

Detecting the Genomic Signature of Divergent Selection in Presence of Gene Flow

M. J. Rivas¹, S. Domínguez-García¹ and A. Carvajal-Rodríguez^{1,*}

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

Abstract: The study of local adaptation is a main focus of evolutionary biology since it may contribute to explain the current species diversity. The genomic scan procedures permit for the first time to study the connection between specific DNA patterns and processes as natural selection, genetic drift, recombination, mutation and gene flow. Accordingly, the information on genomes from non-model organisms increases and the interest on detecting the signal of natural selection in the DNA sequences of different populations also raises. The main goal of the present work is to explore a sequence-based method for detecting natural selection in divergent populations connected by migration. In doing so, we rely on a recently published statistic based upon the definition of haplotype allelic classes (HAC). The original measure was modified to be more sensitive to intermediate frequencies in non-model species. A linkage-disequilibrium-based method was also assayed and individual-based simulations were performed to test the methods. The results suggest that the HAC-based methods and, specifically, the new proposed method are quite powerful for detecting the footprint of moderate divergent selection. They are also robust to reasonable model misspecification. One obvious advantage of the new algorithm is that it does not require knowledge of the allelic state.



A. Carvajal-Rodríguez

Keywords: Local adaptation, Single nucleotide polymorphism, Selective sweep, Detection of selection, Gene flow, Divergent populations.

INTRODUCTION

The recent development of genomic scan procedures has allowed for the searching of the mark of positive selection in the DNA of some model species [1, 2]. Now we are facing the application of such techniques to non-model organisms [3, 4]. However, the new bunch of information poses various problems to deal with. For example, several genomic patterns will arise due to the combination of different evolutionary processes such as natural selection, genetic drift, recombination, mutation and gene flow. Thus, understanding the connection between such processes and the specific DNA patterns found in genomes is a main objective of current evolutionary biology [5-7]. The detection of signatures of divergent selection in DNA sequences from populations with gene flow will be a key to future studies of ecological speciation scenarios. Although the study of the genomic patterns behind such kind of processes is still in its infancy for most organisms [8] there is an increasing evidence indicating that the footprint of early divergent selection and speciation processes can be widespread in genomes including species that coexist in complete sympatry [9]. Accordingly, as the information on genomes or partial-genomes from non-model organisms increases, it also increases the interest in detecting the signal of selection throughout DNA sequences of these species. The problem is that the processes of divergence and speciation in presence of gene flow are expected to generate

quite heterogeneous patterns of genomic differentiation; hindering the identification of specific selection footprints [10-12]. The existing methods for detecting the signature of selection in structured populations of non-model species use the information of molecular markers to compare the genetic differentiation among populations [13-15] or genotype-environmental correlations ([16, 17] and references therein). Such strategies have as main caveat the high rates of false positives [16, 18-20].

On the contrary, genome-based methods for detecting divergent selection in non-model organisms have not been applicable due to the absence of appropriate sequence and demographic information. Currently, there are several methods for detecting the footprint of selection at genomic level (reviewed in [21-23]) which were mainly developed for model species having enough phylogenetic information on the involved variability (knowledge about ancestral versus derived alleles). Hence, the detection of divergent selection with migration can be complicated because the sustained gene flow between divergent alleles may hide the signal of selection [5, 10, 12, 24-27]. For that reason, it is not expected that methods based on detecting linkage disequilibrium (LD) signature (such as EHH [28]) or frequency spectrum (such as DH [29]) work well under some scenarios of divergent selection and local adaptation.

The main goal in this study is to set up a sequence-based method for detecting local ongoing processes in which natural selection is working in different directions. The involved populations may be 'invaded' by alternative alleles from their neighborhood populations. In doing so, we rely on the defi-

*Address correspondence to this author at the Departamento de Bioquímica, Genética e Inmunología. Universidad de Vigo, 36310 Vigo, Spain; Tel: +34 986813828; E-mail: acraaj@uvigo.es

dition of haplotype allelic classes (HAC, see Material and Methods section and [30]). A previous HAC-based statistic, called Svd, takes advantage from the assumption that a putatively selected allele will be linked more often with the major (the most frequent) alleles than with the minor alleles of neutral positions. Consequently, Svd is able to detect ongoing processes of positive selection [30]. Thus, this method does not perform specific LD measures but a normalized comparison of variances of haplotype classes. The partition into classes is performed depending on the presence of the putatively selected allele. This is interesting since selective sweeps may produce additional patterns of variation other than those measured by LD [31]. Because of the flexibility provided by the HAC method, it seems appropriate for scenarios undergoing disruptive selection and gene flow. It is worth mentioning that the original statistic does not perform well when the frequency of the selected allele is intermediate [30] and also has the drawback of needing information on the allelic state (ancestral or derived). Moreover, it does not provide a clear indication about the optimal window size to use when working at genomic level. In consequence, we have modified the original measure to be more sensitive to intermediate frequencies and applicable to non-model species without information about the state of the alleles. This new measure is called SvdM. We have also automatically chosen as optimal window size that one giving the highest score. Both Svd and SvdM have an expected value of equal or less than 0 under the standard neutral model. However, in the presence of selection, the value of the statistics should be higher than 0 (although some demographic scenarios can also generate higher values). Individual-based simulations were performed to test the original Svd and its modification as well as a LD-based method implemented in the Omega-Plus software [32]. We chose the LD-based method because it has been reported as the best one under equilibrium and non-equilibrium conditions [21]. However, it requires the selection to be strong enough and is quite dependent on the timing of the selective sweep and on the number of segregating sites [33].

To perform the simulations as real as possible, a model that resembles a biological scenario of two populations or microhabitats connected by gene flow was implemented while undergoing divergent selection. There are several known examples of adaptation to contrasting environments with ongoing gene flow, such as the intertidal marine snail *L. saxatilis*, that is a well-known example of ecomorphological diversification [34] but also the wild populations of *Salmo Salar* [35] or Lake whitefish species (*Coregonus spp. Salmonidae* [36]) or even tree species as Cork Oak (*Quercus suber* [37]).

The most favorable conditions for the rapid formation of ecotypes, under local adaptation with gene flow, imply moderate selection pressures and few loci with large effects rather than many with small effects [12, 38]. Because simulating biological systems as real as possible has proven to be a useful strategy [38], we incorporate relevant demographic information from *L. saxatilis* such as migration rates and population sizes estimated from field data. Then we use the sequence samples obtained from the simulations to check the HAC-based methods both at the initial steps (hundred of generations) of the divergent selection process and also when

the number of generations has been large enough for equilibrium to be reached (5-10N generations).

