WHEN CAN GENETIC INFORMATION BE USED TO MEASURE INTER-POPULATION MOVEMENT?

By

JULIE M. BRENNAN, B.SC., M.SC.

A thesis submitted to
The Faculty of Graduate Studies and Research
In partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Department of Biology
Carleton University
Ottawa, Ontario

On

December 1, 2006

Your file Votre référence
ISBN: 978-0-494-27089-9
Our file Notre référence
ISBN: 978-0-494-27089-9

ABSTRACT

I asked which, if any, of F_{ST}, G_{ST}, Θ, R_{ST}, private alleles, and two versions of the genetic assignment test was suitable for calculating inter-population movement. In experimental simulations, I manipulated inter-population movement rate and calculated the difference (bias) between the amount of movement based on these genetic measures and actual inter-population movement for several demographic and genetic scenarios. F_{ST}-based estimates were reasonable in situations with at least 20% of the population sampled, at least 10 loci with at least 6 alleles each, a movement rate of 0.04 to 0.08 per generation, and at least 90 generations since isolation. Θ and G_{ST}'-based estimates were reasonable for the same parameters but limited to situations with non-overlapping generations and a movement rate less than 0.04. Estimates based on R_{ST}, private alleles and both genetic assignment tests did not reliably quantify movement.

In a partial test of the validity of the simulation model results, I compared the bias calculated from empirical studies to those predicted by the corresponding set of simulations. Seventy percent of test cases had an empirical bias within 1 standard deviation of the mean simulation bias. Estimates based on F_{ST} and Θ were more reliable than those based on the frequency of private alleles, R_{ST}, and the genetic assignment test. Θ-based estimates were more reliable when movement rate was less than 0.04. Estimates based on the frequency of private alleles underestimated movement. Bias decreased with percent of population sampled and number of alleles in some situations; bias was also lower if based on populations with non-overlapping generations. This suggests that the conclusions drawn from the simulation results require further study.

Overall, my results suggest that it is generally impractical to use genetic information to quantify inter-population movement reliably at this time. I did identify situations where estimates of movement were reliable. However, the reliability of estimates based on F_{ST} and Θ varied with movement rate itself, creating a situation where the movement rate must already be known to determine whether and which genetic methods can be used to estimate movement.

ACKNOWLEDGEMENTS

My experience in Lenore Fahrig's Landscape Ecology Lab here at Carleton has been a truly enjoyable academic and educational one. Lenore, you foster an atmosphere of collegiality that brings out the best in your students. You have been a patient guide, using your clear thinking and focus to help me become a better scientist. Following your favourite advice "just work it out" – which I did so many times over – has given me a tremendous sense of accomplishment in completing this project. Your support for all my undertakings has allowed me to become a more well-rounded academic. For all this, I thank you.

I also gratefully acknowledge the contribution of Lutz Tischendorf who expertly wrote the C++ code for the simulation model. Lutz, I do not think I can say thank you enough times for your generosity in giving me your time and skill; I guess we did our small part to "save the world." You were, and are, a good friend – always ready for a philosophical discussion, a relaxing ski, or an inspiring concert.

I would also like to thank the past and present members of Landscape Ecology Lab for teaching me about so many topics over the years (intellectual or otherwise), sharing much laughter, and making my time here at Carleton so memorable. In the early days we endured the windowless Tory basement with all its uniquely unpleasant aromas and sounds to end up in a beautiful state-of-the-art geomatics lab. Of the many wonderful people I've had the pleasure of meeting, there are several who I want to acknowledge specifically. Darren Bender, Dan Bert, Jeff Bowman, Naomi Cappuccino, Laurie Carr, Tom Contreras, Felix Eigenbrod, Sara Gagné, Brett Goodwin, Jeff Holland, Jochen

Jaeger, Kathy Lindsay, Rachelle McGregor, Shealagh Pope, Allyson Quinlan, Trina Rytwinski, Adam Smith, and Melissa Vance always had time to discuss ideas, read drafts, lend support, or be a friend. Thanks to Kringen Henein and Rebecca Tittler for being great teaching colleagues, friends, and a mutual support network. We could always count on Dave Omond for everything from the latest weather reports to organizing so many wonderful camping trips. Thank you to Dave Ladd for your wit and creativity and to Neil Charbonneau for your clever impersonations and awful puns. Michael Runtz, thank you for the great birding/wolf-finding nature walks with one of the best.

I would like to thank the members of my supervisory committee, Mark Forbes and Scott Findlay, for contributing their time and expertise to help improve this thesis. I am also grateful for the funding I received from Lenore Fahrig, as well as for scholarships from the Natural Sciences and Engineering Research Council of Canada (NSERC), the Fonds pour la Formation de Chercheurs et l'Aide à la Recherche (Fonds FCAR, Quebec), and Carleton University.

There are some unsung contributors to this project. Somewhere on the order of 150 billion virtual beings were born, grew up, back-packed through cyberspace to find themselves, reproduced, and died to provide me with data. To all those 0's and 1's: I owe you so much.

Finally, I am fortunate to have a loud personal cheering section for the endeavours I undertake. Virginia, you helped make Ottawa my home and were always there for me. My family, you are a constant source of encouragement and support. You are the best role models anyone could ask for. Eric, what a ride it has been! A whole lot

of "life" happened while I was writing this thesis. Through all of it, your support, friendship, and love have been constant companions on my intellectual journey. Thank you for inspiring me with all you are and for being an unwavering "biology groupie". The completion of this thesis is a lifelong dream come true; to share the success with you makes it priceless.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: GENERAL INTRODUCTION

Movement of organisms between populations is linked to the survival, and thus the conservation, of these populations (Brown & Kodric-Brown, 1977; Frankham, 1995). However, it is often challenging to estimate inter-population movement accurately. Evaluating and improving the accuracy of movement measures has been the focus of numerous theoretical, empirical, and review studies for decades.

Researchers have used both direct and genetically-based (indirect) methods to quantify movement (Slatkin 1985a). Direct measures estimate inter-population movement based on observing the movement of individuals between populations (e.g., based on capture-mark-recapture). Several indirect measures infer movement based on the genetic makeup of populations. Both direct and genetic measures have advantages and disadvantages (reviewed in Slatkin 1985a; Koenig et al. 1996; Bossart & Pashley Prowell, 1998a; Neigel 2002). However, with technological advances in molecular genetics (for reviews, see Mitton 1994; Parker et al. 1998; Manel et al. 2003), and continued refinements of the genetically-based measures of movement (see below), genetic estimates of movement have become much more prevalent (Figure 1.1). Furthermore, genetic information often can be measured at a lower cost and with less disturbance of organisms than marking and recapturing individuals.

Over the last several years, researchers have used a variety of methods to estimate movement indirectly from genetic data (Figure 1.1). Among these are several measures of genetic differentiation based initially on Wright's (1951) formulation of F_{ST}, which was subsequently extended to include a series of additional terms for multiple alleles

(Nei's F_{ST}, 1973), observed heterozygosity and sample size (Nei's G_{ST}, 1987), a correction for the number of populations (Nei's G_{ST}', 1987), and multiple loci and uneven sample sizes (Weir & Cockerham's Θ, 1984). Θ is currently the most commonly used estimate of genetic differentiation in the literature. Weir and Cockerham's paper had been cited 3,112 times as of April 2006, compared to 2,265 citations of Nei (1973), published 11 years earlier (*Web of Science* database). The original F_{ST} and its more recent estimates apply to allozyme loci. Slatkin (1995) proposed R_{ST} for microsatellite loci. Slatkin (1985b) also proposed a method based on the frequency of private alleles, i.e., those found in only one population.

In theory, genetic differentiation is inversely related to the amount of effective inter-population movement. Based on Wright's Island model (Wright 1931, 1951), the number of effective migrants per generation (Nm) can be estimated from genetic differentiation (F_{ST}) (Slatkin 1985a; 1987). However, in practice, this approach may be limited because the theoretical models are based on assumptions often either violated or not testable. The assumptions of the Island model are that there is genetic equilibrium between drift and gene flow, demographic equilibrium, non-overlapping generations, no mutation, no selection, equality of gene flow and migration, random migration with respect to alleles, and linkage equilibrium.

These limitations have created controversy in the literature about the validity of genetic methods in measuring movement, leading to cautionary statements about their use (Lewontin 1985; Slatkin 1985b; Crochet, 1996; Bohonak et al 1998; Bossart & Pashley Prowell, 1998a and b; Steinberg & Jordan, 1998; Whitlock & McCauley 1999;

Bohonak & Roderick 2001; Neigel 2002). These issues would be especially relevant in the situation where genetic differentiation is used to estimate movement between formerly panmictic populations that have been recently isolated, for example, by road construction or some other form of habitat fragmentation.

This debate about the effect of violations to the assumptions has led researchers to consider alternative methods that do not make these assumptions. In particular, individual-based assignment tests (Paetkau et al 1995; Rannala & Mountain 1997; Waser & Strobeck 1998; Pritchard et al. 2000; Paetkau et al. 2004) have been growing in use over the last decade with citations increasing steadily each year (*Web of Science* database). The assignment test developed by Paetkau et al. (1995) is the one most widely used in empirical studies (Guinand et al. 2002). Movement is calculated based on the number of individuals whose genotypes have a higher likelihood of belonging to other populations than the one in which they were found (Paetkau et al. 1995; Waser & Strobeck 1998). In a later version of the test (Paetkau et al. 2004), movement is calculated based on the individual's having a statistically significantly higher likelihood of belonging to another population than the one in which it is found. Because both versions of the test make fewer assumptions than previous methods (the tests only assume that the populations are at Hardy-Weinberg equilibrium and linkage equilibrium), they are expected to provide more reliable estimates of movement (Waser & Strobeck 1998).

Another issue that may limit the use of genetic methodologies in a conservation context is that it is sometimes difficult to obtain a large enough sample of individuals

and/or loci per individual to get a reliable estimate of genetic differentiation. Genetically-derived measures of movement based on small sample sizes may not reflect actual movement adequately even when all the assumptions are met (Ruzzante 1998; Whitlock & McCauley 1999; Kalinowski 2005).

The combined quantitative impact of these assumption violations, sampling issues, and the choice of methodology on the reliability of movement estimates based on genetic differentiation is not known. There may be instances when the relationship between the genetically based estimate of movement and actual movement is robust to deviations from some of the assumptions or when one method is more appropriate than another.

Identifying these appropriate situations or methods would provide practical information for an empirical researcher 1) to determine the potential bias in their estimate of movement given a particular parameter combination (e.g., sample size, number of loci, etc) and genetic methodology and 2) to decide whether genetic information can be used to estimate inter-population movement reliably in the researcher's particular context. Identifying these situations using only the current empirical literature is not possible. Many studies use genetically based indirect methods to measure movement, but very few compare their findings systematically to direct measures of movement on the same populations.

The purpose of this project was to determine under what circumstances genetic information can be used as a quantitative measure of inter-population movement. To do so I first developed a stochastic, individual-based simulation model in which both the

demography and the genetics of the populations were known. In experimental simulations I manipulated the movement rate between two hypothetical populations and examined the effect of percent of the population sampled, number of loci and alleles, population size, time since isolation, and whether generations overlapped or not on the relationship between the genetically-based estimate of movement and actual movement.

I began by examining the relative effects of these factors on the reliability of estimates of movement based on Θ, the most commonly used method for measuring movement (Chapter 2). I then expanded the study to determine if estimates based on any of the other methods (F_{ST}, G_{ST}, G_{ST}', R_{ST}, private alleles, and the genetic assignment test) were more reliable in quantifying inter-population movement (Chapter 3). Finally, I examined the concordance of the model results with those found in empirical studies (Chapter 4).

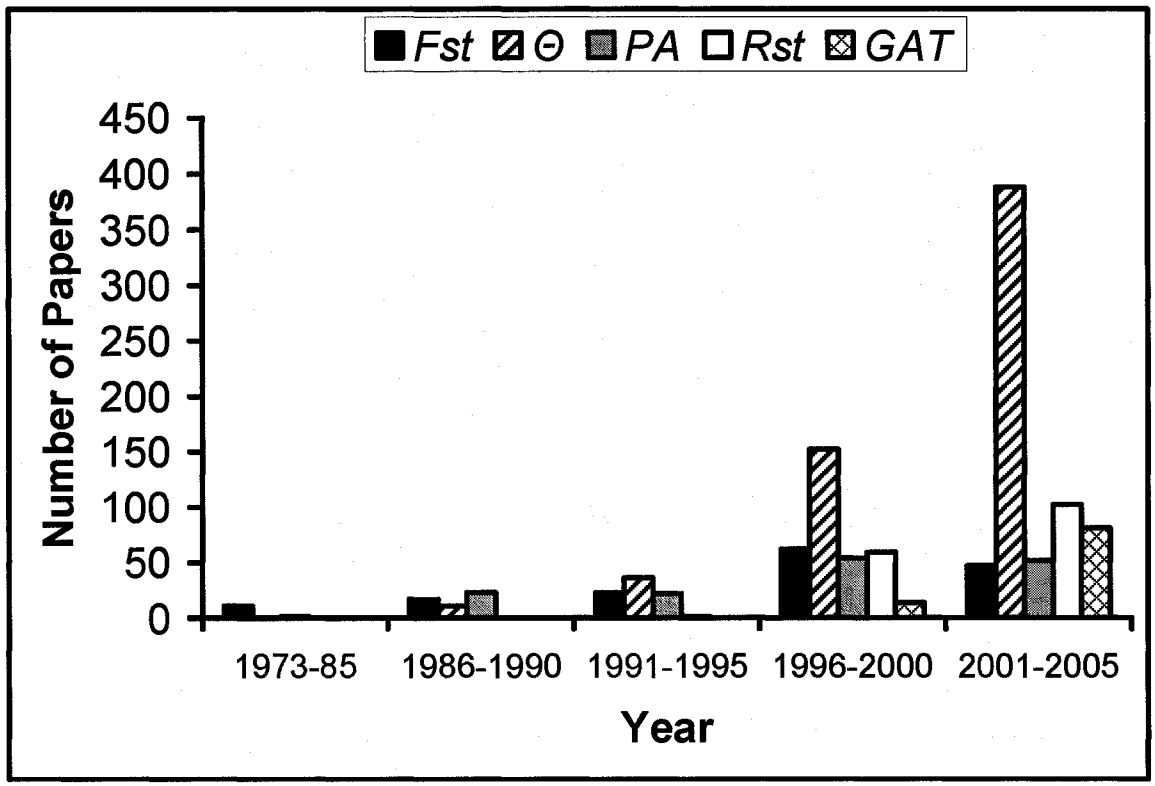


Figure 1.1. Number of empirical papers over time that measure movement indirectly based on F_{ST}, Θ, private alleles (PA), R_{ST}, and the genetic assignment test (GAT).

CHAPTER 2: WHEN CAN GENETIC DIFFERENTIATION BE USED TO ESTIMATE INTER-POPULATION MOVEMENT RATE?

Chapter summary

Several reviews suggest a potential for using genetic information to quantify inter-population movement. This could have advantages over mark and recapture methods, including lower cost and reduced disturbance of rare or endangered organisms. The theoretical basis for this approach is that genetic differentiation between populations is inversely related to inter-population movement. However, in practice this approach may be limited because the theoretical models are based on assumptions often either violated or not testable. Further, sample sizes and the number of loci available for sampling may be inadequate to reflect movement even when the assumptions are met. To determine the effects of these limitations on quantifying inter-population movement I developed a stochastic, individual-based simulation model in which both the demography and the genetics of the populations were known. In experimental simulations I manipulated the movement rate between two populations and calculated genetic differentiation (Θ), the effective number of immigrants (Nm), and the difference between actual movement and Nm. I examined the effect of percent of the population sampled, number of loci and alleles, population size, time since isolation, and whether generations overlapped or not on the relationship between Nm and actual movement. Generally Nm was not an accurate quantitative estimator of movement; it usually overestimated movement by anywhere from 2 to 8 fold. The degree of over-estimation increased exponentially with increasing actual movement rate. However, Nm was a reasonable estimate of movement rate for

simulations based on at least 20% of the population sampled, at least 10 loci with at least 6 alleles per locus, non-overlapping generations, at least 90 generations since isolation, and, especially, when actual movement rate was less than 4% per generation. This last requirement implies a Catch-22 situation, where the researcher needs to first be confident that the movement rate is low (less than 4%) before attempting to estimate it using genetic information. I found the direct relationship between Θ and movement to be potentially more reliable but not across all parameter combinations. I suggest further research on the theoretical relationship between Θ and movement that accounts for these parameters.

Introduction

Movement of organisms between populations is linked to the survival, and thus the conservation, of populations. Movement ensures that populations that have gone extinct are recolonized and/or that populations with low numbers are "rescued" from becoming extinct (Brown & Kodric-Brown, 1977) or from inbreeding depression (Frankham, 1995). However, inter-population movement is often difficult to estimate.

Researchers have used measures of genetic differentiation and distance to estimate movement between populations (Slatkin 1985a, 1987). With recent technological advances in molecular genetics (for reviews, see Mitton 1994; Parker et al. 1998; Manel et al. 2003), this approach has advantages over more traditional methods (e.g., mark recapture). Often, genetic information can be measured at a lower cost and with less disturbance of organisms.

In theory, genetic differentiation is inversely related to the amount of effective inter-population movement. Based on Wright's Island model (Wright 1931, 1951), the number of effective migrants (*Nm*) can be estimated from genetic differentiation (F_{ST}) (Slatkin 1985a). However, there is disagreement in the literature on the reliability of such movement estimates (Lewontin 1985; Slatkin 1985a; Crochet, 1996; Bohonak et al 1998; Bossart & Pashley Prowell, 1998a and b; Steinberg & Jordan, 1998; Whitlock & McCauley 1999; Bohonak & Roderick 2001; Neigel 2002). The assumptions of the Island model are that there is genetic equilibrium between drift and gene flow, demographic equilibrium, non-overlapping generations, no mutation, neutrality to selection, equality of gene flow and migration, random migration with respect to alleles, and linkage equilibrium. One of the main objections to using genetic differentiation to estimate actual movement is that these assumptions are often either violated or not testable. This would be especially true in the situation where the genetic differentiation is used to estimate movement between formerly panmictic populations that have been recently isolated, for example, by road construction or some other form of habitat fragmentation. The combined *quantitative* impact of these violations on the reliability of movement estimates based on genetic differentiation is not known. There may be instances when the relationship is robust to deviations from some of the assumptions.

Another issue in a conservation context is that it is sometimes difficult to obtain a large enough sample of individuals and/or loci per individual to get a reliable estimate of genetic differentiation. Genetically-derived measures of movement based on small

sample sizes, may not reflect actual movement adequately even when all the assumptions are met (Ruzzante 1998; Whitlock & McCauley 1999; Kalinowski 2005).

The usefulness of genetic methodologies to estimate movement rates between populations hinges on the precision and accuracy of estimates of movement based on genetic differentiation when samples sizes are limited and/or the assumptions (above) are not met. In this study I determine the effects of sample size and number of loci sampled on the reliability of estimates of inter-population movement based on genetic differentiation. This involves determining the difference between the actual amount of inter-population movement and that determined using the estimate of F_{ST} for different samples sizes and number of loci. I chose Θ (Weir & Cockerham 1984) to estimate F_{ST}. I also evaluate the effects of the number of alleles, whether generations overlap, the number of generations since isolation, population size, and the actual amount of inter-population movement on how well *Nm* quantifies actual movement. I also examine the relationship between Θ and the actual movement rate. These results are meant to provide a practical tool for an empirical researcher 1) to decide whether or not genetic information can be used to estimate inter-population movement reliably and 2) to determine the potential bias in their estimate of movement given a particular parameter combination (e.g., sample size, number of loci, etc).

Methods

Estimating genetic differentiation and inter-population movement

Estimation of genetic differentiation. I used Θ (Weir & Cockerham 1984) as my estimate of inter-population genetic differentiation, F_{ST}. Θ updates the initial formula for

F_{ST} introduced by Wright (1951). The equation for Θ includes terms that address the effects of multiple alleles and loci, the number of populations, sample sizes (both even and uneven), and observed heterozygosity. It is also currently one of the more commonly used estimates of genetic differentiation in the literature; Weir and Cockerham's paper has been cited 3,112 times based on the *Web of Science* database as of April 2006.

Estimation of movement. Wright (1931) related genetic differentiation to effective inter-population movement (i.e., that resulting in gene flow) by:

$$(2.1) \quad F_{ST} \approx \frac{1}{1+4Nm}$$

where N is the effective population size and m is the proportion of the population migrating per generation. This relationship makes several assumptions (Table 2.1). Slatkin (1985a) suggested using F_{ST} to estimate the number of individuals moving between populations, or Nm, by rearranging the previous equation:

$$(2.2) \quad Nm \approx \frac{1}{4} * \left(\frac{1}{F_{ST}} - 1 \right).$$

The simulation model

I designed an object-oriented individual-based simulation model with a hierarchical structure (Figure 2.1). The individual-based design was appropriate for two main reasons. Each individual's genotype and dynamics (movement, mating, aging and death) were governed by probability. The resulting stochasticity introduced a level of reality to the model not found in a deterministic model. Secondly, the individual-based

design allowed me to calculate statistics and genetic differentiation measures based both on sub-sampling the population and using all individuals within the population.

Each simulation began with an initially panmictic population that was subsequently subdivided into two populations (Figure 2.2). At each generation, individuals could potentially reproduce, die, age, and move to the other population (Figures 2.3a-d), with particular probabilities (Table 2.2). At the end of each generation I recorded the actual amount of movement, calculated the genetically-based estimate of the amount of movement, and calculated the difference (or standardized difference, see section "*Model Output*" below) between these two values. The offspring and surviving adults, if any, were carried over to the next generation. The sequence was iterated for 100 generations. Thus one simulation began with an isolation event creating two populations from a formerly panmictic one and ended 100 generations later. I incorporated the equations in Weir and Cockerham (1984) to calculate Θ based on the raw genetic data produced in the model runs.

Model Input Parameters

The parameters and the values used are shown in Table 2.2. The ranges in values were based on empirical studies. I chose the intervals within these ranges based on preliminary sensitivity analyses. I considered all levels of each parameter listed in a complete factorial experimental design. This resulted in 204,120 combinations. Since 100 replicates were run for each combination (below), the total number of simulation runs was 20,412,000.

The demographic parameters, while subject to stochastic variation, were not varied systematically among simulation runs (Table 2.2). I set the population sizes at the beginning of the simulation but stochastic variation in the sex ratio, number of offspring, deaths and migrants contributed to variation in the population sizes over the 100 generations. This variation was not density dependent. Population size increased over time in some runs, stayed the same in others, and decreased in still others, but the average population size across replicate runs was constant over the 100 generations.

At initialization there were on average an equal number of males and females in the populations (Figure 2.3b). When the individuals mated, a male was drawn randomly for each female in the population (Figure 2.3c). Any adult individual could potentially mate and males could potentially mate with multiple females (polygynous mating system). Immigrants and residents had the same probability of mating. The number of offspring produced was drawn from a Poisson distribution with a mean of two (Figure 2.3d). Thus, on average, 50% of the population (the females) produced two offspring. The death rate was set to counter the now doubled population size in one of two ways. When generations were non-overlapping (an assumption of the Island model), all adults (50% of the population, on average) were "killed." If the generations were overlapping, i.e., adults could coexist with juveniles, 50% of the individuals of any age in the population were "killed" at random.

I varied movement probabilities from complete population isolation (0) to panmixia (0.5) across simulations. The number of adults moving between the two populations was on average the same (symmetrical movement). Thus, in this closed

system, movement did not result in any net change in population size. When generations overlapped, adults could potentially survive multiple generations. In these instances I established two scenarios where 1) movement could only occur once in an adult's life regardless of lifespan or 2) movement could potentially occur once per generation for as many generations as the individual survived.

Individuals were initialized with a set of loci and a set of alleles for each locus (Figure 2.3b). The numbers reflected combinations found in empirical studies, from a one locus-two allele model to genotypes with 15 loci and 10 alleles. At initialization I assigned each allele in equal proportions, on average. Mutation was not included.

Isolated populations were expected to reach an equilibrium between drift and gene flow at a rate that depends on the population size (Crow & Aoki 1984). I thus also varied initial population sizes. I set the duration of the simulations at 100 generations from isolation to explore the effect of this approach to equilibrium on the relationship between actual movement and the genetically derived measure. I chose this duration as it was reasonable on an ecological time scale.

I calculated Θ and Nm based on sample sizes from two to 20 percent of the populations, and also included 100 percent for comparison. The sampling was done as a percentage to be equally representative regardless of population size. The sampling scenarios were independent meaning the results for different sample sizes were from separate simulations, and not multiple sub-samples from the same group of individuals.

I repeated each simulation 100 times for each of the above parameter combinations, based on preliminary simulations in which I looked for consistency in

mean values among sets of simulations. Additional trials did not yield significant changes in the mean value.

