MOLECULAR ECOLOGY

Molecular Ecology (2015) doi: 10.1111/mec.13360

SPECIAL ISSUE: DETECTING SELECTION IN NATURAL POPULATIONS: MAKING SENSE OF GENOME SCANS AND TOWARDS ALTERNATIVE SOLUTIONS

Detection of selective sweeps in structured populations: a comparison of recent methods

ALEXANDRA I. VATSIOU,*† ERIC BAZIN* and OSCAR E. GAGGIOTTI*†

Laboratoire d'Ecologie Alpine, UMR CNRS 5553, Université Joseph Fourier, Grenoble, France, †Scottish Oceans Institute, East Sands, University of St Andrews, St Andrews KY16 8LB, UK

Abstract

Identifying genomic regions targeted by positive selection has been a long-standing interest of evolutionary biologists. This objective was difficult to achieve until the recent emergence of next-generation sequencing, which is fostering the development of large-scale catalogues of genetic variation for increasing number of species. Several statistical methods have been recently developed to analyse these rich data sets, but there is still a poor understanding of the conditions under which these methods produce reliable results. This study aims at filling this gap by assessing the performance of genome-scan methods that consider explicitly the physical linkage among SNPs surrounding a selected variant. Our study compares the performance of seven recent methods for the detection of selective sweeps (iHS, nSL, EHHST, xp-EHH, XP-EHHST, XPCLR and hapFLK). We use an individual-based simulation approach to investigate the power and accuracy of these methods under a wide range of population models under both hard and soft sweeps. Our results indicate that XPCLR and hapFLK perform best and can detect soft sweeps under simple population structure scenarios if migration rate is low. All methods perform poorly with moderate-to-high migration rates, or with weak selection and very poorly under a hierarchical population structure. Finally, no single method is able to detect both starting and nearly completed selective sweeps. However, combining several methods (XPCLR or hapFLK with iHS or nSL) can greatly increase the power to pinpoint the selected region.

Keywords: accuracy, genome-scan methods, haplotype structure, positive selection

Received 2 May 2015; revision received 27 July 2015; accepted 25 August 2015

Introduction

Population geneticists and evolutionary biologists have a long-standing interest in understanding the ecological and genetic mechanisms that allow species to adapt to local environmental conditions. The recent advent of next-generation sequencing (NGS) (Shendure & Ji 2008) and the high density SNP arrays it generates has allowed rapid advances in this field and has fostered the emergence of the population genomics approach (Luikart *et al.* 2003). This new paradigm is focused on the use of genomewide data to distinguish between locus-specific effects (mainly selection but also mutation, and recombination) and genomewide effects such as genetic drift. It has proven particularly useful to detect signatures of selection and has been used to uncover genes involved in local adaptation, disease susceptibility, resistance to pathogens and other phenotypic traits of interest to plant and animal breeders.

At the genetic level, local adaptation involves a process whereby directional selection induced by local environmental conditions will favour the spread of genetic variants associated with beneficial phenotypic traits. If selection is strong at the level of an individual locus, the selected variant will increase in frequency. Additionally, selection will modify the pattern of diversity around the selected locus through genetic hitchhiking (Smith & Haigh 1974; Barton 2000). This process, known as a selective sweep, has been extensively

Correspondence: Oscar E. Gaggiotti, Fax: +44 1334 463443; E-mail: oeg@st-andrews.ac.uk

studied using models of isolated populations (Smith & Haigh 1974; Sabeti *et al.* 2002; Kim & Nielsen 2004; Hermisson & Pennings 2005; Pennings & Hermisson 2006a, b; Voight *et al.* 2006) but much less studied under structured population scenarios. In this latter case, analyses focused on either an universally favoured mutation that spreads from its deme of origin to other demes (Slatkin & Wiehe 1998; Barton 2000; Bierne 2010) or on a scenario where the new selected variant is favoured in one part of the species range but counter-selected in the other half (Bierne 2010). However, there is a third scenario still poorly understood but frequently assumed by studies of local adaptation, particularly in humans. Under this scenario, a selected variant is favoured in one part of the species range and is neutral elsewhere (e.g. lactase persistence, skin pigmentation, high altitude adaptation; Jeong & Di Rienzo 2014).

Several so-called genome-scan methods have been proposed for the detection of positive selection from dense SNP maps. The most widely used and thoroughly evaluated type of methods is based on Lewontin & Krakauer (1973) approach and is focused on single-locus F_{ST} (Beaumont & Nichols 1996; Beaumont & Balding 2004; Foll & Gaggiotti 2008). These methods implicitly or explicitly assume that SNPs are physically unlinked and are most effective when neutral genetic differentiation is low (Price *et al.* 2008) and/or when the selective sweep is close to fixation (Pickrell *et al.* 2009). Other methods are specifically aimed at detecting selective sweeps by focusing on the distribution of genetic variation along a chromosome within a population when selection is acting, as predicted by the theory of genetic hitchhiking (Fay & Wu 2000; Kim & Stephan 2002; Nielsen *et al.* 2005). These methods are applicable to isolated populations, and their behaviour has been extensively studied (Jensen *et al.* 2005; Zhang *et al.* 2005; Zeng *et al.* 2007).

A third type of genome-scan methods considers explicitly the physical linkage among SNPs surrounding a selected variant, either by focusing on patterns of long-range haplotype homozygosity (Sabeti *et al.* 2002; Voight *et al.* 2006) or by modelling the effect of linkage on multilocus genetic differentiation (Chen *et al.* 2010). These methods are more recent, and their properties have not been extensively investigated. Moreover, although they are focused on either a single population (Sabeti *et al.* 2002; Voight *et al.* 2006; Ferrer-Admetlla *et al.* 2014) or on pairs of populations (Sabeti *et al.* 2007; Chen *et al.* 2010; Fariello *et al.* 2013), they are being used to study structured populations consisting of many subpopulations without a clear understanding of how migration and complex population structure may affect their power and error rates. Thus, the objective of this study is to carry out a thorough evaluation of the performance of these methods under various scenarios of population structure. We focus mainly on the case where the selected variant is beneficial in part of the species range and neutral elsewhere, as it is the underlying scenario envisaged by many recent studies of adaptation (Lao *et al.* 2007; Hancock *et al.* 2008; Foll *et al.* 2014). Additionally, we consider both hard and soft selective sweeps. These two scenarios differ in the origin of the selected variant. In a hard selective sweep, the favoured allele appears through de novo mutation, while in a soft sweep, it is already segregating at low frequency in the population (standing genetic variation) or it arises from recurrent mutations (Hermisson & Pennings 2005; Pennings & Hermisson 2006a,b; Pritchard *et al.* 2010).

In the present analysis, we compare the performance of seven recent methods to detect selective sweeps. We incorporate in the analysis, methods that were developed to study a single population, a pair of populations or multiple populations. We explain in detail the ability of each method to capture the signal of selection left by both hard and soft sweeps under different scenarios of structured populations and a range of parameter values (migration and selection). The principle is to examine these methods on the same simulated data sets and draw conclusions about how the different model parameters affect their performance as described by power and false discovery rate (FDR). The goal of this analysis is to guide scientists in the choice of the methods that is better suited for their biological model.

Material and methods

Genome-scan methods

We focus our study on seven methods for which software is readily available: Integrated Haplotype Score (iHS) (Voight *et al.* 2006), Number of Segregating sites by Length (nSL) (Ferrer-Admetlla *et al.* 2014), Extended Haplotype-based Homozygosity Score Test (EHHST) (Zhong *et al.* 2010), Cross-population Extended Haplotype Homozygosity (xp-EHH) (Sabeti *et al.* 2007), Cross-population Extended Haplotype-based Homozygosity Score Test (XP-EHHST) (Zhong *et al.* 2011), Cross-population Composite Likelihood Ratio (XPCLR) (Chen *et al.* 2010) and hapFLK (Bonhomme *et al.* 2010; Fariello *et al.* 2013). They all use SNP data but propose different statistics to detect selection. In what follows, we will highlight their main differences, but we also include more technical details about all these methods in SI.

The methods we evaluate use different summary statistics that try to capture different genetic patterns consistent with the action of positive selection. We can distinguish three groups of methods:

1 Methods based on the decay of *haplotype* homozygosity as a function of recombination distance (iHS, nSL and xp-EHH): the underlying rationale of these methods is that selected alleles will have unusually long-range linkage disequilibrium given their frequency in the population.
2 Methods based on the decay of *genotype* homozygosity around a target SNP (EHHST and XP-EHHST): the underlying rationale is similar to that of the previous group, but in this case, homozygosity is measured in terms of mean homozygosity across all individuals in the sample instead of homozygosity of a region with respect to all chromosomes in the sample as in the previous group.
3 Methods based on the extent of multilocus genetic differentiation among populations around a target SNP (XPCLR and hapFLK): the underlying rationale is that genetic differentiation around a selected variant will be much larger than expected under drift, but instead of using single-locus measures of differentiation, it calculates differentiation for all SNPs within a window centred around the target SNP.

