Inferring individual inbreeding and demographic history from segments of identity by descent in *Ficedula* flycatcher genome sequences

Marty Kardos, Anna Qvarnström, Hans Ellegren
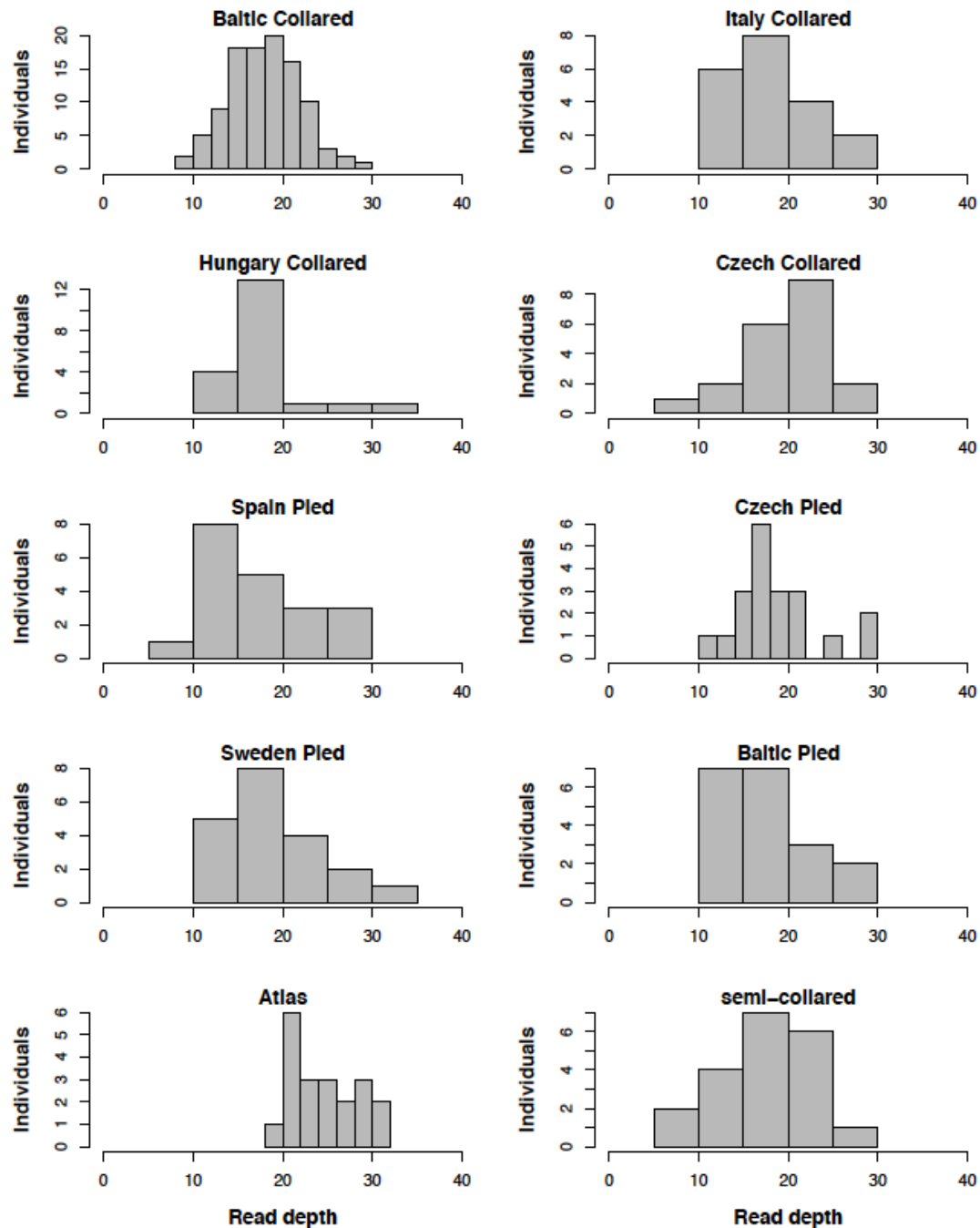
Supplementary Information

Figure S1. Histograms of mean sequencing read depth at SNP positions across individuals in each of the study populations. Mean read depth was calculated for each individual in each population after filtering the SNPs as described in the main text.
