Joint inference of adaptive and demographic history from temporal population genomic data

Vitor A. C. Pavinato[1,2,3], Stéphane De Mita[4], Jean-Michel Marin[2], Miguel de Navascués[1,5*]

[1]CBGP, INRAE, CIRAD, IRD, Montpellier SupAgro, Univ Montpellier, Montpellier, France

[2]UMR Institut Montpellirain Alexander Grothendieck, Universit de Montpellier, France [¡- because of the myfont command the letters with accents are missing]

[3]Entomology Dept., CFAES, The Ohio State University, Wooster, USA

[4]UMR Interactions Arbres-Microorganismes, INRAE, France

[5]Human Evolution, Department of Organismal Biology, Uppsala University, Uppsala, Sweden

[*]corresponding author: miguel.navascues@inrae.fr

March 12, 2021

Abstract

Disentangling the effects of selection and drift is a long-standing problem in population genetics. Theoretical works based on simulations show that the signal of selection may bias demographic inference when it is pervasive. Ideally, interactions between selection and demography should be considered in the estimation of parameters of demographic and adaptive models. One potential approach is to co-estimate demography and selection parameters using simulation-based likelihood-free methods such as Approximate Bayesian Computation (ABC). We propose a framework based on ABC via Random Forests to jointly infer demographic and selection parameters from temporal population genomic data (e.g. experimental evolution, monitored populations, ancient DNA). The proposed framework allows the separation of demography (census size, N) from genetic drift (effective population size, N_e) and the estimation of the genome-wide effect of selection as the scale mutation rate of beneficial mutations (θ_b). We applied this approach to a dataset of feral populations of *Apis mellifera* collected in California, and we estimated parameters consistent with the biology and the recent history of this species.

Keywords— Temporal data, Population genomics, Machine learning, Adaptation

Introduction

One aim of population genomics is to understand how demography and natural selection shape the genetic diversity of populations. A classical approach is to assume that demography (migration, population subdivision, population size changes) leaves a genome-wide signal, whereas selection leaves a localizes signal close to where the causal mutation is located. Many methods follow this approach to infer demography or selection (reviewed by Beichman et al., 2018; Casillas and Barbadilla, 2017). Methods for demographic inference assume that most of the genome evolves without the influence of selection and that any deviation from the mutation-drift equilibrium observed in the data was caused by demographic events (Beichman et al., 2018). For selection, the majority of the methods search for locus-specific signals left by the beneficial mutation on nearby neutral mutations (Tajima, 1989; Fay and Wu, 2000; Kim and Nielsen, 2004) (low genetic diversity and high differentiation) to localize the beneficial mutation, assuming constant population size (Nielsen, 2005; Pool et al., 2010).

Conducting demographic and selection inference separately may have some shortcomings. First, there is the assumption that the signal left by demography is little affected by selection because selection is rare. However, linked selection can affect neutral and weakly selected sites that are far from the mutation targeted by selection (Sella et al., 2009; Neher, 2013) and selection can be pervasive (Sella et al., 2009; Lange and Pool, 2018). In addition, some methods for selection scans are not robust to misspecifications of demographic history. Consequently, an unspecified bottleneck or population increase, for example, can inflate the type I error rate of genome scans (Jensen et al., 2005, 2007; Schrider et al., 2016).These findings highlight the necessity of inferential methods that jointly accounts for the multiple evolutionary forces that act on populations (Lin et al., 2011; Li et al., 2012; Bank et al., 2014).

It is difficult to obtain the likelihood of models including demography and selection (but see Vitalis et al., 2014). Methods that rely on simulations provide alternatives to the use of likelihood functions (Csilléry et al., 2010; Schrider and Kern, 2018). One of the first works that proposed such strategy addressed the inference of local adaptation (Bazin et al., 2010). With coalescent simulations of an island model, they estimated demographic parameters and inferred the number of loci under selection. In their simulations, selection was modelled as differential locus-specific migration rate in which a selected locus had lower migration rates than neutrally-evolved loci. However, locus-specific migration rate or effective population size (as in Roux et al., 2016; Fraïsse et al., 2020) represent approximations of the selection process. Forward-in-time (FWT) simulation allows more realistic models of selection. These were used to make inferences on N_e in the presence of selection by Sheehan and Song (2016) (for selective sweep and balancing selection) and Johri et al. (2020) (background selection). However, these works rely on simulations of few independent loci-not more than 50Kbp-which prevents the modeling of genome-wide effects of selection as the reduction of effective population size due to the variance of reproductive success of individuals (e.g. Santiago and Caballero, 1995). Nevertheless, this strategy brought new insights into the dynamics of selection. For instance, Laval et al. (2019) estimated the number of past selective sweeps that occurred in the human genome in the past 10,000, their intensity, and their age. Besides some limitations, these works exemplify the power of likelihood-free methods to the inference of the complex interaction between demography and selection.

Most population genetic studies use samples collected at one time point to infer the neutral processes (mutation, recombination, random genetic drift) and selection throughout the history of populations. Temporal data allows a better understanding of recent evolutionary processes (Feder et al., 2021; Dehasque et al., 2020) because it contain information about the allele frequency changes through time. By tracking the allele frequency changes over time, it is possible to estimate the relative role of selection and drift, consequently, temporal data has the potential to give us a better understanding of the interaction between drift and selection (see for example,

Buffalo and Coop, 2019b,a).

Here, we propose the use of ABC to jointly estimate demography and positive selection from temporal genomic data. In our framework, we use agent-based, forward-in-time simulations which allowed the modelling of the genome-wide and linked selection effects of beneficial mutations. Until recently, the use of such computationally demanding simulations for ABC inference was unrealistic since a great number of simulations are required to achieve accuracy in ABC (Frazier et al., 2018). However, with the introduction of Random Forests (ABC-RF), it was possible to reduce the computational burden as fewer simulations are required to achieve reliable estimates (Pudlo et al., 2016; Raynal et al., 2019). While many methods focus on the detection of targets of selection, our work addresses the inference of parameters that characterizes the genome-wide signal of demography and selection. Our genome-wide estimates showed to be reasonably accurate for a wide rate of adaptation and strength of selection. We were able to separate the estimates of N_{e}, which reflects the short-term drift from the population census size N. We also estimated the rate of influx of new beneficial mutations as measured by genome-wide scaled mutation rate of beneficial mutations. We first evaluated the performance of an ABC-RF approach with forward-in-time simulations, and finally, we applied this framework to the analysis of a real time-series population genomics dataset of the feral population of *Apis mellifera* (Cridland et al., 2018). Our results were consistent with the biology of the species and with events that occurred recently in the history of the analyzed populations, taking into account the limitations of our approach.

Methods

Inference model

We assumed a closed population (no migration) of N diploid individuals that evolved under a Wright-Fisher model with selection. Selection only acted on *de novo* beneficial mutations. Beneficial mutations entered the population per generation with a rate μ_{b}. The rate of *de novo* beneficial mutations that enter the population was independent of their selective strength. The scaled mutation rate of the beneficial mutations θ_{b} depends on the product of $N\mu_{\mathrm{b}}$ which can be viewed as the waiting time for the appearance of a new beneficial mutation in the population. Populations with high θ_{b} receive new beneficial mutations every generation (Karasov et al., 2010).

