The consequences of the unlikely but critical assumption of stepwise mutation in the population genetic software, MSVAR

Faurby, Søren; Pertoldi, Cino

The consequences of the unlikely but critical assumption of stepwise mutation in the population genetic software, MSVAR

Søren Faurby[1,2] and Cino Pertoldi[1,3,4]

[1]*Department of Bioscience, Aarhus University, Aarhus, Denmark,*
[2]*Department of Ecology and Evolutionary Biology, University of California,*
Los Angeles, California, USA, [3]*Department 18/Section of Environmental Engineering,*
Aalborg University, Aalborg, Denmark and [4]*Aalborg Zoo, Aalborg, Denmark*

ABSTRACT

Background: MSVAR is a software package for simulating and analysing microsatellite evolution. It assumes that all mutations lead to alleles exactly one repeat longer or shorter than the original (the stepwise mutation model) but many actual mutations lead to size changes of several repeats.

Question: How reliable are the conclusions of microsatellite analyses using the software MSVAR if we assume realistic violations of its stepwise mutation assumption?

Mathematical methods: Simulate microsatellite evolution using forward simulations, under various demographic scenarios. Use different degrees of departure from the pure stepwise mutation model. Then analyse the results using a GLM approach.

Key assumption: Regardless of starting conditions, microsatellite size attains equilibrium or near equilibrium conditions after $5N_e$ generations (where N_e is effective population size).

Results: Absolute values of past and present N_e, the magnitude of N_e change or the time frame of the demographic changes from the output of MSVAR cannot be trusted in the presence of realistic deviations from the assumed (stepwise) mutation model. Ancestral population sizes will be severely overestimated while current populations will be underestimated but to a lesser extent. The extent of bias produced by a given degree of model deviation is only slightly influenced by demography. The relative differences in N_e between populations of the same species will therefore generally be reliable.

Keywords: demographic changes, effective population size, microsatellites, N_e, stepwise mutation model.

Correspondence: S. Faurby, Department of Bioscience, Aarhus University, Ny Munkegade 114, 8200 Aarhus, Denmark. E-mail: soren.faurby@biology.au.dk

INTRODUCTION

Reconstructing the historical and ongoing processes of genetic diversification is beneficial to studies of theoretical and applied conservation biology (Pertoldi et al., 2007). However, the complex dynamics of the late Pleistocene/Holocene make it difficult to disentangle the genetic consequences of natural climatic and habitat changes from those of human-mediated habitat alterations and direct exploitation of natural populations. Recent progress in biostatistics and mathematics has increased our potential to infer population genetic processes via the development of theoretical models. For example, these models can be used to estimate historical and current genetic effective population size (N_e), the degree of genetic isolation and rates of gene flow, past population expansions or declines, and cryptic bottlenecks. Microsatellites are an important source of information for various questions regarding population demography. Microsatellite loci have been increasingly used in population genetics studies, because they generally are neutral and highly polymorphic, which makes them suitable for estimating N_e. Their main advantages are their high mutation rates that increase the amount of information that can be gained by each microsatellite and the relatively low cost and ease of analysing individuals once the markers are developed (Morin et al., 2009; Haasl and Payseur, 2011).

Although many models have been developed to describe the mutation process of microsatellites (Bhargava and Fuentes, 2010), most programs that analyse microsatellite data use the Stepwise Mutation Model [SMM (Ohta and Kimura, 1973)]. This model assumes that all mutations result in either one more or one less repeat than the original, and that insertions and deletions are equally likely. The SMM is computationally simple and provides straightforward expectations for the distribution of allele sizes. However, it has become increasingly clear that the vast majority of microsatellites do not evolve according to this model (e.g. Dieringer and Schlötterer, 2003; Ellegreen, 2004; Seyfert et al., 2008). The frequency of deviations appears high but also very variable, and estimated values for humans and zebrafish have ranged from 0.10 to 0.63 and from 0.05 to 0.75 respectively (Ellegreen, 2000, 2004; Huang et al., 2002). Insertions or deletions of a single repeat are generally the most common, but larger insertions or deletions make up a significant proportion of mutations. Furthermore, mutations are often not symmetric, with either deletions or insertions being the most common in different taxa or microsatellites (Bhargava and Fuentes, 2010).

The focus of this paper is MSVAR, a widely used program that uses microsatellites to detect changes in population size in an individual population, based on a single contemporary dataset (Beaumont, 1999; Storz and Beaumont, 2002). Using several assumptions, MSVAR detects whether N_e has remained constant, expanded or declined over time, and estimates both the original and current N_e and the time when N_e began to change. The two most important assumptions are the mutation model, which is assumed to be the SMM, and the absence of population genetic substructure (i.e. the sampled individuals are assumed to come from a closed panmictic population).

Published studies using MSVAR suggest an apparent bias. Although the program has been used approximately 200 times in published studies, the analysis of human demography from the initial program description (Beaumont, 1999) is one of only a few studies indicating a population expansion; almost all other studies infer population declines (Girod et al., 2011, Table S1). This might be partially explained by the fact that the program has mainly been used for species in which population declines are expected, but the results still appear suspicious.

Several studies have recently shown that, when MSVAR is used, genetically structured populations can create a false signal of population decline (Nielsen and Beaumont, 2009; Chikhi et al., 2010; Peter et al., 2010). However, many studies that have used MSVAR have been conducted on organisms where the sampled individuals do appear to consist of closed panmictic populations, and population declines are still consistently found (e.g. Johnson et al., 2009; Pruett et al., 2010). A reanalysis of other studies by Chikhi et al. (2010) also showed that population structure cannot be the sole explanation for these inferred bottlenecks. This suggests misspecification of the mutational model may be the source of additional bias.

Mutations of a size greater than one single repeat can create a distribution of allele sizes that is similar to the allelic distribution observed in a declining population. The combined effects of drift and stepping stone mutations under a constant population size will generate a more or less bell-shaped distribution of allele sizes. Deviation from such a bell shape will therefore mean either a deviation from a constant population size or from stepping stone mutations. A population decline can create deviations when alleles are randomly lost from the population, while similar patterns can appear if alleles are not lost but rather never occurred, i.e. if some alleles were created by non-stepping stone mutations. Frequent deviations from SMM could therefore lead to false inferences of population declines (Chikhi et al., 2010). However, the magnitude of this error is very difficult to determine without simulations. Unlike errors caused by wrongly assuming a closed panmictic population, which may be specific to programs assuming closed panmictic populations, errors caused by the application of inappropriate mutation models would also be relevant to many other analyses, such as IMa (Hey and Nielsen, 2007; Hey, 2010) or some implementations of ABC approaches (e.g. Peter et al., 2010).

IMa can also be used to analyse sequence data, and simulations have investigated the importance of model violations regarding such use (Strasburg and Rieseberg, 2008). Violations were found to be relatively important when IMa was used to analyse data that were simulated with a more complex and realistic model than the one applied by IMa. This may be comparable to simulation of microsatellites under a more complex model than SMM. The effects of microsatellite mutation model violations were recently analysed as part of a general study of the power of MSVAR (Girod et al., 2011), but the authors only analysed two specific degrees of deviations from the stepwise mutation rate under a single model and thus could not determine what magnitude of errors MSVAR can handle without problems. Furthermore, the authors only investigated whether using the wrong model resulted in strong evidence of a population decline; they did not estimate the magnitude of the error.

In the present study, we analysed how MSVAR behaved in the presence of deviations from the SMM, to help understand the importance of mutational model violations in creating the apparent bias in reported population size changes. We investigated both the frequency and distribution of mutations larger than one. The goal of this study was to estimate the degree of deviation from the SMM that was required to generate results indicating a false population decline, as well as to examine the extent to which useful information (e.g. which populations are largest or which populations have experienced the largest population decline) could be extracted from the analysis, even in cases where model violations led to unreliable estimated values.

