GENETIC ASSIGNMENT TESTS AS INDICES OF INTERPOPULATION DISPERSAL

Peter M. Waser\textsuperscript{1}, Dept of Biological Sciences, Purdue University, W.Lafayette, IN 47907 USA
Andrew F. Waser, 6317 Stair Rd, Lafayette, IN 47905 USA
Curtis Strobeck, Dept of Biological Sciences, University of Alberta, Edmonton, Alberta T6G 2E9 Canada

1. email: pwaser@bilbo.bio.purdue.edu
**ABSTRACT:** We use Monte Carlo simulations to assess whether interpopulation dispersal rate might be inferred using a recently-described genetic "assignment test". The method assigns individuals to putative source populations based on the expected frequencies of their genotypes in those populations. Individuals whose genotypes are more likely in populations other than the one they are found in are said to be "misassigned". Some such individuals are immigrants, "misassigned" to the population that they emigrated from. We simulate populations linked by dispersal and explore the relationship of the proportion of individuals that are misassigned ($P_{ma}$) to the per capita probability of dispersal between populations ($m$). We explore the robustness of this relationship to a series of genetic assumptions (style and rate of mutation) and demographic assumptions (population size, number of populations, proportion of population sampled, annual mortality rate, sex ratio of dispersers, mating system). The relationship between $P_{ma}$ and dispersal rate is influenced by the number of loci available to the investigator and by the assumed mutation rate, but relatively robust to variation in all other assumptions. The assignment test is unlikely to provide a useful index of dispersal rate when that rate is high, but when populations are discrete and dispersal between them is uncommon (a situation that commonly faces ecologists), dispersal rate is indexed by $P_{ma}$ in an approximately linear fashion. The index
responds quickly to changes in interpopulation dispersal rate. As a result, this approach shows promise as a means of estimating contemporary dispersal rates, relatively uninfluenced by historical patterns of gene flow.

Key words: dispersal, migration, gene flow, metapopulations, microsatellites, assignment test

INTRODUCTION

Dispersal between populations can rescue declining populations, sustain metapopulations in the face of local extinction, and maintain genetic variation within and between populations. Models of population size and persistence, as well as models of genetic population structure, are sensitive to assumptions about rates of dispersal or gene flow.

Nevertheless, “direct” measures of gene flow (Slatkin 1985, 1994), obtained by following the movements of marked individuals throughout their lives, have been notoriously difficult to obtain. Rates of interpopulation movement are generally too low and the logistic difficulties in marking adequate numbers of young animals are too severe to obtain reliable dispersal measurements (Doak et al. 1992, Hanski and Kuussaari 1995, Lima and Zollner 1996). Indeed, most
data on dispersal rates and distances are inherently censored because of the limited size of the study area (Barrowclough 1978, Baker et al. 1995, Koenig et al. 1996).

Rates of movement between populations have also been estimated by “indirect” means from the spatial distribution of alleles (Neigel 1997). Most often, this approach relies on some variant of Wright’s $F_{st}$ and its relationship to the expected number of migrants per generation $Nm$ in the island model, $F_{st} = 1/(1+4Nm)$. In practice, however, this approach has suffered both practical and conceptual limitations. Until recently, the overriding practical limitation has been a lack of allelic variation adequate to detect differentiation between populations, especially at a local scale. Prominent among the conceptual limitations has been the assumption inherent in Wright’s approach that the populations analyzed are at genetic equilibrium. Whether this assumption is close enough to validity that we can trust $F_{st}$ - based estimates of gene flow remains a matter of debate (Bossart and Prowell 1998, Bohonak et al. 1998, Davies et al. 1999, Bohonak 1999). In any case, many questions regarding gene flow, in particular those of relevance to conservation, are at base questions about current rates rather than historical ones, and $F_{st}$ - based estimates do not distinguish the two (Neigel 1997, Sork et al. 1999).
Recently, indirect studies of gene flow have begun to tap the huge reservoirs of information present in widely-distributed, highly variable loci such as microsatellites. Microsatellite-based studies of genetic structure have detected population differentiation on a considerably smaller scale than was possible with allozymes, sometimes over a few kilometers or less (Taylor et al. 1994, Dallas et al. 1995, Lade et al. 1996, Pope et al. 1996, Becher and Griffiths 1998, Balloux et al. 1998, Keyghobadi et al. 1999, Mossman 1999). Microsatellites are beginning to be used to investigate interpopulation movement, as indexed by genetic distance (Barker et al. 1997, Paetkau et al. 1995, 1997, 1998, Keyghobadi et al. 1999) or by estimating Nm from $F_{st}$ (Barker et al. 1997, Ishibashi et al. 1997, Sork et al. 1999, Paetkau et al. 1999). Where both have been used in the same populations, microsatellite-based analyses have been considerably more powerful in detecting differentiation than those based on allozymes (Scribner et al. 1994, Barker et al. 1997, Estoup et al. 1998, Paetkau et al. 1999).

