MOLECULAR ECOLOGY

Molecular Ecology (2016) **25**, 121–134 doi: 10.1111/mec.13323

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

Bayesian inference of selection in a heterogeneous environment from genetic time-series data

ZACHARIAH GOMPERT
Department of Biology, Utah State University, Logan, UT 84322, USA

Abstract

Evolutionary geneticists have sought to characterize the causes and molecular targets of selection in natural populations for many years. Although this research programme has been somewhat successful, most statistical methods employed were designed to detect consistent, weak to moderate selection. In contrast, phenotypic studies in nature show that selection varies in time and that individual bouts of selection can be strong. Measurements of the genomic consequences of such fluctuating selection could help test and refine hypotheses concerning the causes of ecological specialization and the maintenance of genetic variation in populations. Herein, I proposed a Bayesian nonhomogeneous hidden Markov model to estimate effective population sizes and quantify variable selection in heterogeneous environments from genetic time-series data. The model is described and then evaluated using a series of simulated data, including cases where selection occurs on a trait with a simple or polygenic molecular basis. The proposed method accurately distinguished neutral loci from non-neutral loci under strong selection, but not from those under weak selection. Selection coefficients were accurately estimated when selection was constant or when the fitness values of genotypes varied linearly with the environment, but these estimates were less accurate when fitness was polygenic or the relationship between the environment and the fitness of genotypes was nonlinear. Past studies of temporal evolutionary dynamics in laboratory populations have been remarkably successful. The proposed method makes similar analyses of genetic time-series data from natural populations more feasible and thereby could help answer fundamental questions about the causes and consequences of evolution in the wild.

Keywords: Bayesian data analysis, fluctuating selection, genome scan, nonhomogeneous hidden Markov model, Wright-Fisher model

Received 4 May 2015; revision received 9 July 2015; accepted 10 July 2015

Introduction

There has been a long-standing interest among evolutionary biologists in identifying and analysing the alleles that underlie functional trait variation and adaptation (e.g. Lewontin 1974; Nielsen 2005; Martin & Orgogozo 2013; but see Rockman 2012 for a criticism of this research programme). We now know that natural selection occurs frequently in wild populations, but often varies in strength, direction and form over space and time (Endler 1986; Grant & Grant 2002; Siepielski *et al.* 2009; Bell 2010; Weese *et al.* 2010; Siepielski *et al.* 2013). Spatial variation in selection commonly occurs when populations occupy divergent environments and can result in local adaptation or speciation (Endler , 1977, 1980; Schluter 2000; Benkman *et al.* 2001; Nosil 2004; Yeaman & Jarvis 2006; Fournier-Level *et al.* 2011). Temporal variation in selection is also common and can be driven by changes or variability in the biotic or abiotic environment within or across years or generations (e.g. Schemske & Horvitz 1989; Grant & Grant 2002; Weese *et al.* 2010; Bergland *et al.* 2014). Thus, our search for the molecular targets of selection should include

Correspondence: Zachariah Gompert,
E-mail: zach.gompert@usu.edu

genetic loci subject to variable selection in heterogeneous environments. Indeed, quantifying selection on these loci could help resolve fundamental questions about the causes of ecological specialization (e.g. Agrawal *et al.* 2010; Poisot *et al.* 2011; Anderson *et al.* 2013; Gompert *et al.* 2015) and the maintenance of genetic variation in populations (e.g. Gillespie 1991; Hahn 2008; Leffler *et al.* 2012; Huang *et al.* 2014).

Numerous statistical methods have been developed to identify the molecular targets of selection and to estimate the strength of selection they experience. Many of these methods identify possible targets of selection based on exceptional patterns of genetic variation within or among populations (e.g. Lewontin & Krakauer 1973; Fu 1997; Sabeti *et al.* 2002; Beaumont & Balding 2004; Foll & Gaggiotti 2008; Gompert & Buerkle 2011b). Some also incorporate hypothesized environmental sources of heterogeneous selection or known trait-locus associations when testing for selection (Hancock *et al.* 2008; Coop *et al.* 2010; Günther & Coop 2013; Berg & Coop 2014). These methods capture the cumulative effects of selection over many generations, but inference of selection can be partially confounded by the unknown demographic history of the populations (Kelley *et al.* 2006; Excoffier *et al.* 2009).

Alternatively, patterns of allele frequency change within or across generations in natural or experimental populations can be analysed to infer the targets and intensity of selection (e.g. Burke *et al.* 2010; Illingworth & Mustonen 2011; Mathieson & McVean 2013; Bergland *et al.* 2014; Gompert *et al.* 2014a; Foll *et al.* 2015). Time-series methods are not confounded by past demographic events, but samples from many generations or multiple replicate populations are required to confidently distinguish the effects of selection from genetic drift (Lewontin & Krakauer 1973; Baldwin-Brown *et al.* 2014; Gompert *et al.* 2014a; Schlötterer *et al.* 2014). Several methods now exist that use hidden Markov models to infer the targets and strength of selection from such genetic time-series data (Bollback *et al.* 2008; Malaspinas *et al.* 2012; Mathieson & McVean 2013; Foll *et al.* 2015), but these approaches can only be used to quantify selection that varies in space, not in time. Alternative methods have been proposed to study temporal variation in selection, either based on the expected site-frequency spectrum (Mustonen & Lässig 2007; Huerta-Sanchez *et al.* 2008) or based on (cyclic) fluctuations in allele frequencies (Mueller *et al.* 1985; O'Hara 2005; Bergland *et al.* 2014). However, statistical methods to infer the strength of selection when selection varies in both space and time have not been developed.

Here I propose a new statistical method to fill this analytical gap. The approach attempts to quantify the selection experienced by many genetic loci based on patterns of allele frequency change across multiple populations and generations (Fig. 1; as in, e.g. Bollback *et al.* 2008; Malaspinas *et al.* 2012; Mathieson & McVean 2013; Foll *et al.* 2015). Importantly, I assume that the strength and direction of selection for each population and generation is a function of some measurable aspect of the biotic or abiotic environment. This approach is related to the method developed by Coop *et al.* (2010) to infer selection from allele frequency–environment correlations, but the proposed method instead makes inferences based on the association between allele frequency changes and the environment each generation. The variance effective population size is also allowed to vary each generation, which could be important for better modelling eco-evolutionary dynamics (e.g. Cortez & Ellner 2010; Cortez & Weitz 2014).

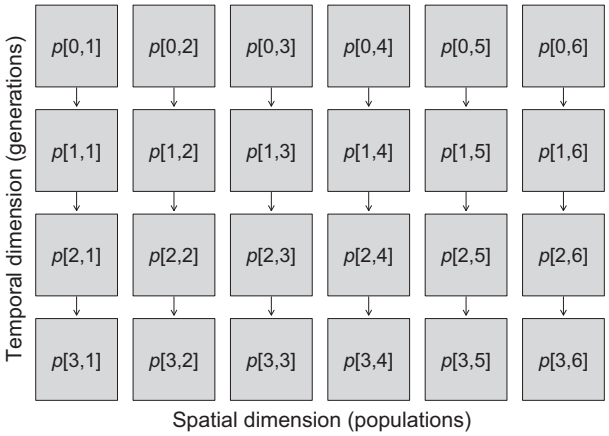

Fig. 1 This figure provides a graphical overview of the proposed model. Boxes denote the population allele frequencies in a population (column) and generation (row). Arrows show transitions between successive generations caused by genetic drift or selection. In this model, the populations are not connected by gene flow.

The proposed approach has two important limitations, which are shared with most other methods. First, no attempt is made to parse the direct versus indirect effects of selection on each locus. Direct selection occurs when variation at a locus affects fitness, whereas selection has an indirect effect on any locus where alleles are statistically associated with (i.e. correlated with) expected fitness (Gompert *et al.* 2014a). Such indirect effects of selection arise because of linkage disequilibrium, which will often but not always reflect patterns of physical linkage (as in models of linked selection and genetic hitchhiking; Maynard-Smith & Haigh 1974; Charlesworth *et al.* 1993; Gillespie 2000; Hahn 2008). While methods to parse the direct versus indirect effects of phenotypic selection are well developed (Lande & Arnold 1983), similar approaches cannot be used here, as the number of genetic loci will often be large relative

to the number of observations. Thus, the proposed method simply estimates the combined influence of direct and indirect selection on each locus. Second and perhaps more important, as is well known, a one-to-one mapping of process and pattern does not exist in evolutionary genetics, and thus, many different processes could give rise to the same patterns of allele frequency change. Consequently, inferred selection coefficients must be interpreted cautiously. Nonzero estimates of selection are consistent with the hypothesis of selection on a given locus, but should not be viewed as strong support for this hypothesis without other evidence.