MATERIALS AND METHODS

Microsatellite simulations

We simulated data using four different unbounded mutation models using custom R scripts (see evolutionary-ecology.com/data/2763Appendix.pdf): (1) Unbiased Mixed Mutations (UMM), (2) Biased Mixed Mutations (BMM), (3) Unbiased Geometric Mutations (UGM), and (4) Biased Geometric Mutations (BGM). UMM is given by:

$$X_i \text{ with rate } 1 - \mu$$
$$X_i + 1 \text{ with rate } \tfrac{1}{2}\mu \times (1 - \text{NonSMM})$$
$$X_{i+1} = X_i - 1 \text{ with rate } \tfrac{1}{2}\mu \times (1 - \text{NonSMM}) \qquad (1)$$
$$X_i + D \text{ with rate } \tfrac{1}{2}\mu \times \text{NonSMM}$$
$$X_i - D \text{ with rate } \tfrac{1}{2}\mu \times \text{NonSMM}$$

where X_{i+1} is the length of a microsatellite for individual X at generation $i + 1$, μ is the mutation rate, NonSMM is the proportion of mutations not following the SMM model, and D is a discrete uniform distribution with a lower bound of 2 and an upper bound of 10. BMM is similar to UMM, except that it is asymmetrical and assumes that most stepwise mutations are insertions and most multi-step mutations are deletions. It is given by:

$$X_i \text{ with rate } 1 - \mu$$
$$X_i + 1 \text{ with rate } \tfrac{5}{6}\mu \times (1 - \text{NonSMM})$$
$$X_{i+1} = X_i - 1 \text{ with rate } \tfrac{1}{6}\mu \times (1 - \text{NonSMM}) \qquad (2)$$
$$X_i + D \text{ with rate } \tfrac{1}{6}\mu \times \text{NonSMM}$$
$$X_i - D \text{ with rate } \tfrac{5}{6}\mu \times \text{NonSMM}$$

UMM and BMM both assume that mutations that are not stepwise are equally likely to be of any length in a specified range. The other two models assume that the lengths of insertions or deletions follow a geometric distribution. Mutations of increasing size are therefore progressively more rare, but there is no upper limit to the length of a single mutation. BGM and UGM are similar, but BGM assumes that insertions are more likely than deletions. UGM is given by:

$$X_i \text{ with rate } 1 - \mu$$
$$X_{i+1} = X_i + G \text{ with rate } \tfrac{1}{2}\mu \qquad (3)$$
$$X_i - G \text{ with rate } \tfrac{1}{2}\mu$$

and BGM by:

$$X_i \text{ with rate } 1 - \mu$$
$$X_{i+1} = X_i + G \text{ with rate } \tfrac{5}{6}\mu \qquad (4)$$
$$X_i - G \text{ with rate } \tfrac{1}{6}\mu$$

where G is a geometric distribution with a probability of $1 - $ NonSMM. UGM is the model generally used by programs such as SIMCOAL to model non-stepwise mutations (Laval and Excoffier, 2004). This model assumes that the majority of non-stepwise mutations are small, which most likely will mean that the effect of each non-stepwise mutation is relatively small. UMM has to our knowledge not previously been used but was chosen to highlight that the effect of a given NonSMM is heavily dependent on the mutation model. UGM and BGM assume that the vast majority of all multi-step mutations consist of only two repeats. UMM and BMM on the other hand assume that there is no correlation between mutation frequency and mutation size for multi-step mutations. Empirical studies (e.g. Seyfert *et al.*, 2008) generally find that while such a correlation does exist, the proportion of mutations larger than two repeats is usually larger than expected, under the UGM and BGM. UMM/BMM and UGM/BGM can therefore be seen as the worst and best case scenario respectively for the effects of a given frequency of non-stepwise mutations. The contraction bias chosen in BMM and BGM with a five-fold difference in deletion and insertion rates is within the observed frequencies [and identical to the estimates from 229 mutations in dimorphic microsatellites in *Caenorhabditis elegans* in Seyfert *et al.* (2008)], although higher than many reported values (e.g. Ellegreen, 2000; Xu *et al.*, 2000).

For all four models, we analysed 19 values of NonSMM, ranging from 0.00 to 0.90 in steps of 0.05. UMM is identical to UGM when NonSMM is equal to 0, but both sets were analysed, since the codes used to simulate the UMM were slightly different from the codes used to simulate UGM and we wished to ensure that this fact did not influence our results.

For all 76 combinations (4 models each with 19 values of NonSMM) we investigated 37 different evolutionary scenarios, resulting in a total of 2886 different combinations. For all combinations, we performed a single set of simulations. Scenarios consisted of different ancestral population sizes and either constant population sizes or mild to moderate population increases or declines (evolutionary-ecology.com/data/2763Appendix.pdf, Table S1); all population size changes were simulated as exponential. In all cases, we simulated 20 unlinked microsatellites. After each simulation, 50 individuals were selected with replacements from the simulated population to be analysed with MSVAR. These sample sizes were larger than the usual sample sizes of empirical studies using MSVAR, which have a median of 30 individuals and 11.5 microsatellites (Girod *et al.*, 2011), but the amount of data in genetic studies increases over time and future studies will likely use more microsatellites than this median. Furthermore, we wished to simulate larger datasets to increase the information available and thereby decrease the importance of prior distributions.

We simulated a haploid population of twice the mentioned diploid size; all references to population size in this paper will refer to the diploid population size. Simulations were conducted in R v.2.12.1 (R Development Core Team, 2010) using forward simulations with custom codes (evolutionary-ecology.com/data/2763Appendix.pdf). Simulations were started with a population that was monomorphic for a single allele in all microsatellites, and were run for a length of five times the number of N_e generations before any changes in population size were simulated (10 times the simulated haploid generation size). The vast majority of coalescence events should have occurred during this period, but coalescence is a stochastic process and we acknowledge that our approach may be slightly biased if not all coalescence occurred during the simulated generations; in this case, our approach could generate a signal of slight population increase.

In all cases, we assumed a constant mutation rate of 10^{-3}, without variation in the mutation rate between microsatellites, although such variation would likely be found in natural populations. MSVAR v.1.3 (Storz and Beaumont, 2002) incorporates variation in the microsatellite mutation rate among loci into its analysis, thus we did not believe that the inclusion of such variation would influence our results substantially. We chose not to include this variation because we wanted to reduce the risk of chains getting stuck in local optima, which would slow down convergence time. However, it should be noted that while MSVAR can handle variation in mutation rate between loci, it (like most other programs) cannot handle the interaction between mutation rate and N_e, which recently has been suggested to occur as a consequence of higher mutation rate in heterozygote sites (Amos et al., 2008). If such interaction has a detectable effect, it will provide an additional source of error for MSVAR, which we have not taken into account in this analysis. We chose a relatively high mutation rate since higher mutation rates generally increase the information available for each marker and thereby reduce the effects of prior distributions leading to a more precise estimate of the effects of model violations.

MSVAR analysis

We used MSVAR v.1.3 in all analyses; for a description of the algorithm, see Storz and Beaumont (2002). MSVAR v.1.3 is substantially different from the software originally described by Beaumont (1999) and, while the overall conclusions of the results obtained in this investigation will likely be transferable, caution is advised when applying results from these simulations to interpretations of results from the original MSVAR program. We used relatively flat priors for ancestral population size (N_a), current population size (N_c), and time since decline/expansion (T_a) ($\mu = 3$, $\sigma = 3$, $\beta = 0$, $\tau = 1$), but highly informative priors for mutation rate ($\mu = -3$, $\sigma = 0.1$, $\beta = 0$, $\tau = 0.1$) (for definitions of these parameters, see Storz and Beaumont, 2002). The starting conditions for all parameters were the means of their prior distribution. Our results assume that the mutation rate is known virtually without error, which is unlikely to be true. However, as pointed out by Girod et al. (2011), very high precision in the priors of mutation rate is required to generate precise posterior distributions of the natural parameters, which we will focus on in the analysis. More plausible wider priors for mutation rate would generate an extra source of variation, but this is not expected to change the underlying dynamics analysed in this paper.

