*Conservation Genetics* **3:** 167–173, 2002.



*Review*

# Is $F_{ST}$ obsolete?

Joseph E. Neigel
*Department of Biology, Box 42451, University of Louisiana at Lafayette, Lafayette, Louisiana 70504 (E-mail: jneigel@louisiana.edu)*

Received 15 October 2001; accepted 13 November 2001

*Key words:* $F_{ST}$, dispersal, gene flow, *Nm*, population structure

**Abstract**

Since the introduction of allozyme methods in the mid 1960s it has been a standard practice to report Wright's measure of population subdivision, $F_{ST}$, for surveys of genetic variation. Its widespread use has provided us with a sense of what values can be expected in particular situations and how they can be interpreted. With some theoretical justification, $F_{ST}$ has also been used to estimate rates of gene flow. However there are conditions under which $F_{ST}$ is inappropriate for gene flow estimation and can lead to incorrect or even absurd conclusions. These pitfalls have prompted critics to suggest that $F_{ST}$ has failed to deliver what its proponents have promised and should be abandoned. A further challenge has been the development of new methods that offer even greater promise. Thus it is reasonable to ask if perhaps it is time to retire $F_{ST}$ and turn to new and more powerful methods for the inference of gene flow from genetic markers. Here I will argue that although gene flow should be estimated by more powerful approaches whenever practical, $F_{ST}$ remains a useful measure of the average effects of gene flow and will continue to be used for comparative purposes.

## What is $F_{ST}$?

Several definitions of $F_{ST}$ can be found in the literature. Wright's original definition (Wright 1951) is based on the inbreeding coefficient: the probability of alleles that are identical-by-descent (from an ancestral population) being combined in zygotes. This is the most fundamental definition of $F_{ST}$ because inbreeding coefficients are defined simply by the pattern of mating. With this definition it is possible to *estimate* $F_{ST}$ from allele frequencies, but $F_{ST}$ is not *defined* by allele frequencies and is independent of the particular characteristics of individual loci or alleles. Later definitions that were intended to be more practical are based explicitly on parameters of allele frequency distributions, such as the proportion of variance in allele frequencies that is among populations:

$$F_{ST} = \frac{\mathrm{Var}(p)}{\bar{p}(1-\bar{p})}$$

Definitions of $F_{ST}$ that differ from Wright's have been given distinct names, including $G_{ST}$ (Nei 1973), $\theta$ (Weir and Cockerham 1984), $N_{ST}$ (Lynch and Crease 1990), and $R_{ST}$ (Slatkin 1995). Some definitions lend themselves to a general analysis of variance approach that can be been extended to all forms of genetic variation (Excoffier 1992). However if they do not correspond to inbreeding coefficients they will not necessarily have the same relationship to gene flow as $F_{ST}$.

## $F_{ST}$ as a measure of gene flow

The use of $F_{ST}$ as an indirect measure of gene flow is suggested by Wright's island model of population subdivision (Slatkin 1985). Wright considered several models of population structure, although his "infinite islands model" (Wright 1951, 1965) is most often cited as the basis of gene flow estimation. It assumes an infinite number of equal-sized island populations that exchange migrants at a constant rate. The island

populations can be treated as replicates that are characterized by just two parameters: population size ($N$) and the migration rate ($m$). The strength of genetic drift is proportional to $1/N$, while the strength of gene flow is proportional to $m$. Wright showed that at equilibrium:

$$F_{ST} = \frac{1}{(4Nm + 1)}$$

Slatkin and Barton (1989) reviewed the theoretical literature on the relationship between $F_{ST}$ and gene flow and investigated the robustness of estimates based on this relationship. They found support for the generality of the island model from both analytical theory and from their simulations in which estimates of $Nm$ were reasonably close to their true values. Their conclusion was that $F_{ST}$ provided useful estimates of average gene flow in populations at equilibrium. However a decade later two commentaries were published that stridently argued against the use of $F_{ST}$ to estimate gene flow. Bossart and Prowell (1998) suggested that in spite of the "difficulties associated with interpreting patterns of molecular variation and model-based estimates . . . these cautions have not been widely embraced by the scientific community". They pointed to publications in the journal *Evolution* between 1996 and 1997, which they claimed showed that "analyses of population structure rely almost exclusively on antiquated methodology". The methodology they referred to included the use of allozymes as genetic markers as well $F_{ST}$. This pessimistic view was soon challenged (Bohonak et al. 1998), with additional statistics to show that papers in *Molecular Ecology* utilized DNA markers more often than allozymes, and very few papers (2/67) in either *Evolution* or *Molecular Ecology* interpreted $Nm$ estimates literally. The second paper that criticized $F_{ST}$-based estimation of gene flow appeared the following year and focused on mathematical and statistical problems (Whitlock and McCauley 1999). It included results from simulation of an island model in which estimates of both $F_{ST}$ and $Nm$ were subject to extremely high variance. They concluded, "it is rare that $F_{ST}$ can be translated into an accurate estimate of $Nm$" and stated that estimates of $Nm$ were only likely to be correct within a few orders of magnitude. However with the parameters they chose for their simulation ($Nm$ = 50), it is not surprising that inaccurate estimates resulted. As discussed below, the accuracy of $F_{ST}$-based estimates of gene flow is critically dependent on demographic and genetic parameters. This explains how such disparate conclusions could be reached by Slatkin and Barton (1989) on one hand and Bossart and Prowell (1998) or Whitlock and McCauley (1999) on the other.

In spite of its known limitations, estimates of $F_{ST}$ are often consistent with biologically-informed expectations. For example, empirical estimates of gene flow based on $F_{ST}$ are strongly rank-correlated with independent assessments of vagility (Bohonak 1999). Thus while $F_{ST}$ does not always provide an accurate estimate of gene flow it is unlikely that these estimates are usually wrong by orders of magnitude. Fortunately, because we do understand the limitations of $F_{ST}$ we can evaluate the appropriate assumptions and avoid the most serious kinds of misuse.

## The Infinite Island model

A common criticism of $F_{ST}$-based estimation of gene flow is that it assumes the infinite island model of population structure, which is clearly unrealistic. The model stipulates an infinite number of populations, no selection, no mutation and gene flow that is unaffected by the geographic distance between populations. No one has seriously argued that natural populations have these characteristics; the model is just a convenient abstraction that isolates the opposing effects of genetic drift and gene flow. However the model is relevant to the interpretation of data from real populations because it is possible to relax its assumptions without greatly altering the relationship between $F_{ST}$ and $Nm$. The number of populations doesn't really have to be infinite (or even a very large number), mutation and selection are only likely to be important when populations are very large, and even if gene flow is limited by distance the overall value of $F_{ST}$ is expected to be similar to that obtained for the infinite island model (Slatkin and Barton 1989). Another simplification of the island model is that the parameters $N$ and $m$ are the same for every population; it is a model of the overall effects of gene flow among a group of populations rather than a model of the effects of gene flow between particular pairs of populations. If estimates of other parameters are needed, the island model is inappropriate. Unfortunately, it has not been a trivial matter to develop estimates of gene flow for more complex "landscape" models (Felsenstein 1982), although there has recently been some progress. The infinite island model has provided a robust, albeit coarse guide to how overall levels of gene flow influence overall $F_{ST}$. Use of the island model in this way

neither requires nor implies that its assumptions are thought to be true.

