Inferring Population Decline and Expansion from Microsatellite Data: a Simulation-Based Evaluation of the MSVAR Method

Christophe GIROD,* Renaud VITALIS,§ Raphaël LEBLOIS,** and Hélène FREVILLE*§§

*UMR 7179 (CNRS – MNHN), F-91800 Brunoy, France

§Centre National de la Recherche Scientifique – Institut National de la Recherche Agronomique, UMR CBGP (INRA – IRD – CIRAD – Montpellier SupAgro), Campus International de Baillarguet, F-34988 Montferrier-sur-Lez Cedex, France

** Institut National de la Recherche Agronomique, UMR CBGP (INRA – IRD – CIRAD – Montpellier SupAgro), Campus International de Baillarguet, F-34988 Montferrier-sur-Lez Cedex, France

§§CEFE, UMR 5175 CNRS, F-34293 Montpellier cedex 05, France
**Running title:** Genetic inference of demography

**Keywords:** bottleneck, demographic history, expansion, likelihood-based inference, microsatellite DNA.

**Corresponding author:**

Hélène FREVILLE
Centre d’Ecologie Fonctionnelle et Evolutive
UMR 5175 CNRS
1919 route de Mende
34293 Montpellier Cedex 5
Phone: +33 (0)4 67 61 33 02
Fax: +33 (0)4 67 41 21 38
E-mail: helene.freville@cefe.cnrs.fr
ABSTRACT

Reconstructing the demographic history of populations is a central issue in evolutionary biology. Using likelihood-based methods coupled with Monte Carlo simulations, it is now possible to reconstruct past changes in population size from genetic data. Using simulated datasets under various demographic scenarios, we evaluate the statistical performance of MSVAR, a full-likelihood Bayesian method that infers past demographic change from microsatellite data. Our simulation tests show that MSVAR is very efficient at detecting population declines and expansions, provided the event is neither too weak, nor too recent. We further show that MSVAR outperforms two moment-based methods (the $M$-ratio test and BOTTLENECK) for detecting population size changes, whatever the time and the severity of the event. The same trend emerges from a compilation of empirical studies. The latest version of MSVAR provides estimates of the current and the ancestral population size, and the time since the population started changing in size. We show that, in the absence of prior knowledge, MSVAR provides little information on the mutation rate, which results in biased estimates and/or wide credibility intervals for each of the demographic parameters. However, scaling the population size parameters with the mutation rate and scaling the time with current population size, as coalescent theory requires, significantly improves the quality of the estimates for contraction but not for expansion scenarios. Finally, our results suggest that MSVAR is robust to moderate departures from a strict stepwise mutation model.
INTRODUCTION

Inferring past demography is a central concern in evolutionary biology and applied ecology. Characterizing past variations in population size is crucial, e.g., for understanding the impact of past climatic fluctuations on the current distribution of species (JACOBSEN et al. 2005; ELMER et al. 2009; Hu et al. 2009), and for the conservation of endangered species (FRANKHAM et al. 2002). Characterizing the demographic history of a species by direct approaches requires the monitoring of census data, which can be extremely difficult, not to say impossible, particularly in long-lived species. Yet, variations in census numbers of individuals also affect the dynamics of the genes carried by these individuals. A powerful alternative to direct approaches is therefore to use the recent advances in population genetic theory, which allow inferences on past demography from the observed distribution of genetic variation in natural populations (LAWTON-RAUH 2008).

Until recently, most of these indirect methods relied on summary statistics calculated from genetic data, and tests for departure from their theoretical distribution under a given demographic and mutational model (CORNUEt and LUIKART 1996; Schneider and Excoffier 1999; Garza and Williamson 2001). For instance, Cornuet and Luikart’s (1996) approach relies on the rationale that rare alleles, which contribute marginally to the heterozygosity, are more likely to be lost following a bottleneck. A transient excess in heterozygosity, as compared to that expected at equilibrium given the observed number of alleles in the sample, can therefore be used as a proxy to detect a bottleneck (Luikart and Cornuet 1998). Conversely, a transient heterozygosity deficiency may provide evidence for a population expansion (Cornuet and Luikart 1996; Leblois et al. 2006). In the same line of ideas, Garza and Williamson (2001) proposed a test to detect past population declines, based on the ratio \( M \) of the number of alleles to the range in allele size observed at
microsatellite loci. Because they are easy to implement and do not require time-consuming computations, these moment-based methods have been used in many empirical studies (see, e.g., SPENCER et al. 2000; COMPS et al. 2001; COLAUTTI et al. 2005). However, these methods suffer from a limited statistical power because they do not make full use of the data. Furthermore, they do not provide any estimate of the severity and the duration of the bottleneck.

Likelihood-based methods coupled with Monte Carlo sampling offer a powerful alternative to these moment-based methods (FELSENSTEIN 1992; GRIFFITHS and TAVARÉ 1994; EMERSON et al. 2001). They rely upon the computation of the likelihood of a sample configuration, i.e. the probability to observe the allele counts or the DNA polymorphic sites in that sample, given a demographic and mutational model. The parameters of interest of the underlying model are then estimated by maximizing the likelihood of the observed data. Likelihood-based methods that have been developed for inferring past demographic changes from the observed distribution of genetic variation include, e.g., BATWING (WILSON and BALDING 1998; WILSON et al. 2003), BEAST (DRUMMOND and RAMBAUT 2007), IM and IMa (HEY and NIELSEN 2004, 2007), LAMARC (KUHNER 2006) and MSVAR (BEAUMONT 1999). These methods differ not only with respect to the underlying demographic model, but also to the markers used (microsatellites, DNA sequences, etc.). Yet, because the computational burden required to evaluate statistical power and accuracy is particularly high, only few studies have attempted to test these methods (WILSON et al. 2003; ABDO et al. 2004; ROUSSET and LEBLOIS 2007; CHIKHI et al. 2010; STRASBURG and RIESEBERG 2010).

Among those methods, the one developed by BEAUMONT (1999), implemented in the software package MSVAR and further improved by STORZ and BEAUMONT (2002) and STORZ et al. (2002), has been increasingly used in the past few years to infer past demographic
changes (Table S1). MSVAR assumes a demographic model consisting in a single isolated population, which has undergone a linear or exponential change in effective population size at some time in the past. This method is designed to analyze multilocus microsatellite data that evolve according to a stepwise mutation model (SMM) (Ellegren 2004). MSVAR uses a Markov chain Monte Carlo (MCMC) method to sample from the posterior distribution of the model parameters (i.e., the current effective population size, the ancestral effective population size before the demographic change, the time at which the latter occurred, and the mutation rate of microsatellite loci).

Although MSVAR has been widely used, the statistical performance of the method has never been extensively evaluated. In his original article, Beaumont (1999) simulated a handful of data sets with known mutational and demographic parameters, and then evaluated the performance of the method for detecting demographic events and its sensitivity to the shape (linear or exponential) of the demographic change. Yet, the precision of the estimation of the model parameters was not evaluated. Furthermore, the performance of MSVAR with respect to the severity of demographic change, the time since the population started changing in size, and the mutation model has not been studied yet.

Here, we therefore aimed at evaluating the statistical performance of MSVAR (i) in detecting population declines and expansions and (ii) in providing accurate estimates of the model parameters, as a function of the severity of the demographic change, the time since it occurred, and the mutation model. To that end, we performed stochastic simulations to generate microsatellite datasets under different demographic scenarios and mutation models, and then analyzed these simulated data with MSVAR. In light of our results, we comment upon the published empirical studies that used MSVAR, and provide some guidelines for future studies.
**METHODS**

**Demographic model:** The demographic model implemented in **Msvar** (Beaumont 1999; Storz and Beaumont 2002) considers an isolated panmictic population of size $N_0$ at sampling time ($t = 0$). Going backward in time, the population size $N(t)$ changes deterministically (either linearly or exponentially) to an ancestral size $N_1$ at time $t = T_a$ and then remains constant at $N_1$ for $t > T_a$ (Beaumont 1999). In the following, we will only consider an exponential change in population size, with:

$$N(t) = N_0 \left( \frac{N_1}{N_0} \right)^{\frac{t}{T_a}},$$

for $0 < t < T_a$, and $N(t \geq T_a) = N_1$. For simplicity, the time is measured in units of generations, and population sizes are expressed as numbers of diploid individuals.

