Separating the Wheat From the Chaff: Patterns of Genetic Differentiation in High Gene Flow Species

R. S. Waples

In many marine species, high levels of gene flow ensure that the genetic signal from population differentiation is weak. As a consequence, various errors associated with estimating population genetic parameters that might normally be safely ignored assume a relatively greater importance. This fact has important implications for the use of genetic data to address two common questions in fishery conservation and management: (1) How many stocks of a given species are there? and (2) How much gene flow occurs among stocks? This article discusses strategies to maximize the signal:noise ratio in genetic studies of marine species and suggests a quantitative method to correct for bias due to a common sampling problem. For many marine species, however, genetic methods alone cannot fully resolve these key management questions because the amount of migration necessary to eliminate most genetic evidence of stock structure (only a handful of individuals per generation) will generally be inconsequential as a force for rebuilding depleted populations on a time scale of interest to humans. These limitations emphasize the importance of understanding the biology and life history of the target species—first, to guide design of the sampling program, and second, so that additional information can be used to supplement indirect estimates of migration rates based on genetic data.

Conservation of marine organisms is a challenging enterprise for several reasons. The habitat is immense—salt water covers over 70% of the world's surface—and presents a number of difficulties for observing, studying, and collecting marine species. Marine diversity and endemism are high, particularly among higher taxa, and associated conservation issues cover a broad spectrum. A particular concern is that aquatic organisms are the only remaining species on earth harvested in appreciable numbers from the wild for human consumption (Ryman et al. 1995). World catch of aquatic species is dominated by catches of finfish, but substantial harvest also occurs for many other groups, including crustaceans, mollusks, echinoderms, algae, mammals, and reptiles. Managing these harvests in a sustainable way has long been a concern of human societies. In fact, conservation biology as a discipline can trace its roots to efforts to conserve living natural resources exploited by humans (Primack 1993).

One outcome of these management efforts for aquatic species has been development of the stock concept. Numerous definitions of a stock can be found in the literature (reviewed by Carvalho and Hauser 1994), but in general the term refers to a group of organisms whose demographic/genetic trajectory is largely independent from other such groups. The stock concept is popular with fishery managers because efforts both for exploitation and conservation can be most effective when it is possible to identify and focus on individual stocks.

A challenge to application of the stock concept in the marine realm is the capability of many marine species for long-distance dispersal. Although ocean current patterns, sea floor topology, and other geographic features provide opportunities for isolation and differentiation of some species, for the most part the oceans of the world lack obvious barriers to migration and dispersal. Many marine species have evolved extended pelagic larval stages (Scheltema 1971; Waples 1987) and/or impressive migratory capabilities as adults (Harden Jones 1968; NRC 1994) to take advantage of the opportunities for dispersal in the marine realm.

Studies that directly evaluate dispersal in marine organisms (e.g., egg and larval

From the Conservation Biology Division, Northwest Fisheries Science Center, National Marine Fisheries Service, 2725 Montlake Boulevard East, Seattle, WA 98112.
I am grateful to Paul Bentzen, Joe Felsenstein, Mike Ford, Stewart Grant, Dennis Hedgecock, Yuri Kartavtsev, Jerry Pella, Alexander Pudovkin, Nils Ryman, and Elizabeth Thompson for insightful discussion and comments. Address correspondence to Dr. Waples at the address above or e-mail: robin.waples@noaa.gov. This paper was delivered at a symposium entitled "Conservation and Genetics of Marine Organisms" sponsored by the American Genetic Association at the University of Victoria, Victoria, BC, Canada, June 7, 1997.

 89:438–450

Table 1. The proportion of total gene diversity that occurs among populations ($\hat{F}_{ST}$) in 113 fish species

	Number of species	Mean number of populations	$\hat{F}_{ST}$ Mean	$\hat{F}_{ST}$ Median
Marine	57	6.4	0.062	0.020
Anadromous	7	13.1	0.108	0.081
Freshwater	49	5.9	0.222	0.144

Based on data in Ward et al. (1994), who summarized published reports that met the following criteria: a minimum of two subpopulations, a minimum of 15 randomly chosen loci, a minimum of 15 individuals per subpopulation, and availability of allele frequency data.

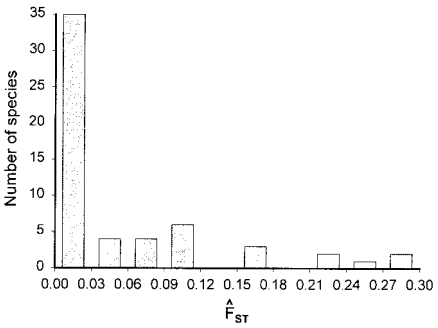

Figure 1. Distribution of $\hat{F}_{ST}$ values across 57 species of marine fishes, based on data in Ward et al. 1994. Although the mean F_{ST} for marine fishes is 0.062, the median value is only 0.02 (Table 1).

surveys, mark/recapture; reviewed by Pawson and Jennings 1996) can provide valuable insights, but they often are logistically difficult and have been relatively few in number. Furthermore, such studies provide evidence for the movement of individuals but do not necessarily elucidate the genetic consequences of dispersal. For these reasons, there has been considerable interest in the use of genetic data to provide indirect estimates of the importance of dispersal in marine organisms. In general, these data support the proposition that levels of migration and gene flow are relatively high in marine species (Shaklee and Bentzen, in press). For example, Ward et al. (1994) reviewed population genetic data for fishes and found that mean $\hat{F}_{ST}$ (a measure of the proportion of total gene diversity that is allocated among populations) was 0.062 for marine species, considerably lower than for anadromous or freshwater species (Table 1). [Here and throughout this paper, the "hat" indicates an estimate ($\hat{F}_{ST}$), while absence of a "hat" (F_{ST}) indicates a parameter.] Moreover, this mean value is inflated by relatively high ($\hat{F}_{ST}$) values for a few marine species (Figure 1). A more useful indicator is the median $\hat{F}_{ST}$, which is much smaller (0.020). Sixty percent of the marine fishes surveyed by Ward et al. (1994) have $\hat{F}_{ST}$ values less than 0.03, indicating relatively low levels of genetic differentiation and consistent with relatively high levels of gene flow among populations or stocks. Several articles in this volume also show relatively weak genetic differentiation in marine species (Gold and Richardson 1998; Graves 1998; Hoelzel 1998).

There are, however, some inherent limitations to the usefulness of genetic data for examining stock structure in high gene flow species, and these limitations have not been treated thoroughly in the literature. The difficulties arise because the genetic "signal" indicating stock structure is relatively weak for high gene flow species, and consequently various sources of noise in the analysis assume relatively greater importance (and require more rigorous attention) than would normally be the case. The noise can affect both the accuracy and precision of the estimates of population genetic parameters. Estimators are inaccurate (biased) if they exhibit systematic, directional departures from the true value of a parameter. Even if the estimator is unbiased, it still may be of limited use if its precision is low.

In this article I evaluate these limitations as they apply to two questions that have occupied those interested in marine conservation biology for many years: (1) How many stocks of a given species are there? and (2) How much gene flow occurs among stocks? For high gene flow species, each of these questions presents significant technical and statistical challenges. After first outlining some common problems, I will suggest strategies to maximize the value of genetic data for marine species and then give an example that demonstrates a quantitative way to deal with a common sampling problem.

The Stock Identification Problem

A common approach to the stock identification problem is to sample two or more putative stocks and examine each sample for a set of traits (genetic, morphological, meristic, phenotypic, etc.). Although this article focuses on the analysis of genetic markers, some of the same principles apply to other characteristics as well. Typically the next step in stock identification is to perform a statistical test to determine whether differences among the samples are statistically significant. Depending on the result of the test, a decision is made whether to manage for one or more stocks (Figure 2). If the test is significant, the null hypothesis of no differences among stocks is rejected and they are managed separately; if no significant differences are found, they are managed jointly.

This approach is appealing in its simplicity, but it is not without problems. The most basic difficulty is that there is little reason to expect a direct relationship between statistical significance and biological significance (Waples 1991b). There can be a disconnect between the two on either branch of the flow diagram. For example, the literature contains numerous examples of the failure of any particular class of markers to distinguish populations (or even species) that are known to be discrete (Allendorf et al. 1987). Taylor and Gerrodette (1993) and Dizon et al. (1995) have argued persuasively about the dangers of deciding (on the basis of a nonsignificant test result) to manage for a single stock, unless one has first evaluated the power of the test to detect differences between stocks if they do exist. The situation in which biologically significant differ-

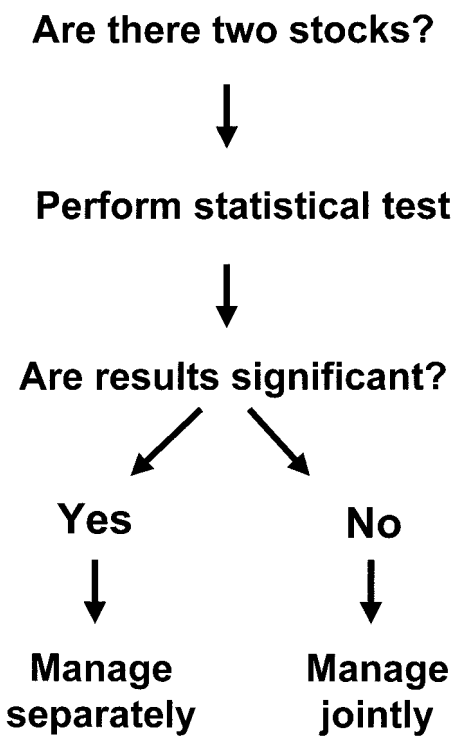

Figure 2. Flow diagram showing a typical approach to a common fishery management question: Are there one or two stocks?

ences exist but are not detected statistically leads to a type II error, and a manager following the right branch in Figure 2 risks making this error with respect to stock identification.

