5b1c	U5a1a2	U5a1h	*RSRS (Ancestral)*	U5a1d	U5a1c	U5a1a1	U5b2a	U5b2, U5b2c	U5b2a2
13,630	ND5 (Thr—1st position codon)								A						
13,637	ND5 (Gln—2nd position codon)								A				G	C	G
14,182	ND6 (Val—1st position codon)					C			T				C		C
14,518	ND6 (Gly—1st position codon)								A						
14,793	CYB (His—2nd position codon)	G	G				G	G	A	G	G				
15,218	CYB (Thr—1st position codon)						G	G	A	G	G				
15,497	CYB (Gly—1st position codon)								G						
15,511	CYB (Asn—3rd position codon)								T						
15,924	tRNA-Thr								A						
16,114	HVS-I								C						
16,129	HVS-I				G			G	G			G	G	G	G
16,187	HVS-I	C	C		C		C	C	C					C	
16,189	HVS-I						T		T			C!	T!!		
16,192	HVS-I		T					T	C			T		T	
16,223	HVS-I	C	C		C		C	C	T		C			C	C
16,230	HVS-I	A			A		A	A	G		A	A	A	A	A
16,239	HVS-I							T	C						
16,256	HVS-I		T					T	C		T				
16,270	HVS-I								C					T	T

Table 2 (continued)

		hierBAPS Group													
		I	II	III	IV	V	VI	VII		IX	X	XI	XII	XIII	XIV
		U5a2, U5a2b, U5a2c, U5a2d	U5a2e	U5a1,U5a1g, U5a1i	U5b1, U5b1a, U5b1d, U5b1f, U5b1i, U5b3	U5b1 + T16189C! U5b1b, U5b1c	U5a1a2	U5a1h	*RSRS (Ancestral)*	U5a1d	U5a1c	U5a1a1	U5b2a	U5b2, U5b2c	U5b2a2
16,278	HVS-I		C				C	C	C		C		C	C	C
16,294	HVS-I								C						
16,311	HVS-I		C!					T	T		T			T	T
16,320	HVS-I								C		T				
16,362	HVS-I								T						
16,398	HVS-I								G						A
16,399	HVS-I						G	G	A		G				
16,465	HVS-I								C						
16,519	HVS-I							T	T						

	hierBAPS Group									
	XV	XVI	XVII	XVIII	XIX	XX	XXI	XXII	XXIII	XXIV
	U5b2b	U5b2b1	U5b1b1a	U5b1b1	U5b2a1a + C16311T!	U5b2a1a2	U5a2a	U5a2a2a	U5b1e1	U5b1e1 (+T8337C)
Number of shared mitochondrial polymorphisms within each U5 hierBAP group (n)	33	36	20	32	32	41	18	40	33	38
Nucleotide Position										
146	T	T		T	T	T		T	T	T
150	T	T		T	T	T			T	T
151										
152						T		T		
195	T	T			T	T		T		
247	G	G		G		G		G	G	G
523		A		A		A		A		
524		C		C		C		C	d	d
769	G	G	G	G	G	G	G	G	G	G
1303										
1700										
1721	T	T			T	T				
2757									G	G
3027										

Table 2 (continued)

	hierBAPS Group									
	XV	XVI	XVII	XVIII	XIX	XX	XXI	XXII	XXIII	XXIV
	U5b2b	U5b2b1	U5b1b1a	U5b1b1	U5b2a1a+C16311T!	U5b2a1a2	U5a2a	U5a2a2a	U5b1e1	U5b1e1 (+T8337C)
3107						d		d		d
3192										
3197	C	C		C	C	C	C	C	C	C
3212										
3552										
3591										
3768										
4592										
4732					G	G				
5452						T				
5495										
5656			G	G					G	G
7146	A	A	A	A	A	A	A	A	A	A
7256	C	C	C	C	C	C	C	C	C	C
7521		G	G	G	G	G	G	G	G	G
7768	G	G	G	G	G	G			G	G
7853								A		
8337										C
8701	A	A	A	A	A	A	A	A	A	A
8705						C				
9477	A		A	A	A	A	A		A	A
9540	T	T	T	T	T	T		T	T	T
10,283									G	G
10,398	A	A	A	A	A	A	A	A	A	A
10,810	T	T	T	T	T	T	T	T	T	T
10,873	T	T	T	T	T	T	T	T	T	T
10,915	T	T	T	T	T	T	T	T	T	T
10,927				C						
11,296										
11,653	G	G								
11,914	G	G	G	G	G	G		G	G	G
11,938										
12,308	G	G	G	G	G	G	G	G	G	G

Table 2 (continued)

	hierBAPS Group									
	XV	XVI	XVII	XVIII	XIX	XX	XXI	XXII	XXIII	XXIV
	U5b2b	U5b2b1	U5b1b1a	U5b1b1	U5b2a1a+C16311T!	U5b2a1a2	U5a2a	U5a2a2a	U5b1e1	U5b1e1 (+T8337C)
12,346										
12,372	A	A	A	A	A	A	A	A	A	A
12,406								A		
12,616									C	C
12,618			A	A						
12,634	G									
12,705							C			
13,105	A	A	A	A	A	A	A	A	A	A
13,145								A		
13,276							A			
13,617	C	C	C	C	C	C	C	C	C	C
13,630	G	G								
13,637	G		C		G	G				
14,182	C	C		C	C	C			C	C
14,518								G		
14,793							G	G		
15,218										
15,497		A								
15,511					C	C				
15,924						G				
16,114								A		
16,129	G	G				G		G	G	G
16,187	C	C		C	C	C		C	C	C
16,189		T			T!!	T		T		
16,192		C!				C!			C!	C!
16,223	C	C		C	C	C		C	C	C
16,230	A	A		A	A	A		A	A	A
16,239										
16,256								T		
16,270		T						T	T	T
16,278	C	C		C	C	C		C	C	C
16,294								T		
16,311	T			T	C!			T		T

Table 2 (continued)

	hierBAPS Group									
	XV	XVI	XVII	XVIII	XIX	XX	XXI	XXII	XXIII	XXIV
	U5b2b	U5b2b1	U5b1b1a	U5b1b1	U5b2a1a + C16311T!	U5b2a1a2	U5a2a	U5a2a2a	U5b1e1	U5b1e1 (+T8337C)
16,320										
16,362										
16,398										
16,399										
16,465								T		T
16,519										T

[a] The Ancestral state is represented by the RSRS sequence. Blank cells indicate that the nucleotide position was not a factor in determining the hierBAPS group. All BAPS groups also contain the following mutations: 825 T,1018G,2758G,2885 T,3594C,4104A,4312C,8468C,8655C,10664C,10688G,11467G,12705C,13276A,13506C,13650C. Mutations are reckoned in forward evolutionary time direction in reference to the RSRS sequence. In case of a transversion, the derived allele is shown in lowercase instead of uppercase. Exclamation mark signifies back mutation to the ancestral sequence RSRS. (!) for single mutation and (!!) for double back mutation. Yellow-colored boxes indicate mutations that are diagnostic for particular haplogroup or subclade as per Phylotree. *ATP* ATP synthase, *CO* Cytochrome c oxidase, *CYB* Cytochrome b

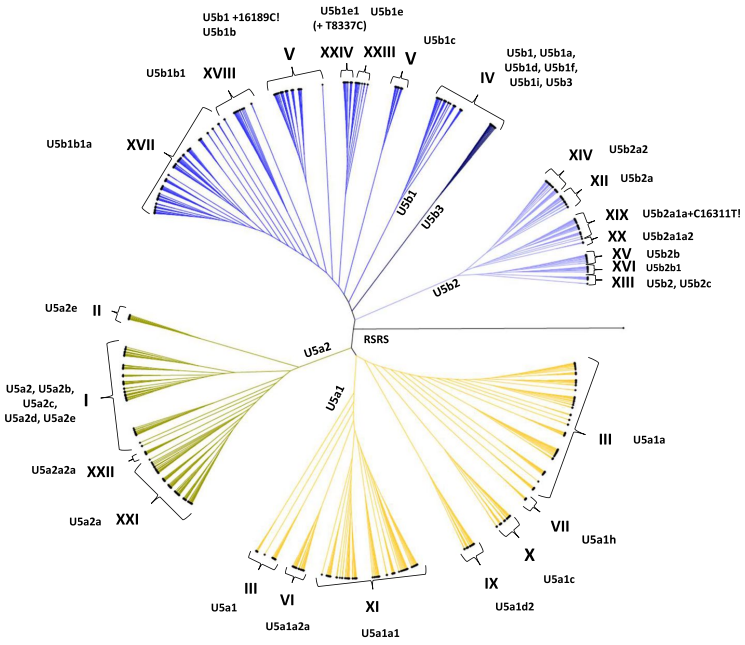

Fig. 1 A phylogenetic tree with 23 hierBAPS groups of haplogroup U5 mitogenome sequences. Roman numerals denote the hierBAPS subclades. The hierBAPS subclades were superimposed on a phylogenetic tree, generated using maximum likelihood analysis, to help visualize the phylogenetic relationships of each sequence. The yellow coloring represents the U5a subhaplogroup while the blue coloring represents the U5b subhaplogroup. RSRS is the Reconstructed Sapiens Reference Sequence

detailed hierBAPS groups, and specific subclades could be identified in accordance with the nomenclature in Phylotree.

The subclades represented by each of the 23 hierBAPS groups were mapped onto a ML phylogeny to determine how well they cohered with the phylogenetic branches produced with this method (Fig. 1). These branches could be subdivided into four main clusters guided by the *Level 1*:4 group analysis: A (U5a1), B (U5a2), C (U5b1 + U5b3), and D (U5b2). Within these main clusters, subclades with nested groups were III (U5a1) and V (U5b1 + T16189C! + T16192C!, U5b1b, U5b1c). Subclade III also had several nested subclades, including VI (U5a1a2a), XI (U5a1a1), IX (U5a1d2), X (U5a1c), and VII (U5a1h), while subclade V consisted of the nested subclades XXIII (U5b1e) and XXIV (U5b1e1 + T8337C).

The ML phylogeny generated from only the coding region of the mitogenome sequences had a similar conformation to that based on whole mitogenome sequences (Figure S2). However, the hierBAPS algorithm was able to identify more specific hierBAPS groups ($n = 23$) for the whole mitogenome sequences compared to the ML phylogeny based on coding-region sequences ($n = 18$). The coding-region hierBAPS groups and their corresponding whole mitogenomes equivalents are shown in Table S1. This table indicates that the hierBAPS groupings are less specific without the non-coding region of the mitogenome sequence.

Geographic distribution of haplogroup U5 and its subclades

To better understand how the U5 phylogeny related to the geographical sources of the mitogenome sequences comprising it, we marked the geographic region from which each mtDNA originated using different colors (Fig. 2). The geographical distribution of the sequences is tabulated in Table S2. For the purposes of this study's focus on northern Europe, the regions are defined by geographic location as follows: Africa (Burkina Faso, Berber, Fulbe, and Fulani ethnic groups), Western Europe (Ireland, Germany, United Kingdom), Southern Europe (France, Italy, Spain, Sardinia), Scandinavia (Denmark, Norway, Sweden), Finland, Saami (includes Saami from Scandinavia and Finland), Central Europe

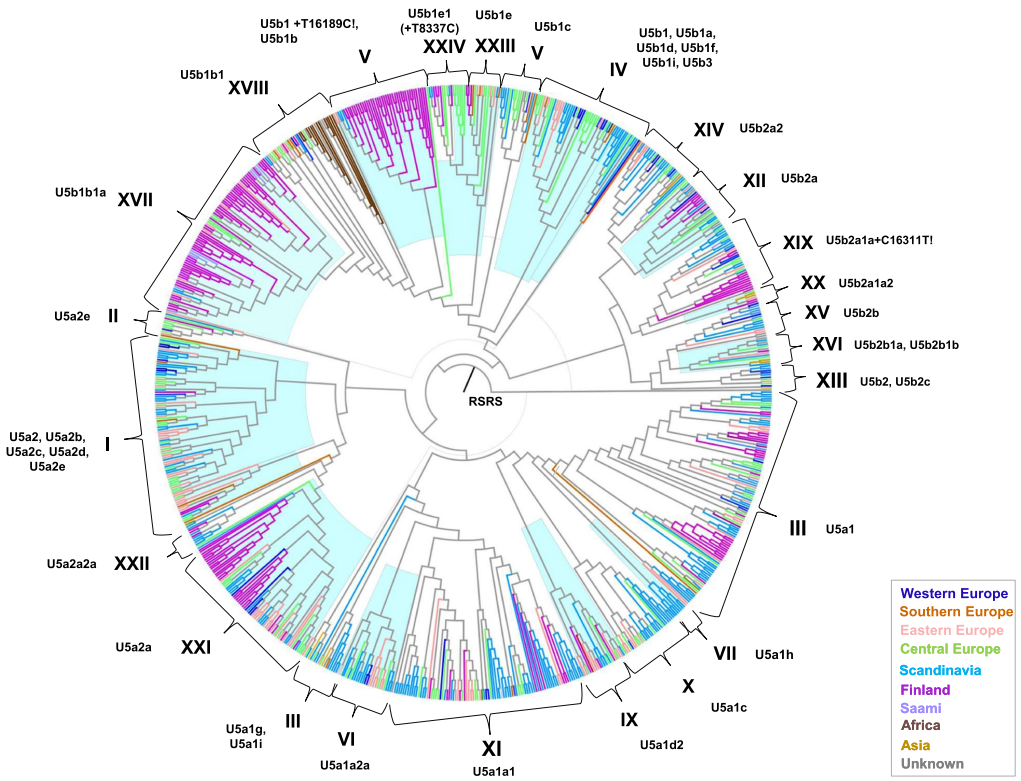

Fig. 2 A phylogenetic tree of haplogroup U5 mitogenome sequences. hierBAPS groups are separated by a light blue watercolor. Roman numerals denote the hierBAPS subclades and their representative Phylotree based subhaplogroups. The geographic regions are defined as: Africa (Burkina Faso, Berber, Fulbe, Fulani), Western Europe (Ireland, Germany, United Kingdom), Southern Europe (France, Italy, Spain, Sardinia), Scandinavia (Denmark, Norway, Sweden), Finland, Saami (includes Saami from Scandinavia and Finland), Central Europe (Czech Republic, Hungary (Roma), Poland, Serbia, Slovenia, Slovakia), Eastern Europe (Baltic, Belarus, Caucasus, Russia), Asia (India, Iran). Unknown origins are colorized in grey

(Czech Republic, Hungary (Roma), Poland, Serbia, Slovenia, Slovakia), Eastern Europe (Baltic, Belarus, Caucasus, Russia), Asia (India, Iran).

