The Consistencies of Y-Chromosomal and Autosomal Continental Ancestry Varying among Haplogroups

Article - January 2016
DOI: 10.4103/2349-5014.197925

4 authors:

Chuan-Chao Wang
Xiamen University
165 PUBLICATIONS 738 CITATIONS

Lei Shang
Anhui USTC iFlytek
4 PUBLICATIONS 16 CITATIONS

Ivy Hui-Yuan Yeh
Nanyang Technological University
32 PUBLICATIONS 103 CITATIONS

Ryan Lan-Hai Wei
Xiamen University
32 PUBLICATIONS 761 CITATIONS

Some of the authors of this publication are also working on these related projects:

Millet and beans, language and genes. The dispersal of the Transeurasian languages View project

Human Y-chromosome haplogroup Q-L275 View project
ORIGINAL ARTICLES

- The relevant physical trace in criminal investigation
- A Gold Nanoparticle-Enhanced Surface Plasmon Resonance Aptasensor for the Detection of TNT
- Meta-analysis of the association between serotonin transporter polymorphisms and sudden infant death syndrome

BRIEF COMMUNICATION

- The consistencies of Y chromosomal and autosomal continental ancestry vary among haplogroups
The Consistencies of Y-Chromosomal and Autosomal Continental Ancestry Varying among Haplogroups

Chuan-Chao Wang1,2, Lei Shang1, Hui-Yuan Yeh4, Lan-Hai Wei5

1Department of Genetics, Harvard Medical School, Boston, MA, USA, 2Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany, 3Key Laboratory of Forensic Genetics, Institute of Forensic Science, Ministry of Public Security, Beijing, China, 4School of Humanities and Social Sciences, Nanyang Technological University, Singapore, 5State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Contemporary Anthropology, Collaborative Innovation Center for Genetics and Development, School of Life Sciences, Fudan University, Shanghai, China

Abstract

The Y-chromosome has been widely used in ancestry inference based on its region-specific haplogroup distributions. However, there is always a debate on how informative such a single marker is for inferring an individual’s genetic ancestry. Here, we compared genetic ancestry inferences at continental level made by Y-chromosomal haplogroups to those made by autosomal single-nucleotide polymorphisms in 1230 samples of Affymetrix Human Origins dataset. The highest ancestry proportions of a majority of individuals match the highest average continental-ancestry proportions in haplogroups A, B, D, H, I, K, L, T, O, and M. The high consistencies have not been observed in haplogroups E, C, G, J, N, Q, and R, but in some of their sublineages, such as E1a, E1b1a1, E1b1b1b1a, E2b1a, J1a2b, Q1a1a1, Q1a2a1a, R1b1a2a1a, and R2. Although the consistencies of Y-chromosomal and autosomal continental ancestry vary among haplogroups, Y-chromosome could provide valuable clues for individual’s continental ancestry.

Key words: Ancestry inference, autosomal single-nucleotide polymorphism, Y-chromosome

INTRODUCTION

With the advantages of lack of recombination, strict paternal inheritance, small effective population size, low mutation rate, sufficient markers, and population-specific haplotype distribution, Y-chromosome has been widely used in anthropology, population genetics, and forensic genetics to understand population genetic structure, population history, and forensic identifications.1-3 Y-chromosome has also inspired widespread public interest to trace paternal ancestors and been commercially used by many companies. A very famous example is the Y-chromosomal type of Genghis Khan, which was supposed to belong to the “star cluster” under haplogroup C3*-M217 (xM48)2 and has gained extensive attention and attracted numerous consumers to get tested. However, as Y-chromosome is only a single marker and suffers from severe genetic drift, such simple ancestry analyses tend to overlook the contribution of the vast majority of an individual’s ancestors to his/her genome.3,4

There are also many alternative ancestry inference methods, such as testing mitochondrial DNA (mtDNA),5 genome-wide short tandem repeat (STR)5 or single-nucleotide polymorphism (SNP),6 and ancestry informative markers (AIMs).7 The mtDNA is maternally inherited and has been widely used to trace maternal history. Genome-wide STRs, SNPs, and AIMs are usually applied to inferring a detailed composition of an individual’s ancestry. However, some recent genome-wide studies have revealed frequent discrepancies between ancestry inferences using mtDNA versus autosomal SNPs.8,9 The mtDNA case reminds us to rethink how much ancestry information that Y-chromosome could give and the accuracy of Y-chromosomal ancestry inference compared to that of genome-wide ancestry.

