The Precision and Power of Population Branch Statistics in Identifying the Genomic Signatures of Local Adaptation

Max Shpak [1,*], Kadee N. Lawrence [1], John E. Pool [1,*]

[1]Laboratory of Genetics, University of Wisconsin–Madison, Madison, WI, USA

*Corresponding authors: E-mails: shpak.max@gmail.com; jpool@wisc.edu.

Accepted: April 29, 2025

Abstract

Population branch statistics, which estimate the degree of genetic differentiation along a focal population's lineage, have been used as an alternative to F_{ST}-based genome-wide scans for identifying loci associated with local selective sweeps. Beyond the population branch statistic (PBS), the normalized PBSn1 adjusts focal branch length with respect to outgroup branch lengths at the same locus, whereas population branch excess (PBE) incorporates median branch lengths at other loci. PBSn1 and PBE were proposed to be more specific to local selective sweeps as opposed to geographically ubiquitous selection. However, the accuracy and statistical power of branch statistics have not been systematically assessed. To do so, we simulate genomes in representative large and small populations with varying proportions of sites evolving under genetic drift or (approximated) background selection, with local selective sweeps or geographically parallel selective sweeps. We then assess the probability that local selective sweep loci are correctly identified as outliers by F_{ST} and by each of the branch statistics. We find that branch statistics consistently outperform F_{ST} at identifying local sweeps. Particularly when parallel sweeps are introduced, PBSn1 and PBE correctly identify local sweeps among their top outliers more frequently than PBS. Additionally, we evaluate versions of these statistics based on maximal site differentiation within a window, finding that site-based PBE and PBSn1 are particularly effective at identifying local soft sweeps. These results validate the greater specificity of the rescaled branch statistics PBE and PBSn1 to detect population-specific positive selection, supporting their use in genomic studies focused on local adaptation.

Key words: population branch statistics, fixation index, selective sweeps, local adaptation, population genetic simulation.

Significance

Population branch statistics are widely used in genome-wide scans to identify loci associated with local adaptation. This simulation study finds that branch statistics are more accurate than F_{ST} at identifying local selective sweeps under a wide range of demographic parameters and models of evolution. It also demonstrates that rescaled branch statistics have improved ability to distinguish local adaptation from other models of natural selection, and that site-level versions of these metrics are particularly useful in detecting local adaptation from standing genetic variation. These findings will provide empirical studies a more informed basis for selecting statistics for use in genome-wide scans for loci involved in local adaptation.

Introduction

One of the cornerstones of evolutionary genetics is understanding the processes that drive genetic differences within and among populations. A research area of particular interest is determining which observed differences in allele frequency between populations are driven by local adaptation versus genetic drift or background selection (BGS). Local adaptation is of special importance in part because population-specific selective sweeps can contribute to the initial stages of allopatric or parapatric speciation by generating genetic incompatibilities between regional variants. Furthermore, the genomic signatures of local selective sweeps can provide insights into the molecular basis and genetic architecture of adaptive phenotypic traits as they relate to regional variation.

Among the most notable and well-documented examples of local adaptation include grasses that tolerate high heavy metal concentrations in contaminated soils (Antonovics and Bradshaw 1970; Macnair 1987), mouse coat color variants adapted to different substrates by selection for camouflage (Nachman et al. 2003; Hoekstra et al. 2006), the evolution of body shape and coloration of sticklebacks in response to different species of predator (Colosimo et al. 2005; Miller et al. 2015; Gygax et al. 2018), the evolution of toxin resistance in garter snakes that co-occur in regions with highly toxic newts as prey (Brodie et al. 2002), and the independent adaptation to hypoxic environments for humans and other large mammals living at high altitudes (Yi et al. 2010; Julian and Moore 2019; Witt and Huerta-Sanchez 2019). *Drosophila melanogaster* and its close relatives have also provided important model systems for genetic studies of local adaptation, including for studies focused on adaptation to high altitudes (Lack et al. 2016; Sprengelmeyer and Pool 2021; Sprengelmeyer et al. 2022), upper latitudes (Adrion et al. 2015; Svetec et al. 2015; Siddiq and Thornton 2019), and novel food sources (Yassin et al. 2016).

One general approach to identifying and characterizing local adaptations is based on quantifying the differences in allele frequencies among populations at different loci, on the assumption that highly divergent loci are potential genomic signatures of local adaptation. A variety of statistical methods have been proposed and used to leverage this information. Some approaches identify deviations from neutrality in a focal population by comparisons to outgroup populations, e.g. the maximum frequency of derived mutations test of Li (2011), which identifies excess derived alleles in a population with respect to an outgroup. More recent studies (e.g. Xue et al. 2021) have applied machine-learning algorithms to allele frequency distributions to identify ad classify selective sweeps as soft (i.e. involving contributions from more than one preexisting haplotype) versus hard.

Alternative sweep detection approaches use maximum likelihood calculations to optimize and infer demographic parameters including branch lengths, e.g. the cross-population likelihood ratio (XP-CLR) of Chen et al. (2010), which compares likelihood ratios for allele frequencies under neutrality to models with selective sweeps. This method was extended and generalized to a three-population 3P-CLR model by Racimo (2016), which calculates likelihood ratios based on a tree topology as well as demographic parameters. Such methods make explicit assumptions about evolutionary models and so require the estimation of underlying parameters, e.g. 3P-CLR leverages estimates of ancestral allele frequencies and split times.

In contrast, model-free approaches, such as those based on the relative magnitudes of genetic differentiation among populations, require no parameter estimation and are independent of specific assumptions about underlying evolutionary processes and history. The present study focuses on evaluating the performance of model-free statistics derived from the fixation index (F_{ST}) under a broad range of demographic and selection regimes. These model-free methods have the advantage of versatility and simplicity, while identifying their potential shortcomings can point to scenarios where model-based methods may be more appropriate.

Fixation Index and Population Branch Statistics

Genome-wide scans for selection identify outliers in the distribution of differences in allele frequency between a focal population and its outgroup(s) using various measures of genetic distance. Among the most widely used is Wright's F_{ST}, defined as the ratio of among population variance to the total variance across populations. The quantity was originally derived for single loci and sites, but can be generalized as a statistic for genomic regions (e.g. Reynolds et al. 1983). Many studies have used high F_{ST} between two populations at specific loci in comparison to the rest of the genome or chromosome region to infer local adaptation (e.g. Akey 2009; Amato et al. 2009; Kapun et al. 2020).

There are several drawbacks to using F_{ST} to measure population divergence, such as its nonlinearity and nonadditivity (making comparisons across multiple populations difficult). Importantly, when comparing very large numbers of sites in a genome-wide scan, it is inevitable that a significant fraction will be divergent among populations due to genetic drift, resulting in some fraction of loci that are falsely identified as having evolved under positive directional selection. Further, the use of F_{ST} leaves the problem of determining which population in a divergent pair has experienced local adaptation.

Particularly to address the latter problem, several statistics have been proposed that use a rescaled F_{ST}-based value across population triplets (a focal population and two

outgroup populations) to estimate population-specific differentiation. Among these is the population branch statistic (PBS), first proposed in Yi et al. (2010) as an estimate of the branch length of the focal population. The distance metric for population pairs is a log-transformed (Cavalli-Sforza 1969) function of F_{ST}, which, in contrast to F_{ST} itself, accumulates additively with time under the simplest population split models:

$$T = -\log(1 - F_{ST}) \quad (1)$$

Defining a three-population tree (Fig. 1a) with focal population A, a relatively closely related population B, and a "population outgroup" C, we can estimate the corresponding branch lengths of a,b,c. Because the pairwise distances T are approximately additive, the branch length a from the three pairwise T values can be estimated as:

$$\mathrm{PBS}_A = \frac{T_{AB} + T_{AC} - T_{BC}}{2} \quad (2)$$

A similar statistic, known as the locus-specific branch length (LSBL) was proposed earlier by Shriver et al. (2004); LSBL also estimates the focal population branch length using Eq. (2), but with F_{ST} in place of T. Although untransformed F_{ST} values are not additive quantities, LSBL would be expected to give qualitatively similar results to PBS, especially with rank-based approaches such as outlier quantiles.

PBS as an estimator of focal branch length has been used in a number of recent studies to identify sites or regions of the genome evolving under selective sweeps (e.g. Jiang and Assis 2020), with the logic that loci with high PBS compared to other loci in the genome are strong candidates for selective sweeps since population divergence.

To the extent that studies aim to identify population-specific positive selection in particular, a potential disadvantage to PBS is its inability to distinguish between cases where only the focal branch is especially long from cases where all population branches are long. While sites under strong local directional selection will have high PBS, the same is true at sites undergoing similar, parallel selective sweeps in all three populations. Depending on the goals of the study, such loci may or may not be of interest, as they may often reflect instances of positive selection unrelated to adaptation to local environments, such as arms races driven by meiotic drive or reproductive competition. Furthermore, BGS (Charlesworth et al. 1993, 1995) may have a relatively widespread influence on genetic diversity in at least some regions of the genome (e.g. regions with low recombination rates and/or a high density of functional sites), and frequency changes driven by BGS may also lead all populations to have unusually long branches at certain loci. Consequently, PBS may be effective at identifying instances of recent natural selection generally while being less effective at discriminating between population-specific versus species-wide selection pressures (although such predictions remain to be tested).

To address the above concerns regarding PBS, two branch statistic rescalings have been proposed with the specific purpose of identifying loci with population-specific elevations in genetic differentiation. The normalized PBS, i.e. PBSn1, was introduced by Malaspinas et al. (2016; see their supplementary section S16, Supplementary Material online) and also featured in subsequent studies (Crawford et al. 2017, Vicuna et al. 2019). This metric rescales PBS_A with respect to the total tree length, e.g.

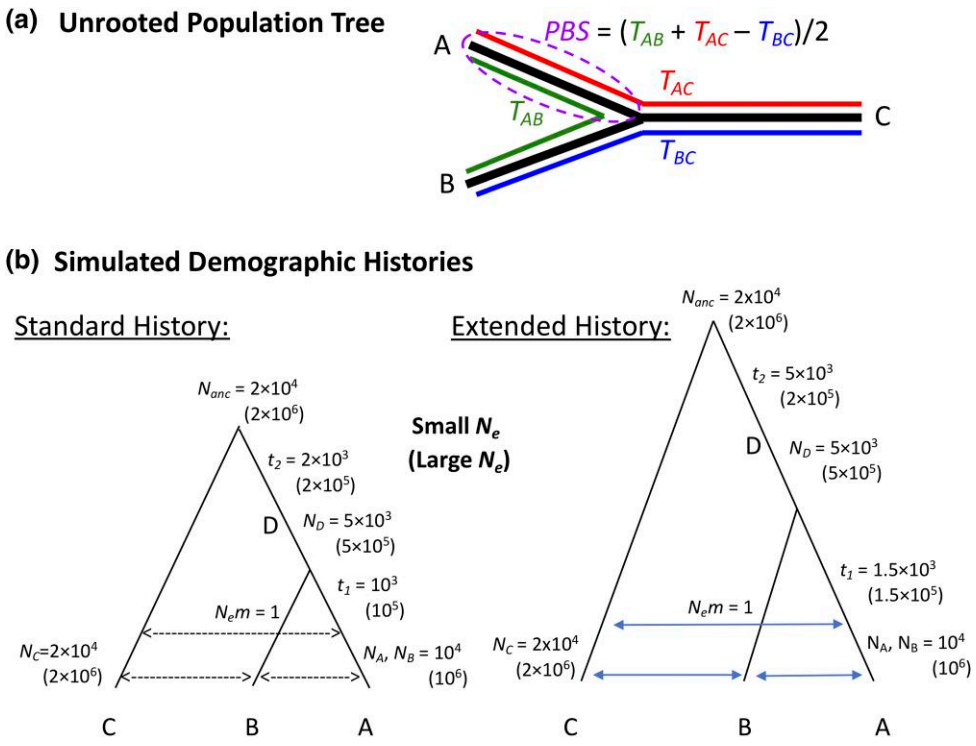
$$\mathrm{PBSn1}_A = \frac{\mathrm{PBS}_A}{1 + \mathrm{PBS}_A + \mathrm{PBS}_B + \mathrm{PBS}_C} \quad (3)$$

This rescaling has the potential to discriminate between cases where PBS_A is high in comparison to the lineages of populations B and/or C from cases where B or C also have long branches. Consequently, PBSn1 is expected to give a lower false positive rate in identifying local sweeps at loci that are evolving under parallel (i.e. separately occurring) selective sweeps in multiple populations. The addition of 1 in the denominator prevents division by zero and provides a baseline (although arbitrary) normalization term so that cases where all three branches are extremely short but the focal population is several times longer is not scored as significant in the same way that an especially long focal branch in particular would be (e.g. without the 1, a case of branch lengths $a,b,c = 0.02,0.01,0.01$ and $a,b,c = 20,10,10$ would be treated equivalently).

