Supplementary Figures and Tables

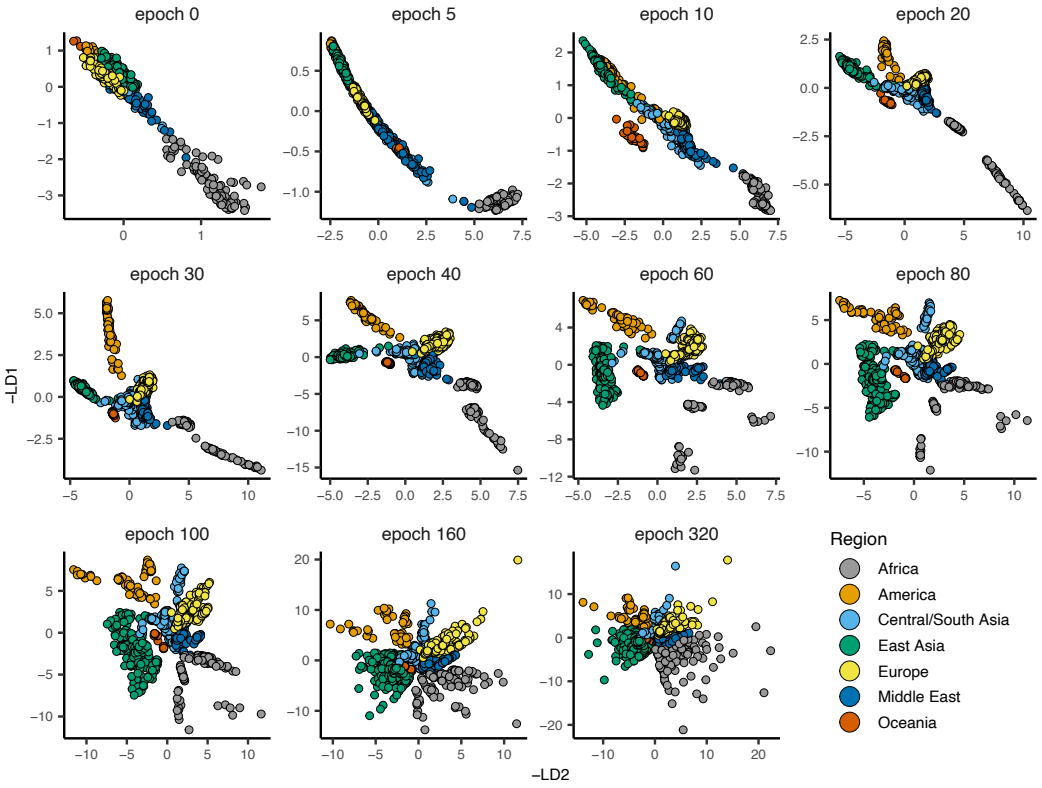

Figure S1: Latent spaces output during model training for HGDP data. Here `patience` was increased to 500 to show the overfitting behavior of `popvae`'s latent space. In this run validation loss was minimized at epoch 59.

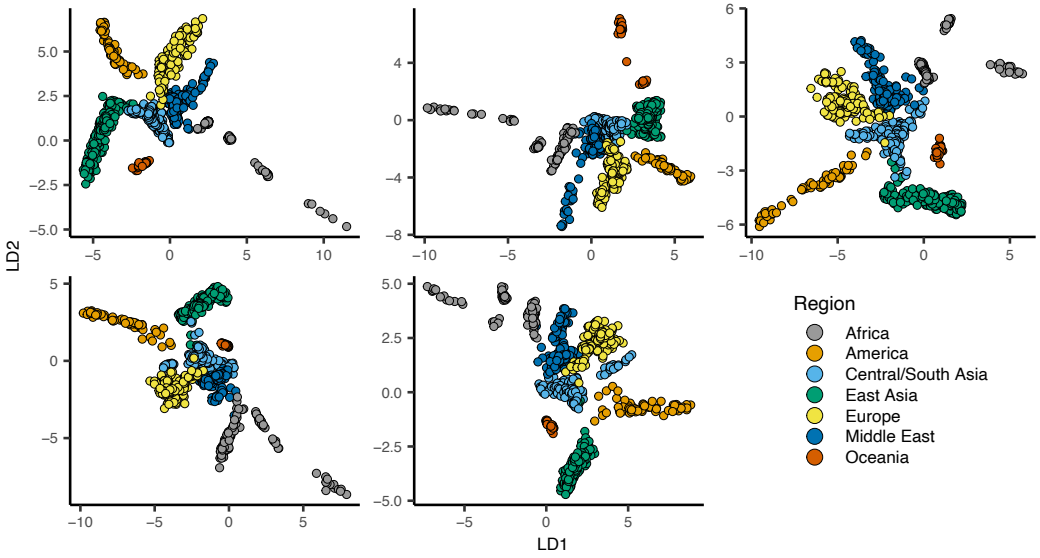

Figure S2: `popvae` latent spaces from runs with default hyperparameters and different random seeds.