MATERIAL AND METHODS

Statistic for Haplotype Allelic Class (HAC) Patterns Under Divergence

For each haplotype, the HAC-based statistics compute a distance with respect to a reference configuration. This distance is called the haplotype allelic class (HAC [30]). The reference configuration is represented as the haplotype carrying only the major frequency alleles of its constituting SNPs. Therefore, the HAC of a given haplotype will be the number of minor frequency alleles it carries [30]. Haplotypes with the same HAC distance will be grouped. Given a candidate SNP, we may assume that if the new (derived) allele is at highest frequency then it is the positively selected one. Consequently, the data can be partitioned into those haplotypes carrying the major allele of the SNP under evaluation and those carrying the minor allele. We can compute the variance V_1 of the HAC distances for the haplotypes in the partition with the major allele and similarly V_2 in the partition with the minor allele. The summary statistic developed by Labuda and co-workers [30] is based on the normalized variance difference:

$$Svd = \frac{(V_2 - V_1)}{S} \times f$$

where f is the frequency of the derived allele and S is the number of SNPs considered in the variance estimation (window size). This statistic, while efficient for strong ongoing positive selection, would fall short when the selective sweep is at low frequency (i.e., the selected allele has not reached intermediate frequencies [30]). Another issue is that Svd needs to distinguish between derived and ancestral alleles, but this is not always possible when working with non-model species. To solve these problems, the Svd statistic was modified to be independent of the state, ancestral or derived, of the selected allele and at the same time to have its highest power at intermediate frequencies. This might be of special interest when the evolutionary scenarios involve divergent selection and migration. We call the new statistic SvdM:

$$SvdM = 4 \frac{(V_2 - V_1)}{S} \times f \times (1 - f) = 4Svd \times (1 - f)$$

Note that when computing these statistics throughout the genome, the maximum value is returned. However, other measures could be assayed such as for example the average of positive Svd (SvdM) values.

Simulations

To test the ability of the Svd and SvdM methods for detecting the footprint of selection, the program GenomePop [39] was used to simulate two populations of facultative hermaphrodites under divergent selection and migration. Each individual consisted of a diploid chromosome of length 1Mb of biallelic loci. There was only one selected locus that had the derived allele as beneficial in population 1 and the ancestral allele as beneficial in population 2. Thus, natural selection acts in opposite directions in the two populations. In population 1 the favored derived allele was initially at low

frequency (1/1000), while in population 2 the favored ancestral allele was initially fixed. In population 1 (selective allele at low frequency) we checked that the allele was not lost in the first generations, if so we discarded that run. The fitness model was $w = (1-hs)$ where h had a value of 0.5 in the heterozygote and 1 in the homozygote. Ancestral alleles had always coefficient $s = 0$ ($w = 1$), while if the derived allele was favorable (population 1) the coefficient was $s = -0.15$ (if $h = 1$, $w = 1.15$); and if the derived was non-favorable (population 2) then $s = 0.15$ (if $h = 1$, $w = 0.85$). Within each population the mating was at random and both populations were connected by migration with $Nm = 10$ migrants per generation. The position of the selective locus was at the center (relative position 0.5) of the chromosome, although some extra cases (positions 0, 0.01, 0.1 and 0.25) were performed to assess the effect of locating the selective locus between the extreme and the center of the chromosome.

Concerning other evolutionary parameters such as population mutation $\theta = 4N\mu$ and recombination $\rho = 4Nr$ rates, we simulated different scenarios (Table 1). The population selection rate $\alpha = 4Ns$ was 0 (neutral cases), 600 (weak selection) or 6000 (strong selection). It was assumed that the long-term simulations, number of generations $t = 5N$ or $10N$, reached equilibrium. Due to computational efficiency, strong selection was studied only for the short-term (non-equilibrium) cases. For each selective case assayed, 1000 replicates of the corresponding neutral case were also run. After the evolutionary process finished, 50 sequences were sampled from each population. These sequences were analyzed using an in-house C++ implementation of the methods Svd and SvdM (available upon request to AC-R) and the program OmegaPlus [32], in order to compare the behavior of the distinct methods for detecting natural selection in a local adaptation scenario.

Table 1. Parameter values in the simulations.

N	t	s	θ	ρ
1000	100	0	12	0
10000	500	± 0.15	60	4
	5000			12
	10000			60

N: Population size; *t*: number of generations; *s*: coefficient of selection; θ : mutation rate; ρ : recombination rate.

Phasing Errors

To check the robustness of the methods to inaccuracy in the haplotype phase, some simulations ($t = 10N$) were performed as follows: we compared the number of SNPs between each sample of individuals (diploids) and a sample of gametes from such individuals. There is always a possibility that some SNPs were lost when sampling the gametes (because the rare allele was in the discarded chromosome). In the case that a SNP was lost in this way, the state of the allele was changed in order to recover the SNP. Therefore, the same SNP number as in the individuals' sample, was main-

tained at the prize of introducing phasing error in some of the gametes. In equilibrium populations, this implied error percentages of about 2% in most of the haplotypes and up to 20% in a few of them.

Data Analysis and Statistical Significance

As the simulated data correspond to two populations, different analyses can be performed. For example, the populations can be analyzed separately and afterwards the data can be joined to just evaluate one metapopulation. In the latter we can study both, every SNP or just those that are shared by the two populations. The difference will depend on the rate of gene flow. In our case the results were very similar so, when considering the metapopulation scenario, we focused on the whole set of SNPs. Hence, as already mentioned, the maximum value of the statistics was considered but the average of positive values (since the neutral expectation is negative or zero) or indeed a combination of maximum and average of positives, could also have been used.

Neutrality was rejected when the value obtained with the statistic (Svd, SvdM or OmegaPlus) was higher than a critical value. We used the 95th percentile of each statistic under the simulated neutral scenarios as threshold value. To study the neutral data with the HAC-based methods we fixed the window size to the value obtained when previously analyzed the candidate data under the automatic sliding window mode (which in our implementation uses as window size that giving the maximum value of the statistic).