Although this phylogenetic tree cannot be interpreted as an exhaustive representation of every known U5 sequence, it nevertheless provided important insights into the way that the hierBAPS groups, each representing U5 subhaplogroups, are regionally related. It also demonstrated that the hierBAPS algorithm, along with ML phylogenetic visualization, can be utilized as a starting point for understanding the divergence of mtDNA haplogroups in evolutionary and geographical terms.

Haplogroup U5b

The phylogenetic groupings produced with the hierBAPS algorithm demonstrated that some sequences specifically clustered by geographic region (Table S3). For example, subclade V contained Central European and Scandinavian branches, including subhaplogroup U5b1c (Age: 12.1 kya; 95% CI:7.7–19.7), and subhaplogroups U5b1 + 16189C! and U5b1b (Ages: 17.6; 95% CI: 10.4–25.9 and 15.4; 95% CI: 19.5–23.2, respectively) (Table S4). Subhaplogroup U5b1e1 (age: 6.4; 95% CI: 4.2–10.1) mtDNAs were also mainly present in Central and Eastern European populations [25]. By contrast, subhaplogroup U5b1e1 sequences were nested between two branches containing Finnish and Scandinavian/Central European mtDNAs, respectively, implying that they were related to both of them.

Subhaplogroup U5b1 branched off between subclade XVIII, which includes 33% of sequences from Africa (subhaplogroup U5b1b1) (age: 12.5 kya; 95% CI:

8.8–18.2) and subclade XVII, comprised of mostly Saami and Finnish mtDNAs (subhaplogroup U5b1b1a) (age: 4.1 kya; 95% CI: 2.7–6.2) sequences. The shared ancestry of U5b1b1 mtDNAs in both the Saami and African populations confirmed findings from an earlier study suggesting that the divergence of these subhaplogroups occurred in southwestern Europe in the Franco-Cantabrian refuge during the Last Glacial Maximum [52]. Subclade XVIII sequences later spread to other African ethnic groups, including the Fulbe, Mande, and other nomadic or pastoral peoples which were part of the former Ghana Empire of Western Africa [53].

A detailed overview of subclade XVII (subhaplogroup U5b1b1a), including the phylogenetic results and the countries in which they occur, is shown in Fig. 3, with age estimate confidence intervals being shown in Table S5. U5b1b1a is found in Finns, Saami, Poles, Belarussians, and Yakuts of eastern Russia, although the vast majority of these mtDNAs appear in the Saami and Finns. While a number of U5b1ba and U5b1b1a1 haplotypes in the Saami and Finns are similar, the Saami have U5b1b1a3 mtDNAs with the A16335G mutation that Finnish populations lack, suggesting they arose in this ethnic group.

We also found other U5 branches within Northern Europe has an estimated age older than subclade XVII (subhaplogroup U5b1b1a) (age: 4.1 kya; 95% CI: 2.7–6.2). For example, subclade XII (subhaplogroup U5b2a) (age: 22.8 kya; 95% CI: 16.3–32.1) mtDNAs were shared by Finns and Scandinavians, while subclade XXI (subhaplogroup U5a2a) (age 17.5 kya; 95% CI: 11.7–25.6) were shared by Finns and Saami. This pattern suggests the presence of subhaplogroups other than U5b1b1 among Saami populations, which may have arrived with populations from Finland or the Scandinavian peninsula.

Haplogroup U5a

Subclade VII (subhaplogroup U5a1h) (age: 1.4 kya; 95% CI: present day-3.7), which includes 45 common polymorphisms, exhibited haplotypes with the diagnostic G1303A, C3192T, T3591A, T4592C, C11296T, C11938T, G12618A, and C16239T motif as well as other polymorphisms which are present in other U5a1 mtDNAs. Subclade U5a1h was present in six samples from Denmark and one from Yorkshire, England, indicating a probable maternal lineage of Viking Age Danish settlers in northwestern England [54].

As expected, hierBAPS groups that occurred earlier in the phylogenetic tree were less geographically specific than later-occurring hierBAPS groups. Subclade III (Age: 20.1; 95% CI: 15.3–28.3), was the most geographically diversified hierBAPS group, and included U5a1 mtDNAs from Southern Europe, Scandinavia, Finland, Central Europe, and Eastern Europe (Fig. 4; Table S6). Subclade X (subhaplogroup U5a1c) (Age: 10.7 kya (95% CI: 5.7–18.1), contained mostly sequences from Eastern and Central Europe with some coming from Denmark. The sequences of U5a1b (Age: 11.0 kya (95% CI: 8.0–16.3) contained several geographic regions, with its distal haplotypes being mostly Scandinavian, Finnish and Eastern European in origin, A similar trans-European clustering was observed for haplogroup U5a1a1 (subclade XI) (Age: 10.9 kya (95% CI: 7.8–16.0)).

Interestingly, subhaplogroups U5a1g (Age: 11.2 kya (95% CI: 6.5–18.6) and U5a1i (Age: 11.9 kya (95% CI: 6.2–19.8), which are found in Iran (Qashqai), India, the Caucasus, and Russia, point to the dispersal of some U5a lineages into eastern regions, as well. These subhaplogroups lacked the extended daughter lineages observed in other subhaplogroups of U5a1. This finding suggested that these lineages did not diversify as successfully as did U5a1b and U5a1a, or else the current sampling of global populations is sufficiently incomplete so as not to reveal any derivative branches. In either case, there is also a lack of daughter haplogroups for U5a1i and U5a1g in Phylotree [2].

Comparative data analysis

To obtain further information about the dispersal of U5 mtDNAs, we aggregated sequence data for this haplogroup from published sources (Table S7 and Supplemental Material 1) and projected them onto a Eurasian map (Fig. 5). Overall, U5 mtDNAs were most prevalent among Saami populations of Norway, Sweden, Finland, and the Kola Peninsula (between 40–64.8%). They were next most frequent in Uralic speakers, mostly Finns (23.1% in higher latitudes of Finland to 15.6% in the southernmost part of the country) [27, 31, 35, 55], and then Estonians, Karelians (16.0%) [31], Mordovians (15.9%) [31], and Russians from the Pskov Oblast (19.2%) [36], the latter region having long barrow burials pointing to early Finnic tribe settlements in the ninth-tenth century [56]. In addition, north-dwelling Norwegians (19.0%) [57–60] and Swedes (16.6%) [31, 55] had a moderate frequency of U5 mtDNAs. As a whole, Finns had a higher proportion of U5 mtDNAs than Scandinavians [37].

We found almost exclusively U5b1 sequences (9/10) in the Saami, and this finding is consistent with previous studies showing that the majority of Saami U5 sequences belonged to this subhaplogroup (about 40–65%, depending on the country) [30, 32, 61]. Even so, we observed a single Saami sequences in subclade XXI (U5a2a), which appears to have separate evolutionary origin from those from the younger subclade XVII (U5b1b1a). It is therefore possible that U5 mtDNAs in the Saami have two sources, the first being Southern Europe via the Franco-Cantabrian

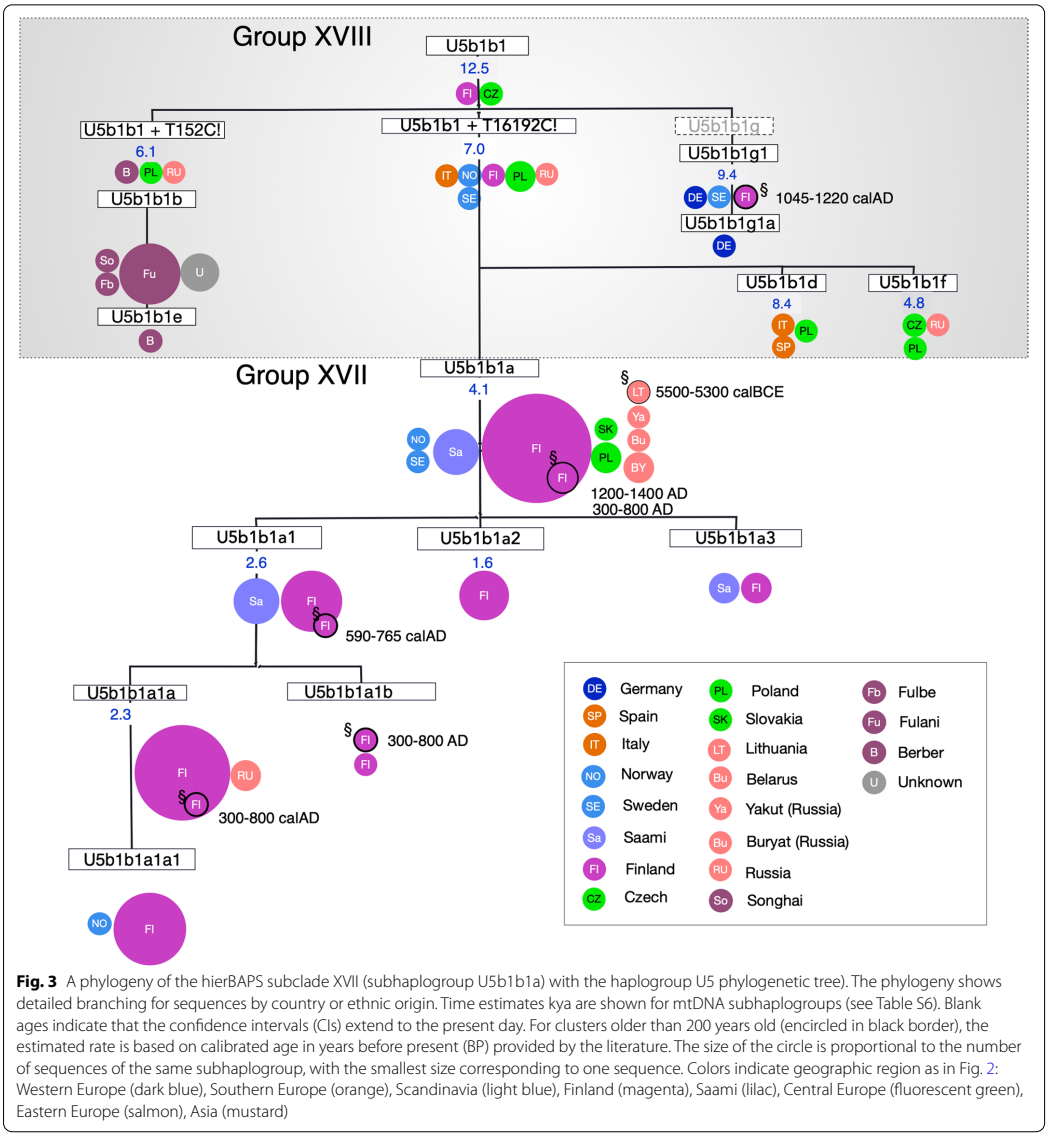

Fig. 3 A phylogeny of the hierBAPS subclade XVII (subhaplogroup U5b1b1a) with the haplogroup U5 phylogenetic tree). The phylogeny shows detailed branching for sequences by country or ethnic origin. Time estimates kya are shown for mtDNA subhaplogroups (see Table S6). Blank ages indicate that the confidence intervals (CIs) extend to the present day. For clusters older than 200 years old (encircled in black border), the estimated rate is based on calibrated age in years before present (BP) provided by the literature. The size of the circle is proportional to the number of sequences of the same subhaplogroup, with the smallest size corresponding to one sequence. Colors indicate geographic region as in Fig. 2: Western Europe (dark blue), Southern Europe (orange), Scandinavia (light blue), Finland (magenta), Saami (lilac), Central Europe (fluorescent green), Eastern Europe (salmon), Asia (mustard)

refuge (U5b1), and the other from Finland and/or Central Europe (U5a2) [subclade XXI]). With regard to the U5a2 sequence, it was detected in a Saami from Finland, and may have entered Northern Europe during 8th to ninth century migrations from Estonia [62].

