Inference of Selection from Genetic Time Series Using Various Parametric Approximations to the Wright-Fisher Model

Cyriel Paris, Bertrand Servin, and Simon Boitard[1]
Génétique, Physiologie Systèmes d'Elevage, INRA, Centre Occitanie-Toulouse, 31326 Castanet-Tolosan, France
ORCID IDs: 0000-0001-5141-0913 (B.S.); 0000-0002-2499-5052 (S.B.)

ABSTRACT Detecting genomic regions under selection is an important objective of population genetics. Typical analyses for this goal are based on exploiting genetic diversity patterns in present time data but rapid advances in DNA sequencing have increased the availability of time series genomic data. A common approach to analyze such data is to model the temporal evolution of an allele frequency as a Markov chain. Based on this principle, several methods have been proposed to infer selection intensity. One of their differences lies in how they model the transition probabilities of the Markov chain. Using the Wright-Fisher model is a natural choice but its computational cost is prohibitive for large population sizes so approximations to this model based on parametric distributions have been proposed. Here, we compared the performance of some of these approximations with respect to their power to detect selection and their estimation of the selection coefficient. We developed a new generic Hidden Markov Model likelihood calculator and applied it on genetic time series simulated under various evolutionary scenarios. The Beta with spikes approximation, which combines discrete fixation probabilities with a continuous Beta distribution, was found to perform consistently better than the others. This distribution provides an almost perfect fit to the Wright-Fisher model in terms of selection inference, for a computational cost that does not increase with population size. We further evaluated this model for population sizes not accessible to the Wright-Fisher model and illustrated its performance on a dataset of two divergently selected chicken populations.

Detecting the molecular basis of adaptation in natural species is one of the major questions in population genetics. Most studies in this context rely on contemporary genomic data and exploit several typical signatures left by the spread of a benefical allele in a population: decreased heterozygosity, extended linkage disequilibrium (LD), increased differentiation with other populations, see Vitti *et al.* (2013) for a review.

Genomic time series provide an even more direct access to past allele frequency changes, allowing a more precise estimation on the onset or the intensity of selection. With the combined advances in sequencing technologies, ancient DNA processing and cryoconservation, this type of data are becoming accessible to many researchers and are found in various contexts. On a long time scale, time series of ancient DNA have been used for example to detect the genetic basis of adaptation in humans from Eurasia (Mathieson *et al.*, 2015) or to infer allele frequency trajectories at several causal variants or quantitative trait loci in domestic horses (Fages Antoine *et al.*, 2019). On a shorter time scale, genetic time series are provided by the sequencing of samples conserved in biobanks during decades. This approach allowed for instance to assess the recent evolution of genetic diversity in Holstein cattle and to predict the selective value that could be obtained by using semen from 'historical' bulls in this breed (Doekes *et al.*, 2018). Controlled evolution experiments also provide a lot of interesting temporal genetic data and concern a large set of organisms. Such experiments generally imply intentional selection for a given trait of interest, *e.g.*, thermal tolereance

in Drosophila (Tobler *et al.*, 2015) or intra-muscular ultimate pH in chicken (Bihan-Duval *et al.*, 2018).

Following the increased production of genetic time series, several methods allowing researchers to exploit such data for selection inference have been developed in the two last decades (Malaspinas 2015). Most of them infer selection at a given variant based on temporal observations at this variant only, *i.e.*, without exploiting LD information (but see (Terhorst *et al.*, 2015; He *et al.*, 2019)). In this context, several statistical approaches have been proposed to model the link between the full temporal trajectory of a population allele frequency, which is informative about selection intensity at the locus, and the (partial) genetic samples observed at some specific dates along this trajectory. These approaches include approximate Bayesian computation (ABC) (Foll *et al.*, 2015; Sackman *et al.*, 2019), Bayesian path augmentation (Schraiber *et al.*, 2016) or hidden Markov models (HMM) (Bollback *et al.*, 2008).

The HMM approach arises quite naturally when analyzing genetic time series, as it accounts both for the Markovian property of allele frequency trajectories over time and for the fact that these trajectories are latent variables. Besides, efficient algorithms have been developed for likelihood evaluation and parameter estimation in such models (Rabiner 1989). Following this approach, one important choice concerns the modeling of the underlying allele frequency trajectory at a locus. In the initial work of Bollback *et al.* (2008), this process was modeled by the Wright-Fisher (WF) diffusion and the partial differential equation (PDE) resulting from this diffusion between each pair of consecutive sampling dates was solved using a numerical scheme. More recently, Song and Steinrücken (2012) derived an analytical solution of this PDE based on an infinite series expansion and integrated this solution into the HMM framework to detect selection on ancient DNA (Steinrücken *et al.*, 2014). The diffusion process can also be approximated using Markovian continuous time processes with a relatively low number of discrete spaces (Malaspinas *et al.*, 2012; Ferrer-Admetlla *et al.*, 2016). All these approaches are quite computationally demanding, moreover the WF process converges to the diffusion only for weak selection (Ewens 2004).

Considering the WF model as a reference, an even more natural approach is to compute transition probabilities of the Markov chain using this model directly (Iranmehr *et al.*, 2017; Hubert *et al.*, 2018). However, this implies that the dimension of the transition matrix is equal to the population size so this approach is in practice only possible for small populations. A more efficient strategy is to approximate the WF process by a parametric continuous distribution, whose first two moments fit those of the WF and whose transitions can be computed using a simple analytical formula. This moment-based approximation was implemented by Lacerda and Seoighe (2014) and Terhorst *et al.* (2015) using a Gaussian distribution and by Gompert (2016) using a Beta distribution. However, Tataru *et al.* (2017) showed that these two distributions accurately approximate the WF only for very short time scale and weak selection and that a much closer fit can be obtained by incorporating fixation probabilities in the Beta distribtuion, leading to what they called the Beta with spikes distribution. They illustrated the performance of this distribution to infer branch lengths in a population tree, but did not consider the potential application for selection inference.

Here we compared the performance of these different approximations for inferring a constant selection parameter in a population of constant size. This choice aims to establish a likelihood calculation framework that could be generalized to a more realistic scenario (*e.g.*, time varying selection and population size). To compare performances, we implemented a generic HMM that allows to estimate the selection intensity by maximum likelihood and to detect selected loci by a likelihood-ratio test. Transition probabilities in this model can be computed using either the WF model, the Gaussian approximation, the Nicholson Gaussian approximation, the Beta approximation or the Beta with spikes approximation. We evaluated the quality of inference provided by these different transitions in a large range of selection scenarios and showed an application of the HMM approach to a dataset concerning the experimental divergent selection of two chicken lines (Bihan-Duval *et al.*, 2018).

METHODS

Statistical models for genetic time-series

Hidden Markov Model framework: Let A be a locus of interest accepting two alleles: A_0 is called the (arbitrary) *reference allele* while A_1 is called the *alternative allele*. Considering a diploid organism, we assume each genotype is associated to a particular fitness: the homozygote reference allele fitness is 1, the heterozygote form fitness is $1 + sh$ and the homozygote alternative allele fitness is $1 + s$, where s is the selection parameter and h the dominance parameter. Assume a population has been sampled at n different dates $t_1 = 0, \ldots, t_n = T$ (in generations after the first sampling date) in order to assess the evolution of the frequency of allele A_1. We note $X_{(t)}$ the true frequency of A_1 at date t. At date t_k ($k \in \{1, \ldots, n\}$), we note $X_k = X_{(t_k)}$, n_k the total number of alleles sampled and Y_k the number of A_1 alleles sampled. Finally, we note N_e the haploid effective size of the population.

In this study we will follow a common approach to model such data, first proposed by Bollback *et al.* (2008) and use a Hidden Markov Model (HMM, see Cappé *et al.* (2005) for a general description). This HMM (Figure 1) is a bivariate Markov process $(X_k, Y_k)_{k \geq 1}$ with $X_k \in [0, 1]$ and $Y_k \in \{0, \ldots, n_k\}$. (X_k) is a Markov process with *transition kernel* $\mathbf{Q_k}$ between X_{k-1} and X_k. Given (X_k), observations are independant and the *emission kernel* of Y_k given X_k is noted g_k. The distribution of X_1 is noted ν. This model is fully specified by setting ν, Q_k and g_k, and each of them can be thought of as a probability:

$$\nu(dx) = \mathbb{P}(X_1 \in dx)$$

$$g_k(x, y) = \mathbb{P}(Y_k = y | X_k = x)$$

$$Q_k^{s,N,h}(x, dx') = \mathbb{P}(X_k \in dx' | X_{k-1} = x)$$

In this study we assume that the distribution of X_1 is uniform on $[0, 1]$ ($\nu(dx) = dx$) (but see Discussion). As the observations Y_k are sampled from a population with allele frequency X_k, the distribution of Y_k given X_k is a binomial distribution with parameters n_k and X_k:

$$Y_k | X_k \sim \mathscr{B}(n_k, X_k)$$

$$g_k(x, y) = \binom{n_k}{y} x^y (1-x)^{n_k - y}$$

For a given x, $Q_k(x, .)$ is the distribution of X_k given that the previous frequency X_{k-1} is x. Note that the Markov chain X_k is typically not homogeneous as for example the duration between successive sampling dates may vary. Apart from the time between samples, Q_k also depends on the transition model and on parameters s, N_e and h. One of the main objective of this study is to compare different possible transition models which are described below.
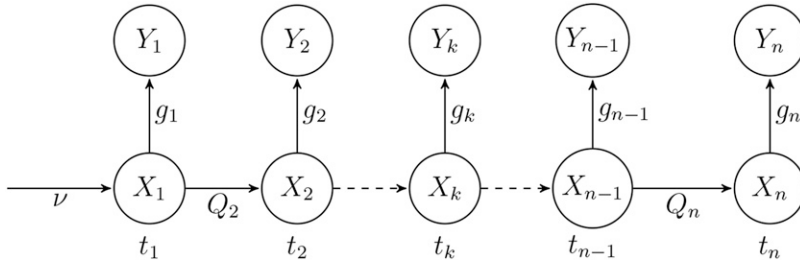
Figure 1 Graphical representation of the Hidden Markov Model for genetic time series.

