HacDivSel: Two new methods (haplotype-based and outlier-based) for the detection of divergent selection in pairs of populations of non-model species.

A. Carvajal-Rodríguez

Departamento de Bioquímica, Genética e Inmunología. Universidad de Vigo, 36310 Vigo, Spain.

Keywords: haplotype allelic class, $F_{ST}$, $G_{ST}$, outlier test, divergent selection, genome scan, non-model species.

*: A. Carvajal-Rodríguez. Departamento de Bioquímica, Genética e Inmunología. Universidad de Vigo, 36310 Vigo, Spain. Phone: +34 986813828

e-mail: acraaj@uvigo.es

Running title: HacDivSel: detection of divergent selection
Abstract

The detection of genomic regions involved in local adaptation is an important topic in current population genetics. There are several detection strategies depending on the kind of genetic and demographic information at hand. A common drawback is the high risk of false positives. In this study, we introduce two complementary methods for the detection of divergent selection from populations connected by migration. Both methods have been developed with the aim of being robust to false positives. The first method combines haplotype information with inter-population differentiation ($F_{ST}$). Evidence of divergent selection is concluded only when both the haplotype pattern and the $F_{ST}$ value support it. The second method is developed for independently segregating markers i.e. there is no haplotype information at hand. In this case, the power to detect selection is attained by developing a new outlier test based on detecting a bimodal distribution. The test computes the $F_{ST}$ outliers and then assumes that those of interest would have a different mode which is detected by a clustering algorithm. The utility of the two methods is demonstrated through simulations and the analysis of real data. The simulation results show power ranging from 60-94% in several of the scenarios while the false positive rate is controlled below the nominal level in every scenario. The analysis of data from intertidal marine snail ecotypes confirms lower number of outliers than previously estimated, maybe as a result of previous false positives or, alternatively, due to our outlier method conservativeness. The software HacDivSel implements both methods.
Introduction

Current population genetics has an important focus on the detection of the signature of natural selection at the molecular level. The detection of the selection effect in a given DNA region matters because it may connect such region with key functionality or with past or ongoing selective events and, in general, may give us a deeper understanding of evolutionary processes. More specifically, it may help to understand the evolutionary mechanisms allowing the species adaptation to local conditions. The study of local adaptation processes implies that some genetic variant is favored by the local environmental conditions. The positively selected locus increase in frequency and the pattern of variation around that locus will change through the process known as selective sweep (Smith and Haigh 1974; Nielsen et al. 2005).

There are many tests designed to detect different kind of effects produced by selective sweeps. Such effects can involve skewed site frequency spectra, high linkage disequilibrium or high rates of genetic divergence (Crisci et al. 2013; Jensen, Foll, and Bernatchez 2016). The information required by those tests is different; they could need knowledge about candidate adaptive loci, the haplotypic phase, the recombination rates, or the ancestral/derived status at each segregating site. This kind of information is often available for model organisms and so, in previous years, most of the effort was focused in humans and other model organisms. In the case of non-model organisms, the most useful methods for studying local adaptation have been those based on measuring genetic differentiation between populations. The idea behind these methods is that the loci involved in local adaptation would be outliers, i.e. would have unusually large values of $F_{ST}$. From its original formulation (LK test, Lewontin and
Krakauer 1973), this technique has been improved in several ways depending on the
summary statistic’s - the $F_{ST}$ or any other differentiation index - expected neutral distribution.
That is, in order to account for more realistic situations, the different methods change the
assumptions of the null demographic model (reviewed in Whitlock and Lotterhos 2015).

Even so, one of the main drawbacks of outlier-based methods is the difficulty of defining an
accurate null model since there are several historical events and demographic scenarios,
other than local selection, that can produce similar $F_{ST}$ patterns. Still in the presence of local
adaptation we can expect different $F_{ST}$ patterns, since the populations involved may be more
or less connected by migration, and this will influence the structure of the genetic variation
both at intra and inter-population levels. Consequently, when using outlier methods based
on the deviation over an expected demographic null model, we always face the risk of
having an excess of false positives (Perez-Figueroa et al. 2010; Bierne, Roze, and Welch 2013;

Another problem concerning outlier-based methods is that they have low power when the
overall $F_{ST}$ is high as it can happen when the genetic basis of the adaptation is polygenic or
when the populations under study are subspecies (De Villemereuil et al. 2014; Whitlock and
Lotterhos 2015). Fortunately, in recent years, the amount of information on genomes of
several species has increased (Ellegren 2014) and consequently new and more sophisticated
methods can now be applied to detect local adaptation in non-model organisms. As
mentioned above, the linkage disequilibrium (LD) is the basis for several computational
methods aiming the detection of selective sweeps (reviewed in Crisci et al. 2012). Some LD-
based methods try to identify the region with maximized LD (Kim and Nielsen 2004; Pavlidis,
Jensen, and Stephan 2010) while others explore the pattern of LD decay from candidate SNPs (Voight et al. 2006; Sabeti et al. 2007). However, only few LD-based methods have considered structured populations as the evolutionary scenario of interest. Still, there are rather different scenarios that can be evaluated under structured populations (Vatsiou, Bazin, and Gaggiotti 2016). Therefore, although LD-based tests can be powerful and robust for detecting selective sweeps in isolated or simple structured population scenarios with low migration rate, they are not, under several other realistic scenarios (Crisci et al. 2012; Crisci et al. 2013; Rivas, Dominguez-Garcia, and Carvajal-Rodriguez 2015; Vatsiou, Bazin, and Gaggiotti 2016).

Moreover, the possibility of observing local adaptation with gene flow depends on the demography and on the genetic basis of the traits involved (Yeaman and Otto 2011). This complicates the performance of the methods under moderate-to-high migration scenarios (Vatsiou, Bazin, and Gaggiotti 2016). Thus, even if haplotype phase information is at hand, specific methods should be developed to detect local adaptation under structured population scenarios using LD-like patterns (Rivas, Dominguez-Garcia, and Carvajal-Rodriguez 2015).

The aim of this paper is to present two complementary methods specialized to detect divergent selection in pairs of populations with gene flow. Both methods are adequate for working with non-model species although the first requires an approximate knowledge of the phase of the SNPs under study. If the SNPs under evaluation are not linked then the second method should be applied. Our working definition for non-model species involves those for which we could have information about the haplotypic phase but we do not have
estimates of recombination rates neither we have information on potentially adaptive loci
nor we know the ancestral/derived status at each segregating site. This definition also
implies that we barely have good knowledge about the demography of the populations
under study and so, we cannot use a simulated neutral distribution to assess significance.

The first method combines haplotype-based information with a diversity-based $F_{ST}$ measure.
It is a sliding-window method that performs automatic decision-making to fix the adequate
window size. The second method is not haplotype-based and performs a two step $F_{ST}$ outlier
test. The first step of the algorithm consists in a heuristic search for different outlier clusters,
the second step is just a conditional LK test that will be performed only if more than one
cluster was found and in this case the test is applied through the cluster with the higher $F_{ST}$
values.

