Schneider, K., White, T. J., Mitchell, S., Adams, C. E., Reeve, R. and Elmer, K. R. (2021) Pitfalls and virtues of population genetic summary statistics: detecting selective sweeps in recent divergences. *Journal of Evolutionary Biology*, 34(6), pp. 893-909.

(doi: [10.1111/jeb.13738](https://doi.org/10.1111/jeb.13738))

The material cannot be used for any other purpose without further permission of the publisher and is for private use only.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

This is the peer reviewed version of the following article:

Schneider, K., White, T. J., Mitchell, S., Adams, C. E., Reeve, R. and Elmer, K. R. (2021) Pitfalls and virtues of population genetic summary statistics: detecting selective sweeps in recent divergences. *Journal of Evolutionary Biology*, 34(6), pp. 893-909, which has been published in final form at: [10.1111/jeb.13738](https://doi.org/10.1111/jeb.13738)

This article may be used for non-commercial purposes in accordance with [Wiley Terms and Conditions for Self-Archiving](#).

<https://eprints.gla.ac.uk/226213/>

Deposited on: 19 November 2021

**The pitfalls and virtues of population genetic summary statistics: detecting
selective sweeps in recent divergences**

Kevin Schneider¹, Tom J. White¹, Sonia Mitchell¹, Colin E. Adams^{1,2}, Richard Reeve¹, Kathryn
R. Elmer^{1*}

¹ *Institute of Biodiversity, Animal Health & Comparative Medicine, College of Medical,
Veterinary & Life Sciences, University of Glasgow, Glasgow, G12 8QQ, UK*

² *Scottish Centre for Ecology and the Natural Environment, Institute of Biodiversity, Animal
Health and Comparative Medicine, College of Medical, Veterinary & Life Sciences, University
of Glasgow, Rowardennan, Glasgow, G63 0AW, UK*

** corresponding author; Kathryn.Elmer@glasgow.ac.uk*

Abstract

During evolution, genomes are shaped by a plethora of forces that can leave characteristic signatures. A common goal when studying diverging populations is to detect the signatures of selective sweeps, which can be rather difficult in complex demographic scenarios, such as under secondary contact. Moreover, the detection of selective sweeps, especially in whole-genome data, often relies heavily on a narrow set of summary statistics that are affected by a multitude of factors, frequently leading to false positives and false negatives. Simulating genomic regions makes it possible to control these demographic and population genetic factors. We used simulations of large genomic regions to determine how different secondary contact and sympatric speciation scenarios affect the footprint of hard and soft selective sweeps in the presence of varying degrees of gene flow and recombination. We explored the ability of an array of population genetic summary statistics to detect the footprints of these selective sweeps under specific demographies. We focussed on metrics that do not require phased data or ancestral sequences and therefore have wide applicability. We found that a newly developed beta diversity measure, $\bar{B}^{GD}$, outperformed all other metrics in detecting selective sweeps and that F_{ST} also performed well. High accuracy was also found in $\Delta\pi$ and genotype distance-derived metrics. The performance of most metrics strongly depended on factors such as the presence of an allopatric phase, migration rates, recombination, population growth, and whether the sweep was hard or soft. We provide suggestions for locating and analysing the response to selective sweeps in whole-genome data.

Keywords: adaptation; beta diversity; demography; genome; negative selection; positive selection; secondary contact; simulation; sympatry; speciation

Introduction

Evolutionary processes commonly act on genomes in a concerted manner and often leave characteristic signatures. Frequently studied examples include positive selection in the form of soft and hard selective sweeps (Hermisson and Pennings 2005; Pritchard et al. 2010), background selection (Charlesworth et al. 1995), recombination (Bamshad and Wooding 2003), and introgression (Hufford et al. 2013; Moest et al. 2020). While these responses and processes leave typical signatures on patterns of diversity across the genome, the signatures can become eroded and are therefore increasingly difficult to capture over time. Many studies have attempted to screen for signatures (such as those of positive selection) in ongoing or recent evolutionary events in nature (Gagnaire et al. 2013; Malinsky et al. 2015; Torres-Dowdall et al. 2015; Lamichhaney et al. 2018; Rettelbach et al. 2019; Jacobs et al. 2020) and in experimental evolution settings (Dunham et al. 2002), such as in the form of evolve-and-resequence experiments (E&R) (Long et al. 2015; Schlötterer et al. 2015).

Because evolutionary responses are variable, complex, and ultimately unknowable in natural settings, simulating genomic regions *in silico* is an approach to dissecting the influence of particular evolutionary forces on genomes. This makes it possible to control evolutionary forces rigorously and determine the effect of particular modifications. Simulation studies have led to important insights about the genomic footprints of evolutionary forces such as natural selection (Lohmueller 2014; Matthey-Doret and Whitlock 2019). Empirical studies on population genomics and evolutionary biology can use simulations to obtain informed estimates of genome signatures under given evolutionary scenarios and compare the simulated and empirical patterns (Enard et al. 2014; Rettelbach et al. 2019).

Despite the incorporation of various demographic models in *in silico* and empirical studies, many demographic scenarios have so far received comparatively little attention. An example is secondary contact, in which two populations become sympatric again after an allopatric phase. This scenario has been shown to be particularly common in recent divergences, many of which are affected by the expansion and retreat of glaciers, including many temperate region animal model systems in ecological and evolutionary genomics (Schluter 1996; Pettengill and Moeller 2012; Gagnaire et al. 2013; Elmer 2016). Other examples include secondary contact after range expansion in ecotypes of the plant model

systems *Mimulus* sp. and *Clarkia* sp. (Streisfeld and Kohn 2005; Pettengill and Moeller 2012; Kenney and Sweigart 2016). Secondary contact scenarios can be rather complex because they include elements of allopatric and sympatric speciation in varying proportions (Jacobs et al. 2020), complicated by processes such as reinforcement (Noor 1999) and character displacement (Brown and Wilson 1956), which accentuate differences accumulated in allopatry. Sympatric speciation, on the other hand, does not include an allopatric phase. Instead, divergence takes place under actual or potential gene flow between incipient species or subpopulations (Foote 2018).

While whole-genome sequence data are the most informative for discerning evolutionary processes, screening such data sets for genomic responses to selection and recent divergences can be challenging. Very large amounts of data are difficult to analyse with model-based approaches, such as likelihood-based and Bayesian sweep-finding methods (Nielsen et al. 2005; Jensen et al. 2008; Pavlidis et al. 2013; Huber et al. 2016), due both to computational limitations and to difficulties in comparing across study systems (Hämälä and Savolainen 2019; He et al. 2019). To reduce sequencing costs and maximise numbers of individuals in ecological and evolutionary model systems, in many cases low coverage sequence data are used to infer genotype likelihoods, in which all possible genotypes at each nucleotide position in the genome are assigned a likelihood value. Currently most model-based methods are not suitable for genotype likelihoods and rely on individual-level called genetic variants, where polymorphisms in individuals are known with high certainty (Nielsen et al. 2011; Korneliussen et al. 2014). Additionally, sequence data of low to moderate coverage are usually not suited for phasing or haplotype inference, which are prerequisites for many metrics to infer selection or other evolutionary processes (Sabeti et al. 2007; Alachiotis et al. 2012; Ferrer-Admetlla et al. 2014). Lastly, many commonly used metrics to infer selection require a polarised site frequency spectrum, which requires knowledge about the ancestral sequence (Fay and Wu 2000) that may or may not be available.

To overcome these limitations, summary statistics are commonly used, often in concert, to screen for characteristic signals of selective sweeps. Population genetic metrics applied across the genome are widely used by evolutionary biologists to determine regions under selection compared to background divergence, such as *Tajima's D* (Tajima 1989), which can be used to detect positive selection, and F_{ST} (Wright 1951), which tests for differentiation between populations but can also be used to screen for differential selection

(Beaumont 2005; Holsinger and Weir 2009). The candidate genomic regions can then be evaluated for functional genes and potential function-altering substitutions (Ng and Henikoff 2006; Barrett and Hoekstra 2011) or combined with evidence from phenotypic data (Stinchcombe and Hoekstra 2008; Barrett et al. 2019) to further infer their evolutionary significance.

A major drawback of applying population genetic summary statistics to regions of the genome is that they can be influenced by processes other than natural selection, such as gene flow, recombination, population structure, and demography (Stajich and Hahn 2005). In the case of F_{ST} for example, several studies have demonstrated how false positives can result from the demographic history of a population (Lotterhos and Whitlock 2014), differential introgression (Gosset and Bierne 2013), hierarchical population structure (Excoffier et al. 2009), variation in recombination rates within the genome (Lotterhos 2019), and background selection (Matthey-Doret and Whitlock 2019). Similarly, *Tajima's D* has been shown to produce biased estimates of selection depending on population history (Stajich and Hahn 2005), population structure (Moeller et al. 2007), and recombination rate variation (Thornton 2005). The influence of demography, gene flow, and genomic structure on other summary statistics, particularly those that have been developed recently, is not well understood (Ravinet et al. 2017).

Metrics that draw upon whole-genome data commonly make use of the following genomic properties: 1) The site frequency spectrum, which is the distribution of allele frequencies at a number of loci in the genome, shifts in response to selection due to genomic hitchhiking (Nielsen et al. 2005) but also can be influenced by demographic factors (Stajich and Hahn 2005). 2) The distribution of genotype distances g_{kl} is a recent approach that essentially captures sequence divergence in particular genomic regions and has to date only been assessed using a small set of evolutionary scenarios (Kern and Schrider 2018). 3) Multilocus genotype frequency metrics, which are similar to haplotype frequency statistics for unphased data (Messer and Petrov 2013; Garud et al. 2015), can also be used to infer sweeps but have been shown to have little power in the case of soft sweeps (Harris et al. 2018; Kern and Schrider 2018). 4) Linkage disequilibrium can be used to detect sweeps, but metrics have primarily relied on phased data (Kelly 1997; Kim and Nielsen 2004), while recently these have been extended to unphased data (Kern and Schrider 2018; Schlamp et al. 2016). However, Kim and Nielsen (2004) pointed out that the relative power to detect sweeps

is weaker using linkage disequilibrium compared to methods based on allele frequencies (e.g. the site frequency spectrum). 5) Differentiation between populations can also be used to infer differential selection but can be confounded by demographic processes (Chen et al. 2010; Lotterhos and Whitlock 2014). Until now, a full assessment of the relative performances of all of these metrics to detect selective sweeps under a range of evolutionary scenarios has not been made.

