Inferring Population Decline and Expansion From Microsatellite Data: A Simulation-Based Evaluation of the Msvar Method

Christophe Girod,* Renaud Vitalis,[†] Raphaël Leblois[‡] and Hélène Fréville*[,§,1]

*UMR 7179 (Centre National de la Recherche Scientifique–Muséum National d'Histoire Naturelle), F-91800 Brunoy, France, [†]Centre National de la Recherche Scientifique–Institut National de la Recherche Agronomique and [‡]Institut National de la Recherche Agronomique, UMR Centre de Biologie pour la Gestion des Populations (INRA–Institut de Recherche pour le Développement–Centre de Coopération Internationale en Recherche Agronomique pour le Développement–Montpellier SupAgro), F-34988 Montferrier-sur-Lez Cedex, France and [§]Centre d'Ecologie Fonctionnelle et Evolutive, UMR 5175, Centre National de la Recherche Scientifique, F-34293 Montpellier Cedex 05, France

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 data sets 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.

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, 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 with respect to the markers used (microsatellites, DNA sequences, etc.). However, 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 (supporting information, Table S1). Msvar assumes a demographic model consisting of 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 expo-

nential) of the demographic change. However, 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 data sets 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 consider only an exponential change in population size, with

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

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: 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 $N_1 = \{1000; 10,000; 100,000\}$ 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 $N_0 = \{1000;$ 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 SMM, as assumed in Msvar. The mutation rate μ was set at 10^{-3}, which is in agreement with estimates from the literature (ELLEGREN 2004).

Then, 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 multistep mutations occur, with a frequency of multistep 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 μ 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 of seven sets of parameters: three mutation models were considered for the stable population scenario (the SMM and the two GSMs), and two 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 canceled 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., Figure S2 in CORNUET et al. 2008). However, we find it more relevant to simulate the data without relying on approximations. Each data set consisted of 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 data sets were 11.5 and 30, respectively.

For each set of parameters, we simulated five microsatellite data sets to have replicates from the same underlying demographic and mutation model. We therefore obtained a total of 120 simulated data sets for the first set of simulations and 35 data sets for the second set. 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 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 version 1.3 of Msvar, which provides separate estimates for N_0, N_1, μ, 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 File S1. The parameter values reported in the following correspond to the mean of N_0, N_1, T_a, and μ, across loci. 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 biprocessors AMD Opteron monocore running at 1.8 GHz to biprocessors Intel Xeon quadcore running at 2.0 GHz. For each of the simulated data sets, 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 RESULTS). 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 3×10^9 steps, with parameter values recorded every 30,000 steps. The first 10% of steps of the chains were discarded as burn-in. For each data set, convergence was assessed by computing the multivariate extension of Gelman and Rubin's diagnostic (BROOKS and GELMAN 1998) on the three independent Markov chains. 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 $\mathbf{M} \equiv \{M_{N_0}, M_{N_1}, M_{T_a}, M_\mu\}$ and standard deviations $\mathbf{V} \equiv \{V_{N_0}, V_{N_1}, V_{T_a}, V_\mu\}$ of Msvar parameters across loci, using the CODA package (PLUMMER *et al.* 2006) implemented in the statistical software R (R DEVELOPMENT CORE TEAM 2009). Although it might be recommended to run more chains to compute Gelman and Rubin's diagnostic (*e.g.*, NETTEL *et al.* 2009 used eight independent chains), the computational burden prevented us from running 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 and (ii) the precision of Msvar estimates of the model parameters. For each of the simulated data sets, we combined the three Markov chains before running the following analyses. The strength of evidence of population expansion *vs.* 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 μ 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 $t_f \equiv T_a/(2N_0)$, 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 data sets.

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 data sets 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 data set 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 GARZA and WILLIAMSON's (2001) M ratio on the 60 data sets 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 θ_1 ($\theta_1 = 4$, 40, and 400 in the scenarios considered) and assuming a strict stepwise mutation model. We considered that an M ratio below the critical value M_c was indicative of a population decline.

RESULTS

Genetic diversity of the simulated data sets, 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 increases 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 between 1.0 and 1.1 indicate reasonable convergence, whereas values >1.1 indicate poor convergence. 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 nonconverged analyses were run again for 3×10^9 steps

TABLE 1

Genetic diversity in the first set of simulated data

Contractions, with $N_0 = 100$ (SMM)

T_a	$N_1 = 1,000$				$N_1 = 10,000$				$N_1 = 100,000$			
	H_e	N_a	A_r	V_a	H_e	N_a	A_r	V_a	H_e	N_a	A_r	V_a
10	0.60 (0.06)	4.5 (0.4)	3.8 (0.6)	2.7 (0.8)	0.86 (0.01)	10.8 (0.4)	11.8 (1.9)	17.2 (5.3)	0.94 (0.01)	23.7 (2.6)	34.6 (6.1)	107.3 (62.8)
50	0.61 (0.04)	4.4 (0.3)	3.7 (0.7)	2.7 (1.0)	0.85 (0.02)	10.2 (0.1)	12.7 (0.9)	21.7 (6.4)	0.93 (0.01)	20.1 (0.6)	37.5 (1.9)	138.7 (15.3)
100	0.55 (0.03)	3.9 (0.5)	3.5 (0.8)	2.8 (1.0)	0.80 (0.01)	8.1 (0.6)	10.9 (1.9)	16.4 (6.1)	0.89 (0.01)	14.8 (1.1)	39.2 (2.8)	171.6 (33.8)
500	0.24 (0.05)	2.3 (0.2)	1.3 (0.2)	0.7 (0.2)	0.56 (0.04)	4.2 (0.4)	6.9 (1.3)	13.5 (5.6)	0.66 (0.05)	5.8 (0.6)	27.1 (2.2)	165.7 (58.6)

Expansions, with $N_1 = 100$ (SMM)

T_a	$N_0 = 1,000$				$N_0 = 10,000$				$N_0 = 100,000$			
	H_e	N_a	A_r	V_a	H_e	N_a	A_r	V_a	H_e	N_a	A_r	V_a
10	0.34 (0.03)	2.7 (0.1)	1.8 (0.2)	1.0 (0.3)	0.31 (0.03)	2.8 (0.1)	1.8 (0.1)	1.0 (0.1)	0.44 (0.02)	3.4 (0.4)	2.4 (0.5)	1.4 (0.4)
50	0.30 (0.03)	3.0 (0.3)	2.0 (0.4)	1.1 (0.3)	0.30 (0.03)	3.2 (0.3)	2.2 (0.3)	1.1 (0.2)	0.36 (0.06)	3.3 (0.4)	2.3 (0.5)	1.3 (0.4)
100	0.32 (0.03)	3.0 (0.1)	2.0 (0.2)	1.1 (0.1)	0.34 (0.05)	3.6 (0.2)	2.6 (0.2)	1.4 (0.2)	0.32 (0.05)	3.5 (0.2)	2.5 (0.3)	1.4 (0.2)
500	0.43 (0.08)	3.5 (0.3)	2.6 (0.3)	1.5 (0.2)	0.52 (0.04)	4.3 (0.4)	3.3 (0.4)	2.0 (0.3)	0.58 (0.03)	4.9 (0.3)	4.0 (0.3)	2.6 (0.2)

SMM, stepwise mutation model; H_e, 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 data sets for each set of demographic parameters (N_0, current effective population size; N_1, ancestral effective population size; T_a, time since the population size change). Standard deviations are indicated in parentheses.

and were thinned to 120,000 output lines by recording parameter values every 25,000 steps. Of these 53 analyses, 20 converged after 3×10^9 steps, which took on average 20 days per chain. Finally, the last 33 nonconverged analyses were run for 1.5×10^{10} steps, which took 60 days per chain on average. Of these 33 analyses, 16 converged after 1.5×10^{10} steps. Therefore, a total of 17 analyses of 120 (14.2%) did not converge after 1.5×10^{10} steps. Most of these nonconverged analyses corresponded to recently and severely bottlenecked populations ($T_a < 500$ and $N_0/N_1 = 0.001$; Table S2). However, visual inspection of the three chains in the nonconverged analyses, as well as the similarity of the marginal posterior distributions, suggested that the chains were close to equilibrium. Therefore, we included the 17 nonconverged analyses in our results. The cumulative computation time for the completion of all the analyses included in our study exceeded 276×10^3 hr (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 Msvar, 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 ≥ 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 < 0.33 indicate false detection of contraction or expansion. In 85 analyses of 120 (70.8%), Bayes factors indicated a change in population size consistent with the simulated scenario with substantial to strong support (BF ≥ 3 and BF ≥ 10, respectively; see Figure 1). Of the 60 Markov chains corresponding to contraction scenarios, 41 (68.3%) indicated a population decline (BF ≥ 3), of which 40 (97.6%) showed strong support (BF ≥ 10). Fifteen of these 40 analyses (37.5%) did not converge. Of the 60 analyses corresponding to expansion scenarios, 44 (73.3%) indicated a population expansion (BF ≥ 3), of which 34 (77.3%) showed strong support (BF ≥ 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
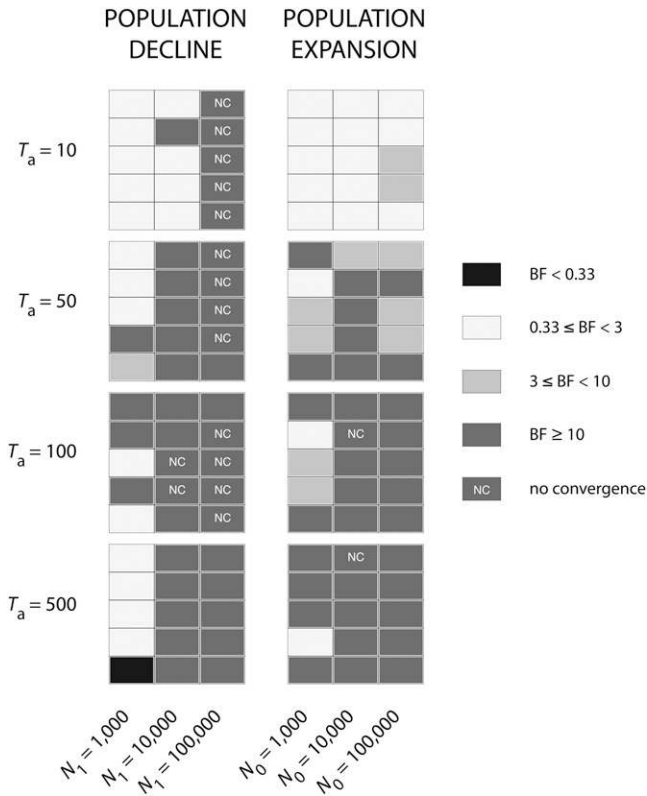
FIGURE 1.—Detection of change in population size with Msvar. For population declines (left) and population expansions (right), the Bayes factors (BF) are given for each set of demographic parameters N_0, N_1, and T_a for each replicated data set (lines). Following JEFFREYS (1961), BF $\geq$ 10 indicate strong support, and BF ranging from 3 to 10 indicate substantial support. BF ranging from 0.33 to 3 indicate no support and values <0.33 indicate false detection of contraction or expansion. Nonconverged (NC) analyses are also indicated.

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), and an *M* ratio below the critical value M_c (GARZA and WILLIAMSON 2001) for the *M*-ratio test.

Bottleneck detected a significant excess of heterozygosity in only 5 of the 60 data sets 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, 4 data sets 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 of 60 data sets corresponding to expansion scenarios (58.3%). GARZA and WILLIAMSON's (2001) *M*-ratio method correctly detected a signal of contraction in 32 of 60 data sets (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 of 30. Recent ($T_a = 10$) and/or weak declines ($N_0/N_1 = 0.1$) were barely detected (6 significant tests of 30).

Estimation of demographic and mutational parameters with Msvar, under a strict SMM: Demographic and mutational parameters were estimated for all data sets, by combining the three Markov chains run for each data set. We assessed the quality of the estimates by examining the marginal posterior distributions of the parameters, compared to their prior distributions. We summarized these results by calculating the modes and the 90% HPD intervals for each data set (Figure S1, Figure S2, Figure S3, and Figure S4) as well as the absolute value of the bias and the average HPD range over the five replicate data sets for each of the 24 demographic scenarios (Figures 3 and 4).

