

Table S1: Pairwise F_{ST} between ancestral groups and Vietnamese set. The ancestral groups contained wild individuals with one ancestry accounting for > 80%. Pairwise F_{ST} was computed using 1.1M random genome-wide SNPs. All p-values were < 0.01.

Group	OB	ER	D	C	AG
ER	0.22				
D	0.55	0.54			
C	0.43	0.40	0.44		
AG	0.43	0.41	0.50	0.39	
Vietnam	0.23	0.01	0.54	0.41	0.41

Table S2: Run time and memory usage of ELAI runs with different sets of parameters and SNP numbers

		Time (h)			Memory (Gb)		
		c			c		
snp	mg	5	15	25	5	15	25
10k	5	1.9	12.4	46.2	0.1	0.9	2.5
	10	1.9	7.2	32.6	0.1	0.9	2.5
	20	1.6	12.2	28.8	0.1	0.9	2.5
	30	1.3	7.6	27.7	0.1	0.9	2.5
100k	5	20.7	164	551	1.2	9.1	24.8
	10	19.2	155	327	1.2	9.1	24.8
	20	18	104	399	1.2	9.1	24.8
	30	16.8	103	309	1.2	9.1	24.8
even ~11k	5	1.7	8.7	41.5	0.1	1.1	2.9
	10	2	14.5	53.7	0.1	1.1	2.9
	20	1.6	13.4	45	0.1	1.1	2.9
	30	1.6	13.5	40.8	0.1	1.1	2.9
*all ~1.1M	5	166.6	1372.8	4462.2	13.6	104	281.6
	10	164.5	1,020	3749.3	13.6	104	281.6
	20	151.4	1309.2	4396.4	13.6	104	281.6
	30	141.6	976.8	4270.8	13.6	104	281.6

* estimated from benchmark of all SNP chunks

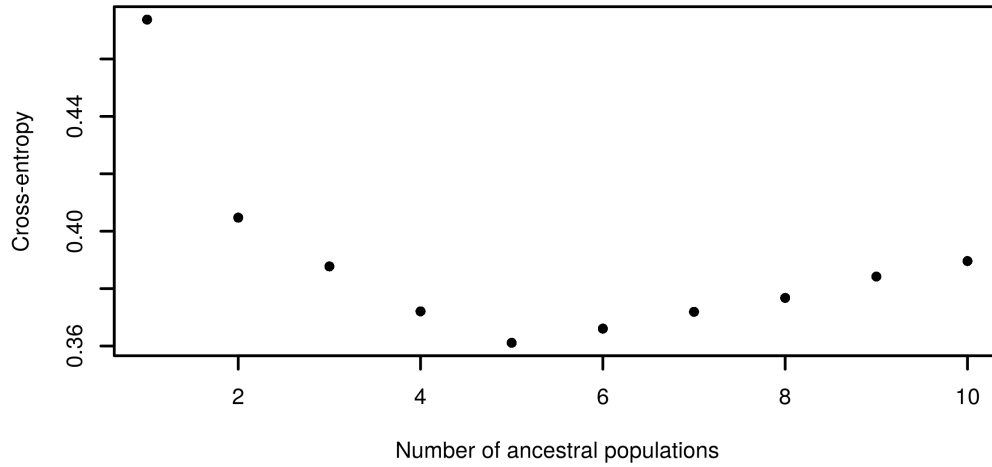


Fig. S1: Cross-entropy criterion for each tested number of ancestral groups (K) on the set of African references and Vietnamese accessions. Cross-entropy values of sNMF runs, with $K = 1$ to 10 and 10 iterations for each K , using 1.1M random genome-wide SNPs, were plotted. Lower cross-entropy implies a better prediction capacity.