Since convergence can be very slow for MCMC algorithms, previous simulation studies of MSVAR (Chikhi et al., 2010; Girod et al., 2011) have used lower numbers of simulations and longer chains to ensure data precision. Due to the much higher number of simulations in our study, such long chains were not possible and we only ran the chains for 3.75×10^8 generations with samples every 250,000 generations and the first third discarded as burn-in. We acknowledge that this means that not all chains reached their equilibrium and, therefore, we incorporated parameters related to convergence of the MCMC chains into our analysis (see next section).

The MSVAR analyses and data simulations in R were run on a total of sixty-four 3.3 GHz Power7 CPU cores, each with 128 GB memory under a AIX 6.1 operating system at CSCAA (Centre for Scientific Computing in Aarhus, Aarhus University).

Test for convergence

We were aware that the chains were too short to expect that all had converged. In fact, the results obtained by Girod *et al.* (2011) suggest that convergence to a stable and well-mixed condition may never be reached when there are severe model violations. All else being equal, the deviations from the true values are likely to decrease over time and, therefore, analyses of chains that have not fully reached their equilibrium will still provide information about the underlying parameters, as long as the priors are sufficiently wide to be effectively uninformative and the chain has run sufficiently long to be effectively independent of the starting conditions.

Since we did not expect full convergence, we needed estimates of the convergence level. We estimated the absolute value of the Geweke test statistics (Geweke, 1992) and the inverse of effective sample size (hereafter referred to as Geweke and ESS^{-1}) using coda (Plummer *et al.*, 2006). The Geweke test statistic is a comparison between the mean of the first and the last part of the chain with well-defined characteristics for large well-mixed chains, but like any metric based on means it is sensitive to extreme values. ESS is a measure related to autocorrelation between samples in the chain and is low for poorly mixed or non-converged chains. We chose the inverse of estimated sample size because changes in ESS are only likely to be important for small ESS, while it will be unimportant whether an ESS is 500 or 1000. The Geweke test failed in a few simulations, and in these cases we set the Geweke test statistic as the largest number produced in any of the analyses. $ESSr^{-1}$ was defined as:

$$ESSr^{-1} = \left(\frac{ESSN_a + ESSN_c}{2} \right)^{-1}$$

and Geweke$r$ as the mean of GewekeN_a and GewekeN_c.

Choice of parameters

We focused on estimates of N_c, N_a, T_a, and the exponential growth rate (r), which was the ratio (N_c/N_a). The analysis by Girod *et al.* (2011) suggested that MSVAR provides more precise estimates of the scaled parameters (e.g. the product of population size and mutation rate as well as $T_a/(2N_c)$). Their analyses were, however, based on simulations with a far less informative prior of mutation rate than the one we used, meaning that the precision of their estimates of the non-scaled parameters were far lower than the ones in our analysis.

Using the package TeachingDemos (Snow, 2010), we calculated the mode (with all values rounded to the first decimal) for each of these four parameters, hereafter referred to as ModeN_c, ModeN_a, ModeT_a, and Mode$r$, as well as the lengths of the 70% highest probability density, hereafter referred to as HPD-lengthN_c, HPD-lengthN_a, HPD-lengthT_a, and HPD-length$r$. We also estimated whether the true values of every estimate were included in the 70% HPD (scored as 1 if the HPD contained the true value and 0 otherwise), which will be referred to as PrecisionN_c, PrecisionN_a, PrecisionT_a, and Precision$r$.

T_a is meaningless as a parameter unless N_c and N_a are different from each other (meaning that $r \neq 0$). Therefore, all cases in which the 70% HPD of r spanned 0 were removed from further analyses of T_a.

Comparisons of the effects of various values of NonSMM

All analyses were performed within a generalized linear modelling (GLM) framework using the command glm() in R v.2.11.1, with effect sizes, standard errors, and P-values estimated using Wald's tests (R Development Core Team, 2010). The first set of analyses were an attempt to define the proportion of non-stepwise mutations required to generate estimates that were systematically different from the parameters without model violations, as well as to define the magnitude of the deviation observed with each NonSMM. Each NonSMM in UMM and BMM was compared to UMM with a NonSMM of 0, and each NonSMM in UGM and BGM was compared to UGM with a NonSMM of 0.

All analyses included five parameters: NonSMM (treated as an indicator variable, i.e. coded as 0 for NonSMM of zero and 1 otherwise), ESS^{-1}, and Geweke as well as the interactions between ESS^{-1} or Geweke and the true values of the studied parameter (N_c, N_a, T_a, or r). The last four parameters were included to incorporate the potential effect of non-convergence, but we only focused on the effects of NonSMM. Mode and HPD-length were analysed with family Gaussian (equivalent to a standard ANCOVA). Precision was analysed using family Quasi-binomial and link logit (similar to a standard logistic regression, but incorporating that the variance can be smaller or higher than expected under a binomial distribution). It should be noted that Wald's test is not meaningful for the Quasi-binomial link functions, in cases where the parameter in question (in this case, Precision) is 0 or 1 for all cases in one group, since there is no single ML optimum in these cases. We consider these tests as statistically significant since the Wald's tests are significant in cases with smaller differences between groups, although standard errors are not reported in these cases.

For modes, we reported the raw effect. For HPD-length, we reported results standardized by dividing the difference by the mean HPD-length for UMM or UGM with a NonSMM of 0. For precision, we reported the estimated differences as percentages calculated with ESS^{-1} and Geweke equal to the average Geweke and ESS^{-1} values in the appropriate analysis.

Test of relative performance of MSVAR

A second set of analyses focused on whether the relative values are trustworthy, regardless of whether the results were biased or not – in other words, whether the differences between estimated parameters are the same as the differences between the true values regardless of whether they are both wrong. These analyses were performed for all NonSMM values in all models. They included estimated Mode with five parameters (the true values, ESS^{-1}, Geweke, as well as the interactions between ESS^{-1} or Geweke and the true values) using family Gaussian, and focused on the slope of the true value in the analysis of Mode. In this analysis, an estimated slope of 1 thus means that the difference between estimated values is on average the same as the difference between the true values, whereas an estimated slope of 0 means that the estimated values do not contain any information of the relative size of the true values. Frequent non-convergence would, if unaddressed, lead to low slopes (near 0), but through simulations we have shown that our incorporation of the additional parameters removes this bias, although non-convergence does produce noise in the results, and therefore the individual slope values should be interpreted with caution (evolutionary-ecology.com/data/2763Appendix.pdf).

Multifactorial analysis

In a third set of analyses, we attempted to determine the full effect of model violations, including information on whether the effects of NonSMM or biased mutation interacted with the effects of population size or population history. These analyses combined all runs of UMM and BMM and all runs of UGM and BGM. Parameters were progressively removed from the model until only significant parameters ($P < 0.05$) remained, although main effects were kept if interactions or higher-order effects were significant. The initial model for r and T_a included 18 parameters: four related to chain convergence (ESS^{-1}, Geweke, and the interactions between them and the parameter in question), two directly related to model violations (NonSMM and $NonSMM^2$), four main effects (Bias [an indicator variable equal to 0 for UMM and UGM and 1 for BMM and BGM], True r, True T_a, and the interaction between True r and True T_a), and eight interactions between NonSMM or $NonSMM^2$ and the four main effects. The initial model of N_a and N_c included 21 parameters, the 18 mentioned already as well as True N_a or True N_c and the interactions with NonSMM and $NonSMM^2$. We used the same families and links as in the initial analyses. Here, we report all values but only discuss the effects of parameters directly related to the effects of NonSMM or Bias.

