File S3: Estimating the β_i's and the Bayes Factors (BF) under the core model using an Importance Sampling algorithm

Here we detail the Importance Sampling algorithm used to estimate the β_i's and the Bayes Factors under the STD model (Figure 1B in the main text) using MCMC samples drawn from the posterior distribution of the core model. Proposed in Coop *et al.* (2010) for the Bayes Factors (Appendix B), this approach is computationally efficient and allows to consider simultaneously any number of covariates. However it suffers from some limitations (see the main text).

For the sake of simplicity, derivations are only presented here for allele count data. Notations are the same as in the main text but locus indices i are omitted.

1 Estimation of the BF's

1.1 Derivation

We hereby elaborate on the results described in the Appendix B by Coop *et al.* (2010) By definition:

$$\mathrm{BF} = \frac{\mathbb{P}\left(M_1 \mid \mathbf{y}, \mathbf{n}, \mathbf{Z}\right)}{\mathbb{P}\left(M_0 \mid \mathbf{y}, \mathbf{n}\right)}$$

Here:

$$\mathbb{P}\left(M_1 \mid \boldsymbol{y}, \mathbf{n}, \boldsymbol{Z}\right) \propto \int \mathbb{P}\left(\boldsymbol{y} \mid \mathbf{n}, \boldsymbol{\alpha}\right) f\left(\boldsymbol{\alpha}^\star \mid \boldsymbol{\Lambda}, \boldsymbol{\pi}, \beta, \boldsymbol{Z}\right) f(\boldsymbol{\Lambda}) f(\boldsymbol{\pi}) f(\beta) d\boldsymbol{\Lambda} d\boldsymbol{\pi} d\beta$$

and

$$\mathbb{P}\left(M_0 \mid \mathbf{y}, \mathbf{n}, \boldsymbol{Z}\right) \propto \int \mathbb{P}\left(\boldsymbol{y} \mid \boldsymbol{n}, \boldsymbol{\alpha}\right) f\left(\boldsymbol{\alpha}^\star \mid \boldsymbol{\Lambda}, \boldsymbol{\pi}\right) f(\boldsymbol{\Lambda}) f(\boldsymbol{\pi}) d\boldsymbol{\Lambda} d\boldsymbol{\pi}$$

