Fig 1: Average assignment errors (*AAE*) as a function of the extent of unbalance in sample sizes when $K = 3$. Three populations with $F_{ST} = 0.1$ in the island model were simulated. The sample size was n_i for population i (i=1, 2, 3), with $n_2 \equiv n_3$ and $n_1 + n_2 + n_3 \equiv 200$. The extent of unbalanced sampling is measured by the sample size ratio, n_1/n_2 (*x* axis). The *AAE* (*y* axis) is shown for 100 replicate datasets analysed by Structure using L (=10, 20, 40, in figure legend) loci with each having 10 alleles, assuming either the default (Dft) or alternative (Alt) ancestry prior, K=3 populations, the admixture and correlated allele frequency models, and other parameter settings as described in the text.
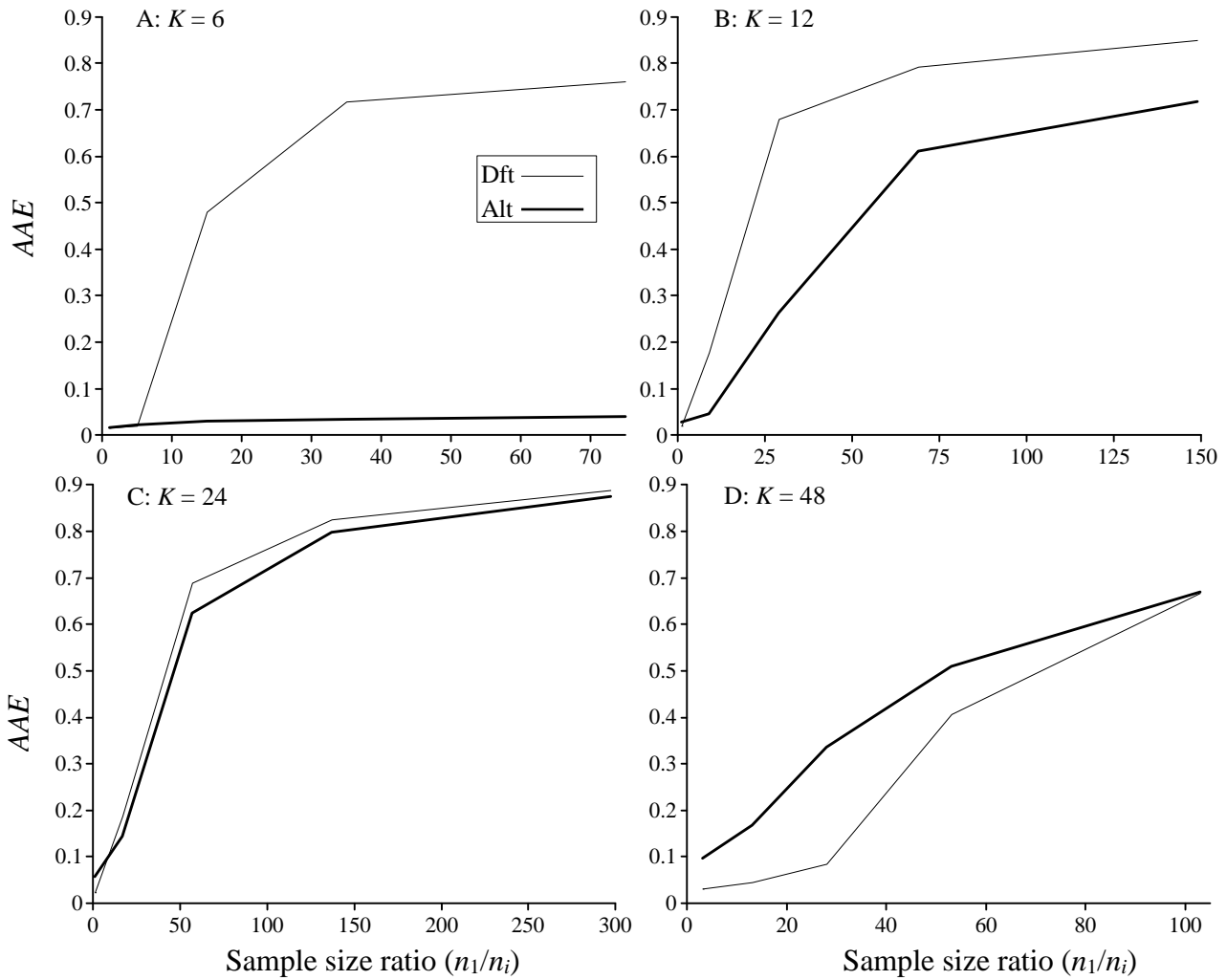
