Report Population Genetics Project

Ancient Connections: An Interactive Tool for Tracing Maternal Lineages through Ancient DNA

Tabea Dittmar[1]

[1]Department of Biology, Box 118, 221 00, Lund University, Sweden

Abstract

Ancient Connections is an interactive tool for exploring shared maternal ancestry between modern individuals and ancient DNA samples based on mitochondrial haplogroups. By integrating curated genetic and archaeological data, the app identifies common maternal ancestors and visualizes their relationships on a temporal timeline. This approach makes genetic history tangible and accessible, offering user-friendly ways to investigate human population history and ancestral ties across time.

Availability: The full source code, dataset, and usage instructions are available at: https://github.com/tbadtmr/AncientConnections_Project
Contact: tabeadittmar@gmail.com

1   Introduction

Understanding human population history has greatly benefited from genetic research, particularly through the analysis of uniparentally inherited markers such as mitochondrial DNA (mtDNA) and the Y-chromosome. These markers enable the reconstruction of direct maternal and paternal lineages over thousands of years. Mitochondrial DNA, passed down exclusively from mothers to their offspring, is especially informative due to its lack of recombination and relatively consistent mutation rate (van Oven & Kayser, 2009; Behar et al., 2012).

Over time, small mutations accumulate in the mitochondrial genome. These mutations form distinct sets of related sequences known as haplogroups, which represent populations sharing similar mitochondrial DNA polymorphisms. Each haplogroup corresponds to a specific branch on the maternal lineage tree and reflects descent from a common maternal ancestor (Kivisild, 2015). Haplogroups are often associated with particular geographic regions and periods in human prehistory, providing important clues about migration patterns, population expansions, and demographic history (Kivisild, 2015; Miyamoto-Mikami & Fuku, 2019).

Recent advances in ancient DNA sequencing have made it possible to analyse genetic material from archaeological remains, offering new insights into the genetic structure of past populations. Resources like the Allen Ancient DNA Resource (AADR), curated by the Reich Lab at Harvard University, now contain thousands of annotated ancient individuals with mitochondrial haplogroups and associated geographic and chronological metadata (Günther & Jakobsson, 2016; Zhao et al., 2009).

Despite the availability of such datasets, few accessible tools allow researchers or the general public to explore shared ancestry between modern and ancient individuals in an intuitive way. Most existing tools emphasize phylogenetic relationships in static trees or through summary statistics, without offering interactivity or visualizing ancient individuals within their historical context.

Therefore, this project introduces *Ancient Connections*, an interactive application for exploring maternal ancestry through mtDNA haplogroups. Users can select two individuals and determine their most recent shared maternal haplogroup. The app searches the AADR dataset for ancient individuals carrying matching or ancestral haplogroups and displays these individuals on a timeline.

By linking genetic and archaeological data in an interactive format, *Ancient Connections* provides a novel and accessible tool for investigating maternal genetic heritage and highlighting where and when shared ancestry may have occurred.

2   Methods

The *Ancient Connections* tool was developed using Python and the Streamlit framework to create an interactive web application. The goal of the tool is to identify and visualize shared maternal ancestry between two selected individuals based on mtDNA haplogroups. It uses Streamlit for the interface, Pandas and NumPy for data handling, and Plotly for visualizing timelines and lineages. Usage instructions are available at: https://github.com/tbadtmr/AncientConnections_Project.

Data Source

The primary dataset consists of curated ancient mtDNA samples obtained from the Allen Ancient DNA Resource (AADR, 2024). These include metadata such as haplogroup, sample identifier, chronological dating, geographic origin, and archaeological context. A phylogenetic tree of mtDNA haplogroups was used to reconstruct ancestral lineages. A set of five modern test users is included for demonstration purposes.

The ancient data was cleaned and manually reviewed to ensure consistency across haplogroup labels and chronological fields. All cleaned data are included in the project's `Data/` folder.