**Simulation study:** In order to test how **Msvar** performed depending upon the nature of the demographic change (decline or expansion), its strength and its time of occurrence, we simulated population declines and expansions for a range of parameter values for the current population size $N_0$, the ancestral population size $N_1$ and the time $T_a$. The computational burden of the method prevented an exhaustive exploration of the parameter space. In a first set of simulations, we therefore concentrated on a set of parameter values that represented a range of situations characterized by weak, moderate and strong changes in population size, with varying time of occurrence. For population declines, we fixed the current population size $N_0 = 100$ in all scenarios, and varied the ancestral population size $N_1 = \{1,000; 10,000; 100,000\}$ and the time since the demographic change $T_a = \{10; 50; 100; 500\}$. For population expansions, we fixed the ancestral population size $N_1 = 100$ in all scenarios, and varied the current population size $N_0 = \{1,000; 10,000; 100,000\}$ and the time since the demographic change $T_a = \{10; 50; 100; 500\}$. A total of 24 sets of demographic parameters were therefore...
considered. For this first set of simulations, each locus evolved according to a strict stepwise mutation model (SMM), as assumed in MSVAR. The mutation rate $\mu$ was set at $10^{-3}$, which is in agreement with estimates from the literature (Ellegren 2004).

Then, in order to test how MSVAR performed depending upon the mutation model, we performed a second set of simulations. The mutation process of microsatellites is complex and highly heterogeneous across loci and organisms (Ellegren 2000, 2004). While some observations of spontaneous mutations support a strict SMM, others suggest that multi-step mutations occur, with a frequency of multi-step changes $p$ varying from 0.04 to 0.74 (Ellegren 2000, 2004). Apart from a strict SMM, we thus simulated microsatellite data under a generalized stepwise model (GSM) with $p = 0.22$, an average value found in the literature (Dib et al. 1996; Ellegren 2000; Estoup et al. 2001; Ellegren 2004) and with $p = 0.74$, the most extreme value reported ever (Fitzsimmons 1998). The mutation rate $\mu$ was set at $10^{-3}$. For that second set of simulations, we considered a population decline scenario (with $N_0 = 100$, $N_1 = 10,000$ and $T_a = 500$), a population expansion scenario (with $N_0 = 10,000$, $N_1 = 100$ and $T_a = 500$), and a stable population scenario, taking the (constant) population size as the harmonic mean of the population size change from 100 to 10,000 for $T_a = 500$ generations, i.e. $N_0 = N_1 = 464$. This second set of simulations therefore consisted in 7 sets of parameters: 3 mutation models were considered for the stable population scenario (the SMM and the two GSMS), and 2 mutation models were considered for each of the declining and expanding population scenarios (the two GSMS).

Microsatellite data were simulated with Simcoal2 (Laval and Excoffier 2004), which generates samples of genes under various demographic models, using a discrete-generation coalescent algorithm. Discrete-generation algorithms produce simultaneous and multiple coalescences, which are cancelled out in the continuous-time approximation of the
coalescent. There might therefore be a slight discrepancy between the coalescence rate in the
discrete-generation algorithm and the continuous-time approximation that is assumed in
MSVAR, particularly for large sample sizes and small effective population size (see, e.g., the
Supplementary Figure S2 in CORNUET et al. 2008). Yet, we find it more relevant to simulate
the data without relying on approximations. Each dataset consisted in a sample of 50 diploid
individuals, genotyped at 10 unlinked microsatellite loci. This sampling scheme is consistent
with empirical studies that inferred past demographic changes using MSVAR: from an
exhaustive survey of the literature (Table S1), we found that the median numbers of
microsatellite loci and sampled individuals across datasets were 11.5 and 30, respectively.

For each set of parameters, we simulated five microsatellite datasets in order to have
replicates from the same underlying demographic and mutation model. We therefore obtained
a total of 120 simulated datasets for the first set of simulations, and 35 datasets for the second
set. (ELLEGREN 2004). For each set, we calculated the mean and standard deviation over the
five replicates of the expected heterozygosity $H_e$ (NEI 1978), the observed number of alleles
$N_a$, the range in allele size $A_r$ and the variance of allele range $V_a$, using ARLEQUIN (EXCOFFIER
et al. 2005).

**Parameterization of MSVAR:** In MSVAR, the posterior distribution of the model
parameters is computed by means of a Markov chain Monte Carlo (MCMC) method using the
Metropolis-Hastings’ algorithm (METROPOLIS et al. 1953; HASTINGS 1970). The likelihood is
calculated from the genealogical history of the sample of genes, represented as a sequence of
events (coalescences and mutations, see BEAUMONT 1999).

We used the version 1.3 of MSVAR, which provides separate estimates for $N_0$, $N_1$, $\mu$
and $T_a$ (STORZ and BEAUMONT 2002). This implementation of BEAUMONT’S (1999) method,
available at http://www.rubic.rdg.ac.uk/~mab/stuff/, relies upon a hierarchical model where
demographic and mutational parameters are allowed to vary among loci. The extent of interlocus variation is set by priors, as described in the Supporting Information. The parameter values reported in the following correspond to the mean of $N_0$, $N_1$, $T_a$ and $\mu$, across loci. In order to test whether the method could retrieve information from the data, we chose relatively flat priors on the mean parameters, including the mutation rate.

**Implementation:** Analyses were run on a Beowulf cluster made of 19 computer nodes, with CPUs ranging from bi-processors AMD Opteron mono-core running at 1.8 GHz to bi-processors Intel Xeon quad-core running at 2.0 GHz. For each of the simulated datasets, three independent MSVAR analyses were performed, with different starting values of the model parameters and different sets of seeds for the random number generator. For the first set of simulations (strict SMM with population size change), each Markov chain was initially run for $10^9$ steps and was thinned to 40,000 output lines by recording parameter values every 25,000 steps. In the absence of convergence, longer chains were run (see the Results section, below). For the second set of simulations (strict SMM with stable population, and GSM with stable, expanding and declining populations), each Markov chain was run for $3 \times 10^9$ steps, with parameter values recorded every 30,000 steps. The first 10% steps of the chains were discarded as burn-in. For each dataset, convergence was assessed by computing the multivariate extension of the Gelman and Rubin’s diagnostic (Brooks and Gelman 1998) on the three independent Markov chains. The Gelman and Rubin’s diagnostic is based on the computation of the ratio of the pooled-chains variance over the within-chain variance, which should be close to 1 if the chains converge to the target distribution. The multivariate Gelman and Rubin’s diagnostic was calculated from the means $\mathbf{M} = \{M_{N_0}, M_{N_1}, M_{T_a}, M_\mu\}$ and standard deviations $\mathbf{V} = \{V_{N_0}, V_{N_1}, V_{T_a}, V_\mu\}$ of MSVAR parameters across loci, using the CODA package (Plummer et al. 2006) implemented in the statistical software R (R
Although it might be recommended running more chains to compute the Gelman and Rubin’s diagnostic (e.g., Nettel et al. 2009 used 8 independent chains), the computational burden prevented us to run more than three chains in the present study.

**Analysis of MSVAR outputs:** MSVAR outputs were analyzed by focusing on two issues: (i) the performance of MSVAR at detecting past demographic changes; (ii) the precision of MSVAR estimates of the model parameters. For each of the simulated datasets, we combined the three Markov chains before running the following analyses. The strength of evidence of population expansion versus population decline (and vice versa) was evaluated using Bayes factors (Jeffreys 1961; Kass and Raftery 1995), as suggested by Beaumont (1999) and Storz and Beaumont (2002). The Bayes factor is a ratio where the numerator is the posterior probability of one model divided by its prior probability and the denominator is the posterior probability of an alternative model divided by its prior probability (Gelman et al. 1995). With identical priors for the population decline and the population expansion models (i.e., identical priors for $N_0$ and $N_1$), the Bayes factor for, e.g., a population decline is the ratio of the posterior probability of a population decline divided by the posterior probability of a population expansion. This ratio can be estimated by counting the number of states in the chain in which the population has declined (i.e., $N_0 / N_1 < 1$), and then dividing this by the number of states in which the population has expanded (i.e., $N_0 / N_1 > 1$) (see Storz and Beaumont 2002).

We estimated the marginal posterior distributions of the model parameters using the LOCFIT package (Loader 1999) implemented in R (R Development Core Team 2009). Point estimates of natural parameters $N_0$, $N_1$, $T_a$ and $\mu$ were computed from the mode of their marginal posterior distribution. The 90% highest probability density (HPD) intervals were
computed with the CODA package. We also estimated the marginal posterior distributions of the scaled parameters $\theta_0 \equiv 4N_0\mu$, $\theta_1 \equiv 4N_1\mu$ and $t_t \equiv T_a / (2N_0)$, and we computed point estimates and 90% HPD intervals for these scaled parameters. For each demographic scenario considered, we calculated the absolute value of the bias for both natural and scaled parameters over the five replicated datasets.