We divided this model in two periods: 1) the burn-in period, with a duration determined by the time necessary to contain all most recent common ancestors (MRCA) for all genomic regions and 2) the inference period, with longitudinal samples. The population census sizes of each period were N_0 (burn-in period) and N (inference period). Beneficial mutations were present since the first generation and every beneficial mutation had a selection coefficient of s higher than zero and all were co-dominant. The values of the selection coefficients s were drawn from a gamma distribution with mean γ.

Individuals were diploid and have a genome of size G in base pairs (bp) that consisted of a single linkage group with per base recombination rate per generation of c_0. Each genome was composed of delimited "neutral" and "non-neutral" regions. Non-neutral regions held both neutral and beneficial mutations. This division can be interpreted as a genomic architecture in which there are genic regions that have a combination of neutral (synonymous intron mutations) and selected (non-synonymous mutation) sites, and intergenic regions (neutral mutations) but also other characteristics of the genome that would make the action of selection heterogeneous along the genome. We chose this simplification because it is general and it was a straightforward way to define independent priors for the relative number of non-neutral to neutral regions, and for the number of beneficial mutations in non-neutral regions. The probability of beneficial mutation to arise in the simulation, or their mutation rate per generation μ_{b}, was determined

by the product of the proportion of non-neutral regions P_{R}, the proportion of beneficial mutation in a non-neutral region P_{B} and the mutation rate per generation μ. Consequently, we scaled the population scaled beneficial mutation rate θ_{b} by the genome size as $\theta_{\mathrm{b}} = 4N\mu_{\mathrm{b}}G$.

We were interested in the drift and selection dynamics during the sampling period. The first sample was taken at t_1, that was the first generation after the burn-in period ended. The second was taken at t_2, after τ generations from t_1. Individuals were sampled following the sample plan II of Nei and Tajima (1981), where individuals were taken before their reproduction and permanently removed from the population. In this way their genotypes did not contribute to the next generation genotypes. Sampled individuals were used for the calculation of the summary statistics for the ABC-RF framework. Figure 1 shows a schematic representation of the model template (and see Table S1 for a summary of the notation).

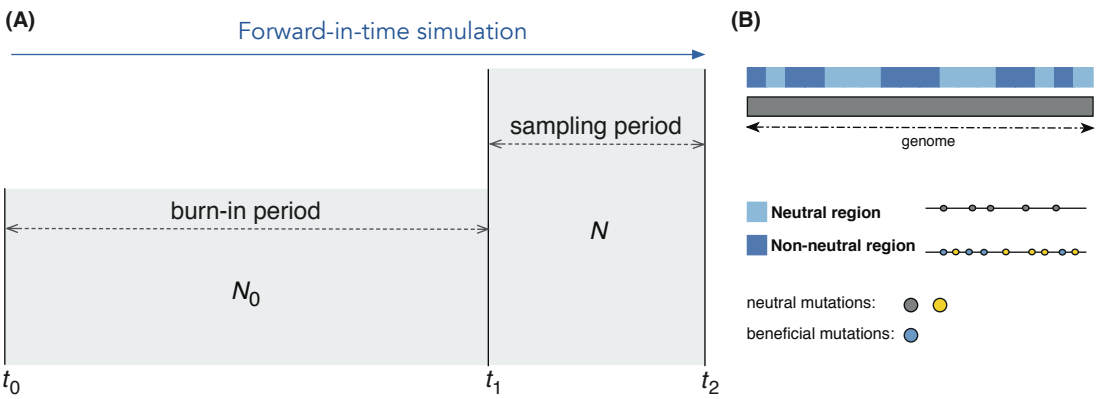

Figure 1: A schematic representation of the model used to simulate temporal population genomic data. (A) the population model that consisted of: 1) the burn-in period, where the number of generations was determined by the time necessary to contain the MRCA for all genomic regions. 2) the sampling period between the two time points, where the inference of demography and selection was made. (B) the genomic architecture model that consisted of: 1) a diploid genome of one linkage group that was divided in neutral and non-neutral regions that were composed of neutral and a combination of neutral and beneficial mutations.

Calculation of summary statistics and latent variables

For each simulation, we calculated summary statistics that: 1) compared the two samples (*e.g.* genetic differentiation F_{ST}), and 2) quantified the diversity within-sample (*e.g.* expected heterozygosity H_{E}). For the later, calculations were obtained for each sample and the pooled sample (when the genetic data of populations were combined in one population). Some summary statistics were calculated genome-wide, for example, global F_{ST}, global H_{E} and the total number of polymorphic sites S; others were calculated SNP-by-SNP as the H_{E}; or they were calculated in windows as S, the nucleotide diversity π, and Tajima's D. For every simulation, we measured the mean, variance, kurtosis, skewness, and 5% and 95% quantiles among all locus-specific or window summary statistics. These statistics informed about the heterogeneity of genome-wide distribution of locus-specific or window summary statistics. We set three window sizes for the window summary statistics: 500, 5 000, and 10 000 bp. Those windows overlapped because each window was composed around every single SNP which put the targeted variation in the middle of the window with the other SNPs in half of the window size on each side of the targeted SNP. The site-frequency spectrum was obtained as a global summary statistics with three different

numbers of discrete classes (bin sizes): 10, 15, and 20 bins (the complete list of summary statistics can be found in Supplementary Methods, section S2.1 List of summary statistics).

We defined five latent variables to measure the combined effects of selection and drift: 1) the effective population size N_e, and 2) the ratio between the effective population size N_e and the population census size N, N_e/N), 3) the number of beneficial mutations under strong selection P, 4) the average selection coefficient of strongly selected mutations $\bar{s}$, and 5) the average substitution load L. Each latent variable produced by the simulation depended on the simulated dynamics of selection and drift.

The effective population size N_e measures the increase of inbreeding at each generation. In this definition, N_e is the size of an ideal population with the same amount of drift as the population under consideration. Defined in these terms, N_e is the inbreeding effective size (Santiago and Caballero, 1995; Walsh and Lynch, 2018). It was calculated in every generation i of the sampling period as:

$$N_{\mathrm{e},i} = \frac{4N}{\sigma^2_{k_i} + 2}$$

$\sigma^2_{k_i}$ being the variance among parents of the number of gametes produced that contributed to offspring in generation i. The N_e for the whole inference period was obtained by calculating the harmonic mean of $N_{\mathrm{e},i}$. The population size of N was kept constant for the whole period as shown above, and it represents a parameter of the simulation. From the N_e we obtained the ratio N_e/N (it measures how the census size reflects the actual effective population size: we expect to have a dramatic reduction on N_e compared to N when beneficial mutations are more frequent).

We also recorded the selection coefficient of all beneficial mutations that arose in every generation i from t_1 to t_2 in each simulation. After, we calculated the fraction of beneficial mutations that were strongly selected (where $s > 1/N_\mathrm{e}$ over all mutations that were segregating in the period). This fraction represented all beneficial mutations that arose in between t_1 and t_2, regardless if they were lost or fixed at any generation of the period, or if their frequency oscillated but never reached fixation. We decided for it because any beneficial mutation that arise can impact the allele frequency trajectories of other mutations (neutral or beneficial). For these mutations, we also calculated the average across all selection coefficients. We also calculated, in every generation of this period, the substitution load L_i as the difference between the total fitness of the individual with the highest fitness $W_{\mathrm{max}i}$ and mean total fitness of the population $\bar{W}_i$ (it measures the overall diversity of beneficial mutations present in the inference period),

$$L_i = \begin{cases} 0, & \text{if } W_{\mathrm{max}i} = 0 \\ \frac{W_{\mathrm{max}i} - \bar{W}_i}{W_{\mathrm{max}i}}, & \text{otherwise} \end{cases}$$

The average substitution load was obtained by averaging all values of L_i.