Estimates of the natural parameters N_0, N_1, T_a, and μ: 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.*, Figure 5, A and B). The estimated 90% HPD limits were therefore broad (Figure 3), ranging from -4 to 8 in $\log_{10}$ scale (Figure S1 and Figure S2).

For contractions, replicated data sets tended to provide more consistent results for old and severe events, 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 (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 (Figure 3 and Figure S1); (ii) for moderate to strong contractions ($N_0/N_1 > 0.1$), the 90% HPD range decreased with increasing T_a; and (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 poorer for expansions, 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 data sets (Figure S2). We noted 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 μ 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 μ of the simulated data sets was set at -3 on a $\log_{10}$ scale, the
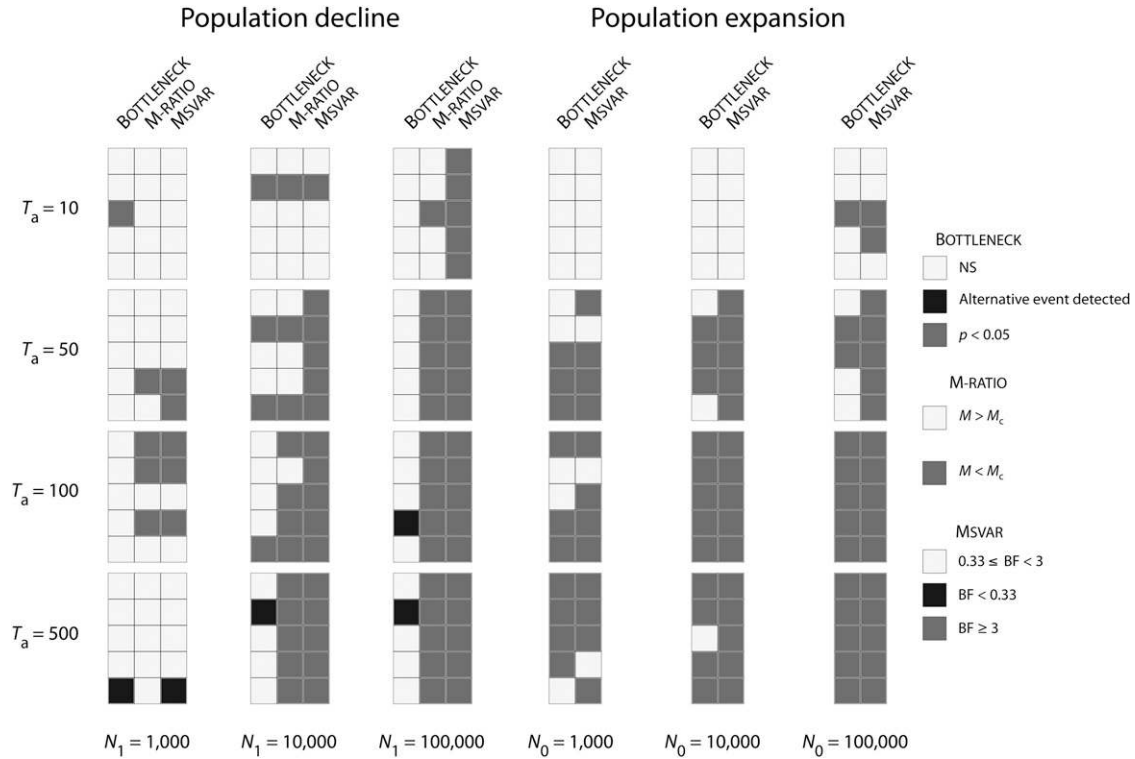
FIGURE 2.—Detection of change in population size with Bottleneck, the *M*-ratio test, and Msvar. For population declines with $N_0 = 100$ (left set of panels), Bottleneck, the *M*-ratio test, and Msvar are compared. For population expansions with $N_1 = 100$ (right set of panels), only Bottleneck and Msvar are compared. Squares with dark shading indicate detection with Msvar, the *M*-ratio test, or Bottleneck; squares with light shading indicate no detection; and solid 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.

mutational parameter μ was therefore systematically underestimated, as already pointed out by MILTON *et al.* (2009). The 90% HPD intervals of the marginal posterior distributions of μ 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 μ 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 θ_0, θ_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 θ_0, θ_1, and t_f were very peaked and departed markedly from the prior distributions

(*e.g.*, Figure 5, C and D), 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 (Figure 4 and Figure S3). In particular, estimates of θ_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 θ_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 compared to those of θ_1. Replicate data sets provided consistent results for θ_1 and t_f, for moderate and strong contractions ($N_0/N_1 < 0.1$) that occurred >10 generations ago ($T_a > 10$). Larger variation across replicate data sets was observed for θ_0.

For expansions, the marginal posterior distributions of the scaled parameters θ_0 and θ_1 were peaked and departed markedly from the prior distributions (data not shown). However, the mode of the marginal posterior distributions for θ_0 departed markedly from the true simulated value, resulting in severe biases (Figure 4 and Figure S4). By contrast, the scaled parameter θ_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
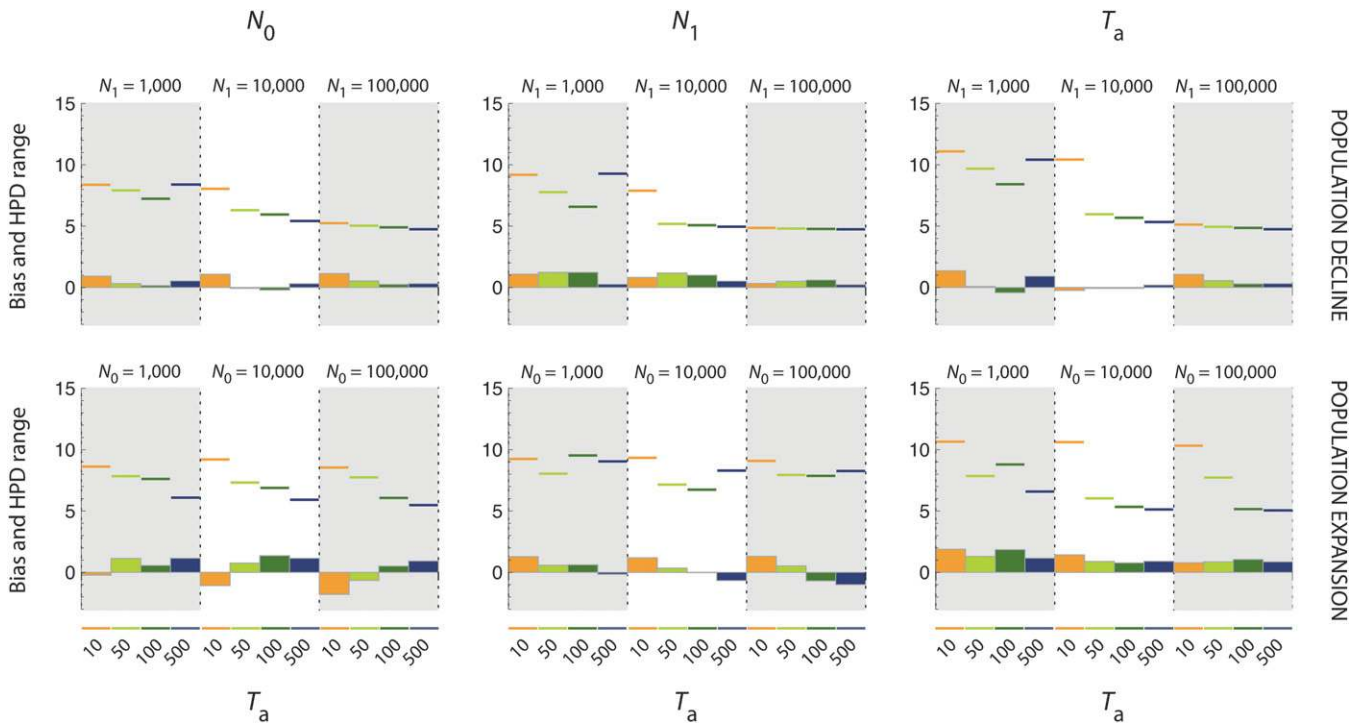
FIGURE 3.—Precision of the estimates of the natural demographic parameters N_0, N_1, and T_a. Bias (histograms) and absolute value of the range of the 90% HPD interval (horizontal colored traits) for natural demographic parameters N_0, N_1, and T_a (from left to right) are presented in a $\log_{10}$ scale. Top, population declines; bottom, population expansions. In each graph, the dotted vertical line separates scenarios of increasing severity ($N_1 = 1000$, $N_1 = 10,000$, and $N_1 = 100,000$ for population declines and $N_0 = 1000$, $N_0 = 10,000$, 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 colors (orange for $T_a = 10$, light green for $T_a = 50$, dark green for $T_a = 100$, and blue for $T_a = 500$).

noted that the marginal posterior distributions of θ_0 and θ_1 were skewed, respectively, to upper and lower values (Figure 4 and Figure S4). The 90% HPD intervals of the marginal posterior distributions for the time parameter t_f were wide in most conditions (Figure 4 and Figure 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 data sets for all scaled parameters in almost all situations, particularly for t_f.

Influence of the mutation model in Msvar: Of the 30 analyses presented in Figure 6 for the GSM, 12 (40%) did not converge after 3×10^9 steps. Out of these, 9 analyses (75%) concerned the data sets 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 data sets and a population expansion for four simulated data sets of five. However, Msvar detected a false signal of population decline for two data sets of five 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 esti-

mates of θ_1 and t_f, a slightly larger bias, and 90% HPD intervals for θ_0 compared to θ_1 in contraction scenarios and poorer estimates, with large variations across replicate data sets, in expansion scenarios (Figure S6). For stable population scenarios, both the strict SMM and the moderate GSM ($p = 0.22$) produced unbiased estimates of θ_0 and θ_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 data sets 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 data sets under a SMM, with only 5 significant tests 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 (CORNUET 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
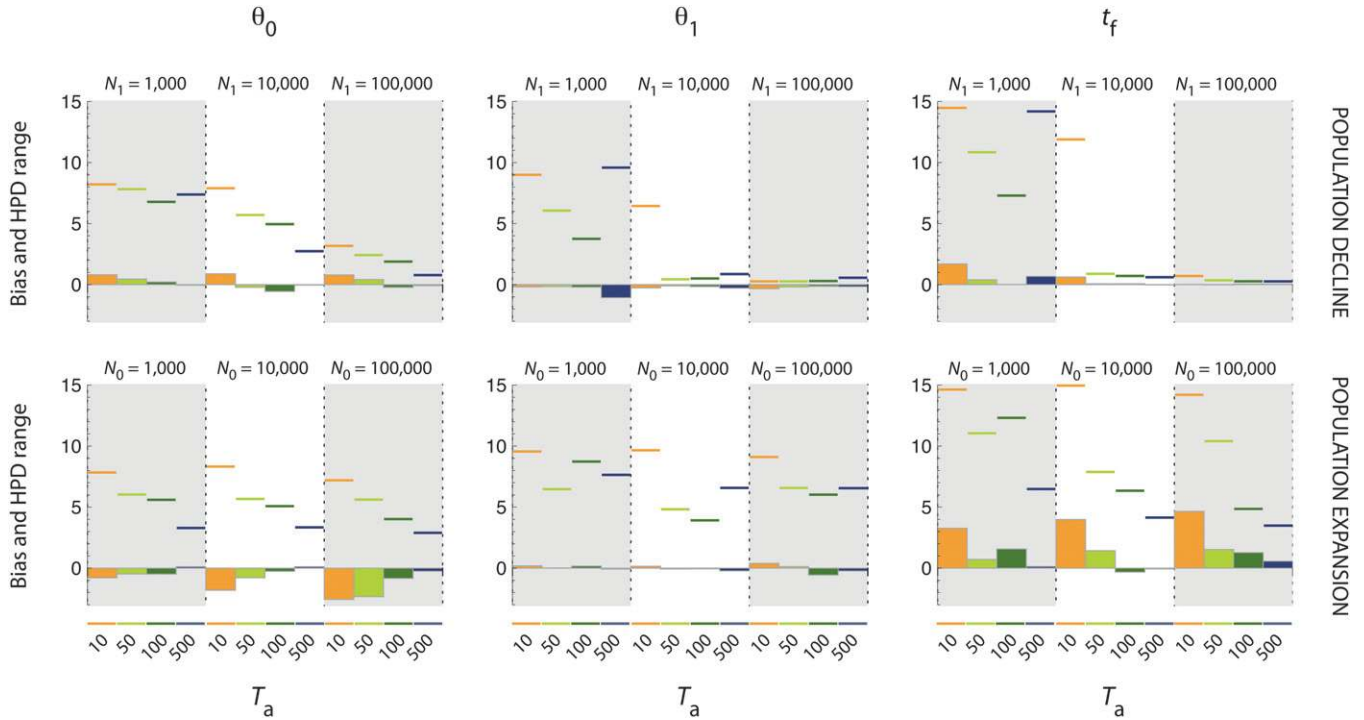
FIGURE 4.—Precision of the estimates of the scaled parameters θ_0, θ_1, and t_f. Bias (histograms) and absolute value of the range of the 90% HPD interval (horizontal colored traits) for scaled parameters θ_0, θ_1, and t_f (from left to right) are presented in a $\log_{10}$ scale. Top, population declines; bottom, population expansions. In each graph, the dotted vertical line separates scenarios of increasing severity ($N_1 = 1000$, $N_1 = 10,000$, and $N_1 = 100,000$ for population declines and $N_0 = 1000$, $N_0 = 10,000$, 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 colors (orange for $T_a = 10$, light green for $T_a = 50$, dark green for $T_a = 100$, and blue for $T_a = 500$).

