The use and abuse of microsatellite DNA markers in conservation biology

Authors: Moss, Robert, Piertney, Stuart B., and Palmer, Stephen C.F.

Source: Wildlife Biology, 9(4) : 243-250

Published By: Nordic Board for Wildlife Research

URL: https://doi.org/10.2981/wlb.2003.011
The use and abuse of microsatellite DNA markers in conservation biology

Robert Moss, Stuart B. Piertney & Stephen C.F. Palmer

Conservation genetics is concerned with the genetic factors that affect extinction risk and the management practices needed to minimise such risk, thereby maintaining populations or species as dynamic entities that can survive environmental change (Frankham, Ballou & Briscoe 2002). The primary causes of extinction risk are often anthropogenic, typically involving remnant populations in fragmented or degraded habitats (Lande 1988, 1999). The contribution of genetic factors to the fate of endangered populations has often, therefore, been considered secondary. Even so, the genetic changes associated with population isolation, fragmentation and concomitant reduction in effective population size are intimately linked with population viability. Two key processes have been highlighted. First, small populations are prone to inbreeding as individuals become related by descent over time. This reduces individual fitness through inbreeding depression, itself brought about by increased homozygosity, the unmasking of deleterious recessive alleles and reduced genetic variation (Crnokrak & Roff 1999). Second, small, isolated populations have reduced levels of genetic variation, which compromises their ability to adapt and so survive environmental change. This association, between reduced ge-
Genetic variation, inbreeding and inability to adapt, means that a primary objective in genetic management of populations and species is maintenance of genetic variation (Avise & Hamrick 1996).

Genetic variation within a population is naturally maintained by new genetic variants, alleles, that arise from mutation or immigration. Conversely, genetic variation is lost from a population by natural selection against alleles with lower reproductive fitness in particular environments, or by genetic drift. Low genetic variation can also occur when a biased set of founders forms a new population.

Measuring genetic variation in natural populations, and so identifying populations at risk of extinction, is problematic. A current trend is to characterise genetic variation within populations and species in terms of the number of alleles and their respective frequencies and heterozygosities at individual loci, frequently using molecular markers such as microsatellite DNA loci (Frankham et al. 2002). Such surveys allow comparison between fragmented and continuous, or small and large, populations. Also, because allele and genotype frequencies attain Hardy-Weinberg equilibrium after a single generation of random breeding in a large population, any significant deviation from Hardy-Weinberg predictions should reflect processes such as inbreeding.

A major criticism of this approach, however, lies in the implicit assumption that a population depauperate in microsatellite variation shows a proportionate reduction in the genetic variation associated with the quantitative traits that underpin reproductive fitness and adaptive potential (Reed & Frankham 2001). Certainly, mean heterozygosity should be proportional to the variance for a polygenic trait, if all gene action is additive (Falconer 1989). However, quantitative traits associated with fitness vary continuously due to the contributions of many loci and to genotype-environment interactions. Moreover, pleiotropy, epistasis, dominance, differential selection, different mutation rates and regulatory variation further complicate the structure of fitness-related traits (Reed & Frankham 2001).

Here we use simple models to examine the dynamics of genetic variation under particular regimes of random genetic drift, selection and mutation. They illustrate the dangers of using microsatellites as a surrogate for adaptive genetic variation. Our models are more ecologically meaningful than classical population genetic models.

The usual textbook (e.g. Hartl 1980) model for micro-evolutionary processes such as genetic drift, gene flow and selection concentrates on a single locus with two alleles, A and a, with population frequencies p and q, respectively. It focuses on p, estimating q by difference \( q = 1 - p \). There are two alleles per locus and so the total number of gene copies is 2N, where N is the number of organisms in the constant population. The model ignores sex and assumes independent segregation of alleles. To mimic drift, a binomial distribution (number 2N, probability p, in generation t) is used to generate a random walk in the number of allele A. Hence genetic drift is due to environmental, not demographic, stochasticity. Population size affects drift because there are fewer possible values of A in smaller populations, and so drift is faster.

