

Supplementary Material of "Loter: A software package to infer  
local ancestry for a wide range of species"

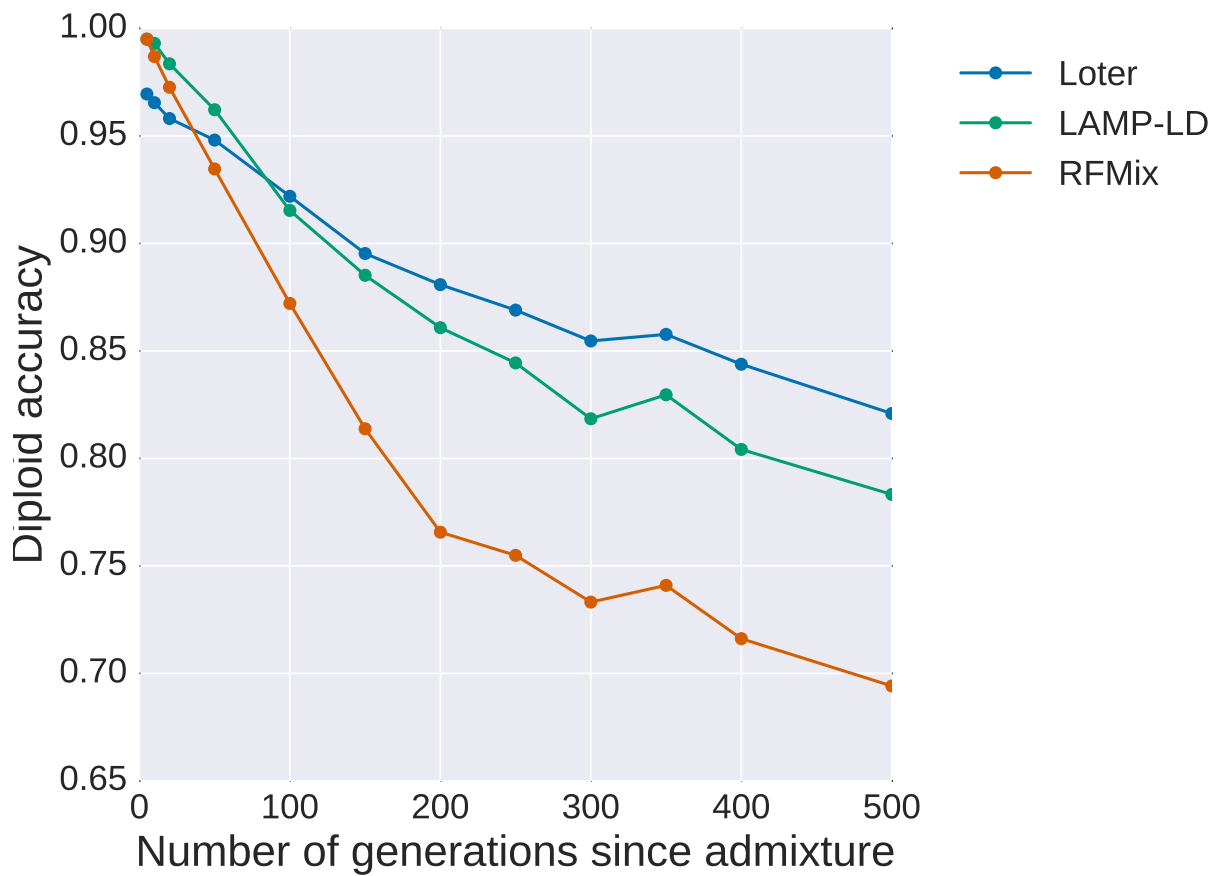


Figure S11: Diploid accuracy obtained with LAMP-LD, Loter, and RFMix when haplotypes of the reference panel are not drawn from the source populations but from populations that have diverged from the source populations. Admixed individuals are simulated by constructing their genomes from a mosaic of true African (YRI) and European (CEU) haplotypes and reference panels are the Luhya sample from Kenya (LWK) and the Toscani sample from Italy (TSI) (?). For performing simulations, true haplotypes are obtained using trio information. For local ancestry inference, haplotypes are obtained with Beagle using individuals that are not used for simulating admixed individuals. For each value of the number of generations since admixture, 1 set of 48 admixed individuals are generated.

