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

email: 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

59 Krakauer 1973), this technique has been improved in several ways depending on the
60 summary statistic's -the F_{ST} or any other differentiation index - expected neutral distribution.
61 That is, in order to account for more realistic situations, the different methods change the
62 assumptions of the null demographic model (reviewed in Whitlock and Lotterhos 2015).

63 Even so, one of the main drawbacks of outlier-based methods is the difficulty of defining an
64 accurate null model since there are several historical events and demographic scenarios,
65 other than local selection, that can produce similar F_{ST} patterns. Still in the presence of local
66 adaptation we can expect different F_{ST} patterns, since the populations involved may be more
67 or less connected by migration, and this will influence the structure of the genetic variation
68 both at intra and inter-population levels. Consequently, when using outlier methods based
69 on the deviation over an expected demographic null model, we always face the risk of
70 having an excess of false positives (Perez-Figueroa et al. 2010; Bierne, Roze, and Welch 2013;
71 De Mita et al. 2013; Lotterhos and Whitlock 2014).

72 Another problem concerning outlier-based methods is that they have low power when the
73 overall F_{ST} is high as it can happen when the genetic basis of the adaptation is polygenic or
74 when the populations under study are subspecies (De Villemereuil et al. 2014; Whitlock and
75 Lotterhos 2015). Fortunately, in recent years, the amount of information on genomes of
76 several species has increased (Ellegren 2014) and consequently new and more sophisticated
77 methods can now be applied to detect local adaptation in non-model organisms. As
78 mentioned above, the linkage disequilibrium (LD) is the basis for several computational
79 methods aiming the detection of selective sweeps (reviewed in Crisci et al. 2012). Some LD-
80 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 (Omegaplus, 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 nvdF_{ST} 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 *nvd*F_{ST} because it combines a normalized variance difference, *nvd*, with the F_{ST} index. The *nvd* part performs a sliding-window approach to identify sites with specific selective patterns. When combined with the F_{ST}, 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^2_{1i} and correspondingly variance S^2_{2i} 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^2_{1i}) 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^2_{2i}) should not be affected by the sweeping effect because it does not carry the favored allele. So, the difference S^2_{2i} - S^2_{1i} 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

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

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

189 HAC distance, *m*, of the sample, and m_1 and m_2, the means corresponding to the partitions
190 P_1 and P_2, respectively. We have the following relationships for the mean *m* and sample
191 variance S^2 (the subscripts 1 or 2 identify each partition; see the Appendix A-2 for details)

192 $$m = \frac{n_1 m_1 + n_2 m_2}{n}; \qquad S^2 - \bar{S} = \frac{n}{n-1}\Delta \qquad (1)$$

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

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

201 $$vd_i = vd_{i0} + 4f_i(1-f_i)(S_{2i}^2 - S_{1i}^2) \qquad (2)$$

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

203 Therefore, the effect of selection upon vd_i is two-fold. The first term of the sum in (2)
204 corresponds to the effect of the difference between means and the second between
205 variances. Clearly, increasing S_{2i} or decreasing S_{1i}, as expected under selection, will increase
206 the value of the statistic. If S_{1i} and S_{2i} are equal then the value of vd_i is independent of the
207 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_{i0} 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_{max}) for vd_i

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

$$vd_i \leq d_{max} \qquad (3)$$

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

$$nvd_i = \frac{vd_{i0} + 4f_i(1-f_i)(S_2^2 - S_1^2)}{d_{max}} \qquad (4)$$

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_{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}.

Sign test

In our definition of *nvd* the term vd_{i0} 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

$$tss = \frac{4(n-1)S^2 - 2\sum_h \mathrm{hac}_{1h}^2}{nL^2} \qquad (5)$$

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

292 procedure has variance $p(1-p)n$ which is larger at intermediate pooled frequency values.
293 Thus, the method has the desired property to be more conservative at intermediate pooled
294 frequencies which minimize the possibility of false positives.

