Turkish Journal of Fisheries and Aquatic Sciences 2: 145-155 (2002) **REVIEW**

Fish Population Genetics and Applications of Molecular Markers to Fisheries and Aquaculture: I- Basic Principles of Fish Population Genetics

Yılmaz Çiftci[1,*], İbrahim Okumuş[2]
[1]Trabzon Central Fisheries Research Institute, P.O. Box 129, 61001, Trabzon, Turkey.
[2] Karadeniz Technical University, Faculty of Marine Sciences, Department of Fisheries, 61530 Çamburnu, Trabzon, Turkey

* Corresponding Author: Tel.: +90. 462. 341 10 52; Fax: +90. 462. 341 11 57;
E-mail: yilmazciftci@ttnet.net.tr

Received 12 December 2002
Accepted 08 March 2003

Abstract

Fisheries management is getting difficult due to over utilization of fish stocks, pollution and various human activities resulting reduction of genetic resources and variations. Therefore, molecular genetic studies of natural population is dependent on the polymorphic neutral markers and offer the possibility of investigation of population structure and provide scientific data for regulation of harvest to protect weaker populations and finally long term management of fisheries resources. The study of the genetic variation in populations and its change, the following of allele frequencies in populations through time and space, is the main subject of population genetics.

This paper emphasizes the importance and development in population genetics and contains a description of procedures used in population genetic studies together with references.

Key Words: population genetics, fisheries management, gene flow, Hardy Weinberg Equilibrium, *HWE*, effective population size, genetic distance

Contents

Introduction

Of all the animals and plants in the aquatic environment fish is the most important source of human food. They are the major source of protein for many people and constitute the main part of the diet in many cultures. Fish and fisheries not only provide a significant portion of the protein available for human consumption, but they are also an economically significant activity, providing jobs and investment opportunities and, for many countries, a means of improving the balance of international trade.

Most of the fish used for human consumption is obtained through exploitation of wild populations. The management of the wild populations comprising commercial or sport fisheries presents genetic problems that are unique to fisheries management. Reduction in the genetic resources of natural fish populations has become an important fisheries management problem. Much of the reduction is due to various human activities. Not only has the genetic diversity of many fish populations been altered, but many thousands of populations and species have been extirpated by pollution, over fishing exploitation, destruction of habitat, blockage of migration routes and other human developments (Ferguson, 1995). In nearly all cases, fishery management has largely been concerned with the immediate resource of interest; that is, the abundance and size of fish available for harvesting (Ward and Grewe, 1995). This short-term focus may be economically advantageous in the short run, but in the long term may cause extinction of the population. Concern with reduction of genetic resources in fish is part of a larger global concern for the genetic resources of the biosphere. For this reason, molecular genetic research should be strongly supported, for it is vital to the long-term management of fisheries resources (Park and Moran, 1995). This approach addresses two slightly different aspects of genetic resources: conservation of gene pools and conservation of genetic diversity.

The study of the genetic variation in populations and its change, the following of allele frequencies in populations through time and space, is the main subject of population genetics. Genetic variation is the raw material in a species and populations, which enables them to adapt to changes in their environment. New genetic variation arises in a population from either spontaneous mutation of a gene or by immigration from a population of genetically different individuals. The number and relative abundance of alleles in a population is a measure of genetic variation.

This fisheries genetics review series attempt to revise current knowledge and practices in fish population genetics and molecular markers and applications of the molecular genetics in fisheries management, conservation and aquaculture. The review is divided into two sections and is going to be published in successive issues of this Journal. Section one covers basic concepts of population genetics from fisheries management point of view, while sections two will focus on recent developments in molecular markers and their role in the management and conservation of natural fish populations and applications to aquaculture.

In this first section, principles of fish population genetics, testing the basic assumptions for population genetic analysis, departures from Hardy-Weinberg expectation, linkage disequilibrium between the loci, estimation of the genetic differentiation within and among populations, using genetic distance, F_{ST} and gene flow analyses, determination of frequencies of genes and genotypes, estimation of effective population sizes and null alleles have been outlined.

The Stock Concept

A fundamental problem for fisheries is the identification of populations of a species and this idea has been brought together with the definition of stock for management (Carvalho and Hauser, 1995). The term stock has been used in various management contexts with little or no genetic content. Several approaches have been advocated to solve this problem. Ihssen *et al*. (1981) defines a stock as “an intraspecific group of randomly mating individuals with temporal or spatial integrity”. Larkin (1981) defined a stock as "a population of organisms which share a common gene pool, is sufficiently discrete to warrant consideration as a self-perpetuating system which can be managed." In fishery management, a unit of stock is normally regarded as a group of fish exploited in a specific area or by a specific method (Carvalho and Hauser, 1995).

