

Supplementary Information

Inferring the probability of the derived *versus* the ancestral allelic state at a polymorphic site

Peter D. Keightley¹⁼ and Benjamin C. Jackson

Figure S1. Bias affecting element 19 of a 20 element uSFS as a function of the transition:transversion ratio (κ) simulated. Three different models were used to estimate the uSFS (JC, K2 and R6) and the results are compared to maximum parsimony. There were 360 replicates with 100,000 neutral sites simulated in each. Nucleotide diversity was 0.01. There were two outgroups. The divergence between the focal species and outgroup 1 was 0.1, and divergence between the common ancestor of the focal species and outgroup 1 and outgroup 2 was 0.1.

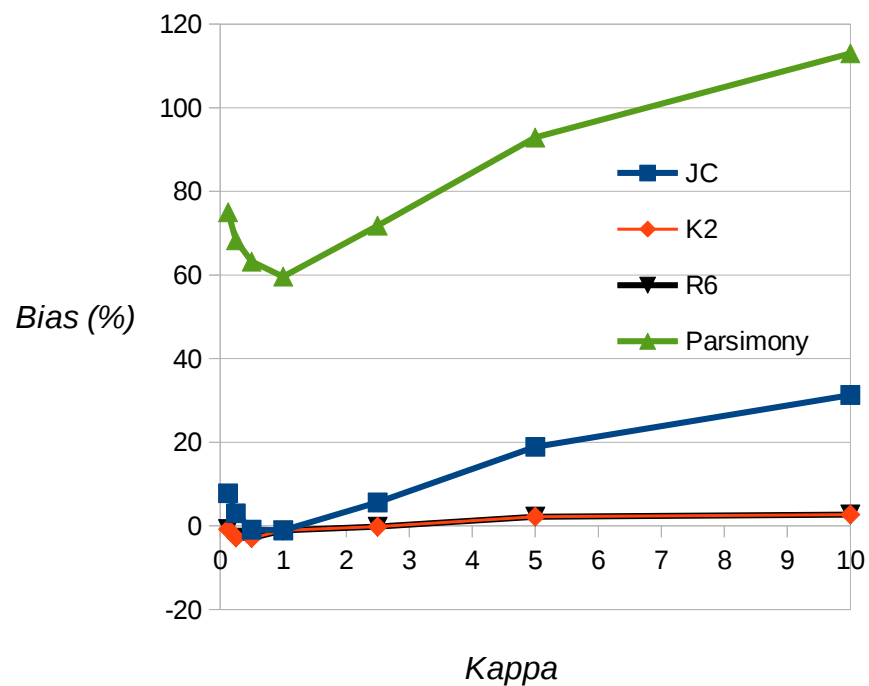


Figure S2. uSFSs for 0-fold degenerate sites of the LWK population (25 individuals) inferred by the 1000 Genomes Project and by the methods described in this paper using two outgroups for three nucleotide substitution models. (A) non-CpG sites. (B) CpG sites.