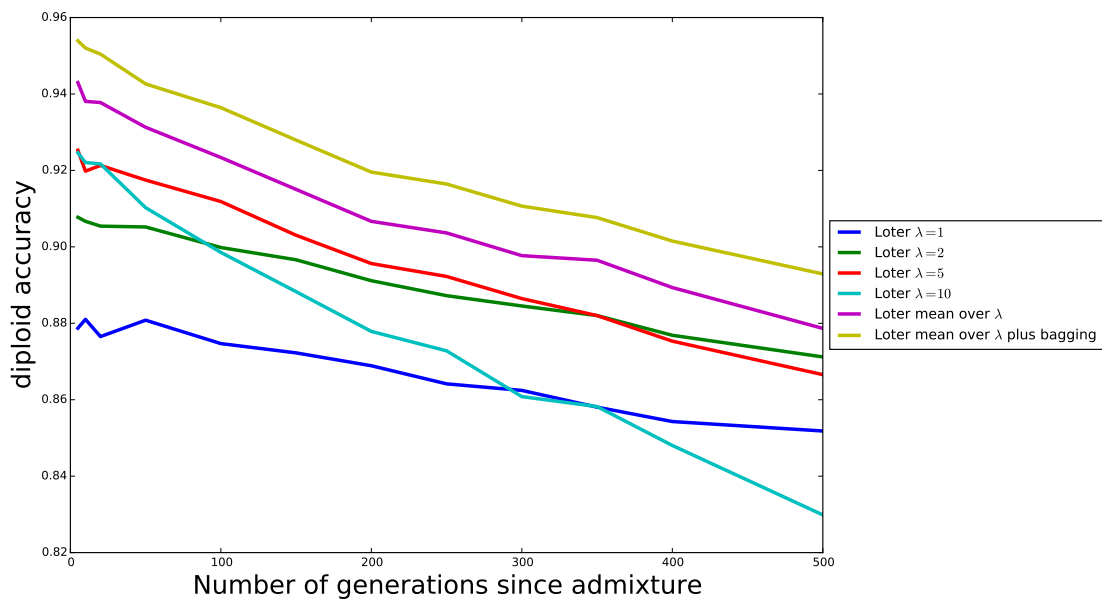


Figure SI2: Diploid accuracy obtained with Loter is improved when bagging and when averaging over multiple values of the regularization parameter. Admixed individuals were simulated by constructing their genomes from a mosaic of true African (YRI) and European (CEU) haplotypes (?) (Figure ??). Diploid accuracies are evaluated for twelve different values of the admixture time corresponding to 5, 10, 20, 50, 100, 150, 200, 250, 300, 350, 400 and 500 generations.

