bioRxiv preprint doi: https://doi.org/10.1101/848911; this version posted November 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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-NC 4.0 International license.

Quantifying the prevalence of assortative mating in a human population

Klaus Jaffe
Universidad Simón Bolívar, Caracas, Venezuela

Abstract

For the first time, empirical evidence allowed to construct the frequency distribution of an index related to the degree of genetic relatedness between the parents of about 0.5 million humans living in the UK. The results show that a large proportion of the population is not the product of parents choosing a mate randomly. Assortative mating leading to offspring, that occurs between genetic related individuals, is very common. High degrees of genetic relatedness, i.e. extreme inbreeding, is rare. The evidence shows that assortative mating is highly prevalent in this large population sample. This novel empirical result suggests that assuming random mating, as widely done in population genetic studies, is not an appropriate approximation to reality.

INTRODUCTION

In most population genetic models, random mating is considered the default option. For the first time, empirical evidence from large human populations as to the generality of this assumption is presented (4). The evidence shows that assortative is highly prevalent in this large population sample. Assortation has been proposed to be necessary and widespread in animal sexual reproduction of most living beings. Assortative mating has been demonstrated to occur among humans (1, 5, 14) and many other animal and plants (7). The physical mechanism behind this assumption is that assortation allows evolution to handle epistasis more effectively than random mating, increasing the error threshold for lethal mutations (11,12), allowing synergy between cooperating genes to be maintained and expanded (9). Assortative mating is also known to accelerate adaptation due to its effect on the Hardy–Weinberg equilibrium (13). As these mechanism are basic features of all complex evolutionary scenarios, this has implications for a wide range of disciplines (6).

The degree to which the non-random mating influences genetic architecture remains unclear as most theoretical studies assume random mating. However, existing simulations showed that assuming non-random mating has a substantial effect on the outcomes (2,10). Empirical research (14) studied genetic variants associated with human height to assess the degree of height-related assortative mating in European-American and African-American populations. The study compared the inbreeding coefficient estimated using known height associated variants with that calculated from frequency matched sets of random variants. The results showed a significantly higher inbreeding coefficients for the height associated variants than from frequency matched random variants ($P < 0.05$), demonstrating height-related assortative mating in both populations.

Recent advances in quantitative genetics allow to access empirical information about the prevalence of mate selection mechanisms acting on a large set of genetic traits in natural populations. Specifically, data in the UK Biobank allowed to construct the frequency distribution of genetic relatedness of the parents of about half a million people living in Britain (4). Here I use this data to assess the extent to which assortative mating, assessed through genetic means, occurs in a human population

bioRxiv preprint doi: https://doi.org/10.1101/848911; this version posted November 23, 2019. 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-NC 4.0 International license.

METHODS

I used data gathered for another purpose (15-16) but which has the information required to build frequency distributions of the genetic relatedness between parents. The form of this frequency distribution allows to discern which of the two mate selection mechanisms, random or assortative, is more prevalent. Random mating should produce an inbreeding coefficients frequency distribution among individuals of a population that matches a truncated Poisson distribution with the median of the genetic variability of that population. Out-breeding should produce a frequency distribution where low inbreeding coefficients are more common. Assortative mating should favor higher inbreeding coefficients, than those expected from populations with only random mating, producing a skewed distribution with a fat tail towards high genetic relatedness. Extreme inbreeding is known to depress the fitness of the affected individuals (15) and should thus be rare.

Genotyped Single Nucleotide Polymorphism (SNPs) from Biobank were used to detect large runs of homozygosity (ROH). Runs of Homozygosity (ROH) are genomic regions where identical haplotypes are inherited from each parent. Once ROHs was detected, an inbreeding measure Froh was calculate for each individual as a measure of autozygosis by dividing the cumulated length of ROH in Mb by an estimate of the length of the human autosome (15).

The percentage of individuals born through assortative mating will affect the form of the frequency distribution of F. A theoretical distribution of the degree of assortment assuming 100% random mating was produced using Monte Carlo matching of two randomly chosen haploid DNA strands in a virtual population of 500 individuals each possessing 100 loci containing one of potentially 50 different alleles. The Model was a simplified version of virtual genomes where ROHs were represented by virtual alleles. Autozygosis was estimated in the simulation through the number of homologous allelic matches producing the inbreeding estimate “Frnd”. The frequency distribution of Frnd was normalized to the number of individuals in the sample (456414).