Unlike the above statistics, another elaboration on PBS, known as Population Branch Excess (PBE; Yassin et al. 2016), incorporates branch length information from additional genomic loci. This method posits that in the absence of local adaptation at a given locus, the relationship between the focal population's branch length at this locus (i.e. PBS_A) and the combined lengths of the two nonfocal population branches (i.e. T_{BC}) should be similar to the relationship between these two quantities observed at most other genomic loci. PBE employs this logic to obtain an expected value for PBS at this locus, based on T_{BC} at this locus and the median values of both PBS and T_{BC} observed at all other analyzed loci in the genome (or else a component of it such as the same chromosome arm). PBE then quantifies the degree to which the observed PBS value for this locus exceeds its expected value, as follows:

$$\mathrm{PBE}_A = \mathrm{PBS}_A - \mathrm{PBS}_{A\_exp} = \mathrm{PBS}_A - \frac{T_{BC} \times \mathrm{med}(\mathrm{PBS}_A)}{\mathrm{med}(T_{BC})}, \quad (4)$$

where $\mathrm{med}(\mathrm{PBS}_A)$ and $\mathrm{med}(T_{BC})$ are the median values of their respective statistics in a distribution across loci in the



(c) Genomic Parameters

Parameter	Small N_e	Large N_e
Mutation Rate	1.22×10^{-8}	5.21×10^{-9} [2.61×10^{-7}]
Crossing Over Rate	1.14×10^{-8}	1.09×10^{-8} [5.45×10^{-7}]
Gene Conversion Rate	5.9×10^{-8}	6.25×10^{-8} [3.13×10^{-6}]
Conversion Tract Length (bp)	100	518
Window Size (kilobases)	100	5

Fig. 1. Illustration of the demographic models and genomic parameters implemented in population genetic simulations. a) Unrooted three-population tree, with A representing the the focal population. *T* values represent the genetic distances between each pair of populations, based on a log transformation of F_{ST} (Eq. (1)). PBS, as the estimated length of the focal population branch, is then an intuitive function of these *T* values. b) The simulated three-population genealogies. N_i represents the size of the *i*th small population, the value in parentheses is the size of the corresponding simulated large population (the lineage D is the ancestral population for A,B before their split, N_{anc} is the size of the population ancestral to all three sampled populations). t_1 and t_2 are the split times for the inner and outer divergences, respectively. The left genealogy was used for the migration-free simulations as well as the first set of migration simulations (a net population migration rate of $N_em = 1$ is represented by the dashed lines). The right genealogy has branch lengths adjusted to generate the same pairwise F_{ST} as in the first genealogy without migration under genetic drift alone. c) The genomic parameters used in simulations for the small and large populations (values in parentheses are the 50× rescalings used in the simulations). Mutation, crossing over, and gene-conversion rates are per-site, per-generation.

genome. The comparison to an expected PBS value may make PBE values somewhat more comparable across studies that involve populations with very difficult demographic histories. For example, if all loci are highly differentiated due to a recent bottleneck in the focal population's history, PBE will adjust all values based on the relatively high expected PBS (although an elevated neutral variance may still be expected). The ratio of branch lengths T_{BC}/med(T_{BC}) also corrects for cases where the nonfocal branches are also long due to selection in those populations as well (whether sweeps or BGS). Or viewed another way, a larger-than-usual value of T_{BC} at this locus increases the value expected for PBS, and thus will tend to yield lower PBE values, which is desirable if the goal is a specific focus on population-specific positive selection.

We also note that multiple branch-oriented statistics have been developed to detect local adaptation from data sets that include greater than three populations (Schmidt et al. 2019; Schlebusch et al. 2020; Cheng et al. 2022). However, this study's focus is restricted to frequency-based statistics that incorporate at most three populations, in part for the sake of clarity, and in part because of the relatively larger number of empirical studies that can generate adequate population genomic data for these methods.

The heuristic considerations that prompted the proposal of PBS, PBSn1, and PBE as measures of population divergence suggest that all branch statistics may have greater power and accuracy in identifying instances of local selective sweeps against a backdrop of neutral evolution than F_{ST}, and that PBSn1 and PBE may outperform PBS when some loci evolve under parallel sweeps while the evolution of other loci is driven by local adaptation. However, their relative efficacy and statistical power have not been systematically examined over a range of demographic and evolutionary parameters in either the papers that introduced them or in subsequent studies.

Most empirical genomic scans for local adaptation have calculated differentiation statistics like those mentioned above on the scale of whole windows, whose lengths may be scaled by physical distance, or perhaps preferably, by the amount of genetic variation each contains. Window sizes are typically chosen with the intention that genetic differentiation within a window will be strongly affected by a local selective sweep. However, when positive selection acts upon an already-segregating variant, which has already recombined onto multiple haplotype backgrounds, a soft sweep may result, in which only a narrow interval of genetic variation is strongly affected. In this scenario, whole window statistics may fail to detect local soft sweeps as outliers, but focusing instead on the maximum SNP F_{ST} value within a window can be more powerful (da Silva Ribeiro et al. 2022). A related approach of using the 90th percentile F_{ST} value within a window has also been deployed (FST_{90}; Fraïsse et al. 2016), as have maximum SNP versions of PBS (Lange et al. 2022) and PBE (Pool et al. 2017). However, as with the whole window population branch statistics, the performance of maximum SNP versions of these metrics in detecting various scenarios of local adaptation has not been assessed.

In this study, we analyze data from population genetic simulations representing a wide range of scenarios in order to assess the above predictions regarding the performance of the above-mentioned whole window and SNP-focused statistics to detect population-specific positive selection. We include simulations with large and small population sizes (motivated by *Drosophila* and humans, respectively), and we investigate models with and without gene flow after population divergence. In each demographic model, we combine simulated replicates into simulated genomes and assess which scenarios are placed in the top (upper 1%) outlier quantile for each statistic, emulating an empirical outlier genome scan. This approach allows us to evaluate how well each statistic differentiates local adaptation not only from neutral evolution but also from BGS (approximated via reduced effective population size) and parallel selective sweeps. Results of these analyses will allow researchers to make more informed choices regarding which statistics to use in future population genomic scans for local adaptation.

Results

We performed population genetic simulations to test the ability of four statistics (F_{ST}, PBS, PBSn1, and PBE) to detect population-specific positive selection, examining three population models in small and large N_e cases motivated by human and *Drosophila* data, respectively. The demographic and genomic parameters used in these models are summarized in Fig. 1b and c. We aggregated simulations from distinct neutral/selection scenarios into "model genomes" (Fig. 2a) to test how reliably each statistic placed the 1% of true local sweep loci in the upper 1% tail, with or without an additional 1% of loci subject to parallel sweeps in all three populations, and with the remaining loci subject to either neutral evolution or BGS (modeled as variably reduced N_e).

Detecting Local Adaptation Against a Neutral Genomic Background

We simulated local complete hard sweeps with selection coefficients $s = 0.025$ and 0.001 in the small and large populations, respectively. When these local sweeps occur amongst neutral loci, the loci under selection can be identified with a high level of accuracy by all statistics, with only slight advantages of the three population statistics over pairwise F_{ST}. As we deviated from this simple scenario, the statistics began to diverge from one another in both their precision based on the model genomes defined above (Fig. 3, supplementary table S1, Supplementary Material online) and their more traditional statistical power based on comparing individual local sweep replicates to neutral or BGS distributions (supplementary table S2, Supplementary Material online).

Precision was reduced in the cases of partial and soft local sweeps in otherwise neutral genomes, with PBS, PBSn1, and PBE often performing similarly, and F_{ST} consistently showing the lowest precision (Fig. 3). In contrast, in model genomes where most loci evolve under BGS rather than genetic drift, the relative performances of the branch statistics became more distinct: here, the precision of PBSn1 and PBE was invariably larger than for PBS in both the large and

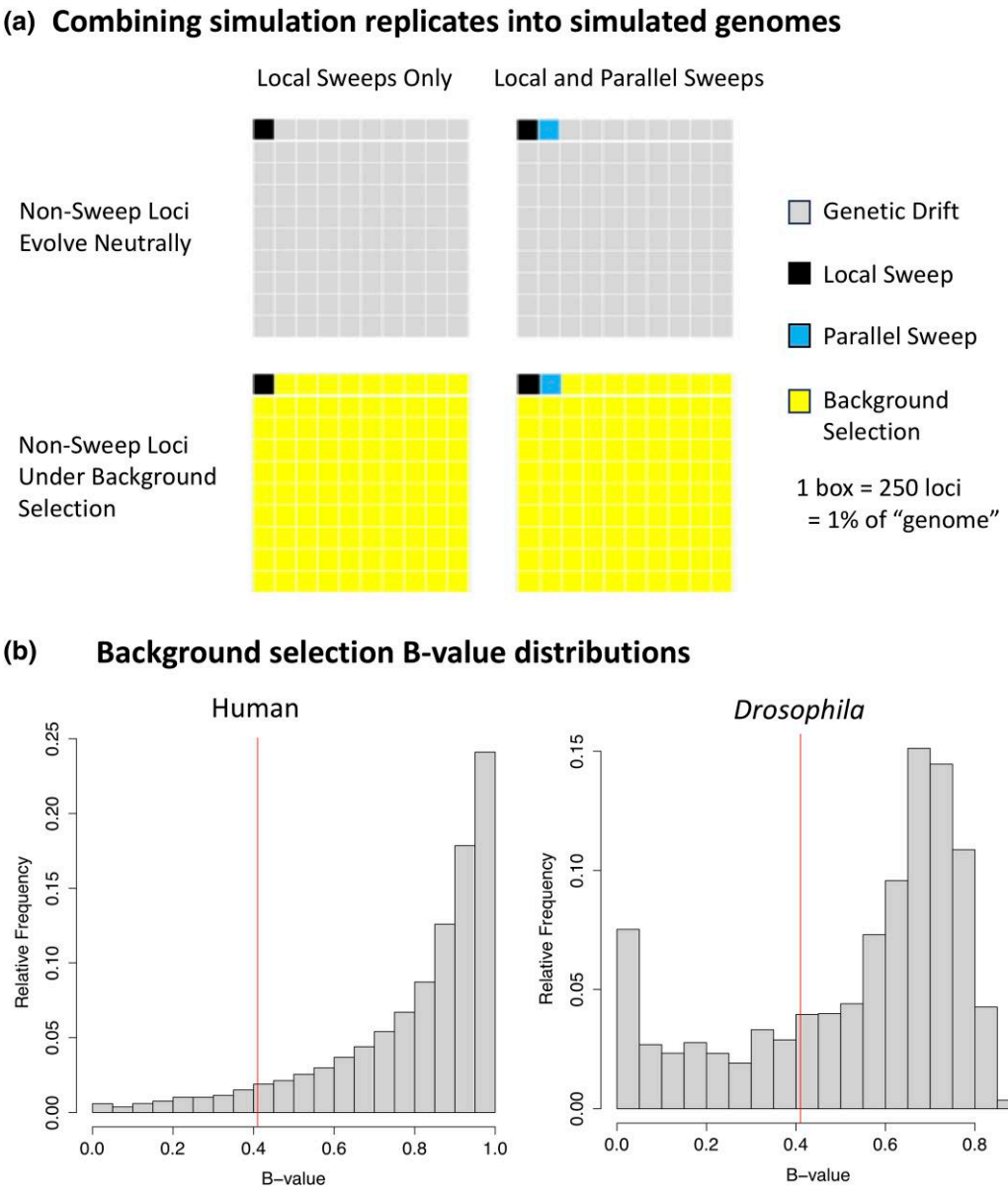


Fig. 2. Simulated genome models used in this study. a) Schematics representing the model genomes, each containing 25,000 simulated loci (representing genomic windows in an empirical scan). The upper left genome has 99% of loci randomly sampled from the 10^6 genetic drift simulations and 1% from the 10^4 local selective sweep simulations. The upper right genome has 98% neutral, 1% local selective sweep simulations, and 1% parallel sweep simulations. The lower left genome has 99% BGS loci and 1% local sweeps. The lower right genome has 98% BGS loci, 1% local sweeps, and 1% parallel sweep loci. b) The autosomal *B*-value distributions for human (left) and *D. melanogaster* (right) genomes are shown, as used here for the simulations of small and large population simulations, respectively. The vertical red line represents the truncation at $B = 0.41$ for each population size scenario, in order to approximate genome scans in which low recombination regions are excluded due to the difficulty in localizing targets.

small populations (Fig. 3). PBE also outperformed PBSn1 in each of these cases, although the difference was more notable in the large populations. In the scenario that yielded the most disparate outcomes, that of a soft sweep from 0.5% initial frequency in a large population with BGS, precision was 1.6% for F_{ST}, 10.8% for PBS, 16.7% for PBSn1, and 31.6% for PBE. In a few of the most challenging cases, the precision of some or all statistics fell considerably. In the large population case, a large number of false positives generated by BGS led some precision estimates to fall below the null expectation of 0.01 (i.e. the true fraction of loci under local sweeps). This could be seen in the case of soft sweeps with higher initial frequency, perhaps because in our simulations, we did not additionally subject sweep