Another important difference between methods lies in whether or not they require phased data and information on the ancestral/derived status at each segregating site. XPCLR is the only method that does not have these requirements. Finally, one last difference among methods that needs to be highlighted refers to the number of populations they consider. iHS, nSL and EHHST are focused on a single population, xp-EHH, XP-EHHST, XPCLR consider two populations, while hapFLK considers an arbitrary number of populations.

Calculation of P values

The first step in the comparison of several methods is to define a common framework for assessing significance, which then allows us to calculate false-positive and false-negative rates as well as power. We used two alternative approaches:

(a)*From the empirical distribution of test scores:* in this case, we calculate the test statistic for all SNPs in the sample. Then using the empirical distribution of test scores, we consider as potentially adaptive all the loci with scores falling in the outlying 5% of the distribution. In the context of a simulation study, we know the truth and, therefore, we can readily identify true and false positives across all synthetic samples so as to calculate error rates and power of each method.
(b)*From a distribution of tests scores generated by neutral simulations*: in this case, we generate a large number of synthetic data sets assuming a particular demographic history (deemed appropriate for the species under study) and calculate the statistic scores for a target SNP. The distribution of test scores is then used as the null distribution and any loci with a test score falling in the outlying 5% of the distribution are considered potentially selected. To compare the performance of the different methods, we also carried out simulations under different selection scenarios and then pooled neutral and selected replicates to estimate power at various false-positive rates. These results are then presented as ROC curves obtained using the R package 'ROCR' (Sing *et al.* 2005).

The most widespread approach to assess significance when analysing real data is based on the empirical distribution (approach a). The reason for this is that in most cases, we do not know with certainty the true demographic history of the species under study. Thus, we present the results of this procedure in the main text and the results of the second procedure in the supplementary information.

Simulations

We generated synthetic data using SimuPOP (Peng & Amos 2008; Peng *et al.* 2011), a general-purpose, individual-based simulation platform for forward-in-time population genetic modelling. The Python scripts used to carry out the simulations are available at GitHub (https://github.com/alexvat/simulations).

Initially, we simulated three different population structure scenarios, an island model (Wright 1990), a stepping-stone model (Kimura 1953) and a dichotomous population fission model that leads to a hierarchical island structure (Fig. S1, Supporting information). In these cases, we considered four diploid demes, each of constant effective population size $N_e = 2500$. Thus, total population size was 10 000. Table 1 presents a summary of the parameters that were used in the simulations. In the case of the island and the stepping-stone models, every individual migrates to another deme with probability m (0.05, 0.01 or 0.008). In the case of the hierarchical model, migration between demes within the same group (continent) was higher than migration between demes in different groups (see Fig. S1c, Supporting information). In this latter scenario, we start at $t = 0$ with a single population (Z with 10 000 individuals). At $t = 100$ generations, it splits into two subpopulations (Y, Z of size 5000 individuals each), and at $t = 300$, each of the 2 subpopulations (Y, Z) splits into two other subpopulations [(X, Y) and (W, Z), respectively], resulting in four subpopulations at $t > 300$.

Table 1 Parameters that were used in the simulations with simuPOP for the hard and the soft sweep. m1 is the migration rate of populations within the same group in the hierarchical model and m2 the migration rate of populations between different groups

	Population structure	Migration rate (m)	Selective coefficient (s)	Mutation rate	Recombination rate (r)
Hard sweep	Island model	0.008	0.1 ($2N_es$ = 500)	10^{-8}	0.00375 cM/kb
	Stepping stone model	0.01			
		0.05			
		0.008	0.08 ($2N_es$ = 400)		
			0.01 ($2N_es$ = 50)		
	Hierarchical model	m1 = 0.02	0.1 ($2N_es$ = 500)		
		m2 = 0.01			
Soft sweep	Island model	0.01	0.05 ($2N_es$ = 250)		
	Stepping stone model				
	Hierarchical model	m1 = 0.02	0.05 ($2N_es$ = 250)		
		m2 = 0.01			
		m = 0	0.05 ($2N_es$ = 250)		

Following previous analyses (Hanchard *et al.* 2006; Zhong *et al.* 2010, 2011), we considered L = 101 bi-allelic SNPs located in the same chromosome. The recombination rate was ρ = 1.5 (= $4N_er$) so that r = 0.00375 cM/kb leading to a fixed distance of 4 kb between loci. For all the scenarios, neutral loci shared the same mutation rate (10^{-8} per generation).

For each demographic model, we considered two selection scenarios, a hard sweep and a soft sweep. Under a hard sweep, new mutations are easily lost due to genetic drift so that large selection coefficients are needed to minimize stochastic loss. In our case, we used s = 0.1 ($2N_es$ = 500), 0.08 ($2N_es$ = 400) and 0.01 ($2N_es$ = 50). On the other hand, a soft sweep acts upon standing genetic variation, so selection does not need to be very strong to overcome stochastic loss in most simulations. In our case, we used s = 0.05 ($2N_es$ = 250). For the simple structured population cases (island, stepping-stone and hierarchical model with a total of four subpopulations each), we assumed that a selected variant at locus 50 (i.e. the middle of the genomic region) was favoured in only one deme and that it was neutral in all other demes. We assumed a codominant selection model where fitness of the homozygotes for the ancestral allele is 1, fitness of heterozygotes is (1 + s/2), and fitness of homozygotes for the derived allele is (1 + s).

For all scenarios, we used an initialization procedure that samples allele frequencies from an island model at migration–mutation–drift equilibrium. More precisely, all loci were initialized at the beginning of the simulations, t_0 = 0, by sampling the allele frequencies of each locus from a beta distribution with parameters $a = 4N_em^*p$ and $b = 4N_em^*(1-p)$, where p is the frequency in a migrant pool, which was derived from real human SNP data from noncoding regions, m is the migration rate, and N_e is the effective population size (Wright 1931). We started selection after a burn-in (t_1) that allowed the system to reach migration–mutation–drift equilibrium. In the case of the island model, the burn-in period was very short (50 generations) compared to the stepping-stone model (100 generations) and the hierarchical model (500 generations). Figs S2–S4 (Supporting information) in show the steady state reached in terms of equilibrium allele frequencies and LD under each scenario. In the case of hard sweeps, locus 50 was monomorphic at t_0 and all throughout the burn-in period. At t_1, once populations were at equilibrium, a single copy of a new advantageous mutation (the derived allele) was introduced at this locus in deme Y only. All the simulations were carried out until the selected locus was nearly fixed in the selected population. We took samples of populations at different times points where the selected allele frequency exceeds a given threshold (0.1, 0.2, ... , c. 1) to study its influence on the performance of the methods.

In the case of the soft sweep from standing variation, the selected variant was already segregating in the population before the onset of selection. More precisely, we assume that the allele became beneficial after an environmental change, but was neutral under the previous conditions. At $t = t_0$, we set the frequency of the selected allele at locus 50 in the migrant pool to 0.02, 0.1, 0.2 or 0.4. At $t = t1$, when selection started, the average allele frequency of the selected variant over the replicates remained unchanged at these respective values. We generated 1000 replicates for each of these scenarios.

Statistical analysis

Performance of each method was evaluated using the two methods described above which henceforth are referred to as the *empirical distribution* (method a) and

simulated distribution (method b) approaches. The results are similar for both approaches so here we focus on the empirical distribution approach, while the simulated distribution approach is further described in supplementary information.

Given that the aim of all methods is to identify genomic regions under selection and not necessarily to uncover a specific advantageous mutation, we considered that a method succeeded at detecting selection if at least one of the SNPs in a window bounded between SNP 45 and SNP 55 was identified as selected (i.e. a window spanning 20 kb upstream and 20 kb downstream the selected locus). Outlier SNPs outside of this window were considered as false positives. The choice of a 40 kb window (10 SNPs) was decided after investigating the distribution of the scores produced by each method around the selected variant (see Fig. S5, Supporting information) and ensures that the signature of selection is restricted to the window and, therefore, does not lead to wrong estimations of power and FDR. The statistical significance threshold for all tests was defined as the 5% outliers considering the whole region of 101 loci. FDR is rarely measured. Indeed, most previous studies assess performance based on neutral simulations that only allow for the calculation of power and FPR. However, the application of these methods involve multiple testing and, therefore, we measure error rates in terms of FDR at several time points to better characterize the stage of the selective sweep (i.e. initial, intermediate or nearly completed) at which each method performs best.

Results

We first compared the performance of six methods [iHS (Voight *et al.* 2006), nSL (Ferrer-Admetlla *et al.* 2014), EHHST (Zhong *et al.* 2010), xp-EHH (Sabeti *et al.* 2007), XP-EHHST (Zhong *et al.* 2011) and XPCLR (Chen *et al.* 2010)] for the hard sweep scenario under the island (Wright 1990) and stepping-stone (Kimura 1953) models, the two most well-known population models. We then selected the methods that were the most efficient under these conditions and we compared them under the hierarchical island model. In this case, we also included hapFLK (Fariello *et al.* 2013) in the comparison because it is specifically developed for this scenario. Next, we selected the methods that were the most efficient under this latter scenario and subjected them to further scrutiny, using data generated from soft sweep scenarios and more complex stepping-stone models. The results are similar for the two approaches used to compare methods; therefore, we present the results of the *empirical distribution* approach here and those of the *simulated distribution* approach in the supplementary information.