295

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

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

309 The *q*-values (Storey 2003) can be seen as a multiple testing analogs of *p*-values. They have
310 been proposed as an useful approach for evaluating method performance in terms of false
311 discoveries (De Villemereuil *et al.* 2014). Accordingly, we estimate the *q*-value (see Appendix
312 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 *nvd*F_{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 *nvd*F_{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. saxattilis* (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 ($\rho = 4Nr \in \{0, 4, 12, 60, \text{unlinked}\}$) 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).

377 The simulated data is also utilized for the comparison of $nvdF_{ST}$ with OmegaPlus (Alachiotis,
378 Stamatakis, and Pavlidis 2012) and of the EOS test with BayeScan 2.1 (Foll and Gaggiotti
379 2008). We chose OmegaPlus because of its good performance, compared with other
380 haplotype-based methods, under non-equilibrium demographic conditions (Crisci et al.
381 2013). Several combinations of the parameters for OmegaPlus were tested in (Rivas,
382 Dominguez-Garcia, and Carvajal-Rodriguez 2015), here we selected the ones with the best
383 performance. BayeScan was chosen because it is one of the main state-of-the-art outlier-
384 based programs. The parameters for BayeScan were the default ones as this is a
385 conservative setting and we were interested in comparing the false positive rates. Only SNPs
386 shared between populations and with a minimum allele frequency (maf) of 2 per population
387 (4%) were considered.

388

389 ## Results

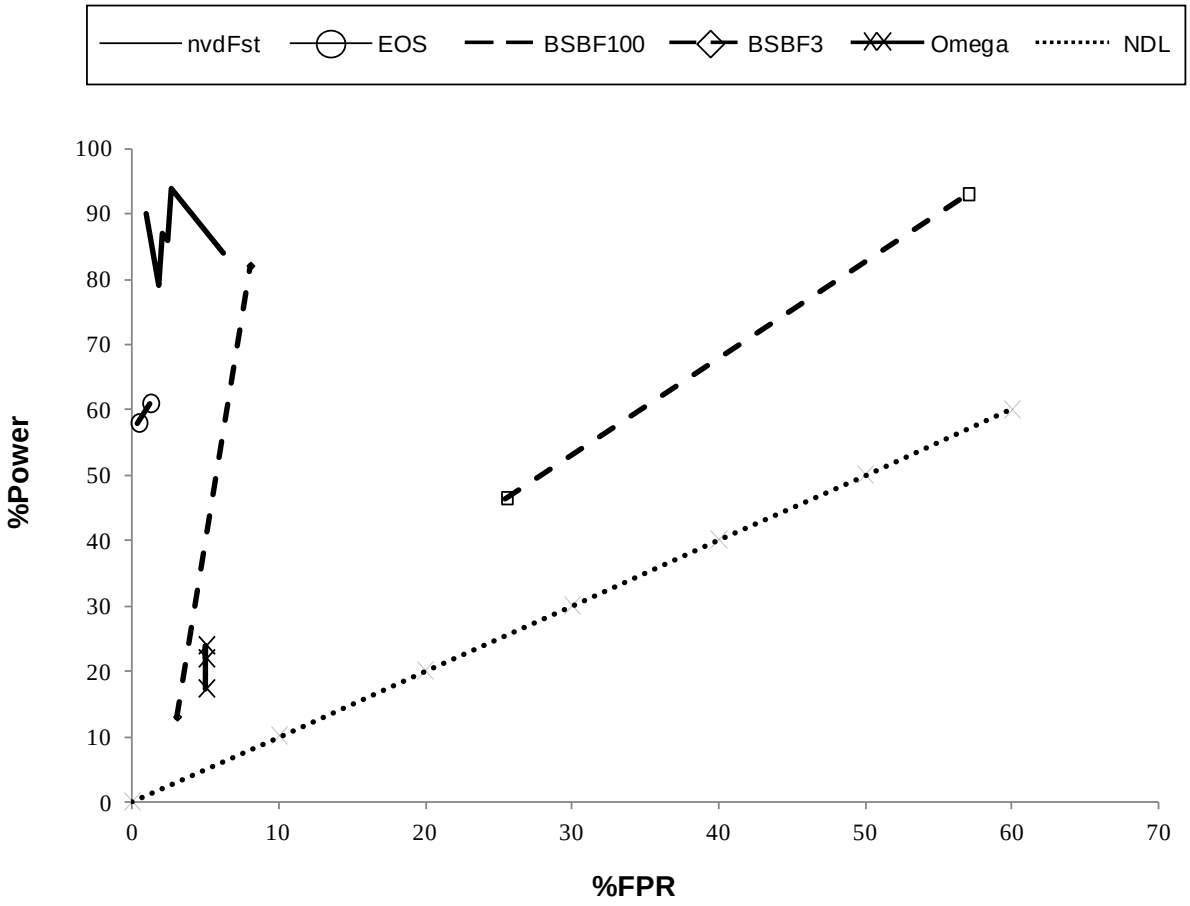
390 In what follows we present the results after applying the tests to simulated and real data.
391 The power of a test (true positive rate) is measured as the % of runs in which selection was
392 detected from simulated selective scenarios and the false positive rate (FPR) is measured as
393 the % of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