## When is $F_{ST}$ accurate?

Much of the argument over the usefulness of $F_{ST}$ for estimates of gene flow has to do with the characteristics of the populations being considered. In most of the simulations used by Slatkin and Barton (1989) the population size was 128. In contrast, $F_{ST}$ is sometimes estimated for populations that span large portions of a species range and are likely to number in the millions or more (e.g. Bossart and Scriber 1995). This is a crucial distinction because estimates of gene flow from $F_{ST}$ are problematic when genetic drift is weak (effective population sizes are large). $F_{ST}$ can be thought of as a dynamic balance between genetic drift and gene flow, and even a small amount of gene flow can overcome weak drift and bring $F_{ST}$ close to zero. Small values of $F_{ST}$ cannot be precisely estimated because they correspond to differences in population allele frequencies that are small relative to differences that arise by chance in samples of populations (Weir and Cockerham 1984; Waples 1998). Because of the non-linearity of the relationship between $F_{ST}$ and $Nm$, when $F_{ST}$ is small the variance in its estimator is transformed into a much larger variance in the estimator of $Nm$ (Waples 1998). These estimation problems with low $F_{ST}$ explain the high variance found in Whitlock and McCauley's simulation (Whitlock and McCauley 1999), in which the true value of $F_{ST}$ was only 0.005 ($N = 100$, $m = 0.5$).

If both genetic drift and gene flow are weak the equilibrium value of $F_{ST}$ may be large enough to provide a precise estimate of gene flow but the observed $F_{ST}$ may be far from this equilibrium. The time required to be near equilibrium is on the order of $1/[2m + 1/(2N)]$ (Crow and Aoki 1984). Large populations that were historically connected but are now completely isolated would be fixed for different alleles at equilibrium; but the time required to approach this equilibrium could be far greater than the age of most species. Estimates of gene flow based on such data are of course inappropriate, and likely to produce absurd results such as non-zero gene flow between species that have long been reproductively incompatible.

When genetic drift and gene flow are weak the effects of other forces such as selection and mutation may be significant. The potential effect of mutation on $F_{ST}$ in large populations is especially important. Although it is often stated that $F_{ST}$ is not very sensitive to mutation rates, this conclusion is only valid when effective population sizes are small (i.e., 1000 or less) and mutation rates are much lower than migration rates (Crow and Aoki 1984). However when large populations are surveyed for markers such as mitochondrial sequences or microsatellite loci that have high mutation rates, the effect of mutation on $F_{ST}$ may be considerable. High mutation rates will lower $F_{ST}$ (Crow and Aoki 1984; Neigel 1997; Balloux 2000), although this is counter to the notion that microsatellites are "high resolution" markers (Bossart and Prowell 1998). This results in a upward bias in estimates of gene flow when highly polymorphic markers are used (Hedrick 1999). Differences in mutation rate among loci will also lead to high interlocus variance in $F_{ST}$ estimates. Whether this represents a limitation in the application of population genetic models to the estimation of gene flow or a problem in the way $F_{ST}$ is estimated depends on one's definition of $F_{ST}$ (Neigel 1997).

The scenario depicted in Figure 1 illustrates a crucial difference between $F_{ST}$ defined as an inbreeding coefficient and $F_{ST}$ defined in terms of allelic differences. In this scenario there are two populations and one locus with two major allelic variants: *A* and *B*. These major variants are only distantly related and are distinguished by a large number of independent point mutations. All of the *A*-alleles are found in one population: the A-population, while all of the *B*-alleles are found in the other population: the B-population. Each major allelic variant is represented by four minor variants that are distinguished by single point mutations, and each has a frequency of 0.25 in the population where it occurs. As shown in the figure, in each population all alleles are identical-by-descent with respect to the ancestral population. Thus by Wright's original definition, $F_{ST}$ is 1 which would correspond to an $Nm$ estimate of zero. However if $F_{ST}$ is defined in terms of the frequencies of the minor variants, its value would be 0.14 and the estimate of $Nm$ would be 1.5. It is likely that a survey of restriction site variation would distinguish only the major variants while a DNA-sequence survey would detect all of the minor variants and thus lead to these very different estimates of $F_{ST}$ and $Nm$. This example also illustrates the effect of mutations on $F_{ST}$. Infrequent mutations would distinguish only the two major alleles while more frequent mutations would generate minor variants and lead to lower values for estimates of $F_{ST}$ based on allele frequencies.

