Demographic Cognitive Patterns Revealed from Human Genome

Levon Yepiskoposyan
Laboratory of Ethnogenomics, IMB, NAS RA
Yerevan, Armenia
e-mail: lepiskop@gmail.com

ABSTRACT

Here, we provide some examples of extracting demographic cognitive patterns from human genome data through applying traditional and recently developed statistical and bioinformatics tools. These methods were used to make inferences on the demographic history of Armenians (as a template population) based on uniparentally (through paternally inherited Y chromosome and maternally inherited mithochondrial DNA) and biparentally (through autosomes) transmitted genetic information.

Keywords

Human genome, genetic history, demographic cognitive patterns.

1. INTRODUCTION

A typical human genome comprising ca. 3,200,000 nucleotides differs from the reference sequence at 4.1 million to 5.0 million sites. Although >99.9% of variants consist of single nucleotide polymorphisms (SNPs) and short insertions and deletions (indels), structural variants affect more bases: the typical genome contains in average 2,100 to 2,500 structural variants, affecting ≈20 million bases of sequence [1].

Thorough exploration of the patterns in which genetic variants (polymorphisms) are shared among individuals and distinct ethnic or geographic groups provides detailed accounts of population demographic history. Although the bulk of common polymorphisms are shared across the globe, rarer variants are typically confined to geographically closely related populations: ca. 85% of variants are restricted to a single continental group. Using a maximum likelihood approach, it's possible to estimate the proportion of each genome derived from several putative 'ancestral populations' [1].

Here, we show some principal outcomes of various computational methods' application which shed light on main demographic features of the Armenian population based on the genome data. The following crucial questions related to the genetic history of the population were addressed:

(1) time of origin of the Armenian gene pool;
(2) genetic continuity of populations from the Neolithic to modern day.

2. TIME OF ORIGIN

We carried out full-likelihood Bayesian inference of genetic and demographic parameters under population splitting and growth using the BATWING program (Bayesian analysis of trees with internal node generation; URL: http://www.maths.abdn.ac.uk/~ijw), extended from the algorithm presented by Wilson and Balding [2]. BATWING uses a Markov chain Monte Carlo procedure to generate a sequence of genealogical and population trees, with associated model parameter values, consistent with the genetic data (Y-chromosomal SNPs and Short Tandem Repeats – STRs) observed in a sample of individuals. At equilibrium the sequence of trees correctly samples from the posterior probability distribution of trees given the observed data and the assumed underlying genetic and demographic model.

Using BATWING, we fitted a model to the Armenian data that included: (a) exponential growth from an initially constant-size population and (b) binary fission (without subsequent migration) of an initially panmictic population into the six regions that we defined in and around Armenia today.

Posterior distributions of parameters of interest are presented in Fig. 1 [3]. Despite wide credible intervals, the dates for the time to most recent common ancestor, start of growth and the oldest (deepest) population split are clearly separated. The posterior growth rate (mean=0.026, 95% credible interval: 0.010–0.053) indicates a strong growth signal in the data, but with a recent start date under the assumed exponential growth model (mean=4.8 thousand years ago, 95% credible interval: 2.0–11.1), most likely within the last 10,000 years (95.7% support).
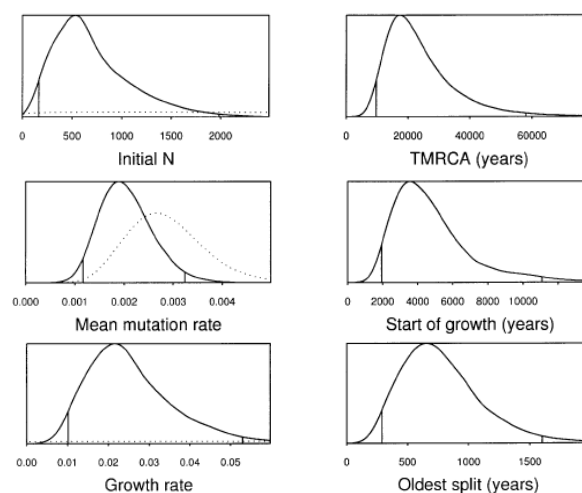

Figure 1. Posterior marginal distributions (solid lines) of key parameters from the BATWING analysis described in the text. Vertical lines 95% credible intervals. Where applicable, prior distributions are also shown (dotted lines). TMRCA time to most recent common ancestor. (from Weale et al [3])

The BATWING results support a signal of population growth starting in the Neolithic (95% credible interval: 2.0–11.1 thousand years ago), and this signal appears robust to the model of population stratification used. This signal does not necessarily reflect the demographic history of the Armenian population per se. It may, for example, be present in other populations that at some time shared common

demographic and genealogical histories with the Armenian sample. However, the date suggested for the start of population growth is consistent with the region's archaeological evidence that suggests an increase in local settlement density dating to the Kuro-Araxian culture c. 3500–2500 BC [4].