The stock concept has two central arguments: that fish species are subdivided into local populations and that there may be genetic differences between local populations which are adaptive. If managers are to include genetic considerations in their decisions, they will need information on the biological differences between discrete local groups of a species and they will need to understand the genetic and ecological processes that influence discreteness (Palumbi, 1996; Ward and Grewe, 1995).

Genetics and Fisheries Management

Genetics and fishery management can interact in several ways. When the genetic population structure of a species is known, the distribution of subpopulations in mixed fisheries can be estimated (Utter, 1991). Regulation of harvest to protect weaker populations can be made based on these distributions. It is important to identify and regulate for genetic changes within a population because of differential harvests because of the drastic and long-term effects they may have on a population.

The genetic study of natural populations is dependent on the availability of polymorphic neutral

markers. Although electrophoresis of proteins has been widely used for the direct study of genetic variation in fish populations, DNA markers are becoming more popular in order to obtain information about gene flow, allele frequencies and other parameters that are crucial in population biology (Neigel, 1997).

Biologically important characteristics of populations, including their size and productive efficiency, are determined by the historically established gene pools (Altukhov and Salmenkova, 1987). Therefore, the population genetic analysis of species in nature is of primary importance in developing an optimal strategy for their effective management. Such a strategy should provide not only for maximum economic benefits but also for continuous maintenance of natural populations. Fisheries biologists must emphasize the importance of elucidating the factors and conditions that permit populations and species to be maintained (Allendorf *et al*., 1987).

Population Genetics

The general goals of population genetic studies are to characterize the extent of genetic variation within species and account for this variation (Weir, 1996). The amount of genetic variation within and between populations can be determined by the frequency of genes and the forces that affect their frequencies, such as migration, mutation, selection and genetic drift (Gall, 1987).

During the last two decades, a large amount of genotype and allele frequency data have been obtained from a large number of species, including many fish species, primarily through the means of protein and DNA base molecular genetic techniques. These studies have shown that most species are subdivided into more or less distinct units that differ genetically from each other (Chakraborty and Leimar, 1987). At this point intraspecific groups of fish have to be described to prevent confusion by terms such as race, tribe, population, subpopulation, stock and subspecies and are intended to reflect the magnitude of differences among such subdivisions (Ihssen *et al*., 1981).

Genetic differences between subpopulations will evolve in the course of time if there is little or no gene flow between them (Chakraborty and Leimar, 1987). Gene flow rates of 10% or less may justify treatment as separate stocks. That means restriction on gene flow may lead to genetic subdivision. In particular, marine species show lower levels of genetic population differentiation than fresh water or anadromous species, probably because there are potentially fewer barriers to migration and gene flow (Carvalho and Hauser, 1995).

Most indirect estimates of gene flow have been based on demographic models in which migration occurs between discrete subpopulations (Neigel, 1997). There are essentially three principle models of population structure, which can result in differentiation of genetic patterns within and between geographic localities (May and Kruger, 1990; Baverstock and Moritz, 1996). These are, first, the "panmictic model": the entire population may consist of single panmictic unit (free exchange); second, "stepping-stone" or "island model", may consist of a series of small subpopulations each largely isolated from other subpopulations (no interchange between subpopulation); Third, the "isolation-by-distance" model: a continuous population, but with organisms exchanging genes only with geographically close areas (local interchange only).

Gene flow among subpopulations is a characteristic attribute of population genetic studies. With high levels of migration and gene flow between populations, the similarity of populations increases (Neigel, 1997). Thus, the first step in understanding the population genetics of a specific species is to consider which model best describes the population structure (Baverstock and Moritz, 1996).

In fact, in some cases, movement between areas will not always result in gene flow for the reason of spatial and temporal isolation (Carvalho and Hauser, 1995). Spatial isolation mechanisms restrict the gene flow among populations because of natal homing and distance. Natal homing can be distinguished by the tendency for adults to return to the same spawning sites in successive years such as in salmonids. Temporal isolation affects gene flow by differences in spawning time that range from days to years.

Fluctuations in environmental conditions and population density may cause considerable variability. Important factors in genetic variability, like selection, migration and genetic drift are affected by human activities. For example, selection, in general, is a process by which the future contribution of some genotypes to the next generation is limited and it is the dominant mode of human interaction with fish population (Nelson and Soule, 1987). Size selective fishing gear, destruction of habitat, alteration of prey availability, pollution stress and other such activities can impose new selection pressures on a stock or may alter the existing selection forces. Also human activities that affect movements of fish (either directly through erection or removal of barriers or indirectly through changes in population density or environmental conditioning) have the potential to alter the genetic diversity of a population (Nelson and Soule, 1987). Human activities related to those changes are habitat alteration, transplantation of exotic stocks, introduction of hatchery-reared strains and over exploitation (Ferguson, 1995).

