Published online 9 October 2020

NAR Genomics and Bioinformatics, 2020, Vol. 2, No. 4 **1**
doi: 10.1093/nargab/lqaa081

aYChr-DB: a database of ancient human Y haplogroups

Laurence Freeman[1], Conrad Stephen Brimacombe[1,2] and Eran Elhaik[1,3,*]

[1]University of Sheffield, Department of Animal and Plant Sciences, Sheffield S10 2TN, UK, [2]University of Bristol, Department of Archaeology and Anthropology, Bristol BS8 1TH, UK and [3]Lund University, Department of Biology, Lund 223 62, Sweden

Received May 15, 2020; Revised August 21, 2020; Editorial Decision September 10, 2020; Accepted September 16, 2020

ABSTRACT

Ancient Y-Chromosomal DNA is an invaluable tool for dating and discerning the origins of migration routes and demographic processes that occurred thousands of years ago. Driven by the adoption of high-throughput sequencing and capture enrichment methods in paleogenomics, the number of published ancient genomes has nearly quadrupled within the last three years (2018–2020). Whereas ancient mtDNA haplogroup repositories are available, no similar resource exists for ancient Y-Chromosomal haplogroups. Here, we present aYChr-DB—a comprehensive collection of 1797 ancient Eurasian human Y-Chromosome haplogroups ranging from 44 930 BC to 1945 AD. We include descriptors of age, location, genomic coverage and associated archaeological cultures. We also produced a visualization of ancient Y haplogroup distribution over time. The aYChr-DB database is a valuable resource for population genomic and paleogenomic studies.

INTRODUCTION

The genomic history of populations is a tapestry of undirected changes as no population remains immutable over time. Whereas coalescent and other reconstruction methods that rely on modern populations are inaccurate and carry a high risk of misinterpretation (1), analyzing the DNA of ancient human populations allows capturing their fine-scale population structure (2) and past events as they were. Combining this evidence with environmental, cultural and other genomic information enables a more accurate representation of the past (3).

The Y-Chromosome contains the largest nonrecombining block in the human genome (4). Using both traditional methods (e.g. PCR) and high-throughput sequencing, haplogroups of ancient individuals are identifiable, facilitating the study of past genetic diversity (3). Combining Y-DNA with radiocarbon dating also provides a means to map Y-chromosomes onto a phylogenetic tree, which can be used to assess whether previous reports of ancestral variation based on modern DNA are supported by ancient samples and if we can find representatives of ancient clades that are rare (5) or no longer exist (6).

Over the past 2 years, ancient Y-chromosomal data have begun to accumulate rapidly. Published data from the period 2007 to 2017 (480 Y chromosomes) was nearly quadrupled within the next 3 years 2018–2020 (1797 Y chromosomes) (Supplementary Table S1). In concert with mitochondrial DNA, Y-Chromosomal DNA has been used to study the origins of present-day and ancient Eurasians (7) along with their languages (8–11) and disease prevalence (3).

Only a handful of ancient DNA databases have been compiled to date, such as the Online Ancient Genome Repository (https://www.oagr.org.au), which primarily stores samples sequenced by the Australian Centre for Ancient DNA, and the AmtDB (12), which predominantly features ancient mtDNA. The lack of a dedicated database focusing on the collection of ancient Y-Chromosomal data has impeded research in the field and prompted us to develop aYChr-DB.

aYChr-DB collates a large proportion of the published Eurasian ancient Y-DNA data over the past 13 years (2007–2020) into an easily accessible archive. The manually curated database not only standardizes the reporting of data and makes haplogroup comparison feasible but also offers socio-cultural annotation. The genomic sequences are available through the source studies.

MATERIALS AND METHODS

Relevant papers were identified by querying PubMed and Google Scholar with the key words ‘ancient Y’, ‘ancient haplogroup’ and ‘ancient DNA’ + ‘Y chromosome’. Both reviews and research articles were selected, with no restrictions on date of publication or journal of publication. Records were then manually curated to remove duplications.