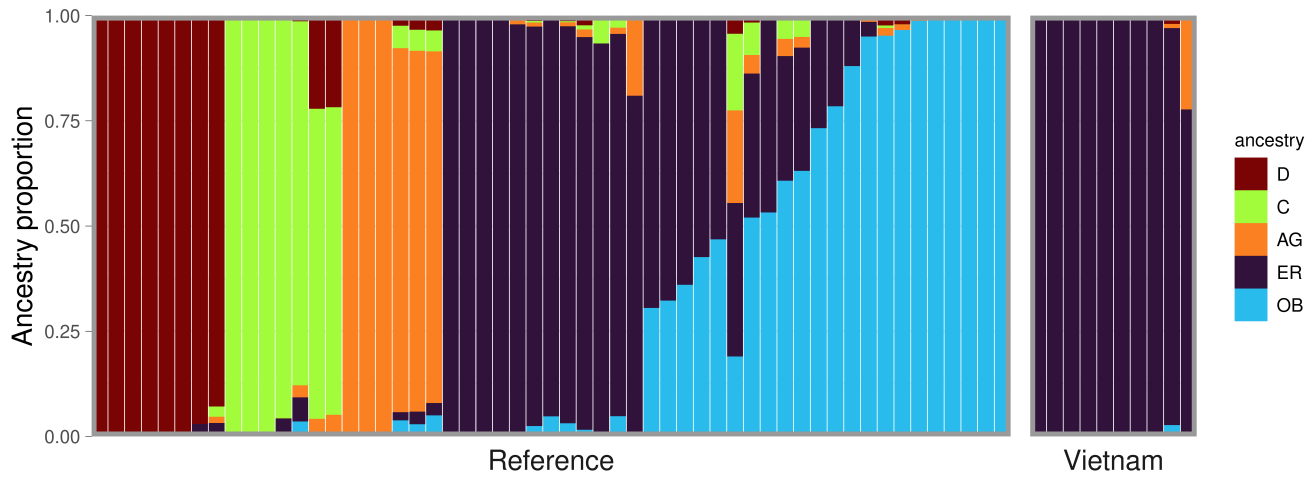


Fig. S2: Ancestry proportions for each African reference individual and tested Vietnamese individual. Five ancestral groups were estimated from the best sNMF run using 1.1M random genome-wide SNPs based on cross-entropy criteria. Proportions of ancestral groups (inferred by color) per individual are presented in each vertical bar. The ancestries were named according to the genetic groups associated with the geographical distribution of the African individuals which had > 80% ancestry proportions: group D (red) from Guinea, Ghana and Côte d’Ivoire, C (green): Cameroon, AG (orange): Gabon and Angola, OB (blue): Uganda and CAR, and ER (purple): DRC. The group names were as proposed previously in Tournebize et al. 2022.

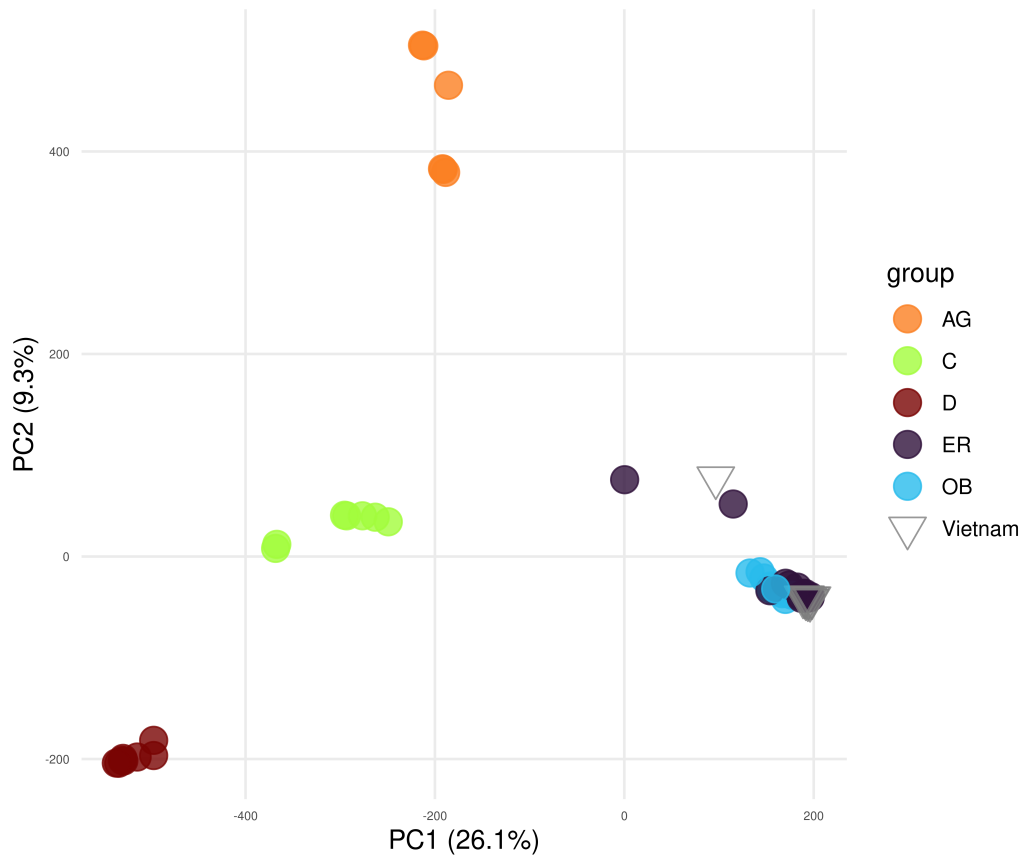


Fig. S3: Projection on PC axes 1 and 2 of 55 African wild accessions and 10 Vietnamese individuals using 1.1M random genome-wide SNPs. The African individuals were labeled by the most contributing ancestry based on the sNMF results. Group ER is in purple, OB in blue, C in green, AG in orange, and D in red), and Vietnamese individuals are in gray dots.