Below I first describe the proposed method, including its theoretical basis and the computational algorithms used to estimate model parameters. I then evaluate the performance of the method with a series of simulated data sets. Simulations are also used to highlight the overall signal in the data independent from the proposed analytical procedure. Based on the results, the proposed method is able to distinguish between neutral and non-neutral loci under a variety of conditions (all involving strong selection). Accurate estimates of the intensity of selection are obtained when fitness has a simple genetic basis and selection is constant or linearly related to the environment, but these are less accurate when fitness is polygenic or nonlinearly associated with environmental variation. The manuscript concludes with a discussion of possible refinements of the proposed method.

Methods

Model

The proposed statistical method is motivated by a diploid Wright-Fisher model with selection (as described in Ewens 2004). Thus, discrete, nonoverlapping generations are assumed. Selection and the variance effective population size (N) are both allowed to vary among populations and generations. Evolutionary dynamics are modelled for a set of L unlinked loci in J populations observed over K generations. I assume that each locus has two alleles denoted a_1 and a_2. Let p_{ijk} be the frequency of the a_1 allele at locus i and population j in generation k. Similarly, let $y_{ijk} \in \{0, \ldots, 2n_{ijk}\}$ be the number of a_1 alleles observed in a sample of n_{ijk} diploid individuals. Thus, the sample allele counts follow a binomial distribution such that,

$$y_{ijk} \sim \text{binomial}(p_{ijk}, 2n_{ijk}). \quad (1)$$

Denote the relative fitness of each genotype as $w_{ijk}^{11} = 1 + 2s_{ijk}$, $w_{ijk}^{12} = 1 + 2hs_{ijk}$ and $w_{ijk}^{22} = 1$, for the a_1a_1, a_1a_2 and a_2a_2 genotypes, respectively, where s_{ijk} is the selection coefficient and h is the heterozygote effect. Herein, I assume $h = \frac{1}{2}$ for simplicity (no dominance), but the proposed method could be modified to allow for $h \neq \frac{1}{2}$. Under this model, the expected value of the population allele frequency in the next generation is,

$$E[p_{ijk+1}|p_{ijk}, \mathbf{w_{ijk}}] = p_{ijk} + \frac{w_{ijk}^{11}p_{ijk}^2 + w_{ijk}^{12}p_{ijk}(1-p_{ijk}) - p_{ijk}\bar{w_{ijk}}}{\bar{w_{ijk}}}. \quad (2)$$

Here $\bar{w_{ijk}}$ is the mean fitness of the population. The population allele frequencies each generation are Markovian random variables with the transition probabilities given by,

$$\Pr(p_{ijk+1}|p_{ijk}, s_{ijk}, N_{jk}) \sim \text{binomial}(E[p_{ijk+1}|p_{ijk}, \mathbf{w_{ijk}}], 2N_{jk}). \quad (3)$$

Because allele frequencies are unobserved and selection and effective population size vary over time, this is a nonhomogeneous hidden Markov model.

My proposed method aims to estimate the parameters from this theoretical model using Bayesian inference. Similar to Foll *et al.* (2015), effective population sizes are estimated first, and then, population allele frequencies and selection coefficients are inferred conditional on these estimates. Specifically, a point estimate of the effective population size for each population and generation is obtained using the Jorde & Ryman's (2007) unbiased estimator,

$$Fs'_{jk} = \frac{1}{2}\frac{\sum_i Fs_{ijk}\left(1 - \frac{1}{4\tilde{n_{ijk}}}\right) - \frac{1}{\tilde{n_{ijk}}}}{\sum_i \left(1 + \frac{Fs_{ijk}}{4}\right)\left(1 - \frac{1}{2n_{ijk+1}}\right)} \quad (4)$$

Here $Fs_{ijk} = \frac{(\hat{p_{ijk}} - \hat{p_{ijk+1}})^2}{\hat{\pi_{ijk}}(1-\hat{\pi_{ijk}})}$ is a measure of allele frequency change between generations that must be corrected to generate an unbiased estimator of the variance effective population size, Fs' (this correction assumes individuals were sampled destructively; Jorde & Ryman 2007). $\tilde{n_{ijk}}$ is the harmonic mean sample size for generations k and k+1, $\hat{p_{ijk}}$ is the sample allele frequency in generation k, and $\hat{\pi_{ijk}}$ is the mean sample allele frequency for generations k and k+1. A genome-average estimate of the variance effective population size is then obtained as $N_{jk} = \frac{1}{Fs'_{jk}}$. Next, following Foll *et al.* (2015), a Bayesian bootstrap approach is used to generate samples from the posterior probability distribution of N_{jk} conditional on the sample allele frequencies (a Bayesian bootstrap is operationally similar to a conventional nonparametric bootstrap, but is instead used to simulate from the posterior distribution of a parameter (Rubin 1981). This procedure ignores selection, which should be fine if most genetic loci are neutral. Even if selection affects most of the sampled genetic loci, one should obtain an accurate estimate of the variance effective population size (which will be depressed because of selection) as long as the

genetic loci are a random sample. However, if specific genetic loci likely under selection are targeted for analysis, a different set of loci or an independent estimate of effective population size should be used (this is an option in the computer software).

Markov chain Monte Carlo can then used to sample the joint posterior distribution for the population allele frequencies and selection coefficients (Gamerman & Hedibert 2006). First, a linear model is specified for the selection experienced by each locus in each generation and population, $s_{ijk} = \alpha_i + \beta_i x_{jk}$. Here x_{jk} is a centred environmental covariate measured at population j in generation k, such as total summer precipitation or the abundance of a resource or competitor. α_i and β_i are regression coefficients that describe the average selection experienced by locus i (α_i) and the relationship between the environmental covariate and the selection experienced by locus i (β_i). Gaussian priors are placed on the regression coefficients: $\alpha_i \sim \text{normal}(0, \sigma_\alpha)$ and $\beta_i \sim \text{normal}(0, \sigma_\beta)$. Small values can be used for the standard deviations to reflect the prior expectation that most loci do not experience substantial selection in a given generation (a related approach is used in Bayesian genomewide association mapping; Stephens & Balding 2009). These regression coefficients are estimated using the Metropolis algorithm (Metropolis *et al.* 1953). While it would be possible to consider multiple covariates and their interactions in a single analysis, the current implementation of the method is restricted to the analysis of one covariate at a time.

Direct Gibbs sampling is used to estimate the population allele frequencies. This is computationally efficient, but requires a few approximations. The posterior distribution for the allele frequency p_{ijk} is proportional to the product of three probability functions,

$$\begin{aligned} &Pr(p_{ijk}|y_{ijk}, p_{ij(k-1,k+1)}, 2n_{ijk}, 2N_{ij(k-1,k)}, s_{ij(k-1,k)}) \propto \\ &\quad Pr(y_{ijk}|p_{ijk}, 2n_{ijk}) Pr(p_{ijk}|p_{ijk-1}, s_{ijk-1}, 2N_{ijk-1}) \\ &\quad Pr(p_{ijk+1}|p_{ijk}, s_{ijk}, 2N_{ijk}). \end{aligned} \tag{5}$$

The first term on the right side of this equation is the binomial probability mass function specified in eqn 1. The second term denotes the transition probability from the allele frequency in the previous generation to that in the focal generation (k). This is often modelled using a binomial probability mass function for the population allele counts. Instead, I approximate this with a beta probability density function for the population allele frequencies, $p_{ijk} \sim \text{beta}\big(E[p_{ijk}]2N_{jk}, \big(1 - E[p_{ijk}]\big)2N_{jk}\big)$, because of its conjugate relationship with the binomial likelihood (i.e. eqn 1). Here $E[p_{ijk}]$ is the expected allele frequency, which is a function of the allele frequency in the previous generation and the selection coefficient (see eqn 2). While a binomial probability function could be used to calculate the transition probability from the allele frequency at generation k to k+1, the binomial parameter would be $E[p_{ijk+1}]$ rather than p_{ijk}. This means that the analytical solution to eqn 5 could only be computed when the selection coefficient (s_{ijk}) is zero. Thus, $2N_{ijk+1}\ p_{ijk+1} \sim \text{binomial}(E[p_{ijk+1}], 2N_{ijk})$ is instead approximated by $2N_{ijk+1}\ (p_{ijk+1} - (E[p_{ijk+1}] - p_{ijk})) \sim \text{binomial}(p_{ijk}, 2N_{ijk})$. These probability functions have the same means, but differ slightly in their variances, particularly when allele frequencies are near 0 or 1. Note that this term is dropped from the Gibbs update for the final generation and that an uninformative beta prior is used for the first generation. With these approximations, samples can be drawn directly from the full conditional distribution for each p_{ijk}, which is beta$(\theta_\alpha, \theta_\beta)$, where

$$\theta_\alpha = y_{ijk} + E[p_{ijk}]2N_{ijk-1} + \big(p_{ijk+1} - (E[p_{ijk+1}] - p_{ijk})\big)2N_{ijk} \tag{6}$$

and

$$\begin{aligned} \theta_\beta = n_{ijk} - y_{ijk} + (1 - E[p_{ijk}])2N_{ijk-1} + \\ (1 - \big(p_{ijk+1} - \big(E[p_{ijk+1}] - p_{ijk}\big)\big))2N_{ijk}. \end{aligned} \tag{7}$$

A C++ program (SPATPG) has been developed to estimate the parameters in this model as described above. The computer program uses the GNU Scientific Library (Galassi *et al.* 2009) for most computations and HDF5 for efficient output (The HDF5 Group 2010). This program is hosted on SourceForge at https://sourceforge.net/projects/spatpg/ and is available from https://gompertlab.wordpress.com/software/. A manual is included with the program which explains how the program is used and provides an example analysis.