395 ### ROC curves

396 The ROC curve is a useful way to explore the performance of any detection statistic because
397 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 $nvdF_{ST}$, EOS, OmegaPlus and BayesCan methods through the α = 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.



410

411 Figure 1. ROC curves for *nvdF*$_{ST}$, EOS, OmegaPlus (Omega) and BayeScan (Bayes factor 3: BSBF3 and Bayes
412 factor 100: BSBF100). FPR: False Positive Rate.

413

414 The method *nvdF*$_{ST}$ is the best performing one since it occupies the upper-left corner.
415 However, the scenario with independent (unlinked) markers is not visible in the *nvdF*$_{ST}$ plot
416 since it has a (0, 0) coordinate (see Table 1 and next section below). For the *nvdF*$_{ST}$ plotted
417 scenarios, the power ranges between 80-90% (y-axis) and FPR (x-axis) is below 5%. On the
418 contrary, OmegaPlus is too close to the NDL diagonal under the same divergent selection
419 scenarios.

420 The BayeScan and EOS methods are plotted only for the unlinked and the weak-linkage (ρ =
421 60; 1.5 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 $nvdF_{ST}$ and EOS under the different simulated scenarios.

Combined Method ($nvdF_{ST}$)

Under a single locus architecture with selection $\alpha = 4Ns = 600$ and migration $Nm=10$, the power of $nvdF_{ST}$ 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 $nvdF_{ST}$ 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 F_{ST} index measure.

458 Table 1. Performance of the combined method (*nvdF*$_{ST}$) with *n* = 1 selective site located at
459 the center of the chromosome or *n* = 5 (see Appendix A-6). Selection was α = 4*Ns* = 600 and
460 migration *Nm* = 10. Mean localization is given in distance kb from the real selective position.

∑	θ	ρ	n	%Power	%FPR (γ = 5%)	*q*-value	Localization (kb)
65	12	0	1	87	2.1	0.0058	±458
63	12	4	1	94	2.7	0.0008	±200
60	12	12	1	90	1.0	0.0003	±33
251	60	0	1	79	1.8	0.0048	±60
232	60	4	1	84	6.2	0.0011	±17
249	60	60	1	86	2.4	0.0002	<±1
282	60	60	5	99	2.4	0.0002	<±1
318	60	∞	1	0	0	-	-

461 ∑: Mean number of shared SNPs per Mb. θ: Mutation rate. ρ: Recombination rate. FPR: false positive rate. *q*-
462 value: mean estimated *q*-value for the significant tests. ∞: Independently segregating sites.

463

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

465 The performance of *nvdF*$_{ST}$ under the strong selection scenario (α = 6000) in the short-term
466 (500 generations) varies between 44% for fully linked to 67% for weak linked markers (Table
467 2). Not surprisingly, the number of segregating sites is considerably reduced. In fact the

468 minimum window size allowed by the program had to be shortened from 51 to 25 to

469 perform the analyses. Notably the false positive rate (FPR) was 0.

