MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 3403–3418 doi: 10.1111/j.1365-294X.2012.05635.x

INVITED REVIEWS AND META-ANALYSES

Reliability of genetic bottleneck tests for detecting recent population declines

M. ZACHARIAH PEERY,* REBECCA KIRBY,* BRENDAN N. REID,* RICKA STOELTING,* ELENA DOUCET-BËER,* STACIE ROBINSON,* CATALINA VÁSQUEZ-CARRILLO,* JONATHAN N. PAULI* and PER J. PALSBØLL†‡

**Department of Forest and Wildlife Ecology, University of Wisconsin-Madison, 1630 Linden Drive, Madison, WI 53706, USA, †Marine Evolution and Conservation, Centre of Evolutionary and Ecological Studies, University of Groningen, PO Box 11103, CC Groningen, The Netherlands, ‡Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA*

Abstract

The identification of population bottlenecks is critical in conservation because populations that have experienced significant reductions in abundance are subject to a variety of genetic and demographic processes that can hasten extinction. Genetic bottleneck tests constitute an appealing and popular approach for determining if a population decline has occurred because they only require sampling at a single point in time, yet reflect demographic history over multiple generations. However, a review of the published literature indicates that, as typically applied, microsatellite-based bottleneck tests often do not detect bottlenecks in vertebrate populations known to have experienced declines. This observation was supported by simulations that revealed that bottleneck tests can have limited statistical power to detect bottlenecks largely as a result of limited sample sizes typically used in published studies. Moreover, commonly assumed values for mutation model parameters do not appear to encompass variation in microsatellite evolution observed in vertebrates and, on average, the proportion of multi-step mutations is underestimated by a factor of approximately two. As a result, bottleneck tests can have a higher probability of 'detecting' bottlenecks in stable populations than expected based on the nominal significance level. We provide recommendations that could add rigor to inferences drawn from future bottleneck tests and highlight new directions for the characterization of demographic history.

Keywords: bottleneck test, extinction, heterozygosity, microsatellite, M-ratio, mutation model, population bottleneck

Received 20 June 2011; revision received 29 March 2012; accepted 31 March 2012

Introduction

Understanding the demographic history of populations and species is a theme that unites ecology and evolutionary biology. From a conservation perspective, a pronounced reduction in population size-often referred to as a 'population bottleneck'—can result in the loss of genetic diversity (Nei *et al.* 1975; Tajima 1989a, 1996; Bouzat 2010), compromise a species' ability to adapt to environmental change (O'Brien & Evermann 1988; Frankham *et al.* 1999; Reusch & Wood 2007) and increase the likelihood of extinction via a host of genetic and demographic processes (Caughley 1994; Keller & Waller 2002). Elevated extinction risk in bottlenecked populations has spurred the development of a rich body of demographic and genetic methods for determining whether a population has undergone a recent reduction in abundance, the most direct approach involving the monitoring of changes in demographic or genetic parameters over time (Williams *et al.* 2002; Schwartz *et al.* 2007; Bonebrake *et al.* 2010). Population and genetic monitoring often require long-term studies and, as a result, the spatial distribution of historically

Correspondence: M. Z. Peery, Fax: 608 890 9922;
E-mail: mpeery@wisc.edu

collected museum specimens (Shaffer *et al.* 1998; Tingley & Beissinger 2009), age ratios of museum specimens (Beissinger & Peery 2007; Green 2008) and archived genetic samples (Wandeler *et al.* 2007; Leonard 2008) are increasingly being used to detect longer term population declines. However, necessary specimens and data are often unavailable, making it infeasible to assess population bottlenecks in this manner for many species (Bonebrake *et al.* 2010).

In cases when long-term monitoring is not possible and historic data or samples are unavailable, genetic methods that require only a single temporal sample constitute an appealing alternative for detecting past population bottlenecks (Tajima 1989b; Cornuet & Luikart 1996; Garza & Williamson 2001). All single-sample population genetic methods are based on detecting deviations from expectations under mutation-drift equilibrium, the most common approaches for species of conservation concern being M-ratio and heterozygosity-excess tests based on multilocus microsatellite genotypes (Box 1). Thus, in practice, bottleneck tests are used to detect declines in abundance, but in principle they reflect the genetic signature of declines in effective population size (N_e). Because genetic bottleneck methods require only a single population sample and may be applied to standard genetic data, they have become a routine part of conservation genetics studies aimed at characterizing recent demographic history. However, rigorous application of bottleneck tests requires that population declines have a high probability of being detected and that bottlenecks are not regularly inferred for stable populations; in other words, tests should have high statistical power and a low probability of resulting in a Type I error (Taylor & Dizon 1997). Failure to detect recent population bottlenecks in endangered species might delay the implementation of needed conservation actions and increase the likelihood of extinction, whereas incorrectly inferring bottlenecks in stable populations could result in the misallocation of conservation resources to the detriment of more imperiled

Box 1. Principles of genetic bottleneck tests

Genetic bottleneck tests are rooted in a wider class of population genetic methods aimed at detecting departures from expectations under mutation-drift equilibrium. Population genetic tests of mutation-drift equilibrium typically contrast two different indices of genetic diversity. One measure is expected to be only marginally affected by the underlying process causing deviations from mutation-drift equilibrium and represents a baseline against which the second, more sensitive, diversity index is compared. Examples of genetic diversity indices employed in mutation-drift equilibrium tests that are expected to be less affected by a population bottleneck initially are nucleotide diversity, heterozygosity and the variance and range in microsatellite allele size. In contrast, a greater reduction is expected in the number of alleles and segregating sites, particularly rare ones. Tajima's *D*, a classic test of mutation-drift equilibrium that was originally developed to detect selection, is often used to infer long-term changes in N_e by contrasting the number of segregating sites with nucleotide diversity in DNA sequences (Tajima 1989b). As low-frequency alleles contribute little to the nucleotide diversity, nucleotide diversity is reduced proportionally less than segregating sites during a population bottleneck, resulting in a positive value of Tajima's *D*.

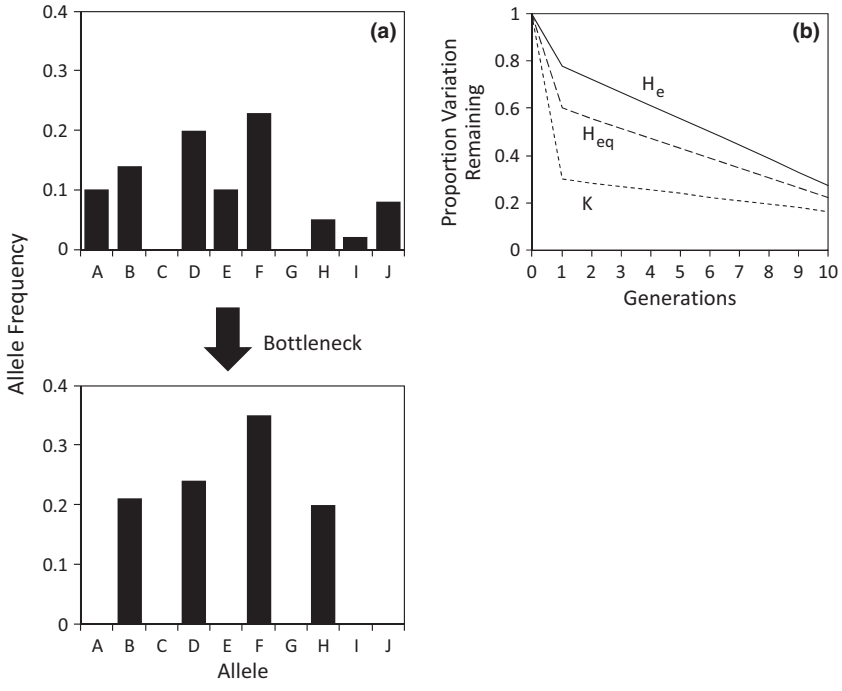
The 'M-ratio test' developed by Garza & Williamson (2001) is based on the ratio of the number of microsatellite alleles (K) to the range in allele size (r; i.e. $M = K/r$). During a bottleneck, the number of alleles is expected to decline faster than the range in allele size as most alleles lost will, by chance, be intermediate in size for loci with at least five alleles (graph a). Accordingly, the M-ratio is expected to be smaller in bottlenecked populations than in equilibrium populations. Tests for reductions in M-ratios are conducted by comparing the mean observed M-ratio (across loci) with an expected distribution generated from simulations under mutation-drift equilibrium, or by using a critical value of 0.68 derived from putatively stable wild populations (Garza & Williamson 2001).

The 'heterozygosity-excess' test developed by Cornuet & Luikart (1996) contrasts heterozygosity expected under Hardy–Weinberg equilibrium with heterozygosity expected under mutation-drift equilibrium calculated from observed number of alleles. The latter measure is more sensitive to the loss of low-frequency alleles during a bottleneck (graph b), and a bottleneck is therefore inferred when heterozygosity under Hardy–Weinberg equilibrium exceeds heterozygosity under mutation-drift (i. e. heterozygosity excess is detected). Coalescent simulations are conducted assuming a constant effective population size to estimate heterozygosity under mutation-drift equilibrium, conditional on the number of alleles in the observed data (only those simulated data sets with the observed number of alleles at each locus are retained). Several statistical methods can be used to test for greater Hardy–Weinberg than mutation-drift heterozygosity, but the Wilcoxon signed-rank test is the most powerful and commonly used approach (Luikart & Cornuet 1998).