Some microsatellite-based studies also take advantage of new statistical approaches to estimating dispersal (Rousset and Raymond 1997, Estoup et al. 1998, Luikart and England 1999, Rousset in press). Among the most powerful of these are likely to be approaches that, unlike those related to
Wright’s $F_{st}$, are based on individual genotypes (Bowcock et al. 1994, Favre et al. 1997, Rannala and Mountain 1997, Shriver et al. 1997, Smouse and Chevillon 1998, Davies et al. 1999, Roques et al. 1999). Several such approaches can trace their ancestry to the idea of genetic “assignment”, in which individual genotypes are assigned to populations where their expected frequency is greatest. The application of this approach to interpopulation dispersal was first recognized by Paetkau et al. (1995).

Paetkau et al. (1995) pointed out that when populations are linked by dispersal, dispersers will often be “misassigned”; the assignment test will (correctly) assign them to the population they were born in, not the population they were sampled in. By extension, the proportion of individuals in a population that are misassigned to another might index the rate of dispersal between those two populations. Of course, some misassigned individuals might be the offspring or grandoffspring of immigrants, and if allele frequencies in the populations are very similar, some individuals will be misassigned by chance. Paetkau et al. found that the proportions of genotypes misassigned among a set of polar bear $Ursus maritimus$ populations reflected geographic proximity and suggested that they might also reflect movement patterns “not obvious from geography” (see also Paetkau et al. 1998, 1999, Waser and Strobeck
1998, Waser et al. in press). Several more recent papers have commented on the potential power (and some of the limitations) of this approach (Davies et al. 1999, Hedrick 1999, Luikart and England 1999, Roques et al. 1999). However, the behavior of this index as an estimator of interpopulation dispersal rate has not been investigated in a systematic way.

In this paper, we use Monte Carlo simulations to explore the expected relationship between interpopulation dispersal rate and the proportion of individuals whose genotypes are misassigned by the assignment test, $P_{\text{MA}}$. Using a simulation of two or more discrete populations linked by dispersal, we also explore the ways in which this relationship is influenced by assumptions about mode and rate of microsatellite mutation, by the proportion of the population that is sampled by the investigator and the number of microsatellite loci available for analysis, and by various aspects of the demography and mating system of the populations that exchange dispersers. We first investigate the relationship of $P_{\text{MA}}$ to the per capita probability of dispersal $m$ under equilibrium conditions, and then explore the behavior of the assignment test under nonequilibrium conditions, immediately after a change in the interpopulation dispersal rate.
METHODS

Our simulations begin with discrete populations, each an array of \( N \) individuals characterized by up to 20 microsatellite loci. Each locus in the simulation has \( k \) possible alleles, and at the beginning of the simulation alleles are drawn randomly from a uniform distribution for all individuals (thus all alleles are approximately equiprobable, and the two populations are genetically similar). We then simulate one thousand generations, each consisting sequentially of mutation, dispersal, mating and death. During this period, many of the initial alleles are lost due to drift, and drift and mutation equilibrate. Following the dispersal step in generation 1000, we record individual genotypes and perform assignment tests. Except where otherwise noted, we assume that only two populations exchange dispersers, that population sizes are relatively small (\( N = 25 \)), and that a modest number of loci (8) are available to the investigator.

During the “mutation” step of the simulation, each gene copy at each locus mutates with a probability \( \mu \). In most simulation runs, mutation occurs according to a stepwise model; as is approximately the case for microsatellites, mutation
consists of a gain or a loss of one tandem repeat unit. There are upper and lower limits on allele size; the first allelic state can mutate only by gaining a repeat unit (with a probability of $\frac{1}{2} \mu$) and the $k^{th}$ allelic state can mutate only by losing one (Weber and Wong 1993, Nauta and Weissing 1996, Gaggiotti et al. 1999). Except where otherwise noted, we set $\mu = 0.001$ and $k = 12$. We also explore the effects of assuming “k-allele” mutation; in these runs, a mutation causes the number of tandem repeats to change randomly to one of the $k$ possible allelic states, each represented with a probability of $1/k$.

During the “dispersal” step of the simulation, each individual emigrates from each population with a probability $m$. We simulate $m$ values of 0, 0.0025, 0.005, 0.01, … 0.16 per generation. Our standard runs assume that males and females emigrate at equal rates. After all emigration has occurred, each disperser is allowed to immigrate into the other population. Immigrants move preferentially into slots left empty by emigrants, but if more immigrants exist than slots, each extra immigrant replaces a random, same-sexed resident. As a result, $m$ is the probability per generation that an animal not only emigrates, but also successfully immigrates into another population. In effect, we are assuming that there is a strict upper limit to population
size; when the number of emigrants from the two populations is not identical, the population with the larger number of emigrants is slightly reduced in size, but the population with more immigrants cannot grow larger than N. The simulation keeps track of each individual’s population of birth, so that immigrants (and their offspring and grandoffspring) can be identified if desired.

