

MOLECULAR ECOLOGY RESOURCES

A new FST-based method to uncover local adaptation using environmental variables.

Journal:	<i>Molecular Ecology Resources</i>
Manuscript ID:	MER-14-0489
Manuscript Type:	Resource Article
Date Submitted by the Author:	22-Nov-2014
Complete List of Authors:	de Villemereuil, Pierre; Université Joseph Fourier, Laboratoire d'Écologie Alpine Gaggiotti, Oscar; Université Joseph Fourier, LECA, UMR CNRS 5553
Keywords:	Molecular Evolution, Landscape Genetics, Adaptation, Ecological Genetics

A new F_{ST} -based method to uncover local adaptation using environmental variables.

Pierre de Villemereuil*, & Oscar E. Gaggiotti*[†]

*: Université Joseph Fourier, Centre National de la Recherche Scientifique,
LECA, UMR 5553, 2233 rue de la piscine, 38400 Saint Martin d'Hères, France

[†]Scottish Oceans Institute, University of St Andrews,
Fife, KY16 8LB, United Kingdom

Keywords: genome scan, local adaptation, environment, F model, Bayesian methods, false discovery rate

Corresponding author: Pierre de Villemereuil, E-mail: bonamy@horus.ens.fr

Abstract

Genome-scan methods are used for screening genome-wide patterns of DNA polymorphism to detect signatures of positive selection. There are two main types of methods: (i) “outlier” detection methods based on F_{ST} that detect loci with high differentiation compared to the rest of the genomes and, (ii) environmental association methods that test the association between allele frequencies and environmental variables. In this article, we present a new F_{ST} -based genome scan method, BayeScEnv, which incorporates environmental information in the form of “environmental differentiation”. It is based on the F model but as opposed to existing approaches it considers two locus-specific effects, one due to divergent selection and another due to other processes such as differences in mutation rates across loci or background selection. Simulation studies showed that our method has a much lower false positive rate than an existing F_{ST} -based method, BayeScan, under a wide range of demographic scenarios. Although it had lower power, it leads to a better compromise between power and false positive rate. We apply our method to Human and Salmon datasets and show that it can be used successfully to study local adaptation. The method was developed in C++ and is available at <http://github.com/devillemereuil/bayescenv>.

Introduction

One of the most important aims of the new field of population genomics (Luikart *et al.*, 2003) is to uncover signatures of selection in non-model species genomes (Radwan and Babik, 2012). Of special interest is the process of local adaptation whereby populations experiencing different environmental conditions undergo an

19 adaptive, selective pressure specific to their local habitat. As a result of this spatially heterogeneous selective
20 pressure, each local population evolve traits that provide an advantage under its local environmental conditions
21 (Kawecki and Ebert, 2004). Many experimental approaches focused on potentially adaptive traits have been
22 developed to test for local adaptation (Blanquart *et al.*, 2013) but only recently it has become possible to make
23 inferences about the genetic architecture of these traits. Indeed, the advent of next generation sequencing (NGS,
24 see Shendure and Ji, 2008) has fostered the development of so-called genome-scan methods aimed at identifying
25 regions of the genome subject to selection, which are being frequently used in studies of local adaptation (Faria
26 *et al.*, 2014).

27 There are two main types of genome-scan methods. The first type detects ‘outlier’ loci using locus-specific
28 F_{ST} estimates, which are compared to either an empirical distribution (Akey *et al.*, 2002), or to a distribution
29 expected under a neutral model of evolution (Beaumont and Balding, 2004; Foll and Gaggiotti, 2008). The
30 rationale used by these methods is that local adaptation leads to strong genetic differentiation between popula-
31 tions, but only at the selected loci (or marker loci linked to them). Thus, loci with very high F_{ST} compared to
32 the rest of the genome are suspected to be under strong local adaptation and are referred to as ‘outliers’ loci.
33 The ‘outlier’ approach was further extended to statistics akin to F_{ST} as in the F_{LK} (Bonhomme *et al.*, 2010)
34 and the X^tX (Günther and Coop, 2013) methods, and also to other unrelated statistics (Duforet-Frebourg
35 *et al.*, 2014). One limitation of these methods is that they are not ideally suited to test hypothesis about the
36 environmental factors underlying the selective pressure. Although they can be combined with other existing
37 methods to achieve this goal (e.g. Gaggiotti *et al.*, 2009), inference has to be carried out in two steps, first
38 identifying outliers and then testing for their association with environmental factors.

39 A second type of methods focuses on environmental variables and aim at associating allele frequency pat-
40 terns to environmental gradients. The underlying rationale is that selective pressures underlying local adaptation
41 should create associations between allele frequencies at the selected loci and the environmental variables under-
42 lying the selective pressure (Gaggiotti *et al.*, 2009; Coop *et al.*, 2010). In the presence of population structure,
43 trying to perform simple linear regression would be a misleading, error-prone approach (De Mita *et al.*, 2013;
44 de Villemereuil *et al.*, 2014). Thus, existing methods account for population structure by modelling the co-
45 variance in allele frequencies across populations (Coop *et al.*, 2010), by considering random effects using latent
46 factors (Frichot *et al.*, 2013) or using Gaussian Random Fields (Guillot *et al.*, 2014). One disadvantage of this
47 type of models is that the parameters that capture the effect of demographic history on genetic differentiation
48 do not have a clear biological interpretation, which in turn makes the rejection of the null model hard to in-
49 terpret in terms of detection of local adaptation. We note that although the elements of the covariance matrix
50 estimated by Coop *et al.* (2010) could in principle be interpreted as parametric estimates of the pairwise and
51 population-specific F_{ST} , this is only true when levels of genetic drift are low (Nicholson *et al.*, 2002; Balding
52 and Nichols, 1995).

It is important to note that, regardless of the type of genome-scan method under consideration, processes other than local adaptation might be responsible for the observed spatial patterns in allele frequency or F_{ST} . These include large differences in mutation rate across loci (Edelaar *et al.*, 2011), hybrid incompatibility following secondary contact (Kruuk *et al.*, 1999) and background selection (Charlesworth, 1998). It is therefore possible that some of the loci identified as outliers are in fact false positives. Accounting for processes other than selection would require introducing parameters that could appropriately capture the effect of these other processes.

Here we present a method that incorporate features of the two types of genome scans described above. The objective is to better discriminate between true and false genetic signatures of local adaptation and simultaneously allow for making inferences about the environmental factors underlying selective pressures. More precisely, our method is based on the Bayesian approach first proposed by Beaumont and Balding (2004) and later extended by Foll and Gaggiotti (2008). The original formulation considers population- and locus-specific F_{ST} 's, which are described by a logistic regression model with three parameters: a locus-specific term, α_i , that captures the effect of mutation and some forms of selection, a population-specific term, β_j , that captures demographic effects (e.g. N_e and migration) and a locus-by population interaction term, γ_{ij} , that reflects the effect of local adaptation. The estimation of the two first terms benefit from sharing information across loci or populations, but this is not the case for the interaction term, which is therefore poorly estimated (Beaumont and Balding, 2004). In practice, therefore, signatures of local adaptation are inferred from the locus-specific effects under the assumption that large positive values reflect adaptive selection. The implicit assumption, therefore, is that background selection and mutation should not have much of an effect on this regression term. In order to relax this assumption and to better estimate the interaction term we introduce environmental data so that $\gamma_{ij} = g_i E_j$, where E_j is the observed value of the environmental variable in population j . We can therefore define three alternative models for each locus (neutral, environment-independent selection, and local adaptation) and evaluate their posterior distribution using a RJMCMC approach as in Foll and Gaggiotti (2008). In what follows we first describe in detail the probabilistic model underlying our Bayesian approach, we then evaluate its performance using simulated data and then present an application using human and salmon datasets. Finally, we discuss the interest of our method, as well as the different and shared features with existing ones.

Statistical model

Modelling allele frequencies using the F model

Our new genome scan approach is based on the F model (Beaumont and Balding, 2004; Foll and Gaggiotti, 2008; Gaggiotti and Foll, 2010) and extends the software BayeScan (Foll and Gaggiotti, 2008) by incorporating environmental data so as to explicitly consider local adaptation scenarios. Full details of the F model are given by Gaggiotti and Foll (2010), so here we only provide a brief description. The core assumptions of the F model

is that all populations share a common pool of migrants, but that their effective sizes and immigration rates are population-specific. Thus, population structure at each locus is described by local F_{ST} 's that measure genetic differentiation between each local population and the migrant pool.