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 data sets was
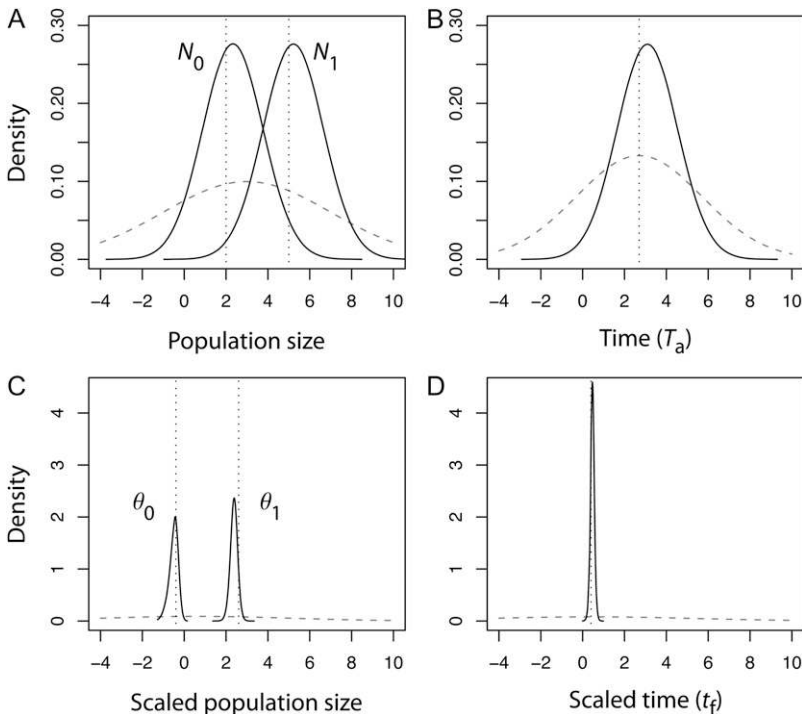

FIGURE 5.—Marginal posterior density of N_0, N_1, and T_a and θ_0, θ_1, and t_f for an ancient and severe population decline. All densities are represented in a $\log_{10}$ scale. (A) Population size natural parameters N_0 and N_1. (B) Time natural parameter T_a. (C) Scaled parameters θ_0 and θ_1. (D) 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 $\log_{10}$ 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 shaded dashed curve in each graph.
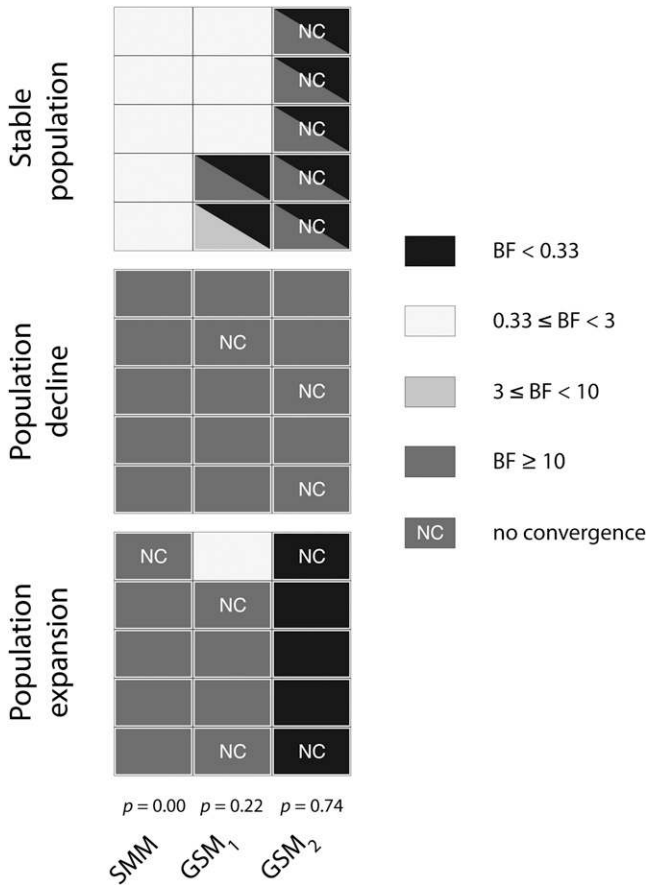
FIGURE 6.—Detection of change in population size with Msvar. The Bayes factors (BF) are given for each of the following demographic scenarios: a stable population (with $N_0 = N_1 = 464$, $T_a = 500$), a declining population (with $N_0 = 100$, $N_1 = 10,000$, and $T_a = 500$), and an expanding population (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 multistep mutation changes: $p = 0.00$ (stepwise mutation model, SMM), $p = 0.22$ (moderate generalized stepwise model, GSM$_1$), and $p = 0.74$ (strong generalized stepwise model, GSM$_2$). 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 indicate strong support, and BF ranging from 3 to 10 indicate substantial support. BF ranging from 0.33 to 3 indicate no support and values <0.33 indicate false detection of contraction or expansion. Nonconverged (NC) analyses are also indicated.

overall higher than in CORNUET and LUIKART (1996) who considered a variable mutation rate across loci and simulations, 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 data sets (32 significant tests 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 θ_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 θ_1. This procedure increased the statistical power of the *M*-ratio test. Since the true value of the parameter θ_1 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 data sets with $M_c = 0.68$ (*e.g.*, as in LEBLOIS *et al.* 2006) resulted in a lower rate of detection (22 significant tests 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 identified only 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 detected only 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
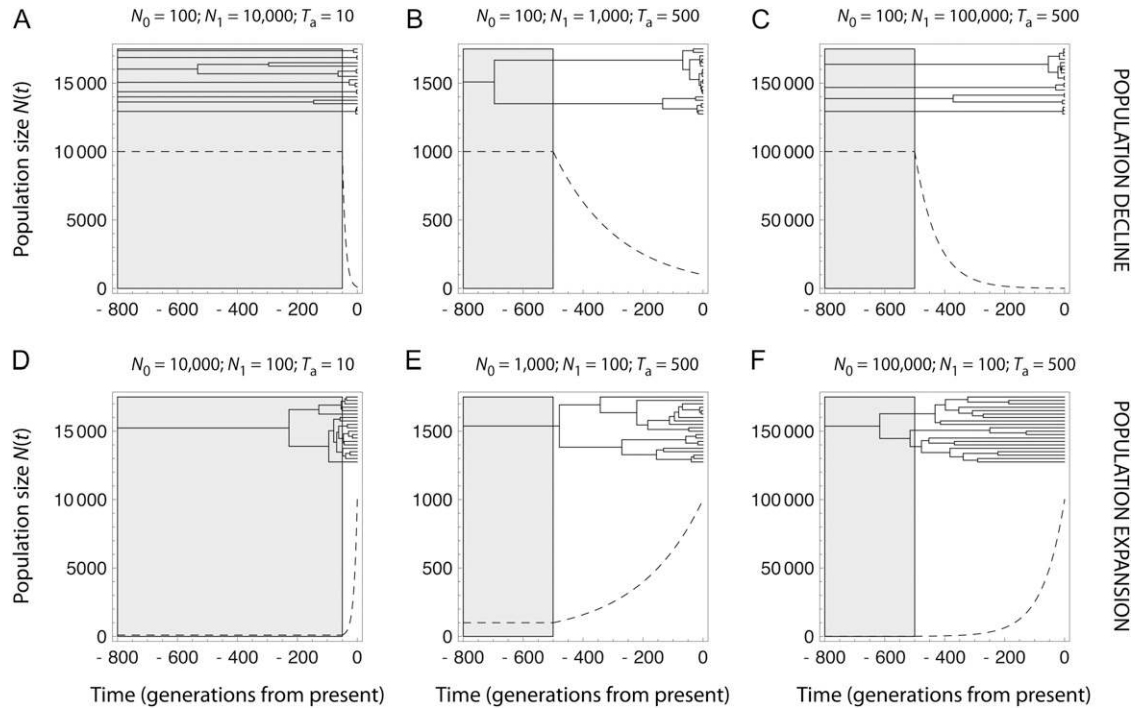
FIGURE 7.—Dynamics of population size changes $N(t)$ corresponding to simulated scenarios and expected gene genealogies. A–C corresponds to population decline and D–F to population expansion (dashed curve). The shaded area in each graph indicates when the ancestral population size is constant; *i.e.*, $N(t) = N_1$. 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.

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 (TAVARÉ *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, 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.*, σ_μ 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 θ_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 τ, from which a sample of n genes is descending. This

number has a known distribution in a constant-size population (TAVARÉ 1984), and LEBLOIS 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 $\Pr(j\,|\,n)$ that a sample of n genes has j ancestors T_a generations ago, we may compute the expected number m of lineages at the time of the population size change as

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

(1)

where $a_{(i)} \equiv a(a+1)\ldots(a+i-1)$, $a_{[i]} \equiv a(a-1)\ldots(a-i+1)$, and $\tau = \int_0^{T_a}(dt'/2N(t')) = ((1-N_0/N_1)/N_0 \, \mathrm{Log}(N_1/N_0))T_a$. 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 θ_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). In contrast, θ_1 was overall precisely estimated (see Figure 4 and Figure 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.*, Figure 7, B and C) at a rate that depends upon θ_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 θ_1 estimates would have 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 θ_1, this latter parameter is generally precisely and accurately estimated. Contrastingly, θ_0 can be precisely and accurately estimated only if the demographic event is severe and ancient. If the change in population size is too recent, provided that it is not too pronounced, θ_0 estimates tend to converge to the true value of θ_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 θ_0 and θ_1.

Scenarios of population expansion: Msvar was also very efficient for detecting expansions. Nevertheless, the estimates of the scaled current population size θ_0 were more severely biased and less precise, 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 Figure 7, A–C, to 7, D–F) and 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 θ_0 decreased with increasing T_a (Figure 4 and Figure 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 θ_0 and t_f reduces to a ridge along a line $\log_{10}(2\mu T_a) = k$ independent of θ_0 (Figure 4b in BEAUMONT 1999), which suggests that Msvar provides information on $2\mu T_a$, rather than on θ_0 in expanding populations. It is worth noting that WAKELEY *et al.* (2001) found similar results in inferring demographic history from single-nucleotide polymorphims (see a comparison in BEAUMONT 2004).

Estimates of θ_1 had a low bias but a large 90% HPD range (Figure 4). Although the marginal posterior distributions of θ_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 θ_1 may be equally likely, provided the genealogy is star shaped.