Mating follows dispersal; in other words we simulate natal dispersal and assume that no breeding dispersal occurs. During the “mating” step, one male and one female are drawn randomly from the population, and an offspring is formed by combining, locus by locus, one of the male’s two gene copies at that locus with one of the female’s. Loci are thus assumed to assort independently, and unless otherwise noted, mating is random with the constraint that sexes are separate. Mating is repeated (with replacement of both sexes) until the number of offspring equals N, at which point adult mortality takes place. Initially, we model nonoverlapping generations; and all adults die. To simulate overlapping generations, we allow each adult to die with a probability d, then choose randomly among the offspring to bring the population size back up to N. The population sex ratio is fixed at 1:1.
Our simulation produces lists of genotypes, each individual tagged with its sex and its "status" (resident, immigrant, offspring of immigrant, etc.), in the ASCII format used by program GENPOP (Raymond and Rousset 1995). We subsequently use these lists in two ways. First, we can randomly sample a proportion $s$ of the adults living in the population during that generation. For most of our simulations, $s = 1.0$, but decreasing the value of $s$ allows us to mimic the situation faced by investigators who cannot sample their populations exhaustively. Second, we perform an assignment test, as follows. (1) We remove the test individual’s genotype from the population it was sampled in and estimate allele frequencies in that population at each locus ($p_{ii}$, $p_{jj}$ ... for alleles $i$, $j$, ... at locus $l$). (2) We determine the genotype’s expected frequency in that population at each locus ($p_{ii}^2$ for homozygotes, $2p_{ii}p_{jj}$ for heterozygotes). (3) We multiply across loci and log-transform the product, producing the “assignment index” for the test genotype in the population it was found in. (4) We perform the same calculations to estimate the genotype’s frequency in the second population (its assignment index in that population). (5) Finally, we assign the genotype to the population in which it has the highest expected frequency.

In estimating allele frequencies, a problem arises when the test genotype contains an apparently unique allele: because
the test genotype is not included when calculating population allele frequencies, the expected frequency of that allele in all the putative source populations is zero. Various “zero-avoidance” procedures have been suggested (Waser and Strobeck 1998, Davies et al. 1999). In these simulations, we used the approach developed by Titterington et al. (1981): we add 1/a copies of every allele observed in either of the putative source populations into both populations. We then calculate the expected frequency of an allele as $p = (f + 1/a)/(n + 1)$, where $f$ is the number of copies of that allele observed in the population, $n$ is the number of gene copies for that locus in the population, and $a$ is the total number of alleles at that locus observed in either population. Further explanation of the assignment test calculation and a downloadable assignment test calculator can be found at [www.biology.ualberta.ca/jbrzusto/Doh.html](http://www.biology.ualberta.ca/jbrzusto/Doh.html).

To check the validity of our simulation, we determined that the heterozygosities and numbers of alleles reached after several hundred simulated generations approximated those expected at mutation-drift equilibrium; that numbers of dispersal events and mutations were normally distributed with means equal to the values of $m$ and $\mu$; that assignment index distributions of the two populations were the same; and that the mean values of the assignment indices of individuals in their own populations equaled their expected values under Hardy-Weinberg equilibrium, $2 \sum p_i \log p_i + (1 - \sum p_i^2) \log 2$. 
After investigating the relationship of $P_{\text{na}}$ to interpopulation dispersal rate under equilibrium conditions (after 1000 generations without any changes in simulation parameters), we modified the simulation to examine the proportion of animals misassigned between a pair of populations in which the interpopulation dispersal rate has recently changed. Our intent was to ask how useful $P_{\text{na}}$ might be as an index of current dispersal rate between populations, shortly after a change in the matrix of habitats separating them. For example, if two initially connected populations have recently been separated by a road, how quickly will it be possible to tell whether the road disrupts dispersal? If corridors are restored between separate habitat fragments, how quickly might we be able to assess the corridors’ effectiveness in linking initially isolated populations? To address such questions, we ran the simulation with one set of parameter values for 1000 generations, then modified the per capita dispersal rate. We calculated $P_{\text{na}}$ the generation before the change in $m$, the generation of the change, and 1, 2, 4, ... 64 generations afterwards.

All analyses are based on 100 replicate runs of the simulation unless otherwise noted (each run generated “data” from two populations). Analyses were performed using
DBASE and SYSTAT. The simulation program was written by AW in PowerBASIC 3.5 (Zale 1997) and is available at [www.bio.purdue.edu/researchlabs/waser/waser.html](http://www.bio.purdue.edu/researchlabs/waser/waser.html)

**RESULTS**

The general properties of assignment indices as a function of interpopulation dispersal rate are illustrated in figure 1. This scatterplot illustrates assignment indices for adults in a single generation sampled from two simulated populations, A and B, each with N = 100. Each point represents the expected frequencies of an individual’s genotype in the two populations, based on eight microsatellite loci. Points below the 45° line represent genotypes that are more likely in population A. Open circles denote individuals sampled from population A, solid circles are individuals sampled from population B. Solid circles that fall below the 45° line, or open circles that fall above it, represent individuals whose genotypes are more common in the population they were not sampled in, so that they are “misassigned” by the assignment test.