The F model uses the multinomial-Dirichlet likelihood for the allele counts $\mathbf{a}_{ij} = (a_{i,j,1}, \dots, a_{i,j,K_i})$ at locus i within population j (where K_i is the number of distinct alleles at the locus i) with parameters given by the migrant pool allele frequencies, $\mathbf{f}_i = (f_{i,1}, \dots, f_{i,K_i})$, and a population- and locus-specific parameter of differentiation, $\theta_{ij} = \frac{1 - F_{ST}^{ij}}{F_{ST}^{ij}}$:

$$\mathbf{a}_{ij} \sim \text{multDir}(\theta_{ij} f_{i,1}, \dots, \theta_{ij} f_{i,K_i}) \quad (1)$$

where *multDir* stands for the multinomial-Dirichlet distribution.

Although, for the sake of simplicity, we only present here the formulation for co-dominant data, the software implementing our approach also allows for dominant data (e.g. AFLP-type of markers) using the same probabilistic model as Foll and Gaggiotti (2008). Note finally that, for bi-allelic co-dominant markers (e.g. SNP markers), the likelihood reduces to a Beta-binomial model. For dominant data such as AFLP, the model is trying to estimate the F_{IS} as well as the allele frequencies, see Foll and Gaggiotti (2008).

The model for θ_{ij} -population structure and selection

Null model of population structure Under the null hypothesis that all loci are neutral, the local differentiation parameter F_{ST}^{ij} will be driven only by local population demography and, hence, should be common to all loci:

$$\log \left(\frac{F_{ST}^{ij}}{1 - F_{ST}^{ij}} \right) = \log \left(\frac{1}{\theta_{ij}} \right) = \beta_j \quad (2)$$

Note that this model assume that β_j is a logit transformation of the bounded values (between zero and one) of F_{ST}^{ij} , which means that the higher β_j is, the higher the F_{ST}^{ij} . This model is thus a logistic regression on the F_{ST}^{ij} 's. Finally, note that a value of 0 for β_j means a $F_{ST}^{ij} = 0.5$.

A high β_j value means that the population j is strongly differentiated from the pool of migrants. This could be due to a lack of immigration from the other populations, a reduced effective size, or a particular spatial structure. For example, a classical non-cyclic stepping-stone structure will lead to an U-shape distribution of the β_j 's, with higher values toward the boundaries of the meta-population.

Alternative model of local adaptation In this model, we focus on a particular signature left by a process of local adaptation. If selection is driven by a putative environmental factor, we expect that genetic differentiation for the locus or loci under selection will be stronger than expected under neutrality for populations with strong environmental differentiation. Any measure of distance between the environmental value of population j and the average environment could serve as a measure of such differentiation. For the sake of simplicity, we here only consider the absolute value (i.e. Manhattan distance), whose major interest is that it does not over-state

the importance of “outlier” environmental values. Furthermore, in order to facilitate the calibration of prior distributions, we only consider standardised environmental values.

To model the effect of local adaptation on locus i , we consider the impact of environmental differentiation E_j of population j on the locus, we thus modify Eq. 2 as follows:

$$\log \left(\frac{F_{ST}^{ij}}{1 - F_{ST}^{ij}} \right) = \beta_j + g_i E_j \quad (3)$$

where g_i quantifies the effect of factor E_j on the locus i .

Alternative model of locus-specific effect Local adaptation due to our focal environmental variable is not the only evolutionary phenomenon that could lead to departures from the neutral model. Other phenomena that could produce such locus-specific effects include local adaptation due to other unknown factors, large differences in mutation rate across loci and background selection (Charlesworth, 2013). Locus-specific effects considered by existing F_{ST} -based genome scan methods incorporate all these confounding effects as well as that of local adaptation. In the particular case of the model originally implemented in BayeScan (Foll and Gaggiotti, 2008) this is achieved using the following parametrisation for local differentiation:

$$\log \left(\frac{F_{ST}^{ij}}{1 - F_{ST}^{ij}} \right) = \alpha_i + \beta_j \quad (4)$$

The difference between the terms α_i in Eq. 4 and g_i in Eq. 3 are twofold. First, whereas the first only tries to account for departures from the neutral model of unknown origin, the second explicitly assumes a scenario of local adaptation due to the effect of environmental variable E_j . Second, the α_i term is only locus-specific, whereas we can view $g_i E_j = \gamma_{ij}$ as a locus-by-population interaction term.

The main advantage of implementing both of the above alternative models is that we can distinguish between departures from the neutral model of unknown origin (using Eq. 4) and departures due to local adaptation caused by a particular environmental factor (using Eq. 3).

Material and Methods

Implementation of the statistical model

Our method uses two types of data: (i) the allele counts \mathbf{a} for each locus in each population sample, and (ii) observed values \mathbf{E} of an environmental variable (one value for each population), which are transformed into environmental differentiation using an appropriate function. We chose the absolute-value Manhattan distance, because it allows to weigh down the effect of outlier (or, say, strongly differentiated) environmental values. Also pilot studies showed that it had better properties than Euclidian (i.e. squared) distance. Note that

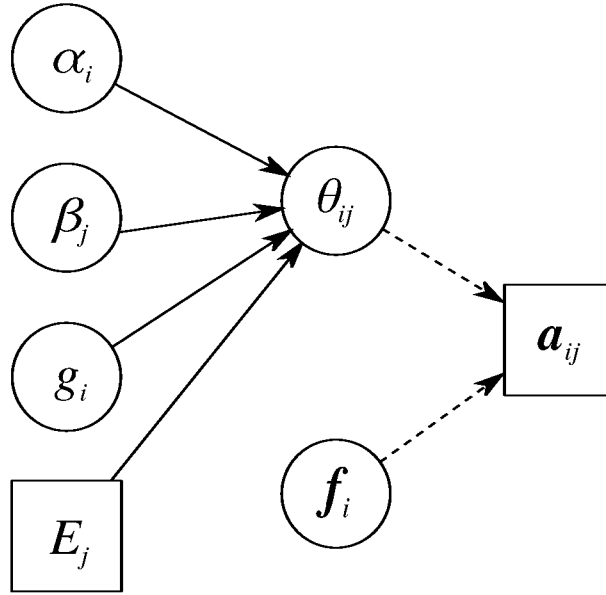


Figure 1: Directed Acyclic Graph (DAG) of the model. Squared nodes denotes know quantities (E for environmental data, and A for genetic marker data). Circled nodes denotes unknown parameters. Plain arrows stands for deterministic relationships, and dashed arrows stands for stochastic relationships.

measuring an environmental distance require to define a reference. The most natural reference would be the mean environmental value, but this is not always the case (see the altitudinal gradient example in Human data analysis). Also, it is strongly advised to standardise the environmental values by dividing by the standard deviation, in order to avoid effect size issues regarding the inference of the parameter g .

As stated in the previous section, there are three alternative models:

M1 Neutral model: β_j

M2 Local adaptation model with environmental differentiation E_j : $\beta_j + g_i E_j$

M3 Locus-specific model: $\alpha_i + \beta_j$

Note that the method of Beaumont and Balding (2004) could not estimate the interaction terms because of the very large number of γ_{ij} parameters and the fact that inferences about them do not benefit from sharing information across populations or loci (but see Riebler *et al.*, 2008). To overcome this problem, we incorporate the environmental data so that the interaction term can be modelled as the product of a known environmental factor specific to population j , and a parameter g_i that quantifies the effect of the environment on locus i . Besides using another source of information (environmental data), the number of parameters that need to be estimated is greatly reduced making inference much easier than with a model with a purely random coefficient for the interaction term.

All 3 models were implemented using an RJMCMC algorithm (Green, 1995). In order to propose relevant values for new parameters during the jumps, the RJMCMC is preceded by pilot runs aimed at both calibrating the

MCMC proposals to reach efficient acceptance rates and approximating the posterior distribution of parameters, as proposed by Brooks (1998) and already implemented in BayeScan (Foll and Gaggiotti, 2008). Our code is based on the source code of BayeScan 2.1 and is written in C++. The source and binaries are available at <https://github.com/devillemereuil/bayescenv>.

Our prior belief in the 3 models is described by 2 parameters: the probability π of moving away from the neutral model and the preference p for **M3** against **M2** as alternative models. We can calculate the prior probability for each model:

$$\begin{aligned} P(\mathbf{M1}) &= 1 - \pi \\ P(\mathbf{M2}) &= \pi(1 - p) \\ P(\mathbf{M3}) &= \pi p \end{aligned} \quad (5)$$

The details of the mathematical calculation of transition between models can be found in the Supplementary Material. Pilot studies showed that using values of p above 0.5 yielded extremely conservative results. Thus, the simulation study (see below) considers only $p = 0.5$ and $p = 0$.

