Supplementary Information

1 Definition of the prior probabilities of jump between models

Recall that we have 3 models from which we would like to infer posterior probabilities:

M1 Neutral model: β_j

M2 Locus-specific model: $\alpha_i + \beta_j$

M3 Local adaptation model with environmental differentiation E_j: $\beta_j + g_i E_j$

Let Π_2 be the prior probability of model 2 and Π_3 the prior probability of model 3. We assume that the probability of going from 1 or 2 (resp. 3) to the model 2 (resp. 3) is equal to Π_2 (resp. Π_3), which leads to the following transition matrix :

$$\begin{pmatrix} \Pi_2 & (1-\Pi_2)(1-\Pi_3) & \Pi_3(1-\Pi_2) \\ \Pi_2 & 1-\Pi_2-\Pi_3 & \Pi_3 \\ \Pi_2(1-\Pi_3) & (1-\Pi_2)(1-\Pi_3) & \Pi_3 \end{pmatrix} \tag{1}$$

If we consider $\Pi_2 = \pi p$ and $\Pi_3 = \pi(1-p)$ where π is the probability of jumping away from model 1 and p the "preference" for model 2 (i.e. the probability of choosing the model 2 instead of model 3, when jumping away from model 1), then we can write:

$$\begin{pmatrix} \pi p & 1-\pi+\pi^2 p(1-p) & \pi(1-p)-\pi^2 p(1-p) \\ \pi p & 1-\pi & \pi(1-p) \\ \pi p - \pi^2 p(1-p) & 1-\pi+\pi^2 p(1-p) & \pi(1-p) \end{pmatrix} \tag{2}$$

Thus, when $\pi \to 0$ (in practice $\pi < 0.5$), the transition between models depends only very slightly on the current state of the model, and the prior probabilities of each models reduce approximately to:

$$\begin{array}{rcl} P(\mathbf{M1}) & = & 1-\pi \\ P(\mathbf{M2}) & = & \pi p \\ P(\mathbf{M3}) & = & \pi(1-p) \end{array} \tag{3}$$

2 Reversible jumps between the models

According to Brooks (1998); Gelman *et al.* (2004), the jump between model l and k should be accepted with a probability $\min(r, 1)$ where:

$$r = \frac{L(Y|\theta_k, M_k)p(\theta_k|M_k)p(M_k)}{L(Y|\theta_l, M_l)p(\theta_l|M_l)p(M_l)} \frac{J_{k\to l}J(u_k|\theta_k, k, l)}{J_{l\to k}J(u_l|\theta_l, l, k)} \left|\frac{\nabla g_{l,k}(\theta_l, u)}{\nabla(\theta_l, u)}\right| \tag{4}$$

where $\theta_\bullet$ is the parameter vector for model $\bullet$, L stands for the likelihood of the data Y given the parameters $\theta_\bullet$ and the model $M_\bullet$ and J is the kernel proposal for the (potentially) new parameter $u_\bullet$. Note also the presence of the proposal of new model $J_{\circ\to\bullet}$.

Because a jump toward one of the two alternative models is proposed at each iterations, $\frac{J_{k\to l}}{J_{l\to k}}$ simplifies to one. Likewise, since the transformations from θ_l to θ_k only consists into setting some parameters to 0, the Jacobian determinant $\left|\frac{\nabla g_{l,k}(\theta_l,u)}{\nabla(\theta_l,u)}\right|$ also simplifies to one.

The most efficient way to propose a value for $u_\bullet$ is to draw from its own posterior. In order to do so, pilot runs are performed before the actual reversible jump MCMC, in order to approximate the posterior of the parameters α_i and g_i for each locus i. The posterior mean and variance for both parameters are approximated and used to parametrise a Normal distribution to propose new values when a jump is proposed. Note that, since g_i cannot be negative, a truncated Normal is used, and the J kernel is modified accordingly in Eq. 4.

3 Hierarchically Structured (HS) scenario

Below (Fig. I) is the schematic representation of the demographic scenario called HS in the main text, showing the fission events and the migration between populations. Note that only some illustrative migration combinations are represented for the sake of simplicity.

