The Limitations of Hypothesis Testing as a
Means of Demographically Delineating Independent Units

Karen K. Martien, Department of Biology 0116, University of San Diego, La Jolla, CA 92093

Barbara L. Taylor, Southwest Fisheries Science Center, P.O. Box 271, La Jolla, CA 92038

ABSTRACT

The goal of many studies of population structure is to describe the underlying spatial

structure of a region, including the number and size of demographically autonomous units and

the location of boundaries between them. Such studies are typically conducted using hypothesis-

testing techniques which were designed simply to determine whether or not the individuals

within a region are panmictic. We examine the efficacy of hypothesis-testing in determining the

demographic spatial structure within a region. We evaluate such an approach by using simulated

data for which the underlying spatial structure is known. We determine the average p-value,

average degree of differentiation, and power to detect differentiation as a function of the number

of putative populations defined by the researcher and assess the degree to which the accuracy of

boundary placement corresponds to these value . Statistical power and average differentiation

are much higher and the average p-value is much lower when a region is divided too coarsely

into only two putative populations than when it is divided in a way that accurately reflects the

underlying population structure. There are two reasons for the inverse relationship between

power and the number of putative populations defined. First, as with any real case, the number

of samples per group increases as the number of groups decreases, resulting in an increase in

power. Second, defining fewer groups results in more distantly related samples being placed in

adjacent groups, thus increasing the degree of differentiation between adjacent groups and

thereby the statistical power to detect that differentiation. Our results call into question the

suitability of hypothesis testing for determining population structure, particularly when the actual

number and location of population boundaries are of interest.

INTRODUCTION

An understanding of the population structure of a species is essential in many biological studies, both academic and applied. In many cases, it is sufficient simply to determine whether or not genetic structuring is present within a particular region; in other words, are the individuals within a region panmictic? There are numerous tests available to answer this question, most of which compare the null hypothesis of panmixia to an alternate hypothesis specified by the researcher. These analytical techniques were designed specifically to address this question, and appear to do so reasonably well (Hudson, 1992).

In many cases, however, researchers are interested not only in determining whether or not a region is panmictic, but also in describing the underlying spatial structure: how many populations are present within the region, how big are those populations, and where are the boundaries between them? In these cases, the interest of the researcher is in identifying demographically independent units between which dispersal is restricted rather than in genetic structure *per se*. Nonetheless, they typically use the same hypothesis-testing approaches developed for use in simply determining whether or not genetic structure is present. In doing so, they are using genetic structure as a proxy for addressing a question of demography. In this paper, we examine the efficacy of hypothesis testing in an analysis of population structure when the actual number and location of population boundaries are of interest.

A major stumbling-block to the use of hypothesis-testing in elucidating spatial structure is that a hypothesis-testing approach requires the researcher to construct in advance a hypothesis regarding the number and location of population boundaries. If populations are discretely distributed, distributional gaps can be good guides to the location of population boundaries. In

addition, dispersal rates between discrete populations are often low enough to allow the

development of a strong phylogeographic signal. Researchers can therefore often use gene trees

to guide the placement of hypothesized boundaries (Brown Gladden *et al.* 1997; Palacios and.

González-Candelas 1997). The placement of hypothesized boundaries is much more difficult

when dealing with a continuously distributed species. There are no obvious dispersal barriers or

distributional gaps to guide boundary placement. Furthermore, dispersal rates between

continuous populations are often high enough to eliminate any phylogeographic signal (Moritz *et*

al. 1997; Bérubé *et al.* 1998.). Gene trees are therefore of little help in determining how to

stratify samples into putative populations.

In an attempt to document the different approaches researchers take to stratifying their

data, we reviewed 29 papers on population structure of continuous species published in the last 8

years. In most cases no justification is given for the strata chosen, nor is it specified whether or

not strata other than those presented in the final results were examined. Thus, a rigorous analysis

of the frequency of different strategies is not possible. Nonetheless, based on both our

examination of the literature and conversations with researchers regarding their methods of

stratification, we have been able to discern three commonly employed approaches. First, data are

often stratified on the basis of political boundaries (Graves *et al.* 1992; Moritz *et al.* 1997). Such

a stratification may be chosen when the origin of samples is only known down to the level of the

geopolitical unit (eg. state or province), or in applied studies when the jurisdictional limits of

agencies charged with the management of a species coincide with political boundaries. Second,

samples are divided so as to ensure equal sample size among all of the putative populations.

Typically these strata are rather large because researchers realize that increasing the number of

samples per strata will increase statistical power and therefore make it more likely that they will

obtain statistically significant results.

A final, and perhaps most common, method of stratifying a data is to simply place

hypothesized boundaries in areas where there are gaps in the distribution of samples. Because

sampling is often difficult, few investigations of population structure have a deliberate 'sampling

design'; rather, samples are gathered opportunistically, with the highest sampling effort

concentrated in areas that are easy to access. This method of stratification can be particularly

misleading when researchers only publish a map of the distribution of samples (which may be

patchy and discontinuous) and no map or description of the actual distribution of the species

(which may be continuous).

Interpretation of the results of a hypothesis test of population structure can be difficult. A

finding of statistically significant differentiation is taken by many researchers as evidence that

their hypothesized structure accurately reflects the underlying spatial structure (eg. Moritz *et al.*

1997; Jagielo *et al.* 1996; Johannesen *et al.* 1996), while others interpret such results simply as

an indication that population genetic structuring is present within the species (Graves and

McDowell 1994; Jansen Van Vuuren and Robinson 1997; Houlden *et al.* 1999). Likewise,

failure to detect significant differentiation is often used to argue that rates of gene flow are high

enough to render the population panmictic (Van Treuren *et al.* 1999; Gosselin *et al.* 1999; Scoles

and Graves 1993; Soroka *et al.* 1997) when in fact the lack of significance may simply indicate

that the hypothesized structure is incorrect or that sample sizes are too low (Dizon et al., 1992;

Taylor and Dizon, 1996).

Appreciating the potential pitfalls in interpreting the results of a hypothesis test requires

an understanding of a few statistical terms and how they relate to detecting population structure.