RESULTS

Two frequency distributions of genetic relatedness of parents are presented. The Frnd distribution of random matches of identical alleles obtained with a Mont Carlo model, mimicking the outcome if all parents of the individuals engaged in random mating; and the Froh distribution of the inbreeding coefficient obtained from data in Biobank. Both are plotted in Figure 1a and b. Clearly, the effect of assortation of the frequency distribution of Froh is evident when contrasting it with the distribution of Frnd simulating random mating.

bioRxiv preprint doi: https://doi.org/10.1101/848911; this version posted November 23, 2019. 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-NC 4.0 International license.

Figure 1. Percentage of the number of individuals from the Biobank database at F value intervals of 0.001 Froh units used as an inbreeding coefficient. The red continuous line represents the actual experimental data, the blue punctuated line are estimates for Frnd from simulations assuming random mating.

Figure 1 a: no data transformation

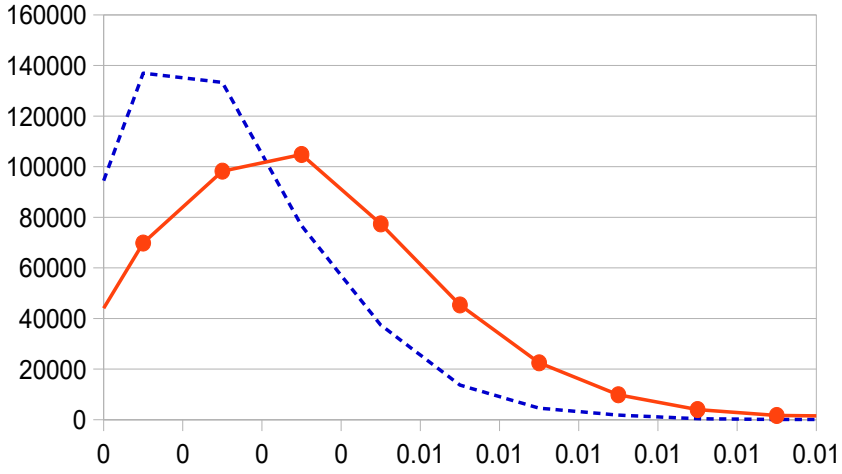
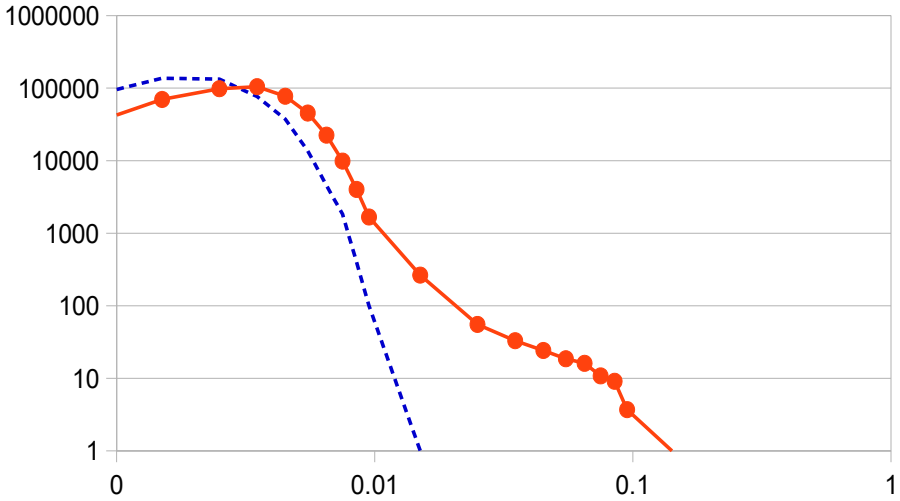


Figure 1 b: log-log transformation



The figure shows that for values of F > 0.009, the frequency of individuals diminishes markedly in the Biobank data confirming that extreme inbreeding is non-adaptive (3, 15).

DISCUSSION

A structural problem in comparing empirical and theoretical data is the matching of parameter values. In our case, the matching between Froh and Frnd. Independent evidence obtained by other means is required (see below) for more robust conclusions. Figure 1 minimize the effect of mismatch errors by presenting a log-log plot. The results show that the inbreeding coefficient from Biobank data between the parents of the individuals studied, produced a maximum peak at Froh 0.04, which coincides with the median value of the population. This value is far higher than that estimated by numerical simulations assuming random mating. Natural human populations have, important genetic homologies due to the fact that they configure a single species with a common ancestor. This fact is not captured by the random simulation model, as data from SNP was also chosen from parts of the genome showing important variance. But clearly, the frequency distribution of actual empirical F data is very different from

bioRxiv preprint doi: https://doi.org/10.1101/848911; this version posted November 23, 2019. 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-NC 4.0 International license.

the simulations which show no fat tail. In Figure 1, 30% of the population has higher F scores than predicted by simulations of random mating. This proportion is probably higher as not all genes relevant to assortation were included in this study. The evidence then is that assortative mating is an adaptive behavior. This means that ignoring assortative mating, as is common in population genetic studies, is not a rational choice.