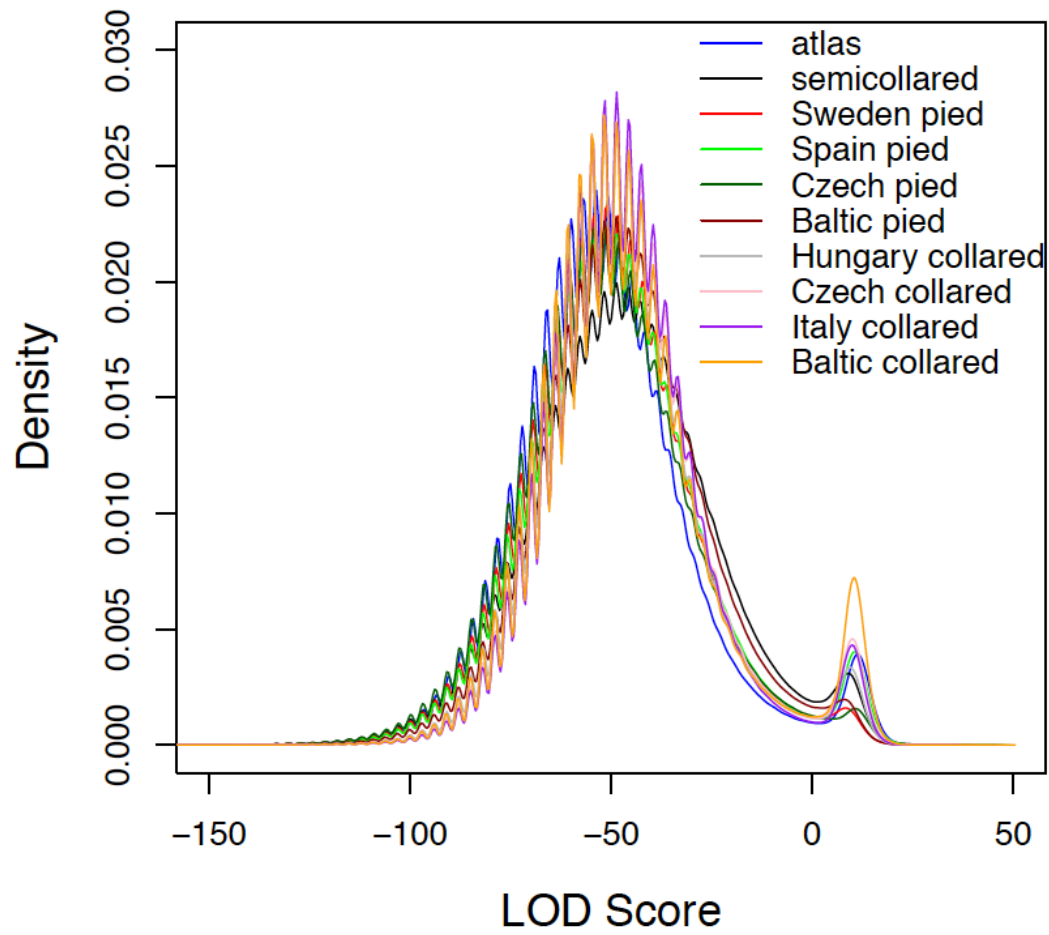
Figure S2. Distribution of LOD scores from the scans for IBD segments in each flycatcher study population. Each study population is represented by a different color line as indicated in the legend.