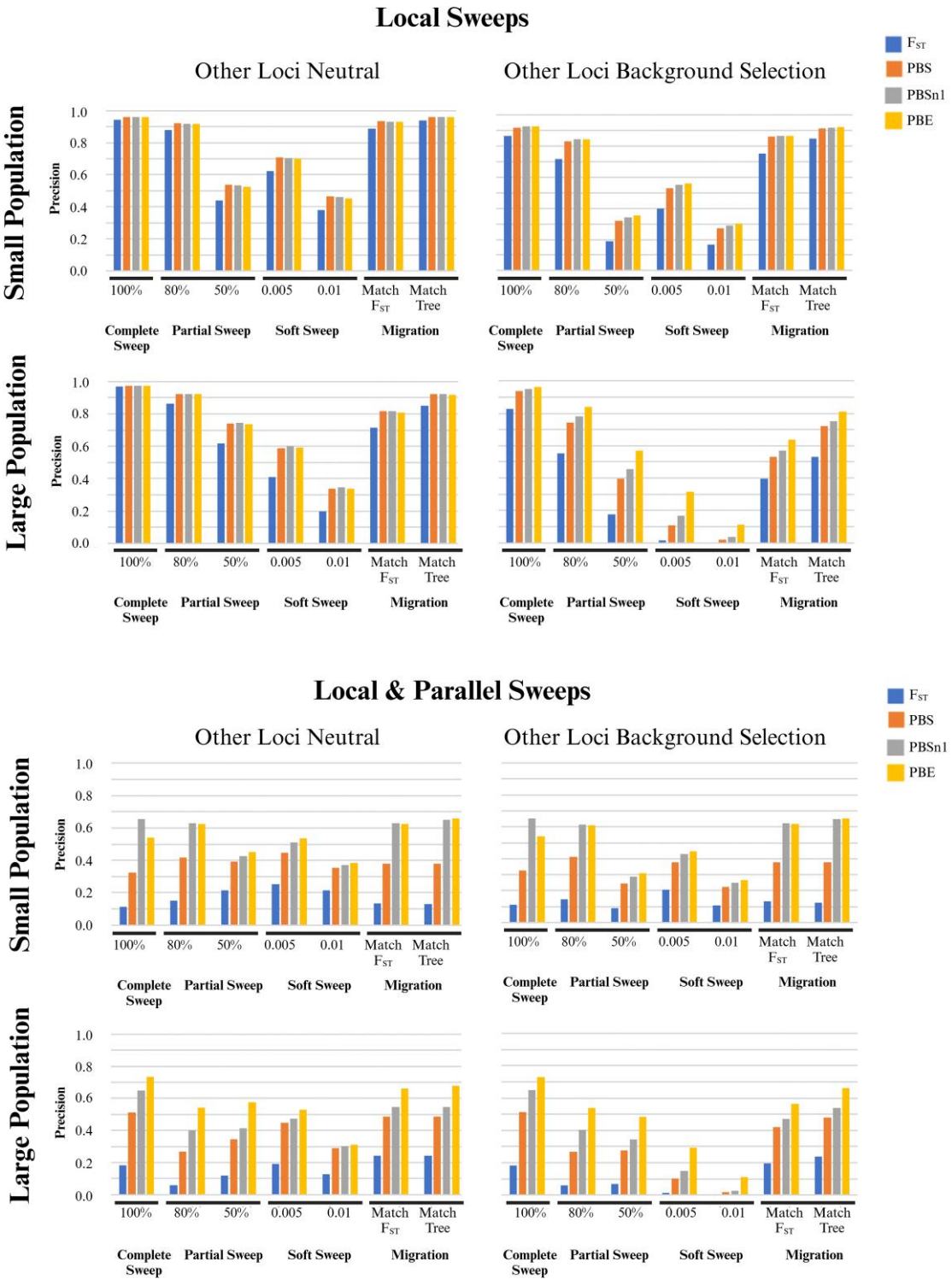


Fig. 3. Population branch statistics show greater precision to detect local selective sweeps than F_{ST} when other loci experience global positive or negative selection. Bar plots show the precision of F_{ST} and the population statistics PBS, PBSn1, PBE with respect to local sweeps, i.e. the fraction of local sweep loci among those contributing to the upper 1% quantile of each statistic. The top four panels show the results for a single selective sweep in the focal population for large and small populations (upper and lower panels), while the lower four panels do the same for parallel sweeps in all three populations. Each set of graphs includes genomic backgrounds with genetic drift (left panels) and with emulated BGS (right panels). Each panel shows results for all modeled selection regimes (hard complete, partial, soft complete sweeps), including migration scenarios involving hard sweeps that might have fixed if not for gene flow. Supplementary table S1, Supplementary Material online provides exact precision values for each scenario.

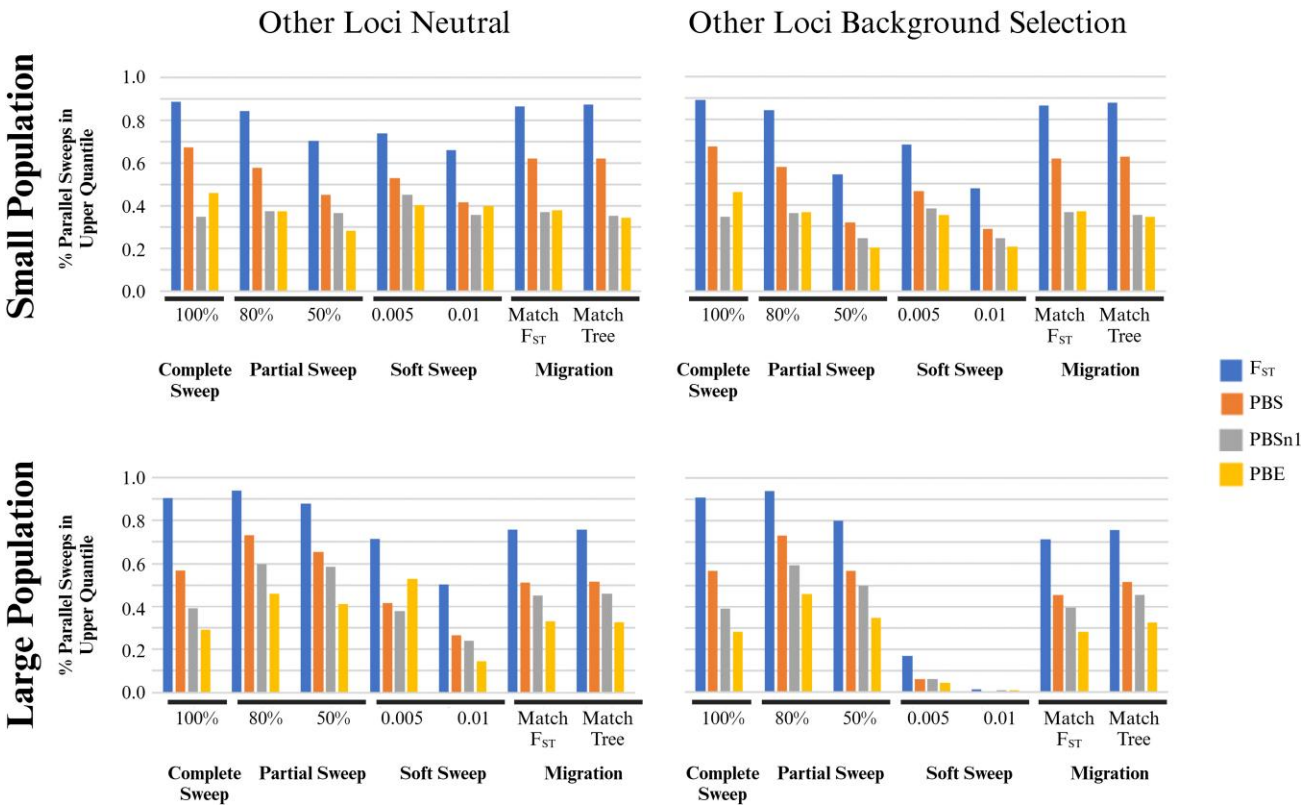


Fig. 4. PBE and PBSn1 are less likely than other statistics to register parallel sweep loci among their top outliers. For model genome scenarios in which 1% of loci are subject to parallel sweeps in all populations, the graphs show the fraction of parallel sweep loci that contribute to the upper 1% quantile of F_{ST} and branch statistic distributions (i.e. the fraction of parallel sweep loci that are false positives) under different demographic and selection parameters.

loci to BGS. Under these scenarios, BGS tends to generate longer branch lengths than genetic drift (especially in the large populations, which have a higher fraction of loci with $B < 0.5$; Fig. 2b). Here, PBE's rescaling of branch lengths with respect to all loci in the "genome" may improve its performance in comparison to PBS and even with respect to PBSn1, which rescales the focal branch length with respect to total tree length at the focal locus but not with respect to an overall genome-wide baseline.

Detecting Local Adaptation in the Presence of Parallel Sweeps and/or BGS

In simulations that included both parallel and local sweeps at 1% of the loci, in otherwise neutrally evolving genomes, the overall frequency of false positives increased for all of the statistics when compared to otherwise equivalent scenarios in which all sweeps were local, albeit to varying degrees (Fig. 3). Here, the advantages of PBE and PBSn1 over PBS became more pronounced, with PBE again showing a larger advantage for the large population scenarios. Whereas, the performance of F_{ST} became especially poor with parallel sweeps. In these scenarios, we expect that parallel sweeps may generate higher F_{ST} values than local sweeps, so parallel sweeps will disproportionately contribute to the upper quantiles of F_{ST} and potentially PBS (which simply looks for a long estimated focal population branch length). The enhanced differentiation of parallel sweeps (and thus their contribution to false positives) can be greater for hard complete sweeps than for partial or soft sweeps (Fig. 4, supplementary table S3, Supplementary Material online), which may contribute to the otherwise surprising observation that in some cases, F_{ST} and to lesser extent PBS had lower precision and power with hard complete sweeps than with certain soft or partial sweeps. This phenomenon is less apparent in the large populations, where the modeled selection coefficients are weaker and the rate of recombination relative to mutation is somewhat higher (see Materials and Methods).

When local and parallel sweeps were both present, and were hard and either complete or else partial at 80% terminal frequency, most false positives were contributed by parallel sweep loci (Fig. 4, supplementary table S3, Supplementary Material online). In contrast, for the cases of soft sweeps from a higher (1%) initial frequency or partial sweeps to 50% terminal frequency, there was a greater contribution from the neutral and BGS loci. This trend is most notable in the case of BGS in the large

populations—e.g. 46% to 94% of parallel sweep loci contributed to upper quantiles under 80% partial sweeps, whereas only about 1% of parallel sweep loci were in the upper quantile in soft sweeps with $P_0 = 0.01$. The proportional contribution of parallel sweeps versus neutral/BGS loci to false positives impacts the relative ability of the population statistics to detect local sweeps. For example, in a small population with both local and parallel sweeps, the precisions for some of the statistics, especially F_{ST} and PBS, are higher for soft sweeps (with $P_0 = 0.005$) and partial sweeps ($P_t = 0.80$) than for complete hard sweeps. This counterintuitive pattern is not seen in the large population simulations, which have lower false positive rates from the parallel sweeps than in the small populations. In contrast, with very soft sweeps ($P_0 = 0.01$) or partial sweeps to only 50% final frequency, there is also a significant relative contribution of neutral or BGS sites to the false positives.

When populations are subject to both parallel sweeps and BGS, similar trends hold, with slightly to moderately reduced precision for each statistic to detect local sweeps compared to cases with parallel sweeps without BGS. In all cases, F_{ST} showed the poorest performance, while PBS lagged behind the other two branch statistics. For small populations, PBSn1 and PBE show similar advantages, while for large populations, PBE displayed consistently greater precision, generally showing similar or greater advantages over PBSn1 and the other statistics than observed with either parallel sweeps or BGS alone. For example, in the case of soft sweeps with $P_0 = 0.005$, PBE at 0.295 gave a precision nearly 2- and 3-fold higher than PBSn1 and PBS, respectively, and 20-fold higher than F_{ST}.

Detecting Local Adaptation in the Presence of Gene Flow

While the above-mentioned results all pertain to simulations with genetic drift but no migration, we also explored cases in which all pairs of populations have exchanged one migrant per-generation (i.e. $N_e m = 1$) since their divergence. Adding this level of migration to the previously studied population history had a relatively weak effect on the precision and power of the statistics to detect hard local sweeps under scenarios lacking any other form of selection (precision based on simulated genomes in Fig. 3 and supplementary table S1, Supplementary Material online; conventional power in supplementary table S2, Supplementary Material online; "match tree"). Here, while migration may have reduced the frequency differences at locally adaptive loci, it also resulted in populations with lower neutral genetic differentiation than in migration-free base scenario. The effect of migration on precision and power is more perceptible under the scenarios in which we extended the divergence times of our populations to emulate the neutral F_{ST} values observed among populations without migration (Fig. 1b), since here allele frequency differences should be reduced at local sweep but not neutral loci compared to the base history (Fig. 3, supplementary tables S1 and S2, Supplementary Material online; "match F_{ST}"). While the effects of migration on its own were not dramatic for either population size or population tree length, its effects synergized with BGS to reduce precision, particularly for the large population size with its more powerful modeled BGS effects (Fig. 3). With parallel sweeps, the performance of F_{ST} and in some cases PBS actually improved with the introduction of migration. To some extent, migration may be blunting the propensity of parallel sweeps to generate false positives (Fig. 4, supplementary table S3, Supplementary Material online), analogous to the outcomes of some partial or soft sweep models discussed above. In general with migration, as observed otherwise, PBE and PBSn1 showed the greatest precision when other types of selection were present, and PBE in particular showed greater advantages in the large populations.

We also evaluated precision statistics for LSBL, an estimate of focal branch length that uses untransformed F_{ST} as a measure of pairwise distances. Since LSBL unsurprisingly performed much like PBS for many selection and demographic regimes, these results are presented in supplementary table S1, Supplementary Material online alongside the other statistics. For local sweeps at the target locus and neutral evolution at the other loci, LSBL and PBS distributions had nearly identical precision values, while diverging somewhat in performance for models with BGS and parallel sweeps. LSBL gave somewhat higher precision for a subset of those scenarios, especially in the small population simulations such as the cases of parallel sweeps with migration. However, there are also scenarios where LSBL had lower statistical precision than PBS, e.g. in large populations with complete hard parallel sweeps (supplementary table S1, Supplementary Material online).

Evaluating the Utility of SNP-Focused Statistics to Detect Local Adaptation

We also analyzed statistics focused on the most extreme site-level differentiation observed within a window, including maximum site versions of F_{ST} and PBS. For PBSn1 and PBE, we focused on site-specific versions of these statistics in which the within-window maxima of different branch terms can come from different sites (in order to boost specificity for local adaptation; see Materials and Methods), but we also computed constrained versions of these statistics in which all terms must be calculated from the same site, denoted PBSn1* and PBE*. We focused on a subset of the above-described models because the primary motivation for site-level approaches is to detect soft local sweeps that may not sufficiently perturb variation on the scale of

whole windows, whereas these approaches are less useful for partial sweeps (da Silva Ribeiro et al. 2022). We found that all of the above statistics were similarly effective at distinguishing selective sweeps from neutral evolution and from BGS. Their precision values resembled those of full window statistics in the case of hard selective sweeps, while greatly outperforming the full window statistics under soft sweeps (Fig. 5, supplementary table S4, Supplementary Material online). Here, maximum site-based statistics only require that a single site is fixed per window, and thus give the same precision regardless of whether selective sweeps are hard or soft. In contrast, the full window statistics require appreciable frequency changes at many sites, which may not be expected if a favored variant had already recombined onto multiple haplotypes before the sweep began.