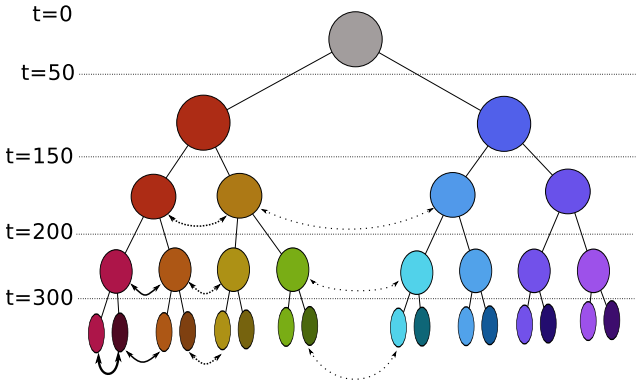


Figure I: Schematic representation of the Hierarchically Structured scenario (HS). Fission events are shown as connectors and migration is denoted using double-arrow (thickness illustrate the strength of migration).

4 Simulation scripts

Four Python scripts are provided as Supplementary Files :

`IM_mono.py` Island model with monogenic selection

`IM_poly.py` Island model with polygenic selection

`SS_poly.py` Stepping Stone model with polygenic selection

`HS_poly.py` Hierarchically Structured model with polygenic selection

These scripts require Python 2.7 and SimuPOP 1.1. Note that the monogenic version is provided only for the Island model, as the modification are identical for the two other models.

5 Environmental gradient

The environmental gradient was designed to be independent from (i.e. not be confounded with) the population structure for the scenarios SS and HS (see Fig. II).

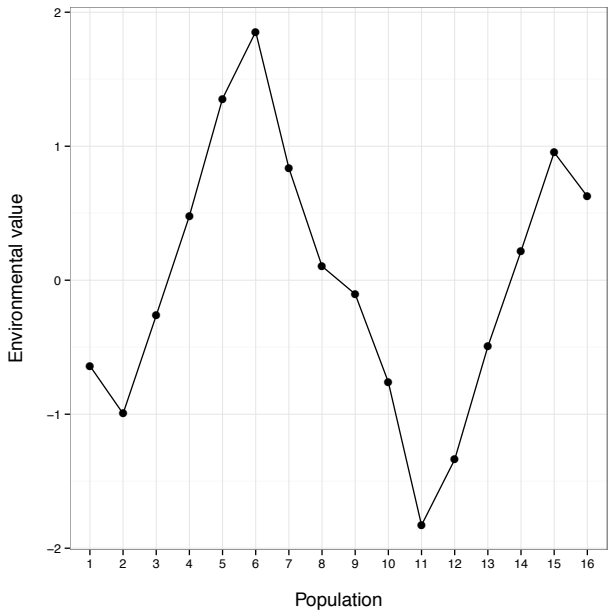


Figure II: Environmental value for all 16 populations in the scenarios SS and HS.

6 Simulation results for the False Positive Rate

The results regarding the False Positive Rate (FPR, Fig. III) were qualitatively comparable to the results regarding the False Discovery Rate (FDR, Fig. 2, main text). Indeed, we again found that the most error-prone method was Bayescan, where BayeScEnv yielded less false positives. For the latter, the parametrisation $\pi = 0.1, p = 0.5$ was, as expected, the most conservative, whereas $\pi = 0.1, p = 0$ was the most laxist.