Hard sweep

Local selective sweeps under simple population structure models. Figure 1 presents the results for a hard sweep under the island model for five different scenarios: (i) $m = 0.008$, $s = 0.01$ ($2N_es = 50$); (ii) $m = 0.008$, $s = 0.08$ ($2N_es = 400$); (iii) $m = 0.008$, $s = 0.1$ ($2N_es = 500$); (iv) $m = 0.01$, $s = 0.1$ ($2N_es = 500$); and (v) $m = 0.05$, $s = 0.1$ ($2N_es = 500$). Both EHHST and XP-EHHST performed poorly under all scenarios (Fig. 1e, g), exhibiting very low power and high FDR (Fig. S6c,e, Supporting information) regardless of the allele frequency of the selected variant. The performance of the four other methods (iHS, nSL, xp-EHH and XPCLR) varies depending on the allele frequency of the favoured variant in the selected population (Y) and the different parameters tested (migration rate and selection coefficient).

As expected, when selection is strong ($2N_es = 500$ or 400) and migration is low ($m = 0.008$ or $2N_es = 50$), the four above-mentioned methods performed quite well at least at one stage of the selective sweep (initial, intermediate or nearly completed; Fig. 1). More precisely, iHS and nSL detected sweeps for which the selected variant was still at low frequency (*c.* 0.1–0.3). The performance of xp-EHH increased slowly as the frequency of the selected allele in the selected population increases and it has a power of *c.* 100% when the selected locus is close to fixation (allele frequency: AF = *c.* 0.9). XPCLR behaved in a similar way, but the performance increased sharply first and remained high until the selected locus approached fixation. The performance of XPCLR was the highest of all methods when the allele frequency was intermediate to high (AF = 0.3, 0.9), but extremely poor when it was low (AF = 0.1, 0.2), in which case iHS and nSL were better methods.

Migration has a strong detrimental effect on the performance of all methods (Fig. 1). Indeed, when migration was high ($m = 0.05$ per generation), the performance of iHS, nSL, xp-EHH and XPCLR was poor. When the selected variant is favoured in one population but neutral elsewhere, migration has a strong homogenizing effect. Therefore, the performance of iHS and nSL decreased because the selected population was swamped by haplotypes carrying the counter-selected variants. Thus, the frequency of the haplotype containing the selected variant decreased and the genetic signal of selection was weakened. On the other hand, the performance of xp-EHH and XPCLR decreased because the nonselected populations were swamped by the haplotype containing the beneficial allele. Thus, with high migration ($m = 0.05$), the beneficial allele

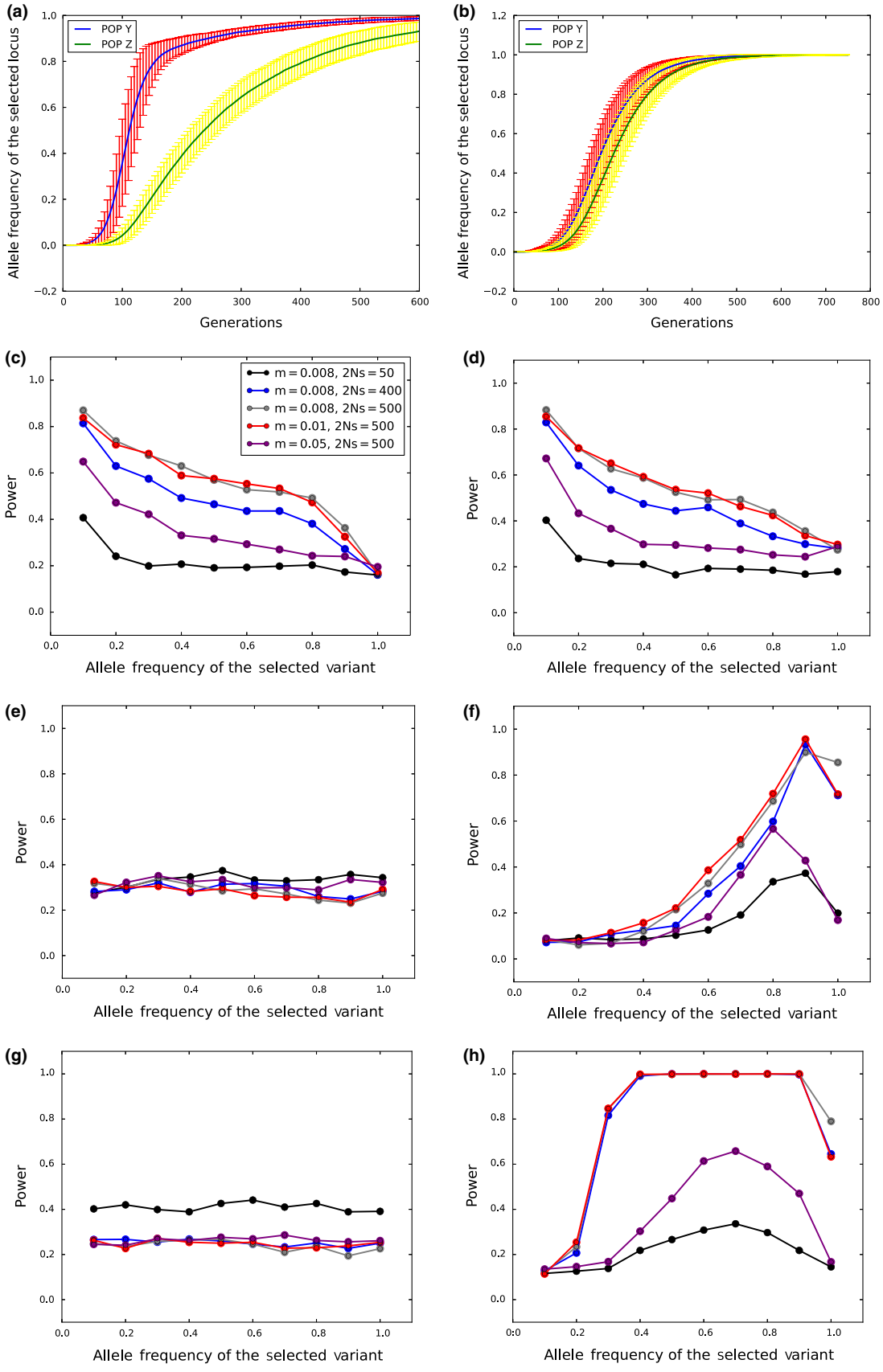

Fig. 1 Results for the island model: (a) trace of the allele frequency of the selected variant in the selected population, Y, and in a neutral population, Z, with migration rate 0.01 per generation; (b) likewise with $m = 0.05$ (the blue/green line represent the mean allele frequency over 1000 simulations, and the vertical lines represent the standard deviation); (c–d) power for each method for the hard sweep under the island model: (c) iHS; (d) nSL; (e) EHHST; (f) xp-EHH; (g) XP-EHHST; and (h) XPCLR. The scenario considers four demes with 2500 individuals each, 101 loci and 0.00375 cM/kb recombination rate, varying the migration rate and selection coefficient (see legend).

spread much faster (than with $m = 0.01$) and the differentiation in frequency of the selected variant between the selected and nonselected populations decreased sharply (Fig. 1a,b). These results hold for both the island and the stepping-stone model (Fig. S7, Supporting information).

Under an isolation-by-distance scenario, the choice of the two populations to include in xp-EHH and XPCLR analyses can affect their performance. To investigate this, we examined the performance of XPCLR, the method with highest power in the previous scenarios, as a function of the distance between the population undergoing selection and the 'neutral' ones for the scenario with $m = 0.01$ and $2N_es = 500$. Figure 2 shows that the larger the distance between the selected and nonselected populations, the lower the power of XPCLR was for intermediate values of the allele frequency of the selected variant. This may seem counterintuitive because larger distance leads to reduced migration and results obtained for the island model suggest that weak migration facilitates the detection of the selection signal. However, we note that XPCLR is based on the multilocus genetic differentiation between a selected and a nonselected population. More precisely, it compares the multilocus differentiation expected around a selected variant with that expected around a neutral variant (c.f. eq. 6 in Chen *et al*. 2010). As distance between the two populations increases, the neutral multilocus differentiation increases strongly and, therefore, the difference in genetic differentiation between neutral and selected regions decreases. This behaviour is similar to that observed for genome-scan methods based on F_{ST} (Price *et al*. 2008). We further studied whether or not selection could be detected when the selected population was not included in the analysis. Interestingly, the selected region is detected when the selected variant has reached intermediate-to-high frequencies in the population right next to a selected one. Thus, in the case of a nearly completed selective sweep, it is possible to wrongly conclude that selection is acting upon one of the two populations when this is not really the case. However, the power of the method decreases sharply when the selected population is not adjacent to one of the two populations included in the analysis.