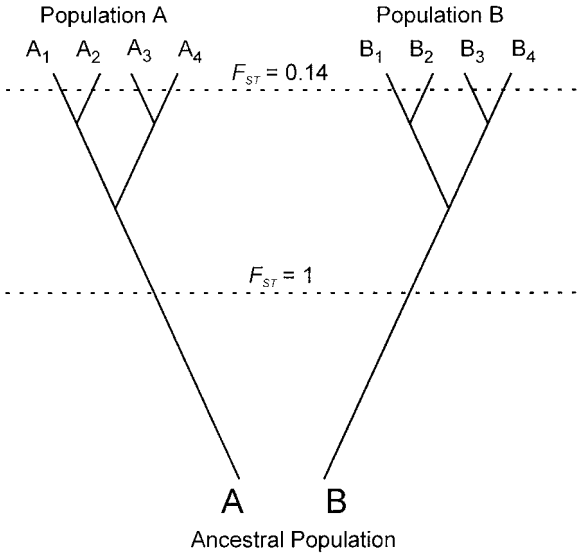


*Figure 1.* Identity-by-descent versus sequence identity for two groups of alleles that distinguish two populations.

Slatkin (1991) showed that $F_{ST}$ could be defined in terms of coalescence times between alleles rather than the frequencies of mutations that distinguish them. This idea was applied to the development of $R_{ST}$ (Slatkin 1995) as an estimator of $F_{ST}$ that is relatively insensitive to mutation rates and thus more appropriate for rapidly mutating microsatellite loci. As expected, the value of $R_{ST}$ tends to be higher than other estimators of $F_{ST}$. However $R_{ST}$ also tends to have a higher variance than other measures of population differentiation and is sensitive to violations of the stepwise mutation model (Balloux et al., 2000; Richard and Thorpe 2001). Presumably the use of $R_{ST}$ to estimate *Nm* trades an increase in accuracy (less bias) with a decrease in precision (more variance).

## New and improved methods

Distributions of population genetic variation contain information that is lost when characterized by a single measure such as $F_{ST}$. This is especially true of variation with phylogeographic structure (Avise, Arnold et al. 1987). This information could be used not only to obtain more accurate estimates of *Nm*, but also to estimate additional parameters of more complex and realistic models of population structure. Changes in the sizes of populations or in levels of gene flow over time are predicted to create recognizable patterns, and these patterns are often thought of as "signatures" of particular kinds of population histories (Avise 2000). For example, it is predicted that gene genealogies will exhibit a "star phylogeny" (Figure 2) in a population that has been expanding in size (Avise, Neigel et al. 1984; Slatkin and Hudson 1991). These descriptive patterns are useful as heuristics and help us understand how we can use population genetic data to reconstruct population histories (Templeton 1998). However it is not clear that they provide reliable evidence of particular kinds of population histories. Often the same pattern may be created by any of several different population histories. For example, a star phylogeny could also result if a once continuous population became fragmented (Figure 2). Furthermore, because of the stochastic nature of population structure, apparently striking (and statistically non-random) patterns can arise simply by chance (Ball, Neigel et al. 1990). To test the interpretation of such a pattern, the null hypothesis should not be simply the absence of a statistically significant pattern but should also include the possibility that the pattern has been generated by something other than the proposed cause.

The problem with matching population genetic "signatures" to particular scenarios is that there are many different population structures or histories that are compatible with a particular set of data. What is needed is a measure of how much the data supports one scenario over others. Likelihood provides such a measure and maximum likelihood methods can be used to choose the model (or parameters of a model) that has the most support from the data (Edwards 1992). Likelihood calculations require a probability model. Early attempts to develop likelihood methods for estimation of gene flow assumed some form for the distribution of allele frequencies among populations that allowed the likelihood of observed distributions to be calculated for different values of gene flow parameters (Barton, Halliday et al. 1983; Wehrhahn and Powell 1987; Slatkin and Barton 1989). More recently, coalescent models (Tavarae 1984) have been used to calculate the likelihood of observed genetic data purely in terms of underlying population genetic parameters such as mutation rate, effective population size, and migration rate. These models eliminate the need to reduce the information in the data to either a single statistic or a single generalized distribution. The early successes of this approach have been impressive. Beerli and Felsenstein (1999, 2001) have developed a method that simultaneously estimates the effective sizes of multiple populations as well as rates of migration