Box 1 *Continued*

Box 1. Graph

Illustration of (a) an increase in heterozygosity excess following a bottleneck and (b) a reduction in M-ratios, modified from Garza & Williamson (2001) and Luikart & Cornuet (1998). (a) The number of alleles is reduced proportionally more than the range of alleles size (r) following a bottleneck such that the M-ratio (calculated as K/r) is reduced, and (b) heterozygosity at mutation-drift equilibrium (H_{eq}) conditional on the number alleles (K) in the population initially experiences a greater proportionate reduction than heterozygosity under Hardy–Weinberg equilibrium (H_e) because rare alleles are lost faster than heterozygosity.



populations or species. Achieving adequate power and reasonably low Type I error rates is particularly important given that the results of genetic bottleneck tests are increasingly being incorporated into assessments of threatened species (NMFS 2010; USFWS 2010a, b).

Initial assessments indicated that genetic bottleneck tests have reasonable power to detect population declines, given adequate sample sizes of individuals and loci (Cornuet & Luikart 1996; Luikart & Cornuet 1998; Garza & Williamson 2001; Williamson-Natesan 2005). However, simulation studies have only evaluated the ability to detect very large proportional declines in N_e (10- to 1000-fold) or declines to very few individuals (N_e = 10–50), and empirical assessments often evaluated test performance in populations experiencing extreme declines (e.g. Mexican wolves, *Canis lupus mexicanus*). In contrast, most research on species of conservation concern aims to detect declines that are less severe and that are not obvious from ecological data, but may nevertheless affect the species' persistence (Traill *et al.* 2010; Flather *et al.* 2011). Bottleneck detection is further complicated by the fact that several factors including the timing and duration of the bottleneck, immigration and the amount of pre-bottleneck genetic diversity can influence and potentially obscure genetic signals of population declines (Cornuet & Luikart 1996; Garza & Williamson 2001; Williamson-Natesan 2005). Thus, bottleneck tests may be less likely to detect population declines in species of conservation concern than previously suggested. Indeed, bottleneck tests have failed to detect well-known population collapses in Scandinavian lynx (*Lynx lynx*; Spong & Hellborg 2002), California sea otters (*Enhydra lutris nereis*; Aguilar *et al.* 2008) and Amur tigers (*Panthera tigris altaica*; Henry *et al.* 2009). In

all three cases, global population size had been reduced to tens of individuals, putting these species at a high risk of extinction from demographic and genetic consequences of reduced population size.

Genetic bottleneck tests require making assumptions about microsatellite evolution to generate expected distributions for test statistics (Cornuet & Luikart 1996; Garza & Williamson 2001). Microsatellites are generally modelled as evolving according to a two-phase mutation model that consists of two parameters, the proportion (p_g) and mean size (δ_g) of multi-step mutations (Box 2). However, results can be highly sensitive to assumed parameter values, and incorrect assumptions about these parameter values can lead to an erroneous inference that a bottleneck occurred (Boxes 2 and 3). As parameter values generally are unknown for the species of interest, values are typically 'borrowed' from species for which these parameters have been estimated or estimated indirectly from allele frequency distributions observed in stable populations (Piry *et al.* 1999; Garza & Williamson 2001). Moreover, parameter values used in bottleneck tests are typically derived from studies of microsatellite mutations in humans, and the extent to which these values apply generally to vertebrates is uncertain given the high level of interlocus and interspecific variation in microsatellite evolution (Ellegren 2000, 2004). However, several studies involving a diversity of vertebrate species have recently been published that allow for assessments of (i) how well typically assumed values reflect microsatellite evolution; and (ii) the effects of differences between assumed and true parameter values on inferences from bottleneck tests.

Here, we reviewed the use of M-ratio and heterozygosity-excess tests in the literature and used simulation analyses to assess how effectively these tests detect recent population declines in vertebrates. We limited our assessment to vertebrates because of the large number of studies that have used genetic bottleneck tests, as well as variability among higher order taxa in ploidy

Box 2. Microsatellite mutation models and their effects on bottleneck tests

Genetic bottleneck tests require making assumptions about microsatellite evolution to generate expected distributions for test statistics (Cornuet & Luikart 1996; Garza & Williamson 2001). Mutation models considered in bottleneck tests range from an infinite alleles model where each new mutation results in a unique allele, to a stepwise mutation model where each new mutation results in the loss or addition of one repeat microsatellite motif. However, microsatellites are generally believed to mutate according to an intermediate, two-phase model where most mutations result in the addition or loss of a single repeat motif and a smaller proportion of mutations result in the addition or loss of a larger number of repeats (Di Rienzo *et al.* 1994). Under this model, the proportion of mutations that involve multi-step repeats can be denoted by the parameter p_g and the mean size of multi-repeat mutations is described by the parameter δ_g. Under the assumption that the size of multi-step mutation follows a two-sided geometric distribution, the probability that a multi-step mutation involves x number of repeats can be expressed as

$$P(x) = \frac{1}{2(\delta_g - 1)}\left(1 - \frac{1}{\delta_g - 1}\right)^{|x|-2}, |x| \geq 2$$

following Williamson-Natesan (2005). The computer program BOTTLENECK (Piry *et al.* 1999), the *de facto* program for conducting heterozygosity-excess tests, requires the specification of the variance of δ_g, which can be expressed as $\sigma_g^2 = 2\delta_g^2 - 3\delta_g + 2$ under the geometric distribution (Williamson-Natesan 2005).

Recommended values for the proportion of multi-step mutations and the mean size of multi-step mutations in M-ratio tests (p_g = 0.10 and δ_g = 3.5 respectively) were inferred by selecting values that resulted in M-ratios in simulated equilibrium populations that matched M-ratios from stable wild populations (Garza & Williamson 2001). The basis for the recommendation of p_g = 0.10 and δ_g = 3.1 (equivalent to $\sigma_g^2 = 12$ under a geometric distribution) for heterozygosity-excess tests is not immediately clear (Piry *et al.* 1999). Erroneous assumptions about these mutation model parameters can lead to incorrect inferences from genetic bottleneck tests (Luikart & Cornuet 1998; Garza & Williamson 2001; Williamson-Natesan 2005). For M-ratios, if the underlying mutation model results in more and larger multi-repeat mutations than assumed, a greater number of 'missing alleles will occur between the smallest and largest alleles than expected, and M-ratios estimated for a stable population could match expectations for a bottlenecked population, resulting in a Type I error (Garza & Williamson 2001). In contrast, Type I errors are more likely to occur in heterozygosity-excess tests if the proportion of multi-step mutations is overestimated (Williamson-Natesan 2005).

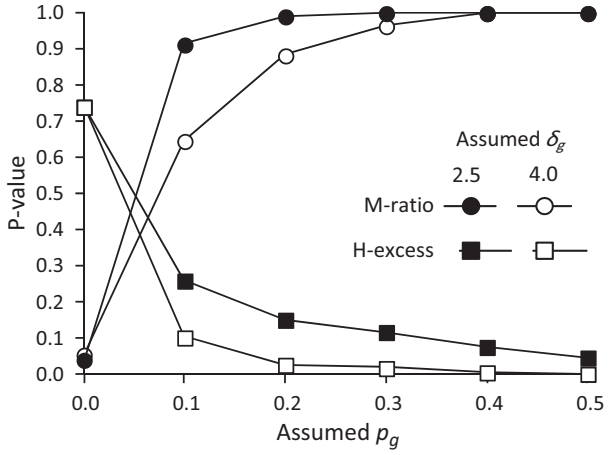
Box 3. Testing for bottlenecks in marbled murrelets

We illustrate some of the potential issues associated with uncertainty in mutation models in genetic bottleneck testing using a case study based on marbled murrelets (*Brachyramphus marmoratus*), a seabird on the US federal threatened species list that nests primarily in old-growth forests in the Pacific Northwest. Genetic bottleneck tests constitute an appealing approach for detecting population declines in a geographically isolated population of murrelets in central California because at-sea monitoring conducted from 1999 to 2010 failed to detect a population decline, despite a century of extensive nesting habitat loss and impacts from a variety of other environmental factors (Peery *et al.* 2004, 2006, 2007). Blood samples were collected from 270 live murrelets and genotyped at 12 microsatellite loci as described in Peery *et al.* (2008), and a bottleneck was tested for using heterozygosity-excess and M-ratio methods.