We used a uniform Dirichlet prior for the allele frequencies $\mathbf{f}_i \sim \text{Dir}(1, \dots, 1)$. The priors for the hyperparameters α and β , were Normal with mean -1 and variance 1. When only β is included, this corresponds to a mean prior F_{ST}^{ij} of roughly 0.27. Since under a local adaptation scenario the parameter g is only expected to be positive, it was assigned a uniform prior between 0 and 10. Because our model and code are based on BayeScan, this new model is also able to handle dominant data in estimating the F_{IS} while performing the scan; see Foll and Gaggiotti (2008) for further details.

Statistical tests

Using the posterior probability of model **M2** for the locus i , $P_i(\mathbf{M2}|\mathbf{a}_i, E)$, we can calculate the posterior error probability (PEP, Käll *et al.*, 2008) for locus i as:

$$\text{PEP}_i = 1 - P_i(\mathbf{M2}|\mathbf{a}_i, E) \quad (6)$$

In order to calculate the q -value (Storey, 2003; Muller *et al.*, 2006), we rank the PEP_i from the lowest to the highest value, and define the q -value for locus i as the average PEP for all loci having a PEP lower, or equal to, PEP_i :

$$q_i = \frac{1}{i} \sum_{j, j \leq i} \text{PEP}_j \quad (7)$$

Note that, because we calculate the average using only PEPs that are lower than PEP_i , we have $q_i \leq \text{PEP}_i$ for all i . The equality only holds for the minimal PEP(s). Käll *et al.* (2008) advocate the use of the q -value because it is optimal in the sense of Bayesian classification theory (see also Storey, 2003). In order to let the user the freedom to decide, our code outputs both PEPs and q -values.

Both of these test statistics are strongly related to the control of False Discovery Rate (FDR, Storey, 2002) during multiple testing. Contrary to the commonly used False Positive Rate (FPR), which is the probability of declaring a locus as positive given that it is actually neutral, the FDR is the proportion of the positive results that are in fact false positives. Note that the PEP is a “locally” (i.e. regarding only the focal locus) inferred measure of the FDR (see Käll *et al.*, 2008), whereas the q -value is based on inferring what the FDR would be when stating that the focal locus, and all the loci with a lower score should be considered as positives. In this sense, for the same significance α threshold, using the PEP is more conservative than using the q -value. In the following, we will focus on the q -value of the local adaptation model. Indeed, since we have a strong uncertainty regarding the biological origin of the locus-specific effect α_i , we can consider it as a “nuisance” parameter in this particular testing framework.

Simulation analysis

In order to test our new model against the previous version implemented in BayeScan (Foll and Gaggiotti, 2008), we performed a simulation study. We modelled 16 populations each with 500 individuals genotyped at 5,000 loci, among which 1 (monogenic scenario) or 50 (polygenic scenario) were selected. We modelled three kinds of population structure: (i) a classical Island Model (IM), (ii) a one-dimension Stepping-Stone (SS) model and (iii) a hierarchically structured (HS) model.

For all scenarios, we simulated 16 populations of 500 individuals. The genome was composed of 10 chromosomes along which were regularly spread 500 bi-allelic SNPs (5,000 loci in total). The selected loci, one for the monogenic case and 50 for the polygenic case, were randomly distributed across the genome. The recombination rate was set to 0.002 (one recombination between two adjacent loci, *per* population and *per* generation). The mutation rate was set to 10^{-7} *per* generation at every locus. The allele frequencies were initialised using a Beta-binomial distribution truncated between 0.1 and 0.9 to avoid too many monomorphic loci. Since each markers were independently initialised, our simulations yielded negligible linkage disequilibrium. For the Island Model (IM) and Stepping Stone model (SS), we directly initialised all 16 populations. For the Hierarchically Structured (HS) scenario, we initialised the ancestral population, which, following successive and temporally spaced-out fission events, gave rise to 2, 4, \dots , 16 populations. This hierarchical structure is reinforced by preferential migration between related populations. More details regarding migration and population history are available in the Supplementary Material. This model is very close to that used by de Villemereuil *et al.* (2014). It should be particularly difficult for our method, because all populations are equally differentiated (i.e. the β_j parameters are expected to be roughly the same across populations), but a phylo-geographic covariance exists between related populations, which is not explicitly accounted for by our probabilistic model.

The distribution of the environmental variable was chosen to be independent from the population structure gradient (see Fig. II in SI). This environmental variable was already standardised for its use in BayeScEnv. A

logit-like transformation of this variable was used to define selective pressure:

$$s_j = s_0 \frac{1 - e^{-E_j}}{1 + e^{-E_j}} \quad (8)$$

where s_0 , the strength of the selection was chosen to be 0.1 for the monogenic case and 0.02 for the polygenic case. The individual fitness was calculated in a multiplicative fashion:

$$W = (1 + s_j)^{n_{11}} (1 - s_j)^{n_{00}} \quad (9)$$

Note that this fitness function assumes co-dominance, with a heterozygous fitness of 1. The simulations were performed using the SimuPOP Python library (Peng and Kimmel, 2005) and the scripts are available online. Our simulated datasets were analysed using our C++ code and version 2.1 of BayeScan (Foll and Gaggiotti, 2008).

As stated above, our simulation yielded negligible linkage disequilibrium. Consequently, we considered as true positives only the loci subject to selection. Over the 100 generated datasets for each scenario, we computed the realised FDR, FPR and power yielded by the non-environmental method (BayeScan) and our new environmental method (BayeScEnv). For the latter, we also compared several parameterisations using a prior probability π of jumping away from the neutral model of 0.1 (equivalent to the default prior odds used by BayeScan, which is 10) or of 0.5, as well as a preference for the locus-specific model p of 0.5 (environmental and locus-specific are equiprobable) or 0 (the locus-specific model is forbidden and only the environmental model is tested against the neutral one).

HGDP SNP data analysis

In order to test our new method against a real dataset, we focused on 26 Asian populations from the Human Genome Diversity Panel (HGDP) SNP Genotyping data (<http://www.hagsc.org/hgdp/files.html>). This data set consists into 660,918 SNP markers genotyped using Illumina 650Y arrays. After cleaning the dataset from mitochondrial and sex-linked markers, we removed all markers with minor allele frequency below 5%. This left us with a total of 446,117 SNPs. For all populations, we obtained environmental variables from the BIOCLIM database (<http://worldclim.org/bioclim>): mean annual temperature, precipitation, as well as altitudinal data. We used these 3 environmental variables with BayeScEnv, and compared the results with BayeScan (thus, with no environmental variable). After standardisation of the environmental variables, we computed environmental differentiation from the mean for temperature and precipitation, and from the sea level for elevation. Gene ontology enrichment test for the detected genes was performed using the “SNP mode” of the Gowinda software (Kofler and Schlötterer, 2012). The prior odds for BayeScan was of 10 for this analysis. BayeScEnv prior parameters π and p were respectively set to 0.1 and 0.5 for this analysis.

Atlantic salmon data analysis

We downloaded genetic markers and environmental data on the Atlantic Salmon (*Salmo salmar*) from the Dryad database (Bourret *et al.*, 2014). The data included 3118 SNP markers obtained using Expressed Sequence Tags (EST) and Genome Complexity Reduction (GCR, see Bourret *et al.*, 2013b). The dataset consists of 23 populations from North American coasts. Again, the dataset was cleaned by removing markers with minor allele frequency below 5%, leading to a final dataset of 2078 markers. Environmental data comprised 53 variables, including information relative to local temperature, precipitations, river and geological properties. A PCA was performed. For further analyses, we retained the 3 first axes which were respectively correlated with temperature, precipitations and river properties, as indicated by Bourret *et al.* (2013a). Since PCA scores are already standardised values, we used these variables as such. For the sake of simplicity, we called the first axis “temperature”, the second “precipitation” and the third “river properties”. The prior odds for BayeScan was of 10 for this analysis. We carried out BayeScEnv analyses using $p = 0.5$ and $p = 0$, whereas π was 0.1.

Results

Simulation results

By definition, a threshold value of α used to decide whether q -values are significant or not is expected to yield a FDR of α on the long run, when the model is robust and priors are calibrated. In order to test the robustness of both methods, we used a simulation study with different priors.

As shown in Fig. 2, BayeScan was less well calibrated yielding higher FDRs than BayeScEnv under all scenarios and for both monogenic and polygenic selection. Additionally, for BayeScEnv, the implementation using $\pi = 0.1$ was fairly well calibrated (i.e. the curve is close the grey line in Fig. 2) under the IM scenario (for both monogenic and polygenic versions) and under the polygenic version of the HS scenario. This implementation was much more conservative than the one using $\pi = 0.5$. For $\pi = 0.1$ and $p = 0$, the FDRs were closer to those yielded by BayeScan but still lower.

