Table S1: Root mean square error (RMSE) for 6 HGDP data sets as a function of the regularization parameter.

α	0	1	10	100	1000
RMSE	0.046 [0.035,0.064]	0.044 [0.035,0.057]	0.041 [0.035,0.052]	0.041 [0.031,0.061]	0.055 [0.033, 0.095]

Table S2: Choice of K for `sNMF` using the cross-entropy criterion (simulated data).

	Dir(1, 1, 1)	Dir(.5, .5, .5)	Dir(.1, .1, .1)	Dir(.2, .2, .05)	Dir(.2, .2, .5)	Dir(.05, .05, .01)
$K = 2$	0.713	0.703	0.682	0.662	0.706	0.645
$K = 3$	**0.707**	**0.691**	**0.660**	**0.642**	**0.697**	**0.624**
$K = 4$	0.708	0.692	0.661	0.644	0.699	0.626
$K = 5$	0.710	0.694	0.663	0.645	0.700	0.628

Dir: Dirichlet distribution used to simulate "true" admixture coefficients using 3 ancestral populations.

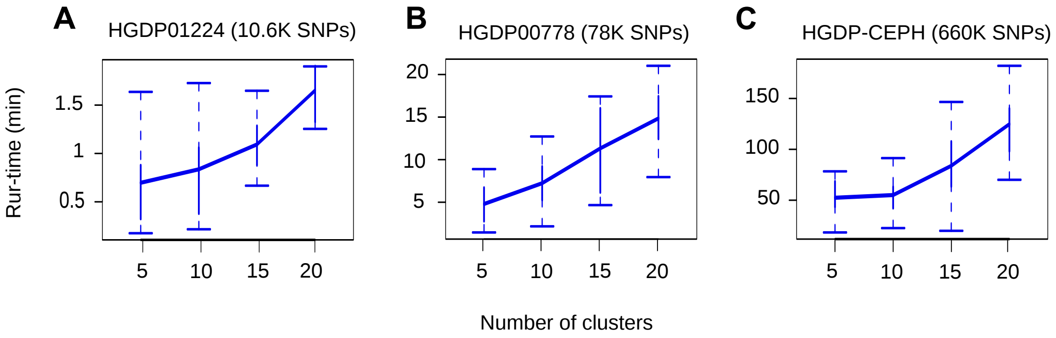

Figure S1. Run-times for `sNMF`. Time is expressed in unit of minutes. A) Run-time analysis for Harvard HGDP panel 01224 (10.6K SNPs). B) Run-time analysis for Harvard HGDP panel 00778 (78K SNPs). C) Run-time analysis for the HGDP-CEPH data (660K SNPs).