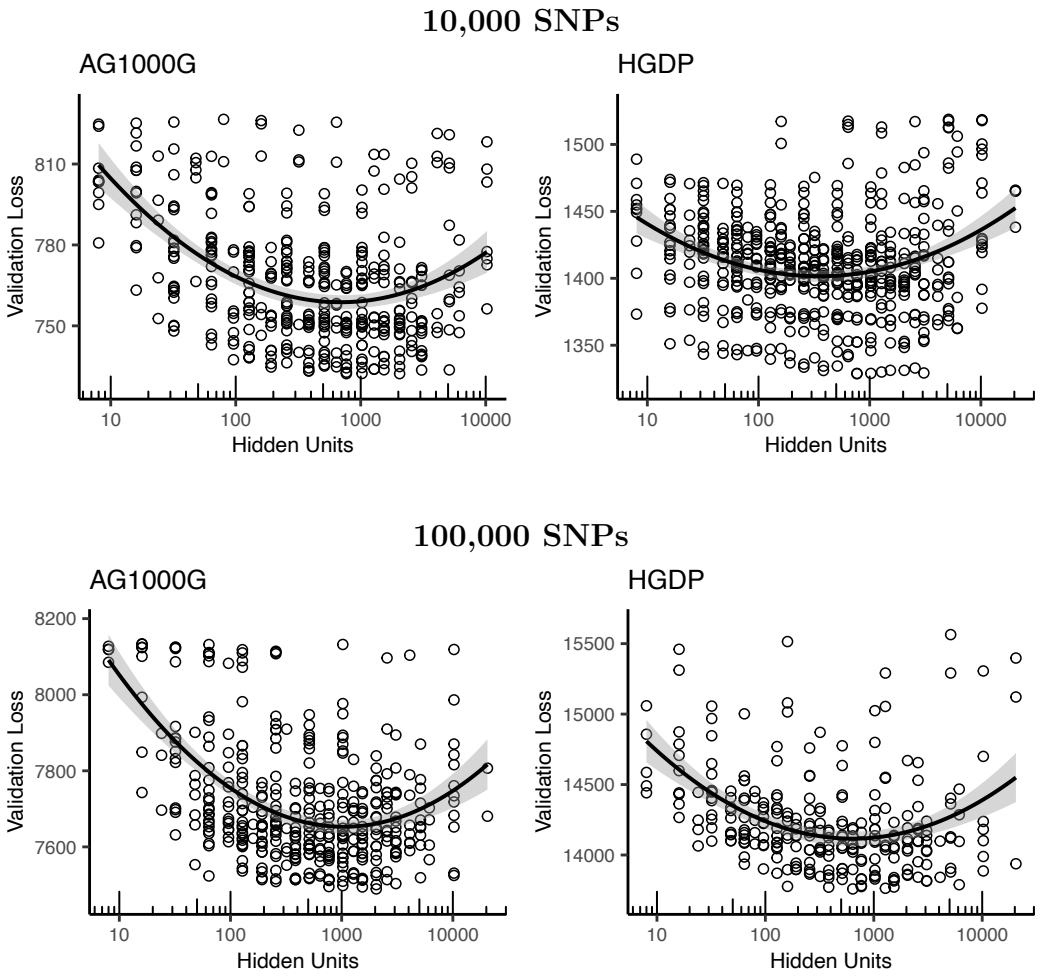

Figure S3: Validation loss as a function of the number of hidden units in a network for models fit to 100,000 SNPs selected randomly from *Anopheles* chromosome 3R and human chromosome 1. See table 1 for model rankings. Curves are quadratic least-squares model fits.

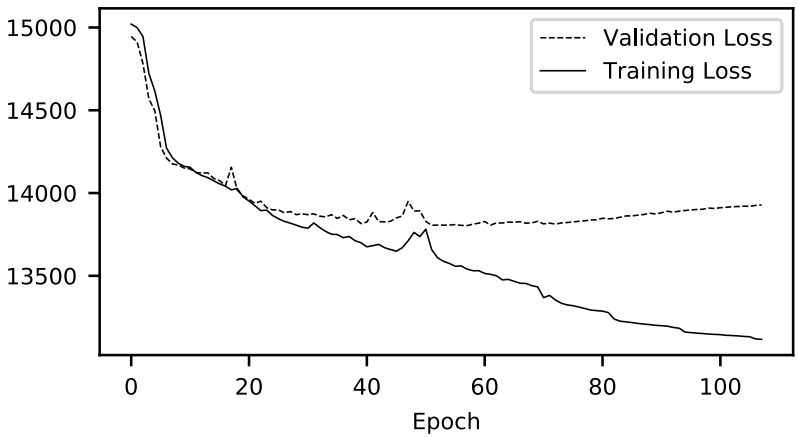

Figure S4: Example training history plot of showing training and validation loss by epoch during model training.

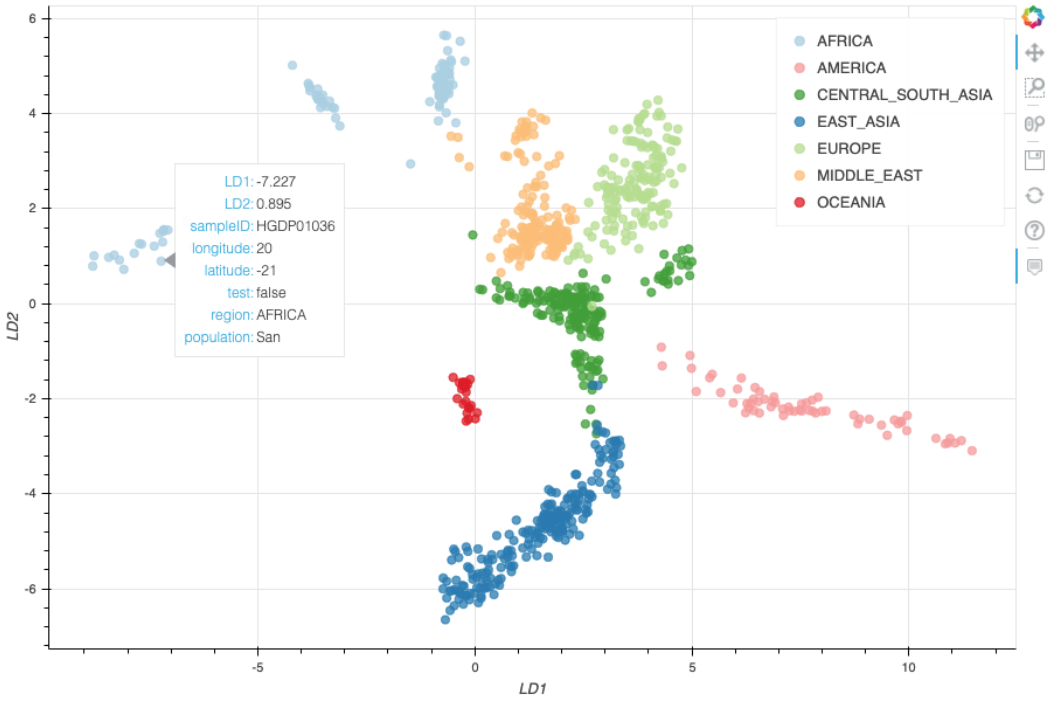

Figure S5: Example interactive plotting with scroll-over metadata.