In genomes where equal fractions of loci evolve under local and under parallel sweeps, the values of maximum site F_{ST} and PBS were generally equivalent at the parallel sweep and local sweep loci, and hence could not distinguish scenarios. This pattern is expected because (in our simulations) the specific sites within a window targeted by parallel sweeps are different in the three populations; therefore, both local and parallel complete sweeps result in at least one site with a population-specific allele fixed in each population, corresponding to the same maximum site F_{ST} and PBS values in either case. Because these are equal for the two scenarios, the precision of these two statistics at identifying local sweeps is at or below 0.5, performances comparable to what is observed for the window-based statistics (Fig. 5, supplementary fig. S4, Supplementary Material online). As observed for full window statistics, BGS is more likely to produce an extreme value for site F_{ST} than for site branch statistics (especially in the larger population with its higher B-values), since F_{ST} is equally sensitive to frequency changes in lineage A or B.

The site-constrained statistics PBSn1* and PBE*, like PBS, are maximized at the 1,0,0 focal-outgroup frequency configuration and therefore responded equivalently under local and parallel sweeps, resulting in precision values that are equal to those for PBS apart from stochastic differences due to sampling (see supplementary table S4, Supplementary Material online). In contrast, the site-unconstrained maximum site statistics PBSn1 and PBE (in which the component branch lengths are maximized over potentially different sites in a genomic window) consistently distinguished local from parallel sweeps, having statistical precision values approaching or exceeding 0.90 in model genomes containing both local and parallel sweeps (Fig. 5, supplementary fig. S4, Supplementary Material online). Although in either of these models, a fixed difference in the focal population exists, the site-unconstrained versions of PBSn1 and PBE are diminished under parallel sweeps because of the presence of sites in the same window with large nonfocal branch lengths, resulting in greater specificity for detecting local sweeps. We note that such a difference in performance would not be expected if parallel sweeps instead targeted different alleles at the same site.

Evaluating Window and SNP-Focused Statistics Under More Complex Demography

The demographic models used above were intentionally simplified to better disentangle the effects of individual variables. We complemented that approach by also simulating the more complex out of Africa (OOA) demographic model of Gutenkunst et al. (2009), which includes population growth, stronger population bottlenecks, and differences in migration rates across pairs of populations. Here, we focused on the detection of local sweeps that were soft ($P_0 = 0.01$) or partial ($P_t = 0.50$), in the presence of approximated BGS, and with or without parallel sweep loci. For the soft sweep scenario, each statistic performed better under the OOA history than under the simpler history above, potentially reflecting a "hardening" of soft sweeps due to the lower initial size (1,000) of the focal European population. For both window and SNP-focused statistics, we again found that the three population branch statistics outperformed F_{ST}, and that the rescaled metrics PBSn1 and PBE generally had similar or better performance than PBS. Distinct from most results from the simple history, window and SNP versions of PBSn1 showed advantages under the OOA history for detecting either variety of local sweep when parallel sweeps were also present (Fig. 6, supplementary table S5, Supplementary Material online), emphasizing the important interplay of demography and selection in determining the power to detect local adaptation.

Discussion

Population branch statistics were introduced as alternatives to F_{ST} in genome-wide scans for local selective sweeps on heuristic grounds, without explicit assessment of their performance. The present study provides simulation-based evidence favoring their use over F_{ST} as well as supporting the use of rescaled branch statistics PBSn1 and PBE rather than PBS (or LSBL) for a wide range of evolutionary models. Generally, only in the most trivial scenario of a small number of loci experiencing local selective sweeps while all others evolve neutrally do F_{ST}-based scans or PBS have precision and power comparable to the rescaled population branch statistics PBSn1 or PBE. Otherwise, the rescaled branch statistics were consistently more robust in contributing to the upper quantiles when potentially confounding selective processes such as BGS or parallel sweeps were introduced.

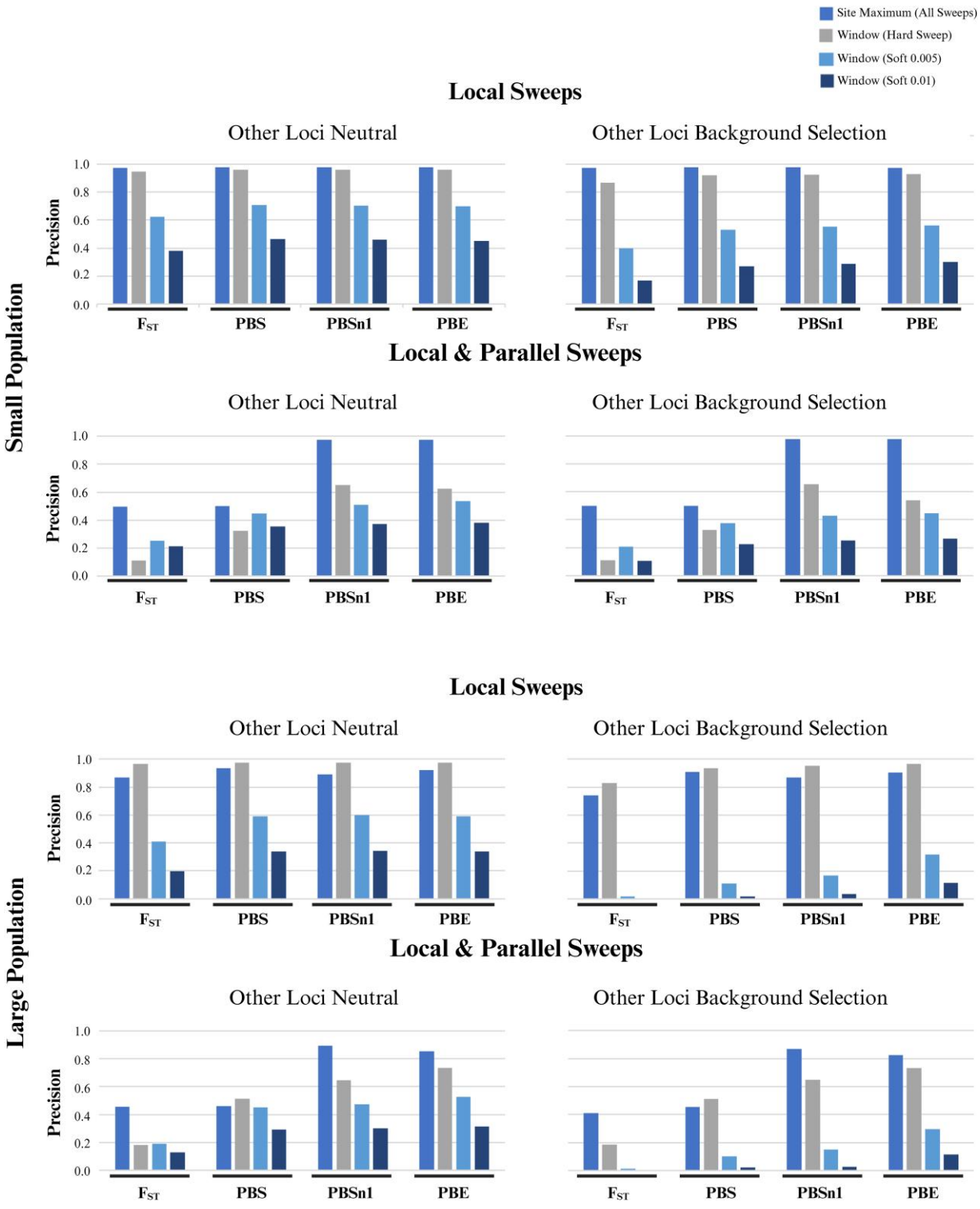


Fig. 5. Unconstrained maximum site PBSn1 and PBE have elevated precision for identifying local soft sweeps. When other loci evolve neutrally or under BGS, the site-level metrics F_{ST_MaxSNP}, PBS_{MaxSNP}, $PBSn1_{MaxSNP}$, and PBE_{MaxSNP} outperform full window metrics in detecting local soft sweeps, and have at least comparable precision for local hard sweeps. When some fraction of loci evolve under local sweeps and others under parallel sweeps, the precision of F_{ST_MaxSNP} and PBS_{MaxSNP} do not exceed 0.5 (indicating inability to distinguish these scenarios), while the site-unconstrained scaling of $PBSn1_{MaxSNP}$ and PBE_{MaxSNP} allows them to effectively distinguish these models, resulting in precision values near or above 0.9. Note that precision values for the site maximum statistics are the same for hard and soft sweeps (since they achieve their same maximal values in either case), unlike the window-based statistics.

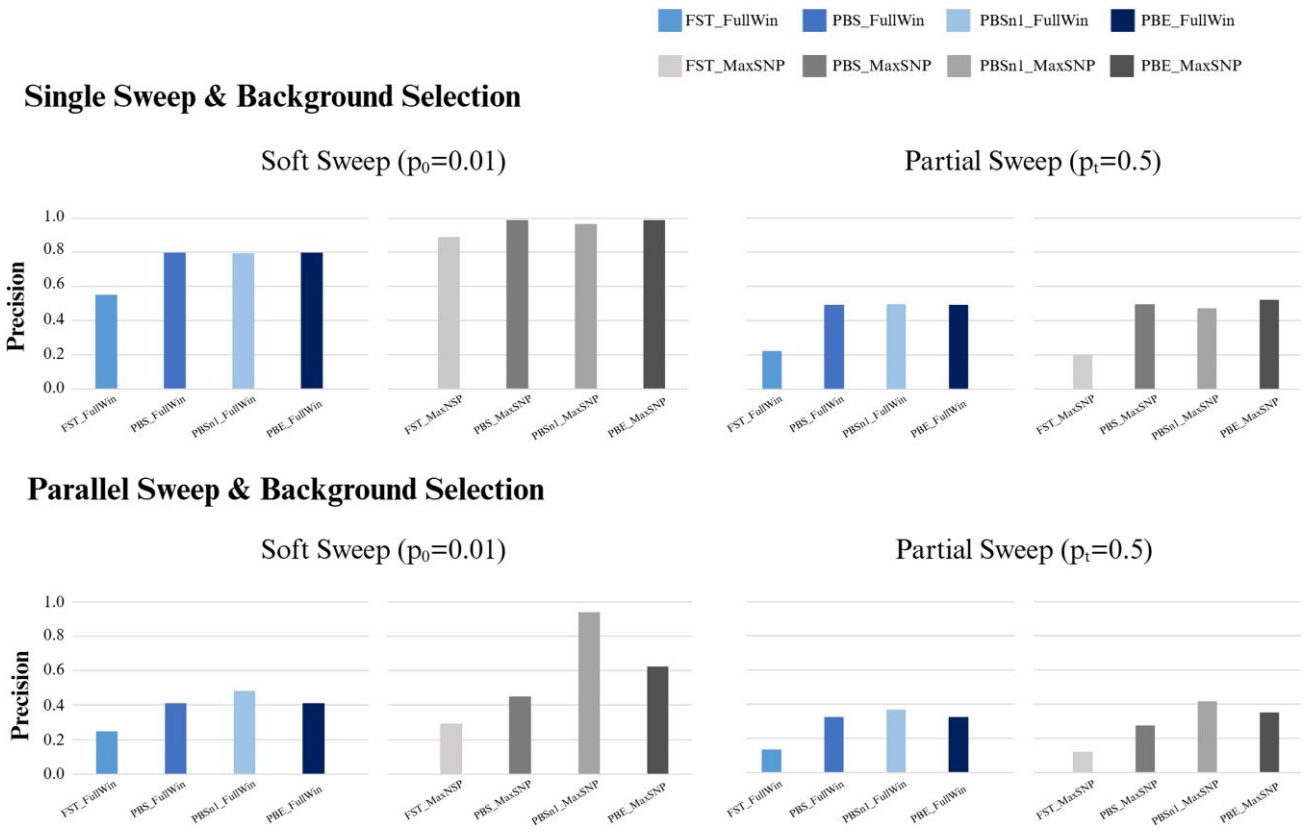


Fig. 6. Precision of window and SNP-focused statistics under a complex human demographic model generally recapitulates patterns observed from simpler histories. Here, the demographic estimates of Gutenkunst et al. (2009) were simulated. Two of the more challenging-to-detect sweep models were chosen for this analysis: a soft sweep starting from 1% frequency (left panels) and a partial (hard) sweep ending at 50% frequency (right panels). Local sweeps in the European population were simulated, with other parallel sweep loci either absent (top panels) or present (bottom panels). Approximated background selelection via N_e reduction was also simulated in all of these cases. Results are shown both for whole window statistics (blue bars) and for SNP-level statistics (gray bars). Some previously observed patterns hold under this history as well, including the consistent advantage of population branch statistics over F_{ST}. Whereas in the presence of parallel sweeps, window and SNP-level PBSn1 metrics show greater precision than PBE, whereas in the above analyses, the reverse was more often true.

