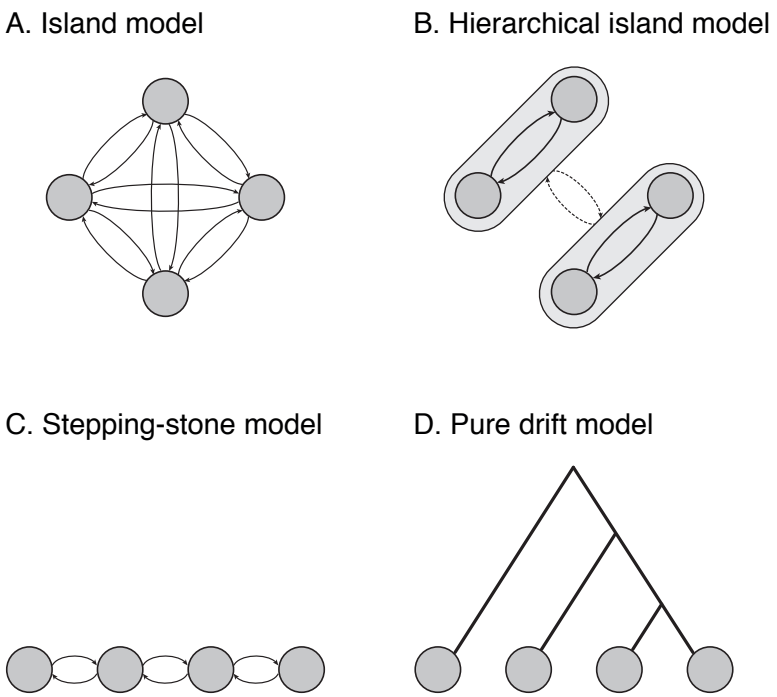

Figure S1 (A) Schematic representation of an island model. The actual data were simulated with n_d = 100 demes, each made of N = 250 diploid individuals (500 genes). Fifty diploid individuals (100 genes) were sampled per deme, in 9 demes. The migration rate (m = 0.003, plain arrows) was fixed to achieve the desired value of F_{ST} = 0.24, using equation 6 in Rousset (1996). (B) Schematic representation of a hierarchical island model. The actual data were simulated with 10 groups of 10 demes, each made of N = 250 diploid individuals (500 genes). Fifty diploid individuals (100 genes) were sampled per deme, in 3 groups of 3 demes. The migration rate within (m = 0.017, plain arrows) and among groups (m = 0.0003, dashed arrows) were fixed to achieve the desired values of F_{SC} = 0.05, F_{CT} = 0.05 and F_{ST} = 0.24, using equations A8–A10 in Excoffier *et al.* (2009). (C) Schematic representation of a stepping-stone model. The actual data were simulated with n_d = 100 demes, each made of N = 250 diploid individuals (500 genes). Fifty diploid individuals (100 genes) were sampled per deme, in 9 demes.The migration rate was fixed (m = 0.028, plain arrows), by trial and error, to achieve the desired value of F_{ST} = 0.24. (D) Schematic representation of a pure drift model. The actual data were simulated with 9 demes, diverging sequentially as depicted. The sample characteristics (number of individuals, number of sampled demes) were the same as in (A–C), and the divergence time (24 generations) between any two successive splits was tuned in order to achieve an overall F_{ST} of ≈ 0.24. In (A–D) 10,000 neutral markers were simulated.

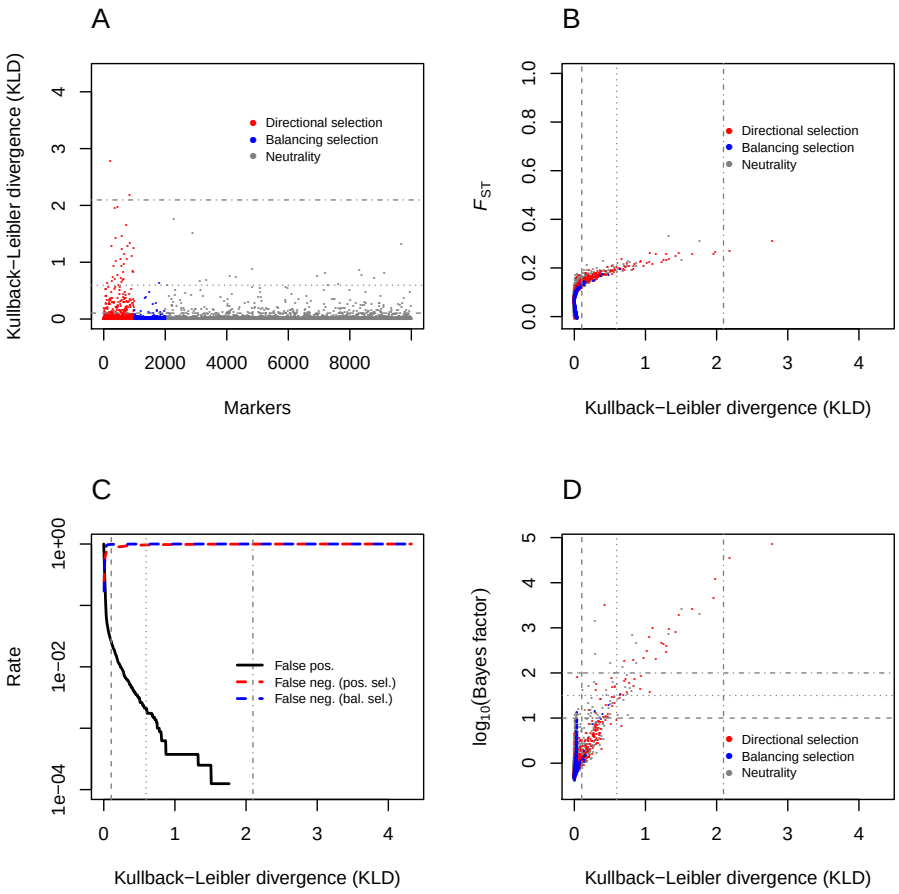

Figure S2 Analysis of the allele count data from dataset 1. (A) Kullback–Leibler divergence (KLD) measure between the posterior of δ_j and its centering distribution for all simulated loci. Loci under positive selection are depicted in red, loci under balancing selection in blue, and neutral markers are in grey. (B) F_{ST} as a function of the KLD measure for all loci. (C) False positive (neutral loci detected as outliers) and false negative (selected loci not detected as outliers) rates as a function of the KLD measure. (D) Relationship between the Bayes factor $\log_{10}$(BF) from the BAYESCAN analysis of dataset 1 and the KLD. The horizontal lines in (A) and the vertical lines in (B–D) indicate the KLD thresholds corresponding to the 95%-, the 99%- and the 99.9%-quantile of the of the KLD distribution from the pod analysis of dataset 1. In (D), the horizontal lines indicate the $\log_{10}$(BF) = 1, $\log_{10}$(BF) = 1.5 and $\log_{10}$(BF) = 2 thresholds, which correspond to ‘‘strong’’, ‘‘very strong’’ and ‘‘decisive’’ support, respectively, following Jeffreys’ (1961) scale of evidence.

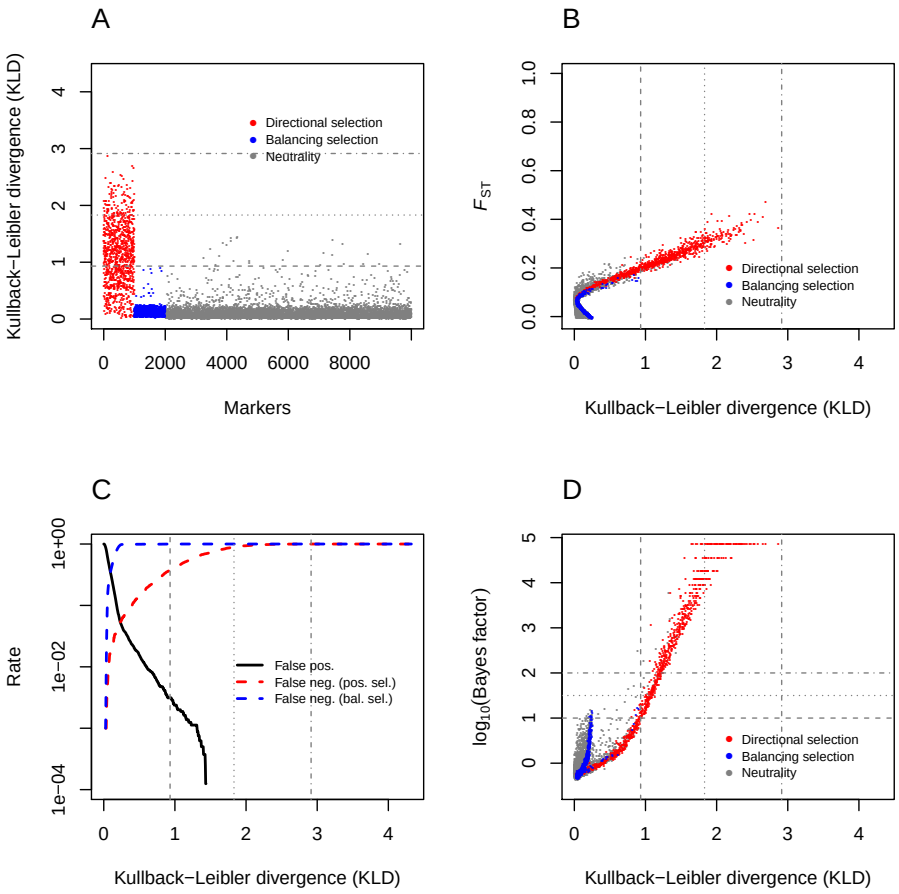

Figure S3 Analysis of the allele count data from dataset 2. (A) Kullback–Leibler divergence (KLD) measure between the posterior of δ_j and its centering distribution for all simulated loci. Loci under positive selection are depicted in red, loci under balancing selection in blue, and neutral markers are in grey. (B) F_{ST} as a function of the KLD measure for all loci. (C) False positive (neutral loci detected as outliers) and false negative (selected loci not detected as outliers) rates as a function of the KLD measure. (D) Relationship between the Bayes factor $\log_{10}$(BF) from the BAYESCAN analysis of dataset 2 and the KLD. The horizontal lines in (A) and the vertical lines in (B–D) indicate the KLD thresholds corresponding to the 95%-, the 99%- and the 99.9%-quantile of the of the KLD distribution from the pod analysis of dataset 2. In (D), the horizontal lines indicate the $\log_{10}$(BF) = 1, $\log_{10}$(BF) = 1.5 and $\log_{10}$(BF) = 2 thresholds, which correspond to ‘‘strong’’, ‘‘very strong’’ and ‘‘decisive’’ support, respectively, following Jeffreys’ (1961) scale of evidence.

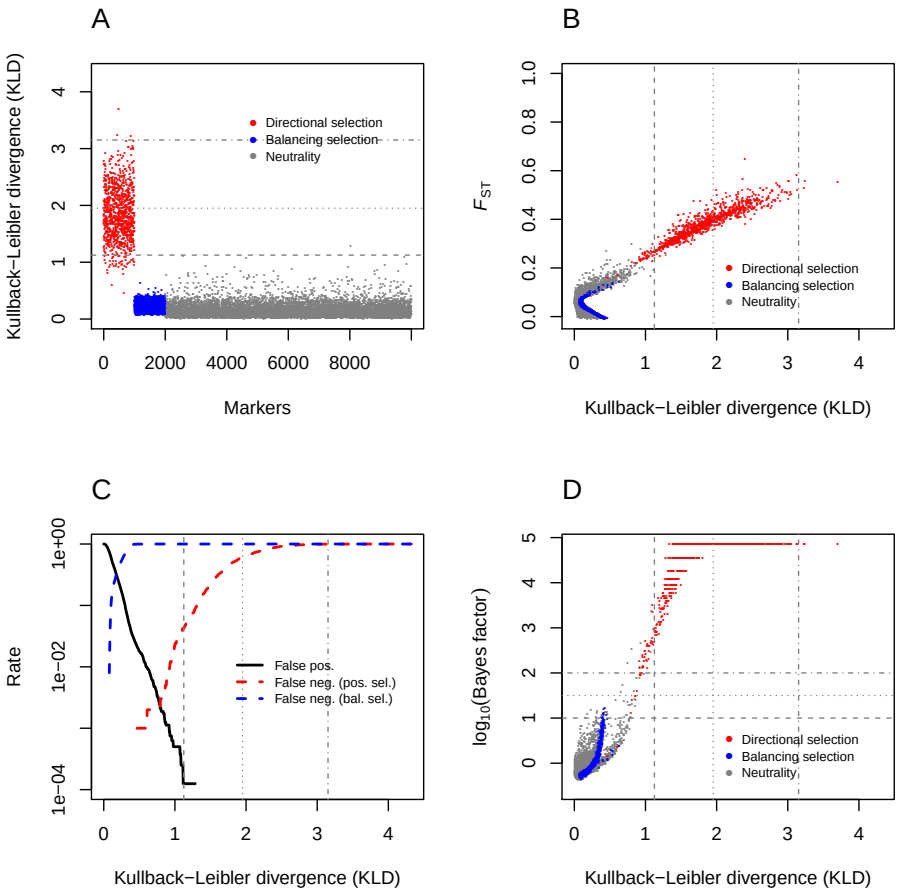

Figure S4 Analysis of the allele count data from dataset 3. (A) Kullback–Leibler divergence (KLD) measure between the posterior of δ_j and its centering distribution for all simulated loci. Loci under positive selection are depicted in red, loci under balancing selection in blue, and neutral markers are in grey. (B) F_{ST} as a function of the KLD measure for all loci. (C) False positive (neutral loci detected as outliers) and false negative (selected loci not detected as outliers) rates as a function of the KLD measure. (D) Relationship between the Bayes factor $\log_{10}$(BF) from the BAYESCAN analysis of dataset 3 and the KLD. The horizontal lines in (A) and the vertical lines in (B–D) indicate the KLD thresholds corresponding to the 95%-, the 99%- and the 99.9%-quantile of the of the KLD distribution from the pod analysis of dataset 3. In (D), the horizontal lines indicate the $\log_{10}$(BF) = 1, $\log_{10}$(BF) = 1.5 and $\log_{10}$(BF) = 2 thresholds, which correspond to ‘‘strong’’, ‘‘very strong’’ and ‘‘decisive’’ support, respectively, following Jeffreys’ (1961) scale of evidence.