RESULTS

Long-term Simulations: Detection of Weak Selection

Under equilibrium conditions ($t = 10,000$) the results of both populations were very similar, so we present only the results for population 1 (Table 2). The results with $t = 5,000$ also had a similar pattern as that shown in (Table 2). The HAC-based methods (Svd and SvdM) show acceptable performance with detection power of about 60-79%. An exception corresponds to the non-recombinant cases where performance is quite poor. Recall (see Material and Methods) that we force the selective allele to be present in the first twenty generations which in general suffices to avoid its loss by random drift during the initial steps of the evolutionary process. Note however that, in non-recombinant cases, it is easier that haplotypes carrying the selective variant became quickly fixed or even lost during the evolutionary process, which explains the worst performance in this case. Regarding the window size, a clear relationship appears with respect to the mutation rate. The lower the mutation rate the lower the window size that produces the highest HAC statistic and vice versa. In fact, when increasing five times the mutation rate, the window size expands approximately by this same factor (e.g. 411/81 for θ 60/12 in Table 2). The effect of recombination is not so clear although a pattern of slightly more detection under the higher recombination rate seems to occur, especially for SvdM.

The data were also analyzed with the OmegaPlus program using different parameterizations and finally choosing the one giving the best results (last column in Table 2). However, not surprisingly, the performance was not good with, at

Table 2. Performance for the different selection detection methods in the long-term ($t = 10,000$) and weak selection ($\alpha = 600$) cases. Data from population 1 (initial frequency of favored allele 1/1000). Values correspond to percentages of detection through replicates. Threshold for each test is between brackets.

S	θ	ρ	%Svd	%SvdM	%Omega
81	12	0	35 (1.43)	27 (2.03)	4.86 (8.42)
82	12	4	70 (0.67)	61 (0.86)	4.60 (8.46)
78	12	12	67 (0.45)	69 (0.47)	4.32 (9.40)
411	60	0	44.5 (6.64)	37.5 (9.55)	19.84 (164.84)
408	60	4	79 (3.51)	71 (4.28)	14.58 (212.08)
374	60	60	63 (0.84)	78 (0.57)	23.95 (232.36)

S: Window size in the selective case for the HAC-based methods. θ : Mutation rate. ρ : Recombination rate. OmegaPlus conditions: Grid 1000; Minwin 1000; Maxwin 20000.

the best, powers of about 20%. This occurs because of two reasons: first, the selection corresponds to $\alpha = 600$ while OmegaPlus requests for values about ten times higher [33] and second, the timing of the selective sweep may not imprint a clear footprint in this kind of scenarios as needed by OmegaPlus. The performance under short-term conditions was not better, again as expected, because in such scenarios the selective sweep is at its very initial steps. Thus, from herein we skip the LD-based method and just focus on the HAC-based ones.

As stated in the Material and Methods section, we have tentatively checked the effect of haplotype phasing inaccuracy on these data. Our findings show that low average error percentages of 2-5% had no qualitative effect as we obtain values very similar to those in (Table 2). Additionally, singletons were discarded, as it has been already recommended for the Svd method [30].

When the two populations were considered as one (metapopulation scenario) the statistic SvdM performs quite well (Table 3). However, Svd performs worse than under the two-population scenario. This is not surprising since SvdM is more sensitive to intermediate frequencies and we may expect intermediate frequencies for some loci when considering jointly two diverging populations. Note that the relationship between the window size (S) and the mutation rate (θ) still holds and a five-fold increase in the mutation rate produces the corresponding expansion in the window size. The effect of recombination seems clearer now, showing that the higher its rate, the higher the power of detection.

Short-term Simulations: Detection of Weak and Strong Selection

Under these conditions (100 or 500 generations) there is an ongoing conflict between the increase in frequency of the favored allele in population 1 while arriving the maladaptive one from population 2. At the same time in population 2, the favored allele decreases in frequency due to an increasing gene flow of the deleterious one from population 1. Thus, the detection of selection occurs only in population 1 where the favored allele is increasing in frequency. The percentages of detection in population 2 are below 5-10% and, therefore, we are giving results only from population 1 (except when the metapopulation setting is considered).

Table 3. Performance for the HAC-based selection detection methods in the long-term ($t = 10,000$) and weak selection ($\alpha = 600$) cases. Data from populations 1 and 2 joined (metapopulation scenario). Values correspond to percentages of detection through replicates. Threshold for each test is between brackets.

S	θ	ρ	%Svd _{metapop}	%SvdM _{metapop}
103	12	0	19 (0.64)	42 (1.0)
102	12	4	40 (0.38)	64 (0.54)
97	12	12	50 (0.22)	80 (0.29)
523	60	0	11 (3.80)	43 (5.83)
511	60	4	39 (2.70)	75 (2.02)
449	60	60	69 (0.40)	94 (0.40)

S: Window size in the selective case. θ : Mutation rate. ρ : Recombination rate.

The previously observed relationship between the ratio of the window size and the mutation rate does not longer hold. Now, more complex interactions appear involving also the recombination rate (Tables 4 and 5). When weak selection and low mutation rate are considered under the first 100 generations, the detection is at best of 60% under the metapopulation scenario (Table 4). This is not surprising since there is too little time and not enough variation for the effect of selection to leave the adequate footprint in the sequences. With higher mutation rate and some recombination, the percentages are slightly improved, especially for SvdM. This fact is expected because SvdM is more sensitive to medium-low frequencies of the favored allele.

The situation does not get much better when $t = 500$. Although it improves for the cases with recombination and higher mutation rate resulting in detection power about 60-70% (Table 5). Additionally, in the case without recombination, Svd performs clearly better than SvdM. This is probably due to the extreme linkage favoring higher frequencies or the opposite, the eventual loss of alleles during the evolutionary process. The SvdM method is more susceptible to this situation since the medium-low frequency alleles have more probability to be lost.

Table 4. Performance for the HAC-based selection detection methods in the very short-term ($t = 100$) and weak selection ($\alpha = 600$) cases. Data from population 1 (initial frequency of favored allele 1/1000) or from the metapopulation (last column). Values correspond to percentages of detection through replicates. Threshold for each test is between brackets.

S	θ	ρ	% Svd	%SvdM	%SvdM _{metapop}
48	12	0	15 (0.22)	54 (0.13)	59 (0.08)
50	12	4	41 (0.16)	40 (0.19)	37 (0.13)
33	12	12	35 (0.11)	26 (0.14)	21 (0.10)
104	60	0	14 (0.17)	5 (0.26)	9 (0.21)
113	60	4	46 (0.18)	59 (0.17)	37 (0.13)
182	60	60	45 (0.31)	57 (0.27)	60 (0.18)

S: Window size in the selective case. θ : Mutation rate. ρ : Recombination rate.

Table 5. Performance for the HAC-based selection detection methods in the short-term ($t = 500$) and weak selection ($\alpha = 600$) cases. Data from population 1 (initial frequency of favored allele 1/1000) or from the metapopulation (last column). Values correspond to percentages of detection through replicates. Threshold for each test is between brackets.