Model output

At the end of each generation, and for each parameter combination, I calculated the following: 1) actual population sizes, 2) actual amount of inter-population movement (referred to as "actual movement" from here on), 3) Θ, 4) estimated amount of inter-population movement, Nm , calculated from Θ (see equation 2.1) (referred to as "estimated movement" or "*Nm*" from here on), 5) difference between *Nm* and actual movement (bias), and 6) bias divided by actual movement (standardized bias) (Table 2.2). Standardized bias adjusted for the fact that the actual number of individuals moving increased with movement rate and also depended on the population size. Thus, if *Nm* and actual movement were equal for all 100 simulations (no bias), the bias (and standardized bias) would equal 0. If *Nm* was 2 times the actual movement, the standardized bias would be 1. If estimated movement was half actual movement, the standardized bias was -0.5. The standardized bias was negative if *Nm* underestimated actual movement and positive if *Nm* overestimated it. When actual movement was zero, the standardized bias was undefined. I thus excluded movement rates of zero from the analyses of standardized bias. This reduced the number of combination from 204,120 to 194,400. I considered estimates of movement based on Θ (*Nm*) to be reliable if (1) the average bias was between -0.5 and 1 (i.e., estimated movement was no less than half and no more than twice the size of actual movement) and (2) the range in bias over the 100 repeats of the

simulations was between -0.6 and 1.5. I considered these criteria reasonable given that my results are intended to help an empirical researcher determine whether or not to use genetic information to estimate inter-population movement.

Results

*Relationship between estimated movement (*Nm*) and actual movement*

Nm was not an accurate quantitative estimator of actual movement. While inter-population movement estimated from Θ (i.e., *Nm*) and actual movement were positively associated over most of the parameter combinations, this relationship was not linear. Figure 2.4a shows average *Nm* vs. actual movement for the parameter combination that gave the best relationship. Here, the calculated *Nm* was based on all individuals (sample size = 100%), 15 sampled loci, 100 generations since the isolating event (genetic equilibrium had been reached), and generations were non-overlapping. The population size was 500. Even for this relative best combination, the relationship did not fall on the 1 to 1 line, which is what I would expect if *Nm* were a good estimate of actual movement. Thus *Nm* did not directly estimate actual movement. Furthermore, as actual movement increased, *Nm* increasingly overestimated it. At movement rates above 0.2 (vertical dotted line on Figure 2.4), the overestimates were extreme. Many of the other combinations (e.g., fewer individuals or loci sampled) produced extremely high overestimates over most of the movement range, resulting in a very poor association (e.g., 4% sampled, Figure 2.4b).

Effects of parameter values on the reliability of Nm

As expected, the average standardized bias was closer to zero more often with increasing percentage sampled, number of loci, and number of alleles (Figures 2.5a-c). There was also a trend for a lower bias for simulations where the generations did not overlap and during time periods further from the isolation event (Figure 2.5d-e). The standardized bias decreased only slightly with increasing population size (Figure 2.5f).

As movement rate increased so did the standardized bias (Figure 2.5g). However, the degree of increase depended on the other parameter levels. Generally, the overestimation increased more sharply when there was a lower percent of the population and fewer loci sampled, fewer alleles per locus, and for populations with overlapping generations (Figure 2.6a-d).

Can Nm *be a reliable estimator of movement?*

There were 20,284 (10.4%) of the total 194,400 parameter combinations where the mean standardized difference between *Nm* and actual movement was between -0.5 and 1 (Figure 2.5h). However, the range in the standardized bias was large over the 100 repeated simulations for several combinations. The number of reliable combinations dropped to 4,899 (2.5%) when I considered only those with a standardized bias between a minimum of -0.6 and a maximum of 1.5. Finally, there were 2,500 (1.5%) combinations if I only considered those where less than the whole population was sampled. These are shown in Table 2.3.

Bias was consistently lower for simulations based on non-overlapping generations, at least 20% of the population sampled, 10 loci, 6 alleles, a movement rate of

0.01 to 0.04, and 90 generations since isolation (Table 2.3). However, there were many interactions among the parameters. For example, if the percentage sampled was lower than 16%, then at least 15 loci with 10 alleles were required, and this produced reliable estimates over a narrower range of movement rates (0.01-0.03). If generations overlapped, then reliable estimates of movement were only produced for populations sizes of at least 2,000, with the genetic measure calculated based on 15 loci and 10 alleles, at least 20% of the population sampled, at least 70 generations since isolation, and over a movement rate range of 0.01 to 0.03. In general, if one parameter was not at its "optimal" value, the other parameters needed to be at, or close to, their "optima" to produce a reliable estimate of actual movement.

An assessment of the direct relationship between Θ and the actual movement rate

Several authors have pointed out that using Θ directly to estimate movement may be more appropriate than using *Nm* (e.g., Cockerham & Weir 1993; Weir & Hill 2002). However, the direct relationship between Θ and actual movement was not linear (Figure 2.7). I attempted some simple transformations in an effort to linearize the relationship. None of these produced a linear relationship across all parameter values.

Discussion

Overall, *Nm* was a poor quantitative estimator of actual movement. The bias in the estimation of movement changed with the parameter values in a non-linear way. However, simulation with 10 or more loci, over 16% of the population sampled, 6 or more alleles per locus, at least 90 generations since isolation, for a species with non-

overlapping generations, and, especially, when actual movement rate was 0.04 or less produced relatively reliable estimates of actual movement.

The researcher faces different challenges in deciding to use genetic information to estimate inter-population movement depending on which parameter is "sub-optimal." Simulations with smaller sample sizes, either in terms of numbers of individuals (fewer than 10%) or number of loci sampled (10 or fewer), and fewer alleles (2-6) did not reflect actual movement adequately even when all the assumptions were met. However, these parameters are the ones that a researcher can potentially manipulate. Whether generations overlapped also had a relatively large effect on bias, but this can not be modified by a researcher since it is an aspect of the species biology. If generations did overlap, only combinations with the relative "best" values produced reliable estimates of movement. Therefore, while having non-overlapping generations is an assumption of the Island model, there are instances where *Nm* can produce reliable estimates of movement for populations with overlapping generations. Finally, the highest simulated movement rate that produced a mean standardized bias between -0.5 and 1 was 0.18 (when 100% of the population was sampled). This movement rate was only 0.04 if the percent of the population sampled was $\leq 20\%$ (more realistic as 100% of the population sampled is not generally feasible). Further, the range in movement rate at which reliable estimates of actual movement were produced decreased when the parameters were not at their "optimal" values. Researchers thus would face the challenge that unless they had some prior knowledge of movement rate, they would have no idea of the potential bias in their estimate. This leads to the rather unfortunate conclusion that in order to decide whether Θ

can be used to estimate actual movement in a particular situation, the researcher needs to already have an estimate of actual movement.

The average *Nm* generally did increase with actual movement, though not in a 1:1 relationship. This result suggested that using genetic differentiation (*Nm*) at least as a *qualitative* or comparative tool to assess actual movement is justified under certain circumstances. This supports conclusions regarding the qualitative value of *Nm* discussed in Crochet (1996), Whitlock and McCauley (1999), and Bohonak et al (1998) for example. However, due to the non-linearity (Figures 2.4 and 2.5) coupled with variability in the bias (Figure 2.6), detecting changes in actual movement would be nearly impossible.

My model included several simplifying assumptions. I assumed gene flow and movement were equivalent, as residents were equally likely to mate as immigrants. The model was not spatially explicit. The system was closed with movement taking place only between the two populations. I did not consider the effects of extinction, colonization, and bottlenecks explicitly. The order of the birth, death, movement, and reproduction did not vary among simulations and only adults moved. Generally, if I relaxed the above mentioned assumptions or included additional effects (e.g., uneven sample sizes), the main difference in my results would have been to further de-couple *Nm* from actual movement and to increase bias and variability, thus further compromising the value of *Nm* as a measure of inter-population movement.

Implications and conclusions

In practical terms, my results suggest that it is not feasible to use *Nm* as a quantitative estimator of movement for two reasons. First, to produce reliable estimates over a range of population sizes, sample size should represent no less than 20% of the population, the genetic measure should be based on at least 10 loci with at least 6 alleles each, there should be at least 90 generations since isolation, and the species should have non-overlapping generations. Also, and most importantly, the suspected movement rate must be lower than 0.04. If one of these variables is somewhat "sub-optimal," the values of the others must be optimal. For example, estimates based on small sample sizes or on few loci can be misleading, even if all the assumptions are met. Real studies, particularly studies of endangered species, are likely to have low sample sizes, recently isolated populations, and overlapping generations. Second, the high variability over the 100 simulation runs for each parameter combination suggests that individual estimates may not be of any value even if the parameter values are optimal and the average bias across all 100 runs for that parameter combination was low.

One of the important parameters affecting bias was the movement rate itself. If a researcher already knew what the rate of movement was, there would be little point in estimating it using *Nm*. However, if all the other parameter values are known and optimal, the researcher could conclude that a calculated *Nm* of less than 0.04 may be a reasonably reliable estimate of actual movement. If the value of *Nm* is larger than 0.04, then it is not reliable, but it may indicate that actual movement is fairly high, which may be a practical qualitative result.

My ability to estimate movement accurately was only marginally improved by relating Θ (rather than Nm) to actual movement. Examining the direct relationship between Θ and actual movement has been suggested (Weir & Hill 2002) in part because conversion (based on equation 2.2) of Θ values close to 0 lead to very high values of Nm (see Waples 1998; Whitlock & McCauley 1999; Neigel 2002). I could not determine a straightforward adjustment to Θ to correct the bias in estimating movement. I suggest that future studies are needed to further develop the theoretical relationship between Θ and actual movement that accounts for the important factors affecting this relationship. These included movement rate, sampling (numbers of individuals and/or loci), number of alleles, and whether generations overlap or not.

Table 2.1. Assumptions made in the Island Model when relating F_{ST} (in this case Θ) to Nm (Wright 1931).

Assumptions	**How addressed in my model**
genetic equilibrium between drift and gene flow	pattern examined over 100 generations, approaching equilibrium
generations do not overlap	pattern compared for overlapping vs. non-overlapping generations
no mutation	assumed
demographic equilibrium	assumed but subject to stochastic variation
gene flow = migration	assumed but subject to stochastic variation
neutral to selection	assumed
migration random with respect to alleles	assumed
linkage equilibrium	assumed

Table 2.2. Parameters and values used in the simulation experiments.

INPUT	
Parameters	**Value**
Held constant for all simulations:	
Number of offspring per female:	2 (drawn from Poisson distribution)
Death rate:	0.5 (individuals chosen at random)
Sex ratio:	1:1
Number of populations:	2
Number of repetitions of simulation	100
Duration of simulation:	100 generations from isolation
Varied among simulations:	
Initial population sizes:	500, 1000, 2000
Sample percentages:	2, 4, 6, 8, 10, 12, 16, 20, 100%
Number of loci:	1, 5, 10, 15
Number of alleles:	2, 6, 10
Movement probability:	0 to 0.5 by 0.01; 0 to 0.2 used in the analyses, see results
Overlapping generations:	no/yes
Once in a lifetime movement:	no/yes
OUTPUT	
Produced for each input parameter combination:	The following output was reported for each of these parameter combinations and represents the descriptive statistic over the 100 repeated simulations:
Population sizes Actual movement Genetic measure *(Θ)* Estimated movement *(Nm)* Bias *(Nm-actual movement)* Standardized bias *((Nm-actual movement)/actual movement)*	• average • minimum • maximum • variance • standard error • 95% confidence interval

Table 2.3. Parameter combinations and ranges (see Table 2.2) that consistently produced a reliable estimate of actual movement. For these combinations (a) the mean standardized bias was between -0.5 and 1, (b) the bias of each of the 100 individual simulations was between -0.6 and 1.5, and (c) populations were sub-sampled.

Population sizes	Number of loci	Number of alleles	Percentage sampled	Movement range	Generations since isolation
Overlapping generations; individuals can move more than once in lifetime					
2000	15	10	16	0.01-0.04	80-100
			20	0.01-0.03	70-100
				0.04	90-100
Overlapping generations; individuals can move only once in lifetime					
2000	15	10	20	0.01-0.03	100
Non-overlapping generations					
500	5	10	20	0.01-0.03	100
	10	6	20	0.01-0.05	90-100
		10	16	0.01-0.03	100
			20	0.01-0.04	80-100
	15	6	16	0.01-0.03	100
			20	0.01-0.05	100
		10	8	0.01-0.02	100
			10	0.01-0.03	100
			12	0.01-0.03	100
			16	0.01-0.04	100
			20	0.01-0.05	90-100
1000	10	6	16	0.01-0.04	90-100
			20	0.01-0.04	90-100
		10	8	0.01-0.02	70-100
			10	0.01-0.03	30-100
			12	0.01-0.03	30-100
			16	0.01-0.04	30-100
			20	0.01-0.05	20-100
	15	6	6	0.01-0.02	100
			8	0.01-0.03	90-100
			10	0.01-0.03	70-100
			12	0.01-0.03	40-100
			16	0.01-0.04	30-100
			20	0.01-0.05	30-100
		10	6	0.01-0.02	80-100
			8	0.01-0.03	30-100
			10	0.01-0.04	30-100
			12	0.01-0.04	30-100
			16	0.01-0.05	20-100
			20	0.01-0.05	20-100

... Table 2.3., continued

Population sizes	Number of loci	Number of alleles	Percentage sampled	Movement range	Generations since isolation
2000	5	10	10	0.01-0.02	90-100
			12	0.01-0.02	50-100
			16	0.01-0.03	30-100
			20	0.01-0.03	30-100
	10	6	10	0.01-0.02	80-100
			12	0.01-0.02	40-100
			16	0.01-0.03	30-100
				0.04	80-100
			20	0.01-0.04	30-100
		10	6	0.01-0.02	30-100
			8	0.01-0.02	30-100
			10	0.01-0.03	30-100
			12	0.01-0.03	20-100
			16	0.01-0.04	20-100
				0.05-0.06	80-100
			20	0.01-0.06	20-100
	15	6	8	0.01-0.02	40-100
			10	0.01-0.02	40-100
			12	0.01-0.03	30-100
			16	0.01-0.05	20-100
			20	0.01-0.05	20-100
				0.06-0.07	100
		10	4	0.01-0.02	100
			6	0.01-0.02	20-100
			8	0.01-0.03	20-100
			10	0.01-0.04	20-100
			12	0.01-0.04	20-100
			16	0.01-0.06	20-100
			20	0.01-0.06	10-100

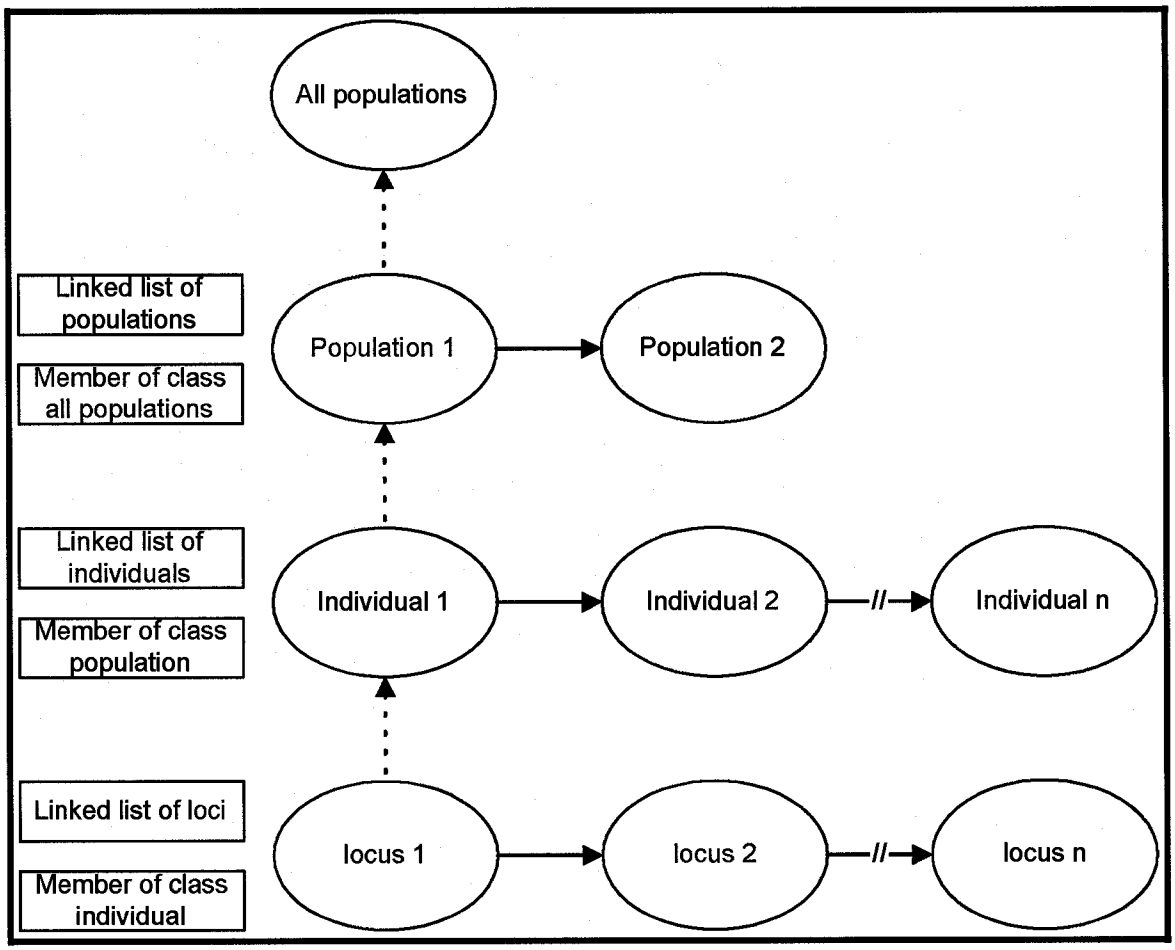


Figure 2.1. Hierarchy of objects in the simulation model. Each object is part of a list and is a member of the class above.

Figure 2.2. Overview of a simulation for a given combination of parameter values. The simulation begins with a panmictic population that becomes divided at time 0. In each generation, the two populations exchange individuals with a given probability. Individuals also age, reproduce, and die. At each generation the output variables in Table 2.2 are generated. The simulation is repeated 100 times, generating 100 sets of data for each generation and parameter combination. Summary statistics are generated from these 100 points.

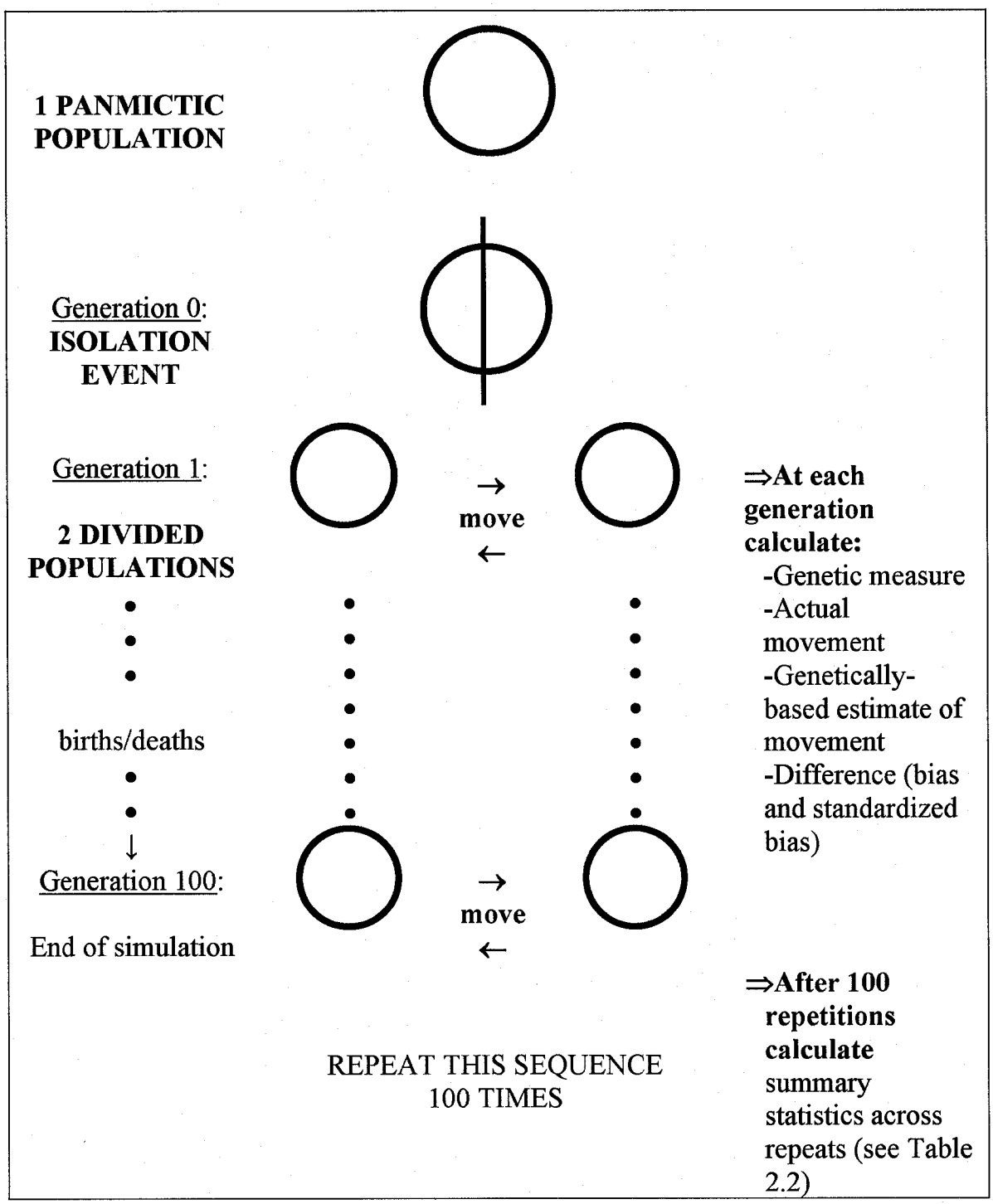
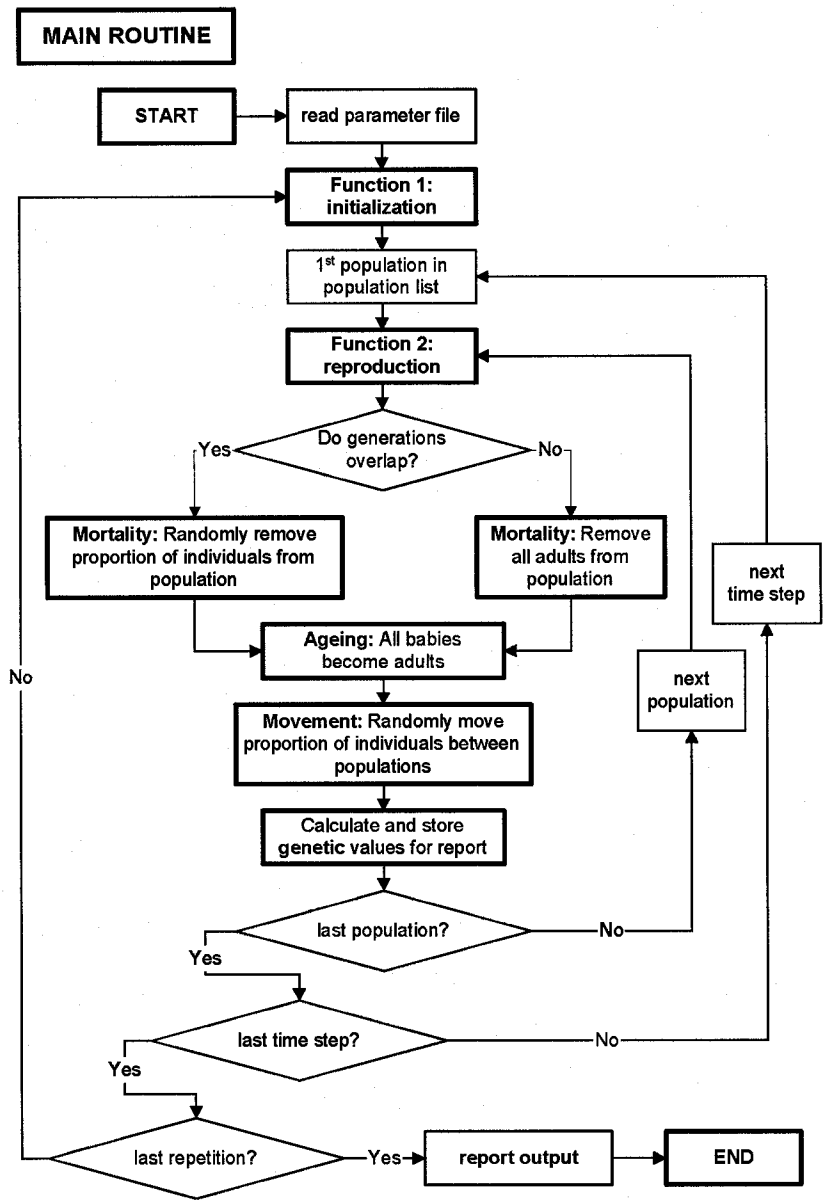
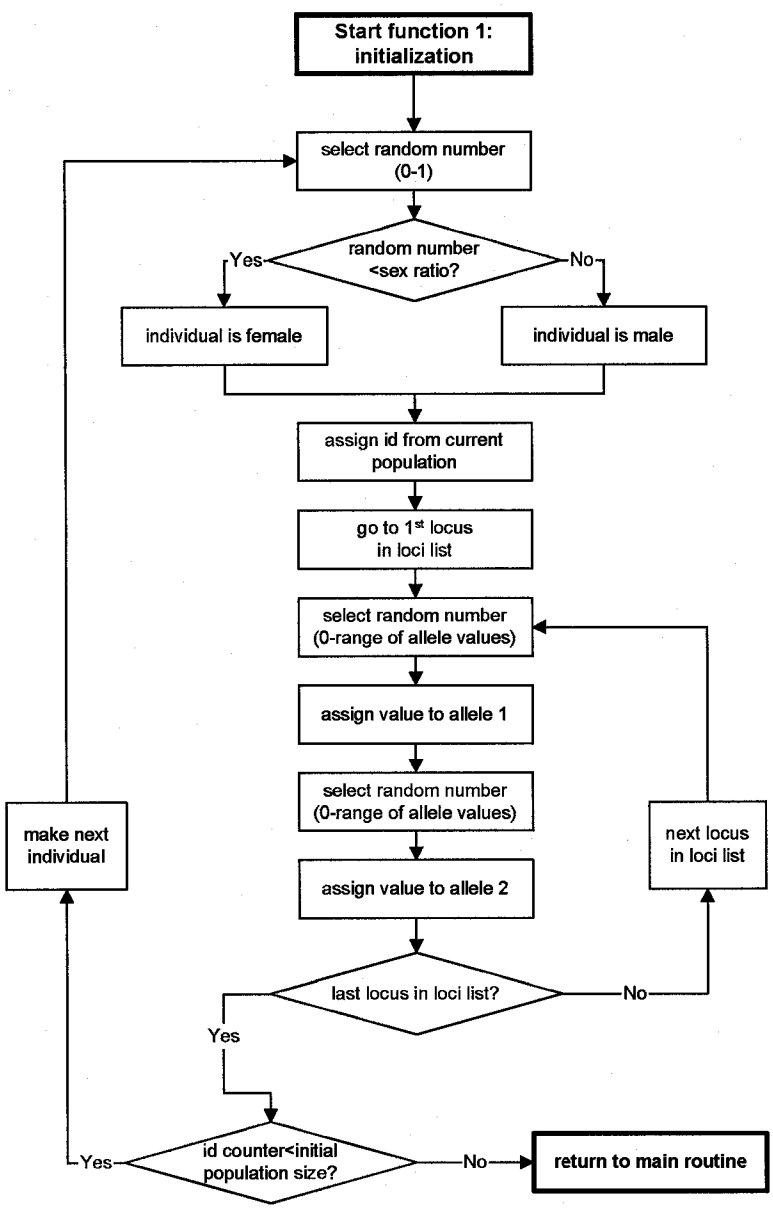


Figure 2.3. Flow diagrams of the simulation steps. The model comprises main routine (A) and several subroutines: (B) initialization; (C) reproduction; (D) actual mating event and assignment of parental genotypes to offspring. Note, the mortality pattern depends on whether or not the generations overlap and movement pattern depends on whether or not generations overlap and whether an individual can move more than once in its lifetime (see text).