In the case of the hierarchical island model (Fig. 3), we focus on five methods (iHS, nSL, xp-EHH, XPCLR and hapFLK) discarding EHHST and XP-EHHST because they performed very poorly under the simple population structure scenarios considered above (island and stepping-stone model with four populations). For the two-population tests (xp-EHH and XPCLR), we investigated the power of the methods both when the selected and nonselected sampled populations were in the same group (continent) and when they were in different groups. Note that migration between populations in the same group is higher ($m = 0.02$) than between those in different groups ($m = 0.01$). The overall pattern of performance as a function of allele frequency of the selected variant is similar to that observed under the simpler spatial structure scenarios. However, the baseline power of all methods is largely reduced. More specifically, the power of iHS and xp-EHH was decreased to *c*. 70%, with an FDR *c*. 30% for the allele frequencies at which they performed optimally under the simpler spatial scenarios. On the other hand, the performance of XPCLR remained high with power *c*. 90% and FDR lower than 20%. Nevertheless, such high performance is achieved for a narrower range of allele frequencies (0.6, 0.7) than for the simple spatial structure scenarios tested before (AF: 0.3–0.9). As it was expected, when comparing populations from the same geographic group (Y-X), the power of the methods was more strongly reduced (*c*. 10% for xp-EHH and *c*. 20% for XPCLR) than when populations belonged to different groups. HapFLK exhibited the best performance for a wide range of allele frequencies but was outperformed by xp-EHH and XPCLR for very high allele frequencies.

Local selective sweeps in a heterogeneous environment. We explore a scenario akin to that considered by previous studies of genetic sweeps in structured populations (e.g. Bierne 2010). More precisely, we simulated a stepping-stone scenario with a large number of populations (52) undergoing a hard selective sweep in a heterogeneous environment where the new mutation is beneficial in half of the species range and detrimental in the other half. We simulated 52 populations with 500 individuals each, a genomic region comprising 101 loci with a recombination rate of 0.00375 cM/kb per generation, a selection coefficient of 0.05 ($2N_es = 50$) and a migration rate of 0.05 per generation. Locus 50 was initially fixed for allele 0 in all populations, and after equilibrium, a

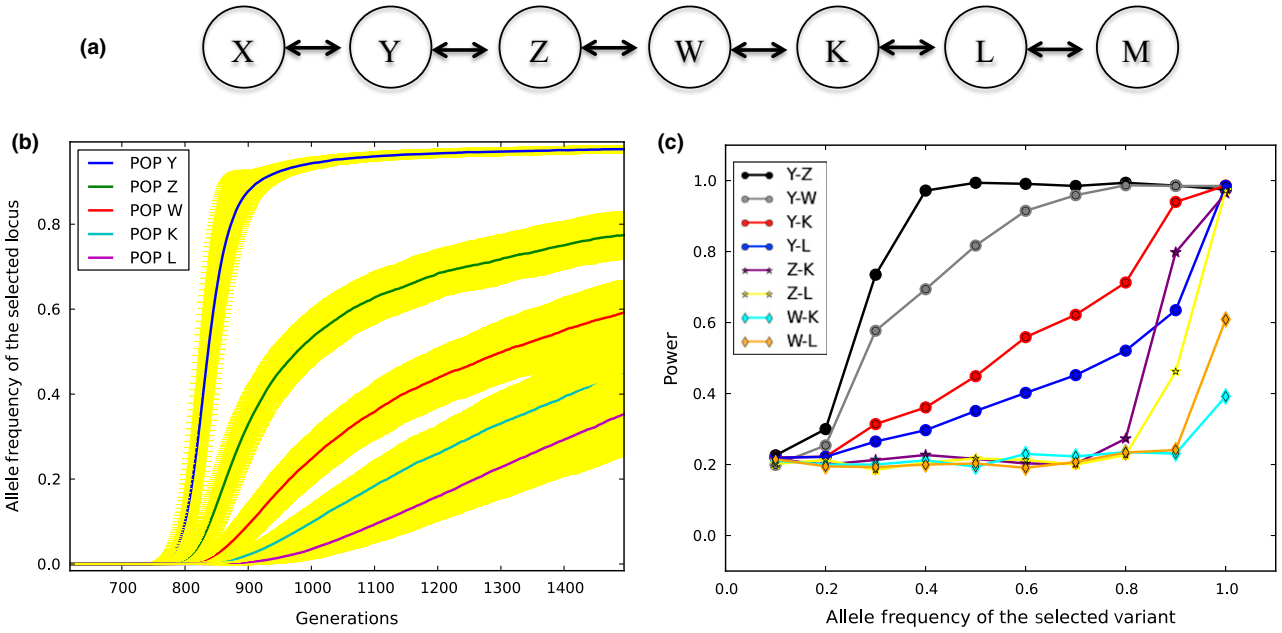

Fig. 2 Effect of distance from selected population on XPCLR: (a) graphical description of the stepping-stone model with 7 populations with 2500 individuals each, 101 loci, selection coefficient 0.1 ($2N_es$ = 500), migration rate 0.01 and recombination rate 0.00375 cM/kb. Selection is present in population Y; and (b) trace of the allele frequency of the selected locus for all pairs of populations except from the boundary ones. The lines represent the mean allele frequency over the 1000 simulations and the vertical lines represent the standard deviation; and (c) power of XPCLR for the case of the hard sweep for the different pairs of populations.

de novo advantageous mutation was introduced in the far left deme. The new mutant was favoured in habitat 1 (populations 1–25) and was counter-selected in habitat 2 (populations 26–50) (Fig. 4b). To avoid computational burden due to the very large number of populations studied here, we evaluated performance using 100 simulations instead of the 1000 used for the simpler scenarios. However, as shown in Fig. S5 (Supporting information), this reduced number of replicates does not have an impact on the outcome of the analysis. All methods were tested, but we only present results for XPCLR and hapFLK because all other methods have negligible power under this scenario.

The power of hapFLK was almost maximal (99.9%), but its error rate was very high too (FDR 43.3%). All 50 populations except the boundary ones were included in the hapFLK analysis. However, in the case of XPCLR, which can only analyse two populations at a time, we focused on pairs of populations and evaluated the effect of distance between them on the performance of the test. Figure 4(a) shows the XPCLR results for analyses using population 1 (i.e. the far left population) as objective and each one of the other populations as reference. Results were obtained after 40 000 generations since the appearance of the mutation. The results show that XPCLR can detect selection only when the reference population is near the boundary between the two habitats (a similar pattern is observed when using demes 13 or 25 as objective populations; Fig. S8, Supporting information). The FDR follows the inverse pattern of the power, and this holds true for all the populations in habitat 1 (Fig. S8, Supporting information). XPCLR does not perform well when populations from the same habitat are compared because after 40 000 generations, the sweep is complete in all demes belonging to habitat 1 (Fig. 4b) and multilocus differentiation around the selected allele has disappeared (Fig. 4c). When the reference population is in habitat 2 and far from the boundary with habitat 1, XPCLR does not perform well either, as the genetic differentiation of the neutral background increases strongly with distance from the objective population (Fig. 4d) and this decreases the power to detect selection using multilocus differentiation. Thus, we conclude that caution is needed when using XPCLR to study scenarios involving genetic clines or secondary contact zones. Nevertheless, it is worth mentioning that this method may be useful to identify the transition zone where the change in selection regime is observed.

Soft sweep

In the case of soft sweeps from standing variation, the most crucial parameter influencing the power of the

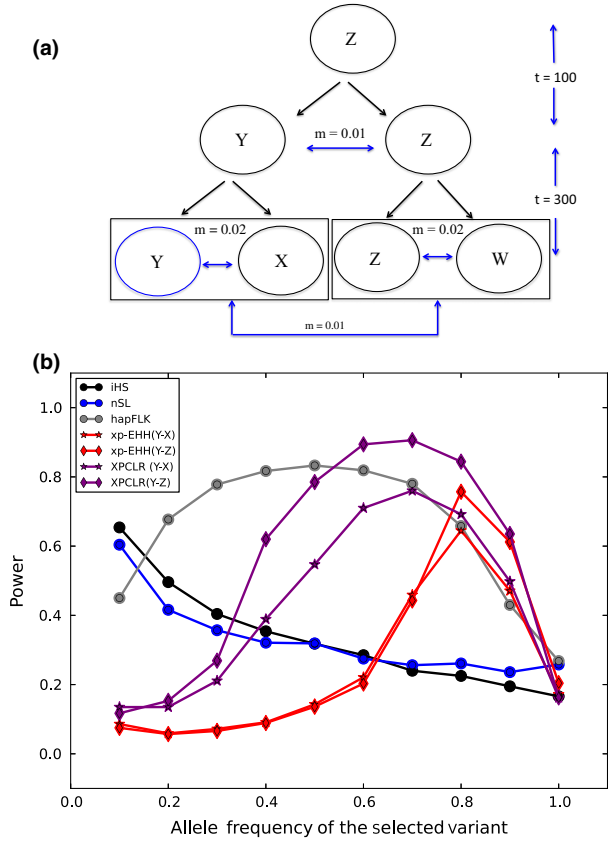

Fig. 3 Results for the hierarchical island model and hard sweep scenario: (a) graphical representation of the population structure of the hierarchical model. Selection is present in only one of the demes (Y); and (b) power for iHS (black), nSL (blue), hapFLK (grey), xp-EHH (red) and XPCLR (purple). Each of the four demes has 2500 individuals. We used 101 loci, migration rate between populations within continents 0.02 and between continent 0.01, selection coefficient 0.1 ($2N_es$ = 500) and 0.00375 cM/kb as recombination rate. In the case of xp-EHH and XPCLR, the comparison of demes in the same (Y–X) and different (Y–Z) continents is also shown.