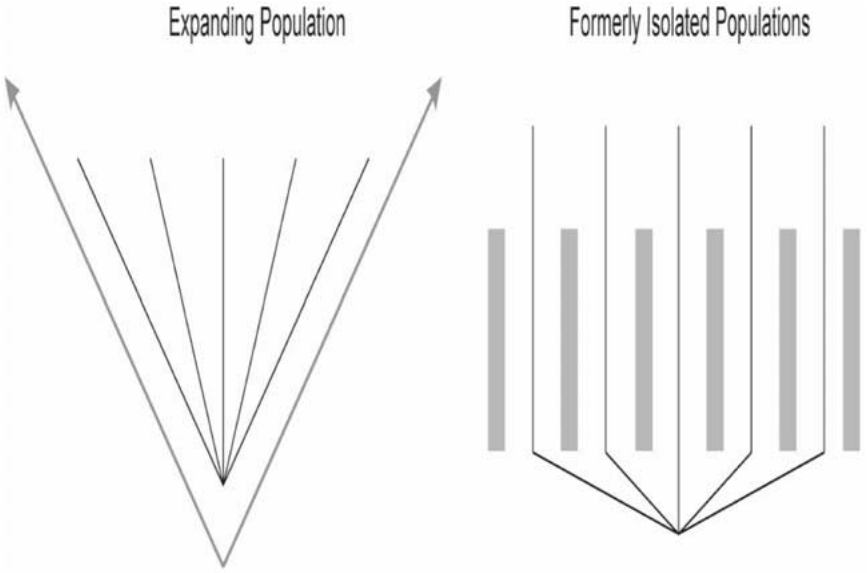


*Figure 2.* Star phylogenies generated either by population expansion or past population subdivision.

between every pair of populations. This method is implemented in the program MIGRATE, which is available as a free, open-source download from http://evolution.genetics.washington.edu/lamarc.html. There has also been rapid progress in the development of methods to infer historical changes in population structure (Nielsen and Slatkin 2000; Nielsen, Mountain et al. 1998; O'Ryan, Harley et al. 1998). These methods would be more appropriate than $F_{ST}$ for estimates of gene flow between large populations that currently experience low gene flow.

Over the past few years it has become clear that analysis of data by likelihood methods and coalescent models could overcome many of the limitations of $F_{ST}$ for estimation of gene flow, and it has even been predicted that these methods will replace the use of $F_{ST}$ (Beerli and Felsenstein 2001). However at present, this approach is limited by its computational requirements so that it is difficult to estimate multiple parameters from data sets that represent large numbers of individuals, populations, and loci. This also limits the speed with which these methods can be tested with data from simulations. Because likelihood calculations are sensitive to violations of the underlying probability model, it will be important to examine the robustness of these estimates as it has been for $F_{ST}$.

**Direct estimates of gene flow**

Estimates of $F_{ST}$ based on genetic markers reflect the cumulative effects of genetic drift and gene flow for groups of populations. Because of this "averaging", they are not appropriate for estimates of gene flow between specific pairs within a network of populations or for estimates of instantaneous rates of gene flow (Felsenstein 1982). This is not to suggest that pair-wise values of $F_{ST}$ are not useful. $F_{ST}$ can be used simply as a relative measure of population divergence without assuming any particular relationship to gene flow. Correlations between pair-wise estimates of $F_{ST}$ and geographic distance can also be used to detect "isolation-by-distance", the dependence of gene flow on geographic distance (Slatkin 1993). However $F_{ST}$ between populations within a network may be affected by indirect as well as direct connections. A classic example is provided by Wright's "Continent-Island" model (Wright 1931, 1940), in which there is no gene flow between "islands" but $F_{ST}$ among island populations is lowered by gene flow from the "continent" to the "islands". If an estimate of contemporary gene flow between specific populations is needed, direct estimates of gene flow are more appropriate. Direct estimates may be based on standard mark-and-recapture methods, or genetic markers can be used to identify the source populations of migrant individuals or gametes (Devlin and Ellstrand 1990; Davies, F. X. Villablanca et al. 1999). These methods provide a