The importance of genetic variation to population adaptability in changing environments or under stressful conditions has long been recognized (Allendorf *et al*., 1987). The loss of genetic variation, due to prolonged selection, loss of heterozygosity due to (random) inbreeding or isolation may result in a

decrease of the potential adaptability of a population (Ferguson *et al*., 1995).

The dangers inherent in subdivision of fish populations are that inbreeding and genetic drift will lead to fixation of genes, loss of fitness (vigour, viability, fecundity, resistance to disease) and ultimately extinction of local populations (Ferguson *et al*., 1995). Stress such as exploitation may augment these dangers through effects on effective population size (Nelson and Soule, 1987).

A logical conclusion for the above argument might be that each discrete group of fish should be identified as a stock and managed separately, with the goal of harvesting it in such a way as to preserve all genetic variation (Allendorf *et al*., 1987). Maintaining the maximum level of genetic variation both in wild and cultured stocks is vital for the preservation of genetic resources (Nelson and Soule, 1987). Therefore, excessive loss of genetic variability, especially through inbreeding, should be avoided. For long term preservation of genetic diversity, large population sizes are required so that the loss of genetic variability due to genetic drift and selection and gain of genetic variability by mutation can be maintained in equilibrium (FAO, 1981).

Testing the Basic Assumptions for Population Genetic Analysis

Frequencies of Genes and Genotypes

To describe the genetic constitution of a group of individuals it is necessary to specify their genotypes and enumerate each genotype. Suppose for simplicity that a certain autosomal locus, A with two different alleles, A_1 and A_2, was studied among the (N) individuals. There would be three possible genotypes, A_1A_1, A_1A_2 and A_2A_2. The genetic constitution of the group would be fully described by the proportion or percentage of individuals that belonged to each genotype. These proportions or frequencies are called genotype frequencies (P) and would be formulated as follows (Crow and Kimura, 1970):

$P_{A1A1} = N_{(A1A1)} / N$;

$P_{A2A2} = N_{(A2A2)} / N$;

$P_{A1A2} = N_{(A1A2)} / N$

The genetic constitution of a population, referring to the genes it carries, is described by the array of gene frequencies; that is, by specification of the alleles present at every locus and the number and proportions of the different alleles at each locus. Thus frequencies (P) for alleles A_1 and A_2 are determined as follows (Crow and Kimura, 1970):

$P_{A1} = (N_{(A1A1)} + ½\ N_{(A1A2)}) / N \Rightarrow P_{A1} = P_{A1A1} + ½\ P_{A1A2}$

$P_{A2} = (N_{(A2A2)} + ½\ N_{(A1A2)}) / N \Rightarrow P_{A2} = P_{A2A2} + ½\ P_{A1A2}$

Heterozygosity and Polymorphic Loci

One useful measure of genetic diversity is population heterozygosity (H), defined as the mean percentage of loci heterozygous per individual (or equivalently, the mean percentage of individuals heterozygous per locus).

Estimation of heterozygosity for a locus (H_{obs}) can be obtained from raw data that consists of observed diploid genotypes and is formulated as the proportion of observed heterozygotes (h) at a given locus: $H_{obs} = h/N$, where N is the total individuals scored for that locus. Heterozygosities may also be estimated from observed frequencies of alleles (rather than genotypes), assuming the population is in Hardy-Weinberg equilibrium. Thus;

$$H_{exp} = 1 - \sum_{i=1}^{k} P_i^2,$$

where P_i is the frequency of the i^{th} allele and k the number of alleles at a locus (Nei, 1978).

Other common measures of population variability for genetic marker data are the mean number of alleles per locus and the percentage of polymorphic loci (L). To avoid an expected positive correlation between L and sample size, a locus is usually considered polymorphic only if the frequency of the most common allele falls below an arbitrary cut-off, typically 0.99 or 0.95 (Avise, 1994). Polymorphism corresponds then to the proportion of polymorphic loci: $L = x/l$, where x is the number of polymorphic loci and l the total number of loci studied and is commonly expressed as a percentage.

Hardy-Weinberg Equilibrium

If analyses are to be made using allele frequencies, rather than genotypic frequencies, it is necessary to ensure the populations are in Hardy-Weinberg equilibrium. This law states that in a large random mating population with no selection, mutation or migration, the allele frequencies and the genotype frequencies are constant from generation to generation and that, furthermore, there is a simple relationship between the gene frequencies and the genotype frequencies (Hardy, 1908; Weinberg, 1908; in Wright, 1969). The relationship is this: if the gene frequencies of two alleles among the parents are p and q, then the genotype frequencies among the progeny are p^2, $2pq$ and q^2, thus;

	Genes in parents		Genotypes in progeny		
	A_1	A_2	A_1A_1	A_1A_2	A_2A_2
Frequencies	p	q	p^2	$2pq$	q^2

Where p is the allelic frequency of A_1 and q is the allelic frequency of A_2 (Falconer, 1989). A population with these genotype frequencies is said to

be in Hardy-Weinberg equilibrium (*HWE*) at the locus under investigation (Guo and Thompson, 1992).