Implementation

The model was simulated with the software SLiM v3.1 (Haller and Messer, 2017; Haller et al., 2018). For the calculation of the inbreeding effective size, we needed to activate an optional SLiM 3.1 behavior to track the pedigrees of each individual in the population. It allowed us to obtain the number of each parent gamete and the population variance of the number of gametes. For the calculation of the generation substitution load, we used a SLiM built-in function that allowed us to obtain the vector of fitness of all individuals in the population. The cached fitness was the sum of all fitness determined by each beneficial mutation.

Each simulation was produced by using different combinations of the model's parameters: 1) the mutation rate per bp per generation μ, 2) the per-base recombination rate per generation c_0, 3) the mean γ of a gamma distribution with the shape parameters equal to the mean, from which the selection coefficients s of each beneficial mutation in the simulation were sampled, 4) the number of non-neutral genomic regions P_{R}, 5) the parameter that determines the probability of beneficial mutation in non-neutral regions P_{B}, 6) the population census size of the burn-in period N_0, and, finally, 7) the population size of the inferential period N.

We set SLiM to output genotypic data of samples of individuals as single nucleotide polymorphisms (SNPs), at t_1 and t_2, in the VCF file format. Using bcftools (Li, 2011), custom R function (R Core Team, 2020) and EggLib (De Mita and Siol, 2012), SLiM outputs were processed and summary statistics calculated. We implemented a pipeline in an R script that automates the sampling of the prior values, runs each simulation, manipulates the VCF files, calculates the summary statistics, and organizes the final reference table. This script was also produced to facilitate the model test with few simulations and the job submission in a HPC node(s). The main R and additional scripts are available on Zenodo (Pavinato, 2021). In this pipeline, for every simulation, a row of the reference table was produced by combining the model parameters used to launch a SLiM simulation, latent variables, and the summary statistics.

ABC-RF

In this work, we propose the use of Random Forests (RF) in ABC procedure, where the parameter estimation is a machine learning problem (Pudlo et al., 2016; Raynal et al., 2019).

A reference table was generated with the model described above. In each simulation, each individual had a genome of size 100 Mb that was divided into 2,000 fragments of 50,000 bps. A number of these fragments were randomly set as either neutral or non-neutral, based on the parameter P_{R}. In the inference period, 100 individual genotypes were sampled at t_1 and t_2 after $\tau = 10$ generations. For all model parameters, values of each simulation were sampled from a log-uniform distribution with range: 1 to 2 000 for N_0 and N, 10^{-10} to 10^{-6} for μ, 5×10^{-10} to 5×10^{-7} for r, 10^{-5} to 1 for P_{B}, and 10^{-3} to 1 for γ. And uniform distribution with range 0 to 1 for P_{R} (Figure S1 shows the prior distribution for all model parameters and latent values).

The raw reference table produced by the pipeline was processed to remove missing data. Missing data were present in several summary statistics of simulations with low genetic diversity, that can be produced, for example, by low mutation rate, by small population size, by selection, or by the combination of these parameters. Missing data were also present in the entire row of a simulation if the combination of population size, mutation, and especially recombination rate produced simulations that were memory intense, which caused the simulation to crash. A final reference table containing 55,634 simulations with 405 summary statistics was used to train the ABC-RFs. Independent RFs were obtained for each parameter and latent variable using R package abcrf (Pudlo et al., 2016; Raynal et al., 2019). Each RF was obtained by growing 1 000 trees. The RF were grown with the default parameters. Average genetic load, L and P were logit transformed prior the training. For these latent variables and for $\bar{s}$, simulations with $L = 0$, $P = 0$ or $\bar{s} = 0$ were also excluded from the training set, which reduced it to 36,026 simulations for L, and with 29,264 simulations for $P = 0$ and $\bar{s}$. For the other parameters and latent variables, we performed log transformation prior training and use the reference table containing all simulations.

The performance of each trained Random Forest was evaluated with *out-of-bag* (OOB) estimates (Breiman, 2001). These estimates were produced by the trained model for the data used for training. Regression trees that compose the actual RF are grown using part of the data, selected randomly from the initial set of simulations (here, the rows of the reference table). As a consequence, for each simulation, there is a subset of trees that were grown without the data from that simulation. The estimate from that subset of trees is called OOB estimate and

with it, cross-validation of the trained model is done without splitting the reference table in the training and testing sets. We calculated the mean squared error (MSE) and the correlation coefficient (R^2) between the true and the OOB estimated values obtained with the function regAbcrf implemented in the R package abcrf.

Alternative estimates of N_e from temporal data

We compared the the ABC-RF N_e estimates with estimates obtained with the global F_{ST} between temporal genomic samples (Frachon et al., 2017). This estimator is defined as:

$$\hat{N}_e = \frac{\tau(1 - \hat{F}_{ST})}{4\hat{F}_{ST}}$$

where τ accounts for the time-interval, in generations, between the first and the last samples used to estimate the F_{ST}, and $\hat{F}_{ST}$ is the the Weir and Cockerham's F_{ST} estimator (Weir and Cockerham, 1984). The N_e from the F_{ST} was calculated for all simulations used to train the random forest. We calculated the mean squared error (MSE) and the squared correlation coefficient of linear regression (R^2) between the observed (true) and the F_{ST}-based N_e estimated values of all simulations. We also evaluated the performance of each estimator by calculating the MSE for simulations within a specific range of values of θ_b (local MSE estimates). By comparing the changes in MSE values of each estimator as a function of θ_b we could better understand how the amount of selection affected each estimator.

Analysis of temporal genomic data of feral populations of *Apis mellifera*

To give an example of an application of our framework, we analyzed a subset of whole-genome sequencing data of feral populations of *Apis mellifera* from California (Cridland et al., 2018). Eight out of fourteen sites in this work were composed of samples from museum and contemporary collections: 1) Avalon site in Catalina Island, Los Angeles county, 2) Arcata and Blue Lake sites in Humboldt county, 3) Placerita Canion Nature Area in Los Angeles county, 4) Sky Valley and Idyllwild in Riverside County, 5) La Grange, Stanislaus county, 6) Stebbins Could Canyon Reserve, Solano county and 7) UC Davis Campus, Yolo county. This dataset contains pairs spanning 104 years (as in the Avalon site, Catalina Island, Los Angeles county) and pairs spanning only 15 years (as in the Placerita Canyon Nature Area, Southern California, and Idyllwild, in Riverside county). For the temporal samples from Riverside County, we only used the two samples collected in May 1999 in Idyllwild as the first sample, and combined all samples collected in September 2014 (in Idyllwild and Sky Valley) as the second sample.

