F1000Research

CORRESPONDENCE

High Frequency Haplotypes are Expected Events, not Historical Figures [version 1; referees: 1 approved, 2 approved with reservations]

Elsa G. Guillot, Murray P. Cox

Statistics and Bioinformatics Group, Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand

Abstract
Cultural transmission of reproductive success states that successful men have more children and pass this raised fecundity to their offspring. Balaresque and colleagues found high frequency haplotypes in a Central Asian Y chromosome dataset, which they attribute to cultural transmission of reproductive success by prominent historical men, including Genghis Khan. Using coalescent simulation, we show that these high frequency haplotypes are consistent with a neutral model, where they commonly appear simply by chance. Hence, explanations invoking cultural transmission of reproductive success are statistically unnecessary.

Open Peer Review

Referee Status: ☑

	Invited Referees		
	1	**2**	**3**
version 1 published 04 Sep 2015	? report	☑ report	? report

1 **Sohini Ramachandran**, Brown University USA

2 **Heather Norton**, University of Cincinnati, USA USA

3 **Nick Patterson**, Broad Institute of MIT and Harvard USA

Discuss this article

Comments (0)

Cultural transmission of reproductive success states that successful men have more children and pass this raised fecundity on to their offspring. Observed in modern human populations from genealogies and surname studies[1], in a genetic setting cultural transmission of reproductive success would cause particular male lines to dominate on the Y chromosome. Balaresque and colleagues[2] examined a Y chromosome dataset from Central Asia to determine whether they could reconstruct historic instances of this behavior. Screening 8 microsatellites on the Y chromosome in 5,321 Central Asian men (distribution in Figure 1), they identified 15 haplotypes that are carried by more than 20 men (grey bars). The authors described these haplotypes as 'unusually frequent,' but did not provide any statistical support for this statement. These lineages were subsequently connected by the authors to prominent historical figures, including Genghis Khan and Giocangga.

However, in any given haplotype frequency distribution, a number of haplotypes are expected to occur at high frequency simply by chance. In neutrally evolving systems, haplotype frequency distributions follow a Zipfian power law[3]: most lineages are carried by only a few men (Figure 1, left side), while a small number of lineages are carried by many men (Figure 1, right side). The Y chromosome distribution observed by Balaresque and colleagues closely follows such a power law, thus providing strong preliminary evidence that their Y chromosome dataset may be selectively neutral.

To more explicitly test whether the observed high frequency haplotypes are actually unusually frequent, we simulated genetic data under the standard coalescent, a neutral model that does not include cultural transmission of reproductive success. We modeled the evolution of 5,321 Y chromosomes, each carrying 8 fully linked microsatellites, to match the observed data. The code for these simulations, including full details of parameter values, is available online (http://elzaguillot.github.io/Allele-Frequency-Spectrum-simulations).

Simulations were first run across a sweep of θ values to find the best match with the power law distribution observed in the Central Asian Y chromosome dataset. The least squares fit between observed and simulated distributions was minimized at $\theta = 131$. In one million simulations run at this value, we found that 27.2% of the simulations contained at least 15 haplotypes carried by more than 20 men, thus illustrating that high frequency haplotypes like those observed among Central Asian Y chromosomes are relatively common, even when cultural transmission of reproductive success is not acting. The Y chromosome haplotype frequency distribution observed by Balaresque and colleagues falls within the 95% confidence intervals of our simulations (Figure 1, red shading).