A deviation from Hardy-Weinberg proportions indicates selection, population mixing or non-random mating and its detection is one of the first steps in the study of population structure (Rousset and Raymond, 1995). Such deviations are usually tested using various methods, which can be divided into two separate groups. In the first are the traditional goodness of fit tests like x^2 tests and the likelihood ratio *G* statistics (Sokal and Rohlf, 1981; Hernandez and Weir, 1989). Such tests usually involve a simple comparison between the observed and expected frequencies of each distinct genotype in a sample from the population. The major limitation of the goodness of fit tests, however, is that they are usually unsuitable when the observed frequencies of some of the classes i.e. genotypes, are small (<5) or absent (Sokal and Rohlf, 1981; Koray, 1993).

The second group of tests, called "exact tests", were developed from Fisher's exact test for contingency tables (Haldane, 1954: in Hernandez and Weir, 1989). Exact tests are very useful even when many rare alleles are present (Rousset and Raymond, 1995) and they, therefore, have been used for population genetic analysis of hypervariable markers such as microsatellite loci (Estoup *et al*., 1993; Di Rienzo *et al*., 1994). The disadvantage of exact tests, however, is that they generally require the enumeration of all possible samples which keep the same allelic frequencies and sample sizes as in the observed sample. As such, this test usually demands prohibitive computing time. The problem is even more obvious as the number of alleles or sample size increase.

Guo and Thompson (1992) described two methods to estimate the exact significance levels (P-values) for a test of *HWE*, which are relatively simple, computationally fast and easily applicable to multi-allelic loci. One is based on a conventional Monte Carlo method and the other on the Markov Chain method. The great power of these approaches has been demonstrated from both simulated and real data (see Guo and Thompson, 1992, for more details). The GENEPOP package of computer programs (Raymond and Rousset, 1995) uses the Markov Chain method of Guo and Thompson (1992). This is the only generally available program usable with the number of alleles. It can be used throughout the study with default values of 1000 dememorisations, 50 batches and 1000 iterations, due to limitation of computing time. The program gives the probability of rejection of H_o (i.e. *HWE*) and the standard error of this value. These values per locus are combined using Fisher's combination of *P* values (Sokal and Rohlf, 1981). The exact probability test used in GENEPOP represents an alternative to bootstrapping and jack-knifing. Exact test estimates the probability value of departure from the null hypothesis, whereas bootstrapping and jack-knifing provide an estimate of the confidence interval around the observed value.

The program also calculates values of F_{IS} (Weir and Cockerham, 1984). This parameter measures the reduction in heterozygosity due to non-random mating within the subpopulation and thus helps to detect departures from Hardy-Weinberg by measuring the amount of heterozygote deficiency or excess observed in the sample. Values significantly greater than zero indicate an excess of homozygotes possibly resulting from inbreeding, population admixture or failure to detect heterozygotes. Conversely negative F_{IS} indicates an excess of heterozygotes and outbreeding.

Linkage Disequilibrium

Looking for relationships between alleles is a reasonable step after observing the *HWE*. Two populations brought into contingence may have genotype frequencies in *HWE* after a single generation of random mating.

The linkage (or gametic) disequilibrium, *D*, measures the lack of fit of observed two-locus gametic frequencies to those anticipated based on the product of the single locus allelic frequencies. That is, the frequency of an A_1B_2 gamete (loci *A* and *B*) in the population should be equal to frequency of the A_1 allele multiplied by the frequency of the B_2 allele (May and Kruger, 1990).

There are two major phenomena responsible for linkage disequilibrium or non-random association of alleles between two loci on a chromosome. They are epistatic natural selection and random genetic drift. The former increases the frequencies of favourable combinations of alleles in a population and stable linkage disequilibrium is expected; the latter causes random fluctuation of gamete frequencies in the population and, hence, increases the variance of the linkage disequilibrium coefficient. Such random fluctuation would be enhanced if the population were divided into subpopulations or if mating were not random in the population (Ohta, 1982).

Furthermore, if the population at issue went through a bottleneck which reduced the effective population size (N_e) to a small number of breeding adults, it might be expected to see a significant linkage disequilibrium value for several generations (May and Kruger, 1990).