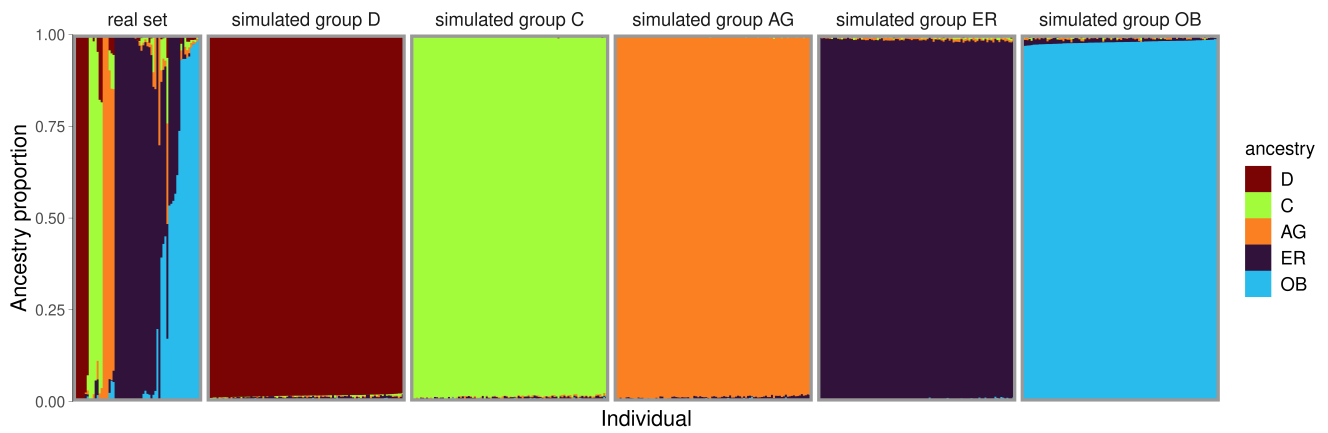


Fig. S4: Individual ancestry proportions of five simulated ancestral groups compared to the real dataset. Each of the five simulated groups consisted of 100 genotypes, corresponding to the ancestral groups as classified in Result 1. The barplot shows the proportion of the five ancestries (group ER in purple, OB in blue, C in green, AG in orange, and D in red) per real and simulated individual, estimated by sNMF using random 100K SNPs on chromosome 1. The average ancestral coefficient in the simulated individuals was 0.99 ± 0.004 .