When the per capita dispersal rate m = 0, the allele frequencies in the two populations are free to diverge through drift and, not surprisingly, genotypes from the same populations tend to cluster in tight, discrete clusters
When $m$ is increased to 0.0025 – representing one migrant, on average, every four generations – genotypes from the two populations cluster much closer together (fig. 1b, note change in scale). Nevertheless, drift keeps the population clusters distinct. Replicate runs of the simulation (and even successive generations within a particular simulation run) produce slightly different clusters of points, but only a small proportion of replicates contain misassigned individuals.

With $m = 0.005$, misassigned genotypes become more common. In the simulation run illustrated (fig. 1c) there is one black point well below the 45° line. Because the simulation keeps track of individuals’ dispersal status, we know that this misassigned individual was born in population A and immigrated into population B, and in fact was the only disperser in either direction during this particular generation.

Three other genotypes in figure 1c are misassigned, but are very close to the 45° line; these belong to descendants of recent immigrants. Note that the offspring of an immigrant will have genotypes represented by a point intermediate
between those of its two parents. This suggests that its
genotype will usually be close to the 45° line, and that it
will itself be misassigned roughly half the time. By the
same reasoning, the grandchild of an immigrant will have a
genotype intermediate between those of its parents, and
still less likely to fall on the “wrong side” of the line.
Only a small fraction of grandoffspring of immigrants, and a
still smaller fraction of great-grandoffspring, will be
misassigned.

Not surprisingly, further increases in dispersal rate bring
the two populations’ clusters still closer together, and
misassigned individuals become more common (fig. 1d). It is
worth noting, however, that even when m = 0.01 (with N =
100, this equates to one disperser in each direction per
generation) the two populations’ clusters have by no means
fused, and the assignment test correctly identifies the
source of most individuals.

The examples illustrated in figure 1 show that the
proportion of misassigned animals $P_{\text{MA}}$ is greater when
populations are more tightly linked by dispersal, for two
reasons. First, $P_{\text{MA}}$ increases because more members of the
population are immigrants, and the assignment test correctly
“misassigns” them to the population they emigrated from,
rather than the population they were sampled in. But
second, $P_{ma}$ also increases because populations that are more tightly linked by dispersal are less differentiated. Animals that are recent descendents of immigrants are also more common at high dispersal rates, and these animals may be misassigned as well.

These trends are illustrated more quantitatively in figure 2, which plots the proportion of misassigned animals (+/- SE based on 100 simulation runs) as a function of dispersal rate when assignment tests are based on 8 microsatellite loci and population size is 25. As expected, the proportion of animals that are immigrants into the population they are sampled in and (correctly) misassigned to the other is linearly related to the dispersal rate. However, $P_{ma}$ increases with increasing $m$ in an approximately exponential fashion. At low dispersal rates, $P_{ma}$ increases approximately linearly with $m$, but at high dispersal rates, it saturates. This relationship is intuitively reasonable: when dispersal is common enough to homogenize the two populations, $P_{ma}$ would be expected to approach 0.5.

Figure 2 also shows that misassigned animals that are not immigrants are likely to be their recent descendents. For a given dispersal rate, the number of misassigned offspring of immigrants is approximately the same as the number of
misassigned immigrants (each immigrant has on average two offspring, and about half of these are misassigned).

The proportion of immigrants' grandoffspring that are misassigned is related to $m$ in a slightly more complex way (fig. 2). The relationship departs from linearity at high dispersal rates (because most grandoffspring of immigrants also have immigrant parents or are immigrants themselves). It also departs from linearity at very low dispersal rates (when populations are distinct enough, immigrants' grandoffspring are never misassigned). However, the departure from linearity at low $m$ is slight. At moderate dispersal rates the proportion of immigrants' grandoffspring that are misassigned increases almost linearly with $m$ (perhaps because each immigrant has on average 4 grandoffspring, and roughly a quarter of those grandoffspring have genotypes unusual enough to result in misassignment). Stated another way, at low to moderate rates of dispersal, most misassigned animals are immigrants or their recent descendents.

The shape of the relationship of $P_{MA}$ to $m$ suggests that, when more than 10–20% of a population’s residents emigrate per generation, the assignment test is unlikely to provide quantitative information on interpopulation dispersal. Nevertheless, when dispersal rates are sufficiently low,
there is a strong and approximately linear relationship between \( P_{\text{na}} \) and \( m \). The details of this relationship depend on the number of microsatellite loci that are available for analysis (fig. 3).

Figure 3 demonstrates that increasing the number of loci available for analysis markedly increases the range over which \( P_{\text{na}} \) indexes \( m \) in an approximately linear fashion. If only 4 loci are available, the assignment test provides little information unless dispersal rates are less than one in one hundred. On the other hand, with 16-20 loci available, \( P_{\text{na}} \) tracks dispersal rate linearly up to rates on the order of one in 20 animals. The slope of the relationship is flatter when the analysis is based on more loci. With only 4 loci (and if \( m \) can be assumed to be < 0.01), \( P_{\text{na}} \approx 20m \), while with 16 - 20 loci (and \( m < 0.05 \)), \( P_{\text{na}} \approx 5m \). With 8 loci available for analysis (and assuming \( \mu = 0.001 \)), the relationship is approximately linear up to a dispersal rate of about 1 animal in 50, and within this range \( P_{\text{na}} \) is approximately 10 times the dispersal rate.