What is not so widely appreciated is that the converse can also be true: that is, not all statistically significant test results indicate biologically important differences. In fact, given enough data, statistically significant differences can be expected to occur routinely in comparisons of geographic samples, since at least some departures from complete panmixia will generally occur. Smith (1969) pointed this out long ago for human populations, and Waples (1989) made a similar point for temporally spaced samples. Basing natural resource decisions solely on the basis of statistically significant differences among stocks can lead to management as separate stocks when there is no strong biological basis for doing so, and managers following the left branch of Figure 2 risk making this type of error. Although it can be argued that this type of error will generally lead to a conservative approach to resource management and therefore should not be a serious concern to conservation biologists, such errors can have profound economic, social, and political consequences that represent a cost to society in other ways (primarily through foregone consumption of the resource and lost economic, social, and cultural benefits associated with the harvest and consumption). Furthermore, any consistent pattern of errors of this type, resulting in biologically unnecessary restrictions on human activities, will stiffen the resolve of skeptics and make it more difficult to accomplish sound resource management in the future. For these reasons, it is important to ensure that both types of errors with respect to stock identification—errors of commission as well as omission—receive careful scrutiny.

Errors that can occur in rejecting the null hypothesis can best be understood by a closer examination of the left branch of Figure 2. The null hypothesis applicable to the stock identification question is typically stated as follows: "H_0: There are no differences in measured traits among populations." This is equivalent to saying that the samples being compared have been drawn from the same population. Rejection of the null hypothesis can be associated with four different states of nature (Table 2), only one of which (outcome D) involves biologically important differences between stocks. Only in this case does rejection of the null hypothesis not lead to any statistical, biological, or management errors. Outcome A is a type I error (rejection of the null hypothesis when it is true), which occurs by chance with probability set by selection of the alpha level (typically 0.05) for the statistical test.

Table 2. Four outcomes that can occur when the null hypothesis of no population differences is rejected by a statistical test

Outcome	H_0	Population differences	Sampling	Error
A	True	None	Random	Type I
B	False[a]	None	Nonrandom	H_0 rejected for wrong reason
C	False	Biologically insignificant	Random	May lead to management error
D	False	Biologically meaningful	Random	None

[a] Test is statistically significant not because of differences between populations, but because sampling assumptions implicit in the null hypothesis were violated.

The remaining two outcomes—rejecting the null hypothesis when it is false but for biologically uninteresting reasons—lead to errors that do not appear to have been formally treated in the statistical literature. In outcome C, differences exist between the populations but are too small to be biologically meaningful for the question at hand. In this case, an error can occur if the test result triggers an action (e.g., management as separate stocks) that is not supported based on the actual biological differences. The probability of making this type of error is related to the power of the test (the probability of detecting differences between populations when they exist). Unfortunately, statistical power is determined not only by the magnitude of the differences between populations, but also the data richness (sample size, number of samples, and number of independent characteristics measured). Only the former is biologically meaningful for stock identification, but the latter can have a profound influence on the power of the statistical test. The probability of finding that small differences are statistically significant increases considerably if overall tests are used that combine information across multiple traits. Because of these factors, whether a statistical test yields a significant result does not by itself provide much biologically useful information, nor does the particular P value associated with the test (Berger and Selke 1987 and associated comments; Hilborn 1997).

Whether statistically significant differences are biologically meaningful will depend on the context. For example, the human sex ratio at birth is slightly skewed toward males (1.04:1 to 1.06:1 for British and American datasets; Cavalli-Sforza and Bodmer 1971; Lewontin 1995). Sample sizes are huge, and there is no question that the difference in rates of male and female births, although small in absolute terms, is real and statistically highly significant. The difference may be important to consider in some applications (e.g., in compiling actuarial tables), and changes in the sex ratio as the population ages can also provide important biological and evolutionary insights. However, for other purposes the difference is inconsequential. Although any departure from a 1:1 sex ratio will reduce effective population size (N_e), the reduction is trivial for a sex ratio this close to unity (N_e = 205.9 for a population of 106 males and 100 females versus N_e = 206 for a population with 103 of each sex).

Outcome B represents a more serious source of error for genetic studies of marine species. As in outcome C, the null hypothesis is false, but not because there are differences between the populations; rather, it is false because assumptions about sampling implicit to the null hypothesis have been violated. Statistical tests used to evaluate this null hypothesis based on discrete genetic data generally involve the assumption that sampling is multinomial—that is, that the samples have been drawn randomly from a population of infinite size. Violation of this assumption can lead to a significant test result even if there are no differences between populations. An error occurs in this situation if the scientist or manager falsely assumes that the significant test result indicates that multiple stocks exist. This type of error is important to consider because it is almost always the case that the sampling protocols violate the assumptions of the null hypothesis. To see why, consider the definition of a random sample: one in which every individual in the population has an equal chance of appearing. In artificial situations, this can be achieved by blindly drawing numbers or colored balls from a large container, or by using a random number algorithm to select individuals from an imaginary population in a computer model. However, sampling from bi-

ological populations is typically constrained in time and space, and often there are individuals in the population that have no chance of appearing in the sample. In most cases, therefore, the question is not whether the assumption about random sampling is violated, but how badly, and what are the consequences.

The other assumption about sampling inherent in most statistical tests—that the samples are drawn without replacement from an infinitely large population, or with replacement from a finite population- –is also routinely violated. In practice, most biological samples are taken without replacement, but since all real populations are finite, none of these are multinomial samples. Fortunately, these departures will tend to reduce sampling error compared to the null expectation, so they will not in general inflate the probability of a category B result.

There is, however, a more serious way in which biological sampling can be "random" but still depart from the assumptions of the standard statistical tests and, perhaps, lead to biologically invalid conclusions. This problem was first pointed out by Allendorf and Phelps (1981), who used as an example a single panmictic population of fish in a lake that randomly return to spawn in either of two tributary streams. Allendorf and Phelps considered two pairs of adults, one pair spawning in stream A and the other in stream B. If these four adults were sampled, it would not be unusual to find substantial frequency differences between the pairs, but the differences would generally not be statistically significant because of the small sample size involved. Often, however, it is not reproductive adults that are sampled but their juvenile progeny. If a researcher were to return the next year and sample a larger number (say 50) of juvenile fish resulting from the pairwise matings in each of the two streams, she would typically find the allele frequency differences in the two sets of parents were passed on (and inflated through an episode of genetic drift) to the offspring. However, in this case, with a relatively large sample size from each stream, the allele frequency differences could easily be statistically highly significant. As a result, the researcher would conclude (wrongly) that the two streams supported reproductively isolated populations (or stocks) of fish. This scenario (which I will term the "Allendorf–Phelps effect") will be discussed in more detail in the next section.

In summary, it should not be surprising, and is not necessarily biologically meaningful, to find statistically significant genetic differences among geographic samples. Rejecting the null hypothesis of no differences does not in itself represent a resolution of the stock identification problem: four different outcomes are possible when the null hypothesis is rejected, only one of which is associated with biologically meaningful differences among populations.

The Gene Flow Problem

For many marine species (especially those harvested by humans), levels of gene flow are a particularly important conservation issue. A common management questions is: If we deplete population (or stock) A through overharvest, will it be replenished by recruitment from elsewhere and, if so, how quickly? This question has been raised for a large number of marine species across many diverse taxa [for example, red drum (*Sciaenops ocellatus*; Gold and Richardson 1998); bluefin tuna (*Thunnus thynnus*; NRC 1994); walleye pollock (*Theragra chalcogramma*; Bailey et al., in press); loggerhead turtles (*Caretta caretta*; Bowen et al. 1992); cetaceans (Hoelzel 1991); green abalone (*Haliotis fulgens*; Tegner and Butler 1985); and spiny lobsters (*Panulirus marginatus*; Shaklee and Samollow 1984)].

A typical approach to answering this question is to estimate the rate of exchange among populations based on genetic data. The most common method is to estimate the parameter F_{ST} and use this result to estimate the migration parameter mN_e from the relationship (Wright 1943)

$$F_{ST} \approx 1/(1 + 4mN_e). \qquad (1)$$

The term mN_e—the product of migration rate (m) and effective population size—is the number of genetically effective migrants per generation received by each population. A graph of the parametric relationship between F_{ST} and mN_e based on equation (1) is shown in Figure 3A. In theory, this method provides a promising way of addressing the key question about how quickly an overharvested population will be replenished by migration from other populations. In practice, there are some serious, and often underappreciated, limitations to the usefulness of genetic data for this purpose with high gene flow species.