In populations from Western, Southern and Central Europe, none of the four major subhaplogroups (U5a1, U5a2, U5b1, U5b2) represented more than 50% of the U5 mtDNAs found in those regions. This distribution implies that a greater diversity of U5 subhaplogroups is present in these areas. Since U5a has been most prevalent in Mesolithic Eurasia at approximately 65% [28, 48] and appears to be widespread, it is less clear as to whether this subhaplogroup had a west-to-east

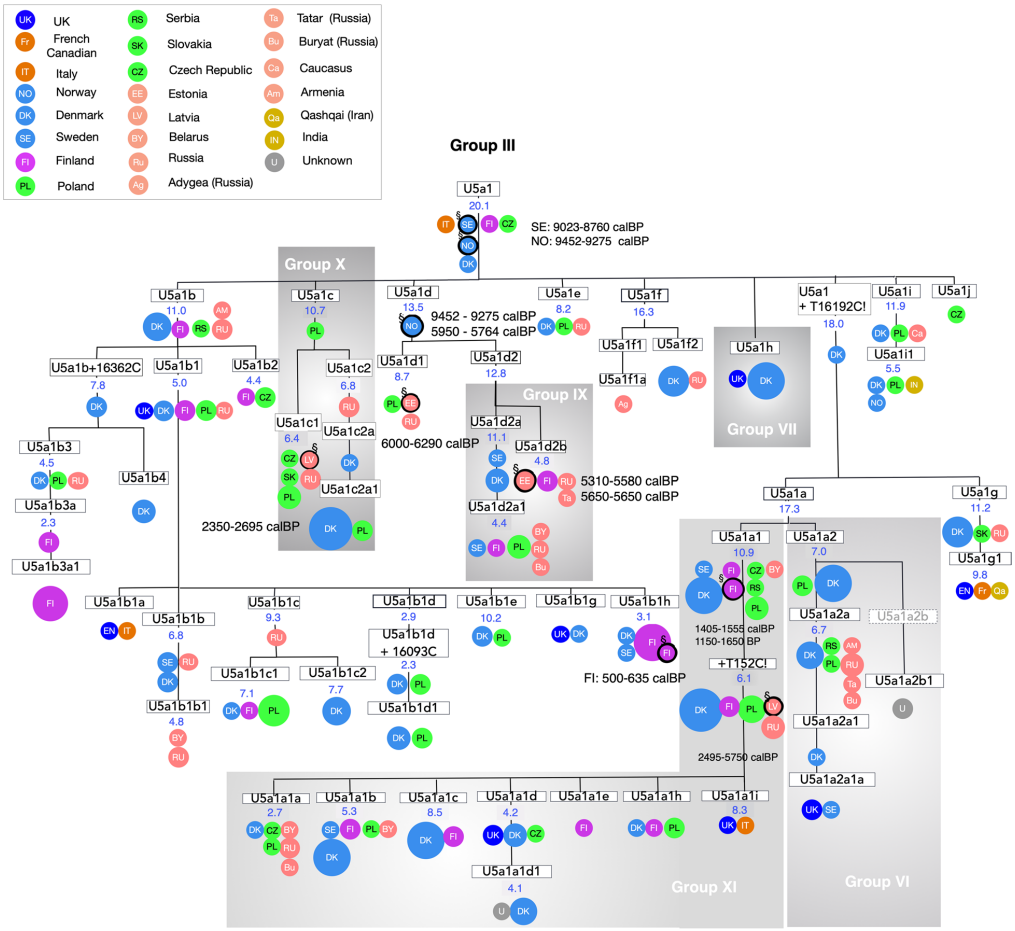

Fig. 4 Phylogeny of the hiers BAPS subclade III (subhaplogroup U5a/U5a1) in the U5 phylogenetic tree). The phylogeny shows detailed branching for each sequence by country or ethnic origin. Time estimates are provided in kya (Table S7). Blank ages indicate confidence intervals (CIs) that extend to the present day. For clusters older than 200 years old (encircled in black border), the estimated rate provided is based on calibrated age in years before present (calBP) provided by the literature. The size of the circle is proportional to the number of sequences of the same subhaplogroup, with the smallest size corresponding to one sequence. Colors indicate geographic region as in Fig. 2: Western Europe (dark blue), Southern Europe (orange), Scandinavia (light blue), Finland (magenta), Saami (lilac), Central Europe (fluorescent green), Eastern Europe (salmon), Asia (mustard)

or east-to-west dispersal. Since we found evidence of its earliest haplogroups across Europe, it is more likely that dispersal happened in both directions.

Subhaplogroup U5b1b diverged and spread in different directions from Europe. According to our survey of GenBank sequences and the results of Achilli and colleagues [52], nearly all African U5 sequences belong to subhaplogroup U5b1b. Its dispersal across North Eurasia and into North Africa suggests that U5b1 had the broadest dispersal of the U5 subhaplogroups.

Discussion

When applied to a dataset of 873 human U5 mitogenome sequences, a combination of hierBAPS clustering with ML analysis accurately reconstructed phylogenetic

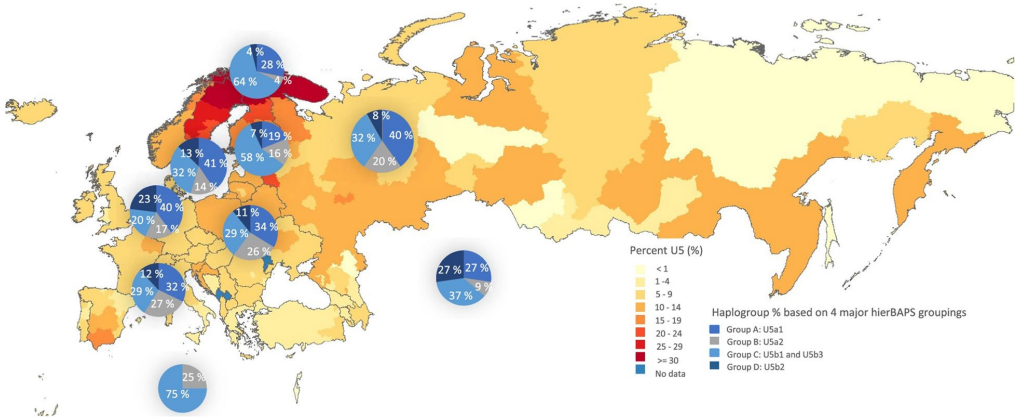

Fig. 5 The frequency of haplogroup U5 mtDNAs in global populations based on the literature (see Table S3). The proportions of each subhaplogroup are listed, based on the four major hierBAPS groups from the FamilyTreeDNA's U5 project. The sample sizes for each data set were as follows: Western Europe ($n=537$), Scandinavia ($n=397$), Sami ($n=78$), Finland ($n=344$), Southern Europe ($n=124$), Central Europe ($n=166$), and Eastern Europe ($n=157$). Countries within Asia ($n=11$) and Africa ($n=4$) were combined due to their small sample sizes, with the los frequency of U5 mtDNAs being supported by the literature (Table S3)

branches that were consistent with the haplogroup U5 phylogeny presented in Phylotree. The findings support the view that the spread of U5 mtDNAs in Northern Europe was skewed from west-to-east through U5b, although some subhaplogroups of U5a found in Northern Europe appear to have been dispersed in both west-to-east and east-to-west directions.

Compared to using Haplogrep2 alone, the hierBAPS groups provide a less tedious, yet accurate method for clustering several haplogroups to investigate population history questions requiring multiple levels of analytical refinement of mtDNA haplogroups. For population genomics, in which several individual sequences are considered simultaneously, this method of mitogenome sequence characterization provides an additional layer for identifying nested genetic population structures separated by allelic patterns. Combining hierBAPS with an ML tree also allows an understanding of similar groups from an evolutionary inference point. To our knowledge, this is the first study to incorporate a hierBAPS analysis with ML phylogenetic tree in a human mtDNA study to investigate historical and evolutionary relationships.

The hierBAPS-ML application

Studies of non-human species that utilized a hierBAPS-based phylogeny vary with respect to the description of the relationships between subclades and the genetic material being analyzed, for example, mtDNA [63] or chloroplast DNA and genomic markers [64–66]. These studies are typically supplemented by additional analyses, such as admixture and estimates of genetic diversity, or the addition of other biomarkers in the population, to draw inferences about their geographical dispersal [64–66].

A recent human mtDNA study used the non-hierarchical version of BAPS in its analysis to identify the origin and genetic affinities of Hill Tribes in Thailand with respect to other Asian populations [67], although a phylogenetic analysis was not undertaken in this study. After mapping the hierBAPS group within each specific population, the authors concluded that, although geographic neighbors were included within the same BAPS groups, it was not possible to draw any conclusions about the regional ancestry of the Hill Tribes. Similarly, mtDNA HVS-I sequences in African Brazilians have been analyzed using the same approach, although this analysis utilized hierBAPS to assess only basic population genetic structure, not the phylogenetic relationships among the sequences or the nested phylogenetic structure that hierBAPS provides [68].

While these studies assessed the genetic structure of the study populations, they were specifically limited in the ability to make evolutionary inferences about the lineages present in them. The incorporation of a rooted ML phylogeny facilitates making temporal inferences about the branching structure by mapping the progression of polymorphisms from an ancestral point-of-reference to the clusters found by the BAPS algorithm.

One of the greatest advantages of integrating hierBAPS algorithmic clustering with phylogenetic analysis is that it

quickly disentangles relationships between large groups of similar sequences that would otherwise be difficult to interpret using haplogroup nomenclature alone. We have observed that the ability to distinguish between similar sequences was more specific when the mtDNA non-coding region was included, and less specific when it was removed. This outcome was expected, considering the high number of mutations that occur in the non-coding region of the human mitogenome [1]. Thus, with respect to mtDNA diversity, the greater the allelic information provided to the hierBAPS algorithm, the more detailed the resulting clustering.

U5 Sequences in Northern Europe

The hierBAPS-ML analysis of haplogroup U5 was especially enhanced when combined with geographic information, age estimates, and U5 demographics. The results of this analysis confirmed a previous study of haplogroup U5 [25], which documented that subhaplogroup U5b1 expanded into Central and Southern Europe before it spread into Western Europe. Our results build upon this earlier study by focusing on the high frequency of U5 mtDNAs within the populations of the Scandinavian Peninsula and Finland, and exploring the geographic sources of the sequences that appear within the phylogeny of U5.

The hierBAPS-ML phylogeny showed that populations from Finland, Scandinavia, North Africa, and Central and Eastern Europe share several U5 subclades/hierBAPS groups. A previous study by Tambets and coworkers [32] found that the geographical source of the Saami-specific U5b1b1 subhaplogroup was difficult to discern. While haplogroup diversification in Southern and Western Europe indicated a west-to-east migration, the observation that the Saami-specific lineages were also present in Uralic-speaking populations of Eastern Europe [32] suggested that U5b1b may have arisen in and spread with these groups [32]. Our results supports the view that U5b1b divergence likely occurred via a scenario in which one subhaplogroup (U5b1b1) became prominent among African populations after hunter-gatherers crossed the Strait of Gibraltar [52]. The other subhaplogroup, U5b1b1a (subclade XVII) became prominent farther north in Scandinavia with the spread of U5b1b1, which eventually gave rise to the "Saami motif" [32, 37]. Furthermore, our phylogenetic tree showed that both lineages were distantly related to the younger subhaplogroups U5b1c and U5b1e1 in Central and Eastern Europe. This finding confirms that the migration of U5b1 mtDNAs likely occurred from west to east rather than the opposite direction.

Studies of the maternal lineages of Saami populations have focused on haplogroups U5b1b1 and V because they are found at the highest frequencies in these and other Scandinavian populations [32, 69]. While U5b1b1 comprises the vast majority of Saami U5 mtDNAs, other haplogroups in Saami populations may potentially have Southern and Central European sources. Lahermo and colleagues found a single U5b sequence, likely U5b3 based on its having the T16304C polymorphism, that was shared by Saami, Finns, and eastern-dwelling circumarctic populations [35]. Our analysis shows that this subhaplogroup is present among modern populations from Southern and Central Europe in addition to Scandinavia, indicating that it has a wide distribution. In fact, U5b3 is found at its highest frequency in Sardinia (3%), although it is the least frequent major U5 subhaplogroup in Europe (< 1% in most populations) [51].

Due to its proximity to Atlantic moisture, the Norwegian shelf was deglaciated between the local LGM and 14–10 thousand calibrated years before present (cal BP) [70], allowing migration from Southern Europe into Northern Europe to occur at that time. While Southern Europe became habitable for settlement during the Last Glacial Maximum, archeological evidence suggests there another co-existing refuge in the so-called "periglacial zone" was located in Ukraine and the West Siberian Plain [71]. Geological evidence supports this view, as ice retreat from the eastern portion of the Fennoscandinavian Ice Sheet led to the formation of large ice-dammed lakes separating the Baltic countries and Russia from Scandinavia [72], preventing early human migrations there. The Baltic Ice Lake persisted until approximately $11{,}620 \pm 100$ cal BP when dissipated, and before the time by which several U5 lineages had already started to expand [73].

Of these lineages, U5a2 constitutes a larger proportion of the U5 sequences in Eastern Europe, while there are also daughter branches of Group XXI containing Scandinavian, Finnish, and one Saami sequence. The earliest dispersals of U5a2 appear to have occurred in Central and Eastern Europe, with later dispersals into Scandinavia/Finland. We also note that some early U5a1 subhaplogroups (namely U5a1g and U5a1i) occur in the east. This second Ukrainian/Pontic refuge is a possible source of some U5 lineages having an eastern geographic origin.

The high frequency of U5, particularly U5b1 among the Saami, appears to be the result of genetic drift [31, 35, 74]. This interpretation is supported by a number of studies based on SNPs, and microsatellite markers which show a high level of linkage disequilibrium among the Saami [74–77] compared to surrounding Scandinavian populations. Most genetic studies further indicate that the Saami population formed as the result of several migration events into Fennoscandia through the coastal edges of land, after which the limited population size had minimally expanded over a long period of time [35, 78].