**Detection of population size change with BOTTLENECK and the $M$-ratio test:**

Finally, for the first set of simulations (strict SMM with population size change), we compared the performance of MSVAR to detect genetic signatures of demographic changes with the two most widely used moment-based methods available for microsatellite data. First, we analyzed the datasets using the method developed by CORNUET and LUIKART (1996) and implemented in the software package BOTTLENECK v.1.2 (CORNUET and LUIKART 1996). Wilcoxon signed-rank tests were performed to determine if a dataset exhibited a significant number of loci with heterozygosity excess as expected in bottlenecked populations (LUIKART et al. 1998), or with heterozygosity deficiency as expected in expanding populations (CORNUET and LUIKART 1996). Second, we calculated the GARZA and WILLIAMSON’s (2001) $M$-ratio on the 60 datasets corresponding to population declines. We compared empirical values of the $M$-ratio to 95% critical values ($M_c$) derived from 10,000 simulations of stable populations using the program CRITICAL_M. Simulations were performed using the true value of $\theta_1$ ($\theta_1 = 4, 40$ and $400$ in the scenarios considered), and assuming a strict stepwise mutation model. We considered that a $M$-ratio below the critical value $M_c$ was indicative of a population decline.
RESULTS

Genetic diversity of the simulated datasets, under a strict SMM: For the first set of simulations (strict SMM with population size change), the expected heterozygosity $H_e$, the number of alleles $N_a$ and the range in allele size $A_r$ are reported in Table 1. For contraction scenarios, $H_e$ ranged from 0.24 to 0.94. $N_a$ ranged from 2.3 to 23.7 and $A_r$ varied from 1.3 to 39.2. In agreement with theoretical expectations, $H_e$, $N_a$ and $A_r$ increased with $N_1$, the genetic diversity in the current population being sustained by large ancestral populations. Furthermore, genetic diversity decreased with increasing $T_a$, the loss of genetic diversity being more pronounced for long contraction events. For expansion scenarios, $H_e$ ranged from 0.30 to 0.58. $N_a$ ranged from 2.7 to 4.9 and $A_r$ varied from 1.8 to 4.0. In agreement with theoretical expectations, $H_e$, $N_a$ and $A_r$ increased with increasing $T_a$ since the number of mutations that segregate in the population increase with the age of the expansion event. We also observed a tendency for genetic diversity to increase with increasing $N_0$, although this trend was not clear-cut.

MCMC convergence: In the following, we used Gelman et al.’s (2004) rule of thumb, which suggests that values of the multivariate Gelman and Rubin’s convergence diagnostic comprised between 1.0 and 1.1 indicate reasonable convergence, whereas values above 1.1 indicate poor convergence. Out of the 120 analyses of the first set of simulations, 67 converged after $10^9$ steps (Table S2). The average computational time of these chains was 1.5 days for expansions and 3 days for contractions. The 53 non-converged analyses were run again for $3 \times 10^9$ steps and were thinned to 120,000 output lines by recording parameter values every 25,000 steps. Out of these 53 analyses, 20 converged after $3 \times 10^9$ steps, which took on average 20 days per chain. Finally, the last 33 non-converged analyses were run for $1.5 \times 10^{10}$ steps, which took 60 days per chain on average. Out of these 33 analyses, 16
converged after $1.5 \times 10^{10}$ steps. Therefore, a total of 17 analyses out of 120 (14.2%) did not converge after $1.5 \times 10^{10}$ steps. Most of these non-converged analyses corresponded to recently and severely bottlenecked populations ($T_a < 500$ and $N_0 / N_1 = 0.001$; Table S2). Yet, visual inspection of the three chains in the non-converged analyses, as well as the similarity of the marginal posterior distributions, suggested that the chains were close to equilibrium. Therefore, we included the 17 non-converged analyses in our results. The cumulative computation time for the completion of all the analyses included in our study exceeded $276 \times 10^3$ hours (33.5 years). There was a significantly positive correlation between the time of convergence and the average range in allele size $A_r$ in the sample for both contractions (Spearman’s $\rho = 0.82; p < 0.001$) and expansions (Spearman’s $\rho = 0.49; p < 0.001$).

**Detection of demographic events with M$\textsc{svar}$, under a strict SMM:** Bayes factors (BF) were computed for each of the 120 analyses of the first set of simulations, and interpreted following Jeffreys (1961): BF $\geq 10$ indicate strong support, BF ranging from 3 to 10 indicate substantial support, BF ranging from 0.33 to 3 indicate no support and values smaller than 0.33 indicate false detection of contraction or expansion. In 85 analyses out of 120 (70.8%), Bayes factors indicated a change in population size consistent with the simulated scenario with substantial to strong support (BF $\geq 3$ and BF $\geq 10$, respectively; see Figure 1). Out of the 60 Markov chains corresponding to contraction scenarios, 41 (68.3%) indicated a population decline (BF $\geq 3$), out of which 40 (97.6%) showed strong support (BF $\geq 10$). Fifteen of these 40 analyses (37.5%) did not converge. Out of the 60 analyses corresponding to expansion scenarios, 44 (73.3%) indicated a population expansion (BF $\geq 3$), out of which 34 (77.3%) showed strong support (BF $\geq 10$). Two of these 34 analyses (5.9%) did not converge. Overall, all the ancient ($T_a \geq 50$) and severe demographic changes ($N_0 / N_1 \leq 0.01$ for contractions and $N_0 / N_1 \geq 100$ for expansions) were detected with substantial to
strong support (Figure 1). By contrast, recent declines and expansions \( (T_a = 10) \) were largely undetected \( (BF < 3) \), except for strong contractions \( (N_0 / N_1 = 0.001) \). Moreover, weak contractions \( (N_0 / N_1 = 0.1) \) were largely undetected whatever their time of occurrence and one false expansion was even detected for an ancient and weak bottleneck \( (BF < 0.33, T_a = 500, N_0 / N_1 = 0.1) \).

**Comparison of MSVAR with moment-based methods:** Because Bayes factors cannot be formally compared to \( P \)-values, we were not able to use the same criterion for detecting population size change with MSVAR, BOTTLENECK and the \( M \)-ratio test. Therefore, we reported in Figure 2 the criteria that are generally used in empirical studies: \( BF \geq 3 \) for MSVAR; the result of the Wilcoxon signed-rank tests at the \( \alpha = 0.05 \) level for BOTTLENECK (CORNUET and LUIKART 1996); a \( M \)-ratio below the critical value \( M_c \) (GARZA and WILLIAMSON 2001) for the \( M \)-ratio test.

BOTTLENECK detected a significant excess in heterozygosity in only 5 out of the 60 datasets corresponding to contraction scenarios (8.3%). Ancient events \( (T_a = 500) \) were never detected whatever their severity. Moreover, there was no clear relationship between the rate of detection of population decline and the severity of the event. Finally, 5 datasets corresponding to ancient contractions \( (T_a \geq 100) \) showed significant heterozygote deficiency, hence supporting population expansions. Contrastingly, BOTTLENECK detected a significant deficiency in heterozygosity in 35 out of 60 datasets corresponding to expansion scenarios (58.3%). The GARZA and WILLIAMSON’S (2001) \( M \)-ratio method correctly detected a signal of contraction in 32 out of 60 datasets (53.3%). The rate of detection was higher for ancient \( (T_a \geq 50) \) and moderate-to-severe population declines \( (N_0 / N_1 \geq 0.01) \), with 26 significant tests out of 30. Recent \( (T_a = 10) \) and/or weak declines \( (N_0 / N_1 = 0.1) \) were barely detected (6 significant tests out of 30).
Estimation of demographic and mutational parameters with MSVAR, under a strict SMM: Demographic and mutational parameters were estimated for all datasets, by combining the three Markov chains run for each dataset. We assessed the quality of the estimates by examining the marginal posterior distributions of the parameters, as compared to their prior distributions. We summarized these results by calculating the modes and the 90% HPD intervals for each dataset (Figures S1, S2, S3 and S4) as well as the absolute value of the bias and the average HPD range over the five replicate datasets for each of the 24 demographic scenarios (Figures 3 and 4).