In contrast, advantages of either PBE or PBSn1 over the other were more context-specific. PBE displays generally (although not universally) better performance across our simplified histories, especially for window statistics in the large popualtion scenarios, including under the potentially more biologically realistic models of partial and soft sweeps, and in genomes where BGS plays an important role in shaping genetic differentiation (Fig. 3, supplementary table S1, Supplementary Material online). Whereas in the more complex OOA demographic model, PBSn1 showed advantages in the presence of parallel sweeps (Fig. 6, supplementary table S5, Supplementary Material online). The variation in results under these models emphasizes that the optimal choice of PBE versus PBSn1 may depend on the histories of the populations being studied.

Because PBE and PBSn1 differ methodologially and arithmetically in their rescaling of population branch lengths, they may be differentially impacted by longer nonfocal branch lengths due to drift or else nonlocal positive or negative selection, or by nonuniform migration levels among the three populations. PBE uses information across loci to compare focal branch length at a given locus to its expected value. Whereas, PBSn1 (which is calculated using only the variation at a single locus) uses an arbitrary scaling factor of 1.0 in the denominator (Eq 2), which may serve to dampen noise from excessively small denominators. Conceivably, the precision of PBSn1 relative to PBE might vary if other values of this scaling factor were used.

Our analysis of maximum site-based population branch statistics indicates that site-unconstrained PBSn1 and PBE, which have yet to be applied in any empirical study, are particularly effective at identifying local selective sweeps and distinguishing their effects on polarized frequency changes from those of parallel sweeps. The motivation for including site maximum differentiation statistics is strongest in the case of local soft sweeps, for which they show clear

superiority to whole window metrics (Fig. 5, da Silva Ribeiro et al. 2022). However, site maximum approaches are less useful against partial hard sweeps, in which case many sites may have at most moderate frequency changes (da Silva Ribeiro et al. 2022). Because in empirical studies, one generally does not know which types of sweeps to expect a priori, it may often be advisable to employ a combination of full window and site maximum approaches. However, it is important to note that site maximum outlier identification requires larger sample sizes than for full window differentiation (da Silva Ribeiro et al. 2022). Furthermore, outliers for maximum site statistics alone may need extra scrutiny, since they could conceivably be driven by erroneous data at a single site. An alternative approach is to look for multiple sites, but not whole windows, with elevated differentiation. Such an approach, as implemented by the FST_{90} statistic which quantifies the 90th percentile of SNP F_{ST} values within a window (Fraïsse et al. 2016), might be expected to sacrifice some sensitivity to the softest sweeps in favor of greater robustness to data errors.

Given the lack of prior results, our study represents a useful starting point in evaluating the relative performance of population branch statistics. There are several potential extensions and new directions that can be explored to provide a more comprehensive comparison of the methods used to detect local selective sweeps from genomic data. As hinted by our OOA results, one obvious generalization is to demographically expand models that incorporate methods accommodating more than three populations (e.g. Cheng et al. 2022), as well as a wider array of demographic models such as stronger population bottlenecks or more temporally heterogeneous migration rates.

We intentionally focused on a model of gene flow with a moderate population migration rate $Nm = 1$ that was symmetric among populations in order to evaluate the robustness of our results to moderate gene flow. These models could readily be extended to asymmetric regimes of migration with source-sink effects, as was done for our OOA simulations. Another possible scenario for future examination entails admixture from a population outside of the focal deme and its two outgroups (Huerta-Sanchez et al. 2013; Racimo et al. 2017). If some alleles that reached high frequency due to selective sweep in an outside population introgress into the focal population, population branch statistics might identify these as instances of local selective sweeps, in addition to identifying sweeps that occurred subsequently in the target population. Distinguishing these scenarios requires information about local ancestry as well as allele frequency, signals which are integrated by tools such as the Ohana package of Cheng et al. (2022).

Such ancestry-based approaches underscore both the limitations and the strengths of model-free methods of inferring non-neutral evolution such as rank comparisons of F_{ST} or branch statistics. Population branch statistics have the advantages of being easily calculated and independent of assumptions about underlying evolutionary models. On the other hand, model-based inference methods have the potential to provide and leverage further information about the populations studied than branch statistics, provided that the model assumptions and parameter estimates involved are accurate. Therefore, it would be instructive for future studies to compare inferences of local sweeps from branch statistic magnitudes to the likelihood of neutral models estimated from methods such as Racimo's (2016) 3P-CLR method for population triads. In other contexts, summary statistics such as those examined here might be integrated into more complex analysis methods such as approximate Bayesian computation and machine learning.

Here, we used reduced population size as a proxy for the effects of BGS on genetic differentiation. While reasonable in the context of this study's goals, it would be of interest to systematically assess how well the *B*-value rescaling approximates BGS by simulating selection against deleterious alleles (as has been done for human-scale N_e by Boyko et al. 2008; Ewing andJensen 2016; Marsh and Johri 2024) and comparing patterns of genetic differentiation generated by BGS to those generated by the drift approximation. Such simulations may not currently be computationally feasible in large, *Drosophila*-like populations, due to both the extended chromosomal scale that may be relevant (e.g. Comeron 2014) and the likelihood that population size rescaling (as often conducted for simulations with very large N_e) would fail to yield accurate BGS simulations.

Our simulation results suggested a strong potential of BGS to generate outliers in F_{ST}-based statistics in the *Drosophila*-like simulations in particular. However, we note that the empirical analysis of Booker et al. (2020) found that the relationship between crossing-over rate and F_{ST} between *D. melanogaster* populations was roughly similar to that expected under a simulated model without BGS, particularly above the female crossing-over rate threshold of 1 cM/Mb that our *B*-value cutoff approximates. In light of this contrast between prediction and data, further study seems necessary to clarify the models of BGS at work in *D. melanogaster* and their influence on genetic differentiation.

Even in an era featuring increasingly complex population genetic methods for detecting local adaptation, we suggest that there is value in simpler statistics that have relatively straightforward interpretation with regard to genetic variation. In this context, we suggest that rescaled population branch statistics such as PBE and PBSn1 are relatively powerful and efficient tools for identifying the genetic signatures of local adaptation in genome-wide scans, particularly when evaluated in both full window and maximum site contexts. The results of this study argue in favor of their

expanded use in the toolkit of those seeking evidence of local adaptation in natural populations, particularly in cases where other forms of positive or negative selection may impact genetic differentiation. Furthermore, these statistics and their underlying logic may continue to provide useful building blocks for more advanced methodologies in the future.

Methods

We simulated evolution in a three-population model where the evolutionary dynamics of genotype frequencies are driven by (at different loci in various combinations) neutral genetic drift, BGS against deleterious mutations, local selective sweeps, and parallel selective sweeps in all three populations. Unless otherwise indicated, simulations were run using Python 3.8.8 scripts and all data analyses were performed using R 4.1.1.

Population Demographic History

We chose two base demographic models for small and large populations, respectively, using approximated and simplified population histories motivated by humans and *Drosophila melanogaster* as representative examples, with respective effective sizes of $N_e \sim 10^4$ and 10^6. We set the outer and inner population split times at 2×10^3 and 10^3 generations before present for the small population case and 2×10^5 and 10^5 generations for the large population case (as shown in Fig. 1b). These time intervals approximate, at least in orders of magnitude, the estimated split times between Africa versus Eurasia and Eurasia versus the Americas for humans (Schiffels and Durbin 2014) and between sub-Saharan Africa versus the Mediterranean and the Mediterranean versus north-central Europe for the flies (Sprengelmeyer et al. 2020), while allowing symmetric histories (in terms of coalescent time units) to be investigated in each case.

We applied a reduced population size for the internal branch representing the common ancestor of populations A and B (labeled in Fig. 1b as population D, which is half the size of its descendant populations), to roughly recapitulate the genetic drift involved in both of the motivating species' founder event expansions from Africa into Eurasia (note that this bottleneck was much stronger in human evolutionary history). We emphasize that we are not attempting to model the precise demographic details of any particular species' history with these models, rather, we use these approximate and simplified topologies to capture the qualitative differences between large and small population models.

To make large population forward simulations computationally feasible in light of the very high number of generations needed, we modeled evolution using a rescaling approximation where a 50-fold smaller N_e population was simulated in combination with 50-fold higher rescaling of selection coefficients, mutations rates, and other populations parameters, an approach similar to the one used by Lange and Pool (2018). The demographic and genetic parameters are listed in Fig. 1c for both the small and large populations (based on ancestral $N_e = 2 \times 10^4$ and 2×10^6, respectively, simulated with a corresponding 50× rescaling in the large population). We used this rescaling in both the forward time (selection) and backward time (coalescent drift) simulations for consistency when simulating evolution in the large populations.

In addition to simulating evolution on the simplified demographic histories outlined above, which facilitate identification of the effects of specific parameters including population size and migration on each statistic, we also simulated selective sweeps on a more complex demography—specifically, a three population OOA model using the parameter estimates from Gutenkunst et al. (2009), using parameter values reported in the documentation of popsim (Adrion et al. 2020). In these simulations, we treated the western European population as the focal population for local sweeps, with Han Chinese and Yoruba (west Africa) as the first and second outgroups, respectively. Thus, in terms of the labels used in our other histories, A = Western Europe, B = Han Chinese, and C = Yoruba.

Among the most salient differences between our small population model and OOA is the much deeper split time for the former. While the Europe-China estimated split time at 848 generations closely resembles the small population model A,B split at 1,000, Gutenkunst et al. estimate 5,600 generations for the original OOA Eurasia split (vs. 2,000 generations for our more distant outgroup C). The OOA model also includes much stronger population bottlenecks than in our other simulations—both the original OOA split and those for Europe and (especially) China are smaller than their ancestral populations by an order of magnitude (as opposed to a factor of 2× in our other models). Additionally, the Guntenkunst OOA demographic model includes per-generation growth rates of 0.4% and 0.055% for Europe and China, respectively, in contrast to the constant population sizes in our large and small demographic models.

Parameterizing the Simulations

We generated genomic windows so that in our simulations a "locus" is defined not as a single site, but as a chromosome region. We simulated 100 and 5 kb regions for the small and large populations, respectively, following previous simulation studies (e.g. Lange and Pool 2016; da Silva Ribeiro et al. 2022), and equal or close to window sizes from empirical genome scan studies (e.g. Voight et al. 2006; Granka et al. 2012; Lack et al. 2015; Pool et al. 2017). The 100 kb window size also reflects the lengths

of typical to large protein-coding genes and their flanking regions, as well as a distance at which linkage disequilibria decay to modest levels in typical human recombination environments (McVean 2002). The 20-fold smaller window length for the large populations also represents a compromise between the approximately 10-fold higher population mutation rate and the roughly 40-fold higher population recombination rate estimated for fruit fly versus human populations. While constant length windows are used for convenience in this simulation study, we note that in empirical studies, it is often preferable to specify window size with respect to the diversity each window contains, as opposed to a constant physical window size.

Per-site mutation, recombination, and gene-conversion rates were based on the mean autosomal estimates in *D. melanogaster* from Comeron et al. (2012) and for humans from Schiffels and Durbin (2014), Williams et al. (2015), and Jonsson et al. (2017). Figure 1c summarizes the genomic parameters used in the simulations, including the per-generation/per-site mutation rates, crossing-over rates, as well as the gene-conversion rates and conversion tract lengths. The estimated mutation rate in the OOA model of Gutenkunst et al. (2009) was 2.35e-8, which is somewhat higher than the estimate of 1.22e-8 that we used in our small population models. For consistency with the rest of the parameters in Gutenskunst et al.'s analysis, we used this higher estimated rate when simulating the OOA genealogy.

In purely neutral simulations, genetic drift was modeled as a backward time coalescent process using the msprime package 1.2.0 (Baumdicker et al. 2022) for the demographic models and parameters summarized in Fig. 1b and c. The msprime commands were embedded in Python script "wrappers" that calculated allele frequencies and the population statistics. We also used msprime to simulate the OOA demography.

Background Selection

The effects of BGS (Charlesworth et al. 1993, 1995) were simulated without explicitly modeling deleterious mutations or negative selection as such. Rather, we modeled BGS implicitly by simulating the effects of negative selection on the genetic diversity of linked neutral variation. While BGS has been simulated for smaller populations (e.g. Boyko et al. 2008, Marsh and Johri 2024), we used the drift-based model as an approximation because for the large populations in particular, simulating BGS explicitly is not practical—it would require modeling negative selection in the full (not rescaled) population sizes of $\sim 10^6$. An additional order of magnitude higher generation times would have to be simulated to approximate an equilibrium distribution (which would be time-intensive even for the smaller populations and prohibitive for large N_e).

Charlesworth (2012) demonstrated that the reduction in linked neutral genetic variation due to BGS is equivalent to that under a reduction of effective population size by a factor B which depends on the recombination rate in a genomic window (i.e. regions of low recombination correspond to B-values much less than one). As examples, rescaling N_e (and therefore expected coalescent times) with B-values to model the effects of BGS on neutral genetic variation was applied by Jiang and Assis (2020) to estimates of F_{ST} among populations evolving with BGS, and in the calculations of Huber et al. (2015) of allele frequency spectra under BGS. There is evidence that modeling BGS as drift with reduced N_e is inadequate for demographic inference because of its failure to generate equivalent allele frequency spectra (Ewing and Jensen 2016), and BGS is known to change the shape of genealogies (Nicolaisen and Desai 2013). Hence, drift-based approximations would seem inappropriate for studies making quantitative inferences about the BGS process itself. In contrast, because the B-value approximation does capture the reduction in nucleotide diversity due to BGS as well as the expected differences in allele frequencies between populations, it should recapitulate the resulting increases in allele frequency differentiation between populations to a sufficient degree for a study not focused on parameter estimation.