The textbook model contains no explicit reproduction or death. Nonetheless, it seems to represent demography like that of an annual plant, each generation reproducing once and then dying. It seems inappropriate for many natural systems, including grouse populations. Therefore, our models explicitly involve the effects of demographic stochasticity on survival and reproduction.

**Methods and models**

**Demographic model with explicit survival and recruitment**

We begin with a simple demographic model of grouse numbers, starting with a population of 100 grouse in autumn, 50 of which survive to reproduce next spring. To maintain a constant population, each surviving adult must rear on average one recruit per year (this differs from the number of young reared, for it includes only recruits to the breeding population). Over-winter survival is enacted by applying to each individual in the autumn population a binomial survival probability of 0.5. Each survivor then produces recruits to the next generation, the number being taken at random from a Poisson distribution with mean one. The total number of grouse fluctuates due to demographic stochasticity, and it might become extinct or expand indefinitely.

**The basic model**

We apply the same model to the alleles at a single locus, making extra assumptions. Given a fixed number of 100 autumn grouse, there are 200 copies of alleles A or a. Focusing on allele A, we start the model with 100 copies of A. If the number of A falls to zero (extinction), the number of a must be 200 (fixation) and vice versa. Hence one unit of genetic heterogeneity is lost when A numbers zero or 200. The model parameters (i.e. survival probability, recruitment probability and population size) can, of course, be varied.
Population size and the harmonic mean
We are interested in how population size affects drift and selection, but have two population sizes, autumn and spring. In a fluctuating population, the mean number of generations taken by a neutral allele to drift to fixation/extinction should be proportional to the harmonic mean of the population size (Wright 1938, Gillespie 1998). We therefore use the harmonic mean of population size in autumn (2N) and spring (2N x over-winter survival) to represent the size of constant populations.

In reality, population size varies from year to year. To model this, we calculated the frequency \( p_t \) of A after reproduction in year \( t \), the population size in year \( t+1 \), and reset the number of A = \( p_t (2N_{t+1}) \). The harmonic mean of all autumn and spring population sizes throughout each simulation represented the size of fluctuating populations.

Loss of variation
The average rate of loss of genetic or microsatellite variation due to drift was simulated by starting the basic model with a frequency for A of 0.5. This was done repeatedly, the model restarting after each fixation/extinction. We measured the mean number of years to fixation or extinction and its inverse, the number of fixations/extinctions per year. Simulations continued until the standard error of the mean number of years to fixation/extinction was less than 10% of the mean, or until 40,000 years had been simulated.

Selection, heterozygosity and protected polymorphisms
Selection for or against A could be enacted by varying the average survival or recruitment rate of A. The above model does not incorporate heterozygosity and so, for a diploid organism like grouse, it implies that A dominates a. To represent recessive A alleles we incorporated heterozygosity into the model, defining the frequency of homozygous A alleles in year \( t \) as \( p_t^2 \) and the frequency of heterozygotic A alleles as \( 2p_tq_t \) [i.e. \( 2p_t(1-p_t) \)]. This approximated a population that remained at Hardy-Weinberg equilibrium.

Drift or unidirectional selection implies that A or a will eventually become fixed, with loss of genetic variation. This can be preserved if neither A nor a becomes fixed because the heterozygotic genotype is protected through frequency-dependent selection. Several such mechanisms are known. We illustrate the principle with models that give the heterozygote a selective advantage over the two homozygotes.

Mutation and immigration
A starting frequency of 0.5 for each run is conventional when simulating fixation/extinction of alleles already in a population. Mutations, however, are likely to occur singly. We therefore assumed that a mutation affecting reproduction occurred in a single individual in a population previously homozygous for allele a. Hence the starting number of allele A was one. The same model can be used for immigration of a single heterozygous individual into a homozygous population.

Results
The results from several iterations of the same model often differed widely due to random drift. Each quoted result was the average of many iterations for the modelled locus. For heuristic purposes, we assume that the modelled locus is typical of an idealised genome. Many of the results are expressed as rates, for example the mean number of loci at which A became fixed/extinct each year. These rates can also be regarded as probabilities, for example the probability that a locus with the modelled characteristics will become fixed/extinct each year.