470 Concerning weak selection in long-term scenarios (Table 2, α = 140) the power varies

471 between 49-52% with false positive rate between 2.2 and 5.7%.

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

∑	θ	ρ	α	*t*	%Power	%FPR (γ = 5%)	*q*-value	Localization (kb)
112	60	0	6000	500	44	0	0	±66
32*	60	4	6000	500	63	0	0.0014	±5
62	60	60	6000	500	67	0	0.0008	±93
165	60	0	140	5,000	49	3.6	0.0280	±33
156	60	4	140	5,000	52	5.7	0.0219	±14
135	60	60	140	5,000	49	2.2	0.0054	±6

475 ∑: mean number of shared SNPs per Mb. θ: Mutation rate. ρ: recombination rate. *t*: number of generations.
476 FPR: false positive rate. *q*-value: mean estimated *q*-value for the significant tests. *: only 40 runs having a
477 minimum of 25 SNPs.

478

479

480 **Extreme Outlier Set Test (EOS)**

481 We applied the EOS test under the single locus architecture with selection α = 600 and
482 migration *Nm*=10. As desired, the test is very conservative with false positive rates below the
483 nominal 0.05 in every case (Table 3). Not surprisingly for an outlier-based method, the test
484 has no power if the markers are strongly linked (ρ from 0 to 12) or under a polygenic setting
485 (row with *n* = 5 in Table 3). However, in the case of independent SNPs but also in the case of
486 maps with 250-300 SNPs/Mb under 1.5 cM/Mb the power rises up to 60%. Therefore, the
487 EOS test is complementary to $nvdF_{ST}$ having its maximum power when the latter has its
488 minimum.

489

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 α = 600 and *Nm* = 10. Mean localization is given in distance kb from the real selective position.

∑	θ	ρ	n	%Power EOS	%FPR (γ = 5%)	q'-value	Localization (kb)
65	12	0	1	0	0	-	-
63	12	4	1	0.2	0	0.46	±3
60	12	12	1	1.1	0	0.45	±77
251	60	0	1	0.7	0	0.10	0
232	60	4	1	1.3	0	0.20	±150
249	60	60	1	58	0.4	0.5	<±1
282	60	60	5	1.6	0.4	0.49	±5
318	60	∞	1	61	1.2	3×10^{-6}	±7

∑: mean number of shared SNPs per Mb. θ: mutation rate. ρ: recombination rate. FPR: false positive rate. q'-value: mean corrected (see appendix A-4) estimated *q*-value in the significant tests. ∞: 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 (3×10^{-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, α = 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 ρ = 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 α = 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.

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

∑: Mean number of shared SNPs per Mb. θ: Mutation rate. ρ: 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 $nvdF_{ST}$ in the short term (500 generations) with a single selective site. Selection was α = 600 and *Nm* = 50. Mean localization is given in distance kb from the real selective position.

∑	θ	ρ	%Power	%FPR (γ = 5%)	*q*-value	Localization (kb)
116	60	0	56	0	0.0098	±152
180	60	4	59	0	0.0042	±123
178	60	60	34	0.1	0.0096	±4

∑: Mean number of shared SNPs per Mb. θ: Mutation rate. ρ: 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).

Island	Unique	Only with Jutholmen	Only with Ramsö	Only with Saltö	Shared all	Total
Jutholmen	27 (59)	__	2 (13)	0 (16)	2 (9)	31 (97)
Ramsö	24 (86)	2 (13)	__	1 (21)	2 (9)	29 (129)
Saltö	6 (134)	0 (16)	1 (21)	__	2 (9)	9 (180)

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).