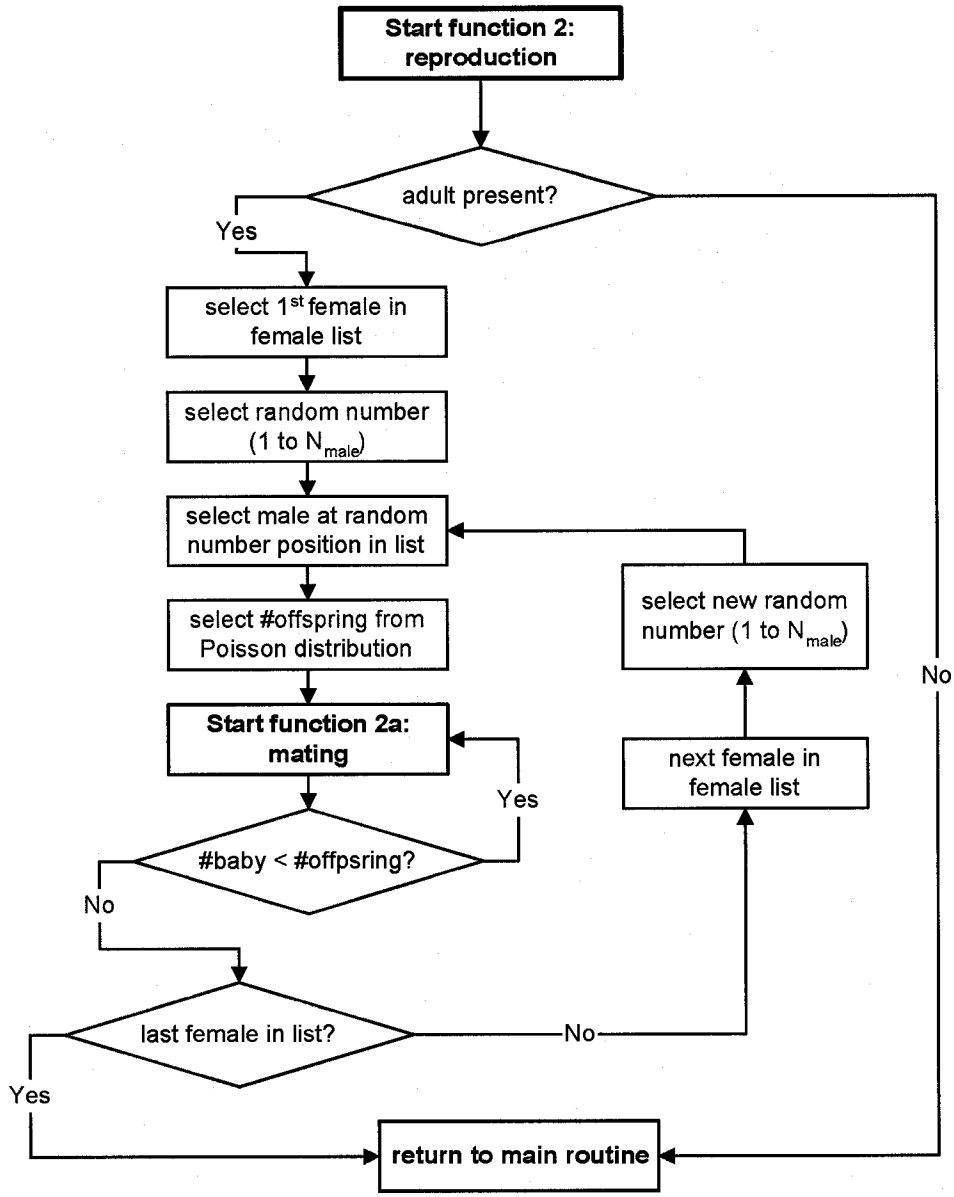
A. main routine



B. initialization routine



C. reproduction routine



D. actual mating and assignment of parental genotypes to offspring routine

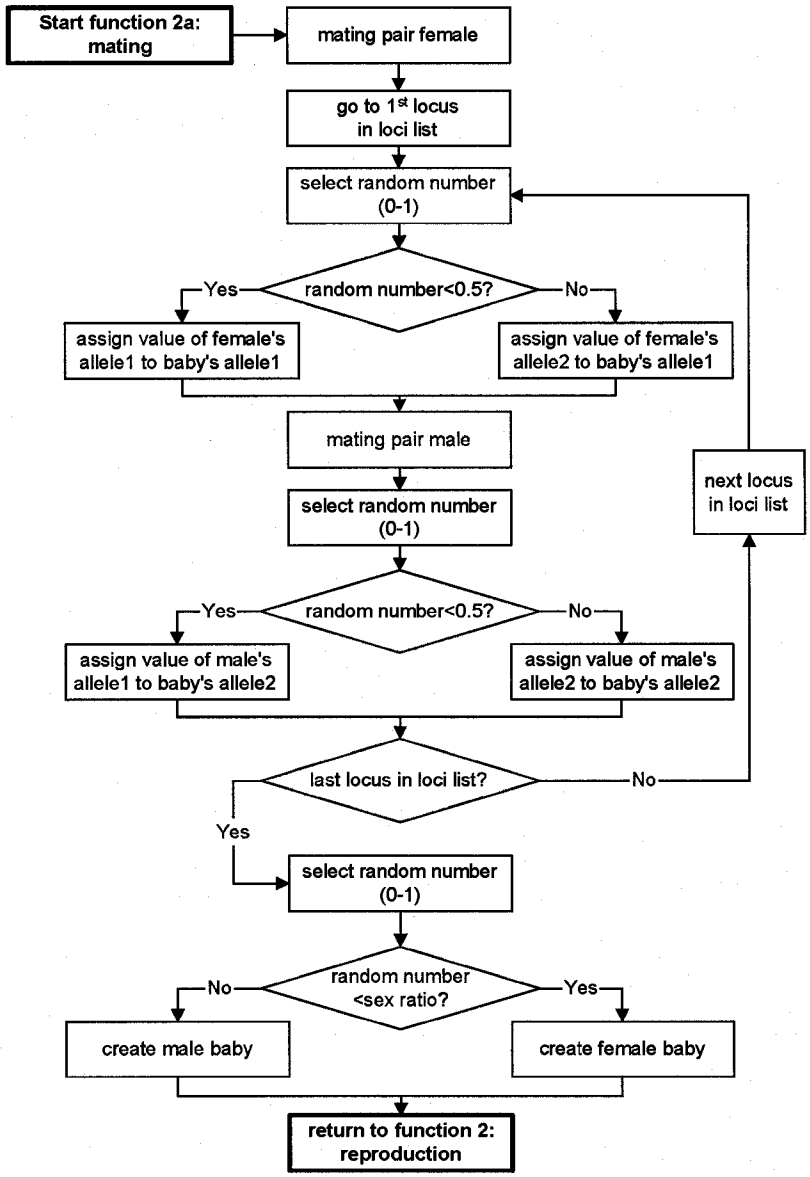


Figure 2.4. Plot of the relationship between estimated movement based on Θ (Nm) vs. actual movement for (A) a population size of 500 (percent sampled = 100%, loci number = 15, allele number = 10, generations since isolation = 100, and generations = non-overlapping). The values represent the averages over 100 replicate simulations. I plotted both axes on a log scale to show the pattern while including a wide range of values. The vertical dashed line indicates the movement rate above which the relationship is consistently unreliable for all parameter combinations. Panels B and C represent the relationship between mean Nm and an actual movement of ≤ 0.2 for the same combination as in panel A but for (B) selected values of percent of population sampled and (C) number of loci.

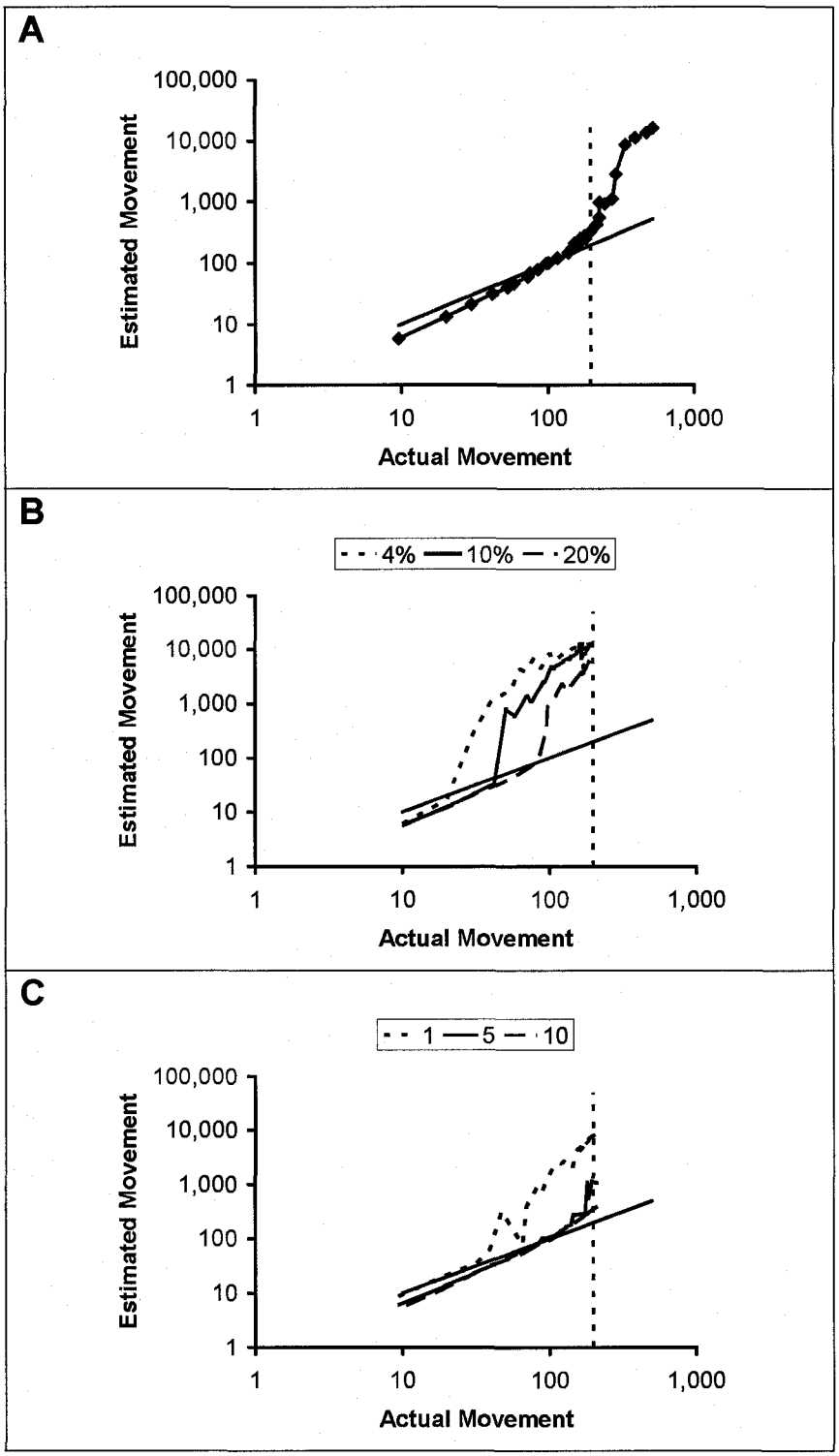
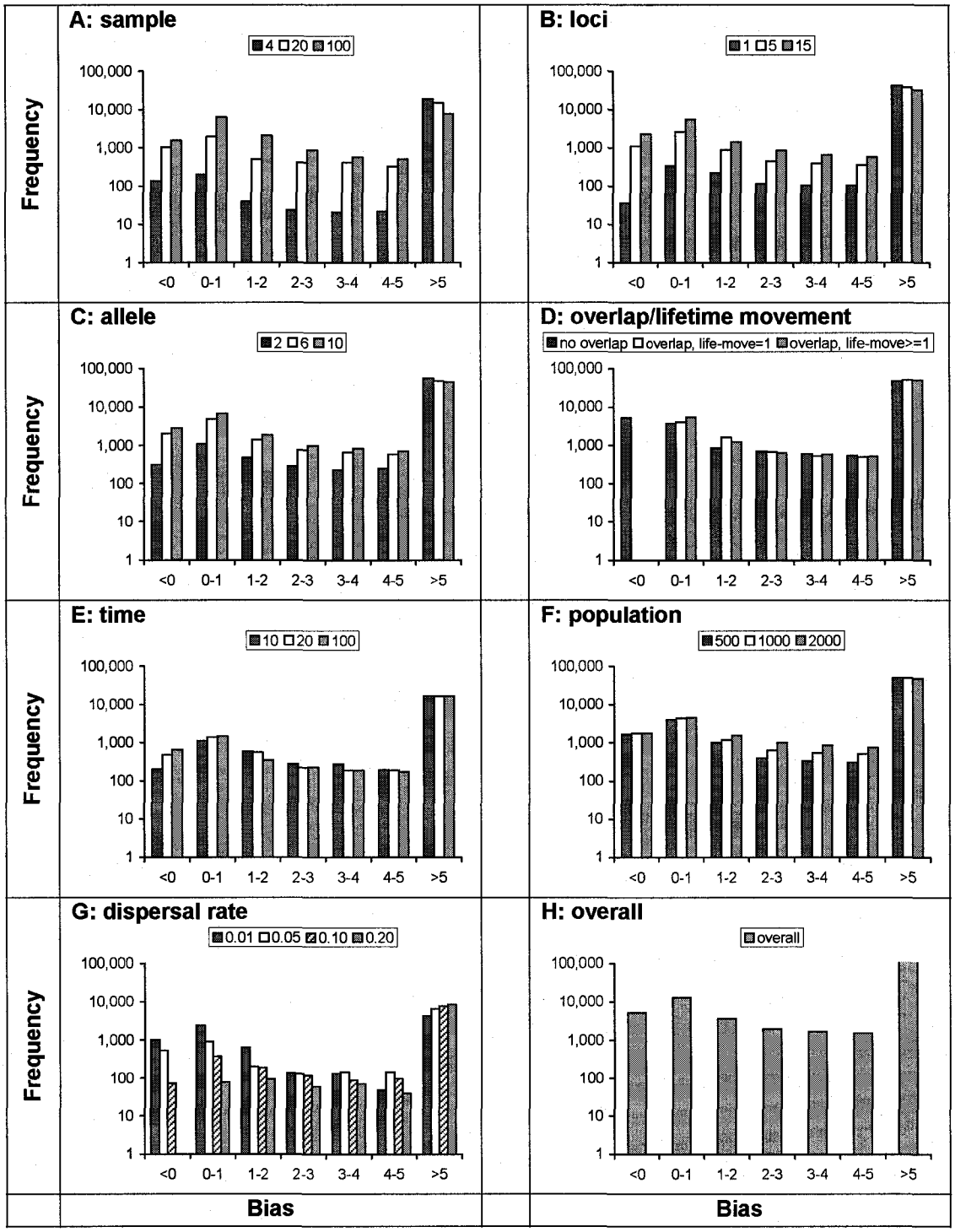


Figure 2.5. Frequency histograms of the mean standardized bias (i.e., [estimated movement-actual movement]/actual movement) for a selection of the parameter values shown in Table 2.2. Movement rates were ≤ 0.2 (to the left of the vertical dotted line in Figure 2.4). The bars in panels A to G show how many times the mean of the 100 replicate simulations were within a particular range when all other parameters were pooled.



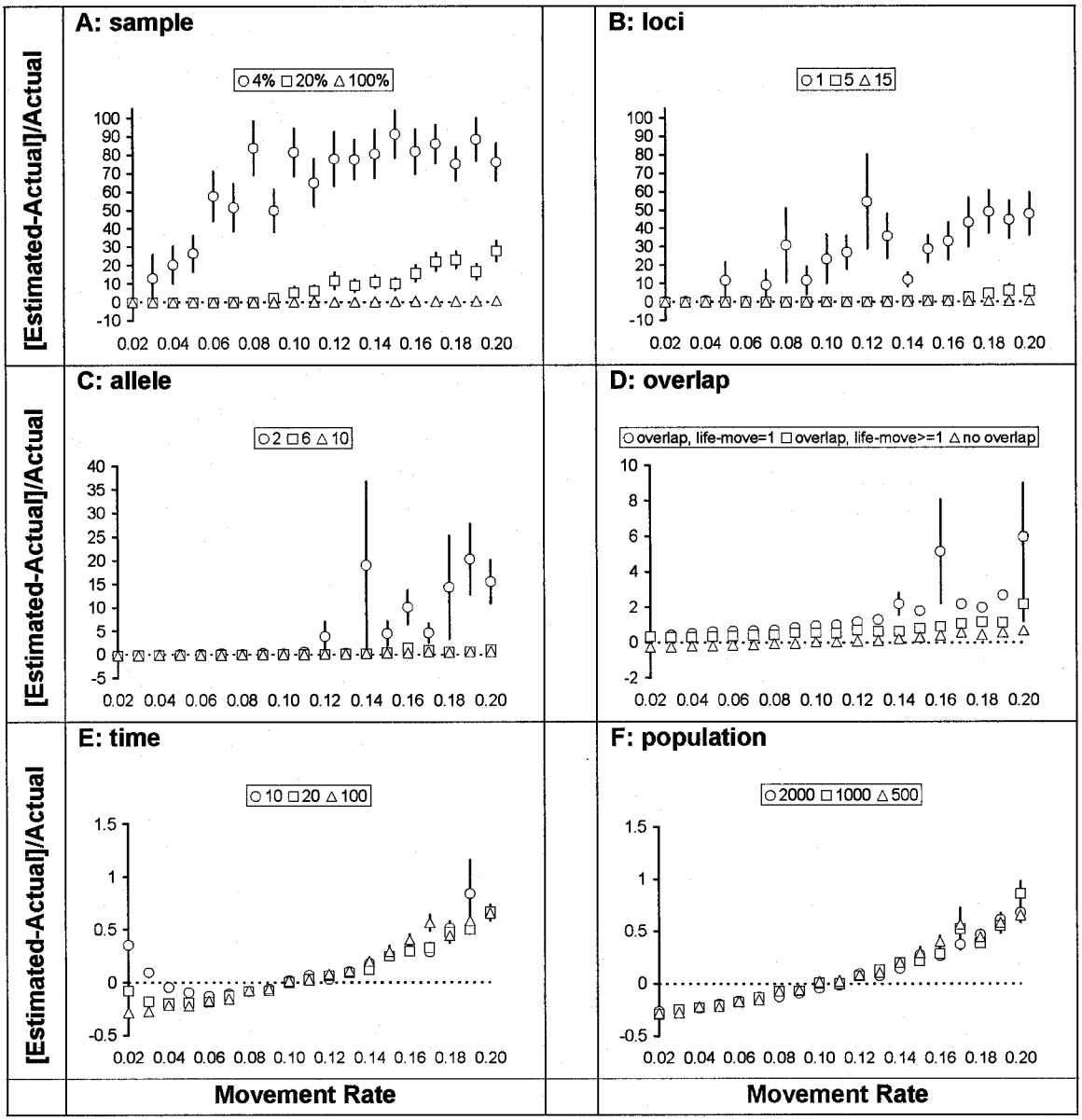


Figure 2.6. Standardized bias (± 1SE) vs. the actual movement rate. For all plots shown except F, population sizes depicted were 500. The triangle in each panel is the default combination (sample = 100%; 15 loci; 10 alleles; non-overlapping generations, and 100 generations since isolation). Panels A-F show how changing (A) sample sizes, (B) number of loci, (C) number of alleles, (D) life histories, (E) population size, and (F) time since isolation affect the standardized bias for the given movement rate.

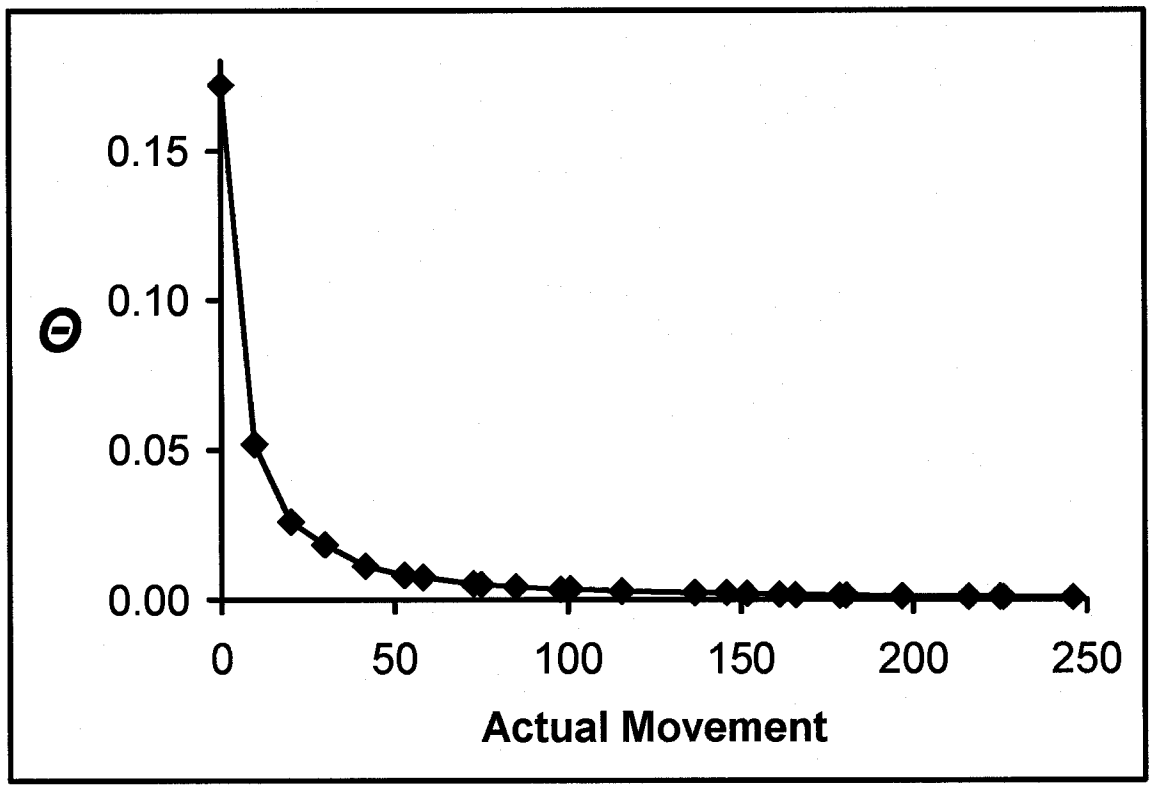


Figure 2.7. Relationship between Θ and actual movement for the same parameter combination as presented in Figure 2.4a.