In this HMM framework, the likelihood is:

$$L(y_1, \ldots, y_n; s, N_e, h) = \int \ldots \int \nu(dx_1) \prod_{k=2}^{n} Q_k^{s,N_e,h}(x_{k-1}, dx_k)$$
$$\times \prod_{k=1}^{n} g_k(x_k, y_k)$$

A direct calculation of this likelihood is very costly because the integral is over $[0, 1]^n$. The usual way to tackle this problem is to introduce the forward measure α_k, which is the joint probability of the k first observations and the $k - th$ state:

$$\alpha_k(y_1, \ldots, y_k, dx) = \mathbb{P}(Y_1 = y_1, \ldots, Y_k = y_k, X_k \in dx)$$

This forward measure is computed using the following recursion (Cappé *et al.*, 2005), generally referred as the forward algorithm:

$$\alpha_1(y_1, dx) = \mathbb{P}(Y_1 = y_1, X_1 \in dx) = g_1(x, y_1)\nu(dx)$$
$$\alpha_{k+1}(y_1, \ldots, y_{k+1}, dx') = \int_{x \in [0,1]} \alpha_k(y_1, \ldots, y_k, dx) \quad (1)$$
$$\times Q_{k+1}(x, dx')g_{k+1}(x', y_{k+1})$$

The likelihood is obtained by integrating the last forward measure α_n over all possible states X_n, using the following relation:

$$L(y_1, \ldots, y_n; s, N_e, h) = \int_{x \in [0,1]} \alpha_n(y_1, \ldots, y_n, dx)$$

Using this approach, the likelihood calculation involves only n integrals over $[0, 1]$ instead of one integral over $[0, 1]^n$. For a later use, let's also define the log likelihood:

$$l(y_1, \ldots, y_n; s, N_e, h) = \log(L(y_1, \ldots, y_n; s, N_e, h))$$

This likelihood forms the basis of statistical inference on the population parameters N_e, s and h. In this study, we focus on the inference on the selection parameter s and will therefore assume N_e and h to be known. We estimate the selection parameter s by maximum likelihood:

$$\hat{s} = \underset{s}{\text{argmax}}\, l(y_1, \ldots, y_n; s, N_e, h)$$

To build a test for the null hypothesis ($s = 0$), we use the log likelihood ratio statistic:

$$\lambda(y_1, \ldots, y_n; N_e, h) = 2(l(y_1, \ldots, y_n; \hat{s}, N_e, h)$$
$$- l(y_1, \ldots, y_n; 0, N_e, h)) \quad (2)$$

Under the null, λ asymptotically follows a χ^2 with one degree of freedom, a property that allows to derive p-values which are useful quantities in particular in a multiple testing context.

Having described the general framework used in this study, we now turn to its implementation that requires specifying the transition kernel $Q_k(x, .)$. One objective of this work was to investigate different models for this transition kernel. A reference model in this context is the one-locus Wright-Fisher (WF) model under diploid selection pressure (Ewens 2004). However, this model has computational limitations and, in the HMM famework, must generally be approximated. We first present the WF model and then several continuous approximations of this model based on a common approach: the method of moments.

Wright-Fisher transition model: The WF model is a discrete time model that assumes random mating and non overlapping generations. Let N_e be the haploid (constant) effective population size. Under this model, $X_{(t)}$ can take a discrete number of values in $\left\{0, \frac{1}{N_e}, \ldots, \frac{N_e - 1}{N_e}, 1\right\}$ and the transition kernel of this process is a transition matrix with size $(N_e + 1) \times (N_e + 1)$. Let $\mathbf{P}$ designate the transition matrix of this process in one generation. Then, $\mathbf{P} = \{P_{ij}\}_{i,j \in \{0, \ldots, N_e\}}$ and

$$P_{ij} = \mathbb{P}\left(X_{(t+1)} = \frac{j}{N_e} \middle| X_{(t)} = \frac{i}{N_e}\right)$$
$$= \binom{j}{N_e} f(i/N_e)^j (1 - f(i/N_e))^{N_e - j}$$

where f is the diploid fitness function (Ewens 2004):

$$\forall x \in [0, 1], f(x) = \frac{(1 + s)x^2 + (1 + sh)x(1 - x)}{(1 + s)x^2 + 2(1 + sh)x(1 - x) + (1 - x)^2}$$
$$= \frac{x(1 + sh + s(1 - h)x)}{1 + 2shx + s(1 - 2h)x^2} \quad (3)$$

Note that this function can easily incorporate the possibility of mutations (Ewens 2004) but it is not considered here for the sake of simplicity.

Given this one step transition matrix, the transition matrix of the HMM is obtained by taking $\mathbf{P}$ to the required power ($Q_k \sim P^{t_k - t_{k-1}}$):

$$\forall i, j \in \{0, \ldots, N_e\}, Q_k\left(\frac{i}{N_e}, \frac{j}{N_e}\right) = \left(P^{t_k - t_{k-1}}\right)_{ij} \quad (4)$$

In the WF model, the effective population size N_e is a critical parameter affecting computational properties. More precisely, the transition matrix memory occupation is $O(N_e^2)$ and the forward algorithm computation time is at least $O(N_e^2)$. Moreover, the transition matrix $\mathbf{Q_k}$ is obtained by taking a $(N_e + 1) \times (N_e + 1)$ matrix to the power $t_k - t_{k-1}$, with a complexity in $O(\Delta t N_e^3)$. It can thus be very costly for large N_e and large intersample time. To overcome this issue, one approach is to approximate the Wright-Fisher process with a continuous space process such that integral calculations of the forward

algorithm are made over a discretization of $[0, 1]$, with a grid size (i.e., a number of discretization points) much smaller than N_e. We followed this approach here and approximated the WF process using the method of moments.

Parametric approximations based on the method of moments: The principle of the method of moments consists in modeling the transition kernel Q_k with a parametric distribution, matching its mean and variance to those of the WF model. The advantages of using parameteric distributions are (i) to be able to compute transitions faster than raising the transition matrix to a power equal to the number of generations elapsed, (ii) to use smaller transition matrix in the forward algorithm. This approach was first proposed by Lacerda and Seoighe (2014) with a Gaussian distribution and was then used in several other studies (Terhorst *et al.*, 2015; Tataru *et al.* 2017; Gompert 2016). One prerequisite of the method of moments is to know the moments of a WF model for any value of s, h and N_e. Under selection ($s \neq 0$) no analytical formula of these moments exists but they can be approximated using a recursion inspired by (Lacerda and Seoighe 2014; Terhorst *et al.*, 2015; Tataru *et al.* 2017). In the WF model, moments at generation $t + 1$ can be derived from those at generation t:

$$\mathbb{E}\left[X_{(t+1)}\right] = \mathbb{E}\left[f\left(X_{(t)}\right)\right]$$

$$Var\left[X_{(t+1)}\right] = \left(1 - \frac{1}{N}\right) Var\left[f\left(X_{(t)}\right)\right]$$

$$+ \frac{\mathbb{E}\left[f\left(X_{(t)}\right)\right]\left(1 - \mathbb{E}\left[f\left(X_{(t)}\right)\right]\right)}{N}$$

For models with selection, the fitness function f (Equation (3) is non-affine so this recursion cannot be solved analytically. However, using a Taylor expansion of f around $\mathbb{E}[X_{(t)}]$, it is possible to build a recursion allowing the approximation of $\mathbb{E}[X_{(t)}]$ and $Var[X_{(t)}]$. Let μ_t and σ_t^2 be these approximations at generation t, they can be computed as follows (see Supplemental File S1 for details):

$$\mu_0 = \mathbb{E}\left(X_{(0)}\right)$$

$$\sigma_0^2 = Var\left(X_{(0)}\right)$$

$$\mu_{t+1} = f(\mu_t) + \frac{f''(\mu_t)}{2}\sigma_t^2$$

$$\sigma_{t+1}^2 = \frac{1}{N}\left(f(\mu_t) + \frac{f''(\mu_t)}{2}\sigma_t^2\right)\left(1 - f(\mu_t) - \frac{f''(\mu_t)}{2}\sigma_t^2\right)$$

$$+ \left(1 - \frac{1}{N}\right)f'(\mu_t)^2\sigma_t^2$$

$$(5)$$

Note that as f is a rational fraction, its derivatives can be calculated exactly. To simplify notations in the following, we further define $m_k = \mu_{t_k - t_{k-1}}$ and $z_k^2 = \sigma_{t_k - t_{k-1}}^2$.

Having established approximations to the Wright-Fisher moments in the general case, different parametric distributions can be chosen to approximate the transition kernel. We show on Figure S1 examples of approximations using models presented below.

Gaussian Model (Ga) The first possibility, considered in the original description of the method of moments approach (Lacerda and Seoighe

2014), is to use the Gaussian distribution (Cavalli-Sforza *et al.*, 1964; Nicholson *et al.*, 2002). Indeed, when $\frac{t}{N_e}$ and $N_e s$ are small, the rescaled Wright-Fisher process converges to a Gaussian process and the HMM transition kernel is:

$$X_{k+1}|X_k \sim \mathcal{N}\left(m_k, z_k^2\right)$$

One of the main issues with the Gaussian model is that $X_{(t)}$ can take values outside $[0, 1]$. To tackle this issue we chose to reject values distributed outside $[0, 1]$ and appropriately rescale the Gaussian density so that it integrates to one.

Nicholson Gaussian Model (NG) Another option to tackle the Gaussian distribution support issue, is to allocate the mass of the distribution outside $[0, 1]$ to discrete fixation probabilities at 0 (i.e., $P(X_{(t)} = 0) := P(X_{(t)} \leq 0)$) and 1 (i.e., $P(X_{(t)} = 1) := P(X_{(t)} \geq 1)$) (Nicholson *et al.*, 2002). Hence, the resulting distribution is a mixture of a Gaussian continuous density on $[0, 1]$ and a discrete distribution at 0 and 1.