598 Table 7. Summary of EOS analysis for the between ecotypes *Littorina saxatilis* data (Ravinet
599 *et al.* 2016).

Island	Nonoutliers	Outliers not in EOS	EOS	F_{ST}	F_{ST_EOS}	$pval_{EOS}$	$qval_{EOS}$
Jutholmen	4564	91	31	0.045	0.40	0.004	0.52
Ramsö	4602	82	29	0.064	0.53	0.005	0.63
Saltö	4632	51	9	0.060	0.60	0.002	0.76

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

603

604 Discussion

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

613 For the $nvdF_{ST}$ method, it has been shown that combining haplotype-based and F_{ST}
614 differentiation information is a quite powerful strategy for detecting divergent selection.
615 However, the $nvdF_{ST}$ algorithm does not perform well when the whole set of markers is
616 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.

637 **Position Effect**

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

647

648 **High Migration Scenario**

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

654 In the case of the EOS test, there is an obvious tradeoff between the astringency of the
655 cutoff point for the outlier set and the migration rate. Our cutoff depends on the F_{ST} upper-
656 bound which is a function of the number of populations, the sample sizes and the minimum
657 allele frequency. However, with higher migration the upper bound of the F_{ST} should be lower

658 because as migration increases the number of populations becomes virtually one. Therefore,
659 a possible solution to improve the efficiency of EOS under high migration would be to
660 update the F_{ST} upper bound as a function of the migration rate.

661

662 **Empirical Data**

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

677 Regarding the shared genomic divergence of the *L. saxatilis* system in Swedish populations,
678 it seems to be a small proportion of the total genomic divergence (Hollander, Galindo, and
679 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 *nvd*F_{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

Alachiotis, N., A. Stamatakis, and P. Pavlidis. 2012. OmegaPlus: a scalable tool for rapid detection of selective sweeps in whole-genome datasets. Bioinformatics **28**:2274-2275.

Bierne, N., D. Roze, and J. J. Welch. 2013. Pervasive selection or is it…? why are FST outliers sometimes so frequent? Molecular Ecology **22**:2061-2064.

Bourret, V., M. P. Kent, C. R. Primmer, A. Vasemägi, S. Karlsson, K. Hindar, P. McGinnity, E. Verspoor, L. Bernatchez, and S. Lien. 2013. SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (Salmo salar). Molecular Ecology **22**:532.

Carvajal-Rodriguez, A. 2008. GENOMEPOP: A program to simulate genomes in populations. BMC Bioinformatics **9**:223.

Crisci, J. L., Y.-P. Poh, A. Bean, A. Simkin, and J. D. Jensen. 2012. Recent Progress in Polymorphism-Based Population Genetic Inference. Journal of Heredity.

Crisci, J. L., Y.-P. Poh, S. Mahajan, and J. D. Jensen. 2013. The Impact of Equilibrium Assumptions on Tests of Selection. Frontiers in Genetics **4**.