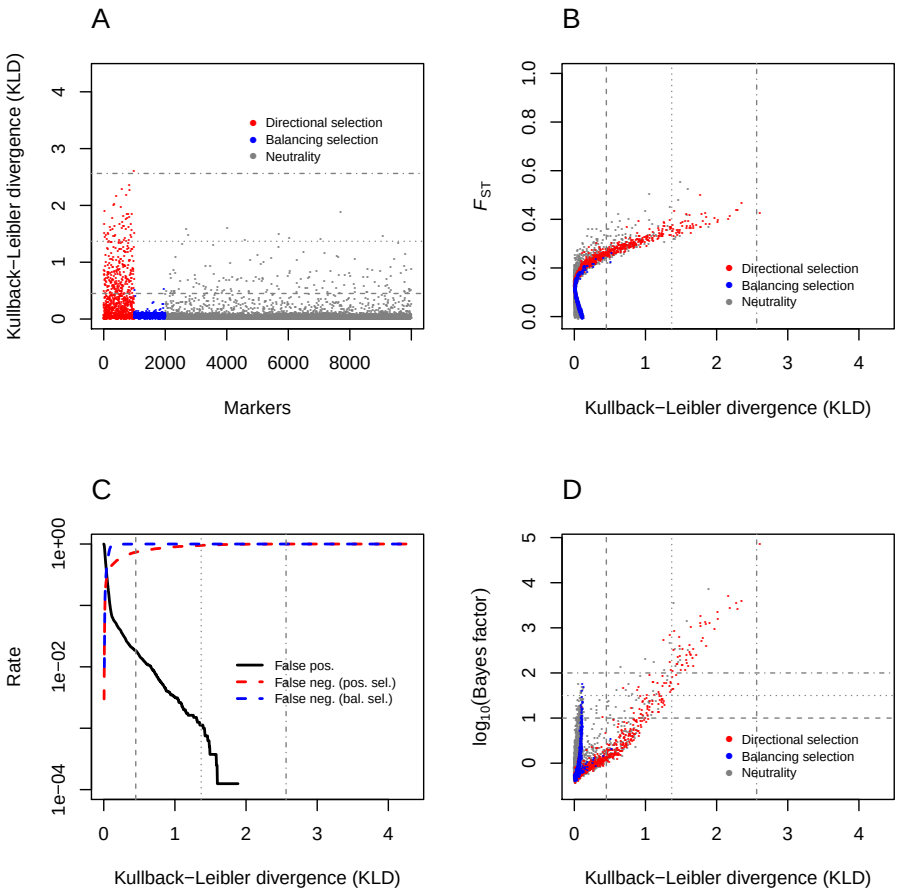

Figure S5 Analysis of the allele count data from dataset 4. (A) Kullback–Leibler divergence (KLD) measure between the posterior of δ_j and its centering distribution for all simulated loci. Loci under positive selection are depicted in red, loci under balancing selection in blue, and neutral markers are in grey. (B) F_{ST} as a function of the KLD measure for all loci. (C) False positive (neutral loci detected as outliers) and false negative (selected loci not detected as outliers) rates as a function of the KLD measure. (D) Relationship between the Bayes factor $\log_{10}$(BF) from the BAYESCAN analysis of dataset 4 and the KLD. The horizontal lines in (A) and the vertical lines in (B–D) indicate the KLD thresholds corresponding to the 95%-, the 99%- and the 99.9%-quantile of the of the KLD distribution from the pod analysis of dataset 4. In (D), the horizontal lines indicate the $\log_{10}$(BF) = 1, $\log_{10}$(BF) = 1.5 and $\log_{10}$(BF) = 2 thresholds, which correspond to ‘‘strong’’, ‘‘very strong’’ and ‘‘decisive’’ support, respectively, following Jeffreys’ (1961) scale of evidence.

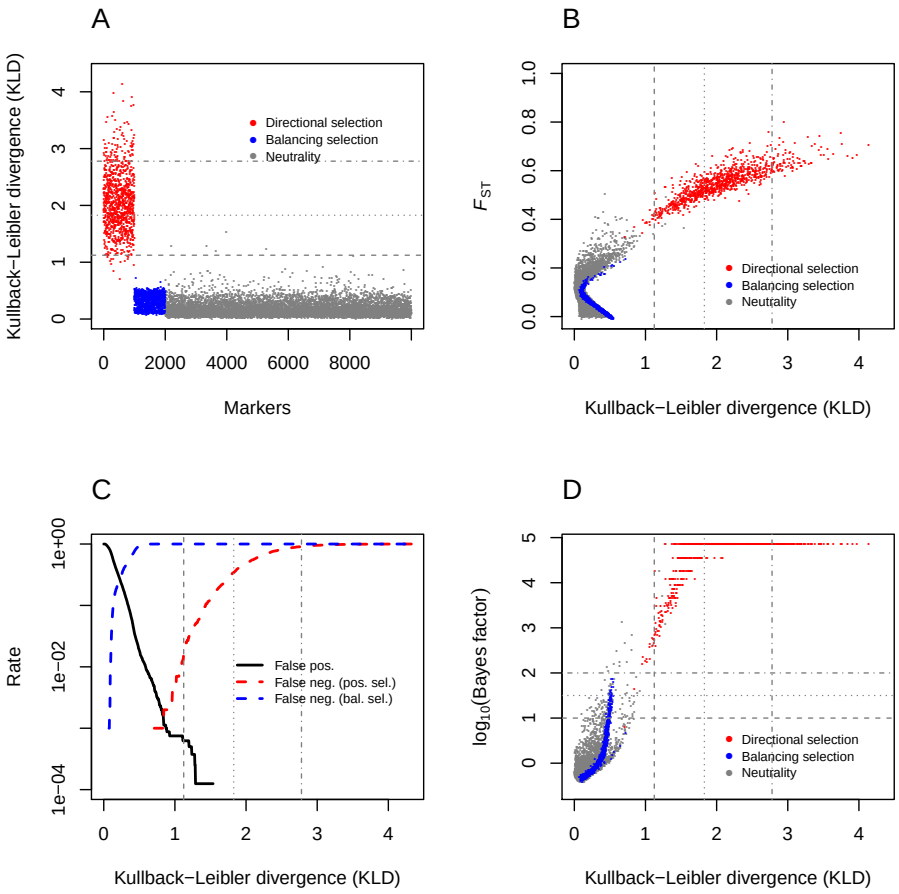

Figure S6 Analysis of the allele count data from dataset 6. (A) Kullback–Leibler divergence (KLD) measure between the posterior of δ_j and its centering distribution for all simulated loci. Loci under positive selection are depicted in red, loci under balancing selection in blue, and neutral markers are in grey. (B) F_{ST} as a function of the KLD measure for all loci. (C) False positive (neutral loci detected as outliers) and false negative (selected loci not detected as outliers) rates as a function of the KLD measure. (D) Relationship between the Bayes factor $\log_{10}$(BF) from the BAYESCAN analysis of dataset 6 and the KLD. The horizontal lines in (A) and the vertical lines in (B–D) indicate the KLD thresholds corresponding to the 95%-, the 99%- and the 99.9%-quantile of the of the KLD distribution from the pod analysis of dataset 6. In (D), the horizontal lines indicate the $\log_{10}$(BF) = 1, $\log_{10}$(BF) = 1.5 and $\log_{10}$(BF) = 2 thresholds, which correspond to ‘‘strong’’, ‘‘very strong’’ and ‘‘decisive’’ support, respectively, following Jeffreys’ (1961) scale of evidence.

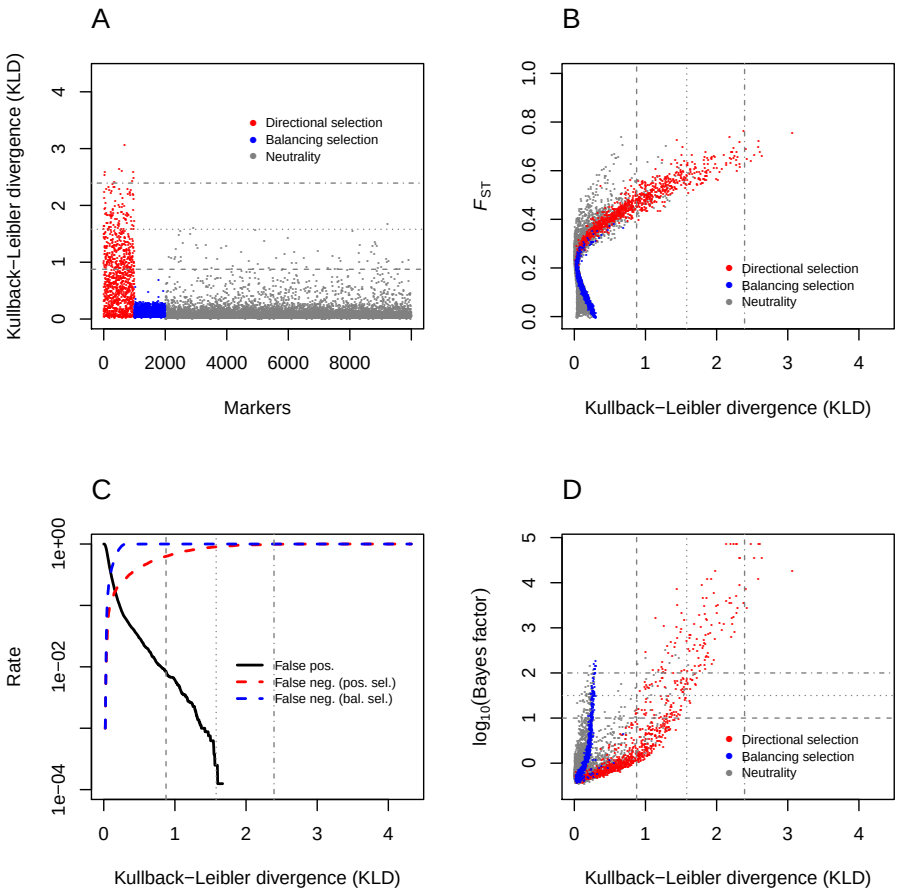

Figure S7 Analysis of the allele count data from dataset 7. (A) Kullback–Leibler divergence (KLD) measure between the posterior of δ_j and its centering distribution for all simulated loci. Loci under positive selection are depicted in red, loci under balancing selection in blue, and neutral markers are in grey. (B) F_{ST} as a function of the KLD measure for all loci. (C) False positive (neutral loci detected as outliers) and false negative (selected loci not detected as outliers) rates as a function of the KLD measure. (D) Relationship between the Bayes factor $\log_{10}$(BF) from the BAYESCAN analysis of dataset 7 and the KLD. The horizontal lines in (A) and the vertical lines in (B–D) indicate the KLD thresholds corresponding to the 95%-, the 99%- and the 99.9%-quantile of the of the KLD distribution from the pod analysis of dataset 7. In (D), the horizontal lines indicate the $\log_{10}$(BF) = 1, $\log_{10}$(BF) = 1.5 and $\log_{10}$(BF) = 2 thresholds, which correspond to ‘‘strong’’, ‘‘very strong’’ and ‘‘decisive’’ support, respectively, following Jeffreys’ (1961) scale of evidence.