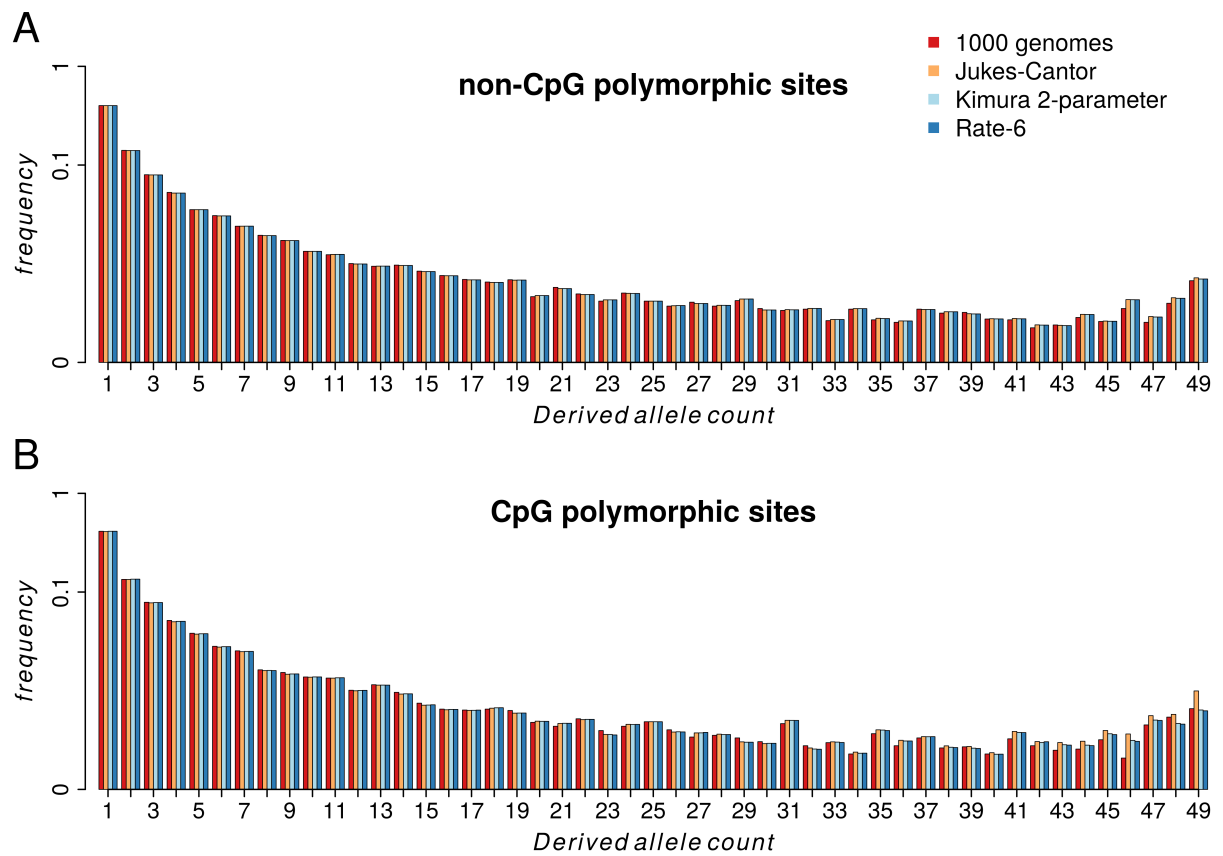


Figure S3. uSFs for 4-fold degenerate sites for the full sample of 99 individuals from the LWK population, inferred by the 1000 Genomes Project and by the methods described in this paper using two outgroups for three nucleotide substitution models. (a) non-CpG sites. (b) CpG sites. The methodology is as for Figure 7 in the main text.

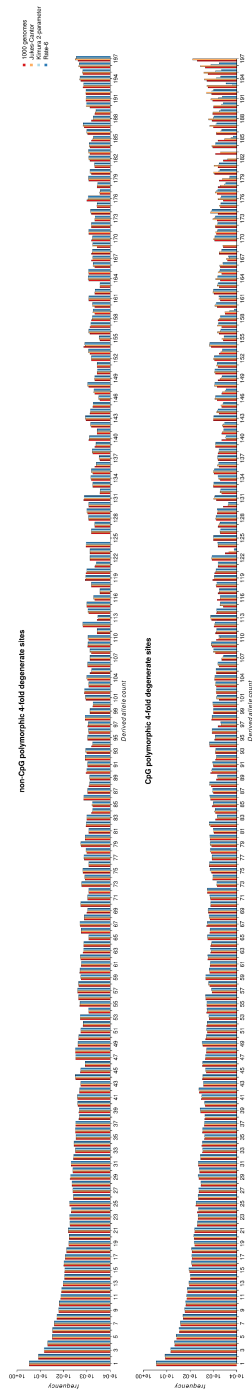


Figure S4. Comparison of uSFSs inferred using a single outgroup (orangutan) and two outgroups (macaque and orangutan) for 4-fold degenerate sites for a sample of 50 individuals from the LWK population of the 1000 Genomes Project. uSFSs were inferred using the R6 model. (a) non-CpG sites. (b) CpG sites.