728 Crow, J. F., and M. Kimura. 1970. An Introduction to Population Genetics Theory. Harper &
729 Row, New York.
730 Cheverud, J. M. 2001. A simple correction for multiple comparisons in interval mapping
731 genome scans. Heredity **87**:52-58.
732 De Mita, S., A.-C. Thuillet, L. Gay, N. Ahmadi, S. Manel, J. Ronfort, and Y. Vigouroux. 2013.
733 Detecting selection along environmental gradients: analysis of eight methods and
734 their effectiveness for outbreeding and selfing populations. Molecular Ecology
735 **22**:1383.
736 De Villemereuil, P., É. Frichot, É. Bazin, O. François, and O. E. Gaggiotti. 2014. Genome scan
737 methods against more complex models: when and how much should we trust them?
738 Molecular Ecology **23**:2006-2019.
739 Debarre, F., S. Yeaman, and F. Guillaume. 2015. Evolution of Quantitative Traits under a
740 Migration-Selection Balance: When Does Skew Matter?*. The American Naturalist
741 **0**:S000.
742 Devlin, B., and N. Risch. 1995. A Comparison of Linkage Disequilibrium Measures for Fine-
743 Scale Mapping. Genomics **29**:311-322.
744 Dharmadhikari, S. W., and K. Joag-Dev. 1989. Upper bounds for the variances of certain
745 random variables. Communications in statistics-theory and methods **18**:3235-3247.
746 Ellegren, H. 2014. Genome sequencing and population genomics in non-model organisms.
747 Trends in Ecology & Evolution **29**:51-63.
748 Ferretti, L., S. E. Ramos-Onsins, and M. Pérez-Enciso. 2013. Population genomics from pool
749 sequencing. Molecular Ecology **22**:5561-5576.
750 Foll, M., and O. Gaggiotti. 2008. A genome-scan method to identify selected loci appropriate
751 for both dominant and codominant markers: a Bayesian perspective. Genetics
752 **180**:977-993.
753 Hollander, J., J. Galindo, and R. K. Butlin. 2015. Selection on outlier loci and their association
754 with adaptive phenotypes in Littorina saxatilis contact zones. Journal of Evolutionary
755 Biology **28**:328-337.
756 Hudson, R. R. 2002. Generating samples under a Wright-Fisher neutral model of genetic
757 variation. Bioinformatics **18**:337-338.
758 Hussin, J., P. Nadeau, J.-F. Lefebvre, and D. Labuda. 2010. Haplotype allelic classes for
759 detecting ongoing positive selection. BMC Bioinformatics **11**:65.
760 Jensen, J. D., M. Foll, and L. Bernatchez. 2016. The past, present and future of genomic scans
761 for selection. Molecular Ecology **25**:1-4.
762 Johannesson, K. 2015. What can be learnt from a snail? Evolutionary Applications:n/a.
763 Kim, Y., and R. Nielsen. 2004. Linkage disequilibrium as a signature of selective sweeps.
764 Genetics **167**:1513-1524.
765 Lewontin, R. C. 1988. On measures of gametic disequilibrium. Genetics **120**:849-852.
766 Lewontin, R. C., and J. Krakauer. 1973. Distribution of Gene Frequency as a Test of Theory of
767 Selective Neutrality of Polymorphisms. Genetics **74**:175-195.
768 Lin, K., H. Li, C. Schlotterer, and A. Futschik. 2011. Distinguishing positive selection from
769 neutral evolution: boosting the performance of summary statistics. Genetics
770 **187**:229-244.