We model the effects of BGS on neutral genetic variation by using the distribution of B-values estimated for chromosome regions in *D. melanogaster* in Comeron (2014) and for the human genome in McVicker et al. (2009). In each simulation, a value of B is randomly drawn from the frequency distribution (from the *Drosophila* autosome distribution for large population simulations, the human autosome distribution for small populations) and used as a multiplier to rescale the effective population size. The evolution of genetic diversity in a region of genome with this value of B is simulated as genetic drift using msprime with $N_e^* = BN_e$ in place of N_e. This BGS approximation was not implemented for the loci modeled with selective sweeps. Although the effect of the strong positive selection modeled here on genetic differentiation may be stronger in some cases than that of BGS, the absence of a modeled N_e reduction at sweep loci may somewhat reduce the power to detect local sweeps reported here. However, this effect may be somewhat mitigated by the exclusion of regions with strong BGS from our simulated genomes, as described below. Whereas, when simulating selective sweeps in the OOA demographic model, we did include BGS when implementing recapitation at the selected locus for the sake of greater realism.

Certain chromosome regions, particularly those near centromeres in *Drosophila*, have very low B, i.e. < 0.01 (Fig. 2b), resulting in very small rescaled effective population size and greatly inflated F_{ST} for simulated loci in those intervals. We corrected this potential artifact by

removing *B*-values corresponding to regions where the sex-averaged crossing-over rate $r < 0.5$ cM/Mb, analogous to the removal of low recombination regions such as those around autosomal centromeres in many genome-wide scans of *D. melanogaster* (e.g. Pool et al. 2017). Low recombination regions are often excluded from genome scans because of (1) the greater variance in diversity patterns expected under neutrality (Booker et al. 2020), (2) stronger effects of BGS, and (3) the difficulty in localizing outliers due to positive selection in low recombination regions.

To find an appropriate cutoff, we performed a LOESS regression of *D. melanogaster B*-values against recombination rate and identified the value of *B* predicted at $r = 0.5$ by the regression model. This corresponds to $B = 0.41$, so the *B*-value distribution was truncated at $B > 0.41$. For consistency, this was done for both the human and fruit fly distributions, even though human chromosomes had only 7% of regions where $B < 0.41$% versus 30% of fruit fly autosomes. Simulations of BGS as genomically variable drift were run by sampling *B* from these truncated distributions of *B*-values for each simulated genomic window (i.e. simulation replicate).

Selective Sweeps

All simulations of positive selection (including under the OOA demographic models) were performed using the SLiM v. 3.5 software package (Messer 2013; Haller et al. 2019). Unlike the coalescent-based models in msprime, SLiM simulations run in forward time. The SLiM output consists of a tree structure (the sequences and their genealogical history) of the site(s) under selection. To generate the corresponding history of neutral sites that evolve in association with the sites evolving under directional selection, we used msprime to simulate their coalescent history and link them to the tree structure generated in SLiM, a process known as recapitation of the genealogy, as described in Kelleher et al. (2018) and Haller et al. (2019).

Most of the SLiM simulations were initiated by introducing a beneficial mutation at a randomly selected single site in the focal population in the first generation following the split from its outgroup. The simulation runs were conditioned on either the fixation of the beneficial allele or it reaching a specified terminal frequency $P_t < 1.0$. If the beneficial mutation was lost from the population during the simulation, the simulation was terminated and restarted. Thus, only the output tree structures of those replicates that reached the desired terminal frequencies were retained.

Variations on this basic model of selective sweeps included conditioning on intermediate $P_t < 1.0$ (partial sweeps), soft sweeps where the initial frequencies were orders of magnitude greater than $1/(2N)$, and simultaneous selection at the same locus in both the focal population and its outgroups (parallel sweeps). All of the selective sweep scenarios described below were run to generate 10,000 replicates for analysis.

A. *Complete local hard sweeps*: These simulations used selection coefficients $s = 0.025$ and 0.001 in the small and large populations, respectively (with the smaller *s* in the larger N_e populations reflecting the expectation that weaker beneficial mutations can successfully be favored by selection in larger populations). With the population size rescaling by 50× of the large populations, the rescaled selection coefficient used in those simulations was $s' = 0.05$. These selection coefficients were chosen as the smallest *s* that would consistently achieve fixation or near-fixation for most runs within the time since the populations split, i.e. a compromise between the fact that most true selection coefficients may be relatively small versus having to be large enough to generate complete sweeps.

   Because of the high probability that a single beneficial mutation will be lost or fail to be fixed in a simulation with small N_e and $s << 1$, we used an initial frequency $P_0 = 0.001$ rather than $P_0 = 1/(2N_e)$ (set at the time of the focal population's split from its outgroup) in the small N_e scenarios, to reduce the number of times that simulations had to be restarted to achieve fixation. However, in the large N_e scenarios, we used $P_0 = 1/(2N_e)$ to avoid soft sweep effects in the rescaled model populations. Previous studies, e.g. da Silva Ribeiro et al. (2022) have shown that these initial frequencies are sufficiently low to generate hard sweep outcomes (i.e. only one haplotype carrying the beneficial mutation tends to contribute to the sweep). Additionally, because many simulations did not reach fixation by the final generation, rather than restarting all simulations where the $P_t < 1.0$, we implemented a strategy in which simulations with $P_t > 0.95$ were saved and only those haplotypes with the beneficial mutation were sampled for F_{ST} estimation and other analyses. These near-fixations provide a close approximation to a complete sweep, so the approach was also used when modeling parallel and soft sweeps.

B. *Complete local soft sweeps*. Soft sweeps can occur when positive selection acts on alleles that had already achieved a frequency significantly higher than $1/(2N_e)$ in the population through genetic drift, and multiple initial haplotypes contribute to the sweep, contributing to greater haplotype diversity after fixation than hard sweeps (Hermisson and Pennings 2005). We simulated soft sweeps with initial beneficial allele frequencies of $P_0 = 0.005$ and 0.01 in both the large and the small populations, using the same selection coefficients and initiation times at the A,B split as in the complete sweep simulations. These initial frequencies were chosen

because they were expected to generate soft sweep outcomes while remaining plausibly detectable with window statistics (da Silva Ribeiro et al. 2022).

C. *Partial hard sweeps*. In practice, many beneficial alleles may never become fixed, or they may be sampled at time points prior to their fixation in a population. To simulate (hard) partial sweeps, we applied positive selection to an allele in the focal population using the same initial frequencies and selection coefficients as described for complete local sweeps. To target sweeps reaching terminal frequency $P_t = 0.5$, we introduced the beneficial mutations at $t = 1{,}700$ and 3,800 generations, i.e. 300 and 200 generations prior to termination in the small and large populations, respectively; for $P_t = 0.8$, the beneficial mutations were introduced at generation times 1,500 and 3,700. These initiation times were chosen because they generated terminal frequencies close to the desired P_t. The terminal frequencies were selected as meaningfully different from complete sweeps while remaining potentially detectable (da Silva Ribeiro et al. 2022). Simulations were repeated until a terminal allele frequency between 0.45 and 0.55 was observed in the last generation to simulate partial sweeps to 50% (or between 0.75 and 0.85 to approximate $P_t = 0.8$). Unlike for the complete sweeps, haplotypes were sampled randomly from the entire population at the end of each saved simulation, as opposed to only sampling haplotypes carrying the beneficial allele.

D. *Parallel sweeps*. We simulated parallel selective sweeps by introducing a distinct beneficial mutation in all three populations at the same position. For hard complete parallel sweeps, we used the same initial frequencies and selection coefficients as were used for hard complete and partial local sweeps. Beneficial mutations were introduced at $t = 1{,}000$ and $t = 2{,}000$ at randomly selected sites within the same genomic window (locus) in all three populations (corresponding to the time of the A,B population split in the small and large populations, respectively) in order to allow maximum time to reach fixation. As with hard complete local sweeps, the simulations were re-run until the beneficial mutations reached $P_t > 0.95$, and if not fixed, only haplotypes carrying the beneficial mutation were sampled for analysis.

For simplicity, parallel sweeps were always simulated consistently across populations, i.e. if there was a hard complete sweep in the focal population, there were hard complete sweeps in the outgroup populations as well, etc. We also simulated scenarios of soft parallel sweeps and (hard) partial parallel sweeps in all three populations so that the resulting allele frequencies and F_{ST}-based statistics could be compared to those generated by the same positive selection models in local sweeps. The parallel soft and parallel partial sweeps used the same initial frequencies, selection coefficients, initiation times for the introduction of beneficial mutations, and targeted terminal frequencies as in the corresponding local sweep simulations, but applied these conditions to all three populations rather than the focal population alone.

When simulating selective sweeps in the OOA models, we considered two of the most biologically plausible scenarios: a soft sweep from an initial frequency of $P_0 = 0.01$, and a partial (hard) sweep from $P_0 = 1/(2N)$ to a terminal frequency P_t of approximately ~ 0.5 (constrained to 0.45 to 0.55). Additionally, these scenarios were considered because they tend to generate patterns of haplotype diversity with weaker "signal" than hard complete sweeps, for which performance metrics are nearly identical for all F_{ST}-derived statistics and therefore rather trivial. For the same reasons, we assumed BGS rather than pure genetic drift at all other loci.

The soft sweep was modeled by introducing a beneficial mutation in the European population at the time of the Europe-Asia split (848 generations ago); 20 favored alleles were introduced at the time of the split so that $P_0 = 0.01$ with respect to the initial size of the European population at the time of the split (as opposed to its much larger terminal size). Parallel soft sweeps were modeled by introducing a beneficial mutation at the same locus but at a different site in the Yoruba and China populations, also with an initial number of favored alleles so that $P_0 = 0.01$ based on their respective population sizes at 848 generations before present.

Because of the slightly shorter branch for the European population than for the focal population "A" in our simplified small population demography (i.e. 848 generations vs. 1,000), as well as the effects of migration, a selection coefficient of $s = 0.025$ was not sufficient to achieve a terminal frequency $P_t \geq 0.95$ (particularly with migration) in the majority of replicates, so we used a slightly higher selective coefficient of $s = 0.03$. We simulated partial hard sweeps in the OOA model using the same selection coefficient, by introducing the beneficial mutation at 300 generations before present census, so that most simulations would attain the desired terminal frequency of $P_t \sim 0.5$. As with the simplified histories, a single beneficial mutation was introduced at this time in the focal population (local sweeps) and all three populations (parallel sweeps), thus simulating a hard partial sweep.

Simulating Migration

A subset of simulations included migration among all populations in order to determine the extent to which results on the relative efficacy of branch statistics were robust in the presence of gene flow. We considered two demographic

models with migration. One retains the split times of the migration-free models (left tree in Fig. 1b), with a symmetric pairwise migration rate introduced so that $N_e m = 1$ per-generation between all pairs of populations. The migration rates were set so that each population received on average a single migrant per-generation independent of population size (such that m was smaller for small population receiving migrants from a large population than the reverse) to avoid source-sink effects. In the rescaling of the large populations, we set $N_e' m' = 1$ (where $N_e' = N_e/50$, $m' = 50$ m) so that there would be comparable numbers of migrants exchanged among pairs of populations over the same number of coalescent time units. A population migration rate with $Nm = 1$ was chosen to test the robustness of population branch statistics to moderate gene flow, i.e. at values $Nm > 1$ there is little population differentiation (Varvio et al. 1986), whereas for $Nm << 1$ the differentiation is trivially similar to models with no gene flow among populations.

Because migration is expected to reduce pairwise differentiation among populations, the baseline expected F_{ST} at neutrally evolving sites will be lower in the presence of migration than in its absence, so a high or low branch statistic in the absence of migration may not be comparable to values observed in populations evolving with migration. Therefore, we also implemented a longer branch demographic history for a three-population model with migration that generates very similar pairwise F_{ST} among all pairs as the migration-free model. For the same population sizes, we introduced split times at 3×10^3 and 5×10^3 generations before the present for the small population models and 3×10^6, 5×10^6 for the large population models (right tree in Fig. 1b).

Simulations of genetic drift and selection with migration were executed in msprime and SLiM with the same selection coefficients in the small and large populations as for the models of selective sweeps without migration.

Because of migration, the selective sweeps are not expected to be complete, so we conditioned the local sweeps on attaining a terminal frequency $P_t > 0.75$, and imposed the same constraint on all three populations when simulating parallel sweeps in both the large and small N_e scenarios. This threshold value was selected to eliminate cases where the beneficial mutation was nearly lost while allowing for intermediate frequencies due to migration (e.g. when simulating 1,000 replicates of local hard sweeps, we found that 98% of the replicates attained frequencies $P_t > 0.80$). As with the partial sweep simulations, all haplotypes were included in the final sample rather than only those carrying the beneficial mutation because we are not simulating fixation. Beneficial mutations were introduced at the time of the split between populations A,B, as was done for the hard complete sweep models without migration.

In simulations of migration with local sweeps, the beneficial mutation in the focal population was neutral in the two nonfocal populations. Similarly, when modeling parallel selection with migration, the beneficial mutation for each population was neutral with respect to the wild-type background in the other two populations (and thus having lower fitness compared to the beneficial mutation within its home population).

For the OOA genealogy, there are two important qualitative differences in how migration is modeled compared to our simplified small population simulations (in which all populations exchange migrants at equal and constant rates). First, the migration rates vary as predicted by geography in OOA, with the highest migration rate being between Europe and China, intermediate rates for Europe and Africa, and the lowest for Africa and China (as well as a high rate between Africa and the original OOA branch). Additionally, in the Gutenkunst et al. (2009) OOA model, while $m_{ij} = m_{ji}$ are symmetric for each population pair i,j, the population migration rates Nm are not (and moreover, they vary with time because of population expansion).