The first difficulty is that the numerous assumptions of the model are often ignored. The formula cited above applies to

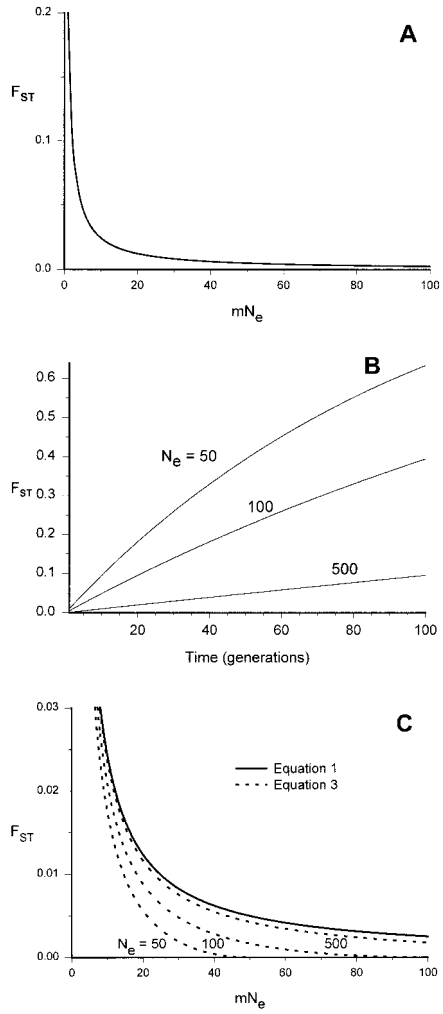

Figure 3. **(A)** The parametric relationship between F_{ST} and mN_e based on equation (1), assuming an island model of migration at equilibrium. **(B)** F_{ST} as an increasing function of time in an isolation model, based on equation (2) for various values of N_e. **(C)** As in (A), but including a comparison of the approximate [equation (1)] and exact [equation (3)] expressions for the relationship between F_{ST}, m, and N_e.

the island model of migration (Wright 1943), which is based on the following assumptions: (1) the number of subpopulations is infinite; (2) N_e is the same (and constant over time) in every subpopulation; (3) breeding is random within subpopulations; (4) generations are discrete; (5) m is constant over time and the same for every subpopulation; (6) m is "small;" (7) alleles are selectively neutral; and (8) there is no mutation. Clearly these assumptions will never be met in any application involving biological populations. Only some of these assumptions have been examined by sensitivity analysis to determine their effects on estimates of mN_e (see Waples 1986 for a review). Alternative formulas are available for several different migration patterns (reviewed by Felsenstein 1976; Slatkin 1985a), and one

of these may be more suitable than the island model for many applications involving marine species. In any case, it is important to recognize that the shape of the curve describing the actual relationship between F_{ST} and mN_e often will differ from that shown in Figure 3A.

Equation (1) also assumes that an equilibrium has been attained between migration and genetic drift. Strictly speaking, this will never occur, since a permanent equilibrium is only attained with an infinite global population size. With populations of finite number and size, F_{ST} in the migration model will eventually collapse to zero as all subpopulations become fixed for the same allele. Nevertheless, the system may attain a quasi-equilibrium for a long period of time, during which the relationship shown in equation (1) approximately holds (Nei et al. 1977). Alternative scenarios that must be considered are that the system has not yet reached equilibrium, or that the subpopulations were at one time connected by migration but are now completely isolated. In the latter case, F_{ST} increases asymptotically over time to its maximum value ($F_{ST} = 1$) according to the following equation (Nei and Chakravarti 1977):

$$F_{ST} = 1 - (1 - 1/(2N_e))^t \approx 1 - e^{-t/(2N_e)}. \quad (2)$$

In equation (2), t is elapsed time in generations after the subpopulations diverged from a common source. Figure 3B shows three curves (corresponding to different values of N_e) of the relationship between F_{ST} and divergence time in an isolation model. Researchers interested in evaluating the robustness of their genetic estimates of migration should realize that any empirical $\hat{F}_{ST}$ value that can be used to estimate mN_e using equation (1) is also consistent with a complete isolation model as described by equation (2). In fact, for any $\hat{F}_{ST}$ value there are an infinite number of combinations of t and N_e that will satisfy equation (2), as well as an infinite number of nonequilibrium scenarios involving migration. One factor that is favorable for the analysis of marine species is that the rate of approach to equilibrium increases with migration rate (Crow and Aoki 1984). However, it still may take tens or hundreds of generations to restore equilibrium after a perturbation.

Another difficulty is that rates of gene flow under equilibrium conditions—even when they can be estimated reliably—will not necessarily provide an accurate picture of the rate of recolonization following depletion of a local stock. With intraspecific competition relaxed or nonexistent, recolonization of an empty niche may occur much faster than would be predicted based on migration rates at equilibrium.

Because the question of interest is whether migration is sufficiently high to reestablish a harvestable surplus in a short period of time, it also seems prudent to evaluate the assumption of the island model that m is "small." Equation (1) is actually an approximation derived from the following exact equation by assuming that m is small enough that the m^2 terms can be ignored (Wright 1943):

$$F_{ST} = (1 - m)^2/[2N_e - (2N_e - 1) \cdot (1 - m)^2]. \quad (3)$$

Equation (3) is more complicated than equation (1) because it does not lead to a simple relationship between F_{ST} and the number of migrants per generation (mN_e). Instead, to estimate mN_e one must estimate two parameters: F_{ST} and either m or N_e. This inconvenient fact has no doubt contributed to the popularity of the simplified equation (1) and the relative obscurity of the exact expression given by equation (3).

Figure 3C shows a plot of equation (1) and three curves for equation (3) corresponding to N_e values of 50, 100, and 500. We can see that if F_{ST} is larger than about 0.03 and N_e is 50 or more, the approximation in equation (1) is very good. For smaller F_{ST} values, however (and especially for $F_{ST} < 0.01$) there can be significant bias in using equation (1) to estimate mN_e unless N_e is large. For example, with $\hat{F}_{ST} = 0.005$, equation (1) yields an estimate of mN_e of 50, while use of the exact equation with $N_e = 50$ gives an estimate of $\hat{m} = 0.42$ and $mN_e = 21$. F_{ST} values less than 0.01 may not be uncommon for marine species; recall that the median $\hat{F}_{ST}$ for marine fishes reported by Ward et al. (1994) was only 0.02.

Finally, even if we ignore the many other possible models of migration and drift and assume that equation (1) is applicable, there still is a significant problem with precision for high gene flow species. Because of the inverse relationship between F_{ST} and mN_e (Figure 3), the same magnitude of error in $\hat{F}_{ST}$ translates into a much larger error in estimating mN_e for high gene flow species than it does if gene flow is more restricted. Although some sources of error in estimating F_{ST} may be proportional to the magnitude of F_{ST} (and hence be smaller for low values of F_{ST}), other sources (e.g., sampling error in estimating allele frequencies) are fixed in magnitude and assume a relatively greater importance for high gene flow species.

The large uncertainties inherent in estimating mN_e for low $\hat{F}_{ST}$ values create a serious problem in addressing the key management question raised in this section. With adequate amounts of data, it may be possible to demonstrate that the lower limit of the confidence interval for $\hat{F}_{ST}$ leads to an estimate of mN_e that is too low to lead to rapid recolonization of a population that has been depleted by overharvest or other factors. For example, assume that the lower limit to a confidence interval for $\hat{F}_{ST}$ is calculated as 0.025, which leads to an upper bound for mN_e of 10 using equation (1). Although migration at this rate has a profound effect on population genetic structure and can affect population dynamics and persistence times over evolutionary time scales, it is far too low to be an appreciable factor in rapidly rebuilding a depleted population except under very unusual circumstances (e.g., recolonization by a few individuals with very high fecundity and a short generation length). Generally the number of migrants per generation must be in the hundreds or thousands to have any significant effect on the rate of stock rebuilding. Unfortunately, in most realistic situations it will be virtually impossible to demonstrate that migration rates are sufficiently large to allow rapid rebuilding. For example, to demonstrate that immigration is no lower than, say, 1,000 migrants per generation would require estimating F_{ST} with such precision that the upper bound on the confidence interval is 0.00025. This is not a realistic expectation. Asymmetry in the power to resolve this key management problem—genetic data may be able to demonstrate differences but generally cannot prove that migration is large enough to warrant management as a single stock—is inherently unsatisfying, both in a scientific sense and from a management perspective.

To summarize, the intrinsic disconnect between the genetic and demographic consequences of dispersal make the gene flow problem a difficult one for marine species. The nature of the relationship between F_{ST} and mN_e is not favorable to precise estimation of migration rates for high gene flow species. It is essentially impossible to demonstrate using genetic data alone that migration rates are high enough to quickly rebuild stocks depleted by ov-

erharvest. In some cases it will be possible to show that estimated rates of dispersal are too low to be consistent with management as a single stock. However, these estimates are prone to a number sources of bias. Furthermore, estimates obtained from equilibrium systems, even if they are accurate, may not provide a reliable indication of the rate of recolonization or dispersal into an empty niche.

Strategies for Maximizing the Value of Genetic Information

We have seen in the preceding sections that the signal:noise ratio problem is inherent to the study of the population genetics of high gene flow species, and this creates problems for both of the key management questions considered above. The genetic signal from stock differentiation is the sum of the genetic differences among the populations of interest. Little can be done to enhance the signal, since it is determined by factors beyond the control of the experiment (effective population size, migration rate, selection, mutation rate, elapsed time). However, a number of strategies can be used to enhance our ability to detect the signal, minimize the noise levels that tend to obscure the signal, or adjust for the noise that cannot be avoided. In this section I discuss these strategies as well as some other approaches that can help to increase the value of genetic information for marine species.

Random Sampling Error

The data at the core of most population genetic analyses (including all of the analyses discussed here) are genotypic or allelic frequencies. Random sampling error is a major source of the noise in estimating these frequencies. This sampling error has two components: intralocus error is a function of the number of individuals sampled, while interlocus error is a function of the number of loci examined.

