

File S2

Parameter estimation

Given neutral molecular data and the breeding experiment data, we fitted a neutral model to the data using Bayesian inference, and then tested for deviation from neutrality using the approach described in the main paper. Here we describe the statistical model, the prior distributions and a Markov Chain Monte Carlo (MCMC) method that was used to parameterize the model.

The statistical model for neutral molecular markers

We model the neutral divergence of the n_P local populations from a common ancestral population as a mixture of n_L independent lineages. The model summarized in this section will be published later in more detail (KARHUNEN and OVASKAINEN 2011).

The allele frequencies in locus j in lineage k are distributed as

$$\mathbf{z}_{kj} \sim \text{Dirichlet}(a_k \mathbf{q}_j).$$

where \mathbf{q}_j denotes the vector of allele frequencies in the ancestral population, and a_k measures the amount of drift experienced by lineage k . The allele frequencies in locus j in local population A are defined as a mixture of the lineage-specific frequencies,

$$\mathbf{p}_j^A = \sum_{k=1}^{n_L} \kappa_k^A \mathbf{z}_{kj}.$$

Here, we assume that the mixture weights κ_k^A sum up to unity over the lineages, $\sum_{k=1}^{n_L} \kappa_k^A = 1$, so that vector \mathbf{p}_j^A is a proper frequency distribution. The genotype (in terms of the neutral marker loci) of each individual in a population A is a multinomial random variable of the allele frequencies, $x_{ij} \sim \text{Multinomial}(2, \mathbf{p}_j^A)$.

The population-population coancestry coefficients depend on the model parameters as

$$\theta_{AB}^P = \sum_{k=1}^{n_L} \frac{\kappa_k^A \kappa_k^B}{a_k + 1}.$$

Note that θ_{AB}^P is defined through probability of IBD for neutral loci, and thus, it does not depend on allele frequencies \mathbf{q}_j in the ancestral generation.

The Directed Acyclic Graph (DAG) that describes the dependencies among model variables (including both neutral marker data and quantitative trait data) is shown in Supplementary Figure 1. While the framework develop in the paper and presented in the DAG allows for the inclusion of an arbitrary level of inbreeding (θ_X^S), we did not account for inbreeding in the present estimation scheme and thus assumed $\theta_X^S = 0.5(1 + \theta_X^P)$.

Prior distributions

To parameterize the model with Bayesian inference, prior distributions need to be defined for the primary parameters of the model: $\boldsymbol{\mu}$, \mathbf{G}^A , \mathbf{V}_E , \mathbf{q} , \mathbf{a} and $\boldsymbol{\kappa}$. We assume the priors

$$\mu_m \sim N(0, 100),$$

$$\mathbf{G}^A \sim \text{Wishart}(\mathbf{I}_m (m + 1)^{-1}, m + 1),$$

$$\mathbf{V}_E \sim \text{Wishart}(\mathbf{I}_m (m + 1)^{-1}, m + 1),$$

$$\mathbf{q}_j \sim \text{Dirichlet}(\boldsymbol{\beta}_j^q),$$

$$\log a_k \sim N(\mu_a, \sigma_a^2),$$

$$\boldsymbol{\kappa}_A \sim \text{Dirichlet}(\boldsymbol{\beta}_A^\kappa).$$

Here the indices j , k , and A refer to loci, lineages and subpopulations, respectively. In the numerical examples of this paper, we assume the values $\boldsymbol{\beta}_j^q = \mathbf{1}_{n_j}$, $\mu_a = 1$, $\sigma_a^2 = 2$. We set the number of lineages equal to the number of populations, and assume that lineage A makes the dominant contribution to population A , i.e. that the matrix $\boldsymbol{\kappa}$ is diagonally dominant. To do so, we let

$$\beta_{AA}^\kappa = 0.8, \text{ and } \beta_{Ak}^\kappa = \frac{0.2}{n_L - 1} \text{ for } k \neq A,$$

and truncate the prior by the requirement that $\beta_{AA}^\kappa > \beta_{Ak}^\kappa$ for all $k \neq A$. The latter specification ensures that label switching is not possible, which property improves the mixing of Markov Chain Monte Carlo (MCMC) algorithm (GELMAN *et al.* 2004).

