UNIVERSITÀ
DI PAVIA

Dipartimento di Biologia e Biotecnologie "Lazzaro Spallanzani"

Laurea Magistralis in Molecular Biology and Genetics

Refining Ukraine's genetic history through high-resolution mitogenome analysis

Un approfondimento sulla storia genetica dell'Ucraina mediante analisi ad alta risoluzione sul mitogenoma

Supervisor:
Prof. Alessandro Achilli

Co-supervisor:
Dr. Nataliia Kozak

Experimental thesis by
Aleksandro Cuci

Academic Year 2023/2024

RIASSUNTO

Comprendere la struttura genetica delle popolazioni umane permette di ottenere importanti informazioni sulla loro storia evolutiva, sui modelli di migrazione e sugli scambi culturali. Situata all'intersezione tra Europa e Asia, l'Ucraina ha da sempre rappresentato un percorso fondamentale per gli spostamenti umani attraverso la steppa eurasiatica. La sua posizione strategica, sulle rive settentrionali del Mar Nero e nei pressi della steppa pontico-caspica, l'ha resa un rifugio cruciale durante l'ultimo Massimo Glaciale, nonché un punto di partenza per la successiva ripopolazione dell'Europa. Nel corso dei millenni, l'Ucraina ha assistito alla nascita delle prime civiltà slave e dei tatari di Crimea, a numerose ondate migratorie e a frequenti trasformazioni culturali e politiche, contribuendo alla formazione di un paesaggio genetico ricco e complesso. Nonostante questo ruolo centrale, regioni specifiche come l'oblast di Donetsk rimangono scarsamente rappresentate negli studi di genetica di popolazione. Questa tesi indaga la variabilità genetica mitocondriale della popolazione di Donetsk attraverso l'analisi combinata di mitogenomi completi e dati genealogici, con l'obiettivo di ricostruire la struttura genetica regionale ed esplorare la continuità delle linee materne attraverso confini storici e geopolitici.

Lo studio ha analizzato una coorte di 104 individui, campionati alla fine degli anni '90, dai quali sono state raccolte informazioni genealogiche dettagliate per risalire ai cosiddetti *Terminal Maternal Ancestors* (TMA), ovvero le ultime antenate materne identificabili attraverso la genealogia. Sebbene la maggior parte dei TMA fosse localizzata all'interno dell'Ucraina, circa un terzo proveniva dalla Russia occidentale, in linea con le dinamiche storiche di migrazione e l'interconnessione culturale tra queste regioni. Il sequenziamento massivo dell'intero genoma mitocondriale è stato eseguito su 94 campioni di alta qualità. Gli indici di diversità hanno mostrato un'elevata diversità a livello di aplotipi (Hd = 1,0) e una bassa diversità nucleotidica (π = 0,00178), una caratteristica genetica tipica di popolazioni che hanno attraversato colli di bottiglia demografici seguiti da rapide espansioni. Questo scenario è stato ulteriormente supportato da valori significativamente negativi del test di Tajima's D, osservati in entrambi i gruppi TMA ucraini e russi, indicativi di un eccesso di varianti rare e coerenti con possibili eventi di deriva genetica. Le analisi comparative basate su Fst, AMOVA e test del chi-quadro hanno mostrato una bassa differenziazione genetica complessiva tra i TMA ucraini e russi, riflettendo un'origine materna condivisa. Tuttavia, la composizione degli aplogruppi ha rivelato alcune differenze interessanti: aplogruppi di origine est-asiatica (D4 e G3) sono stati rilevati esclusivamente tra i TMA ucraini, insieme a sottocladi specifici di aplogruppi euroasiatici occidentali come H10, H11, T2, HV, V e W. Questi dati suggeriscono una maggiore diversità di aplogruppi

all’interno del campione ucraino, potenzialmente legata a storie di mescolanza materna più complesse. L’Analisi delle Componenti Principali (PCA) ha ulteriormente posizionato sia il raggruppamento ucraino che quello russo all’interno del cluster dell’Europa orientale, in prossimità di popolazioni come polacchi, slovacchi e serbi. Le ricostruzioni filogenetiche hanno evidenziato un’elevata eterogeneità mitocondriale e una bassa condivisione di aplotipi, confermando la diversità genetica della regione. Alla luce degli sviluppi geopolitici successivi all'annessione russa del 2022 di alcune aree dell'Ucraina orientale e meridionale, lo studio ha inoltre esaminato la diversità genetica tra individui i cui TMA avevano origine in questi territori annessi, nel resto dell'Ucraina e in diverse regioni della Russia occidentale. L'obiettivo non era valutare gli effetti politici contemporanei, bensì esplorare se eventuali pattern storici di somiglianza o differenziazione genetica potessero corrispondere a queste aree culturalmente e geograficamente significative.

Nel complesso, questo studio fornisce nuove prospettive sulla struttura genetica materna dell’Ucraina sud-orientale. Esso sottolinea il ruolo di Donetsk come regione modellata da millenni di dinamiche demografiche e mette in evidenza il valore dell’integrazione tra dati genetici e genealogici per comprendere i processi storici che hanno plasmato le popolazioni dell’Europa orientale.

TABLE OF CONTENTS

1. ABSTRACT

Understanding the genetic structure of human populations provides critical insights into their evolutionary history, migration patterns, and cultural exchange. Located at the crossroads of Europe and Asia, Ukraine has long served as a key corridor for human movements across the Eurasian steppe. Its strategic location on the northern shores of the Black Sea and near the Pontic-Caspian steppe, made it a crucial refuge during the Last Glacial Maximum and later a launching point for the repopulation of Europe. Over the millennia, Ukraine has witnessed the rise of early Slavic and Crimean Tatar civilizations, waves of migration, and frequent cultural and political shifts, all of which have contributed to a rich and diverse genetic landscape. Despite this central role, specific regions such as Donetsk oblast remain underrepresented in population genetic studies. This thesis investigates the current maternal genetic variation in the Donetsk population through the combined analysis of complete mitochondrial genomes and genealogical data, aiming to reconstruct regional population structure and explore lineage continuity across geopolitical and historical boundaries.

The study analyzed a cohort of 104 individuals, collected in the late 1990s, from whom detailed genealogical information was used to trace Terminal Maternal Ancestors (TMAs). While most TMAs originated in Ukraine, approximately one-third were located in western Russia, an outcome consistent with historical migration dynamics and the cultural interconnection between these regions. High-throughput sequencing of the full mitochondrial genome was performed on 94 high-quality DNA samples. Diversity indices revealed high haplotype diversity (Hd = 1.0) and low nucleotide diversity (π = 0.00178), a genetic signature typically associated with post-bottleneck population expansions. This was further supported by significantly negative Tajima's D values in both Ukrainian and Russian TMA groups, indicating an excess of rare variants and supporting models of genetic drift. Comparative analyses using Fst, AMOVA, and chi-square tests indicated low overall genetic differentiation between Ukrainian and Russian TMAs, reflecting a shared maternal ancestry. However, haplogroup composition revealed subtle distinctions. A few haplogroups of East Asian origin (D4 and G3) were detected exclusively among Ukrainian TMAs, alongside unique subclades of Western Eurasian lineages such as H10, H11, T2, HV, V, and W. These findings point to greater haplogroup diversity in the dataset with Ukrainian TMA, potentially reflecting more complex maternal admixture histories. Principal Component Analysis further placed both Ukrainian and Russian TMAs within the Eastern European cluster, closely aligned with populations such as Poles, Slovaks, and Serbs. Phylogenetic reconstruction illustrated extensive lineage diversity and little haplotype sharing, emphasizing the genetic heterogeneity within the region. In light of the

geopolitical developments following the 2022 Russian annexation of parts of eastern and southern Ukraine, the study also examined genetic diversity among individuals whose TMAs originated from these annexed territories, from the rest of Ukraine, and from various regions of western Russia. The goal was not to assess contemporary political effects, but rather to explore whether historical patterns of genetic similarity or differentiation might correspond to these culturally and geographically significant areas.

Overall, this study contributes new insights into the maternal genetic structure of southeastern Ukraine. It underscores Donetsk's role as a region shaped by millennia of demographic movement and highlights the value of integrating genetic and genealogical data in understanding the historical processes that have shaped Eastern European populations.

2. INTRODUCTION

2.1 INTRODUCTION TO POPULATION GENETICS

Since the earliest stages of human intellectual curiosity, our species has sought to reconstruct its origins and understand the biological threads that connect us to one another and to our ancestors. This quest to answer the fundamental questions "Who are we?" and "Where do we come from?" has driven the development of a multidisciplinary scientific approach, integrating anthropology, archaeology, linguistics, history, and increasingly, molecular biology (Cavalli-Sforza & Feldman, 2003). At the heart of this multidisciplinary effort lies population genetics, a branch of genetics that studies the distribution and changes in allele frequencies within populations under the influence of evolutionary forces such as natural selection, genetic drift, mutation, and gene flow. While natural selection describes the differences in survival and reproductive success associated with individual genotypes, genetic drift refers to the random fluctuations of allele frequencies within a population. Together, these two mechanisms primarily determine whether an allele is transmitted through generations within a population (Wright, 1931).

A major empirical leap came with Karl Landsteiner's discovery of the ABO blood group system in 1901, which became the first widely used genetic marker to study human population variation, providing early evidence of biological diversity between groups (Landsteiner, 1901). This was followed by the work of Ludwik and Hanna Hirszfeld, who in 1919 showed that blood group frequencies varied geographically, laying the groundwork for the idea that genetic markers could trace population structure and migration (Hirszfeld & Hirszfeld, 1919). The field expanded further with the development of protein electrophoresis in the mid-20th century, which allowed scientists like Linus Pauling and colleagues to identify variations in hemoglobin molecules, notably linking the sickle-cell trait to a specific molecular difference (Pauling et al., 1949). This discovery was monumental in illustrating that genetic variation was widespread in populations and often had health and evolutionary significance. One of the most transformative moments in biological sciences came with the discovery of the DNA double helix structure by Watson and Crick in 1953, which provided the physical and chemical foundation for understanding heredity (Watson & Crick, 1953). This breakthrough, together with the subsequent development of powerful molecular techniques such as DNA cloning (Cohen *et al*.,1973), Southern blotting analysis (Southern, 1975), the identification of restriction fragment length polymorphisms (RFLPs) (Botstein *et al*., 1980), the first generation of DNA sequencing methods (Sanger & Coulson, 1975), and polymerase chain reaction (PCR) (Mullis *et al*., 1986), enabled scientists to directly analyze genetic material, shifting the focus of population genetics from

phenotypic and protein-based markers to DNA-level investigations. The most recent advancements in population genetics have come with next-generation sequencing (NGS) technologies, which enable whole-genome analyses at an unprecedented scale (Metzker, 2010). These tools have allowed scientists to uncover ancient admixture events, such as the Neanderthal and Denisovan genetic contributions to modern human genomes (Green *et al*., 2010; Reich *et al*., 2010), and to study fine-scale population structure, migration, and adaptation. Together, these milestones have solidified population genetics as an essential field for reconstructing evolutionary histories, understanding human diversity, and applying this knowledge to medicine, anthropology, and conservation biology.

2.2 GENETIC STUDIES ON EUROPE

Europe has been a central focus of population genomics research, making it the most extensively studied continent in this field. It is also the location where the first Neanderthal remains were discovered, providing early evidence that the region was once inhabited by individuals morphologically distinct from modern humans. Between approximately 430 and 39 thousand years ago (kya), the European population included Neanderthals and their ancestors, who were genetically distinct not only from other archaic human groups such as the Denisovans of Siberia, but also from anatomically modern humans.

Modern humans began migrating out of sub-Saharan Africa around 60 kya, during which they admixed with Neanderthals. The earliest evidence for the presence of modern humans in Europe dates to approximately 45 kya, as confirmed by both genetic and archaeological data. However, individuals predating 40 kya did not contribute genetically to present-day populations. The first individuals in Europe who were genetically linked to contemporary populations appeared shortly after this period. In fact, the oldest genomes carrying ancestry components found in modern Europeans date from around 37 kya to 35 kya (Lazaridis, 2018; Posth *et al*., 2023). Prior to the Last Glacial Maximum (LGM; 25–19 kya), a population known as the "Věstonice cluster" inhabited an area extending from the Russian Steppe to central Europe and was associated with the Gravettian culture (33–26 kya). Despite their cultural unity, this population was not biologically homogeneous. Recent genomic data have revealed two main ancestral sources: individuals related to the Goyet cluster contributed to the Gravettian populations in western and southwestern Europe (Fournol cluster), while those from central and southern Europe descended from pre-30 kya eastern European individuals linked to the Kostenki cluster, thereby giving rise to the Věstonice cluster. This group falls outside the genetic variation seen in modern humans and was eventually replaced by post-LGM hunter-gatherers. These

newer populations carried genetic components more closely related to those of the Near East, known as the Villabruna cluster. By around 15 kya, Europe was inhabited primarily by a relatively homogeneous population of hunter-gatherers, considered the first of the three major ancestral components of contemporary Europeans. These groups, associated with the Epigravettian culture, displayed genetic affinities with Siberian hunter-gatherers in eastern Europe. It has been proposed that the Villabruna cluster entered northeastern Italy via the Balkans and expanded southward across the peninsula (Posth *et al*., 2023).

During the Neolithic Revolution (8–9 kya), a period marked by significant cultural transformations including the domestication of plants and animals, these indigenous hunter-gatherers were joined by agricultural communities migrating from the Near East. These early farmers, who settled across mainland Europe, represent the second principal ancestral component of modern European genomes, also referred to as the Oberkassel cluster. Genetic and archaeological evidence suggests these populations originated in the Fertile Crescent, which includes regions of southern Anatolia, present-day western Iran, Iraq, Syria, Lebanon, Israel, Palestine, Jordan and Egypt (Aneli *et al*., 2021). Their migration likely followed two routes: a Mediterranean coastal path (including Italy) and a continental path through the Balkans and the Danube Valley into central and western Europe.

Later admixture occurred with migrants from the Caucasus who moved into eastern Europe, giving rise to the so-called steppe populations. Around 4.9 kya, during the Early Bronze Age, these populations, closely related to the Yamnaya culture from the Pontic-Caspian-Ural steppe, expanded in both eastern and western directions. In central Europe, this ancestry became prominent in populations associated with the Corded Ware culture (Haak *et al*., 2015). The earliest signs of steppe-related ancestry in southeastern Europe date back to approximately 6.7–6.5 kya, with sporadic occurrences in Bulgaria and low-level presence across the Bronze Age (~5.4–3.1 kya) (Mathieson *et al*., 2018). This migration may have been facilitated by technological innovations such as horseback riding and is associated with the spread of Indo-European languages across Europe. Around 4.6 kya, steppe-derived individuals reached the British Isles in conjunction with the spread of the Bell Beaker Complex, almost entirely replacing the local gene pool within just a few centuries (Olalde *et al*., 2018).

The genetic makeup of present-day Europeans thus reflects the admixture of three primary ancestral populations: Upper Paleolithic hunter-gatherers, early Neolithic farmers from the southeast, and Bronze Age steppe migrants (Richards *et al*., 2016), along with additional contributions from more recent migration events (Raveane *et al*., 2019) (Figure 2.1).

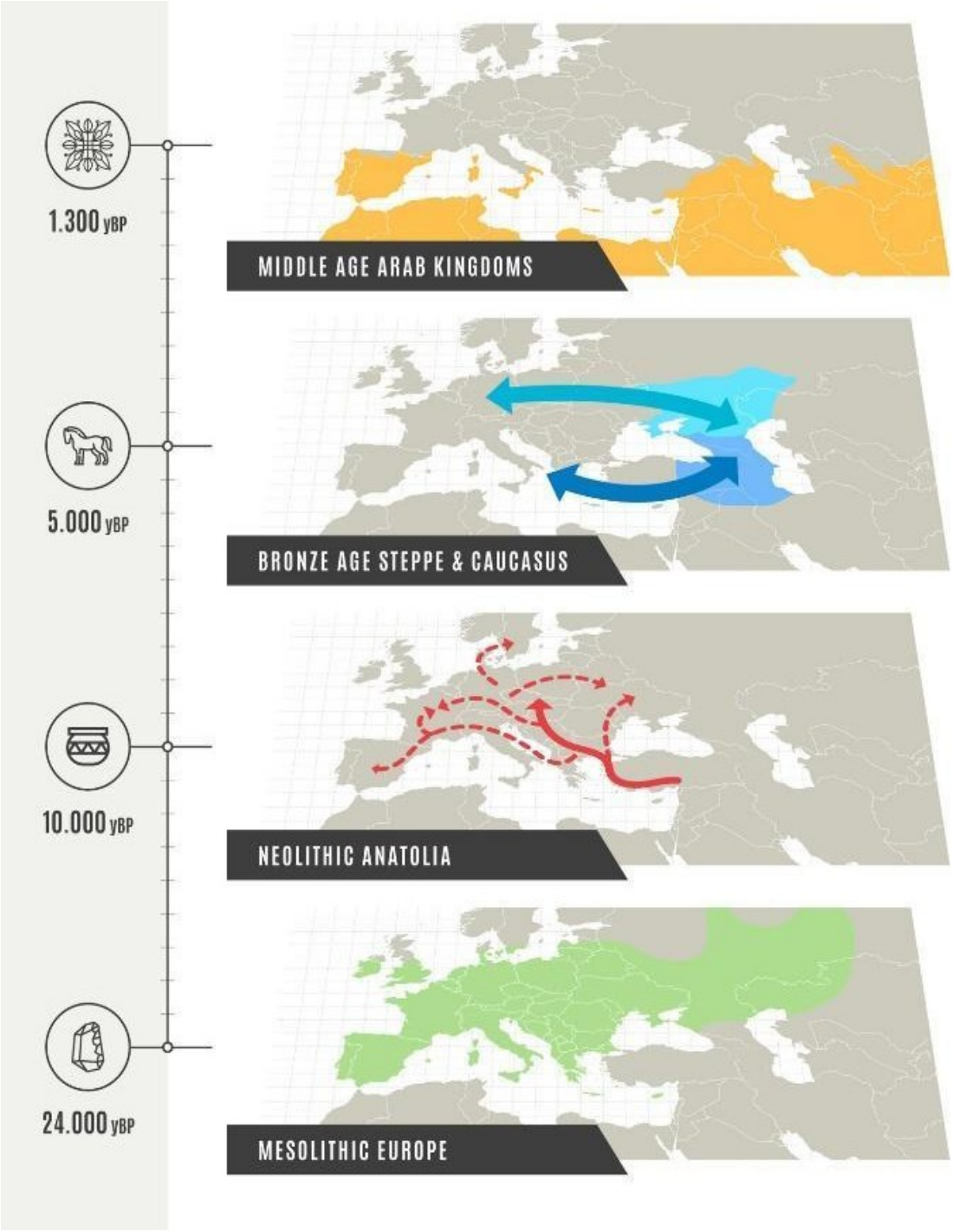


Figure 2.1. *Prehistoric and historical contributions in Europe in chronological order. In green the hunter-gatherers, in red the Neolithic farmers, in blue the two parallel routes of the nomadic shepherds of the steppe towards Nothern Europe and the new signal coming from Anatolia towards Southern Italy. In yellow the genetic heritage of Arab domination in Southern Europe (Press release University of Pavia – credits to A.Corlianò).*

2.3 MITOCHONDRIAL DNA: A TOOL IN POPULATION GENETICS

Over the past three decades, mitochondrial DNA (mtDNA) has been widely used as a powerful tool for inferring the evolutionary and demographic history of populations and species (Avise, 2000). Mitochondria, which are double-membraned organelles present in most eukaryotic cells, house mtDNA in a specialized environment that includes distinct compartments, such as the outer membrane, intermembrane space, inner membrane, cristae and matrix (Figure 2.2A). These structural features play a crucial role in mitochondrial function and, consequently, in the maintenance and inheritance of mtDNA. Mitochondria are primarily responsible for generating energy through oxidative phosphorylation, a process that is carried out by the molecular machinery called the respiratory chain, which produces adenosine triphosphate (ATP), the cell's main energy currency. Beyond energy production, mitochondria also regulate apoptosis, calcium homeostasis, and reactive oxygen species (ROS) generation, all of which are critical for maintaining cellular function (McBride *et al*., 2006). Mitochondria are highly dynamic organelles that constantly undergo fusion and fission, processes essential for maintaining mitochondrial function and quality control (Youle & van der Bliek, 2012). Fusion allows mitochondria to share contents, including enzymes and mitochondrial DNA, helping to mitigate damage, while fission facilitates the removal of defective mitochondria through mitophagy, a specialized form of autophagy that prevents the accumulation of dysfunctional organelles (Twig *et al*., 2008).

In humans, the number of mitochondria per cell varies depending on the cell type and the energy demands of the tissue. This number ranges from none in red blood cells to over 2,000 in liver cells (Figure 2.2B). Within mitochondria, thousands of mtDNA molecules are organized into several hundred nucleoids that can replicate and distribute throughout the mitochondrial network. This dynamic organization suggests that nucleoids function as the fundamental units of mitochondrial inheritance (Garrido *et al*., 2003).

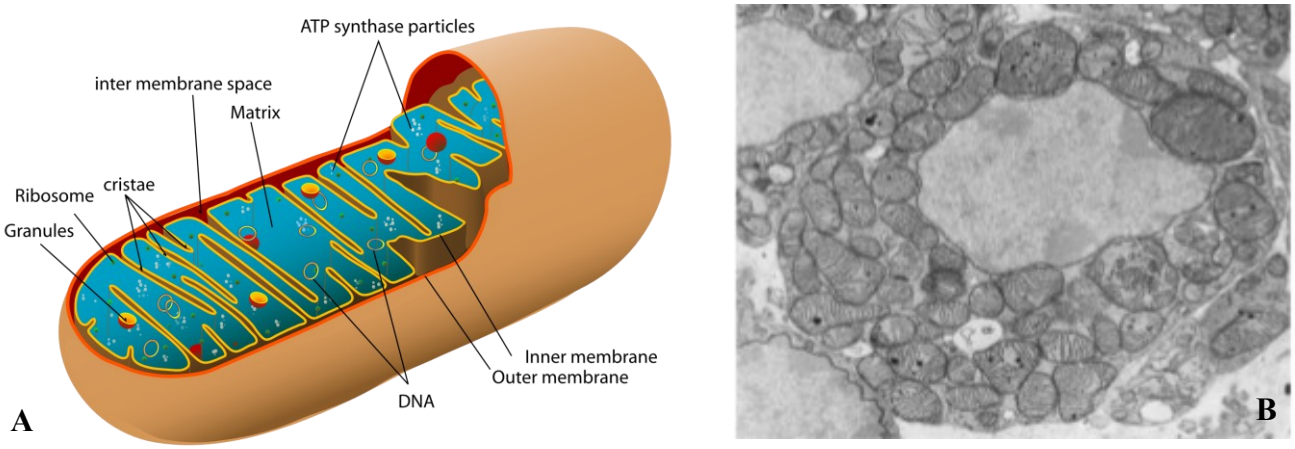


***Figure 2.2**. (A) Structure of the human mitochondrion. (B) Electron microscope cell section view: numerous mitochondria and their internal membranes are visible.*

In most human cells, mitochondria contain 10^3-10^4 copies of this molecule. The first completely sequenced human mtDNA was published in 1981 (Anderson *et al*., 1981). This initial sequence was used as a reference sequence under the name "Cambridge Reference Sequence (CRS)", which was later revised (rCRS) (Andrews *et al*., 1999). It is a circular double-stranded molecule of 16,569 base-pairs (Figure 2.3). It includes 37 genes that encode for two ribosomal RNAs, 22 tRNAs, and 13 polypeptides that form parts of the respiratory chain located in the inner mitochondrial membrane (Iborra *et al*., 2004). The non-coding part mainly consists of a control region, named displacement loop (D-loop). This region is involved in the control of transcription and replication of the molecule. It spans 1,121 bps (nucleotide positions, nps, 16024–576) and it is the most polymorphic region of the mtDNA. Its variability is higher in three hypervariable segments: HVS-I (nps 16024-16400), HVS-II (nps 44-340), and HVS-III (nps 438-576) (Brandstätter *et al*., 2004).