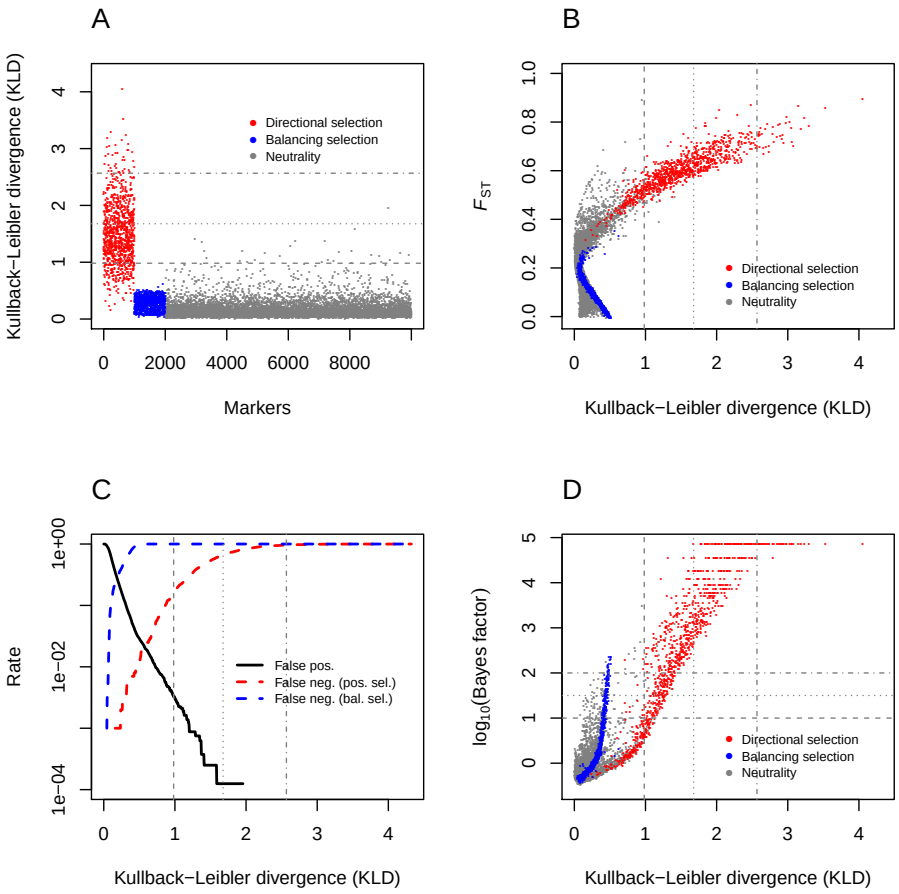

Figure S8 Analysis of the allele count data from dataset 8. (A) Kullback–Leibler divergence (KLD) measure between the posterior of δ_j and its centering distribution for all simulated loci. Loci under positive selection are depicted in red, loci under balancing selection in blue, and neutral markers are in grey. (B) F_{ST} as a function of the KLD measure for all loci. (C) False positive (neutral loci detected as outliers) and false negative (selected loci not detected as outliers) rates as a function of the KLD measure. (D) Relationship between the Bayes factor $\log_{10}$(BF) from the BAYESCAN analysis of dataset 8 and the KLD. The horizontal lines in (A) and the vertical lines in (B–D) indicate the KLD thresholds corresponding to the 95%-, the 99%- and the 99.9%-quantile of the of the KLD distribution from the pod analysis of dataset 8. In (D), the horizontal lines indicate the $\log_{10}$(BF) = 1, $\log_{10}$(BF) = 1.5 and $\log_{10}$(BF) = 2 thresholds, which correspond to ‘‘strong’’, ‘‘very strong’’ and ‘‘decisive’’ support, respectively, following Jeffreys’ (1961) scale of evidence.

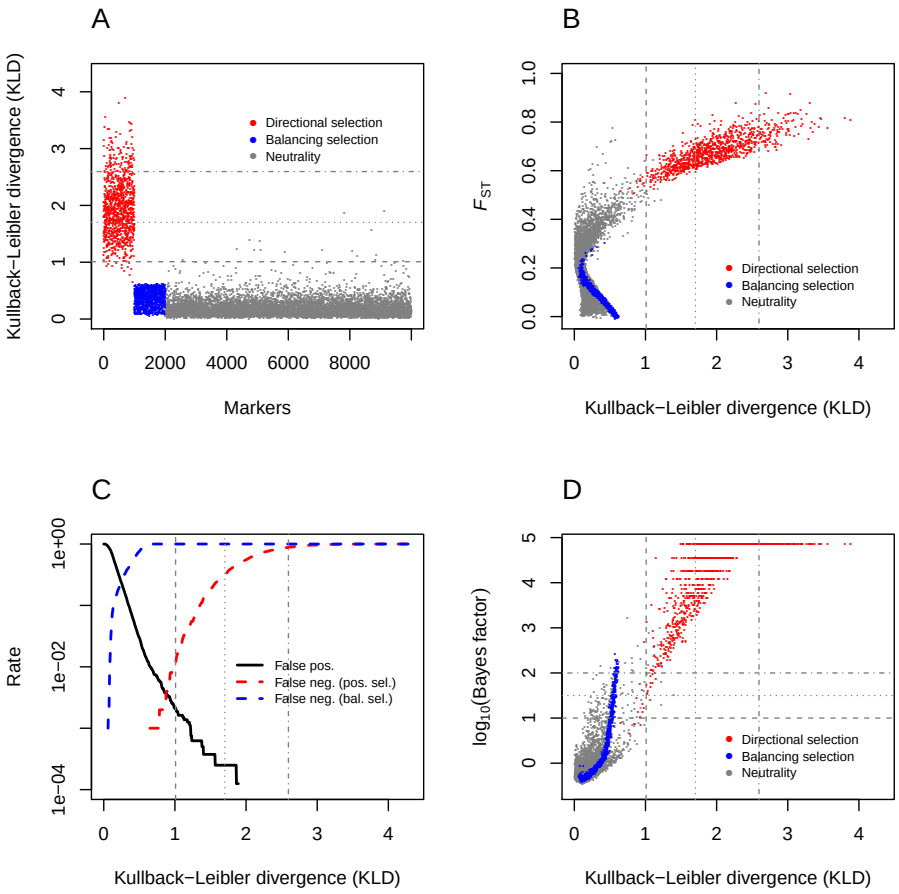

Figure S9 Analysis of the allele count data from dataset 9. (A) Kullback–Leibler divergence (KLD) measure between the posterior of δ_j and its centering distribution for all simulated loci. Loci under positive selection are depicted in red, loci under balancing selection in blue, and neutral markers are in grey. (B) F_{ST} as a function of the KLD measure for all loci. (C) False positive (neutral loci detected as outliers) and false negative (selected loci not detected as outliers) rates as a function of the KLD measure. (D) Relationship between the Bayes factor $\log_{10}$(BF) from the BAYESCAN analysis of dataset 9 and the KLD. The horizontal lines in (A) and the vertical lines in (B–D) indicate the KLD thresholds corresponding to the 95%-, the 99%- and the 99.9%-quantile of the of the KLD distribution from the pod analysis of dataset 9. In (D), the horizontal lines indicate the $\log_{10}$(BF) = 1, $\log_{10}$(BF) = 1.5 and $\log_{10}$(BF) = 2 thresholds, which correspond to ‘‘strong’’, ‘‘very strong’’ and ‘‘decisive’’ support, respectively, following Jeffreys’ (1961) scale of evidence.

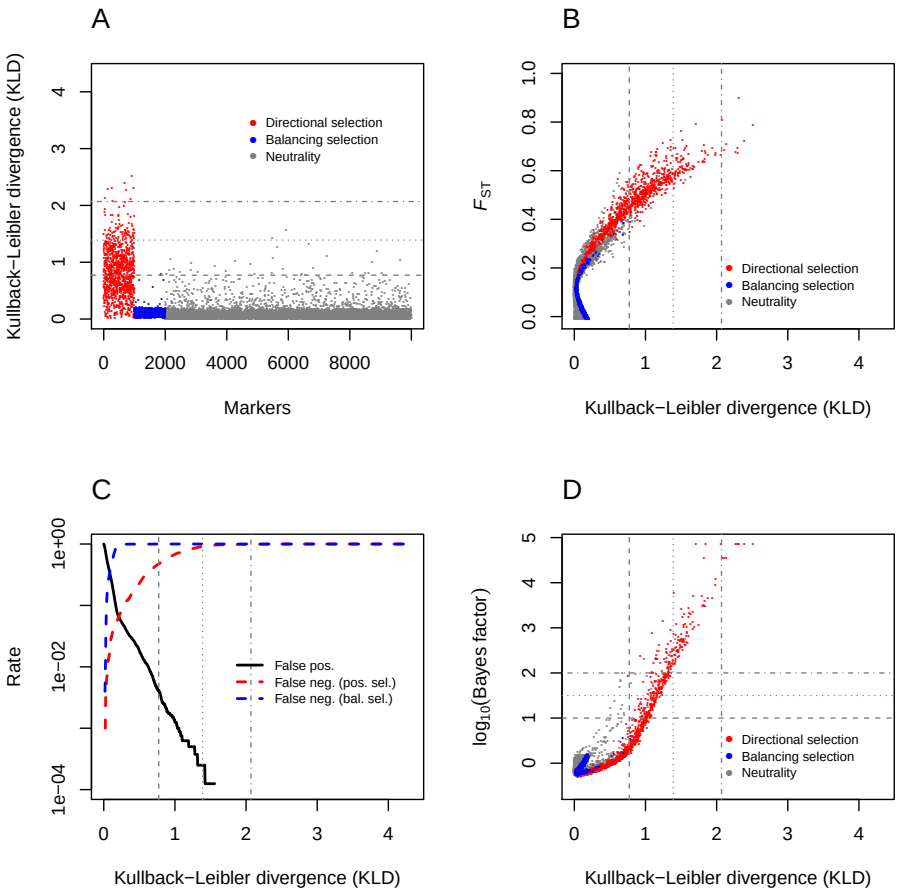

Figure S10 Analysis of the allele count data from dataset 10. (A) Kullback–Leibler divergence (KLD) measure between the posterior of δ_j and its centering distribution for all simulated loci. Loci under positive selection are depicted in red, loci under balancing selection in blue, and neutral markers are in grey. (B) F_{ST} as a function of the KLD measure for all loci. (C) False positive (neutral loci detected as outliers) and false negative (selected loci not detected as outliers) rates as a function of the KLD measure. (D) Relationship between the Bayes factor $\log_{10}$(BF) from the BAYESCAN analysis of dataset 10 and the KLD. The horizontal lines in (A) and the vertical lines in (B–D) indicate the KLD thresholds corresponding to the 95%-, the 99%- and the 99.9%-quantile of the of the KLD distribution from the pod analysis of dataset 10. In (D), the horizontal lines indicate the $\log_{10}$(BF) = 1, $\log_{10}$(BF) = 1.5 and $\log_{10}$(BF) = 2 thresholds, which correspond to ‘‘strong’’, ‘‘very strong’’ and ‘‘decisive’’ support, respectively, following Jeffreys’ (1961) scale of evidence.

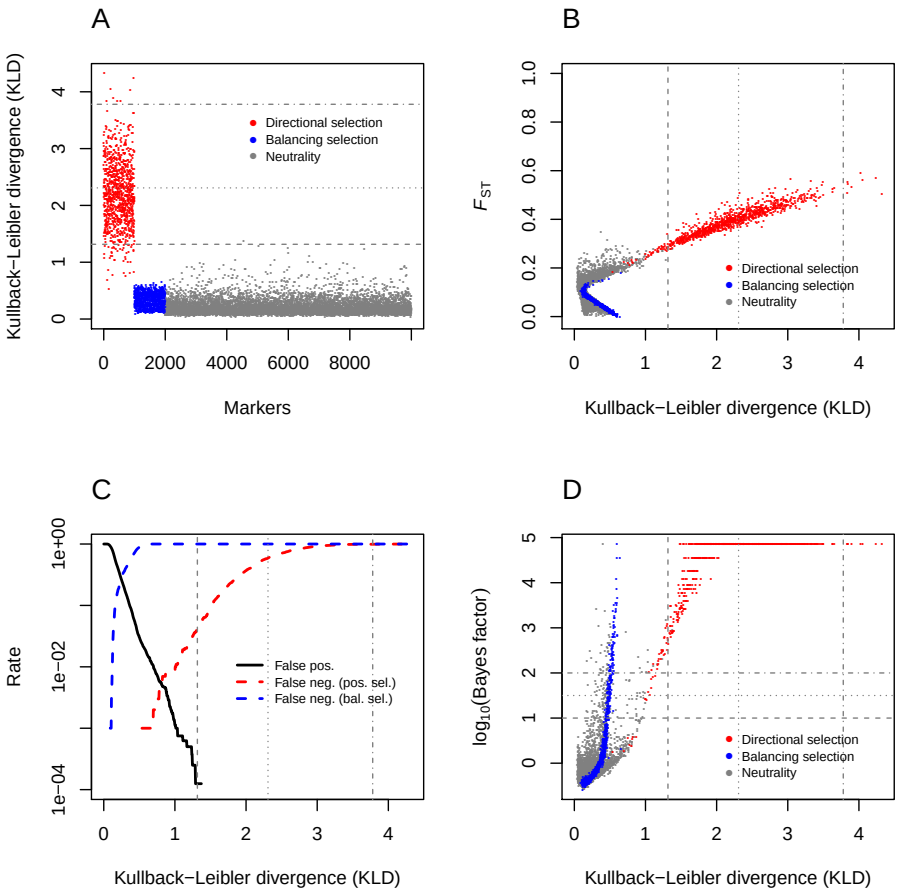

Figure S11 Analysis of the allele count data from dataset 11. (A) Kullback–Leibler divergence (KLD) measure between the posterior of δ_j and its centering distribution for all simulated loci. Loci under positive selection are depicted in red, loci under balancing selection in blue, and neutral markers are in grey. (B) F_{ST} as a function of the KLD measure for all loci. (C) False positive (neutral loci detected as outliers) and false negative (selected loci not detected as outliers) rates as a function of the KLD measure. (D) Relationship between the Bayes factor $\log_{10}$(BF) from the BAYESCAN analysis of dataset 11 and the KLD. The horizontal lines in (A) and the vertical lines in (B–D) indicate the KLD thresholds corresponding to the 95%-, the 99%- and the 99.9%-quantile of the of the KLD distribution from the pod analysis of dataset 11. In (D), the horizontal lines indicate the $\log_{10}$(BF) = 1, $\log_{10}$(BF) = 1.5 and $\log_{10}$(BF) = 2 thresholds, which correspond to ‘‘strong’’, ‘‘very strong’’ and ‘‘decisive’’ support, respectively, following Jeffreys’ (1961) scale of evidence.