Beta Model (Be) A second possible parametric distribution is the Beta, already used in several contexts (Balding and Nichols 1995; Gompert 2016) including the HMM based analysis of genomic time series (Hui and Burt 2015; Tataru *et al.* 2017). Use of this distribution can be motivated by the fact that it takes values on $[0, 1]$ and because it is the stationary distribution of the diffusion approximation under neutrality (Kimura 1964). In addition, the family of Beta distributions encompasses a lot of different shapes that can fit many behaviors for allele-frequency variation. In this model, the HMM transition kernel can be written as:

$$X_{k+1}|X_k \sim Beta(a_k, b_k)$$

with:

$$a_k = m_k\left(\frac{m_k(1 - m_k)}{z_k^2} - 1\right) \quad b_k = (1 - m_k)\left(\frac{m_k(1 - m_k)}{z_k^2} - 1\right)$$

$$(6)$$

Beta with spikes Model (BwS) We finally considered the Beta with spikes model described by Tataru *et al.* (2017). Similarly to the Nicholson Gaussian model described above, this model explicitly accounts for the possibility of allele fixation: it is a mixture of a Beta distribution and two discrete fixation probabilities at 0 and 1. According to Tataru *et al.* (2017), this model is a better approximation of the Wright-Fisher than the Ga and Be models. Let

$$p_{0,k}(x) = \mathbb{P}(X_k = 0|X_{k-1} = x)$$

and

$$p_{1,k}(x) = \mathbb{P}(X_k = 1|X_{k-1} = x).$$

The transition kernel is:

$$Q_k^{s,N_e,h}(x, .) = p_{0,k}(x)\delta_0 + p_{1,k}(x)\delta_1 + (1 - p_{0,k}(x) - p_{1,k}(x))\eta_{a_k^*, b_k^*}$$

where $\eta_{a_k^*, b_k^*}$ is a Beta distribution describing the transition from X_k to X_{k+1} conditional on the fact that X_{k+1} did not fix in 0 or 1. Parameters a_k^* and b_k^* are obtained with a formula similar to (6), replacing abolute moments m_k and z_k by conditional mean and variance m_k^* and z_k^* given that the WF process did not fixed at 0 or 1:

$$m_k^* = \frac{m_k - p_{1,k}}{1 - p_{0,k} - p_{1,k}} \qquad z_k^{2*} = \frac{z_k^2 + m_k^2 - p_{1,k}}{1 - p_{0,k} - p_{1,k}} - m_k^{2*}$$

In the Beta with spikes model, fixation probabilities must be computed in addition to WF moments. Let us define $p_{0,t} = \mathbb{P}(X_{(t)} = 0)$ and $p_{1,t} = \mathbb{P}(X_{(t)} = 1)$. We used the following recursion, proposed by Tataru *et al.* (2015) assuming Beta with spikes transitions:

$$p_{0,0} = 0$$

$$p_{1,0} = 0$$

$$p_{0,t+1} = p_{0,t} + \left(1 - p_{0,t} - p_{1,t}\right) \frac{B\left(a_t^*, b_t^* + N\right)}{B\left(a_t^*, b_t^*\right)}$$

$$p_{1,t+1} = p_{1,t} + \left(1 - p_{0,t} - p_{1,t}\right) \frac{B\left(a_t^* + N, b_t^*\right)}{B\left(a_t^*, b_t^*\right)}$$

where B denotes the Beta function.

Implementation In this study, the likelihood was always evaluated on a grid of $N_e s$ composed of 600 values ranging from -100 to 300. For WF transitions the integrals involved in the likelihood computation using the forward algorithm (1) are finite sums. However for the parametric approximations to the WF these integrals are continuous and must therefore be numerically approximated. Here, this was done using the trapezoidal rule on an integration grid composed of 129 points evenly spaced on $[0, 1]$. Note that the calculation of these transitions is only required once per dataset.

Comparison of HMM transition models
We studied the quality of the transition approximation to the WF process. For this objective, we set the effective population size to $N_e = 100$ haploids to render the comparisons computationally tractable. We compared the Ga, NG, Be and BwS models as approximations to the Wright-Fisher process, first by evaluating their fit to the WF distribution and second by measuring their impact on selection inference.

Approximation of the WF transition: We first studied the quality of approximations of the WF transition kernel. For all parametric distributions, at least part of the WF discrete transition is approximated by a continuous density. Thus, measuring the fit of these approximations requires comparing continuous densities to discrete probabilities. In such a case, classical distances aimed at comparing densities, such as the Kullback-Liebler divergence or the Hellinger distance are not valid. Therefore we used the *Wasserstein distance* defined as the L_1 distance between the cumulative distribution functions of two distributions and which is well defined in all cases considered here. The fit of the Ga, NG, Be and BwS models were evaluated over the parameter set $h = 0.5$, $(t_1 - t_0) \in [1, 20]$, $N_e s \in \{0, 10, 100\}$, and starting allele frequency covering all i/N_e values for $i \in \{0, \ldots, N_e\}$

Consequences for statistical inference: We computed the likelihood profile (see implementation paragraph) for genomic time series randomly simulated with the Wright-Fisher process. We set the initial allele frequency to 0.1 or 0.5 and its selection parameter varying from 0 (neutral case) to 1 (strong selection). We simulated a random sampling of $n_k = 30$ alleles at 10 evenly spaced dates over a total range of $T \in \{9, 45, 90, 180\}$ generations (*i.e.*, $t_k - t_{k-1} \in \{1, 5, 10, 20\}$). We ran 100000 simulations for $s = 0$ and 10000 for other s (see Figure S2).

■ **Table 1 Haploid sample sizes across 6 generations of chicken populations selected for intra-muscular ultimate pH**

Dates (in generations)	0	1	2	3	4	5
pHu+	102	28	42	56	54	60
pHu-	102	34	46	44	54	56

To compare the Ga, NG, Be and BwS models in terms of selection inference, we focused on four summarizing quantities computed from the likelihood: the log-likelihood value at $s = 0$, the maximum likelihood estimator $\hat{s}$, the log-likelihood value at $s = \hat{s}$, and the log-likelihood ratio statistic (2).

Inferring selection using the HMM: As the BwS model was found to be the best approximation to the Wright-Fisher (see Results), we further studied its statistical properties on simulations for higher N_e and applied it to detect selection on a real data set. The NG model, which provided the second best fit to the Wright-Fisher, was also considered for comparison when analyzing simulated data.

Simulation study: To study the statistical properties of NG and BwS models, we created new datasets designed to represent consistent selection dynamics as for the $N_e = 100$ dataset but with larger N_e. To do so, we considered that the diffusion approximation to the WF suggests that selection dynamics are driven by the two compound parameters $N_e s$ and T/N_e. Therefore the new datasets were created by increasing N_e to 1,000 and 10,000 but in each case adjusted T and s so as to keep the same set of $N_e s$ and T/N_e values as for the $N_e = 100$ dataset. Specifically we considered $N_e s \in [0, 100]$ and $T/N_e \in \{0.09, 0.45, 0.9, 1.8\}$. We also kept the same sampling pattern of $n_k = 30$ haploids sampled at 10 evenly spaced dates over the total range T and an initial allele frequency of 0.1 or 0.5.

Real data analysis: We applied our method using the BwS model on a dataset of two experimental populations of chicken divergently selected for the intra muscular ultimate pH (pHu) (Alnahhas *et al.*, 2016; Bihan-Duval *et al.*, 2018). This selection experiment was conducted for 5 generations and at each generation (including the starting population) samples were collected and genotyped for 40,199 markers. Sample sizes in each line (pHu- and pHu+) are given in Table 1.

Our current implementation of the method does not estimate the N_e parameter which is not known in this real dataset. To remedy this issue, we estimated N_e in each line using the NB R package (Hui and Burt 2015). Motivation for using this method to estimate N_e comes from the fact that the underlying model is a HMM with Beta transitions, similar to our general framework. To estimate N_e, we removed trajectories starting near edges (frequencies below 20% or above 80%), running the analysis over 27,669 loci. For our analysis we considered the two lines as two different populations.

As a further validation of the Beta with spikes approximation, we applied the HHH both with Wright-Fisher and Beta with spikes transitions. We removed from the analysis all trajectories fixing in one step as we used the $\chi^2(1)$ distribution to compute p-values. SNP were called significant with a FDR threshold of 5% estimated using the q-value R package (Storey *et al.*, 2015).