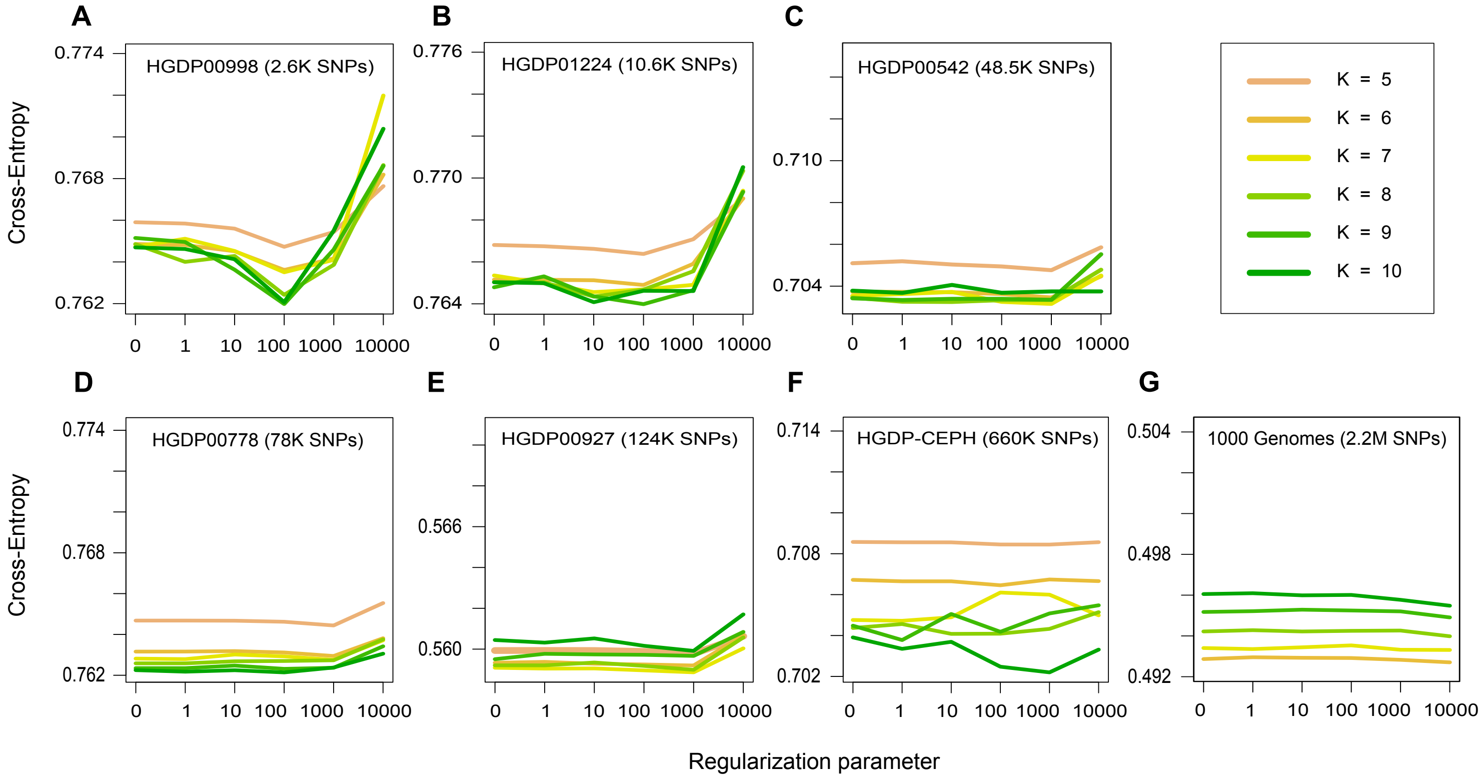

Figure S2. Values of the cross-entropy criterion for `sNMF` algorithms. Minimal values of the cross-entropy criterion over 5 runs of the `sNMF` program for A-E) 5 Harvard HGDP panel, F) the HGDP-CEPH data, and G) The 1000 Genomes Project data set. The number of clusters ranged from 5 to 10, and the values of the regularization parameter ranged from 0 to 10,000.

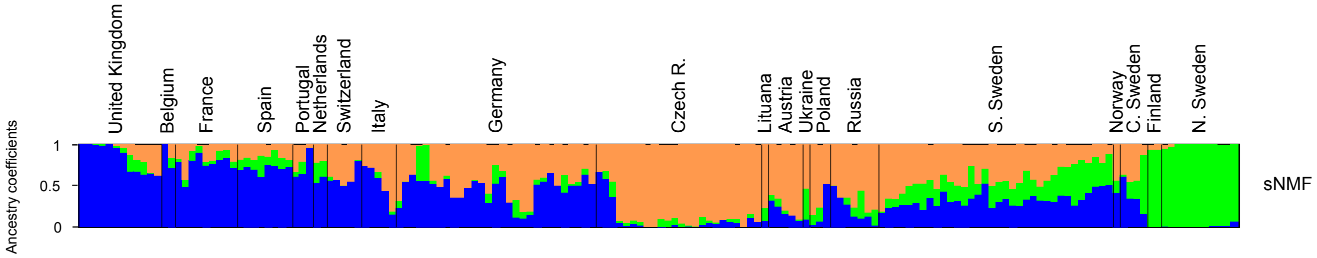

Figure S3. Graphical representation of admixture estimates for European populations of *A. thaliana*. Estimated admixture coefficients using `sNMF` using $K = 3$ and $\alpha = 100$ (cross-entropy $= 0.483$).