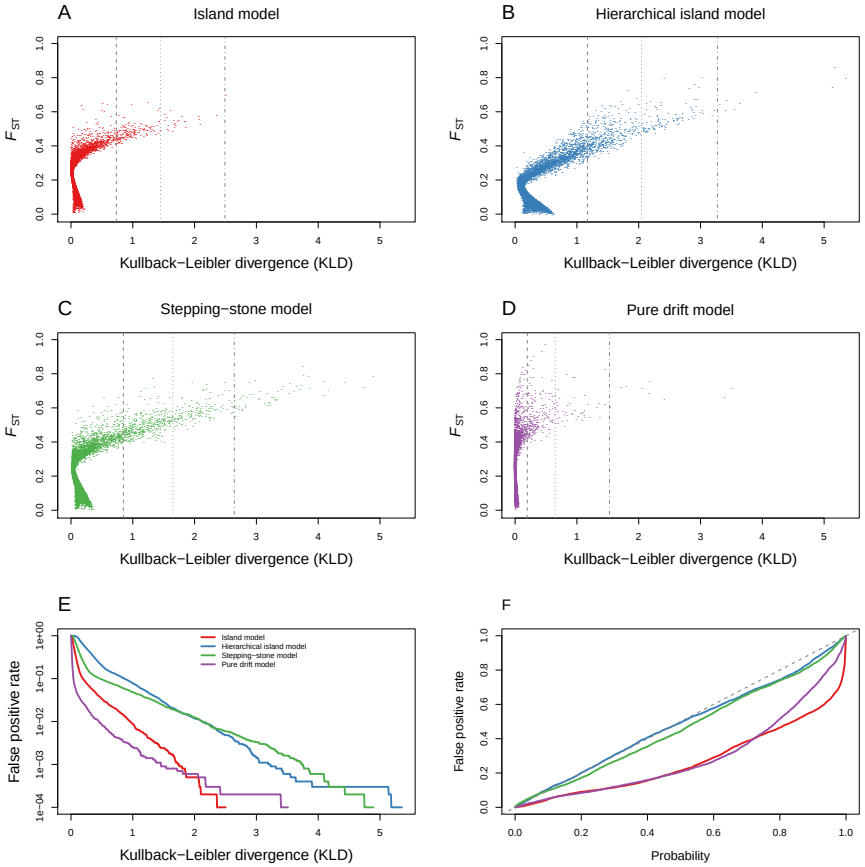

Figure S12 (A–D) SELESTIM analysis of the datasets from Figure S1. (E) False positive rate (neutral loci detected as outliers) as a function of the Kullback–Leibler divergence (KLD) threshold, for the datasets analyzed in (A–D). (F) False positive rate, as a function of the quantile probability. For each dataset analysis, pseudo-observed data (pod) are generated from the joint posterior distribution of the model parameters, using a rejection-sampling algorithm (see File S2). The pod is then analyzed, using the same MCMC parameters (number and length of pilot runs, burn-in, chain length, etc.) as for the analysis of the original data. Each quantile probability defines a KLD threshold, which is used for model choice between selection and neutrality.

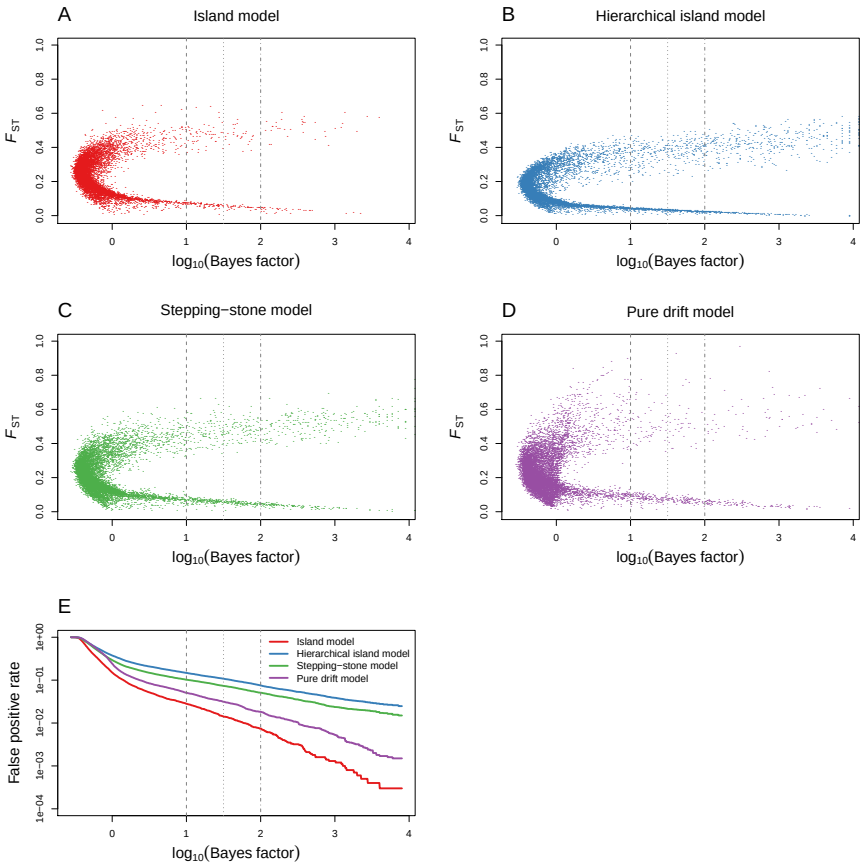

Figure S13 (A–D) BAYESCAN analyses of the datasets from Figure S1, using prior odds of 10 for the neutral model. (E) False positive rate as a function of the $\log_{10}$(BF) threshold. Vertical lines indicate the $\log_{10}$(BF) = 1, $\log_{10}$(BF) = 1.5 and $\log_{10}$(BF) = 2 thresholds, which correspond to ''strong'', ''very strong'' and ''decisive'' support, respectively, following Jeffreys' (1961) scale of evidence.

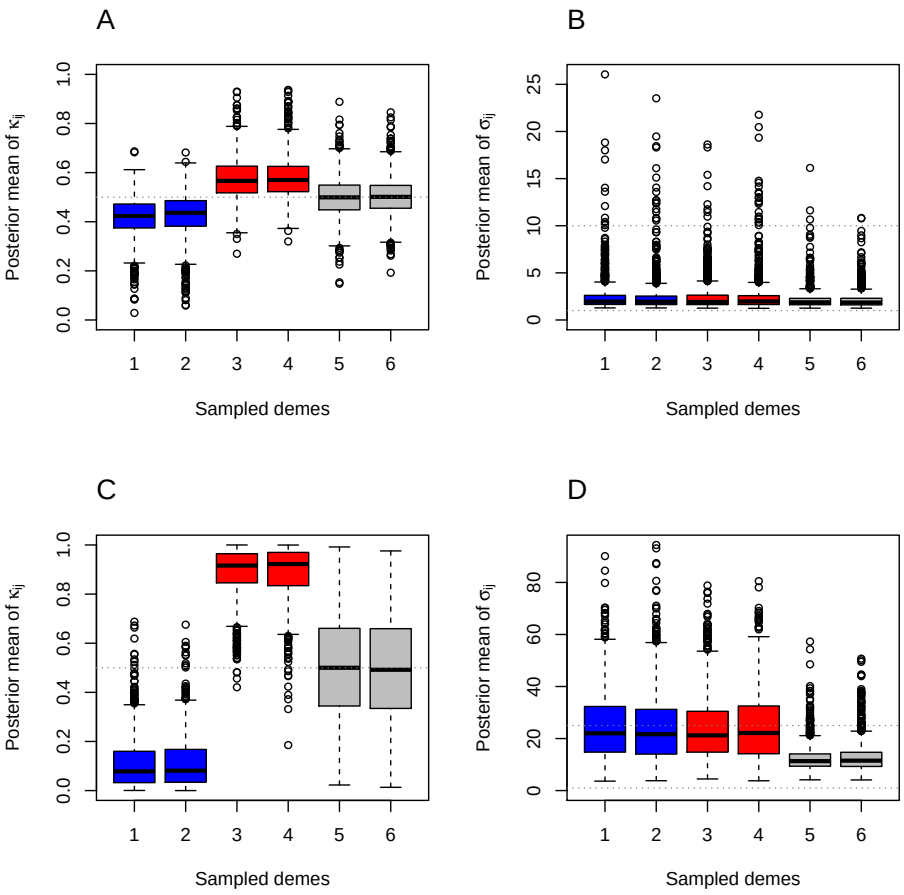

Figure S14 Analysis of the allele count data from datasets 1 and 2. (A) Boxplot representation of the posterior means of the parameters κ_{ij} (that indicate which allele is selected for) for the 1,000 positively selected loci in ''blue'' demes (1–2), ''red'' demes (3–4) and ''uncolored'' demes (5–6) in dataset 1. (B) Boxplot representation of the posterior means of the selection coefficients σ_{ij} for positively selected loci in dataset 1. For ''blue'' demes, the posterior means of the selection coefficients σ_{ij} are conditional upon the ''blue'' allele being selected for (κ_{ij} = 0). For ''red'' demes, the posterior means of the selection coefficients σ_{ij} are conditional upon the ''red'' allele being selected for (κ_{ij} = 1). The horizontal dotted lines indicate the true value of $\sigma_{ij} \equiv 2Ns_{ij}$ (top) and the prior mean σ_{ij} = 1 (bottom). For ''uncolored'' demes, the posterior means of the selection coefficients σ_{ij} are unconditional. (C) Idem as (A) for dataset 2. (D) Idem as (B) for dataset 2.

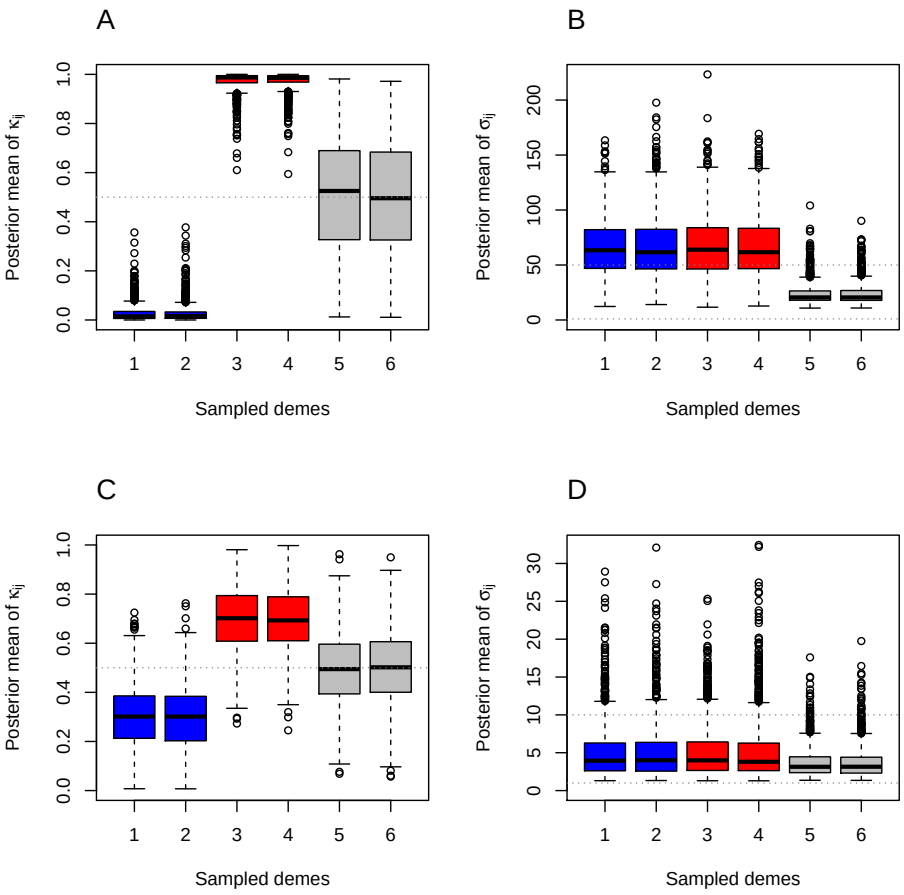

Figure S15 Analysis of the allele count data from datasets 3 and 4. (A) Boxplot representation of the posterior means of the parameters κ_{ij} (that indicate which allele is selected for) for the 1,000 positively selected loci in ''blue'' demes (1–2), ''red'' demes (3–4) and ''uncolored'' demes (5–6) in dataset 3. (B) Boxplot representation of the posterior means of the selection coefficients σ_{ij} for positively selected loci in dataset 3. For ''blue'' demes, the posterior means of the selection coefficients σ_{ij} are conditional upon the ''blue'' allele being selected for (κ_{ij} = 0). For ''red'' demes, the posterior means of the selection coefficients σ_{ij} are conditional upon the ''red'' allele being selected for (κ_{ij} = 1). The horizontal dotted lines indicate the true value of $\sigma_{ij} \equiv 2Ns_{ij}$ (top) and the prior mean σ_{ij} = 1 (bottom). For ''uncolored'' demes, the posterior means of the selection coefficients σ_{ij} are unconditional. (C) Idem as (A) for dataset 4. (D) Idem as (B) for dataset 4.