Inferences about the occurrence of a putative bottleneck were complicated by the sensitivity of results to the assumed mutation model (see graph). For heterozygosity excess, tests were not statistically significant when p_g = 0–0.1, but became increasingly significant with larger values of p_g. The mean M-ratio across loci was 0.90 (range = 0.54–1.00) and well above the commonly used critical value of 0.68 (Garza & Williamson 2001), suggesting that the population did not deviate from mutation-drift equilibrium. However, in contrast to heterozygosity-excess tests, significance declined as p_g increased and M-ratios were statistically significant when $p_g = 0$ (Box 3, Fig. 1). Thus whether neither of the two tests, both tests or one of the two tests was statistically significant depended entirely on the assumed mutation model.

Box 3. Graph

Probability of rejecting the null hypothesis (*P*-value) of mutation-drift equilibrium in marbled murrelets with M-ratio and heterozygosity-excess tests as a function of the assumed proportion of multi-step mutations (p_g) and mean size of multi-step mutations (δ_g).



and microsatellite evolution that likely influence inference from bottleneck tests. Our aim was not to conduct a comprehensive analysis of all factors that might influence statistical power and Type I error rates (Williamson-Natesan 2005), but instead to determine how effectively bottleneck tests as implemented in the literature can identify populations in need of conservation action. First, we determined how effective typical sample sizes are for detecting recent bottlenecks across a range of post-bottleneck N_e and determine if less extreme but potentially important population declines are likely to be detected. Second, we assessed potential biases created by incorrect choices of mutation model parameter values using parameter estimates derived from studies of microsatellite mutations in vertebrates. By highlighting potential issues in studies applying bottleneck tests, we hope to provide guidance for future genetic-based assessments of population declines and increase the reliability of bottleneck tests in conservation research and planning.

Material and methods

Literature review

We reviewed published articles that tested for recent population bottlenecks in vertebrates to determine how frequently bottlenecks were detected in both putatively stable populations and populations known to have declined recently. We also quantified typical sample sizes, assumed values for mutation model parameters, and bottleneck characteristics (e.g. magnitude of the bottleneck) and used these data to parameterize simulation-based assessments of statistical power and Type I error rates. We searched the *ISI Web of Knowledge* database and identified 1247 unique scientific articles citing Cornuet & Luikart (1996), Luikart & Cornuet (1998), Piry *et al.* (1999) or Garza & Williamson (2001) as of February 2010. To limit the scope of our review, we focused our efforts on the 105 studies for which some quantitative information was available about the status of the population (post-bottleneck census population size or percent decline) at the time of sampling (Appendix S1). In sum, the studies reviewed here used genetic bottleneck tests to test for recent declines in a total of 703 populations among 116 vertebrate species. We classified each population as either 'bottlenecked', 'stable' or 'unknown' based on the authors' assessment of the demographic history of the population.

We also reviewed 18 published studies that characterized the size and proportion of multi-step mutations in microsatellites in 15 vertebrate species. We pooled all observed mutations when estimating mutation model parameters from these studies rather than calculating the means of species- or study-specific estimates so that studies of a single species with comparatively small sample sizes did not have a disproportionate influence on overall estimates. Specifically, we estimated p_g by dividing the total number of multi-step mutations by the total number of mutations and estimated δ_g by calculating the average size of all multi-step mutations.

Simulation methods

Simulation methods used to estimate power followed the approach of Williamson-Natesan (2005), where we first simulated populations in mutation-drift equilibrium using standard coalescent simulations (Hudson 1990) and then subjected simulated populations to the bottleneck process using a Wright–Fisher model to project populations forward in time after an instantaneous reduction in effective population size. The mutation process during coalescent and Wright–Fisher simulations was modelled according to a two-phase model, as described in Box 2. Note that hereafter we refer to mutation model parameter values used to generate simulated data for bottlenecked populations as 'true' values.

We estimated statistical power to detect a bottleneck as the proportion of simulated bottlenecked populations for which M-ratios were significantly lower, or heterozygosity excess was significantly greater, than expected for populations in mutation-drift equilibrium. We estimated expected distributions of test statistics for populations in equilibrium using coalescent simulations (i.e. omitting the Wright–Fisher projections described above). Mutation model parameter values used to generate these expected distributions were termed 'assumed' values and represented assumptions made by investigators conducting bottleneck tests. For M-ratio tests, we assessed significance by comparing the mean M-ratio in each bottlenecked population to the lower 95th percentile of the mean M-ratio in 500 equilibrium populations. We estimated the expected distribution for heterozygosity by simulating single-locus microsatellite genotypes from an equilibrium population until 200 replicate genotypes were obtained for each value of K (number of alleles) present in the bottlenecked population. We used a Wilcoxon signed-rank test to test for heterozygosity excess that paired heterozygosity in the bottlenecked population at locus i with the mean heterozygosity from the 200 replicates of the simulated equilibrium population, conditional on the number of alleles at locus i (Luikart & Cornuet 1998). The difference in testing procedures between the two metrics reflected differences in the manner in which the two bottleneck tests are typically implemented in the literature. All tests were one-tailed and only tested for population bottlenecks (i.e. heterozygosity excess or a reduction in the M-ratio).

Parameter values used to conduct simulations quantifying the effects of sample size and bottleneck characteristics are provided in Table 1, and we describe the rationale for these values here. We set both assumed and true p_g and δ_g to estimates derived from the 18 studies of microsatellite mutations we reviewed (0.22 and 3.1 respectively; see below), such that the mutation model was assumed 'correctly' for all simulations evaluating the statistical power of bottleneck tests. We simulated equilibrium (pre-bottlenecked) populations assuming θ was 1, 5 or 10 (where θ was pre-bottleneck genetic diversity, $\theta = 4N_e\mu$ in a diploid organism, and μ was the per generation mutation rate). We considered four post-bottleneck effective population sizes (N_e = 25, 50, 100 and 500) to estimate power to detect a large bottleneck (N_e = 25) as well as a range of smaller bottlenecks that included thresholds where inbreeding depression (N_e = 50) and the loss of additive genetic diversity (N_e = 500) may threaten viability (Franklin 1980; Soulé 1980). Note that the bottleneck scenarios

Table 1 Simulation parameters used to estimate statistical power and Type I error rates as a function of sampling effort, demographic history and mutation model assumptions

Simulation parameter	Power analysis	Type I error rate assessment
Sampling effort		
Number of individuals	**35**, 100	**35**
Number of loci	**8**, 16	**8**
Demographic history		
θ (pre-bottleneck)	1, **5**, 10	**5**
Post-bottleneck N_e	25, 50, 100, 500	n/a
Number of generations since bottleneck	1, 5, 10, 20, 30, 40, 50	n/a
Mutation model		
Assumed proportion multi-step mutations (p_g)	**0.22**	0.10, **0.22**
True proportion multi-step mutations (p_g)	**0.22**	0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7
Assumed mean size multi-step mutations (δ_g)	**3.1**	**3.1, 3.5**
True mean size of multi-step mutations (δ_g)	**3.1**	2.5, 3, 3.5, 4

Values in bold indicate approximate median values in published studies and n/a = not applicable.

considered here represent large percentage declines; for example when $\theta = 5$, a post-bottleneck N_e of 25 represents a 98–99.9% decline and a post-bottleneck N_e of 500 represents a 60–99.4% decline assuming $\mu = 10^{-5}$–10^{-3}. Populations were simulated for 1, 5, 10, 20, 30, 40 and 50 generations forward after the bottleneck to evaluate the effect of bottleneck duration on the ability of bottleneck tests to detect reductions in effective population size. We sampled the approximate median number of individuals and loci used in bottleneck studies (35 individuals and eight loci; see below) from simulated populations and also explored the ability to increase power by increasing both the number of individuals and loci (100 individuals and 16 loci).

Simulations used to assess the effect of erroneous assumptions about mutation model parameters on Type I error rates were conducted similarly, except that simulated populations did not go through a bottleneck and individuals were sampled from populations in mutation-drift equilibrium. Moreover, true and assumed values for p_g and δ_g were allowed to differ to reflect mistaken assumptions about mutation model parameters. Type I error rates were estimated as the proportion of equilibrium populations that yielded significant M-ratio or heterozygosity-excess tests, using the testing methods described above. Deviations of assumed from true parameter values were interpreted to have an impact on tests when Type I error rates were greater than the nominal rate of 0.05.

Parameter values used to conduct simulations assessing the effects of erroneous assumptions about mutation models are presented in Table 1, and we describe the rationale for these values here. In separate simulations, we varied true p_g from 0.1 to 0.7 by increments of 0.10 to reflect the approximate range of values observed among species in studies of microsatellite mutations (see below). We set assumed p_g to 0.10 to represent the most commonly used value in bottleneck studies (see below) and assess the impacts of what are likely typical errors in assumptions about mutation models. We also conducted simulations with assumed p_g set to 0.22 to reflect the estimate derived from the microsatellite mutation studies we reviewed (see below) and determine if using a more realistic value could improve error rates. For simulations evaluating the effects of erroneous assumptions of p_g, we set both true and assumed δ_g to 3.1. We evaluated the impacts of mistaken assumptions about the size of multi-step mutations in a separate set of simulations, where we set true δ_g from 2.5 to 4.0 by increments of 0.5 to reflect the approximate range of values observed in microsatellite mutation studies (see below). We set assumed δ_g to 3.5 to reflect the most commonly used value in M-ratio tests and assess what are likely typical errors in assumptions about mutation models. We also set assumed δ_g to 3.1 to reflect the best estimate from the microsatellite mutation studies (see below) and determine if using a more realistic value could improve error rates. We set both the true and assumed p_g to 0.22 when simulating the effects of mistaken assumptions about δ_g on Type 1 error rates.