Previously, a computer programme described by Weir (1990), which calculates gametic disequilibrium when *HWE* is assumed, was employed. As this programme is limited to a situation of two loci each with two alleles, for multiallelic loci, the 2 alleles represent the common allele and all other alleles are pooled. Thus, an alternative statistical test described by Chakraborty *et al*. (1991) was used to test the non-random association of genes at different loci. GENEPOP (Raymond and Rousset, 1995) tests for compound genotypic disequilibria using the Markov Chain method to predict exact probability. This is

analogous to Fisher's exact test on a 2x2 contingency table which is extended to a "*r*x*k*" contingency table. The hypothesis is the test of a random distribution of k different alleles among r population.

Population Differentiation

Wright's (1951; 1969) F-statistics have proved to be an extremely useful tool for illuminating the pattern and extent of genetic variation residing within and among natural populations of animal and plant species. For a total population that is subdivided into many subpopulations, Wright (1951) defined three F-statistics (correlation between uniting gametes), to relate the deviation from Hardy-Weinberg in the total population (F_{IT}), to the genetic divergence among subdivisions (F_{ST}) and to averaged deviation from Hardy-Weinberg within subdivisions (F_{IS}) (Yang, 1998).

$$(1-F_{IT}) = (1-F_{IS})(1-F_{ST})$$

The following F-statistics are taken from Wright (1951) and Nei (1978) as reviewed by May and Kruger (1990) and Chakraborty and Leimar (1987).

$$F_{IT} = 1 - \overline{H}_I / \overline{H}_T$$

F_{IT} values are seldom used since any type of departure from a single panmictic population will lead to a significant F_{IT} value.

$$F_{IS} = 1 - \overline{H}_I / \overline{H}_S$$

F_{IS} values help us to detect departures from Hardy-Weinberg by measuring the amount of heterozygote deficiency or excess observed in the sample.

$$F_{ST} = 1 - \overline{H}_S / \overline{H}_T$$

F_{ST} values help us to understand the degree of population differentiation within species.

$\overline{H}_S$, average expected heterozygosity within populations; $\overline{H}_I$, average observed heterozygosity within populations; $\overline{H}_T$, average expected heterozygosity in total population.

For a given locus, let P_{ix} be the frequency of i^{th} allele in population x; k the total number of alleles; s the number of populations; and P_{jk}, the frequency of j^{th} allele in population y and define the components of the above averages as follows;

$$H_T = \sum_{j=1}^{k} \sum_{i=j+1}^{k} 2\overline{P}_{ix}\overline{P}_{jx} \quad \text{Where } \overline{P}_{ix} = \frac{1}{s}\sum_{x=1}^{S} P_{ix}$$

$$H_S = \sum_{j=1}^{k} \sum_{i=1}^{k} 2P_{ix}P_{jx} \text{ or } 1 - \sum_{i=1}^{k} P_{ix}^{\,2} \text{ per locus}$$

H_I = Frequency of observed heterozygotes.

These tools are often used by population biologists because they can easily be associated with the inbreeding coefficient, which show the rate of the homozygosity in a population.

To date, there are two ways of estimating the *F*-statistics, one that computes unbiased estimators of gene diversity components (Nei, 1973) and the other with variance components (Weir and Cockerham, 1984). Chakraborty and Leimar (1987) commented that both groups of estimators give the same result for F_{IS} when sample sizes are equal, but Weir and Cockerham's F_{IS} deals with the problem of unequal sample sizes by weighing the F_{IS} obtained from each sample by its sample size.

When handling real populations, it is the Weir and Cockerham (1984) approach that should be used, because this gives non-biased estimates of the original Wright's statistics and they showed that these parameters are related to Wright's statistics as;

$$F = F_{IT},\ \theta = F_{ST},\ f = F_{IS}$$

The following approach adapted by Weir and Cockerham (1984) is analysis of variance of the frequencies for the allele *A* under consideration. The observed components of variance (*a* for between populations; *b* for between individuals within populations; and *c* for between gametes within individuals) have the following expectations:

$$a = p(1-p)\theta\ ,$$

$$b = p(1-p)(F-\theta)\,,$$

$$c = p(1-p)(1-F)\,,$$

Where p is the expected frequency of the allele and is equal to its frequency in the ancestral population. The following estimators for the three parameters are suggested:

$$1-\hat{F} = \frac{c}{a+b+c}\ ,$$

$$\hat{\theta} = \frac{a}{a+b+c}\,,$$

$$1-\hat{f} = \frac{c}{b+c}$$

These calculations are detailed in Weir and Cockerham (1984).