In this regard, Uralic speakers have been shown to have a distinct ancestral component of Siberian origin [79], with the Saami exhibiting a sizable proportion (13%) of East Eurasian ancestry [80].

It is not until the influx of haplogroups accompanying later dispersals during the Neolithic (approximately 11,000 – 6,500 kya) [81] that there is genetic evidence showing that the predominant U5 subhaplogroups had been diluted in Europe [82]. The Neolithic agriculturalists of central Europe carried mainly N1a, but also H, HV, J, K, T, V, and U3 haplogroups [83]. These Neolithic maternal lineages did not extend as successfully far north, where U5 comprises over 50% of maternal lineages among the Saami. Among Finns and Scandinavians, U5 continues to be the second-most frequent haplogroup after H [30, 31, 33].

Given its widespread distribution in Europe and especially northern Europe, there has been speculation about the possible adaptive features of haplogroup U5 mtDNAs. As an example, nonsynonymous substitutions identified in subclades U5a1 and U5a1a1b were found to arise at the time of maximal decrease in temperature, and suggested to reflect adaptive changes to the cytochrome b and ND5 gene in Europeans during the glaciation period [84]. That is, they were surmised to have produced more uncoupled mtDNAs that generate additional heat as a by-product of normal oxidative metabolism [84]. While these are intriguing results, more work is needed to demonstrate that these variants actually have this physiological effect.

From a clinical standpoint, haplogroup U5 has been linked to a number of complex diseases. For example, a case–control study of 406 patients and 183 healthy controls found a favorable statistical association between haplogroup U5 and the risk of cardiovascular infarction, but a higher risk of a low ventricular ejection fraction (< 40%) [85]. Another study found biological mechanisms that supported higher sperm motility among patients with U5 mtDNAs [86]. A third study found that the parent haplogroup U occurred at high frequency among patients with elevated risk for occipital brain infarct [87], with a related study suggesting that the association was due to a high frequency of haplogroup U5 [88]. Given that these findings are largely correlative in nature, verifying these associations and elucidating the mechanism by which this maternal lineage produces disease phenotypes will be needed to clarify the possible role of haplogroup U5 in human health and disease.

In conclusion, the combined hierBAPS-ML based phylogeny analysis provides insights into the phylogeographic partitioning of genetic diversity, providing a panoramic view of the range of subclades present. Further, it can quickly identify large subclades of related subhaplogroups for population studies that require analysis of a large number of individuals. Combined with archeological evidence, linguistic, and sociocultural knowledge, this methodology provides a visual consolidation of both ancestral and derived features of major mtDNA lineages that can enhance our understanding of human migration history.

Abbreviations

ML: Maximum likelihood; HierBAPS: Hierarchical Bayesian Analysis of Population Structure; Kya: Thousand years ago; RSRS: Reconstructed Sapiens Reference Sequence; Rcrs: Cambridge reference sequence; cal BP: Calibrated years before present.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-022-08572-y.

Additional file 1: Supplementary material 1. References to accompany Table S7

Additional file 2: Figure S1. Three analysis levels of hierBAPS groups superimposed onto a maximum likelihood phylogenic tree. **Figure S2.** Coding region only analysis of hierBAPS group identification using mtDNA. The hierBAPS groups have been superimposed on a phylogenetic tree, generated using maximum likelihood analysis to view the phylogenetic relationships of each sequence

Additional file 3: Table S1. Sensitivity analysis of hierBAPS groupings of human mtDNA U5 haplogroup using coding regions only. **Table S2.** Frequencies and proportions of mitochondrial DNA sequences used for hierBAPS-maximum likelihood evolutionary inferences. **Table S3.** The 24-level hierBAPS groups by geographic region. **Table S4.** Age estimates of the representative subclade(s) of the hierBAPS groups. **Table S5.** Age estimates and confidence intervals for Haplogroups in Figure 3. **Table S6.** Age estimates and confidence intervals for Haplogroups in Figure 4. **Table S7.** Percentage of U5 haplogroup within each population based on published literature in order of highest to lowest. **Table S8.** Mitochondrial DNA sequences used in this analysis.

Authors' contributions

The project was conceived by D.K., J.B., and T.G.S. Data collection, assembly, analysis, and drafting of the original draft was conducted by D.K. The hierBAPS groupings were conducted by D.K. and J.B. The geographic information systems map coordinates were provided by T-T. N. The investigation and drafting of the manuscript were supervised by T.G.S. The funding was acquired by A.J. and T.G.S. All authors read and approved the final manuscript.

Funding

This research was supported by the Research Council of Norway, Grant/Award #:287961.

Availability of data and materials

The mitogenome sequence data that are the focus of this study can be obtained from GenBank and the European Nucleotide Archive. Data from the European Nucleotide Archive are listed in project number PRJEB21940. The GenBank accession numbers for mitogenome sequences reported in this paper are listed in Table S8.

Declarations

Ethics approval and consent to participate

This study is based on open-access and publicly available datasets. The respective studies from which these data are derived have gone through standard protocols to obtain approval from the respective ethics committees

for sample collection and analysis, and to obtain informed consent from participants, as outlined in the associated publications.

Consent for publication
Not applicable.

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

Author details
[1]Center for Fertility and Health, Norwegian Institute of Public Health, Oslo, Norway. [2]Department of Global Public Health and Primary Care, Faculty of Medicine, University of Bergen, Bergen, Norway. [3]Department of Method Development and Analytics, Norwegian Institute of Public Health, Oslo, Norway. [4]IT Systems Bergen, Norwegian Institute of Public Health, Bergen, Norway. [5]Department of Anthropology, University of Pennsylvania, Philadelphia, PA, USA.

Received: 25 November 2021 Accepted: 20 April 2022
Published: 7 May 2022

References

1. Soares P, Ermini L, Thomson N, Mormina M, Rito T, Röhl A, et al. Correcting for purifying selection: an improved human mitochondrial molecular clock. Am J Hum Genet. 2009;84(6):740–59.
2. van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat. 2009;30(2):E386-94.
3. Vianello D, Sevini F, Castellani G, Lomartire L, Capri M, Franceschi C. HAPLOFIND: a new method for high-throughput mtDNA haplogroup assignment. Hum Mutat. 2013;34(9):1189–94.
4. Jagadeesan A, Ebenesersdóttir SS, Guðmundsdóttir VB, Thordardottir EL, Moore KHS, Helgason A. HaploGrouper: a generalized approach to haplogroup classification. Bioinformatics. 2021;37(4):570–2.
5. 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;44(W1):W58-63.
6. Eltsov N, Volodko N. mtPhyl-software tool for human mtDNA analysis and phylogeny reconstruction. [Internet]. 2009 [cited 2022 Mar 17]. Available from: https://sites.google.com/site/mtphyl/home
7. Röck AW, Dür A, Van Oven M, Parson W. Concept for estimating mitochondrial DNA haplogroups using a maximum likelihood approach (EMMA). Forensic Sci Int Genet. 2013;7(6).
8. Kong S, Sánchez-Pacheco SJ, Murphy RW. On the use of median-joining networks in evolutionary biology. Cladistics. 2016;32(6):691–9.
9. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol. 2015;32(1):268–74.
10. Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018;35(2):518–22.
11. Malyarchuk B, Litvinov A, Derenko M, Skonieczna K, Grzybowski T, Grosheva A, et al. Mitogenomic diversity in Russians and Poles. Forensic Sci Int Genet. 2017;30.
12. Davidovic S, Malyarchuk B, Aleksic J, Derenko M, Topalovic V, Litvinov A, et al. Mitochondrial super-haplogroup U diversity in Serbians. Ann Hum Biol. 2017;44(5).
13. Sahakyan H, Kashani BH, Tamang R, Kushniarevich A, Francis A, Costa MD, et al. Origin and spread of human mitochondrial DNA haplogroup U7. Sci Rep. 2017;7.
14. Malyarchuk B, Derenko M, Denisova G, Litvinov A, Rogalla U, Skonieczna K, et al. Whole mitochondrial genome diversity in two Hungarian populations. Mol Genet Genomics. 2018;293(5).
15. Davidovic S, Malyarchuk B, Grzybowski T, Aleksic JM, Derenko M, Litvinov A, et al. Complete mitogenome data for the Serbian population: the contribution to high-quality forensic databases. Int J Legal Med. 2020;134(5).
16. Cheng L, Connor TR, Sirén J, Aanensen DM, Corander J. Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. Mol Biol Evol. 2013;30(5):1224–8.
17. Smith JT, Amador S, McGonagle CJ, Needle D, Gibson R, Andam CP. Population genomics of Staphylococcus pseudintermedius in companion animals in the United States. Commun Biol. 2020;3(1):1–11.
18. Suárez-Esquivel M, Hernández-Mora G, Ruiz-Villalobos N, Barquero-Calvo E, Chacón-Díaz C, Ladner JT, et al. Persistence of brucella abortus lineages revealed by genomic characterization and phylodynamic analysis. PLoS Negl Trop Dis. 2020;14(4): e0008235.
19. Tonkin-Hill G, Lees JA, Bentley SD, Frost SDW, Corander J. RhierBAPs: An R implementation of the population clustering algorithm hierbaps. Wellcome Open Res [Internet]. 2018 [cited 2021 Jun 21];3(93). Available from: /pmc/articles/PMC6178908/
20. van Hal SJ, Willems RJL, Gouliouris T, Ballard SA, Coque TM, Hammerum AM, et al. The interplay between community and hospital Enterococcus faecium clones within health-care settings: a genomic analysis. The Lancet Microbe. 2022;3(2):e133–41.
21. Posth C, Renaud G, Mittnik A, Drucker DG, Rougier H, Cupillard C, et al. Pleistocene mitochondrial genomes suggest a single major dispersal of non-africans and a late glacial population turnover in Europe. Curr Biol. 2016;26(6):827–33.
22. Günther T, Malmström H, Svensson EM, Omrak A, Sánchez-Quinto F, Kılınç GM, et al. Population genomics of Mesolithic Scandinavia: Investigating early postglacial migration routes and high-latitude adaptation. Barton N, editor. PLOS Biol [Internet]. 2018 Jan 9 [cited 2021 Jan 11];16(1):e2003703. Available from: https://dx.plos.org/https://doi.org/10.1371/journal.pbio.2003703
23. Juras A, Chyleński M, Ehler E, Malmström H, Żurkiewicz D, Włodarczak P, et al. Mitochondrial genomes reveal an east to west cline of steppe ancestry in Corded Ware populations. Sci Rep. 2018;8(1).
24. 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;90(4):675–84.
25. Malyarchuk B, Derenko M, Grzybowski T, Perkova M, Rogalla U, Vanecek T, et al. The peopling of Europe from the mitochondrial haplogroup U5 perspective. PLoS One. 2010;5(4):10285.
26. Richards MB, Macaulay VA, Bandelt H-J, Sykes BC. Phylogeography of mitochondrial DNA in western Europe. Ann Hum Genet. 1998;62(3):241–60.
27. Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, et al. Tracing european founder lineages in the near eastern mtDNA pool. Am J Hum Genet. 2000;67(5):1251–76.
28. Bramanti B, Thomas MG, Haak W, Unterlaender M, Jores P, Tambets K, et al. Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. Science (80-). 2009;326(5949):137–41.
29. Röhl A, Brinkmann B, Forster L, Forster P. An annotated mtDNA database. Int J Legal Med. 2001;115(1).
30. Dupuy BM, Olaisen B. mtDNA sequences in the Norwegian Saami and main populations. In: Carracedo A., Brinkmann B. BW, editor. Advances in Forensic Haemogenetics [Internet]. 1st ed. Berlin, Heidelberg: Springer; 1996 [cited 2021 Feb 4]. p. 23–5. Available from: https://link.springer.com/chapter/https://doi.org/10.1007/978-3-642-80029-0_6
31. Sajantila A, Lahermo P, Anttinen T, Lukka M, Sistonen P, Savontaus ML, et al. Genes and languages in Europe: An analysis of mitochondrial lineages. Genome Res. 1995;5(1):42–52.
32. Tambets K, Rootsi S, Kivisild T, Help H, Serk P, Loogväli EL, et al. The western and eastern roots of the Saami - the story of genetic "outliers" told by mitochondrial DNA and Y chromosomes. Am J Hum Genet. 2004;74(4):661–82.
33. Kristjansson D, Bohlin J, Jugessur A, Schurr TG. Matrilineal diversity and population history of Norwegians. Am J Phys Anthropol. 2021;176:120–33.
34. Simoni L, Calafell F, Pettener D, Bertranpetit J, Barbujani G. Geographic patterns of mtDNA diversity in Europe. Am J Hum Genet. 2000;66(1):262–78.
35. Lahermo P, Sajantila A, Sistonen P, Lukka M, Aula P, Peltonen L, et al. The genetic relationship between the Finns and the Finnish Saami (Lapps): Analysis of nuclear DNA and mtDNA. Am J Hum Genet. 1996;58(6):1309–22.