771 Lotterhos, K. E., and M. C. Whitlock. 2014. Evaluation of demographic history and neutral
772 parameterization on the performance of FST outlier tests. Molecular Ecology
773 **23**:2178.
774 Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proceedings of the
775 National Academy of Sciences **70**:3321-3323.
776 Nielsen, R., S. Williamson, Y. Kim, M. J. Hubisz, A. G. Clark, and C. Bustamante. 2005.
777 Genomic scans for selective sweeps using SNP data. Genome Res **15**:1566-1575.
778 Pavlidis, P., J. D. Jensen, and W. Stephan. 2010. Searching for footprints of positive selection
779 in whole-genome SNP data from nonequilibrium populations. Genetics **185**:907-922.
780 Perez-Figueroa, A., M. J. Garcia-Pereira, M. Saura, E. Rolan-Alvarez, and A. Caballero. 2010.
781 Comparing three different methods to detect selective loci using dominant markers.
782 Journal of Evolutionary Biology **23**:2267-2276.
783 Ravinet, M., A. Westram, K. Johannesson, R. Butlin, C. André, and M. Panova. 2016. Shared
784 and nonshared genomic divergence in parallel ecotypes of Littorina saxatilis at a local
785 scale. Molecular Ecology:287-305.
786 Renaut, S., A. W. Nolte, S. M. Rogers, N. Derome, and L. Bernatchez. 2011. SNP signatures of
787 selection on standing genetic variation and their association with adaptive
788 phenotypes along gradients of ecological speciation in lake whitefish species pairs
789 (Coregonus spp.). Molecular Ecology **20**:545.
790 Rivas, M. J., S. Dominguez-Garcia, and A. Carvajal-Rodriguez. 2015. Detecting the Genomic
791 Signature of Divergent Selection in Presence of Gene Flow. Current Genomics **16**:203-
792 212.
793 Rolan-Alvarez, E. 2007. Sympatric speciation as a by-product of ecological adaptation in the
794 Galician Littorina saxatilis hybrid zone. Journal of Molluscan Studies **73**:1-10.
795 Rousset, F. 2008. genepop'007: a complete re-implementation of the genepop software for
796 Windows and Linux. Molecular Ecology Resources **8**:103-106.
797 Sabeti, P. C.P. VarillyB. FryJ. LohmuellerE. HostetterC. CotsapasX. XieE. H. ByrneS. A.
798 McCarrollR. GaudetS. F. SchaffnerE. S. LanderK. A. FrazerD. G. BallingerD. R. CoxD. A.
799 HindsL. L. StuveR. A. GibbsJ. W. BelmontA. BoudreauP. HardenbolS. M. LealS.
800 PasternakD. A. WheelerT. D. WillisF. YuH. YangC. ZengY. GaoH. HuW. HuC. LiW. LinS.
801 LiuH. PanX. TangJ. WangW. WangJ. YuB. ZhangQ. ZhangH. ZhaoH. ZhaoJ. ZhouS. B.
802 GabrielR. BarryB. BlumenstielA. CamargoM. DefeliceM. FaggartM. GoyetteS. GuptaJ.
803 MooreH. NguyenR. C. OnofrioM. ParkinJ. RoyE. StahlE. WinchesterL. ZiaugraD.
804 AltshulerY. ShenZ. YaoW. HuangX. ChuY. HeL. JinY. LiuY. ShenW. SunH. WangY.
805 WangY. WangX. XiongL. XuM. M. WayeS. K. TsuiH. XueJ. T. WongL. M. GalverJ. B.
806 FanK. GundersonS. S. MurrayA. R. OliphantM. S. CheeA. MontpetitF. ChagnonV.
807 FerrettiM. LeboeufJ. F. OlivierM. S. PhillipsS. RoumyC. SalleeA. VernerT. J. HudsonP.
808 Y. KwokD. CaiD. C. KoboldtR. D. MillerL. PawlikowskaP. Taillon-MillerM. XiaoL. C.
809 TsuiW. MakY. Q. SongP. K. TamY. NakamuraT. KawaguchiT. KitamotoT. MorizonoA.
810 NagashimaY. OhnishiA. SekineT. TanakaT. TsunodaP. DeloukasC. P. BirdM. DelgadoE.
811 T. DermitzakisR. GwilliamS. HuntJ. MorrisonD. PowellB. E. StrangerP. WhittakerD. R.
812 BentleyM. J. DalyP. I. de BakkerJ. BarrettY. R. ChretienJ. MallerS. McCarrollN.
813 PattersonI. Pe'erA. PriceS. PurcellD. J. RichterP. SabetiR. SaxenaS. F. SchaffnerP. C.
814 ShamP. VarillyD. AltshulerL. D. SteinL. KrishnanA. V. SmithM. K. Tello-RuizG. A.
815 ThorissonA. ChakravartiP. E. ChenD. J. CutlerC. S. KashukS. LinG. R. AbecasisW.