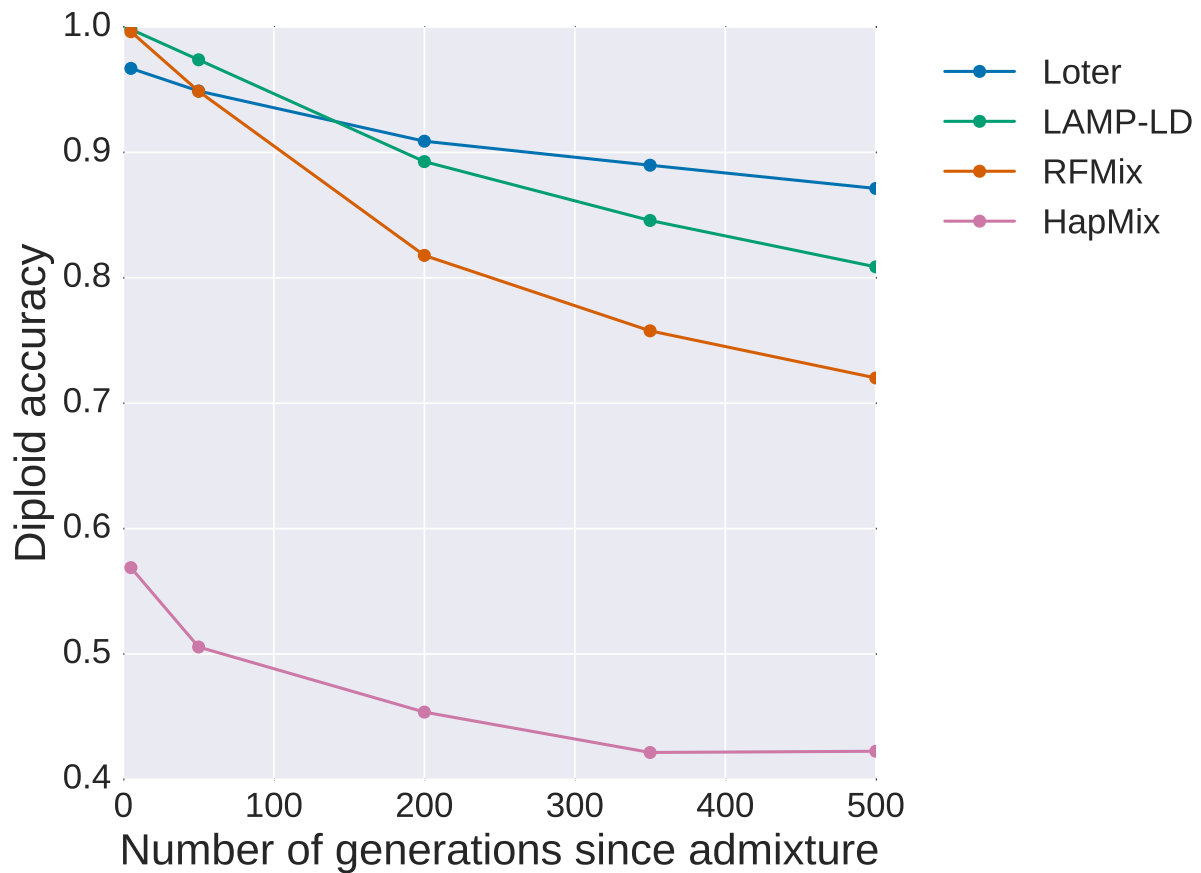


Figure SI3: Diploid accuracy obtained with LAMP-LD, Loter, and RFMix for simulated human individuals as a function of the time since admixture occurred. Admixed individuals are simulated by constructing their genomes from a mosaic of true African (YRI) and European (CEU) haplotypes (?) (Figure ??). For each admixture time, HAPMIX is evaluated using a single simulation of 48 admixed individuals. Other software, which run faster, are evaluated based on the mean diploid accuracy obtained with 20 simulated sets of 48 admixed individuals.