S	θ	ρ	% Svd	%SvdM	%SvdM _{metapop}
38	12	0	46 (0.17)	5 (0.27)	6 (0.28)
31*	12	4	8 (0.15)	3 (0.21)	3 (0.14)
36	12	12	58 (0.15)	36 (0.19)	21 (0.16)
125	60	0	43 (0.20)	21 (0.22)	20 (0.41)
153	60	4	79 (0.15)	70 (0.22)	59 (0.18)
165	60	60	69 (0.22)	67 (0.19)	61 (0.15)

S: Window size in the selective case. θ : Mutation rate. ρ : Recombination rate. *: in the selective case only 37 replicates having a minimum of 25 SNPs.

To explore whether such results would change if selection is stronger, we studied the cases with $\theta = 60$ increasing the selection pressure up to 10 times ($\alpha = 6000$, Table 6). Not surprisingly the percentage of detection is, in general, higher for both methods. However, when recombination is absent, a strange pattern of detection appears with better detection after 100 generations that seems to vanish under 500 generations. This may be caused by a lack of variability since only half of the runs had a minimum number of 25 SNPs we required for the statistics to be applied. For the rest of the cases ($t = 100$ or 500 , $\rho = 4$ or 60) SvdM performed quite well, especially when recombination is high ($\rho = 60$).

Robustness and False Positives

The misspecification of the neutral distribution is a general issue for any selection detection method that needs to simulate the neutral demography. In the case of HAC-based algorithms, we have considered both robustness and the rate of false positives.

Robustness

What happens if the neutral distribution is misspecified by assuming e.g. $\theta = 60$ for a data set that in fact corresponds to $\theta = 12$? By comparison of rows 2 and 5 from (Table 2), it can be appreciated that for low recombination rate ($\rho = 4$) the critical values are very different whether θ is 12 (critical

values 0.7 – 0.9) or 60 (critical values 3.5 – 4). Thus, if the second set of critical values was used with a data set that corresponds to $\theta = 12$, the researcher may have very low detection power (less than 3% in this example). An obvious solution is to estimate adequately the θ parameter. However, this is not always possible as it could be necessary to estimate different parameter values and may have wrong or not enough information at hand. Fortunately there is a simple and easy solution that would work even if the neutral model is misspecified; we can take advantage on the strong dependence of the HAC statistics on the window size. In (Table 2) we see that the optimal window size was also very different depending on the θ parameter ($S = 82$ vs. 408). The strategy proposed all through this work consisted of computing the neutral distribution fixing the window size to the value obtained when computing the HAC statistics for the problem data. And the same applies for this case. When computing the HAC-statistics for the neutral distribution, the window size obtained for the candidate data should be applied. If we do so, the method seems quite robust to the parameter misspecification (80% of detection with the misspecified neutral model, i.e. using $\theta = 60$ instead of 12, in this example).

False Positives

The drawback of getting false positives when using the HAC-based statistics is the reverse to that exposed above for

Table 6. Performance for the HAC-based selection detection methods in the short-term ($t = 100-500$) and strong selection ($\alpha = 6000$) cases. Data from population 1 (initial frequency of favored allele 1/1000) or from the metapopulation (last column). Values correspond to percentages of detection through replicates. Threshold for each test is between brackets.

S	t	θ	ρ	% Svd	%SvdM	%SvdM _{metapop}
155	100	60	0	76 (0.09)	81 (0.15)	61 (0.10)
73	500	60	0	28 (0.08)*	15 (0.16)*	14 (0.15)
198	100	60	4	30(0.45)	81.5 (0.19)	82.5 (0.12)
40	500	60	4	63 (0.038)	65 (0.035)	59 (0.07)
194	100	60	60	30 (0.375)	76 (0.19)	78.5 (0.11)
67	500	60	60	90 (0.14)	96 (0.09)	70 (0.13)

S: Window size in the selective case. t: number of generations. θ : Mutation rate. ρ : Recombination rate. *: only 46 runs having a minimum of 25 SNPs.

robustness. Imagine a candidate set of neutral data so that it could not be real selection at work. However, under some settings, these data can give a high HAC-based optimum e.g. under a window size of 408 as computed for the neutral cases corresponding to $\theta = 60$ and $\rho = 4$ (Table 2). In this case the values of the statistics are higher than zero (mean value 1-1.5 and 95th percentile 3.5 – 4 for Svd and SvdM respectively). Therefore, consider that this neutral set is the problem data and hence the neutral demography is misspecified to be $\theta = 12$, $\rho = 4$. The first concern the researcher has to face occurs because the analysis returns a window size of 408 and when trying to simulate a neutral demography under low mutation rate, $\theta = 12$, most runs will not reach enough number of SNPs to define such window size. Thus, the automatic optimal window size must be computed for that spurious neutral demography. When doing so, an average window size value of $S = 73$ is obtained. Now, if the critical values from this simulated neutral demography are used to analyze the candidate (neutral) data (window size $S = 408$), a large number of false positives is obtained (60-65% in this example). The solution is, as before, to use the same window size for both the problem data and the neutral distribution. Then, if HAC is computed for the candidate data using $S = 73$, the false positive rate is far below 1%. Thus, as a rule of thumb, if we are not able of computing the HAC neutral distribution with the window size provided by the candidate data, we can get the optimal window size for the neutral and compute again the HAC using this new window size. In doing so we are going to be protected from false positives even under model misspecification. Note however that now, due to misspecification of the neutral model, it is necessary to use the simulated neutral data to fix the window size for the problem data.

Positional Effect of the Selective Site

To finish this work, we raise the question about if the position of the selective site in the genome affects the ability of the HAC-methods to detect selective patterns. This kind of question has been rarely considered because a common approach is to put the candidate at the center of the genome (but see [40]). Thus we focused on two matters: first, to study the power of detection and second, to investigate how well is the selective effect localized in the genome depending

on the position. Concerning the detection power, no clear effect was noticed in general. However, under equilibrium conditions and maximum mutation and recombination rates ($\theta = \rho = 60$) there is a positive relationship between the position and the power of detection (Fig. 1). A similar pattern was also detected under non-equilibrium conditions with $\theta = 60$ and $\rho \geq 4$. The maximum power occurs when the selective position is located in the middle and diminishes as the position approach the extremes of the chromosome.