An estimation of F-statistics based on the analysis of variance (ANOVA) approach described by Weir and Cockerham (1984) can be calculated with the computer program GENEPOP (Raymond and Rousset, 1995) either for each pair of samples (GENEPOP, option 6, sub-option 2 and 4) or a single measure for all samples (GENEPOP, option 6, sub-option 1 and 3).

Estimates of Gene Flow

Gene flow is defined as the movement of genes within and between populations and thus it includes

all movement of gametes and individuals that are effective in changing the spatial distributions of genes (Slatkin, 1985).

Basically, gene flow is interpreted as a migration rate (m) illustrating the allele frequencies in a population each generation which are of migrant origin. It is notoriously difficult to monitor gene flow directly so it has to be inferred indirectly from the spatial distributions of genetic markers by statistical approaches (Avise, 1994). In order to measure the rate of migration (m), knowledge of the effective population size (N_e) is required; however, the absolute number of migrants into a population (N_em) is related to the level of genetic differentiation between the source and the native populations.

Analyses of gene flow of natural populations are often based on the 'island' or 'stepping stone' models (Rousset, 1997). Developed by Wright (1951), the 'island model' describes a population divided into several groups with only limited levels of gene flow among them. Using this model, the relationship between genetic differentiation (F_{ST}), local population size and migration was approximately,

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

This relationship has been widely used to estimate gene flow in various species, or more specifically, the effective number of migrants among populations (N_em values).

In 'stepping-stone' model (Kimura and Weiss, 1964) populations are organised in discrete colonies or demes with migration between them. The fundamental difference from Wright's isolation by distance model is the assumption that only migration into adjacent demes is allowed (Neigel, 1997).

An additional way of estimating gene flow is by using "private" or rare alleles (Slatkin, 1985). Slatkin (1993; 1995) showed by simulations that the two methods yield rather similar results. Additionally, the methods have been shown to give rather good estimates of the overall levels of gene flow, even if the assumptions of the island model are violated, such as in a stepping stone model of migration (Olsen *et al*., 1998).

Isolation by Distance

Pair-wise measurements of F_{ST} and genetic distance produce matrices of values. It is possible to compare these matrices against geographical distance or against each other by means of the Mantel test (Mantel, 1967). The Mantel test can be performed using the program ISOLDE of the GENEPOP package (Raymond and Rousset, 1995) with 10,000 bootstraps, in order to test for isolation by distance. The program uses randomization to create a distribution of test values and a correlation coefficient between distances.

Genetic Distance

To demonstrate levels of genetic relatedness (similarity) or genetic distance (D) between pairs of populations within group of populations, a large number of different algorithms are available. In essence the algorithms measure the similarity of allelic frequencies over all loci among populations. The commonly used measures reviewed and demonstrated by May and Kruger (1990) and Swofford *et al*. (1996) are those of Nei (1978), Rogers (1972), Cavalli-Sforza and Edwards (1967) and additionally Reynolds *et al*. (1983).

Nei's distance has been one of the most frequently used measures of genetic distance for molecular data (Nei, 1976; 1987), because D is intended to estimate the average number of gene differences, which is the most basic process of evolution, per locus. Weir (1990) also stated that Nei's distance is appropriate for long-term evolution when populations diverge because of drift and mutation. This genetic distance can be estimated by $D = -\log_e l$, where l is the normalized identity of genes (Nei, 1972; 1976). The l value is computed as reviewed in May and Kruger (1990) by

$$I = \frac{\frac{\sum\sum x_i y_i}{L}}{\left[\frac{\sum\sum x_i^2}{L} \bullet \frac{\sum\sum y_i^2}{L}\right]^{1/2}}$$

where x_i and y_i are the frequencies of the i^{th} allele at a given locus in populations x and y; L, number of loci, x_i^2, y_i^2 and x_iy_i are summed across alleles and loci.

The use of Reynolds' distance measure is dependent on the drift model, just as Nei's D is dependent on the mutation model assuming mutation drift equilibrium. For this reason, Reynolds' D is considered to be an appropriate distance for short-term evolution in the absence of mutation. This method generates a genetic distance coefficient that reflects allele frequency differences. For Reynolds' genetic distance the better approximation is

$D = -\ln(1 - \sigma)$

where, σ is the coancestry coefficient and estimated as,

$$\sigma = 1 - \left(1 - \frac{1}{2N}\right)^t$$

where N, sample size and t, is generation. Enhancement treatments of these two different measures are discussed by Nei (1987) and Reynolds *et al*. (1983).

A genetic similarity value allows an examination of which pairs of populations are the most related and which are the least related. In order to examine the population as a whole, a series of "tree" methods have been described. There are now several cases in which the study of the same or similar sets of individuals, or at least similar sets of populations, gives the same or a

very similar topological tree using different types of markers (Cavalli-Sforza, 1998). Two popular sets of tree methods tend to give consistently different results and, in some cases, even gross differences in topology. Typical representatives of the first set are maximum likelihood (ML) or the numerically convenient short cut called the Unweighted Pair Group method using Arithmetic Averages (UPGMA) and the second set is represented by Neighbor Joining (NJ) methods (Swofford *et al*., 1996).