Address for correspondence: Dr. Chuan-Chao Wang, Department of Genetics, Harvard Medical School, Boston, MA, USA. Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany. E-Mail: wang@shh.mpg.de

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

For reprints contact: reprints@medknow.com

How to cite this article: Wang CC, Shang L, Yeh HY, Wei LH. The Consistencies of Y-Chromosomal and Autosomal Continental Ancestry Varying among Haplogroups. J Forensic Sci Med 2016;2:229-32.
estimation. Here, we presented a comprehensive analysis using Y-chromosomal and genome-wide autosomal SNP data of more than 1200 male individuals from Affymetrix Human Origins dataset to directly and quantitatively assess the consistency of Y-chromosomal and autosomal continental ancestry.

Materials and Methods

The Y-chromosomal and autosomal genotype data for 1230 male individuals were extracted from Affymetrix Human Origins dataset using EIGENSOFT and PLINK. Y-chromosomal haplogroups were classified based on the International Society of Genetic Genealogy phylogenetic tree at January 28, 2015 (http://www.isogg.org/). We used ADMIXTURE v. 1.23 to estimate ancestry proportions for 1230 males with 594,924 autosomal SNPs. Each run involved 100 replicates with different random starting seeds, default 5-fold cross-validation, and varying the number of ancestral populations K from 2 to 12. At K = 8, the samples were well assigned to eight continental regions: Africa, Middle East, Europe, South Asia, East Asia, Siberia, Oceania, and Americas. The average continental-ancestry proportions within each Y-chromosomal haplogroup, standard deviation (SD) of individual continental-ancestry percentages for each continental region in each haplogroup, mean pairwise Euclidean distance (d) within each haplogroup, and consistency scores were all calculated according to Emery et al. The graphical displays for ancestry plot were carried out in R statistical software v3.0.2.

Results

The Human Origin dataset contained male samples of worldwide lineages from Y-chromosomal haplogroups A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, Q, R, S, and T. All the haplogroups, except A, B, K, M, and S, were found in more than one continent. Haplogroups A and B were only discovered in Africa whereas K, M, and S were only present in Oceania. Likewise, haplogroups A and B had predominately African ancestry whereas K, M, and S had predominately Oceanian ancestry. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among populations from East Asia. The East Asian ancestry proportions in haplogroups D and O were extremely high among popu...
Figure 1: (a) Haplogroup-averaged continental-ancestry proportions; (b) Individual continental-ancestry proportions in the male individuals of Affymetrix Human Origins Dataset
Similarly, haplogroup J1a2b was associated with Middle East, and Q1a1a1 could be assigned into East Asia.