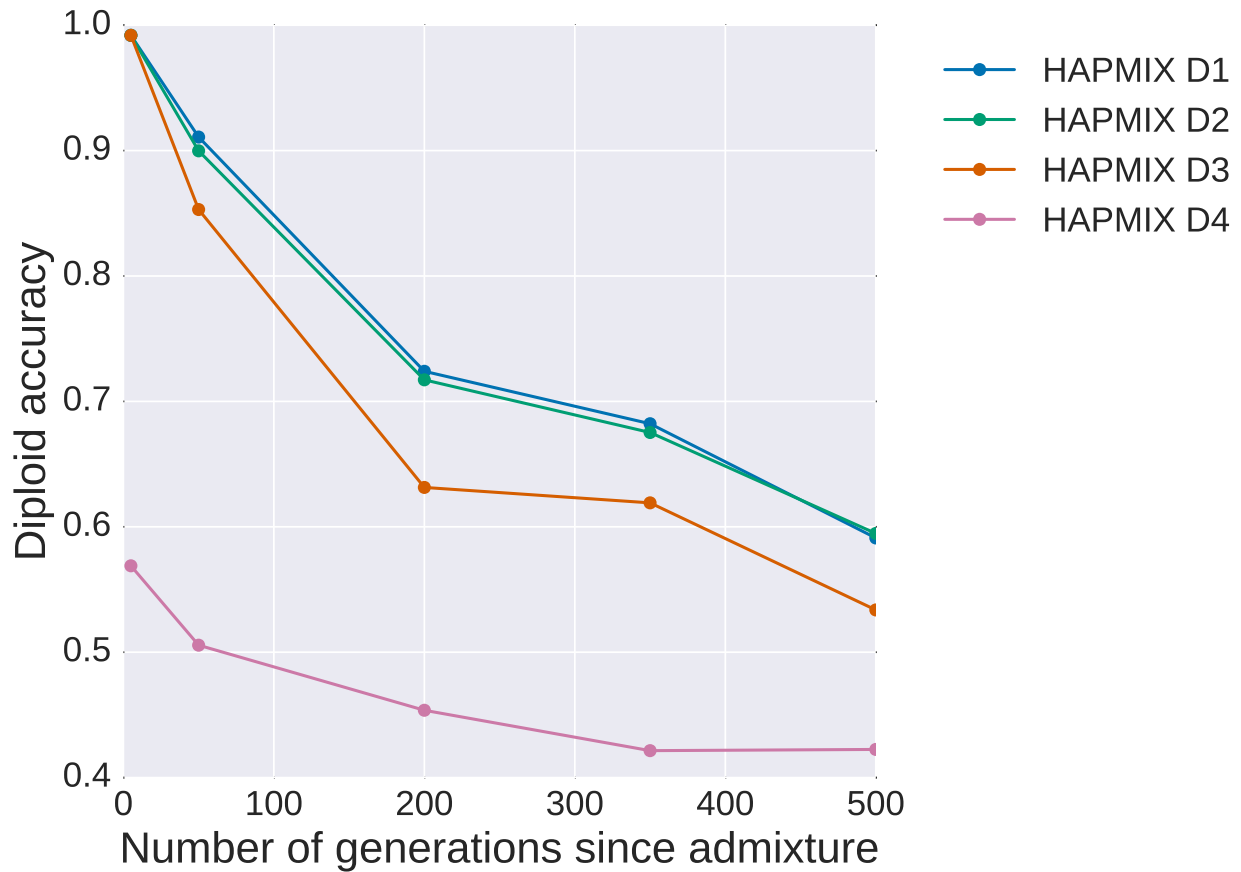


Figure SI4: Diploid accuracy obtained with HAPMIX using four different haplotype sets for LAI. The diploid accuracy of HAPMIX is severely reduced when considering reconstructed haplotypes of admixed individuals instead of true haplotypes. Admixed individuals are simulated by constructing their genomes from a mosaic of true African (YRI) and European (CEU) haplotypes (?) (Figure ??). In set D1, we consider the same true haplotypes (trio-phased) for simulations and inference. In set D2, we consider different haplotypes for simulations and inference but haplotypes are all trio-phased and admixture time is assumed to be known. The set D3 is the same as D2 except that admixture time is unknown and assume to be equal to 6 generations. In set D4, haplotypes used for inference are not true haplotypes but have been reconstructed with Beagle.