The higher FDR for BayeScan and BayeScEnv with $\pi = 0.5$ or $p = 0$ was mainly driven by a higher FPR rather than a lack of power (Fig. 3, see also Fig. III in SI). Notably, BayeScan had a quite high power, higher than that of BayeScEnv. Note, however, that BayeScEnv with $p = 0$ had, as BayeScan, a maximal power in the monogenic scenarios, and was almost as powerful as BayeScan in the polygenic scenarios. However, its FDR was lower (sometimes much lower) than that of BayeScan. This indicates that the incorporation of environmental data reduces the error rate even when the effect of spurious locus-specific effects are not taken into account. More details regarding the FPR are available in the Supplementary Information (Fig. III).

Another traditional way to apprehend the compromise between power and false positives is the so-called ROC curve, plotting power against FPR (Fig. 4). In these plots, the curve that is “more to the left” is preferred

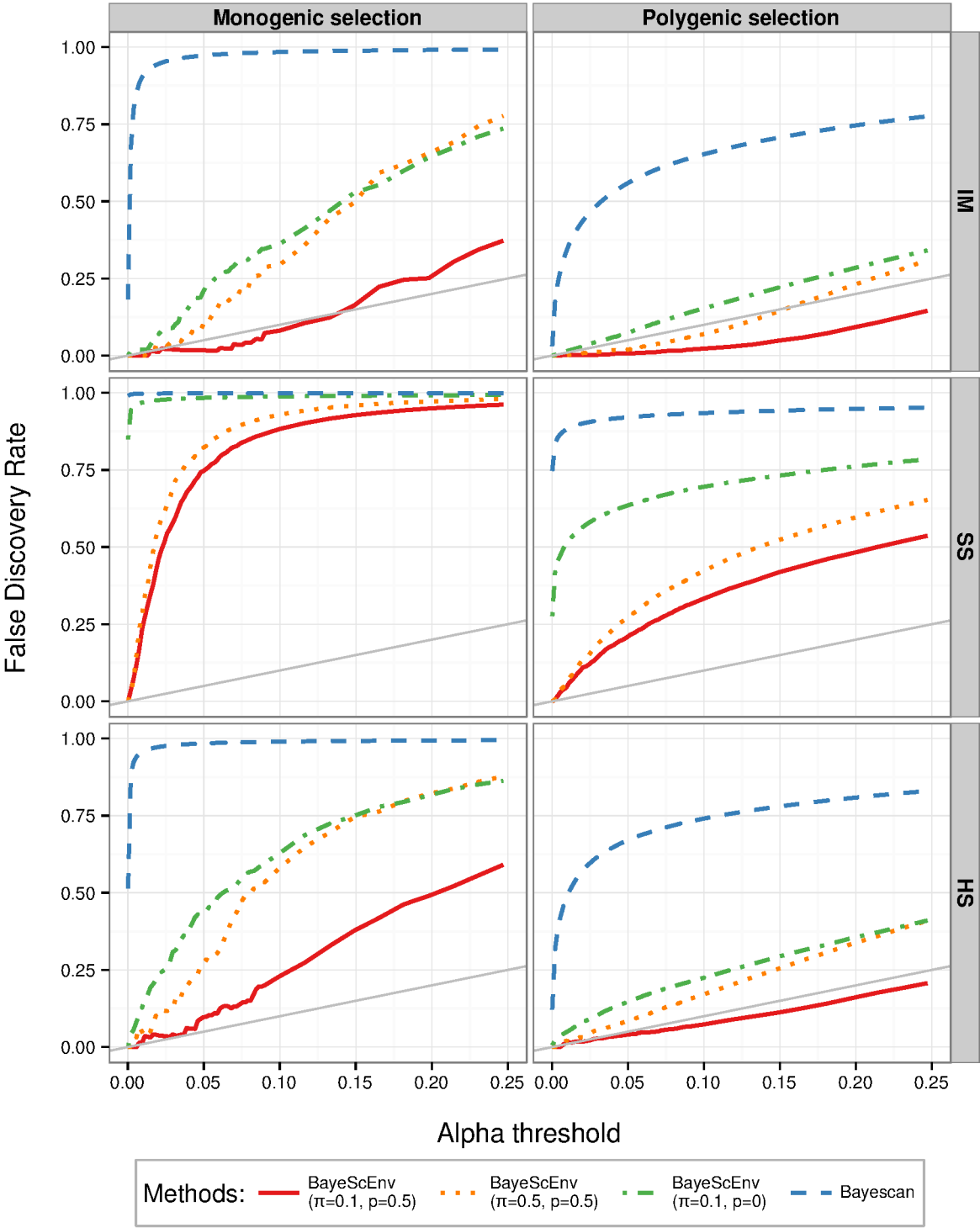


Figure 2: False Discovery Rate (FDR) against significance α threshold for three scenarios (IM: Island model, SS: Stepping-Stone model and HS: Hierarchically Structured model) and monogenic/polygenic selection. The grey line is the expected identity relationship between the FDR and α . The models tested 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.

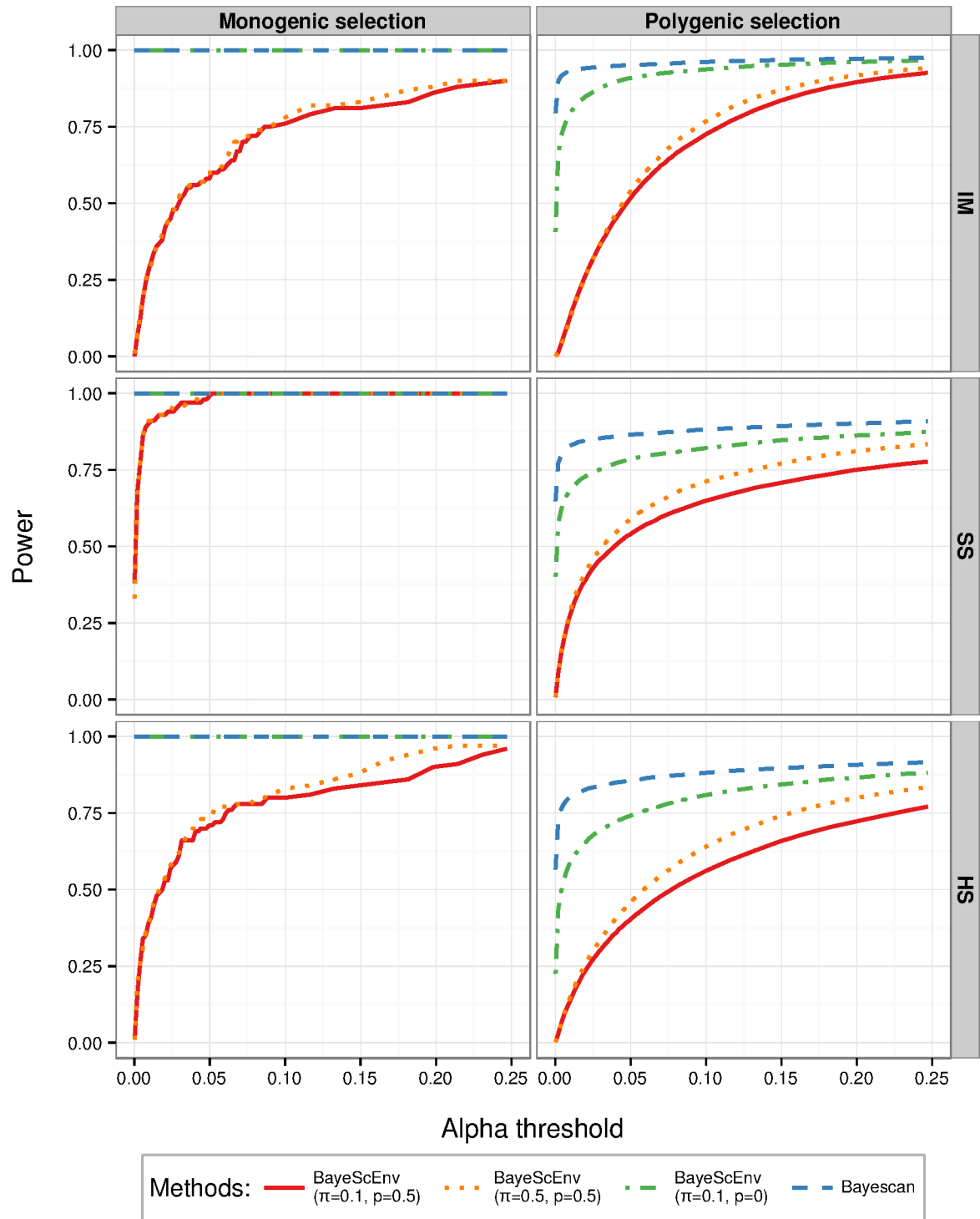


Figure 3: Power against significance α threshold for three scenarios (IM: Island model, SS: Stepping-Stone model and HS: Hierarchically Structured model) and monogenic/polygenic selection. The models tested are BayeScan (blue dashed), and BayeScEnv (orange dotted, green dot-dashed and solid red) with 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.