Estimates of the natural parameters $N_0, N_1, T_a, \mu$: Overall, the marginal posterior distributions of the demographic parameters $N_0, N_1$ and $T_a$ were wide, and departed only slightly from the priors (e.g., Figures 5A and 5B). The estimated 90% HPD limits were therefore broad (Figure 3), ranging from -4 to 8 in log_{10} scale (Figures S1 and S2).

For contractions, replicated datasets tended to provide more consistent results for old and severe events, as compared to recent events ($T_a = 10$) or events of low severity ($N_0 / N_1 = 0.1$) (Figure S1). The precision of the demographic parameter estimates tended to increase with increasing severity of the demographic change (as measured by the ratio $N_0 / N_1$) and the time of the event: (i) the 90% HPD range of the demographic parameter estimates decreased with increasing $N_0 / N_1$ (Figures 3 and S1); (ii) for moderate to strong contractions ($N_0 / N_1 > 0.1$), the 90% HPD range decreased with increasing $T_a$; (iii) for $N_0 / N_1 = 0.1$, the 90% HPD range was the lowest for intermediate values of $T_a$. The absolute value of the bias of $N_0$ estimates tended to be lower than that of $N_1$, and was maximized for recent events ($T_a = 10$).

The quality of the estimates of $N_0, N_1$ and $T_a$ was poor for expansions, as compared to contractions. The marginal posterior distributions were not sharply peaked and did not depart markedly from the priors. The 90% HPD limits were wide and the absolute value of
the bias was high, overall (Figure 3). This was true whatever the severity of the event and its
time of occurrence. It is noteworthy that, with few exceptions, all demographic parameter
estimates differed markedly across replicate datasets (Figure S2). We noticed that for a given
expansion severity, estimates of $N_0$ increased with $T_a$, while estimates of $N_1$ decreased with $T_a$
(Figure S2).

For contractions and expansions, the marginal posterior distributions of $\mu$ departed
only slightly from the prior distributions, whose mean was set at $\alpha_{\mu} = -4$ on a $\log_{10}$ scale.
Because the true mutation rate $\mu$ of the simulated datasets was set at -3 on a $\log_{10}$ scale, the
mutational parameter $\mu$ was therefore systematically underestimated, as already pointed out
by MILTON et al. (2009). The 90% HPD intervals of the marginal posterior distributions of $\mu$
were wide (data not shown).

Finally, we examined the patterns of correlation between natural parameters to assess
the performance of MSVAR to estimate natural parameters separately. We observed strong
correlations between natural parameters of the model. Overall, both $N_0$ and $N_1$ were
negatively correlated with the mutational parameter $\mu$ and there was a positive correlation
between $N_0$ and $T_a$ (Figure S5). The correlations were stronger for more severe events and
more ancient events. Furthermore, the correlations were more pronounced for contractions
than for expansions.

*Estimates of the scaled parameters $\theta_0$, $\theta_1$ and $t_f$: Scaled parameters were overall much
more precisely estimated than the natural parameters for contractions, whereas they were
poorly estimated for expansions. As with the natural parameters, the quality of the estimates
depended upon the severity of the demographic change and its time of occurrence.*

For contractions, the marginal posterior distributions of the scaled parameters $\theta_0$, $\theta_1$
and $t_f$ were very peaked and departed markedly from the prior distributions (e.g., Figures 5C
and 5D), except for contractions of low severity ($N_0 / N_1 = 0.1$). The precision (low bias, narrow 90% HPD interval) increased with increasing severity of the event and time of occurrence (Figures 4 and S3). In particular, estimates of $\theta_1$ and $t_f$ were overall very precise for moderate to severe bottlenecks ($N_0 / N_1 < 0.1$), except for very recent events ($T_a = 10$). Although $\theta_0$ was also well estimated for ancient declines ($T_a > 50$) from moderate to strong severity, the bias and the range of 90% HPD intervals were larger as compared to $\theta_1$. Replicate datasets provided consistent results for $\theta_1$ and $t_f$, for moderate and strong contractions ($N_0 / N_1 < 0.1$), that occurred more than 10 generations ago ($T_a > 10$). Larger variation across replicate datasets was observed for $\theta_0$.

For expansions, the marginal posterior distributions of the scaled parameters $\theta_0$ and $\theta_1$ were peaked and departed markedly from the prior distributions (data not shown). Yet, the mode of the marginal posterior distributions for $\theta_0$ departed markedly from the true simulated value, resulting in severe biases (Figures 4 and S4). By contrast, the scaled parameter $\theta_1$ exhibited low bias in all scenarios (Figure 4), although the 90% HPD intervals were wide, especially for weak and recent expansions ($N_0 / N_1 = 10$; $T_a > 10$). We noticed that the marginal posterior distributions of $\theta_0$ and $\theta_1$ were skewed respectively, to upper and lower values (Figures 4 and S4). The 90% HPD intervals of the marginal posterior distributions for the time parameter $t_f$ were wide in most conditions (Figures 4 and S4) and estimates were severely biased, except for $T_a = 500$. Both the 90% HPD interval and the bias decreased with increasing $T_a$. We observed large variations across replicate datasets for all scaled parameters in almost all situations, particularly for $t_f$.

**Influence of the mutation model in MsVAR:** Out of the 30 analyses presented in Figure 6 for the GSM, 12 (40%) did not converge after $3 \times 10^9$ steps. Out of these, 9 analyses (75%) concerned the datasets generated with the strongest GSM ($p = 0.74$). With data
generated under the moderate GSM \((p = 0.22)\), MSVAR successfully detected a population decline for the five simulated datasets, and a population expansion for 4 simulated datasets out of 5. Yet, MSVAR detected a false signal of population decline for 2 datasets out of 5 that were simulated under a stable population scenario (Figure 6). Under a strong GSM \((p = 0.74)\), MSVAR detected a signal of population decline with strong support \((BF \geq 10)\), whatever the simulated scenario (Figure 6).

The quality of MSVAR estimates of scaled parameters for the moderate GSM \((p = 0.22)\) was very similar to that observed for the strict SMM, with very precise estimates of \(\theta_1\) and \(t_f\), a slightly larger bias and 90% HPD intervals for \(\theta_0\) as compared to \(\theta_1\) in contraction scenarios, and poorer estimates, with large variations across replicate datasets, in expansion scenarios (Figure S6). For stable population scenarios, both the strict SMM and the moderate GSM \((p = 0.22)\) produced unbiased estimates of \(\theta_0\) and \(\theta_1\), but with very large 90% HPD intervals. Note that in the absence of population size change, estimates of \(t_f\) are meaningless. Very consistently, MSVAR produced biased estimates of the model parameters, with very narrow 90% HPD intervals, for all the datasets generated under the strong GSM \((p = 0.74)\) (Figure S6).

**DISCUSSION**

**Comparing MSVAR, BOTTLENECK and the \(M\)-ratio test:** BOTTLENECK performed poorly in detecting population declines from our simulated datasets under a SMM, with only 5 significant tests out of 60. The statistical power of BOTTLENECK for population declines is much lower when microsatellite loci evolve under a strict SMM than under an infinite allele model (CORNUEJ and LUIKART 1996) or a GSM (LEBLOIS et al. 2006). This may partly explain the low performance of BOTTLENECK in our comparative study. Our results for weak
population declines \((N_1 / N_0 = 10)\) are in agreement with previous simulation-based evaluations, given the set of demographic and mutational parameters considered here (see, e.g., Figure 3B in CORNUET and LUIKART 1996). For moderate to severe population declines \((N_1 / N_0 \geq 100)\), however, the rate of detection was lower in our study than in CORNUET and LUIKART (1996). Two possible reasons may explain this discrepancy. First, the average heterozygosity in our simulated datasets was overall higher than in CORNUET and LUIKART (1996) who considered a variable mutation rate across loci and simulations, in order to cover a range of heterozygosities per set of parameters. Second, we simulated an exponential change of population size, whereas CORNUET and LUIKART (1996) assumed an instantaneous reduction of population size in their simulation-based tests. The impact of the shape of the demographic change on the performance of BOTTLENECK has not been studied yet. Consistent with our study, the simulation-based evaluation of BOTTLENECK by LEBLOIS et al. (2006) also showed a low statistical power of the method. Interestingly, BOTTLENECK performed largely better for expansions (58.3\%) than for contractions (8.3\%), given the model parameter values of our study.