To better understand the characteristic patterns of selective sweeps and the relative performance of summary statistics under different demographic scenarios, we performed simulations of how genomes diverge between populations under an array of evolutionary and demographic scenarios. These include positive selection in one population and negative selection of varying degrees in the other population, reflecting differential selection and adaptation as a consequence of ecological specialisation (Poisot et al. 2011). We did this in the versatile simulation environment SLiM (Messer 2013; Haller and Messer 2017), which allows for bespoke adjustment of genetic and evolutionary processes. Sixteen summary statistics were assessed with regard to their ability to detect the selective sweeps that we simulated: π , $\Delta\pi$, $\hat{\theta}_W$, *Tajima's D*, the variance, skewness, and kurtosis of the genotype distance distribution g_{kl} (Kern and Schrider 2018), the multilocus genotype frequency-based *nDiplos*, J_1 , J_{12} , and J_2/J_1 (Messer and Petrov 2013; Garud et al. 2015; Kern and Schrider 2018), the genotypic linkage disequilibrium-based *diploZ_{NS}* and *diplo ω* (Kelly 1997; Kim and Nielsen 2004; Kern and Schrider 2018) as well as *H – scan* (Schlump et al. 2016), F_{ST} , and a recently developed beta diversity measure based on information theory (Reeve et al. 2016) newly adapted to this genomic context, $\bar{B}^{GD}$ (White 2019).

We compared the relative accuracy of the 16 metrics above in predicting the location of the simulated selective sweep under the series of different evolutionary and demographic histories. We calculated accuracies based on 1) the correct identification of the location at which the sweep occurred, detected as peaks or valleys of summary statistics, and 2) significant signals of selection when comparing to distributions of summary statistics derived from neutral simulations. We found large differences in performance among metrics, with some summary statistics outright failing to detect the location of selection under particular demographic scenarios. The newly developed genetic beta diversity measure, $\bar{B}^{GD}$, outperformed all other metrics we assessed, including F_{ST} . Based on our simulation results,

we provide some guidance as to which metrics are useful for sweep prediction, especially in the case of populations with complex, recent histories of divergence.

Materials & Methods

Simulations

The scripts for all simulations were written in the Eidos language (Haller 2016) for simulation in SLiM 2 (Haller and Messer 2017). A total of 144 scenarios were simulated: 96 scenarios that included a selective sweep and 48 scenarios without selection (general outline: Figure 1; all parameter combinations: Supplementary Table S1). These scenarios modelled either sympatric speciation (i.e. speciation-with-gene-flow and no isolation in the initial stages of divergence) or secondary contact (i.e. an allopatric phase during the initial stage of population divergence followed by gene flow at secondary contact). The parameter settings were meant to reflect recent divergence scenarios, for example speciation following colonisations of lakes or other disjunct habitats. Generations in all simulations were modelled as being discrete and the simulation duration was 53,000 generations. Individuals in post-divergence populations are offspring of pre-divergence populations.

To build up genetic variation, all scenarios involved a 50,000-generation pre-divergence period for a population of 1000 individuals (hereafter “population 1”). Sympatric speciation simulations (hereafter “scenario 1”) then involved divergence into two populations initially of 500 individuals, with each population experiencing low levels of gene flow for another 3000 generations until the simulations ended. In some contexts, sympatry with low levels of gene flow rather than panmixia is often called parapatry (Gavrilets 2004), but we will call it sympatry in the following since ecological divergence of geographically sympatric populations typically entails a reduction in gene flow of varying rates (Rundle and Nosil 2005; Nosil 2012), while parapatry could also imply overlapping but not completely sympatric populations in a biogeographic sense (Schuler et al. 2016). Secondary contact experiments involved the same divergence into two populations of 500 individuals; thereafter each population experienced a 1,500-generation allopatric phase, followed by a 1,500-generation period with gene flow until the simulations ended.

Selection: In all scenarios that included selection (the neutral simulations are described below), the pre-divergence population ("population 1") was not subject to any selection, one of the post-divergence populations (hereafter "population 2") experienced relatively strong positive selection (selection coefficient $s = 0.1$; relative fitness of offspring per generation $f = 1 + s = 1.1$ for both homozygote and heterozygote carriers of the selected mutation; see http://benhaller.com/slim/SLiM_Manual.pdf) at two loci (one in a region with low and one in a region with standard recombination rates), while the other population (hereafter "population 3") experienced negative selection of varying strength caused by the same mutations that are positively selected in population 2. Depending on the relative fitness, f , of mutations in population 3, secondary contact scenarios are designated as "scenario 2" ($f = 1 / \text{relative fitness of population 2} = 0.909$), "scenario 3" ($f = (1 / \text{relative fitness of population 2}) - 0.1 = 0.809$), or "scenario 4" ($f = (1 / \text{relative fitness of population 2}) - 0.5 = 0.409$, i.e. substantially reduced fitness). The chosen selection coefficient of $s = 0.1$ has been shown to be realistic and common in a meta-analysis of selection coefficients in natural populations (Thurman and Barrett 2016). The rationale for including stronger negative selection in scenarios 3 and 4 was to mimic the effects of an allele of one population being confronted with a different genomic background in another population upon secondary contact, which can lead to lower fitness in hybrids or even hybrid sterility/inviability (Fang et al. 2012; Crespi and Nosil 2013; Sweigart et al. 2019).

Demography: In all scenarios the migration rates, m , the number of exchanged individuals per generation, were set to 0.1, 0.01, or 0.001 across different simulations. Migration rates here can be interpreted as the fraction of individuals of one population contributing to offspring in the next generation of the other population. All scenarios were modelled either with population sizes held constant at 500 individuals per population after population divergence or population sizes logistically growing from 500 individuals to 2500 individuals per population. Population sizes in population 3 were either the same as population 2 or higher, with 2000 individuals in the case of constant population sizes or growing to 10,000 individuals in the case of increasing population sizes. All combinations of parameter values and scenarios were tested. The effective population sizes and parameter settings used are meant as a compromise between more extreme sizes and values and account for limitations due to computational feasibility.

Selective sweep: Scenarios were modelled with 'population 2' experiencing a hard sweep, a soft sweep, or no selection (neutral evolution). Soft sweeps were modelled as selective sweeps from standing genetic variation in 'population 2', i.e. the mutation was introduced in 400 of 1000 individuals at generation 50,000 (immediately before divergence into populations 2 and 3) with a starting frequency of 0.4 in both post-divergence populations, which ensures that the selected allele is neither lost nor fixed, while at the same time making sure that starting points are equal across repetitions. Hard sweeps were modelled as one single point mutation in one individual in population 2. The onset of selection was always at generation 50,001 and continued to the end of the simulation. The selected mutations were modelled with a dominance coefficient, $h = 1.0$, denoting complete dominance, which is thought to be common in newly arising beneficial mutations (Haldane 1927; Turner 1977; Charlesworth 2017).

Genome structure: Genomes were simulated with a size of 4,000,050 bp. These consisted of two recombination regions of 2,000,025 bp each: one 'standard' recombination rate region (or probability) representative of many fish species, $r = 1e-7$, and one low recombination rate region with $r = 1e-8$ (Stapley et al. 2017). The recombination rates are the probabilities of recombination per nucleotide site per generation. The mutation rate in all simulations was $1e-8$ mutations per position per generation. The selective sweep mutations were always located in the centre of each recombination rate region, at positions 1,000,013 bp (in the standard recombination rate region) and 3,000,038 bp (in the low recombination rate region). The two recombination rate regions were completely independent (i.e. a selective sweep in one region did not affect the patterns in the other region) and can therefore be considered as separate chromosomes.

In total, simulations were replicated 100 times for each of the 144 different combinations of parameter values and scenarios (Supplementary Table 1). The simulated genomes were output as files in both ms (Hudson 2002) and vcf format, depending on analysis software requirements and computational efficiency of the summary statistic analyses.

Summary statistics

We focussed on summary statistics that are computationally efficient, easy to apply, and do not require phased data or ancestral sequence information. Since machine learning methods to infer selective sweeps, such as evoNet (Sheehan and Song 2016) and S/HIC or diploS/HIC

(Schridder and Kern 2016; Kern and Schridder 2018), usually require the simulation of training and testing data sets, drastically increasing the computational investment, we did not include these in the present study. Based on randomly sampling 30 individuals each of population 2 and population 3 from generation 53,000 of the simulations, we calculated summary statistics in 50 non-overlapping windows of 80,001 bp size using the ms or vcf simulation output files. The window size of 80,001 bp was found to lead to good resolution of signatures of selection after initial trials of various window sizes between 50,000 bp and 100,001 bp. Non-overlapping windows were used due to the selected allele being located in the centre of the central window of each recombination region, so problematic edge effects (i.e. mutation at the edge of a window leads to two minor peaks rather than one major peak) of selective signatures on empirical data can be excluded. Non-overlapping windows also allowed for faster computation times. To avoid biases due to definition of windows, we additionally defined windows using a total of 22 SNP-based windows (11 per recombination rate region) holding 21 SNPs per window, which were held constant and are based on the minimum number of SNPs present in each of the simulation runs. The chosen number also ensured that the central window had the selected mutation in its centre and is therefore compatible with the 1-in-1 accuracy method described below. These windows holding 21 SNPs each were calculated based on individual SNP metric values for each position using a custom R script. Individual SNP metric values could not be calculated for the linkage disequilibrium-based metrics *diploZ_{NS}* and *diplo ω* of the diploSHIC software. The lower number of windows using the SNP-based approach than in the physical size approach (11 compared to 25 per recombination rate region, respectively) is expected to increase baseline accuracies for the 1-in-1 accuracy method, as these accuracies constitute proportions of the most extreme values observed.