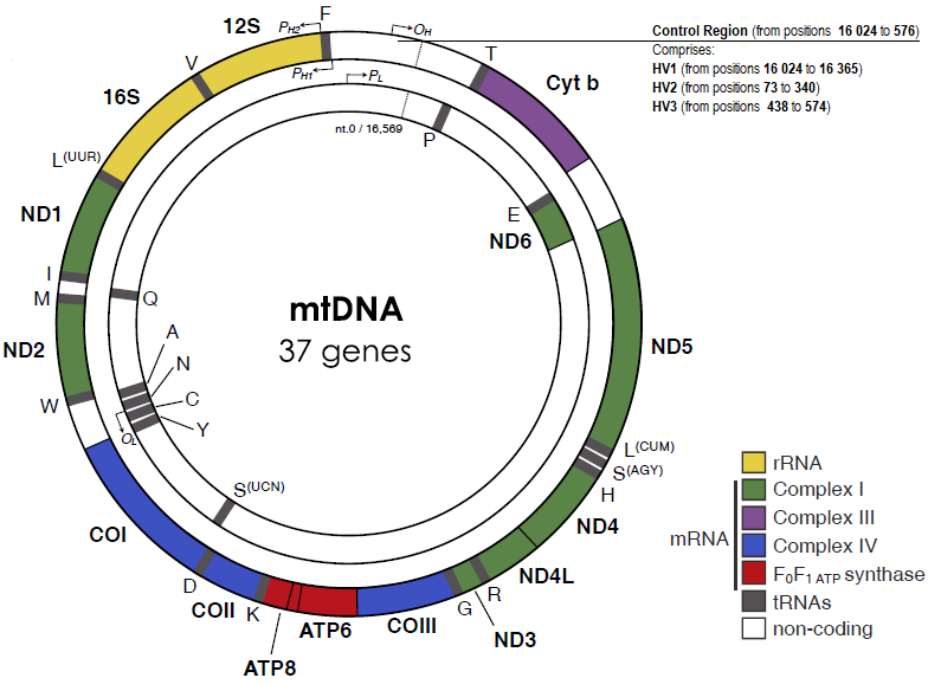


Figure 2.3. *Map of the human mitochondrial genome. Loci are colored according to functional groupings. The non-coding D-loop is shown at the top of the map (in white) (Amorim et al., 2019).*

As a uniparental genetic system, mtDNA is inherited solely through the maternal line (Giles *et al*., 1980). Uniparental systems, such as mitochondrial DNA and the Y chromosome, do not experience meiotic recombination, and their sequence variation arises only from random mutations over time. Moreover, mtDNA has a mutation rate that is 10-20 times higher than the nuclear genome (Brown *et al*., 1979). Thanks to this peculiarity, it is possible to obtain a large amount of information on genetic variability by only sequencing a molecule that is much shorter than the nuclear chromosomes. Furthermore, due to all the characteristics mentioned before, mitochondrial DNA plays a crucial role in phylogenetic and population genetic research, serving as an essential tool for tracing maternal lineages, understanding genetic diversity, and reconstructing evolutionary history.

2.3.1 Mitochondrial phylogeny

Mitochondrial phylogeny provides a detailed reconstruction of human evolutionary history, offering insights into past migrations and population structures. Since mitochondrial DNA (mtDNA) is maternally inherited and evolves more rapidly than the nuclear genome, it serves as a powerful molecular marker for tracing lineage divergence and human dispersal. Genetic variation in mtDNA accumulates over time, creating distinct genetic signatures that can be categorized into haplotypes and haplogroups. A haplotype refers to a specific set of mtDNA variations inherited together from a

common maternal ancestor, while a haplogroup is a larger monophyletic unit encompassing multiple haplotypes that share a common ancestor at a deeper evolutionary timescale (Torroni *et al*., 1993a). Haplogroups are defined by characteristic mutations that have accumulated sequentially along maternal lineages, allowing researchers to construct phylogenetic trees, including the human mitochondrial phylogenetic tree (Figure 2.4) and track the evolutionary relationships between different human populations (van Oven & Kayser, 2009).

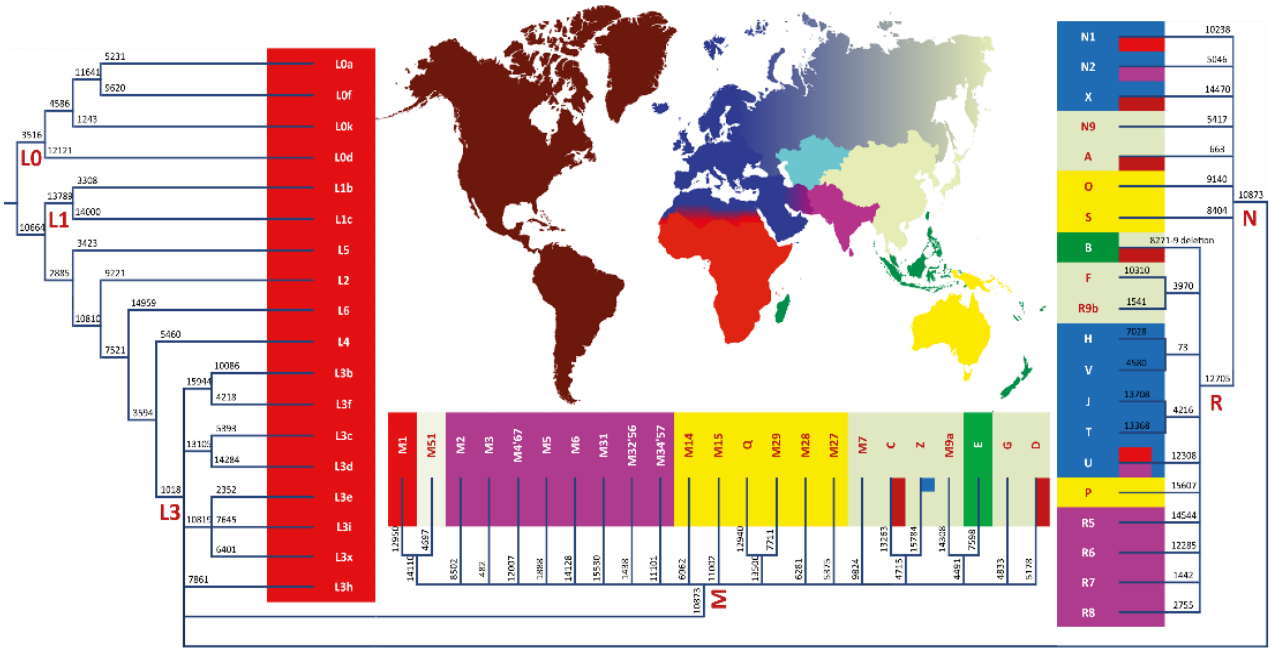


Figure 2.4. *Global distribution of mtDNA haplogroups, with key geographic patterns highlighted in color. (Kivisild, 2015)*

Phylogenetic studies of mtDNA reveal that all modern human mitochondrial lineages trace back to a common ancestor in Africa, often referred to as Mitochondrial Eve, who is estimated to have lived approximately 150,000–200,000 years ago (Ingman *et al*., 2000). The universally accepted mtDNA nomenclature was established with the identification of the first four branches of the phylogenetic tree, named alphabetically (A, B, C, and D), corresponding to Native American populations (Torroni *et al*., 1993b). Subsequent analyses identified haplogroups E, F, and G in Asia and H, I, J, K, T, U, W, X, and V in Europe (Torroni *et al*., 1994a; b, 1996). The root of the tree is restricted to African populations, and the first seven bifurcations represent the strictly sub-Saharan African branches (L0-L6) (Kivisild *et al*., 2004). The migration of modern humans out of Africa between 90–60 kya is directly linked to haplogroup L3, making it the only African lineage with descendants outside the

continent. As humans spread into Eurasia, L3 gave rise to haplogroups M, and N, which later diversified into all non-African maternal lineages (Cerezo *et al*., 2012; Soares *et al*., 2012). Haplogroup M primarily expanded into South and East Asia. Haplogroup N, on the other hand, took a more westward route, contributing to the early European, Central Asian, and Australian populations, as well as leading to haplogroup R, a major lineage that further diversified into H, V, U, JT, and other European haplogroups (Richards *et al*., 2000). In Eurasia, haplogroup U became particularly widespread, with subclades such as U5 being common among European hunter-gatherers (Malmström *et al*., 2009), while haplogroup H, which likely originated in Southwest Asia, became dominant in Neolithic Europe. Additionally, some haplogroups indicate instances of back-migration to Africa. Haplogroups U6 and M1 are found in North Africa and the Mediterranean and they are believed to be the result of an early return migration from Eurasia (Olivieri *et al*., 2006). This global dispersal of mitochondrial haplogroups highlights how a single major migration event, originating from haplogroup L3, ultimately shaped the genetic diversity of modern human populations outside Africa.

2.4 THE DONETSK OBLAST

2.4.1 Geographic and historical overview

Donetsk oblast is located in the southeastern part of Ukraine. It spans 26,517 km², accounting for approximately 4.4% of Ukraine's total land area. It is characterized by vast plains, hills, forests, spoil tips and numerous water bodies, including the Kalmius River, which connects Donetsk to the Sea of Azov and, ultimately, the Black Sea (www.encyclopediaofukraine.com). As the most populous region in the country, it is home to around 4.1 million residents (*State Statistics Committee of Ukraine*).

The origins of Donetsk date back to 1779 during the reign of Empress Catherine the Great, when the settlement of Aleksandrovka was first recorded. In 1869, Welsh industrialist John Hughes established a steel factory and several coal mines, which spurred the area's rapid industrialization. The settlement, initially known as Yuzovka, was later renamed Donetsk in 1961 to reflect industrial expansion of the Soviet era (https://www.britannica.com/place/Donetsk-Ukraine).

Donetsk oblast is part of the Eurasian steppe, which has historically been a transitional zone between nomadic pastoralists and sedentary agricultural communities. Due to its strategic location and geological landscape, Ukraine has always been a crucial transit route for human migrations. The lack of natural barriers has facilitated the transit of the territory, making the region a major gateway from Asia into Europe. Following the Last Glacial Maximum, Ukraine served as a key refugium for

populations repopulating Europe, particularly in the East European Plain (Pshenichnov *et al*., 2013). Historically, Ukraine has been home to over 100 ethnic and national groups (Kubicek, 2008). According to the 2001 census, ethnic Ukrainians represent 78% of the population, while ethnic Russians make up 17%, alongside smaller communities of Greeks, Tatars, and Jews (http://2001.ukrcensus.gov.ua/results/general/nationality/). Recent migration processes in this region were particularly dynamic during the 19th and 20th centuries, driven by urbanization and social modernization under the occupations of the Russian Empire, the Soviet Union, and finally, independent Ukraine (Yakubova, 2014).

Over the past decade, the region has undergone significant socio-political turmoil. In 2014, armed conflict erupted in the Donbass following the intervention of pro-Russian separatist factions supported by the Kremlin. This led to the declaration of the so-called "Donetsk People's Republic" and "Luhansk People's Republic," causing economic instability and infrastructure destruction.

2.4.2 Genetic background of Donetsk oblast

Of the various regions in Ukraine, Donetsk remains one of the least explored from an archaeogenomic perspective. It is widely recognized that Ukraine has played a key role in demographic movements since the Paleolithic. This is largely due to its strategic geographic location, serving as a crucial corridor for early migratory routes from Asia into Europe. Additionally, it functioned as one of the glacial refugia in the East European Plain, facilitating the re-expansion of human populations across Europe after the Last Glacial Maximum. Previous studies (Pshenichnov *et al*., 2013; Malyarchuk *et al*., 2001; Malyarchuk *et al*., 2023) have identified haplogroup H and various sub-lineages of haplogroup U, such as U4 and U5a (U5a1 and U5a1a), as predominant in Ukraine, along with lineages from haplogroups J and T. Subhaplogroups of HV and haplogroups with East Asian origins, including A, B, C, D, and G, were also found, though less frequent. Rare haplogroups in the region included I, N1b, W, X, and V.

These studies highlighted the complex genetic landscape of the Ukrainians, revealing high mtDNA diversity and strong genetic affinities with neighboring populations such as Slovaks, Russians, Poles, Serbs, and Estonians. It is clear that Ukrainians occupy a central position among both Western and Eastern Slavs, reflecting close genetic connections within the Slavic genetic pool. Moreover, Ukrainians showed mtDNA sequences that are similar or identical to those found in both eastern and western European populations, suggesting a considerable heterogeneity among Eastern Slavic populations. Additional studies highlighted that some western Ukrainian populations residing in

mountainous areas share closer genetic ties with Western Slavs than with other Ukrainian groups (Nikitin *et al*., 2009; Kushniarevich *et al*., 2015). This suggests a complex genetic landscape influenced by both geographic and cultural factors.

2.4.3 Geopolitical instability and its effects on Donetsk oblast

Since 2014, the geopolitical situation in Donetsk oblast has changed significantly, which has affected both the local population and the ability to conduct research in the region. Following the annexation of Crimea by the Russian Federation in 2014, armed conflict broke out in parts of eastern Ukraine, including Donetsk, where Russian-backed separatists took control of some areas (United Nations, 2022). The situation became more serious in 2022 when Russia launched a full-scale invasion of Ukraine. In September of that year, Russia held referendums in several occupied regions, including Donetsk, claiming that the local people wanted to join the Russian Federation. These votes took place under military occupation, and there were many reports of pressure on people to vote, including soldiers collecting ballots from homes (Amnesty International, 2022). The results claimed that more than 99% of people in Donetsk supported the annexation, but most of the international community rejected these numbers and called the referendums illegal. According to international law, including the UN Charter and the Geneva Conventions, such actions violate Ukraine's sovereignty (UN General Assembly, 2022). In response, the United Nations passed a resolution declaring the referendums invalid and calling for Russian troops to leave Ukraine (United Nations, 2022). NATO and the European Union also condemned the annexations, stating that the process was not transparent or fair, and that Donetsk remains part of Ukraine (Stoltenberg, 2022). Figure 2.5 illustrates the territorial changes and disputed regions in eastern Ukraine as of September 2022. These political developments have made it harder to conduct genetic studies in the region. For population genetics researchers in particular, the conflict not only affected access to the area, but also caused sudden changes in the original composition of the population due to displacement, which can complicate the interpretation of data. This context is important to keep in mind in order to better understand the genetic makeup of a region like Donetsk, where both scientific and political realities play a role.

Figure 2.5. *Map showing the annexation of southern and eastern Ukraine by Russia as of September 2022, including areas occupied and annexed in 2014 (Crimea) and during the 2022 invasion (Donetsk, Luhansk, Kherson, Zaporizhzhia). Donetsk oblast, a focal point of ongoing conflict and geopolitical dispute, is highlighted as part of the territories claimed by Russia but internationally recognized as Ukrainian.*

3. AIM OF THE WORK

Donetsk oblast, located in southeastern Ukraine along the border with the Russian Federation, holds a unique position both geographically and historically. As part of the vast Eurasian steppe, this region has served as a crucial transit corridor for human populations migrating between Asia and Europe since prehistoric times. Throughout various periods, from the Paleolithic era through the complex socio-political transformations of the modern age, Donetsk has been a melting pot of diverse ethnic groups and cultural influences. This complex background is further underscored by recent geopolitical developments, including the 2022 annexation of the region by Russia, an event that not only reflects contemporary tensions but also revives long-standing historical ties between populations on either side of the border. Despite its rich historical and anthropological significance, Donetsk remains relatively unexplored from a genetic perspective. Understanding the genetic pool of this region is particularly important and the mitogenome analysis offers an opportunity to trace the maternal lineage diversity shaped by millennia of migrations, admixtures, and historical events. This thesis aims to fill this gap by exploring the mitochondrial genetic landscape of present-day individuals from Donetsk Oblast, with the goal of uncovering patterns of lineage diversity and population structure. To complement the genetic data, genealogical information was also considered, providing context on ancestral backgrounds and highlighting lineages of particular interest. Through this integrated approach, the study seeks to deepen our understanding of how historical migrations, admixture, and the region's complex past have shaped the genetic profile of Donetsk within the broader framework of Eurasian diversity.

4. MATERIALS AND METHODS

4.1 SAMPLE COLLECTION

The modern mitogenomes of this study derive from an original collection of 104 samples, provided by our collaborators, Prof. Svetlana Arbuzova from the International Medico-Genetic Centre, Hospital Nol, Donetsk, and Prof. Ornella Semino from the University of Pavia. The samples were collected in the Donetsk oblast (Ukraine) between 1998–1999 from male individuals aged 18 to 75 years. Each participant was asked to complete and sign an informed consent form for the processing of personal data and to fill out a pedigree chart indicating the birthplace of their maternal ancestors for at least two generations (grandmother and great-grandmother). To ensure anonymity, each individual was assigned a unique identification code. The historical contextualization of the study was undertaken by Dr. Yevhen Zakharhcenko, Senior Lecturer in the Department of History of Ukraine at V. N. Karazin Kharkiv National University.

4.2 DNA QUANTIFICATION AND QUALITY CHECK

Out of the original 104 samples, 10 were excluded from the analyses due to missing consent forms and/or sample tubes. The remaining 94 samples had been previously extracted as part of a previous study on the Y chromosome (Semino *et al*., 2000), using the classic phenol-chloroform method and Corex centrifuge tubes. DNA extractions were stored at -20°C in plastic tubes for preservation.

DNA quantification was performed using the Invitrogen Qubit™ 4 fluorometer (Figure 4.1A). This instrument measures DNA concentration based on fluorescence, through the use of specific dyes that selectively bind to nucleic acids (DNA or RNA). Prior to measurement, calibration with two standard DNA solutions is required. Each assay tube is prepared by adding 199 µL of Qubit™ working solution to 1 µL of DNA, followed by brief vortexing and a 5-minute incubation, protected from light. The fluorescence emitted is directly proportional to the amount of DNA present in the sample (https://www.thermofisher.com/it/en/home/industrial/spectroscopy-elemental-isotope-analysis/molecular-spectroscopy/fluorometers/qubit/qubit-fluorometer.html).

To assess the quality and integrity of the extracted DNA samples, capillary electrophoresis was performed, using the Agilent 4150 TapeStation system (Figure 4.1B). This automated platform provides information on DNA quality and fragment size. The analysis was carried out with the

Genomic DNA (gDNA) Kit, which includes a ScreenTape (the capillary support), a buffer, and a DNA ladder. The system is capable of analyzing fragments ranging from 200 to over 60,000 bp. This step was crucial, as DNA concentration alone does not provide information on molecular integrity. For this quality check, three extracted samples were selected based on their DNA concentrations: one with the highest concentration, one with an intermediate value, and one with the lowest. This selection allowed for the evaluation of different possible scenarios regarding DNA degradation. For optimal integrity, DNA fragments should exceed 10 kilobase pairs.

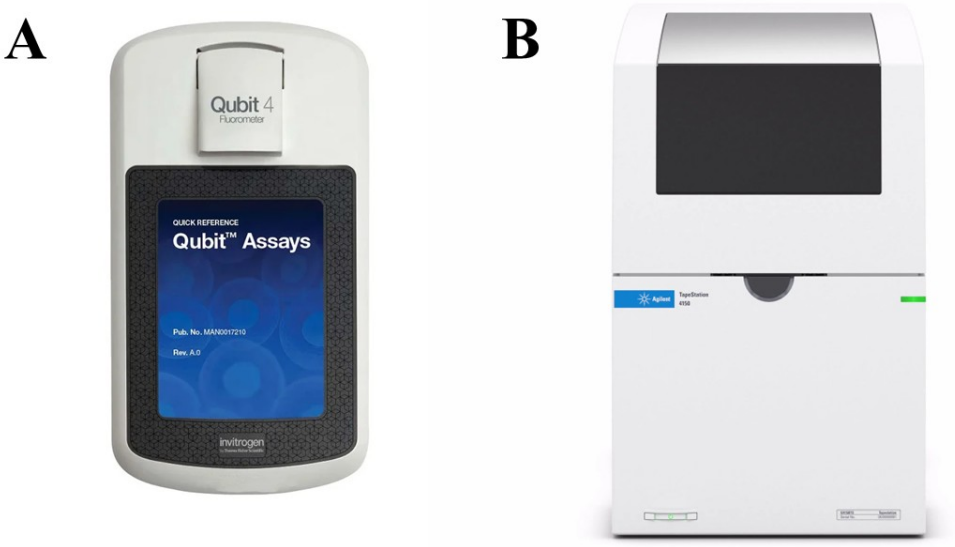


Figure 4.1. *(A) Qubit 4 Fluorometer (https://www.thermofisher.com/it/en/home/industrial/spectroscopy-elemental-isotope-analysis/molecular-spectroscopy/fluorometers/qubit/models.html). (B) Agilent 4150 Tapestation (https://www.agilent.com/en/product/automated-electrophoresis/tapestation-systems/tapestation-instruments/4150-tapestation-system-297322).*

4.3 COMPLETE MITOGENOME AMPLIFICATION

The entire mitochondrial genomes were amplified through Long Range (LR) PCR using the Promega GoTaq® Long PCR Master Mix kit. This mix contains all necessary components for the reaction, including cofactors, buffers and a thermostable Taq polymerase, ensuring efficient amplification across extended cycles and accurate amplification of large DNA fragments, up to 30,000 bp.

