The Effect of Recent Admixture on Inference of Ancient Human Population History

Kirk E. Lohmueller, Carlos D. Bustamante and Andrew G. Clark

Copyright © 2010 by the Genetics Society of America
DOI: 10.1534/genetics.109.113761
FIGURE S1.—Expected SFS in a sample size of 24 chromosomes for Pop A and Pop AA under population growth when admixture occurs 20 or 7 generations ago. Note in both cases, the SFS when admixture occurs 20 generations ago is similar to that when admixture occurs 7 generations ago. (A) \( N_A = N_d \) and (B) \( N_A = 0.1 N_d \). Overall there is a slight excess of singletons when admixture occurs 7 generations ago as opposed to 20 generations ago, since there is less drift in Pop AA with the more recent founding. However, the number of singletons in Pop AA when \( N_A = 0.1 N_d \) is still less than that when \( N_A = N_d \). \( N_A = 20,000; N_d/N_A = 0.5; t_{cur} = 2400 \) generations.
Figure S2.—Expected HCN statistic for population growth (Pop A) and for growth with admixture (Pop AA) when admixture occurred 7 generations ago (instead of 20 generations). Each cell in the matrix is colored according to the proportion of simulation replicates (windows) having the particular configuration of the number of haplotypes and count of the most common haplotype. For example, red cells contain <2.5% of windows and medium-blue cells contain 20-22.5% of simulated windows. Note the excess of windows with fewer haplotypes and where the most common haplotype is at higher frequency in Pop AA relative to Pop A. This is most pronounced when $N_{AA} = 0.1N_A$. Note, the simulations assume $N_A = 20,000; N_H/N_A = 0.5; t_{sw} = 2400$ generations.
Figure S3.— Quantile-Quantile (Q-Q) plot comparing the chi-square goodness of fit $P$-values from data simulated from Pop A (x-axis) and Pop AA (y-axis). Note the excess of lower $P$-values in Pop AA relative to Pop A for both values of $N_{AA}$. These results suggest that the best-fitting growth parameters tend to fit Pop AA (where the true demographic model involves admixture) slightly worse than they do for Pop A (where the true demographic model is a growth model). Here $t_{cur} = 2400$ generations.
FIGURE S4.—Quantile-Quantile (Q-Q) plot comparing the chi-square goodness of fit test $P$-values from data simulated from Pop A (x-axis) and Pop AA (y-axis). Note that there is not as much of an excess of low $P$-values for Pop AA as there was in Supplementary Figure 2. Here $t_{\text{cur}} = 4000$ generations.
FIGURE S5.—The folded SFS for the Yoruba (YRI) and African American (AA) samples in the NIEHS data set. The folded SFS presents the number of SNPs where the minor allele has a given frequency. Note, to allow for missing data, we projected the SFS to a sample size of 20 chromosomes.
FIGURE S6.—Distribution of MLEs for the three growth parameters inferred using the SFS from Pop AA using different size data sets. “Original” denotes the original data sets presented in Figure 5 using a sample size of \( n = 24 \) chromosomes. “Filtered” denotes data sets consisting of a sample size of \( n = 100 \) chromosomes. Here we filtered SNPs so that the number of SNPs per data set was approximately equal to the mean number of SNPs in the original data sets. “Full” denotes the data sets consisting of a sample size of \( n = 100 \) without filtering any SNPs. Solid horizontal bars denote the true parameter values for Pop A (\( N_A = 20,000; N_B/N_A = 0.5; t_{\text{cur}} = 4000 \) generations).