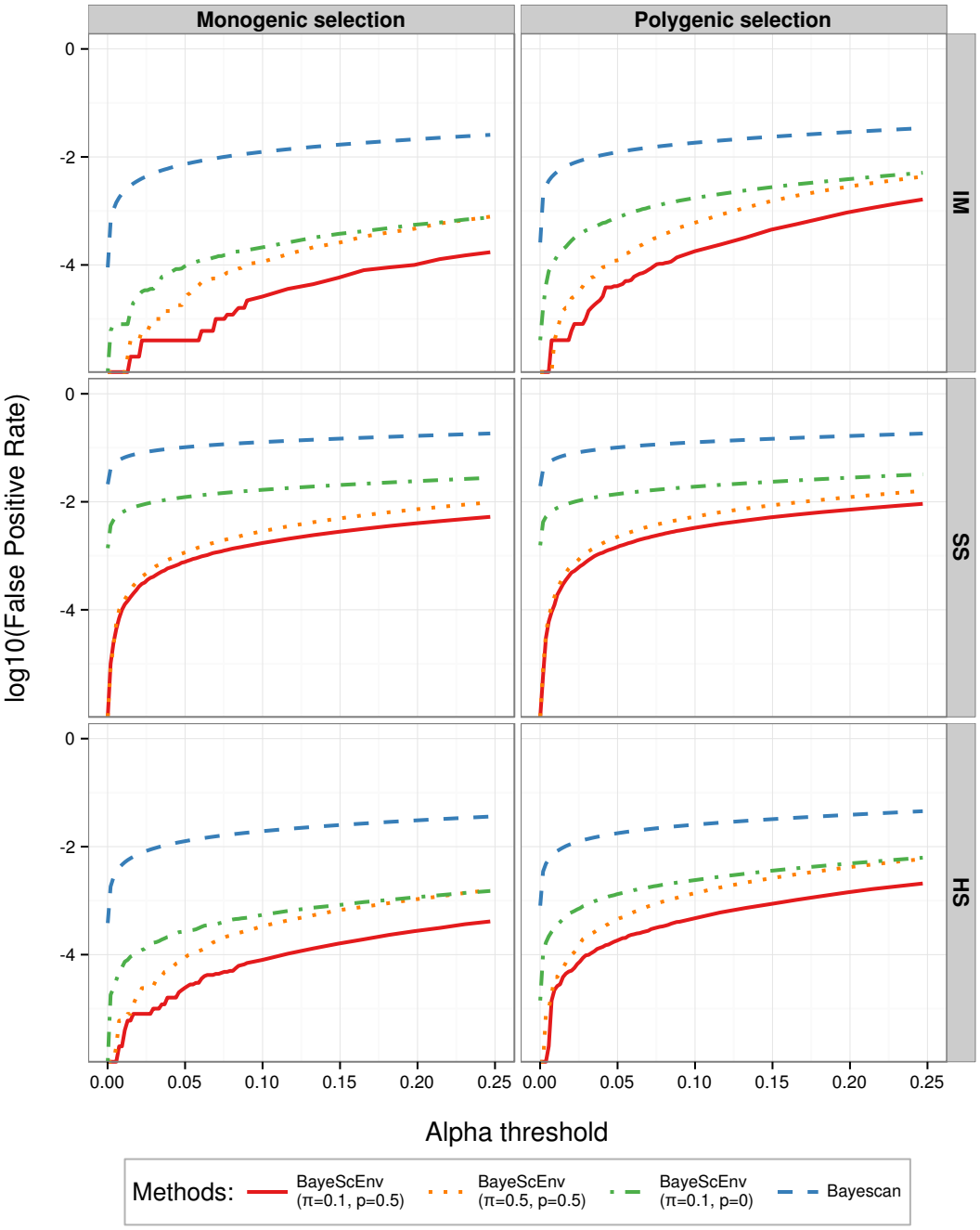


Figure III: False Positive Rate (FPR) against significance α threshold for three scenarios (IM: Island model, SS: Stepping-Stone model and HS: Hierarchically Structured model) and monogenic/polygenic selection. The models tested are Bayescan (blue dashed), and BayeScEnv (orange dotted, green dot-dashed and solid red) with different probabilities π of jumping away from the neutral model and different preferences p for the locus-specific model. Note that $p = 0$ means the environmental model is tested against the neutral one only.

7 List of candidate genes associated with significant GO terms

Below are the list of the genes that fulfil two criteria. First, there is at least a significant SNPs in their neighbourhood, indicating them as potential candidates for local adaptation. Second, at least one of their associated GO terms were found as significantly enriched in candidates compared to the rest of the genome.

- **For the altitude analysis:** SCARB1, SLC12A1, MUCL1, DNM2, MLANA, ATP6V1C2, CLDN12, FBN1, OTUD7A, SLC24A5, NOS1AP, SLC12A8
- **For the temperature analysis:** SYNE2, SPTB, ANKRD46, HAO1, HCK, FOXP1, ONECUT2, CDH15, ATP8A2, FADS2, ESR2, ATP6V1C2, FADS1, NRG1, APBB2, CMYA5, SERPINA6, SLC8A1, PRKG1, LAMA2, SERPINA1

Note that the majority of the significant GO terms were represented by a unique candidate gene for the altitude and temperature analysis. The list of the genes fulfilling the two criteria is not shown for the precipitation analysis and Bayescan, as there are too many of them for such a list to be relevant. The output the Gowinda analyses are provided as data files.