Sampling individuals. The intralocus signal:noise ratio problem is illustrated in Figure 4. Two components that contribute to a raw $\hat{F}_{ST}$ value are plotted: a signal resulting from the parametric value for the populations involved, and noise due to the fact that the parameter is estimated from a sample of individuals rather than measured exactly. The graphs depict the expected (mean) values of both the signal and noise. The signal is based on the relationship between F_{ST} and mN_e shown in equation (1), and the expected contribution to the raw $\hat{F}_{ST}$ value from sampling S individuals is approximately $1/(2S)$ (Chakraborty and Leimar 1987; Workman and Niswander 1970; Wright 1978).

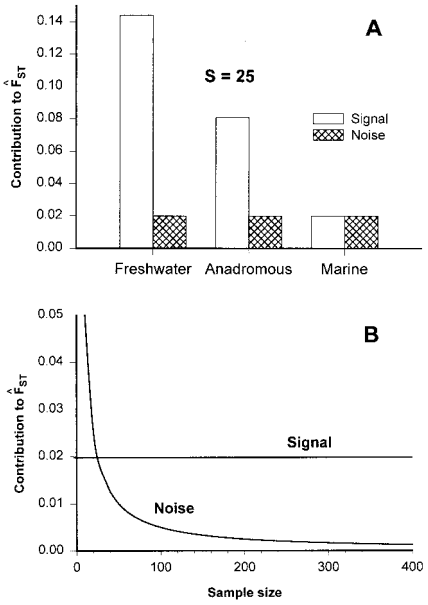

Figure 4. The importance of intralocus sampling error in estimating F_{ST}. **(A)** Relative magnitude of signal (parametric F_{ST}) and noise (intralocus sampling error) for freshwater, anadromous, and marine fishes. For each group, the signal is based on the median F_{ST} value from Table 1, and the noise is the expected magnitude of sampling error from samples of $S = 25$ individuals. **(B)** For marine species (median $F_{ST} = 0.02$), relative and absolute magnitude of sampling error decreases asymptotically as S increases.

Figure 4A shows the expected contributions of signal and noise to a typical raw $\hat{F}_{ST}$ value for freshwater, marine, and anadromous species. In this example, the signal is determined by the median $\hat{F}_{ST}$ value for each group of species (Table 1), and the noise is the sampling error expected from sampling $S = 25$ individuals from each subpopulation—a typical sample size for many DNA studies (and some allozyme studies) of marine organisms. This sampling error causes an upward bias in the raw $\hat{F}_{ST}$ value compared to its true (parametric) value. For a given sample size, this bias represents a larger fraction of the raw $\hat{F}_{ST}$ value for marine fishes than it does for freshwater or anadromous species. This indicates that lack of attention to sampling considerations will affect precision and bias of F_{ST} estimates more strongly for marine species.

This problem can be alleviated to some extent by taking larger samples. Figure 4B shows that with the signal held constant, the absolute (and relative) contribution of sampling error to the raw $\hat{F}_{ST}$ value declines asymptotically as sample size increases. For example, for a typical marine species (parametric $F_{ST} = 0.02$), intralocus sampling error is over twice as large as the signal with $S = 10$ and of equal magnitude to the signal with $S = 25$, but the relative magnitude of sampling error decreases to 50%, 25%, and 12.5% of the signal with successive doubling of S to 50, 100, and 200, respectively. Because the marginal benefits from subsequent increases in sample size continually diminish, there are limits to the benefits of increased sample size. At some point, further increases will become too costly (in terms of increasing logistical difficulties or higher demands on resources) to justify the relatively small additional reduction in bias. Where this critical point lies will vary with attributes of the organism involved and the tolerance for bias and lack of precision in the estimates of population genetic parameters.

Sampling loci. As defined by Wright (1943), F_{ST} is a parameter that has no variance, since it applies to a global population with an infinite number of subpopulations. In any real application, of course, the number of subpopulations (and the global population size) will be finite. In a finite population, each gene locus can be considered an independent realization of an evolutionary process that involves a balance between migration and genetic drift. As a result of this stochastic process, parametric F_{ST} will vary considerably among loci (Nei et al. 1977) even when m and N_e are fixed and other assumptions of the island model hold. This means that estimates of F_{ST} calculated using data for a single gene locus will have a wide confidence interval for the associated estimate of mN_e.

Because the signal (parametric F_{ST}) is not fixed but instead varies across loci, our ability to reliably detect the signal is enhanced by considering data for multiple, independent gene loci. The importance of sampling multiple gene loci cannot be overstated. In fact, Chakraborty and Leimar (1987) argued that there is little reason to make a bias adjustment for sampling individuals unless a large number of loci are used to reduce the standard error of the overall estimate $\hat{F}_{ST}$. Unfortunately, because the interlocus variance of $\hat{F}_{ST}$ depends on a number of factors (number of subpopulations, number of subpopulations sampled, population size, time of divergence, and sample size of individuals), there is not a simple expression for the variance (Long 1986). The most that can be said categorically is that the variance declines asymptotically as the number of loci increases. For an empirical

study, the best way to evaluate the variance (or a confidence interval) for a multilocus $\hat{F}_{ST}$ value appears to be by resampling techniques (Weir and Cockerham 1984; Raymond and Rousset 1995a).

Adjusting for bias. We have seen that taking larger samples reduces the bias caused by intralocus sampling error but does not eliminate it entirely. Unavoidable bias associated with sampling individuals is generally dealt with in one of two ways. The first (and surprisingly common) approach is to assume that the bias can be safely ignored because it will be small compared to the signal. In general, this approach will be reasonable for comparisons of well-differentiated entities (e.g., different species or subspecies), for which the signal can be orders of magnitude larger than intralocus sampling error (e.g., Gorman and Renzi 1979). Nei (1978) stated that there is little difference between standard and unbiased estimates of genetic distance if sample sizes are 50 or more, and Chakraborty and Leimar (1987) argued that because intralocus sampling error from 50 individuals per subpopulation upwardly biases $\hat{F}_{ST}$ estimates by only the magnitude 0.01, this effect can usually be safely ignored.

For many marine species, however, ignoring this bias is risky. The relative importance of the bias depends not only on its magnitude but also on the strength of the signal (the parametric F_{ST}), which typically will be weak for marine species. Furthermore, because of the hyperbolic relationship between F_{ST} and mN_e (Figure 3), a small amount of bias in a low F_{ST} value can have a large effect on the estimate of mN_e. For example, an upward bias in $\hat{F}_{ST}$ of magnitude 0.01 (as would occur from samples of $S = 50$) may have little effect on estimates of mN_e for most anadromous or freshwater species (median $\hat{F}_{ST} = 0.08$–0.14), but for a species with parametric $F_{ST} = 0.005$, ignoring intralocus sampling error will on average downwardly bias the estimate of mN_e from 50 to 16. This indicates that for marine species, it is not only important to reduce bias as much as possible (by taking large samples of individuals), but also to correct for the bias.

A bias correction can be applied by computing the average, or expected, contribution to $\hat{F}_{ST}$ from intralocus sampling error [approximately $1/(2S)$] and subtracting that amount from the raw value of the statistic [see, for example, Weir and Cockerham (1984), Workman and Niswander (1970), and Wright (1978) for bias corrections for $\hat{F}_{ST}$ and Nei (1978) for a bias correction for genetic distance]. The goal is to filter out the noise in the raw statistic to yield a more accurate estimate of the signal.

The major difficulty with this approach is that intralocus sampling error arises from a stochastic process, and its magnitude in any particular application can be described only in a statistical sense. In an empirical study, the actual error in estimating allele frequencies at a gene locus (and hence the error in estimating the parametric F_{ST} value) could be much greater or much less than the theoretical expectation. Thus, although explicitly adjusting for sampling error can lead to an estimate that is unbiased, it will not necessarily yield an estimate that is sufficiently precise. Another difficulty is that in some cases, applying the bias correction for sampling error can lead to a negative $\hat{F}_{ST}$ value. This can occur if genetic differences between populations are small and the actual magnitude of sampling error is smaller than the expected value. In this case, the genetic data provide no evidence that gene flow among populations is restricted.

This lack of precision can be overcome to some extent by sampling multiple gene loci. Averaging across multiple gene loci ensures that the mean value of sampling error in the experiment will be closer to the theoretical expectation than would be the case for a single locus, which in turn means that the bias correction will be more reliable. Use of multiple loci in computation of $\hat{F}_{ST}$ can thus be a powerful way to increase precision of the unbiased estimate. As is the case with increasing sample size, the benefits of adding additional loci are nonlinear, with the greatest proportional improvements in precision associated with increases from one or just a few loci.

Another source of bias arises from the inevitable violation of the assumption in the island model of an infinite number of subpopulations. Even in the unlikely situation in which all other assumptions of the island model are met, the actual number of subpopulations will be finite, and this will downwardly bias the parametric F_{ST} compared to the relationship shown in equation (1). If only a subset of the existing subpopulations are sampled, there will be an additional source of downward bias in $\hat{F}_{ST}$, as well as an additional source of error in estimating the parametric F_{ST} for the population as a whole. Some methods for estimating F statistics and related quantities (e.g., θ in Weir and Cockerham 1984) include corrections for this effect, while others (e.g., G_{ST} in Nei 1973) do not. When estimating mN_e using a method that does not include an adjustment for number of subpopulations, equation (1) can be modified and rearranged to yield the following:

$$\hat{F}_{ST} \approx 1/\{1 + 4mN_e dd_s/[(d-1)(d_s-1)]\};$$

$$m\hat{N}_e = \frac{d-1}{d}\frac{d_s-1}{d_s}\left[\frac{1}{4\hat{F}_{ST}} - \frac{1}{4}\right], \qquad (4)$$

where d and d_s are the total number of subpopulations and the number that are sampled, respectively (Slatkin 1993). The bias in $m\hat{N}_e$ from ignoring this effect will be relatively small unless d and/or d_s are small numbers.