The design of the work is as follows: We begin developing the model for the haplotype-based
method and then, to deal with the case of fully unlinked SNPs, we build up the
algorithm for a, conservative, extreme outlier set test (EOS). In the results section we first
compare in the ROC (receiving operating characteristic) space, the two new methods with an
LD-based (Omegalus, Alachiotis, Stamatakis, and Pavlidis 2012) and with an outlier-based
method (Foll and Gaggiotti 2008). Finally, the EOS test is applied to a recently published real
data set and the results are discussed. We will show that our methods are able of detecting
divergent selection with migration while being robust to false positives.
The nvdFST model

In this section we improve a previous haplotype-based method for detecting divergent selection (Hussin et al. 2010; Rivas, Dominguez-Garcia, and Carvajal-Rodriguez 2015). The new statistic is called nvdFST because it combines a normalized variance difference, nvd, with the FST index. The nvd part performs a sliding-window approach to identify sites with specific selective patterns. When combined with the FST, it allows assessing significance of the candidate sites without the need of simulating neutral demography scenarios. Before developing the nvd formula, we review some concepts related with haplotype allelic classes.

Generalized HAC variance difference

A major-allele-reference haplotype (MARH) is an haplotype that carries only major frequency alleles of its constituting SNPs (Hussin et al. 2010). Then, for a given sample of haplotypes, we can define the mutational distance between any haplotype and MARH as the number of sites (SNPs) carrying a non-major (i.e. minor) allele. Each group of haplotypes having the same mutational distance will constitute a haplotype allelic class. Therefore (with some abuse of notation) we call HAC to the value of the mutational distance corresponding to each haplotype allelic class. That is, every haplotype having the same number of minor alleles belongs to the same HAC class.

Given the definitions above, consider a sample of n haplotypes of length L SNPs. For each evaluated SNP i (i\in[1,L]) we can perform a partition of the haplotypes (and their HAC classes) into \(P_1\), the subset carrying the most frequent (major) allele at the SNP i and \(P_2\) the
subset with the remaining haplotypes carrying the minor allele at $i$. That is, let '0' to be the major allele for the SNP $i$ and '1' for the minor. Then, $P_1$ includes every haplotype carrying the allele '0' for the SNP $i$ and $P_2$ the remaining haplotypes carrying '1' for that SNP. In $P_1$ we have different HAC values depending on the distance of each haplotype from MARH and similarly in $P_2$. Within each subset we can compute the variance of the HACs. Thus, in $P_1$ we have the variance $S_{1i}^2$ and correspondingly variance $S_{2i}^2$ in $P_2$ where $i$ refer to the SNP for which we have performed the partition.

The rationale of the HAC-based methods relies on the sweeping effect of the selectively favored alleles. Therefore, if the SNP $i$ is under ongoing selection then the variance in the partition 1 ($S_{1i}^2$) will tend to be zero because the allele at higher frequency (i.e. the allele of the SNP $i$ in the partition 1) is being favored and the sweeping effect will make the HAC values in this partition to be lower (because of sweeping of other major frequency alleles) consequently provoking lower variance values (Hussin et al. 2010). The variance in the second partition ($S_{2i}^2$) should not be affected by the sweeping effect because it does not carry the favored allele. So, the difference $S_{2i}^2 - S_{1i}^2$ would be highly positive in the presence of selection and not so otherwise. For a window size of $L$ SNPs, the variance difference between $P_2$ and $P_1$ can be computed to obtain a summary statistic called Svd (Hussin et al. 2010) that can be generalized to

$$gSvd_i = \frac{S_{2i}^2 - S_{1i}^2}{L} \times f_i (1 - f_i)^a \times b.$$  

Where $f_i$ is the frequency of the derived allele of the SNP $i$, and the parameters $b$ and $a$ permit to give different weights depending on if it is desired to detect higher frequencies ($a = 0$) or more intermediate ones ($a > 0$) of the derived allele. If $a = 0$ and $b = 1$ the statistic
corresponds to the original Svd and if $a = 1$ and $b = 4$ it corresponds to the variant called SvdM (Rivas, Dominguez-Garcia, and Carvajal-Rodriguez 2015). Note that when taking $a = 1$ it is not necessary to distinguish between ancestral and derived alleles because $f_i$ and $1 - f_i$ are interchangeable.

A drawback in the gSvd statistic is its dependence on the window size as has already been reported for the original Svd (Hussin et al. 2010; Rivas, Dominguez-Garcia, and Carvajal-Rodriguez 2015). Although gSvd is normalized by $L$, the effect of the window size on the computation of variances is quadratic (see Supplementary Appendix A-1 for details) which explains why the normalization is not effective in avoiding a systematic increase of the statistic under larger window sizes. This impact of the window size is important because the partitions $P_1$ and $P_2$ may experience different scaling effects, which would increase the noise in the estimation. The change in the scale due to the window size will also be dependent on the recombination and selection rates. Thus, it is desirable to develop a HAC-based statistic not dependent on the window size. In what follows, the between-partition variance difference is reworked in order to develop a new normalized HAC-based statistic, specially focused on detecting divergent selection in local adaptation scenarios with migration.

**Normalized variance difference (nvd)**

We have seen that for any haplotype sample of size $n$, we can compute the statistic $gSvd$ which basically is a difference between the HAC variances of the partitions $P_1$ and $P_2$. Remarkably, the corresponding HAC means and variances at each partition are related via the general mean and variance in that sample. Consider, for any candidate SNP, the mean
HAC distance, $m$, of the sample, and $m_1$ and $m_2$, the means corresponding to the partitions $P_1$ and $P_2$, respectively. We have the following relationships for the mean $m$ and sample variance $S^2$ (the subscripts 1 or 2 identify each partition; see the Appendix A-2 for details)

$$m = \frac{n_1 m_1 + n_2 m_2}{n}, \quad S^2 - \bar{S} = \frac{n}{n-1} \Delta \quad (1)$$

With $\bar{S} = \frac{(n_{11} - 1)S_{11}^2 + (n_{21} - 1)S_{21}^2}{n-1}$; $n_1$ and $n_2$ are the sample sizes at each partition ($n_1 \geq n_2$ by definition) and $\Delta = \frac{n_1 n_2}{n^2} (m_1 - m_2)^2$.

Using the relationships in (1), it is possible to compute the variance difference as appears in gSvd. However, we can substitute the parameters $b$ and $a$ by $a = 1$ and $b = 4$ as these are the values that permit to ignore the allelic state while maximizing the frequency product. In addition, it is also possible to consider the difference between means term (delta) in order to engage it in the detection of selection (see details in the Appendix). Thus, we finally obtain a new statistic for the variance difference of the candidate SNP $i$

$$vd_i = vd_{i0} + 4f_i(1-f_i)(S_{2i}^2 - S_{1i}^2) \quad (2)$$

where $vd_{i0} = \frac{n\Delta}{n_2-1} \times 4f_i(1-f_i)$

Therefore, the effect of selection upon $vd_i$ is two-fold. The first term of the sum in (2) corresponds to the effect of the difference between means and the second between variances. Clearly, increasing $S_{2i}$ or decreasing $S_{1i}$, as expected under selection, will increase the value of the statistic. If $S_{1i}$ and $S_{2i}$ are equal then the value of $vd_i$ is independent of the variances and just relies on the term $vd_{i0}$ corresponding to the partitions’ mean ($m_1$ and $m_2$).
and the candidate SNP frequencies. Please note that $vd_{10}$ is also the value of $vd_i$ when both variances are 0.

It is worth mentioning that, because the HAC values are bounded by 0 and $L$, the two parts of $vd_i$ are not independent. So that, if we have an extreme value for the HAC mean in the selective partition, say $m_1 = 0$, this implies that $S_1^2 = 0$ since every haplotype has to have a HAC of 0 to get that mean value. Note however that the opposite is not true, a value of $S_1^2 = 0$ does not imply necessarily that $m_1 = 0$.