We used a total of 16 population genetic summary statistics grouped into five families based on the type of genomic information they use. The five families are 1) metrics based on the site frequency spectrum (SFS): nucleotide diversity π (Tajima 1983), the difference in nucleotide diversity between the two post-divergence populations $\Delta\pi$, $\hat{\theta}_w$ (Watterson 1975), and *Tajima's D* (Tajima 1989); 2) metrics based on genotype distance: the variance, skewness, and kurtosis of the genotype distance distribution g_{kl} (Kern and Schridder 2018); 3) metrics based on the distribution of multilocus genotype frequencies: *nDiplos* (the number of distinct genotype vectors), J_1 , J_{12} , and J_2/J_1 (or G_1 , G_{12} , and G_2/G_1 , the multilocus

genotype equivalents of the haplotype-based metrics H_1 , H_{12} , H_2/H_1 ; Messer and Petrov 2013; Garud et al. 2015; Harris et al. 2018); 4) metrics based on linkage disequilibrium (LD): $diploZ_{NS}$ and $diplo\omega$ (the genotypic LD equivalents of Z_{NS} (Kelly 1997) and ω (Kim and Nielsen 2004)) as well as $H - scan$ (Schlamp et al. 2016); and 5) measures of population differentiation: F_{ST} (Wright 1951) and the genetic beta diversity measure $\bar{B}^{GD}$ (Reeve et al. 2016; White 2019; described below).

Metrics based on the SFS (family 1) make use of the distribution of allele frequencies in the genome. Nucleotide diversity π is defined as the number of nucleotide differences per site between two randomly chosen sequences from a population (Tajima 1983). The Watterson estimator $\hat{\theta}_W$ measures the population mutation rate, which is the product of the effective population size and the mutation rate (Watterson 1975). *Tajima's D* is defined as the average number of pairwise differences divided by the number of segregating sites (Tajima 1989). A negative *Tajima's D* indicates an excess of rare alleles, which can be caused by a recent selective sweep. In metrics based on genotype distance (family 2), the genotype distance distribution g_{kl} between individuals k and l is defined as $g_{kl} = \sum_{i=0}^m 1_{x_{ki} \neq x_{li}}$, which is the sum of unequal genotypes in a genotype vector x_k (Kern and Schrider 2018), where the multilocus genotype $x_k = \{x_{k0}, x_{k1}, \dots, x_{km}\}$ denotes the set of genotypes in individual k . In metrics based on the distribution of multilocus genotype frequencies (family 3), $nDiplos$ is simply the number of distinct multilocus genotypes in a population. J_1 is the multilocus genotype homozygosity calculated using all multilocus genotypes. J_{12} is the multilocus genotype homozygosity calculated using the two most frequent multilocus genotypes. J_2/J_1 is the ratio of multilocus genotype homozygosity calculated using all but the most frequent multilocus genotype to J_1 (Messer and Petrov 2013; Garud et al. 2015; Harris et al. 2018). Metrics based on LD (family 4) depend on the association of alleles at different loci in a population. $diploZ_{NS}$ uses multilocus genotypes to estimate the squared correlation of alleles at pairs of polymorphic sites (Kelly 1997; Kern and Schrider 2018). $diplo\omega$ is a related metric that is based on the sum of correlation coefficients between sets of polymorphic sites (Kim and Nielsen 2004), which can also be calculated from multilocus genotypes (Kern and Schrider 2018). $H - scan$ is a method that aims to detect selective sweep signatures by measuring the average length of pairwise homozygosity tracts along the

genome (Schlamp et al. 2016). The differentiation-based metric F_{ST} (family 5) measures the between-population variance in allele frequencies (Wright 1951).

All metrics except $H - scan$ and population differentiation (F_{ST} and $\bar{B}^{GD}$) were estimated using diploS/HIC on ms files for populations 2 and 3 separately and constitute standardised values (as described in Kern and Schrider 2018). For diploS/HIC metrics and $H - scan$, only the results for the population experiencing positive selection (population 2) are shown. $H - scan$ was calculated using the software provided with the original publication (Schlamp et al. 2016). F_{ST} and $\bar{B}^{GD}$ were calculated between populations 2 and 3 based on vcf files. F_{ST} was estimated using vcftools v0.1.17 (Danecek et al. 2011). $\bar{B}^{GD}$ was estimated using the R package rdiversity v2.0 (Reeve et al. 2016; Mitchell, Reeve, and White 2020).

$\bar{B}^{GD}$, a beta diversity measure using genetic distances derived from information theory $\bar{B}^{GD}$ is a measure of between-population diversity (i.e. beta diversity) based on genetic or genomic data (White 2019). $\bar{B}^{GD}$ is calculated from the similarity between individuals (see Reeve et al. 2016, Tables 1 and 2 for further details) and is mathematically independent from within-population diversity (Jost 2010). The similarity between individuals is defined as $z_{ij} = 1 - k \times n_{ij}$, where n_{ij} is the number of changes in sequence between individuals i and j in the genomic window being considered (with homozygotes for the opposite allele having 2 changes and heterozygotes compared to homozygotes having 1 change), and k is a scaling constant less than the inverse of the maximum number of changes observed (so $z_{ij} \geq 0$). In trials, the scaling factor $1/k$ has been set to 100, 1000, or was automatically determined by the maximum distance values. We found a value of $k = 1/1000$ to lead to the highest accuracies in sweep prediction for our data set. The resultant measure, $\bar{B}^{GD}$, derived from $\bar{B}$, the effective number of distinct populations (or communities) (Reeve et al 2016), measures the number of genetically distinct genomic windows (GD = Genetic Distance), which takes a value of 1 when the windows have exactly the same genotypes in both populations (i.e. no differentiation), and a maximum value of 2 when the windows are completely genetically distinct. A parameter q , $0 \leq q \leq \infty$, determines the relative importance of rare and common sequences in the analysis (see Hill (1973) and Reeve et al. (2016) for further details on q). Three values of q were tested (0, 1, and 2). Since the choice of q had a very minor effect on the results (Supplementary Tables S2, S3, S4), we only report $\bar{B}^{GD}$ for $q = 2$ in the main text.

Accuracies of sweep prediction

To determine how well each metric predicted the presence and location of mutations under selective sweeps, accuracies were calculated on the outcome across the 100 replicates in the following ways: 1) "1-in-1 window accuracy": the proportion of replicates in which the window containing the selected mutation (the central window) had the most extreme value (peak or valley/highest or lowest, depending on the metric) in each recombination region; 2) "1-in-3 windows accuracy": the proportion of replicates in which the window containing the selected mutation or its neighbouring one up- and one down-stream windows had the most extreme value, which is designed to also capture less distinct signals of selection. These accuracies therefore represent the proportion of simulation replicates with a summary statistic peak (or valley) in the central window of a recombination region, which was found to be a useful proxy for the observation of summary statistic peaks in the present study. We only report results based on the 1-in-1 window accuracy in the main manuscript, since we found that the relative performance of metrics was essentially the same using the 1-in-3 windows accuracy. For the analyses based on windows holding 21 SNPs, we only calculated 1-in-1 window accuracies (Supplementary Table S4 and Supplementary Figures S4–S7). In Supplementary File S1 and Supplementary Table S5 we also describe and report results from an accuracy metric based on comparisons of simulations with selection to neutral simulations, which are comparable to the proportion of simulation replicates with P values more extreme than in simulations of neutral scenarios. All analyses of accuracy were performed using R v3.5.1 (R Core Team 2018). R scripts, simulated genetic data, and SLiM scripts are deposited at <http://dx.doi.org/10.5525/gla.researchdata.981>.

Results

Patterns of selective sweeps across the genome

When averaging the summary statistics across replicates of each scenario, most metrics showed more or less clear peaks at or around the selected mutations (see Figure 2 for an example; see Supplementary File S2 for plots of all scenarios). These peaks were more distinct in the differentiation measures ($\bar{B}^{GD}$ and F_{ST}) compared to the other metrics used. Measures

based on genotype distance (variance, skewness, and kurtosis of g_{kl}), the multilocus genotype frequency metrics J_1 and J_{12} as well as $H - scan$ also frequently had distinct peaks, while other measures based on multilocus genotype frequencies (J_2/J_1 and $nDiplos$) and LD (in particular $diploZ_{NS}$ and $diplo\omega$) were more prone to noise, with less distinct peaks around the selected mutations (Figure 2; Supplementary File S2). The distinctness of peaks around the genomic region under selection is directly reflected in the sweep prediction accuracies reported in more detail below.

Accuracy in identifying sweep location across scenarios

The ability of the 16 different summary statistics to accurately infer selective sweeps varied dramatically across scenarios. Generally, the more distinct the peaks observed in the graphs were (Figure 2; Supplementary File S2), the higher the accuracy in predicting the true location of the selective sweep. Figure 3 shows results obtained for all four types of scenarios (sympatric speciation and three secondary contact scenarios) and differing population growth and migration parameters, under hard and soft sweep simulations, in the region with standard recombination rate in simulations with equal population sizes. $\bar{B}^{GD}$ outperformed all other metrics in finding the correct location of the mutation under selection. The mean accuracy of F_{ST} was, on average, slightly lower (Supplementary Tables S2–S4). Of the diploS/HIC selection metrics, genotype distance-based metrics performed best, especially the skewness and kurtosis of g_{kl} . Of the metrics based on the SFS, only $\Delta\pi$ had comparable accuracies to g_{kl} metrics. $H - scan$ had somewhat higher accuracies compared to the SFS-based $\hat{\theta}_W$ and other metrics based on LD, which had the lowest accuracies. Metrics based on multilocus genotype frequencies performed better than most LD metrics but not as well as the g_{kl} metrics. Additional results are given in supplementary figures for the low recombination rate region in simulations with equal population sizes and unequal population sizes (Supplementary Figures S1 and S2, respectively) and standard recombination rate region simulations with unequal population sizes (Supplementary Figure S3), with values listed in Supplementary Tables S2, S3. Results for the windows based on holding 21 SNPs mostly agreed with those based on physical size of windows and are given in Supplementary Table S4 and Supplementary Figure S4 for the standard recombination rate region with equal population sizes, Supplementary Figure S5 for the standard recombination rate region with unequal population sizes, Supplementary Figure S6 for the low recombination rate region

with equal population sizes, and Supplementary Figure S7 for the low recombination rate region with unequal population sizes. The SNP-based window approach led to overall higher accuracies compared to the physical window approach due to the lower number of windows used (11 compared to 25 per recombination rate region, respectively). The accuracies increased more in some metrics than in others, particularly in the case of π and the metrics based on multilocus genotype frequencies. However the overall pattern of performance was consistent to the window approach.