Data availability
The implementation of the HMM is freely available at https://github.com/CyrielParis/compareHMM. Observed allele frequencies in the two chicken lines, for all markers and generations, are available
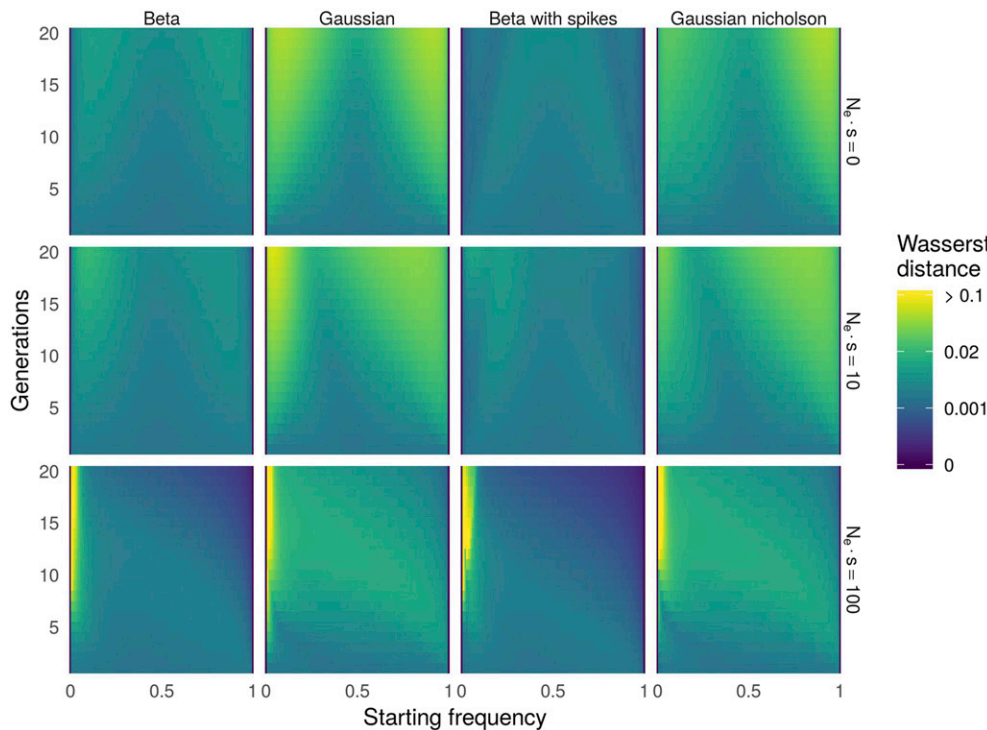
Figure 2 Wasserstein distances to Wright-Fisher transition for four continuous approximations (columns) and for varying starting allele frequencies (x axis of each panel) and number of generations (y axis of each panel). Top panels: Neutral evolution ($N_e s = 0$). Middle panels: Mild selection ($N_e s = 10$). Bottom panels: Strong selection ($N_e s = 100$).

at: http://doi.org/10.5281/zenodo.3461580. Supplemental material available at figshare: https://doi.org/10.25387/g3.9959492.

RESULTS

We have established a general HMM framework that allows to perform statistical inference from genetic time-series data. Within this framework, different transition models can be used and easily compared. In this section we will first present results comparing the quality of approximations of the Gaussian (Ga), Nicholson Gaussian (NG), Beta (Be) and Beta with spikes (BwS) transition kernels on their ability to approximate the Wright-Fisher (WF) kernel. Next, we study in more details the statistical performances of the BwS for detecting selection and estimating its intensity. Finally, we illustrate how the BwS model can be used on real data by re-analyzing an experimental evolution dataset.

Approximation of the Wright-Fisher model

Approximation of the WF transition: We first compared each continuous approximation with the WF using the Wasserstein distance. Figure 2 shows for each model and for different selection intensities the heatmap of the Wasserstein distance between the continuous approximation and the true WF distribution. In principle this distance ranges from 0 which is a perfect match to 1 which is, for an allele frequency, the worst possible distance. For this set of parameters, the Wasserstein distance ranged from 0 to ≈ 0.5. According to this distance, all approximations are valid for intermediate starting allele frequencies and a small number of generations but the accuracy of all approximations decreases as the number of generations increases. The accuracy of the approximation also worsens when the selection parameter increases, especially for small starting allele frequencies. Indeed, the figure shows that for starting frequency < 0.1, generations > 15 and $N_e s = 100$ all models performed badly in approximating the Wright-Fisher model. As the forward algorithm involves an integration over all possible allele frequencies at each sampling date, poor approximations of transitions from small allele frequencies may in principle affect the performance of

HMM inference for any observed trajectory (with or without small frequencies). We evaluate this in the next setion.

Concerning the comparison between models, Figure 2 shows that the Be and BwS models are better approximations than the Ga and NG models, the BwS model being overall slightly better than others. These results are consistent with those obtained by Tataru *et al.* (2017), although they used a different approximation for the moments and a different measure of the distance between distributions.

The error in the approximation can come from the model itself but also from the use of approximated moments. To evaluate what proportion of the distance to the true distribution comes from the moment approximation, we computed the Wasserstein distance of each continuous distribution using the true WF moments (Equation 4, Figure S3). Figure 3 presents this proportion for the BwS model under neutrality (left $N_e s = 0$) and under selection (middle $N_e s = 10$ and right, $N_e s = 100$). Under neutrality the error comes mainly from the continuous approximation. This is also usually the case under selection although for small starting frequencies ($x_1 < 0.1$) and large duration between samples the error in the approximation of moments is the most influential. As getting the true moments of the Wright-Fisher process can be costly especially for large population size, all results presented further are based on approximated moments.

Approximation of the WF likelihood: Next, we evaluated the ability of continuous WF approximations to recover the WF likelihood of genetic time-series data. To do so, we simulated trajectories under different evolutionary scenarios (see Methods) and compared the inference obtained with each continuous model to the one obtained using the Wright-Fisher model.

Simulations of time-series genetic data to evaluate the quality of parametric approximations to the Wright-Fisher were performed under this reference model. Hence, using the HMM with WF transitions should provide the best statistical behavior and be considered a reference inference method. Before assessing the quality of approximations to the
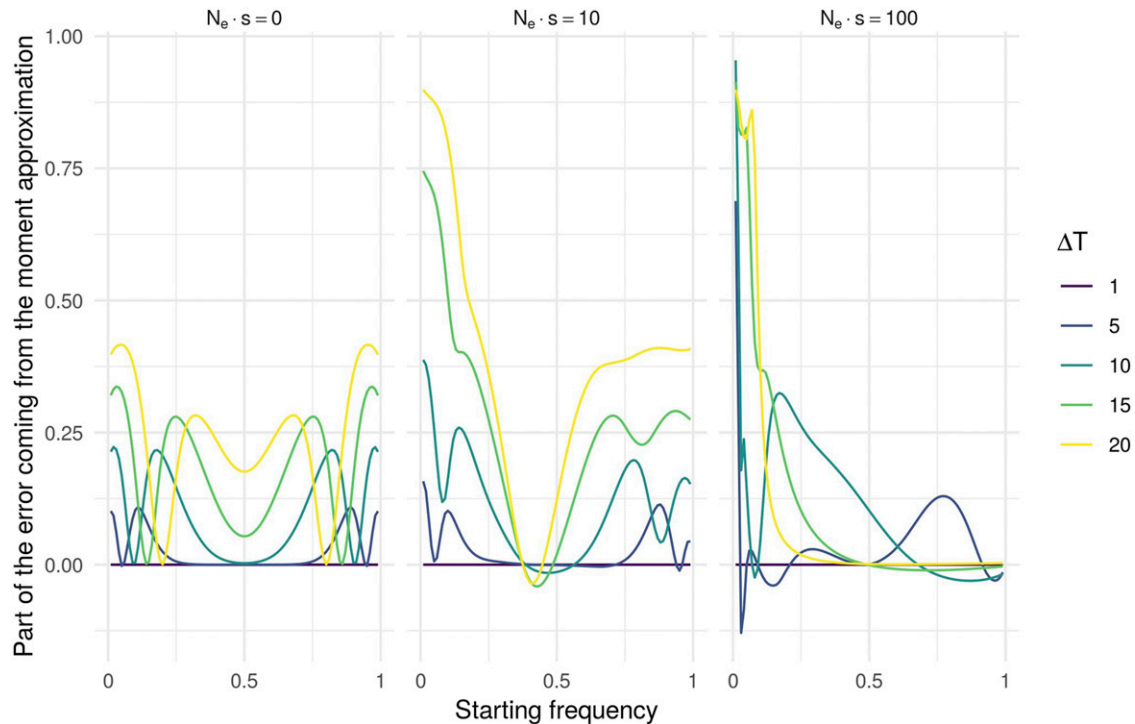
Figure 3 Proportion of the Wasserstein distance of the BwS due to the approximation of moments. Negative values correspond to very rare cases where the transition with approximated moments is better than with true moments.

WF model with respect to their accuracy for computing the WF-based likelihood, it is necessary to ensure that this model itself provides sensible likelihood inference. Results on simulated data showed that this was not always the case: a family of trajectories were found to lead to unreliable results even when using the WF model for inference. These trajectories correspond to cases where the observed allele frequency reached fixation in one time step (*i.e.,* $\forall k > 1, Y_k = 0$ or $\forall k > 1, Y_k = n_k$). In such cases the likelihood function is monotonic in s: the s value that maximizes the probability of fixation at 1 in one time step is $s = \infty$ and similarly the s value that maximizes the probability of fixation at $s = 0$ in one time step is $s = -1$, essentially corresponding to an infinite s for the other allele. In these cases, likelihood inference is not valid. They could be deemed merely pathological but they occur with a non negligible probability that is increasing as the intersample duration or the selection intensity increases. In the following, we excluded such cases when evaluating the statistical properties of inference models. However we studied their occurrence frequency (see below).

Figure 4 presents four different statistics based on the likelihood of simulated data under various scenarios (see Methods), comparing results obtained for each parametric approximation to those obtained using the WF: Figure 4a presents results on the likelihood at $s = 0$, Figure 4b the maximum likelihood estimates (MLE), Figure 4c the likelihood at the MLE (which can be different from the WF one) and Figure 4d the likelihood ratio statistic. For all these statistics and independently of the scenario simulated, the BwS and the WF provided essentially the same inference. Only for the MLE there are few cases where the two transition model disagreed (Figure 4b). However, this is never true for the LMLE (Figure 4c) so when the two models disagree for $\hat{s}$ they still give similar likelihood values at the MLE. Overall this shows that the

few large differences in MLE between BwS and WF most likely correspond to cases where the data are not very informative on the selection intensity. On the contrary, there are many scenarios where the Be and Ga models have very different values than the WF for all statistics considered. Finally, the NG model showed significantly better properties than the Be and Ga, but it provided a less accurate fit to the WF than the BwS model, for all parameters studied.