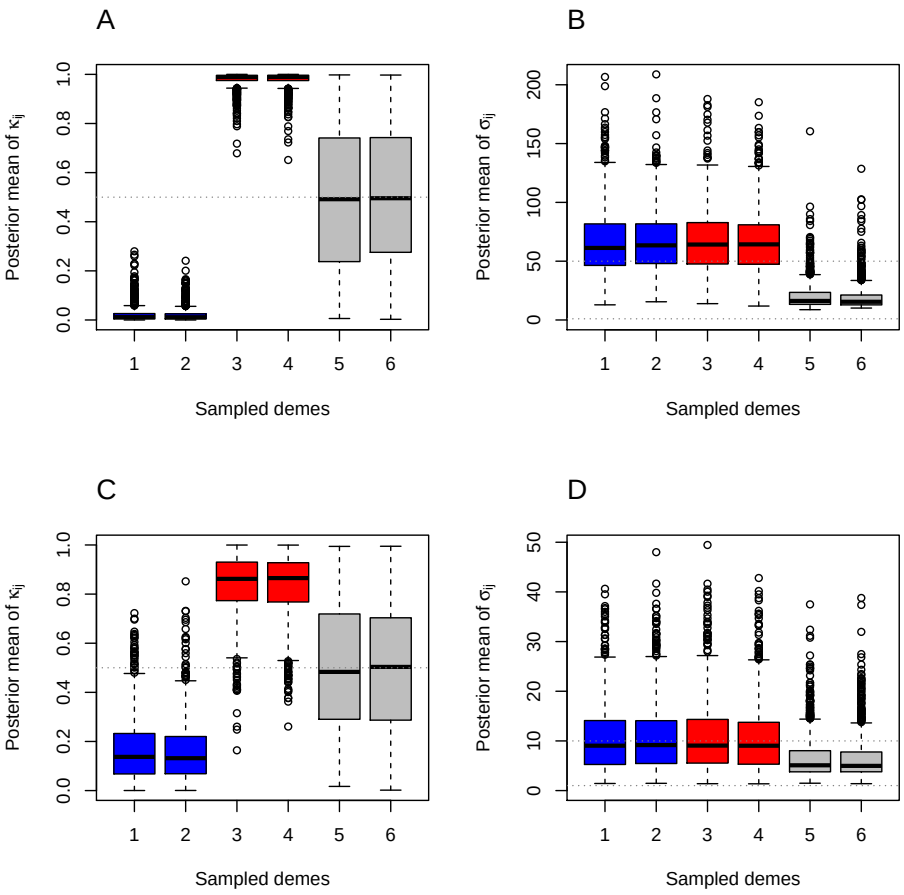

Figure S16 Analysis of the allele count data from datasets 6 and 7. (A) Boxplot representation of the posterior means of the parameters κ_{ij} (that indicate which allele is selected for) for the 1,000 positively selected loci in ''blue'' demes (1–2), ''red'' demes (3–4) and ''uncolored'' demes (5–6) in dataset 6. (B) Boxplot representation of the posterior means of the selection coefficients σ_{ij} for positively selected loci in dataset 6. For ''blue'' demes, the posterior means of the selection coefficients σ_{ij} are conditional upon the ''blue'' allele being selected for (κ_{ij} = 0). For ''red'' demes, the posterior means of the selection coefficients σ_{ij} are conditional upon the ''red'' allele being selected for (κ_{ij} = 1). The horizontal dotted lines indicate the true value of $\sigma_{ij} \equiv 2Ns_{ij}$ (top) and the prior mean σ_{ij} = 1 (bottom). For ''uncolored'' demes, the posterior means of the selection coefficients σ_{ij} are unconditional. (C) Idem as (A) for dataset 7. (D) Idem as (B) for dataset 7.

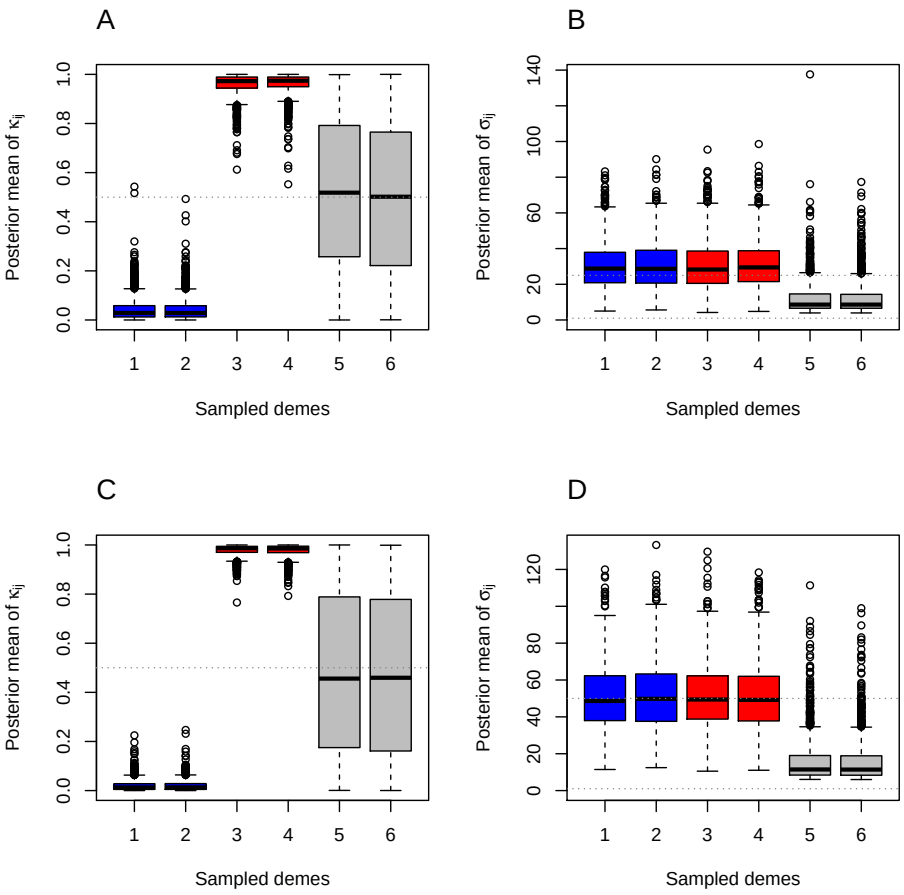

Figure S17 Analysis of the allele count data from datasets 8 and 9. (A) Boxplot representation of the posterior means of the parameters κ_{ij} (that indicate which allele is selected for) for the 1,000 positively selected loci in ''blue'' demes (1–2), ''red'' demes (3–4) and ''uncolored'' demes (5–6) in dataset 8. (B) Boxplot representation of the posterior means of the selection coefficients σ_{ij} for positively selected loci in dataset 8. For ''blue'' demes, the posterior means of the selection coefficients σ_{ij} are conditional upon the ''blue'' allele being selected for (κ_{ij} = 0). For ''red'' demes, the posterior means of the selection coefficients σ_{ij} are conditional upon the ''red'' allele being selected for (κ_{ij} = 1). The horizontal dotted lines indicate the true value of $\sigma_{ij} \equiv 2Ns_{ij}$ (top) and the prior mean σ_{ij} = 1 (bottom). For ''uncolored'' demes, the posterior means of the selection coefficients σ_{ij} are unconditional. (C) Idem as (A) for dataset 9. (D) Idem as (B) for dataset 9.

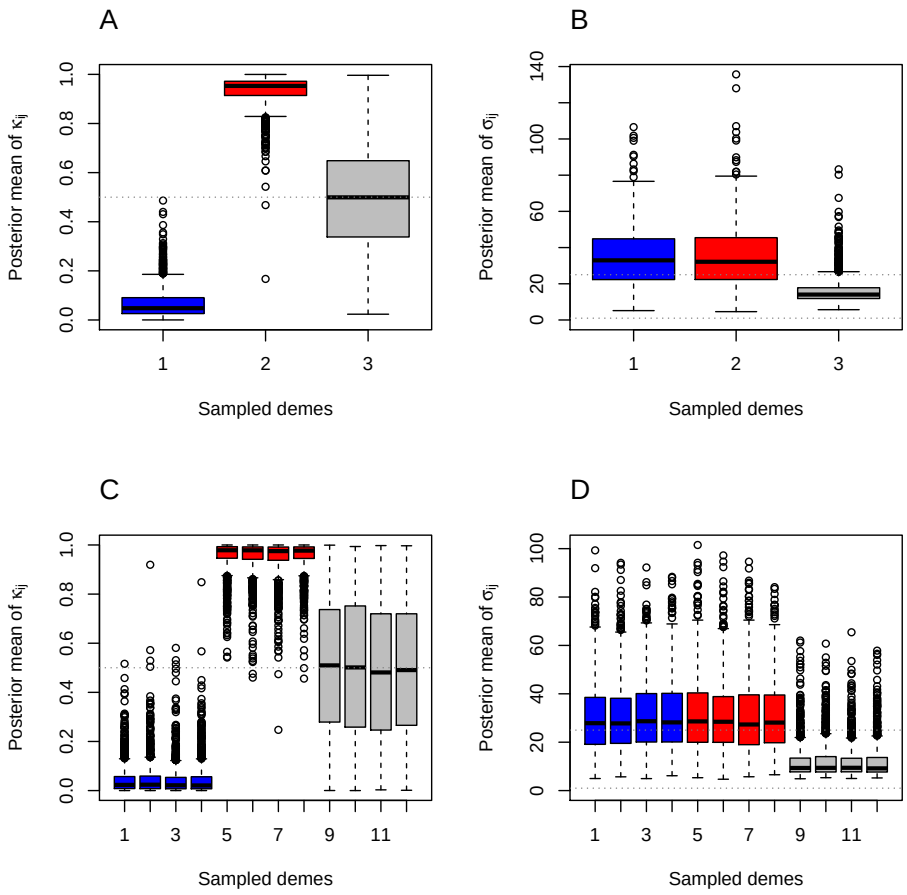

Figure S18 Analysis of the allele count data from datasets 10 and 11. (A) Boxplot representation of the posterior means of the parameters κ_{ij} (that indicate which allele is selected for) for the 1,000 positively selected loci in ''blue'' demes (1–2), ''red'' demes (3–4) and ''uncolored'' demes (5–6) in dataset 10. (B) Boxplot representation of the posterior means of the selection coefficients σ_{ij} for positively selected loci in dataset 10. For ''blue'' demes, the posterior means of the selection coefficients σ_{ij} are conditional upon the ''blue'' allele being selected for (κ_{ij} = 0). For ''red'' demes, the posterior means of the selection coefficients σ_{ij} are conditional upon the ''red'' allele being selected for (κ_{ij} = 1). The horizontal dotted lines indicate the true value of $\sigma_{ij} \equiv 2Ns_{ij}$ (top) and the prior mean σ_{ij} = 1 (bottom). For ''uncolored'' demes, the posterior means of the selection coefficients σ_{ij} are unconditional. (C) Idem as (A) for dataset 11. (D) Idem as (B) for dataset 11.