**Influence of genetic assumptions:** The relationship between \( P_{\text{na}} \) and \( m \) is sensitive not only to the number of loci used for the analysis, but also to the mutation rate (fig. 4a). If one assumes a mutation rate of 0.0005 rather than 0.001,
$P_{MA}$ is approximately 13, rather than 10 times the dispersal rate. If one assumes a mutation rate of 0.0001, lower than most estimates published for microsatellites, $P_{MA}$ gives useful information about interpopulation dispersal only at very low dispersal rates.

In contrast, the relationship between $P_{MA}$ and $m$ is surprisingly insensitive to assumptions about the style of mutation or the number of allelic states available to mutate to. When the simulations are run with stepwise mutation, $P_{MA}$ has the same relationship to dispersal rate whether it is assumed that the number of possible alleles per locus is 12, 25, or 50 (fig. 4b). The relationship is virtually identical if the simulation is run with $k$-allele mutation (fig. 4c). Again, whether there are 12, 25 or 50 allelic states has little influence on the relationship between $P_{MA}$ and $m$.

**Influence of demographic assumptions:** Also surprisingly, the relationship between dispersal rate and $P_{MA}$ is virtually identical over a broad range of population sizes (fig. 5a). The proportion of animals misassigned is influenced by $m$, but not by $Nm$.

Similarly, in many field studies populations are incompletely sampled. Figure 5b shows that this problem as
little effect on the investigator’s ability to estimate dispersal rate from $P_{MA}$. When allele frequencies are based on 25 animals, it does not matter whether these animals represent 100% of a population of 25, or 25% of a population of 100.

In many cases, populations will exchange dispersers with several surrounding populations, rather than just one. We ran simulations in which dispersal occurred randomly among 2, 4 and 8 populations, but only two populations were sampled. When per capita emigration rates remain constant but dispersal occurs randomly among four populations rather than two, the rate of dispersal between any two of the populations is decreased. Moreover, allele frequencies may drift in different directions in the different populations. Both these factors might be expected to increase differentiation. Nevertheless, if we plot $P_{MA}$ between two particular populations against the probability that a member of either one transfers to the other, the relationship changes little when the target areas are also exchanging dispersers with other nearby populations (fig. 5c).

Dispersal is sex biased in most animals, and variations of the assignment test have been proposed as a means of detecting sex-biased dispersal (Favre et al. 1997, Mossman and Waser 1999). However, whether dispersers are members of
one or both sexes does not influence the ability of $P_{na}$ to estimate $m$ (fig. 6a).

Most vertebrates are iteroparous, and overlapping generations influence effective population size and thus might influence interpopulation differentiation. We ran simulations with annual adult death rates of 1.0 (semelparity), 0.75, 0.50 and 0.25; the relationship between $P_{na}$ and $m$ is robust to such changes (fig. 6b) (recall that $m$ is the probability of an individual’s dispersing per generation, not per year, and that in our simulations animals disperse at most once in their lifetimes, as juveniles).

Finally, we examined the sensitivity of assignment test results to assumptions about the mating system. In addition to random mating, we ran simulations with strict monogamy (each male mates with one and only one female, and vice versa, during a breeding season; each couple has the same probability of producing offspring) and extreme polygyny (one, randomly-chosen male in each population fathers all young during a breeding season; mothers are drawn randomly, with replacement, from the pool of all females until all young are produced). The proportion of misassigned animals tends to be slightly lower, and variance between runs slightly higher, when matings are polygynous. On the other
hand, the relationship is little influenced by the difference between monogamy and random mating, and overall the effects of mating system on the ability of $P_{MA}$ to index dispersal are small (fig. 6c).

**Response of $P_{MA}$ to changes in dispersal rate:** When interpopulation dispersal rate changes, the proportion of misassigned animals responds rapidly. Figure 7 shows the trajectories followed by $P_{MA}$ through time when the dispersal rate is changed from a range of initial rates to a new rate of 0.0025, 0.01, or 0.04. $P_{MA}$ responds immediately to a perturbation of $m$, especially when dispersal rate is increased. It rises or falls to the equilibrium value expected for the new dispersal rate within a few tens of generations.

Intuitively, these patterns can be understood in terms of the rapidly decaying “signature” of dispersal in the genotype of immigrants and their descendents. Increasing the per capita probability of dispersal increases the number of immigrating animals in the same generation. Most immigrants are misassigned, thus $P_{MA}$ increases immediately in response to an increase in $m$. By the same token, when $m$ is decreased the number of immigrants immediately decreases, decreasing $P_{MA}$. However, misassigned individuals also include descendents of immigrants from previous generations.
When $m$ is decreased, there is a lag before the number of descendents of immigrants decays to the new, lower levels; during this period, $P_{HA}$ remains inflated because some of these descendents are misassigned.