Interpretation of Bayesian parameter estimates

Bayesian analyses have become common in molecular ecology and evolutionary genetics (e.g. Pritchard *et al.* 2000; Beaumont & Rannala 2004; Csillery *et al.* 2010; Gompert & Buerkle 2011a), but there are a few aspects of Bayesian inference that warrant discussion, particularly in the context of estimates of selection from genetic data. The outcome of a Bayesian analysis is the joint posterior probability distribution of the model parameters (Gelman *et al.* 2004), which here include the variance effective population sizes, the coefficients from the linear model and the selection coefficients. Marginal posterior distributions for individual parameters can be summarized by measures of their central tendency (i.e. the mean, median or mode of the posterior, which often serve as a point estimates) and their dispersion (e.g. quantiles of the posterior that delineate various credible intervals). With the proposed model, posteriors for the regression coefficients should be most informative for identifying loci with the greatest evidence of selection as these parameters

combine information across populations and generations, whereas the selection coefficients provide actual estimates of the selection experienced by a locus in each population and generation. When presenting results in this study (see below), I treat 95% equal-tail probability intervals (ETPIs) for either regression coefficient that do not include 0 as significant evidence that a genetic locus was affected by selection. More (or less) extreme quantiles of the posterior could be used in applications where more (or less) confidence that selection occurred is desired. In general, evidence for selection should be viewed as a continuous quantity, with those loci with the greatest proportion of their posterior density for the selection coefficients (or α and β) showing the greatest evidence of selection, and loci should be prioritized for any follow-up analyses accordingly.

Bayesian parameter inference does not require a correction for multiple tests (Kruschke 2010). However, the prior on the regression coefficients can serve a somewhat analogous, albeit philosophically unrelated, function. In particular, a small standard deviation (small relative to the standard deviation of the covariate) can be used for the prior on α or β to indicate that most loci are expected to experience modest to weak selection, or no selection at all. This will result in some shrinkage of parameter estimates towards zero and thus should be viewed as a conservative procedure relative to the use of an uninformative prior with a large standard deviation (e.g. Guan & Stephens 2011; Zhou & Stephens 2012). As an alternative to Bayesian parameter inference, one could test for selection by computing Bayes factors contrasting a hypothesis of selection with a null hypothesis of no selection. While this approach has been popular in Bayesian population genetics (e.g. Foll & Gaggiotti 2008; Coop *et al.* 2010), I avoid it here for two reasons. First, it encourages categorical thinking and converts what is inherently a continuous variable (the strength of selection on each locus) into a binary null hypothesis test ($s = 0$ versus $s \neq 0$). Second, and perhaps of more practical importance, Bayes factors are very sensitive to one's choice of priors (Gelman *et al.* 2004; Kruschke 2010). This is because the parameter values for the alternative hypothesis are weighted by their prior (not posterior) probabilities. Thus, if selection is weak, one might favour the alternative hypothesis when a prior that assumes mostly weak selection is specified, but not when an uninformative prior that allows for stronger selection is used.

Analysis of simulated data

Data sets were simulated to evaluate the performance of the statistical method and computer software. An initial series of data sets were simulated in R under a Wright-Fisher model with 10 populations, 10 generations of evolution and a constant population size of 1000 diploid individuals. Genomes consisted of 1000 unlinked SNPs with starting allele frequencies drawn from a uniform distribution bounded by 0.05 and 0.95. One hundred individuals were sampled from each population each generation. Ten replicate data sets were simulated under each of four different conditions: (i) pure drift, (ii) constant selection, (iii) linear fluctuating selection and (iv) nonlinear fluctuating selection. No selection occurred in the pure drift simulations. In the other simulations, fitness was determined by the genotype at one of the 1000 SNPs. In the constant selection simulations, the relative fitness values of the genotypes were $w^{11} = 0.8$, $w^{12} = 0.9$ and $w^{22} = 1$, and thus, the selection coefficient (s) was equal to -0.1 for all populations and generations. Fitness was determined by the value of the environment in each population and generation in the fluctuating selection simulations. Environmental values were sampled from a standard normal distribution and were uncorrelated across space and time. The fitness of each genotype was either a linear or nonlinear function of the environmental value (Fig. 2). Note that the latter is a violation of the linear model assumed by the analytical method. All three of these sets of non-neutral simulations assumed strong selection. Strong, but episodic and variable selection has often been documented in phenotypic studies in the wild (e.g. Grant & Grant 2002; Nosil 2004; Siepielski *et al.* 2009), and my main intent for the proposed method is to detect selection under these conditions. With that said, I conducted a second series of simulations to determine whether the method could also detect weak selection and to evaluate the sensitivity of the results with strong selection to sample size. Ten replicate data sets were simulated under both weak constant selection ($s = -0.01$) and weak fluctuating selection ($\beta = 0.01$ rather than 0.1 as shown in Fig. 2). The effect of sample size was investigated for strong linear fluctuating selection by analysing samples of 25 or 50 individuals (rather than 100) for each of the ten replicate data sets.

Selection on complex or quantitative phenotypic characters is common in nature (Endler 1986) and thus QUANTINEMO was used to simulate a third series of data sets under more realistic conditions involving selection on a polygenic quantitative trait (Neuenschwander *et al.* 2008). QUANTINEMO allows for genetically explicit, individual-based simulations with discrete generations and fertility selection. A series of five data sets was simulated under the following conditions. Ten populations were followed for 20 generations, although data were only sampled from generations 10-20. Population cen-

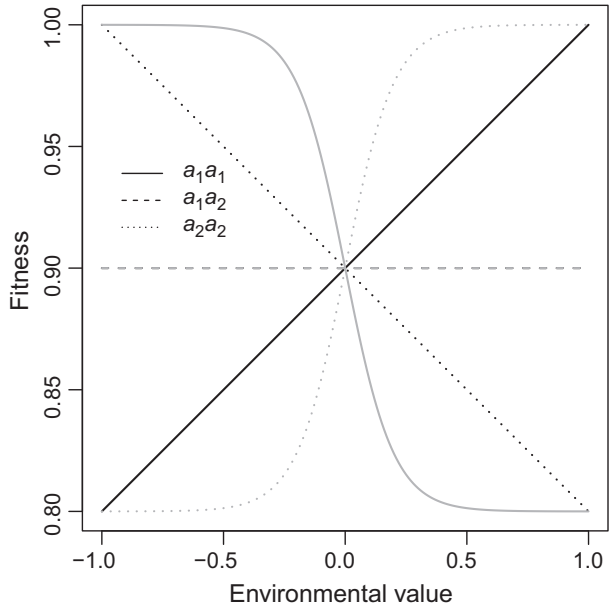

Fig. 2 Lines in this plot show the fitness functions used for the Wright Fisher simulations with linear (black) and nonlinear (grey) fluctuating selection. A logistic function with a slope of ±10 was used to model nonlinear selection. No genotype had a net fitness advantage, but rather alternative homozygotes performed better in different environments. Heterozygotes had a constant and intermediate fitness in all Wright Fisher simulations and across all environments.

sus sizes were set to 1000 diploid individuals, and limited dispersal (m = 0.001) was allowed among populations; an island model was assumed. Fitness was a function of the state of the environment and each individual's trait value. The trait value was determined by 10 bi-allelic SNPs with equal, additive effects (a = 0.2). Stabilizing selection was modelled using a Gaussian fitness function where the optimal phenotypic value was given by the state of the environment. The selection intensity (SD of the fitness function) was set to 1.5. A two-step procedure was used to sample the environmental state for each population and generation, and the optimal phenotypic value was then equated with the environmental state (i.e. the optimal phenotype when the environmental state was 0.3, was 0.3). In the first step, the average environmental state for each population was sampled from a standard normal distribution. Then, values for each generation were sampled from a normal distribution with this mean and a standard deviation of 1.