methods is expected to be the initial allele frequency (IAF) of the selected variant. To investigate this, we examined the power of the methods at the following IAF of the selected variant: 0.4, 0.2, 0.1 and 0.02. Given that the methods did not show sufficient performance with a high migration rate (m = 0.05) under the hard sweep scenario, we examined their behaviour for the soft sweep with a migration rate of 0.01. The results for the island model are presented in Fig. 5 and are identical to those of the stepping-stone model, which are presented in Fig. S9 (Supporting information). The power of iHS and nSL was dramatically reduced (to <50%) under all three scenarios tested. The performance of xp-EHH was good at high allele frequencies (AF = 0.9) before fixation, as in the case of the hard sweep. This holds true for all the different initial allele frequencies that were tested. The performance of XPCLR was good for intermediate and high allele frequencies of the selected locus before fixation, particularly for IAF: 0.2, 0.1 and 0.02.

Next, we investigated the performance of xp-EHH, XPCLR and hapFLK under a hierarchical island model undergoing a soft sweep. The power of all methods drops substantially, being in general below ≈40%, while their FDR is very high (Fig. S10, Supporting information). As opposed to iHS and xp-EHH that are based on long-range haplotype homozygosity, XPCLR and hapFLK are based on multilocus genetic differentiation and, therefore, their performance under this scenario might be improved in the absence of migration. To investigate this possibility, we carried out simulations of this same scenario without migration. The results show that performance of both methods, but especially of hapFLK, improves particularly for high frequencies of the selected variant (Fig. S11, Supporting information).

Discussion

This study aimed at assessing the performance of recent statistical methods that are being used to detect selective sweeps in structured populations. These methods focus on multilocus signatures of selection that include information on linkage disequilibrium. Although they were originally developed to study isolated populations or two population scenarios, they are being applied to all kinds of structured populations (e.g. island, stepping stone, hierarchical). Thus, our objective was to investigate how violations to the underlying model influence their power and error rates.

We compared the performance of seven genome-scan methods (iHS, nSL, EHHST, XP-EHHST, xp-EHH, XPCLR and hapFLK) under subdivided population structures. Some of them such as iHS and xp-EHH have already been widely used (Qanbari *et al.* 2011; Andersen *et al.* 2012; Park *et al.* 2012), while the others, such as XPCLR, nSL and hapFLK, are quite popular but fairly recent and have not yet been extensively scrutinized (Peng *et al.* 2011). We evaluated these methods under a wide range of population structure scenarios undergoing either a hard or a soft selective sweep. Furthermore, we investigated how the power and false discovery rate of the methods are influenced by the allele frequency of the selected variant at the time of sampling.

We mainly focus on a local selective sweep scenario where the sweeping allele is beneficial in one deme and neutral in all the others, a selection scenario that has been frequently used in studies of human populations (Fournier-Level *et al.* 2011) but which has not yet been studied extensively. Previous analyses on subdi-

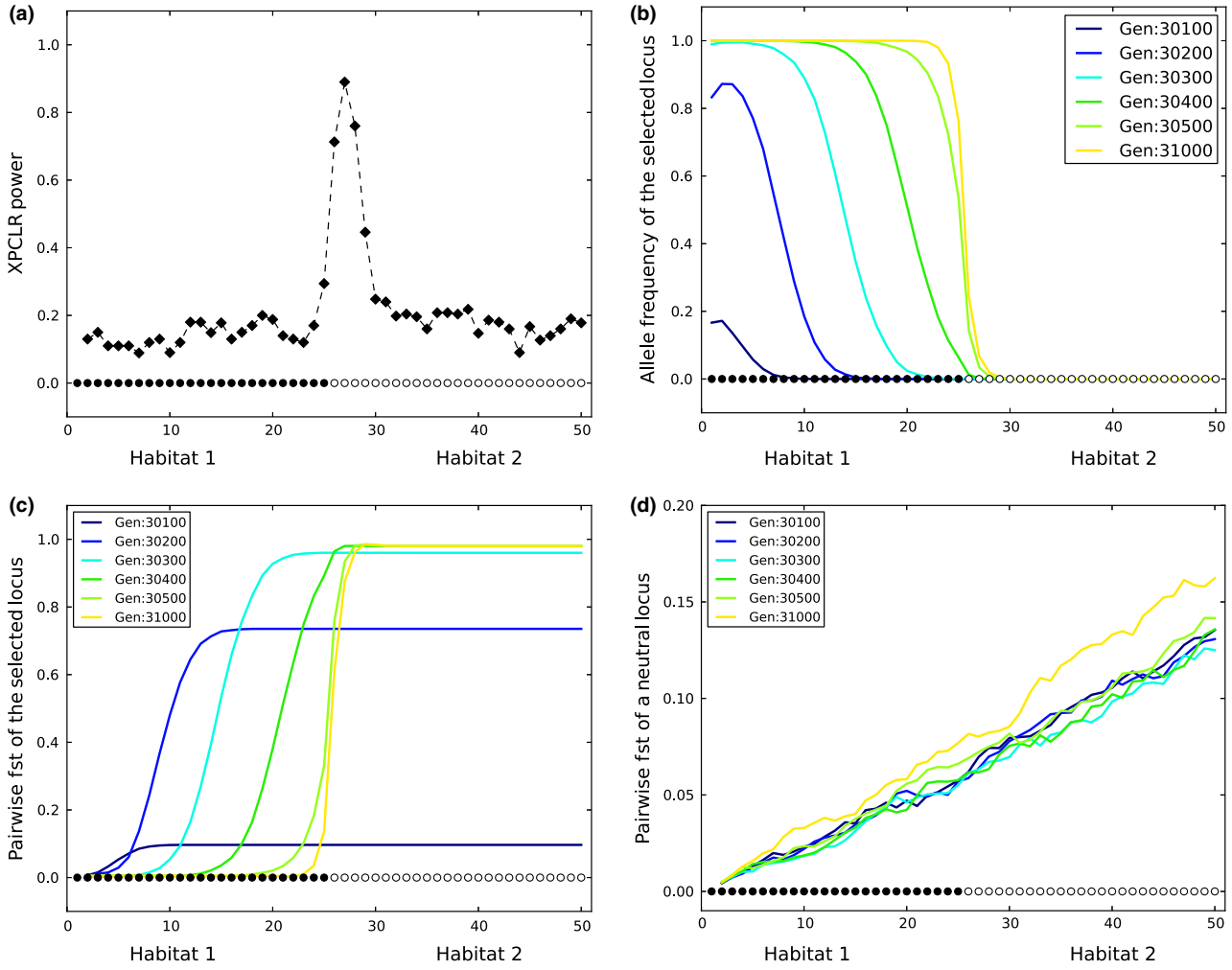

Fig. 4 Results of simulations of the stepping-stone scenario with 52 populations. We simulated 101 loci with a recombination rate of 0.00375 cM/kb. Each population had 500 individuals, the migration rate was 0.05, and the selection coefficient was 0.05 ($2N_e s = 50$). Allele 1 is favoured in populations 1–25 (habitat 1), and allele 0 is favoured in populations 26–50 (habitat 2). (a) Power of XPCLR for analyses with population 1 as the objective population and each one of the other populations as the reference after 40 000 generations since the appearance of the mutation; (b) frequency of the selected allele (at locus 50) across all populations at different times since its appearance in population 1 (number of generations indicated in the legend); (c) pairwise F_{ST} between population 1 and all the others for the selected locus (50); and (d) pairwise F_{ST} between population 1 and all the others for the neutral locus (80).

vided populations have examined the case of global sweeps (Barton 2000; Santiago & Caballero 2005; Bierne 2010) or sweeps where a new variant is beneficial in one part of the species range but detrimental elsewhere (Le Corre & Kremer 2003; Bierne 2010). Here, we investigate in detail the scenario of an allele that is neutral in most of the range but beneficial in one population. A feature of this latter scenario that is shared with models of global sweeps is that migration will ultimately lead to the fixation of the beneficial allele in all populations (Fig. 1b).

