Supplemental Methods

Dataset

Our dataset was composed of publicly available genotypes of 485 horses from 32 breeds (Schaefer et al., 2017), obtained from either the MNEc2M SNP array (https://www.animalgenome.org/cgi-bin/util/vcf) or masked from variants called from whole genome sequences (NCBI BioProjects PRJEB14779, PRJNA273402, and PRJEB10098). Originally, the genomic coordinates were based on the EquCab2 equine genome reference build (Wade et al., 2009). We remapped SNPs to EquCab3 (Kalbfleisch et al., 2018) using a method based on probe sequences (Beeson et al., 2018), resulting in a final set of 1,820,349 biallelic SNP markers, and phased the genotypes with BEAGLE (Browning & Browning, 2007) to predict haplotypes and missing genotypes.

Recombination maps

We generated recombination maps following procedures described previously (Beeson et al., 2019). Using 100 random samples of 40 chromosomes from the entire dataset, we estimated population recombination rates ($4N_er$) with the *LDhat* package program *rhomap* (Auton & McVean, 2007). We split chromosomes into 2,000-SNP intervals with overlaps of 200 SNPs between adjacent windows. For each chromosome, we ran 10,100,000 total iterations with sampling occurring every 2,000 iterations and a burn-in period of 100,000 iterations. Overall, we used the average of 100 rate estimates per inter-SNP interval to generate a breed- and sex-averaged recombination map. We created breed-specific maps for individual breeds with genotypes for 18 or more individuals (Arabian, Belgian, Franches-Montagnes, French Trotter, Icelandic, Lusitano, Maremmano, Morgan, Quarter Horse, Standardbred, Thoroughbred, and Welsh Pony) following the procedure above. We used our effective population size estimates from the original dataset (Beeson et al., 2019) to convert $4N_er$ estimates to centimorgans.

Hotspot prediction

To predict recombination hotspots, we used *LDhot* (Auton et al., 2014), which performs likelihood ratio tests over sliding windows spanning 100 kb, we compared a model of a window with a constant recombination rate to one with a central 1-kb hotspot. For each window, the null distribution came from as many as 1,000 coalescent simulations which were matched for SNP density, background recombination rate, and sample size. Our significance threshold for hotspots was $p < 0.001$, joining adjacent windows with a cutoff of $p < 0.01$. Because several large regions (up to 1 Mb) were called as hotspots but only supported by one inter-SNP interval, we automatically discarded all hotspots greater than 10 kb containing only two SNPs. It is likely that these regions do contain hotspots, but due to lack of data, we cannot define their boundaries precisely.