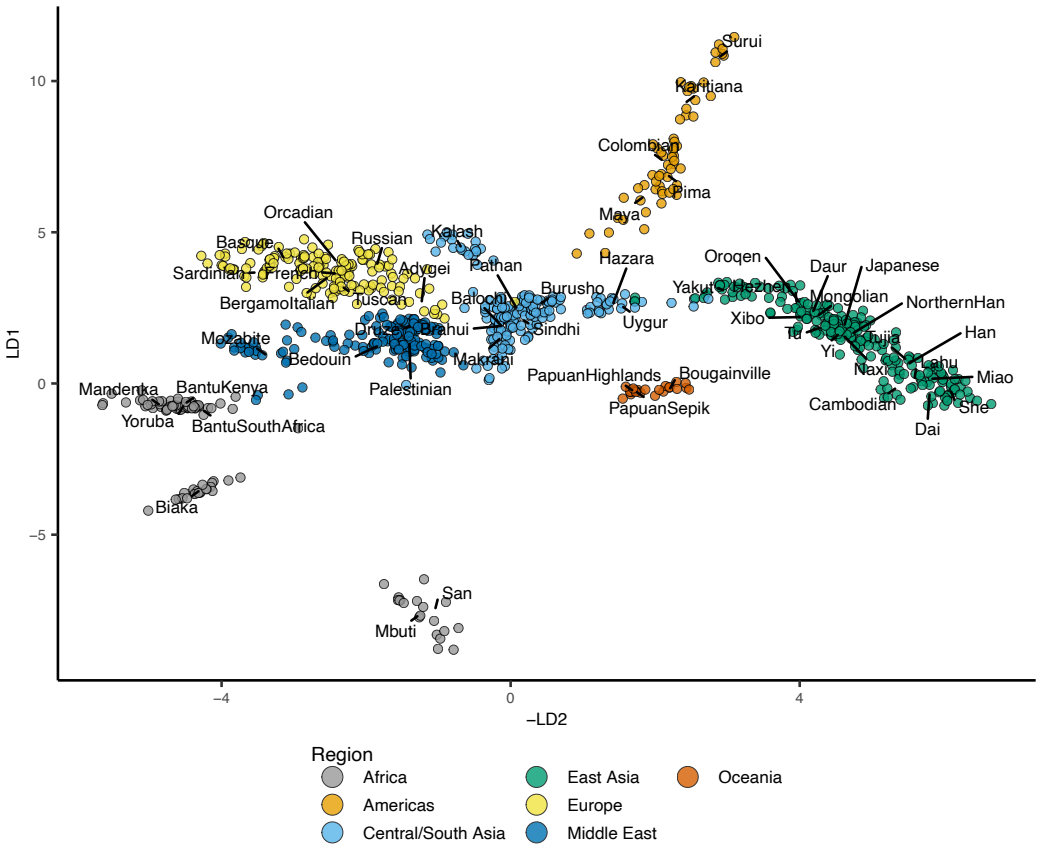

Figure S6: Latent space for 100,000 SNPs from chromosome 1 of the HGDP cohort (see Figure 2), with population centroids labeled.

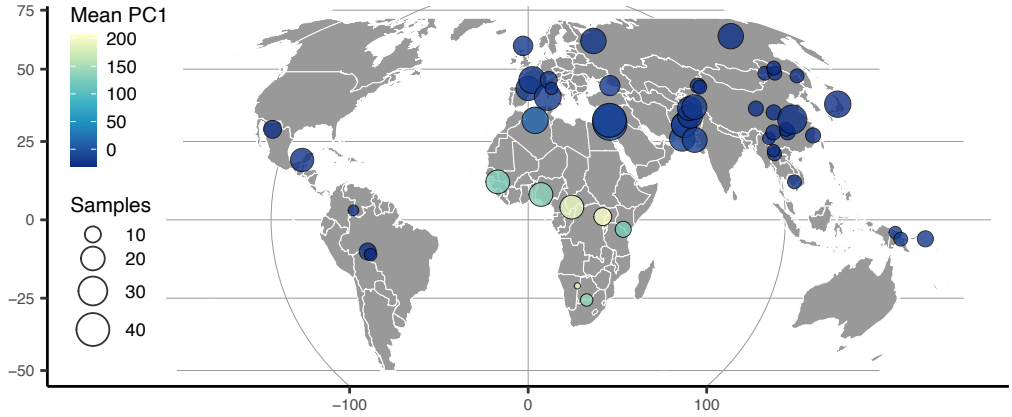

Figure S7: The first PC axis for HGDP SNPs summarized on a map as in Figure 3. Points show approximate population locations and are colored by the mean PC1 coordinate for each HGDP population. Densities show the distribution of PC1 scores for each HGDP region.

