bioRxiv preprint doi: https://doi.org/10.1101/022160; this version posted July 7, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

High Frequency Haplotypes are Expected Events, not Historical Figures

Elsa G. Guillot[1] and Murray P. Cox[1,*]

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

[*] Corresponding author

Contact:

Institute of Fundamental Sciences
Massey University
Private Bag 11 222
Palmerston North 4442
New Zealand

Email: m.p.cox@massey.ac.nz
Phone: +64-6-356 9099 x84747
Fax: +64-6-350 5682

bioRxiv preprint doi: https://doi.org/10.1101/022160; this version posted July 7, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Cultural transmission of reproductive success states that successful men have more children and pass this higher 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] interrogated 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 were 'unusually frequent,' which they defined as those haplotypes carried by more than 20 men (grey bars). The authors connect these lineages to prominent historical figures, including Genghis Khan and Giocangga.

However, high frequency haplotypes are expected 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 observed Y chromosome distribution follows a power law, thus providing strong preliminary evidence that this Y chromosome dataset is selectively neutral. However, 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 details of parameter values, is available online[4]. Simulations were first run across a sweep of θ values to find the best match with the observed Y chromosome data. The least squares fit between observed and simulated distributions was minimized at $\theta = 131$, so one million simulations were run at this value. We found that 27.2% of the simulations contain 15 or more haplotypes carried by more than 20 men, thus illustrating that such high frequency haplotypes are actually quite common. The observed Y chromosome haplotype frequency distribution clearly falls within the 95% confidence intervals of the simulations (Figure 1, red shading).

bioRxiv preprint doi: https://doi.org/10.1101/022160; this version posted July 7, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The most parsimonious explanation is therefore that the high frequency haplotypes observed in Central Asia are simply expected chance events, and an explanation invoking cultural transmission of reproductive success is not necessary to account for them. As no other evidence is presented to support 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.

1 Kolk M: Multigenerational transmission of family size in contemporary Sweden. *Popul Stud* 2014; **68**: 111-129. http://dx.doi.org/10.1080/00324728.2013.819112

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. http://dx.doi.org/10.1038/ejhg.2014.285

3 Berestycki J, Berestycki N, Limic V: Asymptotic sampling formulae for Λ-coalescents. Ann I H Poincaré-Pr 2014; **50**: 715-731. http://projecteuclid.org/euclid.aihp/1403276996

4 http://elzaguillot.github.io/Allele-Frequency-Spectrum-simulations

bioRxiv preprint doi: https://doi.org/10.1101/022160; this version posted July 7, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

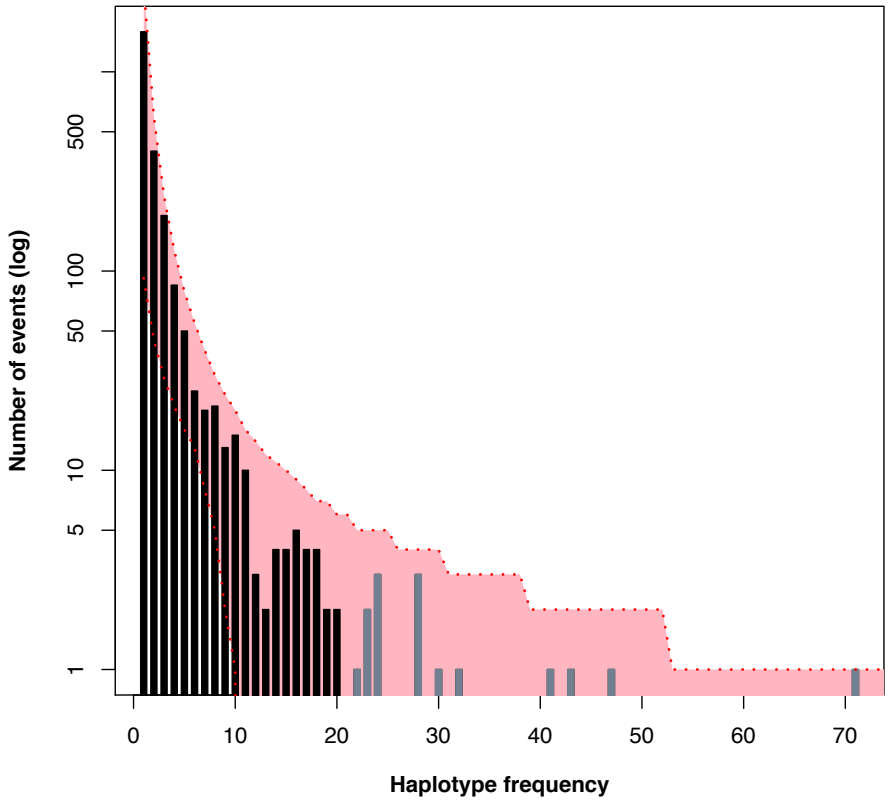

Figure 1 Microsatellite haplotype frequency distribution. The distribution (black and grey bars) is identical to Figure 2 of Balaresque *et al* [2]. Grey bars indicate the 15 haplotypes that are described as 'unusually frequent.' Red shading indicates the 95% confidence intervals of haplotype frequencies from one million simulations. All of the high frequency haplotypes (grey bars) fall within these 95% confidence bounds.