Genomes comprised 10 chromosomes with 100 evenly spaced neutral bi-allelic SNPs per chromosome. Two of the functional SNPs were placed on each of the first five chromosomes at 250 and 750 cMs. Allele frequencies for the initial generation were sampled from uniform distributions bounded by 0.1 and 0.9. A mutation rate of 10^{-4} was assumed. Genetic data were obtained from a sample of 100 individuals in generations 10 through 20, and these data were used for the analysis with the Bayesian method presented in this study. In addition, the population allele frequencies were retained and the selection experienced by each locus in each population and generation was quantified (this included the direct and indirect effects of selection). I determined the total selection experienced at each locus by calculating the slope of the best fit line obtained from regressing expected fitness (based on each individual's multilocus genotype at all non-neutral SNPs) on the genotypic data for each locus.

I analysed all 85 simulated data sets using the analytical method and computer software described above. Lower and upper bounds for the uniform prior on the effective population sizes were set to 20 and 4000, respectively. A standard deviation of 0.1 was used for the priors on the regression coefficients, α and β. Three replicate analyses were conducted for each data set. Each Markov chain Monte Carlo (MCMC) replicate consisted of 25 000 iterations with a 10 000 iteration burn-in and a thinning interval of 15 steps. Effective sample sizes and the Gelman–Rubin potential scale reduction factor were calculated for all regression coefficients to assess MCMC performance. I observed good overall mixing with an average effective sample size of 2723 and 2239 for α and β for the Wright Fisher simulations, respectively (2838 and 2826 for the QUANTINEMO simulations). Moreover, the Gelman–Rubin diagnostic indicated that the MCMC algorithms had likely converged to their stationary distributions (Wright Fisher simulations, α: mean = 1.002, maximum = 1.060; β: mean = 1.008, maximum = 1.110; QUANTINEMO simulations, α: mean = 1.001, maximum = 1.030; β: mean = 1.004, maximum = 1.050).

Results

Variance effective population size estimates obtained from the data sets simulated under the Wright Fisher model were accurate (unbiased), but not very precise (similar results were found by Jorde & Ryman 2007). In particular, estimates of N varied among populations and generations, but the average estimates were close to the true value of 1000: pure drift $\bar{N}$ = 1015.1 (SD 410.8), constant selection $\bar{N}$ = 1019.3 (SD 407.1), fluctuating selection (linear) $\bar{N}$ = 1019.2 (SD 405.8) and fluctuating selection (nonlinear) $\bar{N}$ = 1007.9 (SD 414.0). Moreover, the posterior probability distributions of the effective population sizes were wide, indicating that there was substantial uncertainty about the values of these parameters (the mean width of the 95% equal-tail probability intervals [ETPIs] was 3585.4).

Table 1 Number and proportion (in parentheses) of genetic loci where 95% ETPIs did not include 0 for each parameter. Results are shown for all replicates and for neutral and non-neutral loci (unambiguous true positives are shown in bold font)

Simulation	α neutral	α non-neutral	β neutral	β non-neutral
Neutral	37 (0.004)	NA	142 (0.014)	NA
Constant selection	26 (0.003)	**10 (1.000)**	130 (0.013)	0 (0.000)
Fluctuating selection (linear)	26 (0.003)	0 (0.000)	128 (0.013)	**10 (1.000)**
Fluctuating selection (nonlinear)	33 (0.003)	1 (0.100)	137 (0.014)	**10 (1.000)**
Weak constant selection	36 (0.004)	**0 (0.000)**	145 (0.015)	0 (0.000)
Weak fluctuating selection	40 (0.004)	0 (0.000)	123 (0.012)	**0 (0.000)**
Fluctuating selection (linear, 2N = 50)	5 (0.001)	0 (0.000)	110 (0.011)	**4 (0.400)**
Fluctuating selection (linear, 2N = 100)	0 (0.000)	0 (0.000)	140 (0.014)	**9 (0.900)**
Fluctuating selection (QUANTINEMO)	22 (0.004)	14 (0.280)	45 (0.009)	**46 (0.920)**

Posterior estimates of the regression coefficients, α (constant selection) or β (fluctuating selection), were significantly different than 0 (i.e. the 95% ETPIs did not include 0) for the non-neutral locus in all 10 replicates for the Wright Fisher simulated data sets with strong selection (Table 1). Very low false-positive rates were observed, with regression coefficients differing significantly from 0 for considerably fewer than 5% of the neutral loci (Table 1). Moreover, selection coefficient estimates were similar to the true values, particularly for data sets simulated with constant or linear fluctuating selection (Fig. 3). Along these lines, the root mean square error [RMSE] for s was low (constant selection: RMSE = 0.0206; fluctuating selection (linear): RMSE = 0.0212; fluctuating selection (nonlinear): RMSE = 0.0215). The mean estimate of s at the non-neutral locus for the 10 replicate constant selection data sets was -0.093, and the true value of -0.1 was included in the 95% ETPIs in each case. Likewise, estimates of s were highly correlated with the true value under linear fluctuating selection ($r = 0.983$, $P < 0.0001$) and the true value was included in the 95% ETPIs 96.6% of the time. While the correlation between the true and estimated values of s was also high for the nonlinear fluctuating selection data sets ($r = 0.841$, $P < 0.0001$), the true value was only included in the 95% ETPIs 52.2% of the time.

In contrast, posterior distributions for the regression coefficients for the non-neutral locus in each of the weak selection simulated data sets included 0 (i.e. the 95% ETPIs contained 0; Table 1). However, there was a trend for the point estimates to be in the correct direction (17 of 20 weak selection data sets), and the false-positive rate remained quite low (Table 1). Smaller samples sizes also reduced the method's power to detect the non-neutral locus in each replicate, but to a lesser extent. True-positive rates were 90% and 40% for samples sizes of 50 and 25 diploid individuals per generation per population, respectively (Table 1).

Prior to conducting formal analyses of selection with the proposed method on the QUANTINEMO simulations, I examined the direct output from these simulations to obtain a better qualitative and quantitative understanding of how selection in a heterogeneous environment affects allele frequencies (results are shown for one of the five simulated data sets; Fig. 4). While most loci were affected by selection, the direct effect of selection on non-neutral loci was usually much stronger than indirect effect of selection on neutral loci (e.g. non-neutral mean $|s| = 0.0715$, SD = 0.0416; neutral mean $|s| = 0.0086$, SD = 0.0079; Fig. 4a). Despite substantial differences in the strength of selection experienced by neutral and non-neutral loci, these loci often exhibited fairly similar magnitudes of allele frequency change (e.g. non-neutral mean change = 0.0145, SD = 0.0132; neutral mean change = 0.0082, SD = 0.0065; Fig. 4b). However, patterns of allele frequency change at the non-neutral SNPs were predicted by the state of the environment, while those at neutral SNPs were not (Fig. 4c). Thus, in these simulations, fluctuating selection on a quantitative trait had a much greater affect on the association between environmental variability and patterns of evolutionary change, than on the absolute magnitude of change that occurred.

Accurate but imprecise estimates of variance effective population sizes were obtained for the QUANTINEMO simulated data sets (census size = 1000, $\bar{N} = 949.6$, SD = 429.7, mean width of the 95% ETPIs = 3277.5). Estimates of β differed significantly from 0 for nearly all (92%) of the non-neutral loci in the QUANTINEMO simulations, while estimates of α differed from 0 for a smaller proportion (28%) of non-neutral loci. This high true-positive rate was coupled with a very low false-positive rate (Table 1). With that said, posterior estimates of selection coefficients were only modestly correlated with their true values ($r = 0.153$, $P < 0.0001$), although a much higher correlation was observed when only the non-neutral loci were considered ($r = 0.677$, $P < 0.0001$; Figs 5 and 6). Nonetheless, the RMSE for the selection coefficients was low (0.0197) and the true value of s was included in the 95% ETPIs 95.0% of the