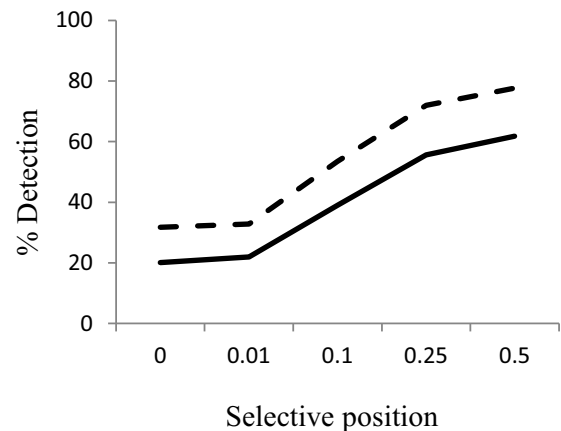


Fig. (1). Power to detect selection under equilibrium conditions ($t = 10,000$) when the selected locus is localized at different positions in the genome. Svd: continuous line. SvdM: dashed line. $S = 390$. $\alpha = 600$. $\theta = \rho = 60$.

Regarding the second question, that is, how well is the localization of the selective site achieved? It is worth mentioning that under neutral conditions, the maximum Svd and SvdM values are expected to be localized at random in the genome so that the distribution of positions is uniform between 0 and 1. Then the expected mean and variance of the maximum HAC-value under neutral conditions are 0.5 and 0.08 respectively. This is exactly what we find in the neutral simulation data (not shown). Therefore, the evaluation of the ability of the methods to locate the selective site is not easy if the site was fixed at the center of the chromosome. We find a similar pattern for both Svd and SvdM statistics. In (Table 7) the absolute value of the difference between the real candidate position and that estimated by SvdM is given.

This difference is called D_{sel} for the selective data and D_{neu} for the neutral one. It can be appreciated that, for sites located at the center of the genome ($P_s = 0.5$), both values D_{sel} and D_{neu} tend to be zero, as expected, because of the above mentioned uniform distribution effect. However, for other localizations (P_s from 0 to 0.25), the real selective position is best localized for the highest recombination cases ($\rho = 60$). Note that in neutral cases, the value of D_{neu} tends to be the difference between the real position (P_s) and the central position. This is expected if, as mentioned, the distribution of positions of the maximum value of the statistic has mean 0.5. For the selective cases, the best localization is attained when $P_s = 0.25$ with a difference of 50Kb ($D_{sel} = 0.05$) with respect to the real one.

Table 7. Difference (D_{sel}) between the real position of the selective site and the localization of the maximum SvdM value. And the same difference in the neutral case (D_{neu}). Populations in equilibrium ($t = 10,000$), $\alpha = 600$, $\theta = 60$.

ρ	P_s	D_{sel}	D_{neu}
0	0	0.448	0.471
4		0.400	0.469
60		0.294	0.494
0	0.01	0.467	0.460
4		0.413	0.462
60		0.283	0.486
0	0.1	0.354	0.374
4		0.301	0.370
60		0.164	0.392
0	0.25	0.219	0.221
4		0.164	0.220
60		0.051	0.246
0	0.5	0.042	0.028
4		0.009	0.030
60		0.012	0.006

DISCUSSION

In the present work we studied a genome-wide sequence-based method for detecting divergent selection with gene flow in non-model organisms. Currently, one of the limitations of the sliding window selection detection methods is to fix the window size. This should be considered because both computational efficiency and the results (robustness and false positives), may be highly compromised by this choice [30, 32]. Here we contribute to solve this question by automating the process of deciding the window size just by choosing the one that gives the maximum value of the statistic for a given data set. Our results indicate that the HAC-

based methods and specifically the one we propose (SvdM) are quite powerful to detect the footprint of moderate divergent selection in presence of gene flow. They are also robust to reasonable model misspecification.

A LD-based detection method has also been tested. The results show that the LD-based omega statistic is not adequate for the kind of scenario assayed, probably due to the intermediate-low level of selection involved and because the selective sweep is incomplete. The latter violates the assumptions of this method [33]. Moreover, the positive selection process may be hidden because it involves two connected populations undergoing selection in opposite directions. For the two HAC-based statistics, the original (Svd [30]) and the new (SvdM) performed similarly, although SvdM seems best fitted for this kind of scenario especially in the long term (Tables 2 and 3). In this case, we were able to detect selection for the different mutation ($\theta = 12$ or 60) and recombination ($\rho = 4, 12$ or 60) values with power between 70-80% or even more when using the metapopulation scenario. In general both HAC-methods worked reasonably well provided that enough variation is at hand. The obvious advantage of SvdM is that it does not require knowledge of the allelic state.

Regarding the short-term scenario, the signal of selection was not detected in population 2. The explanation is that positive selection is detected on the basis of an increase in allelic frequency. In the studied scenario, in population 2, the favored allele is already at maximum frequency so selection cannot be detected at the beginning of the process. Only when the gene flow of the maladapted allele coming from the neighbor population has diminished the frequency of the well-adapted, can selection to be detected as it occurs on the long-term simulations. This could be a problem when working with real data. A promising solution is to perform the analysis at the metapopulation level since in this case there is more power to detect selection in both populations (Tables 4, 5 and 6).

The characterization of patterns of genome divergence is not easy. A particular genome footprint can arise in different ways due to combinations of distinct evolutionary forces. In addition, the setting of an appropriate neutral model is far from easy [41]. We have shown how to apply the HAC-based methods in a robust manner by using a step-wise approach. Basically, we analyze the candidate-data to compute the statistics and fix the window size that will be used in a second step to simulate the neutral distribution in order to get an adequate statistical threshold. When working in this way, the effect of reasonable neutral model misspecification seems to be minimized and the same is true for the type I error.

One remaining issue of sequence-based methods applied at a large scale, is the high impact of the recombination rate on the genomic patterns and, consequently, in the statistics. On the one hand, the combination of selection and migration favors the reduction of recombination to preserve stable clusters of linked alleles [7, 12]. This is not the case in the present work because we defined only one selective site. In fact, under equilibrium conditions, extreme differences as $\rho = 4$ or 60 have little impact on the percentage of detection. On the other hand, the interaction between population subdivision

and selective sweeps critically depends on the recombination rate [41-43]. Indeed recombination seems to have an important effect on the critical thresholds (see Table 2), so that a large misspecification of the recombination rate could produce the failure of the tests. Fortunately, improved methods are coming to allow the precise estimation of recombination from genome wide data [44].

Concerning the migration rate, we have used a value ($Nm = 10$) which is the expected in the ecotype model of *L. saxatilis* [45]. Given this value and the allele effect ($s = 0.15$), the gene flow is far below the migration threshold [12, 42]. Thus, it is not surprising that the favorable allele spreads in population 1 overcoming the homogenizing effect of migration.