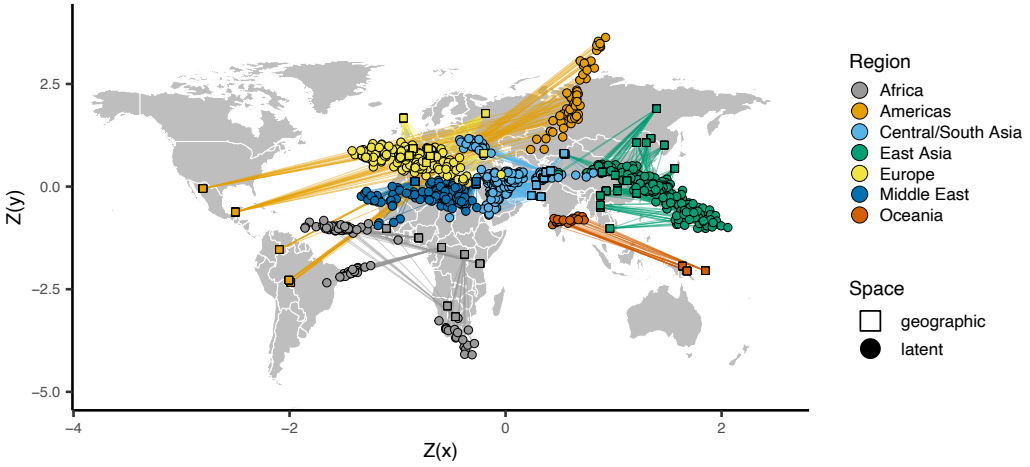

Figure S8: Comparing the VAE latent space with the geography of sampling localities HGDP samples. Circles show z-normalized sample locations in latent space and squares show the corresponding location in geographic space.

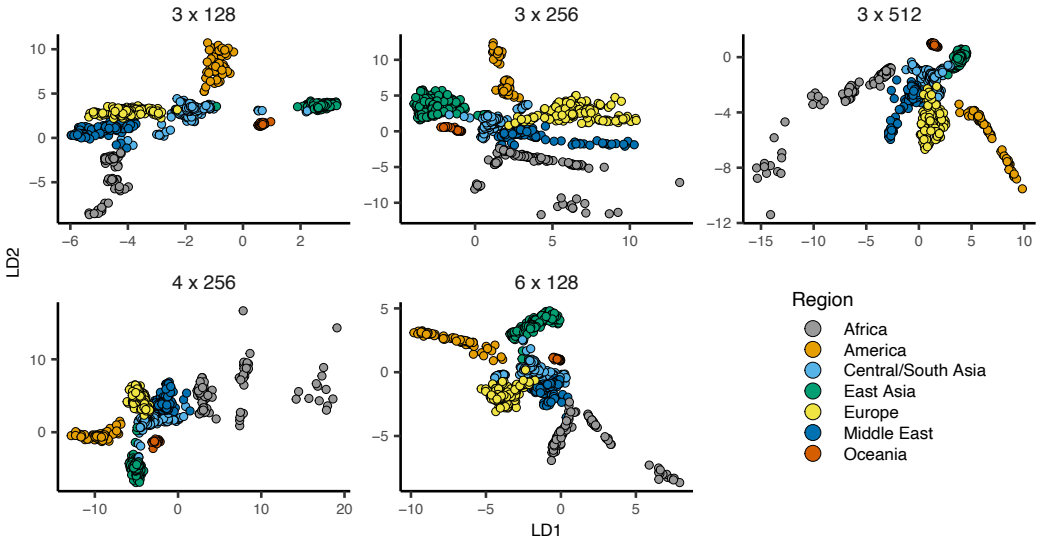

Figure S9: `popvae` latent spaces from runs with the same random seed and the top five network sizes by validation loss. Network sizes are listed as '`depth` x `width`'.

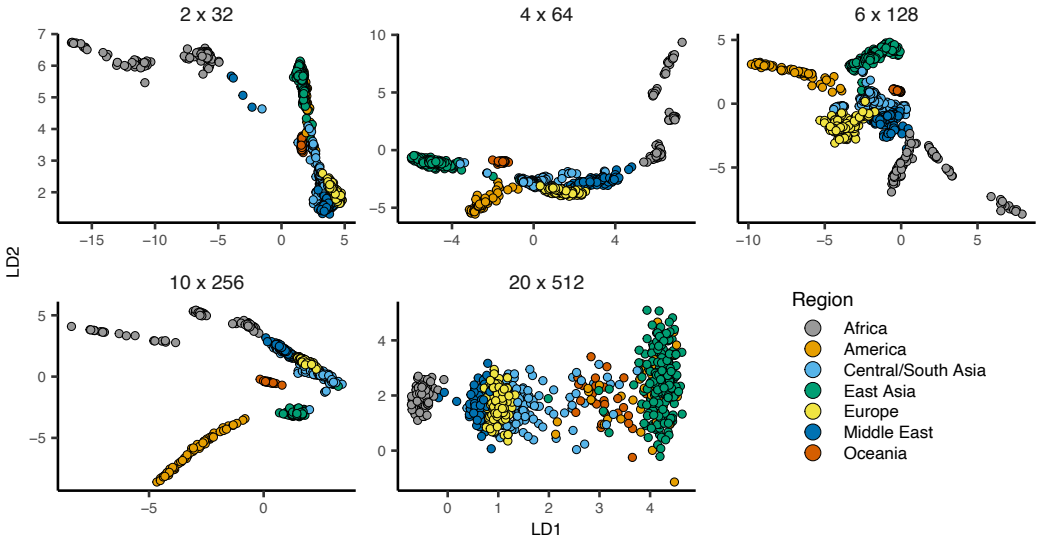

Figure S10: `popvae` latent spaces from models across the range of sizes tested. Network sizes are listed as '`depth` x `width`'.