The test typically consists of comparing an observed measure of differentiation between strata

(the putative populations) to the distribution expected for that measure if the data actually

represent a random sample from a panmictic population. Thus, the test is comparing the null

hypothesis (H_o) that the data come from a single panmictic unit to the alternative hypothesis (H_A)

that the data do not come from a panmictic unit. Despite the fact that the null hypothesis of

panmixia, an equal probability of every individual to mate with any other individual, is extremely

unlikely for large, continuously distributed populations, it is often difficult to reject this null

hypothesis. Consider the case where the data do not come from a single panmictic unit and thus

H_o is false. The probability of our correctly rejecting H_o as false (called statistical power)

depends on two things: sample size and effect size. When examining population structure, the

effect size is the degree of genetic differentiation between populations. The expected level of

population differentiation (F_{st}) is determined by the effective population size (N) and the

effective per generation migration rate (m), as illustrated by Wright's island model (Wright,1931)

$$F_{st} = \frac{1}{4Nm + 1}. \tag{1}$$

The precision of the distribution of expected F_{st} values will depend on the sample size. If the

observed value of F_{st} (effect size) is high enough (eg. point C on Figure 1), then power to reject

the null hypothesis will be high even if precision is low. However, when the effect size is very

low, either due to large population size or high dispersal (point A on Figure 1), statistical power

will also be quite low, even if precision is increased through greater sampling.

We examined the efficacy of hypothesis testing as a means of uncovering the spatial structure within a region. If hypothesis testing is a good method for identifying spatial structure, then we would expect the hypothesis test to be most likely to yield a significant result when the hypothesis corresponds to the actual spatial structure. To reproduce the problem of uncovering spatial structure in a continuously distributed species, we used a stepping-stone model of five populations where dispersal between neighboring populations was known. Thus, the locations of the boundaries between populations and the rates of dispersal across those boundaries were known. We could therefore examine the performance of a hypothesis-testing approach to investigating population structure by dividing the computer-generated data into different hypothesized population structures and determining whether or not the lowest p-value and highest statistical power were obtained when the hypothesized structure corresponded to the actual population structure. We tested hypothesized structures using Analysis of Molecular Variance (AMOVA; Excoffier *et al.*, 1992), an approach that uses an analog of Wright's F_{st} (Wright, 1931) to measure genetic differentiation and assigns statistical significance using a permutation approach. Our results should nonetheless be robust to any method of testing hypotheses about population structure, as will be discussed.

METHODS

We wished to examine the influence that the number of putative populations has on our ability to detect genetic differentiation. To do this, we drew samples from populations of known spatial structure and divided those samples into different numbers of putative populations. We then recorded the probability of detecting differentiation as a function of the number of putative populations. We generated genetic data with a known structure using a computer simulation.

The use of a computer simulation rather than analytical formulae in generating our data allowed

us to incorporate two forms of uncertainty that are inherent to analyses of population structure:

sampling error and temporal variation in the degree of differentiation. Sampling error affects the

precision of our estimate of the actual degree of differentiation between two hypothesized

populations. If we draw three separate samples from a pair of populations and use them to

estimate the degree of differentiation, we will obtain three different estimates of differentiation.

The variance of the three estimates will depend on the number of samples. We were able to

account for this source of uncertainty using computer generated data by drawing multiple

samples from the model populations and averaging our results.

The actual degree of differentiation between two populations will fluctuate over time due

to the inherently stochastic nature of genetic drift and gene flow, even if the sizes of both

populations and the rate of dispersal between them remains constant (Taylor *et al.*, in press;

Figure 2). This temporal variation in the effect size means that a researcher's power to detect

genetic differentiation will also vary depending on when samples are collected. In order to

account for this source of uncertainty in our ability to detect population structure using a

hypothesis-testing approach, we drew samples from our model populations at multiple points in

time so that we could average our results over the range of possible degrees of differentiation that

could result from a given dispersal rate and carrying capacity.

The model we used tracked the evolution of mitochondrial haplotypes in five populations

arranged in a stepping stone manner. The mode was nearly identical to the one used by Taylor *et*

al. (1997) (see Appendix for model details). Dispersal was allowed only between adjacent

populations. Carrying capacity was the same for all five populations, and the rate of dispersal

was equal between all pairs of adjacent populations. At the beginning of each simulation, all

individuals in all populations were assigned the same mitochondrial haplotype. The populations

were then allowed to evolve for 200,000 years. After 150,000 years, the populations reached a

stochastic equilibrium, as evinced by the fact that the average number of haplotypes in the

populations and the average degree of differentiation between them had stabilized. We recorded

the complete haplotype profile (the sequence of each haplotype and its frequency in all five

populations) from the simulation every 500 years beginning at year 150,000. Thus, we obtained

100 haplotype profiles for each combination of carrying capacity and dispersal rate.

   We used the data generated by the stepping stone model to examine the relationship

between the number of groups defined in the initial hypothesized population structure and the

likelihood of finding statistically significant differentiation between those groups. For each

haplotype profile we drew 18 samples at random without replacement from each of the five

populations, for a total of 90 samples. We then divided those 90 samples into two, three or five

equally-sized groups. We calculated the average pairwise genetic differentiation between

adjacent groups for the three hypothesized structures using the statistic Φ_{st}, an analog of Wright's

F_{st} (Excoffier et al., 1992). We used a permutation test (100 permutations) to assign a p-value to

each measure of differentiation. We repeated this procedure five times for each of the 100

haplotype profiles we had generated for a given combination of dispersal rate and carrying

capacity and recorded the average observed p-value. In addition, we recorded two other

quantities derived from the statistical analyses: average differentiation (Φ_{st}) and power.

Statistical power is the probability of correctly rejecting the null hypothesis (panmixia, in this

case) when it is false, and was therefore measured as the proportion of the analyses in which the

p-value was less than 0.05 (i.e. $\alpha = 0.05$). Because our estimates of average p-value, average Φ_{st}

and power are based on 500 different samples taken from 100 different points in time, they take

into account both sampling error and temporal variation in the degree of genetic differentiation.