*To whom correspondence should be addressed. Tel: +46 46 222 9419; Fax: +46 46 222 44 25; Email: eran.elhaik@biol.lu.se

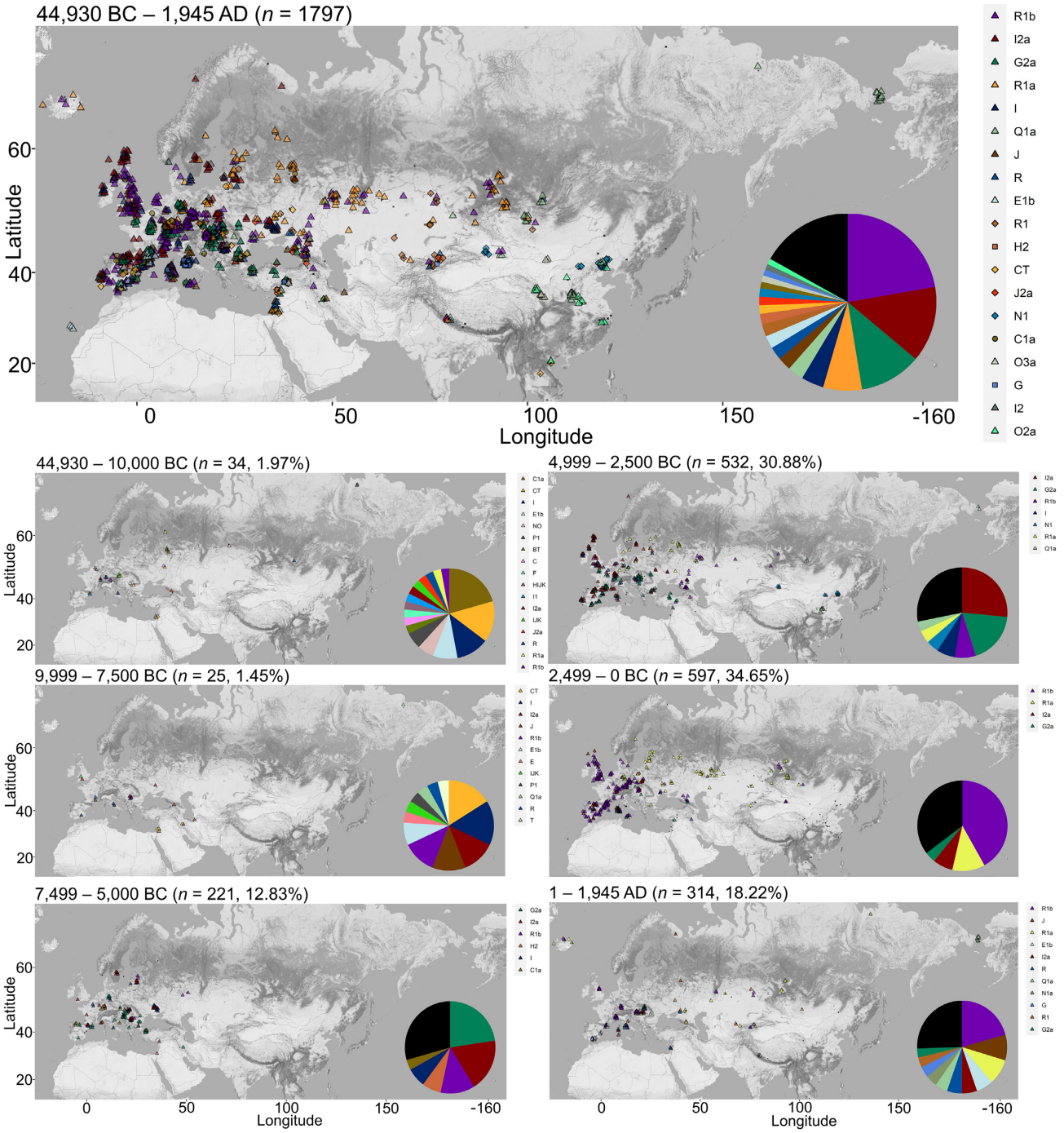

Figure 1. The geographical distribution of 1723 ancient Eurasian haplogroups over time. The location of each archaeological site is marked as a dot. Colored shapes denote the different haplogroups found on the site. A small random variation was used in the plotting to avoid cluttering. Low-frequency haplogroups (<3% in interval maps, <1% in 'all time periods' map) are represented as black wedges in the pie charts and their corresponding locations marked as black crosses on the maps.

Maps were drawn using the ggmap R package (13). aYChr-DB (Supplementary Table S1) is publicly and freely accessible at https://github.com/eelhaik/aYDB.