CHAPTER 3: COMPARISON OF GENETIC METHODS FOR ESTIMATING INTER-POPULATION MOVEMENT

Chapter summary

Several reviews suggest using genetic information to quantify inter-population movement, noting advantages over more traditional measures of movement (e.g., mark and recapture). Genetic methods include calculating movement based on F_{ST}, G_{ST}, Θ, R_{ST}, private alleles, and genetic assignment tests. I asked which, if any, of these methods is suitable for calculating inter-population movement of recently isolated small populations, i.e., the situation of interest in conservation studies. I used a stochastic, individual-based model to simulate the demography and the genetics of hypothetical populations. In experimental simulations I manipulated inter-population movement rate and calculated the amount of movement based on the genetic measures. I examined the effects of percent of the population sampled, number of loci and alleles, population size, time since isolation, and whether generations overlapped or not on the difference between the genetic estimates of movement and actual movement (bias). Estimates based on R_{ST}, private alleles, and the genetic assignment tests did not reliably quantify movement rate. This result is significant because it suggests that the recent increase in use of these methods is misplaced. Interestingly, the original F_{ST} method, first introduced by Wright in 1951 and modified for more than 2 alleles by Nei in 1973, provided a better estimate of inter-population movement than its more recent (assumed to be improved) incarnations, G_{ST}, G_{ST}' and Θ. F_{ST}-based estimates of movement were reasonable in situations with at least 20% of the population sampled, at least 10 loci with at least 6

alleles each, a movement rate of 0.04 to 0.08 per generation, and at least 90 generations since isolation. Θ and G_{ST}'-based estimates of movement were reasonable for the same parameters but limited to situations with non-overlapping generations and a movement rate of less than 0.04. Note that F_{ST} and Θ are only reliable over narrow movement ranges, creating a Catch-22 situation where one must already know the movement rate to determine whether and which genetic methods can be used to estimate movement. I conclude that currently available genetic methods for estimating movement are inappropriate for most studies of species conservation.

Introduction

Movement of organisms between populations is linked to the survival, and thus the conservation, of these populations (Brown & Kodric-Brown, 1977; Frankham, 1995). However, it is often challenging to estimate inter-population movement. Slatkin (1985a, 1987) suggested that measures of genetic differentiation could be used to estimate movement between populations. With technological advances in molecular genetics (for reviews, see Mitton 1994; Parker et al. 1998; Manel et al. 2003), the use of genetic measures to quantify movement has advantages over more traditional methods (e.g., mark recapture). Often, genetic information can be measured at a lower cost and with less disturbance of organisms.

Over the last several years, researchers have used a variety of methods to estimate movement from genetic data. Among these are several measures of genetic differentiation based initially on Wright's (1951) formulation of F_{ST}, which was subsequently extended to include a series of additional terms for multiple alleles (Nei's F_{ST}, 1973), observed

heterozygosity and sample size (Nei's G_{ST}, 1987), number of populations (Nei's G_{ST}', 1987), and multiple loci and uneven sample sizes (Weir and Cockerham's Θ, 1984). Θ is currently one of the most commonly used estimates of genetic differentiation in the literature. Weir and Cockerham's paper had been cited 3,112 times as of April 2006, compared to 2,265 citations of Nei (1973), published 11 years earlier (*Web of Science* database). The original F_{ST} and its more recent estimates apply to allozyme loci. Slatkin (1995) proposed R_{ST} for microsatellite loci. Slatkin (1985b) also proposed a method based on the frequency of private alleles.

The use of estimates of F_{ST} and its derivatives, G_{ST}, G_{ST}', and Θ, to quantify movement has been highly debated in the literature. The main objection is that assumptions (Table 3.1) are either violated or not testable (Lewontin 1985; Slatkin 1985a; Crochet, 1996; Bohonak et al. 1998; Bossart & Pashley Prowell, 1998a and b; Steinberg & Jordan, 1998; Whitlock and McCauley 1999; Bohonak and Roderick 2001; Neigel 2002). These violations would be especially likely in the situation where genetic differentiation is used to estimate movement between formerly panmictic populations that have been recently isolated, for example, by road construction or some other form of habitat fragmentation. However, the combined *quantitative* impact of these violations on the reliability of movement estimates based on genetic differentiation is not known. There may be instances when the relationship is robust to deviations from some of the assumptions.

This debate about the effect of violations to the assumptions has lead researchers to consider alternative methods that do not make these assumptions. In particular,

individual-based assignment tests (Paetkau et al. 1995; Rannala & Mountain 1997; Waser & Strobeck 1998; Pritchard et al. 2000; Paetkau et al. 2004) have been growing in use over the last decade with citations increasing steadily each year (*Web of Science* database). The assignment test developed by Paetkau et al. (1995) is the one most widely used in empirical studies (Guinand et al. 2002). Movement is calculated based on the number of individuals whose genotypes have a higher likelihood of belonging to other populations than the one in which they were found (Paetkau et al. 1995; Waser and Strobeck 1998). In a later version of the assignment test, an individual is considered a migrant if its likelihood of belonging to other populations is higher than a statistical threshold (Paetkau et al. 2004). Because the test makes few assumptions often violated in recently isolated populations (Table 3.1) it is expected to provide more reliable estimates of movement (Waser & Strobeck 1998).

In this study I explore the accuracy of movement estimates calculated from eight methods - F_{ST}, G_{ST}, G_{ST}', Θ, R_{ST}, private alleles, and two versions of the genetic assignment test. These methods differ in the assumptions they make (Table 3.1), how they are calculated, how movement is derived from them (see *Genetic methods and movement* below), and the genetic data from which they are calculated (allozymes or microsatellite DNA). These differences may affect how well they reflect actual inter-population movement. My results are meant to provide a genetic methodological "road map" as a guideline for the appropriate use of population genetic information to measure inter-population movement in the field, given a particular sampling, genetic, and demographic scenario.

Methods

Genetic methods and movement

Genetic methods. I used eight genetic methods to derive estimates of movement: five estimates of genetic differentiation (F_{ST}, G_{ST}, G_{ST}', Θ and R_{ST}), the frequency of private alleles, and two versions of the genetic assignment test (Table 3.1). I calculated a version of F_{ST} based on Nei's (1973) extension of the formula introduced by Wright (1951) to address more than two alleles per locus. G_{ST} includes additional terms that consider the number of observed heterozygotes and the sample size, and G_{ST}' is also independent of the number of samples (Nei 1987). Θ (Weir & Cockerham 1984) includes terms that address the effects of multiple alleles and loci, the number of populations, sample sizes (both even and uneven), and observed heterozygosity. R_{ST} is based on the number of tandem repeats and is applicable to microsatellite loci, which tend to have a higher mutation rate (Slatkin 1995). For the private alleles estimate, I calculated the logarithm of the average frequency of alleles found in only one population (i.e., private alleles), which is predicted to decrease linearly with the logarithm of effective inter-population movement (Slatkin 1985b). Finally, I conducted two versions of the genetic assignment test. Both are based on comparing the gene frequencies of an individual with those of the population in which it was found and with those of a purported alternate source population. In the first version, an individual is assigned to the population for which its log likelihood of occurring is higher (Paetkau et al. 1995; Waser & Strobeck 1998). In the second version (Paetkau et al. 2004) each potential immigrant had to have a

statistically significantly higher likelihood of belonging to another population, as determined using randomization methods.

Estimation of movement from genetic methods. I used Wright's (Wright 1931; Table 3.1) formula for relating the genetic differentiation measures (i.e., F_{ST}, G_{ST}, G_{ST}', and Θ) to effective inter-population movement (i.e., that resulting in gene flow):

$$(3.1) \quad F_{ST} \approx \frac{1}{1+4Nm}$$

where N is the effective population size and m is the proportion of the population migrating per generation. Slatkin (1985a) suggested using F_{ST} to estimate the number of individuals moving between populations, or Nm, by rearranging the previous equation:

$$(3.2) \quad Nm \approx \frac{1}{4} * \left(\frac{1}{F_{ST}} - 1 \right).$$

Note that G_{ST}, G_{ST}', and Θ, are substituted for F_{ST} when these are the measures of genetic differentiation. In the case of R_{ST}, I used

$$(3.3) \quad Nm \approx \frac{d_s - 1}{4d_s} \left(\frac{1}{R_{ST}} - 1 \right) \text{ (equation 15a in Slatkin 1995),}$$

where d_s is the number of populations. In my case d_s is 2, so

$$(3.4) \quad Nm \approx \frac{1}{8} \left(\frac{1}{R_{ST}} - 1 \right).$$

For private alleles, I calculated the movement rate by rearranging equation 3 in Slatkin (1985b) and including his correction for sample size (see p.57, Slatkin 1985b):

$$(3.5) \quad Nm = e^{\left(\frac{\ln\left(\bar{p}(1)\right) + 2.440}{-0.505} \right)} * \left(\frac{25}{n} \right),$$

where $\overline{p}(1)$ is the frequency of private alleles, and n is the sample size.

For both genetic assignment tests I calculated movement by considering individuals to be immigrants if they were assigned to the population other than the one from which the individual was sampled. In the first version, an individual is assigned to the population for which its log likelihood of occurring is higher (Paetkau et al. 1995; Waser & Strobeck 1998). In the second version (Paetkau et al. 2004) potential immigrants also had to be significantly more likely to belong to the other population. I determined this significance for a potential migrant by 1) generating a set of 100 individuals whose genotypes were randomized based on the allele frequencies for each locus in the population in which the individual was found (the frequency excluded the alleles of the potential migrant), 2) calculating the test statistic, A, for each of these 100 randomized individuals, where A is the ratio of the likelihood of the individual's being a resident of the population from which it was sampled over the likelihood of its belonging to the other population, and 3) classifying the individual as a migrant if its A value was greater than that of the highest in the ratio in the randomized population (i.e., $\alpha =$ 0.01).To adjust for sample size, I multiplied the proportion of immigrants in the sample by population size.

The simulation model

I designed an object-oriented individual-based simulation model with a hierarchical structure (Figure 2.1). Each individual's genotype and dynamics (movement, mating, aging and death) were governed by probabilities. The individual-based design

allowed me to calculate statistics and each of the genetic measures based both on sub-sampling the population and using all individuals within the population.

Each simulation began with an initially panmictic population that was subsequently subdivided into two populations (Figure 2.2). At each generation, individuals could potentially reproduce (with or without mutation), die, age, and move to the other population (Figure 2.3a-d), with particular probabilities (Table 3.2). At the end of each generation I recorded the actual amount of movement, calculated each of the genetically-based estimates of the amount of movement, and calculated the difference (or standardized difference, see *Model Output* below and Table 3.2) between these two values. The offspring and surviving adults, if any, were carried over to the next generation. The sequence was iterated for 100 generations. I calculated F_{ST}, G_{ST}, G_{ST}', Θ, R_{ST}, private alleles, and genetic assignment based on the raw genetic data produced in the model runs.

Model Input Parameters

The parameters and the values used are shown in Table 3.2. The ranges in values were based on empirical studies. I chose the intervals within these ranges based on preliminary sensitivity analyses. I ran simulations for all levels of each parameter listed in a complete factorial experimental design. This resulted in 1,016,000 combinations. Since 100 replicates were run for each combination (below), the total number of simulation runs was 101,600,000. For the more computationally intense simulations (i.e., those based on R_{ST}, the frequency of private alleles, and the genetic assignment tests) I examined the effects of the largest sample sizes initially, and then for successively

smaller sizes only if the simulations with larger populations yielded reliable estimates of actual movement. For all methods I repeated each simulation 100 times for each of the above parameter combinations, based on preliminary simulations in which I looked for consistency in mean values among sets of simulations. Additional trials did not yield significant changes in the mean values.

The demographic parameters, while subject to stochastic variation, were not varied systematically among simulation runs (Table 3.2). I set the population sizes at the beginning of the simulation, but stochastic variation in the sex ratio, number of offspring, deaths and migrants contributed to variation in the population sizes over the 100 generations. This variation was not density dependent. Population size increased over time in some runs, stayed the same in others, and decreased in still others, but the average population size across replicate runs was constant over the 100 generations.

At initialization the sex ratio was on average 1:1 (Figure 2.3b). When the individuals mated, a male was drawn randomly for each female in the population (Figure 2.3c), so while all adult females mated, some adult males mated with multiple females and others did not mate. Immigrants and residents had the same probability of mating. The number of offspring produced was drawn from a Poisson distribution with a mean of two (Figure 2.3d). Thus, on average, the females produced two offspring each. The death rate was set to counter the now doubled population size in one of two ways. When generations were non-overlapping, all adults (50% of the population, on average) were "killed." If the generations were overlapping, i.e., adults could coexist with juveniles, 50% of all individuals were randomly chosen and "killed."

I varied movement probabilities from 0 (complete population isolation) to 0.2 across simulations. Preliminary analyses showed that populations were essentially panmictic at movement rates above 0.2. The number of adults moving between the two populations was on average the same (symmetrical movement). Thus, in this closed system, movement did not result in any net change in population size.

Each individual was initialized with two sets of the same number of loci (Figure 2.3b describes one set). The alleles of the first set of loci could not mutate at reproduction. The alleles of the second could mutate based on the probabilities in Table 3.2 (Dallas 1992; Weber & Wong 1993; Jarne & Lagoda 1996). F_{ST}, G_{ST}, G_{ST}', and Θ were calculated based on the non-mutating set. The genetic assignment tests were calculated based on both the mutating and non-mutating sets of loci. Private alleles and R_{ST} were calculated based on the mutating set. The numbers of alleles and loci in the simulations were based on empirical studies, and ranged from a one locus-two allele model to genotypes with 15 loci and 10 alleles. For the methods based on mutating loci I included up to 40 alleles per locus at initialization. Each allele was characterized by a numerical value (e.g., if there were 10 possible alleles, the allele types were numbered 1 to 10). At initiation, the allele types were in equal proportions, on average.

To simulate the microsatellite loci on which R_{ST} was calculated I considered the numerical value describing each allele type (see previous paragraph) to be the number of tandem repeats (e.g., the qualitative designation of allele type 6 above would mean there were 6 tandem repeats at that locus). Mutation was simulated for these microsatellite loci

using a single step model (Kimura & Ohta 1978; reviewed in Freimer & Slatkin 1996) where one repeat was lost or gained with equal probability for each mutation event.

Since population size affects the rate at which isolated populations are expected to reach equilibrium between drift and gene flow (Crow & Aoki 1984), I also varied initial population sizes. I set the duration of the simulations at 100 generations from isolation, to represent recently isolated populations.

I calculated the genetic information based on sample sizes from two to 20 percent of the populations, and also included 100 percent sampling for comparison. The sampling scenarios were independent, meaning the results for different sample sizes were from separate simulations, not sub-samples from the same group of individuals.

Model output

At the end of each generation, and for each parameter combination, I calculated: 1) the actual population sizes, 2) the actual amount of inter-population movement (referred to as "actual movement" from here on), 3) the genetic measure (F_{ST}, G_{ST}, G_{ST}', Θ, R_{ST}, the frequency of private alleles, and the two versions of the genetic assignment test, 4) the estimated amount of inter-population movement based on each of the genetic measures (see *Estimation of movement from genetic methods* above), 5) the difference between the estimated and actual movement (bias), and 6) bias divided by actual movement (standardized bias) (Table 3.2). Standardized bias adjusted for the fact that the actual number of individuals moving increased with movement rate and also depended on the population size. If estimated movement and actual movement were equal for all 100 simulations (no bias), the bias (and standardized bias) would equal 0. If estimated

movement was twice the actual movement, the standardized bias would be 1. If estimated movement was half of the actual movement, the standardized bias was -0.5. The standardized bias was negative if actual movement was underestimated, and positive if it was overestimated. When actual movement was 0, the standardized bias was undefined. I therefore excluded movement rates of 0 from analyses of standardized bias. I considered estimates of movement to be reliable if both (1) the average bias was between -0.5 and 1 (i.e., estimated movement was no less than half and no more than twice the actual movement) and (2) the range in bias over the 100 repeats of the simulations was between -0.6 and 1.5 (i.e., estimated movement between two and half times smaller and two and half times larger than actual movement). I considered these criteria reasonable given that my results are intended to help an empirical researcher determine whether or not to use genetic information to estimate inter-population movement.

Results

Which measure is better and when?

No one genetic measure produced reliable estimates of movement over all parameter combinations. Overall, the most reliable estimates of actual movement of the 8 methods were those derived from F_{ST}. F_{ST} produced reliable estimates of movement in about 10% of the parameter space. In comparison, Θ was reliable in 1.2% of the parameter space and G_{ST} was reliable in only 0.11% of the parameter space. Results based on G_{ST}' were indistinguishable from those based on Θ. Estimates based on R_{ST}, ,private alleles, and the 1995 version of the genetic assignment test did not produce reliable estimates in any of the parameter space where less than 100% of the population

was sampled. Some reliable estimates based on sampled individuals were produced using the 2004 version of the assignment test but these occurred in only 0.02% of the parameter space.

F_{ST}. Estimates of actual movement based on F_{ST} were consistently reliable for simulations with an intermediate range of actual movement rates (0.04-0.08 per generations). There were many interactions among the parameters. In particular, bias was consistently low for the largest range in movement rates (0.03-0.12) and lowest number of generations since isolation (10), when generations were overlapping, population size was large (2,000), at least 20% of the population sampled, and 15 loci with 10 alleles per locus were used. With decreasing sample sizes, numbers of alleles or loci, reliable estimates were produced for narrower ranges in movement rate and a longer time since isolation. For example, with the same parameter combination as above but with a percentage sampled of 12%, reliable estimates were produced for a range of movement rates of 0.03 to 0.06 instead of up to 0.12.

G_{ST}. Estimates based on G_{ST} overestimated movement in the extreme over most of the parameter space (Figure 3.1). Bias was consistently lower for movement estimates derived from G_{ST} for simulations based on non-overlapping generations, at least 20% of the population sampled, 15 loci, 10 alleles, a movement rate of 0.01 to 0.03, and 100 generations since isolation. This relatively narrow range of combinations makes estimates of actual movement based on G_{ST} generally unreliable.

Θ and G_{ST}'. Bias was consistently lower for simulations based on non-overlapping generations, at least 20% of the population sampled, 10 loci, 6 alleles, a

movement rate of 0.01 to 0.04, and 90 generations since isolation. However, as with the F_{ST}-based estimates, there were many interactions among the parameters. For example, if the percentage sampled was less than 16%, then at least 15 loci with 10 alleles were required to produce reliable movement estimates. If generations overlapped, then reliable estimates of movement were only produced for population sizes of at least 2,000. In general, if one parameter was not at its "best" value (e.g., maximum percent sampled, most loci, alleles and generations since isolation, non-overlapping generations), the other parameters needed to be at, or close to, their "best" to produce a reliable estimate of actual movement.

R_{ST}. Estimates of movement based on R_{ST} were not reliable; they overestimated movement in the extreme over most of the parameter space (Figures 3.2a and 3.3a). The variation in bias was also very high over the replicate simulations (Figure 3.3a). Furthermore, the degree of overestimation varied with movement rate and the percentage of the population sampled (Figure 3.2a).

Private alleles. Estimates of movement based on the frequency of private alleles were not reliable; they underestimated movement (Figures 3.2b and 3.3b). For many combinations, there were few private alleles present. Private alleles were absent in over 50% of the replicate runs for simulations based on fewer than 6 alleles, 50 generations since isolation and for movement rates above 0.05. Even when present, for combinations when the whole population was sampled the average frequency of private alleles over the 100 simulation was less than 0.195, with the majority of frequencies falling below 0.03.

The degree of underestimation increased with decreasing percentage of the population sampled (Figure 3.2b).

Genetic Assignment Tests. Estimates of movement based on the 1995 and 2004 versions of the genetic assignment tests were not reliable. Estimates based on the 1995 version overestimated movement (Figures 3.4a and 3.5a). Reliability varied with movement rate and the percentage of the population sampled (Figures 3.4a and 3.5a). The 2004 version of the test over-estimated movement at lower movement rates (0-0.04) and under-estimated movement at higher rates (> 0.04) (Figures 3.4b and 3.5b). There were a limited number of combinations that fit the reliability criteria for the movement rates where the transition between over and under-estimates occurred. For simulations based on a population size of 2000, at least 20% of the population sampled, 15 loci, 25 loci with a mutation rate of 10^{-5}, 25 alleles, and 20 generations since isolation estimates of movement were reliable but only over a very narrow range of movement rates from 0.05 to 0.06. Practically speaking, this transition in the reliability for such a small number of combinations makes estimates of actual movement based this version of the genetic assignment test generally unreliable.

Relationship between estimated movement and actual movement

Overall there was a positive relationship between actual inter-population movement and that estimated from the various genetic measures (Figures 3.1, 3.2, and 3.4) over many of the parameter combinations. However, none of the genetically-based measures of movement were accurate *quantitative* estimators of movement. The estimates failed to quantify actual movement directly because the relationship was non-

linear, too variable, and/or extremely biased depending on the measure used and the movement rate itself (Figures 3.3 and 3.5). While many parameter combinations produced reliable estimates of movement based on F_{ST}, bias varied with movement. Estimates of movement were more variable at very low movement rates than at higher rates; underestimates increased with movement rate (Figure 3.6).

Discussion

F_{ST} provided reliable estimates of movement over more parameter combinations than any other genetic measure studied. These combinations complemented those based on Θ and G_{ST}' by providing reliable estimates at higher movement rates (up to 0.12, depending on the other parameter values) and for more combinations when generations overlapped. Θ and G_{ST}' performed better at very low movement rates of 0.01 to 0.04. At these rates, estimates based on F_{ST} were only reliable after a larger number of generations since isolation. Estimates based on R_{ST}, the frequency of private alleles, and the genetic assignment tests were not reliable at all.

The higher reliability of movement based on F_{ST}, relative to estimates based on its derivatives, G_{ST}, G_{ST}', and Θ, was unexpected. The more recent measures (G_{ST}, G_{ST}', Θ) include additional terms to address the number of observed heterozygotes, sample sizes, and the number of populations, all intended to improve accuracy of the measure. At very low movement rates (≤ 0.04), Θ and G_{ST}' did indeed produce more reliable estimates of movement than F_{ST}, consistent with the simulation results in Weir and Cockerham (1984). These measures would be therefore more reliable in situations where the actual movement rate is thought to be extremely low, as is often the case in research on

evolutionary questions. However, at intermediate movement rates (0.04 to 0.12), the estimates of movement based on F_{ST} were more reliable than the estimates based on its derivatives, G_{ST}, G_{ST}', and Θ. The reason for this difference was that these measures approached 0 faster with increased movement rate than did F_{ST}. Values of F_{ST} and its derivatives near zero produce extreme over-estimates of Nm because of the non-linear relationship based on equation 3.2 above (see Slatkin & Barton 1989; Waples 1998; Whitlock & McCauley 1999; Neigel 2002). Since F_{ST} remains farther from zero at higher movement rates than its derivatives, it produces more reliable estimates of movement rate at these higher values. For many ecological questions about effects of human-caused movement barriers such as habitat destruction or roads, movement rates are likely to be higher than for evolutionary questions. In this case the original F_{ST} measure will provide a more reliable estimate of movement. This does create a catch-22 that, in order know which or whether any of these methods should be used to estimate movement, a researcher would first require some a priori knowledge of movement rate.

The use of microsatellite loci-based estimates of movement has been increasing over the last decade (Jarne & Lagoda 1996; Balloux & Lugon-Moulin 2002; Pearse & Crandall 2004). The high variability and mutation rate for microsatellites are thought to make them more suitable for detecting differentiation among populations in ecological time, as they are better able to respond quickly to recent isolation. However, this expectation was not supported by my simulations. Both R_{ST} and the genetic assignment tests provided poor estimates of movement.

The high variability in R_{ST} for most of the parameter combinations resulted in a large number of values near or equal to 0, which, when converted to *Nm*, overestimated movement in the extreme based on equation 3.4. The values of R_{ST} were typically lower than those of F_{ST}, G_{ST}, G_{ST}', and Θ, especially at lower movement rates. When populations were recently isolated, R_{ST} did not adequately reflect the growing differentiation between the populations.

In the case of the 1995 version of the genetic assignment test, the calculated likelihoods for the individuals were often similar, especially at low movement rates; in many cases only slight changes would produce an opposite classification. The 2004 version of the assignment test was designed to address this artefact by only classifying individuals as migrants if they were more likely to be from another population based on a statistically significant difference in the likelihoods. This adjustment seems to have over-corrected the classification. Even using a less conservative α of 0.05 did not significantly change this conclusion. In practical terms, my results suggest that while there is a qualitative association between estimates based on these genetic assignment tests and actual movement, these methods do not provide a reliable quantitative estimate. Furthermore, my results can probably be extrapolated to other assignment tests (Rannala & Mountain 1997; Pritchard et al. 2000) as they have been shown to have comparable performance (Cornuet et al. 1999; Guinand et al. 2002). These methods have been shown in empirical studies to over-estimate movement rates (e.g., Berry et al. 2004).

Estimates of movement based on private alleles were expected to slightly underestimate movement but still be as reliable as those based on F_{ST} (Slatkin & Barton

1989). However, I found that the estimates were highly negatively biased. Slatkin (1985b) showed that for low frequencies of private alleles (approximately ≤ 0.03) the method of calculating of *Nm* (see equation 3.5 above and Figure 1 in Slatkin 1985b) was not appropriate. In Slatkin's simulations, the curve plotting the frequency of private alleles vs. *Nm* levels off for higher values of *Nm*. *Nm* calculated for this range in frequency of private alleles will underestimate actual movement. This was generally the range in frequency of private alleles produced from my simulations. Thus, for parameter combinations typical of recently isolated populations, the frequency of private alleles is generally too low to reliably reflect actual movement.

My model included simplifying assumptions. I assumed gene flow and movement were equivalent, as residents were equally likely to mate as immigrants. The system was closed with movement taking place only between the two populations. Generally, if I relaxed the above mentioned assumptions or included additional effects (e.g., uneven sample sizes), the main difference in my results would have been to further de-couple the genetically-based estimate of movement from actual movement and to increase bias and variability, thus further compromising the estimate as a measure of inter-population movement.