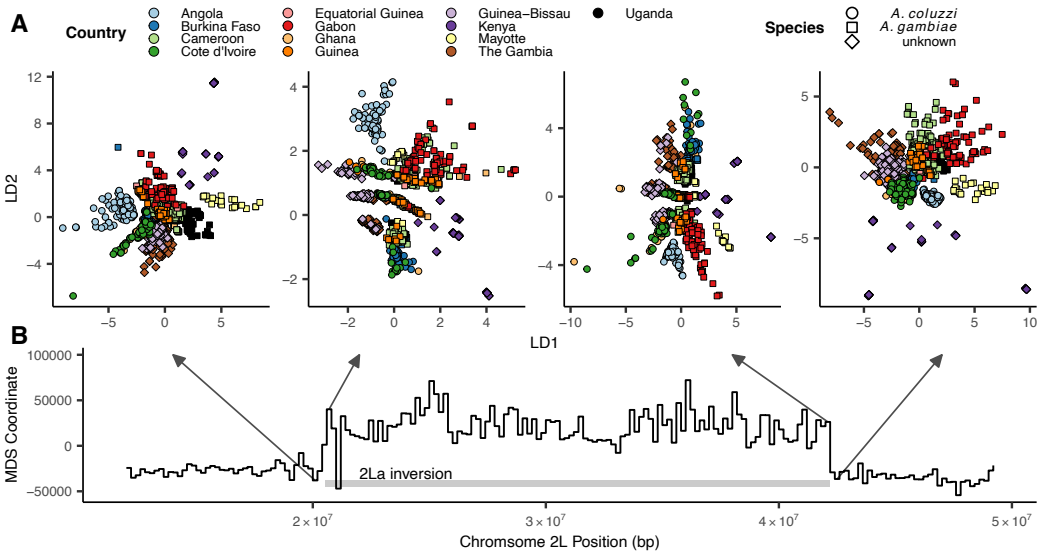

Figure S11: Latent spaces reflect inversion karyotypes at the 2La inversion in *A. gambiae / coluzzii*. A: VAE latent spaces for AG1000G phase 2 samples from windows near the 2La inversion breakpoints, with shapes indicating species and colors the country of origin. "Unknown" species localities include populations from Kenya, Guinea-Bissau, and the Gambia, for which diagnostic PCR markers are inconsistent or fail to amplify. B: Multi-dimensional scaling values showing difference in the relative position of individuals in latent space across windows – high values reflect windows in which samples cluster by inversion karyotype, and low values by species/region.

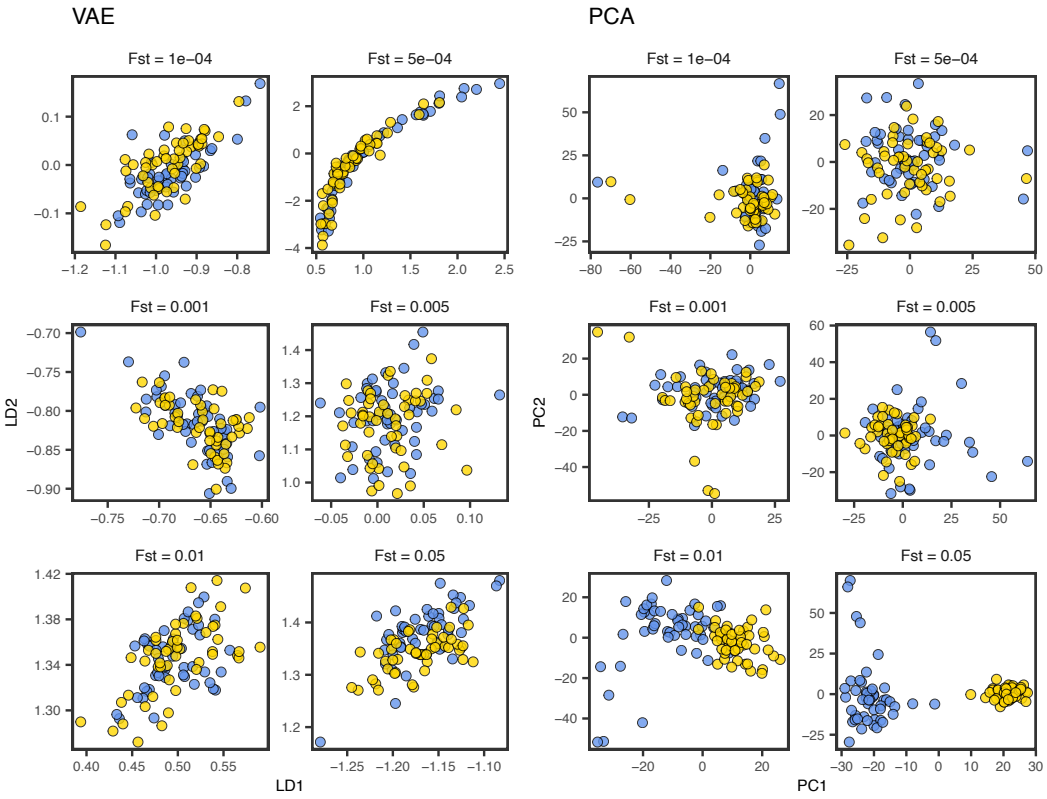

Figure S12: VAE latent spaces from the simulations shown in Figure 7, with `popvae` run at default settings.