36. Malyarchuk B, Derenko M, Grzybowski T, Lunkina A, Czarny J, Rychkov S, et al. Differentiation of mitochondrial DNA and Y chromosomes in Russian populations. Hum Biol. 2004;76(6):877–900.
37. Meinilä M, Finnilä S, Majamaa K. Evidence for mtDNA admixture between the Finns and the Saami. Hum Hered. 2001;52(3):160–70.
38. Ingman M, Gyllensten U. A recent genetic link between Sami and the Volga-Ural region of Russia. Eur J Hum Genet. 2007;15(1):115–20.
39. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 2003;52(5):696–704.
40. Darriba D, Taboada GL, Doallo R, Posada D. JModelTest 2: More models, new heuristics and parallel computing. Vol. 9, Nature Methods. 2012. p. 772.
41. Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst Biol. 2006;55(4):539–52.
42. Minh BQ, Nguyen MAT, Von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol. 2013;30(5):1188–95.
43. Corander J, Marttinen P, Sirén J, Tang J. Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. BMC Bioinformatics. 2008;9.
44. Treangen TJ, Ondov BD, Koren S, Phillippy AM. The harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Genome Biol. 2014;15(11):1–15.
45. Rambaut A, Lam TT, Carvalho LM, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol. 2016;2(1).
46. To TH, Jung M, Lycett S, Gascuel O. Fast Dating Using Least-Squares Criteria and Algorithms. Syst Biol. 2016;65(1).
47. Översti S, Majander K, Salmela E, Salo K, Arppe L, Belskiy S, et al. Human mitochondrial DNA lineages in Iron-Age Fennoscandia suggest incipient admixture and eastern introduction of farming-related maternal ancestry. Sci Rep. 2019;9(1):1–14.
48. Mittnik A, Wang CC, Pfrengle S, Daubaras M, Zariņa G, Hallgren F, et al. The genetic prehistory of the Baltic Sea region. Nat Commun. 2018;9(1):1–11.
49. FamilyTreeDNA. FamilyTreeDNA - The U5 Project. FamilyTreeDNA. 2021.
50. RStudio Team. RStudio: Integrated Development for R. [Internet]. Boston, MA: RStudio, PBC; 2020. Available from: http://www.rstudio.com/.
51. Pala M, Achilli A, Olivieri A, Kashani BH, Perego UA, Sanna D, et al. Mitochondrial Haplogroup U5b3: A Distant Echo of the Epipaleolithic in Italy and the Legacy of the Early Sardinians. Am J Hum Genet. 2009;84(6).
52. Achilli A, Rengo C, Battaglia V, Pala M, Olivieri A, Fornarino S, et al. Saami and Berbers - An unexpected mitochondrial DNA link. Am J Hum Genet. 2005;76(5):883–6.
53. Rosa A, Brehm A, Kivisild T, Metspalu E, Villems R. MtDNA profile of West Africa Guineans: Towards a better understanding of the Senegambia region. Ann Hum Genet. 2004;68(4):340–52.
54. Hadley DM. Viking and native: Re-thinking identity in the Danelaw. Early Mediev Eur. 2002;11(1):45–70.
55. Kittles RA, Bergen AW, Urbanek M, Virkkunen M, Linnoila M, Goldman D, et al. Autosomal, mitochondrial, and Y chromosome DNA variation in Finland: Evidence for a male-specific bottleneck. Am J Phys Anthropol. 1999;108(4):381–99.
56. Tvauri A. Migrants or Natives? The Research History of Long Barrows in Russia and Estonia in the 5th–10th Centuries. 32nd ed. Nuorluoto J, editor. Vol. 32, Slavica Helsingiensia. Helsinki: University of Helsinki; 2007. 247–285 p.
57. Helgason A, Hickey E, Goodacre S, Bosnes V, Stefánsson K, Ward R, et al. mtDNA and the Islands of the North Atlantic: Estimating the proportions of Norse and Gaelic ancestry. Am J Hum Genet. 1998;68(3):723–37.
58. Opdal SHS, Rognum TOT, Vege Å, Stave AKA, Dupuy BMB, Egeland T. Increased number of substitutions in the D-loop of mitochondrial DNA in the sudden infant death syndrome. Acta Paediatr Int J Paediatr. 1998;87(10):1039–44.
59. Passarino G, Cavalleri GL, Lin AA, Cavalli-Sforza LL, Børresen-Dale AL, Underhill PA. Different genetic components in the Norwegian population revealed by the analysis of mtDNA and Y chromosome polymorphisms. Eur J Hum Genet. 1998;10(9):521–9.
60. The Norway DNA Project Group. FamilyTreeDNA - The Norway DNA - Norge Project [Internet]. FamilyTreeDNA. 1998 [cited 2020 Jun 11]. Available from: https://www.familytreedna.com/group-join.aspx?Group=Norway
61. Delghandi M, Utsi E, Krauss S. Saami mitochondrial DNA reveals deep maternal lineage clusters. Hum Hered. 1998;48(2).
62. Kivisild T, Saag L, Hui R, Biagini SA, Pankratov V, D'Atanasio E, et al. Patterns of genetic connectedness between modern and medieval Estonian genomes reveal the origins of a major ancestry component of the Finnish population. Am J Hum Genet. 2021;108(9).
63. Kay C, Williams TA, Gibson W. Mitochondrial DNAs provide insight into trypanosome phylogeny and molecular evolution. BMC Evol Biol. 2020;20(1).
64. Smýkal P, Kenicer G, Flavell AJ, Corander J, Kosterin O, Redden RJ, et al. Phylogeny, phylogeography and genetic diversity of the Pisum genus. Plant Genet Resour Characterisation Util. 2011;9(1):4–18.
65. Afzal-Rafii Z, Dodd RS. Chloroplast DNA supports a hypothesis of glacial refugia over postglacial recolonization in disjunct populations of black pine (Pinus nigra) in western Europe. Mol Ecol. 2007;16(4):723–36.
66. Zachos FE, Frantz AC, Kuehn R, Bertouille S, Colyn M, Niedziałkowska M, et al. Genetic structure and effective population sizes in european red deer (Cervus elaphus) at a continental scale: insights from microsatellite DNA. J Hered. 2016;107(4):318–26.
67. Besaggio D, Fuselli S, Srikummool M, Kampuansai J, Castrì L, Tyler-Smith C, et al. Genetic variation in Northern Thailand Hill Tribes: Origins and relationships with social structure and linguistic differences. BMC Evol Biol. 2007;7(SUPPL. 2):1–10.
68. Gonçalves VF, Carvalho CMB, Bortolini MC, Bydlowski SP, Pena SDJ. The phylogeography of African Brazilians. Hum Hered. 2007;65(1):23–32.
69. Torroni A, Bandelt HJ, Macaulay V, Richards M, Cruciani F, Rengo C, et al. A signal, from human mtDNA, of postglacial recolonization in Europe. Am J Hum Genet. 2001;69(4):844–52.
70. Hughes ALC, Gyllencreutz R, Lohne ØS, Mangerud J, Svendsen JI. The last Eurasian ice sheets - a chronological database and time-slice reconstruction, DATED-1. Boreas. 2016;45(1).
71. Dolukhanov PM. Modern Humans' Expansion in Eurasia: One Flew East. Open Anthropol J. 2008;1(1):26–32.
72. Stroeven AP, Hättestrand C, Kleman J, Heyman J, Fabel D, Fredin O, et al. Deglaciation of Fennoscandia. Quat Sci Rev. 2016;1(147):91–121.
73. Stroeven AP, Heyman J, Fabel D, Björck S, Caffee MW, Fredin O, et al. A new Scandinavian reference 10Be production rate. Quat Geochronol. 2015;29.
74. Ross AB, Johansson Å, Ingman M, Gyllensten U. Lifestyle, genetics, and disease in Sami [Internet]. Vol. 47, Croatian Medical Journal. Medicinska Naklada; 2006 [cited 2021 Jul 5]. p. 553–65. Available from: www.cmj.hr
75. Laan M, Pääbo S. Demographic history and linkage disequilibrium in human populations. Nat Genet. 1997;17(4).
76. Kaessmann H, Zöllner S, Gustafsson AC, Wiebe V, Laan M, Lundeberg J, et al. Extensive linkage disequilibrium in small human populations in Eurasia. Am J Hum Genet. 2002;70(3).
77. Johansson Å, Vavruch-Nilsson V, Edin-Liljegren A, Sjölander P, Gyllensten U. Linkage disequilibrium between microsatellite markers in the Swedish Sami relative to a worldwide selection of populations. Hum Genet. 2005;116(1–2).
78. Sajantila A, Salem AH, Savolainen P, Bauer K, Gierig C, Pääbo S. Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population. Proc Natl Acad Sci U S A. 1996;93(21).
79. Tambets K, Yunusbayev B, Hudjashov G, Ilumäe AM, Rootsi S, Honkola T, et al. Genes reveal traces of common recent demographic history for most of the Uralic-speaking populations. Genome Biol. 2018;19(1):1–20.
80. Huyghe JR, Fransen E, Hannula S, Van Laer L, Van Eyken E, Mäki-Torkko E, et al. A genome-wide analysis of population structure in the Finnish Saami with implications for genetic association studies. Eur J Hum Genet. 2011;19(3).
81. Diamond J, Bellwood P. Farmers and their languages: The first expansions. Vol. 300, Science. 2003.
82. Brandt G, Haak W, Adler CJ, Roth C, Szécsényi-Nagy A, Karimnia S, et al. Ancient DNA reveals key stages in the formation of Central European mitochondrial genetic diversity. Science (80-). 2013;342(6155):257–61.
83. Haak W, Forster P, Bramanti B, … SM-, 2005 U. Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. science.sciencemag.org [Internet]. 2005 [cited 2021 Jul 2];310(5750):1016–8. Available from: https://science.sciencemag.org/content/310/5750/1016.abstract
84. Malyarchuk BA. Adaptive evolution signals in mitochondrial genes of Europeans. Biochem. 2011;76(6).

85. Golubenko M V., Salakhov RR, Makeeva OA, Goncharova IA, Kashtalap V V., Barbarash OL, et al. Association of mitochondrial DNA polymorphism with myocardial infarction and prognostic signs for atherosclerosis. Mol Biol. 2015;49(6).
86. Montiel-Sosa F, Ruiz-Pesini E, Enríquez JA, Marcuello A, Díez-Sánchez C, Montoya J, et al. Differences of sperm motility in mitochondrial DNA haplogroup U sublineages. Gene. 2006;368(1–2).
87. Majamaa K, Turkka J, Kärppä M, Winqvist S, Hassinen IE. The common MELAS mutation A3243G in mitochondrial DNA among young patients with an occipital brain infarct. Neurology. 1997;49(5).
88. Finnilä S, Hassinen IE, Ala-Kokko L, Majamaa K. Phylogenetic network of the mtDNA haplogroup U in northern Finland based on sequence analysis of the complete coding region by conformation-sensitive gel electrophoresis. Am J Hum Genet. 2000;66(3).

Publisher's Note

Paper 2 Supplementary Figures and Materials

Figure S1

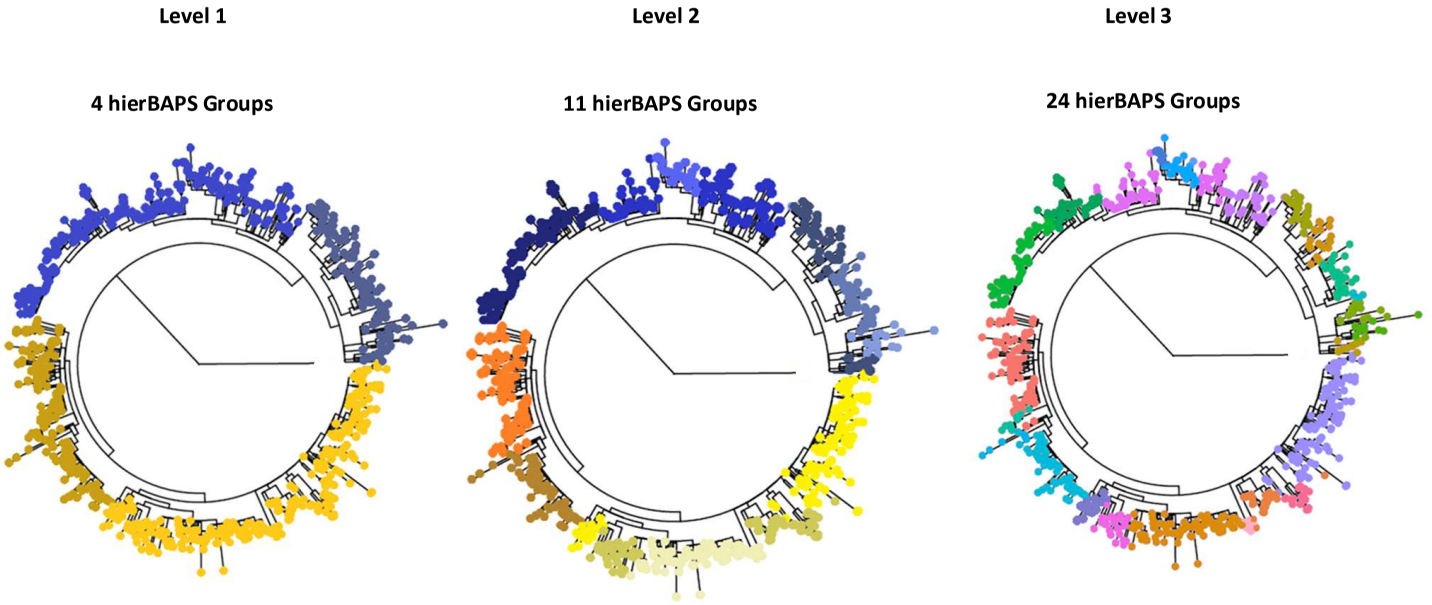

Figure S1. Three analysis levels of hierBAPS groups superimposed onto a maximum likelihood phylogenic tree.