User Input

Users can select two individuals from three sources:

1. Predefined test users
2. Manual input of mtDNA haplogroup and year
3. Ancient individuals from the AADR database, filtered by continent, country, and archaeological period

Ancestral Inference

Once two users are selected, the tool identifies their most recent shared haplogroup using the phylogenetic tree. This is done by tracing each haplogroup's lineage to the root and finding their most recent common node. The tool then searches the ancient database for samples that match either this shared haplogroup or any ancestral haplogroups in the lineage. To ensure temporal validity, only ancient samples that are older than both selected individuals are considered.

Timeline Visualization

All identified shared ancestors are plotted on a horizontal timeline. Each sample is represented as a point, with hoverable details including haplogroup, date range, country, and region. The two user samples are shown at the bottom of the timeline. Ancestors are ordered both by lineage depth and age, allowing a clear top-down structure. Lineage relationships are visualized using smooth Bezier curves. Nodes are color-coded to distinguish users, shared haplogroup matches, and other related ancestors. If no ancient individual in the database matches the exact shared haplogroup, the app introduces a placeholder node to mark the inferred most recent common maternal haplogroup and connects it to the next available ancestor further up the lineage.

3 Features

The application generates three layers of output to visualize maternal lineage connections between two individuals. Before any visualization, a brief description is generated that summarizes information about the most recent shared ancient maternal ancestor found in the dataset. The core visualization is the interactive timeline which displays both users and their shared ancestors in chronological order. Following the plot, the app displays the full maternal lineage for both users, from the most recent shared to each terminal haplogroup. This helps to interpret how far back the connection lies and shows where the lineages diverge within the mtDNA phylogeny.

3.1 Timescale Example Use Cases

To illustrate the tool's application, the following comparisons were conducted using test users and curated ancient samples:

### 3.1.1 Case 1: Test User 1 vs Test User 2

Test User 1 (haplogroup: X2c2) and Test User 2 (K1c1b) represent two modern individuals with distinct maternal lineages. The application identified their most recent common ancestral haplogroup as N and retrieved one ancient individual from the database, PCA0041, who also belongs to this lineage. PCA0041 lived in Central Europe (Poland) during the Middle Ages, between 80 CE and 260 CE, and is shown as the closest shared ancient maternal ancestor in the timeline.
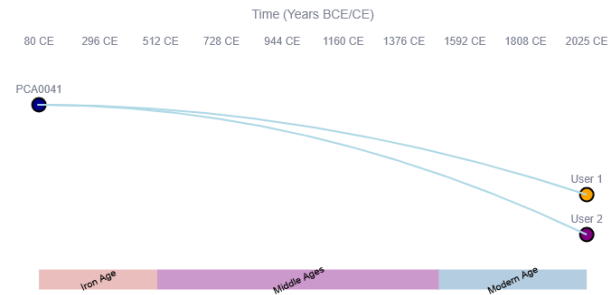
Figure 1. Timeline visualization for Test User 1 and Test User 2. The users are connected to the shared ancestor (PCA0041). Epoch bars provide historical context.

### 3.1.2 Case 2: No Ancient Match (Manual Input)

In this scenario, two mitochondrial haplogroups were entered manually: D2 for User 1 and X1'2'3 for User 2, with estimated dates of 2025 CE and 1000 CE. These haplogroups represent geographically distant maternal lineages, with D2 being associated with northeastern Asia and the Americas, and X1'2'3 found primarily in Eurasia. While both haplogroups descend from the broader L3 branch of the mitochondrial phylogeny, no ancient matching or ancestral haplogroups were found in the database. As a result, the application returned a message informing the user that no match was found. The shared haplogroup is still displayed and indicates that there is a point in the lineage where the individuals share an ancestor. This test case demonstrates how the tool handles cases where no ancient match exists while still conveying shared maternal ancestry and highlighting database coverage limitations.
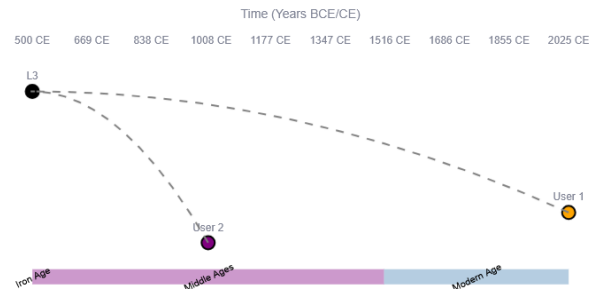
Figure 2. Timeline visualization for two manually entered users (haplogroups D2 and X1'2'3). No ancient ancestor was found in the database, but the shared haplogroup (L3) is shown.