Prediction of sweep location – scenario-dependent performance

The prediction accuracies of demographic scenarios depended substantially on migration rate and the metric. For low migration rates ($m = 0.001$), all four demographic scenarios had similar prediction accuracies (Figure 3; Supplementary Figures S1–S7; Supplementary Tables S2–S4). For intermediate migration rates ($m = 0.01$), secondary contact scenarios with intermediate to strong negative selection in population 3 (scenarios 3 and 4) had the highest prediction accuracies. This was most pronounced for multilocus genotype frequency-based metrics, while differentiation-based metrics and $\Delta\pi$ had more similar prediction accuracies across scenarios in most cases. For high migration rates ($m = 0.1$), the secondary contact scenario with the strongest negative selection in population 3 (scenario 4) had the lowest prediction accuracies across metrics (Figure 3; Supplementary Figures S1–S7; Supplementary Tables S2–S4).

Of all parameters, migration rate had the strongest influence on accuracies of sweep location prediction. Intermediate migration rates ($m = 0.01$) between the two diverging populations led to higher accuracies of sweep prediction for differentiation-based metrics and $\Delta\pi$, with low migration rates ($m = 0.001$) having intermediate accuracies, and high migration rates ($m = 0.1$) having the lowest accuracies (Figure 3; Supplementary Figures S1–S7; Supplementary Tables S2–S4). However, the skewness of g_{kl} , Tajima's D , and $H - scan$ performed best at low migration rates, while the variance and kurtosis of g_{kl} as well as multilocus genotype frequency-based metrics performed equally well at low and intermediate migration rates. For almost all metrics and scenarios, highest migration rates led to the lowest accuracies, especially in the case of secondary contact scenarios with intermediate to strong negative selection in population 3 (scenarios 3 and 4). Differences in migration rate had the strongest effect on differentiation-based measures (F_{ST} and $\bar{B}^{GD}$). At

high migration rates, F_{ST} and $\bar{B}^{GD}$ had similarly low or only slightly higher accuracies compared to the other metrics. Accuracies were also similar in genotype distance-based and differentiation-based metrics at low migration rates and constant population sizes.

The influence of recombination rate on sweep prediction accuracies was comparably high to that of migration rate. Across all metrics, the accuracies were substantially higher in the standard recombination rate region compared to the low recombination rate region (Figure 3; Supplementary Figures S1–S7; Supplementary Tables S2–S4). What is more, for low recombination rates, secondary contact scenarios had lower prediction accuracies than sympatric divergence scenarios, while this was not always true for standard recombination rates. For measures based on differentiation and those based on LD, there was barely any difference between the accuracies of sweep prediction in hard sweep scenarios compared to soft sweep scenarios. However, the influence of sweep type was greater in most other metrics, with higher prediction accuracies in the case of hard sweeps especially in multilocus genotype frequency-based metrics.

Scenarios that included demographic growth led to higher accuracies in sweep prediction across metrics (Figure 3; Supplementary Figures S1–S7; Supplementary Tables S2–S4). Measures based on multilocus genotype frequencies and $\Delta\pi$ had higher prediction accuracies in scenarios with a larger population 3 (the population with negative selection), while for other metrics the prediction accuracies were mostly unchanged (Figure 3; Supplementary Figures S1–S7; Supplementary Tables S2–S4).

Discussion

The pros and cons of different metrics

Our results demonstrated that different summary statistics varied dramatically in their ability to infer simulated selective sweeps under the demographic conditions we evaluated. Differentiation-based measures F_{ST} and particularly the beta diversity measure $\bar{B}^{GD}$, which was evaluated robustly for the first time here, outperformed metrics based on the SFS (π , $\Delta\pi$, $\hat{\theta}_W$, and *Tajima's D*), genotype distance (the variance, skewness, and kurtosis of the genotype distance distribution g_{kl}), multilocus genotype frequencies (*nDiplos*, J_1 , J_{12} , and

J_2/J_1), and LD ($diploZ_{NS}$, $diplo\omega$, and $H - scan$) across the majority of evolutionary and demographic scenarios.

Accuracies based on identifying the correct genomic location where the selective sweep occurred were, on average across scenarios, highest in $\bar{B}^{GD}$ and second highest in F_{ST} , with the best performance at intermediate and the lowest at high migration rates. This was true across sweep types, population sizes, and sympatric and all secondary contact scenarios (i.e. scenarios 1 to 4). This is also evident in the observed distinctness of the peaks in these metrics centred around the selected mutations in a large number of scenarios. Of the metrics that are not based on regions of genomic differentiation, relatively high accuracy in predicting selective sweeps was also found in $\Delta\pi$ and those derived from the distribution of genotype distance. Multilocus genotype frequency-based metrics and $\Delta\pi$ performed best at intermediate migration rates but poorly at other migration rates. In some scenarios, especially at low migration rates, genotype distance-based metrics such as the skewness of g_{kl} performed similarly well as F_{ST} and $\bar{B}^{GD}$. High performance of genotype distance-based metrics (for both hard and soft sweeps) and multilocus genotype frequency metrics (for hard sweeps) has also been observed in other simulation studies and using empirical data (Kern and Schrider 2018). Using SNP-based windows, we observed a slight relative increase in performance for the metrics π and those based on multilocus genotype frequencies, while other metrics were less affected by how windows had been defined. The lowest accuracies were observed using LD-based metrics, especially $diploZ_{NS}$ and $diplo\omega$, which is in concordance with the observation of rather poor performance compared to methods based on allele frequency changes (i.e. SFS and F_{ST}) (Kim and Nielsen 2004). However, other methods based on LD, most of which require phased data, could have higher power (Vatsiou et al. 2016).

The influence of demography, recombination, and sweep type

Our simulations show that gene flow influences how accurately the genomic location of selective sweeps can be predicted in a non-linear way, as both low and high migration rates (low or high m) led to lower prediction accuracies for many metrics. In the case of low migration rates, this is likely a result of increased noise due to increased overall genome-wide differentiation between populations (Ravinet et al. 2017). On the other hand, high migration

rates can lead to the breakdown or slowdown of selective sweeps, even more so in scenarios with strong negative selection in the population without selective sweep (scenario 4). Therefore, in study systems with high ongoing gene flow, as has been observed in recent speciation events in the apple maggot *Rhagoletis pomonella* and in Midas cichlids (*Amphilophus* sp.) (Elmer et al. 2014; Egan et al. 2015; Kautt et al. 2016), the potential for inferring the presence and location of selective sweeps might be reduced and depends on the strength of selection and the presence of negative selection. Secondary contact scenarios, which all included an allopatric phase with zero migration before resuming gene flow after secondary contact, decreased the accuracies of predicting the sweep location when the recombination rate was low. These observations imply that in study systems that experienced secondary contacts, such as has been reported for *Cottus* sp. and *Mimulus* sp. (Nolte et al. 2006; Kenney and Sweigart 2016; Lucek et al. 2018), screens for selection will have reduced power in regions of the genome with reduced recombination. On the other hand, in systems with good evidence for sympatric speciation, as for instance in Midas cichlids, East African crater lake cichlids, and *Howea* palms (Savolainen et al 2006; Malinsky et al. 2015; Kautt et al. 2016; Torres-Dowdall et al. 2017; Osborne et al. 2019; Papadopoulos et al. 2019), screens for selective sweeps in genomic data will have increased power to detect signals of selection in low-recombination regions of the genome. The detection of selective sweeps in genomic regions with standard recombination rates, however, depends more on the particular combination of demographic history, migration rate, and the chosen metric. This shows that, ideally, determining the demographic history of a population, the recombination landscape, and the amount of gene flow should precede selection scans, which is particularly important in systems where the demographic history varies strongly from population to population (e.g. Jacobs et al. 2020).

Accuracies were lower in the low recombination rate region compared to the standard recombination rate region. This is consistent with the observation of false positive signals of selection caused by the decreased local recombination rate due to genomic rearrangements (e.g. inversions) or other influences on genome organisation and structure (Thornton 2005; Lotterhos 2019). A higher number of false positives should therefore also be expected in taxa with overall reduced recombination rates, as for example in opossum *Monodelphis domestica* and in *Oenothera* sp. (Mikkelsen et al. 2007; Rauwolf et al. 2011). The negative effect of lower recombination rate was similarly strong for all metrics, which is to be expected given that all

of these metrics are more or less based on a reduction in local diversity around the mutation under selection or the association of alleles (Pritchard et al. 2010; Messer and Petrov 2013).

As noted by Kern and Schrider (2018), metrics based on multilocus genotype frequencies are expected to have lower power to detect soft selective sweeps compared to hard selective sweeps, which is also what we observed here. This effect was somewhat smaller or was not observed at all for other metrics. Differentiation-based metrics in particular, showed relatively similar prediction accuracies for hard and soft sweeps. This is in concordance with various empirical and theoretical studies based on genome-scale data (Chen et al. 2010; Hohenlohe et al. 2010; Vitti et al. 2013; Pavlidis and Alachiotis 2017). The fact that a substantial number – and in many cases the majority – of sweeps constitute soft sweeps underlines the importance of using metrics that are robust to identifying regions under selection in both hard and soft sweeps (Pritchard et al. 2010; Messer and Petrov 2013; Schrider and Kern 2017). That some metrics are unaffected by soft selective sweeps while other metrics are influenced by both hard and soft sweeps could itself potentially be used to distinguish sweep type. Garud et al. (2015) and Harris et al. (2018) showed that the joint use of the metrics J_{12} and J_2/J_1 (called G_{12} and G_2/G_1 in Harris et al. 2018) enables the detection of and discrimination between hard and soft selective sweeps, which is reflected in our results to some extent. Machine learning approaches such as diploS/HIC (Kern and Schrider 2018), which makes use of the J_{12} and J_2/J_1 metrics as well, are also promising tools for the discrimination between signatures of hard and soft sweeps.