Having said that, it can be shown (see Appendix A-2) that when the variance of the sample (without the partitions) is maximized, we get an upper bound ($d_{\text{max}}$) for $vd_i$

$$d_{\text{max}} = \frac{nL^2}{2(n-2)}$$

so that,

$$vd_i \leq d_{\text{max}} \quad (3)$$

If we divide (2) by $d_{\text{max}}$ we have a normalized variance difference

$$nv_{di} = \frac{vd_{10} + 4f_i(1-f_i)(S_1^2 - S_2^2)}{d_{\text{max}}} \quad (4)$$

The quantity from (4) can be computed for each SNP in a sample of sequences of any given length $L$ and the SNP yielding the maximum $nvd$ may be considered as a candidate for selection. Furthermore, in a two population setting, it is possible to compute (4) for each population or to combine the two populations in a unique sample. The latter is better for our purpose of looking for divergent selection in populations undergoing gene flow. When pooling both populations, the frequencies tend to be intermediate in the divergent selective sites. Therefore, we merge the shared SNPs from the two population samples and then we
compute the normalized variance difference using (4). Note however, that the reference haplotype (MARH) is defined just from one of the populations (by default the population 1).

Because (4) is normalized by $d_{\text{max}}$ the problem of scaling on the window size has been solved. However, the problem of choosing an optimal window size remains. A possible solution is to automate the choice by selecting the size which gives the maximum value for the statistic (Rivas, Dominguez-Garcia, and Carvajal-Rodriguez 2015). Therefore, for every SNP and window size, we consider the SNP having maximum $nvd$ as the candidate for divergent selection between the two populations.

At this point we already have a HAC-based statistic, $nvd$, that is independent of the window size and that should produce higher positive values for pairs of populations undergoing divergent selection. However, even if there is no selection, the maximum $nvd$ value could also be positive. Unfortunately, we ignore the distribution of the statistic and cannot decide if a given maximum is supporting the hypothesis of selection or not. As well, we might not have enough information on the species to simulate its evolution under a given neutral demography. Therefore, we still need to identify whether the value obtained for a given sample, is due to the effect of selection, especially because we desire to put great emphasis on avoiding false positives.

Consequently, we will perform two more measures before giving a diagnostic about the presence of divergent selection. The first is a sign test based on the lower bound of $nvd$, the second is the comparison between the $F_{ST}$ of the SNP having the maximum $nvd$ and the overall $F_{ST}$. 

249
Sign test

In our definition of \( nvd \) the term \( vd_i \) cannot be negative. For that reason, it could happen that with a negative difference in variances and a high mean HAC value in the first partition, as expected under neutrality, the value of \( nvd \) could still be positive. Therefore, we use a lower bound of \( nvd \) to derive the quantity called test selection sign (\( tss \), see Appendix A-3 for details) that would have negative value when the HAC values in the first partition are high.

\[
t_{ss} = \frac{4(n-1)s^2 - 2 \sum_{h} h_{ac}^2}{nL^2}
\]  

Where \( hac_{1h} \) are the HAC values measured at each haplotype \( h \) in the partition 1 and the sum is over the \( n_1 \) sequences in that partition. A negative sign in (5) suggests that the value of \( nvd \) is not the result of divergent selection. Indeed, we require (5) to be positive to count a given candidate as significant.

Now, even if we have a candidate position identified by its high \( nvd \) value and by the positive sign of \( tss \), we still lack a method for obtaining \( p \)-values associated to the sites chosen by the \( nvd \) algorithm. We can solve this problem if we combine the information on the selective candidate SNP, as given by \( nvd \), with the \( F_{ST} \) interpopulation differentiation index at that site. The joint use of these methods produces the combined measure \( nvdF_{ST} \).

Combined method: \( nvdF_{ST} \)

First, it is important to note that, when computing \( nvd \), we have considered only the SNPs shared between both populations in order to avoid low informative loci with high sampling
variance (Whitlock and Lotterhos 2015). Thus, we have an \( nvd \) value that may indicate the presence of divergent selection in a pair of populations connected by migration. The rationale of the \( nvdF_{ST} \) approach is that if divergent selection acts on a specific site then the \( F_{ST} \) at that site would be higher compared to the overall \( F_{ST} \). Then, we proceed as follows, let \( i \) be the candidate site chosen because it has the maximum \( nvd \) value, then we calculate the index \( I_i = F_{STi} - F_{ST} \) comparing the \( F_{ST} \) measured at the candidate site with the overall. The \( F_{ST} \) values were computed following the algorithm in Ferretti et al (Ferretti, Ramos-Onsins, and Pérez-Enciso 2013). To obtain the \( p \)-value we do not perform an LK test (Lewontin and Krakauer 1973) because first, the candidate was not chosen for being an outlier and second, we are considering linked rather than independent sites.

To get the \( p \)-value for a given index \( I_i \), the data is resampled several times (500 by default) to generate an empirical distribution. The expected frequency of each SNP is obtained as the mean frequency between populations as this is the expectation under the homogenizing effect of migration (Crow and Kimura 1970). Then, for each resampling iteration, the probability of a given allele at each population is obtained from a binomial \( B(p,n) \), where \( p \) is the mean allelic frequency at that site and \( n \) the local population sample size. The \( p \)-values correspond to the proportion of times that the resampled indexes were larger than \( I_i \).

For candidates with similar frequencies at both populations we expect low index \( I_i \) and correspondingly high \( p \)-values. When the pooled frequency is intermediate; two situations are possible, first, each population has similar intermediate frequencies which again imply high \( p \)-values; or alternatively, the frequencies can be extreme and opposite at each population. In the latter, \( I_i \) is high and its \( p \)-value low. Note that, for each site, the resampling
procedure has variance $p(1-p)n$ which is larger at intermediate pooled frequency values. Thus, the method has the desired property to be more conservative at intermediate pooled frequencies which minimize the possibility of false positives.

**Effective number of independent SNPs, significance and $q$-value estimation**

We have computed $nvd$ and the $F_{ST}$ index and got a candidate site with its $p$-value. Since $nvd$ was obtained after testing a number of positions on a given window size, it is desirable to apply a multiple test correction for the number of independent SNPs in the window. To roughly estimate the number of independent SNPs, we calculate the linkage disequilibrium measure $D'$ (Lewontin 1988; Devlin and Risch 1995) at each pair of consecutive sites and then store the quantity $r' = 1 - |D'|$ for each pair. The effective number of independent SNPs ($M_{effs}$) between site $w_{ini}$ and $w_{end}$ is then obtained as one plus the summation of the $r'$ values in the interval $[w_{ini}, w_{end}]$. The Šidák correction (Sidák 1967; Cheverud 2001) can now be applied to get the adjusted significance level $c = 1 - (1 - \gamma)^{1/M_{effs}}$ with nominal $\gamma (= 0.05$ by default). Thus, the algorithm $nvdF_{ST}$ would finally suggest a candidate as significant only when the sign as computed in (5) was positive and the $p$-value (as obtained in the previous section) is lower than $c$.