Results

Literature review of the application and effectiveness of bottleneck tests

Our literature review indicated that genetic bottleneck tests were generally conducted with modest sample

sizes of both individuals (median = 38 and 31 for M-ratio and heterozygosity-excess tests respectively) and loci (median = 9 and 8 for M-ratio and heterozygosity-excess tests respectively; Table 2). In cases where demographic information was available, estimated bottlenecks were large both in terms of the percentage decline in census population size (median = 80 and 95% for M-ratio and heterozygosity-excess tests respectively) and the post-bottleneck census population size (median 24 and 100 individuals for M-ratio and heterozygosity-excess tests respectively). Most (86%) of the studies that reported mutation model parameters only explored a single mutation model for both tests. The most commonly assumed p_g was 0.10 for both test types, whereas the most commonly assumed δ_g was 3.5 for M-ratio tests and 3.1 for heterozygosity-excess tests respectively.

Our literature review indicated that genetic bottleneck tests often failed to detect bottlenecks in populations that investigators believed had experienced a reduction in abundance. M-ratios were significantly lower than expected under equilibrium in only 43–56% of 54 bottlenecked populations (n = 22 species; Tables 3 and S1, Supporting information). Moreover, the median M-ratio among 48 bottlenecked populations (n = 18 species) for which this value was available (0.710) was similar to the median M-ratio (0.749) among the 22 putatively stable populations (n = 6 species), and variability was high among bottlenecked populations (range = 0.170–0.986; Table S1, Supporting information). Statistically significant heterozygosity excess was detected in only 26–47% of 91 bottlenecked populations (n = 33 species) we reviewed, depending on whether significance was required for all mutation models considered, or whether only a single mutation model was required to yield a statistically significant result (Tables 3 and S1, Supporting information). Note that population declines often went undetected with both methods despite the fact that many bottlenecks were large. Indeed, the median reduction in census population size reflects declines to effective population sizes of only 3–14 individuals, assuming typical effective to census population size ratios of 0.11–0.14 (Frankham 1995; Palstra & Ruzzante 2008).

Both methods detected bottlenecks in putatively stable populations somewhat more frequently than expected assuming a significance level of 0.05 (Tables 3 and S2, Supporting information). Specifically, the frequency of Type I errors for M-ratio and heterozygosity-excess tests was 0.13 (n = 32 populations of 10 species) and 0.05–0.20 (n = 41 populations of 16 species) respectively, depending on whether one or all mutation models were required to yield a significant result. While these Type I error rates were not excessively high, limited sample sizes could have prevented violations of assumptions from resulting in more Type I errors (see below).

Review of studies of microsatellite mutations in vertebrates

Information associated with 592 mutations detected in the 18 studies of microsatellite evolution in vertebrates is summarized in Table 4. Some 'non-model' species

Table 2 Sampling effort, bottleneck characteristics and mutation models in 105 peer-reviewed genetic bottleneck studies that used either M-ratios or heterozygosity excess to test for recent population declines. For sampling effort and bottleneck characteristics, n reflects the number of populations tested. For two-phase mutation model parameters, n reflects the number of unique tests conducted with different mutation models. θ = genetic diversity, N_c = census population size and n/a = not applicable. References for studies included in this summary are provided in Appendix S1

	M-ratio			Heterozygosity excess		
Summary statistic	Median or mode	10th, 90th percentile	n	Median or mode	10th, 90th percentile	n
Sampling effort						
Median number of individuals	38	16, 100	359	31	15, 100	595
Median number of loci	9	6, 14	359	8	5, 14	595
Bottleneck characteristics						
Median percent decline	80	55, 97	10	95	65, 100	20
Median post-bottleneck N_c	24	7, 500	19	100	11, 500	53
Median θ	5.4	0.3, 29.3	33	n/a	n/a	n/a
Two-phase mutation model parameters						
Modal percent multi-step mutations (p_g)	10	n/a	321	10	n/a	572
Modal mean multi-step mutations size (δ_g)	3.5	n/a	299	3.1	n/a	n/a

had higher mutation rates than usually cited for microsatellites (the highest being 1.8×10^{-2} mutations per generation in barn swallows; *Hirundo rustica*), and it could be argued that mutations at these loci were (i) not generally representative of mutations in microsatellites; or (ii) the result of allelic mismatches caused by incorrect paternity assignments. Indeed, high mutation rates in some species may have reflected ascertainment bias during marker development where the most polymorphic loci were targeted for use in mutation studies. However, we did not detect a relationship between the mutation rate and either the proportion or the mean size of multi-step mutations using linear regression (p_g: $P = 0.35$, $R^2 = 0.12$; δ_g: $P = 0.92$, $R^2 < 0.01$) based on the studies with at least 20 mutations. Moreover, highly polymorphic loci are also frequently used in conservation and population genetic studies; for example the loci used to characterize mutations in kangaroo rats (*Dipodomys spectabilis*) were also used to test for bottlenecks in this species (Busch *et al.* 2007). It is also unlikely that the mutations we considered were an artefact of incorrect paternity assignments because 68% of single-step mutations and 66% of multi-step mutations that could be attributed to an individual parent were inherited from the known rather than the inferred parent. For these reasons, we believe that our estimates of mutation model parameters provided a reasonable representation of microsatellite evolution in vertebrates.

We estimated that $p_g = 0.22$ based on the studies summarized in Table 4, a value that was considerably greater than (i) recommended by Garza & Williamson (2001) and Piry *et al.* (1999) for M-ratio and heterozygosity-excess tests; and (ii) the most commonly used

Table 3 Proportion of bottlenecked and putatively stable populations for which M-ratio and heterozygosity-excess tests were statistically significant derived from 105 peer-reviewed articles. When multiple mutation models were explored, we determined significance based on whether all mutation models yielded significant tests or whether at least one mutation model yielded a significant test. n_{pop} and n_{sp} were the sample sizes of populations and species respectively. The proportion of significant tests was calculated using n_{pop}

Significant tests	M-ratio (n_{pop}, n_{sp})	Heterozygosity excess (n_{pop}, n_{sp})
Bottlenecked populations		
≥1 test	0.56 (54, 22)	0.47 (91, 33)
All tests	0.43 (54, 22)	0.26 (91, 33)
Stable populations		
≥1 test	0.13 (32, 10)	0.20 (41, 16)
All tests	0.13 (32, 10)	0.04 (41, 16)

Table 4 Estimates of microsatellite mutation parameters in 15 vertebrate species derived from 18 studies using known or inferred pedigrees

Common name	Scientific name	Loci	$n_{meioses}$	n_{mut}	n_{multi}	Mutation rate	p_g	δ_g	Ref
Lesser Kestrel	*Falco naumanni*	9	3370	10	4	0.00297	0.40	2.5	Ortego *et al.* (2008)
House Wren	*Troglodytes aedon*	3	12 260	20	10	0.00160	0.50	4.0	Masters *et al.* (2011)
Barn Swallow	*Hirundo rustica*	2	3053	27	6	0.00884	0.22	2.2	Brohede *et al.* (2002)
Superb fairy-wren	*Malurus cyaneus*	2	5980	75	15	0.01250	0.20	3.9	Beck *et al.* (2003)
Gidgee Skink	*Egernia stokesii*	7	1110	13	6	0.01171	0.46	2.8	Gardner *et al.* (2000)
Green Turtle	*Chelonia mydas*	5	9090	24	13	0.00264	0.57	4.0	Fitzsimmons (1998)
Olive Ridley Turtle	*Lepidochelys olivaceae*	2	2812	33	9	0.01174	0.27	3.3	Hoekert *et al.* (2002)
Tiger Salamander	*Ambystoma tigrinum*	6	7906	10	0	0.00126	0.00	n/a	Bulut *et al.* (2009)
Broadnosed Pipefish	*Syngnathus typhle*	4	12 780	26	3	0.00203	0.12	2.3	Jones *et al.* (1999)
Pink Salmon	*Oncorhyncus gorbuscha*	7	3032	11	0	0.00363	0.00	n/a	Steinberg *et al.* (2002)
Kangaroo Rat	*Dipodomys spectabilis*	8	n/p	22	9	0.00810	0.41	2.6	Busch *et al.* (2007)
Zebrafish	*Danio rerio*	2000	176 000	28	19	0.00016	0.68	n/p	Shimoda *et al.* (1999)
Pig	*Sus domestica*	62	24 414	2	0	0.00008	0.00	n/a	Ellegren (1995)
Sheep	*Ovis aries*	246	46 225	5	0	0.00011	0.00	n/a	Crawford & Cuthbertson (1996)
Humans	*Homo sapiens*	151	320 084	284	37	0.00089	0.13	2.6	Brinkmann *et al.* (1998), Kayser & Sajantila (2001), Sajantila *et al.* (1999) and Xu *et al.* (2000)

p_g = the proportion of multi-step mutations, δ_g = the mean size of multi-step mutations, $n_{meioses}$ = number of meiotic events, n_{mut} = number of mutations, n_{multi} = number of multi-step mutations. The per generation mutation rate was estimated by dividing n_{mut} by $n_{meioses}$ and p_g was estimated by dividing n_{multi} by n_{mut}. n/p = not provided by the authors of the study and n/a = not applicable.