Individual vcf files of each site were combined with bcftools (Li, 2011) and a custom R script was used to convert the dataset to an input format to run EggLib custom implementation (in Pavinato, 2021). We tagged samples from the same time point with the same label. The amount of missing genotypes was quantified and a negligible amount ($< 1\%$) were present in all but two populations (populations from Avalon and Placerita had 10% of the total missing genotypes). We did not add missing genotypes in the simulations, but the missing data percentage can be specified in our pipeline. For the simulations, we set the genome size of each individual as 250 Mb (similar to the most recent estimates of *A. mellifera* genome size (Elsik et al., 2014).

The genome was divided into 5,000 fragments of 50,000 bps. These fragments were randomly set as neutral or non-neutral based on the parameter P_R. Dominance coefficients were set as 0.5 for all mutations throughout the simulation. In the sampling period, for each site, the same number of individuals found in each dataset were sampled in the first (t_1) and second time point (t_2). We used a Normal distribution for μ with a mean 3.4×10^{-9} with a standard deviation of 0.5, to have a prior distribution center around the estimated mutation rate for Hymenoptera

(Liu et al., 2017). The per base recombination rate was set as Uniform, ranging from 10^{-8} to 10^{-4}. The genome was represented with a single linkage group. For the population sizes N_0 and N were taken from a Uniform prior distribution ranging from 1 to 10,000 individuals. Other parameters were set with the same prior as described above. We calculated the same summary statistics, but this time with only one window size of 10Kbp and one bin size of 10 for the site-frequency spectrum. The raw reference table produced by the pipeline was processed to remove missing data, and a final reference table containing 162 summary statistics for each population pair, was used to train the ABC-RFs (see Table 1 for the number of simulations of each reference table). The RFs were grown as described above.

Table 1: Populations and number of simulations in the reference table.

Location	Date	N	Simulations
Avalon, Catalina Island, Los Angeles county	1910/2014	2,5	13,953
Blue Lake and Arcata, **Humboldt** county	1966/2015	6,6	14,216
Placerita Canyon Nature Area, Los Angeles county	1999/2014	5,6	14,125
Idyllwild and Sky Valey, **Riverside** county	1999/2014	2,8	13,930
La Grange, **Stanislaus** county	1976/2014	2,6	13,956
Stebbins Cold Canyon Reserve, Los Angeles county	1996/2014	5,5	14,121
UC **Davis** Campus, Yolo county	1968/2015	2,6	13,970

Names highlighted in bold letters corresponds to the population code we used in this work.

Results

ABC-RF framework for joint inference of adaptive and demographic history

The ABC-RF framework jointly estimated parameters informative about adaptive and demographic history in temporal population genomics settings. Independent random forests estimated the population scaled beneficial mutation rate θ_{b}, the population census size N, and the effective population size N_{e} (Figure 2). Trained RFs performed well in predicting N and N_{e} with small MSE and higher R^2 (Figure 2 b and c). But, the trained RF for θ_{b} had a lower performance than the trained RFs for demographic parameters, with high MSE and low R^2 (Figure 2a and b). Still, the estimates were robust for intermediate to higher values of θ_{b}. For the results of other model parameters and latent variables informative about demography and selection, see S3 Supplementary Results.

The automated selection of informative summary statistics is an important feature of ABC-RF. For each tree of a random forest, summary statistics were selected given its ability to split the data. How many times a summary statistics was selected in each RF informs us of their importance for the inference of a given parameter. For the prediction of θ_{b} values, the RF picked more frequently the 5% quantile of Tajima's calculated in the second sample, with the kurtosis and skewness of F_{ST} and D_{a} calculated globally (Figure S4 e). The population size was trained with a combination of within and between sample summary statistics: F_{ST} and D_{a}, with their respective derived statistics were frequently selected (Figure S5 c). For N_{e}, summary statistics that inform about the cumulative divergence between samples as F_{ST} and D_{a}, were frequently selected (Figure S5 d).

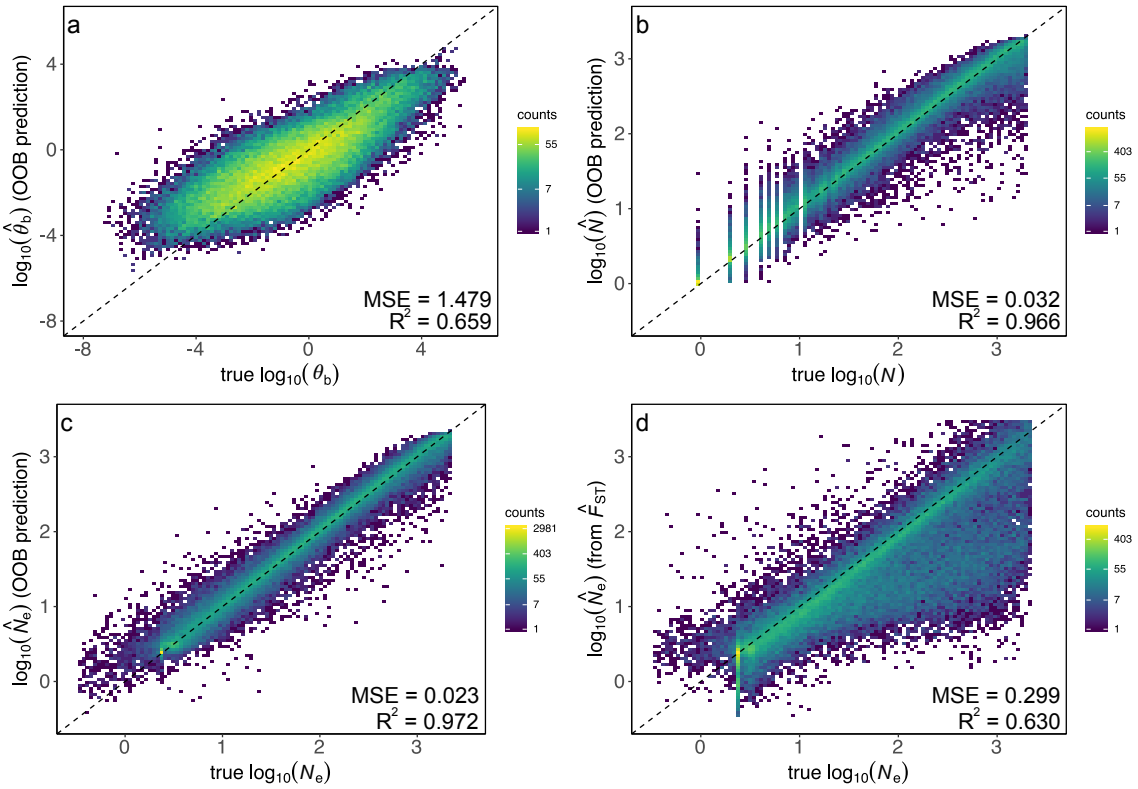

Figure 2: Out-of-bag estimates of ABC-RF trained for the joint inference of demography and selection, and N_e estimates from the temporal F_{ST} to compare with the ABC-RF -based N_e estimates. (a) population census size N; (d) effective population size N_e; (c) population scaled mutation rate of beneficial mutations θ_b; and (d) N_e from temporal F_{ST}

Comparison with available methods to estimate N_e

We compared our ABC-RF N_e estimates with estimates obtained with the temporal F_{ST} (Frachon et al., 2017). The F_{ST}-N_e showed higher MSE and lower R^2 compared to the ABC-RF -based estimates (Figure 2c and d). The impact of the amount of selection on the temporal F_{ST} estimator can be clearly seen. When the beneficial mutations were less frequent, the ABC-RF and the temporal F_{ST} performed well and similarly regardless the strength of selection, with the ABC-RF based estimator with less error than the temporal F_{ST}-based estimator. However, when the frequency of selection started to increase, the N_e estimator based on the temporal F_{ST} started to increase the local MSE (Figure 3).