The $q$-values (Storey 2003) can be seen as a multiple testing analogs of $p$-values. They have been proposed as an useful approach for evaluating method performance in terms of false discoveries (De Villemereuil et al. 2014). Accordingly, we estimate the $q$-value (see Appendix A-4 for details on the calculation) corresponding to each significant $p$-value.
The extreme outlier set test (EOS)

The $nvdF_{ST}$ method assumes the existence of a dense map of linked genetic markers. If the data consists mostly in independent markers this would provoke the failure to detect selection because the HAC-based information does not exist. To deal with this situation, a second method was implemented consisting in a two-step heuristic procedure that performs a conservative test for identifying extreme outliers.

As already mentioned, the variance of the $F_{ST}$ distribution is quite unpredictable under a variety of scenarios. This provokes high rates of false positives associated with the $F_{ST}$ outlier tests. Our heuristic strategy takes advantage of the fact that, independently of the demographic scenario, the involved regions under divergent selection may produce extreme outliers that would be clustered apart from the neutral ones. The subsequent LK test is performed only when this kind of outliers is detected. As $F_{ST}$ estimator we use $G_{ST}$ (Nei 1973).

The rationale of the algorithm is as follows. The first step consists in computing the extreme positive outliers in the sense of Tukey i.e. those sites having a $F_{ST}$ value higher than 3 times the interquartile range (Tukey 1977). The second step identifies different classes inside the extreme outlier set. This is done by a $k$-means algorithm (Vattani 2011; Schubert, Zimek, and Kriegel 2012). Here, a $k$-modal distribution is assumed and all the elements in the outlier set are classified in one of the $k$ classes. The class with lower values is discarded and only the elements, if any, in the upper classes having values higher than a cutoff point are maintained in the set. By default $k = 2$ and two modes $\{0, F_{STu}\}$ were used corresponding to lower (0) and
upper ($F_{STu}$) bounds for the $F_{ST}$ estimator (see Appendix A-5). The cutoff point is defined as

the overall $F_{ST} + F_{STu} / 3$, i.e. the mean plus the square root of the upper-bound for the $F_{ST}$ variance under an asymmetric unimodal distribution (Dharmadhikari and Joag-Dev 1989).

Finally, for each of the candidates remaining in the EOS after the cutoff, the LK test (Lewontin and Krakauer 1973) is performed to compute its $p$-value. The Šidák correction (Sidak 1967; Cheverud 2001) for the number of remaining outliers in the set is applied to get

the significance level.

Software description

Both $nvdF_{ST}$ and the EOS test have been implemented in the program HacDivSel. Complete details of the software can be found in the accompanying manual. We here just mention that the input program files may be in MS (Hudson 2002) or Fasta formats for the haplotype-based test or in Genepop (Rousset 2008) or BayeScan (Foll and Gaggiotti 2008) formats if the data do not include haplotype information. The latter two allow variable sample size. In any case, the data should contain sequence samples from two populations. A typical command line for calling the program in order to analyze a file named $sel.txt$ containing 50 sequences from each population, would be

```
HacDivSel -input sel.txt -sample 50 -candidates 10 -SL 0.05 -output anyname -format ms
```

Where the label, -candidates 10, indicates that the ten highest $nvd$ values should be included in the output. The program would analyze the file and produce as output the highest 10 values and its significance at the 0.05 level for different window sizes after the $nvdF_{ST}$ test. It
also performs the EOS test and gives the candidate outliers, if any, and their significance.

Only the SNPs shared by the two populations are considered. Which imply that there are at least 4 copies of each SNP in the metapopulation.

Simulations

The simulation setting consists in a two population scenario under divergent selection connected by migration. There are several examples of adaptation to divergent environments connected by migration such as the intertidal marine snail *L. saxatilis* (Rolan-Alvarez 2007), some wild populations of *S. salar* (Bourret et al. 2013), lake whitefish species (Renaut et al. 2011) and so on. To perform simulations as realistic as possible, we use some relevant demographic information from *L. saxatilis*, such as migration rates and population size as estimated from field data (Rolan-Alvarez 2007). Concerning selection intensities, we considered moderate selection pressures and few loci with large effects (Thibert-Plante and Gavrilets 2013). Therefore, a model resembling the most favorable conditions for the formation of ecotypes under local adaptation with migration was implemented.

The selective scenario (\(\alpha = 4Ns\)) is divergent so that the allele favored in one population is the deleterious in the other. This simulation model involves low and high mutation rate (\(\theta=4N\mu\)) and different recombination rates (\(p = 4Nr \in \{0, 4, 12, 60, \text{unlinked}\}\)) and extends previous work (Rivas, Dominguez-Garcia, and Carvajal-Rodriguez 2015) by adding new parameter values and demographic scenarios. The whole setting is fully explained in the Supplementary Appendix (Appendix A-6). The simulations were performed using the last version of the program GenomePop2 (Carvajal-Rodriguez 2008).
The simulated data is also utilized for the comparison of $\nu d F_{ST}$ with OmegaPlus (Alachiotis, Stamatakis, and Pavlidis 2012) and of the EOS test with BayeScan 2.1 (Foll and Gaggiotti 2008). We chose OmegaPlus because of its good performance, compared with other haplotype-based methods, under non-equilibrium demographic conditions (Crisci et al. 2013). Several combinations of the parameters for OmegaPlus were tested in (Rivas, Dominguez-Garcia, and Carvajal-Rodriguez 2015), here we selected the ones with the best performance. BayeScan was chosen because it is one of the main state-of-the-art outlier-based programs. The parameters for BayeScan were the default ones as this is a conservative setting and we were interested in comparing the false positive rates. Only SNPs shared between populations and with a minimum allele frequency (maf) of 2 per population (4%) were considered.

**Results**

In what follows we present the results after applying the tests to simulated and real data. The power of a test (true positive rate) is measured as the % of runs in which selection was detected from simulated selective scenarios and the false positive rate (FPR) is measured as the % of runs in which selection was detected from simulated neutral scenarios. The $q$-value (Storey 2003) is an estimate from the results (see Appendix A-4).

**ROC curves**

The ROC curve is a useful way to explore the performance of any detection statistic because it plots, under different conditions, the true positive rate (TPR or power, y-axis) against the
false positive rate (FPR, x-axis). Obviously, the best detection method would have high power and low FPR, which corresponds to the points in the upper-left corner of the ROC space. On the contrary, values close to the diagonal or no-discrimination line (NDL) would be considered as poor performing methods. Recall that we have developed two methods especially intended to be robust to false positives so we expect them to occupy at least the left side of the ROC.

In the ROC space of the Figure 1 we can appreciate the performance of nvdFST, EOS, OmegaPlus and BayesCan methods through the $\alpha = 600$ and $Nm=10$, simulated scenarios. These scenarios include different mutation and recombination rates (see also Table 1 and Appendix A-6). Detailed results for each method and simulation setting are given in the next sections but the main general picture can already be obtained from this ROC curve.
Figure 1. ROC curves for \(nvdF_{ST}\), EOS, OmegaPlus (Omega) and BayeScan (Bayes factor 3: BSBF3 and Bayes factor 100: BSBF100). FPR: False Positive Rate.

The method \(nvdF_{ST}\) is the best performing one since it occupies the upper-left corner.

However, the scenario with independent (unlinked) markers is not visible in the \(nvdF_{ST}\) plot since it has a (0, 0) coordinate (see Table 1 and next section below). For the \(nvdF_{ST}\) plotted scenarios, the power ranges between 80-90% (y-axis) and FPR (x-axis) is below 5%. On the contrary, OmegaPlus is too close to the NDL diagonal under the same divergent selection scenarios.