One reason why BwS and NG models are better approximations to the WF than the Beta and the Gaussian is that they include fixation probabilities. In the HMM framework, fixation events are considered by these models while others fail to correctly interpret vanishing observed allele frequencies. Indeed, when fixation events are unlikely, all models approximate efficiently the Wright-Fisher and the likelihood approximation is good. Fixation events depend mainly on the total duration of sampling T (see Figures S4-S7). For large values ($T/N_e > 0.5$) early fixation events (apart from the pathological cases mentionned above) are not interpreted by the Beta and Gaussian models as fixations. These models will consider that the constant trajectory subsequent to the fixation event is consistent with a neutral process. The resulting maximum likelihood estimator therefore underestimates the true s (Figure 4b). As the maximum is not well found by Beta and Gaussian models, the likelihood-ratio statistic is inconsistent and is not calibrated as a $\chi^2(1)$ distribution (Figure S8).

Inferring selection using the Beta with spikes model

Given that the Beta with spikes model closely mimics the Wright-Fisher model in the HMM framework, we studied the statistical properties of the HMM with BwS transitions for population sizes of 1,000 and 10,000 haploids, which are computationally prohibitive when using the WF transitions. In terms of statistical properties, we studied the ability to detect selective loci and the accuracy in estimating the
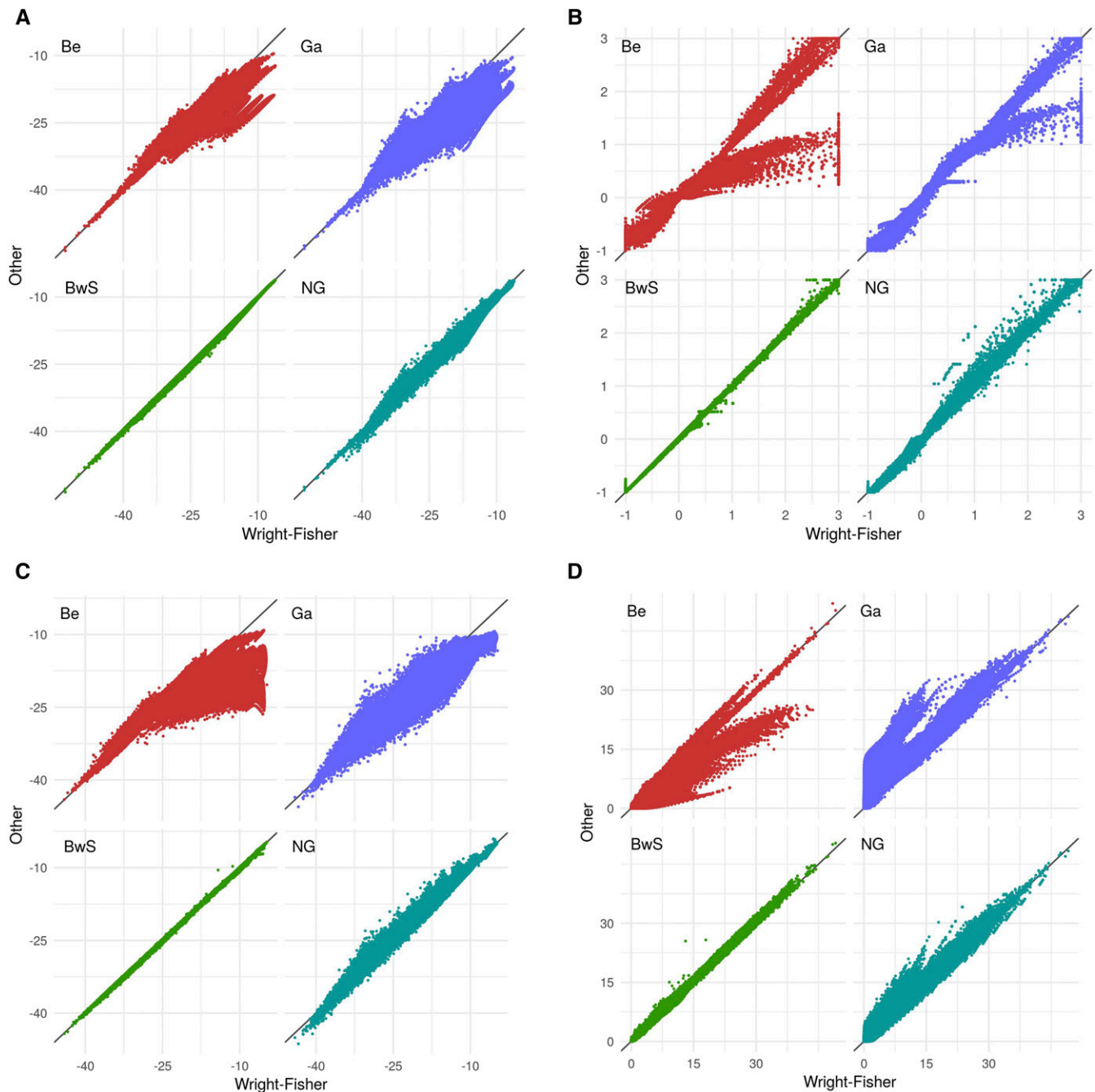
Figure 4 Comparison of four continuous approximations to the WF transition for likelihood-based statistics. In each subplot, Be, Ga, BwS, NG indicate the compared model among Beta model, Gaussian model, Beta with spikes model and Nicholson Gaussian model.

intensity of selection. The same analysis was conducted using the NG model and comparable results are provided in Supplementary Figures S9-S11.

Detection of selection: The HMM framework developed here allows to derive a likelihood-ratio statistic that can be used to test for the null hypothesis $s = 0$. Asymptotically, this statistic follows a $\chi^2(1)$ distribution. However, the asymptotics in the HMM framework are measured in terms of the number of dates sampled. In our simulations, and most likely in typical real datasets, this number is not very high

(10 here). Our first goal was thus to evaluate whether the $\chi^2(1)$ was a good fit to statistics computed under the null hypothesis. Then we evaluated its power under different alternatives.

Calibration under the null: We checked for the good calibration of the BwS under neutrality. Figure 5 shows on the y axis the quantiles of the empirical likelihood ratio distribution and on the x axis, the corresponding quantiles of a $\chi^2(1)$ distribution. Generally the fit of the LR to the theoretical distribution was good although the test would be slightly conservative for large T/N_e.
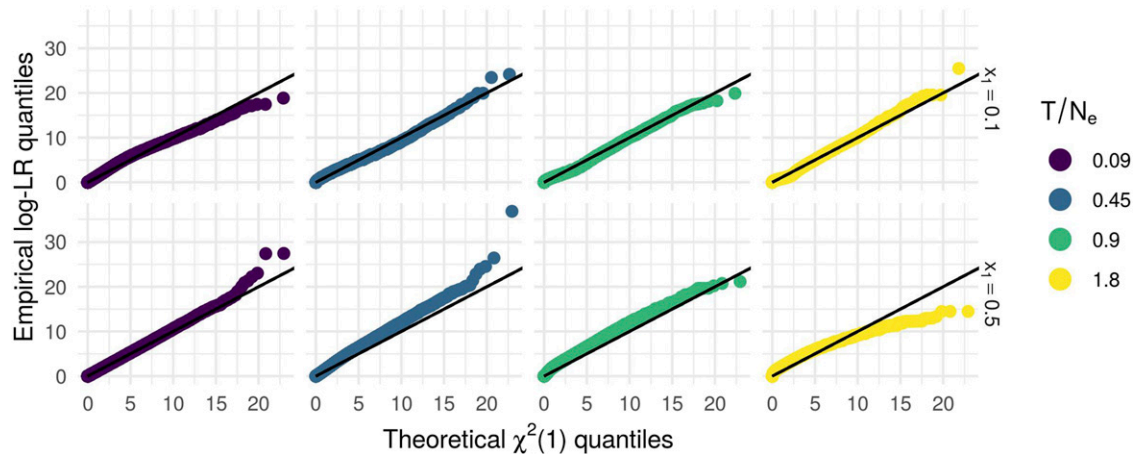
Figure 5 Calibration of $\lambda(y)$ under the null: comparison of empirical *vs.* theoretical $\chi^2(1)$ quantiles for different population sizes, initial allele frequencies and inter-sample times.

Power: With the LR statistic and the χ^2 quantiles, the empirical power of the test can be computed for different scenarios. As in this case we know that all true s parameters are non negative, we considered that all negative $\hat{s}$ having a high LR should not be considered as a success in detecting selection. Indeed, even if the LR is high in this case, the final conclusion of such results would be that the wrong allele is under selection which is even worse than not detecting selection.

Figure 6 shows the power of the test as a function of $N_e s$ (Figure 6a) and as a function of T/N_e (Figure 6b), for two different initial allele frequencies (0.1 and 0.5). As expected, the power increased with $N_e s$ (Figure 6a). We can also see that for larger T/N_e, the power starts to increase for smaller $N_e s$. This can be explained by the fact that in such scenarios, selection has more time to affect the allele frequency and even small $N_e s$ give rise to trajectories that are consistent with selection. However, when starting at an initial frequency $x_1 = 0.5$ the power decreases when T/N_e is too high. In these cases, the probability to observe large variations of allele frequency under neutrality becomes larger and so the ability of the LRT to distinguish neutrality from selection decreases.

Parameter estimation: Figure S12 shows that under the parameters considered, N_e has no influence on estimation, except for large $N_e s$, for which the maximum likelihood estimator has a higher variance for small N_e. In this case, the diffusion approximation is not valid so different N_e are not equivalent for given $N_e s$ and T/N_e.

Figure 7 shows the MLE distribution for different $N_e s$ and T/N_e pooling results from different N_e. It shows that the estimation is more accurate (with lower variance) for large values of T/N_e. Moreover, when fixation events ($T/N_e < 0.5$) are unlikely, the estimation is unbiased. On the other hand, when fixation events are likely ($T/N_e > 0.5$ and $N_e s > 50$), the trajectories that are not filtered out are those that are consistent with lower values than the true s and therefore the MLE underestimates the true s.