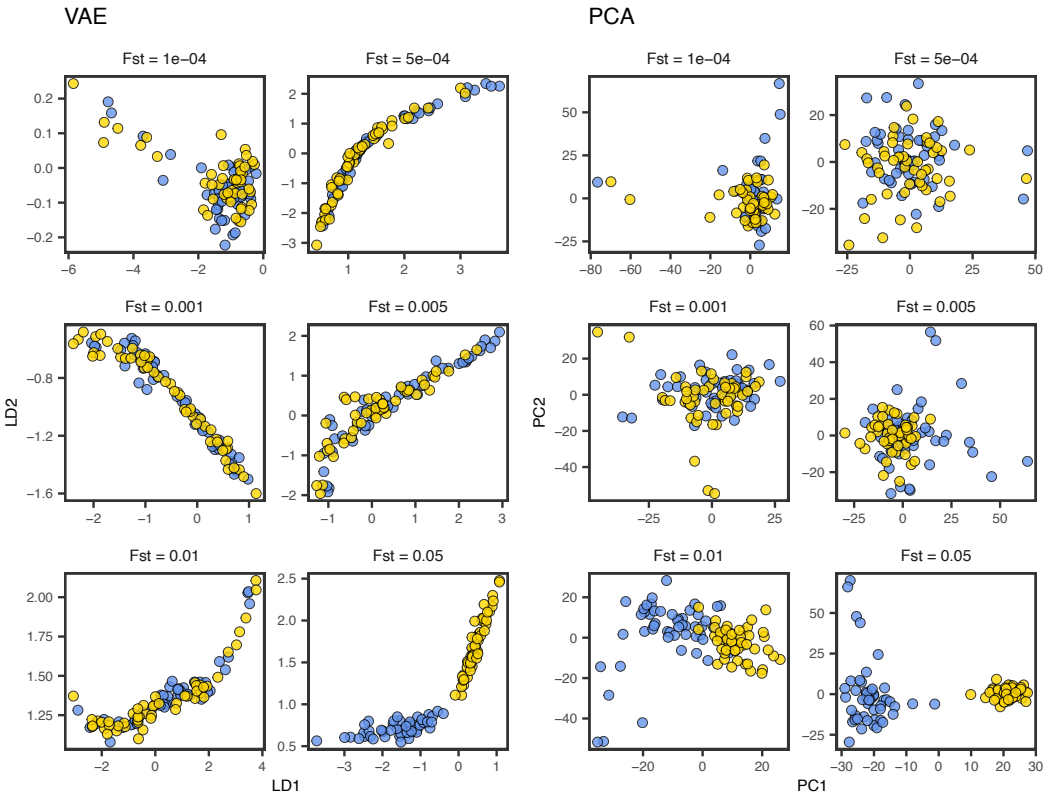

Figure S13: VAE latent spaces from the simulations shown in Figure 7, with `popvae` run with default network size (width 128, depth 6) and patience set to 500.

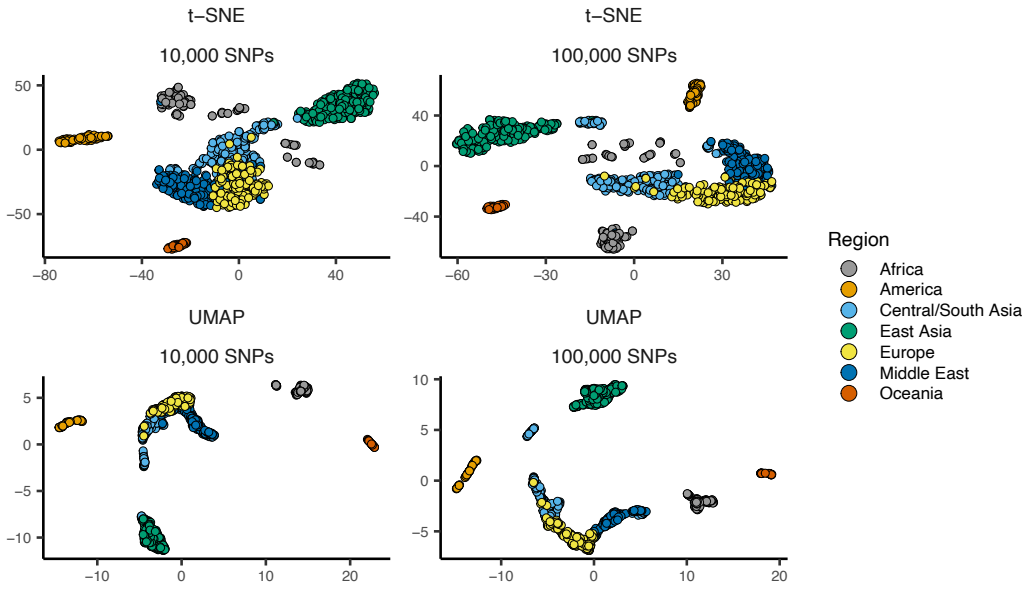

Figure S14: UMAP and t-SNE plots of HGDP samples using 100,000 or 10,000 SNPs. Both methods were run with default settings on the top 15 PC axes.