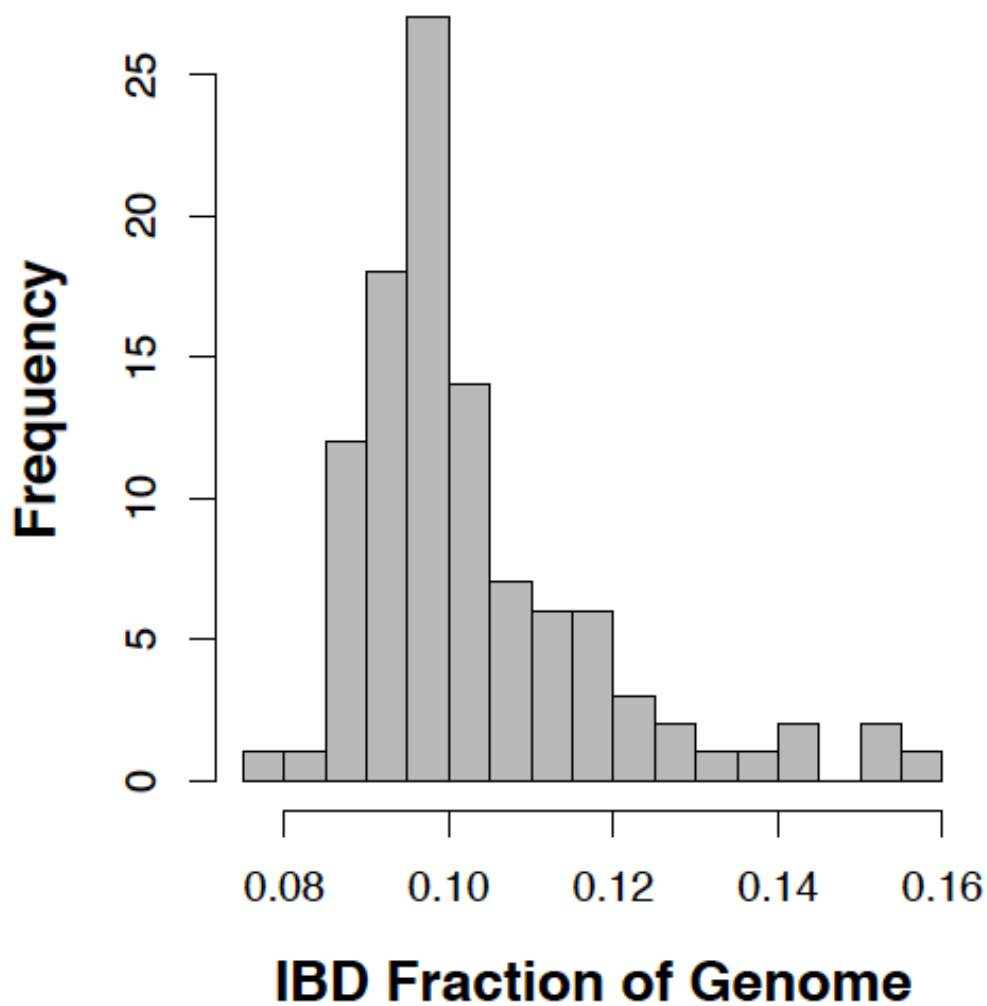
Figure S3. Distribution of the IBD fraction of the genome in 104 Baltic collared flycatchers. The IBD genome fractions were calculated for each individual by summing the physical lengths of all detected IBD segments and dividing by the physical length of the genome assembly.
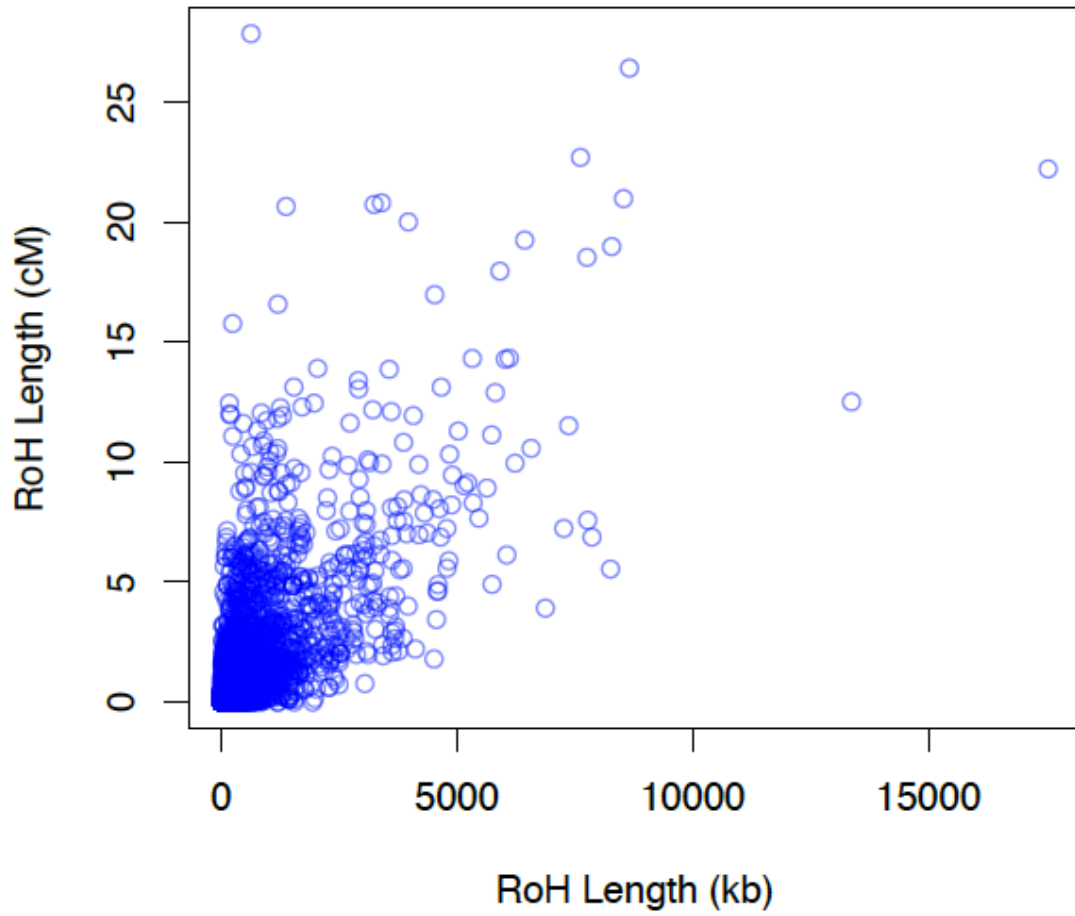
Figure S4. The relationship between the genetic and physical lengths of IBD segments in 104 Baltic collared flycatchers.