Influence of the underlying demographic model: Athough 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 (Figure 7, A–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 (Figure 7, D–F). Therefore, a large proportion of the coalescence events that occur in the growing population take place while $N(t)$ is close to N_1 (compare, *e.g.*, Figure 7C to 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 $N_1 = 10,000$ is 464, which is strictly equal to the harmonic mean of an exponentially growing population with $N_0 = 10,000$ and $N_1 = 100$. Hence, the harmonic mean of a declining population is closer to its current size (N_0) than its ancestral size (N_1), while the reverse is true for expanding populations. Therefore, given the exponential model of population growth, one might expect poor statistical properties of θ_0 estimates in expanding populations, 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 Figure 6 and Figure 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 ≥ 10), even in expanding populations (Figure 6 and Figure 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 (GONSER *et al.* 2000). Indeed, mutations that arise under a strong GSM involve large changes in allele length, which produce some gaps in the distribution of allele 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 Table 1 to Table S3).

Insights from empirical studies: The better performance of Msvar 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 data sets was not substantially different from that observed in empirical studies (compare Table 1 to Table 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. However, 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 θ_0, θ_1, and t_f. 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, to that of 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 (nonhierarchical) 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. However, 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, 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 performances of Msvar 0.4 and 1.3.

Finally, we recommend that inferences about population demographic change with Msvar should 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 Figure 6 and Figure S6), loci that evolve under a strong GSM may invalidate the approach by detecting false signals of population decline whatever the true demographic history. However, 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). Finally, further work is needed to evaluate how Msvar performs when the demography is more complex, *e.g.*, with successions of population declines and expansions.

LITERATURE CITED

Abdo, Z., K. A. Crandall and P. Joyce, 2004 Evaluating the performance of likelihood methods for detecting population structure and migration. Mol. Ecol. **13:** 837–851.

Beaumont, M. A., 1999 Detecting population expansion and decline using microsatellites. Genetics **153:** 2013–2029.

Beaumont, M. A., 2004 Recent developments in genetic data analysis: What can they tell us about human demographic history? Heredity **92:** 365–379.

Beaumont, M. A., and B. Rannala, 2004 The Bayesian revolution in genetics. Nat. Rev. Genet. **5:** 251–261.

Bonhomme, M., A. Blancher, S. Cuartero, L. Chikhi and B. Crouau-Roy, 2008 Origin and number of founders in an introduced insular primate: estimation from nuclear genetic data. Mol. Ecol. **17:** 1009–1019.

Brooks, S., and A. Gelman, 1998 General methods for monitoring convergence of iterative simulations. J. Comput. Graph. Stat. **7:** 434–455.

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.

Colautti, R. I., M. Manca, M. Viljanen, H. A. Ketelaars, H. Bürgi et al., 2005 Invasion genetics of the Eurasian spiny waterflea: evidence for bottlenecks and gene flow using microsatellites. Mol. Ecol. **14:** 1869–1879.

Comps, B., D. Gömöry, J. Letouzey, B. Thiébaut and R. J. Petit, 2001 Diverging trends between heterozygosity and allelic richness during postglacial colonization in the European Beech. Genetics **157:** 389–397.

Cornuet, J. M., and G. Luikart, 1996 Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics **144:** 2001–2014.

Cornuet, J.-M., F. Santos, M. A. Beaumont, C. P. Robert, J.-M. Marin et al., 2008 Inferring population history with DIY ABC: a user-friendly approach to Approximate Bayesian Computation. Bioinformatics **24:** 2713–2719.

Dib, C., S. Fauré, C. Fizames, D. Samson, N. Drouot et al., 1996 A comprehensive map of the human genome based on 5,264 microsatellites. Nature **380:** 152–154.

Drummond, A., and A. Rambaut, 2007 BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. **7:** 214.

Ellegren, H., 2000 Microsatellite mutations in the germline: implications for evolutionary inference. Trends Genet. **16:** 552–558.

Ellegren, H., 2004 Microsatellites: simple sequences with complex evolution. Nat. Rev. Genet. **5:** 435–445.

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. Proc. Natl. Acad. Sci. USA **106:** 13404–13409.

Emerson, B. C., E. Paradis and C. Thebaud, 2001 Revealing the demographic histories of species using DNA sequences. Trends Ecol. Evol. **16:** 707–716.

Estoup, A., I. J. Wilson, C. Sullivan, J.-M. Cornuet and C. Moritz, 2001 Inferring population history from microsatellite and enzyme data in serially introduced cane toads, *Bufo marinus*. Genetics **159:** 1671–1687.

Excoffier, L., G. Laval and S. Schneider, 2005 Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol. Bioinform. Online **1:** 47–50.

Felsenstein, J., 1992 Estimating effective population size from samples of sequences: inefficiency of pairwise and segregating sites as compared to phylogenetic estimates. Genet. Res. **59:** 139–147.

FitzSimmons, N. N., 1998 Single paternity of clutches and sperm storage in the promiscuous green turtle (*Chelonia mydas*). Mol. Ecol. **7:** 575–584.

Frankham, R., D. A. Briscoe and J. D. Ballou, 2002 *Introduction to Conservation Genetics.* Cambridge University Press, New York.

Garza, J. C., and E. G. Williamson, 2001 Detection of reduction in population size using data from microsatellite loci. Mol. Ecol. **10:** 305–318.

Gelman, A., J. B. Carlin, H. S. Stern and D. B. Rubin, 1995 *Bayesian Data Analysis.* Chapman & Hall, London.

Gelman, A., J. B. Carlin, H. S. Stern and D. B. Rubin, 2004 *Bayesian Data Analysis.* Chapman & Hall/CRC, New York.

Gonser, R., P. Donnelly, G. Nicholson and A. Di Rienzo, 2000 Microsatellite mutations and inferences about human demography. Genetics **154:** 1793–1807.

Griffiths, R. C., and S. Tavaré, 1994 Sampling theory for neutral alleles in a varying environment. Philos. Trans. R. Soc. Lond. **344:** 403–410.

Hastings, W. K., 1970 Monte Carlo sampling methods using Markov chains and their applications. Biometrika **57:** 97–109.

Hein, J., M. H. Schierup and C. Wiuf, 2005 *Gene Genealogies, Variation and Evolution. A Primer in Coalescent Theory.* Oxford University Press, Oxford.

Hey, J., and R. Nielsen, 2004 Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. Genetics **167:** 747–760.

Hey, J., and R. Nielsen, 2007 Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. Proc. Natl. Acad. Sci. USA **104:** 2785–2790.

Hu, F. S., A. Hampe and R. J. Petit, 2009 Paleoecology meets genetics: deciphering past vegetational dynamics. Front. Ecol. Environ. **7:** 371–379.

Hufbauer, R. A., S. M. Bogdanowicz and R. G. Harrison, 2004 The population genetics of a biological control introduction: mitochondrial DNA and microsatellite variation in native and introduced populations of *Aphidus ervi*, a parasitoid wasp. Mol. Ecol. **13:** 337–348.

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. Biol. J. Linn. Soc. **84:** 91–101.

Jeffreys, H., 1961 *Theory of Probability*, Ed. 3. Oxford University Press, Oxford.

Kass, R. E., and A. E. Raftery, 1995 Bayes factors. J. Am. Stat. Assoc. **90:** 773–795.

Kuhner, M. K., 2006 Lamarc 2.0: maximum-likelihood and Bayesian estimation of population parameters. Bioinformatics **22:** 768–770.

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.

Lawton-Rauh, A., 2008 Demographic processes shaping genetic variation. Curr. Opin. Plant Biol. **11:** 103–109.

Leblois, R., and M. Slatkin, 2007 Estimating the number of founder lineages from haplotypes of closely linked SNPs. Mol. Ecol. **16:** 2237–2245.

Leblois, R., A. Estoup and R. Streiff, 2006 Genetics of recent habitat contraction and reduction in population size: Does isolation by distance matter? Mol. Ecol. **15:** 3601–3615.

Loader, C., 1999 *Local Regression and Likelihood*. Springer-Verlag, New York.

Luikart, G., and J. M. Cornuet, 1998 Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. Conserv. Biol. **12:** 228–237.

Luikart, G., F. W. Allendorf, J.-M. Cornuet and W. B. Sherwin, 1998 Distortion of allele frequency distribution provides a test for recent population bottlenecks. J. Hered. **89:** 238–247.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, 1953 Equations of state calculations by fast computing machines. J. Chem. Phys. **21:** 1087–1092.

Milton, K., J. D. Lozier and E. A. Lacey, 2009 Genetic structure of an isolated population of mantled howler monkeys (*Alouatta palliata*) on Barro Colorado Island, Panama. Conserv. Genet. **10:** 347–358.

Nei, M., 1978 Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics **89:** 583–590.

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. Am. J. Bot. **96:** 2224–2233.

Nielsen, R., and M. Beaumont, 2009 Statistical inferences in phylogeography. Mol. Ecol. **18:** 1034–1047.

Nordborg, M., 2007 Coalescent theory, pp. 843–877 in *Handbook of Statistical Genetics*, Ed. 3, edited by D. Balding, M. Bishop and C. Cannings. John Wiley & Sons, Chichester, UK.

Peter, B. M., D. Wegmann and L. Excoffier, 2010 Distinguishing between population bottleneck and population subdivision by a Bayesian model choice procedure. Mol. Ecol. **19:** 4648–4660.

Plummer, M., N. Best, K. Cowles and K. Vines, 2006 Coda: output analysis and diagnostics for MCMC. R News **6:** 7–11.

Pope, L. C., A. Estoup and C. Moritz, 2000 Phylogeography and population structure of an ecotonal marsupial, *Bettongia tropica*, determined using mtDNA and microsatellites. Mol. Ecol. **9:** 2041–2053.

R Development Core Team, 2009 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.

Rousset, F., and R. Leblois, 2007 Likelihood and approximate likelihood analyses of genetic structure in a linear habitat: performance and robustness to model mis-specification. Mol. Biol. Evol. **24:** 2730–2745.

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. Mol. Ecol. **9:** 1517–1528.

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 Bayesian model. Mol. Biol. Evol. **19:** 1981–1990.

Strasburg, J. L., and L. H. Rieseberg, 2010 How robust are ''isolation with migration'' analyses to violations of the IM model? A simulation study. Mol. Biol. Evol. **27:** 297–310.

Tavaré, S., 1984 Line-of-descent and genealogical processes, and their applications in population genetics models. Theor. Popul. Biol. **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.

Wakeley, J., R. Nielsen, S. N. Liu-Cordero and K. Ardlie, 2001 The discovery of single-nucleotide polymorphisms—and inferences about human demographic history. Am. J. Hum. Genet. **69:** 1332–1347.

Williamson-Natesan, E. G., 2005 Comparison of methods for detecting bottlenecks from microsatellite loci. Conserv. Genet. **6:** 551–562.

Wilson, I. J., and D. J. Balding, 1998 Genealogical inference from microsatellite data. Genetics **150:** 499–510.

Wilson, I. J., M. Weale and D. J. Balding, 2003 Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities. J. R. Stat. Soc. Ser. A **166:** 155–188.

Wirth, T., F. Hildebrand, C. Allix-Béguec, F. Wölbeling, T. Kubica *et al.*, 2008 Origin, spread and demography of the *Mycobacterium tuberculosis* complex. PLoS Pathog. **4:** e1000160.

GENETICS

Inferring Population Decline and Expansion From Microsatellite Data: A Simulation-Based Evaluation of the Msvar Method

Christophe Girod, Renaud Vitalis, Raphaël Leblois and Hélène Fréville

FILE S1

Supporting Methods

The version 1.3 of MSVAR provides separate estimates for N_0, N_1, μ 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. Hence, the set of parameters of interest is given by $\boldsymbol{\Phi} \equiv \{N_{0i}, N_{1i}, T_{ai}, \mu_i\}_{i=1,k}$ for k loci and the prior distributions of these parameters depend upon hyper-prior distributions. Priors and hyper-priors were specified following STORZ and BEAUMONT (2002). For each locus, the prior distributions of the model parameters were assumed to be log-normal distributions, each with means (on a $\log_{10}$ scale) of $\mathbf{M} \equiv \{M_{N_0}, M_{N_1}, M_{T_a}, M_\mu\}$ and standard deviations (SDs) of $\mathbf{V} \equiv \{V_{N_0}, V_{N_1}, V_{T_a}, V_\mu\}$. Hyper-prior distributions for the means $\mathbf{M}$ were themselves assumed to be normal distributions with means α_{N_0}, α_{N_1}, α_{T_a}, α_μ and SDs σ_{N_0}, σ_{N_1}, σ_{T_a}, σ_μ. Prior means for the current and ancestral population sizes were set equal to the $\log_{10}$-transformed values $\alpha_{N_0} = \alpha_{N_1} = 3$, which amounts to consider population contraction and expansion as equally likely. To allow for uncertainty in these estimates, and to avoid a strong effect of the prior specification, we considered prior distribution with large SDs by setting $\sigma_{N_0} = \sigma_{N_1} = 4$. The prior mean for the time since the population started changing in size was set equal to 500 generations, giving $\alpha_{T_a} = 2.7$. The SD of the mean around the time of demographic change, $\sigma_{T_a} = 3$, was chosen so that more recent and more ancient dates were also supported. In order to test whether the method could retrieve information on the mutation rate from the data, we chose a relatively flat prior on the mean mutation rate per generation with $\alpha_\mu = -4$ and $\sigma_\mu = 2$. Last, the hyper-prior distributions for the SDs $\mathbf{V}$ were assumed to be normal distributions truncated at zero with means $\beta_{N_0} = \beta_{N_1} = \beta_{T_a} = \beta_\mu = 0$ and SDs $\tau_{N_0} = \tau_{N_1} = \tau_{T_a} = \tau_\mu = 0.5$. The parameter values reported in the main text correspond to the mean of N_0, N_1, T_a and μ, across loci.