The number of alleles n_j in neutral marker locus j in the ancestral generation is generally unknown, as some alleles may have disappeared after the lineages have diverged or are not present in the sampled individuals. Due to mathematical properties of the Dirichlet distribution, we may pool all such unobserved alleles into one ‘meta-allele’. Thus, we define n_j as the number of distinct alleles observed in locus j plus one.

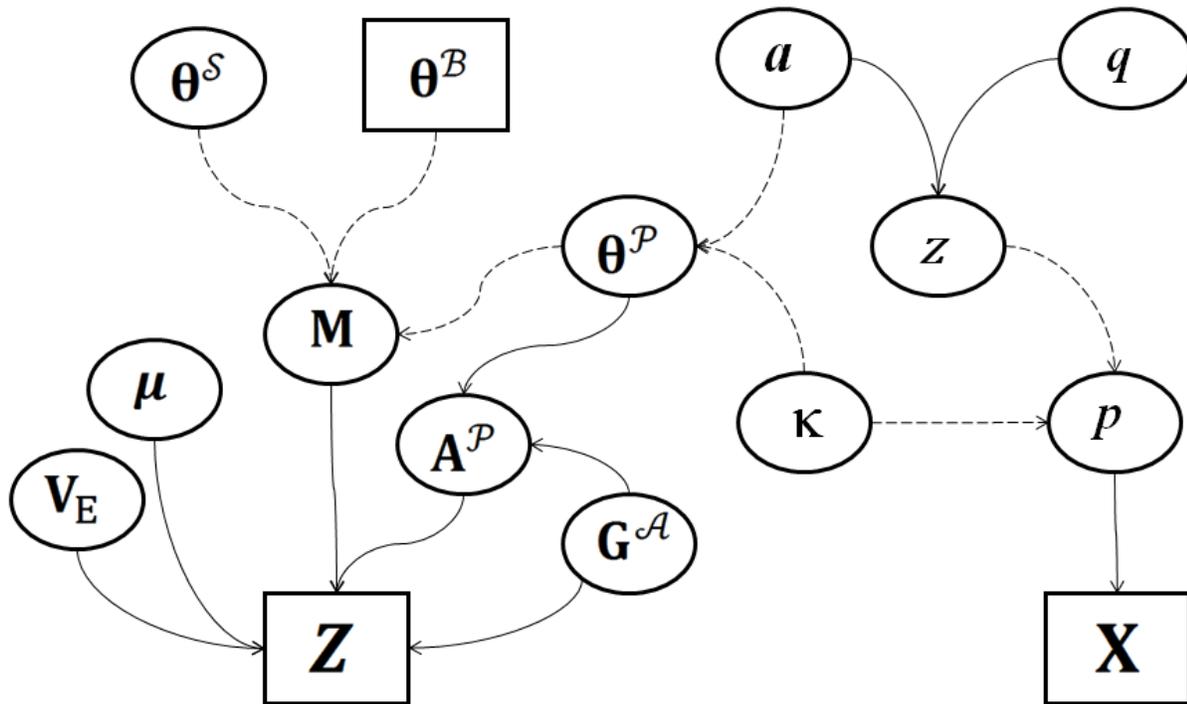
Parameter estimation through MCMC

Each model parameter was sampled from its full conditional while keeping the other parameters fixed. Most parameters were updated using a random-walk Metropolis-Hastings (MH) algorithm, the proposal distributions given below. For these the variances of the proposal distributions were adjusted during the burn-in following Ovaskainen et al. (2008) to give an accept ratio of 0.44 for a single parameter or 0.22 for a vector of parameters.