value in bottleneck studies (i.e. $p_g = 0.10$). The proportion of multi-step repeats varied significantly among species for which at least 20 mutations were characterized based on a chi-square test ($\chi^2_8 = 77.0$, $P < 0.001$), and ranged from 0.12 to 0.68 (Table 4). We estimated that $\delta_g = 3.1$, a value that was identical to the recommendations of Piry *et al.* (1999) for heterozygosity-excess tests, but slightly lower than that Garza & Williamson (2001) recommended for M-ratio tests ($\delta_g = 3.5$). The mean size of multi-step mutations was significantly different among species for which at least nine multi-step mutations were observed based on one-way analysis of variance ($F_{4,73} = 3.74$, $P = 0.007$), and ranged from 2.6 to 4.0 repeats (Table 4).

Statistical power of genetic bottleneck tests

Simulation analyses indicated that genetic bottleneck tests can have limited power to detect population declines, particularly given the limited sample sizes of individuals and loci typically employed in bottleneck tests. For M-ratios, power to detect large reductions in effective population size ($N_e = 25$) 10 generations after the bottleneck was modest (0.63) using approximate median sample sizes and assuming $\theta = 5$, but power to detect smaller declines was considerably lower (0.38–0.08 for $N_e = 50$–500, Fig. 1a). Power to detect bottlenecks was greater when $\theta = 10$, but was still only 0.38 when post-bottleneck $N_e = 100$. Power was very low (≤0.11) for the range of post-bottleneck N_e considered when $\theta = 1$ (Fig. 1a). Power to detect bottlenecks of $N_e = 50$ could be increased to reasonable levels (0.80) by increasing sampling effort to 100 individuals and 16 loci when $\theta = 5$, but remained low for larger bottlenecks when $\theta = 1$ (Fig. 1b). Power to detect bottlenecks with M-ratios was greater when populations were sampled a greater amount of time (up to 50 generations) after the bottleneck, but power was generally very low when populations were sampled only 1 and 5 generations after the bottleneck (Fig. 2a).

Power to detect bottlenecks after 10 generations with heterozygosity-excess tests was low (≤0.27) for all post-bottleneck effective population sizes considered, given the approximate median sample sizes when $\theta = 1$ or 5 (Fig. 1c). Moreover, power was modest (<0.60) for all post-bottleneck effective population sizes even with elevated sampling effort (100 individuals genotyped at 16 loci) when $\theta = 5$ (Fig. 1d). Power to detect declines to $N_e = 25$ and 50 could be increased to reasonable levels (>0.70) with greater sampling effort when $\theta = 1$. It was not possible to estimate statistical power for heterozygosity-excess tests conditional on allele number when $\theta = 10$ because bottlenecks in populations with high genetic diversity often resulted in allele numbers that were lower than allele numbers in the expected data (Fig. S1, Supporting information). Low power for heterozygosity-excess tests was not an artefact of arbitrarily modelling a bottleneck duration of 10 generations; power was <0.40 for bottlenecks lasting from 1 to 50 generations with the approximate median sample size and $\theta = 5$, regardless of post-bottleneck N_e (Fig. 2b).

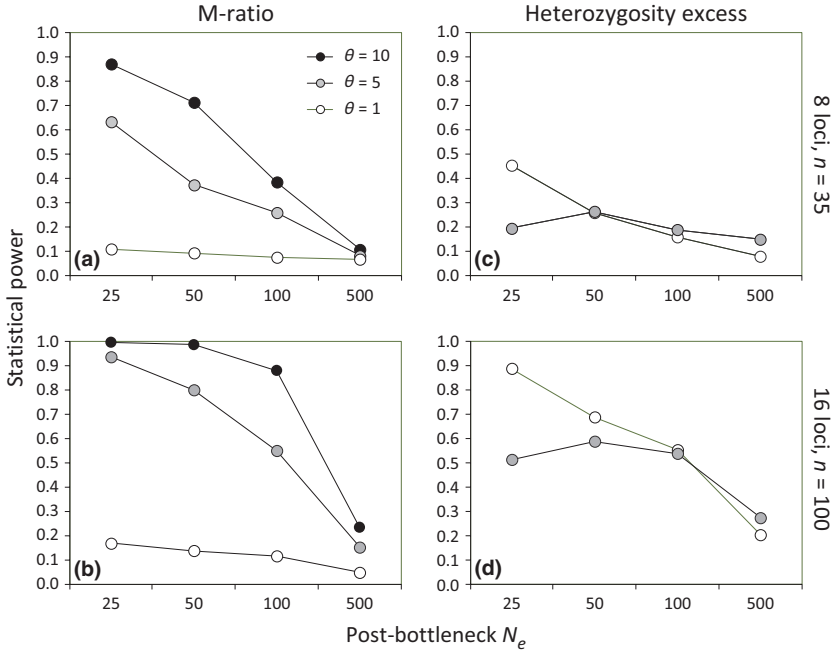


Fig. 1 Statistical power to detect a population bottleneck using M-ratio and heterozygosity-excess tests as a function of post-bottleneck N_e, the number of loci and individuals sampled, and pre-bottleneck genetic diversity (θ). Bottleneck duration was set to 10 generations, and the true and assumed mutation models were identical ($p_g = 0.22$ and $\delta_g = 3.1$).

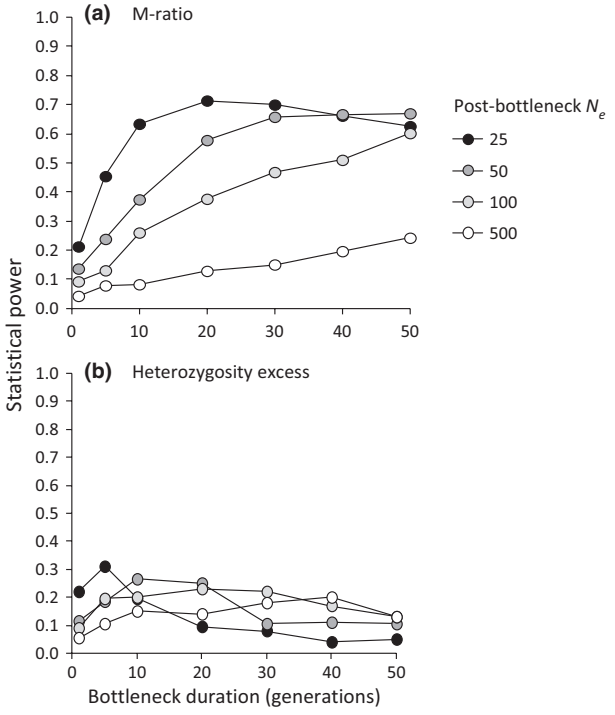


Fig. 2 Probability of detecting a bottleneck (statistical power) as a function of bottleneck duration and post-bottleneck effective population size (post-bottleneck N_e) using M-ratio and heterozygosity-excess tests. The true and assumed mutation models were identical (p_g = 0.22 and δ_g = 3.1), θ = 5 and 8 loci and 35 individuals were sampled for all simulations.

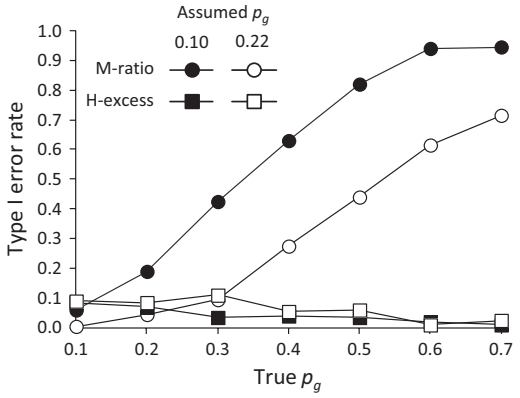


Fig. 3 Probability of detecting a bottleneck in a stable population (Type 1 error rate) with M-ratio and heterozygosity-excess tests as a function of the true and assumed proportion of multi-step mutations (p_g).