Implications and conclusions

Of the eight genetic measures examined, F_{ST} produced the most reliable overall estimates of movement, and Θ or G_{ST}' produced the most reliable estimates when movement rate was very low. In practical terms this suggests that these methods are the most appropriate for estimating inter-population movement in the field. Estimates of

movement based on F_{ST} and Θ or G_{ST}' were consistently reliable for simulations based on at least 20% of the population sampled, at least 10 loci with at least 6 alleles per locus, non-overlapping generations, at least 90 generations since isolation, and for a movement rate of less than 0.04 per generation for Θ or G_{ST}' or for a movement range of 0.04 to 0.08 for F_{ST}. Unfortunately this creates a Catch-22 situation for the researcher, as one must have at least a rough estimate of movement rate to determine which (if any) measure will reliably quantify movement.

In practical terms, it is generally unlikely that there is an appropriate genetic method for estimating movement in any real situation. While the above situations produced reliable estimates of movement based on F_{ST} and Θ or G_{ST}', real studies, particularly studies of endangered species, are likely to have low sample sizes, recently isolated populations, and overlapping generations. Furthermore, methods thought to work better in these conservation studies by making fewer assumptions or using more variable loci, i.e., R_{ST} or the two versions of the genetic assignment test, do not reliably quantify movement at all.

Table 3.1. Assumptions made for each of the genetic models.

Assumptions made	How addressed in my model
Island Model (Wright 1931) assumed for **F_{ST}, G_{ST}, G_{ST}', Θ, R_{ST},** ***and private alleles:***	
genetic equilibrium between drift and gene flow	pattern examined over 100 generations, approaching equilibrium
generations do not overlap	pattern compared for overlapping vs. non-overlapping generations
no mutation	assumed for F_{ST}, G_{ST}, G_{ST}', and Θ; pattern compared for 2 mutation rates (see Table 3.2) for R_{ST}, private alleles, and the genetic assignment test
	assumed but subject to stochastic variation
demographic equilibrium	assumed but subject to stochastic variation
gene flow = migration	assumed
neutral to selection	assumed
migration random with respect to alleles	assumed
linkage equilibrium	
Genetic assignment test (Paetkau et al. 1995; Waser & Strobeck 1998; Paetkau et al. 2004):	
populations at Hardy-Weinberg equilibrium	assumed at equilibrium but subject to stochastic variation
	assumed
linkage equilibrium	

Table 3.2. Parameters and values used in the simulation experiments.

INPUT	
Parameters	**Value**
Held constant for all simulations:	
Number of offspring per female:	2 (drawn from Poisson distribution)
Death rate:	0.5 (individuals chosen at random)
Sex ratio:	1:1
Number of populations:	2
Number of repetitions of simulation	100
Duration of simulation:	100 generations from isolation
Varied among simulations:	
Initial population sizes:	500, 1000, 2000
Sample percentages:	4, 6, 8, 10, 12, 16, 20, 100%
Number of loci:	1, 3, 5, 10, 15
Number of alleles:	2, 6, 10 for all measures 20, 25, 30, 40 for R_{ST}, private alleles, and genetic assignment test
Mutation rate:	10^{-2}, 10^{-5} for R_{ST}, private alleles, and genetic assignment test
Movement probability:	0 to 0.2 by 0.01
Overlapping generations:	no/yes
OUTPUT	
Produced for each input parameter combination: Population sizes Actual movement Genetic measures Genetically-based estimates of movement from each genetic measure Bias *(estimate - actual movement)* for each genetic measure Standardized bias *((estimate - actual movement)/actual movement)* for each genetic measure	The following output was reported for each of these parameter combinations and represents the descriptive statistic over the 100 repeated simulations: • average • minimum • maximum • variance • standard error • 95% confidence interval

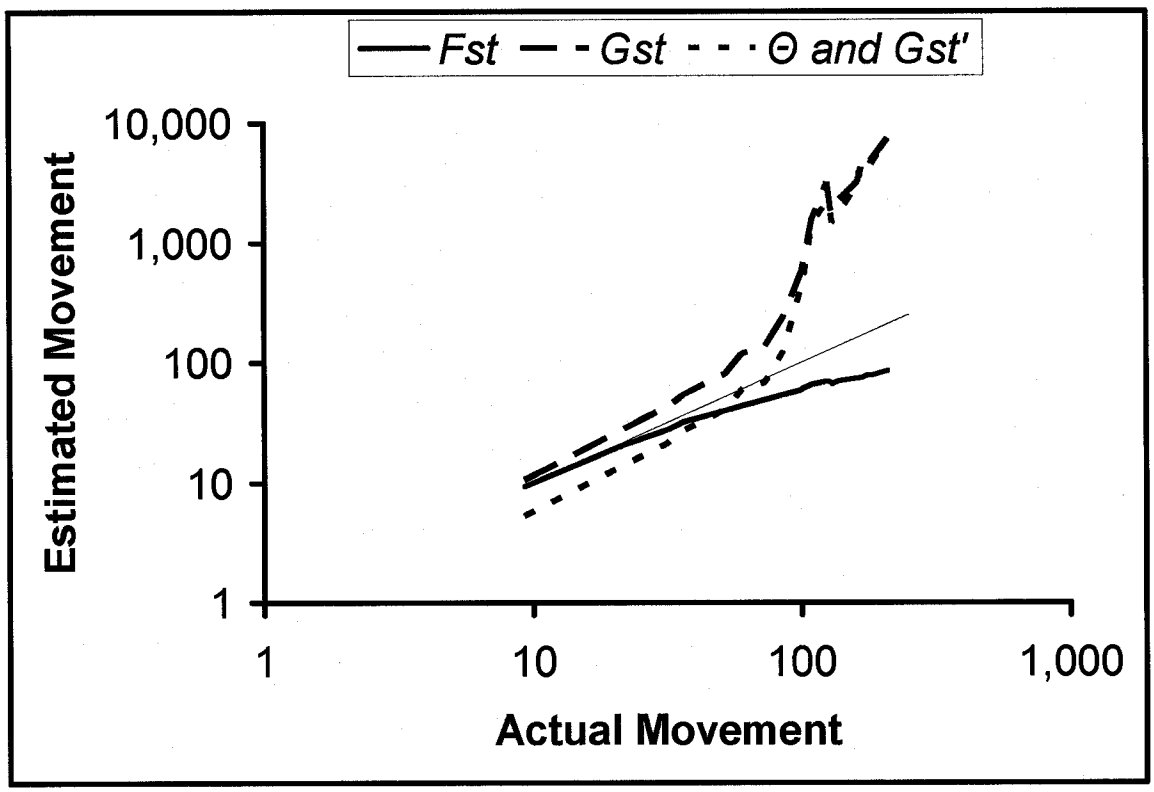


Figure 3.1. Plot of the relationship between estimated movement based on F_{ST}, G_{ST}, G_{ST}', and Θ vs. actual movement. The values represent the averages over 100 replicate simulations. The 1:1 line is the solid light line. The parameter combination depicted was for population sizes of 500, 20% of the population sampled, 15 loci, 10 alleles, non-overlapping generations, and 100 generations since isolation.

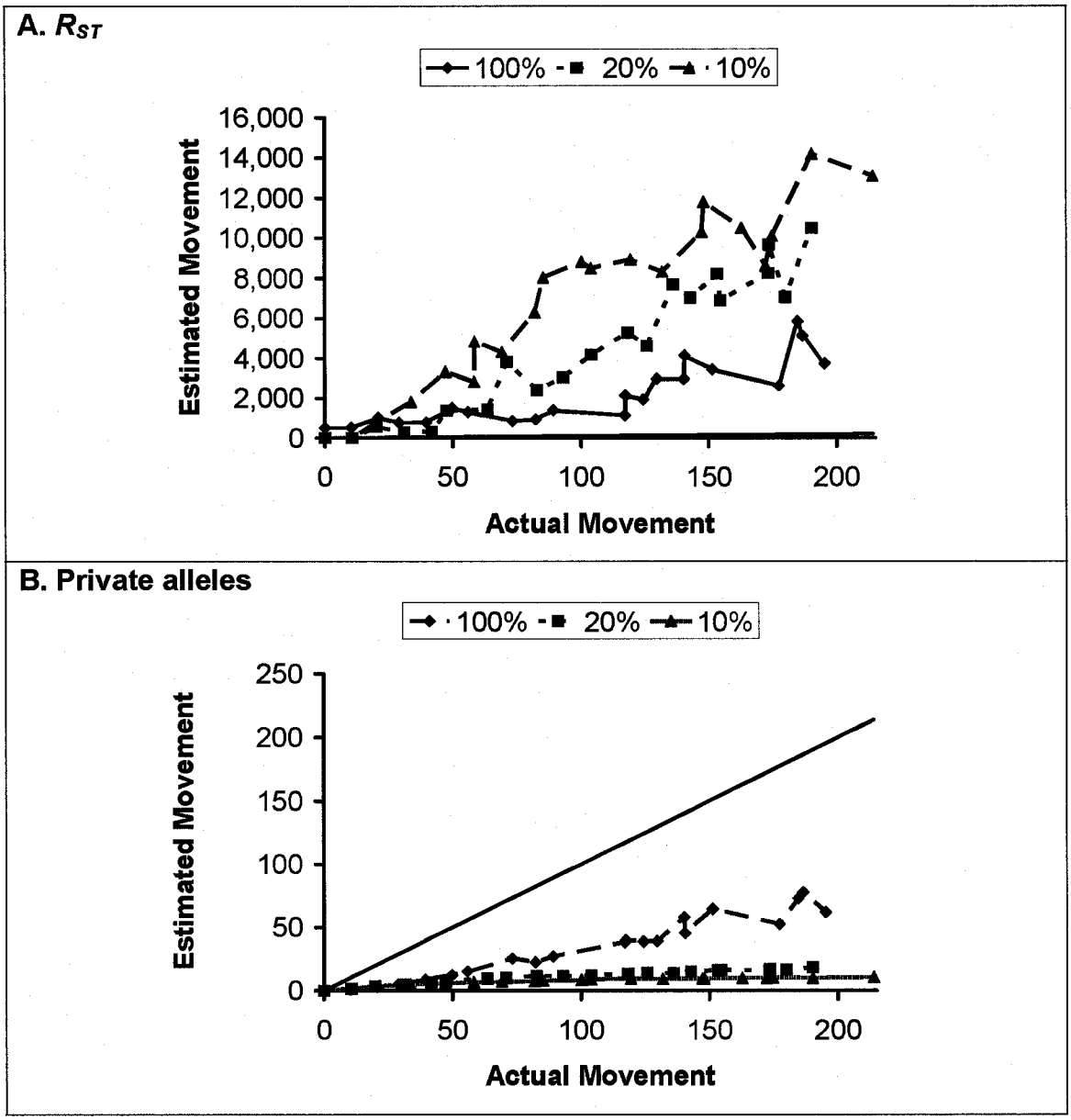


Figure 3.2. Plot of the relationship between estimated movement vs. actual movement for 10, 20 and 100% percent of the population sampled with estimated movement based on (A) R_{ST} and (B) the frequency of private alleles (note: the y-axis scale changes). The 1:1 line (solid dark line) is shown for comparison. For both plots shown the parameter combination depicted was the same as for Figure 3.1 but for 40 alleles and a mutation rate of 10^{-2}.

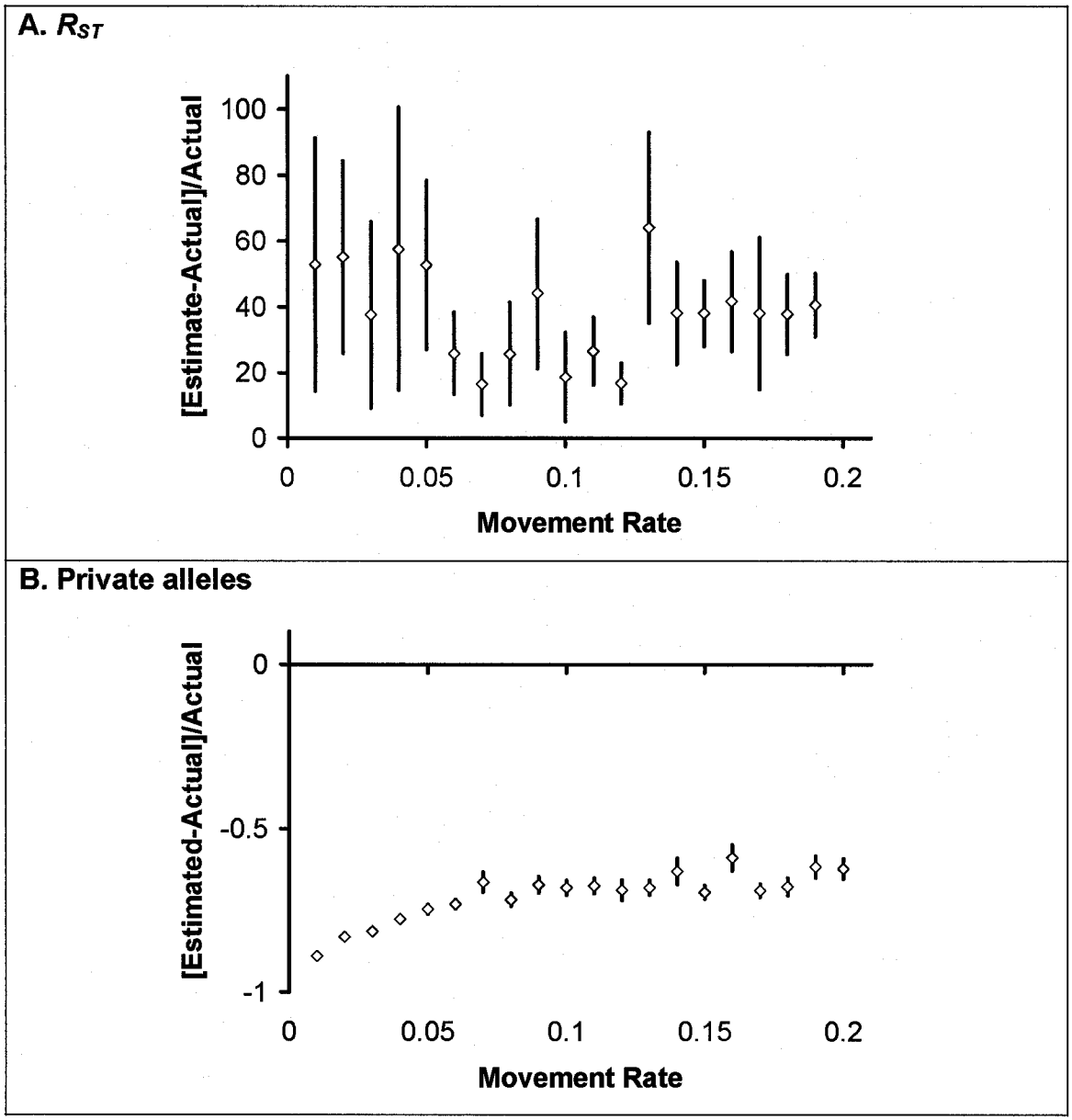


Figure 3.3. Standardized bias (± 1SE) vs. the actual movement rate for estimates based on (A) R_{ST} and (B) the frequency of private alleles (note: the y-axis scale changes). For both plots shown all simulations were based on all individuals sampled and the same parameter combinations depicted in Figure 3.2.

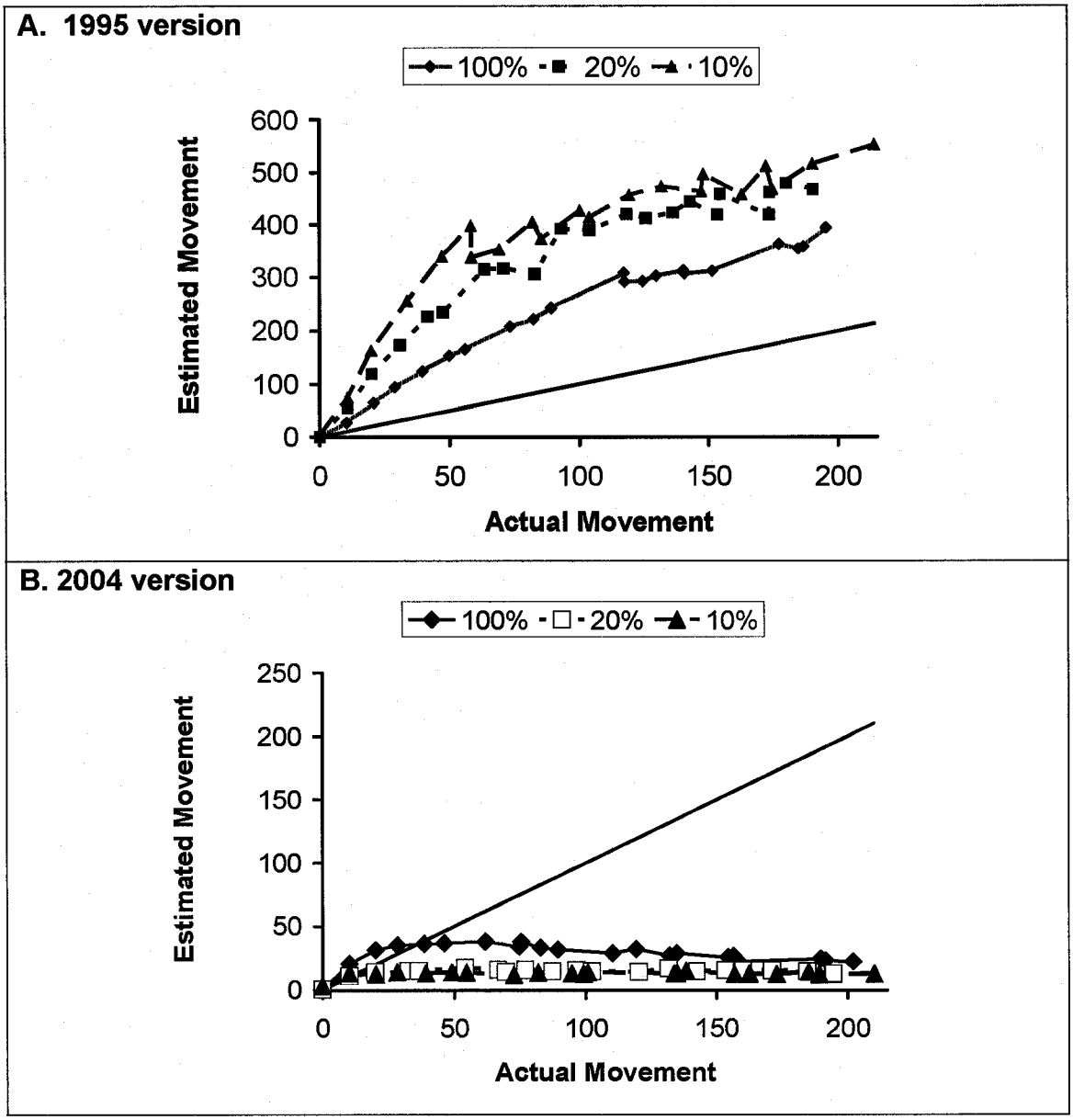


Figure 3.4. Plot of the relationship between estimated movement vs. actual movement for 10, 20 and 100% percent of the population sampled with estimated movement based on (A) the 1995 version of the genetic assignment test (Paetkau et al. 1995) and (B) the 2004 version (Paetkau et al. 2004) (note: the y-axis scale changes). The 1:1 line (solid dark line) is shown for comparison. For both plots shown the parameter combination depicted was the same as for Figure 3.1 but for 40 alleles and a mutation rate of 10^{-2}.

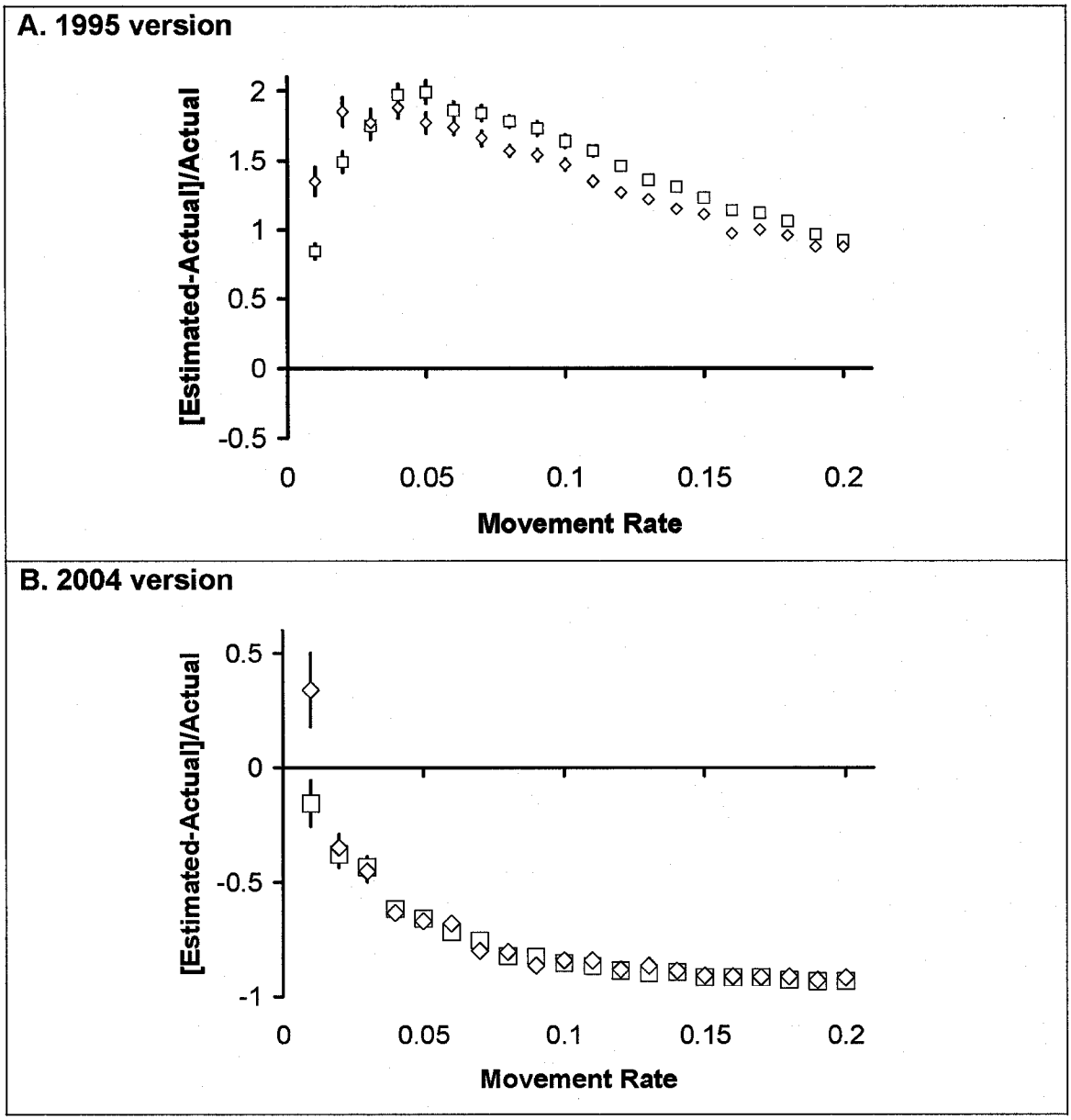


Figure 3.5. Standardized bias (± 1SE) vs. the actual movement rate for estimates based on (A) the 1995 version of the genetic assignment test (based on non-mutating loci [squares] and mutating loci [diamonds]) (Paetkau et al. 1995) and (B) the 2004 version (Paetkau et al 2004) (note: the y-axis scale changes). For both plots shown all simulations were based on all individuals sampled and the same parameter combinations depicted in Figure 3.3.

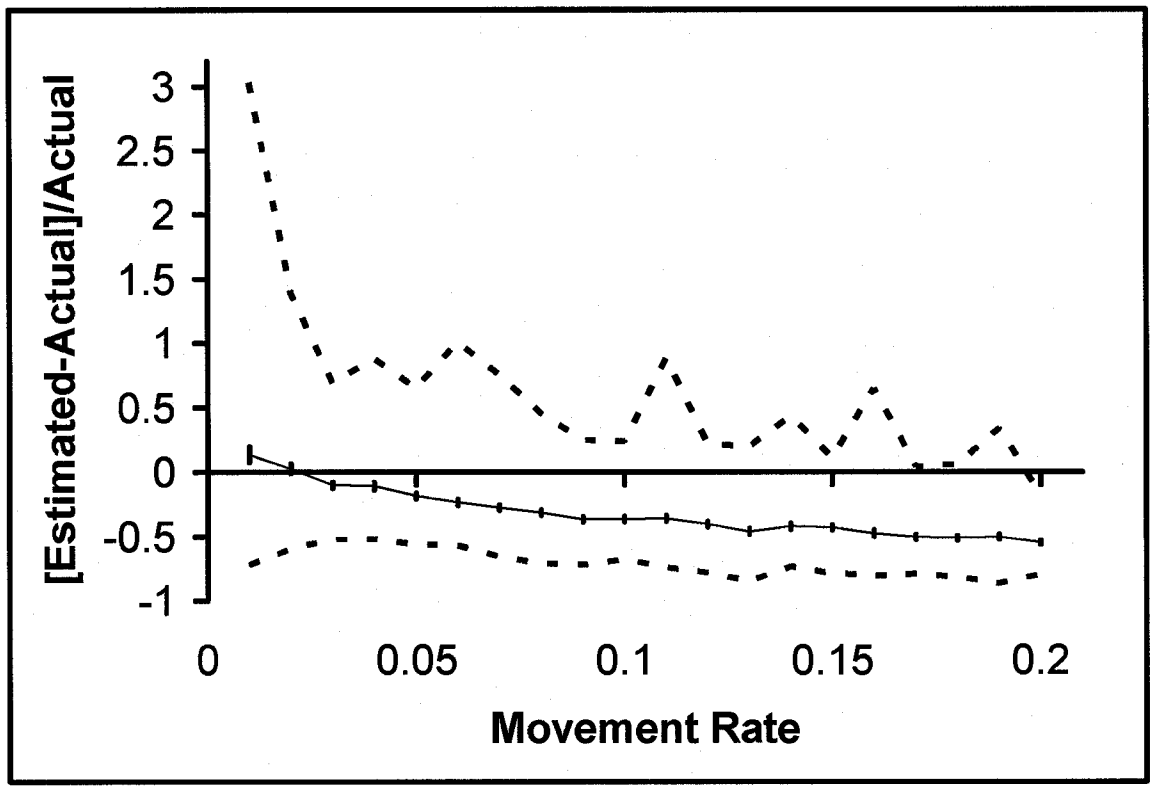


Figure 3.6. Standardized bias (± 1SE) vs. the actual movement rate for estimates based on F_{ST}. The dashed line represents the minimum and maximum over the 100 replicate simulations.