Analysis of temporal genomic data of feral populations of *Apis mellifera*

For the analysis of feral *A. mellifera* populations, we first grew independent RF for each parameter in each population. All populations had the same performance of the ABC-RF estimator for N_e, as they showed similar values of MSE and R^2 (Figure S8). For N, trained RF for Humboldt, Stebbins and Placerita performed similarly well, and they had the lowest MSE and higher R^2 (Figure S7). For θ_b, Riverside had trained RF with the worst performance (Figure S6). Overall, both MSE and R^2 obtained with OOB estimates from simulated data for *Apis mellifera* dataset were comparable to these parameters obtained with OOB estimates for the simulated data for the proof of the method.

Trained RF for N and N_e were able to predict these parameters in all populations, as the inference of the mean posterior value and the posterior distribution differentiated from the mean prior value and distribution (Figure 4 b and c). For N, posterior distribution were wider than for

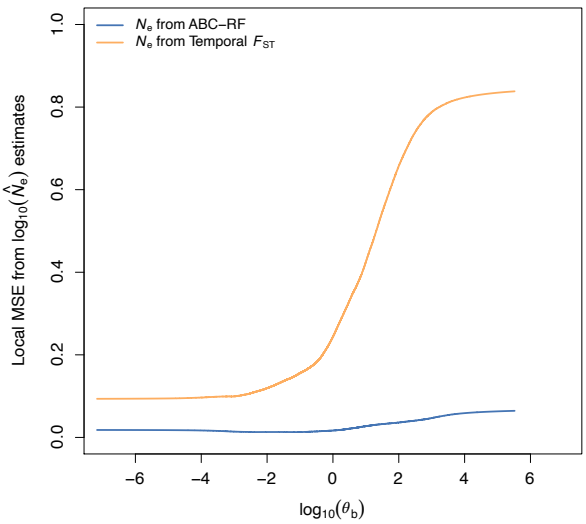

Figure 3: Local MSE of N_e estimates as a function of θ_b. The lines corresponds to the MSE on N_e estimates from ABC-RF and from temporal F_{ST}.

N_e. Trained RF for θ_b, for all populations had a similar posterior mean, except for the Avalon population that had a peak at a lower value (Figure 4 a). But, the posterior distributions were wider and they followed the prior distribution, making it difficult to accurately predict the posterior mean and variance in all populations. It is possible to see together with the posterior mean estimates that the ABC-RF estimates for θ_b were concentrated in lower values (Table S2) in all populations. N_e were also lower, and N_e and N were similar. For the results of OOB estimates of other model parameters and latent variables, and for posterior estimates for these parameters see S3 Supplementary Results.

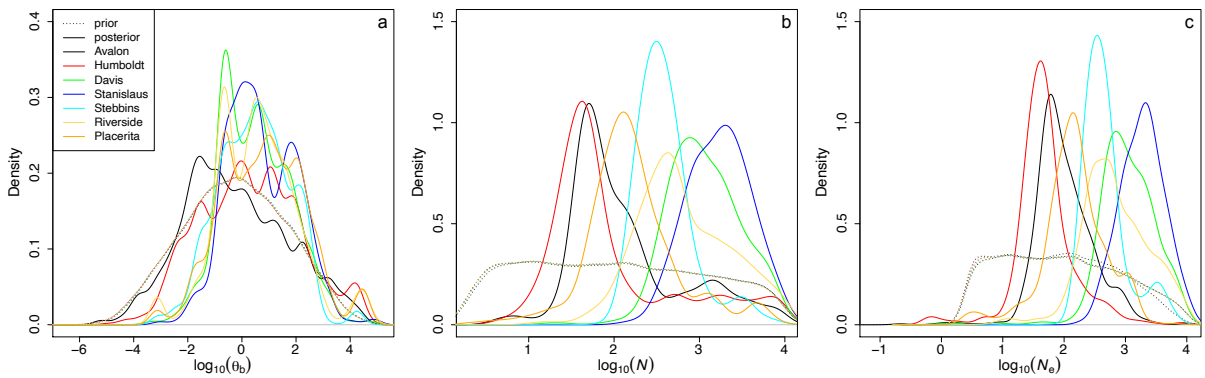

Figure 4: Joint inference of demography and selection for feral *A. mellifera* populations. (a) the population census size N, (b) the effective population size N_e (c) the scale mutation rate of selected mutations θ_b. Dashed and filled lines corresponds to the prior and posterior distributions, respectively

Discussion

Separating demography from drift, and the inference of θ_b

With temporal population genomics data, we can see the evolution in "action" as opposed to single time-point population genomics data (Feder et al., 2021). Consequently, temporal data

have more information about the ongoing process, which make it better for the understanding of the complex interaction between demography and selection (Buffalo and Coop, 2019b; Dehasque et al., 2020; Williams and Pennings, 2020). When samples from more than two time points are available, the footprints of drift and selection in allele frequencies changes can be separated (e.g., Buffalo and Coop, 2019a; Feder et al., 2014). Our results showed that two samples collected in different time points were sufficient for the inference of the genome-wide footprint of adaptive evolution and to separate the demography (population census size N) from drift (effective population size N_{e}).

It is important to stress that N_{e}, as a latent variable, captures the feedback dynamics between drift and linked selection. Selection, either positive or negative, causes a deviation of N_{e} from N. The impact of selection on the genome can extend far from the target of selection since individuals that carry beneficial mutations have more chance to reproduce, their beneficial mutations are more likely to be in the next-generation offspring (Walsh and Lynch, 2018). In this complex dynamic, with many loci under selection which creates a dynamic that cannot be easily described, latent variables actually summarized the by product of drift and selection interactions. With our approach, $\hat{N}_{\mathrm{e}}$ quantifies the drift due both to demographic and selection dynamics.

This genome-wide reduction in N_{e} is not captured when loci are assumed to evolve independently (as in Sheehan and Song, 2016, for example). In contrast, the complexity of linked selection and the genome-wide effect of selection are taken into account by using individual based simulations with whole genome in an ABC approach.

Estimates of genetic load or other genome-wide parameters about selection are obtained when annotated genomic data is available, (Henn et al., 2015) or by conducting experiments on crossing populations (for the genetic load Plough (2016)). However, we were able to obtain estimates of selection parameters only with polymorphism data. A similar approach was used by Laval et al. (2019) to estimate the number of selective sweeps affecting the human genome in the past 10 000 years. But they did not consider the feedback dynamics of selection and demography. Differently, Buffalo and Coop (2019a) measured the genome-wide signature of selection by estimating the covariance of allele frequencies at consecutive time points. This allowed the quantification of the genome-wide contribution of selection to the observed allele frequency changes, even when selection involved many loci of small effect. In this work, we estimated the population scale mutation rate of beneficial mutations θ_{b}, which informs about the diversity of beneficial mutations that existed in the population between the two time points, and the potential speed of adaptation at the genome level (Hermisson and Pennings, 2017). These estimates reflect the potential number of beneficial mutations present between the two time points regardless their impact as determined by their selection coefficients.