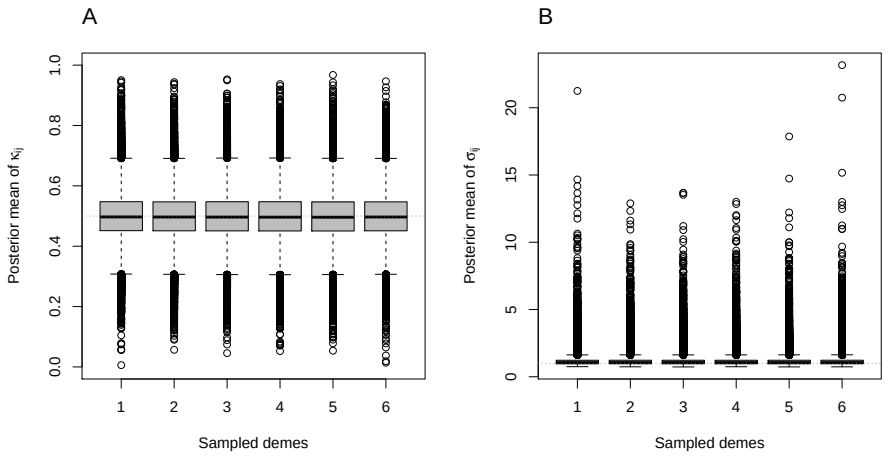

Figure S19 Analysis of the allele count data from a simulation of 50,000 neutral markers. The simulation was performed according to an island model with n_d = 50 ''uncolored'' demes, each made of N = 250 diploid individuals (500 genes). Samples were collected in six demes (50 individuals per deme). The migration rate was chosen to achieve the expected value of F_{ST} = 0.15, using equation 6 in Rousset (1996). The realized value was F_{ST} = 0.153 (multilocus estimate). (A) Boxplot representation of the posterior means of the parameters κ_{ij} (that indicate which allele is selected for) for the 50,000 neutral markers in ''uncolored'' demes (1–6). (B) Boxplot representation of the posterior means of the selection coefficients σ_{ij} for the 50,000 neutral markers (unconditional on κ_{ij}). The horizontal dotted line indicates the prior mean σ_{ij} = 1

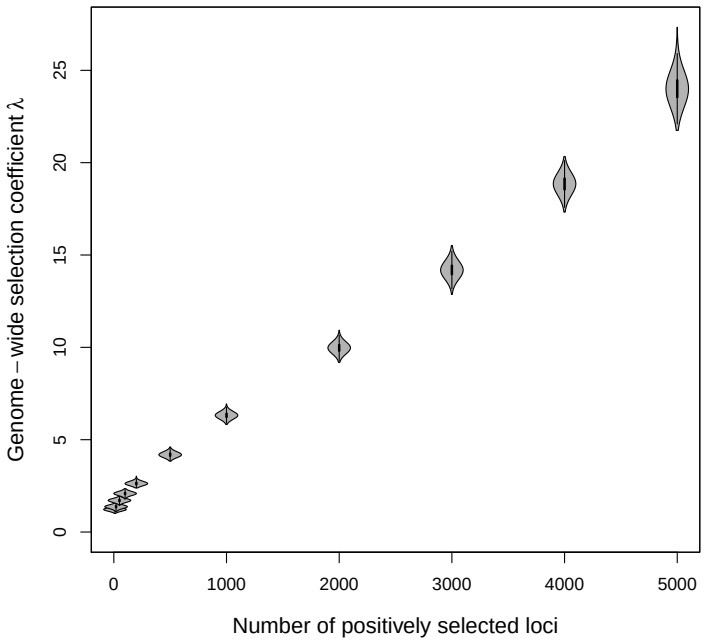

Figure S20 Posterior distributions (violin plot representation) of the genome-wide coefficient of selection λ as a function of the number of positively selected loci. We used the same parameter as for dataset 5 (see Table 1), but varying the proportion of selected loci from 10 to 5,000 out of 10,000 markers (hence, from 0.1% to 50%).

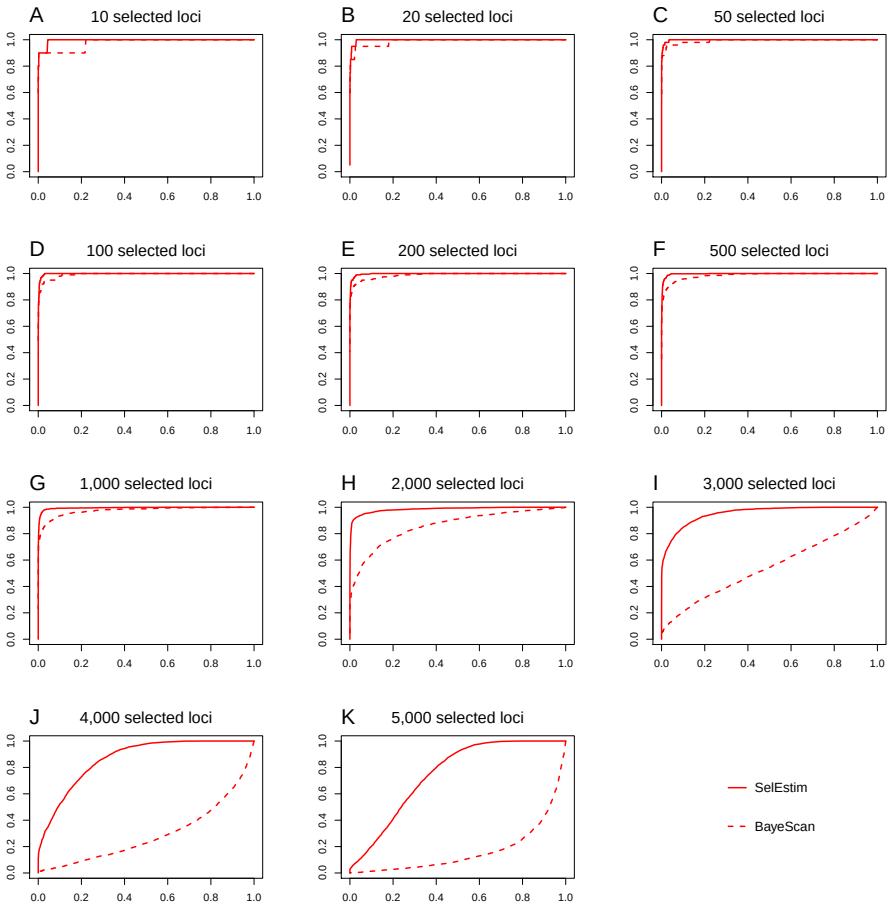

Figure S21 Receiver operating characteristic (ROC) analysis for the same datasets as in Figure S17 (from left to right, top to bottom). In the ROC analysis, the proportion of false positives and true positives is computed for each possible value of the threshold that is used to classify a locus under selection. For SELESTIM, the classifying variable was the KLD between the posterior distribution of the locus-specific coefficient of selection δ_j and its centering distribution, while in the case of BAYESCAN it was the Bayes factor.

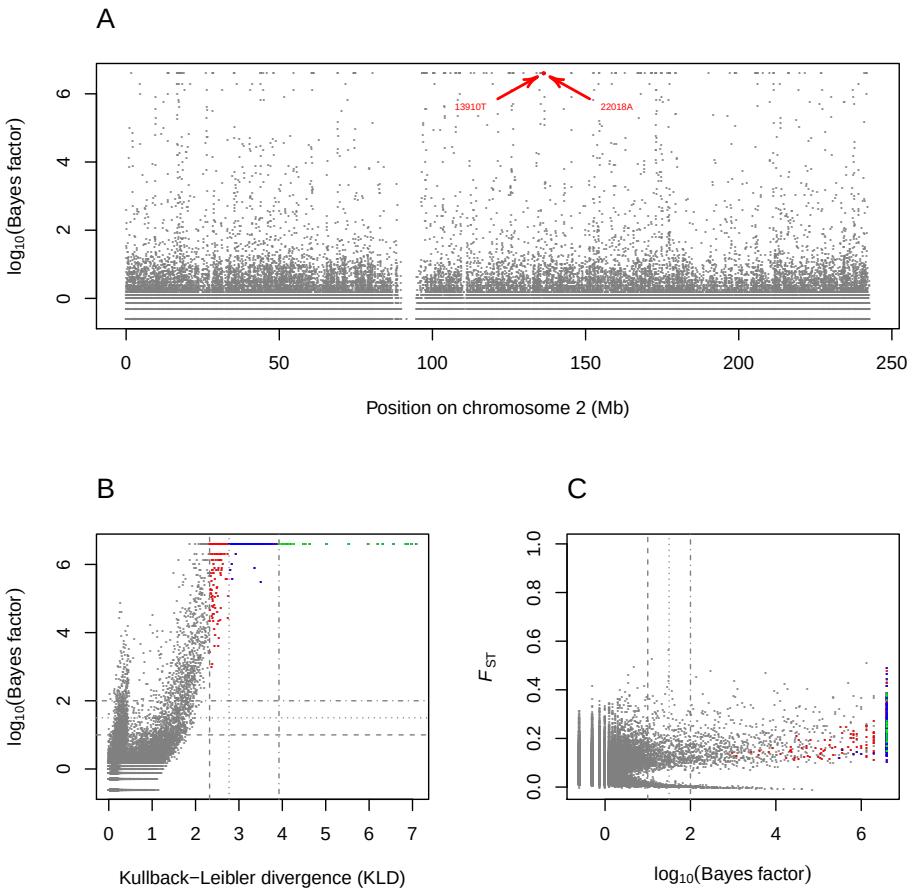

Figure S22 (A) BAYESCAN Bayes factor for the CEPH dataset analyses, along chromosome 2. The alleles -13910T and -22018A associated with lactase persistence are indicated in red. (B) Joint distribution of BAYESCAN Bayes factor and the Kullback–Leibler divergence (KLD) measure for all loci in the dataset. Markers in green have KLD ≥ 3.924, which corresponds to the 99.9%-quantile of the of the KLD distribution from the pod analysis; markers in blue have KLD ≥ 2.772, which corresponds to the 99.5%-quantile of the of the KLD distribution from the pod analysis; markers in red have KLD ≥ 2.324, which corresponds to the 99%-quantile of the of the KLD distribution from the pod analysis. (C) Joint distribution F_{ST} and BAYESCAN Bayes factor for all loci in the dataset.

Table S1 False positive rates using two calibration methods

	False positive rate (KLD)					
	Using McCulloch's (1989) calibration			Using pseudo-observed data		
Dataset	α = 5%	α = 1%	α = 0.1%	α = 5%	α = 1%	α = 0.1%
12	0.002 (1.164)	0.000 (1.959)	0.000 (3.108)	2.764 (0.011)	0.374 (0.045)	0.026 (0.211)
13	0.000 (1.164)	0.000 (1.959)	0.000 (3.108)	1.226 (0.016)	0.076 (0.076)	0.004 (0.349)
14	0.016 (1.164)	0.010 (1.959)	0.002 (3.108)	3.308 (0.035)	0.514 (0.174)	0.048 (0.801)
15	0.008 (1.164)	0.000 (1.959)	0.000 (3.108)	1.880 (0.091)	0.164 (0.434)	0.002 (1.520)
16	0.068 (1.164)	0.008 (1.959)	0.000 (3.108)	1.722 (0.247)	0.194 (0.853)	0.008 (2.019)
17	0.140 (1.164)	0.020 (1.959)	0.000 (3.108)	1.712 (0.374)	0.186 (1.047)	0.022 (1.942)
18	0.182 (1.164)	0.010 (1.959)	0.000 (3.108)	1.478 (0.521)	0.178 (1.179)	0.010 (1.948)

SelEstim analyses of datasets 12–18. Left-hand side: proportion (%) of markers that were classified as outliers, using the threshold KLD = 1.164, 1.959 and 3.108, which equal the KLD between two Bernoulli distributions corresponding to flipping a fair coin and a biased coin that gives a head with probability 0.05, 0.01 and 0.001, respectively. Right-hand side: proportion (%) of markers that were classified as outliers, using the calibration based on pseudo-observed data (pod). For each dataset and each analysis, a rejection sampling algorithm (see File S2) is used to generate a pod from the joint posterior distribution of the model parameters. The quantiles of the KLD distribution from the pod analysis are then used to calibrate the KLD: the (1 - α)%-quantile of the KLD distribution from the pod analysis provides a α%-threshold KLD value, which is then used for model choice between selection and neutrality.

Table S2 False positive rates using two calibration methods

	False positive rate (KLD)					
	Using McCulloch's (1989) calibration			Using pseudo-observed data		
Dataset	α = 5%	α = 1%	α = 0.1%	α = 5%	α = 1%	α = 0.1%
Island model	0.560 (1.164)	0.050 (1.959)	0.000 (3.108)	1.960 (0.734)	0.250 (1.450)	0.010 (2.491)
Hierarchy	5.680 (1.164)	1.260 (1.959)	0.110 (3.108)	5.580 (1.171)	1.140 (2.047)	0.090 (3.274)
IBD	1.900 (1.164)	0.920 (1.959)	0.330 (3.108)	7.830 (0.121)	3.140 (0.655)	0.680 (2.346)
Pure drift	0.160 (1.164)	0.060 (1.959)	0.020 (3.108)	2.680 (0.197)	0.540 (0.649)	0.090 (1.528)

SelEstim analyses of datasets from Figure S19. Left-hand side: proportion (%) of markers that were classified as outliers, using the threshold KLD = 1.164, 1.959 and 3.108, which equal the KLD between two Bernoulli distributions corresponding to flipping a fair coin and a biased coin that gives a head with probability 0.05, 0.01 and 0.001, respectively. Right-hand side: proportion (%) of markers that were classified as outliers, using the calibration based on pseudo-observed data (pod). For each dataset and each analysis, a rejection sampling algorithm (see File S2) is used to generate a pod from the joint posterior distribution of the model parameters. The quantiles of the KLD distribution from the pod analysis are then used to calibrate the KLD: the (1 - α)%-quantile of the KLD distribution from the pod analysis provides a α%-threshold KLD value, which is then used for model choice between selection and neutrality.