**DISCUSSION**

Our simulations indicate that there are both advantages and limitations to using the proportion of animals that are misassigned between populations as an index of interpopulation dispersal rate.

First, the approach depends on the investigator’s having access to large amounts of genetic information, that is, to substantial numbers of highly polymorphic loci. Our simulations suggest that four microsatellite loci would rarely be enough to provide useful information; eight will generally be a minimum, and sixteen will be better.

Second, the approach is likely to be informative only when the per capita rate of dispersal between populations is rather low. Intuitively this is no surprise: populations linked by high rates of dispersal should be genetically similar, so that genotypes would often be assigned to the wrong population by chance. Our simulations suggest that, for an investigator with access to 8 microsatellite loci,
the approach loses most of its power when more than one in 50 animals moves between populations in its lifetime. If the investigator has access to more microsatellite loci, \( P_{MA} \) will saturate at higher dispersal rates, but no matter how many loci are available, the approach will not be useful if per capita probabilities of dispersal are much above one in ten. On the positive side, genetic estimators of dispersal rate are only needed if interpopulation dispersal is too rare to detect by direct observation!

Third, the relationship between \( P_{MA} \) and the dispersal rate is sensitive to the mutation rate of the markers. Mutation rates in microsatellites are most commonly assumed to fall between 0.0001 and 0.001 and may sometimes be even higher (Weber and Wong 1993, Ellegren et al. 1995, Jarne and Lagoda 1996). Our simulations assumed \( \mu = 0.001 \); investigators who have reason to believe that their markers mutate at lower rates will find \( P_{MA} \) less informative and the quantitative relationship between \( P_{MA} \) and \( m \) will differ from that described here. We note however that our simulations with lower mutation rates (\( \mu = 0.0001 \)) led to equilibrium heterozygosity values substantially lower than those reported in most recent studies of microsatellites.
Fourth, we emphasize that our simulations assume that all populations are discrete and that all populations linked by dispersal are sampled. It is unclear how robust the approach described here will be if there are other, unsampled populations contributing immigrants to the sampled populations. On the positive side, estimating interpopulation dispersal rate is of particular conservation importance when populations have become confined to small, discrete areas of suitable habitat in a suboptimal matrix, essentially the circumstances that we simulated.

Fifth, we note that all our simulations assumed that populations or samples of genotypes used in the assignment test were equal in size and that neither has gone through recent bottlenecks. Others (e.g. Davies et al. 1999) have pointed out that the approach we use is biased when one sample is larger, or when one population is more variable. If only a few individuals have been sampled from one population, or if that population has few alleles because of its small size or a recent bottleneck, the few alleles that are represented in the sample will necessarily occur at high frequency. Thus, individuals from the other, larger or better-sampled population will be misassigned to the population with the smaller sample if they have those alleles. Empirical studies have confirmed that a disproportionate number of animals from larger populations
are misassigned to small ones and that the greater the
difference in sample size between populations, the greater
the asymmetry in the direction of misassignments. The
asymmetry vanishes if the genotypes from the larger sample
are randomly subsampled so that both populations are
represented by equal numbers of genotypes (Mossman, 1999).

Our simulations also assumed that dispersal among
populations is bidirectional, that is, that there are no
sources or sinks. When this assumption is correct, the
proportion of animals misassigned from population A to B is
equal on average to the proportion misassigned from B to A.
But where dispersal rates are not symmetrical, the
proportions of animals misassigned in the two directions
will differ. Future work should investigate the possibility
that asymmetries in $P_{HA}$ could provide a viable means of
identifying source and sink populations.

The “zero-avoidance procedure” best used in the assignment
test when the test genotype contains a unique allele remains
a matter of debate. Paetkau et al. (1995) initially added
the test individual’s genotype to all populations prior to
calculating expected genotype frequencies, but this
procedure is clearly biased towards assigning the test
individual to smaller populations. Paetkau et al. (1997,
1998) and Waser and Strobeck (1998) advocated the “leave
one out” procedure from discriminant analysis, in which the test genotype is removed from the population but any unique allele is added back to all populations at a low frequency. However, Rannala and Mountain (1997) have described limitations of the leave-one-out procedure. Therefore, our zero-avoidance procedure in these simulations was essentially that suggested by McLachlan (1992) for the independence model in discriminant analysis. It gives us the Bayesian estimate of the allele frequencies if the a priori distribution of allele frequencies is 1/a (a being the number of observed allelic states). Our procedure is similar, but not identical to that suggested by Rannala and Mountain (1997) which uses the Bayesian estimate of the genotype frequencies if the a priori distribution of allele frequencies is 1/a.