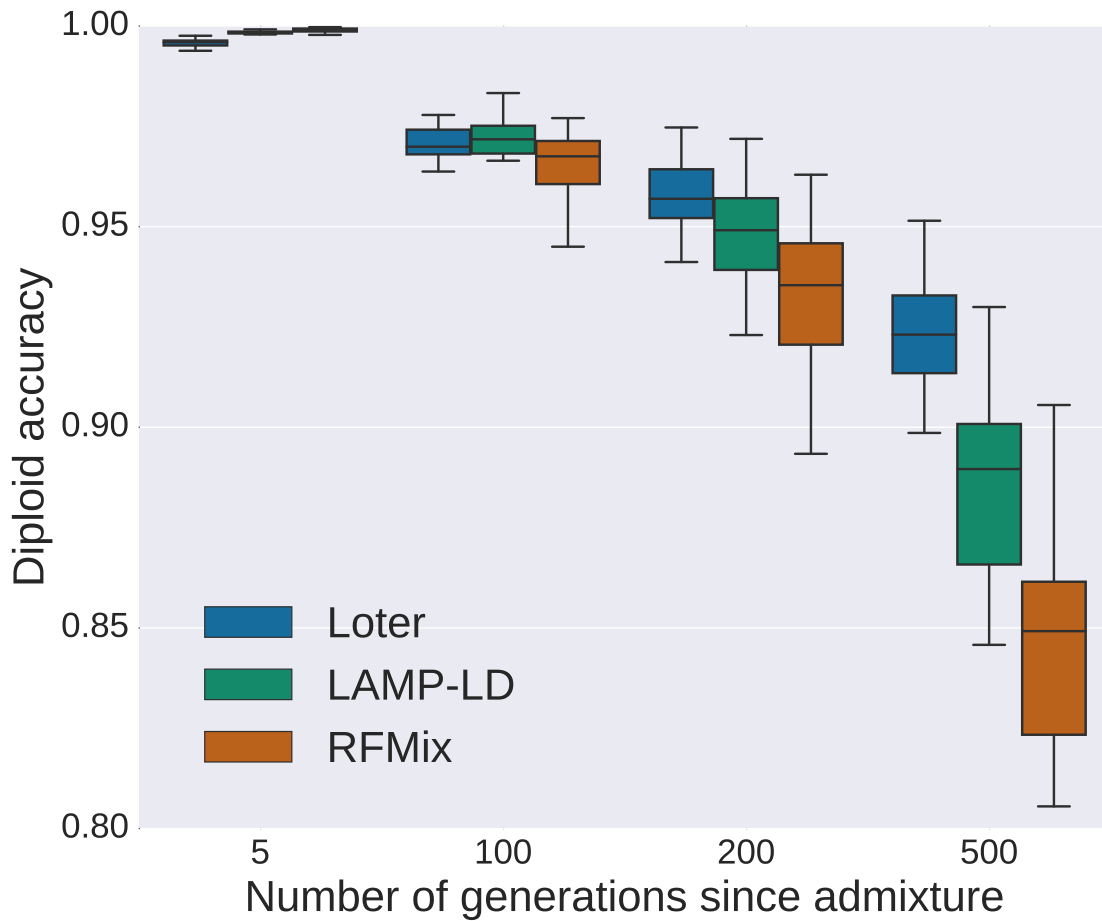


Figure SI5: Diploid accuracy obtained under a 3-way admixture model with LAMP-LD, Loter, and RFMix for simulated admixed human individuals as a function of the time since admixture occurred. Admixed individuals are simulated by constructing their genomes from a mosaic of true African (YRI), European (CEU), and Chinese haplotypes (?). For performing simulations, true haplotypes are obtained using trio information. For local ancestry inference, haplotypes are also reconstructed using trio-based inference and are different from haplotypes used for simulations. For each value of the number of generations since admixture, 20 sets of 20 admixed individuals are generated. Boxplots show the distribution of the 20 values for the mean diploid accuracy.

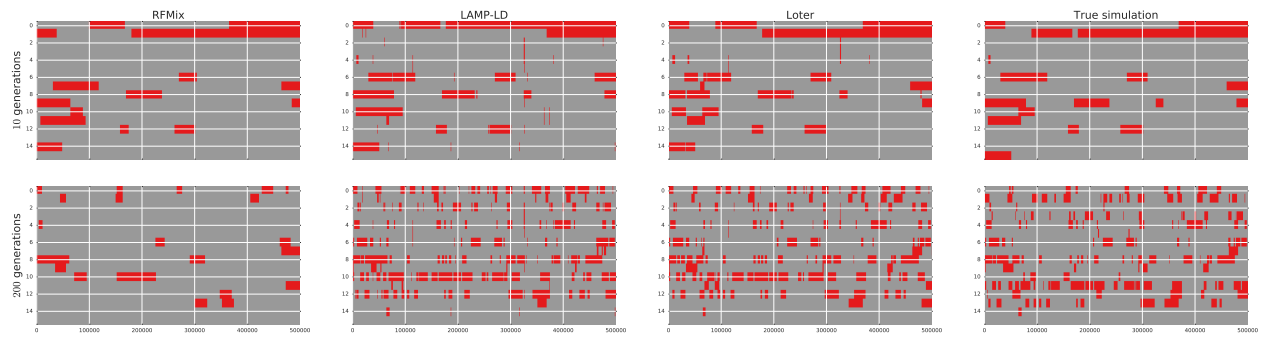


Figure SI6: Ancestry tracts for 20 simulated admixed *Populus* individuals. Grey chunks correspond to *P. trichocarpa* chunks and red chunks correspond to *P. balsamifera* chunks. Two rows correspond to the two haplotypes of a single individual. Ancestry switches between haplotypes are caused by haplotype phasing using Beagle. The presence of spurious and small ancestry chunks contribute to excessively decrease the median length of ancestry chunks in LAMP-LD and Loter.