In general, our results suggest that five (iHS, nSL, xp-EHH, XPCLR, hapFLK) of the seven methods we evaluated are able to identify genomic regions undergoing a selective sweep in one or more of the scenarios we considered. The main difference between this group and the other two methods (EHHST and XP-EHHST) is the nature of the information they use to calculate the test statistic. The first group of five methods uses population level information (either haplotype frequencies or allele frequencies), while the two other methods are based on mean and standard deviation of homozygosity across all individuals in the sample (as opposed to homozygosity of a region with respect to all chromosomes in the sample – see Material and Methods and SI). This could explain their poor performance. More precisely, when there is no migration among populations, as in the scenarios considered by

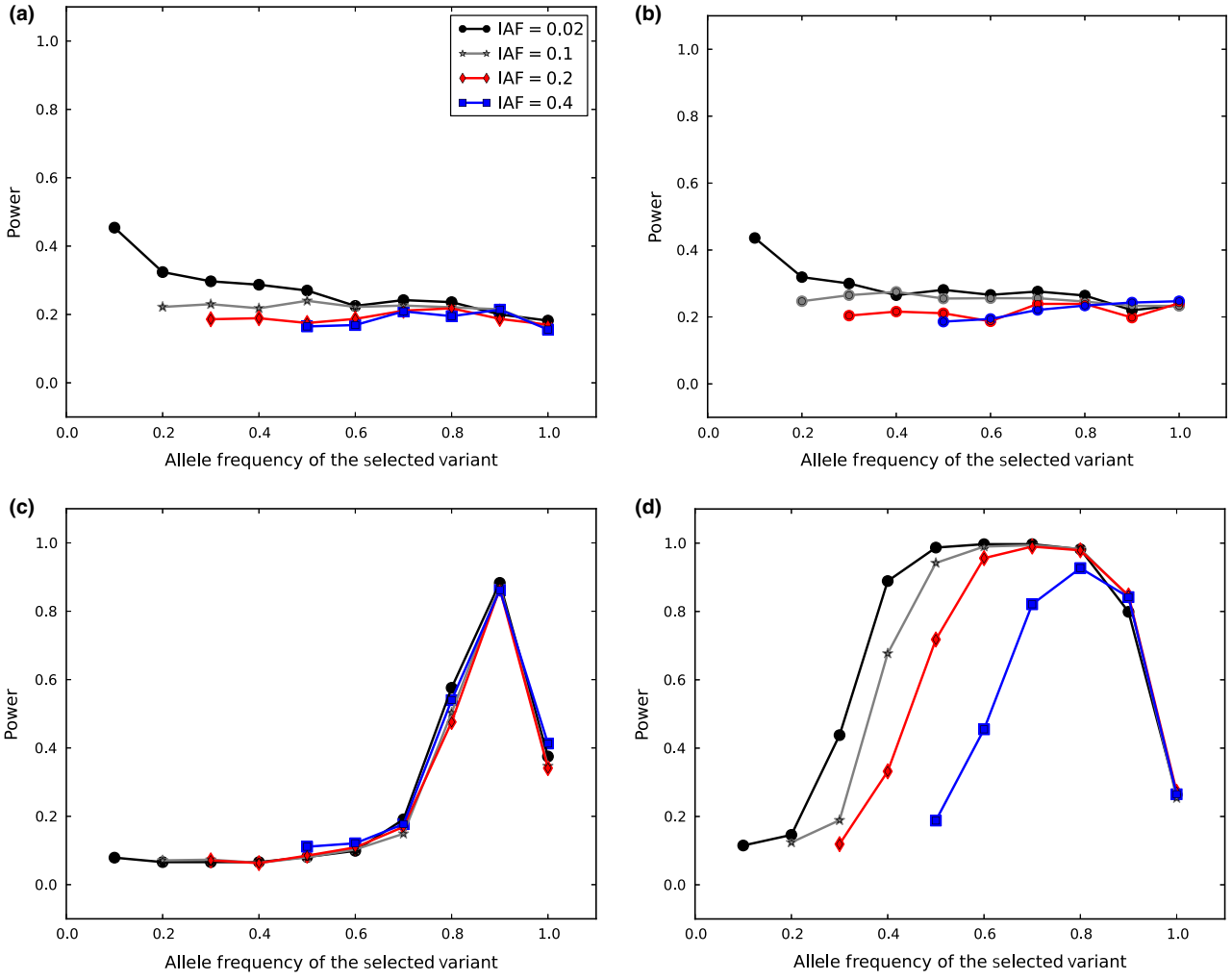

Fig. 5 Power of each method for the case of a soft sweep under the island model. (a) iHS, (b) nSL (c) xp-EHH and (d) XPCLR. Results presented for different initial allele frequencies of the selected variant: 0.02 (black), 0.1 (grey), 0.2 (red), 0.4 (blue). Four demes with 2500 individuals each, 101 loci, migration rate 0.01, selection coefficient 0.05 ($2N_e s = 250$) and 0.00375 cM/kb as recombination rate. Selection is acting only in one deme (Y).

Zhong *et al.* (2010), the homozygosity is high for all individuals in the sample from the selected population and, therefore, its standard deviation is small, which increases the power of the test (Zhong *et al.* 2010). However, in our scenarios, migration is present and, therefore, there is a mixture of individuals with very low and very high homozygosity in the selected population, and thus, the standard deviation of homozygosity is extremely large, decreasing the power of the test. A second general result of our local selective sweep study is that XPCLR (Chen *et al.* 2010) has the best overall performance under the range of scenarios considered in this study. However, it is surpassed by iHS (Voight *et al.* 2006) and nSL (Ferrer-Admetlla *et al.* 2014), when the frequency of the selected variant is low (i.e. for starting selective sweeps ≥0.1 and ≤0.3). XP-EHH performs well for a narrow range of high allele frequencies of the selected variant, as previously shown by Sabeti *et al.* (2007).

In the case of the more complex scenario of a hard selective sweep in heterogeneous environments, only two methods, hapFLK and XPCLR, were relatively efficient at detecting sweeps, but their power was still limited to some particular conditions. HapFLK had high power but also a high FDR. XPCLR, on the other hand, could detect a sweep only if the reference population was located near the boundary between the two habitats. Overall, these results suggest that the applicability of these selection detection methods to study genetic clines and secondary contact zones is limited. Nevertheless, by combining them, it may be possible to identify the genomic region driving the genetic cline and also

the geographic region where the transition between the two selective regimes occurs.

There is a paucity of simulation studies comparing the performance of methods aimed at identifying selective sweeps. However, evaluations of individual methods are presented in the publications that introduce them for the first time. Voight *et al.* (2006) indicate that iHS performs best for intermediate-to-high allele frequencies, while our results show a different pattern with best performance at low frequencies (>0.1 and <0.3). We explain this difference by the homogenizing effect of migration in the subdivided population structures that we investigated. In the case of a local sweep where a variant is favoured in one deme and neutral elsewhere, the selected population is swamped by haplotypes carrying the counter-selected variant. Therefore, the strength of the genetic signal used by iHS decreases. A similar pattern is observed for nSL, another single-population method. The effect of migration on power is also pronounced for the two-population methods (XP-EHH and XPCLR) (c.f. Fig. 1). As time goes by, and when migration is low, the allele frequency of the selected variant (and linked SNPs) increases very rapidly in the selected population but very slowly in the neighbouring populations (Fig. 1a), so power to detect the sweep is high. However, higher migration rates lead to a simultaneous and rapid increase of the selected variant and linked SNPs also in neighbouring populations, which reduces the differentiation and the power to detect selection (Fig. 1b). A similar effect is observed when the selection coefficient is low (0.01), in which case the power decreases dramatically to <45%.

Fariello *et al.* (2013) compare hapFLK with several other methods (F_{ST}, FLK, hapFST and xp-EHH) and show that it performs better than all of them. However, they consider a scenario where there is a single episode of migration throughout the evolutionary history of the population, a scenario applicable to a limited number of species. On the other hand, our analysis assumes continuous migration, a scenario that should be applicable to a wide range of species. In this situation, hapFLK performs well for hard sweeps both in hierarchical and even under simpler population structures (e.g. island model; Fig. S12, Supporting information). However, this is not the case for the soft sweep scenarios. Nevertheless, a great advantage of hapFLK over the other methods is that it is applicable to scenarios with arbitrary number of subpopulations, which makes results independent of the choice of populations included in the analysis. Additionally, hapFLK (and nSL) does not require estimates of recombination rates, and therefore, it is applicable to nonmodel species.