With these limitations in mind, our simulations suggest that the assignment test may indeed be a useful index of interpopulation dispersal, especially when dispersal is rare. The relationship between $P_{HA}$ and $m$ is robust to variation in a surprising number of parameters. The result is that the investigator may use this relationship even when working with a system for which these parameter values are unknown.
For example, we find that neither the style of mutation nor the number of possible allelic states have much influence on the \( P_{na} - m \) relationship. Microsatellite mutation mechanisms and constraints on the range of allele sizes have been the subject of considerable debate (Weber and Wong 1993, Garza et al. 1995, Gaggiotti et al. 1999). However, simulation results suggest that field investigators do not have to wait for molecular biologists to resolve debates over microsatellite mutation mechanisms before they use \( P_{na} \) to estimate \( m \). Similarly, to use the approach the investigator need not know the exact size of the populations or even the proportion of the population that is sampled. The relationship is little influenced by the life history characteristics or the mating system, or by the sex ratio of the dispersers. Whether there are two or many populations exchanging dispersers, the probability of movement between the particular pair of populations that has been sampled will be indexed by \( P_{na} \). For investigators with access to information from eight microsatellite loci, a useful rule of thumb is that \( P_{na} \) is approximately 10 times the per capita dispersal rate, as long as that rate is less than one in fifty.

Furthermore, the rapid response of \( P_{na} \) to changes in interpopulation dispersal rate suggest that it will often be valuable as a means of assessing the impact of habitat
changes. This is especially true if the change in habitat increases dispersal, because the additional immigrants will be detected in the same generation. The effect of a habitat change that blocks dispersal is not so immediately visible, because some descendants of previous immigrants will be misassigned, elevating $P_{na}$ until those animals have been “diluted” by time. Fig 7 suggests that the best way to explore the qualitative impact of a habitat manipulation on interpopulation dispersal is to compare $P_{na}$ immediately before and after the manipulation.

It is important to note that, while the proportion of misassigned animals is related to the interpopulation dispersal rate, this does not mean that misassigned individuals are always, or even usually, immigrants. This study has taken an empirical approach, asking simply whether the proportion of misassigned animals is related in a useful way to the number of dispersers. The answer is clearly yes, but some questions may be most usefully answered by attempting to identify the immigrants themselves. For example, the impact of habitat manipulation on dispersal rates would be most immediately detected through genetic identification of the immigrants alone; $P_{na}$ is less powerful because misassigned animals include not only immigrants, but also residents (mostly recent descendents of immigrants) misassigned “by chance”. How can one determine what
proportion of genetically misassigned animals are true immigrants? One approach would seem to be to use randomization tests to determine what proportion of two populations’ genotypes should be misassigned by chance. The proportion of immigrants would then be $P_{na}$ minus the proportion misassigned by chance (such a procedure is implemented on the web at [www.biology.ualberta.ca/jbrzusto/Doh.html](http://www.biology.ualberta.ca/jbrzusto/Doh.html)). An alternative procedure is to attempt to identify the immigrants directly (Rannala and Mountain 1997, Davies et al. 1999, Sork et al. 1999), an approach in which rapid advances can be expected.

The problems inherent in estimating demographic parameters like interpopulation dispersal rates have constrained many ecological studies, but the hypervariable genetic markers present in every individual contain enormous amounts of information related to lineage and origin. Our simulations suggest that attempts to infer aspects of demography from genetics will be powerfully aided by the development of individual-based statistical approaches that take advantage of this information.

ACKNOWLEDGMENTS
We thank Brad Swanson, Cathy Mossman, David Paetkau, and John Brzustowski for discussion and comments on earlier versions of this manuscript. The work was funded by the National Science Foundation (DEB 9616843).
LITERATURE CITED


FIGURE LEGENDS

Figure 1: Log expected frequencies ("assignment indices") of genotypes drawn from two populations (A and B) in those same two populations in a sample run of the simulation. For both populations, N = 100 and m (the per capita, per generation dispersal rate) = 0, 0.0025, 0.005, or 0.01. Other simulation parameters are at their standard values as described in the text. Each point represents a genotype; "open" genotypes (those sampled from population A) are misassigned if they fall above the 45° line, "solid" genotypes (those sampled from population B) are misassigned if they fall below the line. The proportion of misassigned animals increases as the dispersal rate increases, but even with m = 0.01 (Nm = 1) most genotypes are correctly assigned.

Figure 2: The proportion of animals that are misassigned as a function of the per capita dispersal rate. Points represent means and 95% confidence limits for 200 samples (100 simulation runs x two populations); standard parameter values are described in the text. The proportion of individuals that are immigrants and misassigned is, as expected, linearly related to the dispersal rate. On the other hand, the proportion of all animals (immigrants +
residents) that are misassigned saturates at high dispersal rates. This reflects the fact that not only most immigrants, but also some of their descendents have unusual genotypes and so are misassigned; at high interpopulation dispersal rates, most residents have relatively recent ancestors that were immigrants.

Figure 3: Proportion of animals that are misassigned ($P_{ma}$) as a function of interpopulation dispersal rate. The relationship is influenced by the number of microsatellite loci the investigator has available for analysis. Points represent means for 200 samples; except for number of loci, parameters are set at their standard values as described in the text. When fewer loci are available, $P_{ma}$ saturates at lower dispersal rates and so is less useful.