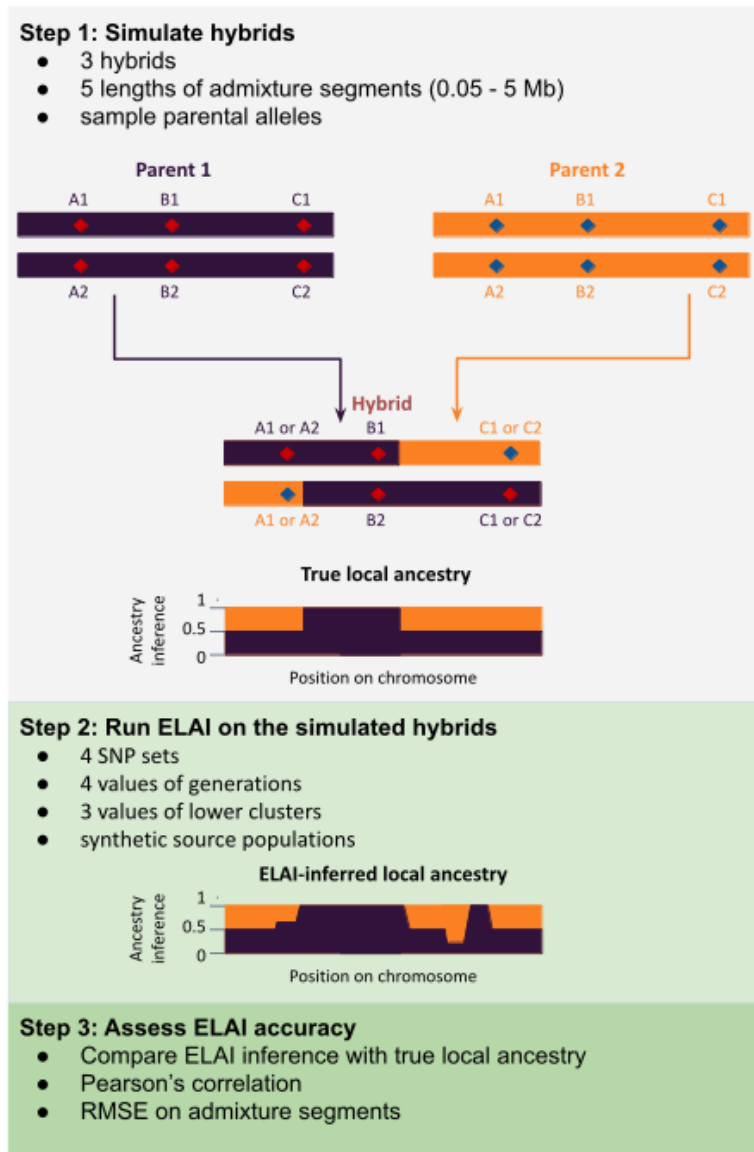


Figure S5: Hybrid simulation steps and test ELAI performance. Step 1: creating 3 hybrids with different admixture patterns of lengths 50 kb, 500 kb, 1 Mb, 2 Mb and 5 Mb, by sampling alleles from two chosen parents. Step 2: running ELAI on the simulated hybrids with different sets of parameters and SNPs, using the synthetic source populations. Step 3: comparing the local ancestry inferred by ELAI and the true inference by Pearson's correlation and assessing the accuracy in known admixture segments by RMSE.