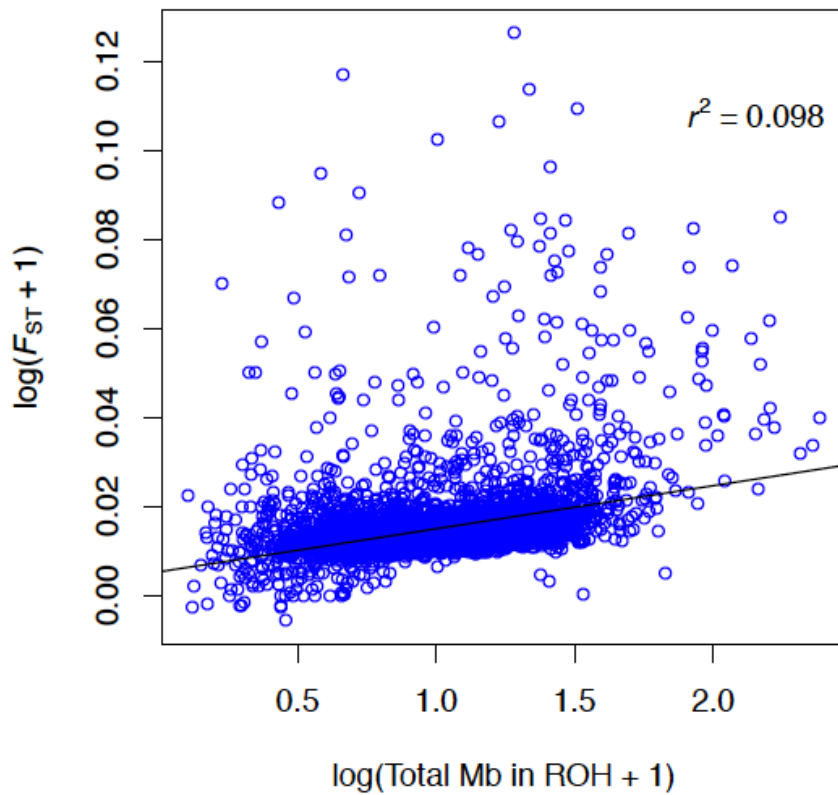
Figure S5. Relationship between F_{ST} and the abundance of ROH. F_{ST} and ROH abundance were each measured in non-overlapping 200 kb windows. F_{ST} (Weir & Cockerham 1984) was estimated between the Czech collared flycatchers and the Czech pied flycatchers. ROH abundance was measured in the Baltic collared flycatchers.