Our ability to detect genetic differentiation depends not only on the degree to which

groups differ, but also on the sample size. Because we are dividing a constant number of

samples into different numbers of strata, the number of samples per strata depends on the number

of strata defined. Consequently, any differences we find in our ability to detect differentiation

between two, three or five groups could be due to either differences in the average effect sizes or

differences in sample sizes, or both. Because researchers investigating population structure also

have a constant number of samples, the number of samples per strata in their investigations will

also depend upon the number of strata defined. Therefore, our analysis should provide an

accurate representation of the relationship between number of strata defined and ability to detect

differentiation. Nonetheless, we wished to determine whether any differences in ability to detect

differentiation were due entirely to differences in the number of samples per strata or if the

average degree of differentiation also differed depending upon the number of strata defined. In

order to do this, we performed an additional analysis in which we drew 48 samples from each of

the five model populations, divided those samples up correctly into five groups, and measured

power as described above. This allowed us to compare the results obtained when the region is

divided up in a way that accurately reflects the population structure to those obtained when the

region is divided too coarsely, while keeping the number of samples per putative population

roughly equal (48 and 45, respectively). Any differences in our ability to detect genetic

differentiation in this analysis will be due entirely to differences in the effect size associated with

different degrees of stratification.

RESULTS

We found that our ability to detect significant genetic differentiation is inversely related

to the number of groups into which we divided the samples (Figures 3a and 3b). When no

dispersal is allowed between adjacent populations, the average p-value is less than 0.001 and

statistical power is one regardless of how samples are divided. For all other dispersal rates we

examined, the average p-value is lowest (Figure 3a) and power to detect differentiation is highest

(Figure 3b) when the samples are divided into just two putative populations rather than being

correctly divided into 5 putative populations. When the annual dispersal rate is 0.001 and

carrying capacity is 300 ($Nm = 1.2$), the average p-value between the five hypothetical

populations is 0.069 and power to detect differentiation is 0.73. However, the average p-value

drops to only 0.005 and power increases to nearly 0.97 when the samples are divided too coarsely,

into only two groups. The average p-value and power are intermediate (0.019 and 0.90,

respectively) when the samples are divided into three groups. Power declines and the average p-

value increases with increasing dispersal, as expected. At the highest dispersal rate examined (d

$= 0.01$), the difference in power between two and five hypothetical populations is slightly greater.

For example, with a carrying capacity of 300 ($Nm = 12$), power is 0.59 for two hypothetical

populations and 0.21 for five hypothetical populations, and the average p-value increased from

0.116 to 0.319 when we increased the number of groups from two to five.

Holding constant the number of samples per hypothetical population reduces the disparity

in average p-values (Figure 3a) and power to detect differentiation (Figure 3b) when samples are

divided into two versus five groups. However, when dispersal rates are high, the probability of

detecting differentiation is still considerably higher when the samples are divided exactly in half

than when they are divided correctly, even after correcting for differences in sample size.

When annual dispersal rates equaled 0.001 or more, the average degree of genetic

differentiation between adjacent putative populations, as measured by Φ_{st}, showed a pattern

similar to that seen for power to detect differentiation: Φ_{st} decreases with increasing dispersal

and is consistently highest when samples are divided into only two hypothetical populations

(Figure 4). However, when no dispersal occurs between adjacent populations, Φ_{st} is extremely

high when the samples are divided into five groups and lowest when they are divided into only

two groups. Φ_{st} drops precipitously for five groups when the annual dispersal rate is increased

from zero to 0.001. Increasing the sample size had no effect on the average value of Φ_{st} observed

between adjacent groups when the samples were divided into five groups.

The relationships between hypothesized structure and power, average p-value and genetic

differentiation were consistent across all three carrying capacities examined. Both power (Figure

5a) and the degree of genetic differentiation (Figure 5b) are highest, and average p-values are

lowest, when carrying capacities are low, as expected.

DISCUSSION

In this study we looked at the relationship between the accuracy of hypothesized

boundary locations and three statistical quantities resulting from a hypothesis test of population

structure: the average degree of statistical significance (measured as the p-value), statistical

power, and the average degree of genetic differentiation. When dispersal is nonzero, all three of

these measures are higher when a region is divided coarsely into just two strata than when the

region is divided up in a way that accurately reflects the underlying population structure. There

are two reasons for this result. The first, and most obvious, is that by dividing a constant number

of samples into more strata, you reduce the number of samples per strata. This reduction in

sample size reduces the power of the pairwise comparisons between strata (Figure 3b). However,

even when sample sizes are equalized, power is still highest when a region is divided exactly in

half, though the disparity is less. Furthermore, a reduction in sample size should not affect the

average degree of genetic differentiation observed between two groups, and therefore does not

explain why average differentiation is also higher when a region is divided coarsely rather than

accurately. This suggests that differences in the number of samples per strata are only partly

responsible for the inverse relationship between power and the number of strata defined by the

researcher.

The second explanation for the increase in power and decrease in average p-value with

coarser grouping is that when a region is divided into only two strata, samples from the two

extremes of the range are placed in adjacent groups. For instance, when the five populations

from our computer simulation are divided exactly in half, we are comparing a group containing

animals from population 1 to a group containing animals from population 5. Populations 1 and 5

are at opposite ends of the range in question, and thus are maximally differentiated. Comparing

groups containing these two distant populations inflates the overall degree of genetic

differentiation, resulting in the inverse relationship we observed between average differentiation

and the number of strata defined. The degree of differentiation describes the effect size in an

analysis of population structure. Because both the p-value and power are functions of effect size,

an increase in the average genetic differentiation results in an increase in power and a lower p-

value (i.e. higher degree of statistical significance). This explains why power is higher and

average p-value lower when we divide our samples into just two strata rather than five, even after

having equalized sample size.

While our results were derived by testing for differentiation via Analysis of Molecular

Variance (AMOVA; Excoffier *et al.*, 1992), our conclusions apply to any hypothesis-testing

approach to examining population structure. The patterns we observed are attributable to the

effect that changing the number of strata has on the number of samples per strata and the degree

of differentiation between strata (the effect size). The number of samples per strata is not a

function of the statistical test used in the analysis, though the degree to which it affects the results

is. The amount by which the degree of differentiation increases when the samples are grouped

more coarsely will depend upon the measure of differentiation used; however, so long as there

are not fixed differences between populations (as when $Nm = 0$), dividing the samples more

coarsely and thereby including in the comparison more distantly related samples will increase the

degree of differentiation, regardless of the measure used. Thus, the patterns we observed

between the number of strata defined by the researcher and power, p-value, and degree of

differentiation should be robust to any hypothesis-testing method of measuring differentiation,

though the strength of the pattern may differ.