The variable importance plot of each parameter shows us the global importance of each summary statistics in the trained Random Forests. For N_{e}, N, and θ_{b} summary statistics calculated from the distribution of locus-specific summary statistics -skewness, kurtosis, mean, variance, 5% and 95% quantiles were more frequently used. Summary statistics derived from the distribution of locus-specific calculated from all segregating loci in the genome inform about the heterogeneity that selection and drift produce genome-wide. For example, a *de novo* a beneficial mutation entered the simulation and was selected, it left a signal of lower diversity around the region it was located. The genome, after selection, contained spots where diversity was high and where it was low, and this heterogeneity was captured by the distribution of locus-specific H_{E}, more specifically the lower tail of the distribution where the values of diversity were lower. The use of the covariance matrix of allele frequencies through Buffalo and Coop (2019a) can be considered as a summary statistic for the the genome-wide heterogeneity that selection and drift left on the genome. It would be interesting to include this matrix as summary statistics for further development of the method.

Comparison with available methods to estimate N_e

We compared the N_e obtained with ABC-RF framework to the N_e obtained with F_{ST} estimator (Skoglund et al., 2014; Frachon et al., 2017). Overall, the F_{ST} -based N_e estimator performed poorly compared to the ABC-RF -based estimator. The lower performance were caused by N_e values that were underestimated when beneficial mutations were more frequent (higher θ_b). Consequently, the N_e estimates from the temporal F_{ST} were strongly affected by selection. When selection was infrequent or rare, both estimators performed similarly, but with the ABC-RF estimator with lower MSE. Positive selection can increase the variance of allele frequency between samples taken in different time points. When selection is infrequent or rare, drift determines most of allele frequency changes between samples; but when selection is pervasive, selection dominates, which cause dramatic and rapid changes in allele frequency, increasing the variance between samples. N_e estimator based on the F_{ST} depends on the difference between samples as function of how the variance of the allele frequencies differ between samples, it is therefore naturally biased by strong and frequent selection. We can assume that the N_e estimator from ABC-RF was insensitive to the amount of selection. It is because we trained the ABC-RF with N_e values from the simulation; in our simulations, N_e was a latent variable that captured the deviation that selection imposed on the number of individuals able to reproduce (selected for); it was not a summary statistic that might be biased by unaccounted factors.

The amount of selection for $\theta_b \sim O(1)$ could be unrealistic in many organisms, but plausible in virus (Feder et al., 2014) and many Arthropod species, with large N_e, which have larger population sizes (except in eusocial insects that have vertebrate-like population sizes Romiguier et al. (2014)). In larger populations sizes selection acts also on weaker and milder beneficial mutations. In those organisms, it might be unreasonable to assume mutation-drift equilibrium given the pervasive role of selection. Consequently, any attempt to estimate demography parameters as N_e without properly accounting for the pervasive role of selection could be biased.

Analysis of temporal genomic data of feral populations of *Apis mellifera*

Overall, the performances of the ABC-RF estimators for selection and demography were similar for all populations. For θ_b, Avalon and Humboldt's population had the posterior distribution tilted toward negative values. It might indicate that θ_b was impacted by the time interval between the first and the second sample. For the Humboldt population, it might also reflect the low genetic diversity of this population, either because of a bottleneck that happened especially in Northern California populations with the arrival of *Varroa*, or because their adaptation to this mite species. *Varroa* is a mite group known for causing serious infestations in bee colonies, to the point to wipe out entire colonies. The arrival of *Varroa* in California caused a severe population reduction, especially in northern populations. We did not take into account the effects of admixture in our model, which might play an important role in most of populations. Admixture can creates heterogeneity in the genome. But, because our model did not include admixture, the higher θ_b values estimated by our model, might actually reflect the genomic heterogeneity's determined by admixture, not by selection.

Estimates of the effective population size, N_e were in the same order of magnitude of other estimates for hymenopterans (Zayed, 2004) for all populations analyzed by the pipeline. Smaller posterior values observed in Humboldt might also reflect the impact of *Varroa* infestation. However, it might also indicate an admixture that occurred with domesticated lineages as a consequence of changes in beekeeping practices. Northern populations, especially from Humboldt County, shared great similarities with bees from domesticated colonies. They also presented ancestry that reflected changes from European M to C lineages. These lineages are less genetically diverse that the African lineage (A). Beekeeping practices might also cause the reduction in N_e, consequently reducing θ_b as pointed above. Southern populations, as pointed by Cridland et al. (2018) showed a higher level of admixture with African lineages. This increased the

genetic diversity of these populations, however, populations from the southern most cites (from Riverside County, Placerita and Avalon, Los Angeles county) did not show the highest values of N_e. The population of Stanislaus County had the highest value of N_e, possibly because it had lower levels of admixture with domesticated lineages.

Both N_e and N had similar estimates. For this, it is important to note that we did not model the specificity of *A. mellifera* reproductive biology. We were aware that our estimate of N_e is overestimated (Nomura and Takahashi, 2012). The use of a model that simulates the biology of the species would possibly give us a different estimate of population size N. N_e similar to N might also reflect some events in the past of these populations, as pointed out by Cridland et al. (2018) and mentioned above: changes in beekeeping practices, the introduction of Africanized bees, the introduction of *Varroa*, and the potential local adaptation of some populations to *Varroa*. The indication that gene flow might be an important force that determined the changes in genetic diversity through time, may be important to also include it in the model. The lack of gene flow in the model is clearly inappropriate for populations that show introgression with African bee lineages during the period considered, which could explain the high mutation rate estimates. Introgressions would also bias some of the results, both for demography and selection estimates.

Our ABC-RF framework also estimated the per site mutation rate per generation μ (Supplementary Results, S18). For all populations, the mean posterior μ exceed the mean μ of the species. The higher estimated values we obtained might reflect the recent admixture events between these populations that were not included in our framework. It is clear that for these dataset, admixture is an important aspect of the species demography that should be modeled to properly separate the effects of selection and drift.

Meanwhile, taking together the population size and selection parameters estimates, it is possible to conclude that in this system adaptation is limited by mutations. The evolution of the reproductive mode that concentrated reproduction on a few individuals came with a cost of reducing N_e. Social insect bees are the few insect groups that show very small N_e (Romiguier et al., 2014). As a consequence, the rate of adaptive evolution θ_b reflects a system limited by an influx of adaptive mutation (Rousselle et al., 2020).