8 Analysis of the simulation scenarios from de Villemereuil *et al.* (2014)

Scenarios In order to compare BayeScEnv with methods other than BayeScan, we used the simulated datasets from de Villemereuil *et al.* (2014) study. For more information regarding these scenarios, please refer to the article. Briefly, four polygenic scenarios were tested:

HsIMM-C Hierarchical scenario with a clinal environment following population structure

HsIMM-U Hierarchical scenario with a random environment strongly correlated with population structure

IMM Isolation with Migration Model

SS Stepping-Stone model with a clinal environment, following population structure

Foreword These scenarios were tested against BayeScan (Foll and Gaggiotti, 2008), BayEnv (Coop *et al.*, 2010) and LFMM (Frichot *et al.*, 2013). Note that these scenarios are very difficult for all methods, thus we do not expect the FDR to be well calibrated. Also, in contrast with the present study, de Villemereuil *et al.* used a prior odds of 100 instead of 10 for BayeScan.
When interpreting the results, it should be remembered that False Discovery Rate (FDR) depends on both the False Positive Rate (FPR) and power. All things being equal, the FDR will be higher if the FPR is higher, and lower if the power is higher.

Results The results are presented in Fig. IV–VI (below). They show that, even in very difficult conditions, BayeScEnv performs honourably. As expected, both the FPR and the power of BayeScEnv are lower than that of BayeScan, resulting in a more conservative method overall. However, in scenarios with low power for all methods, BayeScEnv's lack of power can drastically inflate its FDR (e.g. Fig. IV, red line, IMM model). Interestingly, BayeScEnv with $p = 0$ is more robust in this regard, since its power is generally much higher.
When compared to the other association methods (BayEnv and LFMM), BayeScEnv performed very well in "clinal" scenarios (HsIMM-C and SS), but more poorly in the other scenarios. However when considering the canonical $\alpha = 0.05$ threshold, BayeScEnv's FDR was always lower than at least one of the association methods, except in the IMM scenario.

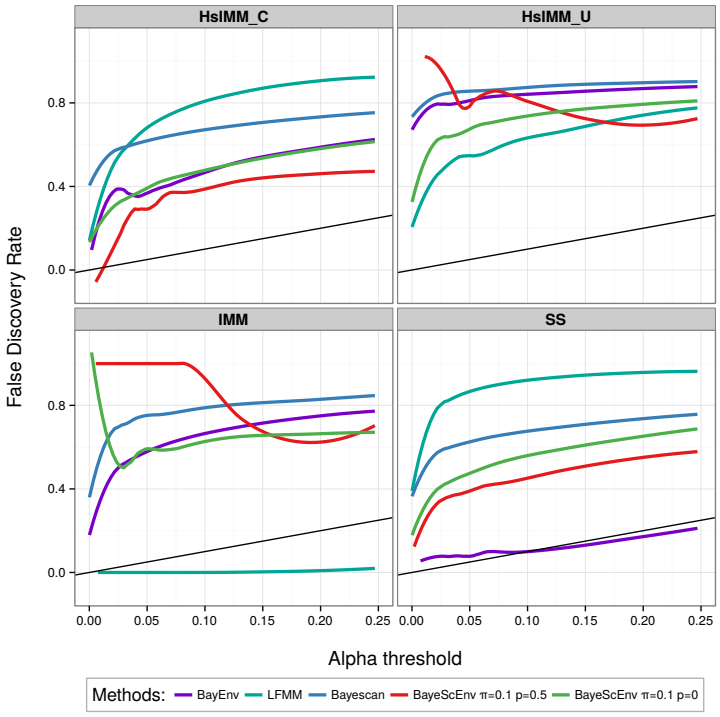


Figure IV: False Discovery Rate (FDR) against significance α threshold for de Villemereuil *et al.* (2014) polygenic scenarios.