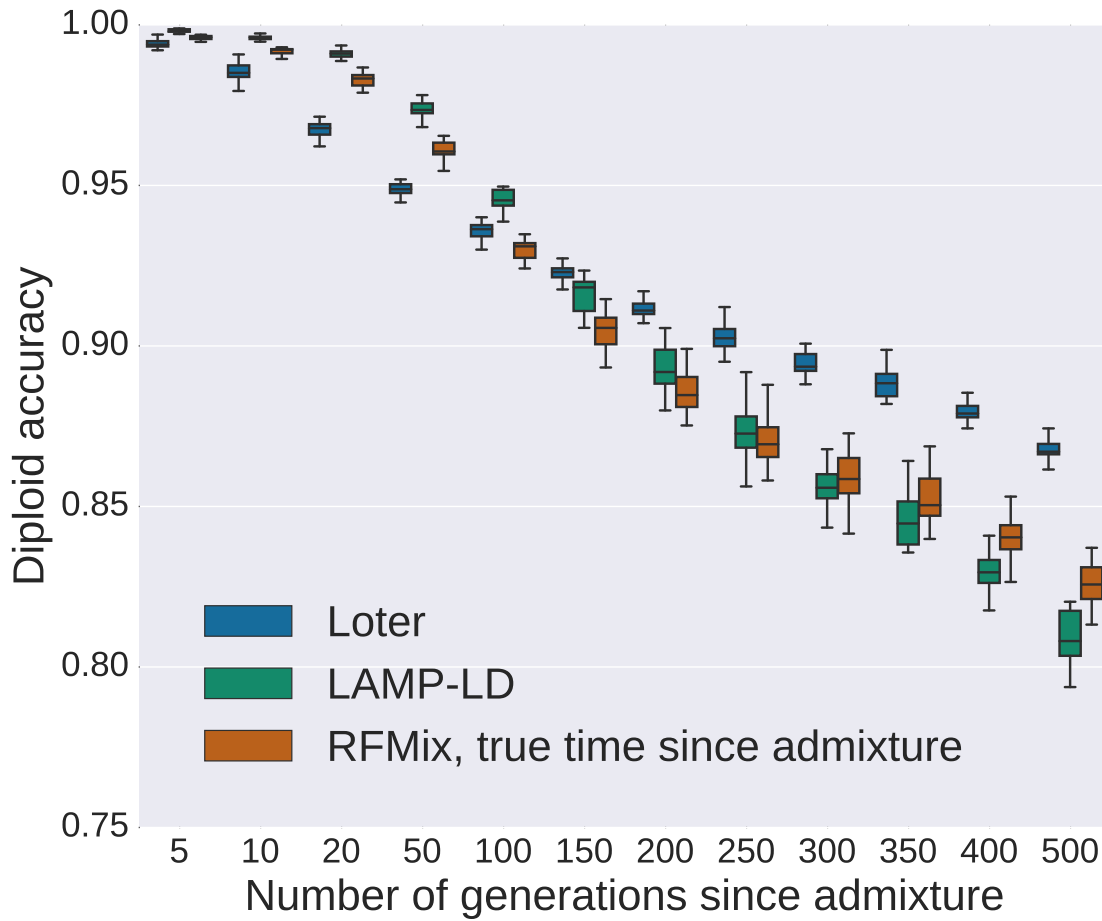


Figure SI7: Diploid accuracy obtained with LAMP-LD, Loter, and RFMix for simulated admixed *Populus* individuals as a function of the time since admixture occurred when true values of the time since admixture are provided to RFMIX. Admixed individuals are simulated by constructing their genomes from a mosaic of *Populus trichocarpa* and *Populus balsamifera* individuals. Individuals are phased using Beagle and two different sets of individuals are used for performing simulations and inference. For each value of the number of generations since admixture, 20 sets of 20 admixed individuals are generated. Boxplots show the distribution of the 20 values for the mean diploid accuracy.