Effects of violations of mutation model assumptions on genetic bottleneck tests

Simulation analyses indicated that heterozygosity-excess tests were reasonably robust to incorrect assumptions about mutation models, but M-ratio tests were quite sensitive to differences between the true and assumed proportion of multi-step mutations (Fig. 3). Specifically, Type I error rates for heterozygosity-excess tests were ≤0.11, regardless of assumptions of p_g, whereas error rates for M-ratio tests increased rapidly with increasing differences between true and assumed values for p_g. The error rate for M-ratio tests was estimated to be 0.20 when p_g was assumed to be 0.10 (the most common value used in bottleneck studies), but the true value for p_g was 0.22 (the overall estimate across microsatellite mutation studies). However, p_g is generally unknown in the species of interest and true p_g may often be >0.22 (Table 4), which can result in even higher error rates (Fig. 3). In an extreme case, the error rate was estimated to be 0.95 when true p_g equalled 0.70 (the approximate estimate for zebrafish, *Danio rerio*) and p_g was assumed to be 0.10. Error rates could be reduced considerably by assuming that p_g = 0.22 instead of 0.10; for example when true p_g = 0.3, error rates could be reduced from 0.43 to 0.10. Nevertheless, error rates were high (≥0.28) when the true p_g was 0.40 or greater, even if p_g was assumed to be 0.22.

Both M-ratio and heterozygosity-excess tests were reasonably robust to incorrect assumptions of δ_g when p_g was assumed correctly. Specifically, Type I error rates did not exceed 0.12 across the range of δ_g observed in studies of microsatellite mutations for either test type (data not shown).

Discussion

Our review suggests that, as typically applied, bottleneck tests often failed to detect declines in populations known to have experienced population reductions and typically have low statistical power as a result of limited sample sizes. Indeed, power to detect even very large bottlenecks based on the level of heterozygosity excess is generally limited, whereas power to detect bottlenecks with M-ratios depends on many factors, but was likely low in many empirical studies of vertebrates of conservation concern. Our results are similar to those of a recent analysis by Girod *et al.* (2011) who found limited power to detect 10- to 1000-fold population declines with heterozygosity-excess tests and 10-fold declines with M-ratio tests using similar sample sizes of loci and individuals. However, Girod *et al.* (2011) used a stepwise mutation model; whereas we showed that power can be low for a more realistic two-phase mutation model. In addition, we showed that power to detect population declines with M-ratio tests was particularly low when populations were sampled one and five generations after the bottleneck and generally did not reach an asymptote until 20 generations under

typical sampling designs. Thus, bottlenecked populations could experience the negative impacts of small population processes for many generations before even the more powerful of the two tests we evaluated has a reasonable probability of detecting the bottleneck.

Studies employing bottleneck tests should, when possible, employ larger samples sizes of individuals and loci than used in the studies we reviewed (median = 8–9 loci and 31–38 individuals). In cases where a bottleneck test fails to reject mutation-drift equilibrium, inference may be improved by estimating power to detect a bottleneck across a range of possible demographic histories given the available sample size. Only one of the 105 studies we reviewed conducted such an assessment. If a test fails to detect a bottleneck but simulations reveal that statistical power was high, inference that a bottleneck did not occur is more robust. Nevertheless, a thorough discussion of alternative biological explanations for non-significance such as the potential effects of immigration or a brief bottleneck is warranted. Quantitative assessments of the possible role of factors such as immigration and bottleneck duration in determining statistical significance can help provide support for alternative hypotheses (Hundertmark & Van Daele 2010). However, none of the studies we reviewed conducted a quantitative analysis of the possible confounding effects of immigration on their findings, although wild populations are rarely completely closed and even small numbers of migrants can mask the genetic signature of bottlenecks (Keller *et al.* 2001; Busch *et al.* 2007). Inference can also be strengthened by comparing genetic variation before and after the bottleneck [e. g. with museum specimens; Beaumont (2003); Leonard (2008)] when such data are available because multi-point sampling methods are generally more powerful in a statistical sense (Ramakrishnan *et al.* 2005). Comparing genetic variation in the population of interest to variation in putatively stable populations of the same species can also add insight (Ennen *et al.* 2010; Hundertmark & Van Daele 2010), although the choice of comparative data needs to be made with care as the evolutionary history of populations can differ in many ways.

We also showed that bottleneck studies often assume values for mutation model parameters that may not be representative of microsatellite evolution in vertebrates and that the proportion of multi-step mutations is, on average, underestimated by a factor of approximately two. Whereas heterozygosity-excess tests are reasonably robust, M-ratio tests can be quite sensitive to mistaken assumptions about the proportion of multi-step mutations which can lead to high probabilities of inferring that bottlenecks occurred in stable populations. Moreover, assuming a single set of mutation model parameter values (as was the case in 86% of bottleneck studies) is not sufficient to infer a bottleneck, given the fact that parameter values appear to vary widely among vertebrates. Inference is stronger when results are consistent across a range of possible values for the proportion of multi-step mutations (Guinand & Scribner 2003). Requiring significance across a full range of values for p_g observed in microsatellite mutation studies (up to 0.68) will, however, likely result in highly stringent tests. Rather, we suggest that future studies explore a range of values up to at least 0.22, given that this value represented the 'average' estimate from studies of microsatellite evolution in vertebrates. Indeed, assuming that $p_g = 0.22$, as opposed to $p_g = 0.10$, reduces Type I error rates for M-ratio tests considerably across a range of values for true p_g, although Type I error rates will still likely be high if true $p_g \geq 0.40$ (Fig. 3). If statistical significance depends on the assumed proportion of multi-step mutations within this range, we believe the most appropriate inference is that the possible occurrence of a bottleneck remains uncertain.

Differences in sensitivities to violations of assumptions about mutation models between M-ratio and heterozygosity-excess tests also have important implications for assessing the timing of population declines with bottleneck tests. In principle, heterozygosity excess is expected to regain mutation-drift equilibrium more rapidly than M-ratios because new alleles arising from mutations do not necessarily increase M-ratios (Garza & Williamson 2001) and M-ratios change more slowly than heterozygosity immediately following a population bottleneck (Williamson-Natesan 2005). As a result, researchers often infer that a bottleneck occurred historically when M-ratio but not heterozygosity-excess tests are significant (Spear *et al.* 2006; Henry *et al.* 2009; Marshall *et al.* 2009), and conversely, that the population decline occurred more recently when heterozygosity excess, but not M-ratios, is significantly different from expectations (Funk *et al.* 2010; Hundertmark & Van Daele 2010). Our simulations indicated that inferring the timing of a population bottleneck based on differences in statistical significance between tests is probably not reasonable because M-ratio and heterozygosity-excess tests are affected differently by violations of assumptions regarding mutation models. Specifically, the probability of incorrectly rejecting the null hypothesis can increase dramatically with the assumed proportion of multi-step mutations for M-ratio tests, but heterozygosity-excess tests are relatively robust to assumptions about this parameter (Fig. 3). Even when mutation model is assumed correctly, power to detect a bottleneck can differ between M-ratio and heterozygosity-excess tests for reasons other than the timing of the decline. For example statistical power for heterozygosity-excess tests may be greater than for M-ratio tests when

pre-bottleneck genetic diversity is low, but heterozygosity-excess tests may be less powerful than M-ratio tests when pre-bottleneck genetic diversity is high (Fig. 1).

Current trends and future directions for characterizing demographic history

The use of summary statistics such M-ratios and heterozygosity to detect population bottlenecks is transitioning to less *ad-hoc* methods for characterizing demographic history. In particular, Bayesian methods that estimate posterior distributions of demographic parameters and evaluate competing models of demographic history are undergoing rapid development and are increasingly being adopted by empirical studies (Beaumont 1999, 2010; Bertorelle *et al.* 2010). Full Bayesian methods make use of coalescence theory and Markov chain Monte Carlo sampling to estimate the likelihood of genealogical trees of sampled genes conditional on model parameters such as effective population size and the timing of demographic changes. Indeed, Bayesian methods implemented in program MSvar (Beaumont 1999; Storz & Beaumont 2002) can have a higher probability of detecting bottlenecks than M-ratio or heterozygosity-excess tests, and estimates of population size and bottleneck duration seem more robust to certain violations of mutation model assumptions (Girod *et al.* 2011). Moreover, Bayesian skyline plots can provide estimates of effective population size between individual coalescent events (Drummond *et al.* 2005; Drummond & Rambaut 2007). Approximate Bayesian Computation (ABC) methods allow for the implementation of more complex population models as they do not rely on computing the full likelihood (Beaumont 2010; Bertorelle *et al.* 2010). Hoffman *et al.* (2011) recently used ABC to reconstruct accurately the demographic history of the heavily exploited Antarctic fur seal (*Arctocephalus gazelle*) when the bottleneck tests described herein failed to detect a decline. Moreover, statistical software packages have recently been developed for both microsatellite and sequence data that make these methods accessible to non-experts (Cornuet *et al.* 2008, 2010; Wegmann *et al.* 2009). Nevertheless, ABC methods for inferring demographic history require careful consideration of a number of factors including alternative models, which summary statistics to use in the estimation procedure and the number of simulations, as well as the sensitivity of results to these choices (Bertorelle *et al.* 2010). Moreover, additional assessments will be necessary to determine how much sampling effort is required to achieve adequate levels of precision in parameter estimates and reliably discriminate among competing models, as well as to understand how sensitive results are to violations of model assumptions.