Analysis of a selection experiment

To illustrate the statistical performance of the HMM approach on a real data set, we applied it to allele frequency trajectories observed at 40,199 SNPs in two lines of chicken divergently selected for the intra muscular ultimate pH (pHu). This dataset was previously analyzed by Bihan-Duval *et al.* (2018) using the FLK Bonhomme *et al.* (2010) and hapFLK Fariello *et al.* (2013) tests: considering each generation of the experiment independently of the others, these authors looked for regions showing a significant excess of allele (for FLK) or haplotype (for hapFLK) frequency differentiation between pHu+, pHu- and the founding population. Thus, this case study offers the opportunity to compare the regions detected under selection by two different approaches: genetic differentiation and time series.

We first estimated population size in the two lines using the NB R package (Hui and Burt 2015) and found $N_e = 157$ haploids in pHu+ and $N_e = 123$ haploids in pHu-. Based on these values, we applied the
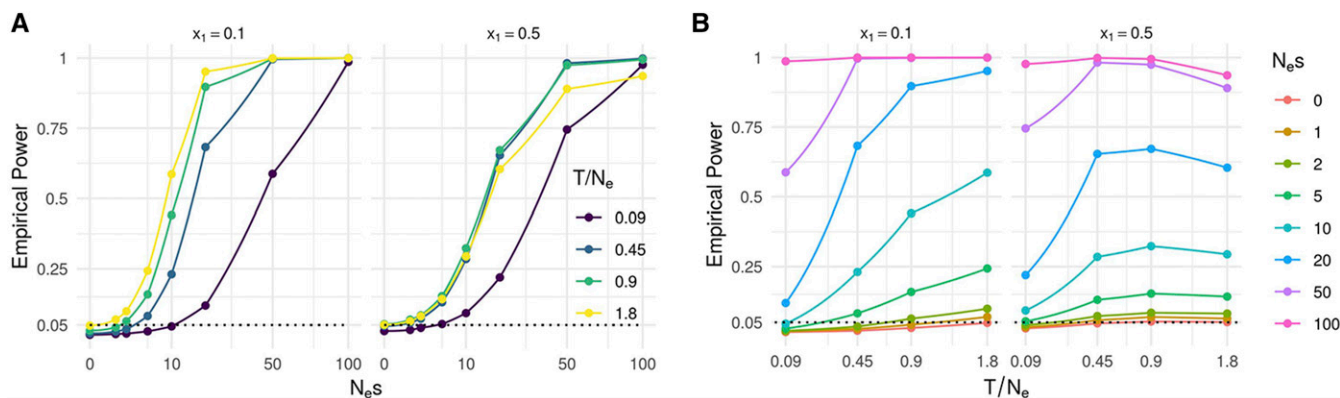
Figure 6 Power of the likelihood ratio test under different simulated scenarios.
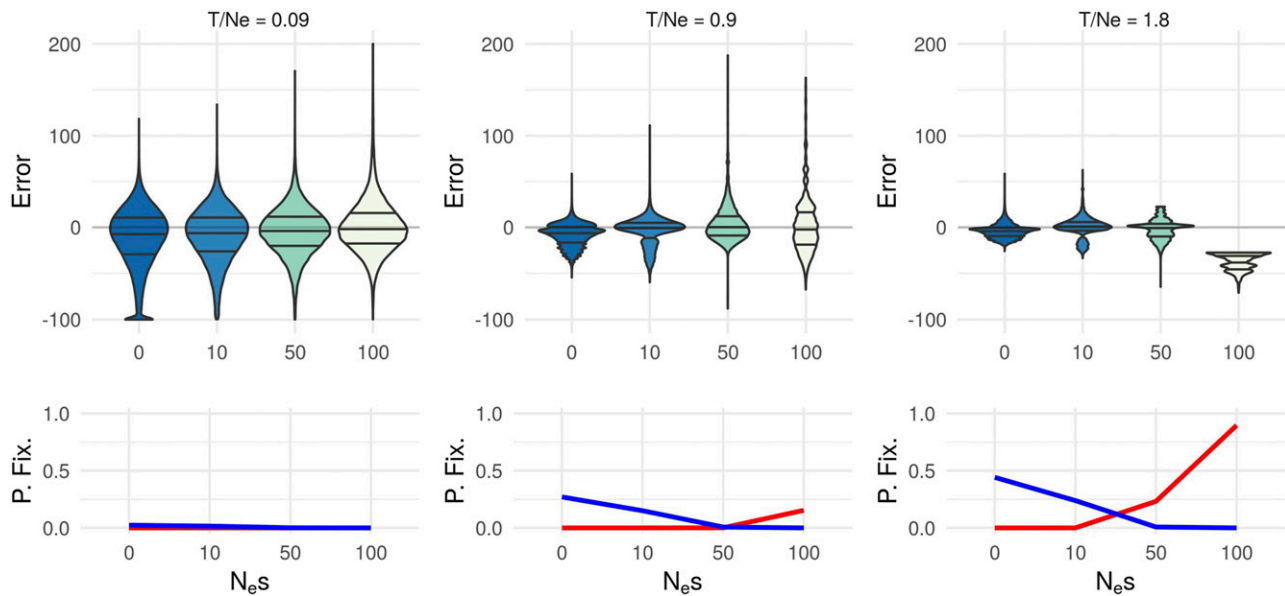
Figure 7 Estimation error distribution in different scenarios for $x_1 = 0.1$. Each column stands for a scaled time parameter $T/N_e \in \{0.09, 0.9, 1.8\}$. The first line indicates the absolute error distribution in each scenario and the second line represents in each case, the proportion of rejected trajectories (red lines for fixations in 1 and blue lines for fixations in 0).

HMM in each line and detected a total of 39 significant SNPs under selection at a FDR of 5% (Table S1). We also conducted the analysis using the Wright-Fisher model and obtained exactly the same results (Figure S13). These SNPs were clustered in 12 genomic regions (Table 2, Figure S14: eight of them were detected only in the phu+ line, one was detected only in the phu- line and three others were detected in both lines. Moreover, for almost all significant SNPs, the allele that was selected in one line (positive value of $\hat{s}$) was also counter selected in the other line (negative value of $\hat{s}$) (Figure 8). This is consistent with the experimental design of divergent selection between the two lines.

Among the 12 regions detected by the time series approach, six were detected by FLK or (and) hapFLK and six were new. SNPs in these six new regions were associated to an allele frequency significantly increasing or decreasing in one line but rather constant in the other line (Figure S15): in such cases, genetic differentiation was elevated but did not result in a significant FLK or hapFLK p-value. In contrast, SNPs detected only with FLK were generally associated to a moderate allele frequency change in both lines, that was not found significant when analyzing each of them independently (Figure S16). Overall, the

differentiation and time series approaches were thus found complementary. Note also that the number of regions detected (*i.e.*, the detection power) with the time series approach was very similar to that obtained by FLK. Detection power was significantly higher with hapFLK, which was related to the use of haplotype information.

DISCUSSION

In this study we have described a general statistical framework for the analysis of genetic time series. It is built upon combining a Hidden Markov Modeling approach, first proposed by Bollback *et al.* (2008) in the context of genetic time series and parametric approximations to the Wright Fisher process. Within this framework we have shown that using the Beta with spikes distribution proposed by Tataru *et al.* (2015) to approximate the Wright Fisher transitions provided a statistical inference comparable to that of the Wright-Fisher model while having a computational cost that does not depend on population size. We have shown this framework allows to perform hypothesis testing by relying on the asymptotic distribution of the likelihood-ratio statistic and that the maximum likelihood

■ **Table 2 Significant regions (for HMM)**

Region	number of SNP	common with	begin	end
1:a	1	hapFLK	7,422,723	7,422,723
1:b	21	FLK & hapFLK	20,580,348	27,053,505
2:a	4	hapFLK	24,035,796	31,975,586
3:a	5	none	16,910,143	18,598,504
3:b	1	hapFLK	104,194,096	104,194,096
4:a	1	none	63,446,574	63,446,574
6:a	1	none	6,356,771	6,356,771
13:a	1	hapFLK	14,544,320	14,544,320
14:a	1	none	1,718,020	1,718,020
17:a	1	FLK	8,225,007	8,225,007
21:a	1	none	4,259,090	4,259,090
22:a	1	none	4,635,167	4,635,167
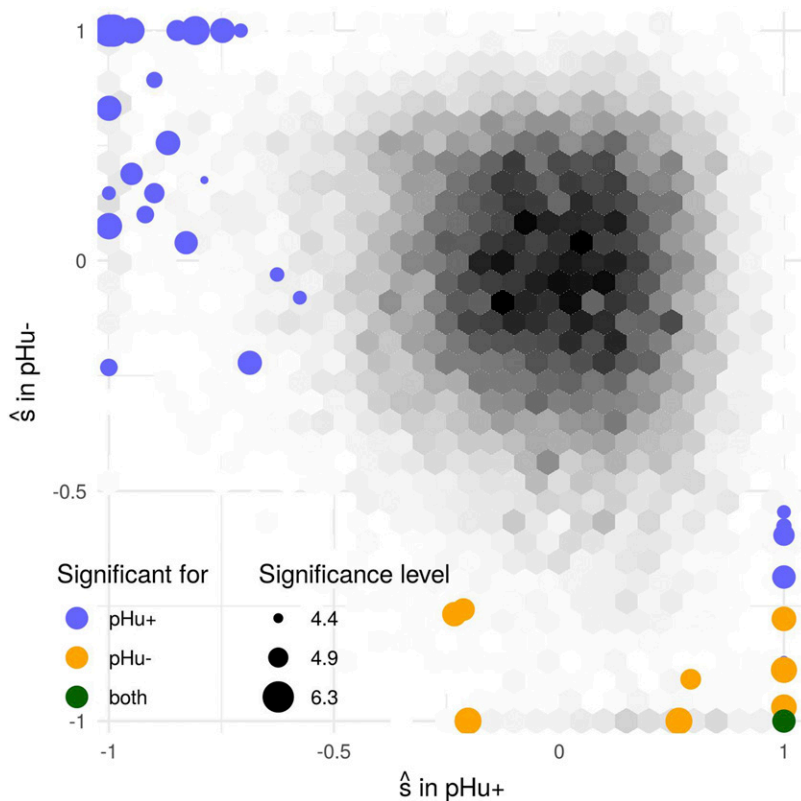
Figure 8 Distribution of ŝ in pHu- vs. pHu+. Levels of gray represent the density of ŝ, significant SNP are represented in colors, the size representing the $-\log_{10}$ associated p-value.

estimate of the selection intensity parameter had generally good statistical properties.