Recently the genome-wide variation in 173 Armenians has been analyzed and compared with 78 other worldwide populations [5]. Based on the results of this study Armenians form a distinctive cluster comprising the Near Eastern, European, and the Caucasus populations. Armenian diversity can be explained by several mixtures of Eurasian populations that occurred between ca. 3000 and ca. 2000 BCE, a time span characterized by significant population movements after the domestication of horse, invention of chariots, and the rise of mighty civilizations in the Near East. However, genetic signals of population mixture abruptly cease after ca. 1200 BCE when Bronze Age civilizations in the Eastern Mediterranean world suddenly collapsed. Armenians have since remained genetically virtually isolated from neighboring populations. Modern Armenians have higher genetic affinity to Neolithic Europeans than other present-day Near Easterners, and that about one third of Armenian ancestry may originate from an ancestral population that is best represented by Neolithic Europeans.

TreeMix package [6] was applied to draw a pattern of genetic relationships based on representative Near East populations plus Armenians and Turks. TreeMix uses a model that allows for both population splits and gene flow to better capture historical relationships between populations. A tree showing the relationships among population groups is presented in Fig. 2.

The TreeMix results revealed gene flow from the Iceman (5300-year-old individual discovered on the Italian part of the Ötztal Alps) to Armenians, accounting for about 29% of their ancestry. The graph structure appeared robust in 100 bootstrap replicates with the first migration (highest weight and lowest P-value), always leading from the Iceman to Armenians.
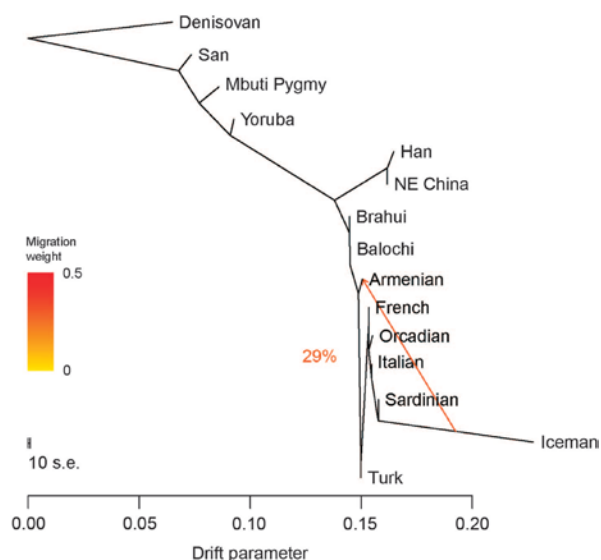

Figure 2. Inferred population tree with one mixture event. The migration arrow is coloured according to its weight; the weight is correlated with the ancestry fraction and shows that 29% of Armenian ancestry is derived from a population related to ancient Europeans. (from Haber et al [2016])

3. GENETIC CONTINUITY

To check the hypothesis of genetic succession of inhabitants in the area of South Caucasus, we have analyzed the whole mitochondrial genomes of 206 modern and 52 ancient individuals representing, respectively, contemporary Armenians and the former population of the East Armenia spanning eight millennia [7]. The results clearly indicate genetic continuity of human maternal gene pool since Neolithic times despite well documented cultural shifts in the South Caucasus.

A scenario of genetic continuity is supported by two previous studies that included low-coverage genomic data from a few ancient individuals from the South Caucasus: Allentoft et al. observed genetic similarities between Bronze Age individuals (ca. 3,500 years BP) and modern Armenians [8], and Lazaridis et al. showed a similarity between Chalcolithic (ca. 6,000 years BP) and Bronze Age (ca. 3,500 years BP) individuals excavated in Armenia [9]. Moreover, Jones et al. presented results implying that such continuity might extend even further back in time: it appears that Upper Palaeolithic Caucasus hunter-gatherers and Mesolithic individuals from the South Caucasus (Georgia) are genetically close to modern Caucasian groups, albeit also displaying their own genetic component [10].