The utility of differentiation-based metrics

Differentiation-based metrics have found wide application in screens for positive selection (Chen et al. 2010; Narum and Hess 2011; De Villemereuil and Gaggiotti 2015). Despite its popularity and good performance in a range of situations, the use of F_{ST} has received criticism as well (Bierne et al. 2013; Cruickshank and Hahn 2014; Lotterhos and Whitlock 2014). The drawbacks of F_{ST} have been one of the motivations for the application of $\bar{B}$ to genetic data, as $\bar{B}^{GD}$ (White 2019). The better performance of $\bar{B}^{GD}$ compared to F_{ST} likely derives from two desirable properties that $\bar{B}^{GD}$ has but F_{ST} does not have. Firstly, $\bar{B}^{GD}$ was observed to be monotonically increasing with increasing time or strength of divergence (White 2019). F_{ST} , on the other hand, is not monotonic with increasing divergence time and strength (Wang 2012; White 2019). Secondly, $\bar{B}$ has the property that low within-population diversity does

not have an effect on the magnitude of $\bar{B}$ (Reeve et al. 2016), or that alpha diversity is mathematically independent from beta diversity (Jost 2010). Rather, only between-population diversity or differentiation affects the magnitude of $\bar{B}$. F_{ST} , however, is substantially affected by low within-population diversity between the two populations, potentially leading to false positive differentiation outliers, as has been shown for a number of studies (Cruickshank and Hahn 2014). These two properties should make $\bar{B}^{GD}$ the better choice if one wants to truly measure differentiation between populations, although further model-based and empirical assessments will be needed.

However, despite the good performance of differentiation-based measures, which could be expected given they *de facto* constitute comparisons to a reference, which is not true for the other (single-population) metrics, differentiation-based metrics are not always appropriate. In particular, $\bar{B}^{GD}$ and F_{ST} as well as $\Delta\pi$ require at least two (related) populations, one of which should be a reference population that does not experience selection at the same genomic location. This means that the signatures of selection will be lost if the selected mutation is also selected for in the reference population (i.e. in the case of global selection). In fact, such cases can even lead to reduced differentiation in the selected regions compared to neutral regions, which would be even more pronounced in low-migration scenarios (Zheng and Wiehe 2019). It is also possible for beneficial mutations in one population to be neutral in the other population, even though such cases might be less common than previously thought (Kern and Hahn 2018). In many study systems such a reference population, for example an ancestral population in a population pair from which a derived population diverges, is not available. An ideal scenario would be a population pair under divergence with distinct ecological preferences that lead to adaptation driven by differential selection. Examples for such a scenario are the divergences of benthic and pelagic ecomorphs in three-spined stickleback (*Gasterosteus aculeatus*) (Schluter and McPhail 1992; Bolnick and Lau 2008) or Arctic charr (*Salvelinus alpinus*) (Elmer 2016; Jacobs et al. 2020). Another potential drawback of differentiation-based metrics is the fact that it has been shown that hidden or hierarchical population structure can negatively affect the detection of natural selection based on differentiation measures (Excoffier et al. 2009). This implies that clear knowledge about the population structure at each level is required for sound application of differentiation-based metrics. It has also been pointed out that differences in sample size

among populations, the dependency on minor allele frequencies, and inconsistencies among estimators can decrease the reliability of F_{ST} (Berner 2019), which is why allele frequency differentiation (AFD) has been proposed as an alternative. We note that a detailed assessment of the effects of background selection, which has been shown to affect estimates of F_{ST} (Matthey-Doret and Whitlock 2019), and other confounding factors on the detection of selective sweeps, especially using $\bar{B}^{GD}$, should be performed in future simulation studies.

Potential caveats

In the current study, we did not account for some factors of empirical data that could potentially affect the performance of summary statistics to detect selective sweeps. This includes poor data quality, insufficient coverage, and inaccuracies (e.g. erroneously called SNPs), which have been shown to negatively impact population genetic estimates (Crawford and Lazzaro 2012).

A number of metrics were not included in this simulation study. Those excluded mostly were ones that belong to the same families of metrics that we included here, i.e. are based on or depend on the SFS, genotype distance, multilocus genotype frequencies, LD, or population differentiation. The decision only to include metrics that do not require phased data or a polarised SFS stems from the fact that phased data and ancestral sequences are not readily available and/or are difficult to obtain for most empirical studies, especially in non-model ecological and evolutionary study systems. The results presented in this paper highlight the importance of knowing the properties of different metrics under different demographic scenarios.

Suggested steps for locating selective sweeps

Drawing from our simulation results, empirical studies, and theory, we suggest an eight-step process that can help to locate and make sense of selective sweeps in whole-genome data (Figure 4):

- 1) Collect knowledge on the underlying demographic history of the studied populations to create a basis for subsequent analyses.
- 2) Genomic properties and parameters (i.e. recombination rate or the location of inversions) of a study system can often be estimated using available data (and in some cases even the same kind of data) to further inform downstream analyses.

3) Based on the knowledge gained in 1), choose a handful of metrics that proved suitable for a given demographic scenario. One of the most important observations from our exploration is that many metrics such as multilocus genotype frequency-based metrics and $\Delta\pi$ are most powerful at intermediate migration rates while genotype distance-based and differentiation-based metrics show a more consistent performance across migration rates. A combination of genotype distance-, multilocus genotype frequency-, differentiation-based metrics, and $\Delta\pi$ could be a powerful way of screening for selection in unphased genomic data. It should also be noted that, given the low accuracies of identifying the correct sweep location of many metrics under various demographic scenarios, complete overlap of candidate genomic regions under selection inferred using different methods can be considered unlikely. However, even partial overlap of candidate regions across metrics is stronger evidence for a selective sweep than that based on only one metric. Since metrics of the same family (e.g. metrics based on multilocus genotype frequencies) showed a correlated or collinear response to selection under a given demographic scenario, metrics in empirical studies should rather be chosen from multiple metric families that have proven to be powerful for a given scenario. In this context, the development of composite measures of selection that capture the combined signatures of different metrics (Lotterhos et al. 2017; Verity et al. 2017) will be an important future avenue for detecting selective sweeps.

4) If available, recombination rate estimates across the genome from 2) can further narrow down the choice of metrics for screening for selective sweeps. As this is not feasible in many cases, taking into account global recombination rate estimates of the same or related species can be useful when choosing suitable metrics. An important result is that, at low recombination rates, *H* – *scan* and multilocus genotype frequency-based metrics were more powerful than most other single-population metrics and could complement differentiation-based and genotype distance-based metrics.

5) Perform selection scans using the chosen population genetic metrics.

6) When potential locations of selective sweeps are identified, knowledge from 2) can be used to rule out factors that could lead to false positive signals of selection (e.g. inversions or other genomic rearrangements). However, the presence of inversions, for example, cannot rule out selective sweeps or other forms of selection. In fact, it has been shown in several study systems, from Atlantic cod (*Gadus morhua*) and rough periwinkles (*Littorina saxatilis*) to seep monkeyflowers (*Mimulus guttatus*), that inversions can harbour genes with high importance

for adaptation (Kirubakaran et al. 2016; Puzey et al. 2017; Morales et al. 2019; Refoyo-Martínez et al. 2019). In many cases, however, the location of inversions cannot be determined due to insufficient and/or unsuitable data.

7) The candidate genomic regions can then be analysed and explored, for example with regard to the functional genes or regulatory regions they contain. Sets of outlier genes in candidate regions can be analysed for gene ontology term or similar functional enrichment (Gaudet and Dessimoz 2017). Using available genome information, SNPs in these regions can be annotated, for example with software such as SnpEff (Cingolani et al. 2012). It can also be helpful to analyse allelic shifts of these SNPs in more detail (potentially also from other populations or related study systems). If available, candidate sweep regions could be analysed for overlap with quantitative trait loci in the same or related taxa (Stinchcombe and Hoekstra 2008; Jacobs et al. 2017).

8) Ultimately, while not currently feasible for most ecological and evolutionary non-model organisms, the last step could involve validating candidate sweep regions, genes, and alleles, for example using analyses of gene expression (Rahi et al. 2017; McGirr and Martin 2018), gene editing and knockouts (Fernandez i Marti et al. 2018; Chao et al. 2019), or computational methods for predicting functional changes such as modified enzyme function or protein structure (Tang and Thomas 2016; Lee et al. 2017).

Our results demonstrate that, if the conditions are met, differentiation-based approaches are a powerful way to detect selective sweeps in whole-genome data. In particular the measure $\bar{B}^{GD}$ proposed here, a beta diversity measure derived from information theory principles, proved to outcompete the other metrics used, making it a promising method for screening for selective sweeps and differentiation. Lastly, other recently described metrics (Kern and Schrider 2008), mostly those based on genotype distance, were shown here to be the most powerful non-differentiation metrics for screening for selective sweeps in unphased whole-genome data.

Acknowledgements

We thank A. Jacobs for valuable comments on earlier versions of the manuscript. The project was supported by a Fisheries Society of the British Isles (FSBI) PhD Studentship awarded to KS with KRE and CEA.

Author contributions

KS, KRE, and RR designed the study. KS performed the simulations. KS analysed the data, with contributions from TJW. TJW and SM wrote the beta diversity R package and associated beta diversity scripts. KS wrote the manuscript, with contributions from all authors. All authors read and approved the final version of the manuscript.

Data availability

All simulation and analysis scripts that are necessary to reproduce the results are available on the University of Glasgow Enlighten repository DOI: 10.5525/gla.researchdata.981. The beta diversity scripts are incorporated into the R package rdiversity v2.0 (Reeve et al. 2016; Mitchell, Reeve, and White 2020).

References

- Alachiotis, N., Stamatakis, A., & Pavlidis, P. (2012). OmegaPlus: a scalable tool for rapid detection of selective sweeps in whole-genome datasets. *Bioinformatics*, 28(17), 2274-2275.
- Bamshad, M., & Wooding, S. P. (2003). Signatures of natural selection in the human genome. *Nature Reviews Genetics*, 4(2), 99-110.
- Barrett, R. D., & Hoekstra, H. E. (2011). Molecular spandrels: tests of adaptation at the genetic level. *Nature Reviews Genetics*, 12(11), 767-780.
- Barrett, R. D., Laurent, S., Mallarino, R., Pfeifer, S. P., Xu, C. C., Foll, M., ... & Hoekstra, H. E. (2019). Linking a mutation to survival in wild mice. *Science*, 363(6426), 499-504.
- Beaumont, M. A. (2005). Adaptation and speciation: what can F_{st} tell us? *Trends in Ecology & Evolution*, 20(8), 435-440.