Figure S2

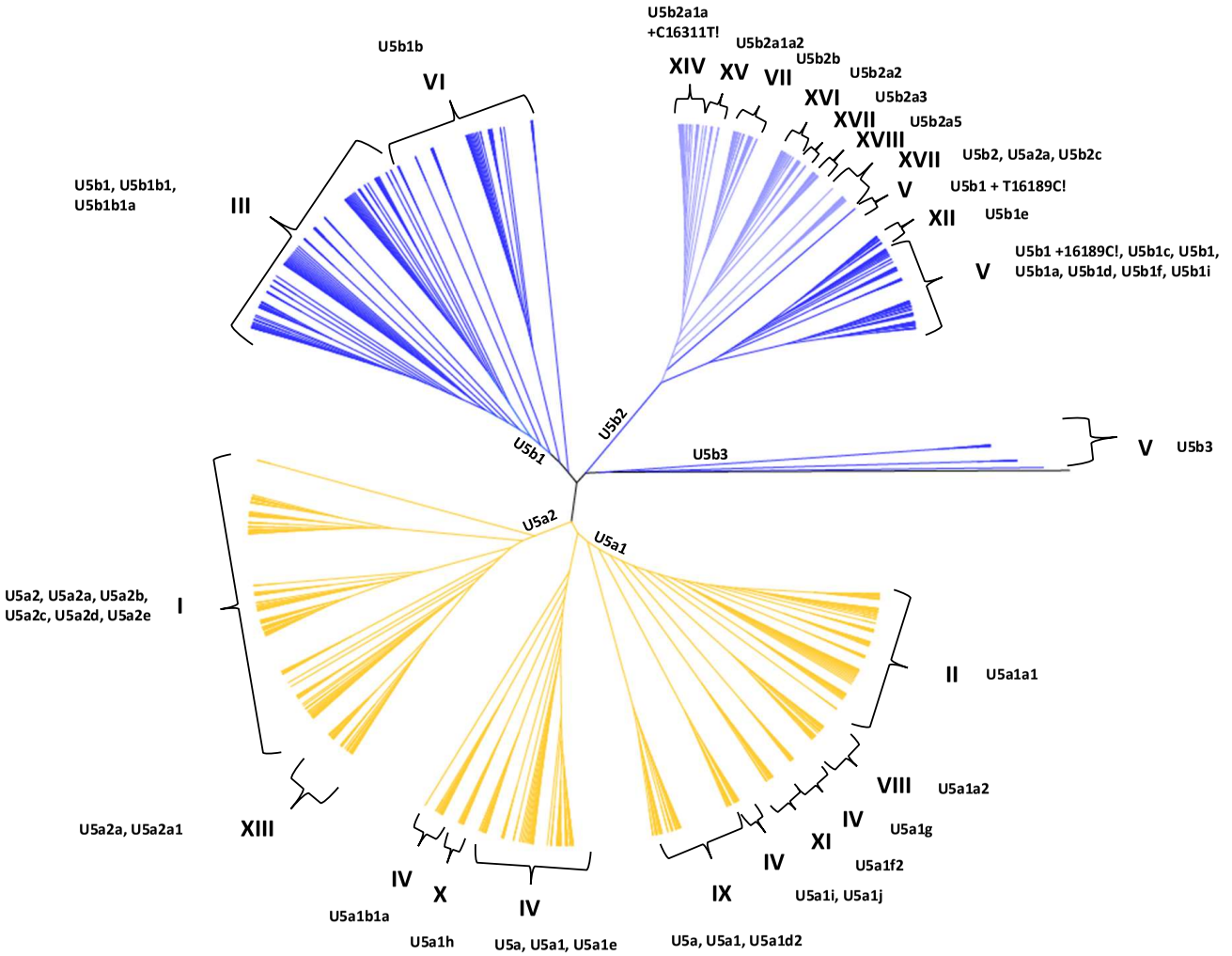

Figure S2. Coding region only analysis of hierBAPS group identification using mtDNA. The hierBAPS groups have been superimposed on a phylogenetic tree, generated using maximum likelihood analysis to view the phylogenetic relationships of each sequence.

Supplementary Material 1. References to accompany Table S7.

Achilli A et al. 2007. Mitochondrial DNA variation of modern Tuscans supports the Near Eastern origin of Etruscans. Am. J. Hum. Genet. 80. doi: 10.1086/512822.

Al-Zahery N et al. 2003. Y-chromosome and mtDNA polymorphisms in Iraq, a crossroad of the early human dispersal and of post-Neolithic migrations. Mol. Phylogenet. Evol. 28. doi: 10.1016/S1055-7903(03)00039-3.

Alfonso-Sánchez MA et al. 2008. Mitochondrial DNA haplogroup diversity in basques: A reassessment based on HVI and HVII polymorphisms. Am. J. Hum. Biol. 20. doi: 10.1002/ajhb.20706.

Alfonso-Sánchez MA et al. 2006. Sequence polymorphisms of the mtDNA control region in a human isolate: The Georgians from Swanetia. J. Hum. Genet. 51. doi: 10.1007/s10038-006-0381-x.

Ahmic, A., Pojskic, N., Silajdzic, E., & Hadziselimovic, R. 2013. A preliminary study of the Paleolithic and Neolithic contribution the European mtdna flow in shaping the genetic structure of recent Bosnian population. *European Scientific Journal*, *9*(36).

Ahmić A, Hadžiselimović R, Silajdžić E, Mujkić I, Pojskić N. 2019. MtDNA variations in three main ethnic populations in Tuzla Canton of Bosnia and Herzegovina. Genet. Appl. 3. doi: 10.31383/ga.vol3iss1pp14-23.

Álvarez-Iglesias V et al. 2009. New population and phylogenetic features of the internal variation within mitochondrial DNA macro-haplogroup R0. PLoS One. 4. doi: 10.1371/journal.pone.0005112.

Baasner A, Schäfer C, Junge A, Madea B. 1998. Polymorphic sites in human mitochondrial DNA control region sequences: Population data and maternal inheritance. Forensic Sci. Int. 98. doi: 10.1016/S0379-0738(98)00163-7.

Babalini C et al. 2005. The population history of the Croatian linguistic minority of Molise (southern Italy): A maternal view. Eur. J. Hum. Genet. 13. doi: 10.1038/sj.ejhg.5201439.

Belledi M et al. 2000. Maternal and paternal lineages in Albania and the genetic structure of Indo-European populations. Eur. J. Hum. Genet. 8. doi: 10.1038/sj.ejhg.5200443.

Belyaeva O et al. 2003. Mitochondrial DNA Variations in Russian and Belorussian Populations. Hum. Biol. 75. doi: 10.1353/hub.2003.0069.

Barbarić L et al. 2020. Maternal perspective of Croatian genetic diversity. Forensic Sci. Int. Genet. 44. doi: 10.1016/j.fsigen.2019.102190.

Bermisheva MA et al. 2004. Phylogeographic analysis of mitochondrial DNA in the Nogays: A strong mixture of maternal lineages from Eastern and Western Eurasia. Mol. Biol. 38:516–523. doi: 10.1023/B:MBIL.0000037003.28999.45.

Bermisheva MA, Tambets K, Villeins R, Khusnutdinova EK. 2002. Diversity of mitochondrial DNA haplogroups in ethnic populations of the Volga-Ural region. Mol. Biol.

36. doi: 10.1023/A:1021677708482.

Bertranpetit J et al. 1995. Human mitochondrial DNA variation and the origin of Basques. Ann. Hum. Genet. 59. doi: 10.1111/j.1469-1809.1995.tb01606.x.

Bini C et al. 2003. Different informativeness of the three hypervariable mitochondrial DNA regions in the population of Bologna (Italy). Forensic Sci. Int. 135. doi: 10.1016/S0379-0738(03)00167-1.

Bogácsi-Szabó E et al. 2005. Mitochondrial DNA of ancient Cumanians: Culturally Asian steppe nomadic immigrants with substantially more western Eurasian mitochondrial DNA lineages. Hum. Biol. 77. doi: 10.1353/hub.2006.0007.

Bonné-Tamir B et al. 2003. Maternal and paternal lineages of the Samaritan isolate: Mutation rates and time to most recent common male ancestor. Ann. Hum. Genet. 67. doi: 10.1046/j.1469-1809.2003.00024.x.

Bosch E et al. 2006. Paternal and maternal lineages in the Balkans show a homogeneous landscape over linguistic barriers, except for the isolated Aromuns. Ann. Hum. Genet. 70. doi: 10.1111/j.1469-1809.2005.00251.x.

Brakez Z et al. 2001. Human mitochondrial DNA sequence variation in the Moroccan population of the Souss area. Ann. Hum. Biol. 28. doi: 10.1080/030144601300119106.

Brandstätter A, Klein R, Duftner N, Wiegand P, Parson W. 2006. Application of a quasi-median network analysis for the visualization of character conflicts to a population sample of mitochondrial DNA control region sequences from southern Germany (Ulm). Int. J. Legal Med. 120. doi: 10.1007/s00414-006-0114-x.

Brandstätter A, Niederstätter H, Pavlic M, Grubwieser P, Parson W. 2007. Generating population data for the EMPOP database-An overview of the mtDNA sequencing and data evaluation processes considering 273 Austrian control region sequences as example. Forensic Sci. Int. 166. doi: 10.1016/j.forsciint.2006.05.006.

Brehm A, Pereira L, Kivisild T, Amorim A. 2003. Mitochondrial portraits of the Madeira and Açores archipelagos witness different genetic pools of its settlers. Hum. Genet. 114. doi: 10.1007/s00439-003-1024-3.

Brisighelli F et al. 2012. Uniparental Markers of Contemporary Italian Population Reveals Details on Its Pre-Roman Heritage. PLoS One. 7. doi: 10.1371/journal.pone.0050794.

Calafell F, Underbill P, Tolun A, Angelicheva D, Kalaydjieva L. 1996. From Asia to Europe: Mitochondrial DNA sequence variability in Bulgarians and Turks. Ann. Hum. Genet. 60. doi: 10.1111/j.1469-1809.1996.tb01170.x.

Cali F et al. 2001. MtDNA control region and RFLP data for Sicily and France. Int. J. Legal Med. 114. doi: 10.1007/s004140000169.

Cardoso S et al. 2011. The maternal legacy of Basques in northern navarre: New insights into the mitochondrial DNA diversity of the Franco-Cantabrian area. Am. J. Phys. Anthropol. 145. doi: 10.1002/ajpa.21532.

Cardoso S et al. 2010. Variability of the entire mitochondrial DNA control region in a human isolate from the pas valley (Northern Spain). J. Forensic Sci. 55. doi: 10.1111/j.1556-4029.2010.01440.x.

Caruana J. 2012. Population Genetics of Western Mediterranean Islands - Malta: a case study. University of Manchester.

Černý V, Hájek M, Čmejla R, Brůžek J, Brdička R. 2004. mtDNA sequences of Chadic-speaking populations from northern Cameroon suggest their affinities with eastern Africa. Ann. Hum. Biol. 31. doi: 10.1080/03014460412331287182.

Cherni L et al. 2005. Female gene pools of Berber and Arab neighboring communities in Central Tunisia: Microstructure of mtDNA variation in North Africa. Hum. Biol. 77. doi: 10.1353/hub.2005.0028.

Cocoş R et al. 2017. Genetic affinities among the historical provinces of Romania and Central Europe as revealed by an mtDNA analysis. BMC Genet. 18. doi: 10.1186/s12863-017-0487-5.

Coia V et al. 2005. Brief Communication: mtDNA variation in North Cameroon: Lack of Asian lineages and implications for back migration from Asia to Sub-Saharan Africa. Am. J. Phys. Anthropol. 128. doi: 10.1002/ajpa.20138.

Coia V et al. 2012. Evidence of high genetic variation among linguistically diverse populations on a micro-geographic scale: A case study of the Italian Alps. J. Hum. Genet. 57. doi: 10.1038/jhg.2012.14.

Comas D et al. 2004. Admixture, migrations, and dispersals in Central Asia: Evidence from maternal DNA lineages. Eur. J. Hum. Genet. 12. doi: 10.1038/sj.ejhg.5201160.

Comas D et al. 1998. Trading genes along the silk road: mtDNA sequences and the origin of central Asian populations. Am. J. Hum. Genet. 63. doi: 10.1086/302133.

Comas D, Calafell F, Bendukidze N, Fañanás L, Bertranpetit J. 2000. Georgian and Kurd mtDNA sequence analysis shows a lack of correlation between languages and female genetic lineages. Am. J. Phys. Anthropol. 112. doi: 10.1002/(SICI)1096-8644(200005)112:1<5::AID-AJPA2>3.0.CO;2-Z.

Côrte-Real HBSM et al. 1996. Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. Ann. Hum. Genet. 60:331–350. doi: 10.1111/j.1469-1809.1996.tb01196.x.

Crespillo M et al. 2000. Mitochondrial DNA sequences for 118 individuals from northeastern Spain. Int. J. Legal Med. 114. doi: 10.1007/s004140000158.

Cvjetan S et al. 2004. Frequencies of mtDNA haplogroups in Southeastern Europe - Croatians, Bosnians and Herzegovinians, Serbians, Macedonians and Macedonian Romani. Coll. Antropol. 28.

Davidovic S et al. 2015. Mitochondrial DNA perspective of serbian genetic diversity. Am. J. Phys. Anthropol. 156. doi: 10.1002/ajpa.22670.