We have also studied the impact of the position of the selective site onto the detection methods. The relative position between selected and neutral sites has great impact on the LD generated by a selective sweep [31]. We were able to localize the selective site when the real position is not in the middle. At best we localized the candidate site in a range of 50Kb with respect to the real one, which is not a bad result for a total chromosome length of 1Mb in a sample size of only 50 sequences [22]. Noticeably, a point is raised about the fact that the precise localization when the selective site is at the middle of the chromosome (Table 7) might be an artifact, since the real position coincides with that expected under the neutral distribution of the statistics. Clearly, the ability to precisely localize the selective positions still needs further improvement [6].

Currently, there is controversy about distinguishing between divergent or directional selection. For example, the LD-based tests can be improved when combined with outlier methods in order to make the distinction between both types of natural selection [46]. In the present work, we are able to discriminate between directional versus (bidirectional) divergent selection at least in the long-term cases, because we detect the signature of selection in both populations separately which is not expected if the selection is acting only in one of the populations. Additionally, the HAC-based tests seem to take advantage of the unusual patterns of variation that the selective sweeps may produce apart from its influence in LD [31]. Unfortunately, it is still necessary to simulate the empirical neutral distribution to assess significance. The development of an automatic test is work in progress.

Finally, the modification of the statistic that we propose may be useful for detecting clues of divergent selection in the chromosomes of non-model organisms. There are various algorithms that already combine sliding-window approaches with F_{ST} ones [23]. It might be of interest to further include the HAC-based methods for detecting the footprint of divergent selection in the sequences at genomic level. Of course it is important not to forget that the ultimate goal is going from the DNA patterns to the evolutionary scenario, where the causes and effects of changes in the genomes can be modeled in order to gain a better knowledge of the mechanisms of evolution and life [7, 23]. In this way, we think that HAC-based methods may help to the progress of the study of the evolutionary processes that underlie the patterns of genomic differentiation.

CONFLICT OF INTEREST

The author(s) confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

We thank J. Hussin who kindly provided the original Svd program in Java and E. Rolán-Alvarez for 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. All authors declare to have no conflict of interest.

REFERENCES

- [1] Sabeti, P.C.; Varilly, P.; Fry, B.; Lohmueller, J.; Hostetter, E.; Cotsapas, C.; Xie, X.; Byrne, E. H.; McCarroll, S. A.; Gaudet, R.; Schaffner, S. F.; Lander, E. S.; Frazer, K. A.; Ballinger, D. G.; Cox, D.R.; Hinds, D.A.; Stuve, L. L.; Gibbs, R. A.; Belmont, J. W.; Bou-dreau, A.; Hardenbol, P.; Leal, S. M.; Pasternak, S.; Wheeler, D. A.; Willis, T. D.; Yu, F.; Yang, H.; Zeng, C.; Gao, Y.; Hu, H.; Hu, W.; Li, C.; Lin, W.; Liu, S.; Pan, H.; Tang, X.; Wang, J.; Wang, W.; Yu, J.; Zhang, B.; Zhang, Q.; Zhao, H.; Zhou, J.; Gabriel, S. B.; Barry, R.; Blumenstiel, B.; Camargo, A.; Defelice, M.; Faggart, M.; Goyette, M.; Gupta, S.; Moore, J.; Nguyen, H.; Onofrio, R. C.; Parkin, M.; Roy, J.; Stahl, E.; Winchester, E.; Ziaugra, L.; Altshuler, D.; Shen, Y.; Yao, Z.; Huang, W.; Chu, X.; He, Y.; Jin, L.; Liu, Y.; Shen, Y.; Sun, W.; Wang, H.; Wang, Y.; Wang, Y.; Xiong, X.; Xu, L.; Waye, M. M.; Tsui, S. K.; Xue, H.; Wong, J. T.; Galver, L. M.; Fan, J. B.; Gunderson, K.; Murray, S. S.; Oliphant, A. R.; Chee, M. S.; Montpetit, A.; Chagnon, F.; Ferretti, V.; Leboeuf, M.; Olivier, J. F.; Phillips, M. S.; Roumy, S.; Sallee, C.; Verner, A.; Hudson, T. J.; Kwok, P. Y.; Cai, D.; Koboldt, D. C.; Miller, R. D.; Pawlikowska, L.; Taillon-Miller, P.; Xiao, M.; Tsui, L. C.; Mak, W.; Song, Y. Q.; Tam, P. K.; Nakamura, Y.; Kawaguchi, T.; Kitamoto, T.; Morizono, T.; Nagashima, A.; Ohnishi, Y.; Sekine, A.; Tanaka, T.; Tsunoda, T.; Deloukas, P.; Bird, C. P.; Delgado, M.; Dermizakis, E. T.; Gwilliam, R.; Hunt, S.; Morrison, J.; Powell, D.; Stranger, B. E.; Whittaker, P.; Bentley, D. R.; Daly, M. J.; de Bakker, P. I.; Barrett, J.; Chretien, Y. R.; Maller, J.; McCarroll, S.; Patterson, N.; Pe'er, I.; Price, A.; Purcell, S.; Richter, D. J.; Sabeti, P.; Saxena, R.; Schaffner, S. F.; Sham, P. C.; Varilly, P.; Altshuler, D.; Stein, L. D.; Krishnan, L.; Smith, A. V.; Tello-Ruiz, M. K.; Thorisson, G. A.; Chakravarti, A.; Chen, P. E.; Cutler, D. J.; Kashuk, C. S.; Lin, S.; Abecasis, G. R.; Guan, W.; Li, Y.; Munro, H. M.; Qin, Z. S.; Thomas, D. J.; McVean, G.; Auton, A.; Bottolo, L.; Cardin, N.; Eyheramendy, S.; Freeman, C.; Marchini, J.; Myers, S.; Spencer, C.; Stephens, M.; Donnelly, P.; Cardon, L. R.; Clarke, G.; Evans, D. M.; Morris, A. P.; Weir, B. S.; Tsunoda, T.; Johnson, T. A.; Mullikin, J. C.; Sherry, S. T.; Feolo, M.; Skol, A.; Zhang, H.; Zeng, C.; Zhao, H.; Matsuda, I.; Fukushima, Y.; Macer, D. R.; Suda, E.; Rotimi, C. N.; Adebamowo, C. A.; Ajayi, I.; Aniagwu, T.; Marshall, P. A.; Nkwodimma, C.; Royal, C. D.; Leppert, M. F.; Dixon, M.; Peiffer, A.; Qiu, R.; Kent, A.; Kato, K.; Niikawa, N.; Adewole, I. F.; Knoppers, B. M.; Foster, M. W.; Clayton, E. W.; Watkin, J.; Gibbs, R. A.; Belmont, J.W.; Muzny, D.; Nazareth, L.; Sodergren, E.; Weinstock, G.M.; Wheeler, D.A.; Yakub, I.; Gabriel, S. B.; Onofrio, R. C.; Richter, D. J.; Ziaugra, L.; Birren, B. W.; Daly, M. J.; Altshuler, D.; Wilson, R. K.; Fulton, L.L.; Rogers, J.; Burton, J.; Carter, N. P.; Clee, C.M.; Griffiths, M.; Jones, M. C.; McLay, K.; Plumb, R. W.; Ross, M. T.; Sims, S. K.; Willey, D. L.; Chen, Z.; Han, H.; Kang, L.; Godbout, M.; Wallenburg, J.C.; L'Archeveque, P.; Bellemare, G.; Saeki, K.; Wang, H.; An, D.; Fu, H.; Li, Q.; Wang, Z.; Wang, R.; Holden, A. L.; Brooks, L. D.; McEwen, J. E.; Guyer, M. S.; Wang, V. O.; Peterson, J. L.; Shi, M.; Spiegel, J.; Sung, L. M.; Zacharia, L. F.; Collins, F.S.; Kennedy, K.; Jamieson, R.; Stewart, J. Genome-wide detection and characterization of positive selection in human populations. *Nature*, **2007**, 449, (7164), 913-8.