We found that the \(M\)-ratio test was more efficient than BOTTLENECK, which is consistent with LEBLOIS et al. (2006), for retrieving signals of population declines from our simulated datasets (32 significant tests out of 60). The rate of detection was higher for ancient and moderate-to-severe declines, while recent and weak declines were barely detected. These results are consistent with previous simulation-based studies that have shown that the \(M\)-ratio test has low statistical power for small \(\theta_1\) values (here, \(\theta_1 = 4\), see GARZA and WILLIAMSON 2001; WILLIAMSON-NATESAN 2005) and for recent population declines (see WILLIAMSON-NATESAN 2005; LEBLOIS et al. 2006). Here, we applied the \(M\)-ratio test by comparing the statistic \(M\) estimated from the data with the expected distribution of that statistic,
conditionally on the true value of $\theta_i$. This procedure increased the statistical power of the $M$-ratio test. Since the true value of the parameter $\theta_i$ is generally unknown in real situations, GARZA and WILLIAMSON (2001) recommended the use of $M_c = 0.68$ as a conservative threshold for the critical value. The reanalysis of our datasets with $M_c = 0.68$ (e.g., as in LEBLOIS et al. 2006) resulted in a lower rate of detection (22 significant tests out of 60), but in similar qualitative trends (higher rate of detection for severe and ancient population declines).

Given the set of demographic and mutational parameters used in our study, and using the decision criteria recommended by the developers of each method, MSVAR clearly outperformed the $M$-ratio test and BOTTLENECK for detecting population size change. While MSVAR correctly detected 68.3% of the declines, the $M$-ratio test and BOTTLENECK detected only 53.3% and 8.3%, respectively, of the declines. Any population decline detected by the $M$-ratio test and BOTTLENECK was also recovered by MSVAR, apart from one case of weak recent decline that was only identified by BOTTLENECK ($T_a = 10$ and $N_0 / N_1 = 0.1$). Therefore, our study does not support the previous claims that the $M$-ratio test and BOTTLENECK are best suited to detect recent population declines, whereas MSVAR is more appropriate to detect ancient contractions (GARZA and WILLIAMSON 2001; WILLIAMSON-NATESAN 2005). Moreover, while MSVAR detected 73.3% of the population expansions, BOTTLENECK only detected 58.3% of the expansions. Any expansion detected by BOTTLENECK was also recovered by MSVAR.

**Performance of MSVAR: what does coalescent theory tell us?** Not surprisingly, we found that the performance of MSVAR to infer past demography strongly depended on the information available in the data, which may be inferred from coalescent theory. Coalescent theory indeed predicts that variations in population size strongly affect the shape of gene
genealogies, which are star-shaped with long terminal branches in expanding populations and shallower in declining populations (Figure 7 and HEIN et al. 2005).

Coalescent theory further shows that only scaled parameters can be directly estimated from the data (TAVARÈ et al. 1997; NORDBORG 2007). Indeed, all parameters in coalescent models are scaled, and the likelihood function in MSVAR makes no exception (BEAUMONT 1999). Hence, inference of unscaled quantities such as population size, or time measured in generations, requires external information. In our study, unscaled parameters were therefore much less precisely estimated than the scaled ones (Figure 5) and were also highly correlated (Figure S5, see also Figure 5 in STORZ et al. 2002). We deliberately chose poorly informative priors, to test the capacity of MSVAR to retrieve information from the data only. In empirical studies, more informative priors of the natural parameters are usually specified. We acknowledge that MSVAR offers a principled approach for providing prior information on the mutation rate, in order to recover posterior densities for natural parameters. Yet it should be borne in mind that precise estimates of unscaled parameters may then largely stem from the specification of the priors. Imagine that analyses were performed using a prior distribution for the mutation rate with very low standard deviation (i.e., $\sigma_\mu$ close to zero). We would then necessarily recover the same level of precision for the natural parameters and the scaled parameters. Yet this improved precision may come at the expense of accuracy, if the prior distribution for the mutation rate departs from its true distribution.

Scenarios of population decline: MSVAR was very efficient for detecting population declines. However, its performance for detecting change in population size and accurately estimating the model parameters was lowest for recent events ($T_a = 10$) of low-to-moderate severity ($N_0 / N_1 \geq 0.01$), as well as for events of low severity ($N_0 / N_1 = 0.1$). This is expected since, for very recent declines, the gene genealogy can barely be distinguished from that
expected in a stable population with population size $N_1$ (Figure 7A). Interestingly, for $N_0 / N_1 = 0.1$, the performance of MSVAR was maximized for intermediate values of $T_a$, particularly with respect to the precision of $\theta_1$ estimates. This might be easily understood by considering that for ancient events ($T_a = 500$) most coalescence events occur while $N(t)$ is close to the current population size (see, e.g., Figure 7B). This might be further quantified by calculating the expected number $j$ of ancestral lineages at (scaled) time $\tau$, from which a sample of $n$ genes is descending. This number has a known distribution in a constant-size population (Tavaré 1984), and Lefebvre and Slatkin (2007) extended Tavaré’s (1984) formula in the case of an exponentially growing or declining population. Using their Equation (2) that gives an expression for the probability $\text{Pr}(j \mid n)$ that a sample of $n$ genes have $j$ ancestors $T_a$ generations ago, we may compute the expected number $m$ of lineages at the time of the population size change, as:

$$m = \sum_{j=1}^n j \text{Pr}(j \mid n) = \sum_{j=1}^n \sum_{i=j}^n \frac{(2i-1)(-1)^{i-j} j_{(i-j)n_{i}}}{j!(i-j)n_{(i)}} e^{-(i-1)\tau/2},$$

where $a_{(i)} = a(a+1)\ldots(a+i-1)$, $a_{[i]} = a(a-1)\ldots(a-i+1)$ and

$$\tau = \int_0^{\tau'} \frac{dt'}{2N(t')} = \frac{(1-N_0/N_i)}{N_0 \log(N_i/N_0)} T_a.$$

For a declining population with $T_a = 500$ and $N_0 / N_1 = 0.1$, we get $m = 1.43$, which confirms that most coalescence events are expected to occur in the current population with this set of parameter values.

For moderate to severe contractions ($N_0 / N_1 \leq 0.1$), both the bias and the 90% HPD range of $\theta_0$ decreased with increasing $T_a$. Using Equation (1), we found that the expected number $m$ of lineages at the time of the event varies between 48.49 and 2.20 for $T_a$ varying from 10 to 500 and for $N_0 / N_1 = 0.01$. This indicates that more coalescence events are expected to occur in the declining population when the event is older (see also Figure 7C).
contrast, $\theta_1$ was overall precisely estimated (see Figures 4 and S3). This is so because, for the scenarios considered here, a large part of the genealogy depends upon the ancestral history, with several lineages coalescing in the ancestral population (see, e.g., Figures 7B-C) at a rate that depends upon $\theta_1$. Had we considered older events ($T_a > 500$), though, thereby decreasing the number of lineages in the ancestral population, it is likely that the precision of $\theta_1$ estimates had declined.

In summary, most scenarios of population decline result in gene genealogies with large times to the most recent common ancestor (TMRCAs). With the set of model parameters considered here, since a large part of gene genealogies depends upon $\theta_1$, this latter parameter is generally precisely and accurately estimated. Contrastingly, $\theta_0$ can only be precisely and accurately estimated if the demographic event is severe and ancient. If the change in population size is too recent, provided that it is not too pronounced, $\theta_0$ estimates tend to converge to the true value of $\theta_1$, and no change of population size is detected. If the difference in population size is weak, then the difference in coalescence rates before and after the event is not sufficient for MSVAR to detect a population size change and to provide precise estimates of $\theta_0$ and $\theta_1$.

Scenarios of population expansion: MSVAR was also very efficient for detecting expansions. Nevertheless, the estimates of the scaled current population size $\theta_0$ were more severely biased and less precise, as compared to scenarios of population decline, for the same relative severity of the event. This may be explained by the fact that expanding populations result in young genealogies with short TMRCAs (compare Figures 7A-C and 7D-F), hence rare mutation events. We found that the absolute value of the bias increased with $N_0$. We further found that both the 90% HPD range and the absolute value of the bias of $\theta_0$ decreased with increasing $T_a$ (Figures 4 and S4). Using Equation (1), we found that the expected number
$m$ of lineages at the time of the event varies between 48.49 and 2.20 for $T_a$ varying from 10 to 500 and for $N_0 / N_1 = 100$. This indicates that the number of coalescence events in the expanding population is expected to increase as $T_a$ increases (see also Figure 7F). More generally, the likelihood surface for expanding populations is complex (Beaumont 1999). In particular, as the genealogies become more star-shaped, the joint posterior distribution of $\theta_0$ and $t_f$ reduces to a ridge along a line $\log_{10}(2\mu T_a) = k$ independent of $\theta_0$ (Figure 4b in Beaumont 1999), which suggests that MSVAR provides information on $2\mu T_a$, rather than on $\theta_0$ in expanding populations. It is worth noting that Wakeley et al. (2001) found similar results in inferring demographic history from single-nucleotide polymorphisms (see a comparison in Beaumont 2004).