File S1

Details on the componentwise Markov chain Monte Carlo algorithm

Here we provide the computational details for the componentwise Markov chain Monte Carlo updates. Our aim is to sample from the joint posterior distribution of $f(\mathbf{M}, \pi, \kappa, \sigma, \delta, \lambda|\mathbf{n})$, which is specified by equation (4) and by the directed acyclic graph (DAG) in Figure 1. To do so, we use a combination of the Metropolis–Hastings algorithm and the Gibbs sampler for generating observations from $f(\mathbf{M}, \pi, \kappa, \sigma, \delta, \lambda|\mathbf{n})$ using outputs from a Markov chain (see, e.g., Gelman *et al.* 2004).

Each Markov chain is initialized with random values of the parameters drawn from their prior densities, except for the parameters p_{ij}, for which the observed frequencies are used, and the parameters π_js, for which the Laplace values are calculated from the dataset frequencies. The updating sequence is as follows: (*i*) all Ln_d parameters p_{ij}; (*ii*) all n_d parameters M_i; (*iii*) all L parameters π_j; (*iv*) the hyperparameter λ; (*v*) all L hyperparameters δ_j; (*vi*) all Ln_d parameters σ_{ij}; (*vii*) all Ln_d parameters κ_{ij}. Since the full posterior distribution of the model can be decomposed as a product over loci and over populations (see equation 4), each update only requires the re-computation of the relevant terms of the distribution $f(\mathbf{M}, \pi, \kappa, \sigma, \delta, \lambda|\mathbf{n})$. This improves the computational efficiency of the algorithm considerably.

The confluent hypergeometric, or Kummer's, functions ${}_1F_1(a; b; z)$ (see, e.g., Abramowitz and Stegun 1965, p. 504) were computed following a procedure proposed by Pearson (Pearson 2009), which is based on the power series definition of the function:

$${}_1F_1(a; b; z) = \sum_{j=0}^{\infty} \underbrace{\frac{(a)_j}{(b)_j} \frac{z^j}{j!}}_{A_j}, \tag{S1.1}$$

where, for some parameter p, the Pochhammer symbol $(p)_j$ is defined as:

$$(p)_0 = 1, \quad (p)_j = p(p+1)\dots(p+j-1), \quad \text{for } j = 1, 2, \dots . \tag{S1.2}$$

The computation of the terms of the power series in equation (S1.1) can then be car-

ried out using the following procedure:

$$\begin{aligned} A_0 &= S_0 = 1, \\ A_{j+1} &= A_j \frac{a+j}{b+j} \frac{z}{j+1}, \\ S_{j+1} &= S_j + A_{j+1}, \quad \text{for } j = 1, 2, \ldots \end{aligned} \tag{S1.3}$$

where A_j represents the $(j+1)$th term of the power series in equation (S1.1), and S_j represents the sum of the first $(j+1)$ terms. The computation was stopped when both $|A_N|/|S_{N-1}| < 10^{-12}$ and $|A_{N+1}|/|S_N| < 10^{-12}$. This criterion is equivalent to truncating the series in equation (S1.1), and requires that two consecutive terms to be small compared to the sum already computed.

Updating p_{ij}: The parameters p_{ij} are updated iteratively in each deme, one locus at a time. In the ith deme, at locus j, one allele is chosen at random from a Bernoulli trial with probability 0.5. The new allele frequency p'_{ij} is chosen as a random variable drawn from a uniform distribution around the current value p_{ij}:

$$p'_{ij} \sim U\left(p_{ij} - \Delta_p, p_{ij} + \Delta_p\right). \tag{S1.4}$$

The size of the interval Δ_p is a constant, which is adjusted during 25 short pilot runs of 1,000 iterations, in order to get acceptance rates between 0.25 and 0.40 (see, e.g., Gilks *et al.* 1996). Since p_{ij} is a frequency comprised between 0 and 1, if p'_{ij} is outside the interval $[0, 1]$, the excess is reflected back into the interval; that is, if $p'_{ij} < 0$ then p'_{ij} is reset to its absolute value $|p'_{ij}|$, and if $p'_{ij} > 1$ then p'_{ij} is reset to $2 - p'_{ij}$. This proposal is symmetric (Yang 2005). The updated allele frequency p'_{ij} is therefore accepted according to the appropriate Metropolis probability, which reads:

$$1 \wedge \frac{\mathcal{L}(p'_{ij}; \mathbf{n}_{ij}) \psi(p'_{ij}; M_i, \pi_j, \kappa_{ij}, \sigma_{ij})}{\mathcal{L}(p_{ij}; \mathbf{n}_{ij}) \psi(p_{ij}; M_i, \pi_j, \kappa_{ij}, \sigma_{ij})}. \tag{S1.5}$$

Equation (S1.5) can be rewritten as

$$1 \wedge \exp\left[\sigma_{ij}\left(\tilde{p}'_{ij} - \tilde{p}_{ij}\right)\right] \frac{p'^{x_{ij}+M_i\pi_j-1}_{ij}(1-p'_{ij})^{(n_{ij}-x_{ij})M_i+(1-\pi_j)-1}}{p^{x_{ij}+M_i\pi_j-1}_{ij}(1-p_{ij})^{(n_{ij}-x_{ij})M_i+(1-\pi_j)-1}}, \quad \text{(S1.6)}$$

where $\tilde{p}'_{ij} \equiv \kappa_{ij}(1-p'_{ij}) + (1-\kappa_{ij})p'_{ij}$.

Updating M_i: The parameters M_i are updated iteratively, one deme at a time. The proposed value M'_i is drawn from a lognormal distribution with median equal to the current value M_i, i.e.:

$$q(M_i \to M'_i) = \frac{1}{M'_i\nu_M\sqrt{2\pi}} \exp\left(\frac{-\ln(M'_i/M_i)^2}{2\nu_M^2}\right), \quad \text{(S1.7)}$$

where ν_M is the standard deviation on the log scale. The standard deviation ν_M is a constant, which is adjusted during 25 short pilot runs of 1,000 iterations, in order to get acceptance rates between 0.25 and 0.40. Because the lognormal jumping rule is asymmetric, a Metropolis–Hastings update is required in which the Metropolis ratio is weighted by the ratio of the forward and reverse proposal densities (which is sometimes referred to as the ''Hastings term'': see, e.g., Gelman *et al.* 2004, p. 291). This means that when some moves are more likely to happen (because of the asymmetry of the proposal distribution), their probability of acceptance is decreased proportionately. Here, the ratio $q(M'_i \to M_i)/q(M_i \to M'_i)$ reduces to M'_i/M_i. In order to avoid computational problems with excessively small or large M_i values, all moves falling outside the interval $[0.0011, 000]$ are discarded (i.e., the chain is kept unchanged). Otherwise, the proposed value M'_i is accepted according to the appropriate Metropolis–Hastings probability, which is:

$$1 \wedge \frac{\left[\prod_{j=1}^{L} \psi(p_{ij}; M'_i, \pi_j, \kappa_{ij}, \sigma_{ij})\right] f(M'_i) q(M'_i \to M_i)}{\left[\prod_{j=1}^{L} \psi(p_{ij}; M_i, \pi_j, \kappa_{ij}, \sigma_{ij})\right] f(M_i) q(M_i \to M'_i)}. \quad \text{(S1.8)}$$

Equation (S1.8) can be rewritten as

$$1\wedge\left[\frac{\Gamma(M_i)}{\Gamma(M'_i)}\right]^L \frac{\prod_{j=1}^{L}\Gamma(M_i\pi_j)\Gamma(M_i(1-\pi_j))_1F_1(M_i\tilde{\pi}_{ij};M_i;\sigma_{ij})p_{ij}^{M'_i\pi_j}(1-p_{ij})^{M'_i(1-\pi_j)}}{\prod_{j=1}^{L}\Gamma(M'_i\pi_j)\Gamma(M'_i(1-\pi_j))_1F_1(M'_i\tilde{\pi}_{ij};M'_i;\sigma_{ij})p_{ij}^{M_i\pi_j}(1-p_{ij})^{M_i(1-\pi_j)}} \tag{S1.9}$$

Updating π_j: The parameters π_j are updated iteratively, one locus at a time. In the ith deme, at locus j, one allele is chosen at random from a Bernoulli trial with probability 0.5. The proposed allele frequency π'_j is chosen as a random variable drawn from a uniform distribution around the current value π_j:

$$\pi'_j \sim U\left(\pi_j-\Delta_\pi,\pi_j+\Delta_\pi\right). \tag{S1.10}$$

The size of the interval Δ_π is a constant, which is adjusted during 25 short pilot runs of 1,000 iterations, in order to get acceptance rates between 0.25 and 0.40. Since π_j is a frequency comprised between 0 and 1, if π'_j is outside the interval $[0,1]$, the excess is reflected back into the interval; that is, if $\pi'_j<0$ then π'_j is reset to its absolute value $|\pi'_j|$, and if $\pi'_j>1$ then π'_j is reset to $2-\pi'_j$. This proposal is symmetric, and the updated allele frequency π'_j is therefore accepted according to the appropriate Metropolis probability, which reads:

$$1\wedge\frac{\left[\prod_{i=1}^{n_\mathrm{d}}\psi(p_{ij};M_i,\pi'_j,\kappa_{ij},\sigma_{ij})\right]f(\pi'_j)}{\left[\prod_{i=1}^{n_\mathrm{d}}\psi(p_{ij};M_i,\pi_j,\kappa_{ij},\sigma_{ij})\right]f(\pi_j)}. \tag{S1.11}$$

Equation (S1.11) can be rewritten as

$$1\wedge\frac{\prod_{i=1}^{n_\mathrm{d}}\Gamma(M_i\pi_j)\Gamma(M_i(1-\pi_j))_1F_1(M_i\tilde{\pi}_{ij};M_i;\sigma_{ij})p_{ij}^{M_i\pi'_j}(1-p_{ij})^{M_i(1-\pi'_j)}}{\prod_{i=1}^{n_\mathrm{d}}\Gamma(M_i\pi'_j)\Gamma(M_i(1-\pi'_j))_1F_1(M_i\tilde{\pi}'_{ij};M_i;\sigma_{ij})p_{ij}^{M_i\pi_j}(1-p_{ij})^{M_i(1-\pi_j)}}, \tag{S1.12}$$

where $\tilde{\pi}'_{ij}\equiv\kappa_{ij}(1-\pi'_j)+(1-\kappa_{ij})\pi'_j$.

Updating λ: The proposed value of the hyperparameter λ' is drawn from a lognormal distribution with median equal to the current value λ, i.e.:

$$q(\lambda \to \lambda') = \frac{1}{\lambda' \nu_\lambda \sqrt{2\pi}} \exp\left(\frac{-\ln(\lambda'/\lambda)^2}{2\nu_\lambda^2}\right), \tag{S1.13}$$

where ν_λ is the standard deviation on the log scale. The standard deviation ν_λ is a constant, which is adjusted during 25 short pilot runs of 1,000 iterations, in order to get acceptance rates between 0.25 and 0.40. Because the lognormal jumping rule is asymmetric, a Metropolis–Hastings update is required in which the Metropolis ratio is weighted by the ratio of the forward and reverse proposal densities. This means that when some moves are more likely to happen (because of the asymmetry of the proposal distribution), their probability of acceptance is decreased proportionately. Here, the ratio $q(\lambda' \to \lambda)/q(\lambda \to \lambda')$ reduces to λ'/λ. In order to avoid computational problems with excessively small or large λ' values, all moves falling outside the interval $[0, 500]$ are discarded (i.e., the chain is kept unchanged). Otherwise, the proposed value λ' is accepted according to the appropriate Metropolis–Hastings probability, which is:

$$1 \wedge \frac{\left[\prod_{j=1}^{L} f(\delta_j|\lambda')\right] f(\lambda'|\Lambda) q(\lambda' \to \lambda)}{\left[\prod_{j=1}^{L} f(\delta_j|\lambda)\right] f(\lambda|\Lambda) q(\lambda \to \lambda')}. \tag{S1.14}$$

Equation (S1.14) can be rewritten as

$$1 \wedge \left(\frac{\lambda}{\lambda'}\right)^{L-1} \exp\left[(\lambda' - \lambda)\left(\frac{\sum_{j=1}^{L} \delta_j}{\lambda\lambda'} - \frac{1}{\Lambda}\right)\right] \tag{S1.15}$$