Figure 4: Sensitivity of simulation results to assumptions about microsatellite genetics. Top: the relationship between $P_{ma}$ and dispersal rate is influenced by the mutation rate that characterizes the microsatellites used in the analysis. This and subsequent figures illustrate means and 95% confidence limits for 200 samples with, except as noted, standard parameter values as described in the text. If the mutation rate is low (0.0001) the relationship saturates at low dispersal rate and so is less useful. Middle: the relationship between $P_{ma}$ and dispersal rate is uninfluenced
by the number of possible allelic states when bounded single-step mutation (ssm) is assumed. Bottom: the relationship between $P_{Na}$ and dispersal rate is again uninfluenced by the number of possible allelic states if we assume that microsatellites mutate randomly to one of $k$ allelic states ($kam$). Comparing the middle and bottom panels indicates that $P_{Na}$ indexes dispersal in a virtually identical way whichever style of mutation one assumes that microsatellites follow.

Figure 5: Sensitivity of simulation results to aspects of sample size. Top: The relationship of $P_{Na}$ to dispersal rate is remarkably uninfluenced by the sizes of the populations exchanging dispersers, at least up to an $N$ of 200. Middle: When the number of individuals sampled is held constant, the proportion of the population that they represent (100% of a population of 25, 50% of a population of 50, or 25% of a population of 100) has little effect on the relationship of $P_{Na}$ to $m$. Bottom: If the populations sampled are exchanging dispersers with other, unsampled populations (so that the total number of populations exchanging dispersers is 4 or 8) $P_{Na}$ still indexes the rate of dispersal between the sampled populations in nearly the same way as when the sampled populations are the only ones exchanging dispersers (total number of populations = 2).
Figure 6: Sensitivity of simulation results to various demographic parameters. Top: The $P_{m\alpha} - m$ relationship is not influenced by the sex ratio of the dispersers (the proportion of dispersers that are male). Middle: the relationship is not influenced by adult mortality rates. Bottom: the relationship is little influenced by the mating system, although whatever the dispersal rate, slightly more individuals tend to be misassigned if the mating system is highly polygynous.

Figure 7: Response of the proportion of animals misassigned to changes in dispersal rate. In all panels, generation “−1” shows the equilibrium $P_{m\alpha}$ when the populations have exchanged dispersers at the “initial” rate for 1000 generations. In generation 0, the dispersal rate is changed to the “final” rate (mimicking a change in the isolation of the populations, for example through the destruction or creation of corridors). Top: final dispersal rate is very low (0.0025). Middle: final dispersal rate is intermediate (0.01). Bottom: final dispersal rate is high (0.04, slightly above the range of dispersal rates within which $P_{m\alpha}$ is linearly related to $m$). Especially when dispersal rate is increased, $P_{m\alpha}$ changes immediately in the expected direction, but it takes several tens of generations for $P_{m\alpha}$ to reach its equilibrium value. Thus a comparison of $P_{m\alpha}$ values before and shortly after a habitat manipulation would
more clearly indicate the direction of a change in dispersal rate than its magnitude.
Figure 1a
Figure 1b

![Graph showing assignment tests and dispersal with index in population A and B, and the line m = 0.0025.](image-url)
Figure 1c
Figure 1d

The figure shows a scatter plot comparing the assignment index in population A against the assignment index in population B. The line $m = 0.01$ indicates a specific relationship or threshold between the two populations.
Figure 2

- proportion misassigned
- immigrants only
- immigrants and offspring
- immigrants, offspring and grandoffspring
- all animals

m (per capita dispersal rate)

proportion of animals misassigned
Figure 3

![Plot diagram with loci and Pma values]
Figure 4a

mutation rate

- ▼ 0.001
- △ 0.0005
- ● 0.0001
Figure 4b

number of alleles, ssm

- ▼ 50
- ▲ 25
- ● 12
Figure 4c

number of alleles, kam

- ▼ 50
- ▲ 25
- ● 12
Figure 5a

population size

- Diamond 200
- Square 100
- Triangle 50
- Triangle 25
- Circle 12
Figure 5b

proportion of population sampled

- ▼ 1
- ▲ 0.5
- ● 0.25

Pma vs m
Figure 5c

number of populations

\( P_{ma} \)

\( m \)
Figure 6a

proportion of male dispersers

- ▼ 1
- ▲ 0.75
- ● 0.5
Figure 6b

annual adult mortality rate

- 1
- ▼ 0.75
- ▲ 0.5
- ● 0.25
Figure 6c

mating system

- random
- polygyny
- monogamy
Figure 7a

initial m (final m = 0.0025)

- ▲ 0.04
- ▲ 0.01
- ● 0.0025

Pma

generations after change in m
Figure 7b

initial m (final m = 0.01)

- ▼ 0.04
- ▲ 0.01
- ● 0.0025

Pma

generations after change in m
Figure 7c

initial m (final m = 0.04)

- ▼ 0.04
- ▲ 0.01
- ● 0.0025