Davidovic S et al. 2020. Complete mitogenome data for the Serbian population: the contribution to high-quality forensic databases. Int. J. Legal Med. 134. doi: 10.1007/s00414-020-02324-x.

Delghandi M, Utsi E, Krauss S. 1998. Saami mitochondrial DNA reveals deep maternal lineage clusters. Hum. Hered. 48. doi: 10.1159/000022789.

Derbeneva O. A., Starikovskaya EB, Volod'ko N V., Wallace DC, Sukernik RI. 2002. Mitochondrial DNA variation in Kets and Nganasans and the early peopling of Eastern Eurasia. Genetika. 38.

Derbeneva O. A., Starikovskaya EB, Volodko N V., Wallace DC, Sukernik RI. 2002. Mitochondrial DNA Variation in the Kets and Nganasans and Its Implications for the Initial Peopling of Northern Eurasia. Russ. J. Genet. 38. doi: 10.1023/A:1021111530654.

Derbeneva Olga A., Starikovskaya EB, Wallace DC, Sukernik RI. 2002. Traces of early Eurasians in the Mansi of northwest Siberia revealed by mitochondrial DNA analysis. Am. J. Hum. Genet. 70. doi: 10.1086/339524.

Destro-Bisol G et al. 2004. The analysis of variation of mtDNA hypervariable region 1 suggests that Eastern and Western Pygmies diverged before the Bantu expansion. Am. Nat. 163. doi: 10.1086/381405.

Di Benedetto G et al. 2001. DNA diversity and population admixture in Anatolia. Am. J. Phys. Anthropol. 115. doi: 10.1002/ajpa.1064.

Dimo-Simonin N, Grange F, Taroni F, Brandt-Casadevall C, Mangin P. 2000. Forensic evaluation of mtDNA in a population from south west Switzerland. Int. J. Legal Med. 113. doi: 10.1007/PL00007715.

Di Rienzo A, Wilson AC. 1991. Branching pattern in the evolutionary tree for human mitochondrial DNA. Proc. Natl. Acad. Sci. U. S. A. 88. doi: 10.1073/pnas.88.5.1597.

Dubut V et al. 2004. mtDNA polymorphisms in five French groups: Importance of regional sampling. Eur. J. Hum. Genet. 12. doi: 10.1038/sj.ejhg.5201145.

Dupuy BM, Olaisen B. 1996. mtDNA sequences in the Norwegian Saami and main populations. In: Advances in Forensic Haemogenetics. Carracedo A., Brinkmann B., BW, editor. Springer: Berlin, Heidelberg pp. 23–25. doi: 10.1007/978-3-642-80029-0_6.

Eduardoff M et al. 2013. Mass spectrometric base composition profiling: Implications for forensic mtDNA databasing. Forensic Sci. Int. Genet. 7. doi: 10.1016/j.fsigen.2013.05.007.

Fadhlaoui-Zid K et al. 2004. Mitochondrial DNA heterogeneity in Tunisian Berbers. Ann. Hum. Genet. 68. doi: 10.1046/j.1529-8817.2004.00096.x.

Falchi A et al. 2006. Genetic history of some western Mediterranean human isolates through mtDNA HVR1 polymorphisms. J. Hum. Genet. 51. doi: 10.1007/s10038-005-0324-y.

FamilyTreeDNA. 2021a. Family Tree DNA - The Saami Project. FamilyTreeDNA. https://www.familytreedna.com/groups/saami/about (Accessed May 31, 2021).

FamilyTreeDNA. 2021b. FamilyTreeDNA - The U5 Project. FamilyTreeDNA.

Fedorova SA, Bermisheva MA, Villems R, Maksimova NR, Khusnutdinova EK. 2003. Analysis of Mitochondrial DNA Lineages in Yakuts. Mol. Biol. 37. doi: 10.1023/A:1025135326954.

Francalacci P, Bertranpetit J, Calafell F, Underhill PA. 1996. Sequence diversity of the control region of mitochondrial DNA in Tuscany and its implications for the peopling of Europe. Am. J. Phys. Anthropol. 100. doi: 10.1002/(SICI)1096-8644(199608)100:4<443::AID-AJPA1>3.0.CO;2-S.

Fraumene C et al. 2006. High resolution analysis and phylogenetic network construction using complete mtDNA sequences in Sardinian genetic isolates. Mol. Biol. Evol. 23. doi: 10.1093/molbev/msl084.

García O et al. 2011. Using mitochondrial DNA to test the hypothesis of a European post-glacial human recolonization from the Franco-Cantabrian refuge. Heredity (Edinb). 106. doi: 10.1038/hdy.2010.47.

Gonzalez AM et al. 2003. Mitochondrial DNA affinities at the atlantic fringe of Europe. Am. J. Phys. Anthropol. 120. doi: 10.1002/ajpa.10168.

Goodacre S et al. 2005. Genetic evidence for a family-based Scandinavian settlement of Shetland and Orkney during the Viking periods. Heredity (Edinb). 95. doi: 10.1038/sj.hdy.6800661.

Gómez-Carballa A, Pardo-Seco J, Amigo J, Martinón-Torres F, Salas A. 2015. Mitogenomes from The 1000 Genome Project reveal new Near Eastern features in present-day Tuscans. PLoS One. 10. doi: 10.1371/journal.pone.0119242.

Graven L et al. 1995. Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large senegalese mandenka sample. Mol. Biol. Evol. 12. doi: 10.1093/oxfordjournals.molbev.a040206.

Grosheva AN, Shneider Y V., Zhukova O V., Morozova IY, Rychkov SY. 2014. Features of the Udmurt mitochondrial gene pool in relation to tribal structure. Russ. J. Genet. 50. doi: 10.1134/S1022795414090063.

Grzybowski T et al. 2007. Complex interactions of the Eastern and Western Slavic populations with other European groups as revealed by mitochondrial DNA analysis. Forensic Sci. Int. Genet. 1. doi: 10.1016/j.fsigen.2007.01.010.

Gubina MA, Damba LD, Babenko VN, Romaschenko AG, Voevoda MI. 2013. Haplotype diversity in mtDNA and Y-chromosome in populations of Altai-Sayan region. Russ. J. Genet. 49. doi: 10.1134/S1022795412120034.

Hedman M et al. 2007. Finnish mitochondrial DNA HVS-I and HVS-II population data. Forensic Sci. Int. 172. doi: 10.1016/j.forsciint.2006.09.012.

Helgason A et al. 2001. mtDNA and the Islands of the North Atlantic: Estimating the

proportions of Norse and Gaelic ancestry. Am. J. Hum. Genet. 68:723–737. doi: 10.1086/318785.

Hernández CL et al. 2014. Human maternal heritage in Andalusia (Spain): Its composition reveals high internal complexity and distinctive influences of mtDNA haplogroups U6 and L in the western and eastern side of region. BMC Genet. 15. doi: 10.1186/1471-2156-15-11.

Hervella M et al. 2014. The Carpathian range represents a weak genetic barrier in South-East Europe. BMC Genet. 15. doi: 10.1186/1471-2156-15-56.

Hofmann S et al. 1997. Population genetics and disease susceptibility: Characterization of central European haplogroups by mtDNA gene mutations, correlation with D loop variants and association with disease. Hum. Mol. Genet. 6. doi: 10.1093/hmg/6.11.1835.

Irwin J et al. 2007. Hungarian mtDNA population databases from Budapest and the Baranya county Roma. Int. J. Legal Med. 121. doi: 10.1007/s00414-006-0128-4.

Irwin J et al. 2008. Mitochondrial control region sequences from northern Greece and Greek Cypriots. Int. J. Legal Med. 122. doi: 10.1007/s00414-007-0173-7.

Jackson BA et al. 2005. Mitochondrial DNA genetic diversity among four ethnic groups in Sierra Leone. Am. J. Phys. Anthropol. 128. doi: 10.1002/ajpa.20040.

Jankova-Ajanovska R et al. 2014. Mitochondrial DNA control region analysis of three ethnic groups in the Republic of Macedonia. Forensic Sci. Int. Genet. 13. doi: 10.1016/j.fsigen.2014.06.013.

Jarczak J et al. 2019. Mitochondrial DNA variability of the Polish population. Eur. J. Hum. Genet. 27. doi: 10.1038/s41431-019-0381-x.

Kasperavičiute D, Kučinskas V, Stoneking M. 2004. Y chromosome and mitochondrial DNA variation in Lithuanians. Ann. Hum. Genet. 68. doi: 10.1046/j.1529-8817.2003.00119.x.

Kittles RA et al. 1999. Autosomal, mitochondrial, and Y chromosome DNA variation in Finland: Evidence for a male-specific bottleneck. Am. J. Phys. Anthropol. 108:381–399. doi: 10.1002/(SICI)1096-8644(199904)108:4<381::AID-AJPA1>3.0.CO;2-5.

Karachanak S et al. 2012. Bulgarians vs the other European populations: A mitochondrial DNA perspective. Int. J. Legal Med. 126. doi: 10.1007/s00414-011-0589-y.

Kivisild T et al. 1999. Deep common ancestry of indian and western-Eurasian mitochondrial DNA lineages. Curr. Biol. 9. doi: 10.1016/S0960-9822(00)80057-3.

Kivisild T et al. 2004. Ethiopian mitochondrial DNA heritage: Tracking gene flow across and around the gate of tears. Am. J. Hum. Genet. 75. doi: 10.1086/425161.

Kouvatsi A, Karaiskou N, Apostolidis A, Kirmizidis G. 2001. Mitochondrial DNA sequence variation in Greeks. Hum. Biol. 73. doi: 10.1353/hub.2001.0085.

Kovacevic L et al. 2014. Standing at the gateway to Europe - The genetic structure of Western Balkan populations based on autosomal and haploid markers. PLoS One. 9. doi: 10.1371/journal.pone.0105090.

Krings M et al. 1999. mtDNA analysis of Nile River Valley populations: A genetic corridor or a barrier to migration? Am. J. Hum. Genet. 64. doi: 10.1086/302314.

Kushniarevich A et al. 2013. Uniparental Genetic Heritage of Belarusians: Encounter of Rare Middle Eastern Matrilineages with a Central European Mitochondrial DNA Pool. PLoS One. 8. doi: 10.1371/journal.pone.0066499.

Lahermo P et al. 2000. MtDNA polymorphism in the Hungarians: Comparison to three other Finno-Ugric-speaking populations. Hereditas. 132. doi: 10.1111/j.1601-5223.2000.00035.x.

Lahermo P et al. 1996. The genetic relationship between the Finns and the Finnish Saami (Lapps): Analysis of nuclear DNA and mtDNA. Am. J. Hum. Genet. 58:1309–1322. /pmc/articles/PMC1915079/?report=abstract (Accessed February 11, 2021).

Lappalainen T et al. 2008. Migration waves to the baltic sea region. Ann. Hum. Genet. 72. doi: 10.1111/j.1469-1809.2007.00429.x.

Larruga JM, Díez F, Pinto FM, Flores C, González AM. 2001. Mitochondrial DNA characterisation of European isolates: The Maragatos from Spain. Eur. J. Hum. Genet. 9. doi: 10.1038/sj.ejhg.5200693.

Lehocký I, Baldovič M, Kádaši Ľ, Metspalu E. 2008. A database of mitochondrial DNA hypervariable regions I and II sequences of individuals from Slovakia. Forensic Sci. Int. Genet. 2. doi: 10.1016/j.fsigen.2007.12.008.

Lembring M, Van Oven M, Montelius M, Allen M. 2013. Mitochondrial DNA analysis of Swedish population samples. Int. J. Legal Med. 127:1097–1099. doi: 10.1007/s00414-013-0908-6.

Lutz S, Weisser HJ, Heizmann J, Pollak S. 1998. Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. Int. J. Legal Med. 111. doi: 10.1007/s004140050117.

Maca-Meyer N et al. 2003. Y chromosome and mitochondrial DNA characterization of Pasiegos, a human isolate from Cantabria (Spain). Ann. Hum. Genet. 67. doi: 10.1046/j.1469-1809.2003.00045.x.

Mairal Q et al. 2013. Linguistic isolates in Portugal: Insights from the mitochondrial DNA pattern. Forensic Sci. Int. Genet. 7. doi: 10.1016/j.fsigen.2013.08.009.

Malyarchuk BA, Derenko M V. 2001. Mitochondrial DNA variability in Russians and Ukrainians: Implication to the origin of the Eastern Slavs. Ann. Hum. Genet. doi: 10.1046/j.1469-1809.2001.6510063.x.

Malyarchuk BA et al. 2002. Mitochondrial DNA variability in Poles and Russians. Ann. Hum. Genet. 66. doi: 10.1017/S0003480002001161.

Malyarchuk B et al. 2004. Differentiation of mitochondrial DNA and Y chromosomes in Russian populations. Hum. Biol. 76:877–900. doi: 10.1353/hub.2005.0021.

Malyarchuk BA et al. 2003. Mitochondrial DNA variability in Bosnians and Slovenians. Ann. Hum. Genet. 67. doi: 10.1046/j.1469-1809.2003.00042.x.

Malyarchuk BA, Vanecek T, Perkova MA, Derenko M V., Sip M. 2006. Mitochondrial DNA variability in the Czech population, with application to the ethnic history of Slavs. Hum. Biol. 78. doi: 10.1353/hub.2007.0014.

Malyarchuk BA et al. 2008. Reconstructing the phylogeny of African mitochondrial DNA lineages in Slavs. Eur. J. Hum. Genet. 16. doi: 10.1038/ejhg.2008.70.