**Discussion**

We directly compared the genetic ancestry revealed by Y-chromosomal haplogroups to those inferred from genome-wide autosomal SNPs in a worldwide dataset. The continental-ancestry compositions varied among individuals of the same Y-chromosomal haplogroup judging from high SDs. About 70% of the Y-chromosomal haplogroups could be assigned to be associated with certain continents due to the high continent-specific ancestry proportions. The highest ancestry proportions of a majority of individuals match the highest average continental-ancestry proportions in haplogroups A, B, D, H, I, K, L, T, O, and M. Although the high consistencies have not been observed in haplogroups E, C, G, J, N, Q, and R, some of their sublineages, such as E1a, E1b1a1, E1b1b1b1a, E2b1a, J1a2b, Q1a1a1, Q1a2a1a1, R1b1a2a1a, and R2 corresponded well with certain continents. The Y-chromosome seemed like to give higher prediction accuracy for individual ancestries than the mtDNA. This phenomenon might be caused by sex-biased migrations, which refers to a higher female migration rate in human populations. A series of studies have revealed that the among-population components of genetic variation are higher for the Y-chromosome than for the mtDNA, indicating that the Y-chromosomes tend to be more localized geographically. The Y-chromosome in a way could provide valuable clues for individual’s continental ancestry, but it probably neglected many other detailed ancestry information. One, or at most two, top ancestry components could be well represented by majority of Y-chromosomal haplogroups, whereas other ancestry information is lost. For instance, the highest South Asian ancestry proportions have been detected in individuals of haplogroup H. Meanwhile, East Asia, Europe, and Middle East each have contributed more than 10% of genetic ancestry to many individuals of haplogroup H, which could not be reflected by such a single Y-chromosomal marker. In addition, the Y-chromosomal haplogroup classifications in this study were not very informative. The rough assignment might lose some information of a certain lineage and probably have resulted in bias conclusion. For example, sublineages of haplogroup C have distinct geographic distributions. However, we do not have enough markers in this dataset to identify the detailed phylogeny of haplogroup C individuals, resulting in the inconclusive ancestry inference of this haplogroup.

**Acknowledgments**

CCW is supported by the Max Planck Society and Harvard Medical School.

**Financial support and sponsorship**

Nil.

**Conflicts of interest**

There are no conflicts of interest.

**References**

1. Jobling MA, Tyler-Smith C. The human Y chromosome: An evolutionary marker comes of age. Nat Rev Genet 2003;4:598-612.
2. Zerjal T, Xue Y, Bertorelle G, Wells RS, Bao W, Zhu S, et al. The genetic legacy of the Mongols. Am J Hum Genet 2003;72:717-21.
3. Shriver MD, Kittles RA. Genetic ancestry and the search for personalized genetic histories. Nat Rev Genet 2004;5:611-8.
4. Emery LS, Magnaye KM, Bigham AW, Akery JM, Bamshad MJ. Estimates of continental ancestry vary widely among individuals with the same mtDNA haplogroup. Am J Hum Genet 2015;96:183-93.
5. Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, et al. Genetic structure of human populations. Science 2002;298:2381-5.
6. Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanov K, et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature 2014;513:409-13.
7. Shriver MD, Smith MW, Jin L, Marcini A, Akery JM, Deka R, et al. Ethnic-affiliation estimation by use of population-specific DNA markers. Am J Hum Genet 1997;60:957-64.
8. Poetsch M, Wiegand A, Harder M, Blöhm R, Rakotomavo N, Freitag-Wolf S, et al. Determination of population origin: A comparison of autosomal SNPs, Y-chromosomal and mtDNA haplogroups using a Malagasy population as example. Eur J Hum Genet 2013;21:1423-8.
9. Price AL, Zaitlen NA, Reich D, Patterson N. New approaches to population stratification in genome-wide association studies. Nat Rev Genet 2010;11:459-63.
10. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559-75.
11. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res 2009;19:1655-64.
12. R Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2013.
13. Seielstad MT, Minch E, Cavalli-Sforza LL. Genetic structure of human populations. Science 1996;274:1292-7.
14. Lippold S, Xu H, Ko A, Li M, Renaud G, Butthof A, et al. Human paternal and maternal demographic histories: Insights from high-resolution Y chromosome and mtDNA sequences. Invest Genet 2014;5:13.
15. Zhong H, Shi H, Qi XB, Xiao CJ, Jin L, Ma RZ, et al. Global distribution of Y-chromosome haplogroup C reveals the prehistoric migration routes of African exodus and early settlement in East Asia. J Hum Genet 2010;55:428-35.