Population decline
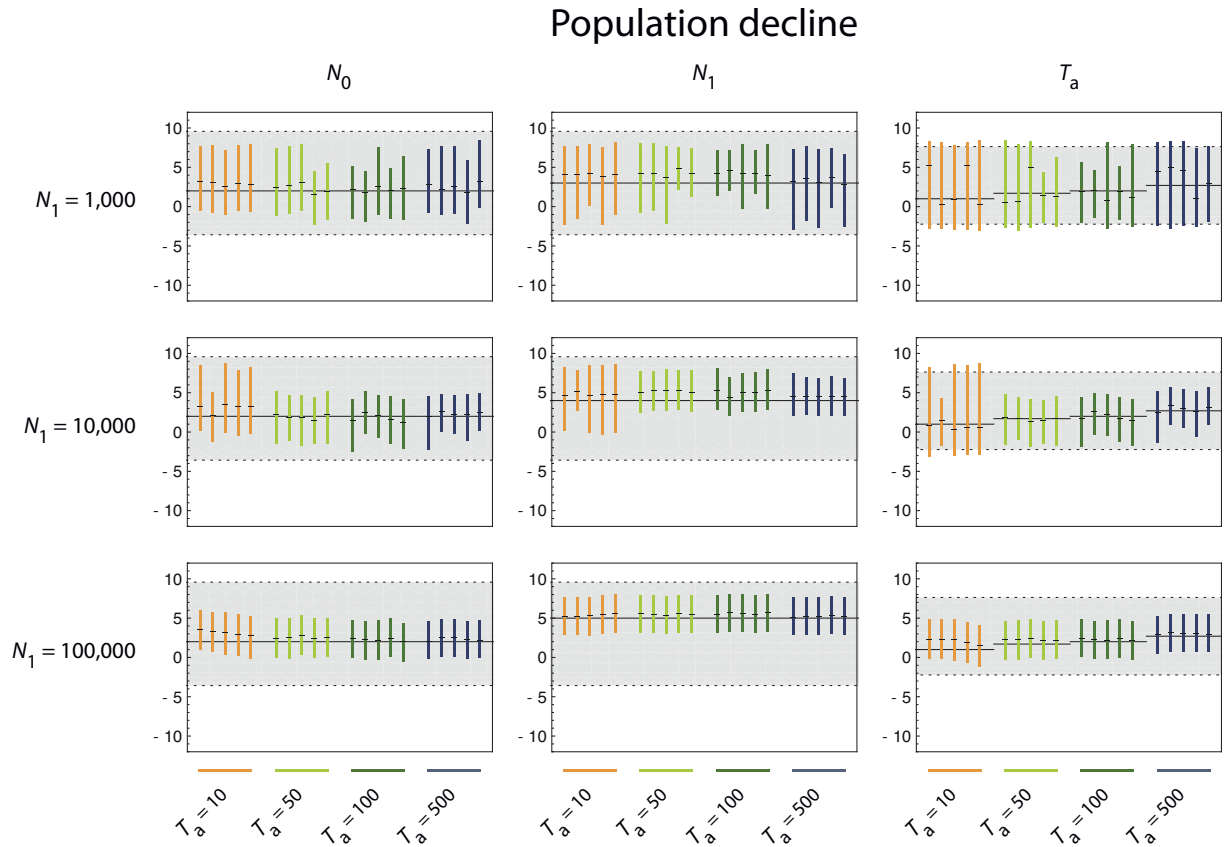
FIGURE S1.—Inference of the natural demographic parameters N_0, N_1 and T_a in declining populations. In each graph, the coloured vertical lines represent the 90% HPD interval in a $\log_{10}$ scale. The black horizontal trait over each coloured line represents the mode of the marginal posterior distribution of the parameters. Each colour stands for a value of the (unscaled) time parameter T_a (orange for $T_a = 10$, light green for $T_a = 50$, dark green for $T_a = 100$ and blue for $T_a = 500$). The grey area within the dotted lines in each graph represents the 90% support of the prior distribution of the parameters. The black continuous horizontal line gives the true value of the parameters.
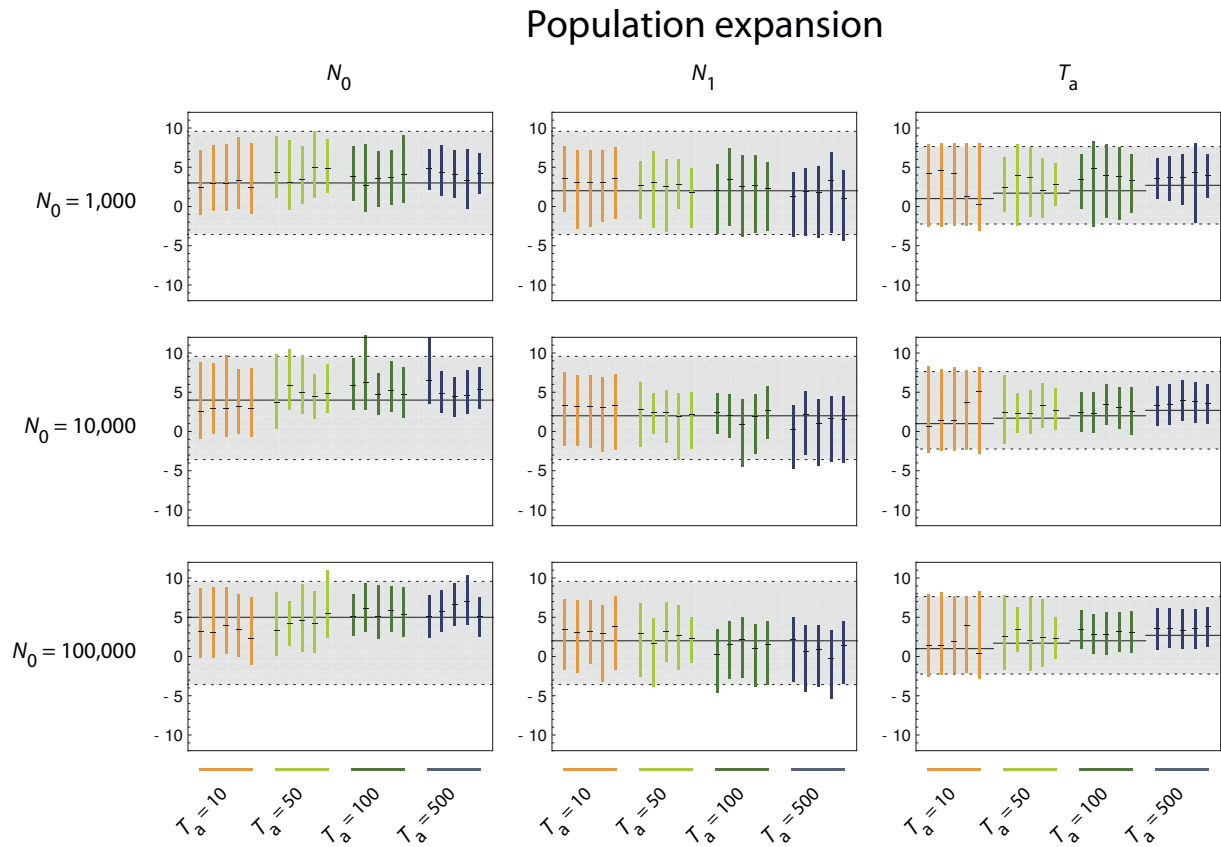
FIGURE S2.—Inference of the natural demographic parameters N_0, N_1 and T_a in expanding populations. In each graph, the coloured vertical lines represent the 90% HPD interval in a $\log_{10}$ scale. The black horizontal trait over each coloured line represents the mode of the marginal posterior distribution of the parameters. Each colour stands for a value of the (unscaled) time parameter T_a (orange for $T_a = 10$, light green for $T_a = 50$, dark green for $T_a = 100$ and blue for $T_a = 500$). The grey area within the dotted lines in each graph represents the 90% support of the prior distribution of the parameters. The black continuous horizontal line gives the true value of the parameters.

Population decline
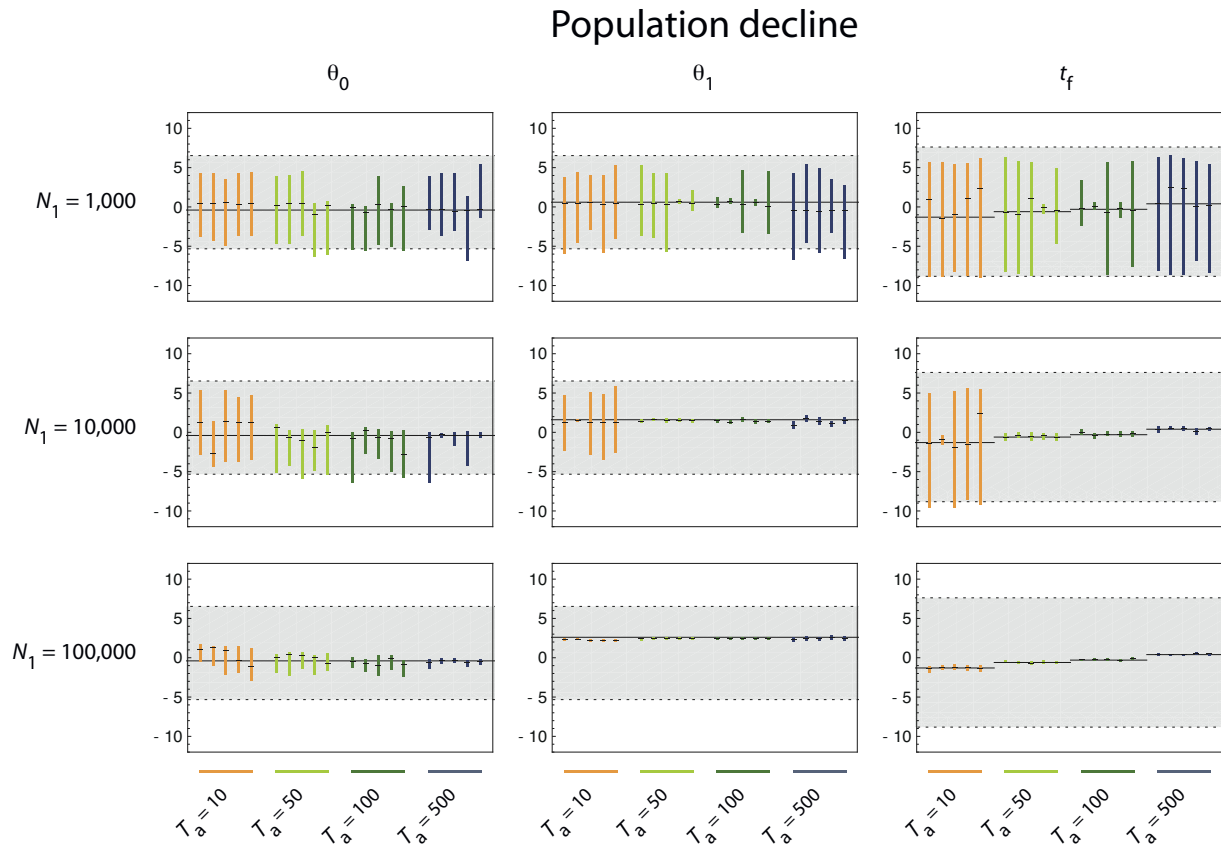
FIGURE S3.—Inference of the scaled demographic parameters θ_0, θ_1 and t_f in declining populations. In each graph, the coloured vertical lines represent the 90% HPD interval in a $\log_{10}$ scale. The black horizontal trait over each coloured line represents the mode of the marginal posterior distribution of the parameters. Each colour stands for a value of the (unscaled) time parameter T_a (orange for $T_a = 10$, light green for $T_a = 50$, dark green for $T_a = 100$ and blue for $T_a = 500$). The grey area within the dotted lines in each graph represents the 90% support of the prior distribution of the parameters. The black continuous horizontal line gives the true value of the parameters.