DNA data. The increasing accessibility of a large amount and variety of DNA data in recent years has considerably enhanced our ability to detect population genetic structure. Direct benefits of DNA data for the key conservation issues considered here include the following: (1) In some fish species of economic and conservation interest [e.g., striped bass (*Morone saxatilis*), Atlantic salmon (*Salmo salar*), American shad (*Alosa sapidissima*)], allozymes provide little basis for informed fishery management because they have very low levels of variation. In contrast, highly variable DNA markers have been identified for each of these species (Epifanio et al. 1995; McConnell et al. 1995; Wirgin et al. 1991). (2) DNA methods provide an opportunity to survey genetic variation at more gene loci than are available using allozymes alone. Combining DNA and allozyme data can therefore be important in reducing the interlocus sampling variance of $\hat{F}_{ST}$ values. Many DNA markers (especially microsatellites) are also highly variable, and this can increase overall power of resolution, provided that appropriate methods can be developed to address statistical challenges posed by large numbers of rare alleles (Chakraborty 1992; O'Connell and Wright 1997). (3) Because mitochondrial DNA (mtDNA) is haploid and clonally inherited, the signal from genetic drift is stronger than for nuclear gene loci. However, the same sample size provides only half as many mtDNA genes for analysis as it does nuclear genes, and there is no opportunity to sample multiple independent gene loci with mtDNA. As a result of these various factors, the net effect of using mtDNA versus nuclear DNA on the signal:noise ratio problems discussed here is complex. (4) DNA methods can provide access to samples that would

be difficult or impossible to obtain for allozyme studies—for example, because of requirements of allozyme analyses for lethal sampling and rapid freezing of samples in the field at ultracold temperatures. (5) DNA data are amenable to genealogical analysis that may provide insight into the importance of the historical biogeographical process to current population genetic structure [see Grant and Bowen (1998) for an example involving marine fishes]. (6) DNA analyses based on the polymerase chain reaction (PCR) can make use of archived material (e.g, scale collections) to provide a historical dimension to population genetic analyses (e.g., Miller and Kapuscinski 1997; Purcell et al. 1996).

There has also been a general expectation on the part of many that DNA methods will provide dramatically higher resolution for situations in which sufficient polymorphic allozyme markers are available but they fail to clearly resolve stock structure. Unfortunately, this expectation does not appear to be well founded. The underlying biological processes (migration and genetic drift) should affect all neutral, nuclear gene loci in a similar fashion, so the intrinsic genetic "signal" from all nuclear DNA markers should be the same magnitude. Although microsatellites have much higher mutation rates than allozymes or most other DNA markers, in theory this should not lead to higher $\hat{F}_{ST}$ values because the relationship shown in equation (1) is independent of mutation rate and the number of alleles (Crow and Aoki 1984). In fact, Crow and Aoki (1984) found through simulations that the equilibrium value of G_{ST} in the finite island model is slightly lower with a mutation rate of 10^{-3} (a typical rate for microsatellites) than with a mutation rate more typical of allozyme loci (10^{-7}).

Empirical studies that have compared levels of differentiation detected by allozyme and different types of DNA markers have found mixed results (Bentzen et al. 1996; Pogson et al. 1995; Scribner et al., in press; Zhang et al. 1993). Explanations that have been proposed to account for the lack of concordance in some studies include natural selection acting on some classes of markers, different rates of approach to equilibrium determined by different mutation rates, and uncertainty about which mutation model is best for use with microsatellite data.

Summary. In summary, bias caused by intralocus sampling error can be reduced by taking larger samples, but it still may remain a significant fraction of the raw F_{ST} value for high gene flow species. Interlocus sampling error does not lead to bias, but it can severely limit precision (and therefore the usefulness) of estimates based on just a few gene loci. Assaying a large number of independent gene loci is important in obtaining precise estimates of F_{ST} and mN_e for high gene flow species. Use of a large number of loci also allows the magnitude of intralocus sampling error to be estimated more precisely, which means that the bias adjustment can be more effective.

Nonrandom Sampling

In addition to collecting large samples of individuals and loci, it is important to obtain samples that satisfy as nearly as possible the assumptions of the model used—typically, that sampling is random with respect to the entire population. Our ability to randomly sample a population is only as good as our understanding of the species' behavior, ecology, and life history. Sex-biased dispersal patterns, ontogenetic movements of individuals, susceptibility to capture, and demographic parameters such as age structure and sex ratio can all affect the ability to obtain a random sample.

Nonrandom sampling can lead to two problems in interpreting genetic data. First, nonrandom samples can be biased if certain types of individuals appear preferentially in the sample. For example, many fishery sampling methods are size selective, meaning that large or small individuals are more likely to be taken. This can lead to bias if the attributes being monitored are correlated with size. Any number of behavioral attributes of individuals might also affect their likelihood of capture, and this could also lead to bias if these behaviors are associated with the genetic characteristics of interest.

Second, even if factors such as size or behavior are not associated in any consistent way with the genetic traits being measured, this type of nonrandom sampling can increase the noise level because only part of the population has been sampled. For example, if size and age are strongly correlated, as is typically the case in marine species, size-selective sampling methods will tend to overrepresent some cohorts at the expense of others. In this case, the genetic data will depend heavily on only a subset of the parents in a generation as a whole, and genetic drift will be a larger factor in determining the results than it would if the generation as a whole were sampled randomly. This will upwardly bias the estimate of F_{ST} unless an adjustment is made.

The first step in addressing this problem is to understand the biology of the species involved. If the generation as a whole cannot easily be sampled randomly, but cohorts can be sampled individually, it may be possible to adjust for bias due to the sampling methods. For example, Waples (1990a,b) developed an explicit correction factor that makes it possible to use standard population genetic models (based on discrete generations) with data for individual cohorts of Pacific salmon. Even if samples contain multiple cohorts, adjustments may be possible if the individuals can be aged so that cohorts can be reconstructed from one or more samples (Jorde and Ryman 1995; Waples 1991a). In special cases, other explicit adjustments may be possible (see example below).

If there is reason to suspect that sampling was not random, but explicit adjustments are not possible, the most effective strategy is to replicate sampling in space and/or time. Adding additional spatial samples can make the results more robust to violations of the assumptions of the migration/drift model used. For example, it will rarely be the case that the island model, which assumes completely symmetrical migration, accurately describes the population structure of any marine species. More typically, nearby populations will exchange migrants more frequently than distant ones, and if this is the case additional spatial samples will increase the ability of the study to detect finer scale population structure. Also, stratified sampling may provide an adequate approximation to random sampling in situations where it is impossible to satisfy the assumption exactly.

Temporally replicated samples provide an extra dimension to the analyses that can be extremely important in evaluating possible sources of bias and sampling artifacts in the data. For example, if differences among geographic samples are relatively modest but consistent over time, then one can be much more confident that they represent a bona fide genetic signal rather than some artifact. On the other hand, if relatively large differences are regularly found among geographic samples, but the pattern of relationships is not consistent over time, this suggests that the biology of the organism is imperfectly understood, that there are unrecognized complexities in the processes involved, or that the data are flawed.

Allendorf–Phelps effect. The example of the Allendorf–Phelps effect described earlier involved a panmictic population of fish in a lake that assorted randomly to spawn in different tributaries. This model is easily extended to apply to anadromous Pacific salmonids. For example, a reasonable null hypothesis regarding salmon populations within a river basin might be that adults return at random to a stream to spawn, with no preference for their natal stream. If spawning adults are sampled from various streams, then a standard chi-square (or similar) test comparing their frequencies is an adequate test of the null hypothesis. If, however, their juvenile progeny are sampled rather than the adults themselves, the chances of a statistically significant test are inflated (the Allendorf–Phelps effect) because the sampling does not conform to assumptions implicit in the null hypothesis (specifically, that the individuals sampled have been drawn randomly from the global population). This same phenomenon may be more generally applicable to genetic studies of marine species. For example, many marine fishes (e.g., cod, herring, pollock) are characterized by spawning aggregations that regularly occur in the same geographic areas every year. The conservation question of interest is whether these spawning aggregations represent discrete stocks. In this case, as in the salmon example, the null hypothesis of panmixia cannot be directly addressed by sampling juveniles derived from different spawning aggregations.

Fortunately, it is possible to adjust the null hypothesis to account for the additional complexities of sampling associated with the Allendorf–Phelps effect. Figure 5 depicts the sampling scheme for the adjusted null hypothesis, which accounts for an episode of founder effect/genetic drift as well as intralocus sampling error in estimating allele frequencies. The chi-square value obtained from a standard contingency test of allele frequencies in two samples is mathematically equivalent to the following quantity (Waples 1989):

$$\text{standard } \chi^2 = \frac{(X_1 - X_2)^2}{\bar{X}(1 - \bar{X})/S}. \qquad (5)$$

In equation (5), X_1 and X_2 are allele frequencies in the two samples and $\bar{X}$ is the weighted mean allele frequency across both samples. To account for the departures from the standard null hypothesis created by the Allendorf–Phelps effect, we can adjust the denominator of equation (5) to include a term that accounts for the episode of founder effect/genetic drift (Waples RS, unpublished data):

$$\text{adjusted } \chi^2 \approx \frac{(X_1 - X_2)^2}{\bar{X}(1 - \bar{X})(1/S + 1/N_b)}. \qquad (6)$$

In equation (6), the $1/N_b$ term represents the effective number of breeders responsible for the juveniles that were sampled. If this number varies across spawning aggregations, then the harmonic mean of the individual N_b values should be used. Note that N_b will in general not be the same as N_e, which is the effective population size per generation. The Allendorf–Phelps effect will generally be important only when the parents involved represent just part of a population (or part of a generation) for the organism of interest.