816 GuanY. LiH. M. MunroZ. S. QinD. J. ThomasG. McVeanA. AutonL. BottoloN. CardinS.
817 EyheramendyC. FreemanJ. MarchiniS. MyersC. SpencerM. StephensP. DonnellyL. R.
818 CardonG. ClarkeD. M. EvansA. P. MorrisB. S. WeirT. TsunodaT. A. JohnsonJ. C.
819 MullikinS. T. SherryM. FeoloA. SkolH. ZhangC. ZengH. ZhaoI. MatsudaY. FukushimaD.
820 R. MacerE. SudaC. N. RotimiC. A. AdebamowoI. AjayiT. AniagwuP. A. MarshallC.
821 NkwodimmahC. D. RoyalM. F. LeppertM. DixonA. PeifferR. QiuA. KentK. KatoN.
822 NiikawaI. F. AdewoleB. M. KnoppersM. W. FosterE. W. ClaytonJ. WatkinR. A. GibbsJ.
823 W. BelmontD. MuznyL. NazarethE. SodergrenG. M. WeinstockD. A. WheelerI. YakubS.
824 B. GabrielR. C. OnofrioD. J. RichterL. ZiaugraB. W. BirrenM. J. DalyD. AltshulerR. K.
825 WilsonL. L. FultonJ. RogersJ. BurtonN. P. CarterC. M. CleeM. GriffithsM. C. JonesK.
826 McLayR. W. PlumbM. T. RossS. K. SimsD. L. WilleyZ. ChenH. HanL. KangM. GodboutJ.
827 C. WallenburgP. L'ArchevequeG. BellemareK. SaekiH. WangD. AnH. FuQ. LiZ. WangR.
828 WangA. L. HoldenL. D. BrooksJ. E. McEwenM. S. GuyerV. O. WangJ. L. PetersonM.
829 ShiJ. SpiegelL. M. SungL. F. ZachariaF. S. CollinsK. KennedyR. Jamieson, and J. Stewart.
830 2007. Genome-wide detection and characterization of positive selection in human
831 populations. Nature **449**:913-918.
832 Schubert, E., A. Zimek, and H.-P. Kriegel. 2012. Local outlier detection reconsidered: a
833 generalized view on locality with applications to spatial, video, and network outlier
834 detection. Data Mining and Knowledge Discovery **28**:190-237.
835 Sidak, Z. 1967. Rectangular Confidence Regions for the Means of Multivariate Normal
836 Distributions. Journal of the American Statistical Association **62**:626-633.
837 Smith, J. M., and J. Haigh. 1974. The hitch-hiking effect of a favourable gene. Genetical
838 research **23**:23-35.
839 Storey, J. 2003. The positive false discovery rate: A Bayesian interpretation and the q-value.
840 The Annals of Statistics **31**:2013-2035.
841 Thibert-Plante, X., and S. Gavrilets. 2013. Evolution of mate choice and the so-called magic
842 traits in ecological speciation. Ecol Lett.
843 Tukey, J. W. 1977. Exploratory data analysis. Addison-Wesley, Reading, Mass.
844 Vatsiou, A. I., E. Bazin, and O. E. Gaggiotti. 2016. Detection of selective sweeps in structured
845 populations: a comparison of recent methods. Molecular Ecology **25**:89-103.
846 Vattani, A. 2011. k-means Requires Exponentially Many Iterations Even in the Plane. Discrete
847 & Computational Geometry **45**:596-616.
848 Voight, B. F., S. Kudaravalli, X. Wen, and J. K. Pritchard. 2006. A map of recent positive
849 selection in the human genome. PLoS Biol **4**:e72.
850 Whitlock, M., C., and K. Lotterhos, E. 2015. Reliable Detection of Loci Responsible for Local
851 Adaptation: Inference of a Null Model through Trimming the Distribution of FST. The
852 American Naturalist **186**:S24-S36.
853 Whitlock, M. C. 2015. Modern Approaches to Local Adaptation. The American Naturalist
854 **186**:S1.
855 Yeaman, S. 2015. Local Adaptation by Alleles of Small Effect*. The American Naturalist
856 **186**:S74-S89.
857 Yeaman, S., and S. P. Otto. 2011. Establishment and Maintenance of Adaptive Genetic
858 Divergence under Migration, Selection, and Drift. Evolution **65**:2123-2129.
859 Yeaman, S., and M. C. Whitlock. 2011. The Genetic Architecture of Adaptation under
860 Migration-Selection Balance. Evolution **65**:1897-1911.

861 Zeng, K., S. Shi, and C. I. Wu. 2007. Compound tests for the detection of hitchhiking under
862 positive selection. Mol Biol Evol **24**:1898-1908.
863
864
865