CHAPTER 4: USING GENETIC INFORMATION TO MEASURE INTER-POPULATION MOVEMENT: RECONCILING MODELS AND EMPIRICAL DATA

Chapter summary

In Chapters 2 and 3 I asked which, if any, of F_{ST}, G_{ST}, Θ, R_{ST}, private alleles, and two versions of the genetic assignment test was suitable for calculating inter-population movement of recently isolated small populations using a simulation model. The current study is a partial test of the predictions generated from the simulations. I compared the difference (bias) between the direct and genetically-based estimates of movement in the empirical studies to those predicted by the corresponding set of simulations. The results showed partial validation for the model but were limited by the fact that there was little replication in the empirical studies. Seventy percent of the standardized empirical biases were within 1 standard deviation of the predicted mean of the simulations. As predicted, estimates of movement based on F_{ST} and Θ were more reliable than those based on R_{ST}, private alleles, and the genetic assignment test (Paetkau et al. 1995 version), although not significantly so. As predicted, Θ-based estimates were more reliable when the direct estimate of movement was less than 0.04. Estimates based on the frequency of private alleles underestimated movement and those based on R_{ST} generally overestimated it, also as predicted. Estimates based on the genetic assignment test underestimated movement, contrary to the model predictions. Finally, bias based on F_{ST} and Θ decreased with increased percentage sampled, number of alleles, and for populations with non-overlapping generations, as predicted. However, the strength and direction of this relationship depended on the actual movement rate. The partial support of the model

predictions suggests that the model results may be reliable, especially given the limitations in the empirical data set. More empirical studies are needed that include both direct and genetically-based quantitative estimates of movement.

Introduction

Measuring inter-population movement accurately is often challenging. Researchers have used both direct and genetically-based (indirect) methods to quantify movement (Slatkin 1985a). Direct measures estimate inter-population movement based on observing the movements of individuals (e.g., based on capture-mark-recapture). Indirect measures infer movement based on the genetic makeup of populations. Both types of measures have advantages and disadvantages (reviewed in Slatkin 1985a; Koenig et al. 1996; Neigel 2002) but indirect estimates based on genetic measures have been growing in popularity because genetic information is often less costly to obtain and results in less disturbance of organisms than marking and recapturing individuals.

Over the last several years, empirical researchers have used a variety of genetic methods to estimate movement, with each one assumed to improve upon the last. F_{ST} is based on Nei's (1973) extension of the formula introduced by Wright (1951) to address more than two alleles per locus. G_{ST} includes additional terms that consider the number of observed heterozygotes and the sample size, and G_{ST}' is also independent of the number of samples (Nei 1987). Θ (Weir & Cockerham 1984) includes terms that address the effects of multiple alleles and loci, the number of populations, sample sizes (both even and uneven), and observed heterozygosity. R_{ST} is based on the number of tandem repeats and is applicable to microsatellite loci, which tend to have a higher mutation rate (Slatkin

1995). Slatkin (1985b) also proposed a method based on the frequency of private alleles (those found in only one population). The genetic assignment test (Paetkau et al. 1995, Waser & Strobeck 1998, Paetkau et al. 2004)) makes fewer assumptions in the calculation of movement from genetic information. Use of these methods in empirical studies of movement has changed with time (Figure 1.1). Θ is currently the most commonly used estimates of genetic differentiation in the literature. However, use of the genetic assignment tests and R_{ST} has been increasing steadily since their introduction.

There is much controversy in the literature, however, about the validity of genetic methods in measuring movement, leading to cautionary statements about their use (Lewontin 1985; Slatkin 1985b; Crochet, 1996; Bohonak et al 1998; Bossart & Pashley Prowell, 1998a and b; Steinberg & Jordan, 1998; Whitlock and McCauley 1999; Bohonak and Roderick 2001; Neigel 2002). This is particularly true for populations that are of interest for conservation. Yet, with advances in molecular genetics (for reviews, see Mitton 1994; Parker et al. 1998; Manel et al. 2003) and refinements of the genetic methods (e.g., Paetkau et al 1995; Rannala & Mountain 1997; Waser & Strobeck 1998; Pritchard et al. 2000; Peatkau et al. 2004), researchers continue to use genetic information to measure movement.

Researchers who have used genetic methodologies to characterize the amount of movement between populations often relate their findings to plausible movement scenarios. They sometimes look for consistency with direct measures of individual movement but more often they compare their findings with expected movement inferred from the spatial arrangement of populations (e.g., distance, barriers) or from an

organism's behaviour or anatomy. The degree of consistency between the genetic and direct measure of movement varies in part because each captures different aspects of movement (Slatkin 1985a; Koenig et al. 1996; Bossart & Pashley Prowell, 1998a) and because it is potentially affected by other factors such as the genetic method used or sample size (Chapters 2 and 3).

Previously I showed that genetic methodologies generally provided poor quantitative estimates of inter-population movement based on my simulation model of the demography and the genetics of hypothetical populations. In the demographic and genetic scenarios simulated, estimates based on R_{ST} (Slatkin, 1995) and the genetic assignment test (Paetkau et al. 1995; Waser and Strobeck 1998) vastly overestimated actual movement. Estimates based on private alleles (Slatkin 1985b) greatly underestimated movement. In general, F_{ST} (Nei 1973) outperformed G_{ST} (Nei 1987), G_{ST}' (Nei 1987), and Θ (Weir and Cockerham 1984), especially at intermediate movement rates (0.04 to 0.08 per generation). Θ and G_{ST}'-based estimates of movement were more reasonable at lower movement rates per generation (less than 0.04).

The goal of the previous simulation study was to predict when genetically-based estimates could be used to measure movement over a broad set of demographic and genetic scenarios. The point of the current study is to test the predictions from the model using empirical comparisons of direct and genetically-based measures of movement. Based on my previous findings I make three predictions about the difference between the direct and genetically-based estimates of movement in the empirical studies: (1) estimates of movement based on F_{ST} and Θ will be more reliable than those based on the

other measures; further, the reliability will interact with movement rate itself; (2) estimates based on the frequency of private alleles will tend to underestimate movement as compared to those based on R_{ST} and the genetic assignment test, which will overestimate movement (Peatkau et al. 1995 version); estimates based on the later version of the genetic assignment test (Paetkau et al. 2004) will vary with movement rate, overestimating it a low rates and underestimating it at higher rates (3) the difference between the direct and the estimates of movement based on F_{ST} and Θ will be lower in empirical studies where a higher percentage of the population is sampled, populations have non-overlapping generations, and when more loci with more alleles are sampled. Because systematic comparisons are rare in the empirical literature, the test is limited to a small disparate subset of the comprehensive set of circumstances evaluated in the model in Chapters 2 and 3.

Methods

Empirical information

I searched the *Web of Science* database for all studies published between 1973 and 2006 that measured inter-population movement based on both genetic and direct measures. The latter was based on either mark-recapture/re-sight or on the tracking of individuals. Total inter-population movement per generation was based on the rate per generation calculated from the empirical sample multiplied by population size. The genetically-based estimates of movement were derived from one or more of the following methods: F_{ST}, G_{ST}, Θ, R_{ST}, private alleles, and/or the 1995 version of the genetic assignment test. For estimates based on F_{ST}, G_{ST}, Θ, R_{ST}, or private alleles, Nm

represented the number of individuals moving between populations per generation. N was the effective population size and m was the proportion of the population migrating per generation. For estimates based on the 1995 version of the genetic assignment test, the individuals assigned to a population other than the one from which they were sampled were considered immigrants. I used the 1995 version of the model as this was the one used in the associated empirical study. The genetic and direct information was sometimes found in separate articles (Table 4.2). I only included studies where this information was collected from the same populations.

For each empirical study, I conducted a corresponding set of simulations. Each simulation provided a prediction of the expected difference between the direct estimate of movement and that measured based on the genetic information. When movement was different from zero, I also standardized this difference by dividing it by the simulated direct estimate of movement. This adjusted for the fact that the actual number of individuals moving increased with movement rate and also depended on the population size. I then compared this expected difference to the difference between movement measured directly and based on the genetic information reported in the empirical study. I refer to these differences as simulation bias and empirical bias respectively (or standardized simulation bias and standardized empirical bias) from here onward. Each comparison represented an individual test case of the model predictions. All statistical analyses were conducted using *SPSS* (SPSS Inc. 2004) and *Gpower* (Erdfelder et al. 1996).

To test the three predictions I outlined in the introduction I examined the direction and size of the standardized empirical bias (i.e., [genetic estimate – direct estimate of movement] / direct estimate of movement) for all cases where the direct estimate of movement was above 0. When the direct estimate of movement was 0, I analyzed the only the size of the empirical bias (i.e., genetic estimate – direct estimate of movement) because in these cases bias could only be positive. Standardized bias was undefined if the direct estimate of movement was 0.

Model information

The model was a stochastic individual-based simulation model with a hierarchical structure. It simulated the demography and the genetics of two-hypothetical populations. The values of these simulated demographic and genetic parameters were based on those reported in the empirical study (Table 4.1). A detailed description of the model is included elsewhere (*Methods* in Chapters 2 and 3).

Actual movement rate was based on the rate determined from the direct measure of movement in the empirical study. Movement between populations could be either symmetrical or asymmetrical as indicated by the direct measure. In cases where only the number of individuals moving was provided, I calculated the rate as the number of individuals observed moving between a particular pair of populations divided by the sample size. Population sizes and sample sizes were set to reflect those provided in the empirical studies. If population sizes were not provided I approximated the values based on other available information such as density.

I did not vary the number of offspring per female, death rate, or sex ratio among simulations (Table 4.1). While I could potentially vary these, in the absence of information from the empirical studies I decided to use default values that would maintain the average population size at that provided in the empirical study. At initialization the sex ratio was on average 1:1. The mating system was polygynous. Immigrants and residents had the same probability of mating. The number of offspring produced was drawn from a Poisson distribution with a mean of two. Thus, on average, the females produced two offspring each. The death rate was set to counter the now doubled population size in one of two ways. When generations were non-overlapping, all adults (50% of the population, on average) were "killed." If the generations were overlapping, i.e., adults could coexist with juveniles, 50% of all individuals were randomly chosen and "killed." If the populations were recently isolated I ran the simulation for the reported number of generations since isolation. If not, I set the duration of the simulations at 100 generations. Preliminary analyses suggested that the difference between actual and estimated movement (simulation bias) would not vary significantly after this many generations.

The numbers of loci and alleles in the simulations were based on those reported in the empirical studies. The mutation rate and model for microsatellite loci was not provided in any of the empirical studies. I ran simulations based on a rate of 10^{-5} (Table 4.1; Dallas 1992; Weber and Wong 1993; Jarne and Lagoda 1996). Mutation was simulated using a single step model (Kimura & Ohta 1978; reviewed in Freimer &

Slatkin 1996) where one repeat was lost or gained with equal probability for each mutation event.

I calculated simulation bias and standardized simulation bias (if movement rate differed from 0) for each parameter combination from the empirical study (Table 4.1). I ran the simulation 100 times for each combination, thus generating a mean, range (minimum and maximum), and standard deviation for each of these two values (Table 4.1). Each simulation represented an individual pair of populations from the empirical studies except when a study only presented global values of the genetic estimate of movement over all populations. In these cases I simulated a pair of populations with the global movement rate.

Results

Out of several hundred potentially relevant empirical papers only 20 included both genetic and direct quantitative estimates of movement for the same populations (Table 4.2). These studies covered a range of taxa and habitat types from around the world. Most used F_{ST} (7), Θ (8), or R_{ST} (4) as the basis for the genetic estimates of movement. The one empirical study using a genetic assignment test used the 1995 version (i.e., Paetkau et al. 1995). In several studies there were more than 2 populations and multiple methods applied (Table 4.2) resulting in 222 pairs of populations for which movement information was derived for all 20 studies combined. Of these, 178 had no movement according to the direct study, leaving 44 that had non-zero direct estimates of movement. These 44 represented only 11 of the 20 studies. Θ and F_{ST} were used more often than the other methods (34 and 5 of the 44 comparisons respectively).

There was a positive relationship between the mean standardized simulation bias and the standardized empirical bias (slope = 0.044). However, this relationship was not significant (r^2 = 0.003; Figure 4.1). With a sample size of only 44, the power of this relationship was also very low (0.063). The sample size that would have been needed to result in a power of 0.80 (default from Cohen 1977) was 2,619. However, most of the standardized empirical bias estimates (30/44 or 70%) did fall within 1 standard deviation of the predicted mean standardized simulation bias (G-test of independence, $G_{1,2}$ = 14.5, p = 0.001; Figure 4.2).

Estimates of movement based on F_{ST} and Θ were more reliable than those based on R_{ST}, private alleles and the genetic assignment test, as predicted. I considered standardized biases to be reliable if they ranged between -0.6 and 1.5, (see *Methods* in Chapters 2 and 3). Two of 5 and 16 of 34 cases were reliable for F_{ST} and Θ respectively. Only 1 of the 2 cases based on private alleles and none of the 3 based on R_{ST} (2 cases) and the genetic assignment test (1 case) were reliable. However, the difference in the number of reliable cases for F_{ST} and Θ (18 of 39) vs. the other 3 methods combined (1 of 5) was not significant (G-test of independence, $G_{1,1}$ = 1.34, p = 0.247). The power was low for this test (0.25); 212 cases would have been required to have a power of 0.80.

Empirical estimates of movement based on Θ were more reliable (i.e., standardized bias was between -0.6 and 1.5) at very low movement rates (< 0.04), as predicted from the simulation results in Chapters 2 and 3. There were significantly more reliable estimates when the direct estimate of movement rate was below 0.04 (13 of 20) than above 0.04 (3 of 13) ($G_{1,1}$ = 5.77, p = 0.016; Figure 4.3). When the direct estimate

of movement rate was 0, the bias in estimates of movement based on Θ ($n = 64$) was significantly lower than those based on the remaining measures ($F_{1,175} = 7.55, p = 0.007$; Figure 4.4), as predicted.

Both estimates based on private alleles underestimated movement, as predicted. The 3 cases based on R_{ST} (2 cases) and the genetic assignment test (1 case) also underestimated the direct movement, contrary to the model prediction. However, the estimates for the 2 cases based on R_{ST} were within 1 standard deviation of the predicted mean standardized simulation bias (Figure 4.2). Further, when the direct estimate of movement rate was 0, the mean overestimate of 25.1 based on R_{ST} ($n = 64$) was higher than for any method combined ($F_{1,175} = 10.26, p = 0.002$; Figure 4.4), even once I had removed the one *Nm* value of 25,000 (i.e., where $R_{ST} = 0$) for one of the 64 cases (Table 4.2).

I predicted that estimates of movement based on F_{ST} and Θ would be increasingly reliable with percentage of the population sampled, number of loci, and number of alleles. The dotted lines in Figures 4.5 and 4.6 show the predicted change in the range in bias with increasing percent sampled, number of loci and alleles. Negative bias values are expected to increase, and positive bias values decrease, with increasing percent sampled, number of loci and alleles. Standardized bias was lower when based on a higher percentage of the population sampled, as predicted. Positive standardized bias values decreased with percentage sampled (slope = -0.212) but not significantly ($F_{1,18} = 0.036, p = 0.851$, power = 0.054; Figure 4.5a). Negative standardized bias increased with percentage sampled (slope = 0.003) but not significantly ($F_{1,17} = 0.132, p = 0.721$, power

= 0.064; Figure 4.5a). Power was low for these 2 relationships; 3,927 and 984 cases respectively would have been required to result in a power of 0.80. However, when the direct estimate of movement rate was 0, bias actually increased slightly with percent sampled (slope = 0.144), but not significantly ($F_{1,107} = 3.563, p = 0.620$; Figure 4.6a).

Empirical estimates of movement based on F_{ST} and Θ were not significantly more reliable with increased number of loci. While negative standardized bias increased slightly (slope = 0.035; $F_{1,17} = 1.064, p = 0.317$, power = 0.164; sample size required: 128; Figure 4.5b) as predicted, positive standardized bias values actually increased (slope = 3.136), though not significantly ($F_{1,18} = 0.121, p = 0.732$; Figure 4.5b). When the direct estimate of movement was 0, the empirical bias increased with number of loci, but not significantly ($F_{1,107} = 0.312, p = 0.578$, Figure 4.6b).

As predicted, empirical bias decreased significantly when the genetic estimates of movement were based on loci with a larger number of alleles, as predicted (slope = -0.902; $F_{1,107} = 9.136$, p = 0.003, power = 0.850; sample size required: 87; Figure 4.6c). However, standardized bias increased with number of alleles, though not significantly; positive bias values increased (slope = 15.208; $F_{1,18} = 1.674, p = 0.212$; Figure 4.5c) and negative values decreased (slope = -0.023; $F_{1,17} = 2.509, p = 0.132$; Figure 4.5c).

Estimates of movement based on F_{ST} and Θ were more reliable for populations with non-overlapping generations than for populations with overlapping generations, as predicted. Standardized empirical bias was lower on average if generations did not overlap (Figure 4.5d), but not significantly ($F_{1,37} = 1.453, p = 0.236$, power = 0.217; required sample size: 398 for power = 0.80). When the direct estimate of movement was

0, empirical bias was significantly lower for cases where generations did not overlap ($F_{1,107} = 6.911, p = 0.01$; Figure 4.6d).

Discussion

The results partially validated the model. Seventy percent of the standardized empirical biases were within 1 standard deviation of the predicted mean of the simulations. As predicted, estimates of movement based on F_{ST} and Θ were more reliable than those based on R_{ST}, private alleles, and the 1995 version of the genetic assignment test, although not significantly. Θ-based estimates were more reliable when the direct estimate of movement was less than 0.04, as predicted. Overall, estimates based on the frequency of private alleles underestimated movement and those based on R_{ST} overestimated it, as predicted. Standardized bias based on F_{ST} and Θ decreased with increasing percent sampled and for populations with non-overlapping generations, although not significantly. When the direct estimate of movement was 0, estimates based on F_{ST} and Θ were significantly less biased with increasing number of alleles and for populations with non-overlapping generations, as predicted. Contrary to the model predictions, however, estimates based on the genetic assignment test underestimated movement. Standardized bias actually increased with increasing number of loci and alleles. Bias also increased with percent of the population sampled and number of loci when the direct estimate of movement was 0. None of these increases was significant, however.

The lack of complete support for the simulation predictions was due, in part, to the limited number of studies that systematically compared direct and genetically-based

quantitative estimates of movement. All of the analyses that showed non-significant results had power estimates below 0.1, which is considered very low (Cohen 1977). What this indicates is that the lack of significance in the analyses is not conclusive. The sample sizes required to achieve a reasonably high power for these effect sizes were in the hundreds and even in the thousands. Sample sizes were so low in some instances that some predictions could not be tested statistically; there were movement estimates for only 1 case based on the genetic assignment test and for 2 based on G_{ST} (Table 4.2). Overall, the current number of cases was thus too low to conclude that the lack of significance was meaningful.

In addition to the limiting effect of small sample size, the tests of the predictions were confounded by an uneven distribution in the values for specific variables. There was sufficient variation in percent of population sampled (0.6-100%), number of loci (5-33), number of alleles (2-25), and movement rate (0-0.336) to examine their effect on empirical bias (Table 4.2); there were several cases for each of F_{ST} (50), Θ (98), and R_{ST} (66). However, for many cases, changes in one variable were accompanied by confounding changes in another. For example, for estimates of movement based on Θ, for cases with fewer than 10 loci, the median number of alleles per locus was 10 (for 25 of 27 cases the number of alleles per locus was 10); for estimates based on 10 loci or more, the median number of alleles was 6 (for 6 of 8 cases the number of alleles per locus was 6) (Figures 4.5b and c). Thus, the positive effect (lower bias) expected from increasing the number of alleles was potentially being countered by a negative effect of decreasing the number of loci on which the genetic estimate was based. Furthermore, almost all (30

of 34) estimates based on Θ were based on samples of less than 6% of the population; estimates based on F_{ST} were based on 22 to 100% of the population. Thus, the apparent decrease in bias with increased percent of the population sampled (Figure 4.5a) may be due to either increasing the percent sampled or to using a more reliable method. Percent of the population sampled, number of loci, number of alleles, and movement rate are all predicted to have some effect on bias. However, because their values were not varied in a factorial design I could not test their independent effects. Attempts to control for the variation were unsuccessful because of the small sample sizes.

There were additional potential limitations. The simulations were run based on parameter settings provided by data from the empirical studies. In the absence of information from the empirical studies, however, I used default values for the number of generations since isolation, mutation rate, death rate, birth rate, and sex ratio which may not accurately reflect the true values for the organism. Inaccuracies in the parameter values will lead to incorrect estimates of bias based on the simulations. This can produce misleading discrepancies between the simulation and empirical values.

Another potential source of this inaccuracy was the estimates of movement rate. First, direct estimates of movement in the empirical studies were based on samples, unlike in their associated simulations for which there was no measurement error. The empirical estimate of movement rate was based on the proportion of the sample that moved per generation. Variation in sample size may alter that rate. Further, direct measures are more likely to underestimate actual movement because rare movement is often recorded as zero. Second, the direct estimate of movement was based on marking

and recapturing an organism. However, there was a variety of taxa (Table 4.2) followed for varying lengths of time and based on different marking techniques, all of which could affect the accuracy of direct estimates of movement (for a review see Turchin 1998). Finally, even in the absence of measurement error, the direct measure of movement may overestimate the genetically-based estimate of movement. In the model I assumed that movement was equivalent to gene flow. However, this is not necessarily the case in the empirical studies if fewer immigrants reproduce than residents; this would result in the genetically-based estimate's underestimating actual movement.

Another potential source for discrepancies between the empirical and simulation estimates may result from incorrect estimates of population sizes. If the true population size is underestimated and this value used in the simulation, the simulated genetic estimate of movement will be smaller than it should be. This is because over generations the relative effect of drift is higher in smaller populations for a given movement rate as compared to larger populations. This would lead to apparently increased differentiation between the simulated populations and produce a genetically-based estimate of movement that is lower than it would be in the empirical study.

Finally, there were also potential limitations because of the model's design. The model is based on some simplifying assumptions that may amplify discrepancies between the simulation and empirical estimates of bias. In particular, not all of the populations in the empirical studies were closed from other, untested, populations. This may result in genes from outside the study system increasing the apparent genetic difference between the populations and producing an underestimate of movement rate. The net effect would

be an increase in bias in the empirical studies. In comparison, the bias predicted from the simulation results would be lower. However, the fact that 70% of the empirical bias values were within 1 standard deviation of the simulation mean suggests that including these assumptions had only a limited effect on discrepancies between the empirical and simulation results.

Implications and conclusions

Because of the lack of power and limited range of parameter values tested, it is clear that more studies are needed that compare direct and genetic estimates of movement. This would be especially encouraged for conservation-related situations, i.e., those where isolation is recent. The studies in the current review were generally isolated over a much longer term. Furthermore, future comparisons should focus on situations where both versions of the genetic assignment test are used to generate an estimate of movement. In only 1 case were quantitative direct and genetic assignment test-based estimates systematically compared for the same populations and only using the older version of the test. As newer, and assumed to be improved, genetically-based estimates of movement emerge (particularly genetic assignment tests), these should be systematically compared to direct estimates of movement.

Regardless of the degree of similarity between the simulation and empirical results, one conclusion is the same for both. The difference between the direct and genetically-based estimates of movement was not constant for varying situations (e.g., sample size, number of loci, populations with overlapping vs. non-overlapping generations, etc.). The degree of reliability of movement estimates based on Θ varied

with movement rate, creating a Catch-22 situation for the researcher. The researcher would thus require prior knowledge of the movement rate in order to decide if their estimate of movement rate is reliable. This suggests that it is impractical to quantify inter-population movement based on genetic methods at this time.

Table 4.1. Parameters and values used in the simulation experiments based on the information provided in the empirical studies.