Demographic extinction versus loss of genetic variation
In our randomly fluctuating demographic model of N grouse, the probability that a population will become extinct before it doubles in number \( P_{\text{dx}(N)} \) is equal to the probability that it will double in number. Having doubled in number, the probability of demographic extinction before a further doubling in number is \( P_{\text{dx}(2N)} \), and so on. Hence, the total probability of extinction of a population of N individual grouse is

\[
P_{\text{dx}(N)} + P_{\text{dx}(2N)} + P_{\text{dx}(4N)} + \ldots
\]

where

\[
P_{\text{dx}(N)} > P_{\text{dx}(2N)} > P_{\text{dx}(4N)} > \ldots
\]

In the genetic version of the same model, there are 2N alleles and the rate at which loci with two alleles become fixed is equal to \( 2P_{\text{dx}(2N)} \). At very low population sizes \( P_{\text{dx}(N)} \gg P_{\text{dx}(2N)} \), so that the probability of demographic extinction is greater than the rate of loss of genetic or microsatellite heterozygosity (Fig. 1).

Drift in models with different vital rates
In our drift model, recruitment of allele A averages just enough to compensate for its mortality, the converse of survival. Survival and recruitment, however, can
differ, giving different rates of turnover, the proportion of alleles newly recruited each year. We simulated drift in populations with three different sets of vital rates (see Fig. 1A) at various population sizes (2N). In each case, the mean number of years to fixation/extinction was directly proportional to the harmonic mean of population size. The slopes of the regression lines represented the rate at which the number of years to fixation/extinction increased with population size. The mean time to fixation/extinction was shorter in populations with faster turnover rates and the difference grew as population size increased.

A number plotted against its inverse describes a hyperbola, so the relationship between the rate of loss of heterozygosity and population size was hyperbolic (see Fig. 1B) and accelerated with decreasing population size, especially when populations were small.

For the rest of the paper we use a model with a survival rate of 0.5 per individual and a recruitment rate of about 1.0 per survivor, varying this somewhat to mimic different selection pressures.

**Protected polymorphisms**

The mean time that a deleterious recessive allele persisted in model populations depended upon the strength of any frequency-dependent selection. When the heterozygote had a very small selective advantage over both homozygotes, its mean time to fixation/extinction was indistinguishable from that due to drift alone (Fig. 2) until population size (2N) reached 6,000-7,000. As the selective advantage of the heterozygote increased, so the mean time to fixation/extinction became longer than expected from drift alone. At population sizes (2N) of a few hundred, the mean time to fixation for the two models with the biggest heterozygote advantages was so long that no fixations occurred in 40,000 simulated years.

In small model populations, however, drift overwhelmed the stabilising effects even of strong frequency-dependent selection, heterozygosity was lost and deleterious recessive alleles became fixed with increasing frequency.

---

Figure 1. Loss of genetic variation in relation to the harmonic mean population size for three different sets of vital rates, i.e. survival rates of 0.1 (△), 0.5 (×) and 0.9 (■); recruitment per survivor 9, 1 and 0.1111, respectively. A) shows the mean number of years that an allele took to become fixed or extinct, and B) shows the mean number of fixations/extinctions per locus per 100 years. Slopes differed significantly (ANCOVA: P < 0.05).

Figure 2. Effect of frequency-dependent selection upon the mean number of years to fixation/extinction, in relation to the harmonic mean population size. With drift alone (×; survival rate = 0.5 and recruitment rate = 1.0), the mean number of years that an allele took to become fixed/extinct was linearly related to population size. The four examples of frequency-dependent selection vary from strong to weak (recruitment rate for homozygotes of 0.73 (△), 0.91 (■), 0.991 (▲), 0.9991 (○), for heterozygotes 1.3 and 1.1, 1.01, 1.001, respectively, survival rate 0.5 throughout). In each example the effects of selection differed significantly (ANOVA: P < 0.05, at each population size) from those of drift alone, at or below the maximum population size shown on the x-axis, except that for ▲ a significant difference occurred at a population size of 6,667 but not at 4,000 or less. In each case 40,000 years were simulated.
Fluctuating populations and the harmonic mean