Under the adjusted null hypothesis (random distribution of breeders followed by an episode of reproduction and genetic drift before sampling), the quantity in equation (6) should be distributed approximately as chi square and can form the basis of a test to determine whether factors other than drift and sampling error must be invoked to explain the results. This result can be generalized to multiple loci and multiple alleles.

A similar adjustment can be made for $\hat{F}_{ST}$ values that are upwardly biased by the Allendorf–Phelps effect. We have seen that, under the standard null hypothesis (parametric $F_{ST} = 0$), the raw $\hat{F}_{ST}$ value will be inflated by intralocus sampling error, which has an average magnitude of $1/(2S)$. Accounting for the episode of founder effect/drift associated with the Allendorf–Phelps effect leads to the following (Waples RS, unpublished data):

$$E(\hat{F}_{ST})(\text{adjusted null hypothesis}) \approx 1/(2S) + 1/(2N_b). \qquad (7)$$

As noted for equation (6), the N_b of interest may be much less than the effective number of breeders in the entire spatial/temporal spawning aggregation. Hedgecock (1994) argued that various factors in the marine environment may greatly increase the variance in reproductive success among individuals of highly fecund species, resulting in effective sizes that are orders of magnitude lower than the number of spawning adults. An important component of Hedgecock's argument is that this variance remains large throughout the whole life cycle, because the effect of family size variance on N_e is measured at the reproductive stage of the offspring from each family. In contrast, all that is required for the Allendorf–Phelps effect to be important is that the juveniles sampled be derived from a relatively few adults. This might occur in any of a number of ways in sampling eggs, larvae, or juveniles of a marine species with large spawning aggrega-

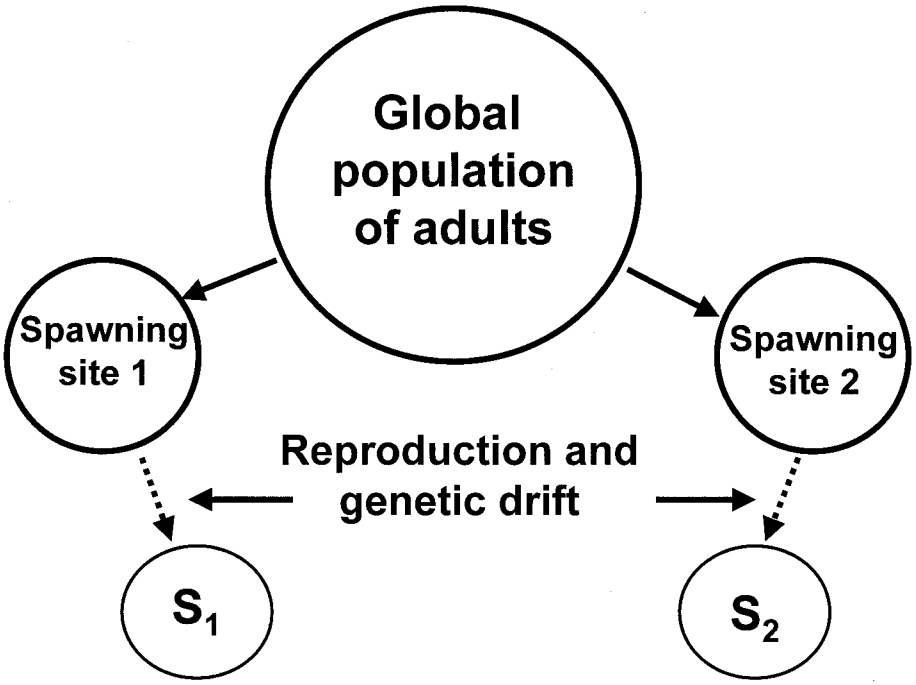

Figure 5. Departures from assumptions of the standard null hypothesis that lead to the Allendorf–Phelps effect. A panmictic global population assorts randomly into different spawning areas. However, samples for genetic analysis (S_1 and S_2) are taken not of reproducing adults but rather their juvenile progeny. The episode of founder effect/genetic drift upwardly biases F_{ST} values and may lead to the (wrong) conclusion that the populations are reproductively isolated. This effect can be adjusted for quantitatively; see text for discussion.

tions, even if the "Hedgecock effect" is not maintained over the full life cycle [see Ruzzante et al. (1996) for an example and some discussion].

As an example, consider a high gene flow species with parametric $F_{ST} = 0.01$, and assume that samples of $S = 50$ juveniles are taken from each of three large spawning aggregations (N and N_b both $> 10^6$). Assume further that at two localities, the samples were essentially random draws from the (very) large total number of offspring produced by the spawners, while in the third locality the researchers by chance obtained their sample from a larval cloud produced by only $N_b = 10$ spawners. The harmonic mean of $N_b = 10^6$, 10^6, and 10 is only 30, so based on equation (7) we calculate the expected value of $\hat{F}_{ST}$ after accounting for sampling and the Allendorf–Phelps effect as

$$E(\hat{F}_{ST}) \approx F_{ST} + 1/(2S) + 1/(2N_b)$$
$$= 0.01 + 0.01 + 0.0167$$
$$= 0.0367.$$

In this case, the raw $\hat{F}_{ST}$ value is over three times as large as the parametric (true) value, and the bias created by the Allendorf–Phelps effect is significantly larger than either the signal or the intralocus sampling error.

A difficulty in applying the correction suggested here for the Allendorf–Phelps effect is obtaining a reliable estimate of N_b. The genetic data gathered to evaluate population genetic structure may provide some insight in this regard. Analysis of gametic disequilibrium (Hill 1981) provides a way of estimating N_b from a single sample. Although a number of issues related to bias and precision must be considered in using gametic disequilibrium to estimate N_b, it can be a powerful tool for identifying cases in which the sample has been produced by a small number of breeders (Waples 1991a). If multiple samples are available, a combination of the temporal and disequilibrium methods can be used to increase the precision of estimates of N_b (Waples 1991a). Herbinger et al. (1997) used a slightly different approach—estimating the incidence of full- and half-sib relationships using microsatellite DNA data— to evaluate N_b in samples of Atlantic cod (*Gadhus morhua*) larvae.

Even if N_b cannot be estimated precisely, equation (7) provides a way of evaluating the likelihood that the Allendorf–Phelps effect has been important. For example, in the above example, after adjusting the raw $\hat{F}_{ST}$ value for intralocus sampling bias, the researcher could consider whether the resulting estimate ($0.0367 - 0.01 = 0.0267$) might be inflated by the Allendorf–Phelps effect. After carefully considering the sampling protocols, the biology of the species involved, and the local oceanographic processes, the researcher might be able to conclude with some confidence that the juvenile samples were not produced by fewer than (say) $N_b = 100$ effective breeders at any locality. This would put an upper limit to the bias remaining in the F_{ST} value after adjusting for sampling error. Alternatively, the researcher could take the following approach: (1) begin by calculating the value of $m\hat{N}_e$ associated with the point estimate of F_{ST} (after adjusting for intralocus sampling error, $\hat{F}_{ST} = 0.0267$, which leads to $mN_e = 9$); (2) identify a migration rate (say $mN_e = 100$) that would lead to a different conservation or management action than would occur with $mN_e = 9$; (3) determine how large the Allendorf–Phelps effect would have to be to reduce the estimate of mN_e from 100 to 9; and (4) evaluate the likelihood that N_b could have been small enough to produce an effect of this magnitude. For example, migration at a rate of 100 individuals per generation implies a parametric $F_{ST} = 0.0025$. To have upwardly biased the $\hat{F}_{ST}$ value from 0.0025 to 0.0267 (and downwardly biased mN_e from 100 to 9), the Allendorf–Phelps effect would require $1/(2N_b) = 0.0267 - 0.0025 = 0.0242$, or a harmonic mean N_b of about 21. If the researcher can conclude with confidence that the samples were not drawn from this small a number of breeders, she will have evidence that the Allendorf–Phelps effect has not biased the results so much that it will lead to a faulty management decision.

Methods for Estimating Gene Flow

Several analytical methods can be used with genetic data to provide an indirect estimate of the migration parameter mN_e. It is beyond the scope of this article to consider this subject in detail, but I will briefly summarize some results of other studies. Slatkin and Barton (1989) compared F statistics and the rare allele method (Slatkin 1985b) as estimators of mN_e. They found that, in theory, the two methods are similar in their sensitivity to selection and variation in population structure and also have similar rates of approach to equilibrium. However, they concluded that practical difficulties involved in collecting data on rare alleles make F statistics a better general method for estimating gene flow in natural populations.

Slatkin and Barton (1989) also used simulations to evaluate the relative performance of two methods for estimating F_{ST}: Nei's (1973) G_{ST} and Weir and Cockerham's (1984) θ. One of their findings was increased bias and decreased precision of θ for high levels of gene flow. However, Cockerham and Weir (1993) questioned the theoretical basis for Slatkin and Barton's simulation results and concluded they were in error; their own simulations showed little bias in θ even for migration rates as high as $m = 0.1$. Simulations performed by Chakraborty and Leimar (1987) are consistent with Cockerham and Weir's results, and Chakraborty and Leimar concluded that θ was the method of choice for estimating F_{ST} in high gene flow species.