The results of this study call into question the suitability of hypothesis testing as a means

of investigating population structure, particularly when the actual locations of population

boundaries are of interest. The difficulty comes in the interpretation of the results of a hypothesis

test. A hypothesis test addresses the question, "Is there significant genetic differentiation across

this hypothetical boundary?", while researchers are often asking the question, "Is there a barrier

to dispersal at this location?" Though there certainly is a strong relationship between dispersal

and genetic differentiation, the answers to the aforementioned questions are often not the same.

As demonstrated in this study, significant differentiation is often more likely to be observed

between strata that do not reflect the underlying population structure. Thus, the finding of

significant differentiation among strata simply indicates that population structure is present, not

that the boundary locations specified by the researcher accurately reflect restrictions in dispersal.

Likewise, a lack of significance in a hypothesis test cannot necessarily be taken as an indication

that the hypothesized boundaries do not reflect the actual population structure, nor as an

indication that the samples come from a single, panmictic population. Our results show that

when samples are divided correctly into five putative populations, Φ_{st} decreases precipitously

when the number of migrants per year (Nm) is increased from 0 to 2.4. When no dispersal occurs

between populations, over 96% of the total genetic variation is between populations. When Nm

is 2.4, only 21% of the total genetic variation is between populations. This dramatic decrease

illustrates the capacity of dispersal to prevent the development of genetic differentiation, even

when dispersal rates are low from a demographic point of view.

Even if the dispersal rate between two populations is low enough to allow them to

become genetically differentiated, if sample size is low, the power to detect differentiation

between those two populations will be correspondingly low as well. In this study, with an annual

dispersal rate of 0.001 and a carrying capacity of 300, average statistical power to detect

differentiation between adjacent roups was only 0.73 and the average p-value was 0.069 when

there were only 18 samples per population. However, power increased to 0.92 and average p-

value decreased to 0.014 when the sample size was increased to 48. This result illustrates a fact

that is well known to but often overlooked by researchers: an inability to show statistical support

for a hypothesis does not mean that the null hypothesis (panmixia, in this case) is correct.

How can the results of a hypothesis test of population structure be properly interpreted? A finding of statistically significant differentiation indicates that there is structure within the population at at-least the level of subdivision investigated by the study, though the analysis provides no support for the locations of boundaries used in the analysis. It is also possible (and perhaps likely) that there is further subdivision on a finer scale, even if significant differentiation is not detected on a finer scale. Furthermore, an inability to detect significant differentiation at any scale does not mean that there is no population structure. A lack of significance could be due to small sample size, incorrect boundary placement, historical effects, or dispersal rates high enough to prevent the development of strong differentiation, yet low enough to render populations demographically independent.

When determining whether to use a hypothesis-testing approach to identifying population structure, we recommend that researchers consider the purpose of their study. What is the goal of the study? Are boundaries likely to be taken literally? What are the potential consequences of such a literal interpretation? When the goal is simply to determine whether or not population structure exists and the location of the population boundaries are not of particular interest, then hypothesis-testing is an appropriate tool. However, when the purpose of a study is to identify demographically independent units between which dispersal is restricted, using a hypothesis testing approach is not advisable. Rather than co-opting techniques originally designed for evolutionary studies, we need to develop new approaches to invigatesting population structure

that focus on estimating dispersal rates between putative populations rather than using genetic

differentiation as a proxy for restricted dispersal.

APPENDIX I: Model Details

We used a stepping-stone model that was nearly identical to the one used by Taylor *et al.* (1997). The model was comprised of five populations, all of which had the same carrying capacity. Because we were simulating mitochondrial DNA, only effective adult females were included in the model. Carrying capacities were set at either 100, 300 or 1,000 effective adult females. Population growth followed a birth-death model with the probability of an individual dying equal to 0.2 and the probability of an individual giving birth ranging from 0.2 to 0.25 in a linear density-dependent manner. These rates result in a generation time of approximately 4 years. Dispersal occurred between adjacent populations at an annual rate d. The use of an annual dispersal rate rather than a per-generation rate simplified the model by placing all demographic rates (birth, death, and dispersal) on the same time scale. We examined dispersal rates ranging from 0 to 1% per year. A translation of these annual dispersal rates (d) to the more familiar generational dispersal rates (m) and number of migrants per generation (Nm) is given in Table 1.

We simulated the evolution in these five populations of mitochondrial haplotypes. We chose to simulate mitochondrial sequence data because it is commonly used in studies of population structure and is particularly useful in identifying demographically independent units (Moritz, 1994; Avise, 1995). A brief examination of the literature showed that analyses of population structure using sequence data typically involve anywhere from 250 to 1200 basepairs, yielding between 20 and 60 variable sites. We therefore chose to model 40 variable basepairs of mitochondrial DNA. Unlike Taylor *et al.* (1997), we used a constant mutation rate of 0.0001 for all 40 sites, with transition probabilities between different nucleotides based on those found in the northern right whale dolphin (*Lissodelphis borealis*). Taylor *et al.* (1997) found that when

mutation rates are determined on a site-specific basis using a gamma distribution, many sites

mutated too quickly or too slowly to be phylogenetically informative. The mutation rate we

chose resulted in all variable sites being phylogenetically informative, thereby maximizing the

usefulness of the genetic data for determining population structure. Thus, our results represent a

best-case scenario in terms of the quality of the genetic data available.

For each simulation, all five populations were initialized at carrying capacity with a single

ancestral haplotype and allowed to evolve for 200,000 years. The sequence of the ancestral

haplotype was randomly generated using the nucleotide frequencies found in *L. borealis* (A –

0.334, T – 0.310, C – 0.124, G – 0.233). After 150,000 years, the populations had reached a

stochastic equilibrium. We recorded the haplotype profile (the sequence of each haplotype and

its frequency in all five populations) from each simulation every 500 years between years

150,000 and 200,000 of the simulation, providing us with 100 haplotype profiles, each from a

different point in time, for each combination of dispersal rate and carry capacity. By collecting

data from multiple time points in the simulation, we were able to take into account the temporal

variation in the degree of differentiation between adjacent populations due to genetic

stochasticity (Figure 2).

REFERENCES

Avise JC (1995) Mitochondrial DNA polymorphism and a connection between genetics and

demography of relevance to conservation. *Conservation Biology*, **9**: 686-690.