Statistics Evaluated

At the termination of each simulation run, we calculated F_{ST} among subpopulations based on the Reynolds et al. (1983) estimate of the F_{ST} as well as PBS and PBSn1. We note that in practice, $0 < F_{ST} < 1$ in nearly all of the simulated windows, but we did not truncate values outside of this range for our rank-based statistics in the rare instances when they did occur. The branch length T_{BC} was also output for each simulated replicate, since this is needed alongside PBS values for the subsequent calculation of PBE, which is calculated with respect to an aggregate of loci in a genome (Eq. (4)). We also evaluated LSBL for the same simulation data, with raw F_{ST} values in place of T from Eq. (2) for comparison to the log-rescaled population branch statistics. Sequences of segregating sites generated by the simulations were saved as output.

Most of our analyses focused on full window versions of F_{ST}, PBS, PBSn1, and PBE (Eqs. (1)–(4)), which are calculated in aggregate from the ratio of the sums of the numerator terms and the denominator terms of F_{ST} (as defined in Reynolds et al. 1983) across all sites within a window. For a subset of sweep models, we also computed each statistic based on maximum SNP-level values observed within each window. F_{ST} was calculated for each site in a window, and F_{ST_MaxSNP} is the maximum value in the window, with $PBS_{\_MaxSNP}$ defined in the same way as the largest PBS value across sites in the window.

We initially used "constrained" versions of site-based PBSn1 and PBE, here labeled $PBSn1_{MaxSNP}$* and PBE_{MaxSNP}*, whose component terms were calculated with reference to the same site in a window. Thus,

$PBSn1_{MaxSNP}$* was evaluated from Eq. (3) using outgroup branch lengths PBS_B and PBS_C calculated at the same site as focal PBS_A, and then selecting the maximum site-level PBSn1 value in the window. Similarly, we computed PBE_{MaxSNP}* from Eq. (4) using PBS_A and T_{BC} from the same site, while median T_{BC} and median PBS were calculated from aggregate window PBS and F_{ST} values as a comparison baseline (Pool et al. 2017).

Thus, we have:

$$PBSn1_{MaxSNP}* = \max\left(\frac{PBS_{A,i}}{1 + PBS_{A,i} + PBS_{B,i} + PBS_{C,i}}\right) \quad (5)$$

with PBS values for A, B, and C calculated for the same site *i*. We refer to this maximum site PBSn1 as site-constrained $PBSn1_{MaxSNP}$* elsewhere in the text. Similarly, the site-constrained PBE* where T_{BC} terms are calculated at the same site as PBS is:

$$PBE\__{MaxSNP}* = \max\left(PBS_{A,i} - \frac{T_{BC,i} \times \text{med}(PBS_{A,\text{ window}})}{\text{med}(T_{BC,\text{window}})}\right) \quad (6)$$

These definitions of $PBSn1_{MaxSNP}$* and PBE_{MaxSNP}* share the property of being maximized at the same sites in a genomic window as F_{ST} and PBS. Consequently, they are all maximized at sites where a mutation is fixed in the focal population but absent from others, regardless of the frequency configurations that may be present at other sites in the same window, i.e. a frequency configuration of 1,0,0, the focal and outgroup populations results in maximal site F_{ST}, PBS, PBSn1, and PBE, regardless of how many sites in a window have fixed variants, or whether there also exist highly differentiated sites between the two outgroup populations within this window.

Because the rescaling for PBS involves the term $T = -\log(1-F_{ST})$, we avoid log(0) by replacing the case of a sample of *n*,0 alleles with pseudocounts of *n*,0.5 alleles. In our case, with a sample size of 25 diploid individuals, an allele fixed in one population and absent in the others gives $F_{STMaxSNP} = 0.9898$, $PBS_{MaxSNP} = 4.585$, $PBSn1_{MaxSNP}* = 0.8209$. Meanwhile, $PBE_{MaxSNP}* = PBS_{MaxSNP} = 4.585$ because the site $T_{BC} = 0$ when the variant allele is absent in outgroup populations. Consequently, any selective sweep resulting in fixation will give the same maximum value, which will be identical for local and parallel sweeps because the parallel sweep typically involves different sites at the same locus (genomic window) rather than at the same site in all three populations. This dynamic implies that all of the statistics will have nearly identical precision at detecting local sweeps, and that they will be unable to distinguish local from parallel sweeps under these scenarios.

To circumvent this problem, we introduce revised heuristic definitions for $PBSn1_{MaxSNP}$ and PBE_{MaxSNP} which can distinguish between local and parallel sweeps by virtue of having higher values for the former. Rather than constraining the calculation of each component term to a single site for PBSn1, we normalize the maximum site PBS_A with respect to the maximum site PBS_B and PBS_C within the window (which will usually each occur at different sites) to define a site-unconstrained maximum site PBSn1 as

$$PBSn1_{MaxSNP} = \frac{\max(PBS_{A,i})}{1 + \max(PBS_{A,i}) + \max(PBS_{B,j}) + \max(PBS_{C,k})} \quad (7)$$

where the indices *i,j,k* indicate that the PBS values are maximized at potentially different sites in the genomic window. Similarly, we define a site-unconstrained maximum site PBE as:

$$PBE_{MaxSNP} = \max(PBS_{A,i}) - \frac{\max(T_{BC,j}) \times \text{med}(\max(PBS_{A,k}))}{\text{med}(\max(T_{BC,l}))}, \quad (8)$$

where *i,j* are potentially different sites in the target window that respectively maximize PBS and T_{BC}, while *k,l* denote the sites in other genomic windows that maximize each of those quantities within those windows. These median values are across the site maxima of all windows; hence this formulation accounts for a typical site maximum values of PBS_A and T_{BC} observed across the genome.

These nonsite constrained definitions of maximum site PBSn1 and PBE have the desired property of being larger for local sweeps than for parallel sweeps. In the local sweep case, the scaling terms $\max(PBS_B)$, $\max(PBS_C)$, and $\max(T_{BC})$ are limited to the range of values generated by neutral evolution (or BGS). Whereas, in the parallel sweep case, they all assume their maximum possible site values (at different sites in the locus), which reduces the magnitudes of our unconstrained $PBSn1_{MaxSNP}$ and PBE_{MaxSNP}.

Model Genomes and Statistical Analysis

The msprime simulations, including the recapitated SLiM trees, generate a distribution of multilocus genotypes from which we calculated allele frequencies in the three populations, and F_{ST} at each locus for every population pair. We used the F_{ST} estimator in Reynolds et al. (1983) and Weir and Cockerham (1984) because it incorporates the contribution of sampling variances and generalizes to a multilocus estimator for F_{ST} across multiple sites, allowing F_{ST} to be calculated for genomic "windows." The branch statistics PBS, and PBSn1 for focal population A were calculated from the genomic window F_{ST} using Equations (1) and (2).

Here, each model genome consisted of a collection of simulated replicates that include both a subset of local sweep

replicates as well as replicates from one or more other evolutionary models. We created four types of model genomes representing scenarios where the majority of sites evolved either neutrally or under BGS, while a small fraction experienced selective sweeps (Fig. 2a). The genomes consisted of 25,000 loci (genomic windows from independent simulation replicates) for both the small and large populations, roughly aligning with the full genome size for human and *Drosophila*, respectively. In one scenario, an expected 99% of the loci were sampled from the million replicates of genetic drift simulations, and the remaining fraction from the 10,000 local sweep runs (the actual number of selected loci was determined by a binomial random variable with parameters $n,P = 25000,\ 0.01$, where P is the expected fraction of selected loci). A similar model had an expected 99% of loci generated under a BGS model rather than drift (we do not simulate cases that are mixtures of BGS and drift, because our BGS models include a subset of sites with B close to 1 that essentially evolve as under neutral drift). The other sets of model genomes combined parallel and local sweeps, with an expected 98% of loci sampled from the genetic drift (or BGS) replicates, 1% from local sweeps, and 1% from parallel sweeps (with realized numbers of sites in each category determined by a multinomial random variable with parameters $n,\ P_1,\ P_2 = 25{,}000,\ 0.01,\ 0.01$, with P_1 and P_2 the expected proportion of local and parallel sweep loci). Focal population PBE at each locus was computed from the locus-specific PBS and T_{BC} and from the median PBS and T_{BC} values across all loci (Eq. (3)) within the model genome.

After the completion of each simulation, we sampled 25 diploid individuals from each population and calculated pairwise F_{ST} and the population branch statistics (this number was chosen because samples from wild populations often involve roughly this many individuals per location). We quantified the relative performance of F_{ST}, PBS, PBSn1, and PBE by calculating their probability of identifying local sweep loci. In this context, true positives were defined as local sweep loci in the upper 1% quantile of a statistic's distribution, while false positives were any nonlocal sweep loci in this quantile (i.e. loci evolving under genetic drift, BGS, or parallel sweeps). The ratio of true positives to all positives defined the statistical *precision* of each test. For the genomes with both local and parallel sweeps, we also calculated the fraction of parallel sweep loci that contributed false negatives, and the fraction of false negatives that were from parallel sweeps rather than drift or BGS.

Above, we describe the estimation of false positive rates in a simulated genome context. As a complement to this approach, we also computed the statistical power of F_{ST} and the branch statistics in a simpler and more conventional sense. Here, we separately compared each simulated local sweep replicate against the full set of replicates simulated under a given null model (either genetic drift or BGS). The statistical power was calculated as the fraction of selective sweep replicates that were within a given statistic's upper 1% quantile for the neutral or BGS replicates.

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Acknowledgments

The authors thank the members of the Pool lab, Josep Comeron, and Aaron Ragsdale for helpful discussion. We also thank Ben Haller, Peter Ralph, Jerome Kelleher, and other developers of SLiM and tskit/msprime in helping to resolve initial issues with the simulations.

Author Contributions

M.S. and J.E.P. created the study design and wrote the manuscript, M.S. and K.N.L. ran the simulations and analyzed the output. M.S. and J.E.P. wrote the paper with input from K.N.L.

Funding

This work was supported by National Institutes Health award R35 GM136306 and by National Science Foundation award DEB 1754745 to JEP.

Data Availability

All of the msprime and SLiM scripts used to generate simulated data, along with scripts used to analyze that data, are available at https://github.com/mshpak76/PBS_PBE.

Literature Cited

Adrion JR, Cole CB, Dukler N, Galloway JG, Gladstein AL, Gower G, Kyriazis CC, Ragsdale AP, Tsambos G, Baumdicker F, et al. A community-maintained standard library of population genetic models. Elife. 2020:9:e54967. https://doi.org/10.7554/eLife.54967.

Adrion JR, Hahn MW, Cooper BS. Revisiting classic clines in *Drosophila melanogaster* in the age of genomics. Trends Genet. 2015:31(8):P434–P444. https://doi.org/10.1016/j.tig.2015.05.006.

Akey JM. Constructing genomic maps of positive selection in humans: where do we go from here? Genome Res. 2009:19(5):711–722. https://doi.org/10.1101/gr.086652.108.

Amato R, Pinelli M, Monticelli A, Marino D, Miele G, Cocozza S. Genome-wide scan for signatures of human population differentiation and their relationship with natural selection, functional pathways and diseases. PLoS One. 2009:4(11):e7927. https://doi.org/10.1371/journal.pone.0007927.

Antonovics J, Bradshaw AD. Evolution in closely adjacent plant populations. VIII. Clincal patterns in *Anthoxanthum odoratum* across a mine boundary. Heredity (Edinb). 1970:25(3):249–362. https://doi.org/10.1038/hdy.1970.36.

Baumdicker F, Bisschop G, Goldstein D, Gower G, Ragsdale AP, Tsambos G, Zhu S, Eldon B, Ellerman EC, Galloway JG, et al. Efficient ancestry