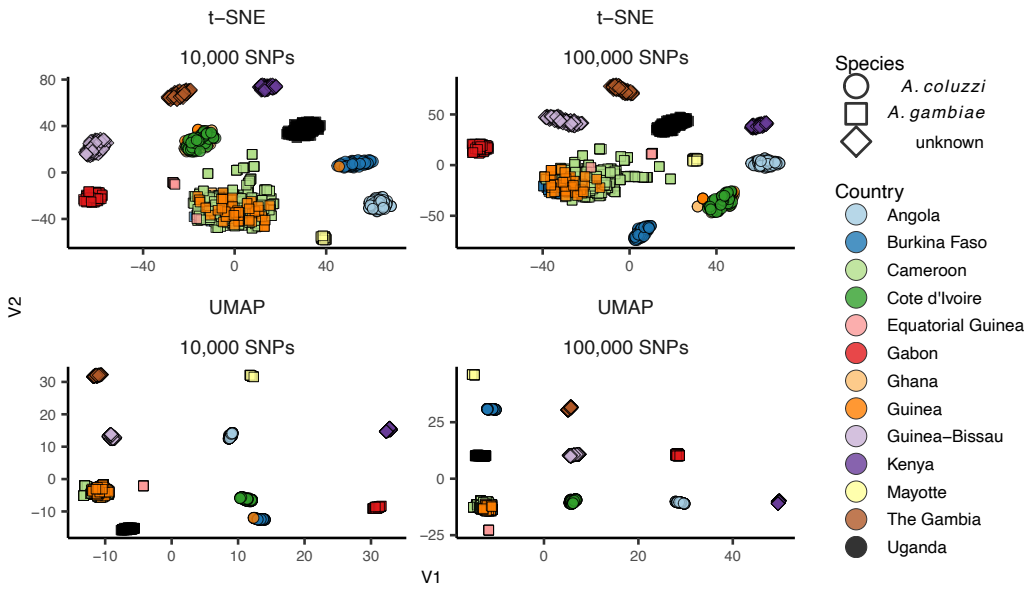

Figure S15: UMAP and t-SNE plots of AG1000G phase 2 samples using 100,000 or 10,000 SNPs at default settings. Both methods were run with default settings on the top 15 PC axes.

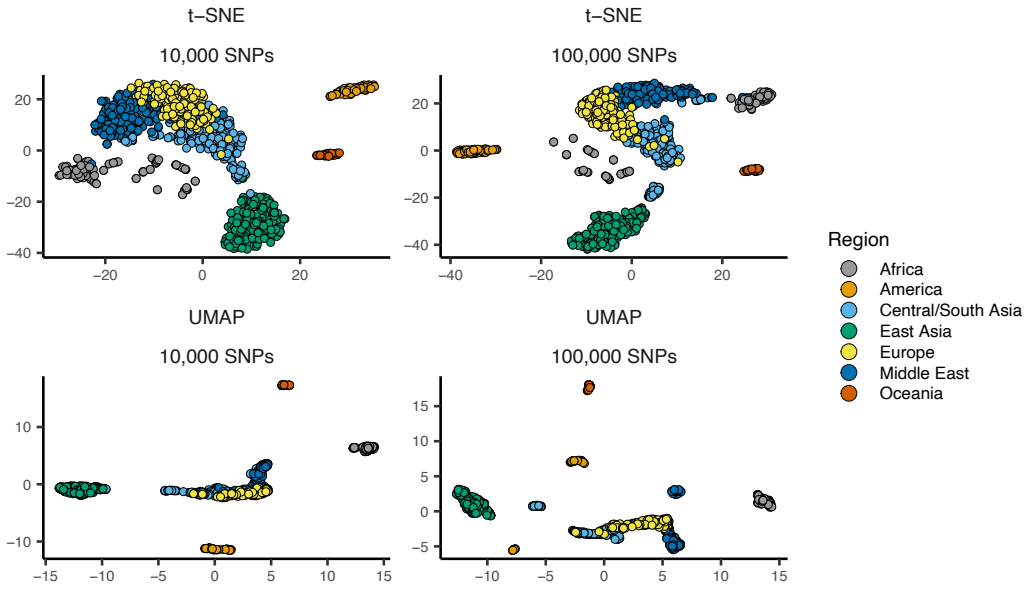

Figure S16: UMAP and t-SNE plots with parameters `n_neighbors=30` and `perplexity=60`. These settings are double the default values and are intended to improve global relative to local structure.

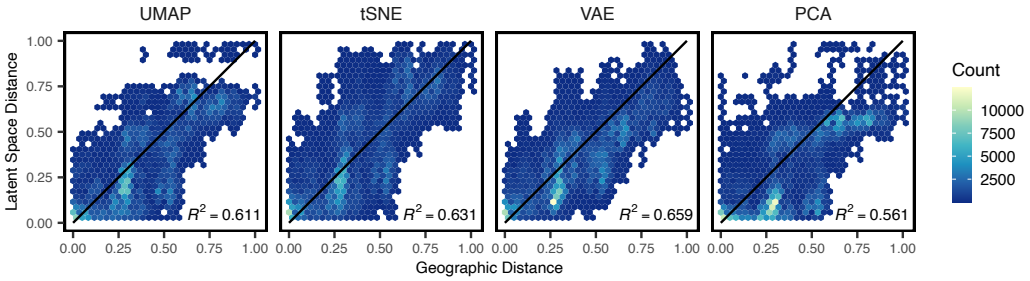

Figure S17: Comparison of relative pairwise distance for Eurasian HGDP samples, with UMAP parameter `n_neighbors=30` and t-SNE parameter `perplexity=60`. These settings are double the default values and are intended to improve global relative to local structure.