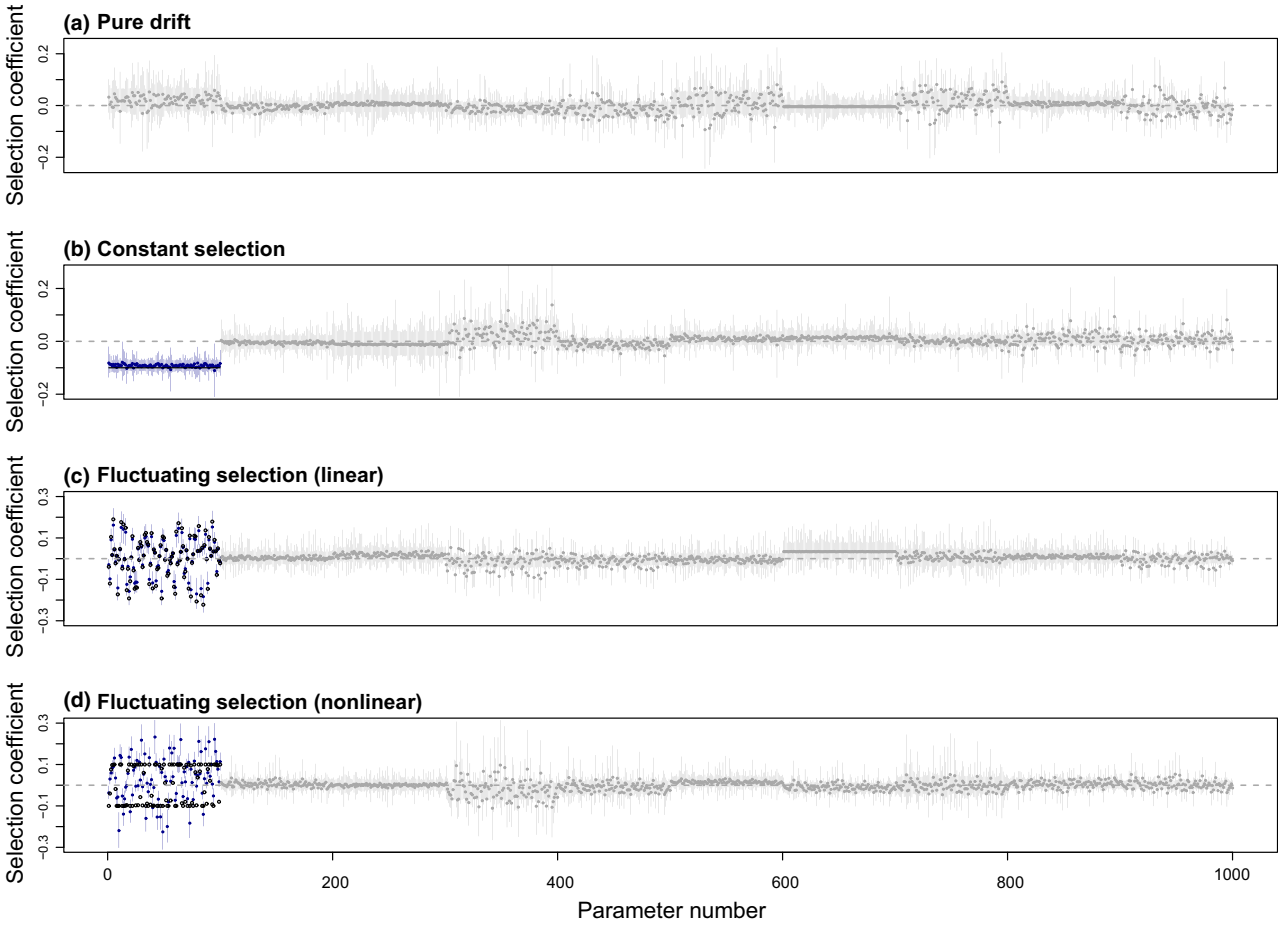

Fig. 3 These plots summarize the posterior probability distributions of selection coefficients for different simulated data sets. Grey (neutral loci) or blue (black in grey-scale image; non-neutral loci) points and lines indicate the posterior median and 95% ETPIs for the selection coefficients for specific loci, populations and generations. All 100 population-by-generation combinations are shown for the non-neutral locus and for a subset of nine representative neutral loci. The true value of s is indicated by a black line (constant selection) or black points (fluctuating selection).

time (this was lower when only considering the non-neutral SNPs; 95% ETPI coverage = 46.0%).

Discussion

A new Bayesian method to detect variable selection from genetic time-series data was proposed and evaluated. The method showed remarkable promise for distinguishing between neutral and non-neutral genetic loci when selection was strong, but not when selection was weak (Table 1). Thus, while other approaches will likely be needed to detect weak selection, the proposed method appears suitable for detecting strong but variable selection, which has frequently been documented in phenotypic studies (Siepielski *et al.* 2009). With that said, even strong phenotypic selection could result in weak selection on individual genes if the selected trait (s) has a polygenic basis. This should make selection more difficult to detect at the molecular level. However, results from the analyses presented here show that strong selection on a polygenic trait can be detected, particularly when all of the functional variants have similar phenotypic effects. Additional information regarding genotype–trait associations could be used to refine the method and improve power when dealing with polygenic phenotypes (e.g. Berg & Coop 2014).

Whereas the proposed method exhibited good and consistent performance with respect to discriminating between neutral and non-neutral genetic loci when selection was strong, the accuracy of the estimated selection coefficients depended on the simulation conditions (Figs 3, 5 and 6). In particular, very accurate estimates of selection coefficients were obtained when fitness was determined by a single gene, and when selection was linearly related to the environment or constant, but not when fitness was polygenic or when selection exhibited a nonlinear relationship with the environment. A nonlinear relationship between the

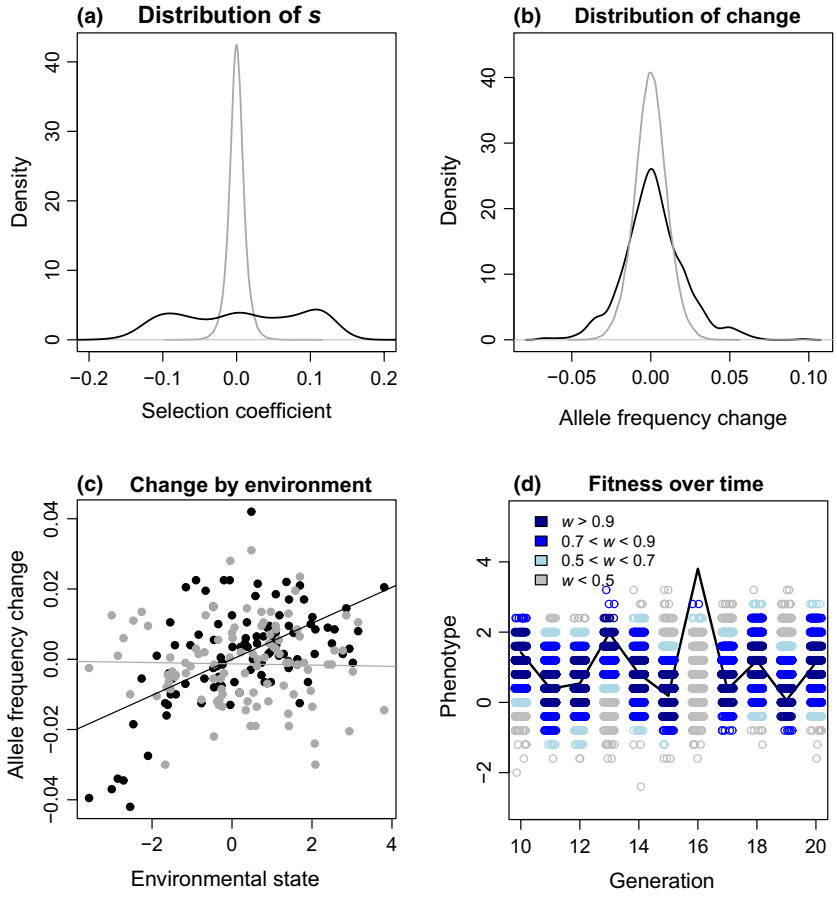

Fig. 4 These plots summarize patterns of selection and allele frequency change for one of the five QUANTINEMO simulations. (a) Density plots depict the distribution of the actual selection experienced by neutral (grey) and non-neutral (black) loci. These selection coefficients capture the effects of direct and indirect selection. (b) Density plots depict the distribution of allele frequency change between generations at neutral (grey) and non-neutral (black) loci. (c) A scatterplot shows the relationship between the environmental state and allele frequency change at neutral (grey) and non-neutral (black) loci. The solid lines show the inferred relationship from linear regression. (d) Points indicate the expected fitness of each individual each generation in a single population. Points were jittered along the x-axis to improve visualization. A line shows the optimal phenotypic value each generation based on the state of the environment.

environment and selection violates the linear model used in the method and should cause some estimates of s to be too high and others to be too low (as seen in Fig. 3d). Unfortunately, such nonlinearities are probably common in nature (this problem and possible solutions are discussed more in the next section).