Advantages and limits of parametric approximations to the Wright Fisher process

Using parametric distributions to approximate Wright-Fisher transitions requires the user to know or approximate the moments of the Wright-Fisher process, including fixation probabilities in the case of the BwS and NG models. We have shown, in the lines of previous work by Lacerda and Seoighe (2014); Terhorst *et al.* (2015) that this could be done using a Taylor expansion of the fitness function (see Supplemental File S1). Despite generally good performance of the approximation of moments, there still exists situations were the approximation performed poorly. For the BwS model, this only happened for low starting allele frequencies ($x_s \leq 0.1$), large number of generations ($t/N_e \gtrsim 0.1$) and strong selection. Note that the poor performance for low starting allele frequencies is relevant for all *underlying true* allele frequency as for computing the likelihood the transition is integrated over all possible starting frequencies in the forward algorithm. So this poor performance could have impacted our simulation results. However, we showed that the WF likelihood was still well approximated by the BwS model. In cases where better approximation of the WF moments might be required, different approaches could be followed. For small population sizes ($N_e \leq 1000$), the WF moments can actually be computed exactly using equation (4) but this computation quickly becomes computationally prohibitive (the cost is approximately $\propto N_e^2$). A solution would be to develop the Taylor expansion some degrees further and possibly include correction terms as was done by Terhorst *et al.* (2015). This could be important for datasets covering large timescales such as in ancient DNA studies where sampling times can be widely spread.

Here, we compared four different parametric distributions for approximating the Wright-Fisher transitions. The Gaussian distribution is a widely used approximation to the evolution of allele frequencies, starting from (Cavalli-Sforza *et al.*, 1964), including for the analysis of genetic time series (Lacerda and Seoighe 2014; Terhorst *et al.*, 2015). Our results confirm what is widely known: this approximation is only valid for small time scales (in units of t/N_e) and when the starting minor allele frequency is high. The Beta distribution has been less used in this context (Siren *et al.*, 2011; Gompert 2016) although we have shown that it offers a better approximation to the transitions than the Gaussian model. However these two models present poor performance when it comes to approximating the likelihood of data simulated under the Wright-Fisher model. This it not true for the Beta with spikes model that provides good approximations both for the transition kernel and the likelihood. Moreover, despite poor approximation of WF transitions for some parameters, the Nicholson Gaussian model generally provided a good fit of the likelihood. The good fit of NG and BwS is most likely due to the fact that they incorporate fixation probabilities that become large when the other models fail (large selection intensity, large intersample time, low minor allele frequency).

The BwS model should therefore be preferred to the other three at least in the context of analyzing genetic time series using Hidden Markov Models. Given the good performance of the BwS model in this framework, it could be an interesting model to use in more general contexts. One interesting extension would be to use the BwS in the context of time series data sampled in multiple populations. However, one important limitation of using the Beta distribution (with or without spikes) for such data are that the multivariate Beta distribution is essentially unusable in practice. Using this class of distribution in a multiple populations context therefore requires developing a specific factorization for calculating the multivariate likelihood. An example of such an extension for the

problem of estimating branch length in population trees can be found in Tataru *et al.* (2015). For such data, the Gaussian model remains much more practical but still suffers from the limitations mentioned above.

In this study, we only focused on comparing parametric distributions to approximate the WF transitions. Another standard approximation to the Wright-Fisher process under selection that was not considered here is the diffusion process. This process has been shown to provide a good approximation to the WF process. However this approximation is only valid for large N_e and small $N_e s$ values and has two additional important limitations. The main limitation is that using the diffusion model in practice implies solving the diffusion equation which is a partial differential equation. One approach to do so is to write the transition distribution as an infinite sum of elementary solutions of the diffusion equation (Song and Steinrücken 2012). This approach was considered in a HMM for genetic time series by Steinrücken *et al.* (2014) but exhibited prohibitive computational cost. Another approach is to consider numerical schemes to approximate the transition density (Bollback *et al.*, 2008; Zhao *et al.*, 2013). These approaches involve numerical integrations and are also computationally demanding. The second limitation of the diffusion process is that it does not naturally incorporate the fixation probabilities at 0 or 1. However, Zhao *et al.* (2013) showed that the discretized diffusion transition density had atoms on these edges that could be used to represent the fixation probabilities. This suggest that the diffusion approximation could be extended to include fixation probabilities. This would require specific developments and evaluation.

Overall, the choice of a particular transition model for genetic time series should depend on the study to analyze. For small N_e, WF calculations are tractable and should be preferred. For larger N_e, the diffusion approximation could be improved but parametric approximations seem more adapted (Lacerda and Seoighe 2014). In the class of parametric distributions, our results show that the Beta with spikes model is probably more flexible as it is adapted to both large N_e and a wider range of $N_e s$.

Inference of selection from genetic time series

In principle, genetic time series can provide a lot of information on selection at a locus. This comes from the fact that the allele frequency trajectory underlying the observed sample is greatly affected by the selection process. Indeed we found that in many scenarios that we simulated likelihood inference provided essentially unbiased estimates of the selection parameter. The power to detect selection was substantial for a wide range of situations when the selection intensity $N_e s$ was greater than 10. For large populations this corresponds to a relatively modest fitness effect. For example for a population of size 10,000 this corresponds to $s = 0.01$, *i.e.*, a 1% increase in fitness for an individual homozygote for the beneficial allele compared to the other homozygote.

Our approach builds upon likelihood inference and in particular maximum likelihood estimation. While this approach was found very efficient in most cases, the maximum likelihood theory could not be applied to allele frequency trajectories where the reference allele was fixed or lost between the first and second sampling dates. Indeed, the best explanation for such events is $\hat{s} = -1$ (for allele loss) or $\hat{s} = \infty$ (for allele fixation), in other words the MLE is reached on an edge of the domain. The occurrence of such events implies that the LR distribution under neutrality is not $\chi^2(1)$. An approach to ensure properly distributed p-values in this situation would rely on performing neutral simulations to obtain empirically the distribution of the LR under the null. However, a more fundamental issue is that the ML estimator gives no accurate information about the true s here, because

sampling dates are too distant to capture how fast the allele frequency increases (or decreases).

We chose to treat these trajectories specifically and excluded them from further analysis. However, the exclusion of fixed trajectories leads to an underestimation of s. This happens because the trajectories that are not excluded correspond to a biased sample of trajectories (those that reached fixation more slowly) and the resulting $\hat{s}$ is lower than the true s. A way to compensate this effect would be to consider the transition process conditionally on the allele not being fixed. This can be done in the method of moments framework by using the conditional moments, and with the diffusion model by modifying the diffusion equation (Ewens 2004). However, this methodology may exclude a potentially large number of loci that can be properly analyzed within our framework, *i.e.*, all trajectories including allele loss or fixation *later than* the second sampling date. In real data, the way to avoid such situations is through the experimental design by chosing an appropriate sampling strategy.

In this study we only examined one sampling strategy: 10 sampled dates equally spread along the total duration with 30 alleles sampled each time. With this strategy, we showed that the statistical quality of the model is driven by the $N_e s$ and T/N_e parameters: if the selection is too weak, increasing the duration considered is needed so as to capture its effect on the allele frequency trajectory. But if the total duration is too large, distinguishing the effects of genetic drift and selection is harder. The consequence of our results is that, for a given dataset, as the statistical accuracy of the estimation and the power of the test depend on the sampling dates distribution and the sample sizes, the interpretation of the results becomes more difficult. Some selected loci will escape detection although they can be potentially important for fitness. Hence, for a given dataset, we would suggest performing simulations (possibly using the scripts we used here) matched for sampling dates and sizes under different selection intensities to help interpreting selection scan results. Moreover, such simulations could actually be used to compare possible sampling strategies (*i.e.*, depending on resources such as cost and sample availability) to optimize the power of the eventual statistical analysis. Deriving general guidelines for the optimization of sampling strategies in the context of genetic time series would require a dedicated study.

One more aspect requiring further research would be the impact of the initial allele frequency distribution used in the HMM. In our simulation study the initial allele frequency was fixed (to 0.1 or 0.5), but we preferred ignoring this information that is unknow in practice and used the uniform distribution. When analyzing real data the allele frequency distribution is not expected to be uniform. Using a more realistic distribution may help the algorithm and make it more confident about its inference. However, this confidence persists even if the given initial distribution is actually wrong, which may then bias the inference. The choice of an informative initial distribution should thus be done with caution. For instance, while the hyperbolic function ($\propto 1/x$) represents a natural choice for the site frequency spectrum, one should be aware that it is only expected for a neutral population with constant size. Another solution would be to estimate the ancestral allele frequency distribution from the data, which is possible if time series are available genome-wide. Finally, note that these considerations really matter only if sample sizes at the oldest sampling dates are small, otherwise the emission probabilities at these dates are informative enough to overcome any prior on the initial distribution.

Our evaluation of the HMM framework focused on its performance to detect and estimate selection. To do so we have assumed that other parameters, in particular effective population size, were known. In addition, we assumed that each parameter was constant over the

sampling duration. In practice most of these assumptions are deemed to be wrong. Regarding the estimation of N_e, in our analysis of the real dataset, we used a two-step approach by first estimating N_e using Hui and Burt (2015) and then s at each locus. The issue with such an approach is that all loci are used to get the N_e estimate. If some of them are under selection, they will exhibit larger allele frequency deviation. The consequence is that the N_e estimate is going to be under estimated and it becomes harder to reject the null hypothesis (genetic drift). Hence, while this two-step approach is practical, it could be improved by jointly estimating N_e and s. The framework described here can in principle be used to do so as the HMM likelihood incorporates both N_e and s. This will require fitting the HMM model on multiple loci jointly as the information on N_e lies in the variance of the allele frequency trajectories over the whole genome. Also, it is now well established that populations of many species have experienced large changes in effective population size over time. There is also many situations where the selection intensity of an allele (expressed by s in our model) will change over time (*e.g.*, time varying environmental conditions, selection for an optimum trait value . . .). In principle the joint model described above to jointly estimate N_e and s could also be extended to allow for parameters to vary in time (*e.g.*, at each interval between sampling dates). The likelihood framework described here can form the basis of such models.