Fig 2: Average assignment errors (*AAE*) as a function of the extent of unbalance in sample sizes for a variable number of populations. A number of *K*=6, 12, 24 and 48 populations, shown in panels A, B, C and D respectively, with a $F_{ST} = 0.2$ in the island model were simulated. The sample size was n_i for population *i*, with $n_2 \equiv n_3 \equiv \ldots \equiv n_K$ and $\sum_{i=1}^{K} n_i \equiv 400, 800, 1600$ and 1500 when *K*=6, 12, 24 and 48, respectively. The extent of unbalanced sampling is measured by the sample size ratio, n_1/n_2 (*x* axis). The *AAE* (*y* axis) is shown for 100 replicate datasets analysed by Structure assuming either the default (Dft) or the alternative (Alt) ancestry prior, a burn-in length of 5×10^5 and a run length of 10^4 iterations, the admixture and correlated allele frequency models, and other parameter settings as described in the text.
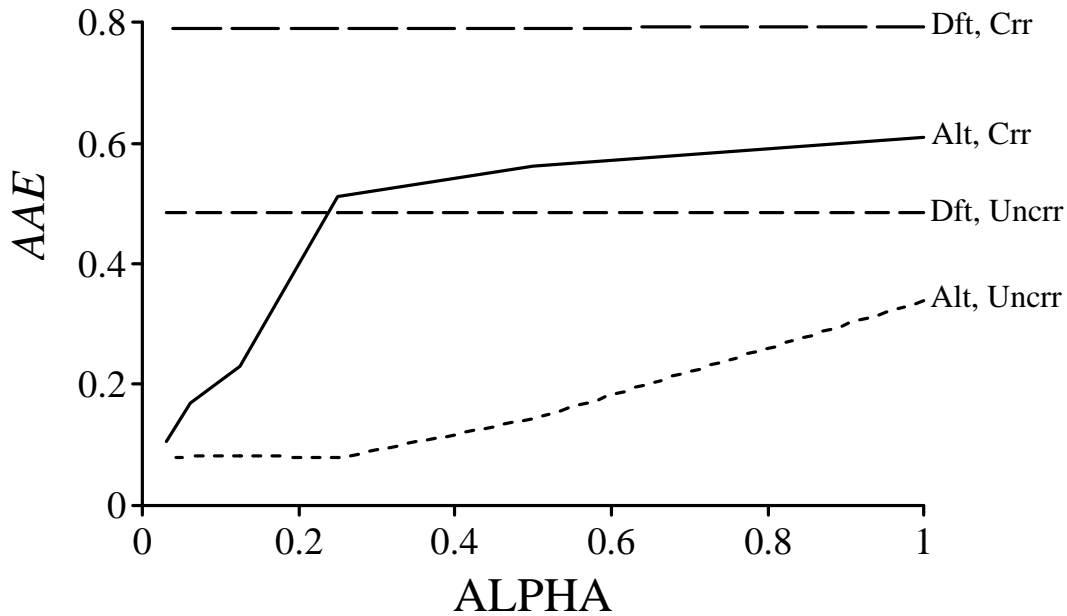
AAE