Perspectives and Limitations

Our model is very simplistic, as it only considered the impact of beneficial mutations, neglecting the effect of background selection and standing variation. Background selection can mimic directional selection because they cause a similar pattern of diversity reduction around the target of selection. However, it was recently shown that background selection only mimics the classical sweep in simplistic models, where the deleterious mutation is localized in a specific region of the genome (Schrider, 2019). For more realistic scenarios, where the concentration of deleterious mutations varies across the genome, background selection does not behave as a classical hard sweep. In an attempt to jointly accommodate the effect of demography and selection on the inference of N_e, Johri et al. (2020) modeled the effect of background selection and developed an ABC-based approach that jointly estimated the distribution of fitness effects (DFE) and N_e. In their simulations, deleterious mutations were modeled more realistically. They also included some beneficial mutations, similar to a classical hard sweep. They showed an unbiased estimate of N_e regardless of the presence of positive and negative selection. Future developments should include a more realistic genomic architecture where both negative and positive mutations can co-occur and explore different concentrations of deleterious mutations. In addition to that, further developments should explore not only scenarios of *de novo* mutations, but selection acting on standing variation. This can be easily achieved with our pipeline and allows for a more general treatment of the selection of soft sweeps. The model can also be expanded to more complex demographic scenarios, including changes in population size and genetic exchange with external

sources (migration). Including such admixtures will be key in the future development of this approach since it is also a source of heterogeneity in the genome and, thus, might influence the performance of the method.

Conclusion

We show that an ABC-RF -based approach is able to jointly infer adaptive and demographic history from temporal population genomics data. This approach allows the quantification of the genome-wide footprint of selection expressed in the scaled mutation rate of beneficial mutations. The ABC-RF N_e is robust to varying degrees of strength of selection and frequency of beneficial mutations. Our ABC-RF -based approach can be applied to temporal population genomics datasets to gain insight about the adaptive and demographic history of natural populations.

Acknowledgements

This project has received funding from the LabEx AGRO (convention ANR-10-LABX-0001-01), CEMEB (convention ANR-10-LABX-0004) and NUMEV (convention ANR-10-LABX-20) through the AAP Inter-LabEx (ABCSelection). This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 791695 (TimeAdapt). S. De Mita was funded by INRAE (*Projet Innovant EFPA*). We are thankful for the Genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul, https://doi.org/10.15454/1.5572369328961167E12), for the High-Performance Computing Center University of Montpellier (HPC-LR) and for the Ohio Super Computer Center (OSC Center (1987)) for providing computing resources.

References

Bank, C., Ewing, G. B., Ferrer-Admettla, A., Foll, M., and Jensen, J. D. (2014). Thinking too positive? Revisiting current methods of population genetic selection inference. *Trends in Genetics*, 30(12):540–546.

Bazin, E., Dawson, K. J., and Beaumont, M. A. (2010). Likelihood-free inference of population structure and local adaptation in a Bayesian hierarchical model. *Genetics*, 185(2):587–602.

Beichman, A. C., Huerta-Sanchez, E., and Lohmueller, K. E. (2018). Using Genomic Data to Infer Historic Population Dynamics of Nonmodel Organisms. *Annual Review of Ecology, Evolution, and Systematics*, 49(1):433–456.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1):5–32.

Buffalo, V. and Coop, G. (2019a). Estimating the genome-wide contribution of selection to temporal allele frequency change. *bioRxiv*, 44(3):285–32.

Buffalo, V. and Coop, G. (2019b). The Linked Selection Signature of Rapid Adaptation in Temporal Genomic Data. *Genetics*, 213(3):1007–1045.

Casillas, S. and Barbadilla, A. (2017). Molecular Population Genetics. *Genetics*, 205(3):1003–1035.

Center, O. S. (1987). Ohio supercomputer center.

Cridland, J. M., Ramirez, S. R., Dean, C. A., Sciligo, A., and Tsutsui, N. D. (2018). Genome Sequencing of Museum Specimens Reveals Rapid Changes in the Genetic Composition of Honey Bees in California. *Genome Biology and Evolution*, 10(2):458–472.

Csilléry, K., Blum, M. G. B., Gaggiotti, O. E., and Francois, O. (2010). Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology and Evolution*, 25(7):410–418.

De Mita, S. and Siol, M. (2012). EggLib: processing, analysis and simulation tools for population genetics and genomics. *BMC Genetics*, 13:27.

Dehasque, M., Ávila-Arcos, M. C., Díez-del Molino, D., Fumagalli, M., Guschanski, K., Lorenzen, E. D., Malaspinas, A.-S., Marques-Bonet, T., Martin, M. D., Murray, G. G. R., Papadopulos, A. S. T., Therkildsen, N. O., Wegmann, D., Dalén, L., and Foote, A. D. (2020). Inference of natural selection from ancient DNA. *Evolution Letters*, 29:753–15.

Elsik, C. G., Worley, K. C., Bennett, A. K., Beye, M., Camara, F., Childers, C. P., de Graaf, D. C., Debyser, G., Deng, J., Devreese, B., Elhaik, E., Evans, J. D., Foster, L. J., Graur, D., Guigo, R., HGSC production teams, Hoff, K. J., Holder, M. E., Hudson, M. E., Hunt, G. J., Jiang, H., Joshi, V., Khetani, R. S., Kosarev, P., Kovar, C. L., Ma, J., Maleszka, R., Moritz, R. F. A., Munoz-Torres, M. C., Murphy, T. D., Muzny, D. M., Newsham, I. F., Reese, J. T., Robertson, H. M., Robinson, G. E., Rueppell, O., Solovyev, V., Stanke, M., Stolle, E., Tsuruda, J. M., Vaerenbergh, M. V., Waterhouse, R. M., Weaver, D. B., Whitfield, C. W., Wu, Y., Zdobnov, E. M., Zhang, L., Zhu, D., Gibbs, R. A., and Honey Bee Genome Sequencing Consortium (2014). Finding the missing honey bee genes: lessons learned from a genome upgrade. *BMC Genomics*, 15:86.

Fay, J. C. and Wu, C.-I. (2000). Hitchhiking Under Positive Darwinian Selection. *Genetics*, 155(3):1405–1413.

Feder, A. F., Kryazhimskiy, S., and Plotkin, J. B. (2014). Identifying signatures of selection in genetic time series. *Genetics*, 196(2):509–522.

Feder, A. F., Pennings, P. S., and Petrov, D. A. (2021). The clarifying role of time series data in the population genetics of HIV. *PLoS Genetics*, 17(1):e1009050.

Frachon, L., Libourel, C., Villoutreix, R., Carrère, S., Glorieux, C., Huard-Chauveau, C., Navascués, M., Gay, L., Vitalis, R., Baron, E., Amsellem, L., Bouchez, O., Vidal, M., LE CORRE, V., Roby, D., Bergelson, J., and Roux, F. (2017). Intermediate degrees of synergistic pleiotropy drive adaptive evolution in ecological time. *Nature Ecology & Evolution*, 1(10):1551–1561.

Fraïsse, C., Popovic, I., Mazoyer, C., Romiguier, J., LOIRE, E., Simon, A., Galtier, N., Duret, L., Bierne, N., Vekemans, X., and Roux, C. (2020). DILS : Demographic Inferences with Linked Selection by using ABC. *bioRxiv*, 162(4):2025–19.

Frazier, D. T., Martin, G. M., Robert, C. P., and Rousseau, J. (2018). Asymptotic properties of approximate Bayesian computation. *Biometrika*, 105(3):593–607.

Haller, B. C., Galloway, J., Kelleher, J., Messer, P. W., and Ralph, P. L. (2018). Tree-sequence recording in SLiM opens new horizons for forward-time simulation of whole genomes. *bioRxiv*.

Haller, B. C. and Messer, P. W. (2017). SLiM 2: Flexible, Interactive Forward Genetic Simulations. *Molecular Biology and Evolution*, 34(1):230–240.