The protocol involved the setup of two parallel PCR reactions, each using a distinct pair of primers targeting specific regions of the mitochondrial genome. In particular, they were designed to produce

two partially overlapping fragments of approximately 8-9 kb each. Each PCR reaction was prepared with a final volume of 25 µL, including the DNA template. The volume of DNA added, ranging from 1 to 5 µL, was adjusted according to the concentration determined during the quantification step.

Table 1. Summary of volumes and initial and final concentrations of each reagent used for the mtDNA amplification in each PCR reaction tube.

Reagent	Volume (µL)	Initial Concentration	Final Concentration
GoTaq® Master Mix	12,5	2X	1X
Primer For	0,5	10 µM	0,2 µM
Primer Rev	0,5	10 µM	0,2 µM
H_2O + DNA	11,5		
Total	25		

Table 2. Primers used to amplify the human complete mitogenome in two overlapping amplicons.

Amplicon	PCR product length (bp)	Primer name	Primer length (nt)	Sequence (5'-> 3')	Melting Temperature (T_M) (°C)
#1	8.637	5.193 For	20	CCCTTATTCCATCCACCCT	60,0
		13.829 Rev	22	AGTCCTAGGAAAGTGACAGCGA	60,44
#2	8.561	13.477 For	23	AGGAATACCTTTCCTCACAGGTT	59,4
		6.151 Rev	20	GTGGTAAGGGCGATGAGTGT	60,0

The used PCR program for the amplification is described in table 3. It consists of two consecutive reaction cycles, to allow the correct amplification of both long-range fragments.

Table 3. Long range PCR reaction protocol for two overlapping fragments amplification.

Step	Temperature (°C)	Time
Initial denaturation	94	2 min
Denaturation	94	30 sec
Annealing	58	30 sec

Elongation	65	10 min
Repeat from denaturation to extension for 20 cycles		
Denaturation	94	30 sec
Annealing	55	30 sec
Elongation	65	10 min
Repeat from denaturation to extension for 10 cycles		
Final extension	72	10 min

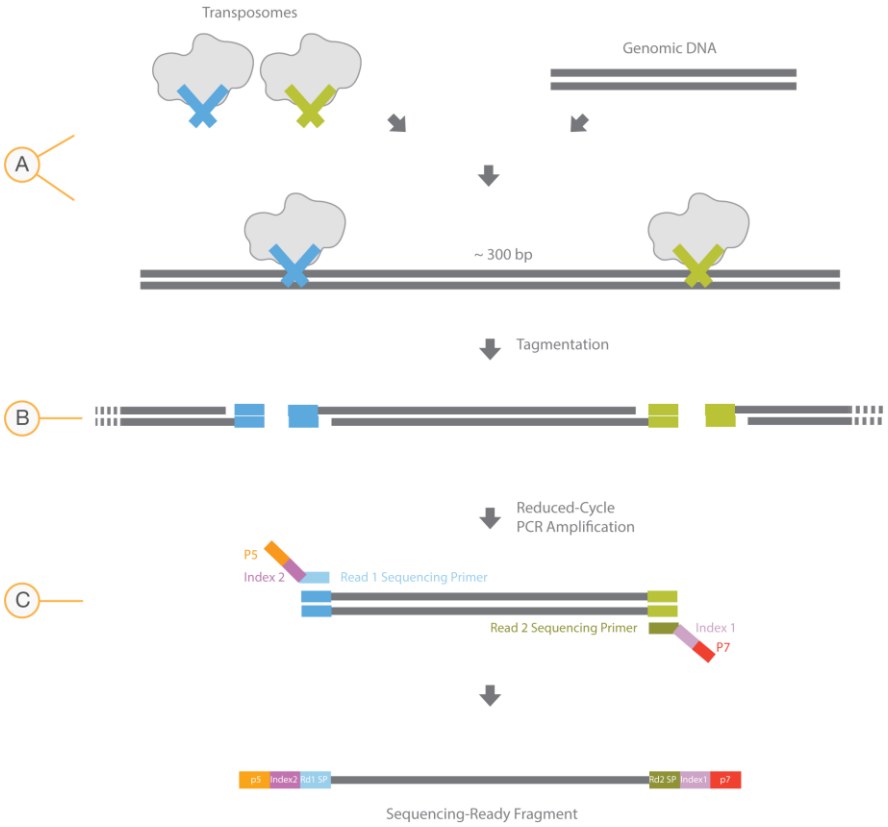
In order to confirm the successful amplification, an agarose gel electrophoresis was carried out. A 1% agarose concentration was selected to facilitate the analysis of large fragments (8–9 kb). Each amplification product (5 µL) was mixed with loading dye and applied onto the gel. Midori Green Advance (Genetics) served as the intercalating agent, and electrophoresis was conducted in a 1X TBE buffer solution. To assess the amplicon size, a Sharpmass™ 1 kb plus ladder (Euro Clone) was used, as it contains bands ranging from 10,000 bp down to 250 bp, making it suitable for evaluating long-range fragments.

4.4 LIBRARY PREPARATION AND NGS ILLUMINA SEQUENCING

Following this verification step, the long-range PCR products were used for NGS library preparation, using the Illumina Nextera® XT DNA library preparation kit (https://emea.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet_nextera_xt_dna_sample_prep.pdf). This system supports the multiplexing of up to 384 samples, enabling the simultaneous sequencing of numerous libraries within a single run on a high-throughput platform. Multiplexing is particularly advantageous when analyzing targeted genomic regions or small genomes, such as mitochondrial DNA. During library preparation, each DNA fragment is tagged with a unique sample-specific barcode index. These indexes allow for accurate identification and separation of reads during downstream analysis. By employing this strategy, the number of samples processed per sequencing run can be significantly increased, thereby enhancing efficiency and reducing overall costs.

4.4.1 Tagmentation

The first step of the library preparation is tagmentation, a process in which DNA is simultaneously fragmented and tagged with adapter sequences in a single enzymatic reaction. This is achieved using a transposase enzyme, which is pre-loaded with adapter oligonucleotides. When the transposase interacts with the DNA, it cuts the molecule and inserts adapter sequences at the ends of the resulting fragments (Figure 4.2). This produces sequencing-ready templates that are further processed in the subsequent PCR amplification step.



A. Nextera XT transposome with adapters combined with template DNA
B. Tagmentation to fragment and add adapters
C. Limited-cycle PCR to add index adapter sequences

***Figure 4.2**. Nextera XT DNA library preparation workflow. (Nextera® XT DNA Library Prep Kit Reference Guide).*

4.4.2 Amplification of tagmented DNA

The tagmentation step is followed by an amplification reaction with a limited cycle PCR (Table 4). This PCR reaction allows the addition of indexes (i5 and i7) and adapters (P5 and P7), taking advantage of complementarity with oligonucleotides added by transposase enzyme (Figure 4.2).

Table 4. Long range PCR reaction protocol for two overlapping fragments amplification.

Step	Temperature (°C)	Time
	68	3 min.
Initial denaturation	98	3 min.
Denaturation	98	45 sec.
Annealing	62	30 sec.
Extension	68	2 min.
Repeat from denaturation to extension for 12 cycles		
Final extension	68	1 min.

4.4.3 Library purification

Amplified libraries were purified using AMPure XP beads, which are paramagnetic (exhibiting magnetism only in a magnetic field), preventing them from clumping or settling out of solution. Each bead consists of a polystyrene core coated with magnetite and a layer of carboxyl groups, allowing negatively charged DNA fragments to reversibly bind to their surface. This purification step removes short DNA fragments, primer dimers, free nucleotides, salts, and residual enzymes from the amplification reaction. AMPure XP beads utilize a specialized buffer system that selectively retains DNA fragments larger than approximately 200 bp, ensuring the removal of unwanted smaller molecules. Following DNA binding, the beads were subjected to two washes with 80% ethanol to eliminate contaminants. After the washing steps, the beads were briefly air-dried to remove residual ethanol before adding resuspension buffer (RSB). This buffer facilitates the elution of purified DNA from the beads, allowing the recovery of high-quality libraries. The final purified libraries were then separated from the beads using a magnetic stand, and the clear supernatant containing the DNA was carefully transferred to a fresh tube. To assess purification efficiency, an optional quality control step was performed by running purified libraries on a 2% agarose gel. The expected result was a smear of

DNA fragments ranging from approximately 200 to 1,000 bp, confirming the successful removal of primer dimers and unwanted short fragments.

4.4.4 Library normalization

Once the libraries had been purified, normalization was performed in order to ensure a homogeneous and balanced representation of each library within the final pool. This process involves quantifying each final library using the Promega Quantus™ Fluorometer and a subsequent dilution of each product to obtain a common concentration. The normalized libraries were then combined into a single tube to create the final sequencing pool. To verify the quality and consistency of the pooled libraries, the final pool was quantified, and the average fragment size was assessed using capillary electrophoresis on the Agilent 4150 TapeStation system with the HS D1000 ScreenTape assay (Figure 4.3).

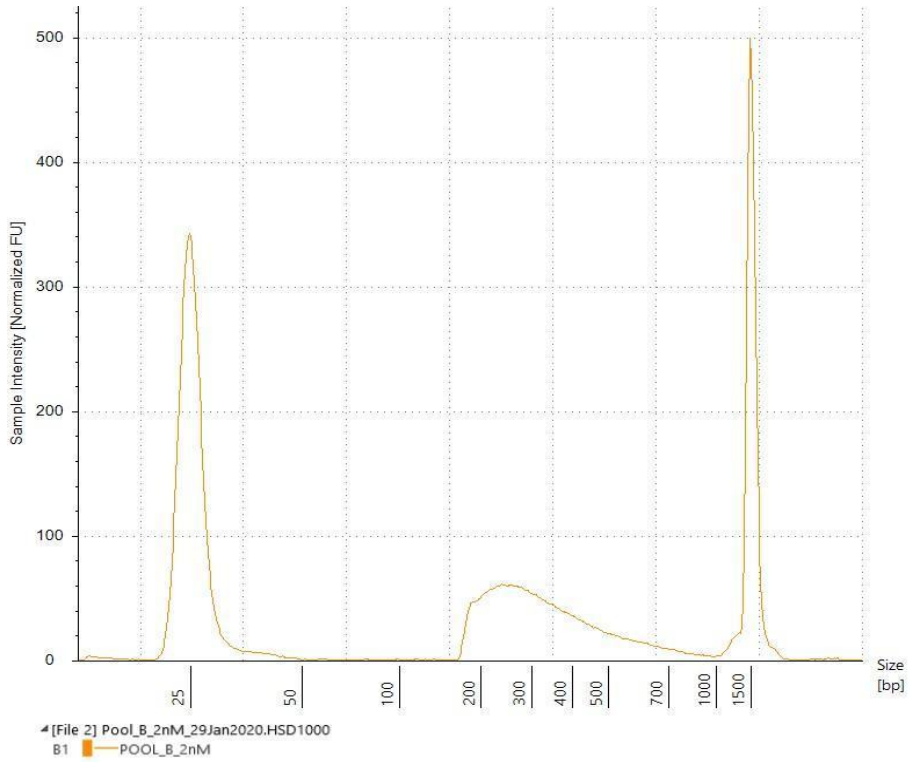


Figure 4.3. *Example of a pooled library analyzed by capillary electrophoresis using the Agilent 4150 TapeStation system.*

The following formula had then been used for the calculation of the nano-molar (nM) concentration of the pool:

nM = [(ng/µl) / (avg bp size*660 g/mol)] * (1.000.000 µl/1 L).

In this equation, the "avg bp size" value represents the mean fragment length of the pooled library in base pairs. The value of 660 g/mol is the molecular weight of a base pair in the double strand molecule of DNA. After calculating the concentration, the pooled library was diluted to a standard working concentration of 4 nM to ensure optimal denaturation conditions. The denaturation step was performed using 0.2 N NaOH. The denatured libraries were further diluted by adding the HT1 Hybridization Buffer, first to a concentration of 20 pM and then to a final loading concentration of 8 pM, reaching a final volume of 600 µL for the sequencing platform. These precise concentration adjustments are crucial to ensure uniform cluster density and high-quality sequencing output.

4.4.5 Illumina sequencing

The sequencing was performed on the Illumina NextSeq 550 platform at the Genomic and Post-Genomic Unit, IRCCS Mondino Foundation in Pavia, with collaborative support from Dr. Stella Gagliardi and Dr. Rosalinda Di Gerlando. This system uses a technology of sequencing by synthesis (SBS) (Figure 4.4). The P5 and P7 adapters, added during the amplification step, are complementary to two oligonucleotides (cP5 and cP7) that are immobilized in an alternating pattern at the base of the flow cell. This glass-based matrix serves as a platform for bridge PCR, a process of clonal amplification. During this reaction, DNA polonies (clusters containing thousands of identical DNA fragments) are generated. Proper dilution of the DNA library is crucial to ensure optimal spacing of fragments within the flow cell, preventing excessive cluster overlap. Once cluster formation is complete, all double-stranded DNA fragments are denatured, and the unbound reverse strands are washed away.

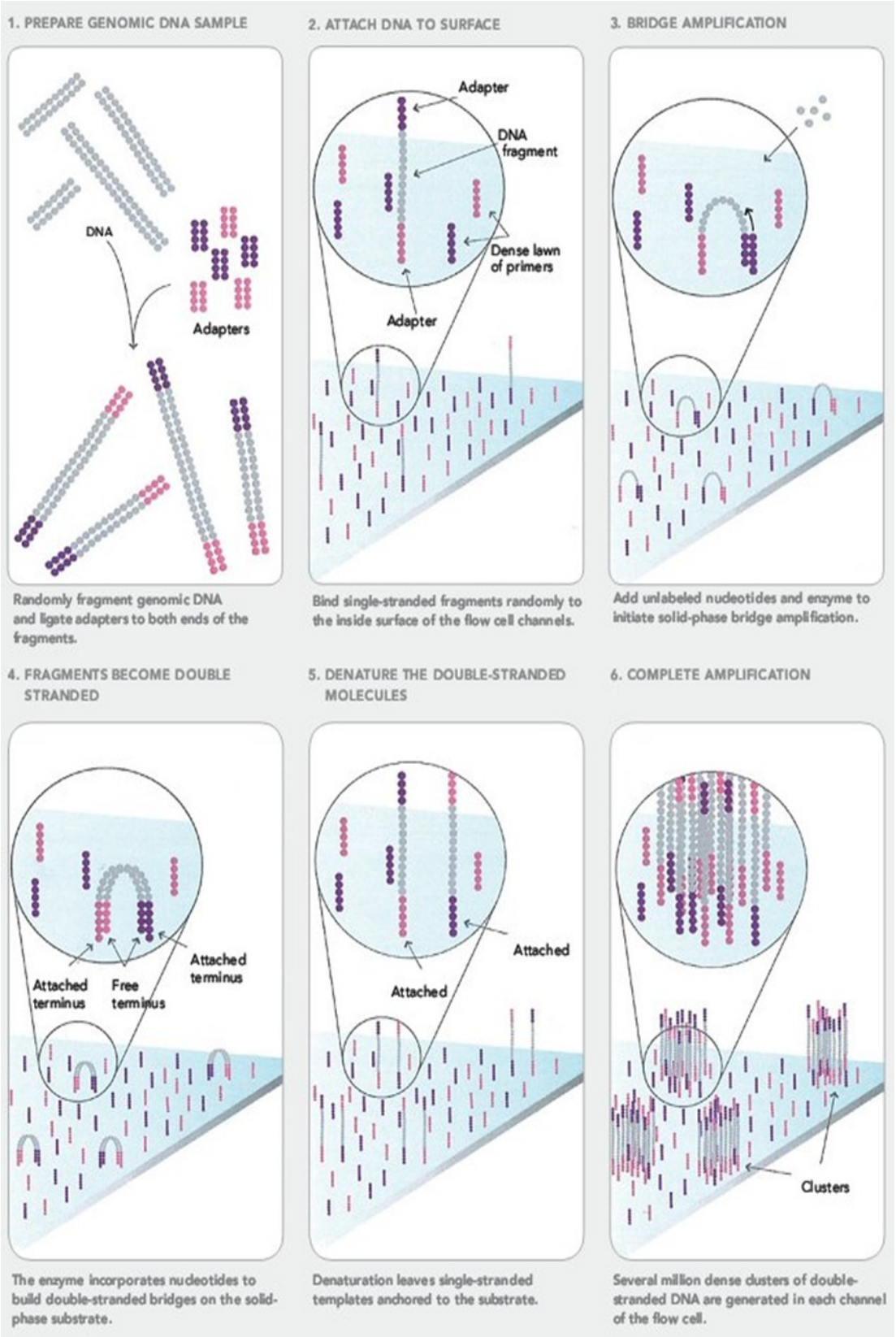


Figure 4.4. *Schematic workflow of the Illumina sequencing technology (Technology Spotlight: Illumina® Sequencing, "Illumina Sequencing Technology").*

The sequencing process, Sequencing by Synthesis (SBS), then begins. During this step, fluorescently labeled nucleotides, each uniquely tagged with a different fluorophore, are introduced simultaneously. However, only the nucleotide complementary to the template strand is incorporated into the growing sequence (https://emea.illumina.com/science/technology/next-generation-sequencing/sequencing-technology/2-channel-sbs.html). Each incorporated base carries a reversible terminator that temporarily halts DNA extension, allowing the emission of a fluorescent signal, which is detected by a photosensitive sensor. After imaging, the terminator is chemically removed, enabling the polymerase to incorporate the next nucleotide, and the cycle repeats (Figure 4.5)

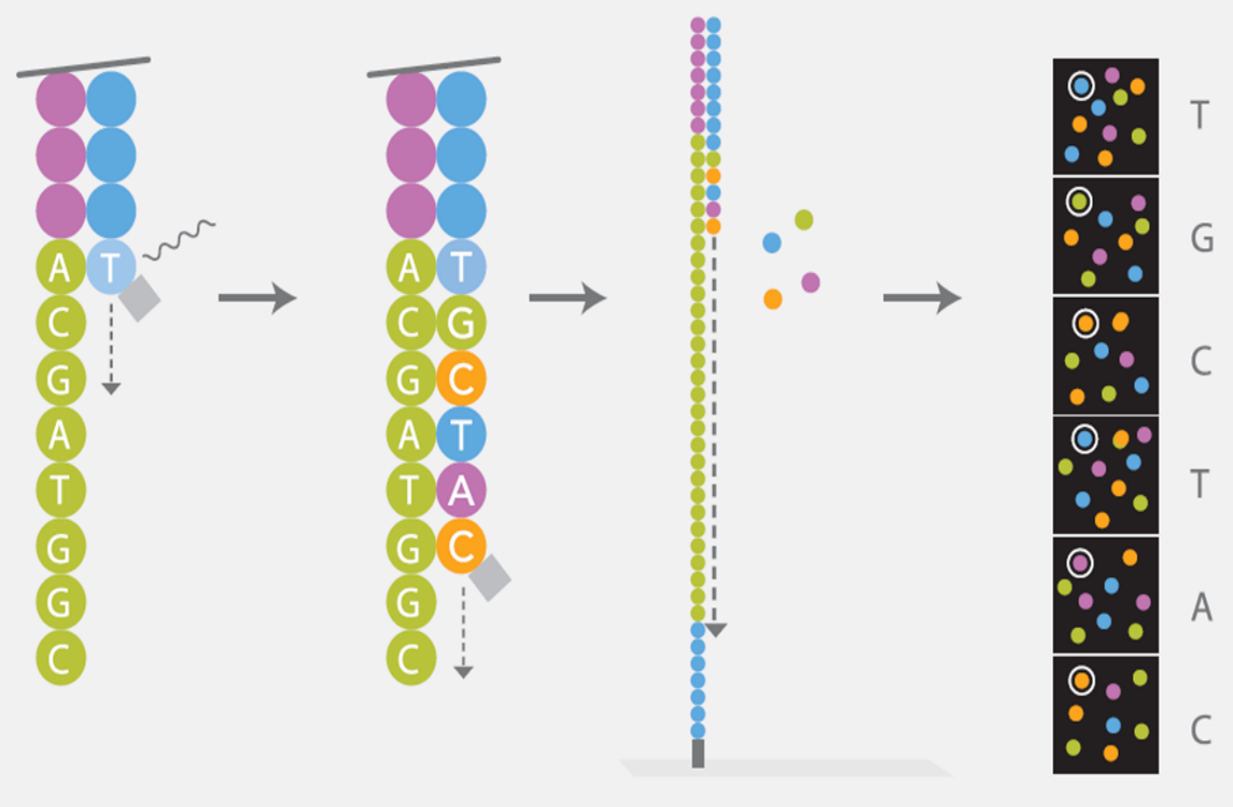


***Figure 4.5**. Illumina sequencing by synthesis (SBS) technology.*

4.5 SEQUENCE ANALYSIS

The analysis of raw data was first conducted using a pipeline developed by Dr. Nicola Rambaldi Migliore (https://github.com/NicRamb/mtDNA_pipe). These scripts generated BAM (Binary Alignment Map) files, which were subsequently processed on the online platform mtdna-server.uibk.ac.at (Weissensteiner *et al*., 2024) to determine final haplotypes aligned to the revised

Cambridge Reference Sequence (rCRS) (Andrews *et al*., 1999). Ambiguous regions in the sequences were manually checked to ensure accuracy.

Three samples were excluded from the analysis due to poor sequence quality, reducing the number of the final dataset to 91 samples. Haplogroup classification was performed using HaploGrep v3.2.1 (Schönherr *et al*., 2023), a tool based on the latest global mitochondrial DNA phylogeny (Dür *et al*., 2021). The resulting fasta files (generated with HaploGrep v3.2.1) were used for subsequent analyses.

4.6 STATISTICAL ANALYSES

To evaluate genetic diversity and population structure, statistical analyses were conducted on the dataset of 91 complete mitochondrial genomes, as well as on two subgroups derived by separating individuals based on their Terminal Maternal Ancestor (TMA) into Ukrainian and Russian TMA groups. Prior to analysis, sequences were aligned using MAFFT v7.520, and the resulting fasta file was used as input for subsequent analyses.