The mutation rate for electrophoretically detectable alleles is low enough that it does not have a major effect on $\hat{F}_{ST}$ estimates based on allozyme data. This is not true of all DNA markers, however. In particular, the mutation rate for microsatellites is high enough and back mutations are common enough that ignoring mutations can seriously bias estimates of gene flow. Slatkin (1995) and Goldstein and Pollock (1997) discussed this issue and suggested methods to account for the bias, and Bentzen et al. (1996) gave an example in which this bias may be a factor in the genetic analysis of a marine species.

Statistical Testing

Numerous statistical tests are available for the analysis of population structure. Researchers studying marine organisms should continue to use these resources to test hypotheses that will further our understanding of the biology of marine species. However, exclusive focus on the results of statistical tests can be misleading. The significance level of a statistical test provides little information by itself; it is also necessary to consider the data richness and the power of the test. Statistical tests will be particularly sensitive to these factors if researchers follow the suggestions in this section for maximizing the signal:noise ratio by intensive sampling efforts. In addition, even relatively minor departures from random sampling assumptions can cause a misleading rejection of the null hypothesis if large amounts of data are collected.

To minimize these potential difficulties, researchers should first evaluate the power of the tests they use. If an evaluation indicates that, because large amounts of data are available, the test under consid-

eration has a high probability of yielding a statistically significant result even if the biological differences among populations are trivially small, then performing the test will be of limited use by itself. A better approach is to focus on estimating the magnitude of the differences between populations and evaluating the biological significance of these differences. This would place more emphasis on the second key management question considered in this article (estimating levels of gene flow among populations or stocks) rather than the first (testing for stock differences). However, the estimation procedure should avoid undue emphasis on the point estimates of F_{ST} and mN_e. Rather the approach should be to describe various hypotheses about migration among populations and, using the empirical data, evaluate their relative probabilities of being correct. Having done this, the researcher or manager will be in a position to evaluate the likely consequences of different conservation actions (e.g., management as one or multiple stocks). The field of statistical decision analysis (Hilborn 1997; Raiffa 1968) provides guidelines for use of this general approach.

Recently there has been a good deal of interest in improving methods for statistical testing of hypotheses about population differentiation (e.g., Hudson et al. 1992; Raymond and Rousset 1995a; Roff and Bentzen 1989). This issue is increasingly important for microsatellite and mtDNA data, which often involve large numbers of alleles (or haplotypes) per locus. Asymptotic tests such as the chi-square test are not well suited to dealing with large numbers of rare alleles, which result in many cells that are empty or at low frequency in the test. It is important that these alternative tests continue to be used and refined, because they allow statistical tests to be extended to a greater range of applications than would otherwise be the case. However, it should be recognized that, although sometimes referred to as "exact" tests, these methods do not deal any more effectively than the chi- square test with factors such as statistical power and nonrandom sampling that complicate the interpretation of test statistics.

Other Approaches

In addition to statistical testing for population differences and estimating levels of gene flow, several other approaches can be useful for studying the population genetic structure of marine species. A gene diversity analysis that partitions genetic variation into various spatial and temporal components can provide insight into population subdivision and temporal stability. Furthermore, collection of temporal and/or spatial replicates allows a quantitative evaluation of the importance of sampling error and the signal:noise ratio. Chakraborty and Leimar (1987) discussed gene diversity analysis in a fishery context.

As noted above, one key to using genetic data to estimate gene flow is to determine whether an isolation model or a migration model is more appropriate and, if the latter, what type of migration model to use. Although an estimate of F_{ST} does not by itself resolve this question, some insight into this issue can be gained by examining the other F statistics. Whereas F_{ST} represents the level of inbreeding in subpopulations compared to the total (global) population (hence the subscript ST), F_{IS} and F_{IT} represent the level of inbreeding in individuals with respect to the subpopulations and the total population, respectively. Long (1986) described an example in which these latter F statistics were used to evaluate different hypotheses about population structure in tribal groups in Papua New Guinea. Estimates of F_{IS} and F_{IT} can also help in evaluating the effects on $\hat{F}_{ST}$ of failing to recognize population structure within sampled populations (Long 1986).

Slatkin (1993) examined several isolation-by-distance models (in which subpopulations receive migrants more frequently from nearby subpopulations than from more distant ones) and showed that using a matrix of pairwise $\hat{F}_{ST}$ values and geographic distances between pairs of populations can reveal information about the importance of historical processes as well as current levels of gene flow. He argued, for example, that failure to find evidence of isolation by distance for species with relatively low $\hat{F}_{ST}$ values may indicate that the species has recently colonized the study area, whereas a lack of isolation by distance in a species with high $\hat{F}_{ST}$ may indicate that there is no ongoing gene flow and an equilibrium model is not appropriate. Slatkin's approach also provides a way of estimating the genetically effective neighborhood size. Shaklee and Bentzen (in press) give some examples of the use of the isolation-by- distance approach in examining the population structure of marine species.

If data for multiple species are available, it may be possible to evaluate from the genetic data alone whether it is reasonable to use an equilibrium model to estimate gene flow. Waples (1987) sampled 10 shorefish species from several mainland and island sites in southern California and Mexico and found that estimates of mN_e based on $\hat{F}_{ST}$ values were strongly correlated (Spearman's rank correlation coefficient = 0.88) with estimated dispersal capabilities based on life-history information and records of larval captures at sea. This is the result that would be expected for neutral genes at migration/drift equilibrium. In contrast, there is no a priori reason to expect this result in an isolation model or a nonequilibrium model involving migration.

Conclusions

Important points that emerge from this article include the following:

We should not rely solely on statistical tests to guide decisions about identification and management of stocks of marine species. It is more important to identify and implement experimental designs to produce genetic information that is useful in choosing among various conservation strategies.

If statistical tests are used, possible departures from the null hypothesis (and associated assumptions) should be carefully evaluated and, if possible, adjusted for quantitatively.

Various strategies (including use of DNA markers) can increase precision and reduce bias in estimates of key genetic parameters, but these methods cannot eliminate all sources of noise. Furthermore, because the amount of migration necessary to obscure most genetic evidence of stock structure (only a handful of individuals per generation) is generally inconsequential as a force for rebuilding depleted populations on a time scale of interest to humans, there is no guarantee that genetic methods alone will provide sufficient precision for key management decisions involving marine species.

The single most effective strategy for dealing with the signal:noise ratio problem is to replicate samples over time. Patterns of genetic relatedness or differentiation that are consistent across a temporal dimension are unlikely to be caused by sampling artifacts.

Given the points above, it is important to make every effort to understand the ecology and life history of the target species. In the absence of such information, not only will a researcher be unable to

make quantitative adjustments for possible sources of bias, she will not even be in a position to know whether bias has occurred, nor will she be able to develop effective strategies to minimize sampling error.

Although they also have limitations, studies designed to measure demographic parameters directly (e.g., through tagging studies, monitoring movement of cohorts in space and time, measuring current patterns, etc.) can be an important complement to indirect genetic studies of population structure.

References

Allendorf F, Ryman N, and Utter F, 1987. Genetics and fishery management: Past, present, and future. In: Population genetics and fishery management (Ryman N and Utter F, ed). Seattle: University of Washington Press; 1–20.

Allendorf FW and Phelps SR, 1981. Use of allelic frequencies to describe population structure. Can J Fish Aquat Sci 38:1507–1514.

Bailey KM, Powers D, Villa G, Nishimura A, Traynor J, and Walters G, in press. Population ecology and structural dynamics of walleye pollock, *Theragra chalcogramma.* In: Bering sea dynamics (Louglin T and Ohtani K, eds). Alaska Sea Grant.

Bentzen P, Taggart CT, Ruzzante DE, and Cook D, 1996. Microsatellite polymorphism and the population structure of Atlantic cod (*Gadus morhua*) in the northwest Atlantic. Can J Fish Aquat Sci 53:2706–2721.

Berger JO and Selke T, 1987. Testing a point null hypothesis: the irreconcilability of P values and evidence. J Am Stat Assoc 82:112–122.

Bowen B, Avise JC, Richardson JI, Meylan AB, Margaritoulis D, and Hopkins-Murphy SR, 1992. Population structures of loggerhead turtles (*Caretta caretta*) in the northwestern Atlantic Ocean and Mediterranean Sea. Conserv Biol 7:834–844.

Carvalho GR and Hauser L, 1994. Molecular genetics and the stock concept in fisheries. Rev Fish Biol Fish 4:326–350.

Cavalli-Sforza LL and Bodmer WW, 1971. The genetics of human populations. San Francisco: W.H. Freeman.

Chakraborty R, 1992. Sample size requirements for addressing the population genetic issues of forensic use of DNA typing. Hum Biol 64:141–159.

Chakraborty R and Leimar O, 1987. Genetic variation within a subdivided population. In: Population genetics and fishery management (Ryman N and Utter F, ed). Seattle: University of Washington Press; 89–120.

Cockerham CC and Weir BS, 1993. Estimation of gene flow from F-statistics. Evolution 47:855– 863.

Crow JF and Aoki K, 1984. Group selection for a polygenic trait: estimating the degree of population subdivision. Proc Natl Acad Sci USA 81:6073–6077.