RESULTS

Comparisons of the effects of various NonSMM values

Modes

ModeN_a increased with increasing NonSMM (Fig. 1). For UMM/BMM, the increase was concave, with ModeN_a essentially constant for NonSMM values higher than 0.5. This was significant for all comparisons using BMM and all NonSMMs higher than 0.05 using UMM. For UGM/BGM, the increase was almost linear with increasing NonSMM and significant for NonSMMs higher than 0.2.

ModeN_c decreased slightly with increasing NonSMM (Fig. 1), but the effect was generally non-significant, suggesting minor systematic effects of NonSMM on ModeN_c.

Since the effect on ModeN_c was minor, the effect on Mode$r$ was mimicking the effect on ModeN_a. Little systematic variation was found in T_a, although a small increase in T_a was observed for high NonSMM.

Precision

The effects of increasing NonSMM on PrecisionN_a, PrecisionN_c, Precision$r$, and PrecisionT_a were very similar, although the pattern was less evident for PrecisionT_a since values were low, even with a NonSMM of 0 (Fig. 2).

Precision decreased very fast for UMM/BMM and was highly significant for all non-zero NonSMM values. Increasing NonSMM above 0.1 had little effect, as nearly all precision was already lost. For UGM and BGM, the loss of precision was slower and all precision was not lost until NonSMM was around 0.5.