Hence

$$\begin{aligned}
\mathrm{BF} &= \frac{\int \mathbb{P}\left(\boldsymbol{y} \mid \mathbf{n}, \boldsymbol{\alpha}\right) f\left(\boldsymbol{\alpha}^\star \mid \boldsymbol{\Lambda}, \boldsymbol{\pi}, \beta, \boldsymbol{Z}\right) f(\boldsymbol{\Lambda}) f(\boldsymbol{\pi}) f(\beta) d\boldsymbol{\Lambda} d\boldsymbol{\pi} d\beta}{\int \mathbb{P}\left(\boldsymbol{y} \mid \boldsymbol{n}, \boldsymbol{\alpha}\right) f\left(\boldsymbol{\alpha}^\star \mid \boldsymbol{\Lambda}, \boldsymbol{\pi}\right) f(\boldsymbol{\Lambda}) f(\boldsymbol{\pi}) d\boldsymbol{\Lambda} d\boldsymbol{\pi}} \\
&= \int f\left(\boldsymbol{\alpha}^\star \mid \boldsymbol{\Lambda}, \boldsymbol{\pi}, \beta, Z\right) \left(\frac{\mathbb{P}\left(\mathbf{y} \mid \mathbf{n}, \boldsymbol{\alpha}\right) f(\boldsymbol{\Lambda}) f(\boldsymbol{\pi}) f(\beta)}{\int \mathbb{P}\left(\boldsymbol{y} \mid \boldsymbol{n}, \boldsymbol{\alpha}\right) f\left(\boldsymbol{\alpha}^\star \mid \boldsymbol{\Lambda}, \boldsymbol{\pi}\right) f(\boldsymbol{\Lambda}) f(\boldsymbol{\pi}) d\boldsymbol{\Lambda} d\boldsymbol{\pi}} \right) f(\boldsymbol{\Lambda}) f(\boldsymbol{\pi}) f(\beta) d\boldsymbol{\Lambda} d\boldsymbol{\pi} d\beta \\
&= \int \frac{f\left(\boldsymbol{\alpha}^\star \mid \boldsymbol{\Lambda}, \boldsymbol{\pi}, \beta, \boldsymbol{Z}\right)}{f\left(\boldsymbol{\alpha}^\star \mid \boldsymbol{\Lambda}, \boldsymbol{\pi}\right)} \left(\frac{\mathbb{P}\left(\mathbf{y} \mid \mathbf{n}, \boldsymbol{\alpha}\right) f\left(\boldsymbol{\alpha}^\star \mid \boldsymbol{\Lambda}, \boldsymbol{\pi}\right) f(\boldsymbol{\Lambda}) f(\boldsymbol{\pi}) f(\beta)}{\int \mathbb{P}\left(\boldsymbol{y} \mid \boldsymbol{n}, \boldsymbol{\alpha}\right) f\left(\boldsymbol{\alpha}^\star \mid \boldsymbol{\Lambda}, \boldsymbol{\pi}\right) f(\boldsymbol{\Lambda}) f(\boldsymbol{\pi}) d\boldsymbol{\Lambda} d\boldsymbol{\pi}} \right) f(\boldsymbol{\Lambda}) f(\boldsymbol{\pi}) f(\beta) d\boldsymbol{\Lambda} d\boldsymbol{\pi} d\beta \\
&= \int \frac{f\left(\boldsymbol{\alpha}^\star \mid \boldsymbol{\Lambda}, \boldsymbol{\pi}, \beta, \boldsymbol{Z}\right)}{f\left(\boldsymbol{\alpha}^\star \mid \boldsymbol{\Lambda}, \boldsymbol{\pi}\right)} f\left(\boldsymbol{\alpha}^\star, \boldsymbol{\Lambda}, \boldsymbol{\pi} \mid \boldsymbol{y}, \boldsymbol{n}, M_0\right) f(\boldsymbol{\Lambda}) f(\boldsymbol{\pi}) f(\beta) d\boldsymbol{\Lambda} d\boldsymbol{\pi} d\beta \\
&= \int_\beta \left(\int \omega\left(\boldsymbol{\alpha}^\star, \boldsymbol{\Lambda}, \boldsymbol{\pi}, \beta, \boldsymbol{Z}\right) f\left(\boldsymbol{\alpha}^\star, \boldsymbol{\Lambda}, \boldsymbol{\pi} \mid \boldsymbol{y}, \boldsymbol{n}, M_0\right) f(\boldsymbol{\Lambda}) f(\boldsymbol{\pi}) d\boldsymbol{\Lambda} d\boldsymbol{\pi} \right) f(\beta) d\beta \\
&= \mathbb{E}_\beta\left[\mathbb{E}_{M_0}\left[\omega\left(\boldsymbol{\alpha}^\star, \boldsymbol{\Lambda}, \boldsymbol{\pi}, \beta, \boldsymbol{Z}\right)\right]\right]
\end{aligned}$$

Note that denoting (as in File S1), $\ddot{\boldsymbol{\alpha}}_i = \Gamma^{-1}\left\{\frac{\alpha^{\star}_{ij}-\pi_i}{\sqrt{\pi_i(1-\pi_i)}}\right\}_{(1..J)}$ and $\widetilde{\boldsymbol{\Phi}} = \left\{\phi_j\right\}_{(1..J)} = \Gamma^{-1}\left\{\frac{\beta_i Z_j}{\sqrt{\pi_i(1-\pi_i)}}\right\}$ (where Γ results from the Choleski decomposition of $\boldsymbol{\Omega} = \boldsymbol{\Lambda}^{-1}$, i.e., $\boldsymbol{\Omega} = {}^{t}\boldsymbol{\Gamma}\boldsymbol{\Gamma}$):

$$\omega\left(\boldsymbol{\alpha}^{\star}, \boldsymbol{\Lambda}, \boldsymbol{\pi}, \beta, \boldsymbol{Z}\right) = \frac{f\left(\boldsymbol{\alpha}^{\star} \mid \boldsymbol{\Lambda}, \boldsymbol{\pi}, \beta, \boldsymbol{Z}\right)}{f\left(\boldsymbol{\alpha}^{\star} \mid \boldsymbol{\Lambda}, \boldsymbol{\pi}\right)} = \frac{f\left(\ddot{\boldsymbol{\alpha}}^{\star} \mid \boldsymbol{\Lambda}, \boldsymbol{\pi}, \beta, \boldsymbol{Z}\right)}{f\left(\ddot{\boldsymbol{\alpha}}^{\star} \mid \boldsymbol{\Lambda}, \boldsymbol{\pi}\right)} = e^{\left(\sum_{j=1}^{j=J}\widetilde{\phi_{ij}}\ddot{\alpha}_{ij}\right)-\frac{1}{2}\left(\sum_{j=1}^{j=J}\widetilde{\phi_{ij}}^{2}\right)}$$

Therefore, the Bayes Factor can simply obtained from posterior samples of the parameters $\boldsymbol{\alpha}^{\star}$,$\boldsymbol{\Lambda}$ and $\boldsymbol{\pi}$ obtained under the null model (M_0) that corresponds to core model (Figure 1A).