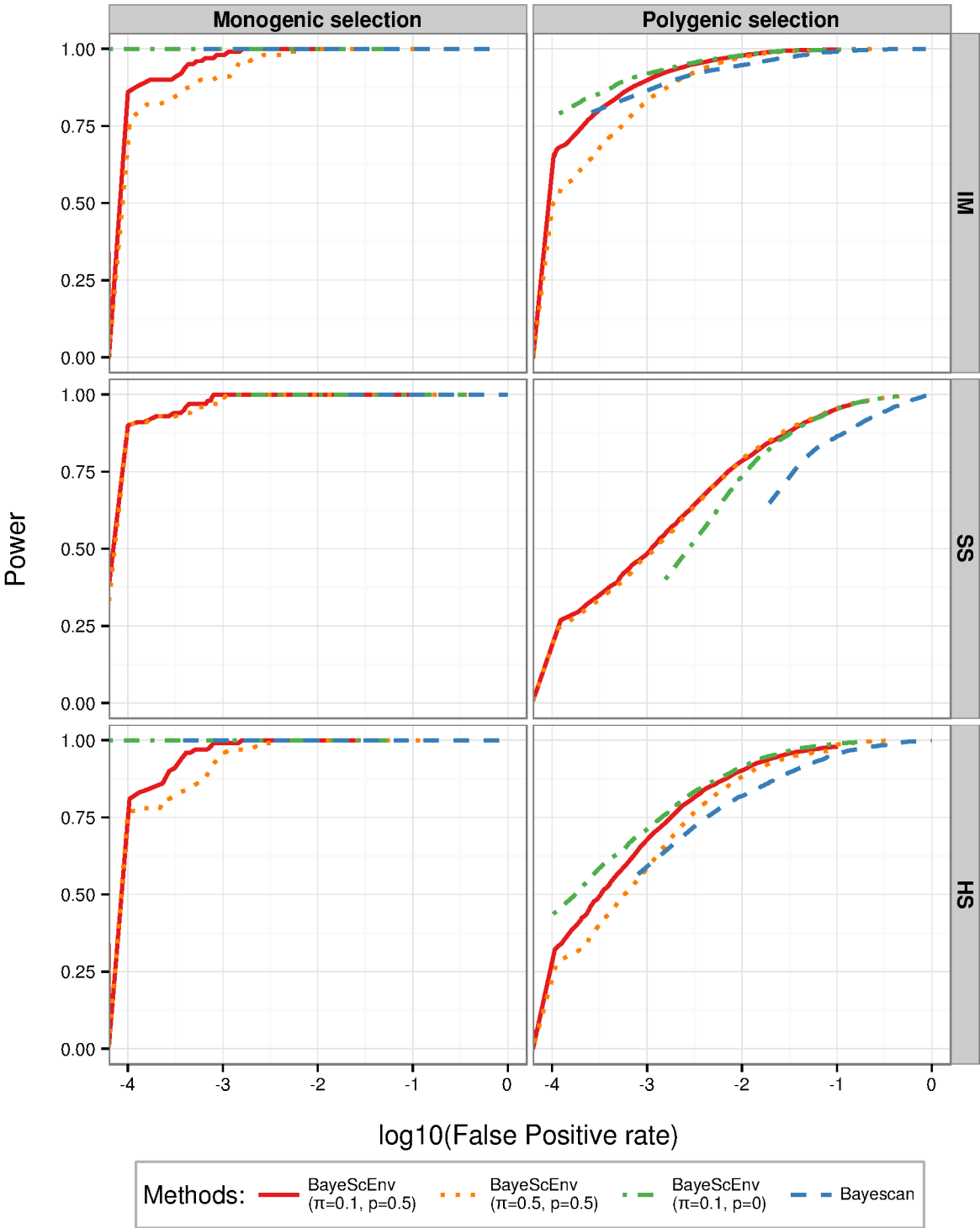


Figure 4: Power against False Positive Rate (FPR), a.k.a. ROC curve, for three scenarios (IM: Island model, SS: Stepping-Stone model and HS: Hierarchically Structured model) and monogenic/polygenic selection. The models tested are Bayescan (blue dashed), and BayeScEnv (orange dotted, green dot-dashed 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.

because this means it offers higher power for a lower FPR. Fig. 4 shows that BayeScEnv with $\pi = 0.1$ and $p = 0$ performed best under the IM and HS scenarios, whereas BayeScEnv with $\pi = 0.1$ and $p = 0.5$ performed better under the “harder” SS scenario. Overall, although BayeScan has higher power to detect local adaptation, it is still too liberal when deciding that a locus is under selection for the scenarios we investigated.

Data analysis on Human adaptation in Asia

The results of the human dataset analyses (Table 1) show a dramatic discrepancy between the two methods. Whereas BayeScan yields a very large number (66,316) of markers considered as significant at the 5% threshold, many fewer markers (154 to 2728) are considered significant by BayeScEnv. Gene Ontology (GO) enrichment tests identified many significant terms (Table 1). Note, however, that in the altitude and temperature analyses they correspond to a small number of genes (11 and 20 respectively, see Table 1). The number of genes is larger for the precipitation analysis (359) and even larger than the number of significant GO terms in the BayeScan analysis (5628).

Regarding the altitude, significant biological processes included the fatty acid metabolism (e.g. SCARB1), skin pigmentation (e.g. MLANA, SLC24A5), kidney activity (e.g. SLC12A1) and oxido-reductase activity (e.g. NOS1AP, which is also involved in cardiac rhythmicity, see Arking *et al.*, 2006). Regarding the temperature, significant biological process included cardiac muscle activity (e.g. SLC8A1) and development (e.g. NRG1, FOXP1), fatty acid metabolism (e.g. FADS1, FADS2) and response to hypoxia (e.g. SLC8A1, SERPINA1). For the precipitation analysis with BayeScEnv, as well as the BayeScan analysis, the number of significant terms was too large for hand-picked examples to be feasible.

The significance results (q -values) are displayed as a Manhattan plot in Fig. 5, along with the above mentioned genes for the altitude and temperature analyses (Fig. 5, A and B). We can see that some other regions of the genome include outlier loci but they correspond to non-coding regions, or are close to genes associated to GO terms that were not significant, or proteins without a known function (e.g. C9orf91, which was the most significant gene in the temperature analysis). Pattern of linkage disequilibrium was visible, which sometimes strongly supported some candidate genes (Fig. 5, A, SLC12A1 and SLC24A5). Finally, comparing BayeScEnv (Fig. 5, A, B and C) and BayeScan analyses (Fig. 5,D), we see that BayeScan yielded too many significant markers for a Manhattan plot to be a useful display of the results. An interesting pattern, is that BayeScan yielded far more outlier markers with maximal certainty (e.g. posterior probability of one) than BayeScEnv. For the present dataset, 22,516 markers had a posterior probability of one, whereas the maximal posterior probability yielded by BayeScEnv was 0.9998. Finally, almost all loci detected using BayeScEnv were also found when using BayeScan (between 98% for altitude to 100% for the two other variables).