The BayeScan and EOS methods are plotted only for the unlinked and the weak-linkage (\(\rho = 60; 1.5\ \text{cM/Mb}\)) marker scenarios. The method BayeScan with a Bayes factor above 3
(BSBF3) is positioned too much at the right side which indicates too high false positive rates for both the unlinked (upper point in BSBF3 line) and the weak-linkage (lower point in BSBF3 line) scenarios. However, with a Bayes factor of 100, BayeScan works very well for the unlinked scenario (upper point in BSBF100 line) but fails when markers are even slightly linked (the lower point in BSBF100 line). The EOS test performance for the slightly linked and unlinked scenarios is not bad in terms of power, and very good in terms of FPR since it occupies the left side of the plot.

In the following sections we detail these and other results for \textit{nvdF}_{ST} and EOS under the different simulated scenarios.

**Combined Method (nvdF\textsubscript{ST})**

Under a single locus architecture with selection $\alpha = 4Ns = 600$ and migration $Nm=10$, the power of \textit{nvdF}_{ST} vary between 79-94% for both medium (60 SNPs/Mb) and high density (250 SNPs/Mb) maps (Table 1 and Figure 1). These results can be compared with published analysis (Rivas, Dominguez-Garcia, and Carvajal-Rodriguez 2015) with the methods Svd, SvdM and OmegaPlus (Alachiotis, Stamatakis, and Pavlidis 2012) for which similar best results were obtained by Svd and SvdM for the same cases with high mutation and recombination (Rivas, Dominguez-Garcia, and Carvajal-Rodriguez 2015). However, recall that the methods Svd, SvdM and OmegaPlus oblige the user to perform simulations of a neutral demography to obtain the $p$-values for the tests, and consequently, the results in the Rivas and coworkers study, were obtained having the exact neutral demography at hand. As it can be appreciated from rows 1 to 6 in Table 1 —that matches the scenarios in (Rivas,
Dominguez-Garcia, and Carvajal-Rodriguez 2015)—the nvdFST performs well without the need of performing additional neutral simulations. Also, the false positive rate and the q-value are low in all the scenarios. The given results are for 10,000 generations; the cases with 5,000 generations were similar.

Under the polygenic architecture (n = 5 in Table 1) at least one candidate is found 99% of the times and more than one, 80% of the time. However, the number of correctly identified sites is quite variable ranging between 1 and 3.

The last row in Table 1 corresponds to the case when all SNPs segregate independently. In this case, the method fails to detect selection which is not surprising because the information from the haplotype allelic classes is absent under linkage equilibrium; the adequate patterns are not found which provokes both a negative in the sign test and a candidate with low FST index measure.
Table 1. Performance of the combined method ($nvdF_{ST}$) with $n = 1$ selective site located at the center of the chromosome or $n = 5$ (see Appendix A-6). Selection was $\alpha = 4N_s = 600$ and migration $Nm = 10$. Mean localization is given in distance kb from the real selective position.

<table>
<thead>
<tr>
<th>$\Sigma$</th>
<th>$\theta$</th>
<th>$\rho$</th>
<th>$n$</th>
<th>%Power</th>
<th>%FPR ($\gamma = 5%$)</th>
<th>$q$-value</th>
<th>Localization (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>65</td>
<td>12</td>
<td>0</td>
<td>1</td>
<td>87</td>
<td>2.1</td>
<td>0.0058</td>
<td>±458</td>
</tr>
<tr>
<td>63</td>
<td>12</td>
<td>4</td>
<td>1</td>
<td>94</td>
<td>2.7</td>
<td>0.0008</td>
<td>±200</td>
</tr>
<tr>
<td>60</td>
<td>12</td>
<td>12</td>
<td>1</td>
<td>90</td>
<td>1.0</td>
<td>0.0003</td>
<td>±33</td>
</tr>
<tr>
<td>251</td>
<td>60</td>
<td>0</td>
<td>1</td>
<td>79</td>
<td>1.8</td>
<td>0.0048</td>
<td>±60</td>
</tr>
<tr>
<td>232</td>
<td>60</td>
<td>4</td>
<td>1</td>
<td>84</td>
<td>6.2</td>
<td>0.0011</td>
<td>±17</td>
</tr>
<tr>
<td>249</td>
<td>60</td>
<td>60</td>
<td>1</td>
<td>86</td>
<td>2.4</td>
<td>0.0002</td>
<td>&lt;±1</td>
</tr>
<tr>
<td>282</td>
<td>60</td>
<td>60</td>
<td>5</td>
<td>99</td>
<td>2.4</td>
<td>0.0002</td>
<td>&lt;±1</td>
</tr>
<tr>
<td>318</td>
<td>60</td>
<td>$\infty$</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

$\Sigma$: Mean number of shared SNPs per Mb. $\theta$: Mutation rate. $\rho$: Recombination rate. FPR: false positive rate. $q$-value: mean estimated $q$-value for the significant tests. $\infty$: Independently segregating sites.

**Short-term Strong Selection and Long-term Weak Selection Scenarios**

The performance of $nvdF_{ST}$ under the strong selection scenario ($\alpha = 6000$) in the short-term (500 generations) varies between 44% for fully linked to 67% for weak linked markers (Table 2). Not surprisingly, the number of segregating sites is considerably reduced. In fact the
minimum window size allowed by the program had to be shortened from 51 to 25 to perform the analyses. Notably the false positive rate (FPR) was 0.

Concerning weak selection in long-term scenarios (Table 2, $\alpha = 140$) the power varies between 49-52% with false positive rate between 2.2 and 5.7%.

Table 2. Performance of the combined method ($nvdF_{ST}$) with a single selective site in the short-term strong ($\alpha = 6000$) and the long-term weak ($\alpha = 140$) selection scenarios. $Nm$ was 10. Mean localization is given in distance kb from the real selective position.

<table>
<thead>
<tr>
<th>$\Sigma$</th>
<th>$\theta$</th>
<th>$\rho$</th>
<th>$\alpha$</th>
<th>$t$</th>
<th>%Power</th>
<th>%FPR ($\gamma = 5%$)</th>
<th>$q$-value</th>
<th>Localization (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>112</td>
<td>60</td>
<td>0</td>
<td>6000</td>
<td>500</td>
<td>44</td>
<td>0</td>
<td>0</td>
<td>$\pm66$</td>
</tr>
<tr>
<td>32*</td>
<td>60</td>
<td>4</td>
<td>6000</td>
<td>500</td>
<td>63</td>
<td>0</td>
<td>0.0014</td>
<td>$\pm5$</td>
</tr>
<tr>
<td>62</td>
<td>60</td>
<td>60</td>
<td>6000</td>
<td>500</td>
<td>67</td>
<td>0</td>
<td>0.0008</td>
<td>$\pm93$</td>
</tr>
<tr>
<td>165</td>
<td>60</td>
<td>0</td>
<td>140</td>
<td>5,000</td>
<td>49</td>
<td>3.6</td>
<td>0.0280</td>
<td>$\pm33$</td>
</tr>
<tr>
<td>156</td>
<td>60</td>
<td>4</td>
<td>140</td>
<td>5,000</td>
<td>52</td>
<td>5.7</td>
<td>0.0219</td>
<td>$\pm14$</td>
</tr>
<tr>
<td>135</td>
<td>60</td>
<td>60</td>
<td>140</td>
<td>5,000</td>
<td>49</td>
<td>2.2</td>
<td>0.0054</td>
<td>$\pm6$</td>
</tr>
</tbody>
</table>