Bérubé M, Aguilar A, Dendanto D, Larsen F, Notarbartolo di Sciara G, Sears R, Sigurjónsson J,

Urban-R. J, Palsbøll, PJ (1998) Population genetic structure of North Atlantic,

Mediterranean Sea and Sea of Cortez fin whales, *Balaenoptera physalus* (Linnaeus

1758): analysis of mitochondrial and nuclear loci. *Molecular Ecology*, 7:585-599.

Brown Gladden JG, Ferguson MM, Clayton JW (1997) Matriarchal genetic population structure

of North American beluga whales *Delphinapterus leucas* (Cetacea: Monodontidae).

Molecular Ecology, **6**:1033-1046.

Dizon A, Lockyer C, Perrin WF, DeMaster DP, Sisson J (1992) Rethinking the stock concept: a

phylogeographic approach. *Conservation Biology*, **6**: 24-36.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of Molecular Variance Inferred From

Metric Distances Among DNA Haplotypes: Application to Human Mitochondrial DNA

Restriction Data. *Genetics*, **131**:479-491.

Gosselin L, Jobidon R, Bernier L (1999) Genetic variability and structure of Canadian

populations of *Chondrostereum purpureum*, a potential biophytocide. *Molecular Ecology*,

8:113-122.

Graves JE, McDowell JR, Beardsley AM, Scoles DR (1992) Stock structure of the bluefish

Pomatomus saltatrix along the mid-Atlantic coast. *Fishery Bulletin*, **90**: 703-710.

Graves JE, McDowell JR (1994) Genetic analysis of striped marlin (*Tetrapturus audax*)

population structure in the Pacific Ocean. *Canadian Journal of Fisheries and Aquatic

Science*, **51**:1762-1768.

Houlden BA, Costello BH, Sharkey D, Fowler EV, Melzer A, Ellis W, Carrick F, Baverstock PR,

Elphinstone MS (1999) Phylogeographic differentiation in the mitochondrial control

region in the koala, *Phascolarctos cinereus* (Goldfuss 1817). *Molecular Ecology*, **8**:999-

1011.

Hudson RR, Boos DD, Kaplan NL (1992) A Statistical Test for Detecting Geographic

Subdivision. *Molecular Biology and Evolution*, **9**:138-151.

Jagielo TH, LeClair LL, Vorderstrasse BA (1996) Genetic variation and population structure of

lingcod. *Transactions of the American Fisheries Society*, **125**:372-386.

Jansen Van Vuuren B, Robinson TJ (1997) Genetic population structure in the yellow mongoose,

Cynictis penicillata. *Molecular Ecology*, **6**: 1147-1153.

Johannesen J, Veith M, Seitz A (1996) Population genetic structure of the butterfly *Melitaea

didyma* (Nymphalidae) along a northern distribution range border. *Molecular Ecology*,

5:259-267.

Latter BHD (1973) The island model of population differentiation: a general solution. *Genetics*,
106:293-308.

Moritz C (1994) Applications of mitochondrial DNA analysis in conservation: a critical review.

Molecular Ecology, **3**:401-411.

Moritz C, Heideman A, Geffen E, McRae P (1997) Genetic popultion structure of the Greater

Bilby *Macrotis lagotis*, a marsupial in decline. *Molecular Ecology*, **6**:925-936.

Palacios C, González-Candelas F (1997) Analysis of population genetic structure and variability

using RAPD markers in the endemic and endangered *Limonium dufourii*

(Plumbaginaceae). *Molecular Ecology*, **6**:1107-1121.

Scoles DR, Graves JE (1993) Genetic analysis of the population structure of yellowfin tuna,

Thunnus albacares, from the Pacific Ocean. *Fishery Bulletin*, **91**:690-698.

Soroka M, Ziclinski R, Polok K, Swierczynski M (1997) Genetic structure of *Dreissena*

polymorpha (Pallas) population in Lake Insko, North-Western Poland. *Polskie Archiwum*

Hydrobiologii, **44**:505-515.

Takahata N, Palumbi SR (1985) Extranuclear differentiation and gene flow in the finite island

model. *Genetics*, **109**:441-457.

Taylor BL, Dizon AE (1996) The need to estimate power to link genetics and demography for

conservation. *Conservation Biology*, **10**:661-664.

Taylor BL, Chivers SJ, Dizon AE (1997) Using statistical power to interpret genetic data to

define management units for marine mammals. *Molecular Genetics of Marine Mammals*,

3:347-364.

Taylor, B. L., S. J. Chivers, S. Sexton, A. E. Dizon. In press. Evaluating Dispersal Estimates

Using mtDNA Data: Comparing Analytical and Simulation Approaches. *Conservation*

Biology.

Van Treuren R, Bijlsma R, Tinbergen JM, Heg D, Van de Zande L (1999) Genetic analysis of

the population structure of socially organized oystercatchers (*Haematopus ostralegus*)

using microsatellites. *Molecular Ecology*, **8**:181-187.

Wright S (1931) Evolution in Mendelian populations. *Genetics*, **16**:97-159.

Table 1. The value of *Nm* corresponding to various combinations of carrying capacity (*K*) and annual dispersal rate (*d*). The per generation migration rate (*m*) was calculated based on a generation time of four years.

d	*m*	*K*		
		100	300	1000
0.002	= 0.008	0.8	2.4	8
0.006	= 0.024	2.4	7.2	24
0.01	= 0.04	4.0	12	40

Figure 1. The expected distribution of genetic distances under the null hypothesis when sample

size is large, yielding a precise estimate of the expectation, and when sample size is small,

yielding an imprecise estimate. Three hypothetical observed values of differentiation are shown

along the x-axis. For case A, we would have low power to detect differentiation regardless of

precision because the effect size is low due to either large population size or high dispersal. In

case C, because the effect size is large, power to reject the null hypothesis will be high even if

precision is low. In case B, the effect size is intermediate and power will depend upon the

precision of the distribution of expected values.

Figure 2. Observed level of differentiation (F_{st}) between two populations over time. The data

were obtained by modeling two populations with carrying capacities of 300 effective adult

females and an annual dispersal rate of 0.002. F_{st} fluctuates considerably, despite the fact that

both carrying capacity and dispersal rate are constant. The dotted line indicates the expected

value of F_{st}, which was calculated using a modification of equation 1 that takes into account the

finite number of populations (Latter, 1973) and the haploid nature of mitochondrial DNA

(Takahata and Palumbi, 1985).