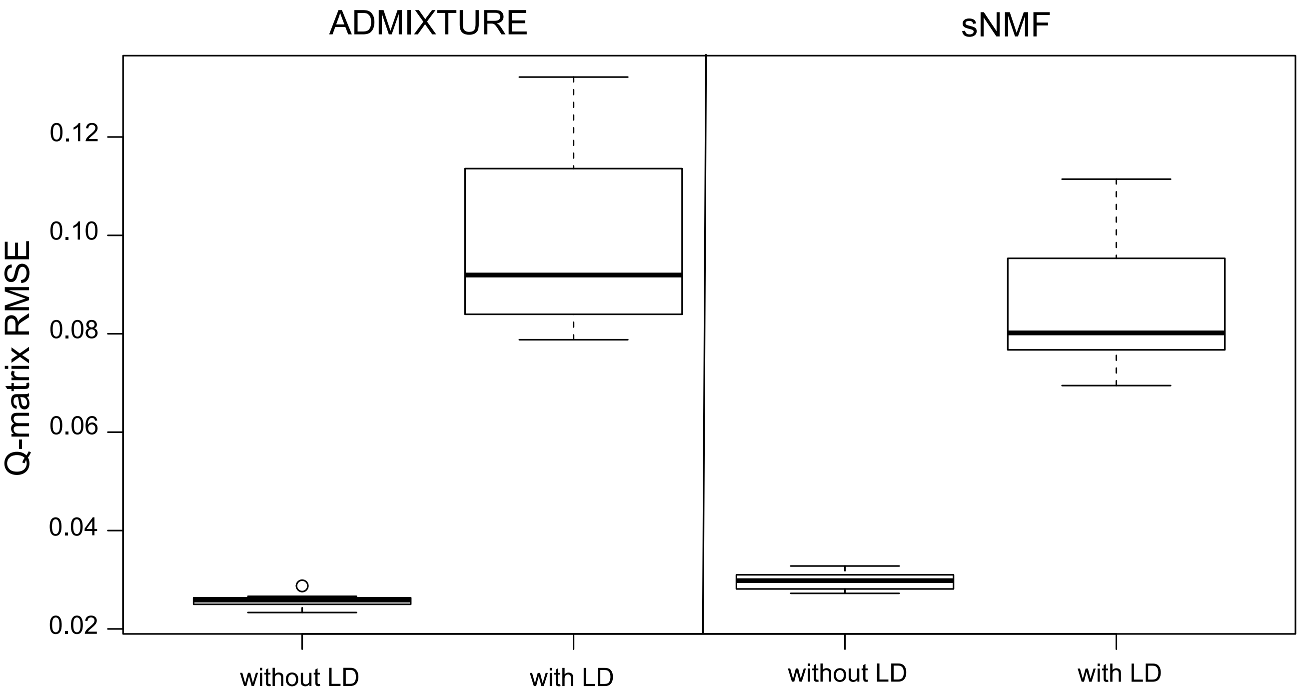

Figure S4. Accuracy of `ADMIXTURE` and `sNMF` in the presence of linkage disequilibrium. RMSEs between estimated Q matrices without and with linkage disequilibrium for `ADMIXTURE` and `sNMF` using $K = 5$ based on subsets of SNPs sampled from the 1000 Genome Project data set.