To assess genetic diversity, nucleotide diversity (π) was calculated for the entire dataset, as well as for the two subgroups (Ukrainian and Russian TMA groups). Nucleotide diversity measures the average pairwise nucleotide differences per site within a population (Nei & Li, 1979) and provides insights into the level of genetic variation present in the dataset. Haplotype diversity (Hd) was also computed to evaluate the uniqueness of haplotypes within each group (Nei, 1987). To quantify genetic differentiation between the Ukrainian and Russian TMA groups, the Fixation Index (F_{ST}) was calculated using the hierfstat (Goudet, 2005) and adegenet (Jombart, 2008) packages implemented in R (version 4.3.0; R Core Team, 2024). F_{ST} measures the proportion of genetic variance that can be attributed to differences among populations and is commonly used to assess population structure (Wright, 1951; Weir & Cockerham, 1984). Additionally, Analysis of Molecular Variance (AMOVA) was conducted using the vegan package (Oksanen *et al*., 2025), specifically the adonis2 function, to assess genetic structuring within and between groups. AMOVA partitions molecular variance at different hierarchical levels and determines whether significant genetic differentiation exists between predefined groups (Excoffier *et al*., 1992). The obtained p-value and R^2 statistic were used to determine whether genetic variation was significantly structured among the groups and to quantify the proportion of the total variance explained by this subdivision. Tajima's D statistic (Tajima, 1989) was performed with the pegas package (Paradis, 2010) to assess the demographic processes within the Ukrainian and Russian Terminal Maternal Ancestor (TMA) groups. This test compares two measures of genetic diversity: the number of segregating sites and the average number of pairwise

nucleotide differences. Negative values of Tajima's D suggest an excess of low-frequency variants, typically associated with population expansion, while positive values indicate lack of rare alleles, suggesting a balancing selection. The associated p-value was derived from a neutral model to determine whether the observed Tajima's D was significantly different from the expectation under neutral evolution, providing insights into the potential demographic history of each group.

In addition to diversity and population structure analyses, chi-square tests were employed to assess whether mitochondrial haplogroup distributions differed significantly between groups. Tests were conducted at two levels of phylogenetic resolution: first using major haplogroups (e.g., H, U, J), and then using all subhaplogroups to capture finer-scale lineage differences. Two main comparisons were performed: one between individuals with Ukrainian and Russian Terminal Maternal Ancestors (TMAs), and another within the Ukrainian TMA group, subdivided geographically into East and West Ukraine. These analyses were aimed to identify potential regional or ancestral structure in haplogroup composition. Given the presence of low-frequency haplogroups in some categories, the chi-square tests were performed with simulated p-values to ensure statistical robustness under conditions of small expected counts.

4.7 PRINCIPAL COMPONENT ANALYSIS (PCA)

Principal Component Analysis (PCA) is a statistical method used to reduce the dimensionality of large datasets while retaining as much variability as possible. It simplifies data by transforming it into a set of new variables, known as principal components (PCs), which represent the directions of maximum variance. PCA analyzes a data table with inter-correlated observations and variables, extracting the most informative patterns of variation and expressing them through orthogonal principal components. It also visualizes the patterns of similarity between observations and variables by displaying them as points in maps (Abdi *et al*., 2010). Due to these features, PCA is widely used in population genetics to identify patterns of genetic variation and structure in large-scale genomic data. PCAs were conducted using the FactoMineR package (Lê *et al*., 2008) in R version 4.3.0. The dataset, consisting of modern individuals from Donetsk oblast along with Western Eurasian populations from GenBank, was imported as a table of haplogroup relative frequencies per population, where populations were treated as observations and haplogroups as variables. These frequency data were used to compute principal components. A total of 11 principal components were retained to explain 80% of the total variance. The PCA results were visualized using ggplot2 (Wickham, 2016), factoextra, and ggbiplot packages.

4.8 PHYLOGENIES

In this thesis, rooted phylogenetic trees were constructed using complete mitochondrial DNA sequences, following the maximum parsimony (MP) approach. The MP principle aims to minimize the total number of evolutionary changes from a common ancestor, resulting in the shortest possible tree. This principle seeks to identify the evolutionary tree that requires the fewest substitutions to account for the observed differences among the sequences. The MP tree in this thesis was built using the mtPhyl software. The software uses the PhyloTree Build 17 worldwide phylogeny as a reference (van Oven & Kayser, 2009). After generating the initial tree, updates were manually made to align it with the latest improvements in mtDNA phylogeny, such as revisions provided by Dür (2021). The resulting tree was visualized using iTOL (Interactive Tree of Life, version 7.1), which enabled graphical representation and annotation of phylogenetic relationships.

5. RESULTS

5.1 THE MATERNAL ORIGIN OF DONETSK OBLAST POPULATION (based on TMAs)

Genealogical data were collected from 104 male individuals who provided a pedigree chart and a written informed consent for genomic research. Genealogical analysis focused on tracing the geographic origins of their terminal maternal ancestors (TMA). These TMAs were predominantly located within Ukraine, but also extended into neighboring Russia and Belarus, reflecting historical demographic dynamics in the region, as visualized on the map in Figure 5.1. Due to their considerable geographic distance, Russian regions such as Omsk, Amur, and Irkutsk oblasts were excluded from the visualization to maintain clarity and focus on the primary study area.

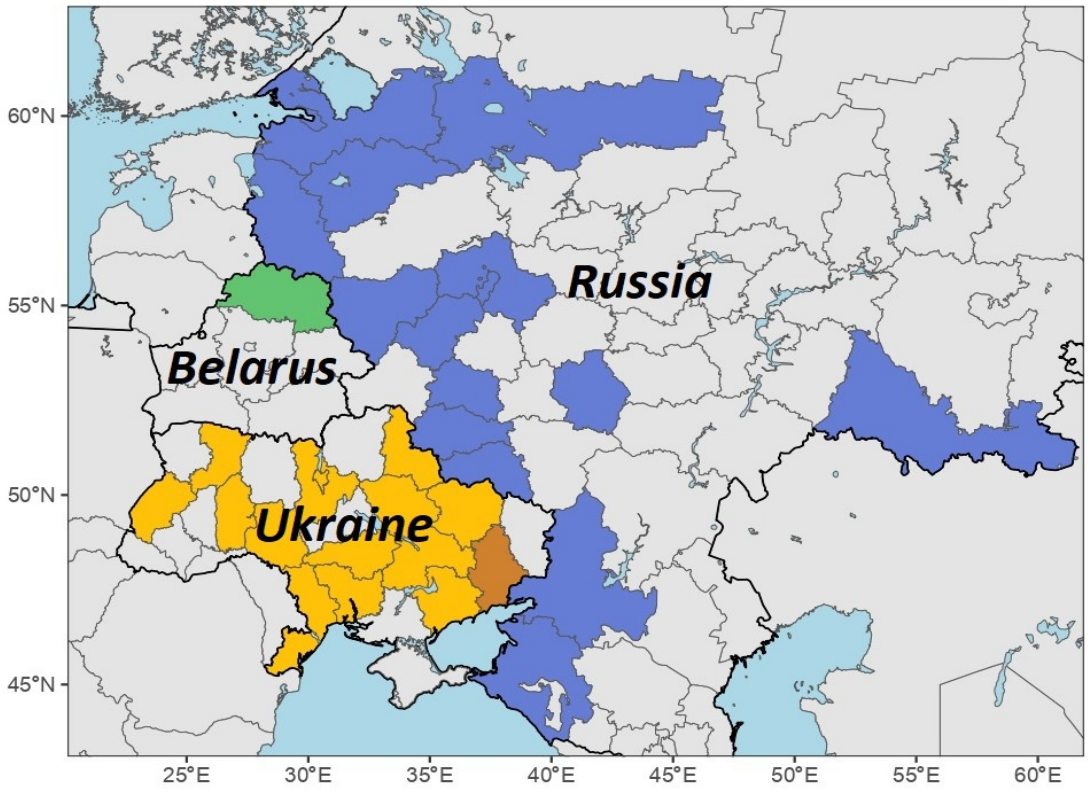


Figure 5.1. *Geographic distribution of the last known terminal maternal ancestors by region. (Donetsk oblast, Ukraine, representing the sampling region).*

Of the 104 individuals, 70 had Ukrainian TMAs. Among these, 33 individuals traced their grandmothers' place of birth to Donetsk oblast. Six individuals each identified Zaporizhzhia and Vinnytsia oblasts as TMA origins, while Kyiv and Kirovohrad oblasts accounted for four individuals each. Three individuals each were linked to Dnipropetrovsk and Kharkiv oblasts, and two individuals were associated with Sumy, Odesa, and Poltava oblasts. Lastly, one individual each traced their TMA to Mykolaiv, Khmelnytskyi, Rivne, Lviv, and Chernihiv oblasts.

Russian TMAs were identified for 33 individuals, while only one person had a Belarusian TMA. Among the Russian regions represented, six individuals each traced their TMA to the Kursk and Rostov oblasts, three to the Oryol and Smolensk oblasts, and two to the Tambov and Pskov oblasts. Additional regions such as Amur, Belgorod, Irkutsk, Kaluga, Krasnodar krai, Leningrad, Moscow, Nizhny Novgorod, Omsk, Orenburg, and Volgograd oblasts were each represented by one individual each.

Following this genealogical analysis, DNA extracts were examined. Of the original 104 samples, 94 DNA extracts were successfully preserved and available for study. Individuals with TMAs originating exclusively from Ukraine (69 individuals) and Russia (25 individuals) were included in further analyses.

5.2 DNA QUALITY CHECK

Capillary electrophoresis was performed to assess the quality of the initial 94 DNA samples. The analysis revealed significant differences in DNA integrity among the samples; three examples are reported in Figure 5.2. These samples were selected based on their concentrations, with UA57 having the lowest DNA amount (29.2 ng/µL), UA12 an intermediate concentration (167 ng/µL), and UA28 one of the highest (322 ng/µL).

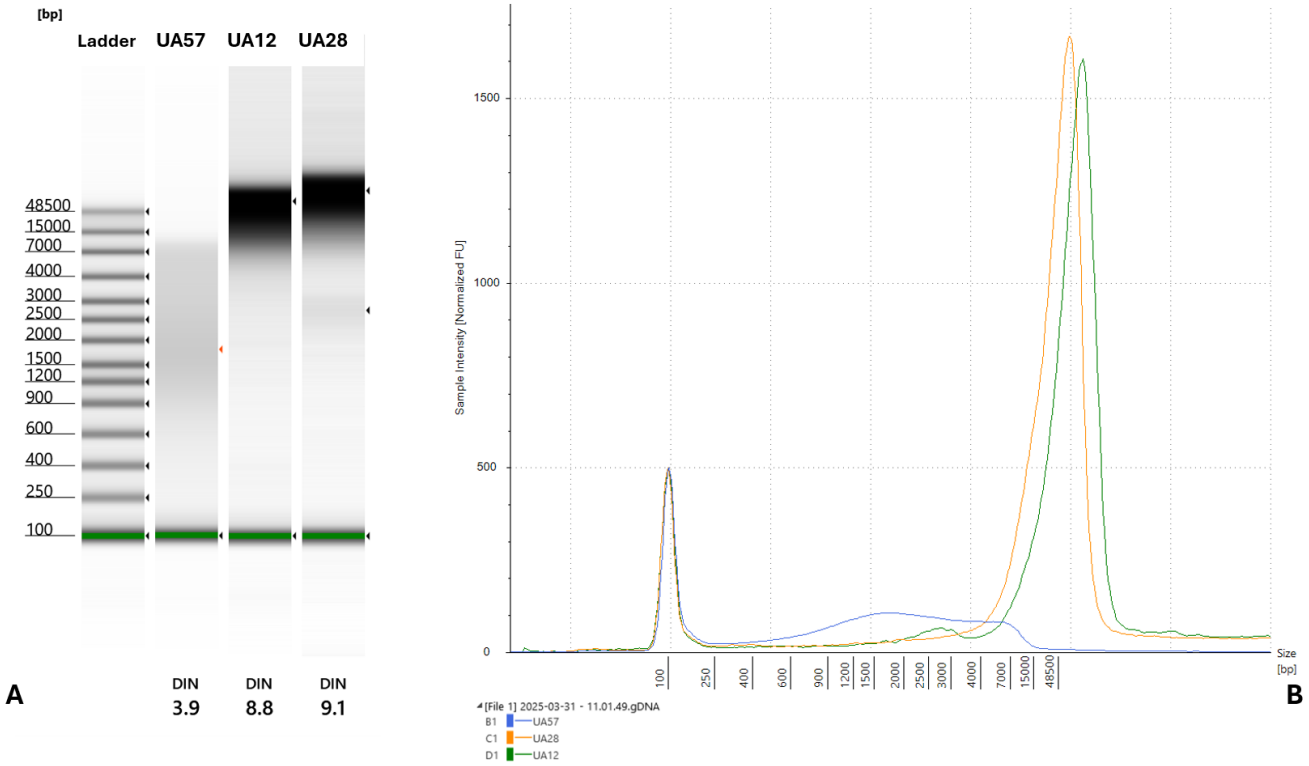


Figure 5.2. *Results of the analysis through the Agilent TapeStation capillary electrophoresis. (A) Digital representation of the DNA fragments distribution. (B) Fluorescence intensity plot showing sample integrity based on fragment size distribution.*

The left image (Figure 5.2A), a virtual gel representation, displays the fragment distribution for each sample alongside a molecular weight ladder. Sample UA57 exhibits a smeared and faint banding pattern, indicating substantial degradation, whereas UA28 and UA12 show well-defined high molecular weight bands, suggesting high-quality DNAs. These observations are quantitatively confirmed by the DNA Integrity Number (DIN) values, where UA57 has a DIN of 3.9, reflecting a high degree of fragmentation, while UA28 and UA12 have DIN values of 9.1 and 8.8, respectively, indicating largely intact DNA. The electropherogram on the right (Figure 5.2B) provides a fluorescence intensity plot, where the x-axis represents fragment size in base pairs and the y-axis presents signal intensity. It is evident that UA57 (in blue) exhibits no distinct peak, with a relatively flat signal at lower fragment sizes, indicating extensive fragmentation. In contrast, UA12 (in green) and UA28 (in orange) both display a prominent peak at fragment sizes exceeding 48,500 bp, suggesting the presence of high-molecular-weight, intact DNA. However, a key distinction between the two is that UA28 presents a single, well-defined peak at this high molecular weight, whereas

UA12 also exhibits a secondary peak in the 2,500–3,000 bp range, indicating an initial degradation process leading to a partial loss of DNA quality.

These observations highlight three distinct DNA integrity scenarios; however, all three samples were sequenced regardless of their quality. This decision was made to assess the impact of DNA fragmentation on sequencing outcomes and to explore whether degraded samples could still yield usable data. Including UA57, despite its extensive fragmentation, allows for a direct comparison between high-quality and degraded samples, providing insights into the potential limitations and biases introduced by DNA degradation.

5.3 THE FINAL MITOGENOME DATASET

After quality checks on the 94 mitogenome sequences, three samples had to be excluded due to quality issues. Specifically, UA11 and UA87 were removed because of very low-quality sequences, while UA72 was excluded due to contamination. UA11 had a Russian TMA, whereas UA72 and UA87 belonged to the group with Ukrainian TMA, respectively from Chernihiv oblast and Vinnytsia oblast. Thus, the final dataset consisted of 91 high-quality mitogenomes. All retained samples have an average depth of coverage above 10, genome coverage greater than 90%, and fewer than 4 heteroplasmic variants.

Table 5 presents a selection of seven representative results from the final mitogenome dataset, including the three samples tested for DNA quality. It is worth noting that UA57, which showed a high DNA degradation, achieved a genome coverage of 99.994% and a quality score of 0.977. These values reflect very high sequencing performance, even when compared to samples with much higher initial DNA integrity, such as UA12 and UA28. This outcome highlights that UA57 still produced reliable and high-quality sequencing data despite substantial degradation observed in the quality check. Table 5 also includes UA72, one of the excluded samples, which had a total of 34 heteroplasmic variants, a value significantly higher than the acceptable threshold and consistent with contamination. These examples emphasize the robustness of the filtering criteria applied and illustrate how the downstream sequencing metrics were essential in validating the quality and reliability of each sample.

Table 5. Example of seven mitogenome data obtained in the present study.

SampleID	TMA	Contamination	Het Number	AvgDepth	Coverage %	Haplogroup	Quality	Range	Haplotype
UA04	Russia	NO	0	128,123	100	J2a1a1	0,966	1-16569	73G 150T 152C 195C 215G 263G 295T 315.1TC 319C 489C 513A 750G 1438G 1850C 2706G 4216C 4769G 7028T 7476T 7789A 8860G 10398G 10499G 11251G 11377A 11719A 12011C 12612G 13708A 13722G 14133G 14766T 15257A 15326G 15452A 16069T 16126C 16145A 16231C 16261T
UA12	Ukraine	NO	0	124,835	100	J1c2	1	1-16569	73G 185A 188G 228A 263G 295T 315.1C 462T 489C 750G 1438G 2706G 3010A 3106A 4216C 4769G 7028T 8860G 10398G 11251G 11719A 12612G 13708A 14766T 14798C 15326G 15452A 16069T 16126C 16519C
UA28	Ukraine	NO	0	145,41	100	H5a8	1	1-16569	263G 456T 750G 1438G 4336C 4769G 8860G 11969A 15326G 16304C
UA57	Ukraine	NO	0	161,779	99,994	J1c3j	0,977	1-16569	73G 185A 263G 295T 315.1C 462T 489C 750G 1438G 2706G 3010A 4216C 4769G 7028T 8860G 10398G 11251G 11719A 12358G 12612G 13708A 13934T 14766T 14798C 15326G 15452A 16069T 16126C 16261T 16311C
UA60	Ukraine	NO	1	170,314	100	H7a1a	1	1-16569	93G 263G 524.1AC 750G 1438G 1719A 4769G 4793G 8860G 9630A 11167G 15001Y 15326G 16261T 16519C
UA72	Ukraine	YES	34	1250	100	H	1	1-16569	73R 146Y 152Y 195Y 263G 315.1C 709R 750G 1438G 1888R 2706R 4216Y 4769G 4917R 7028Y 8697R 8745R 8860G 9899Y 10143R 10463Y 11251R 11719R 12633M 13368R 14281Y 14470W 14602R 14766Y 14905R 15323R 15326G 15452M 15607R 15928R 16126Y 16163R 16186Y 16189Y 16221Y 16294Y 16519C
UA96	Russia	NO	0	143,748	99,994	J1c1b1a	0,987	1-16569	73G 185A 228A 263G 295T 315.1C 462T 482C 489C 750G 1438G 2396T 2706G 3010A 3394C 4216C 4769G 5773A 7028T 7184G 8860G 10398G 10463C 11251G 11719A 12612G 13708A 14766T 14798C 15326G 15452A 15933d 16069T 16126C 16145A 16183C 16189C

5.4 HAPLOTYPE DIVERSITY IN DONETSK OBLAST

The next step was to perform an initial set of statistical analyses to evaluate the genetic diversity within the dataset. Specifically, we calculated both haplotype diversity and nucleotide diversity across the entire cohort, representing the present-day population of Donetsk oblast. To further explore potential patterns of maternal lineage variation, we also assessed these diversity indices within the two defined Terminal Maternal Ancestor (TMA) groups, Ukrainian and Russian TMAs.

Haplotype diversity (Hd) measures the uniqueness of haplotypes within a population, with values close to one indicating greater genetic diversity. The Donetsk oblast population, similar to other previously analyzed European populations (Table 6), exhibited exceptionally high haplotype diversity, with an Hd value of 1. This is indicative of either a large and stable population or a population that has experienced numerous maternal gene flows. The exception to this trend is the Sardinian population, known for its historical isolation and higher levels of inbreeding.

Nucleotide diversity (π), which quantifies the average genetic difference per site when comparing sequence pairs, was found to be 0.00178 for the Donetsk oblast population, and 0.00176 and 0.00183 for the Ukrainian and Russian TMAs, respectively (Table 6). These relatively low values suggest that, despite the high haplotype diversity, the overall genetic divergence among individuals remains limited. Such findings are consistent with a population that may have experienced a bottleneck event followed by rapid expansion, similar to patterns observed in other Eastern and Central European groups, including Ukrainians, Russians, Czechs, Slovaks, Poles, and Hungarians. The expected lower nucleotide diversity of Sardinians reflects an isolated population.

The overall similarity between the Ukrainian and Russian TMA groups suggests comparable levels of genetic diversity between them. These findings support the characterization of the Donetsk oblast population as having high haplotype diversity but low nucleotide diversity, consistent with a population shaped by a complex history of admixture and genetic exchange.

Table 6. Comparison of diversity indices in some European populations based on whole mitogenomes.

Population	N	k	Hd (s.e.)	π (s.e.)
Donetsk oblast	91	91	1	0,00178 (0.00001)
Ukrainian TMA	67	67	1	0,00176 (0.00001)
Russian TMA	24	24	1	0,00183 (0.00001)
Ukrainians*	144	141	0.999 (0.001)	0.0018 (0.0001)
Russians*	376	361	0.999 (0.000)	0.0018 (0.0001)
Czechs*	150	140	0.999 (0.001)	0.0017 (0.0001)
Slovaks*	139	133	0.999 (0.001)	0.0017 (0.0001)
Poles*	300	287	0.999 (0.0003)	0.0019 (0.0001)
Hungarians*	80	78	0.999 (0.002)	0.0018 (0.0001)
Sardinians*	63	50	0.992 (0.004)	0.0015 (0.0001)

*N: Sample size; k: Number of haplotypes; Hd (s.e.): Genetic diversity (standard error); π (s.e.): Nucleotide diversity (standard error); * data from (Malyarchuk et al., 2023)*

Following the analysis of haplotype and nucleotide diversity, we next applied Tajima's D statistic to investigate the demographic processes within the two defined TMA groups: the Ukrainian and Russian TMAs. Tajima's D compares two estimates of genetic diversity (the number of segregating sites and the average number of pairwise differences) to detect deviations from neutral evolution. For both the Ukrainian and Russian TMA groups, Tajima's D was significantly negative: −2.35 ($p = 0.019$) for the Ukrainian group and −2.15 ($p = 0.0314$) for the Russian group (Figure 5.3). These negative values suggest an excess of low-frequency polymorphisms, consistent with the signature of population expansions. The observed patterns of high haplotype diversity and low nucleotide diversity in both groups, along with the negative Tajima's D values, point to demographic expansions, possibly following bottlenecks or founder events. The similar Tajima's D values between the Ukrainian and Russian TMAs imply that both groups may have experienced comparable demographic histories, shaped by post-glacial recolonization, later Slavic migrations, and region-specific settlement dynamics. These findings support the notion of a recent population expansion within each

group, allowing for the possibility of correlations with the historical processes that have shaped the Donetsk oblast region.