and mutation simulation with msprime 1.0. Genetics. 2022:220(3): iyab229. https://doi.org/10.1093/genetics/iyab229.
Booker TR, Yeaman S, Whitlock MC. Variation in recombination rate affects detection of outliers in genome scans under neutrality. Mol Ecol. 2020:29(22):4274–4279. https://doi.org/10.1111/mec.15501.
Boyko AR, Williamson SH, Indap AR, Degenhardt JD, Hernandez RD, Lohmueller KE, Adams MD, Schmidt S, Sninsky JJ, Sunyaev SR, et al. Assessing the evolutionary impact of amino acid mutations in the human genome. PLoS Genet. 2008:30(5):e1000083. https://doi.org/10.1371/journal.pgen.1000083.
Brodie ED Jr, Ridenhour BJ, Brodie ED III. The evolutionary response of predators to dangerous prey: hotspots and coldspots in the geographic mosaic of coevolution between garter snakes and newts. Evolution. 2002:56(10):2067–2082. https://doi.org/10.1111/j.0014-3820.2002.tb00132.x.
Cavalli-Sforza L. Human Diversity. Proceedings of the 12th International Congress of Genetics; Tokyo; 1969:3:405–416.
Charlesworth B. The effects of deleterious mutations on evolution at linked sites. Genetics. 2012:190(1):5–22. https://doi.org/10.1534/genetics.111.134288.
Charlesworth B, Morgan MT, Charlesworth D. The effect of deleterious mutations on neutral molecular variation. Genetics. 1993:134(4): 1289–1303. https://doi.org/10.1093/genetics/134.4.1289.
Charlesworth D, Charlesworth B, Morgan MT. The pattern of neutral molecular variation under the background selection model. Genetics. 1995:141(4):1619–1632. https://doi.org/10.1093/genetics/141.4.1619.
Chen H, Patterson N, Reich D. Population differentiation as a test for selective sweeps. Genome Res. 2010:20(3):393–402. https://doi.org/10.1101/gr.100545.109.
Cheng JY, Stern AJ, Racimo F, Nielsen R. Detecting selection in multiple populations by modeling ancestral admixture components. Mol Biol Evol. 2022:39(1):msab294. https://doi.org/10.1093/molbev/msab294.
Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G Jr, Dickson M, Grimwood J, Schmutz J, Myers RM, Schluter D, Kingsley DM. Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. Science. 2005:307(5717):1928–1933. https://doi.org/10.1126/science.1107239.
Comeron JM. Background selection as baseline for nucleotide variation across the *Drosophila* genome. PLoS Genet. 2014:10(6): e1004434. https://doi.org/10.1371/journal.pgen.1004434.
Comeron JM, Ratnappan R, Bailin S. The many landscapes of recombination in *Drosophila melanogaster*. PLoS Genet. 2012:10(6): e1004434. https://doi.org/10.1371/journal.pgen.1002905.
Crawford JE, Alves JM, Palmer WJ, Day JP, Sylla M, Ramasamy R, Surendran SN, Black WC, Pain A, Jiggins FM. Population genomics reveals that an anthropophilic population of Aedes aegypti mosquitoes in West Africa recently gave rise to American and Asian populations of this major disease vector. BMC Biology. 2017:15:1–6. https://doi.org/10.1186/s12915-017-0351-0.
da Silva Ribeiro T, Galvan JA, Pool JE. Maximum SNP FST outperforms full-window statistics for detecting soft sweeps in local adaptation. Genome Biol Evol. 2022:14(10):evac143. https://doi.org/10.1093/gbe/evac143.
Ewing GB, Jensen JD. The consequences of not accounting for background selection in demographic inference. Mol Ecol. 2016:25(1):135–141. https://doi.org/10.1111/mec.13390.
Fraïsse C, Belkhir K, Welch JJ, Bierne N. Local interspecies introgression is the main cause of extreme levels of intraspecific differentiation in mussels. Mol Ecol. 2016:25(1):269–286. https://doi.org/10.1111/mec.13299.
Granka JM, Henn BM, Gignoux CR, Kidd JM, Bustamante CD, Feldman MW. Limited evidence for classic selective sweeps in African populations. Genetics. 2012:192(3):1049–1064. https://doi.org/10.1534/genetics.112.144071.
Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 2009:5(10):e100695. https://doi.org/10.1371/journal.pgen.1000695.
Gygax M, Rentsch AK, Rudman SM, Rennison DJ. Differential predation alters pigmentation in threespine stickleback (*Gasterosteus aculeatus*). J Evol Biol. 2018:31(10):1589–1598. https://doi.org/10.1111/jeb.13354.
Haller BC, Galloway J, Kelleher J, Messer PW, Ralph PL. Tree-sequence recording in SLiM opens new horizons for forward-time simulation of whole genomes. Mol Ecol Resour. 2019:19(2):552–566. https://doi.org/10.1111/1755-0998.12968.
Hermisson J, Pennings PS. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. Genetics. 2005:169(4):2335–2352. https://doi.org/10.1534/genetics.104.036947.
Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP. A single amino acid mutation contributes to adaptive beach mouse color pattern. Science. 2006:313(5783):101–104. https://doi.org/10.1126/science.1126121.
Huber CD, DeGiorgio M, Hellmann I, Nielsen R. Detecting recent selective sweeps while controlling for mutation rate and background selection. Mol Ecol. 2015:25(1):142–156. https://doi.org/10.1111/mec.13351.
Huerta-Sánchez E, Degiorgio M, Pagani L, Tarekegn A, Ekong R, Antao T, Cardona A, Montgomery HE, Cavalleri GL, Robbins PA, et al. Genetic signatures reveal high-altitude adaptation in a set of Ethiopian populations. Mol Biol Evol. 2013:30(8):1877–8188. https://doi.org/10.1093/molbev/mst089.
Jiang X, Assis R. Population-specific genetic and expression differentiation in Europeans. Genome Biol Evol. 2020:12(4):358–369. https://doi.org/10.1093/gbe/evaa021.
Jonsson H, Sulem P, Kehr B, Kristmundsdottir S, Zink F, Hjartarson E, Hardarson MT, Hjorleifsson KE, Eggertsson HP, Gudjonsson SA, et al. Parental influence on human germline de novo mutations in 1,548 trios from Iceland. Nature. 2017:549(7673):519–522. https://doi.org/10.1038/nature24018.
Julian CG, Moore LG. Human genetic adaptation to high altitude: evidence from the Andes. Genes (Basel). 2019:10(2):150. https://doi.org/10.3390/genes10020150.
Kapun M, Barrón MG, Staubach F, Obbard DJ, Wiberg RAW, Vieira J, Goubert C, Rota-Stabelli O, Kankare M, Bogaerts-Márquez M, et al. Genomic analysis of European *Drosophila melanogaster* populations reveals longitudinal structure, continent-wide selection, and previously unknown DNA viruses. Mol Biol Evol. 2020:37(9): 2661–2678. https://doi.org/10.1093/molbev/msaa120.
Kelleher J, Thornton KR, Ashander J, Ralph PL. Efficient pedigree recording for fast population genetics simulation. PLoS Comput Biol. 2018:14(11):e1006581. https://doi.org/10.1371/journal.pcbi.1006581.
Lack JB, Cardeno CM, Crepeau MW, Taylor W, Corbett-Detig RB, Stevens KA, Langley CH, Pool JE. The *Drosophila* genome nexus: a population genomic resource of 623 *Drosophila melanogaster* genomes, including 197 from a single ancestral range population. Genetics. 2015:199(4):1229–1241. https://doi.org/10.1534/genetics.115.174664.
Lack JB, Yassin A, Sprengelmeyer QD, Johanning EJ, David JR, Pool JE. Life history evolution and cellular mechanisms associated with increased size in high-altitude *Drosophila*. Ecol Evol. 2016:6(16): 5893–5906. https://doi.org/10.1002/ece3.2327.
Lange JD, Bastide H, Lack JB, Pool JE. A population genomic assessment of three decades of evolution in a natural *Drosophila*

population. Mol Biol Evol. 2022:39(2):msab368. https://doi.org/10.1093/molbev/msab368.
Lange JD, Pool JE. A haplotype method detects diverse scenarios of local adaptation from genomic sequence variation. Mol Ecol. 2016:25(13):3081–3100. https://doi.org/10.1111/mec.13671.
Lange JD, Pool JE. Impacts of recurrent hitchhiking on divergence and demographic inference in *Drosophila*. Genome Biol Evol. 2018:10(8):1882–1891. https://doi.org/10.1093/gbe/evy142.
Li H. A new test for detecting recent positive selection that is free from the confounding impacts of demography. Mol Biol Evol. 2011:28(1):365–375. https://doi.org/10.1093/molbev/msq211.
Macnair MR. Heavy metal tolerance in plants: a model evolutionary system. Trends Ecol Evol. 1987:2(12):354–359. https://doi.org/10.1016/0169-5347(87)90135-2.
Malaspinas A-S, Westaway MC, Muller C, Sousa VC, Lao O, Alves I, Bergström A, Athanasiadis G, Cheng JY, Crawford JE, et al. A genomic history of Aboriginal Australia. Nature. 2016:538(7624):207–2014. https://doi.org/10.1038/nature18299.
Marsh JI, Johri P. Biases in ARG-based inference of historical population size in populations experiencing selection. Mol Biol Evol. 2024:41(7):msae118. https://doi.org/10.1093/molbev/msae118.
McVean GA. A genealogical interpretation of linkage disequilibrium. Genetics. 2002:162(2):987–991. https://doi.org/10.1093/genetics/162.2.987.
McVicker G, Gordon D, Davis C, Green P. Widespread genomic signatures of natural selection in hominid evolution. PLoS Genet. 2009:5(5):e1000471. https://doi.org/10.1371/journal.pgen.1000471.
Messer PW. SLiM: simulating evolution with selection and linkage. Genetics. 2013:194(4):1037–1039. https://doi.org/10.1534/genetics.113.152181.
Miller SE, Metcalf D, Schluter D. Intraguild predation leads to genetically based character shifts in the threespine stickleback. Evolution. 2015:69(12):3194–3203. https://doi.org/10.1111/evo.12811.
Nachman MW, Hoekstra HE, D'Agostino SL. The genetic basis of adaptive melanism in pocket mice. Proc Natl Acad Sci U S A. 2003:100(9):5268–5273. https://doi.org/10.1073/pnas.0431157100.
Nicolaisen LE, Desai MM. Distortions in genealogies due to purifying selection and recombination. Genetics. 2013:195(1):221–230. https://doi.org/10.1534/genetics.113.152983.
Pool JE, Braun DT, Lack JB. Parallel evolution of cold tolerance within *Drosophila melanogaster*. Mol Biol Evol. 2017:34(2):349–360. https://doi.org/10.1093/molbev/msw232.
Racimo F. Testing for ancient selection using cross-population allele frequency differentiation. Genetics. 2016:202(2):733–750. https://doi.org/10.1534/genetics.115.178095.
Racimo F, Marnetto D, Huerta-Sanchez E. Signatures of archaic adaptive introgression in present-day human population. Mol Biol Evol. 2017:34(2):296–317. https://doi.org/10.1093/molbev/msw216.
Reynolds J, Weir BS, Cockerham CC. Estimation of the coancestry coefficient: basis for a short-term genetic distance. Genetics. 1983:105(3):767–779. https://doi.org/10.1093/genetics/105.3.767.
Schiffels S, Durbin R. Inferring human population size and separation history from multiple genome sequences. Nat Genet. 2014:46(8):919–925. https://doi.org/10.1038/ng.3015.
Schlebusch CM, Sjödin P, Breton G, Günther T, Naidoo T, Hollfelder N, Sjöstrand AE, Xu J, Gattepaille LM, Vicente M, et al. Khoe-San genomes reveal unique variation and confirm the deepest population divergence in *Homo sapiens*. Mol Biol Evol. 2020:37(10):2944–2954. https://doi.org/10.1093/molbev/msaa140.
Schmidt JM, de Manuel M, Marques-Bonet T, Castellano S, Andrés AM. The impact of genetic adaptation on chimpanzee subspecies differentiation. PLoS Genet. 2019:15(11):e1008485. https://doi.org/10.1371/journal.pgen.1008485.
Shriver MD, Kennedy GC, Parra EJ, Lawson HA, Sonpar V, Huang J, Akey JM, Jones KW. The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. Hum Genomics. 2004:1(4):274–286. https://doi.org/10.1186/1479-7364-1-4-274.
Siddiq MA, Thornton JW. Fitness effects but no temperature-mediated balancing selection at the polymorphic *Adh* gene of *Drosophila melanogaster*. Proc Natl Acad Sci U S A. 2019:116(43):21634–21640. https://doi.org/10.1073/pnas.1909216116.
Sprengelmeyer QD, Lack JB, Braun DT, Monette MJ, Pool JE. The evolution of larger size in high-altitude Drosophila melanogaster has a variable genetic architecture. G3 (Bethesda). 2022:12(3):jkab454. https://doi.org/10.1093/g3journal/jkab454.
Sprengelmeyer QD, Mansourian S, Lange JD, Matute DR, Cooper BS, Jirle EV, Stensmyr MC, Pool JE. Recurrent collection of *Drosophila melanogaster* from wild African environments and genomic insights into species history. Mol Biol Evol. 2020:37(3):627–638. https://doi.org/10.1093/molbev/msz271.
Sprengelmeyer QD, Pool JE. Ethanol resistance in *Drosophila melanogaster* has increased in parallel cold-adapted populations and shows a variable genetic architecture within and between populations. Ecol Evol. 2021:11(21):15364–15376. https://doi.org/10.1002/ece3.8228.
Svetec N, Zhao L, Saelao P, Chiu JC, Begun DJ. Evidence that natural selection maintains genetic variation for sleep in *Drosophila melanogaster*. BMC Evol Biol. 2015:15(1):41. https://doi.org/10.1186/s12862-015-0316-2.
Varvio S-L, Chakraborty R, Nei M. Genetic variation in subdivided populations and conservation genetics. Heredity (Edinb). 1986:57(Pt 2):189–198. https://doi.org/10.1038/hdy.1986.109.
Vicuna L, Fernandez MI, Vial C, Valdebenito P, Chaparro E, Espinoza K, Ziegler A, Bustamante A, Eyheramendy S. Adaptation to extreme environments in an admixed human population from the Atacama Desert. Genome Biol Evol. 2019:11(9):2468–2479. https://doi.org/10.1093/gbe/evz172.
Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. PLoS Biol. 2006:4(3):e72. https://doi.org/10.1371/journal.pbio.0040072.
Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984:38(6):1358–1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x.
Williams AL, Genovese G, Dyer T, Altemose N, Truax K, Jun G, Patterson N, Myers SR, Curran JE, Duggirala R, et al. Non-crossover gene conversions show strong GC bias and unexpected clustering in humans. Elife. 2015:4:e04637. https://doi.org/10.7554/eLife.04637.
Witt KE, Huerta-Sanchez E. Convergent evolution in human and domesticate adaptation to high-altitude environments. Philos Trans R Soc Lond B Biol Sci. 2019:374(1777):20180235. https://doi.org/10.1098/rstb.2018.0235.
Xue AT, Schrider DR, Kern AD; Ag1000g Consortium. Discovery of ongoing selective sweeps within *Anopheles* mosquito populations using deep learning. Mol Biol Evol. 2021:38(3):1168–1183. https://doi.org/10.1093/molbev/msaa259.
Yassin A, Debat V, Bastide H, Gidaszewski N, David JR, Pool JE. Recurrent specialization on a toxic fruit in an island *Drosophila* population. Proc Natl Acad Sci U S A. 2016:113(17):4771–4776. https://doi.org/10.1073/pnas.1522559113.
Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZX, Pool JE, Xu X, Jiang H, Vinckenbosch N, Korneliussen TS, et al. Sequencing of 50 human exomes reveals adaptation to high altitude. Science. 2010:329(5987):75–78. https://doi.org/10.1126/science.1190371.

Associate editor: Carolin Kosiol