Our simulation study also systematically investigates whether or not signals produced by soft selective sweeps from standing variation can be detected. Unsurprisingly, all methods are less efficient under soft sweep than under hard sweep scenarios because multiple haplotypes containing the selected variant segregate in the population. More specifically in the island or stepping-stone models, iHS has very limited power. On the other hand, xp-EHH has high power only for a very small range of high allele frequencies. Interestingly, the initial frequency of the selected variant before the onset of selection has a negligible effect on the performance of iHS and xp-EHH. XPCLR also has high power to detect soft sweeps under simple population structure scenarios, particularly for small and moderate IAF. However, none of the methods performed satisfactorily under the hierarchical population structure with migration, not even hapFLK that was specifically designed for such scenario. Note, however, the performance of XPCLR and hapFLK is greatly increased under the hierarchical scenario in the absence of migration. Thus, XPCLR and hapFLK are the most promising methods for detecting soft sweeps under complex population structures where migration is absent or very low.

As we have shown, no single method is able to detect both starting and nearly completed selective sweeps. Combining several methods (e.g. XPCLR or hapFLK with iHS or nSL) can greatly increase power to detect a wide range of selection signatures. A first step in this direction is presented by Grossman *et al.* (2010) who propose the composite of multiple signals method which combines five different approaches [F_{ST}, xp-EHH, iHS, ΔiHH (measures the absolute integrated haplotype homozygosity) and ΔDAF (accounts for derived alleles at high frequency)].

Although our study suggests that some of these methods are potentially useful to identify selected regions, it is important to keep in mind that the statistical properties of the test statistics they use are unknown and, therefore, assessing significance is based on ad hoc methods that lack statistical rigour. The only exceptions are EHHST and XP-EHHST, which were shown to be asymptotically normal (Zhong *et al.* 2010). However, our study suggests that these two methods are not able to detect selective sweeps under most realistic scenarios. In all other cases, there are two alternative approaches (see Material and Methods). One is based on the empirical distribution of the test statistic, which includes both selected and neutral sites and, therefore, is likely to lead to high false-positive rates. The second approach is based on a simulated distribution and would be preferable in principle. However, it requires very good knowledge about the demographic history of the population under study. Unfortunately, this is almost never the case even for model species. Nevertheless, it is important to note that despite their important

differences, our study suggests that both methods lead to comparable results (compare Figs 1–3, 5 and Figs S13–S23, Supporting information) giving some support for the use of the empirical distribution approach.

Our study represents a substantial evaluation of recent genome-scan methods to detect selective sweeps, and therefore, it should be of broad interest. We note, however, that with the only exception of XPCLR, all these methods are applicable only to model species because they require phased data and information on the ancestral/derived status at each segregating site. However, continued developments in sequencing technology are broadening the range of species that could be studied using these methods. Our systematic comparison of genome-scan methods clarifies the conditions under which they should be applied and will help users to choose the most adequate approach for their study.

Acknowledgements

The authors thank Christelle Melodelima for helpful discussions. This work was supported by the Marie-Curie Initial Training Network INTERCROSSING (European Commission FP7). OEG was further supported by the MASTS pooling initiative (The Marine Alliance for Science and Technology for Scotland).

References

Andersen KG, Shylakhter I, Tabrizi S, Grossman SR, Happi CT, Sabeti PC (2012) Genome-wide scans provide evidence for positive selection of genes implicated in Lassa fever. *Philosophical transactions of the Royal Society of London Series B, Biological Sciences*, **367**, 868–877.

Barton NH (2000) Genetic hitchhiking. *Philosophical transactions of the Royal Society of London Series B, Biological Sciences*, **355**, 1553–1562.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London Series B*, **263**, 1619–1626.

Bierne N (2010) The distinctive footprints of local hitchhiking in a varied environment and global hitchhiking in a subdivided population. *Evolution; International Journal of Organic Evolution*, **64**, 3254–3272.

Bonhomme M, Chevalet C, Servin B *et al.* (2010) Detecting selection in population trees: the Lewontin and Krakauer test extended. *Genetics*, **186**, 241–262.

Chen H, Patterson N, Reich D (2010) Population differentiation as a test for selective sweeps. *Genome Research*, **20**, 393–402.

Fariello MI, Boitard S, Naya H, SanCristobal M, Servin B (2013) Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics*, **193**, 929–941.

Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. *Genetics*, **155**, 1405–1413.

Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R (2014) On detecting incomplete soft or hard selective sweeps using haplotype structure. *Molecular Biology and Evolution*, **31**, 1275–1291.

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

Foll M, Gaggiotti OE, Daub JT, Vatsiou A, Excoffier L (2014) Widespread signals of convergent adaptation to high altitude in Asia and America. *American Journal of Human Genetics*, **95**, 394–407.

Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM (2011) A map of local adaptation in *Arabidopsis thaliana*. *Science*, **334**, 86–89.

Grossman SR, Shlyakhter I, Karlsson EK *et al.* (2010) A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science*, **327**, 883–886.

Hanchard NA, Rockett KA, Spencer C *et al.* (2006) Screening for recently selected alleles by analysis of human haplotype similarity. *American Journal of Human Genetics*, **78**, 153–159.

Hancock AM, Witonsky DB, Gordon AS *et al.* (2008) Adaptations to climate in candidate genes for common metabolic disorders. *PLoS Genetics*, **4**, 32.

Hermisson J, Pennings PS (2005) Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics*, **169**, 2335–2352.

Jensen J, Kim Y, DuMont VB, Aquadro CF, Bustamante CD (2005) Distinguishing between selective sweeps and demography using DNA polymorphism data. *Genetics*, **170**, 1401–1410.

Jeong C, Di Rienzo A (2014) Adaptations to local environments in modern human populations. *Current Opinion in Genetics & Development*, **29**, 1–8.

Kim Y, Nielsen R (2004) Linkage disequilibrium as a signature of selective sweeps. *Genetics*, **167**, 1513–1524.

Kim Y, Stephan W (2002) Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics*, **160**, 765–777.

Kimura M (1953) Stepping-stone'' model of population. *Annual Report to the National Institute of Genetics*, **3**, 62–63.

Lao O, de Gruijter JM, van Duijn K, Navarro A, Kayser M (2007) Signatures of positive selection in genes associated with human skin pigmentation as revealed from analyses of single nucleotide polymorphisms. *Annals of human genetics*, **71**, 354–369.

Le Corre V, Kremer A (2003) Genetic variability at neutral markers, quantitative trait land trait in a subdivided population under selection. *Genetics*, **164**, 1205–1219.

Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, **74**, 175–195.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981–994.

Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C (2005) Genomic scans for selective sweeps using SNP data. *Genome Research*, **15**, 1566–1575.

Park DJ, Lukens AK, Neafsey DE *et al.* (2012) Sequence-based association and selection scans identify drug resistance loci in the Plasmodium falciparum malaria parasite. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 13052–13057.

Peng B, Amos CI (2008) Forward-time simulations of non-random mating populations using simuPOP. *Bioinformatics*, **24**, 1408–1409.
Peng Y, Yang Z, Zhang H *et al.* (2011) Genetic variations in Tibetan populations and high-altitude adaptation at the Himalayas. *Molecular Biology and Evolution*, **28**, 1075–1081.
Pennings PS, Hermisson J (2006a) Soft sweeps II–molecular population genetics of adaptation from recurrent mutation or migration. *Molecular Biology and Evolution*, **23**, 1076–1084.
Pennings PS, Hermisson J (2006b) Soft sweeps III: the signature of positive selection from recurrent mutation. *PLoS Genetics*, **2**, 186.
Pickrell JK, Coop G, Novembre J *et al.* (2009) Signals of recent positive selection in a worldwide sample of human populations. *Genome Research*, **19**, 826–837.
Price AL, Butler J, Patterson N *et al.* (2008) Discerning the ancestry of European Americans in genetic association studies. *PLoS Genetics*, **4**, e236.
Pritchard JK, Pickrell JK, Coop G (2010) The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology*, **20**, R208–R215.
Qanbari S, Gianola D, Hayes B *et al.* (2011) Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle. *BMC Genomics*, **12**, 318.
Sabeti PC, Reich DE, Higgins JM *et al.* (2002) Detecting recent positive selection in the human genome from haplotype structure. *Nature*, **419**, 832–837.
Sabeti PC, Varilly P, Fry B *et al.* (2007) Genome-wide detection and characterization of positive selection in human populations. *Nature*, **449**, 913–918.
Santiago E, Caballero A (2005) Variation after a selective sweep in a subdivided population. *Genetics*, **169**, 475–483.
Shendure J, Ji H (2008) Next-generation DNA sequencing. *Nature Biotechnology*, **26**, 1135–1145.
Sing T, Sander O, Beerenwinkel N, Lengauer T (2005) ROCR: visualizing classifier performance in R. *Bioinformatics*, **21**, 7881.
Slatkin M, Wiehe T (1998) Genetic hitch-hiking in a subdivided population. *Genetical Research*, **71**, 155–160.
Smith JM, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genetical Research*, **23**, 23–35.
Voight BF, Kudaravalli S, Wen X, Pritchard JK (2006) A map of recent positive selection in the human genome. *PLoS biology*, **4**, 72.
Wright S (1931) Evolution in Mendelian Populations. *Genetics*, **16**, 97–159.
Wright S (1990) Evolution in Mendelian populations. 1931. *Bulletin of Mathematical Biology*, **52**, 241–295. discussion 201–247.
Zeng K, Shi S, Wu C (2007) Compound tests for the detection of hitchhiking under positive selection. *Molecular Biology Evolution*, **24**, 1898–1908.
Zhang J, Nielsen R, Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular Biology Evolution*, **22**, 2472–2479.
Zhong M, Lange K, Papp JC, Fan R (2010) A powerful score test to detect positive selection in genome-wide scans. *European Journal of Human Genetics*, **18**, 1148–1159.
Zhong M, Zhang Y, Lange K, Fan R (2011) A cross-population extended haplotype-based homozygosity score test to detect positive selection in genome-wide scans. *Statistics and Its Interface*, **4**, 51–63.