The ability to sequence a large fraction of the genome of non-model species with next-generation sequencing technologies has the potential to improve the characterization of demographic history in several ways. Indeed, these technologies offer the opportunity to sequence hundreds of nucleotides at thousands of loci across individual genomes (Davey *et al.* 2011), which will increase power to detect recent bottlenecks and characterize demographic history by increasing the number of loci that can be analysed with methods described above (Allendorf *et al.* 2010). Next-generation sequencing technologies now make it possible to identify large numbers of single nucleotide polymorphism (SNP) loci, but it remains unclear if SNPs will improve the power of bottleneck tests because the loss of alleles in bottlenecked populations generally results in monomorphic loci (Morin *et al.* 2004). However, sequencing large fractions of genomes makes it possible to apply at least two newer classes of genetic methods that are based on haplotype data to questions of demographic history. The first of these involves linkage tract-based methods, where linkage tracts represent regions of DNA that are identical by descent (IDB) between gene copies present in two individuals (Albrechtsen *et al.* 2009; Gusev *et al.* 2009; Pool *et al.* 2010). The length of linkage tracks is a function of recent effective population size because gene copies are more likely to be identical by descent in small populations where, on average, individuals are more closely related than in large populations. Indeed, a recent study was able to estimate effective population size based on the distribution of linkage-tract lengths in an isolated human population only 23 generations ago (Gusev *et al.* 2012). The second approach involves the use of the allele frequency spectrum, which is a measure of the observed frequency distribution of alleles at a large number of loci [Pool *et al.* (2010); see also Luikart *et al.* (1998) for a similar test using microsatellites]. In principle, a bottleneck can be inferred if the allele frequency spectrum in a sample of individuals taken from a population of interest contains fewer rare alleles than expected. To the best of our knowledge, no study has yet tested for population bottlenecks in a species of concern using either allele frequency spectrum or linkage tract data derived from next-generation sequencing, but we suspect that the use of such methods in conservation will increase dramatically over the coming years.

Acknowledgements

BNR was funded in part by a STAR Fellowship from the US Environmental Protection Agency, EDB was funded by a National Science Foundation IGERT Fellowship and PJP was funded in part by a Hrdy Fellowship in Conservation Biology at Harvard University. We thank four anonymous reviewers for helpful comments on earlier drafts of this work.

References

Aguilar A, Jessup DA, Estes J, Garza JC (2008) The distribution of nuclear genetic variation and historical demography of sea otters. *Animal Conservation*, **11**, 35–45.

Albrechtsen A, Korneliussen TS, Moltke I, Hansen TV, Nielsen FC, Nielsen R (2009) Relatedness mapping and tracts of relatedness for genome-wide data in the presence of linkage disequilibrium. *Genetic Epidemiology*, **33**, 266–274.

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. *Nature Reviews Genetics*, **11**, 697–709.

Beaumont MA (1999) Detecting population expansion and decline using microsatellites. *Genetics*, **153**, 2013–2029.

Beaumont MA (2003) Estimation of population growth or decline in genetically monitored populations. *Genetics*, **164**, 1139–1160.

Beaumont MA (2010) Approximate Bayesian Computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics*, **41**, 379–406.

Beck NR, Double MC, Cockburn A (2003) Microsatellite evolution at two hypervariable loci revealed by extensive avian pedigrees. *Molecular Biology and Evolution*, **20**, 54–61.

Beissinger SR, Peery MZ (2007) Reconstructing the historic demography of an endangered seabird. *Ecology*, **88**, 296–305.

Bertorelle G, Benazzo A, Mona S (2010) ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Molecular Ecology*, **19**, 2609–2625.

Bonebrake TC, Christensen J, Boggs CL, Ehrlich PR (2010) Population decline assessment, historical baselines, and conservation. *Conservation Letters*, **3**, 371–378.

Bouzat JL (2010) Conservation genetics of population bottlenecks: the role of chance, selection, and history. *Conservation Genetics*, **11**, 139–147.

Brinkmann B, Klintschar M, Neuhuber F, Hühne J, Rolf B (1998) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *American Journal of Human Genetics*, **62**, 1408–1415.

Brohede J, Primmer CR, Møller A, Ellegren H (2002) Heterogeneity in the rate and pattern of germline mutation at individual microsatellite loci. *Nucleic Acids Research*, **30**, 1997–2003.

Bulut Z, McCormick CR, Gopurenko D, Williams RN, Bos DH, DeWoody JA (2009) Microsatellite mutation rates in the eastern tiger salamander (*Ambystoma tigrinum tigrinum*) differ 10-fold across loci. *Genetics*, **136**, 501–504.

Busch JD, Waser PM, DeWoody JA (2007) Recent demographic bottlenecks are not accompanied by a genetic signature in banner-tailed kangaroo rats (*Dipodomys spectabilis*). *Molecular Ecology*, **16**, 2450–2462.

Caughley G (1994) Directions in conservation biology. *The Journal of Animal Ecology*, **63**, 215–244.

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

Cornuet JM, Santo F, Beaumont MA *et al.* (2008) Inferring population history with DIYABC: a user-friendly approach to Approximate Bayesian Computations. *Bioinformatics*, **24**, 2713–2719.

Cornuet JM, Ravigné V, Estoup A (2010) Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics*, **11**, 401.

Crawford AM, Cuthbertson RP (1996) Mutations in sheep microsatellites. *Genome Research*, **6**, 876–879.

Davey JW, Hohenlohe PA, Etter PD *et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, **12**, 499–510.

Di Rienzo A, Peterson AC, Garza JC, Valdes A-M, Slatkin M, Freimer NB (1994) Mutational processes of simple sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences*, **91**, 3166–3170.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.

Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, **22**, 1185–1192.

Ellegren H (1995) Mutation rates at porcine microsatellite loci. *Mammalian Genome*, **6**, 376–377.

Ellegren H (2000) Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in Genetics*, **16**, 551–558.

Ellegren H (2004) Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics*, **5**, 435–445.

Ennen JR, Kreiser BR, Qualls CP (2010) Low genetic diversity in several gopher tortoise (*Gopherus polyphemus*) populations in the DeSota National Forest, Mississippi. *Herpetologica*, **66**, 31–38.

Fitzsimmons NN (1998) Single paternity of clutches and sperm storage in the promiscuous green turtle (*Chelonia mydas*). *Molecular Ecology*, **7**, 575–584.

Flather CH, Hayward GD, Beissinger SR, Stephens PA (2011) Minimum viable populations: is there a 'magic number' for conservation practicioners? *Trends in Ecology & Evolution*, **26**, 307–316.

Frankham R (1995) Effective population-size adult-population size ratios in wildlife-a review. *Genetical Research*, **66**, 95–107.

Frankham R, Lees K, Montgomery ME, England PE, Lowe EH, Briscoe DA (1999) Do population size bottlenecks reduce evolutionary potential? *Animal Conservation*, **2**, 255–260.

Franklin IR (1980) Evolutionary change in small populations. In: *Conservation Biology: An Evolutionary-Ecological Perspective* (eds Soulé ME, Wilcox BA), pp. 135–148. Sinauer Associates, Sunderland, Massachusetts.

Funk WC, Forsman ED, Johnson M, Mullins TD, Haig SM (2010) Evidence for recent population bottlenecks in northern spotted owls (*Strix occidentalis caurina*). *Conservation Genetics*, **11**, 1013–1021.

Gardner MG, Bull CM, Cooper SJB, Duffield GA (2000) Microsatellite mutations in litters of the Australian lizard *Egernia stokesii*. *Journal of Evolutionary Biology*, **13**, 551–560.

Garza JC, Williamson EG (2001) Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology*, **10**, 305–318.

Girod C, Vitalis R, Lebois R, Freville H (2011) Inferring population decline and expansion from microsatellite data: a simulation-based evaluation of the MSvar methods. *Genetics*, **188**, 165–179.

Green RE (2008) Demographic mechanism of a historical bird population collapse reconstructed using museum specimens. *Proceedings of the Royal Society B*, **275**, 2381–2387.

Guinand B, Scribner KT (2003) Evaluation of methodology for detection of genetic bottlenecks: inferences from temporally replicated lake trout populations. *Comptes Rendus Biologies*, **326**, S61–S67.

Gusev A, Lowe JK, Stoffel M *et al.* (2009) Whole population, genome-wide mapping of hidden relatedness. *Genome Research*, **19**, 318–326.

Gusev A, Palamara PF, Aponte G *et al.* (2012) The architecture of long-range haplotypes shared within and across populations. *Molecular Biology and Evolution*, **29**, 473–486.

Henry P, Miquelle D, Sugimoto T, McCullough DR, Caccone A, Russello MA (2009) *In situ* population structure and *ex situ* representation of the endangered Amur tiger. *Molecular Ecology*, **18**, 3173–3184.

Hoekert WEJ, Neuféglise H, Schouten AD, Menken SBJ (2002) Multiple paternity and female-biased mutation at a microsatellite locus in the olive ridley sea turtle (*Lepidochelys olivacea*). *Heredity*, **89**, 107–113.