Population expansion
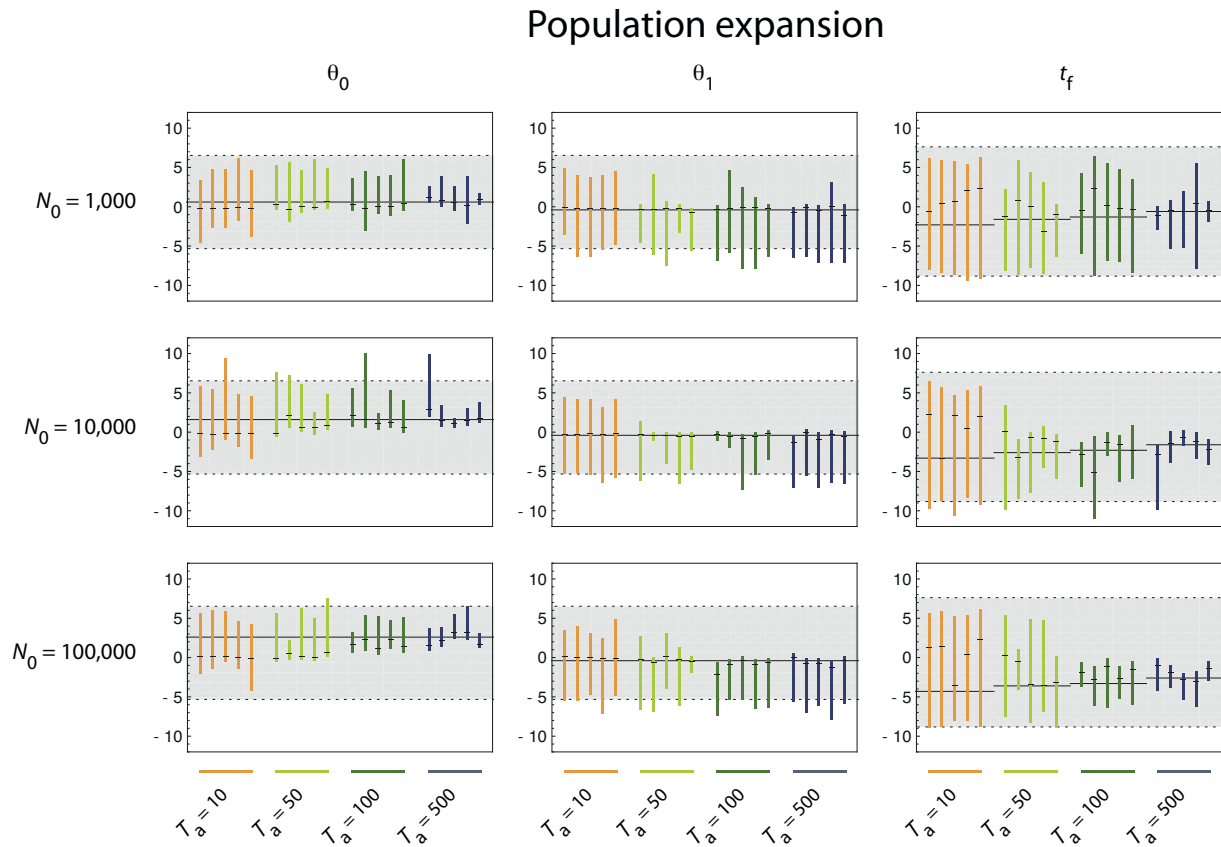
FIGURE S4.—Inference of the scaled demographic parameters θ_0, θ_1 and t_f in expanding populations. In each graph, the coloured vertical lines represent the 90% HPD interval in a $\log_{10}$ scale. The black horizontal trait over each coloured line represents the mode of the marginal posterior distribution of the parameters. Each colour stands for a value of the (unscaled) time parameter T_a (orange for $T_a = 10$, light green for $T_a = 50$, dark green for $T_a = 100$ and blue for $T_a = 500$). The grey area within the dotted lines in each graph represents the 90% support of the prior distribution of the parameters. The black continuous horizontal line gives the true value of the parameters.
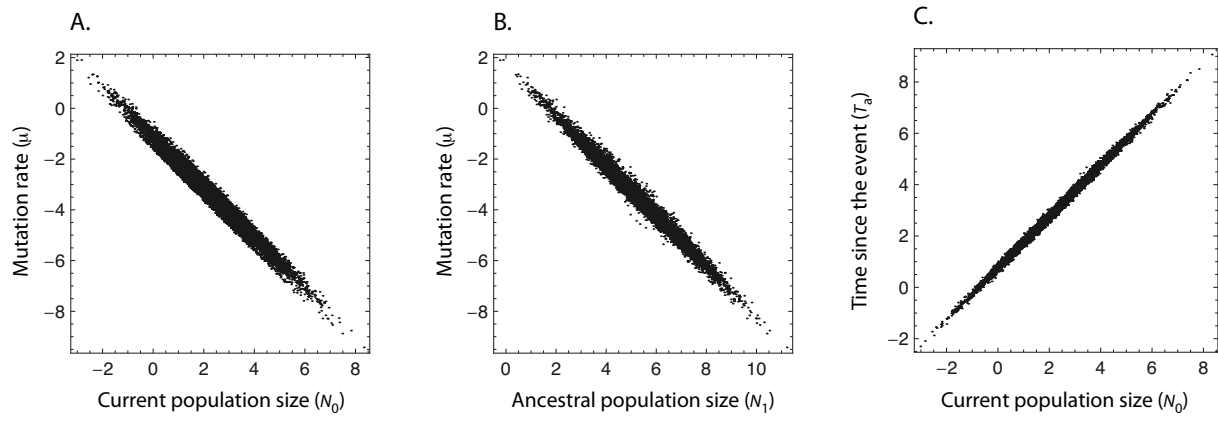
FIGURE S5.—Correlation between natural parameters: (A) current population size N_0 and mutation rate μ, (B) ancestral population size N_1 and mutation rate μ and (C) current population size N_0 and time since event T_a.
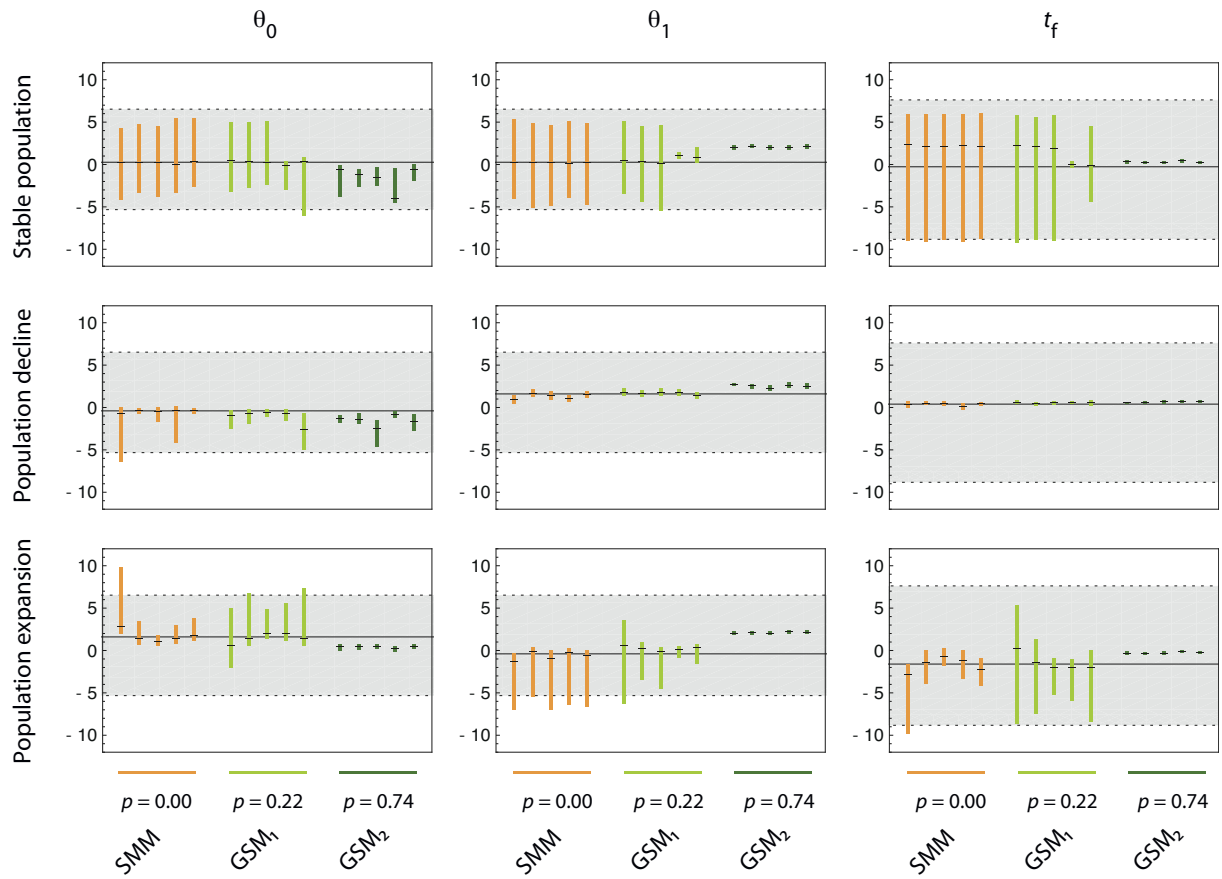
FIGURE S6.—Inference of the scaled demographic parameters θ_0, θ_1 and t_f in a stable (with $N_0 = N_1 = 464$, $T_a = 500$), a declining (with $N_0 = 100$, $N_1 = 10{,}000$ and $T_a = 500$), and an expanding (with $N_0 = 10{,}000$, $N_1 = 100$ and $T_a = 500$) populations. In each graph, the coloured vertical lines represent the 90% HPD interval in a $\log_{10}$ scale. The black horizontal trait over each coloured line represents the mode of the marginal posterior distribution of the parameters. Each colour stands for a value of p, the frequency of multi-step mutations changes: orange for $p = 0.00$ (stepwise mutation model, SMM), light green for $p = 0.22$ (moderate generalized stepwise model, GSM$_1$), dark green for $p = 0.74$ (strong generalized stepwise model, GSM$_2$). The grey area within the dotted lines in each graph represents the 90% support of the prior distribution of the parameters. The black continuous horizontal line gives the true value of the parameters.