The first set ordinarily assumes independent evolution in all branches of the tree, as would be expected for mutation, drift and certain models of variable selection. The second set assumes that the number of mutations has been the minimum possible or if the variable studied is a gene frequency, that it has followed the fastest path from the gene frequencies of one population to those in another (Cavalli-Sforza, 1998). NJ typically produces trees with segments of different length and cannot suggest the position of the tree root, unlike ML or UPGMA. The difference in the length of segments might be due to differences in drift because of different population sizes. This is an advantage of NJ, but the disadvantage is that if there are admixtures of populations that had previously diverged, branches of admixed populations become short and move towards the centre of the tree.

The best test of the validity of a tree uses a statistical method of re-sampling the genes tested, called the bootstrap (Felsenstein, 1985). It usually shows that it would take many more markers than normally employed to generate trees that are highly reproducible in all details of their topology. It can also be applied to estimating the standard errors of branch lengths. For tree reliability, allele frequencies are bootstrapped 1000 times using the SEQBOOT programme of the PHYLIP computer package (Felsenstein, 1993). The resultant bootstrapped frequencies are used to calculate 100 sets of Nei's (Nei, 1972) and Reynolds' (Reynolds *et al*., 1983) distances (using the GENDIST programme), which are then used to generate 100 UPGMA dendrograms with the NEIGHBOR programme. A consensus dendrogram is constructed using the CONSENSE programme from the different dendrograms to assess the reliability of the different nodes.

Estimation of Effective Population Sizes

Using the theory of population genetics, it is possible to derive the effective size of population from the observed heterozygosity under both the Infinite Alleles Model (IAM) and the Stepwise Mutation Model (SMM) (Jarne and Lagoda, 1996).

The IAM and the SMM both have credibility when one considers the likely mechanisms of tandem repeat number change. The models differ in that the SMM predicts that mutation occurs through the gain or loss of single repeat unit, e.g., GT. This means that some mutations will generate alleles already present in the population. Empirical evidence suggests that changes in array length due to slipped-strand mispairing are usually of one repeat unit (Shriver *et al*., 1993). In contrast, the IAM predicts that mutation gives rise to a new allele not previously found in the population, resulting in an infinite number of allelic states (O'Connell *et al*., 1997; Shriver *et al*., 1993). Unequal exchange between long tandem repeat arrays can result in a very large number of different sized alleles as assumed in the IAM.

In spite of the mechanism of repeat number of change, the IAM and SMM provide the upper and lower limit, respectively, of the extent of the variation expected. It is, therefore, fair to test hypotheses about the mutational mechanisms and population dynamics of Variable Number of Tandem Repeat (VNTR) loci by estimating the number of alleles from the observed heterozygosity for both models and comparing the predicted with the observed number of alleles. Both models usually give similar results when heterozygosity is below 0.5, but when heterozygosity is greater than 0.5 the number of alleles in the infinite allele model can be much higher than in the stepwise model for a given heterozygosity (Shriver *et al*, 1993). This means then in certain cases the two models may give different results

In the case of the infinite allele model (IAM), Ewens (1972) found the equilibrium occurrence of the number of alleles is a function of the sample size and the parameter $\varphi = 4\ N_e\mu$ where N_e is the effective size and μ the mutation rate per locus per generation. Accepting the sample size and the number of alleles, can be estimated for a given mutation rate. It can also be showed that, at equilibrium, the heterozygosity under the infinite allele model is expressed as;

$H = \frac{4N_e\mu}{(4N_e\mu + 1)}$, which leads to $N_e = \frac{\left(\frac{H}{1-H}\right)}{4\mu}$

(Crow, 1986). This formula provides estimates of N_e based on heterozygosities, not on the number of alleles.

In the stepwise mutation model, Ohta and Kimura (1973) have shown that the equilibrium heterozygosity can be written as, $H = 1 - \left(\frac{1}{(1+8N_e\mu)^{0.5}}\right)$ which gives; $N_e = \frac{\left[\left(\frac{1}{(1-H)^2}\right) - 1\right]}{8\mu}$.

Unfortunately, in this model, no analytical theory has been developed in order to use the number of alleles to estimate N_e (Shriver *et al*., 1993).

Following the two mutation models, there were two ranges of estimates of effective population size (N_e), which are directly functions of mutation rate estimates. Microsatellites show widely variable rates of instability, with reported recent studies ranging from 10^{-2}-10^{-5} per generation (Weber, 1990; Bruford and Wayne, 1993; Shriver *et al*., 1993; Schlotterer and Pemberton, 1994).