view of gene flow that is complementary, although not necessarily superior to indirect methods such as $F_{ST}$. Rates of gene flow are likely to fluctuate considerably in natural systems and intermittent episodes of high gene flow that would be detected by their long-term effect on $F_{ST}$ could be missed in short-term observations (Slatkin 1985, 1987). $F_{ST}$ is thus better at providing the "big picture" of the overall, cumulative effects of gene flow and its role in evolution.

## A role for $F_{ST}$?

There are now two major alternatives to $F_{ST}$ for the estimation of gene flow. For questions about contemporary gene flow to particular sites, direct estimation based on mark-and-recapture or genetic identification of migrant sources is the preferred alternative. For questions about patterns of gene flow among multiple populations, questions that have usually been addressed with $F_{ST}$, there is now an alternative approach based on likelihoods calculated by coalescent methods. The new likelihood methods are expected to be generally superior to those based on $F_{ST}$, and can be used under conditions when $F_{ST}$ should not be, such as large effective population sizes with low gene flow. With progress in reducing their steep computational requirements and more extensive testing, methods based on likelihood calculations and coalescent models could soon replace $F_{ST}$ for estimation of gene flow (Beerli and Felsenstein 2001). As these new methods become practical, it will be important to compare them with estimates based on $F_{ST}$. From such comparisons will we be able to objectively judge $F_{ST}$ as a measure of gene flow and gauge the progress that we have made.

Even with the development of powerful alternatives, it is likely that we will continue to use $F_{ST}$ as a relative measure of population structure and for comparative estimates of gene flow. $F_{ST}$ has been the standard measure of population structure for several decades and is of fundamental theoretical importance. Furthermore, it is often the only information we have about gene flow from past studies. These older studies will continue to be important as baselines against which changes in population structure can be detected and as part of a growing knowledge base of population genetic variation. Although it may be desirable to reanalyze data from past surveys of variation with new methods, it will not often be possible to do so because the necessary data has not been included in publications and is no longer available from authors (Leberg and Neigel 1999). Fortunately, $F_{ST}$ does impart useful information about gene flow and we are aware of its limitations.

## References


Avise JC (2000) *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge MA.

Avise JC, Arnold J et al. (1987) Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, **18**, 489–522.

Avise JC, Neigel JE, Arnold J (1984) Demographic influences on mitochondrial DNA lineage survivorship in animal populations. *J. Mol. Evol.*, **20**, 99–105.

Ball RM, Neigel JE, Avise JC (1990) Gene genealogies within the organismal pedigrees of random mating populations. *Evolution*, **44**, 360–370.

Balloux F, Brunner H et al. (2000) Microsatellites can be misleading: An empirical and simulation study. *Evolution*, **54**, 1414–1422.

Barton NH, Halliday RB, Hewitt GM (1983) Rare electrophoretic variants in a hybrid zone. *Heredity*, **50**, 139–146.

Beerli P, Felsenstein J (1999) Maximum-likelihood estimation of migration rates and effective population nubmers in two populations using a coalescent approach. *Genetics*, **152**, 763–773.

Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences, USA*, **98**, 4563–4568.

Bohonak AJ (1999) Dispersal, gene flow, and population structure. *Quarterly Review of Biology*, **74**, 21–45.

Bohonak AJ, Davies N, Roderick GK, Villablanca FX (1998) Is population genetics mired in the past? *Trends in Ecology and Evolution*, **13**, 360–360.

Bossart JL, Prowell DP (1998) Genetic estimates of population structure and gene flow: limitations, lessons and new directions. *Trends in Ecology and Evolution*, **13**, 202–206.