730 Berner, D. (2019). Allele frequency difference AFD—an intuitive alternative to FST for
 731 quantifying genetic population differentiation. *Genes*, 10(4), 308.

732

733 Bierne, N., Roze, D., & Welch, J. J. (2013). Pervasive selection or is it...? Why are FST outliers
 734 sometimes so frequent?. *Molecular Ecology*, 22(8), 2061-2064.

735

736 Bolnick, D. I., & Lau, O. L. (2008). Predictable patterns of disruptive selection in stickleback in
 737 postglacial lakes. *The American Naturalist*, 172(1), 1-11.

738

739 Brown, W. L., & Wilson, E. O. (1956). Character displacement. *Systematic Zoology*, 5(2), 49-
 740 64.

741

742 Chao, T., Liu, Z., Zhang, Y., Zhang, L., Huang, R., He, L., ... & Zheng, W. (2019). Precise and rapid
 743 validation of candidate gene by allele specific knockout with CRISPR/Cas9 in wild mice.
 744 *Frontiers in Genetics*, 10, 124.

745

746 Charlesworth, B. (2017). Haldane and modern evolutionary genetics. *Journal of Genetics*,
 747 96(5), 773-782.

748

749 Charlesworth, D., Charlesworth, B., & Morgan, M. T. (1995). The pattern of neutral molecular
 750 variation under the background selection model. *Genetics*, 141(4), 1619-1632.

751

752 Chen, H., Patterson, N., & Reich, D. (2010). Population differentiation as a test for selective
 753 sweeps. *Genome Research*, 20(3), 393-402.

754

755 Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., ... & Ruden, D. M. (2012).
 756 A program for annotating and predicting the effects of single nucleotide polymorphisms,
 757 SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, 6(2),
 758 80-92.

759

760 Crawford, J. E., & Lazzaro, B. P. (2012). Assessing the accuracy and power of population
761 genetic inference from low-pass next-generation sequencing data. *Frontiers in Genetics*, 3,
762 66.
763

764 Crespi, B., & Nosil, P. (2013). Conflictual speciation: species formation via genomic conflict.
765 *Trends in Ecology & Evolution*, 28(1), 48-57.
766

767 Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of
768 speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, 23(13),
769 3133-3157.
770

771 Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... & McVean, G.
772 (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156-2158.
773

774 De Villemereuil, P., & Gaggiotti, O. E. (2015). A new FST-based method to uncover local
775 adaptation using environmental variables. *Methods in Ecology and Evolution*, 6(11), 1248-
776 1258.
777

778 Dunham, M. J., Badrane, H., Ferea, T., Adams, J., Brown, P. O., Rosenzweig, F., & Botstein, D.
779 (2002). Characteristic genome rearrangements in experimental evolution of *Saccharomyces*
780 *cerevisiae*. *Proceedings of the National Academy of Sciences*, 99(25), 16144-16149.
781

782 Egan, S. P., Ragland, G. J., Assour, L., Powell, T. H., Hood, G. R., Emrich, S., ... & Feder, J. L.
783 (2015). Experimental evidence of genome-wide impact of ecological selection during early
784 stages of speciation-with-gene-flow. *Ecology Letters*, 18(8), 817-825.
785

786 Elmer, K. R. (2016). Genomic tools for new insights to variation, adaptation, and evolution in
787 the salmonid fishes: a perspective for charr. *Hydrobiologia*, 783(1), 191-208.
788

789 Elmer, K. R., Fan, S., Kusche, H., Spreitzer, M. L., Kautt, A. F., Franchini, P., & Meyer, A. (2014).
790 Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel routes. *Nature*
791 *Communications*, 5(1), 1-8.

792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822

Enard, D., Messer, P. W., & Petrov, D. A. (2014). Genome-wide signals of positive selection in human evolution. *Genome Research*, 24(6), 885-895.

Excoffier, L., Hofer, T., & Foll, M. (2009). Detecting loci under selection in a hierarchically structured population. *Heredity*, 103(4), 285-298.

Fang, S., Yukilevich, R., Chen, Y., Turissini, D. A., Zeng, K., Boussy, I. A., & Wu, C. I. (2012). Incompatibility and competitive exclusion of genomic segments between sibling *Drosophila* species. *PLoS Genetics*, 8(6), e1002795.

Fay, J. C., & Wu, C. I. (2000). Hitchhiking under positive Darwinian selection. *Genetics*, 155(3), 1405-1413.

Fernandez i Marti, A., & Dodd, R. S. (2018). Using CRISPR as a gene editing tool for validating adaptive gene function in tree landscape genomics. *Frontiers in Ecology and Evolution*, 6, 76.

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(5), 1275-1291.

Foot, A. D. (2018). Sympatric speciation in the genomic era. *Trends in Ecology & Evolution*, 33(2), 85-95.

Gagnaire, P. A., Pavey, S. A., Normandeau, E., & Bernatchez, L. (2013). The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by RAD sequencing. *Evolution*, 67(9), 2483-2497.

Gavrilets, S. (2004). *Fitness landscapes and the origin of species*. Princeton University Press, Princeton, NJ.

823 Gaudet, P., & Dessimoz, C. (2017). Gene ontology: pitfalls, biases, and remedies. In *The Gene*
824 *Ontology Handbook* (pp. 189-205). Humana Press, New York, NY.

825

826 Garud, N. R., Messer, P. W., Buzbas, E. O., & Petrov, D. A. (2015). Recent selective sweeps in
827 North American *Drosophila melanogaster* show signatures of soft sweeps. *PLoS Genetics*,
828 *11*(2).

829

830 Gosset, C. C., & Bierne, N. (2013). Differential introgression from a sister species explains high
831 FST outlier loci within a mussel species. *Journal of Evolutionary Biology*, *26*(1), 14-26.

832

833 Haldane, J. B. S. (1927). A mathematical theory of natural and artificial selection, part V:
834 selection and mutation. In *Mathematical Proceedings of the Cambridge Philosophical Society*
835 (Vol. 23, No. 7, pp. 838-844). Cambridge University Press.

836

837 Haller, B. C. (2016). Eidos: a simple scripting language. URL:
838 http://benhaller.com/slim/Eidos_Manual.pdf.

839

840 Haller, B. C., & Messer, P. W. (2017). SLiM 2: Flexible, interactive forward genetic simulations.
841 *Molecular Biology and Evolution*, *34*(1), 230-240.

842

843 Härmälä, T., & Savolainen, O. (2019). Genomic patterns of local adaptation under gene flow in
844 *Arabidopsis lyrata*. *Molecular Biology and Evolution*, *36*(11), 2557-2571.

845

846 Harris, A. M., Garud, N. R., & DeGiorgio, M. (2018). Detection and classification of hard and
847 soft sweeps from unphased genotypes by multilocus genotype identity. *Genetics*, *210*(4),
848 1429-1452.

849

850 He, Z., Xu, B., Buxbaum, J., & Ionita-Laza, I. (2019). A genome-wide scan statistic framework
851 for whole-genome sequence data analysis. *Nature Communications*, *10*(1), 1-11.

852

853 Hermisson, J., & Pennings, P. S. (2005). Soft sweeps: molecular population genetics of
854 adaptation from standing genetic variation. *Genetics*, *169*(4), 2335-2352.

855

856 Hill, M. O. 1973. Diversity and evenness: A unifying notation and its consequences. *Ecology*,
857 54, 427-432.

858

859 Hohenlohe, P. A., Phillips, P. C., & Cresko, W. A. (2010). Using population genomics to detect
860 selection in natural populations: key concepts and methodological considerations.
861 *International Journal of Plant Sciences*, 171(9), 1059-1071.

862

863 Holsinger, K. E., & Weir, B. S. (2009). Genetics in geographically structured populations:
864 defining, estimating and interpreting F_{ST} . *Nature Reviews Genetics*, 10(9), 639-650.

865

866 Huber, C. D., DeGiorgio, M., Hellmann, I., & Nielsen, R. (2016). Detecting recent selective
867 sweeps while controlling for mutation rate and background selection. *Molecular Ecology*,
868 25(1), 142-156.

869

870 Hudson, R. R. (2002). Generating samples under a Wright–Fisher neutral model of genetic
871 variation. *Bioinformatics*, 18(2), 337-338.

872

873 Hufford, M. B., Lubinsky, P., Pyhäjärvi, T., Devengenzo, M. T., Ellstrand, N. C., & Ross-Ibarra,
874 J. (2013). The genomic signature of crop-wild introgression in maize. *PLoS Genetics*, 9(5).

875

876 Jacobs, A., Carruthers, M., Yurchenko, A., Gordeeva, N., Alekseyev, S., Hooker, O., ... & Elmer,
877 K. R. (2020). Parallelism in eco-morphology and gene expression despite variable evolutionary
878 and genomic backgrounds in a Holarctic fish. *PLoS Genetics* (in press).

879

880 Jacobs, A., Womack, R., Chen, M., Gharbi, K., & Elmer, K. R. (2017). Significant synteny and
881 colocalization of ecologically relevant quantitative trait loci within and across species of
882 salmonid fishes. *Genetics*, 207(2), 741-754.

883

884 Jensen, J. D., Thornton, K. R., & Andolfatto, P. (2008). An approximate Bayesian estimator
885 suggests strong, recurrent selective sweeps in *Drosophila*. *PLoS Genetics*, 4(9).