HPD-length

An increase in NonSMM led to a decrease in HPD-length (Fig. 3). For HPD-lengthN_a, the decrease was relatively slow and almost linear for all models. With UMM/BMM, decreases
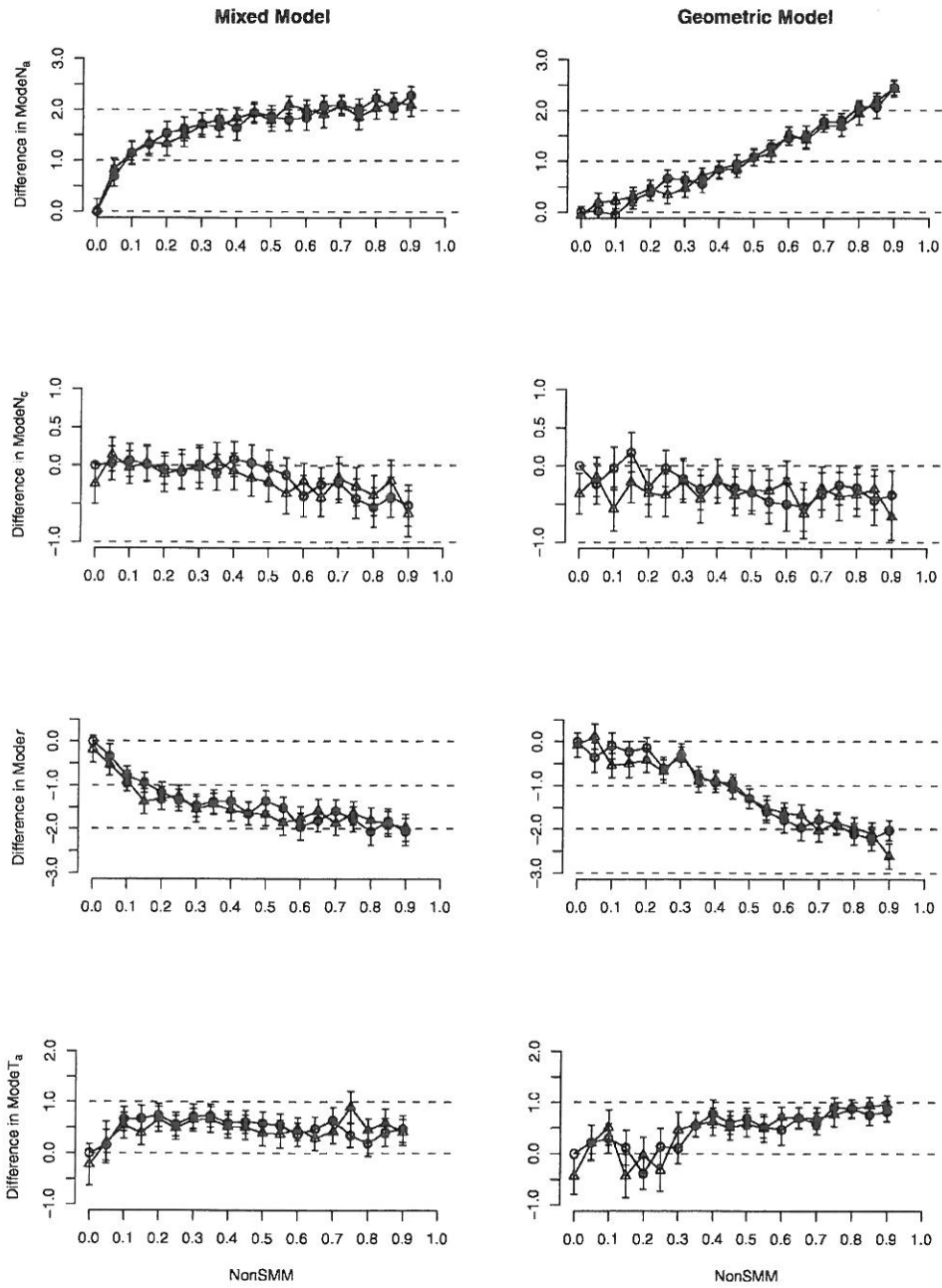
Fig. 1. Effect of NonSMM on modes. Pairwise comparisons of modes between different degrees of NonSMM in both biased and unbiased models, with unbiased models having a NonSMM of 0. Solid symbols denote differences significantly different from 0. Circles are unbiased and triangles are biased. The estimates as well as standard errors are given. Higher values along the y-axis mean that the parameter in question reached higher values at lower stepwise and/or biased models. This and all other figures were constructed using functions from the package Hmisc (Harrell, 2010).
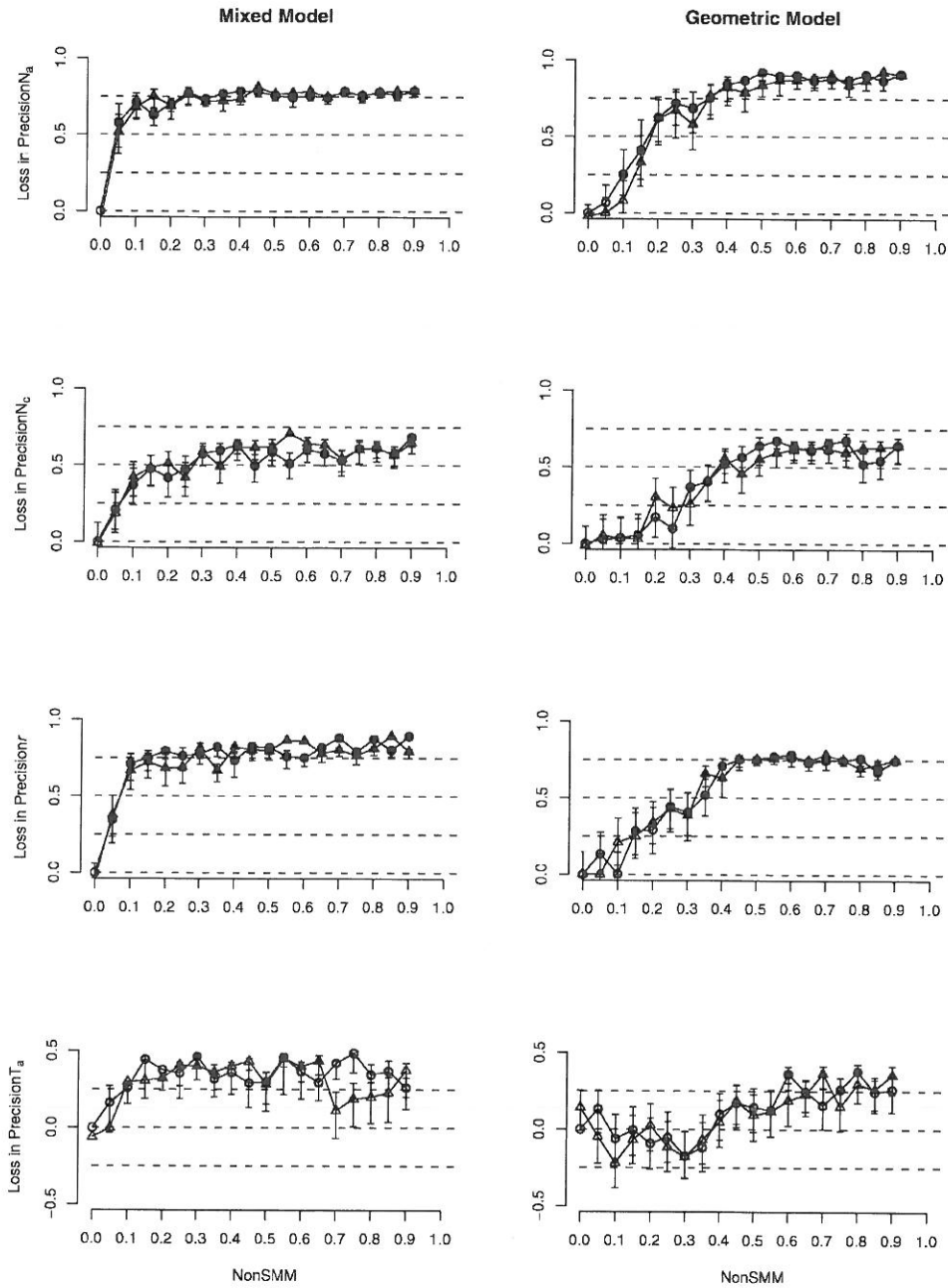
Fig. 2. Effect of NonSMM on precision. Pairwise comparisons of loss in precision between different degrees of NonSMM in both biased and unbiased models, with unbiased models having a NonSMM of 0. Solid symbols denote differences significantly different from 0. Circles are unbiased and triangles are biased. The estimates as well as standard errors are given. In analyses where the precision in one group was zero, the symbol is filled but no standard error is given. Higher values along the y-axis represent a greater loss of precision.
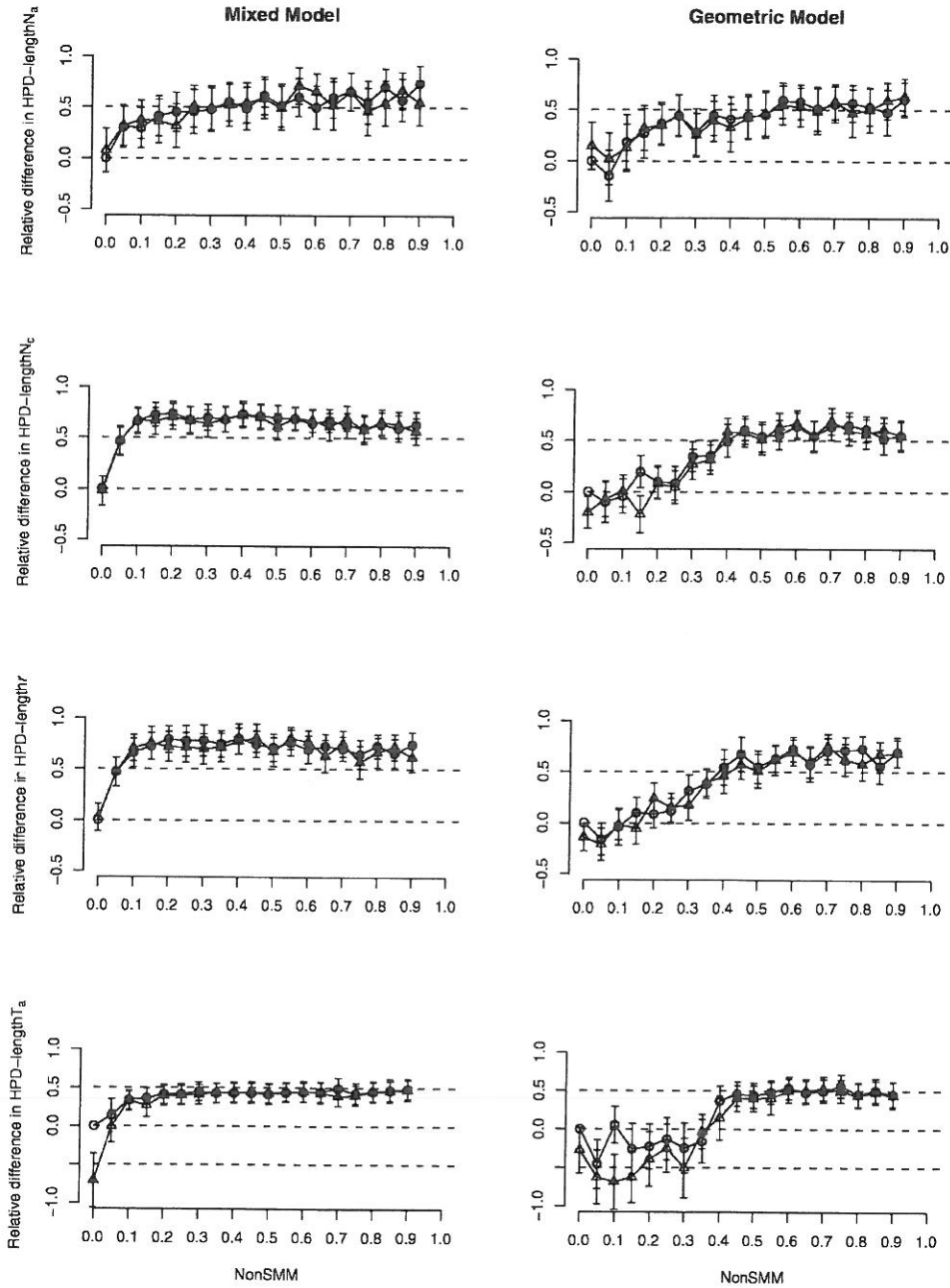
Fig. 3. Effects of NonSMM on HPD-length. Pairwise comparisons of relative difference in HPD-length between different degrees of NonSMM in both biased and unbiased models, with unbiased models having a NonSMM of 0 standardized to the HPD-length of the unbiased model with a NonSMM of 0. Solid symbols denote differences significantly different from 0. Circles are unbiased and triangles are biased. The estimates as well as standard errors are given. Higher values along the y-axis mean a smaller HPD-length.

Fig. 4. Slope comparing the actual N_a, N_c, r or T_a with the modes of the posterior distributions from models including the true value, Geweke, and ESS^{-1}. The slopes as well as standar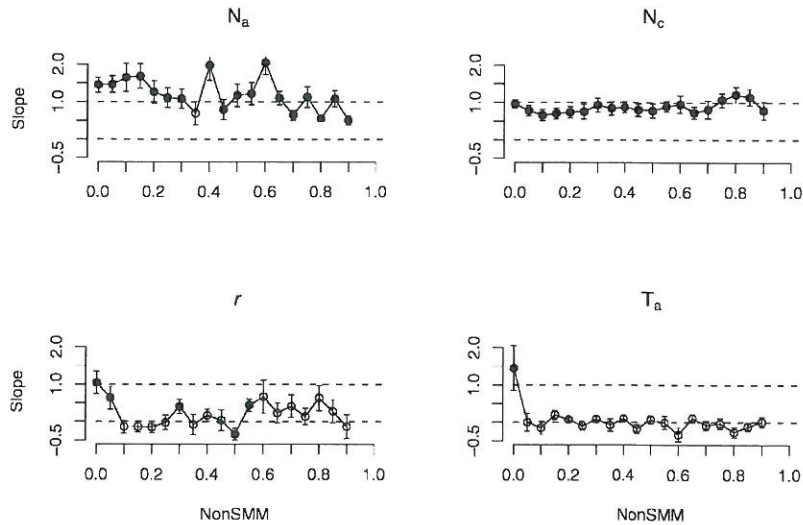d errors are given. Slopes significantly different from 0 represented by solid circles while slopes not significantly different from zero represented by open circles.

in N_c, r, and T_a were almost instantaneous and occurred when NonSMM changed from 0 to 0.1; no additional changes occurred with higher NonSMMs. For UGM/BGM, the change was slower and almost linear with increasing NonSMM, up to 0.45. In all cases, HPD-lengths for higher NonSMM values were around half of the HPD-length with a NonSMM of 0.