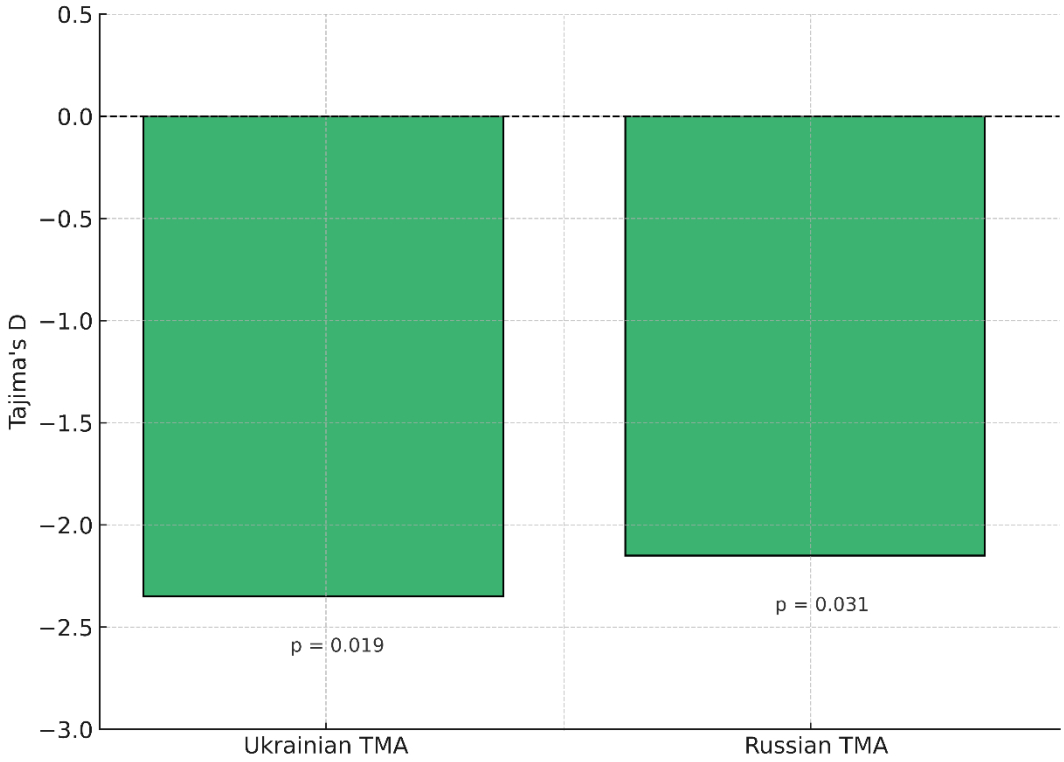


Figure 5.3. *Tajima's D values for the Ukrainian and Russian TMA groups. The negative Tajima's D values for both groups (Ukrainian: −2.35, Russian: −2.15) suggest an excess of low-frequency polymorphisms, consistent with a recent population expansion.*

5.5 GENETIC DIFFERENTIATION BETWEEN UKRAINIAN AND RUSSIAN TMA GROUPS

Thus far, we have focused on examining the genetic diversity within the Ukrainian and Russian TMA groups individually, based on their mtDNA sequences. Now, to directly assess the genetic differentiation between these two groups, we performed a comparison using the Fixation Index (Fst). The calculated Fst value of 1.29×10^{-5} indicates an extremely low level of differentiation between the Ukrainian and Russian TMA groups, suggesting that the two groups share a similar mitogenome pattern. This result aligns with the previous findings from our diversity analyses, reinforcing the notion that these groups are genetically comparable.

Next, we performed an Analysis of Molecular Variance (AMOVA) to further explore the genetic structure of the two groups, still based on their haplotype data. The AMOVA results showed that the

majority of genetic variation is found within each group rather than between them, with only a small proportion (1.6%) of the variation attributed to differences between the Ukrainian and Russian TMA groups ($p = 0.13$). These results further confirm the minimal genetic differentiation between the two mitochondrial gene pools, highlighting that their genetic makeup is largely shared from a female perspective.

5.6 HAPLOGROUP DIVERSITY

Previously, we focused on analyzing the genetic diversity within the Donetsk oblast dataset based on mitochondrial DNA sequences and haplotypes. Now, we shift our focus to comparisons between the haplogroup distributions. Figure 5.4 illustrates the distribution of major haplogroups across the entire dataset. Consistent with previous studies (Malyarchuk *et al*., 2023; Балановский *et al*., 2012), haplogroup H is the most prevalent in this region, accounting for 42.9% of the total. This haplogroup encompasses a wide range of sub-branches, including H1–H3, H5–H7, H10, H11, H13–H15, H23, H34, H36, H41, H54, H55, and H102. Other prominent haplogroups include U5 (11%) and J1 (9.9%). Haplogroups characteristic of western Eurasian populations, such as V and K, were also detected at lower frequencies. Additionally, lineages associated with the Volga region were also identified, such as several sub-branches of haplogroup U, particularly U4 and U5a (including U5a1 and U5a1a). These findings highlight genetic similarities between the Ukrainian population and other European groups. Interestingly, haplogroups D4 and G3, which are of East Asian origin, were observed at a combined frequency of 3.3%. The presence of these haplogroups is consistent with earlier studies of Ukrainian populations, where the East Asian haplogroups D and G, along with Z, were reported at varying frequencies: 1.6% (Pshenichnov *et al*., 2013), 2.5% (Mielnik-Sikorska *et al*., 2013), and 2.8% (Malyarchuk *et al*., 2023). Overall, these results align with existing literature, confirming the extensive mtDNA diversity within Ukraine. This diversity is primarily composed of haplogroups of western Eurasian origin, as demonstrated in earlier works (Pshenichnov *et al*., 2013; Malyarchuk *et al*., 2023; Mielnik-Sikorska *et al*., 2013).

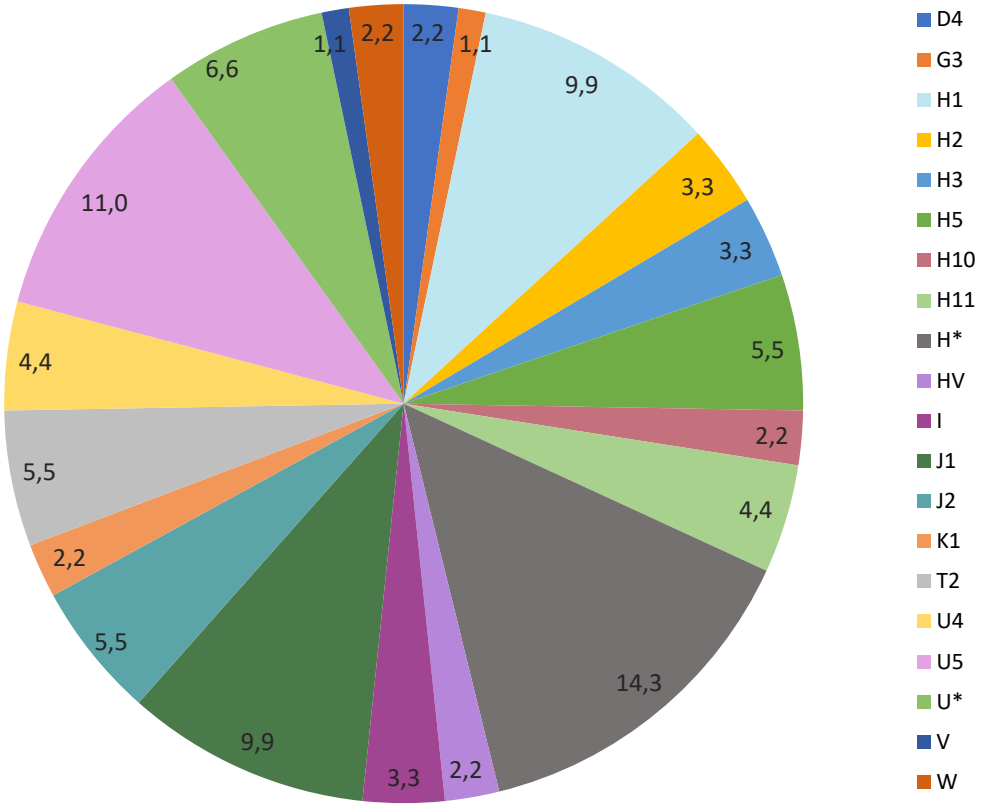


***Figure 5.4**. Haplogroup frequencies (%) in present-day Donetsk oblast population.*

The dataset was further subdivided based on the haplogroups associated with each individual's terminal maternal ancestor (TMA), with the dataset divided into two main groups with Ukrainian and Russian TMAs. Figure 5.5 allows for a clearer comparison of haplogroup distributions between the two subsets.

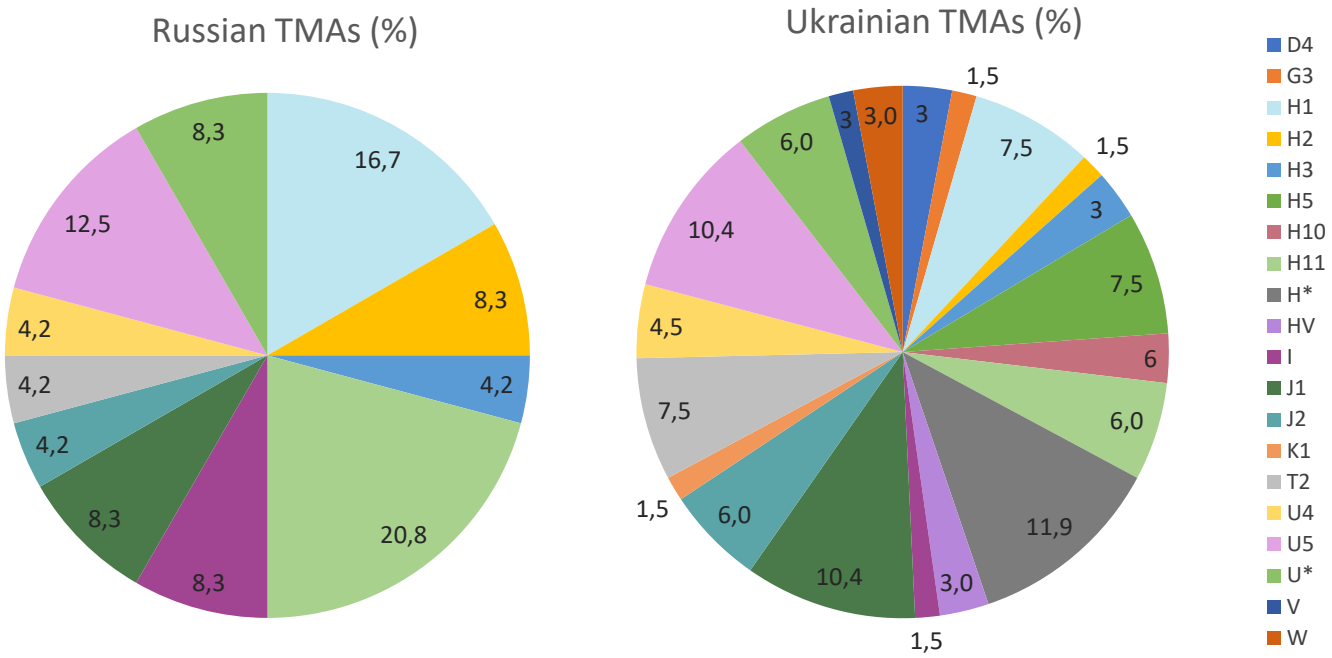


***Figure 5.5**. Haplogroup frequencies (%) among the two distinct TMA groups.*

To better assess the mitochondrial diversity between the two TMAs and potential population-specific haplogroup distributions, we assembled a bar plot with a comparative analysis of haplogroup frequencies between the two TMA subsets (Figure 5.6). The purpose of this analysis was to determine whether certain haplogroups were uniquely associated with one subset or if both groups shared a similar mitochondrial composition. The graph presents the relative frequencies of each major haplogroup, providing a broad overview of the genetic similarities and differences between the two populations.

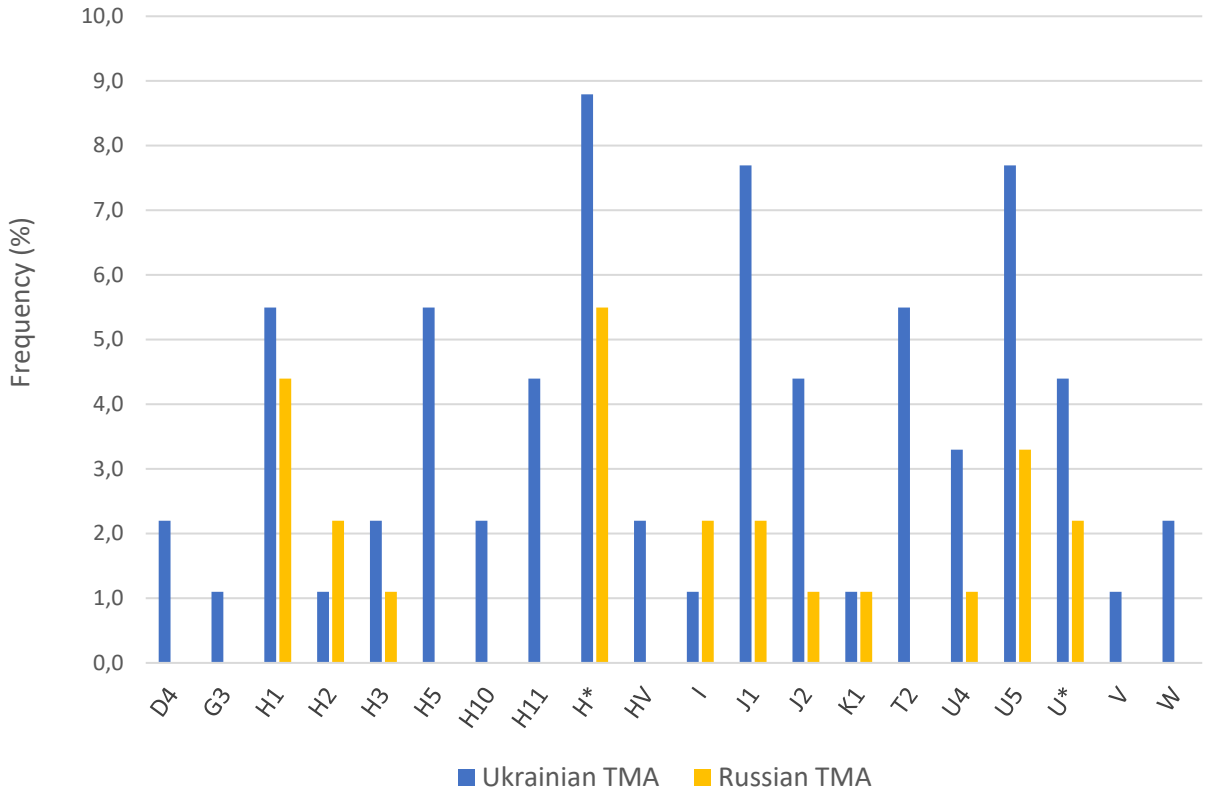


***Figure 5.6**. Comparative analysis of haplogroup frequencies between the two TMA subsets (Ukraine and Russia). H* encompasses haplogroups H6, H7, H13–H15, H23, H34, H36, H41, H54, H55, and H108, while U includes U1–U3, U7, and U8. The graph illustrates variations in haplogroup distribution, highlighting differences in mitochondrial diversity between the two populations.*

The graph illustrates a clear dominance of certain haplogroups in the Ukrainian dataset (blue bars), with haplogroups such as H5, H10, H11, J1, J2, T2 and U5, showing particularly high frequencies compared to the Russian dataset (yellow bars). In contrast, haplogroups such as H1, H2, I and K1 appear more evenly distributed between the two groups, although the Ukrainian dataset still has a higher overall prevalence. Notably, two haplogroups of East Asian origin (D4 and G3) were exclusively present in Ukraine, further suggesting a broader mitochondrial diversity in the Ukrainian lineage. In addition, seven haplogroups of Western Eurasian origin were found solely in Ukraine, reinforcing the idea of greater genetic variability within this population.

To statistically assess the observed differences in haplogroup distribution between the Ukrainian and Russian TMA groups, chi-square tests were performed at two levels of phylogenetic resolution: first using the major haplogroups (e.g., H, U, J), and then using all resolved subhaplogroups. This two-step approach was intended to capture both broad and fine-scale lineage differences between the

groups. The results indicated no statistically significant differences in haplogroup frequencies at either level of comparison. Specifically, the p-value for the major haplogroups was 0.65, and for the subhaplogroup comparison it was 0.16, both well above the conventional threshold for statistical significance. While certain haplogroups appeared to differ in prevalence, such as the exclusive presence of East Asian haplogroups (D4 and G3) in the Ukrainian subset, these differences are likely due to the limited sample size of the Russian TMA group. Therefore, although the observed patterns suggest possible differences in maternal lineage composition, they do not provide statistically robust evidence of population-level differentiation, supporting the conclusion that the Ukrainian and Russian TMA groups are genetically comparable in terms of their mitochondrial lineage composition. Further research with larger and more balanced sample sizes is needed to validate these trends and assess their historical or demographic implications.

5.7 PRINCIPAL COMPONENT ANALYSIS

To explore the genetic relationships reflected in mtDNA haplogroup variation, we performed a principal component analysis (PCA) biplot comparing our Ukrainian dataset (divided into Ukrainian TMA and Russian TMA groups based on Terminal Maternal Ancestors) with various Russian populations. This approach allowed us to visualize genetic affinities between our 91 high-coverage modern mitogenomes and reference datasets. As shown in Figure 5.7, our data were compared with modern mtDNA sequences retrieved from GenBank (Nucleic Acids Research, 2013), including an additional Ukrainian population (labeled as “Ukraine” in the plot) and Russian populations categorized into an East European group and several East Russian groups (e.g., Buryat, Nivkh, Even, Yakut).

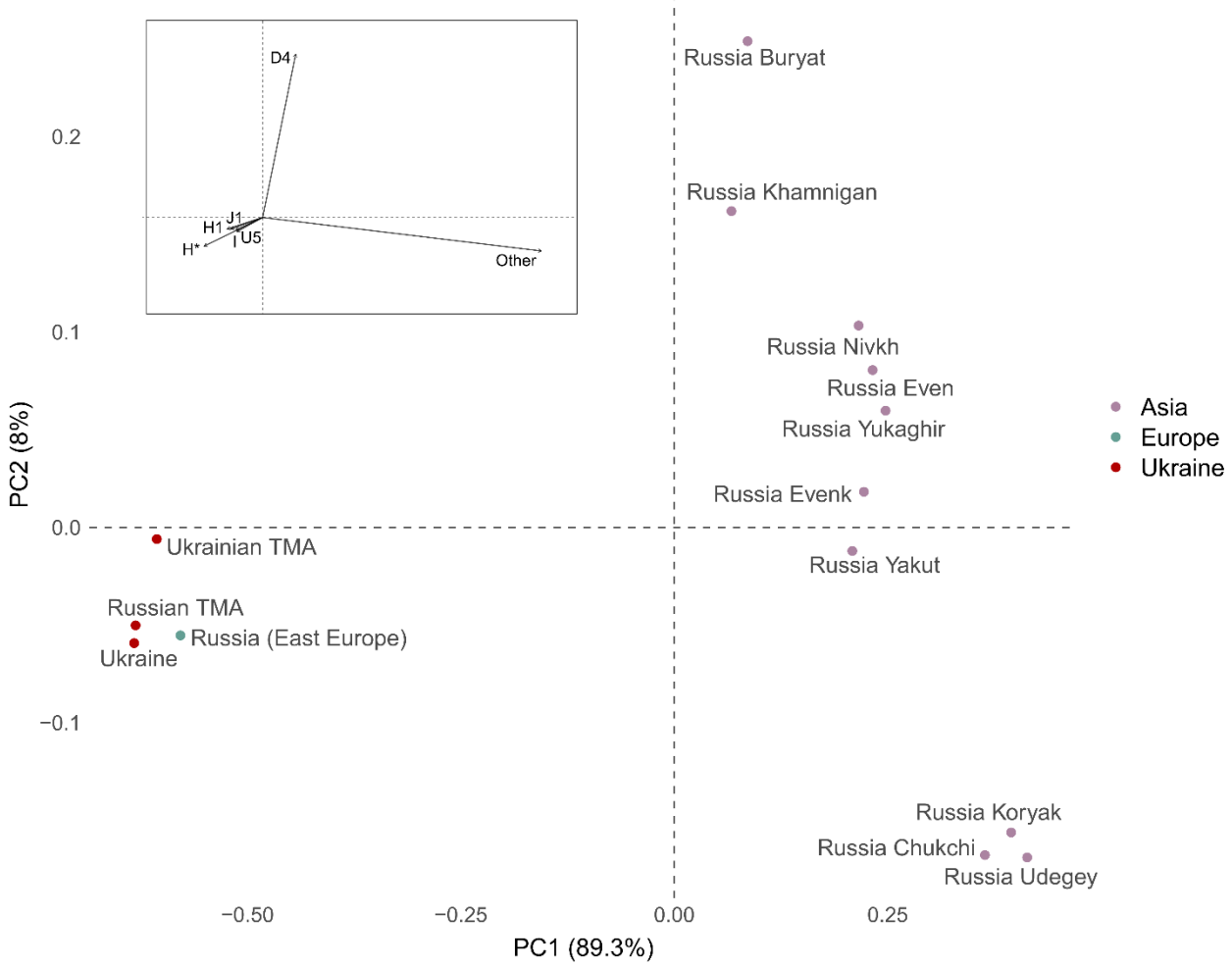


Figure 5.7*. PCA plot based on mitochondrial haplogroup frequencies showing the genetic differentiation between the Ukraine Donetsk dataset (split into Ukrainian TMA and Russian TMA groups), the Ukrainian dataset from GenBank (shown as "Ukraine"), and various East Russian populations. The inset displays the correlation of individual haplogroups with the principal components.*

The PCA reveals clear genetic structuring. Along the first principal component (PC1), which explains 89.3% of the variance, both Ukrainian TMA and Russian TMA samples cluster closely together, positioned near the GenBank-derived Ukrainian dataset, representing a broader East European genetic background. This clustering reflects common maternal lineages within Eastern Europe. In contrast, East Russian populations are widely scattered along both PC1 and PC2 axes, indicating substantial genetic divergence from the Donetsk TMAs and the Eastern European cluster. This segregation illustrates distinct demographic histories and maternal genetic profiles between East Siberian/Far eastern groups and those from eastern Europe. Given this pronounced divergence and

the focus of this study on the Ukrainian genetic variation, these East Russian populations were excluded from subsequent analyses.

After excluding the highly divergent eastern Russian populations, a second PCA was performed to better resolve the genetic affinities of our dataset within a western Eurasian context (figure 5.8). In this analysis, the Russian TMA group was merged with the Russian European population to balance sample sizes and avoid artificial divergence due to the smaller number of individuals (only 24 for the Russian TMA).