TABLE S1

Exhaustive Review of the Studies that used MSVAR for Inference of Population Size Change

Species	s	n	No. of loci	H_e	MSVAR Vers.	Res.	M-ratio	Bottleneck	Reference
Mexican goodeid fish	2	20 − 31	5 − 7	-	v. 0.4	↘	-	-	Bailey *et al.* (2007)
Himalayan brown bear	1	54	6	-	v. 0.4	↘	-	-	Bellemain *et al.* (2007)
Ground beetle	2	54 − 56	9	-	v. 0.4	↘	-	-	Keller *et al.* (2005)
European grayling	14	13 − 48	8	0.62 − 0.69	v. 0.4	↘	-	-	Meldgaard *et al.* (2003)
	4	28 − 52	17	-	v. 0.4	↘	-	-	Koskinen *et al.* (2002a)
Arctic grayling	1	71	-	-	v. 0.4	↘	-	-	Koskinen *et al.* (2002b)
	6	80 − 214	7	-	v. 0.4	↘	-	-	Stamford and Taylor (2005)
Northern pike	16	15 − 50	5	-	v. 0.4	↘	-	-	Jacobsen et al. (2005)
	2	26	5	-		↗			
Japanese eel	1	89	6	-	v. 0.4	↘	-	-	Tseng *et al.* (2003)
American and European eels	4	100	7	-	v. 0.4	↘	-	-	Wirth and Bernatchez (2003)
Fishtail palm	1	143	9	0.67	v. 1.3	↘	-	-	Cibrian-Jaramillo *et al.* (2009)
Lake Victoria Cichlid fishes	3	-	12	-	v. 0.4	↘	-	-	Elmer *et al.* (2009)
Cape Fear Shiner	2	26 − 29	18	-	v. 1.3	↘	-	-	Saillant *et al.* (2004)
Gray snapper	3	50	13	-	v. 1.3	↘	-	-	Gold *et al.* (2009)
Lane snapper	6	50	13	-	v. 1.3	↘	-	-	Karlsson *et al.* (2009)
Drosophila sp.	12	22 − 30	47	0.10 − 0.51	v. 0.4	↘	-	-	Harr and Schlötterer (2004)
Caenorhabditis elegans	1	-	9	-	v. 0.4	↘	-	-	Sivasundar and Hey (2003)
Red deer	3	31 − 33	11	0.37 − 0.50	v. 1.3	↘	-	-	Nielsen et al. (2008)
Mycobacterium complex	8	-	24	-	v. 0.4	↗	-	-	Wirth et al. (2008)

Species									Reference
Sea otter	1	40	24	-	v. 1.3	↘	- a	- a	AGUILAR ET AL. (2008)
Bornean orang-utan	2	26 – 27	14	-	both	↘	-	- a	GOOSSENS ET AL. (2006)
Tanoak	1	447	9	-	v. 0.4	↘	-	- a	NETTEL ET AL. (2009)
Rattlesnake	5	18 – 54	9	0.61 – 0.73	v. 0.4	↘	↘	-	HOLYCROSS ET AL. (2007)
Parasitoid wasp		50	5	-	v. 1.3	↘	-	- a	HUFBAUER ET AL. (2004)
		48	5	-		↗	-	- a	
Cynomolgus macaque	1	81	16	0.66	v. 0.4	=	-	↗	BONHOMME ET AL. (2008)
Madagascar fish-eagle	1	44	22	0.19	v. 0.4	↘	-	=	JOHNSON ET AL. (2009)
African elephant	5	80 – 319	20	-	v. 1.3	↘	-	↘	OKELLO ET AL. (2008)
	1	79	20	0.74		↘	-	=	
Giant panda	2	29 – 32	9	0.49 – 0.56	v. 1.3	↘	-	=	ZHANG ET AL. (2007)
	1	40	9	0.61		↘	-	↗	
Eurasian otter	2	65 – 132	10	0.53 – 0.59	v. 1.3	↘	-	=	HAJKOVA ET AL. (2007)
	2	29 – 58	11	-	v. 0.4	↘	-	=	PERTOLDI ET AL. (2001)
Ethiopian walia ibex	1	24	5	0.35	v. 1.3	↘	-	=	GEBREMEDHIN ET AL. (2009)
Persian wild ass	1	24	12	0.54	v. 1.3	↘	-	=	NIELSEN ET AL. (2007)
Rock ptarmigan	3	17 – 20	6	0.45 – 0.75	v. 1.3	↘	-	=	PRUETT *et al.* (2010)
Cyprinid fish	4	30 – 48	6	0.23 – 0.35	v. 1.3	↘	-	=	SOUSA *et al.* (2008)
	2	21 – 30	6	0.24 – 0.26		=	-	=	
	6	12 – 50	6	0.22 – 0.45	v. 1.3	↘	-	=	SOUSA *et al.* (2010)
Golden-brown mouse lemur	8	15 – 59	8	0.54 – 0.65	both	↘	-	=	OLIVIERI *et al.* (2008)
Bongolava	1	27	8	0.57	both	↘	-	=	

mouse lemur

Species	s	n	No. of loci	H_e	Vers.	Res.			Reference
Danfoss' mouse lemur	1	30	8	0.71	both	↘	-	=	
Milne-Edwards's sportive lemur	2	10	14	$0.53 - 0.55$	both	↘	-	=	CRAUL *et al.* (2009)
Golden eagle	1	172	13	0.48	both	↘	-	=	BOURKE *et al.*(2010)
Yunnan sub-nosed monkey	1	135	10	0.70	v. 1.3	=	-	=	LIU *et al.* (2009)
Drosophila	1	13	17	0.80	v. 0.4	↘	-	↘	DIERINGER ET AL. (2005)
	4	$10 - 30$	17	$0.47 - 0.79$		↘	-	=	
	10	$8 - 30$	17	$0.48 - 0.79$		↘	-	↗	
	1	20	10	-	v. 0.4	↘	-	=	FRYDENBERG ET AL. (2002)
	1	20	10	-		↘	-	↗	
	1	20	10	-		↗	-	=	
	1	20	10	-		↗	-	↗	
	1	20	10	-		=	-	=	
Reed warblers	1	40	9	-	v. 1.3	↘	=	=	PROCHAZKA ET AL. (2008)
Tiger	1	57	30	-	both	↘	↘	↘	MONDOL ET AL. (2009)
European otter	1	6	11	0.77	v. 0.4	=	=	=	RANDI *et al.* (2003)
	7	$3 - 29$	11	$0.45 - 0.74$		↘	=	=	
European grayling	30	$17 - 112$	$13 - 14$	$0.24 - 0.64$	v. 1.3	↘	↘	=	SWATDIPONG *et al.* (2010)
	1	35	13	-	v. 1.3	↘	↘	↘	
Eastern red-blacked salamander	2	$25 - 28$	6	$0.41 - 0.47$	v. 0.4	↘	↘	=	JORDAN *et al.* (2009)
European wolf	2	$30 - 103$	18	-	v. 0.4	↘	↘	=	LUCCHINI *et al.* (2004)
	3	$34 - 115$	18	-		↘	=	=	
African buffalo	2	$33 - 54$	17	$0.81 - 0.82$	v. 1.3	↘	↘	=	HELLER *et al.* (2008)
Howler monkeys	1	50	10	0.58	v. 1.3	↘	=	↗	MILTON *et al.* (2009)

s, number of samples; *n*, number of individuals per sample; No. of loci, number of microsatellite loci used; H_e , expected heterozygosity; Vers., version of MSVAR used in the study (v. 0.4, v. 1.3 or both); Res., MSVAR results, as interpreted by the authors; ↘, evidence for population decline; =, no evidence of population size change; ↗, evidence of population expansion; -

, information not available; all studies using the M-ratio test calculated the critical value M_c (see main text), except that of MILTON *et al.* (2009), which considered $M_c = 0.68$; [a], data not available, since the M-ratio test and/or BOTTLENECK were not applied on the same set of populations as MSVAR. When multiple mutation models have been considered for the M-ratio test and/or BOTTLENECK, only the results assuming a SMM are reported.

Literature Cited

AGUILAR, A., D. A. JESSUP, J. ESTES and J. C. GARZA, 2008 The distribution of nuclear genetic variation and historical demography of sea otters. Animal Conservation **11:** 35-45.

BAILEY, N. W., C. MACIAS GARCIA and M. G. RITCHIE, 2007 Beyond the point of no return? A comparison of genetic diversity in captive and wild populations of two nearly extinct species of Goodeid fish reveals that one is inbred in the wild. Heredity **98:** 360-367.

BELLEMAIN, E., M. A. NAWAZ, A. VALENTINI, J. E. SWENSON and P. TABERLET, 2007 Genetic tracking of the brown bear in northern Pakistan and implications for conservation. Biological Conservation **134:** 537-547.

BONHOMME, M., A. BLANCHER, S. CUARTERO, L. CHIKHI and B. CROUAU-ROY, 2008 Origin and number of founders in an introduced insular primate: estimation from nuclear genetic data. Molecular Ecology **17:** 1009-1019.

BOURKE, B. P., A. C. FRANTZ, C. P. LAVERS, A. DAVISON, D. DAWSON *et al.*, 2010 Genetic signatures of population change in the British golden eagle (*Aquila chrysaetos*). Conservation Genetics **in press**.

CIBRIAN-JARAMILLO, A., C. D. BACON, N. C. GARWOOD, R. M. BATEMAN, M. M. THOMAS *et al.*, 2009 Population genetics of the understory fishtail palm *Chamaedorea ernesti-augusti* in Belize: high genetic connectivity with local differentiation. BMC Genetics **10:** 65-82.

CRAUL, M., L. CHIKHI, V. SOUSA, G. L. OLIVIERI, A. RABESANDRATANA *et al.*, 2009 Influence of forest fragmentation on an endangered large-bodied lemur in northwestern Madagascar. Biological Conservation **142:** 2862-2871.

DIERINGER, D., V. NOLTE and C. SCHLÖTTERER, 2005 Population structure in African *Drosophila melanogaster* revealed by microsatellite analysis. Molecular Ecology **14:** 563-573.

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 Science **106:** 13404-13409.

FRYDENBERG, J., C. PERTOLDI, J. DAHLGAARD and V. LOESCHCKE, 2002 Genetic variation in original and colinizing *Drosophila buzzatii* populations analysed by microsatellite loci isolated with a new PCR screening method. Molecular Ecology **11:** 181-190.

GEBREMEDHIN, B., G. F. FICETOLA, S. NADERI, H.-R. REZAEI, C. MAUDET *et al.*, 2009 Combining genetic and ecological data to assess the conservation status of the endangered Ethiopian walia ibex. Animal Conservation **12:** 89-100.

GOLD, J. R., E. SAILLANT, N. D. EBELT and S. LEM, 2009 Conservation genetics of Gray Snapper (*Lutjanus griseus*) in U.S. Waters of the Northern Gulf of Mexico and Western Atlantic Ocean. Copeia **2:** 277-286.

GOOSSENS, B., L. CHIKHI, M. ANCRENAZ, I. LACKMAN-ANCRENAZ, P. ANDAU *et al.*, 2006 Genetic signature of anthropogenic population collapse in orang-utans. PLoS Biology **4:** 285-291.

HAJKOVA, P., C. PERTOLDI, B. ZEMANOVA, K. ROCHE, B. HAJEK *et al.*, 2007 Genetic structure and evidence for recent population decline in Eurasian otter populations in the Czech and Slovak Republics: implications for conservation. Journal of Zoology **272:** 1-9.

HARR, B., and C. SCHLÖTTERER, 2004 Patterns of microsatellite variability in the *Drosophila melanogaster* complex. Genetica **120:** 71-77.

HELLER, R., E. D. LORENZEN, J. B. A. OKELLO, C. MASEMBE and H. R. SIEGISMUND, 2008 Mid-Holocene decline in African buffalos inferred from Bayesian coalescent-based analyses of microsatellites and mitochondrial DNA. Molecular Ecology **17:** 4845-4858.

HOLYCROSS, A. T., and M. E. DOUGLAS, 2007 Geographic isolation, genetic divergence and ecologicalnon-exchangeability define ESUs in a threatened sky-island rattlesnake. Biological conservation **134:** 142-154.

HUFBAUER, R. A., S. M. BOGDANOWICZ and R. G. HARRISON, 2004 The population genetics of a biological control introduction: mitochondrial DNA and microsatellite variation in native and introduced populations of *Aphidus ervi*, a parasitoid wasp. Molecular Ecology **13:** 337-348.

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.

JOHNSON, J. A., R. E. TINGAY, M. CULVER, F. HAILER, M. L. CLARKE *et al.*, 2009 Long-term survival despite low genetic diversity in the critically endangered Madagascar fish-eagle. Molecular Ecology **18:** 54-63.

JORDAN, M. E., D. A. MORRIS and S. E. GIBSON, 2009 The influence of historical landscape change on genetic variation and population structure of a terrestrial salamander (Plethodon cinereus). Conservation Genetics **10:** 1647-1658.

KARLSSON, S., E. SAILLANT and J. R. GOLD, 2009 Population structure and genetic variation of lane snapper (*Lutjanus synagris*) in the northern Gulf of Mexico. Marine Biology **156:** 1841-1855.