Updating δ_j: The parameters δ_j are updated iteratively, one locus at a time. The proposed value of the hyperparameters δ_j' is drawn from a lognormal distribution with median equal to the current value δ_j, i.e.:

$$q(\delta_j \to \delta_j') = \frac{1}{\delta_j' \nu_\delta \sqrt{2\pi}} \exp\left(\frac{-\ln(\delta_j'/\delta_j)^2}{2\nu_\delta^2}\right), \tag{S1.16}$$

where ν_δ is the standard deviation on the log scale. The standard deviation ν_δ is a constant, which is adjusted during 25 short pilot runs of 1,000 iterations, in order to get acceptance rates between 0.25 and 0.40. Because the lognormal jumping rule is asymmetric, a Metropolis–Hastings update is required in which the Metropolis ratio is weighted by the ratio of the forward and reverse proposal densities. This means that when some moves are more likely to happen (because of the asymmetry of the proposal distribution), their probability of acceptance is decreased proportionately. Here, the ratio $q(\delta'_j \to \delta_j)/q(\delta_j \to \delta'_j)$ reduces to δ'_j/δ_j. In order to avoid computational problems with excessively small or large δ_j values, all moves falling outside the interval $[0, 500]$ are discarded (i.e., the chain is kept unchanged). Otherwise, the proposed value δ'_j is accepted according to the appropriate Metropolis–Hastings probability, which is:

$$1 \wedge \frac{\left[\prod_{i=1}^{n_\mathrm{d}} f(\sigma_{ij}|\delta'_j)\right] f(\delta'_j|\lambda) q(\delta'_j \to \delta_j)}{\left[\prod_{i=1}^{n_\mathrm{d}} f(\sigma_{ij}|\delta_j)\right] f(\delta_j|\lambda) q(\delta_j \to \delta'_j)}. \tag{S1.17}$$

Equation (S1.17) can be rewritten as

$$1 \wedge \left(\frac{\delta_j}{\delta'_j}\right)^{n_\mathrm{d}-1} \exp\left[\left(\delta'_j - \delta_j\right)\left(\frac{\sum_{i=1}^{n_\mathrm{d}} \sigma_{ij}}{\delta_j \delta'_j} - \frac{1}{\lambda}\right)\right] \tag{S1.18}$$

Updating σ_{ij}: The parameters σ_{ij} are updated iteratively in each deme, one locus at a time. In the ith deme, at locus j, the proposed value of the parameters σ'_{ij} is drawn from a lognormal distribution with median equal to the current value σ_{ij}, i.e.:

$$q(\sigma_{ij} \to \sigma'_{ij}) = \frac{1}{\sigma'_{ij}\nu_\sigma\sqrt{2\pi}} \exp\left(\frac{-\ln(\sigma'_{ij}/\sigma_{ij})^2}{2\nu_\sigma^2}\right), \tag{S1.19}$$

where ν_σ is the standard deviation on the log scale. The standard deviation ν_σ is a constant, which is adjusted during 25 short pilot runs of 1,000 iterations, in order to get acceptance rates between 0.25 and 0.40. Because the lognormal jumping rule is asymmetric, a Metropolis–Hastings update is required in which the Metropolis ratio is weighted by the ratio of the forward and reverse proposal densities. This means that

when some moves are more likely to happen (because of the asymmetry of the proposal distribution), their probability of acceptance is decreased proportionately. Here, the ratio $q(\sigma'_{ij} \to \sigma_{ij})/q(\sigma_{ij} \to \sigma'_{ij})$ reduces to σ'_{ij}/σ_{ij}. In order to avoid computational problems with excessively small or large σ_{ij} values, all moves falling outside the interval $[0, 500]$ are discarded (i.e., the chain is kept unchanged). Otherwise, the proposed value σ'_{ij} is accepted according to the appropriate Metropolis–Hastings probability, which is:

$$\frac{\psi(p_{ij}; M_i, \pi_j, \kappa_{ij}, \sigma'_{ij}) f(\sigma'_{ij}|\delta_j) q(\sigma'_{ij} \to \sigma_{ij})}{\psi(p_{ij}; M_i, \pi_j, \kappa_{ij}, \sigma_{ij}) f(\sigma_{ij}|\delta_j) q(\sigma_{ij} \to \sigma'_{ij})}. \tag{S1.20}$$

Equation (S1.20) can be rewritten as

$$\frac{\sigma'_{ij}}{\sigma_{ij}} \exp\left[(\sigma'_{ij} - \sigma_{ij})\left(\tilde{p}_{ij} - \frac{1}{\delta_j}\right)\right] \frac{{}_1F_1(M_i\tilde{\pi}_{ij}; M_i; \sigma_{ij})}{{}_1F_1(M_i\tilde{\pi}_{ij}; M_i; \sigma'_{ij})}. \tag{S1.21}$$

Updating κ_{ij}: The parameters κ_{ij} are updated iteratively in each deme, one locus at a time. In the ith deme, at locus j, the variable κ_{ij}, which indicates which of the two alleles is selected for, is updated using Gibbs sampling based on the conditional posterior distribution:

$$f(\kappa_{ij}|\theta_{[-\kappa_{ij}]}) \propto \psi(p_{ij}; M_i, \pi_j, \kappa_{ij}, \sigma_{ij}) f(\kappa_{ij}), \tag{S1.22}$$

where $\theta_{[-\kappa_{ij}]}$ represents all the model parameters but κ_{ij}. Since κ_{ij} can only take two integer values (0 and 1), it can be shown that:

$$\Pr(\kappa_{ij} = 0|\theta_{[-\kappa_{ij}]}) \propto \frac{1}{2}\left[\frac{\exp\left[\sigma_{ij}p_{ij}\right]}{{}_1F_1(M_i\pi_j; M_i; \sigma_{ij})}\right], \tag{S1.23}$$

and

$$\Pr(\kappa_{ij} = 1|\theta_{[-\kappa_{ij}]}) \propto \frac{1}{2}\left[\frac{\exp\left[\sigma_{ij}(1-p_{ij})\right]}{{}_1F_1(M_i(1-\pi_j); M_i; \sigma_{ij})}\right]. \tag{S1.24}$$

Therefore, the conditional posterior distribution of $\left(\kappa_{ij}|\theta_{[-\kappa_{ij}]}\right)$ from equation (S1.22) can be rewritten as

$$\left(\kappa_{ij}|\theta_{[-\kappa_{ij}]}\right) \sim \text{Bernoulli}\left(\rho\right), \tag{S1.25}$$

where

$$\begin{aligned} \rho &\equiv \frac{\Pr(\kappa_{ij}=0|\theta_{[-\kappa_{ij}]})}{\Pr(\kappa_{ij}=0|\theta_{[-\kappa_{ij}]}) + \Pr(\kappa_{ij}=1|\theta_{[-\kappa_{ij}]})} \\ &= \left[1 + \frac{{}_1F_1(M_i\pi_{ij}; M_i; \sigma_{ij})}{{}_1F_1(M_i(1-\pi_{ij}); M_i; \sigma_{ij})} \exp\left[\sigma_{ij}(1-2p_{ij})\right]\right]^{-1}. \end{aligned} \tag{S1.26}$$

Literature Cited

Abramowitz, M., and I. A. Stegun, 1965 *Handbook of Mathematical Functions*. Dover Publication, Inc., New York.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin, 2004 *Bayesian Data Analysis*. Chapman & Hall, New York, 2nd edition.

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter, 1996 *Markov Chain Monte Carlo in Practice*. Chapman & Hall, New York, 2nd edition.

Pearson, J., 2009 *Computation of Hypergeometric Functions*. Ph.D. thesis, University of Oxford.

Yang, Z., 2005 Bayesian inference in molecular phylogenetics. In O. Gascuel, editor, *Mathematics of Evolution and Phylogeny*. Oxford University Press, Oxford, 63–90.

File S2

Details on the algorithm to sample from the inference model

In order to provide a decision criterion for discriminating between neutral and selected markers, we calibrate the Kullback–Leibler divergence (KLD) using simulations from a predictive distribution based on the observed data set. To that end, we generate pseudo-observed data as follows.

We set the hyperparameters M_i, π_j and λ to their respective posterior means $\bar{M}_i$, $\bar{\pi}_j$ and $\bar{\lambda}$, as estimated from the MCMC. Then we draw δ_j from an exponential distribution $\sim \exp(\bar{\lambda}^{-1})$ and we draw σ_{ij} from an exponential distribution $\sim \exp\left(\delta_j^{-1}\right)$. Last, the parameter κ_{ij} is drawn from a Bernoulli distribution (with parameter the posterior mean $\bar{\kappa}_{ij}$).

We aim at sampling the allele frequency p_{ij} from the distribution with density $f(p_{ij})$ defined by equations 2 and 3 in the main text. Because the cumulative distribution function of the distribution with density $f(p_{ij})$ is not tractable, we use a rejection-sampling algorithm. To that end, we define an instrumental distribution $g(p_{ij}) \sim \mathrm{Beta}(M_i\pi_j, M_i(1-\pi_j))$, with density:

$$g(p_{ij}) = \frac{\Gamma(M_i)}{\Gamma(M_i\pi_j)\Gamma(M_i(1-\pi_j))} p_{ij}^{M_i\pi_j - 1}(1-p_{ij})^{M_i(1-\pi_j)-1} \qquad \text{(S2.1)}$$

We further need to define a constant u, such that $f(p_{ij}) \leq [ug(p_{ij})]$ over the support $[0,1]$. Noting that:

$$\frac{f(p_{ij})}{g(p_{ij})} = \frac{\exp(\sigma_{ij}\tilde{p}_{ij})}{{}_1F_1(M_i\tilde{\pi}_{ij}; M_i; \sigma_{ij})} \qquad \text{(S2.2)}$$

then, if we define $u \equiv \exp(\sigma_{ij})/{}_1F_1(M_i\tilde{\pi}_{ij}; M_i; \sigma_{ij})$ we get:

$$\frac{f(p_{ij})}{ug(p_{ij})} = \exp(\sigma_{ij}(\tilde{p}_{ij} - 1)) \qquad \text{(S2.3)}$$

Since $0 \leq \tilde{p}_{ij} \leq 1$ and $\sigma_{ij} \geq 0$, by definition, we have $\exp(\sigma_{ij}(\tilde{p}_{ij} - 1)) \leq 1$ and therefore $f(p_{ij}) \leq [ug(p_{ij})]$. A straightforward algorithm to sample from the distribution with density $f(p_{ij})$ is then:

(1) Sample x from a beta distribution $\mathrm{Beta}(M_i\pi_j, M_i(1-\pi_j))$ and y from $\mathcal{U}(0,1)$ (the uniform distribution over the unit interval).

(2) Check whether or not $y < f(x)/[ug(x)]$ or equivalently (see equation S2.3) if $\log(y) < \sigma_{ij}(\tilde{p}_{ij} - 1)$:

- If this holds, accept x and set $\tilde{p}_{ij} = x$;
- if not, reject the value of x and repeat the sampling step (1).

(3) Compute $p_{ij} = \tilde{p}_{ij}(1-\kappa_{ij}) + (1-\tilde{p}_{ij})\kappa_{ij}$.

Finally, we draw the allele counts $\mathbf{n}_{ij}$ in the ith deme at the jth locus by a random draw from the binomial distribution $\sim \mathcal{B}(n_{ij}, p_{ij})$. We repeat this procedure for each locus j in each deme i.

This algorithm is computationally efficient, since it avoids computing ${}_1F_1(M_i\tilde{\pi}_{ij}; M_i; \sigma_{ij})$ (see equations 2 and 3 in the main text). However, the efficiency of the algorithm may be very low for large values of σ_{ij}. This is so because the expected number of iterations required until an x is successfully generated is exactly the bounding constant $u \equiv \exp(\sigma_{ij})/{}_1F_1(M_i\tilde{\pi}_{ij}; M_i; \sigma_{ij})$. Therefore, to avoid the algorithm getting stuck in very long loops, we adopt an alternative strategy whenever $u > 10^4$: in such case, we draw x from a beta distribution $\mathrm{Beta}(\alpha, \beta)$ with the same first two moments as the target distribution (equations 2 and 3 in the main text). Little algebra shows that: $\alpha = m_1(m_2 - m_1)/(m_1^2 - m_2)$ and $\beta = \alpha(1/m_1 - 1)$, where

$$m_1 = \tilde{\pi}_{ij}\left(\frac{{}_1F_1(M_i\tilde{\pi}_{ij}+1; M_i+1; \sigma_{ij})}{{}_1F_1(M_i\tilde{\pi}_{ij}; M_i; \sigma_{ij})}\right) \tag{S2.4}$$

and

$$m_2 = \tilde{\pi}_{ij}\left(\frac{M_i\tilde{\pi}_{ij}+1}{M_i+1}\right)\left(\frac{{}_1F_1(M_i\tilde{\pi}_{ij}+2; M_i+2; \sigma_{ij})}{{}_1F_1(M_i\tilde{\pi}_{ij}; M_i; \sigma_{ij})}\right) \tag{S2.5}$$