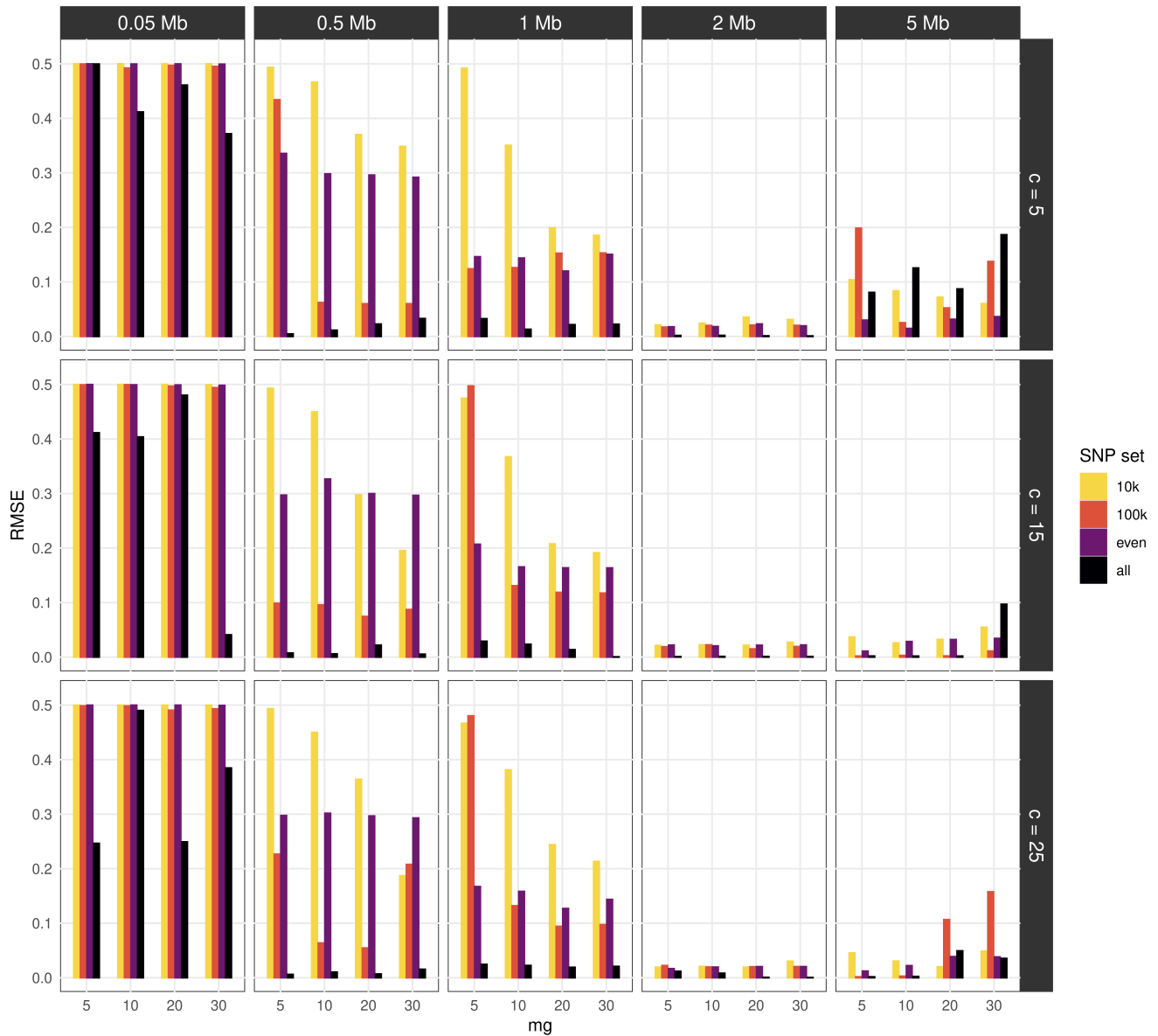


Fig. S6: ELAI errors in detecting true introgression tracts of different lengths. RMSE were computed between the true dosage and ELAI dosages of the simulated hybrids at introgression tracts differed by lengths (0.05, 0.5, 1, 2, and 5 Mb). The panel columns show results of different admixture tracts, and the rows show results in different numbers of lower clusters (5, 15, 25). Each plot presents the RMSE values on the y-axis for ELAI runs with different numbers of generations (5, 10, 20, 30) on the x-axis, and 4 SNP sets (10K SNPs, 100K SNPs, evenly distributed SNPs – 1 SNP/5 kb, and all SNPs - 1M SNPs, represented in yellow, orange, violet and black points, respectively).

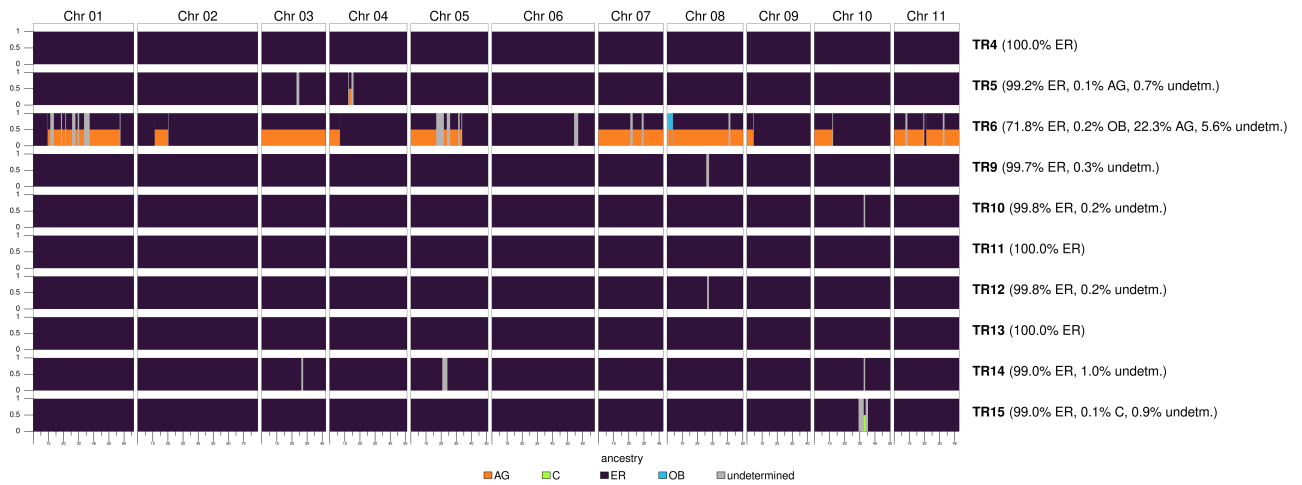


Fig. S7: Local ancestry inference of the 10 Vietnamese accessions. Each row presents the consensus local ancestry dosage (y-axis) along the positions (x-axis) in each genome. The x-axis labels are in the Mb unit. Each row presents the consensus local ancestry dosage (y-axis) along the positions (x-axis) in each genome. The inferred ancestral groups - ER (from Democratic Republic of Congo), OB (from Uganda and Central Africa), AG (from Benin, Gabon, and Angola), and C (from Cameroon and western Central Africa), are denoted by dark purple, blue, and orange, respectively. Uncolored (gray) portions are the undetermined regions. Global ancestry proportion, computed from the local ancestry, for each individual is shown on the label.