$\Sigma$: mean number of shared SNPs per Mb. $\theta$: Mutation rate. $\rho$: recombination rate. $t$: number of generations. FPR: false positive rate. $q$-value: mean estimated $q$-value for the significant tests. *: only 40 runs having a minimum of 25 SNPs.
Extreme Outlier Set Test (EOS)

We applied the EOS test under the single locus architecture with selection $\alpha = 600$ and migration $Nm=10$. As desired, the test is very conservative with false positive rates below the nominal 0.05 in every case (Table 3). Not surprisingly for an outlier-based method, the test has no power if the markers are strongly linked ($\rho$ from 0 to 12) or under a polygenic setting (row with $n = 5$ in Table 3). However, in the case of independent SNPs but also in the case of maps with 250-300 SNPs/Mb under 1.5 cM/Mb the power rises up to 60%. Therefore, the EOS test is complementary to $nvdF_{ST}$ having its maximum power when the latter has its minimum.
Table 3. Performance of the extreme outlier test (EOS) with \( n = 1 \) selective site located at the center of the chromosome or \( n = 5 \) (see Simulations section above). Selection was \( \alpha = 600 \) and \( Nm = 10 \). Mean localization is given in distance kb from the real selective position.

<table>
<thead>
<tr>
<th>( \Sigma )</th>
<th>( \theta )</th>
<th>( \rho )</th>
<th>( n )</th>
<th>%Power EOS</th>
<th>%FPR (( \gamma = 5% ))</th>
<th>( q' )-value</th>
<th>Localization (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>65</td>
<td>12</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>63</td>
<td>12</td>
<td>4</td>
<td>1</td>
<td>0.2</td>
<td>0</td>
<td>0.46</td>
<td>±3</td>
</tr>
<tr>
<td>60</td>
<td>12</td>
<td>12</td>
<td>1</td>
<td>1.1</td>
<td>0</td>
<td>0.45</td>
<td>±77</td>
</tr>
<tr>
<td>251</td>
<td>60</td>
<td>0</td>
<td>1</td>
<td>0.7</td>
<td>0</td>
<td>0.10</td>
<td>0</td>
</tr>
<tr>
<td>232</td>
<td>60</td>
<td>4</td>
<td>1</td>
<td>1.3</td>
<td>0</td>
<td>0.20</td>
<td>±150</td>
</tr>
<tr>
<td>249</td>
<td>60</td>
<td>60</td>
<td>1</td>
<td>58</td>
<td>0.4</td>
<td>0.5</td>
<td>&lt;±1</td>
</tr>
<tr>
<td>282</td>
<td>60</td>
<td>60</td>
<td>5</td>
<td>1.6</td>
<td>0.4</td>
<td>0.49</td>
<td>±5</td>
</tr>
<tr>
<td>318</td>
<td>60</td>
<td>( \infty )</td>
<td>1</td>
<td>61</td>
<td>1.2</td>
<td>3\times10^{-6}</td>
<td>±7</td>
</tr>
</tbody>
</table>

\( \Sigma \): mean number of shared SNPs per Mb. \( \theta \): mutation rate. \( \rho \): recombination rate. FPR: false positive rate. \( q' \)-value: mean corrected (see appendix A-4) estimated \( q \)-value in the significant tests. \( \infty \): independently segregating sites.

Concerning false positives, note —in the last three rows of Table 3— the low false positive rate (FPR) values that are indicating the low percentage of outliers detected as selective in the corresponding neutral scenario (1.2% for independent and 0.4% for linked markers). Thus, the test worked correctly by avoiding false selective sites under the neutral setting. However, the \( q \)-value estimates varied a lot depending on the linkage between markers.
Recall that by $q$-value, we mean the minimum estimated false discovery rate (FDR) that can
be committed when calling significant one test at a given threshold. It can be appreciated
that the $q$-value is very low ($3 \times 10^{-6}$) for independently segregating sites but rise up to 0.5 for
the same scenario when markers are linked. The explanation is that in the case of unlinked
markers the EOS test is quite efficient and just detects the single true selective SNP. On the
contrary, with linked markers, more outliers are detected, which, since there is only one true
selective site, inflates the FDR and the corresponding $q$-values.

**Position Effect**

The ability to locate the position of the selective site increased with the marker density and
the recombination rate (Table 1). The localization is given in kilobases away from the correct
position. The values are averages through the runs. Standard errors are omitted since they
were low (in the order of hundreds of bases or below 5 kb in the worst case of fully linked
markers). The $nvdF_{ST}$ method performs acceptably well when the target site is located at the
centre of the studied region, the selection is not too strong ($\alpha \leq 600$) and the overall
recombination rate is at least 0.3 cM/Mb ($\rho \geq 12$). In these cases, the selective location is
estimated, at worst, within 33 kb of distance from the true location (Table 1). In the case
with strong selection, the localization is bad still under high recombination (Table 2, $\alpha =
6000$). However, this could be due to the lower number of segregating sites (only 62 in Table
2).

Importantly, the localization is also dependent on where the selective site is placed within
the chromosome. The farther from the center the worse the ability to correctly localize the
selective positions (Table 4). In this case, with recombination of 1.5 cM/Mb, the inferred
location changes from an almost perfect localization (<1 kb from Table 1) to distances of 10-
122 kb as the target is shifted away from the center. This issue has already been shown for
other HAC-based methods (Rivas, Dominguez-Garcia, and Carvajal-Rodriguez 2015).

The problem is partially solved, when the recombination is high (1.5 cM/Mb), by using the
EOS test. In such case, the test has high power (67-93%) and good localization of the
selective position. In fact, the position of the selective site is almost perfectly estimated (few
bases or kb) when the true position is not at the extremes. Even if the target sites are at the
extremes the localization is within 40 kb (see cases with $\rho = 60$ in Table 4).

In the case of independent markers with the selective site located at the center (Table 3) the
localization by EOS was perfect in 98% of the replicates. However, in the table the average
appears as ±7 kb because of two runs in which the localization failed by almost 400 kb. These
two runs correspond to lower $F_{ST}$ index values that were marginally considered given the
cutoff for selecting the SNPs. Thus, we could count this two as false positives or
alternatively, by making a bit more astringent the EOS classification in the upper class, we
would have already discarded them. For example, if we change the cutoff from $F_{ST} + F_{STu} / 3$
to $F_{ST} + 1.2F_{STu} / 3$ we decrease the power only from 61 to 59% while getting perfect
localization of the selective SNPs in every run.
Table 4. Performance of \( nvdF_{ST} \) and EOS with a single selective site located at different positions. Selection was \( \alpha = 600 \) and \( Nm = 10 \). Mean localization is given in distance kb from the real selective position. FPRs are the same as in Table 1. \( q \)-value refers to the mean \( q \)-value for the significant \( nvdF_{ST} \) tests.