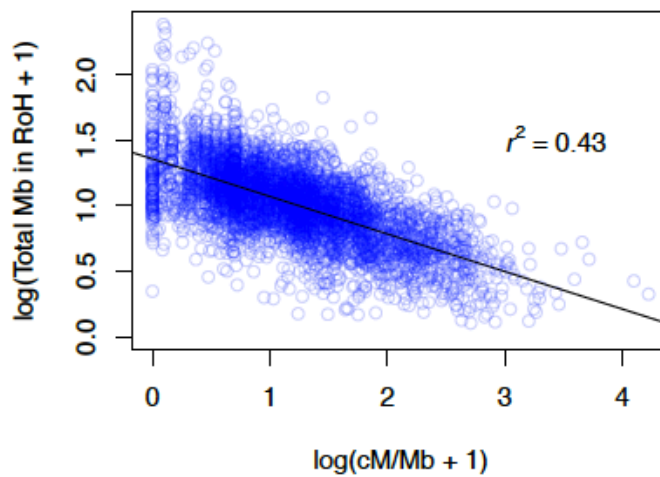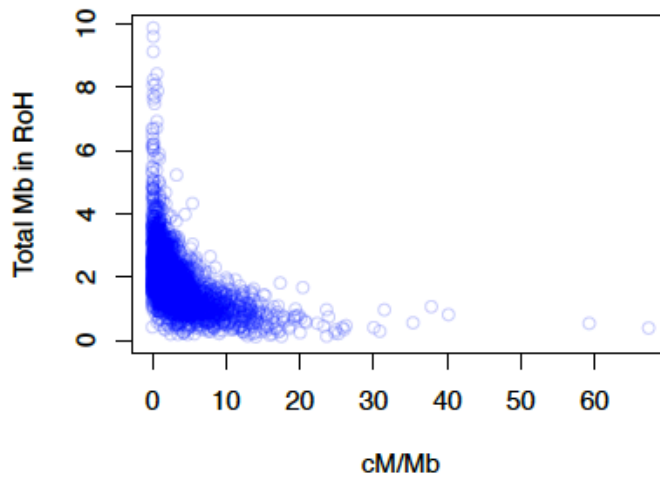
Figure S6. The relationship between the abundance of ROH and the recombination rate among 104 Baltic collared flycatchers. Each point represents the abundance of ROH in 200 kb windows across the collared flycatcher autosomes. Each variable is log transformed in the lower panel. The fitted line and r^2 in the lower panel are from a least squares linear regression.
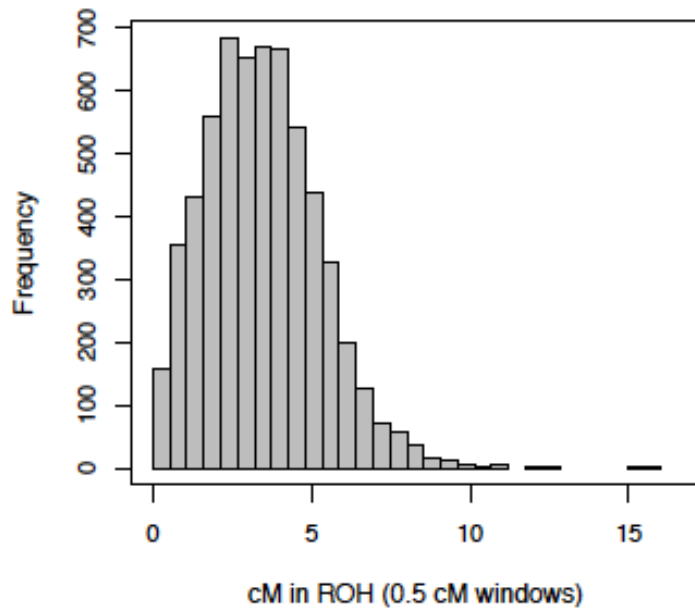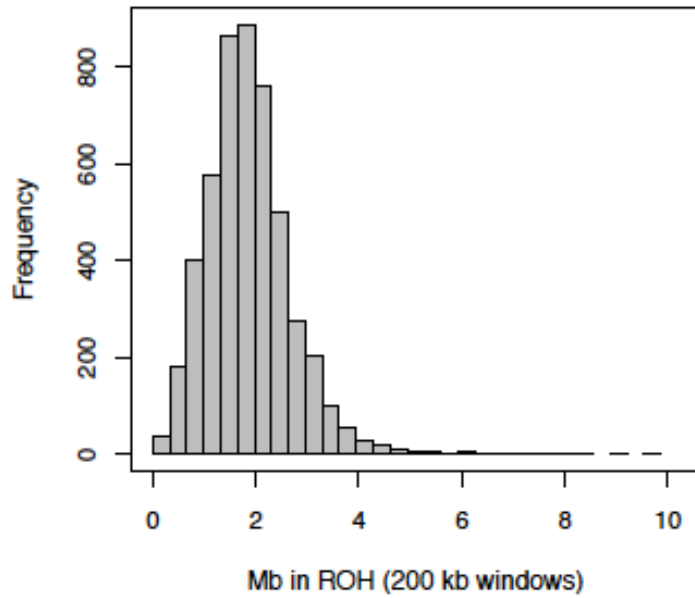
Figure S7. The distribution of the abundance of ROH across the genomes of 104 Baltic collared flycatchers. The upper panel shows a histogram of the total MB in ROH among all non-overlapping 200 kb windows. The lower panel shows a histogram of the total cM in ROH among all non-overlapping 0.5 cM windows.