- [2] Sawyer, S.A.; Parsch, J.; Zhang, Z.; Hartl, D.L. Prevalence of positive selection among nearly neutral amino acid replacements in *Drosophila*. *Pr. Nat. Acad. Sci. USA*, **2007**, *104*(16), 6504-6510.
- [3] Andrews, K.R.; Luikart, G. Recent novel approaches for population genomics data analysis. *Mol. Ecol.*, **2014**, *23* (7), 1661.
- [4] Ellegren, H. Genome sequencing and population genomics in non-model organisms. *Tr. Ecol. Evol.*, **2014**, *29* (1), 51.
- [5] Buerkle, C.A.; Gompert, Z.; Parchman, T.L. The $n = 1$ constraint in population genomics. *Mol. Ecol.*, **2011**, *20* (8), 1575.
- [6] Li, J.; Li, H.; Jacobsson, M.; Li, S.; Sjödin, P.; Lascoux, M. Joint analysis of demography and selection in population genetics: where do we stand and where could we go? *Mol. Ecol.*, **2012**, *21*, 28-44.
- [7] Nosil, P.; Feder, J.L. Genome evolution and speciation: toward quantitative descriptions of pattern and process. *Evolution*, **2013**, *67* (9), 2461.
- [8] Roesti, M.; Gavrilets, S.; Hendry, A. P.; Salzburger, W.; Berner, D. The genomic signature of parallel adaptation from shared genetic variation. *Mol. Ecol.*, **2014**, *23* (16), 3944-3956.
- [9] Seehausen, O.; Butlin, R.K.; Keller, I.; Wagner, C.E.; Boughman, J. W.; Hohenlohe, P.A.; Peichel, C.L.; Saetre, G.-P.; Bank, C.; Brannstrom, A.; Brelsford, A.; Clarkson, C. S.; Eroukmanoff, F.; Feder, J.L.; Fischer, M.C.; Foote, A.D.; Franchini, P.; Jiggins, C.D.; Jones, F. C.; Lindholm, A. K.; Lucek, K.; Maan, M. E.; Marques, D. A.; Martin, S. H.; Matthews, B.; Meier, J. I.; Most, M.; Nachman, M. W.; Nonaka, E.; Rennison, D. J.; Schwarzer, J.; Watson, E.T.; Westram, A. M.; Widmer, A. Genomics and the origin of species. *Nat. Rev. Genet.*, **2014**, *15*(3), 176.
- [10] Michel, A.P.; Sim, S.; Powell, T.H.Q.; Taylor, M.S.; Nosil, P.; Feder, J.L. Widespread genomic divergence during sympatric speciation. *Pr. Nat. Acad. Sci.*, **2010**, *107* (21), 9724-9729.
- [11] Nosil, P.; Funk, D. J.; Ortiz-Barrientos, D. Divergent selection and heterogeneous genomic divergence. *Mol. Ecol.*, **2009**, *18*(3), 375.
- [12] Yeaman, S.; Whitlock, M.C. The Genetic Architecture of Adaptation under Migration-Selection Balance. *Evolution* **2011**, *65*(7), 1897-1911.
- [13] Beaumont, M. A.; Nichols, R. A. Evaluating loci for use in the genetic analysis of population structure. *Pr. Royal Society B-Biol. Sci.*, **1996**, *263*(1377), 1619-1626.
- [14] Bonhomme, M.; Chevalet, C.; Servin, B.; Boitard, S.; Abdallah, J.; Blott, S.; SanCristobal, M. Detecting Selection in Population Trees: The Lewontin and Krakauer Test Extended. *Genetics*, **2010**, *186*(1), 241-262.
- [15] Lewontin, R.C.; Krakauer, J. Distribution of Gene Frequency as a Test of Theory of Selective Neutrality of Polymorphisms. *Genetics*, **1973**, *74*(1), 175-195.
- [16] De Mita, S.; Thuillet, A.-C.; Gay, L.; Ahmadi, N.; Manel, S.; Ronfort, J.; Vigouroux, Y. Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Mol. Ecol.*, **2013**, *22*(5), 1383.
- [17] Endler, J. A. *Natural selection in the wild*. Princeton University Press: Princeton, N.J., **1986**; p xiii, 336 s.
- [18] de Villemereuil, P.; Frichot, É.; Bazin, É.; François, O.; Gaggiotti, O. E. Genome scan methods against more complex models: when and how much should we trust them? *Mol. Ecol.*, **2014**, *23*(8), 2006.
- [19] Perez-Figueroa, A.; Garcia-Pereira, M. J.; Saura, M.; Rolan-Alvarez, E.; Caballero, A. Comparing three different methods to detect selective loci using dominant markers. *J. Evolution. Biol.*, **2010**, *23*(10), 2267-76.
- [20] Vilas, A.; Perez-Figueroa, A.; Caballero, A.A simulation study on the performance of differentiation-based methods to detect selected loci using linked neutral markers. *J. Evolution. Biol.*, **2012**, *25*(7), 1364-1376.
- [21] Crisci, J. L.; Poh, Y.-P.; Mahajan, S.; Jensen, J. D. The Impact of Equilibrium Assumptions on Tests of Selection. *Front. Genetics*, **2013**, *4*, 1.
- [22] Pavlidis, P.; Zivkovic, D.; Stamatakis, A.; Alachiotis, N. SweeD: Likelihood-Based Detection of Selective Sweeps in Thousands of Genomes. *Mol. Biol. Evol.*, **2013**, *30*(9), 2224-2234.
- [23] Vitti, J.J.; Grossman, S.R.; Sabeti, P.C. Detecting Natural Selection in Genomic Data. *Ann. Rev. Genet.*, **2013**, *47*(1), 97-120.
- [24] Feder, J.L.; Egan, S. P.; Nosil, P. The genomics of speciation-with-gene-flow. *Tr. Genet.*, **2012**, *28*(7), 342.