The most parsimonious explanation is therefore that the high frequency haplotypes observed by Balaresque and colleagues in
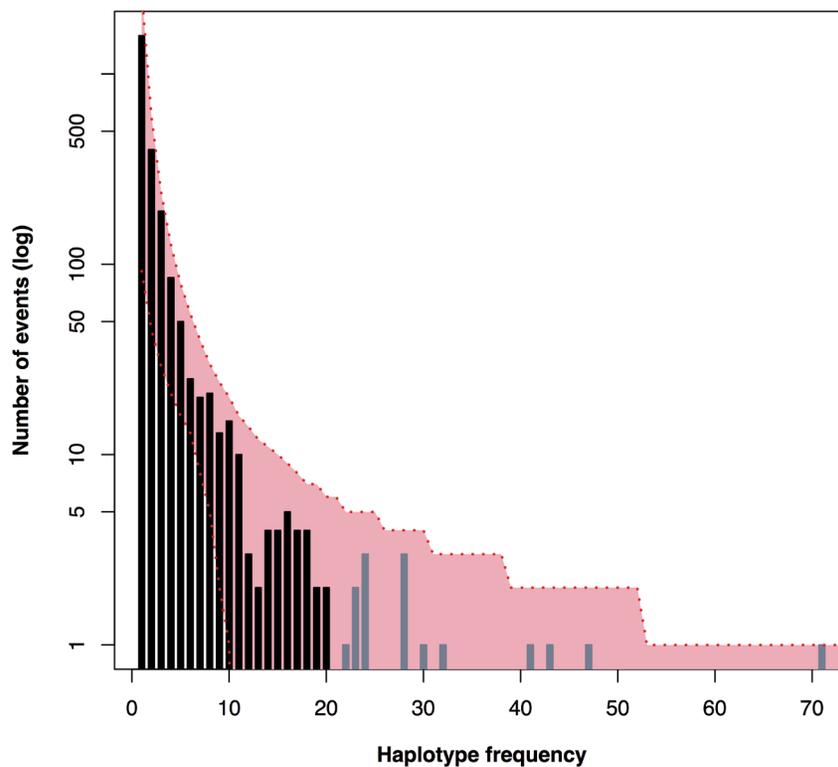
Figure 1. Microsatellite haplotype frequency distribution. The distribution (black and grey bars) is identical to Figure 2 of Balaresque et al[2]. Grey bars indicate the 15 haplotypes that Balaresque and colleagues describe as 'unusually frequent.' Red shading indicates the 95% confidence intervals of haplotype frequencies from one million simulations under a fitted neutral model. All of the high frequency haplotypes (grey bars) fall within these 95% confidence bounds.

Central Asia are simply expected chance events. While we strongly encourage further research into cultural transmission of reproductive success, no statistical evidence has been presented to show that this process has acted on this particular dataset of Central Asian Y chromosomes. As no additional evidence is presented in support of proposed links to famous historical men, these haplotypes instead most likely reflect the chance proliferation of random male lines, probably from historically unrecorded, culturally undistinguished, but biologically lucky Central Asian men.

Software availability
Latest source code for allele frequency spectrum simulations
http://elzaguillot.github.io/Allele-Frequency-Spectrum-simulations

Archived source code as at the time of publication
http://dx.doi.org/10.5281/zenodo.29888[4]

License
Lesser GNU Public License 3.0 https://www.gnu.org/licenses/lgpl.html

Author contributions
EGG conceived the study and carried out the research. EGG and MPC designed the experiments and wrote the manuscript. Both authors were involved in the revision of the draft manuscript and have agreed to the final content.

Competing interests
No competing interests were disclosed.

Grant information
The author(s) declared that no grants were involved in supporting this work.

References

1. Kolk M: **Multigenerational transmission of family size in contemporary Sweden.** *Popul Stud (Camb).* 2014; **68**(1): 111–129.
   PubMed Abstract | **Publisher Full Text**

2. Balaresque P, Poulet N, Cussat-Blanc S, *et al.*: **Y-chromosome descent clusters and male differential reproductive success: young lineage expansions dominate Asian pastoral nomadic populations.** *Eur J Hum Genet.* 2015; In Press.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

3. Berestycki J, Berestycki N, Limic V: **Asymptotic sampling formulae for Λ-coalescents.** *Ann I H Poincaré-Pr.* 2014; **50**(3): 715–731.
   Publisher Full Text

4. Guillot EG, Cox MP: **Allele Frequency Spectrum simulations.** *Zenodo.* 2015.
   Data Source

Open Peer Review

Current Referee Status:

Nick Patterson

Medical and Population Genetics Program, Broad Institute of MIT and Harvard, Cambridge, MA, USA

This short note considers the recent paper by Balaresque *et al*.[1] on the distribution of Y-chromosome haplotypes in Central Asia. Through simulation they show that the haplotype frequency distribution is not very surprising and suggest that the results of[1] are most likely due to chance.