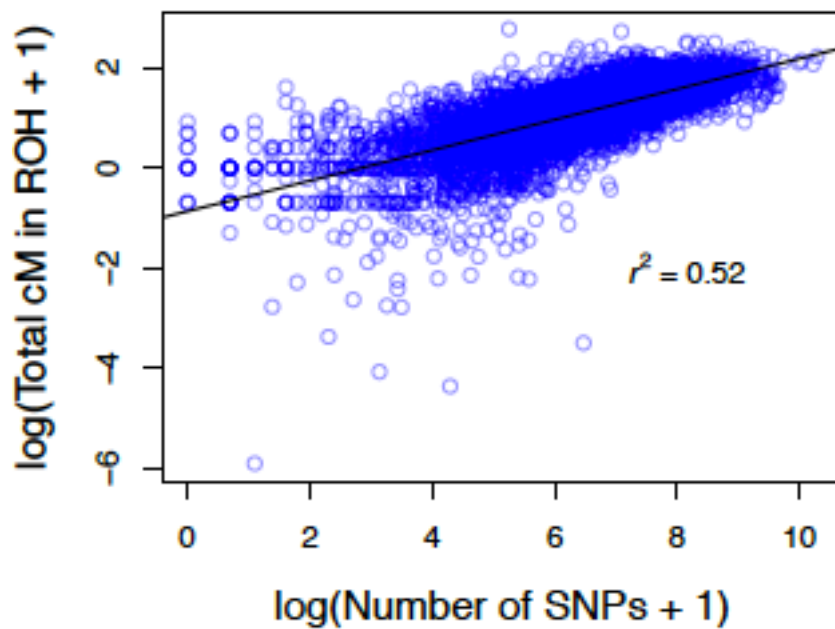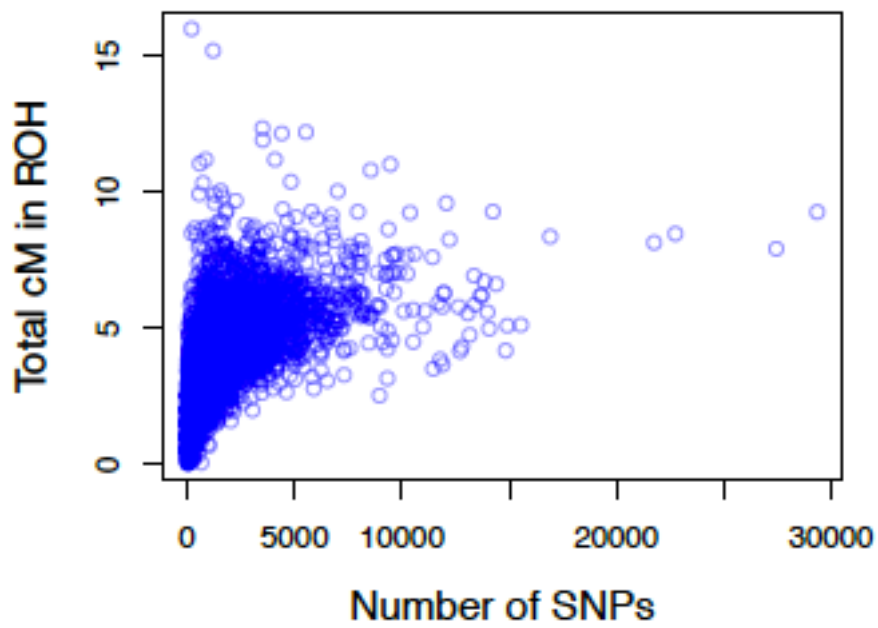
Figure S8. The density of genetically mapped ROH plotted against the number of SNPs in 0.5 cM windows. Log-transformed values, along with a fitted linear regression line are shown in the bottom panel.