<table>
<thead>
<tr>
<th>( \Sigma )</th>
<th>( \theta )</th>
<th>( \rho )</th>
<th>%Power</th>
<th>Position (kb)</th>
<th>( nvdF_{ST} ) ( q )-value</th>
<th>Localization (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>259</td>
<td>60</td>
<td>0</td>
<td>81, 1</td>
<td>0</td>
<td>0.0044</td>
<td>+483, +457</td>
</tr>
<tr>
<td>255</td>
<td>60</td>
<td>0</td>
<td>81, 1.5</td>
<td>10</td>
<td>0.0049</td>
<td>+433, +496</td>
</tr>
<tr>
<td>256</td>
<td>60</td>
<td>0</td>
<td>82, 0.9</td>
<td>100</td>
<td>0.0041</td>
<td>+350, +413</td>
</tr>
<tr>
<td>255</td>
<td>60</td>
<td>0</td>
<td>78, 0.6</td>
<td>250</td>
<td>0.0039</td>
<td>±194, ±185</td>
</tr>
<tr>
<td>230</td>
<td>60</td>
<td>4</td>
<td>75, 2.5</td>
<td>0</td>
<td>0.0014</td>
<td>+324, +127</td>
</tr>
<tr>
<td>226</td>
<td>60</td>
<td>4</td>
<td>77, 3.5</td>
<td>10</td>
<td>0.0016</td>
<td>+326, +142</td>
</tr>
<tr>
<td>233</td>
<td>60</td>
<td>4</td>
<td>80, 1.8</td>
<td>100</td>
<td>0.0017</td>
<td>+227, +140</td>
</tr>
<tr>
<td>229</td>
<td>60</td>
<td>4</td>
<td>83, 1.6</td>
<td>250</td>
<td>0.0009</td>
<td>±123, ±20</td>
</tr>
<tr>
<td>262</td>
<td>60</td>
<td>60</td>
<td>63, 93</td>
<td>0</td>
<td>0.0014</td>
<td>+122, +40</td>
</tr>
<tr>
<td>261</td>
<td>60</td>
<td>60</td>
<td>68, 91</td>
<td>10</td>
<td>0.0014</td>
<td>+113, +34</td>
</tr>
<tr>
<td>257</td>
<td>60</td>
<td>60</td>
<td>81, 84</td>
<td>100</td>
<td>0.0006</td>
<td>±44, ±6</td>
</tr>
<tr>
<td>252</td>
<td>60</td>
<td>60</td>
<td>87, 67</td>
<td>250</td>
<td>0.0004</td>
<td>±10, ±0.06</td>
</tr>
</tbody>
</table>

\( \Sigma \): Mean number of shared SNPs per Mb. \( \theta \): Mutation rate. \( \rho \): Recombination rate. Position: real position of the selective site.
Bottleneck-expansion Scenarios

Bottleneck-expansion scenarios are known to leave signatures that mimic the effect of positive selection. We tested the robustness of the tests by looking for false positives when applying the methods to a bottleneck and expansion scenario under a neutral setting. The bottleneck was simulated by a reduction of one of the populations to 1% of the original size ($N$ from 1000 to 10). Afterwards, the population expansion was implemented by increasing the population size following a logistic growth model (see details in Appendix A-6). The methods performed well, the false positive rate is maintained below the nominal level with 4.6% and 1% for $nvdf_{ST}$ and EOS tests, respectively.

High Migration Scenario

For the short-term (500 generations) scenario with $Nm = 50$ (5%), $nvdf_{ST}$ is still able to detect the effect of selection in spite of the homogenizing effect of migration. The detection power ranges between 34-59% with a false positive rate of 0-0.1% (Table 5). Concerning the EOS test, it works only under weak linkage and with a power of 16%. EOS committed no false positives under this setting.
Table 5. Performance of nvdFST in the short term (500 generations) with a single selective site. Selection was $\alpha = 600$ and $Nm = 50$. Mean localization is given in distance kb from the real selective position.

<table>
<thead>
<tr>
<th>$\sum$</th>
<th>$\theta$</th>
<th>$\rho$</th>
<th>%Power</th>
<th>%FPR ($\gamma = 5%$)</th>
<th>q-value</th>
<th>Localization (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>116</td>
<td>60</td>
<td>0</td>
<td>56</td>
<td>0</td>
<td>0.0098</td>
<td>$\pm$152</td>
</tr>
<tr>
<td>180</td>
<td>60</td>
<td>4</td>
<td>59</td>
<td>0</td>
<td>0.0042</td>
<td>$\pm$123</td>
</tr>
<tr>
<td>178</td>
<td>60</td>
<td>60</td>
<td>34</td>
<td>0.1</td>
<td>0.0096</td>
<td>$\pm$4</td>
</tr>
</tbody>
</table>

$\sum$: Mean number of shared SNPs per Mb. $\theta$: Mutation rate. $\rho$: Recombination rate. FPR: false positive rate. q-value: mean estimated $q$-value for the significant tests.

Empirical Data

We applied the EOS test to analyze a recently published data set from Littorina saxatilis species, concretely the separate-island filtered loci from Ravinet et al. (2016). We have discarded the loci with null allele frequency equal or higher than 0.5. We also discarded those polymorphisms not shared between ecotypes from the same location. Additionally, we required a minimum frequency allele of 4 per metapopulation sample size. Thus, we have excluded about 10-20% of the original individual-island filtered loci. The results of the between ecotypes outlier analysis using EOS are shown in Table 6. We can appreciate that the number of outliers detected as significant is much less than in the original study since we find a total of 69 outliers in the three islands while the number originally found was 406 (RAD loci in Table 2 of Ravinet et al. 2016). This is not surprising given the conservative
nature and low false positive rate of EOS. However, note that we find a 2.9% (2/69) of SNPs
shared by all islands which is quite similar to the 2.2% (9/406) found in the original study.
Considering the islands by pairs, Jutholmen and Ramsö share 2 outliers while Saltö has no
outlier in common with Jutholmen and just 1 with Ramsö.

Table 6. Outliers detected after EOS analysis of the individual-island filtered loci from
*Littorina saxatilis* data. Numbers in parenthesis refer to the results in the original analysis
(Ravinet et al. 2016).

<table>
<thead>
<tr>
<th>Island</th>
<th>Unique</th>
<th>Only with Jutholmen</th>
<th>Only with Ramsö</th>
<th>Only with Saltö</th>
<th>Shared all</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jutholmen</td>
<td>27 (59)</td>
<td>__</td>
<td>2 (13)</td>
<td>0 (16)</td>
<td>2 (9)</td>
<td>31 (97)</td>
</tr>
<tr>
<td>Ramsö</td>
<td>24 (86)</td>
<td>2 (13)</td>
<td>__</td>
<td>1 (21)</td>
<td>2 (9)</td>
<td>29 (129)</td>
</tr>
<tr>
<td>Saltö</td>
<td>6 (134)</td>
<td>0 (16)</td>
<td>1 (21)</td>
<td>__</td>
<td>2 (9)</td>
<td>9 (180)</td>
</tr>
</tbody>
</table>

For the outliers in EOS, the $F_{ST}$ between ecotypes ranges between 0.4-0.6 (Table 7). The $q$-
values are high (0.52 - 0.76) although we already know by the simulations that this may be
indicating linkage between the markers more than an inflated false positive rate (see also De
Villemereuil et al. 2014).
Table 7. Summary of EOS analysis for the between ecotypes *Littorina saxatilis* data (Ravinet *et al.* 2016).