3.1.3 Case 3: Test User and Ancient Database Sample

Test User 2 (Country: Unknown, Time: Present) and I4208 (Country: Cyprus, Time: 3500 BCE) share the common maternal ancestor: I2426. This individual likely lived in the Balkans, Bulgaria, during the Copper Age, between 4450 BCE and 4264 BCE. The shared haplogroup is K1, and the app identifies a series of ancient individuals in the AADR database who belong to this haplogroup or its ancestral branches. Specifically, five individuals with the same haplogroup (K1) are highlighted as dark blue nodes. These samples come from Bulgaria, Germany, and Romania, and span the Neolithic and Copper Age periods. Further back in the lineage, two older samples from Georgia and Italy represent deeper ancestral branches of the K1 lineage. This illustrates how the tool not only identifies a direct shared ancestor but also visualizes the maternal lineage over thousands of years across different regions.
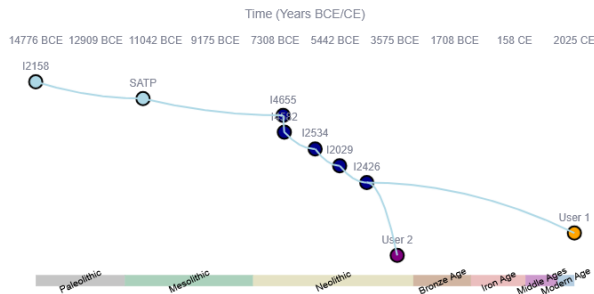
Figure 3. Timeline visualization for Test User 2 and ancient sample I4208. The two individuals are linked through a shared maternal ancestor (I2426), and the lineage is traced through several older ancient samples.

3.2 Lineage Validation

To illustrate the lineage tracing logic, the application's output for two users with different haplogroups is shown below. Both maternal lines are traced back to their most recent common haplogroup and display the full ancestral path for each individual.

This output demonstrates that the app preserves the branching order of mitochondrial haplogroups and correctly identifies the last shared maternal ancestor.

User 1 Haplogroup: X2c2

N → X → X1'2'3 → X2 → X2_G225A → X2c → X2c2

User 2 Haplogroup: K1c1b

N → R → U → U2'3'4'7'8'9 → U8 → U8b'c → U8b → K → K1 → K1c → K1c1 → K1c1b

Shared Haplogroup: N

mt-MRCA → L1'2'3'4'5'6 → L2'3'4'5'6 → L2'3'4'6 → L3'4'6 → L3'4 → L3 → N

Figure 4. Inferred maternal lineages for two users. Both lineages are traced from terminal haplogroups (X2c2 and K1c1b) back to the shared haplogroup **N**, illustrating the correct reconstruction of ancestral relationships.

4 Discussion

The *Ancient Connections app* provides an intuitive interface to explore shared maternal ancestry between modern individuals and ancient DNA samples. By integrating mtDNA haplogroups, phylogenetic relationships, and archaeological data into a unified visualization, the tool offers a simple and educational way to trace genetic lineages over time and space. It helps users contextualize ancestry not just in terms of genetic similarity, but also historical periods and geographic origins and allows for direct exploration of maternal relationships.