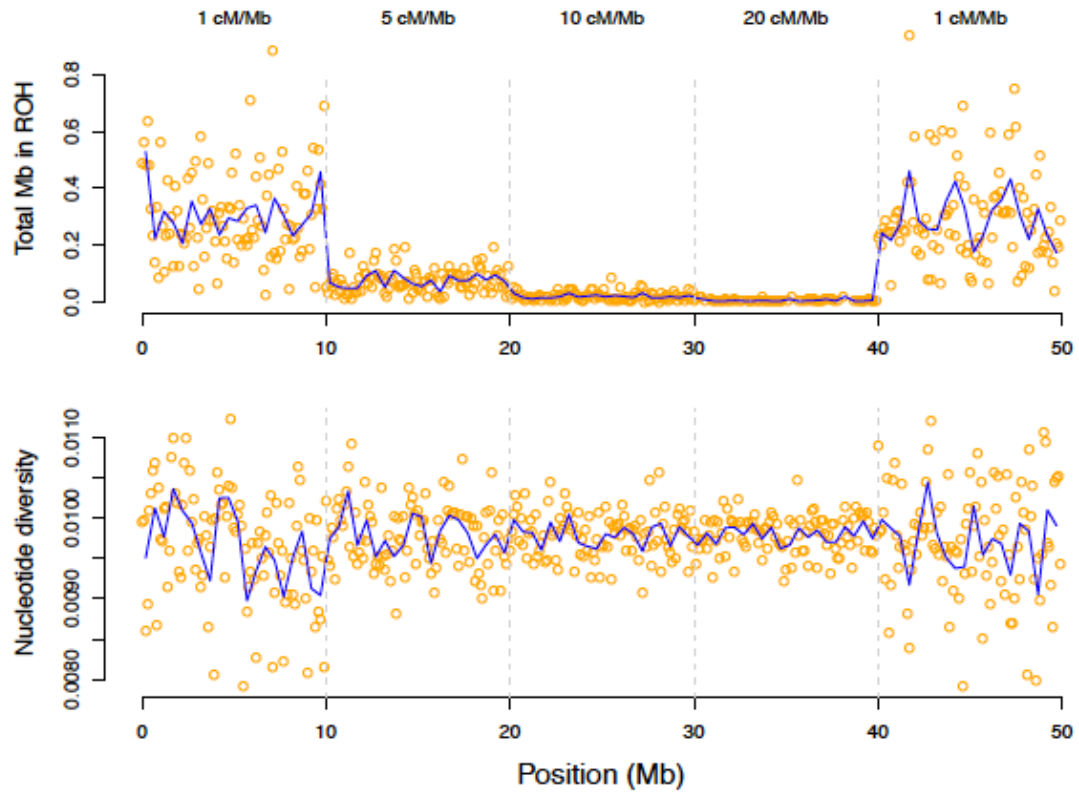
Figure S9 Effect of recombination rate on the detection of ROH in a simulated population. The recombination rate for each 10 Mb chromosome block is given at the top. The upper panel shows the density of ROH in non-overlapping 100 kb windows plotted against physical position (in Mb) on the chromosome. The bottom panel shows nucleotide diversity plotted against physical position in the same 100 kb windows. More ROH are detected in low recombination regions despite no systematic effect of recombination rate on nucleotide diversity and heterozygosity (Figure 7, main text).