Estimates of $\theta_1$ had a low bias but a large 90\% HPD range (Figure 4). Although the marginal posterior distributions of $\theta_1$ were generally peaked, they were flat-tailed on the left. This is so, because large ancestral population sizes are not compatible with the low polymorphism observed in the data. Instead, a large range of small values of $\theta_1$ may be equally likely, provided the genealogy is star-shaped.

**Influence of the underlying demographic model:** Although MSVAR equally detected population declines and expansions, inferences of the demographic parameters were in general more accurate for declines than for expansions. In addition to the above argument from coalescent theory, the exponential model assumed for population size change may partly explain this pattern. For declines, the size of the declining population $N(t)$ decreases sharply at $T_a$ and converges rapidly to $N_0$ (Figures 7A-C). Therefore, most coalescence events occurring in the declining population take place while $N(t)$ is close to $N_0$. For expansions, instead, the size of the expanding population $N(t)$ increases smoothly at $T_a$ before it converges rapidly to $N_0$ (Figures 7D-F). Therefore, a large proportion of the coalescence events that occur in the
growing population take place while $N(t)$ is close to $N_i$ (compare, e.g., Figures 7C and 7F). This can be expressed more formally by considering the harmonic mean of population sizes, which provides the coalescent rate during the change in population size (Hein et al. 2005).

For, say, $T_a = 500$, the harmonic mean of an exponentially declining population with $N_0 = 100$ and $N_i = 10,000$ is 464, which is strictly equal to the harmonic mean of an exponentially growing population with $N_0 = 10,000$ and $N_i = 100$. Hence, the harmonic mean of a declining population is closer to its current size ($N_0$) than its ancestral size ($N_i$), while the reverse is true for expanding populations. Therefore, given the exponential model of population growth, one might expect poor statistical properties of $\theta_0$ estimates in expanding populations, as compared to declining populations.

Robustness of MsVAR to the misspecification of the mutation model: Most importantly, our results suggest that MsVAR is robust to moderate departures from a strict SMM, e.g., a GSM with $p \leq 0.22$, typical of those observed in the literature (see Figures 6 and S6). However, severe departures from a strict SMM (here, a GSM with $p = 0.74$), led MsVAR to detect a signal of population decline with strong support (BF $\geq 10$), even in expanding populations (Figures 6 and S6). This is not surprising since it has been recognized that violation of the assumptions of the SMM might induce severe bias in the inference of demographic history (Gonsert al. 2000). Indeed, mutations that arise under a strong GSM involve large changes in allele length, which produce some gaps in the distribution of alleles types. The large resulting variance of allele range $V_a$ is reminiscent of that observed with population decline (Storz and Beaumont 2002), even in expanding populations (compare Tables 1 and S3).

Insights from empirical studies: The better performance of MsVAR as compared to the $M$-ratio test and Bottleneck also emerged from the empirical studies that inferred past
demographic changes from microsatellite data using MSVAR and at least one of the M-ratio or BOTTLENECK methods (Table S1). We found indeed that MSVAR detected a population decline whenever one of the moment-based methods provided a significant test. By contrast, a large number of population declines that were not detected with any of the moment-based methods were detected with MSVAR. Unfortunately, the scarcity of expansion events detected in the literature (Table S1) prevented any empirical comparison of MSVAR and BOTTLENECK for growing populations. Importantly, the average genetic diversity measured from our simulated datasets was not substantially different from that observed in empirical studies (compare Tables 1 and S1).

Because of the large heterogeneity of the published results, we did not attempt to analyze the quality of MSVAR estimates in empirical studies. Some studies used MSVAR 0.4 (BEAUMONT 1999), hence providing estimates for scaled parameters, and some studies used MSVAR 1.3 (STORZ and BEAUMONT 2002), hence providing estimates for unscaled parameters. Only a handful of studies used both methods, and few provided estimates of the scaled parameters using MSVAR 1.3, as in the present study. Finally, credibility intervals were often not reported or calculated using different methods, which hampered any comparison among studies.

**Recommendation guidelines and conclusions:** Our simulation tests as well as an exhaustive survey of the literature clearly demonstrate that MSVAR outperforms both the M-ratio test and BOTTLENECK for detecting population declines. Our study further shows that MSVAR is also very efficient to detect population expansions and outperforms BOTTLENECK in that respect. Yet, to our knowledge, MSVAR has only scarcely been applied on presumably expanding populations (see, e.g., HUFBAUER et al. 2004; BONHOMME et al. 2008; WIRTH et al. 2008). Hence, we confidently recommend the use of MSVAR for detecting past population size
variation, even if this method is computationally demanding.

Most importantly, in contrast to the $M$-ratio test and BOTTLENECK, MSVAR provides estimates of the parameters that characterize the population demographic history and the mutational model. Using MSVAR 1.3 (STORZ and BEAUMONT 2002), we have shown that the scaled parameters are more precisely estimated than the natural parameters. Although the latter are easier to interpret, our results clearly advocate drawing conclusions from inferences of $\theta_0$, $\theta_1$ and $t_e$. These parameters were precisely estimated for population declines, provided that the change in population size was neither too recent, nor too weak, given the scenarios considered in our study. For expansions instead, both unscaled and scaled parameters were poorly estimated, although the method was efficient for detecting increase in population size. Hence, our results suggest that MSVAR estimates in presumably expanding populations should be taken cautiously. We did not compare the performance of MSVAR 0.4 (BEAUMONT 1999), which provides estimates for scaled parameters, and MSVAR 1.3 (STORZ and BEAUMONT 2002), which provides estimates for unscaled parameters, since the two versions differ by a number of other aspects. MSVAR 0.4 assumes a basic (non-hierarchical) model, where the parameters are not allowed to vary among loci. The parameterization of MSVAR 1.3 by means of a hierarchical model allows for some variation of the parameters among loci, which may provide a means of identifying aberrant loci. Yet, with broad priors on interlocus variation of the model parameters ($V$), MSVAR 1.3 does not fully “borrow strength” from the different loci, e.g., by simply pooling information (multiplying the likelihoods) across loci (BEAUMONT and RANNALA 2004). This generally results in broader posterior distributions in the hierarchical MSVAR 1.3 model, as compared to the basic MSVAR 0.4 model. Hence, although our results suggest that the use of scaled parameters should be preferred, further analyses are required to compare the performance of MSVAR 0.4 and 1.3.
Finally, we recommend inferences about population demographic change with MSVAR to be interpreted cautiously in light of potential departures from the model assumptions. First, MSVAR assumes that microsatellites evolve according to a strict SMM. Although a moderate departure from this mutation model, as classically measured with observations of spontaneous mutations (Ellegren 2000, 2004), does not seem to undermine MSVAR performance (see Figures 6 and S7), loci that evolve under a strong GSM may invalidate the approach by detecting false signals of population decline whatever the true demographic history. Yet, the hierarchical model implemented in MSVAR 1.3 allows for variations in mutation and demographic parameters across loci and may thus limit the potential biases due to misspecifications of the mutation model. Indeed, Storz et al. (2002) argue that loci that strongly depart from the strict SMM shall be given less weight, thereby minimizing their impact on the inference made. Second, MSVAR assumes that populations are isolated. Real populations, however, are in general connected by gene flow. It is now acknowledged that population structure and/or isolation by distance, may result in incorrect inference of population demographic history (Pope et al. 2000; Leblois et al. 2006; Nielsen and Beaumont 2009; Chikhi et al. 2010; Peter et al. 2010). Last, further work is needed to evaluate how Msvar performs when the demography is more complex, e.g., with successions of population declines and expansions.