Figure 3. a) Average p-value in a hypothesis test of population structure and b) power to detect

differentiation as a function of dispersal rate between adjacent populations when samples are

broken into two (♦), three (■) or five (▲) strata. The x-axis shows both annual dispersal rate and,

in parentheses, effective number of migrants per generation. Samples were drawn from five

model populations arranged in a stepping-stone manner, each with a carrying capacity of 300

effective adult females. The solid lines show values obtained with a total of 90 samples (18 from

each population), The dotted line represents results obtained when a total of 240 samples (48 from each population) are divided correctly into five strata, and allows us to determine the impact that number of strata has on power and average p-value while holding the number of samples per strata constant.

Figure 4. Average degree of genetic differentiation, as measured by ϕ_{st}, observed between groups as a function of the dispersal rate between adjacent populations. The x-axis shows both annual dispersal rate and, in parentheses, effective number of migrants per generation. The model populations had carrying capacities of 300 effective adult females. Samples were broken into two ($\blacklozenge$), three ($\blacksquare$) or five ($\blacktriangle$) groups.

Figure 5. a) Power to detect differentiation and b) average value of ϕ_{st} observed between groups as functions of the dispersal rate between adjacent populations and the carrying capacity of populations. The x-axis shows both annual dispersal rate and, in parentheses, effective number of migrants per generation. A total of 90 samples were drawn from the five populations in the model and were divided correctly into five groups. The carrying capacity of each population in the model was 100 (dotted line), 300 (solid line) or 1000 (dashed line). Both average Φ_{st} and power decreased with increasing carrying capacity, as expected.

Figure 1. Will be distributed as hardcopy at JARPN meetings or by FAX if requested) *See last page for the Figure*
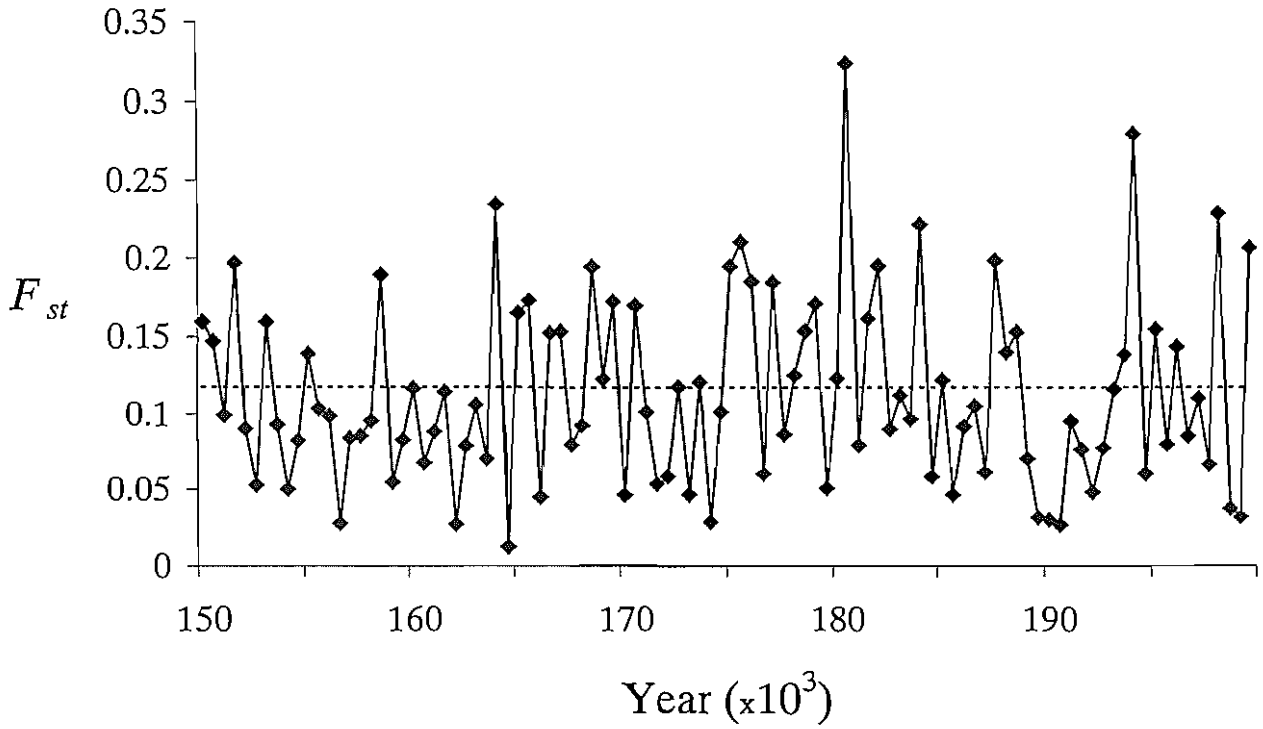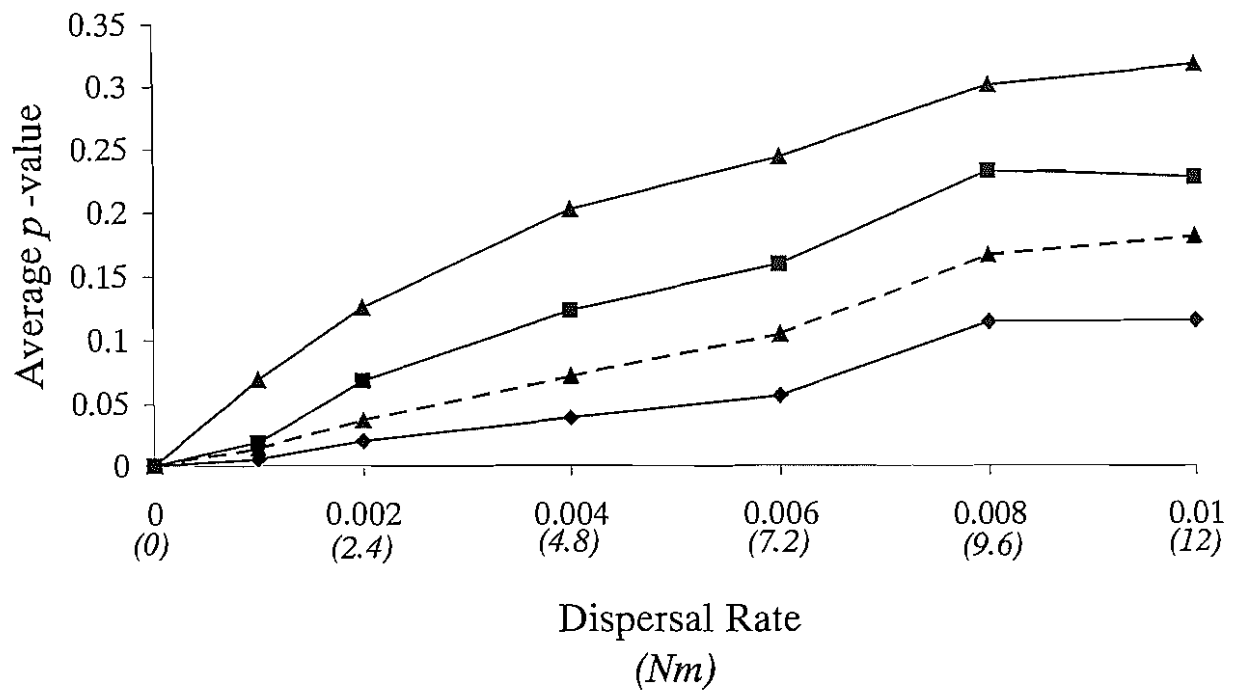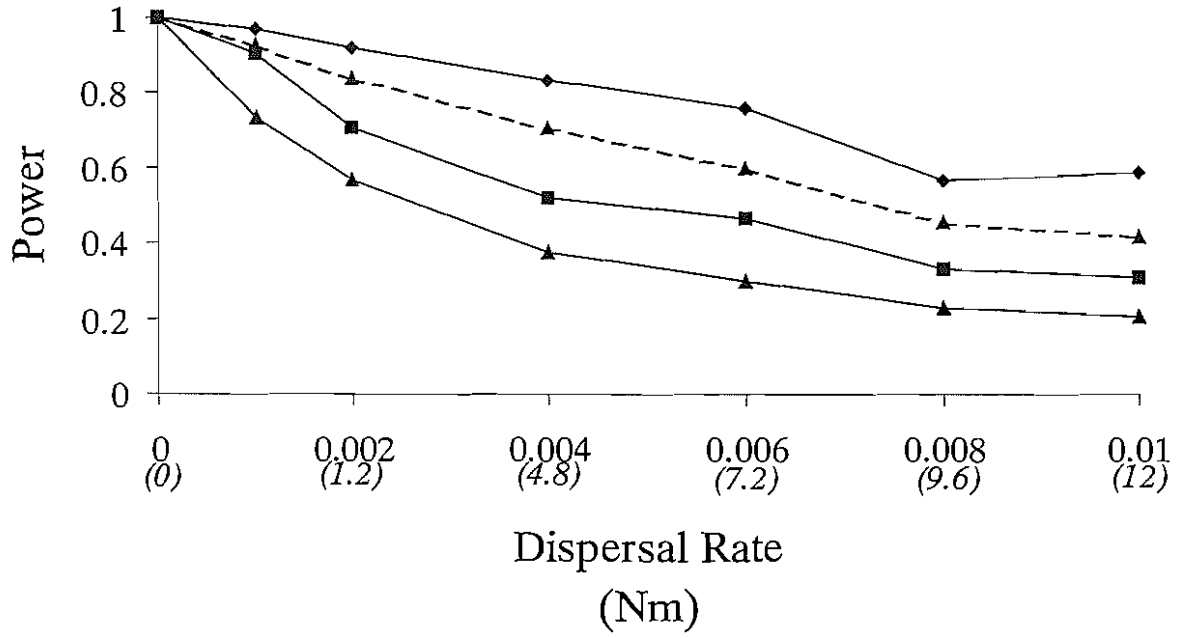(taylor@caliban.ucsd.edu)