All authors contributed to the study design and preparation of the manuscript. A.V. wrote the scripts to run simuPOP and conducted the analyses; O.E.G. was in charge of the overall supervision of the project.

Data accessibility

The code, user manual of the code and an example data set are available on https://github.com/alexvat/simulations.

Supporting information

Additional supporting information may be found in the online version of this article.

Appendix S1. Genome scan methods.

Fig. S1. The three population structure scenarios considered are: (a) the island model, (b) the one-dimensional stepping-stone model, and (c) the hierarchical island model. Selection is present only in population Y while the other populations are neutral.

Fig. S2. Burn-in period for the island model with four demes with 2500 individuals each, 101 loci and 0.00375 cM/kb recombination rate, migration rate 0.01 and selection coefficient 0.1 ($2Nes = 500$).

Fig. S3. Burn-in period for the stepping stone model with four demes with 2500 individuals each, 101 loci and 0.00375 cM/kb recombination rate, migration rate 0.01 and selection coefficient 0.1 ($2Nes = 500$).

Fig. S4. Burn-in period for the hierarchical model with four demes with 2500 individuals each, 101 loci, migration rate 0.02 between populations within the same group and 0.01 between groups, selection coefficient 0.1 ($2Nes = 500$) and 0.00375 cM/kb as recombination rate.

Fig. S5. Mean score (y-axis, blue line) for each of the 101 loci (x-axis) over the 1000 replicates for the following values of the allele frequency of the selected variant: *c.* 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.

Fig. S6. FDR for each method for the case of the hard sweep unde *r* the island model: (a) iHS; (b) nSL; (c) EHHST; (d) xp-EHH; (e) XP-EHHST; (f) XPCLR.

Fig. S7. Power for each method for the case of the hard sweep under the stepping stone model: (a) iHS; (b) nSL; (c) EHHST; (d) xp-EHH; (e) XP-EHHST; (f) XPCLR.

Fig. S8. Results of simulations of the stepping stone scenario with 52 populations with 500 individuals each, migration rate 0.05, selection coefficient 0.05 ($2Nes = 50$), 101 loci and recombination rate 0.00375 cM/kb.

Fig. S9. Power for the case of the soft sweep under the stepping stone model and for varying initial allele frequencies of the selected variant: 0.02 (black), 0.1 (grey), 0.2 (red), 0.4 (blue).

Fig. S10. Power and FDR for the case of the soft sweep under the hierarchical model with four final demes, iHS (black), hapFLK (blue), xp-EHH (grey), and XPCLR (red).

Fig. S11. Power and FDR for the case of the soft sweep in the hierarchical island model with no migration and with IAF 0.1. Results obtained with XPCLR (grey) using populations in different (Y–Z) continents and hapFLK (black).

Fig. S12. Power of hapFLK for the case of the hard sweep in the island model using 100 simulations. Results obtained from 4 populations (2500 individuals each), 101 loci, 0.1 ($2Nes$ = 500) selection coefficient and 0.00375 cM/kb as recombination rate.

Fig. S13. ROC curves (x-axis is on a log scale) for the island model with four demes with 2500 individuals each, 101 loci and 0.00375 cM/kb recombination rate, migration rate 0.01 and selection coefficient 0.1 ($2Nes$ = 500) for (a) iHS (b) nSL (c) EHHST (d) xp-EHH (e) XPEHHST and (f) XPCLR for all the different allele frequencies of the selected locus (50) in the selected population (Y).

Fig. S14. ROC curves (x-axis is on a log scale) for the island model with four demes with 2500 individuals each, 101 loci and 0.00375 cM/kb recombination rate, migration rate 0.05 and selection coefficient 0.1 ($2Nes$ = 500) for (a) iHS (b) nSL (c) EHHST (d) xp-EHH (e) XPEHHST and (f) XPCLR for all the different allele frequencies of the selected locus (50) in the selected population (Y).

Fig. S15. ROC curves (x-axis is on a log scale) for the island model with four demes with 2500 individuals each, 101 loci and 0.00375 cM/kb recombination rate, migration rate 0.008 and selection coefficient 0.1 ($2Nes$ = 500) for (a) iHS (b) nSL (c) EHHST (d) xp-EHH (e) XPEHHST and (f) XPCLR for all the different allele frequencies of the selected locus (50) in the selected population (Y).

Fig. S16. ROC curves (x-axis is on a log scale) for the island model with four demes with 2500 individuals each, 101 loci and 0.00375 cM/kb recombination rate, migration rate 0.008 and selection coefficient 0.08 ($2Nes$ = 400) for (a) iHS (b) nSL (c) EHHST (d) xp-EHH (e) XPEHHST and (f) XPCLR for all the different allele frequencies of the selected locus (50) in the selected population (Y).

Fig. S17. ROC curves (x-axis is on a log scale) for the island model with four demes with 2500 individuals each, 101 loci and 0.00375 cM/kb recombination rate, migration rate 0.008 and selection coefficient 0.01 ($2Nes$ = 50) for (a) iHS (b) nSL (c) EHHST (d) xp-EHH (e) XPEHHST and (f) XPCLR for all the different allele frequencies of the selected locus (50) in the selected population (Y).

Fig. S18. ROC curves (x-axis is on a log scale) for the stepping stone model with four demes with 2500 individuals each, 101 loci and 0.00375 cM/kb recombination rate, migration rate 0.01 and selection coefficient 0.1 ($2Nes$ = 500) for (a) iHS (b) nSL (c) EHHST (d) xp-EHH (e) XPEHHST and (f) XPCLR for all the different allele frequencies of the selected locus (50) in the selected population (Y).

Fig. S19. ROC curves (x-axis is on a log scale) for the stepping stone model with four demes with 2500 individuals each, 101 loci and 0.00375 cM/kb recombination rate, migration rate 0.05 and selection coefficient 0.1 ($2Nes$ = 500) for (a) iHS (b) nSL (c) EHHST (d) xp-EHH (e) XPEHHST and (f) XPCLR for all the different allele frequencies of the selected locus (50) in the selected population (Y).

Fig. S20. ROC curves (x-axis is on a log scale) for the stepping stone model with four demes with 2500 individuals each, 101 loci and 0.00375 cM/kb recombination rate, migration rate 0.008 and selection coefficient 0.1 ($2Nes$ = 500) for (a) iHS (b) nSL (c) EHHST (d) xp-EHH (e) XPEHHST and (f) XPCLR for all the different allele frequencies of the selected locus (50) in the selected population (Y).

Fig. S21. ROC curves (x-axis is on a log scale) for the stepping stone model with four demes with 2500 individuals each, 101 loci and 0.00375 cM/kb recombination rate, migration rate 0.008 and selection coefficient 0.08 ($2Nes$ = 400) for (a) iHS (b) nSL (c) EHHST (d) xp-EHH (e) XPEHHST and (f) XPCLR for all the different allele frequencies of the selected locus (50) in the selected population (Y).

Fig. S22. ROC curves (x-axis is on a log scale) for the stepping stone model with four demes with 2500 individuals each, 101 loci and 0.00375 cM/kb recombination rate, migration rate 0.008 and selection coefficient 0.01 ($2Nes$ = 50) for (a) iHS (b) nSL (c) EHHST (d) xp-EHH (e) XPEHHST and (f) XPCLR for all the different allele frequencies of the selected locus (50) in the selected population (Y).

Fig. S23. ROC curves (x-axis is on a log scale) for the hierarchical model with four demes with 2500 individuals each, 101 loci, migration rate 0.02 between populations within the same group and 0.01 between groups, selection coefficient 0.1 ($2Nes$ = 500) and 0.00375 cM/kb as recombination rate for (a) iHS (b) nSL (c) xp-EHH (d) XPCLR and (e) hapFLK for all the different allele frequencies of the selected locus (50) in the selected population (Y).

Fig. S24. Evaluation of power of XPCLR under the hard sweep scenario and an island population structure with four demes with 2500 individuals each, 101 loci and 0.00375 cM/kb recombination rate, m = 0.008 and selection s = 0.1 ($2Nes$ = 500) coefficient with (a) 100 simulations (black) and 1000 simulations (grey). We observe that the difference between the two assessments is negligible, thus 100 simulations are sufficient to draw robust conclusions.