*Bioinformatics*, 33(18), 2024, 1366–1368

doi: 10.1093/bioinformatics/btx315

Advance Access Publication Date: 12 March 2024

Application Note

|  |
| --- |
| Genome analysis  **ANCIENTLINK\_Y: Tracing Y Chromosome Ancestry**  Tan Hooi Min Grahn1,\*  1Department of Biology, Box 118, 221 00, Lund University, Sweden.  \*To whom correspondence should be addressed.  Associate Editor: Inanc Birol  Received on September 30, 2023; revised on January 15, 2024; accepted on March 1, 2024  Abstract  **Motivation:** The Y chromosome serves as a genetic thread connecting individuals across generations, offering insights into ancient migrations and common ancestry. By addressing a research niche, a user-friendly app is designed to explore the ancestral connections between individuals through their Y chromosome haplogroups.  **Results:** ANCIENTLINK\_Y allows users to type in two haplogroups, then lists out their most common ancestral haplogroup and identifies their most ancient common ancestors within the AADR (Ancient Ancestral DNA Repository) database. The timeline visualization plots these ancestral connections, illustrating the genetic journeys and shared heritage between individuals.  **Availability:** git@github.com:hooimin7/PopGenApp.git  **Contact:** ho4588ta-s@student.lu.se |

# Introduction

Understanding human ancestry has long been a fascination, driven by the desire to uncover the intricate web of connections that tie us together. The Y chromosome, contains the largest nonrecombining block in the human genome passed down through male lineage, provides a unique avenue for exploring ancestral links. By examining specific markers on the Y chromosome, known as haplogroups, researchers can trace migration patterns, genetic diversity, and common ancestry (Consortium, 2002). The integration of Y-DNA data with radiocarbon dating offers a valuable method to construct phylogenetic trees of Y chromosomes. This approach enables the assessment of ancestral variation based on both modern and ancient DNA samples. It also allows for the identification of ancient clades that may be rare or have become extinct over time (Esposito *et al.*, 2018; Kivisild, 2017). There are several tools integrate ancient DNA data with modern haplogroup information, such as : pathPhynder (Martiniano *et al.*, 2022), AncientDNA.info (Chen and Nedoluzhko, 2023), Allen Ancient DNA Resource (AADR) (Mallick *et al.*, 2023), aYChr-DB (Freeman *et al.*, 2020). These resources contribute significantly to our understanding of ancient DNA and its implications for human evolution and history. There is a website named FamilyTreeDNA, <https://www.familytreedna.com/> and users can purchase the services by sending in their cheek swab. However, ANCIENTLINK\_Y is designed to provide nonprofitable services and acts as a user-friendly interface to allow users with least computational knowledge to explore their heritage on their paternal line. ANCIENTLINK\_Y enables users to explore closer connections in our Y-DNA matching database.

# Methods

In folder 01\_PopGenApp, a raw data chrY\_hGrpTree\_isogg2016.txt (International Society of Genetic Genealogy. Y-DNA Haplogroup Tree 2019-2020, Version: 15.73, 2020) was used to build a tree and an output file (outfile.txt) was created, and the paths were written from each leaf to the root. Another raw data AADR Annotation.xlsx (Mallick *et al.*, 2023) was read into .tsv format using pandas (v 2.1.4) package, Date mean in Before Present (BP) in years before 1950 common era (CE) (herein date mean) and Y haplogroup (manual curation in ISOGG format) (herein Y haplogroup) were extracted and kept in a new file named AADRsubset.tsv. A python script used pandas read a tsv file (AADRsubset.tsv) into a data frame. It initialized an empty dictionary (date means) to store mean dates for haplogroups. The script read outfile.txt line by line, splitting each line into haplogroups. For each haplogroup, it checked if it was in the data frame, fetching its mean date if present. The script then added a root haplogroup to date means place holder with a date mean of 200000. Date means was converted to a data frame (date\_means\_df) and written to HaplogroupBPE.tsv. RootHaplogroupBPE.tsv was created from HaplogroupBPE.tsv where the value in data\_mean\_df column were subtracted from 1950. This was a way to convert the dates from years before 1950 CE to years after 1950 CE. While in folder 03\_CheckCE, AADRsubset.tsv was used to filter either CE or before common era (BCE) for selecting the oldest date mean for each haplogroup. The cleaned, sorted data frame with calculated data mean values was then written to a new .tsv file named ‘FinalMergedHaplo.tsv’ (Fig. 1). Before realizing an interactive app, a test model (in 02\_Plot folder) was ran by accepting two users’ inputs, the RootHaplogroupBPE.tsv dataset and the outfile.txt to produce a network graph.