Figure 2.

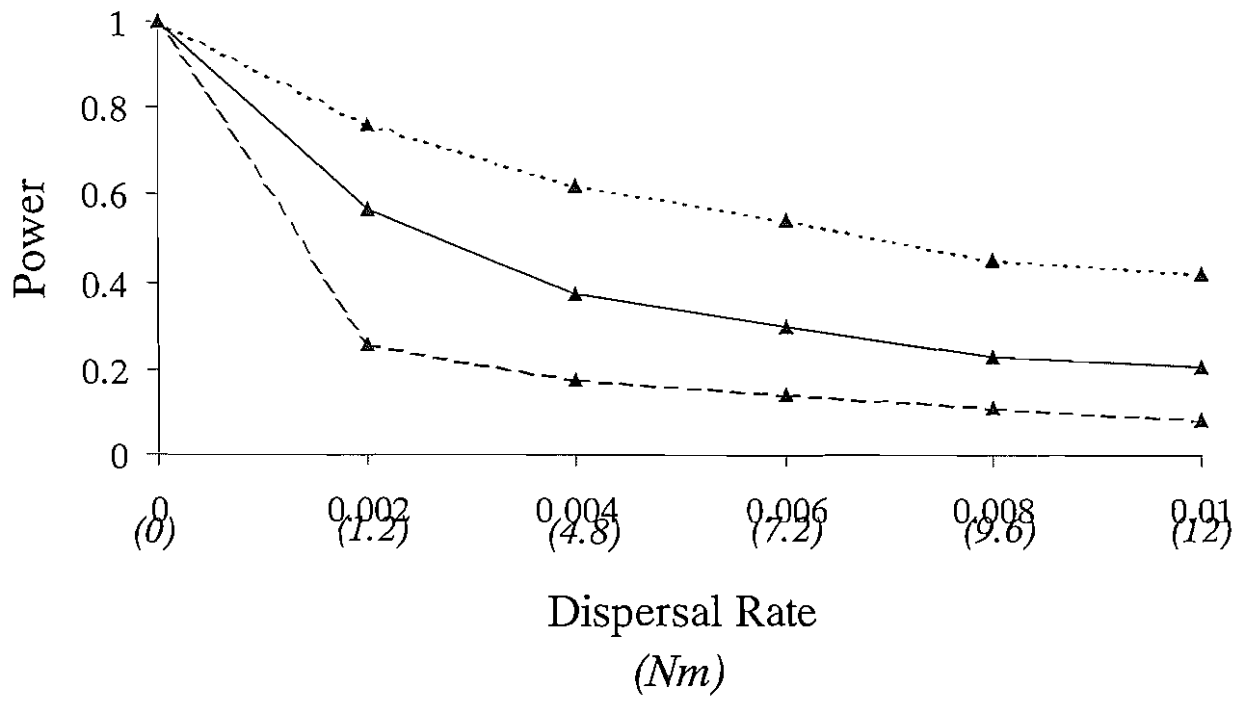

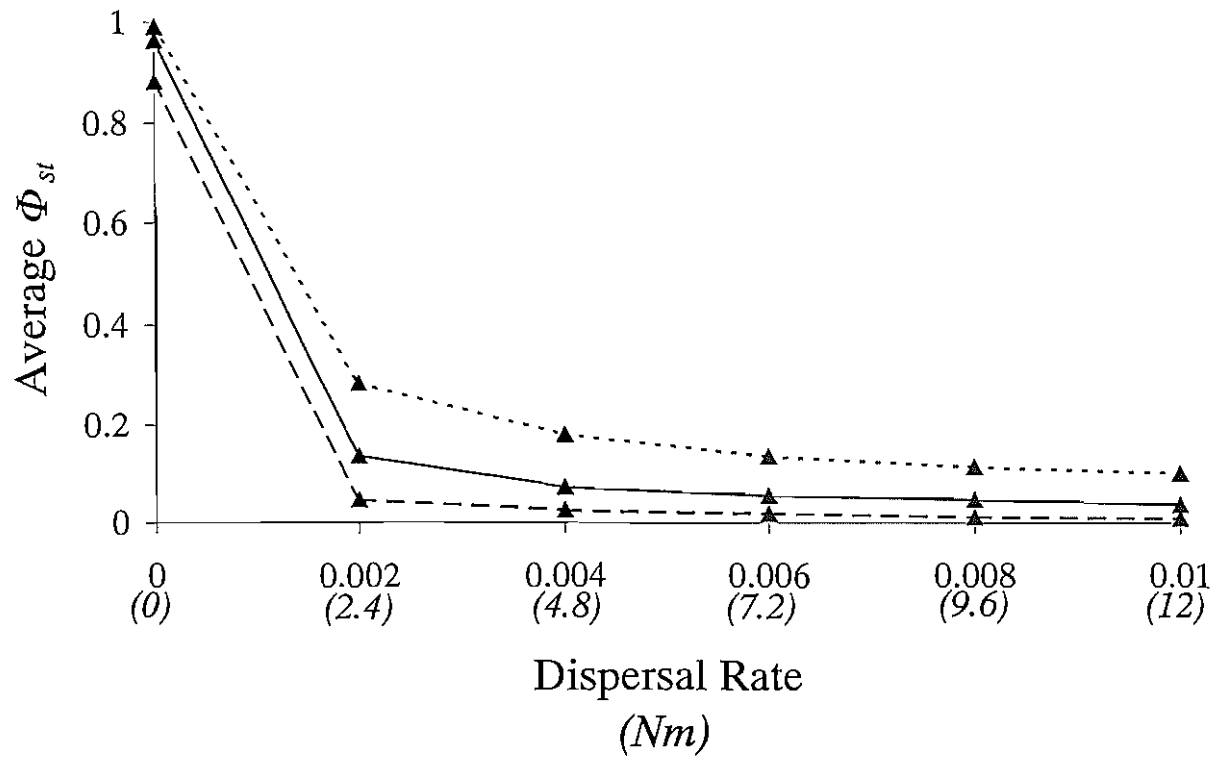F_{st} plotted against Year ($\times 10^3$)

Figure 3a.

Figure 3b.

Figure 4.
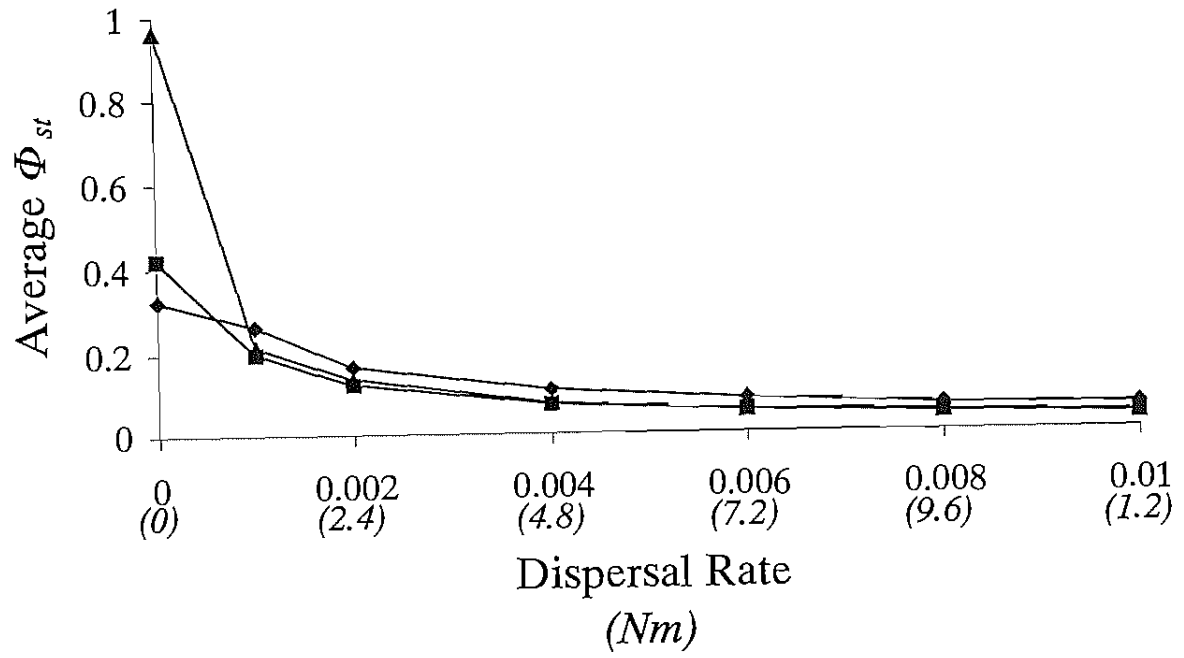
Figure 5a.

Figure 5b.

Figure 1

Figure 1