Test of relative success of MSVAR

Modes in UMM

The results for UMM (Fig. 4) were almost identical to the results for BMM (evolutionary-ecology.com/data/2763Appendix.pdf, Table S3). For all four parameters in both models, the slope at NonSMM of 0 was close to the theoretically expected value of 1, suggesting that either chain convergence was not a problem or that our incorporation of ESS^{-1} and Geweke enabled us reliably to remove the problem. The slopes for both N_a and N_c were high and significant for all values of NonSMM. For N_c, the slope was independent of NonSMM. For N_a, the slope tended to decrease with increasing NonSMM, down to around 0.5 for very high NonSMM values. The slope for r, although generally positive and significant in some cases, was very low for NonSMM values above 0.05. The slope for T_a dropped immediately, suggesting that no or very limited information was available on the relative T_a of different populations, even at a NonSMM of 0.05.

Modes in UGM

The results for UGM (Fig. 5) were very similar to the results for BGM (evolutionary-ecology.com/data/2763Appendix.pdf, Fig. S4). For all four parameters in both models, the
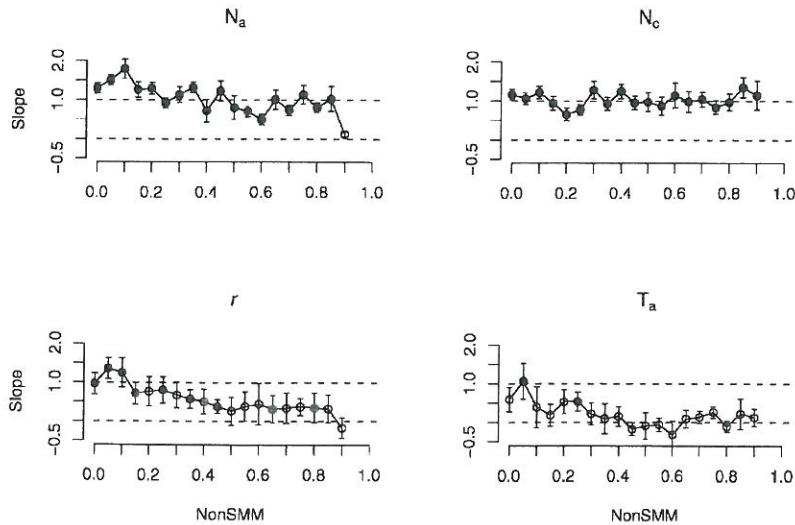
Fig. 5. Analyses of N_a, N_c, r and T_a in UGM. The slope comparing the actual N_a, N_c, r or T_a with the modes of the posterior distributions from models including the true value, Geweke, and ESS^{-1}. The slopes as well as standard errors are given. Slopes significantly different from 0 represented by solid circles while slopes not significantly different from zero represented by open circles.

slope at NonSMM of 0 was close to the theoretical expected value of 1, suggesting that either chain convergence was not a problem or that our incorporation of ESS^{-1} and Geweke enabled us to reliably remove the problem. For N_a, there was a very limited decrease in the slope for NonSMM values between 0 and 0.85, and a slope close to zero for a NonSMM of 0.90. The slope of N_c was essentially independent of NonSMM. For r, the slope decreased with increasing NonSMM and although the slope remained consistently positive until a NonSMM reached a value of 0.9, it was non-significant for all NonSMM values higher than 0.45. The slope dropped drastically for T_a; it remained positive for NonSMM up to 0.45 for UGM and 0.35 for BGM, and was significantly positive for NonSMM of 0.05 and 0.25 for UGM and 0.10 and 0.20 for BGM.

Multifactorial analyses

Because the results for Modes (Table 1) as well as Precision and HPD-length (evolutionary-ecology.com/data/2763Appendix.pdf, Tables S2 and S3) were highly complex, only the results relating to modes will be discussed further. ModeN_a was unaffected by contraction bias but highly affected by NonSMM. The effect of NonSMM was, however, highly complex and non-linear and included several interactions with the true N_a as well as demographic history of the population. One of the most important effects on these interactions is that ModeN_a was more highly influenced by NonSMM when true N_a was small.

ModeN_c was not as influenced by NonSMM as ModeN_a was, but the effects were equally complex. The most important effect was that ModeN_c is underestimated in small but not in large populations as a consequence of NonSMM. Furthermore, there was

a very small but significant effect for UGM/BGM, suggesting that contraction bias led to minor underestimations of N_c.

Mode$r$ was strongly (although non-linearly) influenced by NonSMM. As shown in Figs. 4 and 5, the multidimensional results show that the importance of True r for Mode$r$ decreases with increasing NonSMM. For ModeT_a, the results were less complex. Interestingly for UBM/BMM, the model suggests that ModeT_a was independent of the true T_a. For UGM/BGM, the results were comparable to the results for r with the importance of True T_a for ModeT_a decreasing with increasing NonSMM.

DISCUSSION

Importance of results

The frequently reported bias in mutational direction (e.g. Bhargava and Fuentes, 2010) was found to have a very small effect, since UGM was almost identical to BGM, and UMM almost identical to BMM. This suggests that this bias is of limited importance for analyses of microsatellite data. There was, however, a large difference between UGM/BGM and UMM/BMM, with the effects of a given NonSMM consistently smaller for UGM/BGM than for UMM/BMM. This was likely because the average indel size was smaller for UGM/BGM than for UMM/BMM.

To determine whether the results suggest that model violations are a major problem, it is vital to understand how large of a deviation from a pure SMM can generally be expected. Two mutation matrices derived from *Daphnia* or roundworms showed values of around 0.70 (Seyfert et al., 2008) and both showed large systematic errors in all analysed parameters. Other studies have reported lower but highly variable NonSMM values. For humans, a NonSMM of 0.10–0.15 is often found (e.g. Ellegreen, 2000; Xu et al., 2000), although values as high as 0.63 have also been reported (Huang et al., 2002), while estimates for zebrafish range from 0.05 to 0.75 (for a full discussion, see Ellegreen, 2004). The effect of the size of a given NonSMM may therefore lie somewhere between the reported UGM/BGM and UMM/BMM models.

Mode$r$, and probably ModeN_a, will be systematically biased in almost all cases, whereas systematic bias in ModeN_c may or may not occur in individual studies. Precision will likely be affected in analyses of empirical data and based on our results it is unlikely that reported values of 70% HPD contain the true values. Information on the absolute values could theoretically be retrieved if precise knowledge of the NonSMM values could be obtained. However, since this parameter appears to be highly variable between and within species, it may be necessary to restrict comparisons to those between populations analysed with the same microsatellites.

Fortunately, even though absolute information cannot safely be recovered from MSVAR data, relative data can be recovered for N_c and N_a, whereas even relative information about T_a and probably also r may be very difficult to retrieve. Care should be taken, because the reduction in HPD-length with increasing NonSMM means that non-overlapping HPDs may not always be taken as evidence for different parameter values. Furthermore, the highly complex interactions between modes and effective population size and demography may make even relative comparisons dangerous in some cases, and it may therefore be safer to interpret the output of MSVAR as ordinal rather than interval data.