ACKNOWLEDGEMENTS

We are grateful to L. Chikhi for sharing an unpublished manuscript, as well as to Mark A. Beaumont and two anonymous reviewers for their constructive criticism of the present manuscript. This work was supported by the Plan Pluri-Formation “Evolution et structure des Ecosystèmes” (2004-2008) from the Muséum national d’Histoire naturelle (MNHN), by a
Nouragues Research grant from the CNRS programme Amazonie I (2007), by the CNRS programme Amazonie II (2008-2011) and by the ANR programme BLANC “EMILE” NT09-611697. This work is part of Christophe Girod’s PhD, who was supported by a grant from the French Ministry of Research (2007-2010). Part of this work was carried out by using the resources of the Computational Biology Service Unit from the MNHN (CNRS UMS 2700).

LITERATURE CITED


CHIKHI, L., V. C. SOUSA, P. LUISI, B. GOOSSENS and M. A. BEAUMONT, 2010 The confounding effects of population structure, genetic diversity and the sampling scheme on the detection and quantification of population size changes. Genetics 186:
983-995.


ELMER, K. R., C. REGGIO, T. WIRTH, E. VERHEYEN, W. SALZBURGER et al., 2009 Pleistocene desiccation in East Africa bottlenecked but did not extirpate the adaptive radiation of Lake Victoria haplochromine cichlid fishes. Proceedings of the National Academy of


FELSENSTEIN, J., 1992 Estimating effective population size from samples of sequences: inefficiency of pairwise and segregating sites as compared to phylogenetic estimates. Genetical Research 59: 139-147.


JACOBSEN, B. H., M. M. HANSEN and V. LOESCHCKE, 2005 Microsatellite DNA analysis of northern pike (Esox lucius L.) populations: insights into the genetic structure and demographic history of a genetically depauperate species. Biological Journal of the
Linnean Society 84: 91-101.


LAVAL, G., and L. EXCOFFIER, 2004 SIMCOAL 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history Bioinformatics 20: 2485-2487.


METROPOLIS, N., A. W. ROSENBLUTH, M. N. ROSENBLUTH, A. H. TELLER and E. TELLER,


Nettel, A., R. S. Dodd and Z. Afzal-Rafii, 2009 Genetic diversity, structure, and demographic change in tanoak, Lithocarpus densiflorus (Fagaceae), the most susceptible species to sudden oak death in California. American Journal of Botany 96: 2224-2233.


R Foundation for Statistical Computing, Place Published.

ROUSSET, F., and R. LEBLOIS, 2007 Likelihood and approximate likelihood analyses of 
genetic structure in a linear habitat: performance and robustness to model mis-

SCHNEIDER, S., and L. EXCOFFIER, 1999 Estimation of past demographic parameters from the 
distribution of pairwise differences when the mutation rates vary among sites: 
application to human mitochondrial DNA. Genetics 152: 1079-1089.

SPENCER, C. C., J. E. NEIGEL and P. L. LEBERG, 2000 Experimental evaluation of the 
usefulness of microsatellite DNA for detecting demographic bottlenecks. Molecular 

STORZ, J. F., and M. A. BEAUMONT, 2002 Testing for genetic evidence of population 
expansion and contraction: an empirical analysis of microsatellite DNA variation 
using a hierarchical Bayesian model. Evolution 56: 154-166.

STORZ, J. F., M. A. BEAUMONT and S. C. ALBERTS, 2002 Genetic evidence for long-term 
population decline in a savannah-dwelling primate: inferences from a hierarchical 

STRASBURG, J. L., and L. H. RIESEBERG, 2010 How robust are “isolation with migration” 
alyses to violations of the IM model? A simulation study. Molecular Biology and 
Evolution 27: 297-310.

TAVARÉ, S., 1984 Line-of-descent and genealogical processes, and their applications in 
population genetics models. Theoretical Population Biology 26: 119-165.

TAVARÉ, S., D. J. BALDING, R. C. GRIFFITHS and P. DONNELLY, 1997 Inferring coalescence 
times from DNA sequence data. Genetics 145: 505-518.


TABLE 1
Genetic diversity in the first set of simulated data

A. Contractions, with $N_0 = 100$ (SMM)

<table>
<thead>
<tr>
<th>$T_a$</th>
<th>$N_i = 1,000$</th>
<th></th>
<th></th>
<th></th>
<th>$N_i = 10,000$</th>
<th></th>
<th></th>
<th></th>
<th>$N_i = 100,000$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$H_a$</td>
<td>$N_a$</td>
<td>$A_r$</td>
<td>$V_a$</td>
<td>$H_a$</td>
<td>$N_a$</td>
<td>$A_r$</td>
<td>$V_a$</td>
<td>$H_a$</td>
<td>$N_a$</td>
<td>$A_r$</td>
<td>$V_a$</td>
</tr>
<tr>
<td>10</td>
<td>0.60</td>
<td>4.5</td>
<td>3.8</td>
<td>2.7</td>
<td>0.86</td>
<td>10.8</td>
<td>11.8</td>
<td>17.2</td>
<td>0.94</td>
<td>23.7</td>
<td>34.6</td>
<td>107.3</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.4)</td>
<td>(0.6)</td>
<td>(0.8)</td>
<td>(0.01)</td>
<td>(0.4)</td>
<td>(1.9)</td>
<td>(5.3)</td>
<td>(0.01)</td>
<td>(2.6)</td>
<td>(6.1)</td>
<td>(62.8)</td>
</tr>
<tr>
<td>50</td>
<td>0.61</td>
<td>4.4</td>
<td>3.7</td>
<td>2.7</td>
<td>0.85</td>
<td>10.2</td>
<td>12.7</td>
<td>21.7</td>
<td>0.93</td>
<td>20.1</td>
<td>37.5</td>
<td>138.7</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.3)</td>
<td>(0.7)</td>
<td>(1.0)</td>
<td>(0.02)</td>
<td>(0.1)</td>
<td>(0.9)</td>
<td>(6.4)</td>
<td>(0.01)</td>
<td>(0.6)</td>
<td>(1.9)</td>
<td>(15.3)</td>
</tr>
<tr>
<td>100</td>
<td>0.55</td>
<td>3.9</td>
<td>3.5</td>
<td>2.8</td>
<td>0.80</td>
<td>8.1</td>
<td>10.9</td>
<td>16.4</td>
<td>0.89</td>
<td>14.8</td>
<td>39.2</td>
<td>171.6</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.5)</td>
<td>(0.8)</td>
<td>(1.0)</td>
<td>(0.01)</td>
<td>(0.6)</td>
<td>(1.9)</td>
<td>(6.1)</td>
<td>(0.01)</td>
<td>(1.1)</td>
<td>(2.8)</td>
<td>(33.8)</td>
</tr>
<tr>
<td>500</td>
<td>0.24</td>
<td>2.3</td>
<td>1.3</td>
<td>0.7</td>
<td>0.56</td>
<td>4.2</td>
<td>6.9</td>
<td>13.5</td>
<td>0.66</td>
<td>5.8</td>
<td>27.1</td>
<td>165.7</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.2)</td>
<td>(0.2)</td>
<td>(0.2)</td>
<td>(0.04)</td>
<td>(0.4)</td>
<td>(1.3)</td>
<td>(5.6)</td>
<td>(0.05)</td>
<td>(0.6)</td>
<td>(2.2)</td>
<td>(58.6)</td>
</tr>
</tbody>
</table>
B. Expansions, with \( N_i = 100 \) (SMM)

<table>
<thead>
<tr>
<th>( T_a )</th>
<th>( N_0 = 1,000 )</th>
<th>( N_0 = 10,000 )</th>
<th>( N_0 = 100,000 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( H_a )</td>
<td>( N_a )</td>
<td>( A_r )</td>
</tr>
<tr>
<td>10</td>
<td>0.34</td>
<td>2.7</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.1)</td>
<td>(0.2)</td>
</tr>
<tr>
<td>50</td>
<td>0.30</td>
<td>3.0</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.3)</td>
<td>(0.4)</td>
</tr>
<tr>
<td>100</td>
<td>0.32</td>
<td>3.0</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.1)</td>
<td>(0.2)</td>
</tr>
<tr>
<td>500</td>
<td>0.43</td>
<td>3.5</td>
<td>2.6</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.3)</td>
<td>(0.3)</td>
</tr>
</tbody>
</table>

SMM, stepwise mutation model; \( H_a \), expected heterozygosity; \( N_a \), number of alleles; \( A_r \), allele size range; \( V_a \), variance of allele size range; estimates of genetic diversity are averaged over the five simulated datasets for each set of demographic parameters (\( N_0 \), current effective population size; \( N_i \), ancestral effective population size; \( T_a \), time since the population size change; standard deviations are indicated below the mean, into parentheses.
FIGURE LEGENDS

FIGURE 1.—Detection of change in population size with MSVAR. For population declines (left panel) and population expansions (right panel), the Bayes Factors (BF) are given for each set of demographic parameters \( N_0, N_1 \) and \( T_a \) for each replicated dataset (lines). Following Jeffreys (1961), \( BF \geq 10 \) indicates strong support, and BF ranging from 3 to 10 indicates substantial support. BF ranging from 0.33 to 3 indicates no support and values smaller than 0.33 indicate false detection of contraction or expansion. Non-converged (NC) analyses are also indicated.