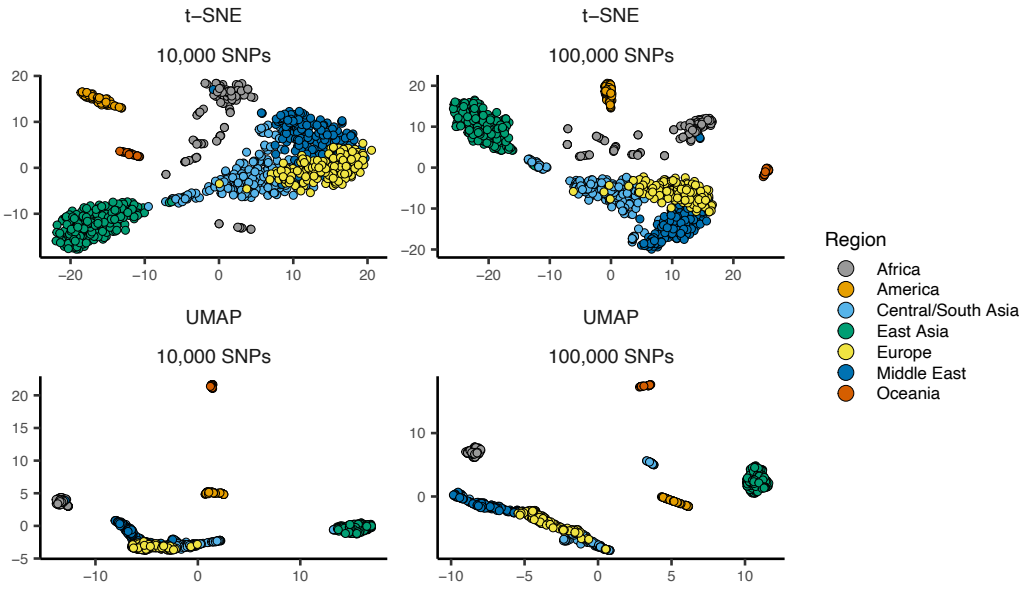

Figure S18: UMAP and t-SNE plots with parameters `n_neighbors=30` and `perplexity=60`. These settings are double the default values and are intended to improve global relative to local structure.

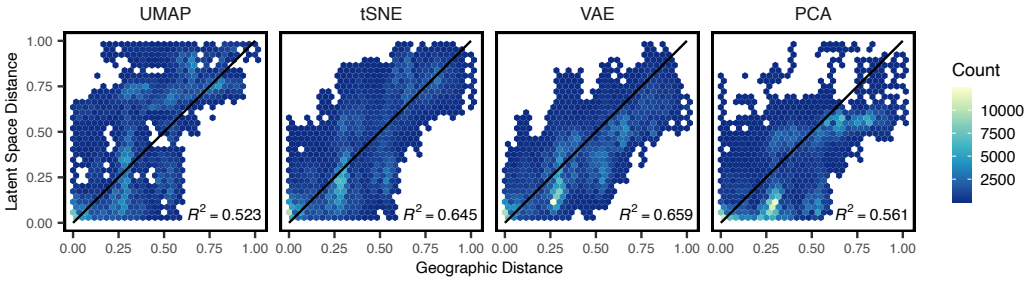

Figure S19: Comparison of relative pairwise distance for Eurasian HGDP samples, with UMAP parameter `n_neighbors=45` and t-SNE parameter `perplexity=90`. These settings are triple the default values and are intended to improve global relative to local structure.

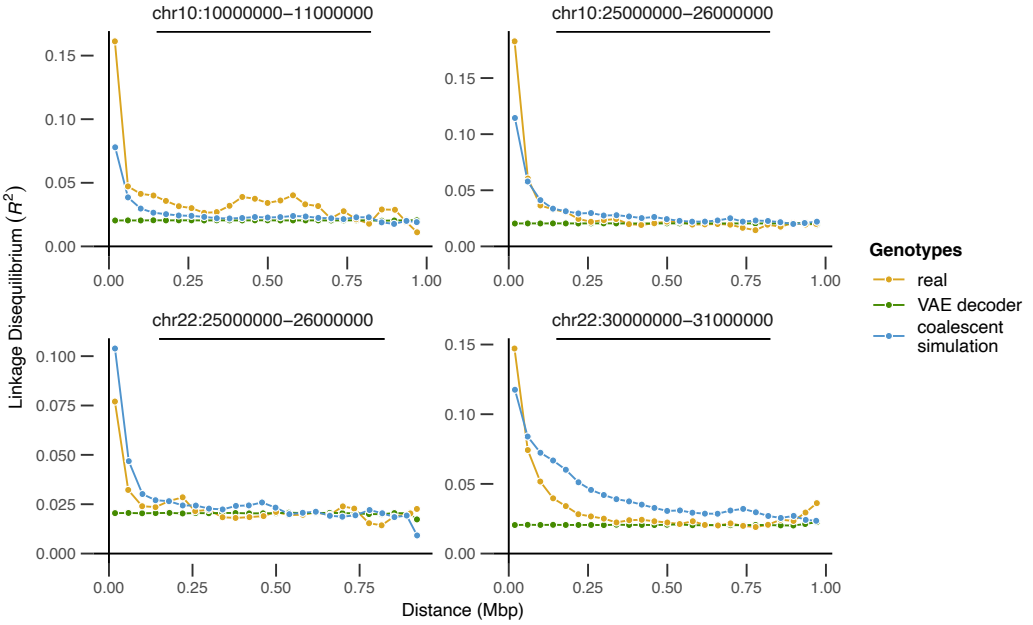

Figure S20: Comparing LD decay curves across real, simulated, and VAE decoder genotypes for four different regions of the genome. Points show the mean LD for all pairs of variants in each of 25 distance bins.

	HGDP			**AG1000G**		
SNPs	Depth	Width	Loss	Depth	Width	Loss
10,000	4	256	1394.231	3	256	750.677
	6	128	1394.48	6	128	750.859
	6	64	1394.504	4	128	751.0646
	10	64	1394.663	4	256	751.1514
	3	256	1394.976	6	64	751.7088
100,000	6	128	13955.76	4	256	7603.105
	3	256	13968.39	6	128	7606.528
	4	256	13971.75	3	256	7613.279
	3	512	13980.04	6	256	7614.232
	3	128	13992.27	4	128	7615.816
500,000	6	128	70087.32	10	128	37836.90
	10	64	70191.43	6	128	37848.67
	10	128	70203.22	6	64	37860.76
	6	64	70221.66	10	64	37872.49
	4	128	70357.73	4	128	37888.68

Table S1: Comparing validation loss across network sizes. Depth is the number of layers, width is the number of hidden units per layer, and loss is the mean validation loss across 5 random starting seeds for each network. Networks are ranked by loss for each dataset. SNPs were selected randomly from human chromosome 1 and *Anopheles* chromosome 3R.