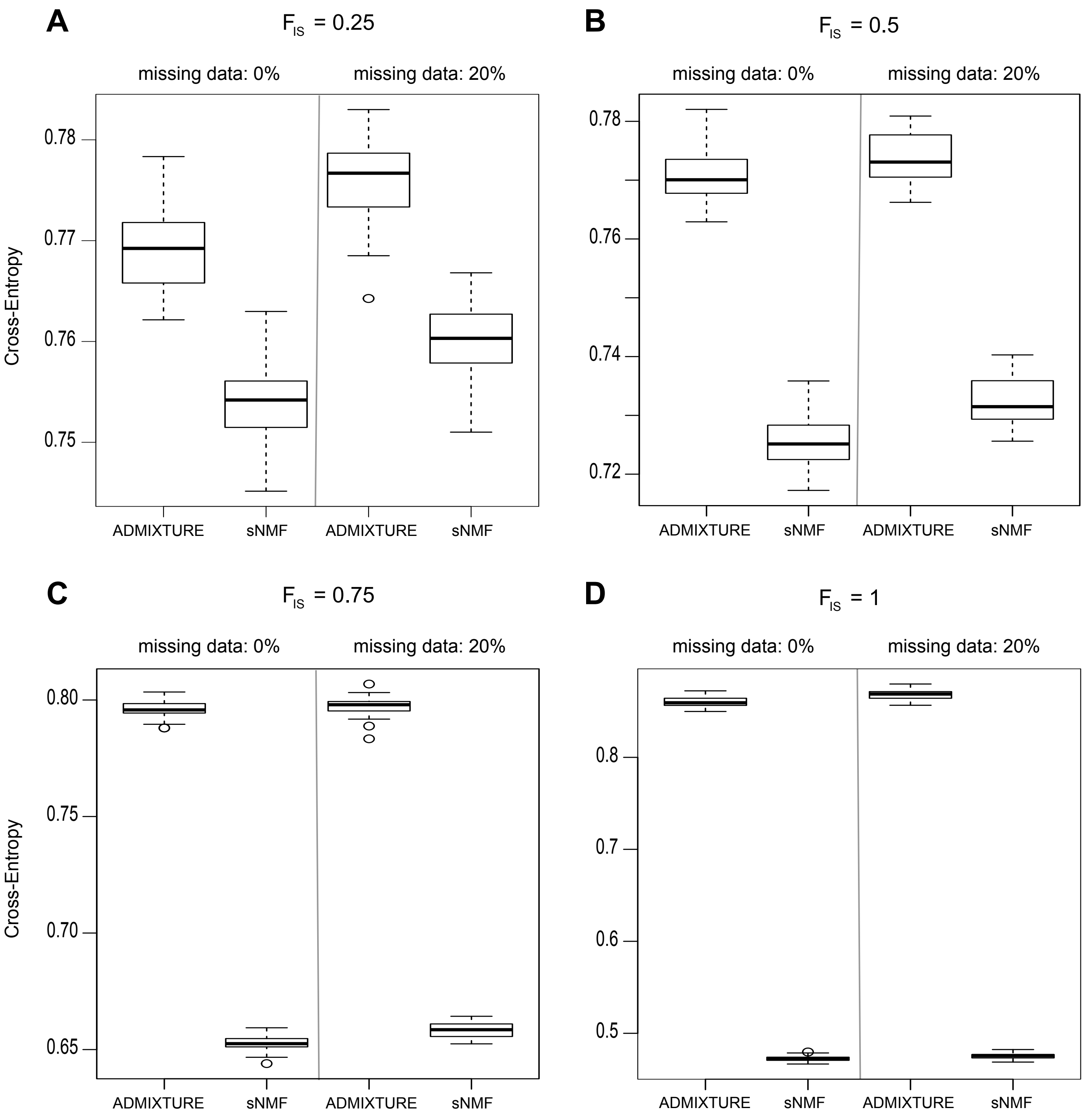

Figure S5. Accuracy of `ADMIXTURE` and `sNMF` in the presence of related individuals. Cross-entropy criterion for `ADMIXTURE` and for `sNMF`. Simulations mimicked the population structure of European populations of *Arabidopsis thaliana*. A-B) Moderate levels of inbreeding, $F_{\text{IS}} = 25 - 50\%$, C-D) Strong levels of inbreeding, $F_{\text{IS}} = 75 - 100\%$.