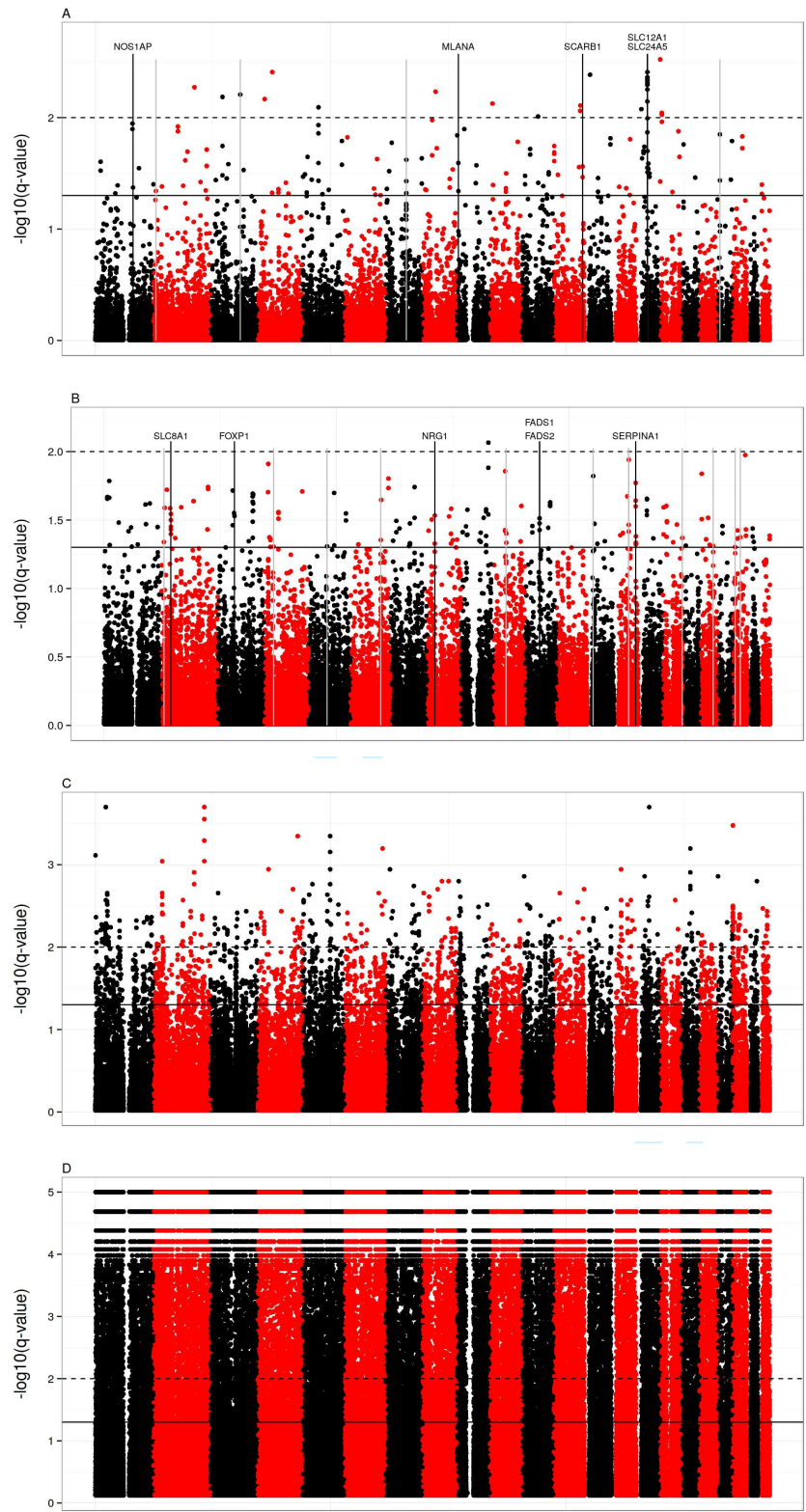


Figure 5: Manhattan plot of the q -values for the Human dataset when using BayeScEnv with altitude (A), temperature (B), precipitations (C) or when using BayeScan (D). For altitude and temperature (A and B), genes mentioned in the text are displayed using black lines and genes associated with a significant GO term using grey lines. Top “stripes” for BayeScan (D) are artefacts due to finite number of iterations in RJMCMC (e.g. 0, 1, 2, 3... iterations outside of model 3), corresponding to determined posterior probabilities when divided by the total number of iterations.

Method	Variable	Nr of significant SNPs	Nr of significant GO terms	Nr of genes associated with a significant GO term
BayeScEnv	altitude	154	32	11
	temperature	170	103	20
	precipitation	2728	439	359
BayeScan	—	66,316	469	5628

Table 1: Results from the BayeScan and BayeScEnv on the Human dataset. FDR significance threshold was set to 5%. The total number of tested markers was 446,117.

Method	Variable	Number of significant	
		with $p = 0.5$	with $p = 0$
BayeScEnv	temperature	8	62
	precipitations	5	45
	river properties	0	46
BayeScan	—	238	

Table 2: Results from the BayeScan and BayeScEnv on the salmon dataset. FDR significance threshold was set to 5%. The total number of tested markers was 2078.

Data analysis of the Atlantic Salmon

The results of the salmon dataset analyses (Table 2) are again a clear demonstration of the discrepancy between the methods. Whereas BayeScan yields 238 SNPs significant at the 5% level, BayeScEnv yields between 0 and 8 significant markers with $p = 0.5$ and between 45 and 62 with $p = 0$. Thus, in agreement with the theoretical expectations and the simulation studies, we obtain more candidates loci with $p = 0$. All loci obtained when using BayeScEnv (for any variable and any value of p) were found when using BayeScan. Unfortunately, the Atlantic salmon genome is poorly annotated and, therefore, it was not possible to carry out a gene ontology enrichment analysis.

Discussion

Features and performance of the method

The method we introduce in this paper, BayeScEnv, has several desirable features. First, just as BayeScan, it is a model-based method. This means that the null model can be understood in terms of a process of neutral evolution. One can thus predict what the method is able to fit or not. Second, we explicitly model a process of local adaptation caused by an environmental variable. Third, in order to render the model more robust, we account for departures from neutrality unrelated to the environmental variables under consideration. These departures can be due to another process of local adaptation (i.e. not caused by the investigated environmental variable), to large differences in mutation rates across loci, to background selection (Charlesworth, 2013) or complex spatial effects.

Our simulation results, as well as our data analyses show that when compared to BayeScan, BayeScEnv has a better control of its false discovery rate under various scenarios (Fig. 2), yielding fewer, but more reliable

candidate markers. Obviously, this has a cost in terms of absolute power (Fig. 3), but BayeScEnv still performs better than BayeScan in terms of the compromise between true and false positives (Fig. 4). Note that the false discovery rate also is a measure of this compromise since, everything else being equal, increasing power should decrease the false discovery rate. Although losing power might seem disappointing, the control of the compromise between true and false positives is of utmost importance when performing such genome scans.

Another issue is the typical tendency of BayeScan to yield maximal certainty for a non negligible (and sometimes important) number of loci (c.f. Fig. 5, D). On the other hand, the only case of maximal certainty for BayeScEnv was observed in the simulation analysis of the monogenic case (i.e. with a strong signal). This can be partly explained by the competition between the two non neutral models. However, even in the case where $p = 0$, the information about the environmental variable is probably too constraining for maximal certainty to occur easily.

Finally, the parametrisation of BayeScEnv allows for a fine and intuitive control of the false positive rate and power. For example, setting p to 0 increases both power and false positive rate, whereas setting $p = 0.5$ will allow for a more conservative test. Our simulation results show that, as long as the species under study has moderate to large dispersal abilities (c.f. hierarchical structure or island model), the former parametrisation will be more efficient, whereas for species with low dispersal abilities (c.f. stepping stone model) the latter should be preferred. Overall, the issue of which parameter set to use depends essentially on the demography of the species, as well as one's risk aversion.

We note that BayeScan was recently extended to consider species with hierarchical population structure (BayeScan3, Foll *et al.*, 2014). With BayeScan3, it is now possible to study widely distributed species covering several continents or geographic regions. It is also possible to better focus on local adaptation by considering groups that include pairs of populations inhabiting different environments such as low and high altitude habitats. Thus, BayeScan3, allows for the consideration of categorical environmental variables. Our new approach on the other hand allows the study of local adaptation processes related to continuous environmental variables in species with a more restricted range. Additionally, although we do not explicitly model more complicated demographic scenarios, our new method is robust to deviations from the island model thanks to the inclusion of a locus-specific effect unrelated to environmental conditions.

How to define the ‘environmental differentiation’?

To model local adaptation, we compute an “environmental differentiation” as a measure of distance (here the Manhattan, absolute value distance) to a reference value. Although this reference can conveniently be chosen as the mean of the environmental values, other kind of reference may be more biologically relevant. It is of crucial importance to correctly define this environmental differentiation and its associated reference, as this can have consequences for the interpretation of the results. For example in the case of analysis focused on elevation, it

seems appropriate to use sea level as the reference. Indeed, given the kind of environmental variables elevation is a proxy for (e.g. partial pressure of oxygen, temperature, solar radiation, etc.), for most systems we would consider the sea level as a neutral environment rather than the differentiated one.

Any univariate measure of environmental differentiation can be used to describe the environment of the populations (e.g. Euclidian distance), but we found that the Manhattan distance was the most efficient in our preliminary studies (results not shown). Another way to account for environmental differentiation is to use Principal Component Analysis (PCA), providing each axis to BayeScEnv as a description of the distance between environments. Despite this practice being an elegant way to summarise environmental distance between populations, it also has the drawback of making it more difficult to identify the “causal” variable. However, this issue concerning the use of either “raw” variables or PC axes is applicable to any environmental association method.