Table 1. Determinants of modes N_a, N_e, r or T_a

Mode N_a	Mixed	Geometric
Intercept	−1.06 (0.16) ***	−0.82 (0.14) ***
ESS⁻¹	49.26 (6.63) ***	12.67 (3.72) ***
ESS⁻¹:True N_a	−14.17 (2.31) ***	−4.05 (1.39) **
Geweke	N.S.	N.S.
Geweke:True N_a	N.S.	N.S.
NonSMM	6.55 (0.37) ***	2.07 (0.68) **
NonSMM²	−3.36 (0.29) ***	2.95 (0.73) ***
True N_a	1.50 (0.05) ***	1.24 (0.05) ***
True N_a:NonSMM	−0.79 (0.09) ***	0.05 (0.25) N.S.
True N_a:NonSMM²	N.S.	−0.91 (0.27) ***
Bias	N.S.	N.S.
Bias:NonSMM	N.S.	N.S.
True r	0.03 (0.06) N.S.	−0.15 (0.04) ***
True r:NonSMM	0.55 (0.21) **	0.47 (0.17) **
True r:NonSMM²	−0.76 (0.22) ***	−0.91 (0.27) ***
True T_a	−0.08 (0.04)*	−0.03 (0.01) N.S.
True T_a:NonSMM	0.11 (0.07) N.S.	N.S.
True r:True T_a	0.11 (0.07) N.S.	0.13 (0.02) ***
True r:True T_a:NonSMM	0.13 (0.02) ***	N.S.

Mode N_c	Mixed	Geometric
Intercept	−0.49 (0.11) ***	−0.58 (0.16) ***
ESS⁻¹	−3.24 (0.28) ***	−3.14 (0.42) ***
ESS⁻¹:True N_e	1.01 (0.16) ***	1.00 (0.18) ***
Geweke	0.013 (0.004)***	N.S.
Geweke:True N_e	−0.004 (0.002)*	N.S.
NonSMM	−2.01 (0.27) ***	−3.13 (0.39) ***
NonSMM²	−0.08 (0.21) N.S.	1.32 (0.28) ***
True N_c	1.11 (0.03) ***	1.18 (0.05) ***
True N_c:NonSMM	0.54 (0.07) ***	0.55 (0.10) ***
True N_c:NonSMM²	N.S.	N.S.
Bias	N.S.	−0.26 (0.07) ***
Bias:NonSMM	N.S.	0.37 (0.14) **
True r	−0.71 (0.07) ***	−0.79 (0.06) ***
True r:NonSMM	−1.41 (0.19) ***	−0.97 (0.09) ***
True r:NonSMM²	0.40 (0.16) *	N.S.
True T_a	−0.04 (0.02) N.S.	0.10 (0.04) *
True T_a:NonSMM	0.04 (0.05) N.S.	−0.22 (0.07) **
True r:True T_a	0.16 (0.03) ***	0.32 (0.03) ***
True r:True T_a:NonSMM	0.26 (0.06) ***	0.34 (0.02) ***

Moder	Mixed	Geometric	Mode T_a	Mixed	Geometric
Intercept	−0.54 (0.09) ***	−0.18 (0.13)[N.S.]	Intercept	4.32 (0.08) ***	2.34 (0.17) ***
Geweke	0.04 (0.005) ***	N.S.	Geweke	0.03 (0.004) ***	0.02 (0.005) **
ESS^{-1}	−5.01 (0.42) ***	−4.51 (1.23) ***	ESS^{-1}	−4.52 (0.23) ***	N.S.
NonSMM	**−5.80 (0.35) *****	**−5.80 (0.35) *****	NonSMM	0.20 (0.26)[N.S.]	**4.17 (0.73) *****
NonSMM^2	**3.98 (0.37) *****	**3.98 (0.37) *****	NonSMM^2	**−1.03 (0.24) *****	**−2.79 (0.72) *****
True r	0.37 (0.07) ***	0.37 (0.07) ***	True T_a	−0.01 (0.04)[N.S.]	0.50 (0.08) ***
True r:NonSMM	**−2.01 (0.26) *****	**−2.01 (0.26) *****	True T_a:NonSMM	0.06 (0.07)[N.S.]	**−1.65 (0.38) *****
True r:NonSMM^2	**1.69 (0.28) *****	**1.69 (0.28) *****	True T_a:NonSMM^2	N.S.	**1.30 (0.38) *****
Bias	−0.20 (0.08) *	**−0.20 (0.08) ***	Bias	**−0.06 (0.03) ***	N.S.
Bias:NonSMM	**0.33 (0.16) ***	**0.33 (0.16) ***	Bias:NonSMM	N.S.	N.S.
True T_a	0.04 (0.03)[N.S.]	0.04 (0.03)[N.S.]	True r	0.54 (0.07) ***	0.08 (0.02) ***
True T_a:NonSMM	N.S.	N.S.	True r:NonSMM	**−0.92 (0.13) *****	N.S.
True r:True T_a	0.18 (0.03) ***	N.S.	True r:True T_a	0.06 (0.07)[N.S.]	N.S.
True r:True T_a:NonSMM	N.S.	N.S.	True r:True T_a:NonSMM	**0.37 (0.07) *****	N.S.

* $0.05 < P < 0.01$, ** $0.01 < P < 0.001$, *** $P < 0.001$; [N.S.] $P > 0.05$.

Note: The point estimates as well their standard errors are shown. The model assumes that the y variable (e.g. ModeN_a) is equal to a linear combination of all variables, i.e. $y = \text{Intercept} + a_1 x_1 + a_2 x_2, \ldots$, where the a's are the parameter estimate sizes given in the table and the x's are the values for each parameter in each run. The estimates as well as their standard errors are given. Parameters not included in the final model in any of the analyses on each row are removed from the table for clarity. Significant parameters related to the effect of deviations from a SMM are shown in bold.

The conclusions of this study seem to contradict the conclusions reached in a previous study, in which MSVAR was investigated and found to be robust to model-violations (Girod *et al.*, 2011). These differences stem partially from different data interpretations and partially from the choice of parameters. Of ten populations that were stimulated according to a UGM with a NonSMM of 0.22, Girod *et al.* (2011) found false population declines in two of five populations simulated under constant population size, while true population increases (simulated with an *r* of 2) were found in two of three populations (five were simulated, two did not converge). A false positive rate of 40% in detecting population declines may arguably be highly problematic. Furthermore, a NonSMM of 0.22 is lower than expected in many cases. Finally, their results are based on the UGM being appropriate, while we suspect that larger indels are substantially more common in true populations than expected under this model, since the relative frequency of indels of 3–5 repeats is generally higher than expected under the UGM (e.g. Ellegreen, 2000; Xu *et al.*, 2000; Seyfert *et al.*, 2008). Our simulations show that MSVAR is more vulnerable to a given NonSMM when the average size of a NonSMM mutation is larger and the use of UGM may therefore underestimate the effects of model violations.

Potential sources of error

A full estimation of the importance of priors for the results would require repeating the entire analysis with several separate sets of priors, but this is not realistically possible, due to the already very high computer demands. However, the importance of such effects is suggested by our analyses, as these effects could be partly responsible for the complex interactions between NonSMM and population sizes or demographic parameters. For instance, ModeN_a is relatively unaffected by true N_a for very high NonSMM values, likely because the upper values for N_a are partly dictated by the priors (Figs. 4 and 5). This suggests that the estimated values in the tables and figures (the exact change in parameter value for a given NonSMM) are only directly comparable with other studies if they have the same priors as well as the same signal strength (requiring the same sample and microsatellite number). The priors used in this study are wider than those used in many empirical studies. For empirical studies using the same priors for N_c and N_a but assuming a lower standard error in the prior, the bias in estimation of *r* may be slightly less (although the estimates will instead be more highly influenced by the priors).