INPUT	
Parameters	**Value**
Held constant for all simulations:	
Number of offspring per female:	2 (drawn from Poisson distribution)
Death rate:	0.5 (individuals chosen at random)
Sex ratio:	1:1
Number of populations:	2
Number of repetitions of simulation	100
Varied among simulations:	
Movement rate:	Based on empirical information
Initial population sizes:	Based on empirical information
Sample percentages:	Based on empirical information
Number of loci:	Based on empirical information
Number of alleles:	Based on empirical information
Mutation rate:	Default: 10^{-5} for R_{ST}, private alleles, and genetic assignment test
Overlapping generations:	Based on empirical information (no/yes)
Duration of simulation (number of generations since isolation):	Based on empirical information; 100 generations if not reported
OUTPUT	
Produced for each input parameter combination:	
• Simulation bias *(genetic estimate - actual movement)*	
• Standardized simulation bias *([genetic estimate - actual movement]/actual movement)* if movement rate $\neq 0$	

Table 4.2. Studies and information used in the analyses. Note, PA: Private alleles; GAT: Genetic assignment test (Paetkau et al.'s 1995 version). Movement rates and population or sample sizes denoted with "*" were uneven for at least one comparison. "Global" genetic estimates of movement indicates that the estimate was reported for all populations combined, not for each pair compared; simulations were run based on the associated mean values. For population pairs where Θ or R_{ST} equaled 0, the genetic estimate of movement was assigned a default of 25,000. Allele values and population sizes in italics were based on estimates and not reported explicitly in the empirical study.

Source	Species and location of study	Direct estimate of movement rate per generation	Genetic estimate of movement	Number of loci (number of alleles)	Population size (number of populations)	Sample size
Non-overlapping generations:						
			Based on F_{ST}			
Geenen et al. (2000)	Green Emerald damselfly (*Lestes viridis*); Belgium	0	0.70-60.09	5 (2-3)	*400* (8)	18, 32*
Roeloffs & Riechert (1988)	Cooperative spider (*Agelena consociata*); Gabon	0	0.23 (global)	5 *(2)*	50 (54)	30
Varvio-Aho & Pamilo (1981)	Water-strider (*Gerris lacustris*); Finland	0.008,0.04	5.85, 9.75	7 *(3)*	67-220 (2 and 2)	40-60*
			Based on G_{ST}			
Rosenberg (1989)	Weidemeyer's admiral butterfly (*Limenitis weidemeyerii*); USA	0	6.69-10.17	33 *(2)*	100 (2 and 2)	20-60*

Source	Species and location of study	Direct estimate of movement rate per generation	Genetic estimate of movement	Number of loci (number of alleles)	Population size (number of populations)	Sample size
			Based on Θ			
Colas et al. (1997); Freville et al. (2001)	Cliff dwelling plant (*Centaurea corymbosa*); France	0	0.28-4.75	5 (2)	498 (6)	17-57*
Funk, Greene, et al. (2005); Funk, Blouin, et al (2005)	Columbia spotted frog (*Rana luteiventris*); USA	0-0.14	0.8-124.75; 25,000	6 (5-16)	300 (11)	12
Ingvarsson et al. (1997)	Mycophagous beetle (*Phalacrus substriatus*); Sweden	0.366	4.90 (global)	7 (5)	85 (42)	5
Lewis et al. (1997); Brookes et al. (1997)	Butterfly (*Plebejus argus*); Wales	0	6.16 (global)	12 (2-5)	432, 921, 2001* (3)	25, 31, 35*
Watts et al. (2004)	Damselfly (*Coenagrion mercuriale*); UK	0-0.049*	1.55-138.64	14 (2-16)	*400-2500** (7)	50
			Based on more than one measure			
Favre et al. (1997)	Shrew (*Crocidura russula*); Switzerland	0.04 (mean)	Θ: 1.51 (global) GAT: 1.50 (global)	8 *(25)*	500 (3)	172
Overlapping generations:						
			Based on F_{ST}			
Dobson (1994); Zammuto & Millar (1985)	Columbian ground squirrel (*Spermophilus columbianus*); Canada	0	3.98-58.35	6 (2-3)	200 (6)	13-82*
Pope (1992)	Red howler monkey (*Alouatta*	0-0.06*	11.65	10	200	107, 30*

Source	Species and location of study	Direct estimate of movement rate per generation	Genetic estimate of movement	Number of loci (number of alleles)	Population size (number of populations)	Sample size
	seniculus); Venezuela			(2)	(2)	
Schilthuizen & Lombaerts (1994)	Land snail (*Albinaria corrugata*); Greece	0, 0.195	0.22, 8.37	6 (2-4)	29 (20) 283, 521*(2)	29 283, 521*
Schwartz & Armitage (1980)	Yellow-bellied marmots (*Marmota flaviventris*); USA	0.02-0.05	3.32 (global)	8 (2)	12 (9)	12
			Based on R_{ST}			
Roemer et al. (2001)	Island fox (*Urocyon littoralis)*; USA	0	1.68	10 (2-3)	12-80 (2)	12
			Based on more than one measure			
Adams & Hutchings (2003)	Brook charr (*Salvelinus fontinalis*); Canada	0-0.004*	*PA*: 9.38 (global) R_{ST}: 0.04-124.88	5 (2-25)	4500 (8)	30-53*
Blundell et al. (2002)	River otter (*Lontra canadensis*); USA	0.07	*PA*: 4.49 R_{ST}:2.19	9 (2-10)	*200* (2)	31, 39*
Edwards et al. (2004)	Desert tortoise (*Gopherus agassizii*); USA	0	*PA*: 5.5 (global) R_{ST}:2.9-393.3, 25,000	7 (8-27)	*500* (9)	8-38*
Hoezel et al. (2001); Lewis et al. (1996)	Southern elephant seal (*Mirounga leonina*); Argentine/USA	0	Θ: 9.75 *PA*: 4.90	7 (2-9)	5000 (2)	28, 32*
Wilson et al. (2004); Hutchings & Gerber (2002)	Brook charr (*Salvelinus fontinalis*); Canada	0.011,0.08*	Θ: 25,000 *PA*: 28.90	16 (2-16)	730 (2)	200, 398*

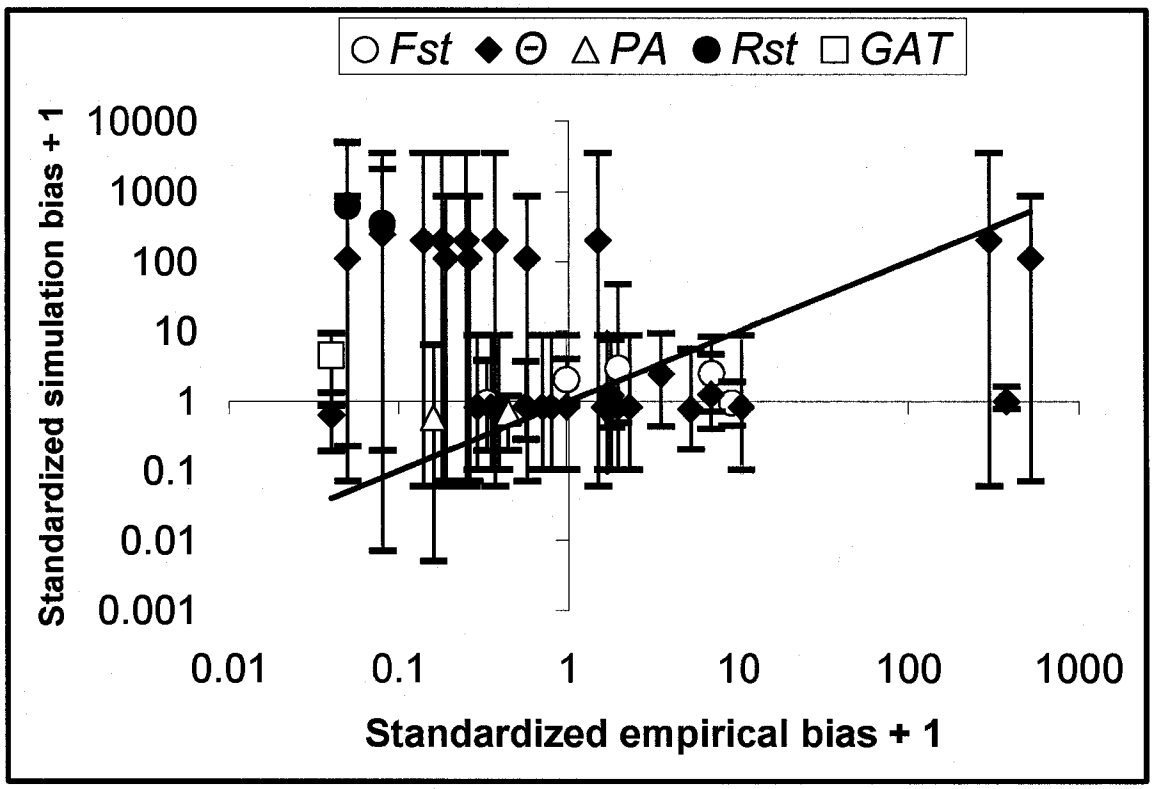


Figure 4.1. Mean standardized simulation bias vs. standardized empirical bias based on F_{ST}, Θ, private alleles (PA), R_{ST}, and the genetic assignment test (GAT, Paetkau et al. 1995 version). To show the full range of values, I added 1 to each bias (to remove negative values) and plotted them on a log scale. Thus, a value of 1 represents no bias. The simulation bias shown represents the mean over 100 replicate simulations; the bars represent the minimum and maximum over the 100 replicates. The solid line is the 1:1 line.

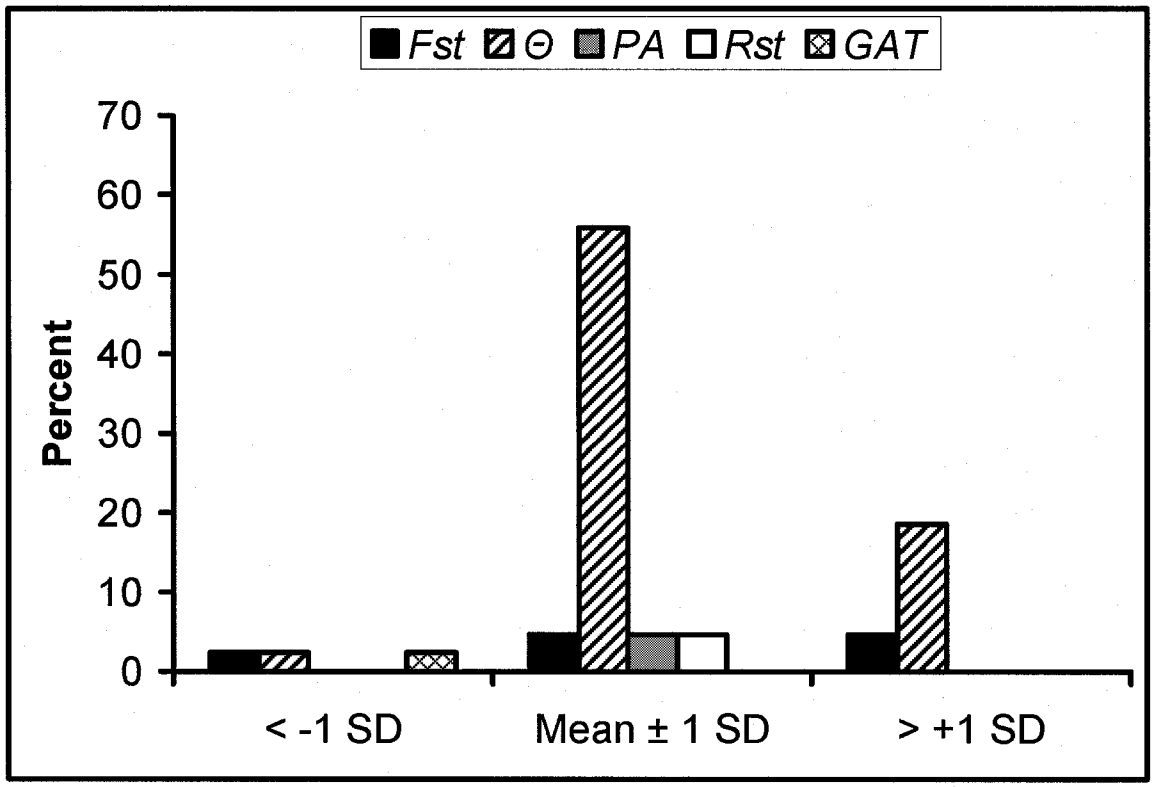


Figure 4.2. Percentage of standardized empirical bias estimates that fell within 1 standard deviation of the mean standardized simulation bias estimates for each of the genetic measures used. The means and standard deviations were calculated based on 100 replicate simulations of each test case parameter combination.

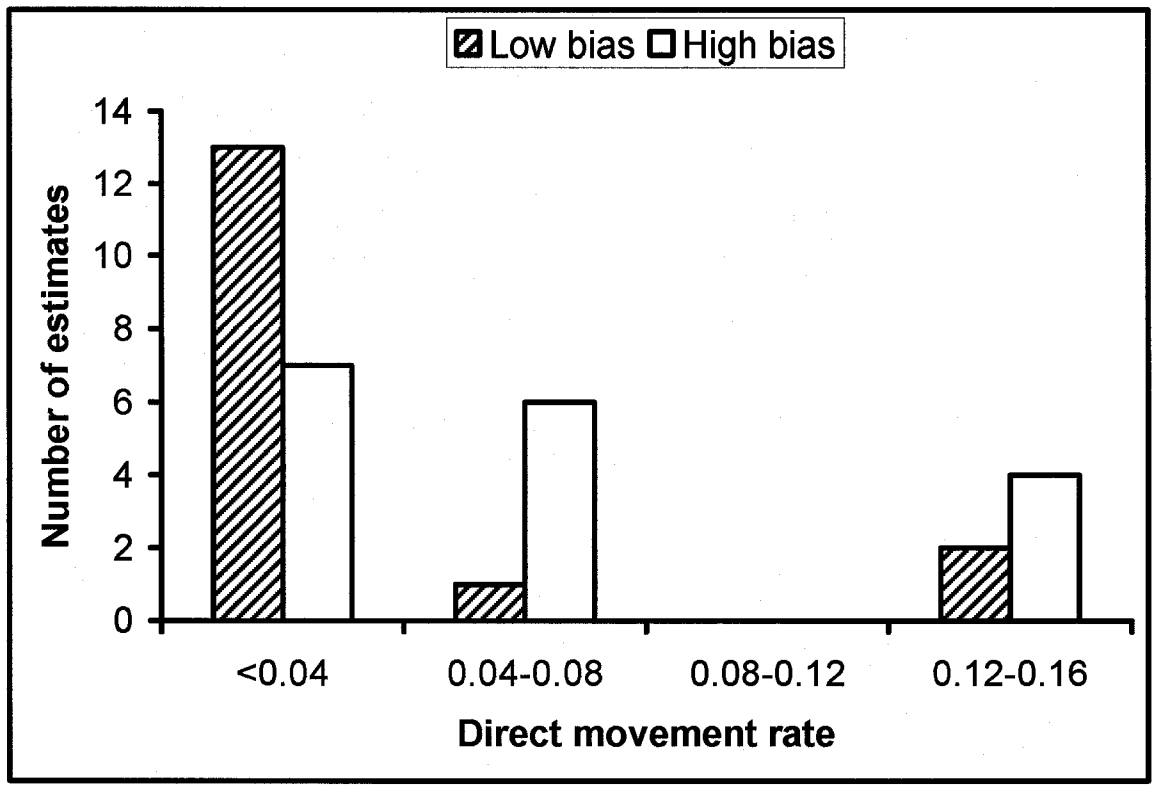


Figure 4.3. Reliability of movement estimates based on Θ vs. the direct empirical estimate of movement rate. Bias was considered low if it ranged between -0.6 and 1.5 (see *Methods* in Chapters 2 and 3). When movement rates were uneven between two populations, I plotted the average rate.

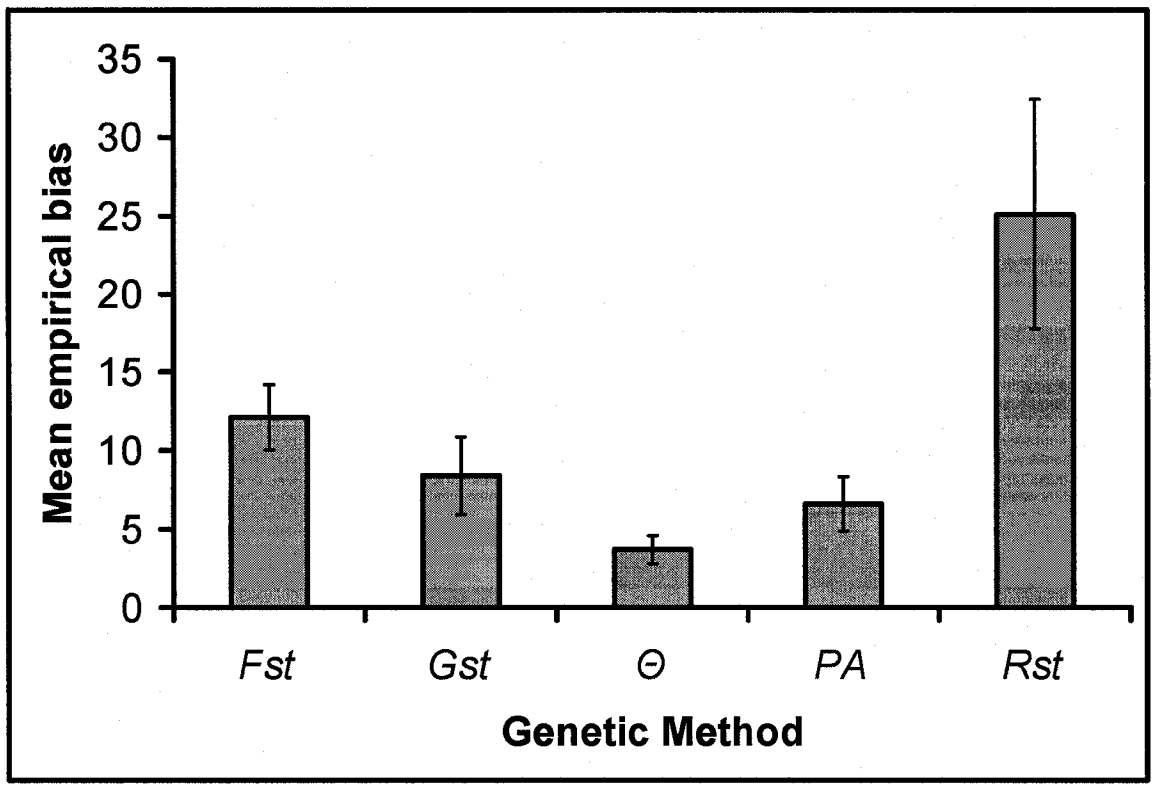


Figure 4.4. Mean empirical bias (i.e., genetic estimate of movement - direct estimate of movement) for each of the genetic methods, ± 1 SE for cases where the direct movement rate was 0. These cases were analyzed separately because bias could only be positive. PA: Private alleles

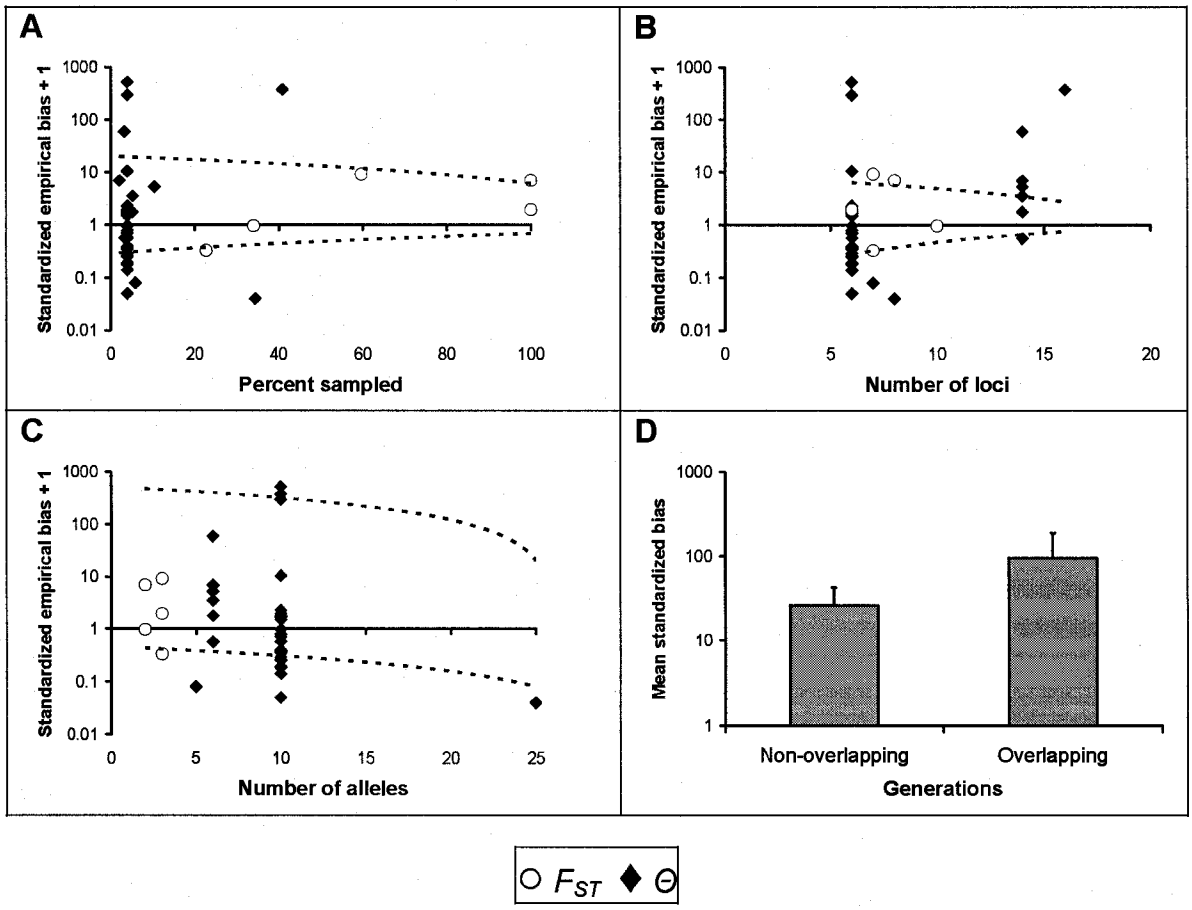


Figure 4.5. Standardized empirical bias based on F_{ST} and Θ vs. (A) the percentage of the population sampled, (B) the number of loci on which the genetic measure was based, (C) the number of alleles per locus, and (D) overlapping vs. non-overlapping generations. To show the full range of values, I added 1 to each bias (to remove negative values) and plotted them on a log scale. Thus, a value of 1 represents no bias. Dotted lines represent the linear trend for the maximum and minimum standardized bias based on the corresponding simulations. In (D) I have plotted the mean standardized empirical bias for all cases where generations overlapped ($n = 4$) vs. those that did not overlap ($n = 34$), +1 SE.

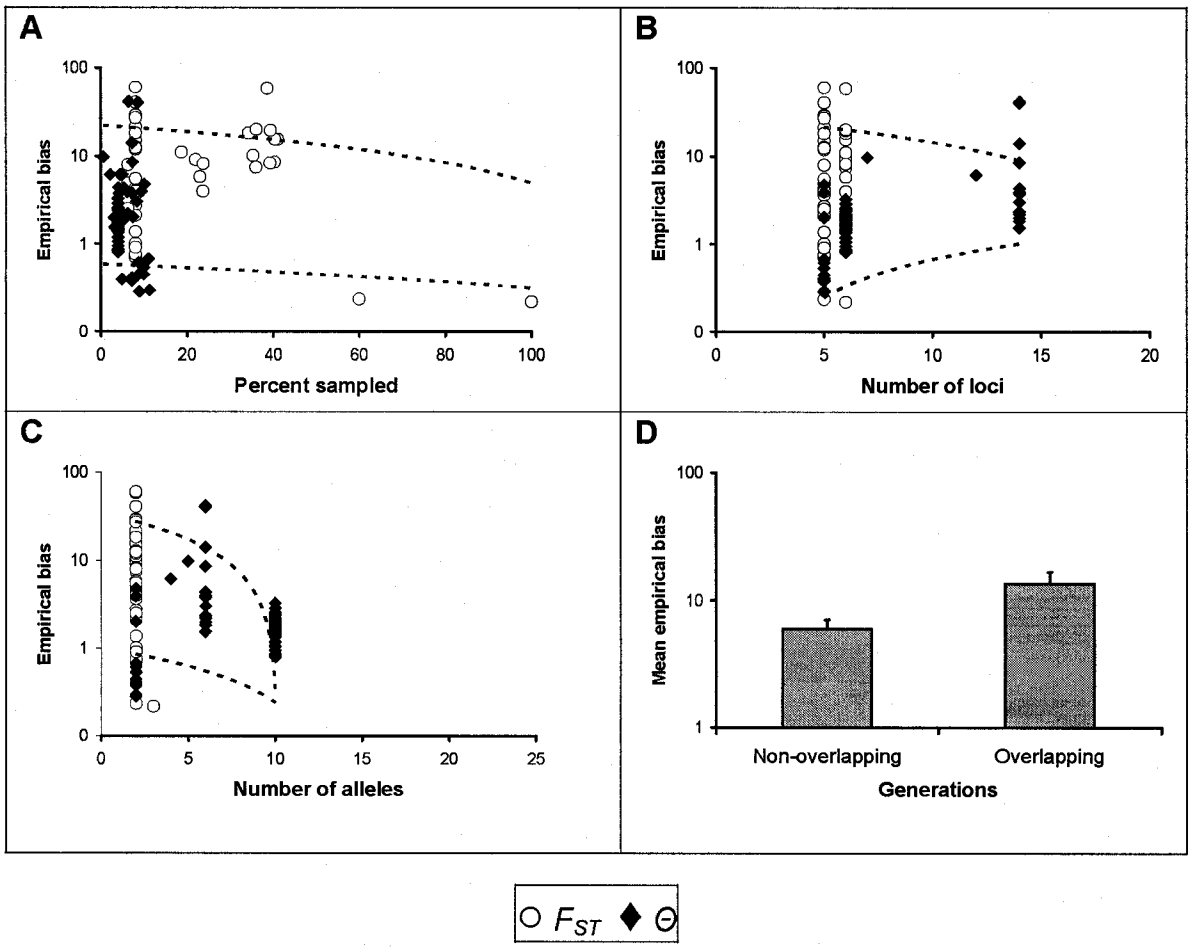


Figure 4.6. Empirical bias based on F_{ST} and Θ vs. (A) the percentage of the population sampled, (B) the number of loci on which the genetic measure was based, (C) the number of alleles per locus, and (D) overlapping vs. non-overlapping generations. To show the full range of values I plotted them on a log scale. Dotted lines represent the linear trend for the maximum and minimum bias based on the corresponding simulations. In (D) I have plotted the mean standardized empirical bias for all cases where generations overlapped (n = 17) vs. those that did not overlap (n = 92), +1 SE. Note, because the direct estimate of movement was 0 for all cases presented here, all empirical bias values were positive.