A Bayesian skyline plot (BSP) based on all modern and ancient mitochondrial genomes analyzed together revealed four putative demographic events (geometric mean of 4.4 with 95% highest posterior density intervals between 4 and 6 as obtained from the results of extended skyline plot analysis). The plot indicates a small but noticeable decrease in the effective female population size (N_e) around 25 kya during the Last Glacial Maximum (LGM), which is followed by a rapid (roughly 10-fold) population increase until around 10 kya (Figure 3).
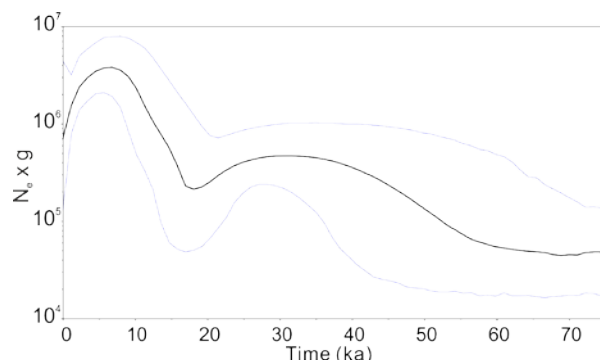

Figure 3. Population Demographics: Bayesian skyline plot. Values on the Y axis represent the effective female population size (Ne) x generation time (g) (from Margaryan et al [7])

This demographic trajectory is in accordance with previously published results based on data from European Mesolithic and Palaeolithic individuals [20]. Interestingly, Ne appears to be declining around 5 kya, although the large confidence interval makes this conclusion tentative. This result was not observed in previous studies based on smaller samples size and modern data alone [11]. The timing of this putative decline coincides with the formation of complex societies during the Bronze Age in the region [12]. This could have increased susceptibility to diseases such as plague, which was present in both Central Asia and Europe during the early Bronze Age [13]. Another possibility is that the society formation of Bronze Age populations could have reduced the effective female population size without affecting the census population sizes. Factors like populations size fluctuations, increased selection, variation in family size, and changing

population sub-structuring can all affect the estimates of effective population size [14]. However, it has previously been noted that recent population declines on BSP plots should be interpreted with caution as it may be an artifact of population structure [15].

Approximate Bayesian computation (ABC) analyses were applied to test five possible demographic model scenarios (Figure 4), simulating 1,000,000 datasets from each model. For the modern group we used a combined (n = 206) contemporary Armenian population.

The cross-validation of the ABC model selection allowed to easily distinguish between the genetic continuity scenario (model 1) and the rest. We used two statistical tests, marginal density p value and Tukey depth p value, to assess the fit of our five models to the observed data.

Based on comparison of the marginal densities, the analysis favor model 1 (posterior probability of 89% and Bayes factor of 8.1), which assumes genetic continuity between the ancient group and the modern Armenians. This result suggests that there were no major genetic shifts in the mtDNA gene pool in Eastern Armenia across the last 7,800 years.
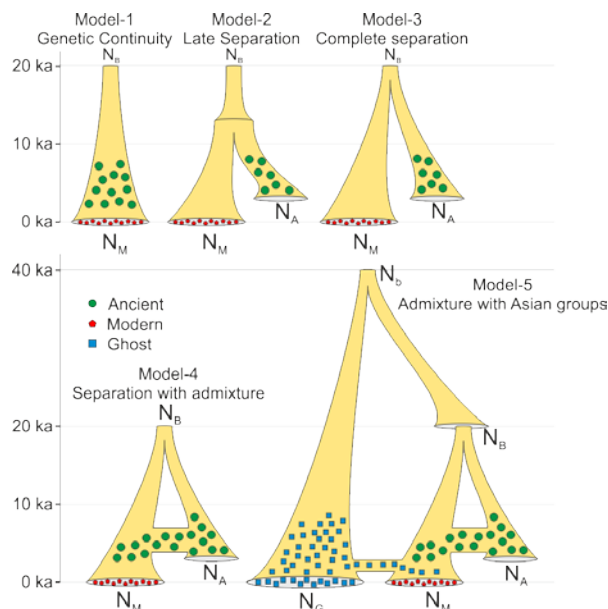

Figure 4. Population Demographics: Schematic representation of the five demographic models used for coalescent simulations. Population size of: modern Armenians (N_M); ancient group (N_A); "ghost" Asian population (N_G); first bottleneck (20 kya; N_B); second bottleneck (40 kya; N_b) (from Margaryan et al [7])

4. CONCLUSION

Human genome (extracted from modern and ancient DNA samples) consists of abundant information sufficient to reveal demographic cognitive patterns and make inferences on the principal events in the genetic history of a given population: the time of reproductive isolation of its gene pool; the timing and extent of admixture with neighboring groups; genetic continuity of populations inhabiting the area under consideration; the directions of ancient migrations and possible genetic contacts with other populations of the area.