- [25] Savolainen, O.; Lascoux, M.; Merila, J. Ecological genomics of local adaptation. *Nat. Rev. Genet.*, **2013**, *14*(11), 807.
- [26] Stöltig, K.N.; Nipper, R.; Lindtke, D.; Caseys, C.; Waeber, S.; Castiglione, S.; Lexer, C. Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. *Mol. Ecol.*, **2013**, *22* (3), 842.
- [27] Yeaman, S.; Otto, S. P. Establishment and Maintenance of Adaptive Genetic Divergence under Migration, Selection, and Drift. *Evolution*, **2011**, *65*(7), 2123-2129.
- [28] Sabeti, P.C.; Reich, D.E.; Higgins, J.M.; Levine, H.Z.; Richter, D. J.; Schaffner, S.F.; Gabriel, S.B.; Platko, J.V.; Patterson, N. J.; McDonald, G. J.; Ackerman, H. C.; Campbell, S. J.; Altshuler, D.; Cooper, R.; Kwiatkowski, D.; Ward, R.; Lander, E. S. Detecting recent positive selection in the human genome from haplotype structure. *Nature*, **2002**, *419*(6909), 832-7.
- [29] Zeng, K.; Shi, S.; Wu, C.I. Compound tests for the detection of hitchhiking under positive selection. *Mol. Biol. Evol.*, **2007**, *24*(8), 1898-908.
- [30] Hussin, J.; Nadeau, P.; Lefebvre, J.-F.; Labuda, D. Haplotype allelic classes for detecting ongoing positive selection. *BMC Bioinformatics*, **2010**, *11*(1), 65.
- [31] McVean, G. The structure of linkage disequilibrium around a selective sweep. *Genetics* **2007**, *175*(3), 1395-406.
- [32] Alachiotis, N.; Stamatakis, A.; Pavlidis, P. OmegaPlus: a scalable tool for rapid detection of selective sweeps in whole-genome datasets. *Bioinformatics*, **2012**, *28*(17), 2274-2275.
- [33] Pavlidis, P.; Jensen, J. D.; Stephan, W. Searching for footprints of positive selection in whole-genome SNP data from nonequilibrium populations. *Genetics*, **2010**, *185*(3), 907-22.
- [34] Rolan-Alvarez, E. Sympatric speciation as a by-product of ecological adaptation in the Galician Littorina saxatilis hybrid zone. *J. Molluscan Studies*, **2007**, *73*(1), 1-10.
- [35] Bourret, V.; Kent, M. P.; Primmer, C. R.; Vasemägi, A.; Karlsson, S.; Hindar, K.; McGinnity, P.; Verspoor, E.; Bernatchez, L.; Lien, S. SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (*Salmo salar*). *Mol. Ecol.*, **2013**, *22*(3), 532.
- [36] Renault, S.; Nolte, A.W.; Bernatchez, L. Mining transcriptome sequences towards identifying adaptive single nucleotide polymorphisms in lake whitefish species pairs (*Coregonus* spp. Salmonidae). *Mol. Ecol.*, **2010**, *19*, 115.
- [37] Modesto, I. s.; Miguel, C. I.; Pina-Martins, F.; Glushkova, M.; Veloso, M.; Paulo, O. v.; Batista, D. Identifying signatures of natural selection in cork oak (*Quercus suber* L.) genes through SNP analysis. *Tree Genetics Genomes*, **2014**, *10*(6), 1645-1660.
- [38] Sadedin, S.; Hollander, J.; Panova, M.; Johannesson, K.; Gavrilets, S. Case studies and mathematical models of ecological speciation. 3: Ecotype formation in a Swedish snail. *Mol. Ecol.*, **2009**, *18*(19), 4006-23.
- [39] Carvajal-Rodriguez, A. GENOMEPOP: A program to simulate genomes in populations. *BMC Bioinformatics*, **2008**, *9*(1), 223.
- [40] Lin, K.; Li, H.; Schlotterer, C.; Futschik, A. Distinguishing positive selection from neutral evolution: boosting the performance of summary statistics. *Genetics*, **2011**, *187*(1), 229-44.
- [41] Cutter, A. D.; Payseur, B. A. Genomic signatures of selection at linked sites: unifying the disparity among species. *Nat. Rev. Genet.*, **2013**, *14* (4), 262.
- [42] Bierne, N. The distinctive footprints of local hitchhiking in a varied environment and global hitchhiking in a subdivided population. *Evolution*, **2010**, *64*(11), 3254.
- [43] Kim, Y.; Maruki, T. Hitchhiking Effect of a Beneficial Mutation Spreading in a Subdivided Population. *Genetics*, **2011**, *189*(1), 213-226.
- [44] Lopes, J. S.; Arenas, M.; Posada, D.; Beaumont, M. A. Coestimation of recombination, substitution and molecular adaptation rates by approximate Bayesian computation. *Heredity*, **2013**, *112*(3), 255.

- [45] Perez-Figueroa, A.; Cruz, F.; Carvajal-Rodriguez, A.; Rolan-Alvarez, E.; Caballero, A. The evolutionary forces maintaining a wild polymorphism of *Littorina saxatilis*: model selection by computer simulations. *J. Evolution. Biol.*, **2005**, *18*(1), 191-202.
- [46] Goicoechea, P.G.; Petit, R.J.; Kremer, A. Detecting the footprints of divergent selection in oaks with linked markers. *Heredity*, **2012**, *109*(6), 361.

Received on: November 16, 2014

Revised on: February 23, 2015

Accepted on: March 09, 2015