886

887 Jost, L. (2010). Independence of alpha and beta diversities. *Ecology*, 91(7), 1969-1974.
888

889 Kautt, A. F., Machado-Schiaffino, G., & Meyer, A. (2016). Multispecies outcomes of sympatric
890 speciation after admixture with the source population in two radiations of Nicaraguan crater
891 lake cichlids. *PLoS Genetics*, 12(6).
892

893 Kelly, J. K. (1997). A test of neutrality based on interlocus associations. *Genetics*, 146(3), 1197-
894 1206.
895

896 Kenney, A. M., & Sweigart, A. L. (2016). Reproductive isolation and introgression between
897 sympatric *Mimulus* species. *Molecular Ecology*, 25(11), 2499-2517.
898

899 Kern, A. D., & Hahn, M. W. (2018). The neutral theory in light of natural selection. *Molecular*
900 *Biology and Evolution*, 35(6), 1366-1371.
901

902 Kern, A. D., & Schrider, D. R. (2018). diploS/HIC: an updated approach to classifying selective
903 sweeps. *G3: Genes, Genomes, Genetics*, 8(6), 1959-1970.
904

905 Kim, Y., & Nielsen, R. (2004). Linkage disequilibrium as a signature of selective sweeps.
906 *Genetics*, 167(3), 1513-1524.
907

908 Kim, Y., & Stephan, W. (2002). Detecting a local signature of genetic hitchhiking along a
909 recombining chromosome. *Genetics*, 160(2), 765-777.
910

911 Kirubakaran, T. G., Grove, H., Kent, M. P., Sandve, S. R., Baranski, M., Nome, T., ... & Sonesson,
912 A. (2016). Two adjacent inversions maintain genomic differentiation between migratory and
913 stationary ecotypes of Atlantic cod. *Molecular Ecology*, 25(10), 2130-2143.
914

915 Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: analysis of next generation
916 sequencing data. *BMC Bioinformatics*, 15(1), 356.
917

918 Lamichhaney, S., Han, F., Webster, M. T., Andersson, L., Grant, B. R., & Grant, P. R. (2018).
919 Rapid hybrid speciation in Darwin's finches. *Science*, 359(6372), 224-228.
920

921 Lee, J., Freddolino, P. L., & Zhang, Y. (2017). Ab initio protein structure prediction. In *From*
922 *Protein Structure to Function with Bioinformatics* (pp. 3-35). Springer, Dordrecht.
923

924 Lohmueller, K. E. (2014). The impact of population demography and selection on the genetic
925 architecture of complex traits. *PLoS Genetics*, 10(5), e1004379.
926

927 Long, A., Liti, G., Luptak, A., & Tenaillon, O. (2015). Elucidating the molecular architecture of
928 adaptation via evolve and resequence experiments. *Nature Reviews Genetics*, 16(10), 567.
929

930 Lotterhos, K. E. (2019). The effect of neutral recombination variation on genome scans for
931 selection. *G3: Genes, Genomes, Genetics*, 9(6), 1851-1867.
932

933 Lotterhos, K. E., Card, D. C., Schaal, S. M., Wang, L., Collins, C., & Verity, B. (2017). Composite
934 measures of selection can improve the signal-to-noise ratio in genome scans. *Methods in*
935 *Ecology and Evolution*, 8(6), 717-727.
936

937 Lotterhos, K. E., & Whitlock, M. C. (2014). Evaluation of demographic history and neutral
938 parameterization on the performance of FST outlier tests. *Molecular Ecology*, 23(9), 2178-
939 2192.
940

941 Lucek, K., Keller, I., Nolte, A. W., & Seehausen, O. (2018). Distinct colonization waves underlie
942 the diversification of the freshwater sculpin (*Cottus gobio*) in the Central European Alpine
943 region. *Journal of Evolutionary Biology*, 31(9), 1254-1267.
944

945 Malinsky, M., Challis, R. J., Tyers, A. M., Schiffels, S., Terai, Y., Ngatunga, B. P., ... & Turner, G.
946 F. (2015). Genomic islands of speciation separate cichlid ecomorphs in an East African crater
947 lake. *Science*, 350(6267), 1493-1498.
948

949 Matthey-Doret, R., & Whitlock, M. C. (2019). Background selection and F_{ST} : consequences for
 950 detecting local adaptation. *Molecular Ecology*, 28(17), 3902-3914.

951

952 McGirr, J. A., & Martin, C. H. (2018). Parallel evolution of gene expression between trophic
 953 specialists despite divergent genotypes and morphologies. *Evolution Letters*, 2(2), 62-75.

954

955 Messer, P. W. (2013). SLiM: simulating evolution with selection and linkage. *Genetics*, 194(4),
 956 1037-1039.

957

958 Messer, P. W., & Petrov, D. A. (2013). Population genomics of rapid adaptation by soft
 959 selective sweeps. *Trends in Ecology & Evolution*, 28(11), 659-669.

960

961 Mikkelsen, T. S., Wakefield, M. J., Aken, B., Amemiya, C. T., Chang, J. L., Duke, S., ... & Jurka, J.
 962 (2007). Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding
 963 sequences. *Nature*, 447(7141), 167-177.

964

965 Mitchell, S., Reeve, R., & White, T. J. (2020). rdiversity. R package version 2.0.
 966 <https://github.com/boydorr/rdiversity>

967

968 Moeller, D. A., Tenaillon, M. I., & Tiffin, P. (2007). Population structure and its effects on
 969 patterns of nucleotide polymorphism in teosinte (*Zea mays* ssp. *parviglumis*). *Genetics*,
 970 176(3), 1799-1809.

971

972 Moest, M., Van Belleghem, S. M., James, J. E., Salazar, C., Martin, S. H., Barker, S. L., ... &
 973 Steiner, F. M. (2020). Selective sweeps on novel and introgressed variation shape mimicry loci
 974 in a butterfly adaptive radiation. *PLoS Biology*, 18(2), e3000597.

975

976 Morales, H. E., Faria, R., Johannesson, K., Larsson, T., Panova, M., Westram, A. M., & Butlin,
 977 R. K. (2019). Genomic architecture of parallel ecological divergence: beyond a single
 978 environmental contrast. *Science Advances*, 5(12), eaav9963.

979

980 Narum, S. R., & Hess, J. E. (2011). Comparison of FST outlier tests for SNP loci under selection.
 981 *Molecular Ecology Resources*, 11, 184-194.

982

983 Ng, P. C., & Henikoff, S. (2006). Predicting the effects of amino acid substitutions on protein
 984 function. *Annu. Rev. Genomics Hum. Genet.*, 7, 61-80.

985

986 Nielsen, R., Williamson, S., Kim, Y., Hubisz, M. J., Clark, A. G., & Bustamante, C. (2005).
 987 Genomic scans for selective sweeps using SNP data. *Genome Research*, 15(11), 1566-1575.

988

989 Nielsen, R., Paul, J. S., Albrechtsen, A., & Song, Y. S. (2011). Genotype and SNP calling from
 990 next-generation sequencing data. *Nature Reviews Genetics*, 12(6), 443-451.

991

992 Nolte, A. W., Freyhof, J., & Tautz, D. (2006). When invaders meet locally adapted types: rapid
 993 moulding of hybrid zones between sculpins (*Cottus*, Pisces) in the Rhine system. *Molecular*
 994 *Ecology*, 15(7), 1983-1993.

995

996 Noor, M. A. (1999). Reinforcement and other consequences of sympatry. *Heredity*, 83(5), 503-
 997 508.

998

999 Nosil, P. (2012). *Ecological speciation*. Oxford University Press, Oxford, UK.

1000

1001 Osborne, O. G., Ciezarek, A., Wilson, T., Crayn, D., Hutton, I., Baker, W. J., ... & Savolainen, V.
 1002 (2019). Speciation in *Howea* palms occurred in sympatry, was preceded by ancestral
 1003 admixture, and was associated with edaphic and phenological adaptation. *Molecular Biology*
 1004 *and Evolution*, 36(12), 2682-2697.

1005

1006 Papadopulos, A. S., Igea, J., Smith, T. P., Hutton, I., Baker, W. J., Butlin, R. K., & Savolainen, V.
 1007 (2019). Ecological speciation in sympatric palms: 4. Demographic analyses support speciation
 1008 of *Howea* in the face of high gene flow. *Evolution*, 73(9), 1996-2002.

1009

1010 Pavlidis, P., & Alachiotis, N. (2017). A survey of methods and tools to detect recent and strong
 1011 positive selection. *Journal of Biological Research-Thessaloniki*, 24(1), 7.

1012

1013 Pavlidis, P., Živković, D., Stamatakis, A., & Alachiotis, N. (2013). SweeD: likelihood-based
 1014 detection of selective sweeps in thousands of genomes. *Molecular Biology and Evolution*,
 1015 30(9), 2224-2234.

1016

1017 Pettengill, J. B., & Moeller, D. A. (2012). Phylogeography of speciation: allopatric divergence
 1018 and secondary contact between outcrossing and selfing *Clarkia*. *Molecular Ecology*, 21(18),
 1019 4578-4592.

1020

1021 Poisot, T., Bever, J. D., Nemri, A., Thrall, P. H., & Hochberg, M. E. (2011). A conceptual
 1022 framework for the evolution of ecological specialisation. *Ecology Letters*, 14(9), 841-851.

1023

1024 Pritchard, J. K., Pickrell, J. K., & Coop, G. (2010). The genetics of human adaptation: hard
 1025 sweeps, soft sweeps, and polygenic adaptation. *Current Biology*, 20(4), R208-R215.

1026

1027 Puzey, J. R., Willis, J. H., & Kelly, J. K. (2017). Population structure and local selection yield high
 1028 genomic variation in *Mimulus guttatus*. *Molecular Ecology*, 26(2), 519-535.

1029

1030 Rahi, M. L., Amin, S., Mather, P. B., & Hurwood, D. A. (2017). Candidate genes that have
 1031 facilitated freshwater adaptation by palaemonid prawns in the genus *Macrobrachium*:
 1032 identification and expression validation in a model species (*M. koombooloomba*). *PeerJ*, 5,
 1033 e2977.

1034

1035 Rauwolf, U., Greiner, S., Mráček, J., Rauwolf, M., Golczyk, H., Mohler, V., ... & Meurer, J.
 1036 (2011). Uncoupling of sexual reproduction from homologous recombination in homozygous
 1037 *Oenothera* species. *Heredity*, 107(1), 87-94.

1038

1039 Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., ... & Westram, A. M.
 1040 (2017). Interpreting the genomic landscape of speciation: a road map for finding barriers to
 1041 gene flow. *Journal of Evolutionary Biology*, 30(8), 1450-1477.