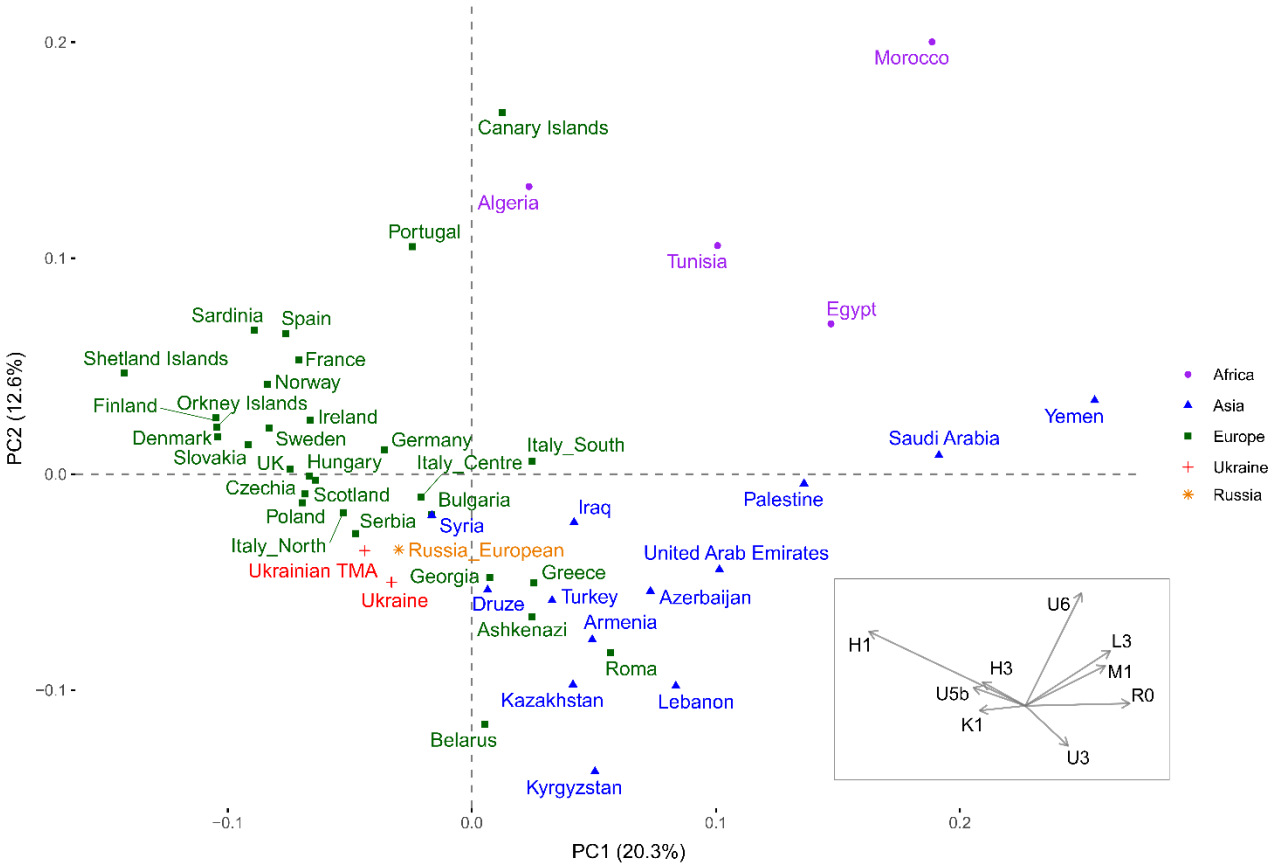


Figure 5.8. *PCA plot based on mitochondrial haplogroup frequencies showing the genetic relationships between our Ukrainian dataset and various Western Eurasian populations. The Ukrainian TMA group is shown in red, while the Russian TMA is merged with the Russian European group and they are shown in orange. The inset displays the correlation of individual haplogroups with the principal components.*

The plot shows that the PC1 (20.3% variance) is the primary axis separating European populations from North African and certain Asian groups. North African populations, such as Morocco, Algeria, Tunisia, and Egypt, are clearly displaced along PC1, driven by the presence of haplogroups, such as

U6 and L3, which are known markers of African maternal lineages. Several Middle Eastern and Central Asian populations, including Yemen, Saudi Arabia, and Kazakhstan, are also pulled away from the European cluster, reflecting distinct haplogroup compositions and maternal ancestries. PC2 (12.6% variance) captures additional variation within Western Eurasia, reflecting subtle differences between European and Mediterranean populations.

Regarding the focus of our study, both the Ukrainian TMA and the merged Russian component cluster closely with European populations, particularly Eastern Europeans such as Poland, Slovakia, and Serbia, reflecting shared maternal genetic backgrounds. The results support the interpretation that the mitochondrial lineages present in the Donetsk region are consistent with long-standing genetic continuity in Eastern Europe, and show no significant deviation that would suggest distinct maternal ancestry outside of this context.

5.8 PHYLOGENETIC ANALYSES

Maximum parsimony analysis of the 91 mitogenomes from our dataset confirms extensive mtDNA diversity within the region, represented by 80 distinct haplogroups and sub-haplogroups (Figure 5.9).

One notable observation is that the tree reflects how haplogroups H and U play a central role in its topology, with their branches and sub-branches accounting for a substantial portion of the tree structure. Haplogroup H appears highly diversified, with numerous subclades such as H1, H2, and H5 distributed radially throughout the tree, indicating widespread maternal ancestry and the presence of multiple lineage expansions. Haplogroup U is similarly well represented, with tightly clustered sub-haplogroups including U4, U5, and U2e, reflecting a more structured phylogenetic pattern that may point to historical founder effects or persistent regional continuity.

Each sample in the tree is annotated by the TMA, allowing for a visual distinction between lineages traced to Ukrainian TMAs (highlighted in blue) and Russian TMAs (highlighted in yellow). This color coding proves that while the majority of lineages are of Ukrainian maternal origin, Russian TMAs are interspersed across various haplogroups, indicating overlapping maternal ancestries and potential gene flow between populations.

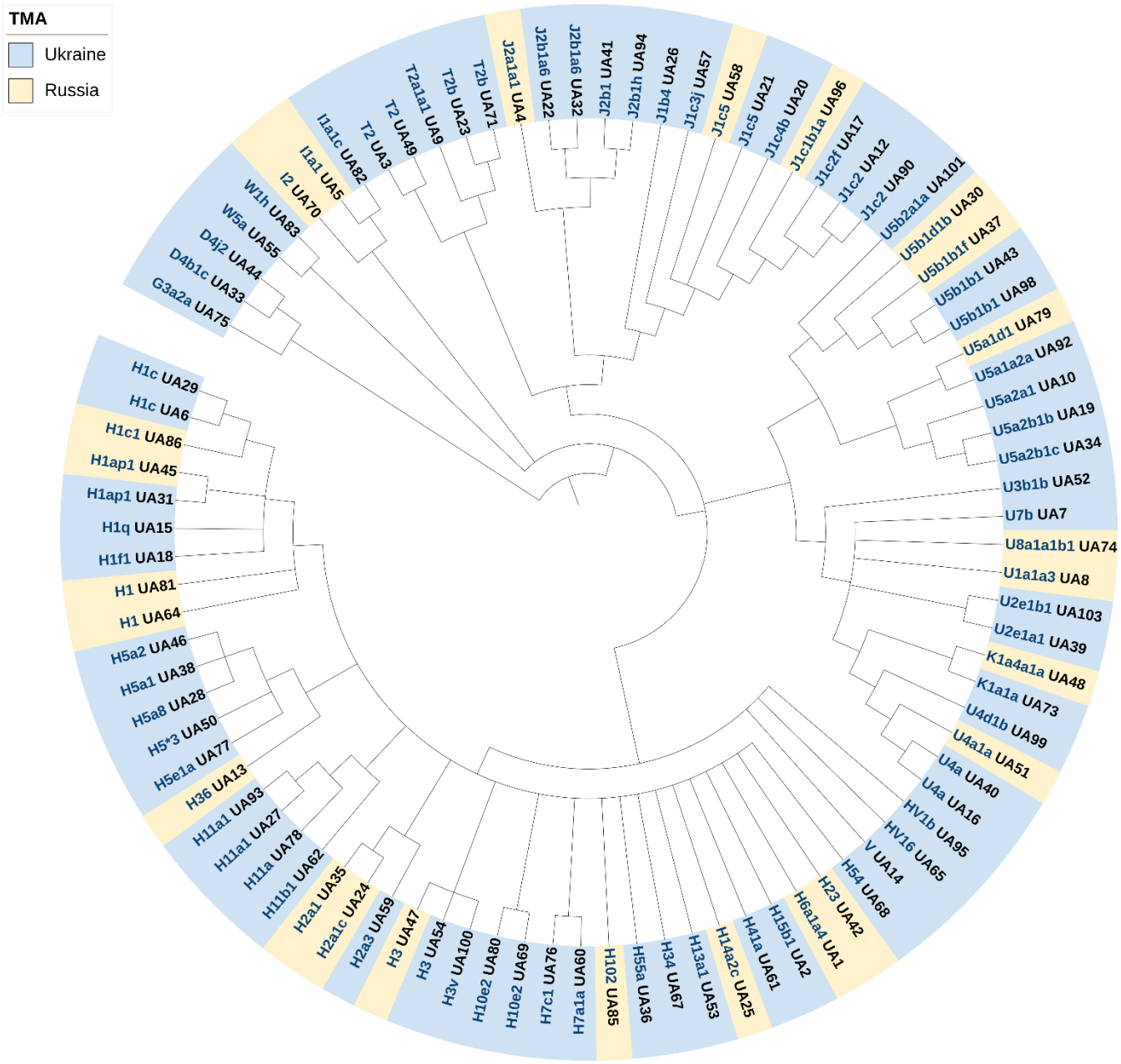


Figure 5.9. *Maximum parsimony tree representing the complete dataset of the Donetsk Oblast population in Ukraine. The individuals are highlighted based on their TMA.*

Although the comparison of major mitochondrial haplogroups between Ukrainian and Russian TMAs revealed substantial overlap (Figure 5.6), reflecting a common maternal genetic background, finer resolution at the subhaplogroup level uncovers a more nuanced picture. It is evident that only a limited number of subhaplogroups are shared between the two groups. Specifically, haplogroup J1c5 is present in both sample UA58 and UA21, H1ap1 is found in UA45 and UA31, and H3 appears in UA47 and UA54. These common lineages suggest some degree of maternal lineage continuity across

TMA origins. However, the overall pattern points to a greater diversification at the sublineage level. This observation underscores the value of high-resolution mtDNA analysis in revealing subtle genealogical differences that might be obscured when considering only broader haplogroup classifications.

5.9 FOCUS ON THE MITOGENOMES WITH UKRAINIAN TMAs

5.9.1 Geographic overview of the Ukrainian Terminal Maternal Ancestors

After examining the genealogical documents related to the origins of the 91 mtDNAs, we identified 67 with TMAs from Ukraine and 24 with TMAs from Russia. Among the individuals whose mtDNA was sequenced and had TMAs from Ukraine, approximately 45% were from Donetsk oblast. To illustrate the birthplaces of the Ukrainian TMAs, a map was generated (Figure 5.10), that shows the extensive geographic distribution of the mtDNAs analyzed in this study. The map confirms a significant concentration of TMAs in Donetsk oblast. Other notable groups are observed in central and southern Ukraine, particularly around major cities such as Vinnytsia, Odesa, and Kharkiv. These distributions suggest that mitochondrial lineages are not evenly spread across the country but rather exhibit localized clustering, potentially influenced by historical migration patterns, demographic shifts, or regional population densities. In addition, the presence of TMAs in diverse locations supports the idea of a well-distributed genetic heritage, reinforcing the comprehensive nature of this study.

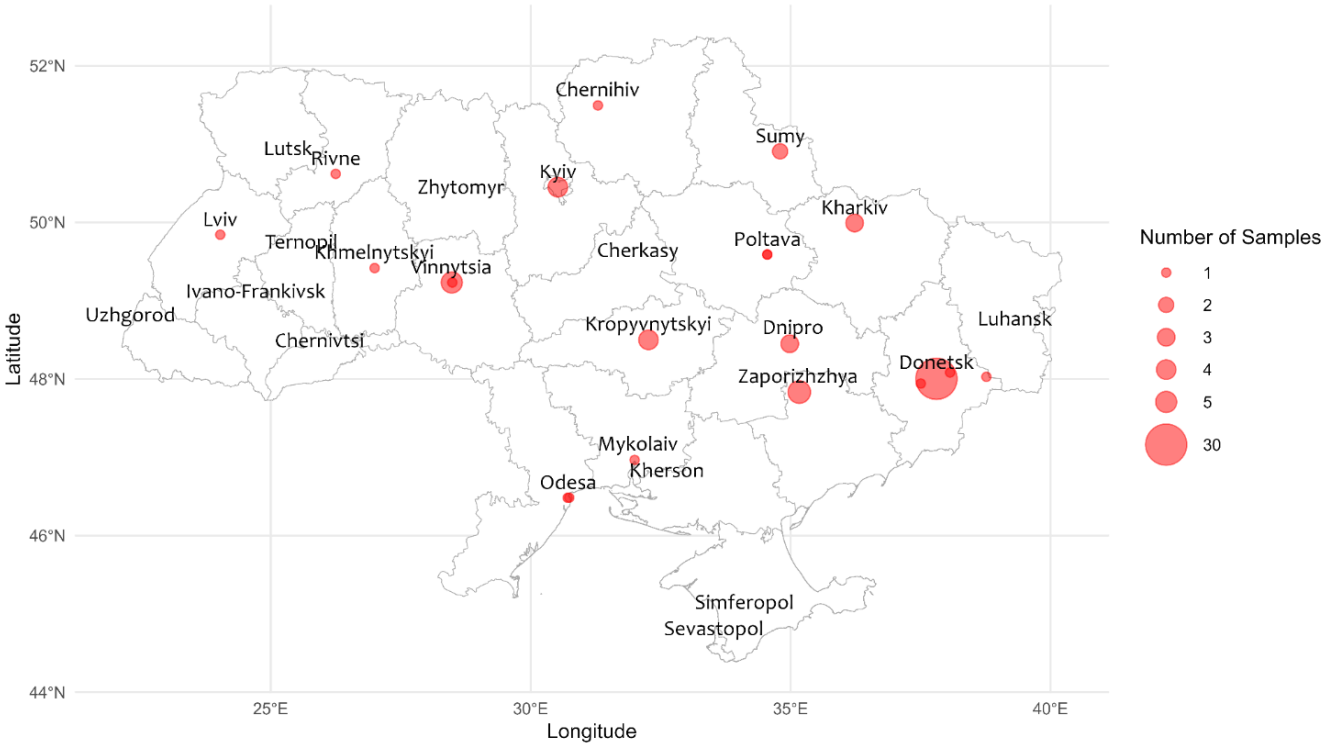


Figure 5.10. *Geographic distribution of TMAs from Ukraine, with birthplace localities represented by red dots. The size of each dot corresponds to the number of probands traced to a specific location or region. When exact birthplaces were available, they are shown accordingly; otherwise, the dots are positioned at the administrative center of the respective region.*

By mapping the birthplaces of individuals' terminal maternal ancestors, we establish a spatial context for the genetic data analyzed in this study. Building on this distribution, the following sections apply statistical and population genetic approaches to explore potential patterns and differences within Ukraine and in comparison, to neighboring populations.

5.9.2 Comparison of mitogenome diversities between eastern and western Ukraine

To further investigate regional variation within the Ukrainian dataset, the country was subdivided into two groups based on a major geographic landmark: the Dnieper River. This river has historically served as a natural barrier and socio-political boundary within Ukraine, shaping patterns of settlement, trade, and migration over centuries. In population genetics, such geographic features are often considered potential drivers of genetic differentiation, as they may influence gene flow and lead to distinct population structures on either side. For this analysis, the western group comprised 19 individuals with TMAs from the oblasts of Khmelnytskyi, Kirovohrad, Kyiv, Lviv, Mykolaiv, Odesa,

Rivne, and Vinnytsia. The eastern group included 48 individuals whose TMAs were from the oblasts of Dnipropetrovsk, Donetsk, Kharkiv, Poltava, Sumy, and Zaporizhzhia. To visualize this division, a map was generated highlighting two mtDNA distributions (Figure 5.11), with the Dnieper River serving as the dividing line. Two pie charts are positioned on each side of the map to summarize haplogroup frequencies per region, with accompanying labels indicating the number of individuals in each group. To explore potential genetic structure associated with this east–west division, we first examined the distribution of mitochondrial haplogroups between these two Ukrainian clusters.

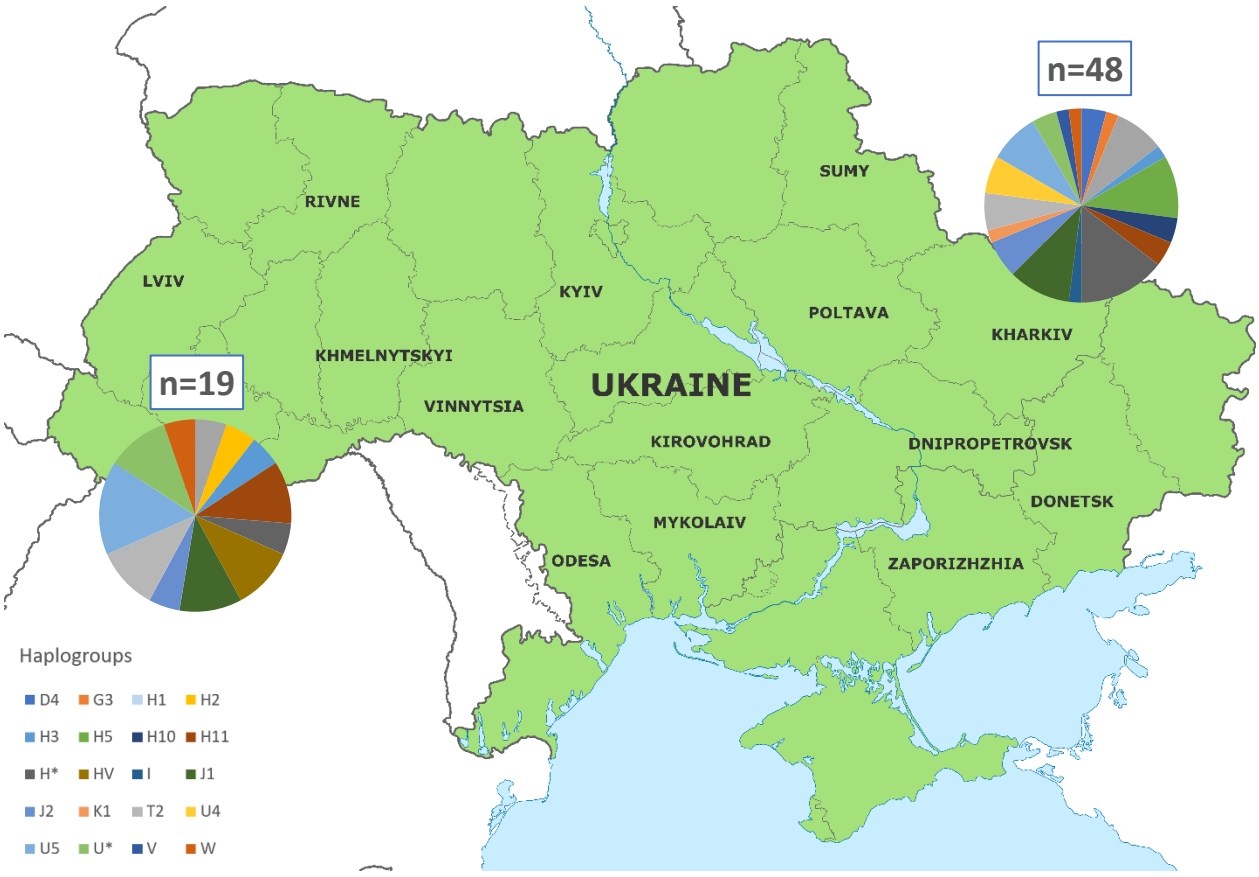


Figure 5.11. *Map showing the geographic subdivision of Ukrainian TMAs into eastern and western regions. The map includes the names of all Ukrainian regions represented in the dataset, with pie charts displaying haplogroup frequencies for each division. The western group includes 19 individuals, while the eastern group includes 48. Regions straddling the Dnieper, such as Kyiv, Dnipropetrovsk, and Zaporizhzhia, were classified based on historical, cultural, and geopolitical contexts. Dnipropetrovsk and Zaporizhzhia oblasts were considered part of Eastern Ukraine, while Kyiv oblast was assigned to Western Ukraine.*

To assess whether this east–west division is reflected in the mitochondrial haplogroup distributions, a chi-square test was conducted. The result ($\chi^2 = 18.921$, $p = 0.50$) indicated no statistically significant difference between the two groups. While this suggests a broad genetic continuity across the Dnieper

divide, the absence of significance does not preclude the existence of subtle regional trends. For a clearer visual representation of the haplogroup patterns, a bar chart comparing the frequencies in the two regions has been included (Figure 5.12).