REFERENCES

[1] The 1000 Genomes Project Consortium, "A global reference for human genetic variation", *Nature*, 2015, 526: 68-74.

[2] Wilson IJ, Balding DJ (1998) Genealogical inference from microsatellite data. Genetics 150:499–510

[3] Weale ME, Yepiskoposyan L, Jager RF, Hovhannisyan N, Khudoyan A, Burbage-Hall O, Bradman N, Thomas M. Armenian Y chromosome haplotypes reveal strong regional structure within a single ethno-national group. Human Genetics, 2001, 109: 659-674.

[4] Badalyan RS (1986) Early Bronze culture in the territory of Shirak. Thesis, Yerevan (in Russian).

[5] Haber M, Mezzavilla M, Xue Y, Comas D, Gasparini P, Zalloua P, Tyler-Smith C Genetic evidence for an origin of the Armenians from Bronze Age mixing of multiple populations. *European Journal of Human Genetics,* 2016, 24:931-936.

[6] Pickrell JK, Pritchard JK: Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet,* 2012; 8: e1002967.

[7] Margaryan A, Derenko M, Hovhannisyan H, Malyarchuk B, Heller R, Khachatryan Z, Avetisyan P, Badalyan R, Bobokhyan A, Melikyan V, Sargsyan G, Piliposyan A, Simonyan H, Mkrtchyan R, Denisova G, Yepiskoposyan L, Willerslev E, Allentoft M E. Eight millennia of matrilineal genetic continuity in the South Caucasus. *Current Biology*, Published online 29 June 2017. DOI: 10.1016/j.cub.2017.05.087

[8] Allentoft ME, Sikora M, Sjogren K-G, Rasmussen S, Rasmussen M, Stenderup J, Damgaard PB, Schroeder H, Ahlstrom T, Vinner L, et al. Population genomics of Bronze Age Eurasia. *Nature*, 2015, 522: 167–172.

[9] Lazaridis I, Nadel D, Rollefson G, Merrett DC, Rohland N, Mallick S, Fernandes D, Novak M, Gamarra B, Sirak K, et al. Genomic insights into the origin of farming in the ancient Near East. *Nature*, 2016, 536: 419–424.

[10] Jones ER, Gonzalez-Fortes G, Connell S, Siska V, Eriksson A, Martiniano R, McLaughlin RL, Gallego Llorente M, Cassidy LM, Gamba C, et al. Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat Commun*, 2015, 6, 8912.

[11] Schonberg A, Theunert C, Li M, Stoneking M, and Nasidze I. High-throughput sequencing of complete human mtDNA genomes from the Caucasus and West Asia: high diversity and demographic inferences. *Eur J Hum Genet*, 2011, 19: 988–994.

[12] Smith AT, Badalyan R, Avetisyan P, Zardaryan M, Hayrapetyan A, Minc L, and Monahan B. Early complex societies in Southern Caucasia: a preliminary report on the 2002 investigations by Project ArAGATS on the Tsakahovit Plain, Republic of Armenia. *Am J Archaeol*, 2004, 108: 1–41.

[13] Rasmussen S, Allentoft ME, Nielsen K, Orlando L, Sikora M, Sjogren K-G, Pedersen AG, Schubert M, Van Dam A, Kapel MO, et al. Early divergent strains of Yersinia pestis in Eurasia 5,000 years ago. *Cell*, 2015, 163: 571–582.

[14] Jobling M, Hollox E, Hurles M, Kivisild T, and Tyler-Smith C. Human Evolutionary Genetics, Second Edition (Garland Science), 2013.

[15] Heller R, Chikhi L., and Siegismund HR. The confounding effect of population structure on Bayesian skyline plot inferences of demographic history. *PLoS ONE*, 2013, 8, e62992.