In Figure 3, the solid symbols represent constant populations with no year-to-year variation in size, and the open symbols denote populations whose size fluctuated through a limit cycle with a period of eight years and a nine-fold amplitude (peak numbers/trough numbers at the same season). Results from otherwise similar constant and cyclic populations showed much the same relationship with population size, provided that this was expressed as the harmonic mean, as expected from standard theory (Wright 1938, Gillespie 1998). The harmonic mean is biased towards smaller populations such that, in these examples, the harmonic mean size of each cycling population was 0.55 of its arithmetic mean.

Mutation and immigration

Selection tends to eliminate deleterious alleles, and yet they remain common in many populations. We illustrate two ways in which this can occur. First, deleterious mutations, recessive or dominant, can become fixed by drift at small population sizes (Fig. 4). Now homozygous, such alleles will remain in the population, even if it increases in size, until they are ousted by a beneficial mutation or immigrant allele.

Second, frequency-dependent selection allows deleterious recessive alleles to become established even at large population sizes (Fig. 5). 'Established' in this context does not mean 'fixed', but that the mutation is sufficiently frequent to be protected within a polymorphism.

We represent immigration of a single heterozygous individual by the same model as mutation. A notable result is that, as with mutations, most immigrant genes were eliminated from the population by drift, even if they were somewhat advantageous. Naturally, the proportion of mu-
tants/immigrants that became established increased as their selective advantage increased (results not shown).

**Discussion**

**Neutral versus adaptive variation**

The models show how neutral molecular markers and adaptive genetic variants can respond differently to selection and drift. Hence molecular markers, such as microsatellite DNA polymorphisms, should not be used uncritically as surrogates for adaptive genetic variation when identifying populations as genetically depauperate, or at increased risk of extinction.

Such caution is mirrored in recent reviews comparing various measures of genetic variation from empirical studies. Butlin & Treganza (1998) found no significant correlation between molecular marker heterozygosity and the coefficient of variation for additive genetic variation of sexually selected traits among 20 different species. Similarly, Reed & Frankham (2001) carried out a meta-analysis on 71 data sets and found only weak correlation ($r = 0.217$) between molecular and quantitative measures of genetic variation, and no correlation at all when analysis was restricted to traits considered to be the best indicators of adaptive potential.

In our models, differences between neutral and adaptive genetic variation depend on the relative strengths of selection versus drift. In populations (2N) of a few hundred or less, weak selection is overwhelmed by drift, and so the expected amount of variation for a polygenic trait, determined by many weakly-selected alleles, is largely a result of mutation-drift balance (Foley 1992), provided that the constituent genes act additively. Under these specific conditions, microsatellite variation might on average reflect genetic variation. The random nature of drift, however, ensures that this will not apply to all small populations, nor necessarily to any particular study population. Moreover, characters influenced by many loci should have a bigger mutational input so that, in small populations, polygenic traits might retain more variation and recover variability more frequently following a population bottleneck than marker surrogates (Lynch 1996).

Weakly-selected genes show similar dynamics to microsatellite markers even in model populations of relatively large sizes. In reality, however, the best-known weakly-selected genes are largely loci, such as allozymes, that may influence individual fitness, but are unlikely to contribute to the polygenic traits thought to underpin evolutionarily important processes such as adaptation.

**Effective population size**

Our results are for a population size of 2N alleles; in principle this represents a population of N grouse. But real grouse populations fluctuate widely in numbers, and past numbers are likely to have influenced present genetic variation. An appropriate measure of population size is the harmonic mean, which is biased towards lower values. Also, in studies of real birds, about 20% of the breeding population produced about 50% of the offspring (Newton 1989). For this and other reasons, natural populations are likely to comprise individuals that share genes by descent, and the effective population size is likely to be smaller than the population size observed in the field. Such considerations indicate that the number of real grouse equivalent to N model grouse might be 5-10N, which implies that the population size of 2N used in our Figures should be multiplied by about 3-4 when considering real animals.