When stabilizing selection occurs on a polygenic trait, selection on each locus depends on the genetic composition of the population (i.e. the allele frequencies at the other loci and patterns of linkage disequilibrium; Lehman & Joyce 1993; Barton & Keightley 2002). For example, an allele that increases a trait value would be favoured by selection if allele frequencies at other loci yield an expected phenotype less than the optimum, but selected against in the same environment if allele frequencies at other loci result in an expected phenotype greater than the optimum (Brodie 2000). Such complexities are not captured by the model and could explain the reduced performance of the method when applied to the data sets simulated with QUANTINEMO (e.g. Fig. 6). But, even under these conditions, inferred and true selection coefficients were correlated and estimates were of the correct order of magnitude. Thus, the method should be useful in natural systems.

The analytical and computational performance of the proposed method is similar to that of other methods that have been used to estimate selection coefficients from genetic time series. In particular, Foll *et al.* (2015) recently compared likelihood-based and approximate Bayesian computation methods for estimating selection from genetic time-series data. The methods considered were also derived from the Wright Fisher model described here, but assumed constant selection through time (Malaspinas *et al.* 2012; Foll *et al.* 2015). Power and error for these methods when applied to cases of strong selection were quite similar to that reported here (see, e.g., table 2 in Foll *et al.* 2015). Thus, the main advantage of the proposed method relative to these is that it allows the strength and direction of selection to vary from generation to generation.

Markov chain Monte Carlo analysis of the simulated data sets took approximately 3 hours each running on a single processor on a standard computer. The run time should increase linearly with the number of loci, populations and generations sampled. Once effective population sizes have been estimated, the population allele frequencies and selection coefficients for each locus

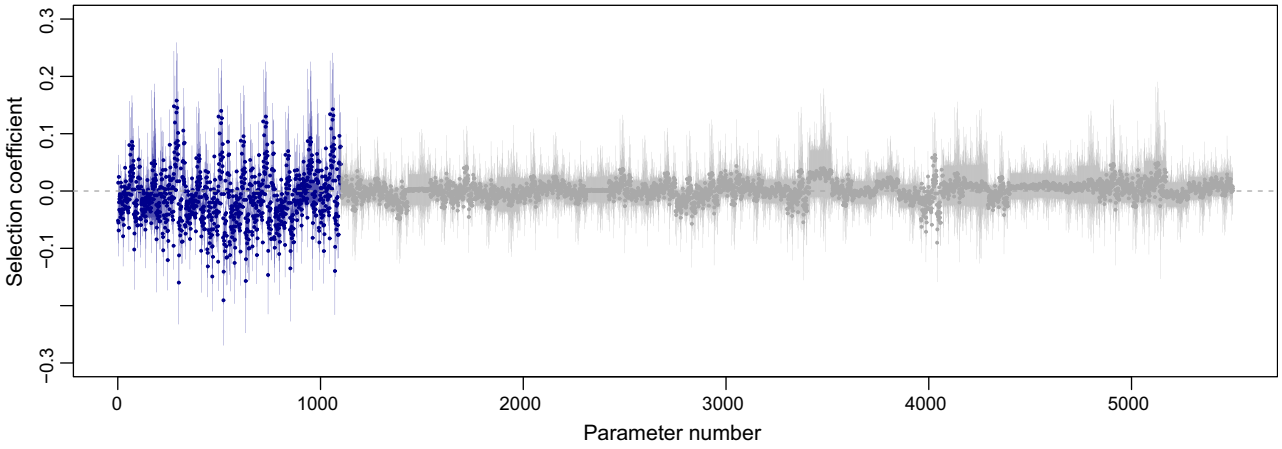

Fig. 5 The plot summarize the posterior probability distributions for selection coefficients for one of the data sets generated with QUANTINEMO. Grey (neutral loci) or blue (black in grey-scale image; non-neutral loci) points and lines indicate the posterior median and 95% ETPIs for the selection coefficients for specific loci, populations and generations. All 110 population-by-generation combinations are shown for the 10 non-neutral loci and for a subset of 40 representative neutral loci.

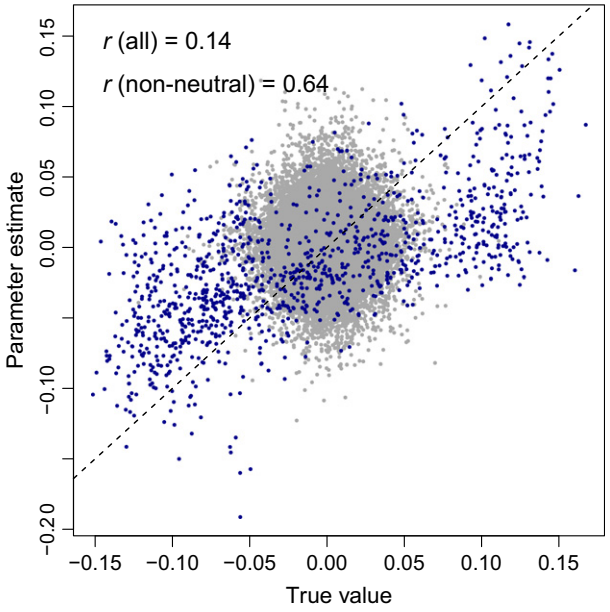

Fig. 6 This scatterplot shows the relationship between the true and estimated selection experienced by each locus for one of the five fluctuating selection data sets simulated with QUANTINEMO. This includes direct selection and indirect selection caused by statistical associations between neutral and non-neutral loci. Neutral loci are shown in grey and non-neutral loci are shown in blue (black in grey-scale image). A dashed black line shows the expected one-to-one relationship. Correlation coefficients between true and estimated values for this data set are reported for all loci and for the subset of non-neutral loci.

could be estimated independently. While I have not yet taken advantage of this, this means that the program could be run in parallel and thus could easily be scaled up to analyse genomic data (i.e. millions of SNPs) in a reasonable amount of time. I intend to implement this option in a forthcoming version of the computer program.

Limiting assumptions and possible refinements

Like most statistical methods used to make inferences from population genetic data, the proposed method requires various assumptions that limit its utility and that should be noted. Perhaps most important, the method assumes selection has been a linear function of a single environmental variable. This simple model could capture nature sufficiently well in some systems, but will likely fail in others. A nonlinear model, such as a logistic function, could be used instead, but it is not clear that this would be better overall (i.e. it would likely be better in some cases and worse in others). While this issue could potentially be solved by specifying a flexible (highly parameterized) model relating selection to the environment (e.g. a spline function), this would probably come at a sizable computational cost. Likewise, it would be straightforward to model selection using multiple environmental variables, and this would certainly better approximate many real systems (e.g. Benkman & Parchman 2009; Orsini *et al.* 2012; Egea-Serrano *et al.* 2014). But this too would likely degrade MCMC mixing and thus complicate the analysis. With that said, multiple features of the environment could readily be considered by analysing a composite variable from an ordination method (e.g. principal component analysis) or by contrasting estimates of selection obtained when fitting models with different environmental variables. An approach analogous to the latter was successfully used by Hancock *et al.* (2008) to

distinguish among more and less important ecological sources of selection in humans. Finally, while I have assumed that the fitness of the heterozygote is exactly intermediate, this assumption is not necessary and could be relaxed.

The proposed model also assumes that populations evolve independently. In other words, gene flow among populations is ignored. While even low rates of gene flow can alter long-term evolutionary dynamics or equilibrium conditions (e.g. Slatkin 1987), limited gene flow will have less of an affect on allele frequency change between generations. In particular, as long as the product of the migration rate and the allele frequency difference between migrant and nonmigrant individuals is substantially less than the allele frequency change expected by drift or selection, gene flow can probably be ignored. Note that the QUANTINEMO data sets included gene flow ($m = 0.001$) among populations, and the proposed method still performed well. Nonetheless, methods do exist to infer gene flow and (constant) selection from genetic time series (Malaspinas *et al.* 2012), and gene flow could be included in the proposed analytical framework. Similarly, the proposed method assumes each locus evolves independently; this assumption would be more difficult to relax. Note that the method does allow for indirect effects of selection (i.e. linked selection or selection arising from correlations between neutral and non-neutral loci independent of physical linkage), but that data on physical linkage and linkage disequilibrium among loci are not incorporated in any of the calculations. While this assumption of independence will rarely be met in nature (nor was it met in the QUANTINEMO simulations), the method can still be used as long as the selection coefficients are interpreted as measures of total effect of selection on a locus.

A final limitation of the proposed method is that it does not account for ascertainment bias or uncertainty in genotypes. The ascertainment process is difficult to model in a likelihood-based framework, but could readily be incorporated using an approximate Bayesian computation approach (as in Foll *et al.* 2015). Because of sequencing errors and finite coverage, modern DNA sequencing approaches do not provide perfect knowledge of genotypes. While one could simply assume the most likely genotype at each locus is the correct genotype, this is not ideal. Alternatively, population genetic parameters can be inferred while treating genotypes as random variables (Li *et al.* 2011; Gompert *et al.* 2012; Buerkle & Gompert 2013; Gompert *et al.* 2014b). Imputed genotypes can be dealt with in the same way. A model for genotype uncertainty will be added to a future version of the software released with this work.