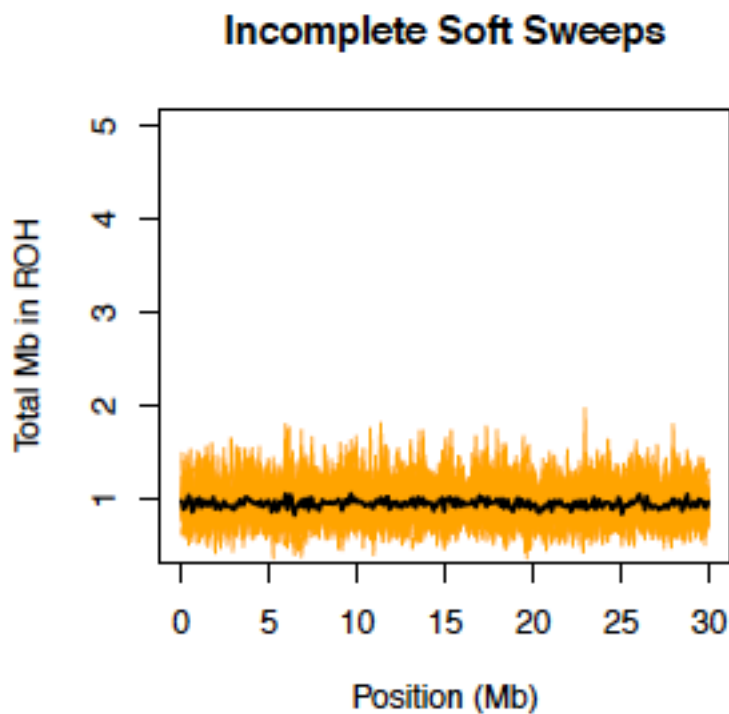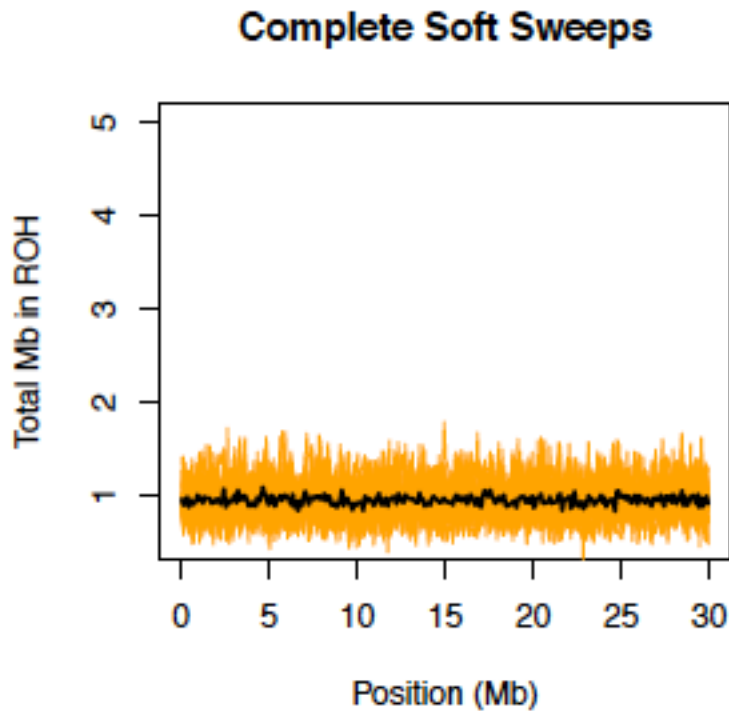
Complete Soft Sweeps

Incomplete Soft Sweeps

Figure S10. Effects of soft selective sweeps on the abundance of ROH. Results from simulations of complete selective sweeps are shown in the top panel, and results from incomplete selective sweeps are in the lower panel. The orange lines represent the total Mb in ROH in 100 kb windows for each of 20 replicate simulations. The black links represent the mean Mb in ROH across all simulation repetitions in the same 100 kb windows.

Table S1. Number SNPs and mean minor allele frequency (MAF) for each study population.

Population	SNPs	Mean MAF
collared Baltic (Öland)	8797090	0.151
collared Italy	7901418	0.151
collared Hungary	7579594	0.151
collared Czech Republic	7921335	0.152
pied Spain	4966019	0.169
pied Czech	5826728	0.167
pied Sweden	6284854	0.158
pied Baltic (Öland)	5634016	0.161
Atlas	6780466	0.173
semi-collared	4586377	0.156

Citations

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, 1358-1370.