1042

1043 R Core Team (2018). R: A language and environment for statistical computing. R Foundation
 1044 for Statistical Computing, Vienna, Austria. URL <https://www.r-project.org>.
 1045

1046 Reeve, R., T. Leinster, C. Cobbold, J. Thompson, N. Brummitt, S. Mitchell, and L. Matthews.
 1047 2016. How to partition diversity. *arXiv*, 1404.6520:1-9.
 1048

1049 Refoyo-Martínez, A., da Fonseca, R. R., Halldórsdóttir, K., Árnason, E., Mailund, T., & Racimo,
 1050 F. (2019). Identifying loci under positive selection in complex population histories. *Genome*
 1051 *Research*, 29(9), 1506-1520.
 1052

1053 Rettelbach, A., Nater, A., & Ellegren, H. (2019). How linked selection shapes the diversity
 1054 landscape in *Ficedula* flycatchers. *Genetics*, 212(1), 277-285.
 1055

1056 Rundle, H. D., & Nosil, P. (2005). Ecological speciation. *Ecology Letters*, 8(3), 336-352.
 1057

1058 Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., ... & Schaffner, S. F.
 1059 (2007). Genome-wide detection and characterization of positive selection in human
 1060 populations. *Nature*, 449(7164), 913-918.
 1061

1062 Savolainen, V., Anstett, M. C., Lexer, C., Hutton, I., Clarkson, J. J., Norup, M. V., ... & Baker, W.
 1063 J. (2006). Sympatric speciation in palms on an oceanic island. *Nature*, 441(7090), 210-213.
 1064

1065 Schlamp, F., van der Made, J., Stambler, R., Chesebrough, L., Boyko, A. R., & Messer, P. W.
 1066 (2016). Evaluating the performance of selection scans to detect selective sweeps in domestic
 1067 dogs. *Molecular Ecology*, 25(1), 342-356.
 1068

1069 Schlötterer, C., Kofler, R., Versace, E., Tobler, R., & Franssen, S. U. (2015). Combining
 1070 experimental evolution with next-generation sequencing: a powerful tool to study adaptation
 1071 from standing genetic variation. *Heredity*, 114(5), 431-440.
 1072

1073 Schluter, D. (1996). Ecological speciation in postglacial fishes. *Philosophical Transactions of*
 1074 *the Royal Society of London. Series B: Biological Sciences*, 351(1341), 807-814.

1075

1076 Schluter, D., & McPhail, J. D. (1992). Ecological character displacement and speciation in
 1077 sticklebacks. *The American Naturalist*, 140(1), 85-108.

1078

1079 Schuler, H., Hood, G. R., Egan, S. P., & Feder, J. L. (2006). Modes and mechanisms of
 1080 speciation. *Reviews in Cell Biology and Molecular Medicine*, 2(3), 60-93.

1081

1082 Schrider, D. R., & Kern, A. D. (2016). S/HIC: robust identification of soft and hard sweeps using
 1083 machine learning. *PLoS Genetics*, 12(3).

1084

1085 Schrider, D. R., & Kern, A. D. (2017). Soft sweeps are the dominant mode of adaptation in the
 1086 human genome. *Molecular Biology and Evolution*, 34(8), 1863-1877.

1087

1088 Sheehan, S., & Song, Y. S. (2016). Deep learning for population genetic inference. *PLoS*
 1089 *Computational Biology*, 12(3), e1004845.

1090

1091 Stajich, J. E., & Hahn, M. W. (2005). Disentangling the effects of demography and selection in
 1092 human history. *Molecular Biology and Evolution*, 22(1), 63-73.

1093

1094 Stapley, J., Feulner, P. G., Johnston, S. E., Santure, A. W., & Smadja, C. M. (2017). Variation in
 1095 recombination frequency and distribution across eukaryotes: patterns and processes.
 1096 *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1736), 20160455.

1097

1098 Stinchcombe, J. R., & Hoekstra, H. E. (2008). Combining population genomics and quantitative
 1099 genetics: finding the genes underlying ecologically important traits. *Heredity*, 100(2), 158-
 1100 170.

1101

1102 Streisfeld, M. A., & Kohn, J. R. (2005). Contrasting patterns of floral and molecular variation
 1103 across a cline in *Mimulus aurantiacus*. *Evolution*, 59(12), 2548-2559.

1104

1105 Sweigart, A. L., Brandvain, Y., & Fishman, L. (2019). Making a murderer: the evolutionary
 1106 framing of hybrid gamete-killers. *Trends in Genetics*, 35(4), 245-252.

1107

1108 Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics*,
 1109 105(2), 437-460.

1110

1111 Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA
 1112 polymorphism. *Genetics*, 123(3), 585-595.

1113

1114 Tang, H., & Thomas, P. D. (2016). Tools for predicting the functional impact of
 1115 nonsynonymous genetic variation. *Genetics*, 203(2), 635-647.

1116

1117 Thornton, K. (2005). Recombination and the properties of Tajima's D in the context of
 1118 approximate-likelihood calculation. *Genetics*, 171(4), 2143-2148.

1119

1120 Thurman, T. J., & Barrett, R. D. (2016). The genetic consequences of selection in natural
 1121 populations. *Molecular Ecology*, 25(7), 1429-1448.

1122

1123 Torres-Dowdall, J., Henning, F., Elmer, K. R., & Meyer, A. (2015). Ecological and lineage-
 1124 specific factors drive the molecular evolution of rhodopsin in cichlid fishes. *Molecular Biology*
 1125 *and Evolution*, 32(11), 2876-2882.

1126

1127 Torres-Dowdall, J., Pierotti, M. E., Härer, A., Karagic, N., Woltering, J. M., Henning, F., ... &
 1128 Meyer, A. (2017). Rapid and parallel adaptive evolution of the visual system of Neotropical
 1129 Midas cichlid fishes. *Molecular Biology and Evolution*, 34(10), 2469-2485.

1130

1131 Turner, J. R., & JRG, T. (1977). Butterfly mimicry: the genetical evolution of an adaptation.
 1132 *Evolutionary Biology*, 10,163-206.

1133

1134 Vatsiou, A. I., Bazin, E., & Gaggiotti, O. E. (2016). Detection of selective sweeps in structured
 1135 populations: a comparison of recent methods. *Molecular Ecology*, 25(1), 89-103.

1136

1137 Verity, R., Collins, C., Card, D. C., Schaal, S. M., Wang, L., & Lotterhos, K. E. (2017). MINOTAUR:
 1138 A platform for the analysis and visualization of multivariate results from genome scans with
 1139 R Shiny. *Molecular Ecology Resources*, 17(1), 33-43.
 1140
 1141 Vitti, J. J., Grossman, S. R., & Sabeti, P. C. (2013). Detecting natural selection in genomic data.
 1142 *Annual review of genetics*, 47, 97-120.
 1143
 1144 Wang, J. (2012). On the measurements of genetic differentiation among populations.
 1145 *Genetics Research*, 94(5), 275-289.
 1146
 1147 Watterson, G. A. (1975). On the number of segregating sites in genetical models without
 1148 recombination. *Theoretical Population Biology*, 7(2), 256-276.
 1149
 1150 White, T. J. (2019). Developing new measures of genetic diversity (Unpublished Master's
 1151 thesis). University of Glasgow, Glasgow, UK.
 1152
 1153 Wright, S. (1951). The genetical structure of populations. *Annals of Eugenics*, 15(1), 323-354.
 1154
 1155 Zheng, Y., & Wiehe, T. (2019). Adaptation in structured populations and fuzzy boundaries
 1156 between hard and soft sweeps. *PLoS Computational Biology*, 15(11).
 1157

Figure 1 – General outline of simulated scenarios. a) Sympatric divergence scenario where gene flow occurs directly after the population split and continues until the samples are taken. In this case, population sizes remain constant after divergence. b) Sympatric divergence scenario with logistic population growth. c) Secondary contact scenario where there is no gene flow directly after the population split. Gene flow continues after secondary contact of populations 2 and 3 until the samples are taken. All combinations of parameters were included in the simulations for a total number of 144 scenarios, including 96 selective sweep scenarios and 48 neutral scenarios.

Figure 2 – Traces of summary statistics for an example simulation scenario (average across 100 replicates): Scenario 2 (secondary contact), with hard sweeps and constant, unequal population sizes with a migration rate of 0.01. Summary statistics a) $varg_{kl}$ and $kurtg_{kl}$ show clear peaks centred around the locations of the selective sweeps (dotted lines). b) J_1 , J_{12} , and $H - scan$ show relatively clear peaks around the sites of the selected mutations. c) $\Delta\pi$ varies across the genome and shows only subtle peaks, while the signals of F_{ST} around the mutation under selection are distinct. d) $\bar{B}^{GD}$ shows distinct peaks around the sweep location. The left part of each panel is the standard recombination region (white background) while the right part is the low recombination region (grey background). The graphs shown are derived by averaging across 100 simulation replicates. For plotting, the absolute values of $\Delta\pi$

1181 multiplied by 10 and the values of $H - scan$ divided by 1,000,000 were used. $\bar{B}^{GD}$ is shown
1182 for $q = 0$, $q = 1$, and $q = 2$, but the graphs are, however, indistinguishable due to almost
1183 identical $\bar{B}^{GD}$ values.
1184

Figure 3 – A heat map of prediction accuracies of the location of selective sweeps for the standard recombination rate region and equal population sizes. The colours in the heat map are based on accuracy values (see scale at the top right of the plot, with the highest accuracies, 1, in dark red and the lowest accuracies, 0, in dark blue), which are the proportions of 100 simulation replicates in which the correct window was identified as containing the selected mutation (1-in-1 window approach). For $\bar{B}^{GD}$, the chosen q value was 2. scen1 = scenario 1 (sympatric divergence); scen2 = scenario 2 (secondary contact scenario with low negative selection in population 3); scen3 = scenario 3 (secondary contact scenario with intermediate negative selection in population 3); scen4 = scenario 4 (secondary contact scenario with strong negative selection in population 3); gro = logistically increasing population size; con = constant population size; soft sweep = soft selective sweep scenario;

1197 hard sweep = hard selective sweep scenario; migration = migration rate between populations
1198 2 and 3; pop. growth = population growth (gro or con); migr. rate = migration rate.
1199

1200

Suggested steps for locating selective sweeps

1201

1202 Figure 4 – Suggested steps for detecting the genomic position of selective sweeps. Drawing
1203 from the simulation results, empirical studies as well as theory, we suggest an eight-step
1204 process that can help to locate and make sense of selective sweeps in whole-genome data.

1205

1206

1207

1208 Supplementary files available from *Journal of Evolutionary Biology*