KELLER, I., L. EXCOFFIER and C. R. LARGIADER, 2005 Estimation of effective population size and detection of a recent population decline coinciding with habitat fragmentation in a ground beetle. Journal of Evolutionary Biology **18:** 90-100.

KOSKINEN, M. T., T. O. HAUGEN and C. R. PRIMMER, 2002a Contemporary fisherian life-history evolution in small salmonid populations. Nature **419:** 826-830.

KOSKINEN, M. T., I. KNIZHIN, C. R. PRIMMER, C. SCHLÖTTERER and S. WEISS, 2002b Mitochondrial and nuclear DNA phylogeography of *Thymallus* spp. (grayling). Molecular Ecology **11:** 2599-2611.

LIU, Z., B. REN, R. WU, L. ZHAO, Y. HAO *et al.*, 2009 The effect of landscape features on population genetic structure in Yunnan snub-nosed monkeys (*Rhinopithecus bieti*) implies an anthropogenic genetic discontinuity. Molecular Ecology **18:** 3831-3846.

LUCCHINI, V., A. GALOV and E. RANDI, 2004 Evidence of genetic distinction and long-term population decline in wolves (*Canis lupus*) in the Italian Apennines. Molecular Ecology **13:** 523-536.

MELDGAARD, T., E. E. NIELSEN and V. LOESCHCKE, 2003 Fragmentation by weirs in a riverine system: A study of genetic variation in time and space among populations of European grayling (Thymallus thymallus) in a Danish river system. Conservation Genetics **4:** 735-747.

MILTON, K., J. D. LOZIER and E. A. LACEY, 2009 Genetic structure of an isolated population of mantled howler monkeys (*Alouatta palliata*) on Barro Colorado Island, Panama. Conservation Genetics **10:** 347-358.

MONDOL, S., K. U. KARANTH and U. RAMAKRISHNAN, 2009 Why the Indian subcontinent holds the key to global tiger recovery. PLoS Genetics **5:** e1000585.

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. Am. J. Bot. **96:** 2224-2233.

NIELSEN, E. K., C. R. OLESEN, C. PERTOLDI, P. GRAVLUND, J. S. F. BARKER *et al.*, 2008 Genetic structure of the Dansih red deer (*Cervus elaphus*). Biological Journal of the Linnean Society **95:** 688-701.

NIELSEN, R. K., C. PERTOLDI and V. LOESCHCKE, 2007 Genetic evaluation of the captive breeding program of the Persian wild ass. Journal of Zoology **272:** 349-357.

OKELLO, J. B. A., G. WITTEMYER, H. B. RASMUSSEN, P. ARCTANDER, S. NYAKAANA *et al.*, 2008 Effective population size dynamics reveal impacts of historic climatic events and recent anthropogenic pressure in African elephants. Molecular Ecology **17:** 3788-3799.

OLIVIERI, G. L., V. SOUSA, L. CHIKHI and U. RADESPIEL, 2008 From genetic diversity and structure to conservation: genetic signature of recent population declines in three mouse lemur species (*Microcebus* spp.). Biological Conservation **141:** 1257-1271.

PERTOLDI, C., M. M. HANSEN, V. LOESCHCKE, A. B. MADSEN, L. JACOBSEN *et al.*, 2001 Genetic consequences of population decline in the European otter (*Lutra lutra*): an assessment of microsatellite DNA variation in Danish otters from 1883 to 1993. Proceedings of the royal society B **268:** 1775-1781.

PROCHAZKA, P., E. BELLINVIA, D. FAINOVA, P. HAJKOVA, A. ELHALAH *et al.*, 2008 Immigration as a possible rescue of a reduced population of a long-distant migratory bird: Reed warblers in the Azraq Oasis, Jordan. Journal of Arid Environments **72:** 1184-1192.

PRUETT, C. L., T. N. TURNER, C. M. TOPP, S. V. ZAGREBELNY and K. WINKER, 2010 Divergence in an archipelago and its conservation consequences in Aleutian Island rock ptarmigan. Conservation Genetics **11:** 241-248.

RANDI, E., F. DAVOLI, M. PIERPAOLI, C. PERTOLDI, A. B. MADSEN *et al.*, 2003 Genetic structure in otter (*Lutra lutra*) populations in Europe: implications for conservation. Animal Conservation **6:** 93-100.

SAILLANT, E., J. C. PATTON, K. E. ROSS and J. R. GOLD, 2004 Conservation genetics and demographic history of the endangered Cape Fear shiner (*Notropis mekistocholas*). Molecular Ecology **13:** 2947-2958.

SIVASUNDAR, A., and J. HEY, 2003 Population genetics of Caenorhabditis elegans: the paradox of low polymorphism in a widespread species. Genetics **163:** 147-157.

SOUSA, V., F. PENHA, M. J. COLLARES-PEREIRA, L. CHIKHI and M. M. COELHO, 2008 Genetic structure and signature of population decrease in the critically endangered freshwater cyprinid *Chondrostoma lusitanicum*. Conservation Genetics **9:** 791-805.

SOUSA, V., F. PENHA, I. PALA, L. CHIKHI and M. M. COELHO, 2010 Conservation genetics of a critically endangered Iberian minnow: evidence of population decline and extirpations. Animal Conservation **13:** 162-171.

STAMFORD, M. D., and E. B. TAYLOR, 2005 Population subdivision and genetic signatures of demographic changes in Arctic grayling (*Thymallus arcticus*) from an impounded watershed. Canadian Journal of Fisheries and Aquatic Sciences **62:** 2548-2559.

SWATDIPONG, A., C. R. PRIMMER and A. VASEMÄGI, 2010 Historical and recent genetic bottlenecks in European grayling, *Thymallus thymallus*. Conservation Genetics **11:** 279-292.

TSENG, M.-C., W.-N. TZENG and S.-C. LEE, 2003 Historical decline in the Japanese eel *Anguilla japonica* in northern Taiwan inferred from temporal genetic variations. Zoological Studies **42:** 556-563.

WIRTH, T., and L. BERNATCHEZ, 2003 Decline of north Atlantic eels: a fatal synergy? Proceedings of the royal society of London B **270:** 681-688.

WIRTH, T., F. HILDEBRAND, C. ALLIX-BÉGUEC, F. WÖLBELING, T. KUBICA *et al.*, 2008 Origin, spread and demography of the *Mycobacterium tuberculosis* complex. PLoS Pathogens **4:** e1000160.

ZHANG, B., M. LI, Z. ZHANG, B. GOOSSENS, L. ZHU *et al.*, 2007 Genetic viability and population history of the Giant Panda, putting an end to the "evolutionary dead end"? Molecular Biology and Evolution **24:** 1801-1810.

TABLE S2

Values of the Gelman - Rubin Shrink Factor

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

	$N_1 = 1,000$			$N_1 = 10,000$			$N_1 = 100,000$		
$T_a = 10$	1.01	-	-	1.01	-	-	2.09	2.10	1.12 NC
	1.01	-	-	1.08*	1.09	1.01	1.85	2.03	1.47 NC
	1.03	-	-	1.09	-	-	2.73	2.56	2.18 NC
	1.01	-	-	1.09	-	-	1.56	1.32	2.14 NC
	1.02	-	-	1.02	-	-	2.24	1.26	1.25 NC
$T_a = 50$	1.03	-	-	2.57	1.19	1.08	1.98	1.92	1.15 NC
	1.01	-	-	1.49	1.28	1.03	1.75	1.32	2.41 NC
	1.02	-	-	1.67	1.5	1.05	2.58	4.30	1.81 NC
	1.05	-	-	1.28	1.27	1.03	1.67	1.56	1.18 NC
	1.01	-	-	2.56	1.26	1.07	2.85	1.32	1.04
$T_a = 100$	1.03	-	-	2.4	1.47	1.02	1.97	1.66	1.02
	1.01	-	-	1.2	1.08	-	1.53	3.10	1.29 NC
	1.01	-	-	3.17	1.32	1.47 NC	2.37	4.59	1.44 NC
	1.04	-	-	1.69	1.63	1.33 NC	1.79	3.58	2.24 NC
	1.04	-	-	1.14	1.15	1.02	2.80	2.68	1.52 NC
$T_a = 500$	2.18	1	-	1.04	-	-	1.70	1.04	-
	1	-	-	1.04	-	-	1.26	1.12	1.01
	1	-	-	1.07	-	-	1.17	1.01	-
	1.02	-	-	1.63	1.27	1.03	1.10	-	-
	1.06	-	-	1.03	-	-	1.16	1.04	-

B. Expansions, with $N_1 = 100$ (SMM)

	$N_0 = 1,000$			$N_0 = 10,000$			$N_0 = 100,000$		
$T_a = 10$	1	-	-	1.02	-	-	1.11	1.00	-
	1.01	-	-	1.02	-	-	1.06	-	-
	1.01	-	-	1.02	-	-	1.06	-	-
	1.07	-	-	1.01	-	-	1.00	-	-
	1.01	-	-	1.01	-	-	1.00	-	-
$T_a = 50$	1.16	1.02	-	1.06	-	-	1.03	-	-
	1.01	-	-	1.3	1.21	1.05	1.04	-	-
	1.03	-	-	1.37	1.02	-	1.18	1.04	-
	1.08	-	-	1.07	-	-	1.02	-	-
	1.62	1.09	-	1.05	-	-	1.10	1.30	1.10
$T_a = 100$	1.04	-	-	1.05*	1.04	-	1.10	-	-
	1.02	-	-	1.08*	1.13	1.88 NC	1.20	1.08	-
	1	-	-	1.05	-	-	1.28	1.03	-
	1.01	-	-	1.61	1.43	1.01	1.21	1.07	-

	1.11	1.08	-	1.09	-	-	3.13	1.08	-
$T_a = 500$	1.07	-	-	1.13	3.71	1.97 NC	1.23	1.02	-
	1.08	-	-	1.22	1.06	-	2.11	1.02	-
	1.06	-	-	1.03	-	-	1.67	1.07	-
	1.02	-	-	1.03	-	-	2.90	1.16	1.06
	1.05	-	-	1.05	-	-	1.07	-	-

N_0, ancestral effective population size; N_1, current effective population size; T_a, time since the population size change. For each value of N_1 or N_0, the first column corresponds to a run length of 10^9 steps, the second column to 3 x 10^9 steps (if performed) and the third one to 1.5 x 10^{10} steps (if performed); *, chains that have not reached stationarity using the GELMAN - RUBIN diagnostic plot, despite having GELMAN - RUBIN shrink factor < 1.10; NC, non-converged Markov chains.

TABLE S3

Genetic Diversity in the Second Set of Simulated Data

	Declining population				Stable population				Expanding population			
	H_e	N_a	A_r	V_a	H_e	N_a	A_r	V_a	H_e	N_a	A_r	V_a
SMM	0.56	4.2	6.9	13.5	0.54	3.8	3.0	2.0	0.52	4.3	3.3	1.98
$p = 0.00$	(0.04)	(0.4)	(1.3)	(5.6)	(0.07)	(0.3)	(0.2)	(0.3)	(0.04)	(0.4)	(0.4)	(0.30)
GSM	0.51	3.9	9.5	30.9	0.57	4.8	4.8	4.4	0.52	5.7	5.3	4.2
$p = 0.22$	(0.04)	(0.2)	(1.6)	(16.2)	(0.08)	(0.7)	(1.4)	(2.3)	(0.04)	(0.1)	(0.4)	(0.6)
GSM	0.56	4.44	32.9	323.0	0.62	6.7	16.9	43.5	0.62	12.1	20.5	38.2
$p = 0.74$	(0.08)	(0.59)	(14.6)	(211.8)	(0.03)	(0.7)	(1.5)	(6.2)	(0.06)	(0.8)	(1.6)	(4.9)

H_e, 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 parameters; standard deviations are indicated below the mean, into parentheses. Data were simulated for a declining population ($N_0 = 100$; $N_1 = 10,000$; $T_a = 500$), a stable population ($N_0 = N_1 = 464$; $T_a = 500$), and an expanding population ($N_0 = 10,000$; $N_1 = 100$; $T_a = 500$), under three mutation models: a strict stepwise mutation model (SMM), a moderate GSM with a frequency of multi-step changes set to $p = 0.22$, and a strong GSM with a frequency of multi-step changes set to $p = 0.74$.