Conclusions

Patterns of allele frequency change over time have generated particularly valuable insights into evolutionary processes (e.g. Fisher & Ford 1947; Lenski *et al.* 1991; Woods *et al.* 2006; Barrick *et al.* 2009; Teotonio *et al.* 2009; Burke *et al.* 2010). Indeed, analyses of these temporal patterns nicely complement genome scans for selection based on spatial patterns of genetic variation (e.g. Ellegren *et al.* 2012; Gompert *et al.* 2012; Hohenlohe *et al.* 2012; Jones *et al.* 2012; Soria-Carrasco *et al.* 2014). Moreover, genetic time series are a particularly fruitful source of information for understanding the causes and molecular targets of selection in variable or changing environments (e.g. Grant *et al.* 2004; Barrett *et al.* 2008). For example, Bergland *et al.* (2014) recently found evidence or repeated oscillations in allele frequencies at hundreds of loci associated with seasonal fluctuations in the environment. Such results suggest that balanced polymorphisms resulting from spatial or temporal variation in selection might be common and that these processes could contribute to the high levels of genetic variation observed in many natural populations(Gillespie 1991; Bell 2010; Leffler *et al.* 2012, but also see Hallsson & Björklund 2012; Huang *et al.* 2014). Answering this and other questions requires analytical methods for detecting and quantifying selection in environments that are heterogeneous across space and through time. Such methods have been lacking, but the statistical method described in this study will help fill this analytical gap and thereby facilitate genomic studies of evolutionary change over time.

Acknowledgements

I would like to thank L. Bernatchez, J. Jensen and M. Foll for organizing this special issue, and two anonymous reviewers for providing constructive comments that improved this manuscript. Computing, storage and other resources from the Division of Research Computing in the Office of Research and Graduate Studies at Utah State University are gratefully acknowledged. This research was funded by Utah State University.

References

Agrawal AA, Conner JK, Rasmann S (2010) Tradeoffs and negative correlations in evolutionary ecology. In: *Evolution Since Darwin: The First 150 Years* (eds Bell MA, Futuyma DJ, Eanes WF, Levinton JS), pp. 243–268. Sinauer Associates, Inc., Sunderland, Massachusetts.

Anderson JT, Lee CR, Rushworth CA, Colautti RI, Mitchell-Olds T (2013) Genetic trade-offs and conditional neutrality contribute to local adaptation. *Molecular Ecology*, **22**, 699–708.

Baldwin-Brown JG, Long AD, Thornton KR (2014) The power to detect quantitative trait loci using resequenced,

experimentally evolved populations of diploid, sexual organisms. *Molecular Biology and Evolution*, **31**, 1040–1055.
Barrett RDH, Rogers SM, Schluter D (2008) Natural selection on a major armor gene in threespine stickleback. *Science*, **322**, 255–257.
Barrick JE, Yu DS, Yoon SH et al. (2009) Genome evolution and adaptation in a long-term experiment with Escherichia coli. *Nature*, **461**, 1243–1249.
Barton NH, Keightley PD (2002) Understanding quantitative genetic variation. *Nature Reviews Genetics*, **3**, 11–21.
Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.
Beaumont MA, Rannala B (2004) The Bayesian revolution in genetics. *Nature Reviews Genetics*, **5**, 251–261.
Bell G (2010) Fluctuating selection: the perpetual renewal of adaptation in variable environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**, 87–97.
Benkman CW, Parchman TL (2009) Coevolution between crossbills and black pine: the importance of competitors, forest area and resource stability. *Journal of Evolutionary Biology*, **22**, 942–953.
Benkman CW, Holimon WC, Smith JW (2001) The influence of a competitor on the geographic mosaic of coevolution between crossbills and lodgepole pine. *Evolution*, **55**, 282–294.
Berg JJ, Coop G (2014) A population genetic signal of polygenic adaptation. *PLoS Genetics*, **10**, e1004412.
Bergland AO, Behrman EL, O'Brien KR, Schmidt PS, Petrov DA (2014) Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in *Drosophila*. *PLoS Genetics*, **10**, e1004775.
Bollback JP, York TL, Nielsen R (2008) Estimation of $2n_e$s from temporal allele frequency data. *Genetics*, **179**, 497–502.
Brodie ED III (2000) Why evolutionary genetics does not always add up. In: *Epistasis and The Evolutionary Process* (eds Wolf JB, Brodie ED III, Wade MJ), pp. 3–19. Oxford University Press, New York, New York.
Buerkle CA, Gompert Z (2013) Population genomics based on low coverage sequencing: how low should we go? *Molecular Ecology*, **22**, 3028–3035.
Burke MK, Dunham JP, Shahrestani P, Thornton KR, Rose MR, Long AD (2010) Genomewide analysis of a long-term evolution experiment with Drosophila. *Nature*, **467**, 587–590.
Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on the neutral molecular variation. *Genetics*, **134**, 1289–1303.
Coop G, Witonsky D, Di Rienzo A, Pritchard JK (2010) Using environmental correlations to identify loci underlying local adaptation. *Genetics*, **185**, 1411–1423.
Cortez MH, Ellner SP (2010) Understanding rapid evolution in predator-prey interactions using the theory of fast-slow dynamical systems. *The American Naturalist*, **176**, E109–E127.
Cortez MH, Weitz JS (2014) Coevolution can reverse predator-prey cycles. *Proceedings of the National Academy of Sciences USA*, **111**, 7486–7491.
Csillery K, Blum MGB, Gaggiotti OE, Francois O (2010) Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology and Evolution*, **25**, 410–418.
Egea-Serrano A, Hangartner S, Laurila A, Räsänen K (2014) Multifarious selection through environmental change: acidity and predator-mediated adaptive divergence in the moor frog (*Rana arvalis*). *Proceedings of the Royal Society of London B: Biological Sciences*, **281**, 20133266.
Ellegren H, Smeds L, Burri R et al. (2012) The genomic landscape of species divergence in Ficedula flycatchers. *Nature*, **491**, 756–760.
Endler JA (1977) *Geographic Variation, Speciation, and Clines*. Princeton University Press, Princeton, New Jersey.
Endler JA (1980) Natural selection on color patterns in Poecilia reticulata. *Evolution*, **34**, 76–91.
Endler JA (1986) *Natural Selection in the Wild*. Princeton University Press, Princeton, New Jersey.
Ewens WJ (2004) *Mathematical Population Genetics: I. Theoretical Introduction*, vol. 27. Springer Science & Business Media, New York, New York.
Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity*, **103**, 285–298.
Fisher RA, Ford EB (1947) The spread of a gene in natural conditions in a colony of the moth *Panaxia dominula* L. *Heredity*, **1**, 143–174.
Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**, 977–993.
Foll M, Shim H, Jensen JD (2015) WFABC: a Wright–Fisher ABC-based approach for inferring effective population sizes and selection coefficients from time-sampled data. *Molecular Ecology Resources*, **15**, 87–98.
Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM (2011) A map of local adaptation in *Arabidopsis thaliana*. *Science*, **334**, 86–89.
Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitch hiking and background selection. *Genetics*, **147**, 915–925.
Galassi M, Davies J, Theiler J et al. (2009) *GNU Scientific Library: Reference Manual*. Network Theory Ltd.
Gamerman D, Hedibert FL (2006) *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman and Hall, New York, New York.
Gelman A, Carlin J, Stern H, Rubin D (2004) *Bayesian Data Analysis*, 2nd edn. Chapman and Hall, London.
Gillespie JH (1991) *The Causes of Molecular Evolution*. Oxford University Press, Oxford.
Gillespie JH (2000) Genetic drift in an infinite population: the pseudohitchhiking model. *Genetics*, **155**, 909–919.
Gompert Z, Buerkle CA (2011a) Bayesian estimation of genomic clines. *Molecular Ecology*, **20**, 2111–2127.
Gompert Z, Buerkle CA (2011b) A hierarchical Bayesian model for next-generation population genomics. *Genetics*, **187**, 903–917.
Gompert Z, Lucas LK, Nice CC, Fordyce JA, Forister ML, Buerkle CA (2012) Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species. *Evolution*, **66**, 2167–2181.
Gompert Z, Comeault AA, Farkas TE et al. (2014a) Experimental evidence 589 for ecological selection on genome variation in the wild. *Ecology Letters*, **17**, 369–379.
Gompert Z, Lucas LK, Buerkle CA, Forister ML, Fordyce JA, Nice CC (2014b) Admixture and the organization of genetic diversity in a butterfly species complex revealed through

common and rare genetic variants. *Molecular Ecology*, **23**, 4555–4573.