Note that the environmental variables must be standardised (i.e. mean-centred and reduced) so that the values of the parameter g are commensurate with the values of α and β . Also the priors of g in BayeScEnv are calibrated for the case of standardised environmental values.

Comparison with other environmental association methods

There are several genome scan approaches that incorporate environmental information. Some are mechanistic (e.g. BayEnv, Coop *et al.*, 2010) while others are phenomenological (e.g. LFMM, Frichot *et al.*, 2013 and the method of Guillot *et al.*, 2014). These methods perform a logistic regression between allele frequencies and environmental values, while controlling for population structure. Only BayEnv has an explicit model of neutral evolution underlying its model of population structure (Nicholson *et al.*, 2002; Balding and Nichols, 1995). LFMM uses Latent Factors inferred from the genotypic data themselves to control for population structure, whereas Guillot *et al.* (2014) use Random Gaussian Fields to take into account spatial effects. A limited analysis using simulated datasets from de Villemereuil *et al.* (2014), available in SI, show that even in difficult conditions, BayeScEnv can compete with other environmental association methods.

Interestingly, all these methods can account for covariance between populations, while BayeScEnv assumes that all populations are independent, exchanging gene flow only through the migrant pool, thus not preferentially with any other population. However, the inclusion of a locus-specific effect unrelated to the environment helps to take into account spatial effects due to deviations from the underlying demographic model. In this regard, our simulation study showed that the method was fairly robust to isolation-by-distance and hierarchically structured scenarios. Nevertheless, we note that BayeScEnv is best suited for species with high dispersal abilities such as marine species and anemophilous plants.

Another point that distinguishes BayeScEnv from these methods is that it does not assume any particular functional form for the relationship between environmental values and allele frequencies. While existing

association methods all assume a clinal pattern, BayeScEnv only assumes that genetic differentiation increase exponentially with environmental differentiation. This allows for a more diverse family of relationships between allele frequencies and environment. For example, a scenario where the same allele is favoured at the margins and counter selected in the middle of the species range can be studied with BayeScEnv but would certainly represent a problem for the other association methods. Such scenario would arise when the target of the selection in extreme environments are plasticity genes (Morris *et al.*, 2014), or regulatory genes for the response to stress. Alternatively, the chosen environmental variable might actually be a proxy of another selective variable that takes similar values when the former takes very low or very high values. For example, environments with very low or very high temperatures are often also arid environments. Another difficult scenario that BayeScEnv would be able to detect is one in which two populations with very similar values for the environmental factor have very different alleles frequencies at a locus and both experience environmental conditions very different from those of the other populations. Such patterns are difficult to relate to local adaptation and might most likely be caused by high drift in extreme environments, due to reduced population sizes. Yet such signal should in principle be readily captured by the β parameter. One particular selective scenario that would explain such pattern would be one of positive frequency-dependent selection modulated by the environment (i.e. only extreme environments would undergo such selection), as expected in the case of Mullerian mimicry with differential predator pressure (Borer *et al.*, 2010). Nevertheless, the number of species where such a scenario would be biologically plausible is limited. In any case, loci showing such a pattern can be easily identified by post-hoc inspection of their allele frequencies in the different populations. It would then be possible to label these loci as false positives if frequency-dependent selection is deemed an unlikely scenario. All these scenarios are illustrated in the SI.

Finally, BayeScEnv is one of the very few genetic-environment association methods that can be used with dominant data (but see also Guillot *et al.*, 2014).

Data analysis

When confronted with real datasets, BayeScEnv typically returned fewer significant markers than BayeScan. This is explained by the focus on searching for outliers linked to a specific environmental factor and to the lower false positive rate of our approach. In accordance with this expectation, its application to the Human dataset identified several genomic regions that gene ontology enrichment identified as relevant to potential local adaptation to altitude or temperature. We emphasise that this study was not meant to exhaustively and rigorously investigate local adaptation in Asian Human populations. However, our results tend to demonstrate that the candidates yielded by BayeScEnv have a biological interpretation. For example, skin pigmentation and cardiac activity could clearly be involved in responses to increased solar radiation and depleted oxygen availability at high elevation. While kidney activity seems less obviously linked to elevation, the gene SLC12A1 is involved in the production of the Na⁺,K⁺,Cl⁻ cotransporter production in the kidney, which is responsible

for the re-absorption of these molecules back into the blood. This could be linked to an adaptation to the risk of respiratory alkalosis (Hickam *et al.*, 1956; Sutton *et al.*, 1979) at high altitudes. Much of the ontologies linked to temperature were potentially confounded with adaptation to altitude, such as the response to hypoxia and cardiac muscle activity. Also, fatty acid metabolism was associated to both altitude and temperature. The role of fatty acid metabolism could be compact energy storage, thermal isolation or increased membrane fluidity in response to cold (Los and Murata, 2004). Of course, the biological functions described here do not account for all the signals yielded by BayeScEnv (see Fig. 5, A and B), other genomic significant regions include genes with less obvious biological function regarding local adaptation, non coding regions and proteins without a known function.

Finally, the analysis using the precipitation variable yielded too many significant markers for a detailed analysis of the biological functions involved. This may not necessarily be due to a confounding effect of the spatial structure (the Human Asian populations being structured mainly from West to East, while the Eastern climate is characterised by strong precipitations during the monsoon), since precipitation may behave as a surrogate for several environmental variables.

A comparison with the genes detected as linked to adaptation to altitude in humans by Foll *et al.* (2014) shows little overlap: only one gene associated to altitude (HEMK1, involved in tyrosine metabolism) and one gene associated to temperature (FBP2, involved in fructose metabolism) were in common with this previous study. For the genes associated with precipitation, 9 genes overlap, but the original set of genes is considerably larger. It is worth noting that although the overlap in terms of candidate genes between the two studies is limited, the metabolic processes they identified as allowing humans to live at high altitude do overlap. For example, one of the genes that our study associates with altitude (SCARB1), was not identified by Foll *et al.* but is involved in fatty acid metabolism, which was also identified as an important process in their study. An explanation for the discrepancies between the two analyses could be the very different datasets used. Foll *et al.* used two group of populations from different continents (America and Asia), each with one high and one low altitude population, a sampling design specially well adapted for the study of local adaptation to high altitude. Here we used 26 populations from Asia, that were not sampled with the specific objective of studying adaptation to high altitude. Because when studying local adaptation the geographical scale matters a lot (Brachi *et al.*, 2013), the observed differences are not surprising. Additionally, our dataset contains half as many loci as that of Foll *et al.*.

The Atlantic salmon genome is poorly annotated so we could not identify genes associated to the observed outlier loci. However, the discrepancy between the number of candidates yielded by BayeScan and BayeScEnv was still quite impressive in this case. Also, when using the parametrization $p = 0$, we obtained almost an order of magnitude more candidates (though our simulations tend to demonstrate that this was probably at a cost of a larger false positive rate).

Conclusion

The main improvement introduced by our new method, BayeScEnv over existing F_{ST} -based genome-scan approaches is the possibility of focusing on the detection of outlier loci linked to genomic regions involved in local adaptation and better distinguishing between the signal of positive selection and that of other locus-specific processes such as mutation and background selection. Although it does not explicitly model complex spatial effects, the consideration of two different locus-specific effects make it more robust to potential deviations from the migrant pool model. This is reflected in its much lower FDR when compared to BayeScan.

Our new formulation also allows for an improved control of the true/false positives compromise through the parameter p , which describes our preference for the model that includes a locus-specific effect unrelated to the environmental factor over the model that includes environmental effects. Although we recommend using $p = 0.5$, lower values (including 0) could be used if population structure is weak or maximising power is more important than reducing the false positive rate.

With this new method, there are now three alternative formulations of genome scan methods based on the F-model. BayeScan detects a wide range of locus-specific effects including background selection. Although its false discovery rate is higher than that of the two extensions, it is able to detect regions of the genome subject to purifying selection. The hierarchical version of this original formulation, BayeScan3, allows the study of local adaptation due to categorical environmental factors. Finally, our new method, BayeScEnv, is more appropriate to detect genomic regions under the influence of selective pressures exerted by continuous environmental variables. Thus, all three methods are complementary and jointly cover scenarios applicable to a wide range of species

Acknowledgement

We thank M. Foll for providing the source code of BayeScan and for clarifying several issues related to the code, J. Renaud for his help on getting the average altitude out of the HGDP latitude/longitude data, S. Schoville for the BIOCLIM data, E. Bazin for his help on the HGDP data analysis and V. Bourret for his help on the Salmon dataset. PdV was supported by a doctoral studentship from the French *Ministère de la Recherche et de l'Enseignement Supérieur*. OEG was supported by French ANR grant No 09 GENM 017 001 and by the Marine Alliance for Science and Technology for Scotland (MASTS).