Conclusion

Genome-wide assessment of the genetic diversity of a species is becoming more and more accessible in particular thanks to the development of cost effective genotyping and sequencing techniques. With the increase in genomic data, the evolution of genetic diversity in time becomes accessible either as a by-product of data accumulation or through dedicated projects in artificial populations (*e.g.*, experimental evolution) or natural settings (*e.g.*, ancient DNA sequencing). As genomic time series data become more widespread, their analysis requires dedicated methods that have good statistical and computational properties. In this study we have established a general statistical framework to this aim that we believe can contribute to further developments for the analysis of genetic time series.

LITERATURE CITED

Alnahhas, N., C. Berri, M. Chabault, P. Chartrin, M. Boulay *et al.*, 2016   Genetic parameters of white striping in relation to body weight, carcass composition, and meat quality traits in two broiler lines divergently selected for the ultimate pH of the pectoralis major muscle. BMC Genet. 17: 61. https://doi.org/10.1186/s12863-016-0369-2

Balding, D. J., and R. A. Nichols, 1995   A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. Genetica 96: 3–12. https://doi.org/10.1007/BF01441146

Bihan-Duval, E. L., C. Hennequet-Antier, C. Berri, S. A. Beauclercq, M. C. Bourin *et al.*, 2018   Identification of genomic regions and candidate genes for chicken meat ultimate pH by combined detection of selection signatures and QTL. BMC Genomics 19: 294. https://doi.org/10.1186/s12864-018-4690-1

Bollback, J. P., T. L. York, and R. Nielsen, 2008   Estimation of 2Nes from temporal allele frequency data. Genetics 179: 497–502. https://doi.org/10.1534/genetics.107.085019

Bonhomme, M., C. Chevalet, B. Servin, S. Boitard, J. Abdallah *et al.*, 2010   Detecting Selection in Population Trees: The Lewontin and Krakauer Test Extended. Genetics 186: 241–262. https://doi.org/10.1534/genetics.110.117275

Cappé, O., E. Moulines, and T. Ryden, 2005   *Inference in hidden Markov models. Springer series in statistics.* Springer science & Business Media, Berlin. https://doi.org/10.1007/0-387-28982-8

Cavalli-Sforza, L. L., I. Barrai, and A. W. F. Edwards, 1964   Analysis of Human Evolution Under Random Genetic Drift. Cold Spring Harb. Symp. Quant. Biol. 29: 9–20. https://doi.org/10.1101/SQB.1964.029.01.006

Doekes, H., R. Veerkamp, P. Bijma, S. Hiemstra, and J. Windig, 2018   Value of the Dutch Holstein Friesian germplasm collection to increase genetic variability and improve genetic merit. J. Dairy Sci. 101: 10022–10033. https://doi.org/10.3168/jds.2018-15217

Ewens, W. J., 2004   *Mathematical population genetics*, Vol. 27. Springer, New York. https://doi.org/10.1007/978-0-387-21822-9

Fages, A., K. Hanghøj, N. Khan, C. Gaunitz, A. Seguin-Orlando *et al.*, 2019   Tracking Five Millennia of Horse Management with Extensive Ancient Genome Time Series. Cell 177:1419–1435: e31. https://doi: 10.1016/j.cell.2019.03.049.

Fariello, M. I., S. Boitard, H. Naya, M. SanCristobal, and B. Servin, 2013   Detecting signatures of selection through haplotype differentiation among hierarchically structured population. Genetics 193: 929–941. https://doi.org/10.1534/genetics.112.147231

Ferrer-Admetlla, A., C. Leuenberger, J. D. Jensen, and D. Wegmann, 2016   An approximate markov model for the Wright–Fisher diffusion and its application to time series data. Genetics 203: 831–846. https://doi.org/10.1534/genetics.115.184598

Foll, M., H. Shim, and J. D. Jensen, 2015   WFABC: a W right–F isher ABC-based approach for inferring effective population sizes and selection coefficients from time-sampled data. Mol. Ecol. Resour. 15: 87–98. https://doi.org/10.1111/1755-0998.12280

Gompert, Z., 2016   Bayesian inference of selection in a heterogeneous environment from genetic time-series data. Mol. Ecol. 25: 121–134. https://doi.org/10.1111/mec.13323

He, Z., X. Dai, M. Beaumont, and F. Yu, 2019   An MCMC-based method for Bayesian inference of natural selection from time series DNA data across linked loci. bioRxiv.

Hubert, J.-N., T. Zerjal, and F. Hospital, 2018   Cancer-and behavior-related genes are targeted by selection in the Tasmanian devil (Sarcophilus harrisii). PLoS One 13: e0201838. https://doi.org/10.1371/journal.pone.0201838

Hui, T.-Y. J., and A. Burt, 2015   Estimating Effective Population Size from Temporally Spaced Samples with a Novel, Efficient Maximum-Likelihood Algorithm. Genetics 200: 285–293. https://doi.org/10.1534/genetics.115.174904

Iranmehr, A., A. Akbari, C. Schlötterer, and V. Bafna, 2017   Clear: Composition of Likelihoods for Evolve and Resequence Experiments. Genetics 206: 1011–1023. https://doi.org/10.1534/genetics.116.197566

Kimura, M., 1964   Diffusion Models in Population Genetics. J. Appl. Probab. 1: 177–232. https://doi.org/10.2307/3211856

Lacerda, M., and C. Seoighe, 2014   Population genetics inference for longitudinally-sampled mutants under strong selection. Genetics 198: 1237–1250. https://doi.org/10.1534/genetics.114.167957

Malaspinas, A.-S., 2015   Methods to characterize selective sweeps using time serial samples: an ancient DNA perspective. Mol. Ecol. 25: 24–41. https://doi.org/10.1111/mec.13492

Malaspinas, A.-S., O. Malaspinas, S. N. Evans, and M. Slatkin, 2012   Estimating allele age and selection coefficient from time-serial data. Genetics 192: 599–607. https://doi.org/10.1534/genetics.112.140939

Mathieson, I., I. Lazaridis, N. Rohland, S. Mallick, N. Patterson *et al.*, 2015 Genome-wide patterns of selection in 230 ancient Eurasians. Nature 528: 499–503. https://doi.org/10.1038/nature16152

Nicholson, G., A. V. Smith, F. Jónsson, Ó. Gústafsson, K. Stefánsson *et al.*, 2002 Assessing population differentiation and isolation from single-nucleotide polymorphism data. J. R. Stat. Soc. Series B Stat. Methodol. 64: 695–715. https://doi.org/10.1111/1467-9868.00357

Rabiner, L. R., 1989 A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE 77: 257–286. https://doi.org/10.1109/5.18626

Sackman, A. M., R. B. Harris, and J. D. Jensen, 2019 Inferring demography and selection in organisms characterized by skewed offspring distributions. Genetics 211: 1019–1028. https://doi.org/10.1534/genetics.118.301684

Schraiber, J. G., S. N. Evans, and M. Slatkin, 2016 Bayesian inference of natural selection from allele frequency time series. Genetics 203: 493–511. https://doi.org/10.1534/genetics.116.187278

Siren, J., P. Marttinen, and J. Corander, 2011 Reconstructing Population Histories from Single Nucleotide Polymorphism Data. Mol. Biol. Evol. 28: 673–683. https://doi.org/10.1093/molbev/msq236

Song, Y. S., and M. Steinrücken, 2012 A Simple Method for Finding Explicit Analytic Transition Densities of Diffusion Processes with General Diploid Selection. Genetics 190: 1117–1129. https://doi.org/10.1534/genetics.111.136929

Steinrücken, M., A. Bhaskar, and Y. S. Song, 2014 A novel spectral method for inferring general diploid selection from time series genetic data. Ann. Appl. Stat. 8: 2203–2222. https://doi.org/10.1214/14-AOAS764

Storey, J. D., A. J. Bass, , A. Dabney, , and D. Robinson, 2015. *qvalue: Q-value estimation for false discovery rate control.*

Tataru, P., T. Bataillon, and A. Hobolth, 2015 Inference under a wright-fisher model using an accurate beta approximation. Genetics 201: 1133–1141. https://doi.org/10.1534/genetics.115.179606

Tataru, P., M. Simonsen, T. Bataillon, and A. Hobolth, 2017 Statistical inference in the Wright-Fisher model using allele frequency data. Syst. Biol. 66: e30–e46. https://doi.org/10.1093/sysbio/syw056

Terhorst, J., C. Schlötterer, and Y. S. Song, 2015 Multi-locus Analysis of Genomic Time Series Data from Experimental Evolution. PLoS Genet. 11: e1005069. https://doi.org/10.1371/journal.pgen.1005069

Tobler, R., J. Hermisson, and C. Schlötterer, 2015 Parallel trait adaptation across opposing thermal environments in experimental Drosophila melanogaster populations. Evolution 69: 1745–1759. https://doi.org/10.1111/evo.12705

Vitti, J. J., S. R. Grossman, and P. C. Sabeti, 2013 Detecting natural selection in genomic data. Annu. Rev. Genet. 47: 97–120. https://doi.org/10.1146/annurev-genet-111212-133526

Zhao, L., X. Yue, and D. Waxman, 2013 Complete Numerical Solution of the Diffusion Equation of Random Genetic Drift. Genetics 194: 973–985. https://doi.org/10.1534/genetics.113.152017

Communicating editor: K. Thornton