RESULTS

aYChr-DB contains 1797 samples (Supplementary Table S1). Multiple descriptors are available for each sample, which are named according to the official/published ID, such as country and location. The age of the sample, where applicable, is provided in both BC and BP calibrated from 1950. Carbon-dated samples are shown as calBC/BP. For samples without published coordinate data, we provide coordinates based on location names and descriptions. The archaeological period of each sample has been assigned based

on age and location. Where given, average genomic coverage has been included. The comments section clarifies additional information on the samples which may be pertinent to database users.

We produced a visualization of the aYChr-DB—for a total of 1723 samples after removing 74 undated samples (Figure 1 and Supplementary Figure S1). The full 1797 samples were included in the main 'all time periods' map. For coherency, haplogroups were trimmed to three letters at most, (i.e. R1a1a1 is shown as R1a). Samples were classified into one of six periods, spanning the range of published dates, using the age or average age of the sample. Several trends are noteworthy. A large proportion (65.5%) of collected ancient samples are dated between 0 and 4999 BC. R1b is the modal haplogroup in the ancient Eurasian samples, accounting for 22.3% of the data. I2a is the second most common at 13.9%, followed by G2a at 11.3% and R1a at 7.1%. That the majority of the samples are located in Europe is likely due to the availability of large depositories and history of archaeological research in this region and its propensity for cool, temperate conditions suitable for the preservation of ancient DNA (14). Over 40% of the samples were found in four countries: Spain (11.4%), Russia (10.4%), Hungary (9.7%) and Italy (9.6%).

The major challenge in our efforts to provide coherent and useful annotation was in ascribing meaningful cultural information to the samples. European prehistoric periods are conventionally defined by technological innovations, excepting the Paleolithic-Mesolithic transition, which is a climate transition. The primary European cultural phases are the Neolithic, Copper Age, Bronze Age and Iron Age, followed by historic periods such as the Romans and Medieval periods. Up to the Bronze Age within Europe and West Asia, this technological framework is useful for geneticists as it often corresponds well with major shifts in population structure because these technologies enabled certain groups to move into adjacent regions. The Iron Age and beyond are characterized by advanced civilizations across Europe and West Asia, while in the colder and less fertile regions of Central and Northeastern Asia, nomadic, and hunter-gatherer lifestyles persisted in a scattering of small populations across a broad expanse of territory (15). These people often possessed iron and bronze technologies but had no sedentary agricultural base and demonstrated high mobility. Their cultures have been challenging to classify archaeologically in terms of any overarching technological or historical framework.

In East Asia, we can observe a parallel, although typically not synchronous development of agriculture, copper/bronze technology and eventually iron (16). The transition to agriculture does correspond with population movement (17,18) and is a pattern demonstrated throughout the region. However, subsequent archaeological transitions are usually referred to through dynastic change rather than technological change (19). This is particularly true within China and adjacent regions, despite migration associated with these technological shifts proven at a genetic level (18).

DISCUSSION

We developed a database of ancient Eurasian Y-Chromosomal haplogroups, collating published data from the last 12 years. We assigned missing descriptors to many samples and provided a socio-cultural annotation, which contributes to the uniqueness and usefulness of this resource. Finally, a geographical visualization of the data provides a convenient review of the samples at discrete intervals.

Version 1.0 of the database includes samples from across Eurasia due to the rarity of ancient Y haplogroups from elsewhere. The database will be updated periodically with recently published Y-Chromosome data. We expect that later updates will provide a denser and more extensive global coverage of published data. We hope that the aYChr-DB will increase the accessibility and availability of ancient Y-DNA data.

SUPPLEMENTARY DATA

Supplementary Data are available at NARGAB Online.

ACKNOWLEDGMENTS

Authors' contributions: E.E. initiated the study. L.F. carried out the analyses. L.F. and C.B. annotated the data. All the authors wrote the paper.

FUNDING

MRC [MR/R025126/1, in part] and Crafoord Foundation [in part] to EE.

Conflict of interest statement. E.E. is a consultant to DNA Diagnostic Centre.