False Discovery Rate When comparing the FDR to other methods, BayeScEnv performs relatively well. Especially, for $p = 0.5$ (red line), its FDR can be the lowest (HsIMM-C) or second best (SS), but it can reach very high values under some scenarios (HsIMM-U and IMM). Surprisingly, the parametrisation $p = 0$ (green) is more stable across scenarios, whereas BayEnv (purple) and LFMM (turquoise) constantly "switch" between best-or-so and poorest-or-so. Overall, BayeScEnv's FDRs with $p = 0$ are lower than those of BayeScan's, at least for the canonical threshold $\alpha = 0.05$.
Large values of FDR for small α's are due to a lack of power (see below), not to a high False Positive Rate.

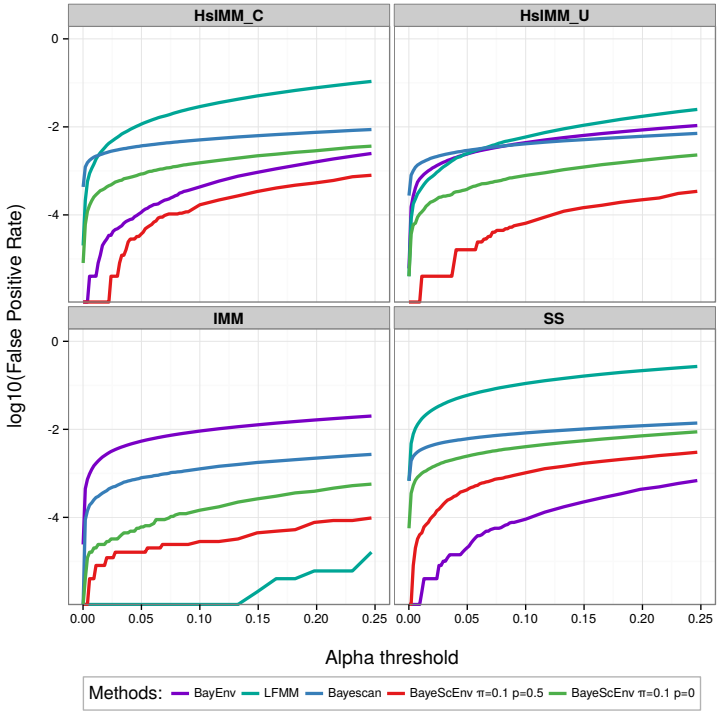


Figure V: False Positive Rate (FPR) against significance α threshold for de Villemereuil *et al.* (2014) polygenic scenarios.

False Positive Rate FPRs are more predictable than FDRs regarding the F-model family: BayeScan (blue) is the most error-prone method, followed by BayeScEnv with $p = 0$ (green) while BayeScEnv with $p = 0.5$ (red) is one of the most conservative methods. Interestingly, the FPRs of the F-model family are more stable than the FPRs of BayEnv (purple) and LFMM (turquoise), which vary greatly across scenarios.

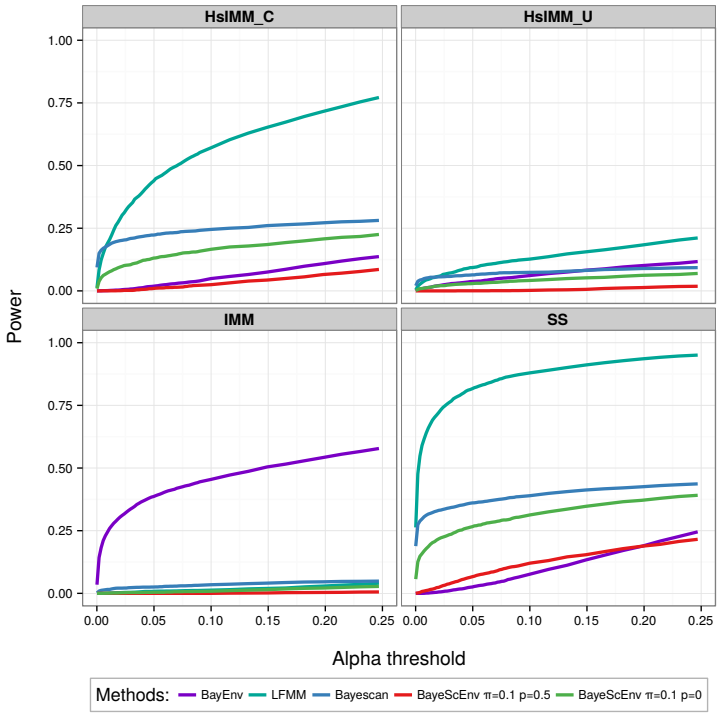
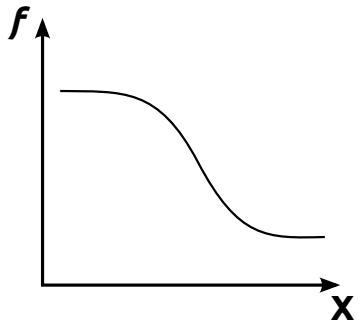