1.2 Computation (as in BayPass)

The Importance Sampling approximation of the BF of a given locus is simply the expectation of $\omega\left(\boldsymbol{\alpha}^{\star}, \boldsymbol{\Omega}, \boldsymbol{\pi}, \beta, \boldsymbol{Z}\right)$ integrated over the core model. Hence, BF might simply be obtained by averaging $\omega\left(\boldsymbol{\alpha}^{\star}, \boldsymbol{\Lambda}, \boldsymbol{\pi}, \beta, \boldsymbol{Z}\right)$ over the MCMC (Coop *et al.*, 2010) and integrating on the whole support of the β parameter, i.e. ($\beta_{\min}$;$\beta_{\max}$). To that end numerical integration is performed over a grid of nint$_\beta$ uniformly distributed values of β.

Let:

- $\beta_p = \frac{\beta_{\max}-\beta_{\min}}{\text{nint}_\beta}$ the grid step
- $\beta_g^{\text{inf}} = \beta_{\min} + (g-1)\beta_p$ and $\beta_g^{\text{inf}} = \beta_{\min} + g\beta_p$ the boundaries of the g$^{\text{th}}$ grid interval
- $\mathbb{P}\left(\beta_g\right) = \int_{\beta_g^{\text{inf}}}^{\beta_g^{\text{sup}}} f(\beta)\,d\beta$ the prior over interval g. If the prior is uniform, then $\mathbb{P}\left(\beta_g\right) = \frac{1}{\text{nint}_\beta}$
- $\omega_g = \int_{\beta_g^{\text{inf}}}^{\beta_g^{\text{sup}}} \omega\left(\boldsymbol{\alpha}_t^{\star}, \boldsymbol{\Lambda}_t, \boldsymbol{\pi}_t, \beta, \boldsymbol{Z}\right) f(\beta) d\beta$ for all g interval.
- $\widehat{\omega_g} = \frac{1}{2}\left(\omega\left(\boldsymbol{\alpha}^{\star}, \boldsymbol{\Lambda}, \boldsymbol{\pi}, \beta_g^{\text{inf}}, \boldsymbol{Z}\right) + \omega\left(\boldsymbol{\alpha}^{\star}, \boldsymbol{\Lambda}, \boldsymbol{\pi}, \beta_g^{\text{sup}}, \boldsymbol{Z}\right)\right)\mathbb{P}\left(\beta_g\right)$ approximates ω_g

If the support β is bounded, then:

$$\begin{aligned}
\text{BF} &= \int_\beta\left(\int \omega\left(\boldsymbol{\alpha}^{\star}, \boldsymbol{\Lambda}, \boldsymbol{\pi}, \beta, \boldsymbol{Z}\right) f\left(\boldsymbol{\alpha}^{\star}, \boldsymbol{\Lambda}, \boldsymbol{\pi} \mid \boldsymbol{y}, \boldsymbol{n}, M_0\right) f(\boldsymbol{\Lambda}) f(\boldsymbol{\pi}) d\boldsymbol{\Lambda} d\boldsymbol{\pi}\right) f(\beta) d\beta \\
&= \int\left(\int_\beta \omega\left(\boldsymbol{\alpha}^{\star}, \boldsymbol{\Lambda}, \boldsymbol{\pi}, \beta, \boldsymbol{Z}\right) f(\beta) d\beta\right) f\left(\boldsymbol{\alpha}^{\star}, \boldsymbol{\Lambda}, \boldsymbol{\pi} \mid \boldsymbol{y}, \boldsymbol{n}, M_0\right) f(\boldsymbol{\Lambda}) f(\boldsymbol{\pi}) d\boldsymbol{\Lambda} d\boldsymbol{\pi} \\
&= \int\left(\sum_{g=1}^{g=\text{nint}_\beta} \omega_g\right) f\left(\boldsymbol{\alpha}^{\star}, \boldsymbol{\Lambda}, \boldsymbol{\pi} \mid \boldsymbol{y}, \boldsymbol{n}, M_0\right) f(\boldsymbol{\Lambda}) f(\boldsymbol{\pi}) d\boldsymbol{\Lambda} d\boldsymbol{\pi} \\
&= \sum_{g=1}^{g=\text{nint}_\beta}\left(\int \omega_g f\left(\boldsymbol{\alpha}^{\star}, \boldsymbol{\Lambda}, \boldsymbol{\pi} \mid \boldsymbol{y}, \boldsymbol{n}, M_0\right) f(\boldsymbol{\Lambda}) f(\boldsymbol{\pi}) d\boldsymbol{\Lambda} d\boldsymbol{\pi}\right)
\end{aligned}$$