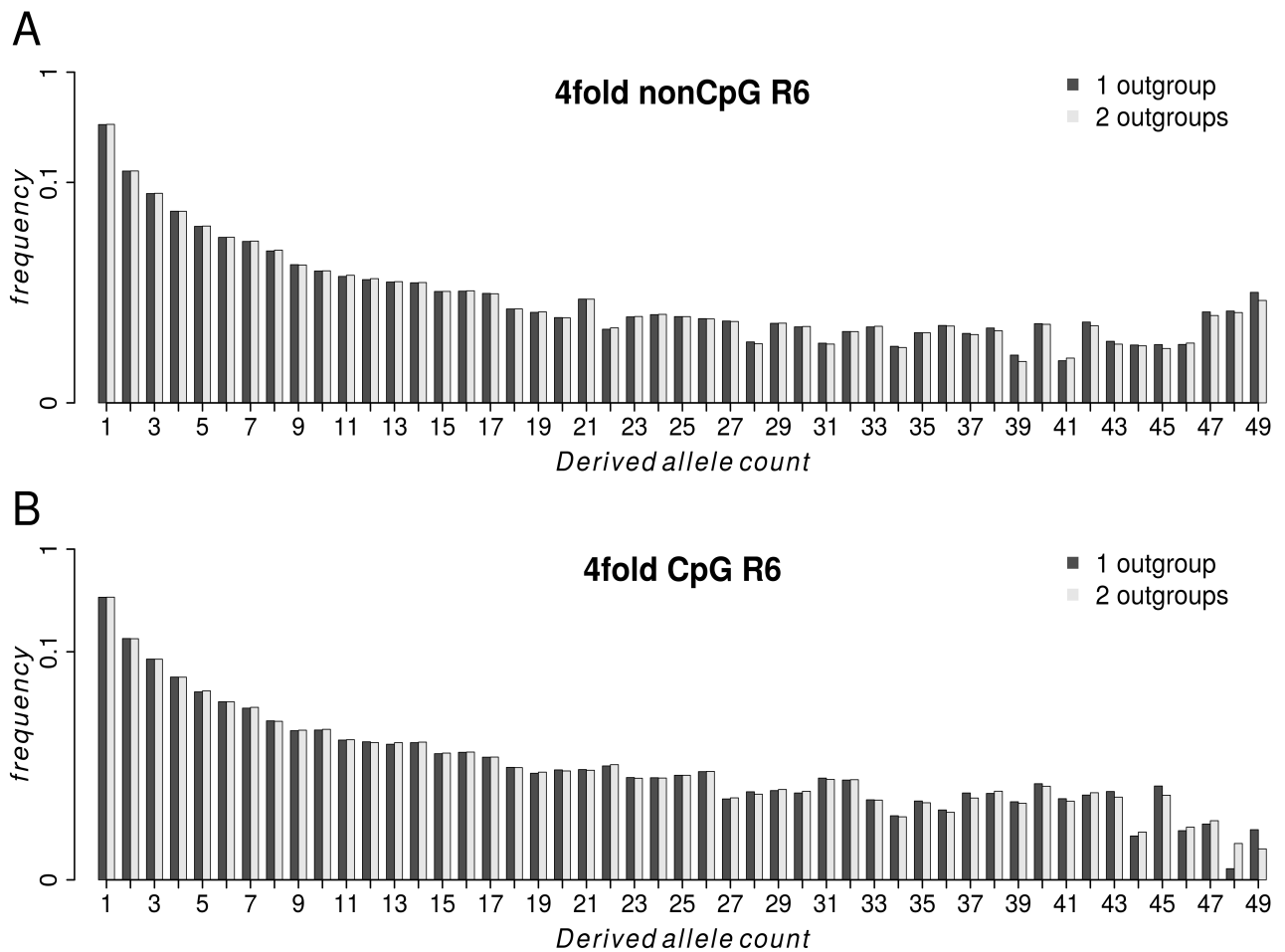


Table S1. Estimated effective population sizes (N) and times of population size change (t), expressed relative to an initial steady-state effective population size N_1 , obtained from a three-epoch demographic model fitted to SFSs for four-fold sites of DPGP phase 2 generated by the present method (assuming the R6 model) or by parsimony.

	N_2/N_1	t_2/N_1	N_3/N_1	t_3/N_1
Present method	0.40	1.0	0.90	0.22
Parsimony	0.10	0.22	0.60	0.29