**Minimum viable population size**

The model indicates some genetic aspects governing minimum viable population size (MVP). There are at least two views of this problem. First, the hyperbolic relationship between the rate of loss of genetic variation and population size resembles a threshold effect (Frankham 1995). Broadly, the loss of adaptive genetic variation, and the fixation of deleterious mutations, became major below a model population size of about 50 (2N), equivalent to about 125-250 real grouse. The sole documented example of inbreeding depression in a wild grouse population occurred when an isolated population of greater prairie chickens *Tympanachus cupido pinnatus* had fallen to about 100-200 birds (Westemeier, Brawn, Simpson, Esker, Jansen, Walk, Kershner, Bouzat & Paige 1998). Also, when exotic bird species were introduced to New Zealand, the chances of successful establishment were much higher when more than 100 individuals were introduced (Green 1997). This apparent agreement between model and reality could well be coincidence.

An important caveat is that population size alone is unlikely to be a good predictor of heterozygosity. For example, different levels of heterozygosity are likely in two small populations, one taken from a large pool of individuals and another, of equal size, kept isolated for a long period. Also, the latter population is likely to contain more deleterious mutations fixed through drift.

A second approach to genetic MVP considers the number of animals necessary to maintain adaptive genetic variation sufficient to allow the population to respond to future environmental changes. Thus the most weakly protected model polymorphism (see legend in Fig.

2) was overwhelmed by drift even in a population (2N) of 4,000. In populations up to this size, the genetic variation for a model polygenic trait, comprising the additive effects of many such alleles, would depend solely upon the balance between mutation, immigration and drift. Only in population sizes (2N) of \( \geq 6,000-7,000 \) would such a trait benefit from frequency-dependent protection.

Calculations (which we did not attempt) use various criteria and conclude that the population size required to maintain adaptive genetic variation (Lynch 1996), and also to ensure against demographic catastrophes (Ewens, Brockwell, Gani & Resnick 1987, Ewens 1990), is several thousand individuals. Hence, even if an inbred wild population of 100-200 grouse can be rescued by new blood in the short term (Westemeier et al. 1998), the current best estimate is that a population of thousands will probably be necessary for long-term survival in a changing environment.

**Microsatellite utility**

Despite concerns about neutral molecular markers as indicators of adaptive genetic variation in natural populations, microsatellites are finding pervasive use in other areas of conservation biology. Frankham et al. (2002) highlight 11 major genetic issues in conservation biology, of which most are not concerned with deleterious effects of reduced variability in natural populations. Microsatellites have proven invaluable in determining levels of population fragmentation and associated levels of gene flow between populations, resolving taxonomic uncertainty and defining management units, forensic analyses and molecular analyses to unravel aspects of species biology. Most promisingly, they can be used to estimate genetic relatedness, which facilitates quantitative genetic analysis in natural populations for which pedigrees are unknown. Such approaches represent perhaps the only way to understand the heritability of quantitative trait loci associated with complex life history traits in real species (Merilä & Sheldon 2000).

**Conclusion**

Genetic variation at molecular markers such as microsatellite DNA polymorphisms is presumably governed by random genetic drift, and so may not reflect variation in polygenic traits that underpin evolutionary potential. Hence, whilst microsatellite data can be used to estimate the extent of inbreeding within natural populations, it is not necessarily straightforward to extrapolate and predict the levels of inbreeding depression and associated reductions in fitness. Nor should microsatellite variation be used as an ersatz measure of the genetic variation that underpins adaptive and evolutionary potential in a changing environment. From limited experience in field (Westemeier et al. 1998) and aviary (R. Moss, unpubl. data), a practical indication of inbreeding depression in small, isolated grouse populations is likely to be a decline in the egg hatching rate. Chick viability might also decline, but may be less useful, being more difficult to study in the field and more affected by extraneous factors.

**Acknowledgements** - we thank Kent E. Holsinger for invaluable advice and many colleagues for useful discussions.

**References**