A diagram of a diagram

Description automatically generated

Fig. 1. Folder organization and files filtering

A screenshot of a computer

Description automatically generated

Fig 2. Introductory text about the ANCIENTLINK\_Y before users starts using the application

A screenshot of a computer

Description automatically generated

Fig. 3. Workflow of ANICIENTLINK\_Y (Graphical User Interface, GUI), runs in conda ‘biopython’ environment

# Results

After verifying the successful run-down of the test model, I developed an interactive app to facilitate the exploration of Y chromosome ancestry (Fig. 2). The app allowed users to select two individuals or samples for analysis. Using genetic data from the AADR database, the app demonstrated the most common ancestral haplogroup shared by the selected individuals. Furthermore, it identified the most ancient common ancestors within the database. ANCIENTLINK\_Y was written in python (v 3.12.2) and could be executed from the terminal using bash and a website was created instantly. Introductory text about the ANCIENTLINK\_Y was published and defined (timeline.xlsx as input file) (Fig. 2). Next, the required inputs were the haplogroup names that could be found in AADR database (Mallick *et al.*, 2023). When the haplogroup names matched either the RootHaplogroupBPE.tsv or the FinalMergedHaplo.tsv (subset of AADR.tsv) files, then these haplogroups were added as nodes to build a network graph. Instantly a file consisted of lineage haplogroup of ISOGG Y-chromosom (named outfile.txt) was called and edges were added to the similar graph according to the outfile. A subgraph was then built based on the user inputs. The final graph was visualized on the website. Control c was typed to quit the app (Fig. 3, 4). Upon selecting two individuals, the app returns their most common ancestral haplogroup, providing insights into their shared paternal lineage. Additionally, it plotted the most ancient common ancestors on a timeline, showcasing the temporal aspect of genetic connections (Fig. 4). A copy of the users’ graph was kept as a html format in the administrator directory. Upon request, users could obtain the graph in html format with build-in special features from ANCIENTLINK\_Y app.

A screenshot of a computer screen

Description automatically generated

Fig 4. Common and the most ancient ancestors are plotted along shared paternal lineage

# Discussion

The ANCIENTLINK\_Y app offers a seamless and intuitive platform designed to delve into the intricate web of ancient connections encoded within the Y chromosome. By leveraging cutting-edge visualization tools, users are empowered to unravel their ancestral haplogroups and trace their lineage back to common progenitors, thus deepening their comprehension of genetic heritage and interconnections with others. Through the simple tap of a finger, individuals can effortlessly navigate through their paternal lineages, charting a course to both recent ancestors and the most ancient forebears.

Currently, the app is integrated with the extensive dataset sourced from the AADR database (Mallick et al., 2023), which chronicles the development of individual haplogroups over thousands of years. An illustrative example of this historical continuum is haplogroup F, which was first recorded in 9161 BCE and later reappears in historical records around 1000 CE, marking significant junctures in human migration and evolutionary history. Consequently, the app offers developers the flexibility to choose between two distinct data inputs, namely RootHaplogroupBPE.tsv or FinalMergedHaplo.tsv (a subset of AADR.tsv), for integration into the graphical user interface (GUI).

Looking ahead, the future trajectory of the ANCIENTLINK\_Y app is poised towards greater user agency and customization. Rather than being confined to a predetermined dataset, users will be empowered to select the specific data they wish to explore, fostering a more personalized and dynamic journey of discovery. By placing the reins of exploration firmly in the hands of the user, the app aims to cultivate a more interactive and enriching experience, enabling individuals to chart their genetic heritage according to their unique interests and curiosities.

# Conclusion

Ancient Connections offers a novel approach to exploring Y chromosome ancestry, allowing users to uncover shared genetic links and ancestral journeys. By leveraging genetic data and visualization techniques, the app facilitates the exploration of human history and the bonds that unite us across time and space.

Acknowledgements

We are grateful to the editor and two anonymous reviewers for their valuable comments on the manuscript, Dr. Eran Elhaik for valuable comment on application and for Lund university for their financial support.

Funding

This work has been supported by the Lund university.

*Conflict of Interest:* none declared.

References

Chen,N. and Nedoluzhko,A. (2023) Ancient DNA: the past for the future. *BMC Genomics*, **24**, 309.

Consortium,T.Y.C. (2002) A Nomenclature System for the Tree of Human Y-Chromosomal Binary Haplogroups. *Genome Res*, **12**, 339–348.

Esposito,U. *et al.* (2018) Ancient Ancestry Informative Markers for Identifying Fine-Scale Ancient Population Structure in Eurasians. *Genes*, **9**, 625.

Freeman,L. *et al.* (2020) aYChr-DB: a database of ancient human Y haplogroups. *NAR Genomics and Bioinformatics*, **2**, lqaa081.

International Society of Genetic Genealogy. Y-DNA Haplogroup Tree 2019-2020, Version: 15.73 (2020).

Kivisild,T. (2017) The study of human Y chromosome variation through ancient DNA. *Hum Genet*, **136**, 529–546.

Mallick,S. *et al.* (2023) The Allen Ancient DNA Resource (AADR): A curated compendium of ancient human genomes. 2023.04.06.535797.

Martiniano,R. *et al.* (2022) Placing Ancient DNA Sequences into Reference Phylogenies. *Mol Biol Evol*, **39**, msac017.