Incomplete chain convergence is a potential source of error in this study, but we believe that we removed most of the effect by including parameters of chain convergence in the analysis, which is supported by our simulations on the behaviour and our treatment of non-convergence (evolutionary-ecology.com/data/2763Appendix.pdf). If there was a systematic residual effect caused by poorer chain convergence at higher NonSMM, as indicated in the study by Girod *et al.* (2011), our analyses of the skewness posterior distributions (evolutionary-ecology.com/data/2763Appendix.pdf) would be expected to show decreasing symmetry with increasing NonSMM but the opposite result was found. It should be noted that our analyses cannot give precise information on whether one or more of our parameters was consistently influenced by poor chain convergence. This is why we have not discussed the absolute performance of MSVAR but only how the performance of MSVAR was influenced by variation in NonSMM.

Implications for previous and future studies

We believe that these simulations have important implications both for the interpretation of previous studies and for the design of future studies. One important aspect is that estimates of T_a, which have previously been discussed in detail by many authors, including ourselves (Pertoldi *et al.*, 2001; Randi *et al.*, 2003; Faurby *et al.*, 2010), actually have little meaning and readers of papers using MSVAR should not focus too much on these estimates. Significantly different N_a and N_c estimates, which have been consistently interpreted as evidence for bottlenecks or population declines (Pertoldi *et al.*, 2001; Randi *et al.*, 2003; Faurby *et al.*, 2010), should similarly be regarded with extreme caution since such differences could very likely be caused by model violations. Readers of MSVAR analyses should instead focus on relative size differences between populations (e.g. Randi *et al.*, 2003; Faurby *et al.*, 2010; Pruett *et al.*, 2010), although caution should be taken if genetic sub-structure is expected to be more pronounced in some populations than in others as such sub-structure may also influence the results of MSVAR (Nielsen and Beaumont, 2009; Chikhi *et al.*, 2010; Peter *et al.*, 2010).

Girod *et al.* (2011) concluded that MSVAR was a very powerful method for identifying bottlenecks, suggesting that researchers who are interested in finding bottlenecks should use the program. However, we disagree with this notion since our analyses indicate that evidence for bottlenecks is likely found even when no bottlenecks have occurred. We believe that the MSVAR program should only be selected for use by researchers interested in contemporary or historical differences in population sizes between populations, and that researchers interested in bottlenecks *per se* need to use programs that incorporate non-stepwise mutations in their models.

ACKNOWLEDGEMENTS

REFERENCES

Amos, W., Flint, J. and Xu, X. 2008. Heterozygosity increases microsatellite mutation rate, linking it to demographic history. *BMC Genetics*, **9**: 72.

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

Bhargava, A. and Fuentes, F.F. 2010. Mutational dynamics of microsatellites. *Mol. Biotechnol.*, **44**: 250–266.

Chikhi, L., Sousa, V.C., Luisi, P., Goossens, B. and Beaumont, M.A. 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.

Dieringer, D. and Schlötterer, C. 2003. Two distinct modes of microsatellite mutation processes: evidence from the complete genomic sequences of nine species. *Genome Res.*, **13**: 2242–2251.

Ellegreen, H. 2000. Heterogeneous mutation processes in human microsatellite DNA sequences. *Nature Genet.*, **24**: 400–402.

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

Faurby, S., King, T.L., Obst, M., Hallerman, E.M., Pertoldi, C. and Funch, P. 2010. Population dynamics of American horseshoe crabs – historic climatic events and recent anthropogenic pressures. *Mol. Ecol.*, **19**: 3088–3100.

Geweke, J. 1992. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In *Bayesian Statistics*, Vol. 4 (J.M. Bernardo, J.O. Berger, A.P. Dawiv and A.F.M. Smith, eds). Oxford: Clarendon Press.

Girod, C., Vitalis, R., Leblois, R. and Freville, H. 2011. Inferring population decline and expansion from microsatellite data: a simulation-based evaluation of the MSVAR method. *Genetics*, **188**: 165–179.

Haasl, R.J. and Payseur, B.A. 2011. Multi-locus inference of population structure: a comparison between single nucleotide polymorphisms and microsatellites. *Heredity*, **106**: 158–171.

Harrell, F.E. 2010. *Hmisc: Harrell Miscellaneous*. R package version 3.8-3. Available at: http://www.inside-r.org/packages/hmisc (accessed 22 June 2011).

Hey, J. 2010. Isolation with migration models for more than two populations. *Mol. Biol. Evol.*, **27**: 905–920.

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

Huang, Q.Y., Xu, F.H., Shen, H., Deng, H.Y., Liu, Y.J., Liu, Y.Z. *et al.* 2002. Mutation patterns at dinucleotide microsatellite loci in humans. *Am. J. Human Genet.*, **70**: 625–634.

Johnson. J.A., Tingay, R.E., Culver, M., Haller, F., Clarke, M.L. and Mindell, D.P. 2009. Long-term survival despite low genetic diversity in the critically endangered Madagascar fish-eagle. *Mol. Ecol.*, **19**: 54–63.

Laval, G. and Excoffier, L. 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.

Morin, P.A., Manaster, C., Mesnick, S.L. and Holland, R. 2009. Normalization and binning of historical and multi-source microsatellite data: overcoming the problems of allele size shift with ALLELOGRAM. *Mol. Ecol. Resources*, **9**: 1451–1455.

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

Ohta, T. and Kimura, M. 1973. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genet. Res.*, **22**: 201–204.

Pertoldi, C., Hansen, M.M., Loeschcke, V., Madsen, A.B., Jacobsen, L. and Baagoe, H. 2001. Genetic consequences of population decline in European Otter *Lutra lutra*: an assessment of microsatellite DNA variation in Danish otters from 1883 to 1993. *Proc. R. Soc. Lond. B*, **268**: 1775–1781.

Pertoldi, C., Bijlsma, R. and Loeschcke, V. 2007. Conservation genetics in a globally changing environment: present problems, paradoxes and future challenges. *Biodivers. Conserv.*, **16**: 4147–4163.

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

Plummer, M., Best, N., Cowles, K. and Vines, K. 2006. CODA: convergence diagnosis and output analysis for MCMC. *R-News*, **6**: 7–10.

Pruett, C.L., Turner, T.N., Topp, C.M., Zagrebelny, S. and Winker, K. 2010. Divergence in an archipelago and its conservation consequences in Aleutian Island rock ptarmigan. *Conserv. Genet.*, **11**: 241–248.

Randi, E., Davoli, F., Pierpaoli, M., Pertoldi, C., Madsen, A.B. and Loeschcke, V. 2003. Genetic structure in otter (*Lutra lutra*) populations in Europe: implications for conservation. *Anim. Conserv.*, **6**: 93–100.

R Development Core Team. 2010. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Seyfert, A.L., Cristescu, M.E., Frisse, L., Schaack, S., Thomas, W.K. and Lynch, M. 2008. The rate and spectrum of microsatellite mutation in *Caenorhabditis elegans* and *Daphnia pulex*. *Genetics*, **178**: 2113–2121.

Snow, G. 2010. *TeachingDemos: Demonstrations for Teaching and Learning*. R package version 2.6. Available at: http://CRAN.R-project.org/package = TeachingDemos (accessed 22 June 2011).

Storz, J.F. and Beaumont, M.A. 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.

Strasburg, J.L. and Rieseberg, L.H. 2008. How robust are 'Isolation with Migration' analyses to violations in the IM model? A simulation study. *Mol. Biol. Evol.*, **27**: 297–310.

Xu, X., Peng, M. and Fang, Z. 2000. The direction of microsatellite mutations is dependent upon allele length. *Nature Genet.*, **24**: 396–399.