Gompert Z, Jahner JP, Scholl CF et al. (2015) The evolution of novel host use is unlikely to be constrained by tradeoffs or a lack of genetic variation. *Molecular Ecology*, **24**, 2777–2793.

Grant PR, Grant BR (2002) Unpredictable evolution in a 30-year study of Darwin's finches. *Science*, **296**, 707–711.

Grant PR, Grant BR, Markert JA, Keller LF, Petren K (2004) Convergent evolution of Darwin's finches caused by introgressive hybridization and selection. *Evolution*, **58**, 1588–1599.

Guan Y, Stephens M (2011) Bayesian variable selection regression for genome-wide association studies and other large-scale problems. *Annals of Applied Statistics*, **5**, 1780–1815.

Günther T, Coop G (2013) Robust identification of local adaptation from allele frequencies. *Genetics*, **195**, 205–220.

Hahn MW (2008) Toward a selection theory of molecular evolution. *Evolution*, **62**, 255–265.

Hallsson LR, Björklund M (2012) Selection in a fluctuating environment leads to decreased genetic variation and facilitates the evolution of phenotypic plasticity. *Journal of Evolutionary Biology*, **25**, 1275–1290.

Hancock AM, Witonsky DB, Gordon AS et al. (2008) Adaptations to climate in candidate genes for common metabolic disorders. *PLoS Genetics*, **4**, e32.

Hohenlohe PA, Bassham S, Currey M, Cresko WA (2012) Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **367**, 395–408.

Huang Y, Wright SI, Agrawal AF (2014) Genome-wide patterns of genetic variation within and among alternative selective regimes. *PLoS Genetics*, **10**, e1004527.

Huerta-Sanchez E, Durrett R, Bustamante CD (2008) Population genetics of polymorphism and divergence under fluctuating selection. *Genetics*, **178**, 325–337.

Illingworth CJR, Mustonen V (2011) Distinguishing driver and passenger mutations in an evolutionary history categorized by interference. *Genetics*, **189**, 989–1000.

Jones FC, Grabherr MG, Chan YF et al. (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, **484**, 55–61.

Jorde PE, Ryman N (2007) Unbiased estimator for genetic drift and effective population size. *Genetics*, **177**, 927–935.

Kelley JL, Madeoy J, Calhoun JC, Swanson W, Akey JM (2006) Genomic signatures of positive selection in humans and the limits of outlier approaches. *Genome Research*, **16**, 980–989.

Kruschke J (2010) *Doing Bayesian Data Analysis: A Tutorial Introduction with R*. Academic Press, Amsterdam.

Lande R, Arnold S (1983) The measurement of selection on correlated characters. *Evolution*, **37**, 1210–1226.

Leffler EM, Bullaughey K, Matute DR et al. (2012) Revisiting an old riddle: What determines genetic diversity levels within species? *PLoS Biology*, **10**, e1001388.

Lehman N, Joyce GF (1993) Evolution in vitro: analysis of a lineage of ribozymes. *Current Biology*, **3**, 723–734.

Lenski RE, Rose MR, Simpson SC, Tadler SC (1991) Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *The American Naturalist*, **138**, 1315–1341.

Lewontin R (1974) *The Genetic Basis of Evolutionary Change*. Columbia University Press, New York, New York.

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

Li Y, Sidore C, Kang HM, Boehnke M, Abecasis GR (2011) Low-coverage sequencing: Implications for design of complex trait association studies. *Genome Research*, **21**, 940–951.

Malaspinas AS, Malaspinas O, Evans SN, Slatkin M (2012) Estimating allele age and selection coefficient from time-serial data. *Genetics*, **192**, 599–607.

Martin A, Orgogozo V (2013) The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution*, **67**, 1235–1250.

Mathieson I, McVean G (2013) Estimating selection coefficients in spatially structured populations from time series data of allele frequencies. *Genetics*, **193**, 973–984.

Maynard-Smith J, Haigh J (1974) Hitch hiking effect of a favorable gene. *Genetical Research*, **23**, 23–35.

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, **21**, 1087–1092.

Mueller LD, Barr LG, Ayala FJ (1985) Natural selection vs. random drift: Evidence from temporal variation in allele frequencies in nature. *Genetics*, **111**, 517–554.

Mustonen V, Lässig M (2007) Adaptations to fluctuating selection in *Drosophila*. *Proceedings of the National Academy of Sciences USA*, **104**, 2277–2282.

Neuenschwander S, Hospital F, Guillaume F, Goudet J (2008) quantiNemo: an individual based program to simulate quantitative traits with explicit genetic architecture in a dynamic metapopulation. *Bioinformatics*, **24**, 1552–1553.

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

Nosil P (2004) Reproductive isolation caused by visual predation on migrants between divergent environments. *Proceedings of the Royal Society of London B: Biological Sciences*, **271**, 1521–1528.

O'Hara RB (2005) Comparing the effects of genetic drift and fluctuating selection on genotype frequency changes in the scarlet tiger moth. *Proceedings of the Royal Society of London B: Biological Sciences*, **272**, 211–217.

Orsini L, Spanier KI, De Meester L (2012) Genomic signature of natural and anthropogenic stress in wild populations of the waterflea *Daphnia magna*: validation in space, time and experimental evolution. *Molecular Ecology*, **21**, 2160–2175.

Poisot T, Bever JD, Nemri A, Thrall PH, Hochberg ME (2011) A conceptual framework for the evolution of ecological specialisation. *Ecology Letters*, **14**, 841–851.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Rockman MV (2012) The QTN program and the alleles that matter for evolution: All that's gold does not glitter. *Evolution*, **66**, 1–17.

Rubin DB (1981) The Bayesian bootstrap. *The Annals of Statistics*, **9**, 130–134.

Sabeti PC, Reich DE, Higgins JM et al. (2002) Detecting recent positive selection in the human genome from haplotype structure. *Nature*, **419**, 832–837.

Schemske DW, Horvitz CC (1989) Temporal variation in selection on a floral character. *Evolution*, **26**, 461–465.
Schlötterer C, Kofler R, Versace E, Tobler R, Franssen S (2014) Combining experimental evolution with next-generation sequencing: a powerful tool to study adaptation from standing genetic variation. *Heredity*, **114**, 431–440.
Schluter D (2000) *The Ecology of Adaptive Radiation*. Oxford University Press, Oxford.
Siepielski AM, DiBattista JD, Carlson SM (2009) It's about time: the temporal dynamics of phenotypic selection in the wild. *Ecology Letters*, **12**, 1261–1276.
Siepielski AM, Gotanda KM, Morrissey MB, Diamond SE, DiBattista JD, Carlson SM (2013) The spatial patterns of directional phenotypic selection. *Ecology Letters*, **16**, 1382–1392.
Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science*, **236**, 787–792.
Soria-Carrasco V, Gompert Z, Comeault AA et al. (2014) Stick insect genomes reveal natural selection's role in parallel speciation. *Science*, **344**, 738–742.
Stephens M, Balding DJ (2009) Bayesian statistical methods for genetic association studies. *Nature Reviews Genetics*, **10**, 681–690.
Teotonio H, Chelo IM, Bradic M, Rose MR, Long AD (2009) Experimental evolution reveals natural selection on standing genetic variation. *Nature Reviews Genetics*, **41**, 251–257.
The HDF5 Group (2010) *Hierarchical data format version 5, 2000–2010*. Available from http://www.hdfgroup.org/HDF5.
Weese DJ, Gordon SP, Hendry AP, Kinnison MT (2010) Spatiotemporal variation in linear natural selection on body color in wild guppies (*Poecilia reticulata*). *Evolution*, **64**, 1802–1815.
Woods R, Schneider D, Winkworth CL, Riley MA, Lenski RE (2006) Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*. *Proceedings of the National Academy of Sciences USA*, **103**, 9107–9112.
Yeaman S, Jarvis A (2006) Regional heterogeneity and gene flow maintain variance in a quantitative trait within populations of lodgepole pine. *Proceedings of the Royal Society of London B: Biological Sciences*, **273**, 1587–1593.
Zhou X, Stephens M (2012) Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, **44**, 821–824.

Z.G. developed the model, wrote the computer code, designed the simulation study, analysed the data and wrote the manuscript.

Data accessibility

Source code, computer scripts and simulated data from the manuscript have been archived in DRYAD (doi: 10.5061/dryad.14853). An up-to-date version of the source code and an instruction manual for method described in this paper are available from SourceForge at http://sourceforge.net/projects/spatpg/.

Supporting information

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

Appendix S1 Reference manual for the computer software based on the methods described in this paper.