0.8 — — — — — — — — — — — Dft, Crr

0.6 — Alt, Crr

— Dft, Uncrr

0.4

- Alt, Uncrr

0.2

0

0   0.2   0.4   0.6   0.8   1

ALPHA

Fig 3: Average assignment errors (*AAE*) as a function of ALPHA values assumed in Structure analysis. A number of $K=12$ populations with a $F_{ST} = 0.2$ in the island model were simulated. The sample size was 690 for population 1, and 10 for each of the other 11 populations. The *AAE* (*y* axis) is shown for 100 replicate datasets analysed by Structure assuming either the default (Dft) or alternative (Alt) ancestry prior, either the correlated (Crr) or uncorrelated (Uncrr) allele frequency model, a burn-in length of 5×10^5 and a run length of 10^4 iterations, the admixture model, different ALPHA values (on the axis), and other parameter settings as described in the text.
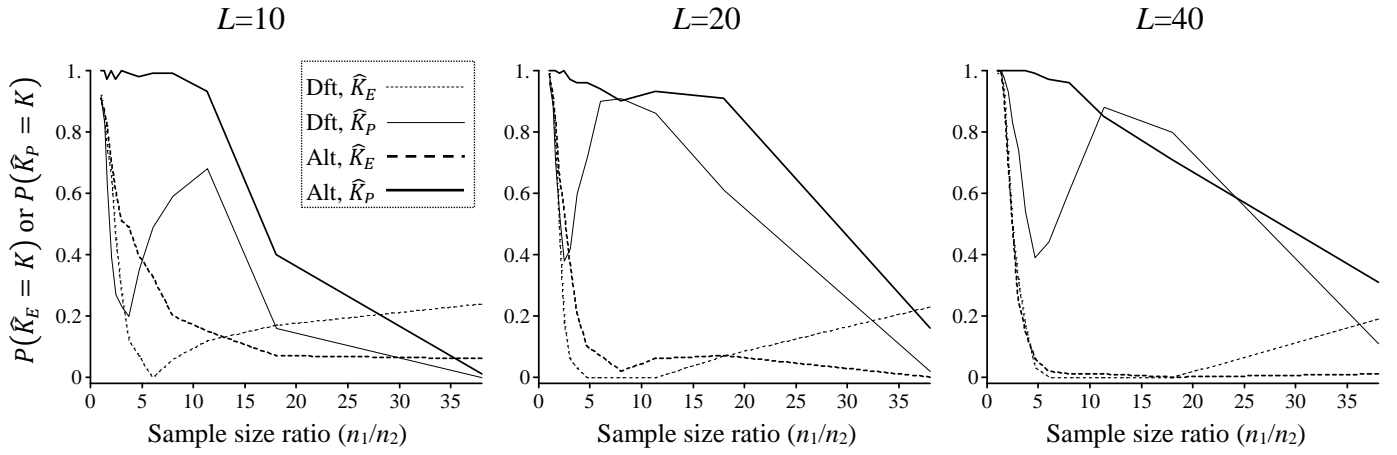
Fig 4: Proportion of correctly inferred number of populations ($P(\widehat{K}_E = K)$ and $P(\widehat{K}_P = K)$) as a function of the unbalance in sample sizes. Three populations with F_{ST}=0.1 in the island model were simulated. The sample size was n_i for population i (i=1, 2, 3), with $n_2 \equiv n_3$ and $n_1 + n_2 + n_3 \equiv 200$. The extent of unbalanced sampling is measured by the sample size ratio, n_1/n_2 (x axis). The accuracy measured by $P(\widehat{K}_E = K)$ or $P(\widehat{K}_P = K)$ (y axis) is shown for 100 replicate datasets simulated at each number of markers (L=10, 20, 40) and at each value of n_1/n_2 and analysed by Structure assuming either the default (Dft) or the alternative (Alt) ancestry prior, an assumed number of 1~5 populations, the admixture and correlated allele frequency models, and other parameter settings as described in the text.
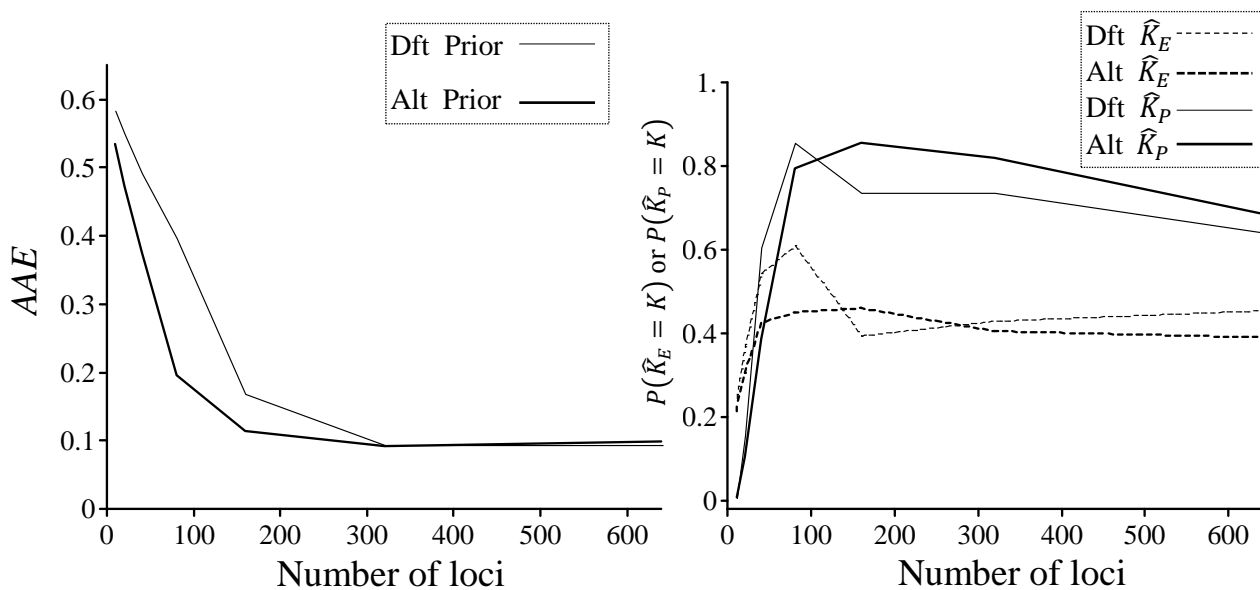
Fig 5: Comparison of Structure analysis results obtained with the default and alternative priors for a human dataset. The original dataset consists of 51 individuals from the Palestinian population in Israel, 13 individuals from the Colombian population in Colombia, and 24 individuals from the Mandenka population in Senegal, and each individual was genotyped at 783 microsatellite loci. A number of 200 datasets were obtained by bootstrapping without replacement from the original dataset, each dataset consisting 51 Palestinian (no bootstrapping), 5 Colombian and 5 Mandenka individuals, each individual at a given number of microsatellite loci (x axes). The left panel shows the AAE (y axis) at each number of loci for the 200 replicate datasets analysed by Structure assuming either the default (Dft) or alternative (Alt) ancestry prior, K=3 populations, and other parameter settings as described in the text. The right panel shows the proportion of correctly inferred number of populations ($P(\widehat{K}_E = K)$ and $P(\widehat{K}_P = K)$) as a function of the number of loci.