Bossart JL, Scriber JM (1995) Maintenance of ecologically significant genetic variation in the tiger swallowtail butterfly through differential selection and gene flow. *Evolution*, **49**, 1163–1171.

Crow JF, Aoki K (1984) Group selection for a polygenic behavioral trait: Estimating the degree of population subdivision. *PNAS USA*, **81**, 6073–6077.

Davies N, Villablanca FX, Roderick GK (1999) Determining the source of individuals: Multilocus genotyping in nonequilibrium population genetics. *Trends in Ecology and Evolution*, **14**, 17–21.

Devlin B, Ellstrand NC (1990) The development and application of a refined method for estimating gene flow from angiosperm paternity analysis. *Evolution*, **44**, 248.

Edwards AWF (1992) *Likelihood*. The Johns Hopkins University Press, Baltimore.

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.

Felsenstein J (1982) How can we infer geography and history from gene frequencies? *J. Theor. Biol.*, **96**, 9–20.

Hedrick PW (1999) Perspective: Highly variable loci and their interpretation in evolution and conservation. *Evolution*, **53**, 313–318.

Leberg PL, Neigel JE (1999) Enhancing the retrievability of population genetic survey data? An assessment of animal mitochondrial DNA studies. *Evolution*, **53**, 1961–1965.

Lynch M, Crease TJ (1990) The analysis of population survey data on DNA sequence variation. *Mol. Biol. Evol.*, **7**, 377–394.

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

Neigel JE (1997) A comparison of alternative strategies for estimating gene flow from genetic markers. *Annual Review of Ecology and Systematics*, **28**, 105–128.

Nielsen R, Mountain JL, Huelsenbeck JP, Slatkin M (1998) Maximum-likelihood estimation of population divergence times and population phylogeny in models without mutation. *Evolution*, **52**, 669–677.

Nielsen R, Slatkin M (2000) Likelihood analysis of ongoing gene flow and historical association. *Evolution* **54**, 44–50.

O'Ryan CO, Harley EH et al. (1998) Microsatellite analysis of genetic diversity in fragmented South African buffalo populations. *Animal Conservation*, **1**, 85–94.

Richard M, Thorpe RS (2001) Can microsatellites be used to infer phylogenies? Evidence from Population Affinities of the Western Canary Island Lizard (*Gallotia galloti*). *Mol. Phylogenet. Evol.*, **20**, 351–360.

Slatkin M (1985) Gene flow in natural populations. *Ann. Rev. Ecol. Syst.*, **16**, 393–430.

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

Slatkin M (1991) Inbreeding coefficients and coalescence times. *Genet. Res. Camb.*, **58**, 167–175.

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, Barton NH (1989) A comparison of three indirect methods for estimating average levels of gene flow. *Evolution*, **43**, 1349–1368.

Slatkin M, Hudson RR (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*, **129**, 555–562.

Tavarae S (1984) Line-of-descent and genealogical processes, and their applications in population genetic models. *Theor. Popul. Biol.*, **26**, 119–164.

Templeton AR (1998) Nested clade analyses of phylogeographic data: Testing hypotheses about gene flow and population history. *Molecular Ecology*, **7**, 381–397.

Waples RS (1998) Separating the wheat from the chaff: Patterns of genetic differentiation in high gene flow species. *J. Heredity*, **89**, 438–450.

Wehrhahn CF, Powell R (1987) Electrophoretic variation, regional differences, and gene flow in the coho salmon (*Onchorhynchus kisutch*) of southern British Columbia. *Can. J. Fish. Aqaut. Sci.*, **44**, 822–831.

Weir BS, Cockerham CC (1984) Estimating F-Statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: $F_{ST}$ not equal 1/(4*Nm* + 1). *Heredity*, **82**, 117–125.

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

Wright S (1940) Breeding structure of populations in relation to speciation. *Am. Nat.*, **74**, 232–248.

Wright S (1951) The genetical structure of populations. *Ann. Eugen.*, **15**, 323–353.

Wright S (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution*, **19**, 395–420.