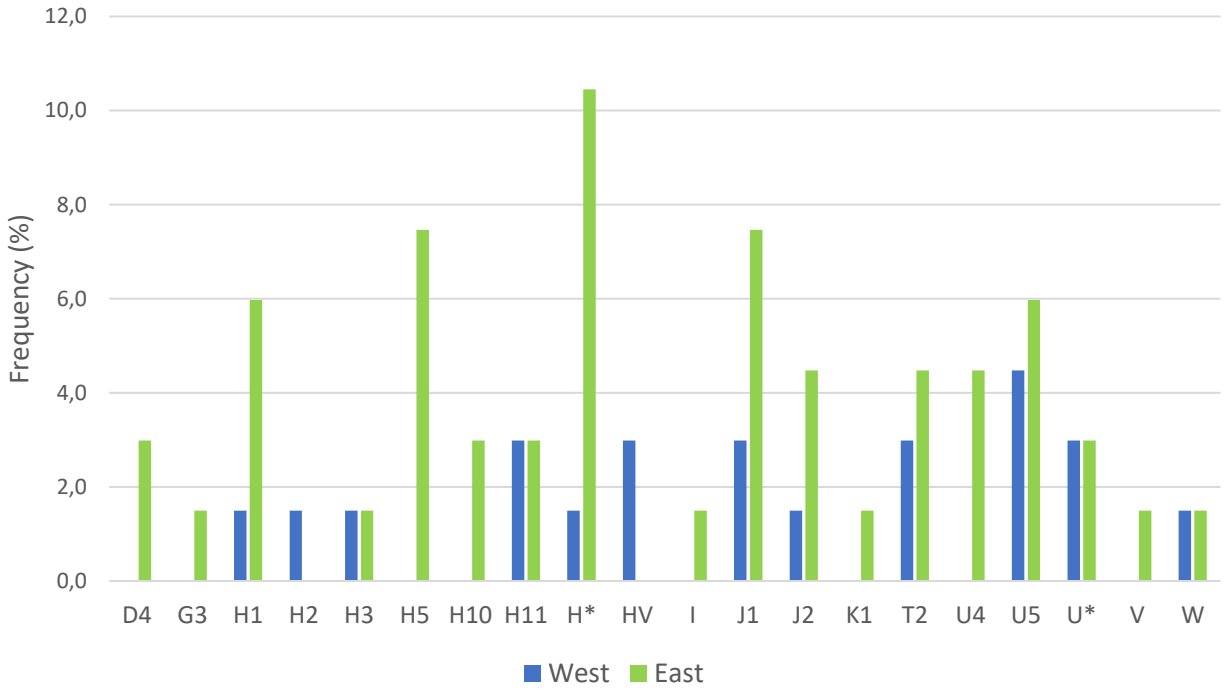


Figure 5.12. *Frequency distributions (%) of mitochondrial haplogroups in western and eastern Ukraine. The chart illustrates relative haplogroup representation across the two regional groups.*

Although the chi-square test revealed no statistically significant difference in mitochondrial haplogroup distributions between western and eastern Ukraine ($p = 0.50$), the corresponding bar plot highlights several notable patterns. Given the larger sample size in the eastern group (48 individuals compared to 19 in the west), overall haplogroup frequencies are more strongly represented for this region. Nonetheless, there are some differences in the relative proportions of specific haplogroups. For instance, H1, H5, H*, J1, and J2 appear more frequently in the eastern cluster, while haplogroups H2 and HV are found exclusively in the western group. Although these variations do not achieve statistical significance, they may reflect subtle regional dynamics or the influence of local maternal lineages, and they provide useful context for interpreting broader population patterns. As such, the visual comparison serves to complement the statistical results and may help guide future research with expanded datasets.

To provide a broader geographic context, a pairwise F_{ST} analysis was performed to assess mitochondrial differentiation between Western Ukraine, Eastern Ukraine, and Russian TMAs. The results, visualized in a heatmap (Figure 5.13), revealed very low F_{ST} values across all comparisons ($\leq 1.5 \times 10^{-5}$), indicating minimal genetic structuring among the three groups. Interestingly, the smallest genetic distance was observed between eastern and western Ukraine, consistent with their geographic proximity and shared historical interactions. Intriguingly, the greatest differentiation is observed between Russia and eastern Ukraine, though the magnitude of this difference remains extremely low. This could be due to sampling bias. Overall, these findings suggest that while minor regional patterns may exist, there is no substantial mitochondrial divergence among the three groups.

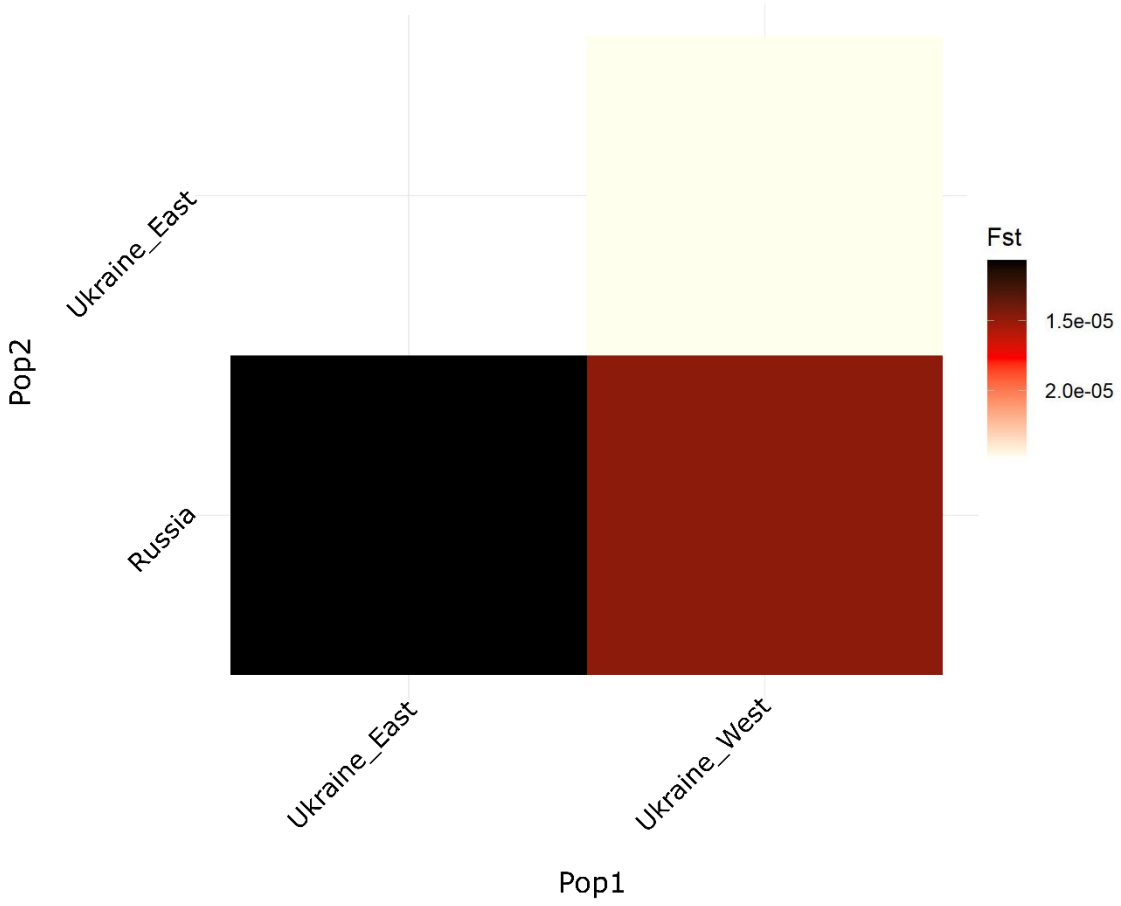


Figure 5.13. *Heatmap showing pairwise F_{ST} values based on complete mitochondrial sequences between Eastern Ukraine, Western Ukraine, and Russian TMAs. The color gradient reflects the degree of genetic differentiation, with darker shades indicating higher F_{ST} values. All values are extremely low ($\leq 1.5 \times 10^{-5}$), suggesting minimal mitochondrial sequence divergence between the groups.*

To further evaluate the distribution of genetic variation, an AMOVA (Analysis of Molecular Variance) was performed on the same three groups. The results confirmed that only 1.7% of the total genetic variance could be attributed to differences between regions, with the remaining 98.3% found within groups ($R^2 = 0.0172$, $p = 0.77$). This very low and statistically non-significant level of between-group variance reinforces the conclusion drawn from the F_{ST} analysis that there is broad mitochondrial continuity across Ukrainian and Russian TMAs, shaped by long-term gene flow rather than distinct population splits.

5.9.3 Comparison of mitogenome diversities after the 2022 Russian Annexation

Following the analysis of regional patterns based on geographic landmarks, we decided to conduct another comparative analysis of mitochondrial haplogroup frequencies among three groups: individuals with TMAs from the Russian annexed regions of Ukraine (Donetsk and Zaporizhzhia, n = 38), those from the rest of Ukraine (n = 29), and those with TMAs located in Russia (n = 24). This subdivision aimed to assess whether the geopolitical and historical distinctions between these areas are reflected in maternal genetic variation. To visualize the sample distribution and haplogroup diversity across the three groups, a new map was generated (Figure 5.14).

As in the previous section, we first evaluated haplogroup frequencies across the three groups to identify broad maternal lineage trends, before analyzing sequence-level variation for finer-scale insights.

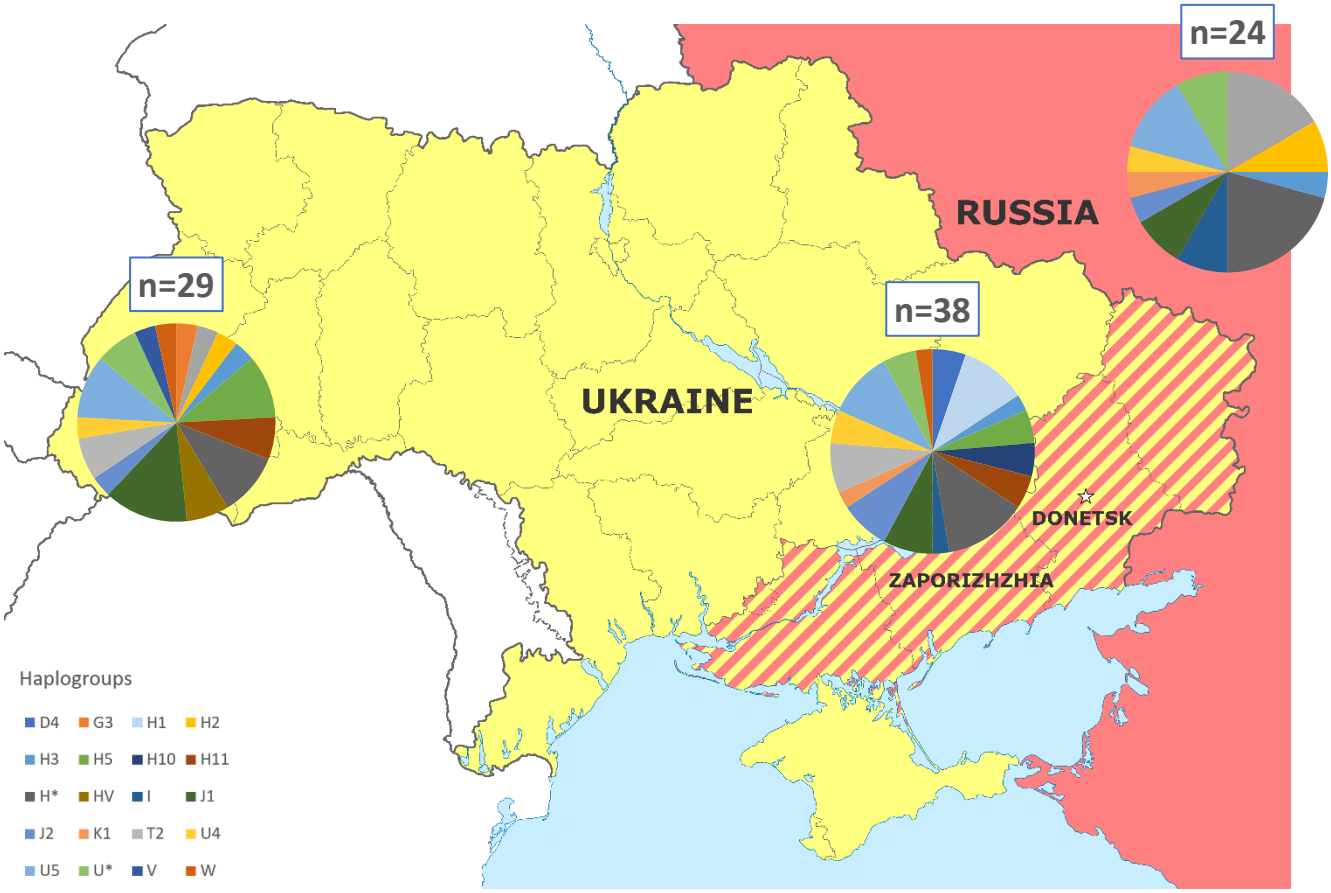


Figure 5.14. *Map showing the geographic subdivision of the dataset into three groups: individuals from the annexed regions of Ukraine (Donetsk and Zaporizhzhia, n = 38), the rest of Ukraine (n = 29), and those with TMAs located in Russia (n = 24). The map includes the names of all regions represented in the dataset, with pie charts displaying the mitochondrial haplogroup frequencies for each group. This classification reflects the geopolitical and historical distinctions between these regions, with the map visually representing the sample distribution and haplogroup diversity across these groups.*

A chi-square test was performed to examine whether there were any significant differences in haplogroup distributions between the three groups. The results indicated no statistically significant difference (χ^2= 33.154, p = 0.74), indicating that mitochondrial haplogroup distributions are broadly similar across the annexed regions, the rest of Ukraine, and Russia. This suggests a general maternal genetic continuity, likely shaped by long-term gene flow and shared historical ancestries across these neighboring regions. Nonetheless, the lack of statistical significance does not preclude the presence of more nuanced regional trends that may be masked by differences in sample size or overlapping lineage histories.

To complement the statistical analysis, a bar plot (Figure 5.15) has been included to provide a more detailed visualization of haplogroup frequency differences among the three populations. While most

lineages are shared and similarly distributed, certain haplogroups show localized peaks or exclusive presence in a single group, suggesting subtle distinctions that may reflect region-specific maternal histories.

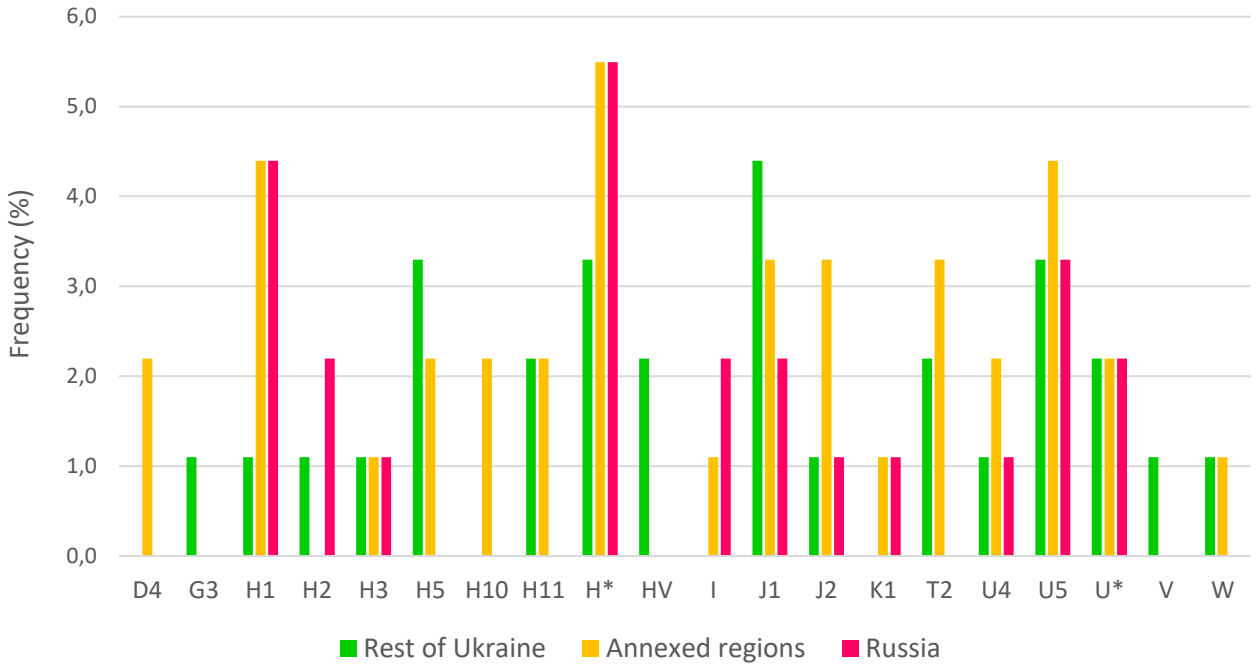


***Figure 5.15**. Frequency distribution (%) of mitochondrial haplogroups in individuals with TMAs from Russian annexed regions of Ukraine (Donetsk and Zaporizhzhia, n = 38), the rest of Ukraine (n = 29), and Russia (n = 24). The chart visualizes the relative distribution of haplogroups across the three groups.*

The bar plot offers a visual overview of haplogroup distributions across the three groups and highlights patterns that, while not statistically significant, offer insight into regional trends. Several haplogroups, namely H1, H3, H*, J1, J2, U*, U4, and U5, are shared across all three groups, forming a common maternal lineage background throughout Ukraine and Russia. However, some haplogroups appear to be restricted to specific regions. For instance, D4 and H10 are found exclusively in the annexed Ukrainian regions, while G3, HV, and V are only present in the rest of Ukraine. Notably, no haplogroups are uniquely found in the Russian TMA group. Additionally, haplogroups I and K1 are shared only between the annexed Ukrainian regions and Russian TMAs, suggesting a possible maternal lineage connection between these groups that is not observed with the rest of Ukraine.

While these patterns may suggest subtle regional signals, it is important to interpret them cautiously. The differences could be influenced by uneven sample sizes, local sampling biases, or the stochastic nature of mitochondrial lineage distribution. Moreover, the absence of statistically significant results

in the chi-square test supports the conclusion that, on a broader scale, haplogroup diversity does not differ markedly between these groups.

To complement the haplogroup analysis and investigate potential sequence-level differences, a pairwise F_{ST} analysis was conducted among the three groups: individuals with TMAs from the annexed Ukrainian regions (Donetsk and Zaporizhzhia), the rest of Ukraine, and Russia (Figure 5.16). The F_{ST} values were uniformly very low across all comparisons ($\leq 1.6 \times 10^{-5}$), indicating minimal mitochondrial differentiation. The closest genetic similarity was observed between the annexed Ukrainian regions and the rest of Ukraine, while the largest differentiation was found between Russia and the rest of Ukraine. This pattern is broadly consistent with geographic and cultural expectations: the two Ukrainian groups, despite recent geopolitical fragmentation, remain closely related both geographically and historically, while the slight increase in differentiation with the Russian group aligns with broader East-West distinctions. Nonetheless, all observed F_{ST} values are extremely low, suggesting that these differences are not biologically meaningful and likely reflect stochastic variation rather than structured genetic divergence. Overall, the results support a picture of mitochondrial continuity across the region, with no evidence of population-level separation based on maternal lineages.

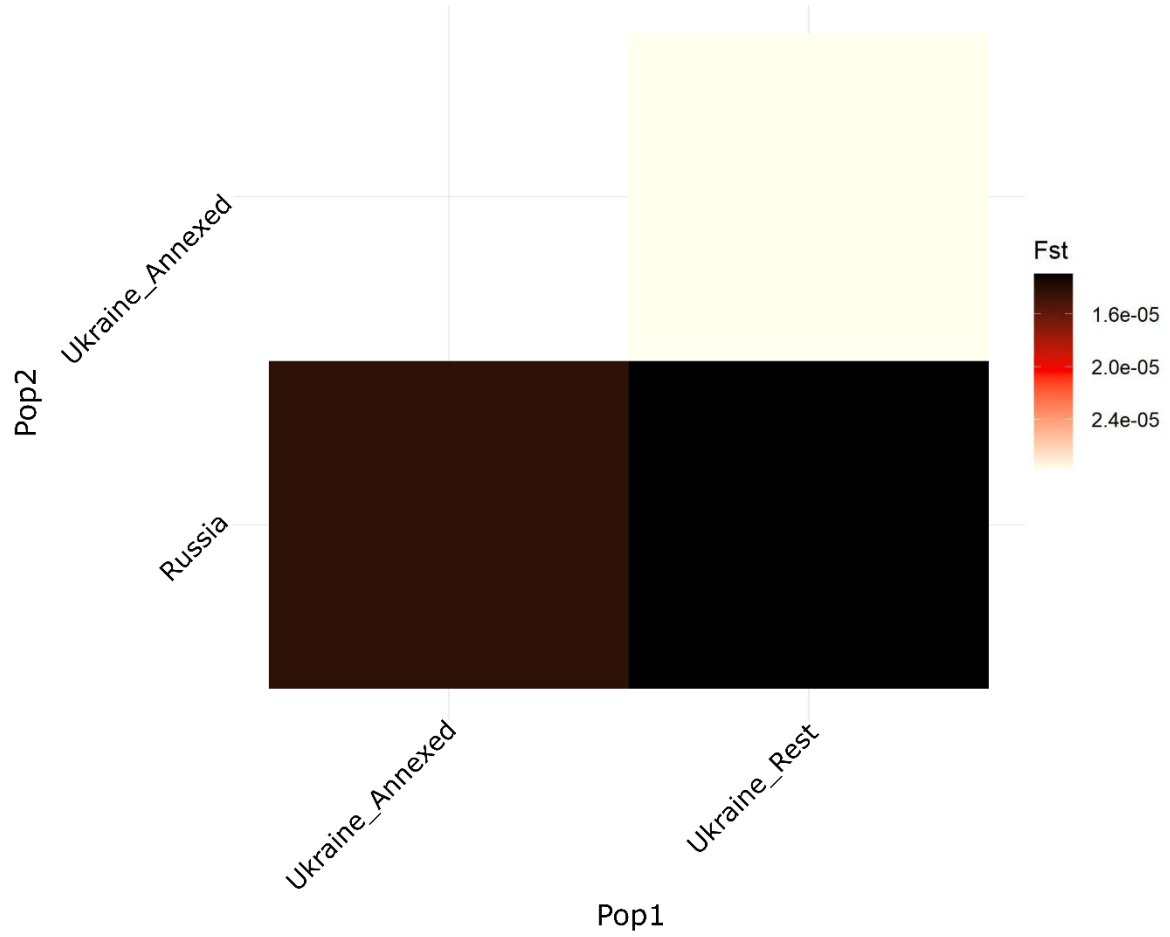


Figure 5.16. *Heatmap showing pairwise F_{ST} values based on complete mitochondrial sequences between annexed Ukrainian regions, the rest of Ukraine and Russian TMAs. The color gradient reflects the degree of genetic differentiation, with darker shades indicating higher F_{ST} values. All values are extremely low (≤1.6 × 10^{-5}), suggesting minimal mitochondrial sequence divergence between the groups.*

To further validate these findings, an Analysis of Molecular Variance (AMOVA) was performed using the same three-group subdivision. The results showed that only 1.4% of the total genetic variance could be attributed to differences among groups ($R^2 = 0.014$), with the vast majority of variation (98.6%) occurring within groups. The associated p-value of 0.93 confirms that this among-group variance is not statistically significant. Together, the F_{ST} and AMOVA results reinforce the conclusion that, despite recent political and historical events, there is no significant mitochondrial differentiation between the annexed regions, the rest of Ukraine, and Russia. These populations appear to share a broadly common maternal genetic background shaped by long-term gene flow and shared ancestral histories.

6. CONCLUSIONS

In recent decades, the field of population genetics has been transformed by the advent of high-throughput technologies such as next-generation sequencing (NGS), which have enabled extraordinary access to large-scale genomic data from both modern and ancient individuals. These advances have allowed for increasingly fine-resolution reconstructions of human population histories, revealing complex patterns of migration, admixture, and isolation. Within this broader context, regions that have historically acted as contact zones, such as eastern Ukraine, hold particular significance. Donetsk oblast, shaped by its position on the Eurasian steppe and a legacy of cultural interchange, provides a compelling case study for understanding how geography, history, and demography intersect to shape genetic diversity.