Hoffman JI, Grant SM, Forcada J, Phillips CD (2011) Bayesian inference of a historical bottleneck in a heavily exploited marine mammal. *Molecular Ecology*, **20**, 3989–4008.

Hudson RR (1990) Gene genealogies and the coalescent process. In: *Oxford Survey in Evolutionary Biology* (eds Futyma D, Antonovics J), pp. 1–42. Oxford University Press, Oxford.

Hundertmark KJ, Van Daele LJ (2010) Founder effect and bottleneck signatures in an introduced, insular population of elk. *Conservation Genetics*, **11**, 139–147.

Jones AG, Rosenqvist G, Berglund A, Avise JC (1999) Clustered microsatellite mutations in the pipefish *Syngnathus typhle*. *Genetics*, **152**, 1057–1063.

Kayser M, Sajantila A (2001) Mutations at Y-STR loci: implications for paternity testing and forensic analysis. *Forensic Science International*, **118**, 116–121.

Keller LF, Waller DM (2002) Inbreeding effects in wild populations. *Trends in Ecology & Evolution*, **17**, 230–241.

Keller LF, Jeffery KJ, Arcese P *et al.* (2001) Immigration and the ephemerality of a natural population bottleneck: evidence from molecular markers. *Proceedings of the Royal Society B*, **268**, 1387–1394.

Leonard JA (2008) Ancient DNA applications for wildlife conservation. *Molecular Ecology*, **17**, 4186–4196.

Luikart G, Cornuet JM (1998) Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conservation Biology*, **12**, 228–237.

Luikart G, Allendorf FW, Cournuet JM, Sherwin WB (1998) Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity*, **89**, 238–247.

Marshall J, Kingsbury B, Minchella D (2009) Microsatellite variation, population structure, and bottlenecks in the threatened copperbelly water snake. *Conservation Genetics*, **10**, 465–476.

Masters BS, Johnson LS, Johnson BGP, Brubaker JL, Sakaluk SK, Thompson CF (2011) Evidence for heterozygote instability in microsatellite loci in house wrens. *Biology Letters*, **7**, 127–130.

Morin PA, Luikart G, Wayne RK (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution*, **19**, 208–216.

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. *Evolution*, **29**, 1–10.

NMFS (2010) *Bi-National Recovery Plan for the Kemp's Ridley Sea Turtle (*Lepidochelys kempii*)*, 2nd revision. National Marine Fisheries Service, Silver Spring, Maryland.

O'Brien SJ, Evermann JF (1988) Interactive influence of infectious disease and genetic diversity in natural poputations. *Trends in Ecology & Evolution*, **3**, 245–259.

Ortego J, Aparicio JM, Cordero PJ, Calabuig G (2008) Characteristics of loci and individuals are associated with mutation rates in lesser kestrels (*Falco naumanni*). *Mutation Research/Fundamentals and Mechanisms of Mutagenesis*, **648**, 82–86.

Palstra FP, Ruzzante DE (2008) Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild population persistence? *Molecular Ecology*, **17**, 3428–3447.

Peery MZ, Beissinger SR, Newman SH, Burkett E, Williams TD (2004) Applying the declining population paradigm: diagnosing causes of poor reproduction in the marbled murrelet. *Conservation Biology*, **18**, 1088–1098.

Peery MZ, Becker BH, Beissinger SR (2006) Combining demographic and count-based approaches to identify source-sink dynamics of a threatened seabird. *Ecological Applications*, **16**, 1516–1528.

Peery MZ, Becker BH, Beissinger SR (2007) Age ratios as estimators of productivity: testing assumptions on a threatened seabird, the marbled murrelet (*Brachyramphus marmoratus*). *Auk*, **124**, 224–240.

Peery MZ, Beissinger SR, House RF *et al.* (2008) Characterizing source-sink dynamics with genetic parentage assignments. *Ecology*, **89**, 2746–2759.

Piry S, Luikart G, Cornuet JM (1999) BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity*, **90**, 502–503.

Pool JE, Hellmann I, Jensen JD, Nielsen R (2010) Population genetic inference from genomic sequence variation. *Genome Research*, **20**, 291–300.

Ramakrishnan U, Hadly EA, Mountain JL (2005) Detecting past population bottlenecks using temporal genetic data. *Molecular Ecology*, **14**, 2915–2922.

Reusch TBH, Wood TE (2007) Molecular ecology of global change. *Molecular Ecology*, **16**, 3973–3992.

Sajantila A, Lukka M, Syvänen A-C (1999) Experimentally observed germline mutations at human micro- and minisatellite loci. *European Journal of Human Genetics*, **7**, 263–266.

Schwartz MK, Luikart G, Waples RS (2007) Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology & Evolution*, **22**, 25–33.

Shaffer HB, Fisher RN, Davidson C (1998) The role of natural history collections in documenting species declines. *Trends in Ecology & Evolution*, **13**, 27–30.

Shimoda N, Knapik EW, Ziniti J *et al.* (1999) Zebrafish genetic map with 2000 microsatellite markers. *Genomics*, **58**, 219–232.

Soulé ME (1980) Thresholds for survival: maintaining fitness and evolutionary potential. In: *Conservation Biology: An Evolutionary-Ecological Perspective* (eds Soulé ME and Wilcox BA), pp. 151–169. Sinauer Associates, Sunderland,, Massachusetts.

Spear S, Peterson C, Matocq M, Storfer A (2006) Molecular evidence for historical and recent population size reductions

of tiger salamanders (*Ambystoma tigrinum*) in Yellowstone National Park. *Conservation Genetics*, **7**, 605–611.

Spong G, Hellborg L (2002) A near-extinction event in lynx: do microsatellite data tell the tale? Available at: http://www.consecol/vol6/iss1/art6.

Steinberg EK, Lindner KR, Gallea J, Maxwell A, Meng J, Allendorf FW (2002) Rates and patterns of microsatellite mutations in pink salmon. *Molecular Biology and Evolution*, **19**, 1198–1202.

Storz JF, Beaumont MA (2002) Testing for genetic evidence of population expansion and contraction: an empirical analysis of microstatellite DNA using a heirarchical Bayesian model. *Evolution*, **56**, 154–166.

Tajima F (1989a) The effect of change in population size on DNA polymorphism. *Genetics*, **123**, 597–601.

Tajima F (1989b) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Tajima F (1996) The amount of DNA polymorphism maintained in a finite population when the neutral mutation rate varies among sites. *Genetics*, **143**, 1457–1465.

Taylor B, Dizon AE (1997) The need to estimate power to link genetics and demography for conservation. *Conservation Biology*, **10**, 661–664.

Tingley MW, Beissinger SR (2009) Detecting range shifts from historical species occurrences: new perspectives on old data. *Trends in Ecology & Evolution*, **24**, 625–633.

Traill LW, Brook BW, Frankham RR, Bradshaw CJA (2010) Pragmatic population viability targets in a rapidly changing world. *Biological Conservation*, **143**, 28–34.

USFWS (2010a) *Draft Revised Recovery Plan for the Northern Spotted Owl*. Strix occidentalis caurina. U.S. Fish and Wildlife Service, Portland, Oregon.

USFWS (2010b) *Mexican Wolf Conservation Assessment*.US Fish and Wildlife Service Region 2, Albuquerque, New Mexico.

Wandeler P, Hoeck PEA, Keller LF (2007) Back to the future: museum specimens in population genetics. *Trends in Ecology & Evolution*, **22**, 634–642.

Wegmann D, Leuenberger C, Excoffier L (2009) Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. *Genetics*, **182**, 1207–1218.

Williams BK, Nichols JD, Conroy MJ (2002) *Analysis and Management of Animal Populations*. Academic Press, San Diego, California.

Williamson-Natesan EG (2005) Comparison of methods for detecting bottlenecks from microsatellite loci. *Conservation Genetics*, **6**, 551–562.

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

M.Z.P. is a conservation biologist that investigates the demography and genetics of endangered species. R.K. studies the molecular and trophic ecology of mammals and birds. B.N.R. studies comparative population dynamics and landscape genetics in freshwater turtles. R.S. is an applied ecologist interested in the conservation of reptiles, amphibians, and birds. S.J.R is a wildlife ecologist using molecular methods to investigate gene flow and disease dynamics in wildlife populations. C.V.-C. uses molecular methods to support the conservation of Neotropical mammals threatened by illegal traffic and exploitation. J.N.P. is a mammalian ecologist interested in a range of applied and basic questions in mammalian population and trophic ecology. P.J.P. is interested in the application of genetic data to conservation, ecology and evolution.

Supporting information

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

Appendix S1. References for 105 studies reviewed to assess the application of genetic bottleneck tests in the conservation literature.

Fig. S1 Mismatch in expected number of alleles in bottlenecked and equilibrium populations.

Table S1 Statistical significance of bottleneck tests in vertebrate populations known to have experienced recent reductions in abundance.

Table S2 Statistical significance of bottleneck tests in putatively stable vertebrate populations.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.