Figure VI: Power against significance α threshold for de Villemereuil *et al.* (2014) polygenic scenarios.

Power Overall power varies greatly across scenarios, HsIMM-U and IMM being the most difficult ones. As expected, the power of BayeScEnv (red and green) is always lower than the power of BayeScan (blue). However, the power of BayeScEnv with $p = 0$ (green), is always comparable to that of BayeScan's. BayeScEnv with $p = 0.5$ (red) is always among the less powerful method. For all scenarios, at least on of the environmental association methods (BayEnv (purple) and LFMM (turquoise)), has greater power than BayeScan and BayeScEnv.

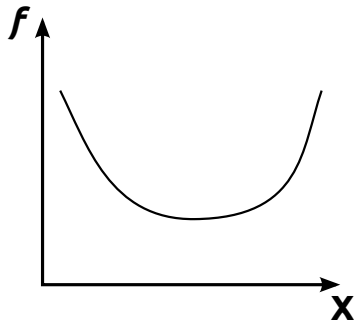
9 Typical frequency patterns detected by BayeScEnv

Consider the environmental variable X_j from which we derived the environmental differentiation using $E_j = |X_j|$. Three main patterns of the population allelic frequencies f_j as a function of this environmental value X_j can be distinguished:



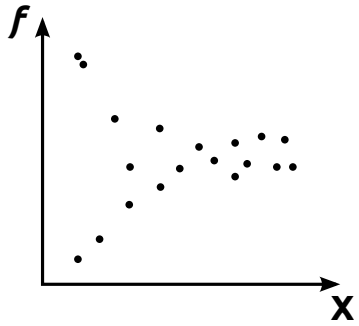
Clinal scenario

This is the canonical scenario, which was simulated in this study and most thoroughly investigated. It is also the scenario tested in all evaluations of environmental association methods.



"Plasticity" scenario

In this scenario, two populations with very different environmental values have similar allele frequencies. It could arise in situations where the environmental variable used is a proxy for a true environmental variable that has a non-monotonic relationship with the environmental variable used (e.g. very high or very low temperatures can both lead to aridity). More interestingly, it is expected, for example, from genes responsible for phenotypic plasticity (Morris *et al.*, 2014).



Difficult scenario

In this scenario populations with similar environmental values have extremely different allele frequencies. This scenario would lead to results that should in principle be interpreted as false positives. Note, however, that such scenario could be explained by positive frequency-dependent selection triggered by the environmental variable.

References

Brooks, S. 1998. Markov chain monte carlo method and its application. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(1): 69–100.

Coop, G., Witonsky, D., Di Rienzo, A., and Pritchard, J. K. 2010. Using environmental correlations to identify loci underlying local adaptation. *Genetics*, 185(4): 1411 –1423.

de Villemereuil, P., Frichot, E., Bazin, E., Franois, O., and Gaggiotti, O. E. 2014. Genome scan methods against more complex models: when and how much should we trust them? *Molecular Ecology*, 23(8): 2006–2019.

Foll, M. and Gaggiotti, O. E. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A bayesian perspective. *Genetics*, 180(2): 977 –993.

Frichot, E., Schoville, S. D., Bouchard, G., and Franois, O. 2013. Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, 30(7): 1687–1699.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. 2004. *Bayesian data analysis*. Text in Statisctical Science. Chapman & Hall/CRC Press, Boca Raton, Florida (US), second edition.

Morris, M. R. J., Richard, R., Leder, E. H., Barrett, R. D. H., Aubin-Horth, N., and Rogers, S. M. 2014. Gene expression plasticity evolves in response to colonization of freshwater lakes in threespine stickleback. *Molecular Ecology*, 23(13): 3226–3240.