- Sampling the additive genetic variance-covariance matrix $\mathbf{G}^{\mathcal{A}}$ was conducted using the MH algorithm. We used the proposal distribution $\text{Wishart}(\mathbf{G}^{\mathcal{A}}/\nu, \nu)$, where the parameter ν was adjusted during the burn-in.
- Sampling the residual variance-covariance matrix \mathbf{V}_E was conducted using the MH algorithm. We used the proposal distribution $\text{Wishart}(\mathbf{V}_E/\nu, \nu)$, where the parameter ν was adjusted during the burn-in.
- Sampling the overall mean $\boldsymbol{\mu}$ and the population specific means $\mathbf{A}^{\mathcal{P}}$. Conditional on the other parameters, the full conditional joint distribution of the parameters $(\boldsymbol{\mu}, \mathbf{A}^{\mathcal{P}})$ follows a multivariate normal distribution. These parameters were thus updated directly (not using the MH algorithm) by drawing a random deviate from the full conditional.
- Sampling drift parameters \mathbf{a} . We used $N(a_k, \delta_{a_k}^2)$ distributions separately for each k to draw proposals for $\log a_k$. The variance parameters $\delta_{a_k}^2$ were adjusted during the burn-in.
- Sampling lineage loadings $\boldsymbol{\kappa}$. We used truncated Dirichlet($\delta_{\boldsymbol{\kappa}_A} \boldsymbol{\kappa}_A$) distributions separately for each A and j to draw proposals for $\boldsymbol{\kappa}_A$. The $\delta_{\boldsymbol{\kappa}_A}$'s are proposal parameters that were adjusted during the burn-in.
- Sampling ancestral allele frequencies \mathbf{q} . We used truncated Dirichlet($\delta_{\mathbf{q}_j} \mathbf{q}_j$) distributions separately for each j to draw proposals for \mathbf{q}_j . The $\delta_{\mathbf{q}_j}$'s are proposal parameters that were adjusted during the burn-in.
- Sampling allele frequencies \mathbf{z} . We used Dirichlet($\delta_{\mathbf{z}_{kj}} \mathbf{z}_{kj}$) distributions separately for each lineage k and locus j . The $\delta_{\mathbf{z}_{kj}}$'s parameters were adjusted during the burn-in.

We implemented the algorithm in Mathematica 7.1 (WOLFRAM RESEARCH 2008), and used 1,000 iterations for the burn-in and 2,000 for sampling the posterior distribution. We note that these sample sizes are rather small, and they were chosen due to logistic constraints (the estimation was conducted for 3,600 data sets in total). We tested for the sufficiency for mixing by examining the chains for convergence and for running a sample of the datasets for 10,000 + 20,000 iterations. Both tests (as well as the Results, which separated well the cases with and without selection, see main text) indicated that the length of the MCMC chain was sufficient for the present purpose. For the case of real data analysis, we naturally recommend a much longer MCMC chain.

Supplementary Figure 1. Directed acyclic graph (DAG) for the Bayesian model. The quantities in the ellipses are the parameter values being estimated. In the single boxes are the observations (the neutral data X and the quantitative trait data Z) and in the double boxes are the fixed quantities. The full arrows give stochastic relationships (i.e. those that have probability distributions associated with them) and the dashed arrows are for deterministic relationships.



References

- ALBERT, A. Y. K., S. SAWAYA, T. H. VINES, A. K. KNECHT, C. T. MILLER *et al.*, 2008 The genetics of adaptive shape shift in stickleback: Pleiotropy and effect size. *Evolution* **62**: 76-85.
- GELMAN, A., J. B. CARLIN, H. S. STERN and D. B. RUBIN, 2004 *Bayesian Data Analysis*. Chapman & Hall/CRC, Boca Raton.
- KARHUNEN, M., and O. OVASKAINEN, 2011 Defining and estimating FST through coancestry. Manuscript.
- WOLFRAM RESEARCH, I., 2008 *Mathematica Edition: Version 7.1*. Wolfram Research, Inc., Champaign, Illinois.