Assortative mating can be achieved by passive and active means. Geographic isolation of sub-populations can achieve similar levels of assortment as that of active mate selection. If assortment is an adaptive feature of humans, however, individuals whose parents engaged in assortative mating will show higher F values and should have a higher fitness. A proof that this is assortation at work can be provided by comparing the average fitness of extreme out-breeders (Froh 0.001) with the random maters (Froh 0.004), those mating assortatively (Froh 0.02-0.03) and those extreme in-breeders (Froh > 0.1). This last group was shown to have phenotypic means between 0.3 and 0.7 standard deviation below the population mean for 7 traits, including stature and cognitive ability, consistent with inbreeding depression estimated from individuals with low levels of inbreeding (1). That is, offspring from assortative maters should have fitness values above the population mean. To demonstrate this, research focus on common fitness enhancing alleles. Science has now the tools to study the effect of homozygosis to understand genetic epistatic synergies favored by assortation (8) among humans and other living creatures.

References

1- Alvarez L, Jaffe K. Narcissism guides mate selection: Humans mate assortatively, as revealed by facial resemblance, following an algorithm of “self seeking like”. Evolutionary Psychology, 2: 177-194, 2004.
2- Chadeau-Hyam, M. Clive J Hoggart, Paul F O'Reilly, John C Whittaker, Maria De Iorio and David J Balding. Fregene: Simulation of realistic sequence-level data in populations and ascertained samples. BMC Bioinformatics 2008, 9:364
3- Clark D W et al. Associations of autozygosity with a broad range of human phenotypes. Nature Comm (2019) 10:4957
4- https://github.com/loic-yengo/Histogram-of-inbreeding-coefficients-in-the-UKB
5- Jaffe K, Chacon G. Assortative mating: Sex differences in mate selection for married and unmarried couples. Human Biol. 67: 111-120, 1995
6- Jaffe K. Extended Inclusive Fitness Theory: Synergy and assortment drives the evolutionary dynamics in biology and economics. SpringerPlus. 5(1) 1-19, 2016, 5:1092
7- Jaffe K. On sex, mate selection and evolution: an exploration. Comments on Theoretical Biology 7(2): 91-107, 2002
8- Jaffe K. On the adaptive value of some mate selection strategies. Acta Biotheoretica 47: 29-40, 1999
9- Jaffe, K. Synergy from reproductive division of labour and genetic complexity drive the evolution of sex. Journal of Biological Physics 44: 317-329, 2018.
10- Matthew C. Keller, Peter M. Visscher, and Michael E. Goddard. Quantification of Inbreeding Due to Distant Ancestors and Its Detection Using Dense Single Nucleotide Polymorphism Data. Genetics.111.130922; 2011.
11- Newberry, Mitchell G et al. “Assortative mating can impede or facilitate fixation of underdominant alleles.” Theoretical population biology vol. 112 (2016): 14-21. doi:10.1016/j.tpb.2016.07.003
12- Ochoa G, Jaffe K. Assortative mating drastically alters the magnitude of error thresholds. Lecture Notes in Computer Science LNCS 4193, Parallel Problem Solving from Nature - PPSN IX. Editors: Runarsson et al. Springer Verlag, Berlin (2006) p890-899.
13- Price GR (1971) Extension of the Hardy-Weinberg law to assortative mating. Ann Hum Genet 35(4):485–490
14- Xiaoyin Li, Susan Redline, Xiang Zhang, Scott Williams, and Xiaofeng Zhu. Height associated variants demonstrate assortative mating in human populations. Sci Rep. 2017; 7: 15689.
15- Yengo,L. Wray N. Visscher P.M. Extreme inbreeding in a European ancestry sample from the contemporary UK population. Nature Communications 10, 3719 (2019).
16. Yengo, L et al. Detection and quantification of inbreeding depression for complex traits from SNP data. Proc Natl Acad Sci U S A. 2017 Aug 8;114(32):8602-8607 (2017)

Acknowledgments: I thank Peter Visher and Loic Yengo Dimbou for access to the data empirical presented here