REFERENCES

1. Brandt,G., Haak,W., Adler,C.J., Roth,C., Szecsenyi-Nagy,A., Karimnia,S., Moller-Rieker,S., Meller,H., Ganslmeier,R., Friederich,S. *et al.* (2013) Ancient DNA reveals key stages in the formation of central European mitochondrial genetic diversity. *Science*, **342**, 257–261.
2. Esposito,S., Das,R., Syed,S., Pirooznia,M. and Elhaik,E. (2018) Ancient Ancestry Informative Markers for Identifying Fine-Scale Ancient Population Structure in Eurasians. *Gene*, **9**, 625.
3. Prohaska,A., Racimo,F., Schork,A.J., Sikora,M., Stern,A.J., Ilardo,M., Allentoft,M.E., Folkersen,L., Buil,A., Moreno-Mayar,J.V. *et al.* (2019) Human disease variation in the light of population genomics. *Cell*, **177**, 115–131.
4. Underhill,P.A. and Kivisild,T. (2007) Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annu. Rev. Genet.*, **41**, 539–564.
5. Elhaik,E., Tatarinova,T.V., Klyosov,A.A. and Graur,D. (2014) The 'extremely ancient' chromosome that isn't: a forensic bioinformatic investigation of Albert Perry's X-degenerate portion of the Y chromosome. *Eur. J. Hum. Genet.*, **22**, 1111–1116.
6. Kivisild,T. (2017) The study of human Y chromosome variation through ancient DNA. *Hum. Genet.*, **136**, 529–546.
7. Morozova,I., Flegontov,P., Mikheyev,A.S., Bruskin,S., Asgharian,H., Ponomarenko,P., Klyuchnikov,V., ArunKumar,G., Prokhortchouk,E., Gankin,Y. *et al.* (2016) Toward high-resolution population genomics using archaeological samples. *DNA Res.*, **23**, 295–310.
8. Haak,W., Lazaridis,I., Patterson,N., Rohland,N., Mallick,S., Llamas,B., Brandt,G., Nordenfelt,S., Harney,E. and Stewardson,K. (2015) Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*, **522**, 207–211.
9. Szécsényi-Nagy,A., Brandt,G., Haak,W., Keerl,V., Jakucs,J., Möller-Rieker,S., Köhler,K., Mende,B.G., Oross,K. and Marton,T. (2015) Tracing the genetic origin of Europe's first farmers reveals insights into their social organization. *Proc. R. Soc. B.*, **282**, 20150339.

10. Haak,W., Balanovsky,O., Sanchez,J.J., Koshel,S., Zaporozhchenko,V., Adler,C.J., Der Sarkissian,C.S., Brandt,G., Schwarz,C., Nicklisch,N. *et al.* (2010) Ancient DNA from European early neolithic farmers reveals their near eastern affinities. *PLoS Biol.*, **8**, e1000536.
11. Zhao,Y.-B., Zhang,Y., Li,H.-J., Cui,Y.-Q., Zhu,H. and Zhou,H. (2014) Ancient DNA evidence reveals that the Y chromosome haplogroup Q1a1 admixed into the Han Chinese 3,000 years ago. *Am. J. Hum. Biol.*, **26**, 813–821.
12. Ehler,E., Novotny,J., Juras,A., Chylenski,M., Moravcik,O. and Paces,J. (2019) AmtDB: a database of ancient human mitochondrial genomes. *Nucleic Acids Res.*, **47**, D29–D32.
13. Kahle,D. and Wickham,H. (2013) ggmap: spatial visualization with ggplot2. *R. J.*, **5**, 144–161.
14. Brandt,G., Szécsényi-Nagy,A., Roth,C., Alt,K.W. and Haak,W. (2015) Human paleogenetics of Europe—the known knowns and the known unknowns. *J. Hum. Evol.*, **79**, 73–92.
15. Koryakova,L. and Epimakhov,A.V. (2014) *The Urals and Western Siberia in the Bronze and Iron ages*. Cambridge University Press, Cambridge.
16. Roberts,B.W., Thornton,C.P. and Pigott,V.C. (2009) Development of metallurgy in Eurasia. *Antiquity*, **83**, 1012–1022.
17. Fuller,D.Q. (2011) Pathways to Asian civilizations: tracing the origins and spread of rice and rice cultures. *Rice*, **4**, 78–92.
18. Lipson,M., Cheronet,O., Mallick,S., Rohland,N., Oxenham,M., Pietrusewsky,M., Pryce,T.O., Willis,A., Matsumura,H., Buckley,H. *et al.* (2018) Ancient genomes document multiple waves of migration in Southeast Asian prehistory. *Science*, **361**, 92–95.
19. Mei,J., Wang,P., Chen,K., Wang,L., Wang,Y. and Liu,Y. (2015) Archaeometallurgical studies in China: some recent developments and challenging issues. *J. Archaeol. Sci.*, **56**, 221–232.