Dizon AE, Taylor BL, and O'Corry-Crowe GM, 1995. Why statistical power is necessary to link analyses of molecular variation to decisions about population structure. Am Fish Soc Symp 17:288– 294.

Epifanio JM, Smouse PE, Kobak CJ, and Brown BL, 1995. Mitochondrial DNA divergence among populations of American shad (*Alosa sapidissima*): How much variation is enough for mixed-stock analysis? Can J Fish Aquat Sci 52:1688–1702.

Felsenstein J, 1976. The theoretical population genetics of variable selection and migration. Annu Rev Genet 10: 253–280.

Gold JR and Richardson LR, 1998. Mitochondrial DNA diversification and population structure in fishes from the Gulf of Mexico and western Atlantic. J Hered 89: 404–414.

Goldstein DB and Pollock DD, 1997. Launching microsatellites: a review of mutation processes and methods of phyologenetic inference. J Hered 88:335–342.

Gorman GC and Renzi J Jr, 1979. Genetic distance and heterozygosity estimates in electrophoretic studies: effects of sample size. Copeia 1979:242–249.

Grant WS and Bowen BW, 1998. Shallow population histories in deep evolutionary lineages of marine fishes: insights from sardines and anchovies and lessons for conservation. J Hered 89:415–426.

Graves JE, 1998. Molecular insights into the population structures of fishes with circumglobal distributions. J Hered 89:427–437.

Harden Jones F, 1968. Fish migration. London: Edward Arnold.

Hedgecock D, 1994. Does variance in reproductive success limit effective population sizes of marine organisms? In: Genetics and evolution of aquatic organisms (Beaumont AR, ed). London: Chapman & Hall; 122–134.

Herbinger CM, Doyle RM, Taggart CT, Lochmann SE, Brooker AL, Wright JM, and Cook D, 1997. Family relationships and effective population size in a natural cohort of Atlantic cod (*Gadus morhua*) larvae. Can J Fish Aquat Sci 54(suppl 1):11–18.

Hilborn R, 1997. Statistical hypothesis testing and decision theory in fisheries science. Fisheries 22(10):19–20.

Hill WG, 1981. Estimation of effective population size from data on linkage disequilibrium. Genet Res Camb 38:209–216.

Hoelzel AR (ed), 1991. Genetic ecology of whales and dolphins. Cambridge: International Whaling Commission.

Hoelzel AR, 1998. Genetic structure of Cetacean populations in sympatry, parapatry and mixed assemblages: Implications for conservation policy. J Hered 89:451–458.

Hudson RR, Boos DD, and Kaplan NL, 1992. A statistical test for detecting geographic subdivision. Mol Biol Evol 9:138–151.

Jorde PE and Ryman N, 1995. Temporal allele frequency change and estimation of effective size in populations with overlapping generations. Genetics 139:1077–1090.

Lewontin RC, 1995. Human diversity. New York: Scientific American Library.

Long JC, 1986. The allelic correlation structure of Gainj-and Kalam-speaking people. I. The estimation and interpretation of Wright's F-statistics. Genetics 112:629–647.

McConnell SK, O'Reilly P, Hamilton L, Wright JM, and Bentzen P, 1995. Polymorphic microsatellite loci from Atlantic salmon (*Salmo salar*): genetic differentiation of North American and European populations. Can J Fish Aquat Sci 52: 1863–1872.

Miller LM and Kapuscinski AR, 1997. Historical analysis of genetic variation reveals low effective population size in a northern pike (*Esox lucius*) population. Genetics 147:1249–1258.

NRC (National Research Council), 1994. An assessment of Atlantic bluefin tuna. Washington, DC: National Academy Press.

Nei M, 1973. Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA 70:3321–3323.

Nei M, 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583–590.

Nei M and Chakravarti A, 1977. Drift variances of F_{ST} and G_{ST} statistics obtained from a finite number of isolated populations. Theor Popul Biol 11:307–325.

Nei M, Chakravarti A, and Tateno Y, 1977. Mean and variance of F_{ST} in a finite number of incompletely isolated populations. Theor Popul Biol 11:291–306.

O'Connell M and Wright JM, 1997. Microsatellite DNA in fishes. Rev Fish Biol Fish 7:331–363.

Pawson MG and Jennings S, 1996. A critique of methods for stock identification in marine capture fisheries. Fish Res 25:203–217.

Pogson GH, Mesa K, and Boutilier RG, 1995. Genetic population structure and gene flow in the Atlantic cod *Gadus morhua*: a comparison of allozyme and nuclear RFLP loci. Genetics 139:375–385.

Primack RB, 1993. Essentials of conservation biology. Sunderland, Massachusetts: Sinauer.

Purcell MK, Kornfield I, Fogarty M, and Parker A, 1996. Interdecadal heterogeneity in mitochondrial DNA of Atlantic haddock (*Melanogrammus aeglefinus*) from Georges Bank. Mol Mar Biol Biotech 5:185–192.

Raiffa H, 1968. Decision analysis: introductory lectures on choice under uncertainty. Reading, Massachusetts: Addison-Wesley.

Raymond M and Rousset F, 1995a. An exact test for population differentiation. Evolution 49:1280–1283.

Raymond M and Rousset F, 1995b. GENEPOP (version 1.2): a population genetics software for exact test and ecumenicism. J Hered 86:248–249.

Roff DA and Bentzen P, 1989. The statistical analysis of mitochondrial DNA polymorphisms: χ^2 and the problem of small samples. Mol Biol Evol 6:539–545.

Ruzzante DE, Taggart CT, and Cook D, 1996. Spatial and temporal variation in the genetic composition of a larval cod (*Gadus morhua*) aggregation: cohort contribution and genetic stability. Can J Fish Aquat Sci 53:2695–2705.

Ryman N, Utter F, and Laikre L, 1995. Protection of intraspecific biodiversity of exploited fishes. Rev Fish Biol Fish 5:417–446.

Shaklee JB and Bentzen P, in press. Genetic identification of stocks of marine fish and shellfish. Bull Mar Sci.

Shaklee JB and Samollow PB, 1984. Genetic variation and population structure in a spiny lobster, *Panulirus marginatus,* in the Hawaiian archipelago. Fish Bull US 82:693–702.

Scheltema RS, 1971. Larval dispersal as a means of genetic exchange between geographically separated populations of shallow-water benthic marine gastropods. Biol Bull 140:284–322.

Scribner KT, Crane PA, Spearman WJ, and Seeb LW, in press. DNA and allozyme markers provide concordant estimates of population differentiation: analyses of U.S. and Canadian populations of Yukon River fall-run chum salmon. Can J Fish Aquat Sci.

Slatkin M, 1985a. Gene flow in natural populations. Annu Rev Ecol Syst 16:393–430.

Slatkin M, 1985b. Rare alleles as indicators of gene flow. Evolution 39:53–65.

Slatkin M, 1993. Isolation by distance in equilibrium and non-equilibrium populations. Evolution 47:264–279.

Slatkin M, 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457–462.

Slatkin M and Barton NH, 1989. A comparison of three indirect methods for estimating average levels of gene flow. Evolution 43:1349–1368.

Smith CAB, 1969. Local fluctuations in gene frequencies. Ann Hum Genet Lond 32:251–260.

Taylor BL and Gerrodette T, 1993. The uses of statistical power in conservation biology: the vaquita and northern spotted owl. Conserv Biol 7:489–500.

Tegner MJ and Butler RA, 1985. A drift tube study of the dispersal potential of green abalone (*Haliotis fulgens*) larvae in the Southern California Bight: implications for recovery of depleted populations. Mar Ecol Prog Ser 26:73–84.

Waples RS, 1986. A multispecies approach to the analysis of gene flow in marine shore fishes (PhD dissertation). San Diego: University of California, San Diego.

Waples RS, 1987. A multispecies approach to the analysis of gene flow in marine shore fishes. Evolution 41: 385–400.

Waples RS, 1989. Temporal stability of allele frequencies: testing the right hypothesis. Evolution 43:1236–1251.

Waples RS, 1990a. Conservation genetics of Pacific salmon. II. Effective population size and the rate of loss of genetic variability. J Hered 81:267–276.

Waples RS, 1990b. Conservation genetics of Pacific salmon. III. Estimating effective population size. J Hered 81:277–289.

Waples RS, 1991a. Genetic methods for estimating the effective size of cetacean populations. Rep Int Whaling Comm 13:279–300.

Waples RS, 1991b. Pacific salmon, *Oncorhynchus* spp., and the definition of "species" under the Endangered Species Act. Mar Fish Rev 53(3):11–22.

Ward RD, Woodward M, and Skibinski DOF, 1994. A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. J Fish Biol 44:213–232.

Weir BS and Cockerham CC, 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370.

Wirgin I, Grunwald C, Garte SJ, and Mesing C, 1991. Use of DNA fingerprinting in the identification and management of a striped bass population in the southeastern United States. Trans Am Fish Soc 120:273–282.

Workman PL and Niswander JD, 1970. Population studies on southwestern Indian tribes. II. Local genetic differentiation in the Papago. Am J Hum Gen 22:24–29.

Wright S, 1943. Isolation by distance. Genetics 28:114–138.

Wright S, 1978. Evolution and the genetics of populations. Vol. 4. Variability within and among natural populations. Chicago: University of Chicago Press.

Zhang Q, Saghai Maroof MA, and Kleinhofs A, 1993. Comparative diversity analysis of RFLPs and isozymes within and among populations of *Hordeum vulgare* ssp. *spontaneum.* Genetics 134:909–916.

Corresponding Editor: Fred W. Allendorf