Henn, B. M., Botigué, L. R., Bustamante, C. D., Clark, A. G., and Gravel, S. (2015). Estimating the mutation load in human genomes. *Nature Reviews Genetics*, 16(6):333–343.

Hermisson, J. and Pennings, P. S. (2017). Soft sweeps and beyond: understanding the patterns and probabilities of selection footprints under rapid adaptation. *Methods in Ecology and Evolution*, 8(6):700–716.

Jensen, J. D., Kim, Y., DuMont, V. B., Aquadro, C. F., and Bustamante, C. D. (2005). Distinguishing between selective sweeps and demography using DNA polymorphism data. *Genetics*, 170(3):1401–1410.

Jensen, J. D., Thornton, K. R., Bustamante, C. D., and Aquadro, C. F. (2007). On the utility of linkage disequilibrium as a statistic for identifying targets of positive selection in nonequilibrium populations. *Genetics*, 176(4):2371–2379.

Johri, P., Charlesworth, B., and Jensen, J. D. (2020). Toward an Evolutionarily Appropriate Null Model: Jointly Inferring Demography and Purifying Selection. *Genetics*, 215(1):173–192.

Karasov, T., Messer, P. W., and Petrov, D. A. (2010). Evidence that Adaptation in Drosophila Is Not Limited by Mutation at Single Sites. *PLoS Genetics*, 6(6):e1000924.

Kim, Y. and Nielsen, R. (2004). Linkage disequilibrium as a signature of selective sweeps. *Genetics*, 167(3):1513–1524.

Lange, J. D. and Pool, J. E. (2018). Impacts of Recurrent Hitchhiking on Divergence and Demographic Inference in Drosophila. *Genome Biology and Evolution*.

Laval, G., Patin, E., Boutillier, P., and Quintana-Murci, L. (2019). A genome-wide Approximate Bayesian Computation approach suggests only limited numbers of soft sweeps in humans over the last 100,000 years. *bioRxiv*, 5:2019.12.22.886234.

Li, H. (2011). Improving SNP discovery by base alignment quality. *Bioinformatics*, 27(8):1157–1158.

Li, J., Li, H., Jakobsson, M., Li, S., Sjödin, P., and Lascoux, M. (2012). Joint analysis of demography and selection in population genetics: where do we stand and where could we go? *Molecular Ecology*, 21(1):28–44.

Lin, K., Li, H., Schlötterer, C., and Futschik, A. (2011). Distinguishing Positive Selection From Neutral Evolution: Boosting the Performance of Summary Statistics. *Genetics*, 187(1):229–244.

Liu, H., Jia, Y., Sun, X., Tian, D., Hurst, L. D., and Yang, S. (2017). Direct Determination of the Mutation Rate in the Bumblebee Reveals Evidence for Weak Recombination-Associated Mutation and an Approximate Rate Constancy in Insects. *Molecular Biology and Evolution*, 34(1):119–130.

Neher, R. A. (2013). Genetic Draft, Selective Interference, and Population Genetics of Rapid Adaptation. *Annual Review of Ecology, Evolution, and Systematics*, 44(1):195–215.

Nei, M. and Tajima, F. (1981). Genetic drift and estimation of effective population size. *Genetics*, 98(3):625–640.

Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review of Genetics*, 39(1):197–218.

Nomura, T. and Takahashi, J. (2012). Effective population size in eusocial Hymenoptera with worker-produced males. *Heredity*, 109(5):261–268.

Pavinato, V. (2021). Tracking-selection. doi:10.5281/zenodo.4599736.

Plough, L. V. (2016). Genetic load in marine animals: a review. *Current Zoology*, 62(6):567–579.

Pool, J. E., Hellmann, I., Jensen, J. D., and Nielsen, R. (2010). Population genetic inference from genomic sequence variation. *Genome Research*, 20(3):291–300.

Pudlo, P., Marin, J.-M., Estoup, A., Cornuet, J. M., Gautier, M., and Robert, C. P. (2016). Reliable ABC model choice via random forests. *Bioinformatics*, 32(6):859–866.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Raynal, L., Marin, J.-M., Pudlo, P., Ribatet, M., Robert, C. P., and Estoup, A. (2019). ABC random forests for Bayesian parameter inference. *Bioinformatics*, 35(10):1720–1728.

Romiguier, J., Lourenco, J., Gayral, P., Faivre, N., Weinert, L. A., Ravel, S., Ballenghien, M., Cahais, V., Bernard, A., Loire, E., Keller, L., and Galtier, N. (2014). Population genomics of eusocial insects: the costs of a vertebrate-like effective population size. *Journal of Evolutionary Biology*, 27(3):593–603.

Rousselle, M., Simion, P., Tilak, M.-K., Figuet, E., Nabholz, B., and Galtier, N. (2020). Is adaptation limited by mutation? A timescale-dependent effect of genetic diversity on the adaptive substitution rate in animals. *PLoS Genetics*, 16(4):e1008668.

Roux, C., Fraïsse, C., Romiguier, J., Anciaux, Y., Galtier, N., and Bierne, N. (2016). Shedding Light on the Grey Zone of Speciation along a Continuum of Genomic Divergence. *PLOS Biol*, 14(12):e2000234–22.

Santiago, E. and Caballero, A. (1995). Effective size of populations under selection. *Genetics*, 139(2):1013–1030.

Schrider, D. R. (2019). Background selection does not mimic the patterns of genetic diversity produced by selective sweeps. *BioRxiv*.

Schrider, D. R. and Kern, A. D. (2018). Supervised Machine Learning for Population Genetics: A New Paradigm. *Trends in Genetics*, 34(4):301–312.

Schrider, D. R., Shanku, A. G., and Kern, A. D. (2016). Effects of Linked Selective Sweeps on Demographic Inference and Model Selection. *Genetics*, 204(3):1207–1223.

Sella, G., Petrov, D. A., Przeworski, M., and Andolfatto, P. (2009). Pervasive natural selection in the Drosophila genome? *PLoS Genetics*, 5(6):e1000495.

Sheehan, S. and Song, Y. S. (2016). Deep Learning for Population Genetic Inference. *PLoS Computational Biology*, 12(3):e1004845.

Skoglund, P., Sjödin, P., Skoglund, T., Lascoux, M., and Jakobsson, M. (2014). Investigating Population History Using Temporal Genetic Differentiation. *Molecular Biology and Evolution*, 31(9):2516–2527.

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

Vitalis, R., Gautier, M., Dawson, K. J., and Beaumont, M. A. (2014). Detecting and measuring selection from gene frequency data. *Genetics*, 196(3):799–817.

Walsh, B. and Lynch, M. (2018). *Evolution and Selection of Quantitative Traits*. Oxford University Press.

Weir, B. S. and Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, 38(6):1358.

Williams, K.-A. and Pennings, P. (2020). Drug Resistance Evolution in HIV in the Late 1990s: Hard Sweeps, Soft Sweeps, Clonal Interference and the Accumulation of Drug Resistance Mutations. *G3: Genes/Genomes/Genetics*, 10(4):1213–1223.

Zayed, A. (2004). Effective population size in Hymenoptera with complementary sex determination. *Heredity*, 93(6):627–630.