Mateu E et al. 1997. A tale of two islands: Population history and mitochondrial DNA sequence variation of Bioko and Sao Tome, Gulf of Guinea. Ann. Hum. Genet. 61. doi: 10.1017/S0003480097006544.

Meinilä M, Finnilä S, Majamaa K. 2001. Evidence for mtDNA admixture between the Finns and the Saami. Hum. Hered. 52:160–170. doi: 10.1159/000053372.

Messina F, Scorrano G, Labarga CM, Rolfo MF, Rickards O. 2010. Mitochondrial DNA variation in an isolated area of Central Italy. Ann. Hum. Biol. 37. doi: 10.3109/03014461003720304.

Metspalu M et al. 2004. Most of the extant mtDNA boundaries in South and Southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. BMC Genet. 5. doi: 10.1186/1471-2156-5-26.

Mergen H, Öner R, Öner C. 2004. Mitochondrial DNA sequence variation in the Anatolian peninsula (Turkey). J. Genet. 83. doi: 10.1007/BF02715828.

Mielnik-Sikorska M et al. 2013. The History of Slavs Inferred from Complete Mitochondrial Genome Sequences Pereira, LMSM, editor. PLoS One. 8:e54360. doi: 10.1371/journal.pone.0054360.

Mikkelsen M, Sørensen E, Rasmussen EM, Morling N. 2010. Mitochondrial DNA HV1 and HV2 variation in Danes. Forensic Sci. Int. Genet. 4. doi: 10.1016/j.fsigen.2009.07.007.

Modi A et al. 2020. The mitogenome portrait of Umbria in Central Italy as depicted by contemporary inhabitants and pre-Roman remains. Sci. Rep. 10. doi: 10.1038/s41598-020-67445-0.

Mogentale-Profizi N et al. 2001. Mitochondrial DNA sequence diversity in two groups of Italian Veneto speakers from Veneto. Ann. Hum. Genet. 65. doi: 10.1017/S0003480001008545.

Morelli L et al. 2000. Frequency distribution of mitochondrial DNA haplogroups in Corsica and Sardinia. Hum. Biol. 72.

Morozova I et al. 2012. Russian ethnic history inferred from mitochondrial DNA diversity. Am. J. Phys. Anthropol. 147. doi: 10.1002/ajpa.21649.

Nasidze I et al. 2005. Genetic evidence for the Mongolian ancestry of Kalmyks. Am. J. Phys. Anthropol. 128. doi: 10.1002/ajpa.20159.

Nasidze I, Quinque D, Rahmani M, Alemohamad SA, Stoneking M. 2006. Concomitant Replacement of Language and mtDNA in South Caspian Populations of Iran. Curr. Biol. 16. doi: 10.1016/j.cub.2006.02.021.

Nasidze I, Stoneking M. 2001. Mitochondrial DNA variation and language replacements in the Caucasus. Proc. R. Soc. B Biol. Sci. 268. doi: 10.1098/rspb.2001.1610.

Opdal SHS et al. Increased number of substitutions in the D-loop of mitochondrial DNA in the sudden infant death syndrome. 87:1039–1044. http://doi.wiley.com/10.1111/j.1651-2227.1998.tb01410.x (Accessed January 11, 2021).

Orekhov V et al. 1999. Mitochondrial DNA sequence diversity in Russians. FEBS Lett. 445. doi: 10.1016/S0014-5793(99)00115-5.

Ottoni C et al. 2009. Human mitochondrial DNA variation in Southern Italy. Ann. Hum. Biol. 36. doi: 10.3109/03014460903198509.

Parson W, Parsons TJ, Scheithauer R, Holland MM. 1998. Population data for 101 Austrian Caucasian mitochondrial DNA d-loop sequences: Application of mtDNA sequence analysis to a forensic case. Int. J. Legal Med. 111. doi: 10.1007/s004140050132.

Passarino G et al. 2002. Different genetic components in the Norwegian population revealed by the analysis of mtDNA and Y chromosome polymorphisms. Eur. J. Hum. Genet. 10:521–529. doi: 10.1038/sj.ejhg.5200834.

Pardiñas AF, Roca A, Garcia-Vazquez E, Lopez B. 2012. Mitochondrial diversity patterns and the Magdalenian resettlement of Europe: New insights from the edge of the Franco-Cantabrian refuge. J. Hum. Genet. 57. doi: 10.1038/jhg.2012.100.

Pereira V, Gomes V, Amorim A, Gusmão L, Prata MJ. 2010. Genetic characterization of uniparental lineages in populations from Southwest Iberia with past malaria endemicity. Am. J. Hum. Biol. 22. doi: 10.1002/ajhb.21049.

Pereira L et al. 2005. High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium. Genome Res. 15. doi: 10.1101/gr.3182305.

Piercy R, Sullivan KM, Benson N, Gill P. 1993. The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. Int. J. Legal Med. 106. doi: 10.1007/BF01225046.

Picornell A, Gómez-Barbeito L, Tomàs C, Castro JA, Ramon MM. 2005. Mitochondrial DNA HVRI variation in Balearic populations. Am. J. Phys. Anthropol. 128. doi: 10.1002/ajpa.10423.

Plaza S et al. 2003. Joining the pillars of hercules: mtDNA sequences show multidirectional gene flow in the Western Mediterranean. Ann. Hum. Genet. 67. doi: 10.1046/j.1469-1809.2003.00039.x.

Pliss L et al. 2006. Mitochondrial DNA portrait of Latvians: Towards the understanding of the genetic structure of Baltic-speaking populations. Ann. Hum. Genet. 70. doi: 10.1111/j.1469-1809.2005.00238.x.

Poetsch M, Wittig H, Krause D, Lignitz E. 2003. Mitochondrial diversity of a northeast German population sample. Forensic Sci. Int. 137. doi: 10.1016/j.forsciint.2003.06.001.

Prieto L et al. 2011. The GHEP-EMPOP collaboration on mtDNA population data - A new

resource for forensic casework. Forensic Sci. Int. Genet. 5. doi: 10.1016/j.fsigen.2010.10.013.

Pshenichnov A et al. 2013. Genetic affinities of Ukrainians from the maternal perspective. Am. J. Phys. Anthropol. 152. doi: 10.1002/ajpa.22371.

Pult I et al. 1994. Mitochondrial DNA sequences from Switzerland reveal striking homogeneity of European populations. Biol. Chem. Hoppe. Seyler. 375. doi: 10.1515/bchm3.1994.375.12.837.

Quintana-Murci L et al. 2004. Where West Meets East: The Complex mtDNA Landscape of the Southwest and Central Asian Corridor. Am. J. Hum. Genet. 74. doi: 10.1086/383236.

Rando JC et al. 1998. Mitochondrial DNA analysis of Northwest African populations reveals genetic exchanges with European, Near-Eastern, and sub-Saharan populations. Ann. Hum. Genet. 62. doi: 10.1017/S0003480099007241.

Richard C et al. 2007. An mtDNA perspective of French genetic variation. Ann. Hum. Biol. 34:68–79. doi: 10.1080/03014460601076098.

Richards M et al. 1996. Paleolithic and neolithic lineages in the European mitochondrial gene pool. Am. J. Hum. Genet. 59:185. doi: 10.1086/516858.

Richards M et al. 2000. Tracing european founder lineages in the near eastern mtDNA pool. Am. J. Hum. Genet. 67:1251–1276. doi: 10.1016/S0002-9297(07)62954-1.

Rousselet F, Mangin P. 1998. Mitochondrial DNA polymorphisms: A study of 50 French Caucasian individuals and application to forensic casework. Int. J. Legal Med. 111. doi: 10.1007/s004140050174.

Sajantila A et al. 1995. Genes and languages in Europe: An analysis of mitochondrial lineages. Genome Res. 5:42–52. doi: 10.1101/gr.5.1.42.

Sajantila A et al. 1996. Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population. Proc. Natl. Acad. Sci. U. S. A. 93. doi: 10.1073/pnas.93.21.12035.

Santos C et al. 2003. Genetic structure and origin of peopling in the Azores Islands (Portugal): The view from mtDNA. Ann. Hum. Genet. 67. doi: 10.1046/j.1469-1809.2003.00031.x.

Schönberg A, Theunert C, Li M, Stoneking M, Nasidze I. 2011. High-throughput sequencing of complete human mtDNA genomes from the Caucasus and West Asia: High diversity and demographic inferences. Eur. J. Hum. Genet. 19. doi: 10.1038/ejhg.2011.62.

Seehausen O et al. 2003. Nuclear markers reveal unexpected genetic variation and a Congolese-Nilotic origin of the Lake Victoria cichlid species flock. Proc. R. Soc. B Biol. Sci. 270. doi: 10.1098/rspb.2002.2153.

Šarac J et al. 2014. Maternal genetic heritage of southeastern europe reveals a new croatian isolate and a novel, local sub-branching in the X2 haplogroup. Ann. Hum. Genet. 78. doi: 10.1111/ahg.12056.

Simoni L, Calafell F, Pettener D, Bertranpetit J, Barbujani G. 2000. Geographic patterns of

mtDNA diversity in Europe. Am. J. Hum. Genet. 66:262–278. doi: 10.1086/302706.

Stenico M et al. 1996. High mitochondrial sequence diversity in linguistic isolates of the Alps. Am. J. Hum. Genet. 59.

Stevanovitch A et al. 2004. Mitochondrial DNA sequence diversity in a sedentary population from Egypt. Ann. Hum. Genet. 68. doi: 10.1046/j.1529-8817.2003.00057.x.

Stoljarova M, King JL, Takahashi M, Aaspõllu A, Budowle B. 2016. Whole mitochondrial genome genetic diversity in an Estonian population sample. Int. J. Legal Med. 130. doi: 10.1007/s00414-015-1249-4.

Sykes B. 2006. Saxons, Vikings, and Celts: The Genetic Roots of Britain and Ireland. In: W. W. Norton and company: New York pp. 147–164.

Tagliabracci A, Turchi C, Buscemi L, Sassaroli C. 2001. Polymorphism of the mitochondrial DNA control region in Italians. Int. J. Legal Med. 114. doi: 10.1007/s004140000168.

Tambets K et al. 2004. The western and eastern roots of the Saami - the story of genetic 'outliers' told by mitochondrial DNA and Y chromosomes. Am. J. Hum. Genet. 74:661–682. doi: 10.1086/383203.

Tetzlaff S, Brandstätter A, Wegener R, Parson W, Weirich V. 2007. Mitochondrial DNA population data of HVS-I and HVS-II sequences from a northeast German sample. Forensic Sci. Int. 172. doi: 10.1016/j.forsciint.2006.12.016.

Tillmar AO, Coble MD, Wallerström T, Holmlund G. 2010. Homogeneity in mitochondrial DNA control region sequences in Swedish subpopulations. Int. J. Legal Med. 124. doi: 10.1007/s00414-009-0354-7.

Tolk H V. et al. 2000. MtDNA haplogroups in the populations of Croatian Adriatic islands. Coll. Antropol. 24.

Tömöry G et al. 2007. Comparison of maternal lineage and biogeographic analyses of ancient and modern Hungarian populations. Am. J. Phys. Anthropol. 134. doi: 10.1002/ajpa.20677.

Tonks S, Winney B, Evseeva I. 2006. Comparison of Sex-Linked and Autosomal Markers in Orkney and Other North European Populations. Genbank data. www.ncbi.nlm.nih.gov (Accessed May 31, 2021).

Turchi C et al. 2008. Italian mitochondrial DNA database: Results of a collaborative exercise and proficiency testing. Int. J. Legal Med. 122. doi: 10.1007/s00414-007-0207-1.

Turchi C et al. 2016. The mitochondrial DNA makeup of Romanians: A forensic mtDNA control region database and phylogenetic characterization. Forensic Sci. Int. Genet. 24. doi: 10.1016/j.fsigen.2016.06.013.

Varesi L et al. 2000. Mitochondrial control-region sequence variation in the Corsican population, France. Am. J. Hum. Biol. 12. doi: 10.1002/(sici)1520-6300(200005/06)12:3<339::aid-ajhb4>3.0.co;2-u.

Vidrová V et al. 2008. Mitochondrial DNA haplogroups in the Czech population compared to

other European countries. Hum. Biol. 80. doi: 10.3378/1534-6617-80.6.669.

Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC. 1991. African populations and the evolution of human mitochondrial DNA. Science (80-.). 253. doi: 10.1126/science.1840702.

Vona G et al. 2001. Mitochondrial DNA sequence analysis in Sicily. Am. J. Hum. Biol. 13. doi: 10.1002/ajhb.1096.

Watson E et al. 1996. mtDNA sequence diversity in Africa. Am. J. Hum. Genet. 59.

Zgonjanin D et al. 2010. Sequence polymorphism of the mitochondrial DNA control region in the population of Vojvodina Province, Serbia. Leg. Med. 12. doi: 10.1016/j.legalmed.2009.10.007.

Zimmermann B et al. 2007. Mitochondrial DNA control region population data from Macedonia. Forensic Sci. Int. Genet. 1. doi: 10.1016/j.fsigen.2007.03.002.

Zupan A, Hauptman N, Glavač D. 2016. The maternal perspective for five Slovenian regions: The importance of regional sampling. Ann. Hum. Biol. 43. doi: 10.3109/03014460.2015.1006678.

Graphic design: Communication Division, UiB / Print: Skipnes Kommunikasjon AS

uib.no

ISBN: 9788230849019 (print)
9788230845561 (PDF)