FIGURE 2.—Detection of change in population size with BOTTLENECK, the \( M \)-ratio test and MSVAR. For population declines with \( N_0 = 100 \) (left set of panels), BOTTLENECK, the \( M \)-ratio test and MSVAR are compared. For population expansions with \( N_1 = 100 \) (right set of panels), only BOTTLENECK and MSVAR are compared. Dark grey squares indicate detection with MSVAR, the \( M \)-ratio test or BOTTLENECK, light grey squares indicate no detection, and black squares indicate detection with MSVAR or BOTTLENECK of the “wrong” event. BOTTLENECK: NS, not significant. \( M \)-ratio test: \( M_c \) is the critical value below which the test is significant. MSVAR: BF, Bayes Factor.

FIGURE 3.—Precision of the estimates of the natural demographic parameters \( N_0, N_1 \) and \( T_a \). Bias (histograms) and absolute value of the range of the 90% HPD interval (horizontal coloured traits) for natural demographic parameters \( N_0, N_1 \) and \( T_a \) (from left to right) are presented in a log_{10} scale. Top: population declines; bottom: population expansions. In each graph, the dotted vertical line separates scenarios of increasing severity (\( N_1 = 1,000, N_1 = 10,000 \) and \( N_1 = 100,000 \) for population declines and \( N_0 = 1,000, N_0 = 10,000 \) and \( N_0 = 100,000 \) for population expansions). For each severity, the time of occurrence of the demographic event \( T_a \) is represented by different colours (orange for \( T_a = 10 \), light green for \( T_a = 50 \), dark green for \( T_a = 100 \) and blue for \( T_a = 500 \)).
FIGURE 4.—Precision of the estimates of the scaled parameters $\theta_0$, $\theta_1$ and $t_f$. Bias (histograms) and absolute value of the range of the 90%-HPD interval (horizontal coloured traits) for scaled parameters $\theta_0$, $\theta_1$ and $t_f$ (from left to right) are presented in a log10 scale. Top: population declines; bottom: population expansions. In each graph, the dotted vertical line separates scenarios of increasing severity ($N_1 = 1,000$, $N_1 = 10,000$ and $N_1 = 100,000$ for population declines and $N_0 = 1,000$, $N_0 = 10,000$ and $N_0 = 100,000$ for population expansions). For each severity, the time of occurrence of the demographic event $T_a$ is represented by different colours (orange for $T_a = 10$, light green for $T_a = 50$, dark green for $T_a = 100$ and blue for $T_a = 500$).

FIGURE 5.—Marginal posterior density of $N_0$, $N_1$ and $T_a$ and $\theta_0$, $\theta_1$ and $t_f$ for an ancient and severe population decline. All densities are represented in a log10 scale. (A) : population size natural parameters $N_0$ and $N_1$, (B) : time natural parameter $T_a$, (C) : scaled parameters $\theta_0$ and $\theta_1$, (D) scaled parameter $t_f$. The scenario corresponds to an ancient ($T_a = 500$) and severe population decline ($N_0 = 100$; $N_1 = 100,000$). The true values of the parameters in a log10 scale ($N_0 = 2$; $N_1 = 5$; $T_a = 2.70$; $\theta_0 = -0.40$; $\theta_1 = 2.60$; $t_f = 0.40$) are indicated by the vertical dotted line in each graph. The prior distributions of the parameters are given by the dashed grey curve in each graph.

FIGURE 6.—Detection of change in population size with MSVAR. The Bayes Factors (BF) are given for each of the following demographic scenario: a stable population (with $N_0 = N_1 = 464$, $T_a = 500$), a declining population (with $N_0 = 100$, $N_1 = 10,000$ and $T_a = 500$), and an expanding population (with $N_0 = 10,000$, $N_1 = 100$ and $T_a = 500$). We considered three different mutation models, which differ from each other by the value of $p$, the frequency of multi-step mutations changes: $p = 0.00$ (stepwise mutation model, SMM), $p = 0.22$ (moderate generalized stepwise model, GSM1), and $p = 0.74$ (strong generalized stepwise model, GSM2). For the stable population scenario, the lower triangle provides the Bayes factor for a population decline (i.e.,
the ratio of the posterior probability of a population decline divided by the posterior probability of a population expansion), and the upper triangle provides the Bayes factor for a population expansion (i.e., the ratio of the posterior probability of a population expansion divided by the posterior probability of a population decline). Following Jeffreys (1961), $BF \geq 10$ indicates strong support, and $BF$ ranging from 3 to 10 indicates substantial support. $BF$ ranging from 0.33 to 3 indicates no support and values smaller than 0.33 indicate false detection of contraction or expansion. Non-converged (NC) analyses are also indicated.

**FIGURE 7.**—Dynamics of population size changes $N(t)$ corresponding to simulated scenarios and expected gene genealogies. (A-C) corresponds to population decline and (D-F) to population expansion (dashed curve). The grey area in each graph indicates when the ancestral population size is constant, i.e. $N(t) = N_1$. Above each curve, the expected gene genealogy for 20 sampled lineages is represented. Expected gene genealogies were obtained by averaging coalescence times over 500,000 simulations of each demographic scenario. The simulations were based on a generation-by-generation coalescent algorithm developed by the authors. Note that some genealogies are incomplete (A and C), some lineages having not coalesced 800 generations from present.
Figure 1

<table>
<thead>
<tr>
<th>$T_a$</th>
<th>$N_1$</th>
<th>$N_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1,000</td>
<td>NC</td>
</tr>
<tr>
<td></td>
<td>10,000</td>
<td>NC</td>
</tr>
<tr>
<td></td>
<td>100,000</td>
<td>NC</td>
</tr>
<tr>
<td>50</td>
<td>1,000</td>
<td>NC</td>
</tr>
<tr>
<td></td>
<td>10,000</td>
<td>NC</td>
</tr>
<tr>
<td></td>
<td>100,000</td>
<td>NC</td>
</tr>
<tr>
<td>100</td>
<td>1,000</td>
<td>NC</td>
</tr>
<tr>
<td></td>
<td>10,000</td>
<td>NC</td>
</tr>
<tr>
<td></td>
<td>100,000</td>
<td>NC</td>
</tr>
<tr>
<td>500</td>
<td>1,000</td>
<td>NC</td>
</tr>
<tr>
<td></td>
<td>10,000</td>
<td>NC</td>
</tr>
<tr>
<td></td>
<td>100,000</td>
<td>NC</td>
</tr>
</tbody>
</table>

POPULATION DECLINE

POPULATION EXPANSION

Legend:
- BF < 0.33
- $0.33 \leq BF < 3$
- $3 \leq BF < 10$
- BF $\geq 10$
- NC: no convergence
Figure 2

<table>
<thead>
<tr>
<th>$T_a$</th>
<th>MSVAR</th>
<th>M-RATIO</th>
<th>BOTTLENECK</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$N_1 = 1,000$ | $N_1 = 10,000$ | $N_1 = 100,000$ | $N_0 = 1,000$ | $N_0 = 10,000$ | $N_0 = 100,000$

$0.33 \leq BF < 3$ | $BF < 0.33$ | $BF \geq 3$
Figure 3
Bias and HPD range

Figure 4
Figure 5
Figure 6

Stable population

Population decline

Population expansion

- BF < 0.33
- 0.33 \leq BF < 3
- 3 \leq BF < 10
- BF \geq 10
- no convergence

\( p = 0.00 \) \( p = 0.22 \) \( p = 0.74 \)
Figure 7

A. $N_0 = 100; N_1 = 10,000; T_a = 10$

B. $N_0 = 100; N_1 = 1,000; T_a = 500$

C. $N_0 = 100; N_1 = 100,000; T_a = 500$

D. $N_0 = 10,000; N_1 = 100; T_a = 10$

E. $N_0 = 1,000; N_1 = 100; T_a = 500$

F. $N_0 = 100,000; N_1 = 100; T_a = 500$