Supplementary figures and table for "Reconstructing the history of polygenic scores using coalescent trees"

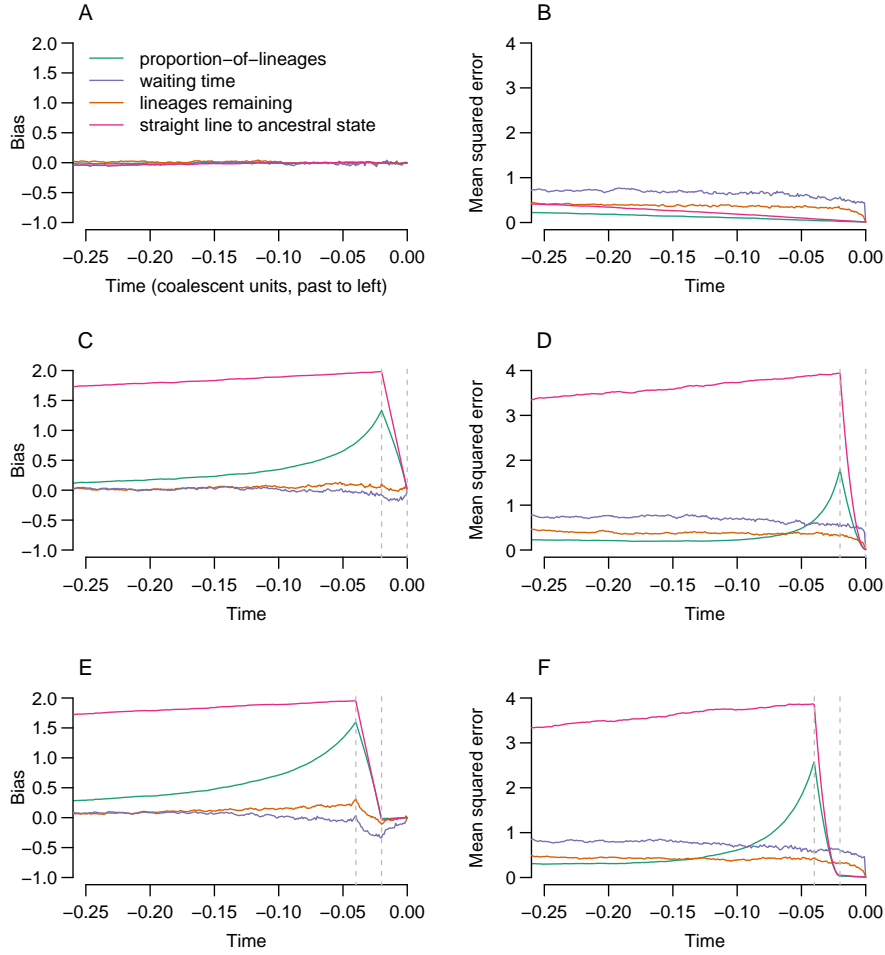Michael D. Edge & Graham Coop

October 2018

Figure S1: Bias and mean squared error (MSE) of the proposed estimators. Simulations and plotting conventions are the same as in Figure 4 of the main text, except that the polygenic scores include 1,000 rather than 100 independent loci, and all lines reflect the true trees—RENT+ was not used to infer trees at these loci. A thousand simulations were performed. In each trial, the polygenic scores were formed from 1,000 loci. The polygenic scores either evolved neutrally (A-B), or with selection leading to an approximate two-standard deviation shift in mean polygenic score (in present-day units), either occuring from .02 coalescent units ago to the present (C-D), or from .04 to .02 coalescent units ago (E-F).
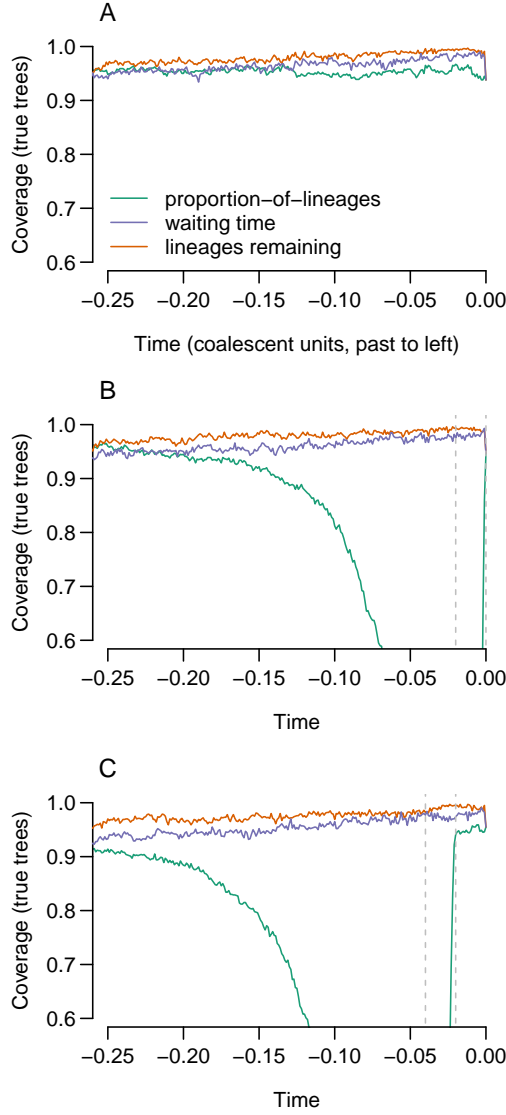
Figure S2: Confidence-interval coverage for nominal 95% confidence intervals based on the proposed estimators. Simulations and plotting conventions are the same as in Figure 5 of the main text, except that the polygenic scores include 1,000 rather than 100 independent loci, and all lines reflect the true trees— RENT+ was not used to infer trees at these loci. All confidence intervals were formed assuming an approximately normal distribution for the estimator, adding ±1.96 standard errors to the estimate. Standard errors were computed by taking the square root of the approximate variance of each estimator. Coverage probabilities are based on 1,000 simulations. Simulations were conducted under either neutrality (A), an approximate two-standard-deviation shift over the last .02 coalescent units (B), or a two-standard-deviation shift from .04 to .02 coalescent units ago (C).

3

		proportion-of-lins		waiting-time		lins-remaining	
		χ^2	perm.	χ^2	perm.	χ^2	perm.
neutral	true trees	.054	.052	.004	.048	.049	.055
sel. .02-now	true trees	1	1	.012	.075	.147	.224
sel. .04-.02	true trees	.985	.977	.017	.076	.034	.078

Table S1: Power/Type I error (with $\alpha = .05$) of various implementations of the T_X statistic in the simulations shown in Figures S1-S2, with 1,000 loci included in each polygenic score and 200 chromosomes sampled in the present. T_X was computed using allele-frequency estimates $0, .01, .02, ..., .1$ coalescent units before the present.