Since the goal of this tool is to provide an easy-to-use and intuitive application, a web application was a fitting choice because it allows users to access it directly through their browser without installation. However, it relies on a stable internet connection and can be less reliable or flexible when built with pre-made frameworks like Streamlit. Additionally, web applications may face performance limitations with large datasets, offer limited UI customization, and raise security concerns depending on how and where they are deployed.

Other suggestions for improvements in future versions are based on dataset-specific limitations. For example, since the database consists mainly of ancient individuals, comparing two ancient samples often leads to situations where no older shared ancestor is present in the data. In such cases, the tool inserts a synthetic marker to indicate the inferred common haplogroup. While this maintains continuity in the lineage, incorporating

more distant or side-branch relatives could enhance the interpretability of such connections and better reflect the complexity of human genetic history.

Another potential extension would be the inclusion of Y-chromosome haplogroup analysis, allowing users to explore both maternal and paternal lineages in parallel. This could provide a more comprehensive view of individual ancestry and broader population dynamics.

This project demonstrates the potential of interactive tools to make complex genetic relationships intuitive and accessible, bridging the gap between population genetics, archaeology, and public engagement.

References

Behar, Doron & van Oven, Mannis & Rosset, Saharon & Metspalu, Mait & Loogväli, Eva-Liis & Silva, Nuno & Kivisild, Toomas & Torroni, Antonio & Villems, Richard. (2012). A "Copernican" Reassessment of the Human Mitochondrial DNA Tree from Its Root (vol 90, pg 675, 2012). American journal of human genetics. 90. 675-84. 10.1016/j.ajhg.2012.03.002.

Eri Miyamoto-Mikami, Noriyuki Fuku, Chapter Six - Variation of Mitochondrial DNA and elite athletic performance, Editor(s): Debmalya Barh, Ildus I. Ahmetov, Sports, Exercise, and Nutritional Genomics, Academic Press, 2019, Pages 129-145, ISBN 9780128161937, https://doi.org/10.1016/B978-0-12-816193-7.00006-3. Retrieved from: https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/mitochondrial-haplogroup

Günther, T., & Jakobsson, M. (2016). Genes mirror migrations and cultures in prehistoric Europe-a population genomic perspective. *Current opinion in genetics & development*, *41*, 115–123. https://doi.org/10.1016/j.gde.2016.09.004

Kivisild T. (2015). Maternal ancestry and population history from whole mitochondrial genomes. *Investigative genetics*, *6*, 3. https://doi.org/10.1186/s13323-015-0022-2

Merheb M, Matar R, Hodeify R, Siddiqui SS, Vazhappilly CG, Marton J, Azharuddin S, Al Zouabi H. Mitochondrial DNA, a Powerful Tool to Decipher Ancient Human Civilization from Domestication to Music, and to Uncover Historical Murder Cases. Cells. 2019 May 9;8(5):433. doi: 10.3390/cells8050433. PMID: 31075917; PMCID: PMC6562384.

Nesheva D. Aspects of ancient mitochondrial DNA analysis in different populations for understanding human evolution. Balkan J Med Genet. 2014 Dec 11;17(1):5-14. doi: 10.2478/bjmg-2014-0019. PMID: 25741209; PMCID: PMC4347471.

Reich Laboratory. (2024). *Allen Ancient DNA Resource (AADR), Version 54.1.* https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data

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

van Oven, M. (2015). PhyloTree Build 17: Growing the human mitochondrial DNA tree. *Forensic Science International: Genetics Supplement Series*, 5, e392–e394. https://doi.org/10.1016/j.fsigss.2015.09.155

Zhao, M., Kong, Q. P., Wang, H. W., Peng, M. S., Xie, X. D., Wang, W. Z., Jiayang, Duan, J. G., Cai, M. C., Zhao, S. N., Cidanpingcuo, Tu, Y. Q., Wu, S. F., Yao, Y. G., Bandelt, H. J., & Zhang, Y. P. (2009). Mitochondrial genome evidence reveals successful Late Paleolithic settlement on the Tibetan Plateau. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(50), 21230–21235. https://doi.org/10.1073/pnas.0907844106