References

Akey JM, Zhang G, Zhang K, Jin L, Shriver MD (2002) Interrogating a high-density SNP map for signatures of natural selection. *Genome Research*, **12**(12), 1805–1814.

- Arking DE, Pfeufer A, Post W, Kao WHL, Newton-Cheh C, Ikeda M, West K, Kashuk C, Akyol M, Perz S, Jalilzadeh S, Illig T, Gieger C, Guo CY, Larson MG, Wichmann HE, Marb  n E, O'Donnell CJ, Hirschhorn JN, K   b S, Spooner PM, Meitinger T, Chakravarti A (2006) A common genetic variant in the NOS1 regulator NOS1ap modulates cardiac repolarization. *Nature Genetics*, **38**(6), 644–651.
- Balding DJ, Nichols RA (1995) A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. *Genetica*, **96**(1-2), 3–12. WOS:A1995RG20800002.
- Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**(4), 969–980.
- Blanquart F, Kaltz O, Nuismer SL, Gandon S (2013) A practical guide to measuring local adaptation. *Ecology Letters*, **16**(9), 1195–1205.
- Bonhomme M, Chevalet C, Servin B, Boitard S, Abdallah J, Blott S, SanCristobal M (2010) Detecting selection in population trees: The lewontin and krakauer test extended. *Genetics*, **186**(1), 241–262.
- Borer M, Van Noort T, Rahier M, Naisbit RE (2010) Positive frequency-dependent selection on warning color in alpine leaf beetles. *Evolution*, **64**(12), 3629–3633.
- Bourret V, Dionne M, Bernatchez L (2014) Data from: Detecting genotypic changes associated with selective mortality at sea in atlantic salmon: polygenic multi-locus analysis surpasses genome scan.
- Bourret V, Dionne M, Kent MP, Lien S, Bernatchez L (2013a) Landscape genomics in atlantic salmon (*salmo salar*): Searching for gene–environment interactions driving local adaptation. *Evolution*, **67**(12), 3469–3487.
- Bourret V, Kent MP, Primmer CR, Vasem  gi A, Karlsson S, Hindar K, McGinnity P, Verspoor E, Bernatchez L, Lien S (2013b) SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range of atlantic salmon (*salmo salar*). *Molecular Ecology*, **22**(3), 532–551.
- Brachi B, Villoutreix R, Faure N, Hautekeete N, Piquot Y, Pauwels M, Roby D, Cuguen J, Bergelson J, Roux F (2013) Investigation of the geographical scale of adaptive phenological variation and its underlying genetics in *arabidopsis thaliana*. *Molecular Ecology*, **22**(16), 4222–4240. WOS:000322886400008.
- 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.
- Charlesworth B (1998) Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution*, **15**(5), 538–543.

- Charlesworth B (2013) Background selection 20 years on: The wilhelmine e. key 2012 invitational lecture. *Journal of Heredity*, **104**(2), 161–171.
- Coop G, Witonsky D, Di Rienzo A, Pritchard JK (2010) Using environmental correlations to identify loci underlying local adaptation. *Genetics*, **185**(4), 1411–1423.
- De Mita S, Thuillet AC, Gay L, Ahmadi N, Manel S, Ronfort J, Vigouroux Y (2013) Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular Ecology*, **22**(5), 1383–1399.
- Duforet-Frebourg N, Bazin E, Blum MGB (2014) Genome scans for detecting footprints of local adaptation using a bayesian factor model. *Molecular Biology and Evolution*, p. msu182.
- Edelaar P, Burraco P, Gomez-Mestre I (2011) Comparisons between QST and FST—how wrong have we been? *Molecular Ecology*, **20**(23), 4830–4839.
- Faria R, Renaut S, Galindo J, Pinho C, Melo-Ferreira J, Melo M, Jones F, Salzburger W, Schluter D, Butlin R (2014) Advances in ecological speciation: an integrative approach. *Molecular Ecology*, **23**(3), 513–521. WOS:000329980000003.
- Foll M, Gaggiotti OE (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A bayesian perspective. *Genetics*, **180**(2), 977–993.
- Foll M, Gaggiotti OE, Daub JT, Vatsiou A, Excoffier L (2014) Widespread signals of convergent adaptation to high altitude in asia and america. *The American Journal of Human Genetics*, **95**(3).
- Frichot E, Schoville SD, Bouchard G, François O (2013) Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, **30**(7), 1687–1699.
- Gaggiotti OE, Bekkevold D, Jørgensen HBH, Foll M, Carvalho GR, Andre C, Ruzzante DE (2009) Disentangling the effects of evolutionary, demographic, and environmental factors influencing genetic structure of natural populations: Atlantic herring as a case study. *Evolution*, **63**(11), 2939–2951.
- Gaggiotti OE, Foll M (2010) Quantifying population structure using the f-model. *Molecular Ecology Resources*, **10**(5), 821–830.
- Green PJ (1995) Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, **82**(4), 711–732.
- Guillot G, Vitalis R, Rouzic Al, Gautier M (2014) Detecting correlation between allele frequencies and environmental variables as a signature of selection. a fast computational approach for genome-wide studies. *Spatial Statistics*, **8**, 145–155.

- Günther T, Coop G (2013) Robust identification of local adaptation from allele frequencies. *Genetics*, **195**(1), 205–220.
- Hickam JB, Wilson WP, Frayser R (1956) Observations on the early elevation of serum potassium during respiratory alkalosis 1. *Journal of Clinical Investigation*, **35**(6), 601–606.
- Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters*, **7**(12), 1225–1241.
- Kofler R, Schlötterer C (2012) Gowinda: unbiased analysis of gene set enrichment for genome-wide association studies. *Bioinformatics*, **28**(15), 2084–2085.
- Kruuk LEB, Baird SJE, Gale KS, Barton NH (1999) A comparison of multilocus clines maintained by environmental adaptation or by selection against hybrids. *Genetics*, **153**(4), 1959–1971. WOS:000084079300040.
- Käll L, Storey JD, MacCoss MJ, Noble WS (2008) Posterior error probabilities and false discovery rates: Two sides of the same coin. *Journal of Proteome Research*, **7**(1), 40–44.
- Los DA, Murata N (2004) Membrane fluidity and its roles in the perception of environmental signals. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, **1666**(1–2), 142–157.
- Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, **4**(12), 981–994.
- Morris MRJ, Richard R, Leder EH, Barrett RDH, Aubin-Horth N, Rogers SM (2014) Gene expression plasticity evolves in response to colonization of freshwater lakes in threespine stickleback. *Molecular Ecology*, **23**(13), 3226–3240.
- Muller P, Parmigiani G, Rice K (2006) FDR and bayesian multiple comparisons rules. *Johns Hopkins University, Dept. of Biostatistics Working Papers*.
- Nicholson G, Smith AV, Jónsson F, Gústafsson O, Stefánsson K, Donnelly P (2002) Assessing population differentiation and isolation from single-nucleotide polymorphism data. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **64**(4), 695–715.
- Peng B, Kimmel M (2005) simuPOP: a forward-time population genetics simulation environment. *Bioinformatics*, **21**(18), 3686–3687.
- Radwan J, Babik W (2012) The genomics of adaptation. *Proceedings of the Royal Society B: Biological Sciences*, **279**(1749), 5024–5028.
- Riebler A, Held L, Stephan W (2008) Bayesian variable selection for detecting adaptive genomic differences among populations. *Genetics*, **178**(3), 1817–1829.

- Shendure J, Ji H (2008) Next-generation DNA sequencing. *Nature Biotechnology*, **26**(10), 1135–1145.
- Storey JD (2002) A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **64**(3), 479–498.
- Storey JD (2003) The positive false discovery rate: A bayesian interpretation and the q-value. *Annals of Statistics*, pp. 2013–2035.
- Sutton RAL, Wong NLM, Dirks JH (1979) Kidney international - abstract of article: Effects of metabolic acidosis and alkalosis on sodium and calcium transport in the dog kidney. *Kidney Int*, **15**(5), 520–533.
- de Villemereuil P, Frichot E, Bazin E, François O, Gaggiotti OE (2014) Genome scan methods against more complex models: when and how much should we trust them? *Molecular Ecology*, **23**(8), 2006–2019.

Data Accessibility

The Python code used to simulate data is available online in the Supplementary Information. The software is available online at GitHub: <http://github.com/devillemereuil/bayescenv>.

Author contributions

PdV and OEG designed the statistical model. PdV modified the C++ code and performed the simulation and data analysis. PdV and OEG wrote the article.