The aim of this thesis was to investigate the genetic variability of the Ukrainian population in order to reconstruct its origins and demographic history within the broader European context, through the integrated analysis of genomic and genealogical data from individuals of the Donetsk oblast. The initial phase of the study employed genealogical records to trace the geographic distribution of Terminal Maternal Ancestors (TMAs) among 104 sampled individuals. This revealed that while the majority of TMAs were located within Ukraine, approximately one-third traced back to western Russia, a finding consistent with the region's shared historical, cultural, and geographic continuity. The complete mitochondrial genome data from 91 high-quality sequences revealed exceptionally high haplotype diversity (Hd = 1), paired with low nucleotide diversity ($\pi = 0.00178$), a combination typically associated with large and expanding populations, or with groups that have undergone historical bottlenecks followed by rapid growth. Demographic inference based on Tajima's D values was consistent with this scenario: both the Ukrainian and Russian TMA groups displayed significantly negative values, indicating an excess of rare variants and supporting models of recent expansion. These findings mirror genetic patterns observed in other Central and Eastern European populations.

Comparative genetic analyses between Ukrainian and Russian TMAs revealed only minimal differentiation, suggesting a substantial degree of shared maternal ancestry. Phylogenetic assessments supported this, revealing a broad diversity of haplogroups without evidence of pronounced lineage structure. Haplogroup analysis confirmed the prevalence of Western Eurasian lineages, particularly subclades of H, U, and J, as expected from prior studies of Ukrainian and Central-Eastern European populations. The presence of East Asian haplogroups, although limited, hints at deeper and more complex migratory histories across the Eurasian steppe.

A key strength of this study lies in its multi-layered comparative analyses. The Ukrainian and Russian TMA groups were directly compared through Fst and AMOVA, which revealed negligible differentiation and indicated that the vast majority of genetic variation resides within rather than between the two groups. This shared maternal ancestry was further reflected in phylogenetic trees, where TMAs from both origins were widely dispersed, showing no clear clustering or lineage-specific divergence. However, despite this general homogeneity, more granular differences emerged when considering haplogroup composition.

Haplogroup H was the most frequent in the dataset (42.9%), followed by subclades of U (especially U5) and J (notably J1 and J2), reflecting trends common across the Slavic and European gene pool. Yet some notable distinctions were observed. For instance, two East Asian haplogroups, D4 and G3, were found exclusively in the Ukrainian TMA group. While their frequencies were low, their presence is significant, as they reflect ancient connections across the Eurasian steppe and may be remnants of historical gene flow from eastern populations. Additionally, several Western Eurasian haplogroups, such as H10, H11, T2 and several rarer U lineages, were found only in Ukrainian TMAs, further suggesting a broader mitochondrial variability within this group. These visual differences in haplogroup composition, however, did not reach statistical significance in chi-square tests, likely due to the limited sample size of the Russian TMA subset. Nevertheless, the patterns hint at potentially richer maternal heterogeneity within the Ukrainian group, possibly shaped by more complex or regionally varied admixture histories.

Principal Component Analysis (PCA) further contextualized the genetic affinities of the Donetsk population, revealing a tight clustering of both Ukrainian and Russian TMAs with Central and Eastern European groups such as Poles, Slovaks, and Serbs, while clearly separating them from North African and Central Asian populations. This reinforces the finding that the Donetsk maternal gene pool is firmly embedded within the broader Eastern European genetic continuum. Phylogenetic reconstruction offered yet another layer of resolution. Among 80 distinct mitochondrial haplogroups identified, very few lineages were shared between Ukrainian and Russian TMAs, highlighting a high degree of individual-level diversity even within a context of overall genetic similarity. This fine-scale diversity, when combined with genealogical data, adds depth to our understanding of maternal lineages and their historical diffusion across the region.

The geographic resolution of the dataset enabled additional layers of comparison. Within Ukraine, 45% of the TMAs originated in Donetsk, which aligns with the high percentage of probands from this region. On the other hand, the remaining TMAs were distributed across a wide range of regions, indicating no strong pattern of maternal localization. A comparison between East and West Ukrainian

TMAs found no statistically significant differences in diversity metrics, but subtle haplogroup variations suggest potential regional structuring, which may become clearer with expanded sampling. Moreover, the analysis of TMAs from annexed regions (i.e., territories affected by the 2022 Russian invasion) revealed a comparable mitochondrial profile to those from both Russia and the rest of Ukraine, underscoring the deep genetic continuity across modern political borders.

Taken together, these results portray Donetsk oblast as a genetically diverse and demographically dynamic region. Despite strong maternal affinities with neighboring Slavic populations, the mitochondrial diversity observed here reveals a complex legacy of migration, admixture, and shared ancestry. Importantly, this work demonstrates how high-resolution mitogenome data, when coupled with genealogical and geographical context, can illuminate both broad and subtle patterns in population structure. By focusing on a region shaped by millennia of human movement, and by recent geopolitical shifts, this study adds valuable insight into the genetic history of Ukraine and contributes to the growing understanding of how political, cultural, and historical forces are reflected in the genetic pool of a population. It also affirms the central role of Ukraine as both a biological and historical bridge within the Eurasian landscape. Future studies with larger sample sizes could further refine these insights and provide a more nuanced understanding of the genetic history of Europe.

7. ONGOING RESEARCH

In parallel with the mitochondrial-focused analysis presented in this thesis, the broader project on the genetic history of the Donetsk population encompasses additional layers of investigation that aim to provide a more comprehensive picture of Ukrainian genomic diversity. Genome-wide data from 45 individuals, selected from the same original cohort and genotyped using the Axiom™ Genome-Wide Human Origins 1 Array, are currently being analyzed to explore biparental inheritance patterns. This high-density SNP array allows for in-depth study of population structure, admixture, and historical demography through autosomal markers, offering a complementary perspective to the uniparental mitogenome data.

Moreover, a diachronic dimension is being added through the integration of ancient DNA data from seven archaeological individuals associated with three culturally distinct groups, Yamna, Catacomb, and Scythian, whose remains were recovered within the territory of modern Ukraine. Radiocarbon dating has confirmed their temporal context, and their genomic sequencing is underway. These ancient genomes will serve as pivotal references for assessing genetic continuity, replacement events, and admixture patterns between prehistoric and modern populations.

Collectively, these ongoing efforts aim to bridge the temporal gap between past and present, refining the reconstruction of population dynamics in Ukraine across millennia. By combining ancient and modern datasets, and incorporating both uniparental and biparental markers, this research initiative strives to deliver a multi-layered narrative of genetic history, one that situates the Donetsk region within the broader Eurasian context and addresses long-standing questions regarding its ancestral origins and demographic transformations.

8. REFERENCES

Abdi, H., Williams, L.J. (2010). Principal component analysis. WIREs Comp Stat, 2: 433-459.

Amnesty International. (2022). Ukraine: Illegitimate results of sham referenda must not enable illegal annexation of occupied areas. https://www.amnesty.org/en/latest/news/2022/09/russia-ukraine-illegitimate-results-of-sham-referenda-must-not-enable-illegal-annexation-of-occupied-areas/

Amorim, A., Fernandes, T., Taveira, N. (2019). Mitochondrial DNA in human identification: a review. *PeerJ*.

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F. (1981). Sequence and organization of the human mitochondrial genome. *Nature*., 290, 457.

Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M., Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet,* 23**,** 147.

Aneli, S., Saupe, T., Montinaro, F., Solnik, A., Molinaro, L., Scaggion, C., ... & Pagani, L. (2022). The genetic origin of Daunians and the Pan-Mediterranean southern Italian Iron Age context. *Molecular Biology and Evolution, 39*(2), msac014.

Avise, J. C. (2000). Phylogeography: the history and formation of species. *Harvard university press*.

Botstein, D., White, R. L., Skolnick, M., & Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American journal of human genetics, 32*(3), 314.

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., ... & Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS computational biology*, 15(4), e1006650.

Bramanti, B., Thomas, M. G., Haak, W., Unterlaender, M., Jores, P., Tambets, K., Antanaitis-Jacobs, I., Haidle, M. N., Jankauskas, R., Kind, C.-J., Lueth, F., Terberger, T., Hiller, J., Matsumura, S., Forster, P., & Burger, J. (2009). Genetic Discontinuity Between Local Hunter-Gatherers and Central Europe's First Farmers. *Science*, *326*(5949), 137–140.

Brandstätter, A., Peterson, C. T., Irwin, J. A., Mpoke, S., Koech, D. K., Parson, W., Parsons, T. J. (2004). Mitochondrial DNA control region sequences from Nairobi (Kenya): inferring phylogenetic parameters for the establishment of a forensic database. *International Journal of Legal Medicine*, 118, 294–306.

Brown, W. M., George, M., Wilson, A. C. (1979). Rapid evolution of animal mitochondrial DNA., *Proc. Natl. Acad. Sci. U.S.A.*76 (4) 1967-1971.

Cavalli-Sforza, L. L., & Feldman, M. W. (2003). The application of molecular genetic approaches to the study of human evolution. *Nature Genetics*, 33(S3), 266–275.

Cerezo, M., Achilli, A., Olivieri, A., Perego, U. A., Gomez-Carballa, A., Brisighelli, F., ... & Salas, A. (2012). Reconstructing ancient mitochondrial DNA links between Africa and Europe. *Genome research, 22*(5), 821-826.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets, *GigaScience*, s13742–015–0047–8.

Cohen, S. N., Chang, A. C., Boyer, H. W., & Helling, R. B. (1973). Construction of biologically functional bacterial plasmids in vitro. *Proceedings of the National Academy of Sciences, 70*(11), 3240-3244.

Dür, A., Huber, N., Parson, W. (2021). Fine-Tuning Phylogenetic Alignment and Haplogrouping of mtDNA Sequences. *International Journal of Molecular Sciences*. 22(11):5747.

Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, 131(2), 479-491.

Fu, Q., Mittnik, A., Johnson, P. L. F., Bos, K., Lari, M., Bollongino, R., Sun, C., Giemsch, L., Schmitz, R., Burger, J., Ronchitelli, A. M., Martini, F., Cremonesi, R. G., Svoboda, J., Bauer, P., Caramelli, D., Castellano, S., Reich, D., Pääbo, S., & Krause, J. (2013). A Revised Timescale for Human Evolution Based on Ancient Mitochondrial Genomes. *Current Biology*, *23*(7), 553–559.

Garrido, N., Griparic, L., Jokitalo, E., Wartiovaara, J., Van Der Bliek, A. M., & Spelbrink, J. N. (2003). Composition and Dynamics of Human Mitochondrial Nucleoids. *Molecular Biology of the Cell*, 14(4), 1583–1596.

Giles, R. E., Blanc, H., Cann, H. M., Wallace, D. C. (1980). Maternal inheritance of human mitochondrial DNA. *Proc. Natl. Acad. Sci. U.S.A.* 77: 6715-6719.

Goudet, J. (2005). Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular ecology notes, 5*(1), 184-186.

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H.-Y., Hansen, N. F., Durand, E. Y., Malaspinas, A.-S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H. A., … & Pääbo, S. (2010). A Draft Sequence of the Neandertal Genome. *Science, 328*(5979), 710–722.

Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., ... & Reich, D. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature, 522*(7555), 207-211.

Hirszfeld, L., & Hirszfeld, H. (1919). Essai d'application des méthodes sérologiques au problème des races. *L'Anthropologie, 29*, 505-537.

Iborra, F. J., Kimura, H., Cook, P. R. (2004). The functional organization of mitochondrial genomes in human cells. *BMC Biol 2*, 9.

Ingman, M., Kaessmann, H., Pääbo, S., & Gyllensten, U. (2000). Mitochondrial genome variation and the origin of modern humans. *Nature, 408*(6813), 708-713.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics, 24*(11), 1403-1405.

Kivisild, T., Reidla, M., Metspalu, E., Rosa, A., Brehm, A., Pennarun, E., ... & Villems, R. (2004). Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. *The American Journal of Human Genetics, 75*(5), 752-770.

Kubicek, P. (2008). *The history of Ukraine*. ABC-CLIO.

Kushniarevich, A., Utevska, O., Chuhryaeva, M., Agdzhoyan, A., Dibirova, K., Uktveryte, I., Möls, M., Mulahasanovic, L., Pshenichnov, A., Frolova, S., Shanko, A., Metspalu, E., Reidla, M., Tambets, K., Tamm, E., Koshel, S., Zaporozhchenko, V., Atramentova, L., Kučinskas, V., … Balanovsky, O. (2015). Genetic Heritage of the Balto-Slavic Speaking Populations: A Synthesis of Autosomal, Mitochondrial and Y-Chromosomal Data. *PLOS ONE*, 10(9), e0135820.

Landsteiner, K. (1901). Ueber Agglutinationser-scheinungen normalen menschlichen Blutes. *Wien Clin Wschr, 14*, 1132-1134.

Lazaridis, I., (2018). The evolutionary history of human populations in Europe. *Current Opinion in Genetics & Development*, *53*, 21-27.

Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: an R package for multivariate analysis. *Journal of statistical software, 25*, 1-18.

Malmström, H., Gilbert, M. T. P., Thomas, M. G., Brandström, M., Storå, J., Molnar, P., ... & Willerslev, E. (2009). Ancient DNA reveals lack of continuity between neolithic hunter-gatherers and contemporary Scandinavians. *Current biology, 19*(20), 1758-1762.

Malyarchuk, B.A., Derenko, M.V. (2023). Mitochondrial gene pool of Ukrainians in the context of variability of whole mitogenomes in Slavic peoples. *Genetika*. 59(1):106-114.

Mathieson, I., Alpaslan-Roodenberg, S., Posth, C., Szécsényi-Nagy, A., Rohland, N., Mallick, S., ... & Reich, D. (2018). The genomic history of southeastern Europe. *Nature, 555*(7695), 197-203.

McBride, H. M., Neuspiel, M., & Wasiak, S. (2006). Mitochondria: More Than Just a Powerhouse. *Current Biology, 16*(14), R551–R560.

Metzker, M. L. (2010). Sequencing technologies—the next generation. *Nature reviews genetics, 11*(1), 31-46.

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.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., & Erlich, H. (1986). Specific Enzymatic Amplification of DNA In Vitro: The Polymerase Chain Reaction. *Cold Spring Harbor Symposia on Quantitative Biology, 51*(0), 263–273.

Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, 76(10), 5269-5273.

Nei, M. (1987). *Molecular Evolutionary Genetics*. Columbia University Press.

Nikitin, A. G., Kochkin, I. T., June, C. M., Willis, C. M., Mcbain, I., & Videiko, M. Y. (2009). Mitochondrial DNA Sequence Variation in the Boyko, Hutsul, and Lemko Populations of the Carpathian Highlands. *Human Biology*, 81(1), 43–58.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R. B., ... & Oksanen, M. J. (2013). Package 'vegan'. *Community ecology package, version, 2*(9), 1-295.

Olalde, I., Brace, S., Allentoft, M. E., Armit, I., Kristiansen, K., Booth, T., ... & Reich, D. (2018). The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature, 555*(7695), 190-196.

Olivieri, A., Achilli, A., Pala, M., Battaglia, V., Fornarino, S., Al-Zahery, N., ... & Torroni, A. (2006). The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa. *Science, 314*(5806), 1767-1770.

Paradis, E. (2010). pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics*, 26(3), 419-420.

Pauling, L., Itano, H. A., Singer, S. J., & Wells, I. C. (1949). Sickle cell anemia, a molecular disease. *Science, 110*(2865), 543-548.

Posth, C., Yu, H., Ghalichi, A., Rougier, H., Crevecoeur, I., Huang, Y., ... & Krause, J. (2023). Palaeogenomics of upper palaeolithic to neolithic European hunter-gatherers. *Nature, 615*(7950), 117-126.

Pshenichnov, A., Balanovsky, O., Utevska, O., Metspalu, E., Zaporozhchenko, V., Agdzhoyan, A., Churnosov, M., Atramentova, L., & Balanovska, E. (2013). Genetic affinities of Ukrainians from the maternal perspective. *American Journal of Physical Anthropology*, *152*(4), 543–550.

R Core Team (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Raveane, A., Aneli, S., Montinaro, F., Athanasiadis, G., Barlera, S., Birolo, G., ... & Capelli, C. (2019). Population structure of modern-day Italians reveals patterns of ancient and archaic ancestries in Southern Europe. *Science Advances, 5*(9), eaaw3492.

Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., Viola, B., Briggs, A. W., Stenzel, U., Johnson, P. L. F., Maricic, T., Good, J. M., Marques-Bonet, T., Alkan, C., Fu, Q., Mallick, S., Li, H., Meyer, M., Eichler, E. E., … & Pääbo, S. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature, 468*(7327), 1053–1060.

Richards, M., Macaulay, V., Hickey, E., Vega, E., Sykes, B., Guida, V., ... & Bandelt, H. J. (2000). Tracing European founder lineages in the Near Eastern mtDNA pool. *The American Journal of Human Genetics, 67*(5), 1251-1276.

Richards, M. B., Soares, P., & Torroni, A. (2016). Palaeogenomics: mitogenomes and migrations in Europe's past. *Current biology, 26*(6), R243-R246.

Sanger, F., & Coulson, A. R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of molecular biology, 94*(3), 441-448.

Semino, O., Passarino, G., Oefner, P. J., Lin, A. A., Arbuzova, S., Beckman, L. E., ... & Underhill, P. A. (2000). The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: AY chromosome perspective. *Science, 290*(5494), 1155-1159.

Schönherr, S., Weissensteiner, H., Kronenberg, F., Forer, L. (2023). Haplogrep 3 – an interactive haplogroup classification and analysis platform. *Nucleic Acids Res*., *51*, W263–W268.

Soares, P., Achilli, A., Semino, O., Davies, W., Macaulay, V., Bandelt, H.-J., Torroni, A., & Richards, M. B. (2010). The Archaeogenetics of Europe. *Current Biology*, *20*(4), R174–R183.

Soares, P., Alshamali, F., Pereira, J. B., Fernandes, V., Silva, N. M., Afonso, C., ... & Pereira, L. (2012). The expansion of mtDNA haplogroup L3 within and out of Africa. *Molecular biology and evolution, 29*(3), 915-927.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of molecular biology, 98*(3), 503-517.

State Statistics Committee of Ukraine (2001). http://2001.UKRcensus.gov.ua/results/general/nationality/donetsk/

Stoltenberg, J. (2022, September 27). Ukraine: US, NATO condemn Kremlin-staged referendums as voting ends. Deutsche Welle. https://www.dw.com/en/ukraine-us-nato-condemn-kremlin-staged-referendums-as-voting-ends/a-63250664

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), 585-595.

Torroni, A., Huoponen, K., Francalacci, P., Petrozzi, M., Morelli, L., Scozzari, R., ... & Wallace, D. C. (1996). Classification of European mtDNAs from an analysis of three European populations. *Genetics, 144*(4), 1835-1850.

Torroni, A., Lott, M. T., Cabell, M. F., Chen, Y. S., Lavergne, L., & Wallace, D. C. (1994a). 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. *American journal of human genetics, 55*(4), 760.

Torroni, A., Neel, J. V., Barrantes, R., Schurr, T. G., & Wallace, D. C. (1994b). Mitochondrial DNA" clock" for the Amerinds and its implications for timing their entry into North America. *Proceedings of the National Academy of Sciences, 91*(3), 1158-1162.

Torroni, A., Sukernik, R. I., Schurr, T. G., Starikorskaya, Y. B., Cabell, M. F., Crawford, M. H., ... & Wallace, D. C. (1993a). mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. *American journal of human genetics, 53*(3), 591.

Torroni, A., Schurr, T. G., Cabell, M. F., Brown, M. D., Neel, J. V., Larsen, M., ... & Wallace, D. C. (1993b). Asian affinities and continental radiation of the four founding Native American mtDNAs. *American journal of human genetics, 53*(3), 563.

Twig, G., Elorza, A., Molina, A. J. A., Mohamed, H., Wikstrom, J. D., Walzer, G., Stiles, L., Haigh, S. E., Katz, S., Las, G., Alroy, J., Wu, M., Py, B. F., Yuan, J., Deeney, J. T., Corkey, B. E., & Shirihai, O. S. (2008). Fission and selective fusion govern mitochondrial segregation and elimination by autophagy. *The EMBO Journal*, *27*(2), 433–446.

UN General Assembly. (2022). Resolution ES-11/4: Territorial integrity of Ukraine: Defending the principles of the Charter of the United Nations. United Nations Digital Library. https://digitallibrary.un.org/record/3986729

United Nations. (2022). General Assembly adopts resolution declaring invalid Ukraine referendums. https://news.un.org/en/story/2022/10/1129492

van Oven, M., & Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat*., *30*: E386-E394.

Watson, J. D., & Crick, F. H. C. (1953). Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature, 171*(4356), 737-738.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*(6), 1358-1370.

Weissensteiner, H., Forer, L., Kronenberg, F., & Schönherr, S. (2024). mtDNA-Server 2: advancing mitochondrial DNA analysis through highly parallelized data processing and interactive analytics. *Nucleic Acids Research, 52*(W1), W102-W107.

Wickham, H. (2016). ggplot2: elegant graphics for data analysis. Springer-Verlag New York.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, *16*(2), 97.

Wright, S. (1951). The genetical structure of populations. *Annals of Eugenics*, *15*(1), 323-354.

Yan, C., Duanmu, X., Zeng, L., Liu, B., Song, Z. (2019). Mitochondrial DNA: Distribution, Mutations, and Elimination. *Cells*. *8*, 379.

Youle, R. J., & Van Der Bliek, A. M. (2012). Mitochondrial Fission, Fusion, and Stress. *Science*, *337*(6098), 1062–1065.

9. WEBSITES

http://2001.ukrcensus.gov.ua/results/general/nationality/

https://www.agilent.com/en/product/automated-electrophoresis/tapestation-systems/tapestation-instruments/4150-tapestation-system-297322

https://www.britannica.com/place/Donetsk-Ukraine

https://emea.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet_nextera_xt_dna_sample_prep.pdf

https://emea.illumina.com/science/technology/next-generation-sequencing/sequencing-technology/2-channel-sbs.html

https://github.com/NicRamb/mtDNA_pipe

https://www.thermofisher.com/it/en/home/industrial/spectroscopy-elemental-isotope-analysis/molecular-spectroscopy/fluorometers/qubit/models.html

https://www.thermofisher.com/it/en/home/industrial/spectroscopy-elemental-isotope-analysis/molecular-spectroscopy/fluorometers/qubit/qubit-fluorometer.html