Hence, with $\widehat{\omega_g^{(t)}} = \frac{1}{2}\left(\omega\left(\boldsymbol{\alpha}_t^{\star}, \boldsymbol{\Lambda}_t, \boldsymbol{\pi}_t, \beta_g^{\text{inf}}, \boldsymbol{Z}\right) + \omega\left(\boldsymbol{\alpha}_t^{\star}, \boldsymbol{\Lambda}_t, \boldsymbol{\pi}_t, \beta_g^{\text{sup}}, \boldsymbol{Z}\right)\right)\mathbb{P}\left(\beta_g\right)$ at iteration t of the MCMC (under the core model), we obtain:

$$\boxed{\widehat{BF} = \sum_{g=1}^{g=\text{nint}_\beta} \frac{1}{\text{niter}}\left(\sum_{t=1}^{t=\text{niter}} \widehat{\omega_g^{(t)}}\right) = \frac{1}{\text{niter}}\sum_{t=1}^{t=\text{niter}}\left(\sum_{g=1}^{g=\text{nint}_\beta} \widehat{\omega_g^{(t)}}\right)}$$

2 Estimation of the β_i's

As for the *BF*, the moments of the posterior distribution of each β_i can be estimate via Importance Sampling as exemplified for the posterior mean below: $\widehat{\beta} = \int \beta f(\beta \mid \text{data}) d\beta$. where:

$$\begin{aligned} f(\beta \mid \text{data}) &\propto f(\boldsymbol{\alpha}^\star \mid \beta, \boldsymbol{\Lambda}, \boldsymbol{\pi}) f(\boldsymbol{\Lambda}) f(\boldsymbol{\pi}) f(\beta) \\ &\propto \omega f(\beta) f(\boldsymbol{\alpha}^\star \mid \boldsymbol{\Lambda}, \boldsymbol{\pi}) f(\boldsymbol{\Lambda}) f(\boldsymbol{\pi}) \end{aligned}$$

Using the same notations as above, and further defining:

- $\beta_g = \frac{1}{2}\left(\beta_g^{\text{inf}} + \beta_g^{\text{sup}}\right)$
- $P_g = \int_{\beta_g^{\text{inf}}}^{\beta_g^{\text{sup}}} f(\beta \mid \text{data}) d\beta$ approximated by $\widetilde{P_g} = \frac{1}{2}\left(f(\beta_g^{\text{inf}} \mid \text{data}) + f(\beta_g^{\text{sup}} \mid \text{data})\right)$
- $b_g = \int_{\beta_g^{\text{inf}}}^{\beta_g^{\text{sup}}} \beta f(\beta \mid \text{data}) d\beta$ approximated by $\widetilde{b_g} = \beta_g \widetilde{P_g}$

$\widetilde{P_g}$ can be estimated from MCMC samples as $\widehat{P_g} = \frac{1}{\text{niter}} \sum_{t=1}^{\text{niter}} \frac{p_g(t)}{\sum_{g=1}^{\text{g=nint}} p_g(t)}$ where $p_g(t) = \omega\left(\boldsymbol{\alpha}_t, \boldsymbol{\Omega}_t, \boldsymbol{\pi}_t, \beta_g^{\text{inf}}, \boldsymbol{Z}\right) f(\beta_g^{\text{inf}}) + \omega\left(\boldsymbol{\alpha}_t, \boldsymbol{\Omega}_t, \boldsymbol{\pi}_t, \beta_g^{\text{sup}}, \boldsymbol{Z}\right) f(\beta_g^{\text{sup}})$

Hence,

$$\widehat{\beta} = \sum_{g=1}^{g=\text{nint}_\beta} \beta_g \frac{1}{\text{niter}} \sum_{t=1}^{\text{niter}} \frac{p_g(t)}{\sum_{g=1}^{\text{g=nint}} p_g(t)} = \frac{1}{\text{niter}} \sum_{t=1}^{\text{niter}} \left(\frac{1}{\sum_{g=1}^{\text{g=nint}} p_g(t)} \sum_{g=1}^{\text{g=nint}} \beta_g p_g(t) \right)$$

References

Coop, G., D. Witonsky, A. D. Rienzo, and J. K. Pritchard, 2010 Using environmental correlations to identify loci underlying local adaptation. Genetics 185: 1411–1423.