<table>
<thead>
<tr>
<th>Island</th>
<th>Nonoutliers</th>
<th>Outliers not in EOS</th>
<th>EOS</th>
<th>$F_{ST}$</th>
<th>$F_{ST, EOS}$</th>
<th>$pval_{EOS}$</th>
<th>$qval_{EOS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jutholmen</td>
<td>4564</td>
<td>91</td>
<td>31</td>
<td>0.045</td>
<td>0.40</td>
<td>0.004</td>
<td>0.52</td>
</tr>
<tr>
<td>Ramsö</td>
<td>4602</td>
<td>82</td>
<td>29</td>
<td>0.064</td>
<td>0.53</td>
<td>0.005</td>
<td>0.63</td>
</tr>
<tr>
<td>Saltö</td>
<td>4632</td>
<td>51</td>
<td>9</td>
<td>0.060</td>
<td>0.60</td>
<td>0.002</td>
<td>0.76</td>
</tr>
</tbody>
</table>

$F_{ST}$: Mean $F_{ST}$ for the analyzed loci. $F_{ST, EOS}$: Mean $F_{ST}$ for the loci included in the extreme outlier set. $pval_{EOS}$: Mean $p$-values across the loci included in the extreme outlier set. $qval_{EOS}$: Mean $q$-values across the loci included in the extreme outlier set.

Discussion

The goal in this work was to develop two methods, haplotype-based and outlier-based, for the detection of divergent selection in pairs of populations connected by migration. We also intended that the methods be robust to false positives. High rate of false positives is a known concern of outlier-based methods (De Mita *et al.* 2013; De Villemereuil *et al.* 2014; Lotterhos and Whitlock 2014) thus, EOS was especially designed to minimize the false positive rate. Additionally, both methods should be useful for non-model species and so, it should not be necessary to perform neutral simulations to obtain critical cut-off values for the candidates.

For the $nvdf_{ST}$ method, it has been shown that combining haplotype-based and $F_{ST}$ differentiation information is a quite powerful strategy for detecting divergent selection. However, the $nvdf_{ST}$ algorithm does not perform well when the whole set of markers is segregating independently. To deal with this latter case, a second method was proposed...
based on the idea that the outliers caused by the effect of divergent selection would cluster apart from those caused by different demography issues. This extreme outlier set test, EOS, was intended to be conservative because the aforementioned tendency of outlier-based methods to produce false positives. Under the simulated scenarios, the EOS test behaves acceptably well when markers are independent or under weak linkage, reaching powers between 60-90% while maintaining false positive rate below the nominal level.

**Polygenic Architecture**

In general, the $F_{ST}$-based methods cannot detect selection in polygenic scenarios (Bierne, Roze, and Welch 2013; De Villemereuil et al. 2014) because those tests are specifically designed for finding larger than average $F_{ST}$ values which are difficult to discover if the frequency differences are slight for the polygenic loci or if the overall $F_{ST}$ is high. On the contrary, the $nvdF_{ST}$ performs even better in this scenario. The explanation for this good performance is that the distributed selective signal facilitates the discovery of the corresponding patterns of divergent selection. These patterns imply high frequency at the target site in one population and low in another. Therefore, the $F_{ST}$ index under the neutral expectation would be very low compared with the observed one and so, still under high overall $F_{ST}$, the test $nvdF_{ST}$ maintains high power under the polygenic setting.
Position Effect

Besides the detection of the signal of selection, we have also inferred the location of the selective site. It has been shown that under $nvdf_{ST}$ the localization is better when the selective site is at the center of the chromosome. The EOS test is not so affected by the position of the selective site. The ability of localizing the selective position is still a pending issue for many of the selection detection methods. There is also plenty of room for improvement under the $nvdf_{ST}$ and EOS methods in this regard. Trying, for example, to further explore the relationship between recombination and the window sizes that yield the highest scores. Indeed, the interplay among divergent selection, recombination, drift and migration should be considered for further improving the efficiency of the methods.

High Migration Scenario

Noteworthy, under high migration, $nvdf_{ST}$ maintains reasonable power (60%). However, the power diminishes with the highest recombination rate. This may occur due to the combined effect of gene flow and recombination that generates intermediate HAC mean values $m_1$ and $m_2$ and similar variances. Indeed, for a given selection intensity, the higher the $Nm$ requires tighter linkage for the establishment of divergent alleles (Yeaman and Whitlock 2011).

In the case of the EOS test, there is an obvious tradeoff between the astringency of the cutoff point for the outlier set and the migration rate. Our cutoff depends on the $F_{ST}$ upper-bound which is a function of the number of populations, the sample sizes and the minimum allele frequency. However, with higher migration the upper bound of the $F_{ST}$ should be lower.
because as migration increases the number of populations becomes virtually one. Therefore, a possible solution to improve the efficiency of EOS under high migration would be to update the $F_{ST}$ upper bound as a function of the migration rate.

**Empirical Data**

Local adaptation may occur, most likely due to alleles with large effect but also under a polygenic architecture (Whitlock 2015; Yeaman 2015). In addition, the geographic structure and the migration-selection balance can generate complex patterns of the distribution of genetic variation (Debarre, Yeaman, and Guillaume 2015). Thus, the natural systems where local adaptation occurs can be of great complexity (Whitlock 2015). The *L. saxatilis* ecotypes are an especially interesting system to study local adaptation in presence of gene flow (Johannesson 2015). This system has an exceptional level of replication at different extent, as country, localities within country, and the micro-geographical level of the ecotypes. In the case of the Swedish populations, the pattern of differentiation can be separated in factors such as, localities and habitat variation among islands —that may be caused by genetic drift— and variation between habitats within localities, that may be caused by divergent selection (Johannesson 2015). There are also different mechanisms by which parallel adaptation may occur, resulting in different predictions about the proportion of shared adaptive variation among localities.

Regarding the shared genomic divergence of the *L. saxatilis* system in Swedish populations, it seems to be a small proportion of the total genomic divergence (Hollander, Galindo, and Butlin 2015; Johannesson 2015; Ravinet et al. 2016). That is, the majority of the genomic
variation linked to the evolution of ecotypes is not shared between the studied islands. The
EOS analysis of the Ravinet et al. data seems to support this finding. At the same time, we
identify far fewer outliers, with Saltö—which is closer to the mainland—having the lowest
number. However, this is the opposite of the result in Ravinet’s study where Saltö had the
highest number of outliers.

In any case, the lower number of outliers that we found in Saltö, could also explain our
results of reduced shared divergence between Saltö and the two other islands. We cannot
rule out that our finding can be an artefact due to the conservativeness of EOS, but,
alternatively, the previous results could be due to an excess of false positives, hiding the
pattern of reduced shared divergence in Saltö.

To conclude, it is worth mentioning that combination of multiple signals from different tests
has been proposed as a way of improving power/false positive rate relationship for the
selection detection methods (Zeng, Shi, and Wu 2007; Lin et al. 2011; Vatsiou, Bazin, and
Gaggiotti 2016). Accordingly, the \( nvdF_{ST} \) test does just that. It combines haplotype and
population differentiation information and may be a helpful tool to explore patterns of
divergent selection when approximate knowledge of the haplotype phase is at hand.

Alternatively, the EOS method is a conservative outlier test useful when the full set of SNPs
is unlinked or under weak linkage. Both strategies can be applied without the need of
performing neutral simulations and have low false positive rate.
Acknowledgements

I thank E. Rolán-Alvarez for useful comments on the manuscript. This work was supported by Ministerio de Economía y competitividad (CGL2012-39861-C02-01 and BFU2013-44635-P), Xunta de Galicia (Grupo con Potencial de Crecimiento, GPC2013-011) and fondos FEDER. The author declares to have no conflict of interest.

Software Accessibility

The computer program HacDivSel implementing the methods explained in this article jointly with the user manual, are available from the author web site http://acraaj.webs.uvigo.es/software/HacDivSel.zip.

Bibliography