But there is more to the analysis of[1] than just the haplotype frequency. Their analysis groups haplotypes into `descent clusters', estimates the time to the most recent common ancestor (TMRCA) and looks into the spatial distribution of the haplotypes. None of this was simulated. There is no formal test applied in[1], but visually the results look to this reviewer very surprising under a scenario where to quote Guillot and Cox the results are
chance proliferation of random male lines... from culturally undistinguished but biologically lucky...men
In autosomal analysis of admixture events[2], overwhelming genetic evidence was found for the Mongol expansion across Eurasia. This by no means shows that the Y-chromosome signal was, at least partially, driven by high status Mongols, but to this reviewer this still seems more likely than not.

References

1. Balaresque P, Poulet N, Cussat-Blanc S, Gerard P, Quintana-Murci L, Heyer E, Jobling MA: Y-chromosome descent clusters and male differential reproductive success: young lineage expansions dominate Asian pastoral nomadic populations.*Eur J Hum Genet*. 2015; **23** (10): 1413-22 PubMed Abstract | Publisher Full Text
2. Hellenthal G, Busby GB, Band G, Wilson JF, Capelli C, Falush D, Myers S: A genetic atlas of human admixture history.*Science*. 2014; **343** (6172): 747-51 PubMed Abstract | Publisher Full Text

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Competing Interests: No competing interests were disclosed.

Heather Norton

Department of Anthropology, University of Cincinnati, USA, Cincinnati, OH, USA

Summary

In this manuscript Guillot and Cox test the claim made by Balaresque *et al*. (2015) that a subset of Y-chromosome haplotypes from Central Asian men occur at "unusually" high frequency, possibly indicating social selection for men carrying these lineages. Using simulations designed to match the data reported by Balaresque *et al*. the authors demonstrate that the reported distribution of Y chromosome haplotypes can be obtained under neutral conditions. This suggests that it is not necessary to invoke a model that includes cultural transmission of reproductive success to explain the observed distribution.

Comment

While the focus of this correspondence article is on the Balaresque data, can the authors briefly comment on other papers that have also investigated cultural transmission of reproductive success—specifically, have there been other studies that report high frequency Y haplotypes in other populations that are **not** consistent with neutrality?

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Referee Report 01 October 2015

doi:10.5256/f1000research.7561.r10223

Sohini Ramachandran

Department of Ecology and Evolutionary Biology, Brown University, Providence, RI, USA

Guillot and Cox present a very interesting criticism of Balaresque *et al*.'s work in press on high frequency haplotypes in Central Asian Y chromosomes, by showing that distributions like those observed by Balaresque and colleagues can be observed using neutral simulations.

I have three comments I would like to see the authors address:
1. The "Genghis Khan reproductive success" hypothesis emerged in Zerjal *et al*.'s work in 2003 and I think it would be helpful for the authors to comment on what analyses in that work support Zerjal *et al*.'s conclusions; can their simulations reproduce what Zerjal *et al*. observed under strictly neutral processes without a high number of mergers in the coalescent process?
2. The phrase "historically unrecorded, culturally undistinguished, but biologically lucky Central Asian men." should be changed to "historically unrecorded but biologically lucky Central Asian men."
3. The authors should provide sample output files for their simulation pipeline for users to analyze, and their code so that the number of simulations run is a user-provided argument. Given that the pipeline can take at least hours and perhaps days to generate the million simulations they studied, a toy example is worth looking at quickly and the authors could allow readers to generate examples more quickly without needing to fiddle with the bash/python/R pipeline on their own.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Competing Interests: No competing interests were disclosed.