CHAPTER 5: GENERAL CONCLUSIONS

Can genetic information be used to quantify inter-population movement?

Using a simulation model, I asked which, if any, of F_{ST} (Nei 1973), G_{ST} (Nei 1987), Θ (Weir and Cockerham 1984), R_{ST} (Slatkin, 1995), private alleles (Slatkin 1985b), and two versions of the genetic assignment test (Paetkau et al. 1995; Waser and Strobeck 1998; Paetkau et al. 2004) was suitable for calculating inter-population movement over a broad set of demographic and genetic scenarios. Estimates of movement based on F_{ST} were reasonable in situations with at least 20% of the population sampled, at least 10 loci with at least 6 alleles each, a movement rate of 0.04 to 0.08 per generation, and at least 90 generations since isolation (Chapter 3). Θ and G_{ST}'-based estimates of movement were reasonable for the same parameters but limited to situations with non-overlapping generations and a movement rate of less than 0.04 (Chapters 2 and 3). This suggests that genetic information could be used to quantify inter-population movement in these particular situations and based on these measures.

There is partial support for results of the simulation model based on empirical studies (Chapter 4). Seventy percent of the 44 test cases had a standardized empirical bias within 1 standard deviation of the mean simulation bias. This suggests that the simulation model provided a somewhat realistic estimate of the range in potential standardized bias values in an empirical study given a particular parameter combination. There was some support for the 3 main predictions based on the simulation results: 1) Estimates based on F_{ST} and Θ were more reliable than those based on the frequency of private alleles, R_{ST}, and the genetic assignment test (Paetkau et al. 1995 version); Θ-based estimates were

more reliable when movement rate was less than 0.04. 2) Estimates of movement based on the frequency of private alleles underestimated movement. 3) Bias decreased with percent of the population sampled and number of alleles in some situations; bias was also lower if based on populations with non-overlapping generations. This partial support suggests that the conclusions drawn from the simulation results in Chapters 2 and 3 are potentially valid but require further study.

In practical terms, however, the results from the experimental simulations and the partial empirical test suggest that it is generally unlikely that genetic information can be used to quantify inter-population movement reliably at this time. Both the simulations (Chapters 2 and 3) and the empirical studies (Chapter 4) showed that the difference between the direct and genetically-based estimates of movement varied among the demographic and genetic situations (i.e., with changes in movement rate, sample size, number of loci, and populations with overlapping vs. non-overlapping generations, etc.). Especially problematic is the fact that bias changed with movement rate itself. This leads to the conclusion that, in order to decide whether F_{ST} or Θ can be used to quantify inter-population movement in a particular situation, the researcher needs to already have an estimate of movement.

Implications and future directions

Two types of situations were not adequately tested because of the lack of power and limited range of parameter values examined in Chapter 4. First, the experimental simulations suggested that genetically-based estimates of movement were reliable when populations were not recently isolated and had non-overlapping generations (Chapters 2

and 3). The empirical test results supported the latter as reliability was higher in situations where generations were non-overlapping (Chapter 4). However, the populations in the studies reviewed had generally been isolated over a much longer period of time. Therefore, I was unable to validate the model in conservation-related situations, i.e., where isolation is recent. Studies of formerly panmictic populations that had been recently isolated by events such as the building of a road would be particularly appropriate to test the conclusions of this simulation.

Second, additional tests should focus on situations where the genetic assignment tests are used to generate an estimate of movement, especially in the context of these conservation-related situations. These methods are expected to work better in these conservation-related situations because fewer assumptions are made or more variable loci (microsatellites) are used. However, based on the simulation results (Chapters 2 and 3), estimates based on the genetic assignment tests did not reliably quantify movement at all. This result is significant because it suggests that the recent increase in use of these methods is misguided (Figure 1.1). Validation of this unexpected result was limited, however, because in only 1 case were quantitative direct and genetic assignment test-based estimates systematically compared for the same sets of populations (Chapter 4). Additional empirical study should focus on comparing direct and genetic estimates of movement in conservation-related situations.

Over the last 20 years, over 1000 studies using a variety of genetic methods to quantify inter-population movement have been published (Figure 1.1). Methods have evolved over this time period as well. As Whitlock and McCauley suggest, there has been

a "minor cottage industry of estimating *Nm* from F_{ST}" (Whitlock & McCauley 1999). There has also been a recent shift towards the use of assignment tests to measure movement (Paetkau et al. 1995; Waser and Strobeck 1998; Rannala & Mountain 1997; Pritchard et al. 2000; Paetkau el al. 2004). There have been several reviews discussing the theoretical merits and limitations of various genetic methods (Lewontin 1985; Slatkin 1985a; Crochet, 1996; Bohonak et al 1998; Bossart & Pashley Prowell, 1998a and b; Steinberg & Jordan, 1998; Whitlock & McCauley 1999; Bohonak & Roderick 2001; Neigel 2002). Several have also compared the performance of different methods and/or examined the effects of factors such as sample size or numbers of loci and alleles (e.g., Slatkin & Barton 1989; Allen et al. 1995; Ruzzante 1998; Cornuet et al. 1999; Whitlock & McCauley 1999; Balloux & Goudet 2002; Bjørnstad & Røed 2002; Kalinowski 2002; Kalinowski 2005). Missing, however, are systematic tests of the performance of these various methods in field situations. A handful of studies have recommended that direct measures should be used in conjunction with genetic estimates as both types of methods are important to assessing movement rate (e.g., Slatkin 1985a; Bossart & Pashley Prowell, 1998a; Whitlock & McCauley 1999; Pocock et al. 2005). A growing number of researchers are including both types of information: they compare genetic distance to geographic distance based on marking and recapturing individuals (e.g., Nève et al. 1996; Pfenniger et al. 1996; Keyghobadi et al. 1999; Miller et al. 2002; Keyghobadi et al. 2003), compare genetic estimates of movement to direct estimates in similar systems (e.g., Merriam et al 1989; Schneider 1999; Blouin-Demers & Weatherhead 2002 paired with Lougheed et al. 1999; Knutsen et al. 2000 paired with Rukke & Midtgaard 1998), or

systematically compare genetic estimates of movement to direct estimates in the same populations (see Table 4.2). Collectively, however, these studies still represent only a small number of the large number of studies that use genetic information to measure movement. Only the latter are appropriate for comparing estimates of movement amount. The trend in the literature appears to be founded on a belief that if we just keep trying to build a better genetic method we will, eventually, succeed. This may or may not be true, but based on the results presented in Chapters 2 to 4, the current newer, assumed to be improved, genetic methods have not necessarily increased the reliability of movement estimates as compared to the older methods. We discuss, theorize about, and actually use genetic methods in the field to measure movement, but we have not sufficiently tested their reliability. Quantifying movement accurately is crucial for answering many questions in areas such as conservation, management, or landscape ecology. To best address the need for an accurate quantitative estimator of movement, we need to continue to build on the theory, but we should not do so without adequate empirical testing.

LITERATURE CITED

Adams B. K. & Hutchings J. A. (2003) Microgeographic population structure of brook charr: a comparison of microsatellite and mark-recapture data. *Journal of Fish Biology* 62:517-533.

Allen P. J., Amos W., Pomeroy P. P., & Twiss S. D. (1995) Microsatellite variation in grey seals (*Halichoerus grypus*) shows evidence of genetic differentiation between two British breeding colonies. *Molecular Ecology* 4:653-662.

Balloux F. & Goudet J. (2002) Statistical properties of population differentiation estimators under stepwise mutation in a finite island model. *Molecular Ecology* 11:771-783.

Balloux F. & Lugon-Moulin N. (2002) The estimation of population differentiation with microsatellite markers. *Molecular Ecology* 11:155-165.

Berry O., Tocher M. D., & Sarre S. D. (2004) Can assignment tests measure dispersal? *Molecular Ecology* 13:551-561.

Bjørnstad G. & Røed K. H. (2002) Evaluation of factors affecting individual assignment precision using microsatellite data from horse breeds and simulated breed crosses. *Animal Genetics* 33:264-270.

Blouin-Demers G. & Weatherhead P. J. (2002) Implications of movement patterns for gene flow in black rat snakes (*Elaphe obsoleta*). *Canadian Journal of Zoology-Revue Canadienne de Zoologie* 80:1162-1172.

Blundell G. M., Ben-David M., Groves P., Bowyer R. T., & Geffen E. (2002) Characteristics of sex-biased dispersal and gene flow in coastal river otters: implications for natural recolonization of extirpated populations. *Molecular Ecology* 11:289-303.

Bohonak A. J., Davies N., Roderick G. K., & Villablanca F. X. (1998) Is population genetics mired in the past? *Trends in Ecology & Evolution* 13:360.

Bohonak A. J. & Roderick G. K. (2001) Dispersal of invertebrates among temporary ponds: are genetic estimates accurate? *Israel Journal of Zoology* 47:367-386.

Bossart J. L. & Prowell D. P. (1998a) Genetic estimates of population structure and gene flow: limitations, lessons and new directions. *Trends in Ecology & Evolution* 13:202-206.

Bossart J. L. & Prowell D. P. (1998b) Reply from J.L Bossart and D. Pashley Prowell. *Trends in Ecology & Evolution* 13:360.

Brookes M. I., Graneau Y. A., King P., Rose O. C., Thomas C. D., & Mallet J. L. B. (1997) Genetic analysis of founder bottlenecks in the rare British butterfly *Plebejus argus*. *Conservation Biology* 11:648-661.

Brown J. H. & Kodric-Brown A. (1977) Turnover rates in insular biogeography - Effect of immigration on extinction. *Ecology* 58:445-449.

Cockerham C. C. & Weir B. S. (1993) Estimation of gene flow from F-statistics. *Evolution* 47:855-863.

Cohen J. (1977) *Statistical power analysis for the behavioral sciences* (revised edition). Academic Press, New York:

Colas B., Olivieri I., & Riba M. (1997) *Centaurea corymbosa*, a cliff-dwelling species tottering on the brink of extinction: a demographic and genetic study. *Proceedings of the National Academy of Sciences of the United States of America* 94:3471-3476.

Cornuet J. M., Piry S., Luikart G., Estoup A., & Solignac M. (1999) New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* 153:1989-2000.

Crochet P. A. (1996) Can measures of gene flow help to evaluate bird dispersal? *Acta Oecologica-International Journal of Ecology* 17:459-474.

Crow J. F. & Aoki K. (1984) Group selection for a polygenic behavioral trait: estimating the degree of population subdivision. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* 81:6073-6077.

Dallas J. F. (1992) Estimation of microsatellite mutation rates in recombinant inbred strains of mouse. *Mammalian Genome* 3:452-456.

Dobson F. S. (1994) Measures of gene flow in the Columbian ground squirrel. *Oecologia* 100:190-195.

Edwards T., Schwalbe C. R., Swann D. E., & Goldberg C. S. (2004) Implications of anthropogenic landscape change on inter-population movements of the desert tortoise (*Gopherus agassizii*). *Conservation Genetics* 5:485-499.

Erdfelder E, Faul F, & Buchner A. (1996) GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers* 28:1-11.

Favre L., Balloux F., Goudet J., & Perrin N. (1997) Female-biased dispersal in the monogamous mammal *Crocidura russula*: evidence from field data and

microsatellite patterns. *Proceedings of the Royal Society of London Series B-Biological Sciences* 264:127-132.

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

Freimer N. B. & M. Slatkin. (1996) Microsatellites: evolution and mutational processes. *Variation in the Human Genome* 197:51-67.

Freville H., Justy F., & Olivieri I. (2001) Comparative allozyme and microsatellite population structure in a narrow endemic plant species, *Centaurea corymbosa* Pourret (Asteraceae). *Molecular Ecology* 10:879-889.

Funk W. C., Blouin M. S., Corn P. S., Maxell B. A., Pilliod D. S., Amish S., & Allendorf F. W. (2005) Population structure of Columbia spotted frogs (*Rana luteiventris*) is strongly affected by the landscape. *Molecular Ecology* 14:483-496.

Funk W. C., Greene A. E., Corn P. S., & Allendorf F. W. (2005) High dispersal in a frog species suggests that it is vulnerable to habitat fragmentation. *Biology Letters* 1:13-16.

Geenen S., Jordaens K., De Block M., Stoks R., & De Bruyn L. (2000) Genetic differentiation and dispersal among populations of the damselfly *Lestes viridis* (Odonata). *Journal of the North American Benthological Society* 19:321-328.

Guinand B., Topchy A., Page K. S., Burnham-Curtis M. K., Punch W. F., & Scribner K. T. (2002) Comparisons of likelihood and machine learning methods of individual classification. *Journal of Heredity* 93:260-269.

Hoelzel A. R., Campagna C., & Arnbom T. (2001) Genetic and morphometric differentiation between island and mainland southern elephant seal populations. *Proceedings of the Royal Society of London Series B-Biological Sciences* 268:325-332.

Hutchings J. A. & Gerber L. (2002) Sex-biased dispersal in a salmonid fish. *Proceedings of the Royal Society of London Series B-Biological Sciences* 269:2487-2493.

Ingvarsson P. K., Olsson K., & Ericson L. (1997) Extinction-recolonization dynamics in the mycophagous beetle *Phalacrus substriatus*. *Evolution* 51:187-195.

Jarne P. & Lagoda P. J. L. (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology & Evolution* 11:424-429.

Kalinowski S. T. (2002) How many alleles should be used to estimate genetic distances? *Heredity* 88:62-65.

Kalinowski S. T. (2005) Do polymorphic loci require large sample sizes to estimate genetic distances? *Heredity* 94:33-36.

Keyghobadi N., Roland J., & Strobeck C. (1999) Influence of landscape on the population genetic structure of the alpine butterfly *Parnassius smintheus* (Papilionidae). *Molecular Ecology* 8:1481-1495.

Keyghobadi N., Roland J., Fownes S., & Strobeck C. (2003) Ink marks and molecular markers: Examining the effects of landscape on dispersal using both mark-recapture and molecular methods. In: *Butterflies: Ecology and evolution taking flight* (eds: C. L. Boggs, W. B. Watt, & P. R. Erlich) pp.169-183. The University of Chicago Press, Chicago.

Kimura M. & Ohta T. (1978) Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proceedings of the National Academy of Sciences of the United States of America* 75:2868-2872.

Knutsen H., Rukke B. A., Jorde P. E., & Ims R. A. (2000) Genetic differentiation among populations of the beetle *Bolitophagus reticulatus* (Coleoptera: Tenebrionidae) in a fragmented and a continuous landscape. *Heredity* 84:667-676.

Koenig W. D., VanVuren D., & Hooge P. N. (1996) Detectability, philopatry, and the distribution of dispersal distances in vertebrates. *Trends in Ecology & Evolution* 11:514-517.

Lewis M., Campagna C., & Quintana F. (1996) Site fidelity and dispersion of southern elephant seals from Patagonia. *Marine Mammal Science* 12:138-147.

Lewis O. T., Thomas C. D., Hill J. K., Brookes M. I., Crane T. P. R., Graneau Y. A., Mallet J. L. B., & Rose O. C. (1997) Three ways of assessing metapopulation structure in the butterfly *Plebejus argus*. *Ecological Entomology* 22:283-293.

Lewontin R. C. (1985) Population genetics. In: *Evolution* (eds P. J. Greenwood, P. H. Harvey, & M. Slatkin) pp. 3-44. Cambridge University Press, New York.

Lougheed S. C., Gibbs H. L., Prior K. A., & Weatherhead P. J. (1999) Hierarchical patterns of genetic population structure in black rat snakes (*Elaphe obsoleta obsoleta*) as revealed by microsatellite DNA analysis. *Evolution* 53:1995-2001.

Manel S., Schwartz M. K., Luikart G., & Taberlet P. (2003) Landscape genetics: combining landscape ecology and population genetic. *Trends in Ecology and Evolution* 18:189-197.

Merriam G., Kozakiewicz M., Tsuchiya E., & Hawley K. (1989) Barriers as boundaries for metapopulations and demes of *Peromyscus leucopus* on farm landscapes. *Landscape Ecology* 2: 227-235.

Miller M. P., Blinn D. W., & Keim P. (2002) Correlations between observed dispersal capabilities and patterns of genetic differentiation in populations of four aquatic insect species from the Arizona White Mountains, USA. *Freshwater Biology* 47:1660-1673.

Mitton J. B. (1994) Molecular approaches to population biology. *Annual Review of Ecology and Systematics* 25:45-69.

Nei M. (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America* 70:3321-3323.

Nei M. (1987) *Molecular Evolutionary Genetics.* Columbia University Press, New York.

Neigel J. E. (2002) Is F_{ST} obsolete? *Conservation Genetics* 3:167-173.

Nève G., Mousson L., & Baguette M. (1996) Adult dispersal and genetic structure of butterfly populations in a fragmented landscape 13. *Acta Œcologica* 17:621-626.

Paetkau D., Calvert W., Stirling I., & Strobeck C. (1995) Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology* 4:347-354.

Paetkau D., Slade R., Burdens M., & Estoup A. (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology* 13:55-65.

Parker P. G., Snow A. A., Schug M. D., Booton G. C., & Fuerst P. A. (1998) What molecules can tell us about populations: choosing and using a molecular marker. *Ecology* 72:361-382.

Pearse D. E. & Crandall K. A. (2004) Beyond F_{ST}: analysis of population genetic data for conservation. *Conservation Genetics* 5:585-602.

Pfenninger M., Bahl A., & Streit B. (1996) Isolation by distance in a population of a small land snail *Trochoidea geyeri*: Evidence from direct and indirect methods. *Proceedings of the Royal Society of London Series B-Biological Sciences* 263:1211-1217

Pocock M. J. O., Hauffe H. C., & Searle J. B. (2005) Dispersal in house mice. *Biological Journal of the Linnean Society* 84:565–583.

Pope T. R. 1992. The influence of dispersal patterns and mating system on genetic differentiation within and between populations of the red howler monkey (*Alouatta seniculus*). *Evolution* 46:1112-1128.

Pritchard J. K., Stephens M., & Donnelly P. (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945-959.

Rannala B. & Mountain J. L. (1997) Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences of the United States of America* 94:9197-9201.

Rekke B. A. & Midtgaard F. (1998) The importance of scale and spatial variables for the fungivorous beetle *Bolitophagus reticulatus* (Coleoptera: Tenebrionidae) in a fragmented forested landscape. *Ecography* 21:561-572.

Roeloffs R. & Riechert S. E. (1988) Dispersal and population-genetic structure of the cooperative spider, *Agelena consociata*, in West African Rainforest. *Evolution* 42:173-183.

Roemer G. W., Smith D. A., Garcelon D. K., & Wayne R. K. (2001) The behavioural ecology of the island fox (*Urocyon littoralis*). *Journal of Zoology* 255:1-14.

Rosenberg R. H. 1989. Genetic differentiation among populations of Weidemeyers admiral butterfly. *Canadian Journal of Zoology-Revue Canadienne de Zoologie* 67:2294-2300.

Ruzzante D. E. (1998) A comparison of several measures of genetic distance and population structure with microsatellite data: bias and sampling variance. *Canadian Journal of Fisheries and Aquatic Sciences* 55:1-14.

Schilthuizen M. & Lombaerts M. (1994) Population structure and levels of gene flow in the Mediterranean land snail *Albinaria corrugata* (Pulmonata: Clausiliidae). *Evolution* 48:577-586.

Schneider J. C. (1999) Dispersal of a highly vagile insect in a heterogeneous environment *Ecology* 80:2740–2749.

Schwartz O. A. & Armitage K. B. (1980) Genetic variation in social mammals: the marmot model. *Science* 207:665-667.

Slatkin M. (1985a) Gene flow in natural populations. *Annual Review of Ecology and Systematics* 16:393-430.

Slatkin M. (1985b) Rare alleles as indicators of gene flow. *Evolution* 39:53-65.

Slatkin M. (1987) Gene flow and the geographic structure of natural populations. *Science* 236:787-792.

Slatkin M. (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139:457-462.

Slatkin M. & Barton N. H. (1989) A comparison of three indirect methods for estimating average levels of gene flow. *Evolution* 43:1349-1368.

SPSS Inc. 2004. SPSS for Windows, Release 13.0. (1 September 2004) SPSS Inc., Chicago.

Steinberg E. K. & Jordan C. E. (1998) Using molecular genetics to learn about the ecology of threatened species: the allure and the illusion of measuring genetic structure in natural populations. In: *Conservation Biology* (eds P. L. Fielder and P. M. Kareiva) pp. 440-460. Chapman & Hall, Toronto.

Turchin P. (1998) *Quantitative analysis of movement: Measuring and modelling population redistribution in animals and plants*. Sinauer Associates, Inc., Massachusetts.

Varvio-Aho S. L. & Pamilo P. (1981) Spatio-temporal microdifferentiation of water-strider (Gerris) populations. *Genetical Research* 37:253-263.

Waples R. S. (1998) Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *Journal of Heredity* 89:438-450.

Waser P. M. & Strobeck C. (1998) Genetic signatures of interpopulation dispersal. *Trends in Ecology and Evolution* 13:43-44.

Watts P. C., Rouquette J. R., Saccheri J., Kemp S. J., & Thompson D. J. (2004) Molecular and ecological evidence for small-scale isolation by distance in an endangered damselfly, *Coenagrion mercuriale*. *Molecular Ecology* 13:2931-2945.

Weber J. L. & Wong C. (1993) Mutation of human short tandem repeats. *Human Molecular Genetics* 2:1123-1128.

Weir B. S. & Cockerham C. C. (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution* 38:1358-1370.

Weir B. S. & Hill W. G. (2002) Estimating F-statistics. *Annual Review of Genetics* 36:721-750.

Whitlock M. C. & McCauley D. E. (1999) Indirect measures of gene flow and migration: $F_{ST} \neq (1/4\ Nm+1)$. *Heredity* 82:117-125.

Wilson A. J., Hutchings J. A., & Ferguson M. M. (2004) Dispersal in a stream dwelling salmonid: Inferences from tagging and microsatellite studies. *Conservation Genetics* 5:25-37.

Wright S. (1931) Evolution in Mendelian populations. *Genetics* 28:114-138.

Wright S. (1951) The genetical structure of populations. *Annals of Eugenics* 15:323-354.

Zammuto R. M. & Millar J. S. (1985) Environmental predictability, variability, and *Spermophilus columbianus* life history over an elevational gradient. *Ecology* 66:1784-1794.

APPENDIX I: SUPPLEMENTAL MATERIAL

The enclosed CD-ROM contains the files needed to run and interpret both versions of the simulation model (one for each version of the genetic assignment test), the source code, and the data on which the results presented in each chapter are based.

Files required to run the simulations and interpret the output:

- Instructions on how to run the simulation

  model information.pdf

- Simulation model that runs the Paetkau et al. 1995 version of the genetic assignment test

  genmodel.exe
  sample input file: parameterfile.par
  sample output files: parameterfile.out/.gen./.rep/.det

- Simulation model that runs the Paetkau et al. 2004 version of the genetic assignment test

  genmodel_gat_random.exe
  sample input file: parameterfile_gat_random.par
  sample output files: parameterfile_gat_random.out/.gen./.rep/.det

Source code for genmodel.exe/genmodel_gat_random.exe (file names are the same for both versions except where indicated):

allelefrequency.h (**genmodel_gat_random** only)
array.h/array.cpp (.cpp for **genmodel_gat_random** only)
defs.h
dfarr.h/dfarr.cpp
diarr.h/diarr.cpp
genutil.h/genutil.cpp
individual.h/individual.cpp

input.h/input.cpp
loci.h/loci.cpp
locus.h/locus.cpp
metapop.h/metapop.cpp
model.cpp
observer.h/observer.cpp
population.h/population.cpp
process.h (**genmodel_gat_random** only)
simul.h/simul.cpp

Data files:

- Chapter 2 data files

  theta-pop500-overlap0
  theta-pop500-overlap1
  theta-pop500-overlap2
  theta-pop1000-overlap0
  theta-pop1000-overlap1
  theta-pop1000-overlap2
  theta-pop2000-overlap0
  theta-pop2000-overlap1
  theta-pop2000-overlap2

- Chapter 3 data files

  allmethod-pop500-overlap0
  allmethod-pop500-overlap2
  allmethod-pop1000-overlap0
  allmethod-pop1000-overlap2
  allmethod-pop2000-overlap0
  allmethod-pop2000-overlap2

- Chapter 4 data files

  lit_review_data.xls