Null Alleles

Null alleles have long been known in protein electrophoresis with reduced or absent expression of a protein product (Utter *et al*., 1987; Murphy *et al*., 1996) and more recently have been observed for microsatellite loci (Callen *et al*., 1993). A null allele can be defined as any allele at a microsatellite locus that is only weakly amplified or not visible after amplification and separation (O'Connell and Wright, 1997) and is recognized, together with population subdivision, as a major factor in depression of observed heterozygosity, compared with that expected on the basis of Hardy-Weinberg equilibrium. Even though the prime reason for null alleles is thought to be deletion and insertion within the priming site of microsatellite DNA, poor DNA preparation and/or mutation within the array may also be responsible (Allen *et al*., 1995; O'Connell and Wright, 1997). For example, Callen *et al*. (1993) demonstrated one null allele was the result of an 8 bp deletion in the DNA flanking the microsatellite coincident with the priming site. Heterozygous individuals, therefore, may be being mistyped as homozygotes. If null alleles are common in the population, mistyping of heterozygous individuals might explain some of the heterozygote deficiencies observed in populations and suggests that caution should be used in comparing levels of heterozygosity among populations differing in the composition of alleles (Bruford and Wayne, 1993).

If it is assumed that null alleles are responsible for the entire heterozygote deficiency, it is a simple matter to calculate the frequency of a null allele (r) by means of an equation developed by Brookfield (1996), $r = \frac{H_{\exp} - H_{obs}}{1 + H_{\exp}}$.

Conclusions

Allelic frequency differences among populations enable the application of genetic stock identification models to determine the contribution of individual stocks. For example, allele distribution at the microsatellite loci can be explained by DNA slippage, which generates different repeat length classes, within the simple sequence. Thus, these different size modes could indicate some sort of genetic heterogeneity within the population. Furthermore such heterogeneity can produce significant departure from *HWE* (Chakraborty *et al*., 1991).

Genetic heterogenity is revealed by the multilocus F_{ST} value. In fact, genetic differences among populations will evolve in the course of time if there is little or no gene flow between them. A low level of gene flow can serve to prevent genetic differentiation. Marine fish stocks exchange between one and two orders of magnitude more migrants per generation than fresh water species (Ward *et al*., 1994). Gene flow rates of a few individuals per generation would mean that populations cannot be distinguished genetically and would appear to be panmictic, yet for fisheries management an exchange of up to 10% between populations may justify their treatment as separate stocks (Hauser and Ward, 1998).

Finally, it can be said that exploitation of marine resources may well be related to underlying genetic differences. Genetic studies can show the fundamental reproductive units of species and require fisheries management policies to take this population structure into account. The complex problem requires agreement between intergovernmental organisations to define and implement policies on the conservation of these natural resources. The management of fish stocks should be based on each population and therefore should be harvested and treated separately in research as well as management policy. Maintaining the maximum level of genetic variations in stocks is vital for the preservation of genetic resources. Therefore, excessive loss of genetic variability should be avoided.

References

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

Allendorf, F., Ryman, N. and Utter, F. 1987. Genetics and fishery management: Past, present and future. N. Ryman and F.M. Utter (Eds.), Population Genetics and Fishery Management. Washington: University of Washington: 1-20.

Altukhov, Y.P. and Salmenkova, E.A. 1987. Stock transfer relative to natural organization, management and conservation of fish populations. N. Ryman and F.M. Utter (Eds.), Population Genetics and Fishery Management. Washington: University of Washington: 333-344

Avise, J.C. 1994. Molecular Markers, Natural History and Evolution. New York: Chapman & Hall.

Baverstock, P.R. and Moritz, C. 1996. Sampling: Project Design. Hillis. D.M., C. Moritz and B. Mable (Eds.), Molecular Systematics, 2nd edition, Massachusetts: Sinauer Associates: 17-28

Brookfield, J.F.Y. 1996. A simple new method for estimating null allele frequency from heterozygote deficiency. Molecular Ecology, 5: 453-455.

Bruford, M.W. and Wayne, R.K. 1993. Microsatellites and their application to population genetic studies. Current Opinion in Genetics and Development, 3: 939-943.

Callen, D.F., Thompson, A.D., Shen, Y., Phillips, H.A., Richards, R.I., Mulley, J.C. and Sutherland, G.R. 1993. Incidence and origin of "null" alleles in the $(AC)_n$ microsatellite markers. American Journal of Human Genetics, 52: 922-927.

Carvalho, G.R. and Hauser, L. 1995. Molecular genetics and the stock concept in fisheries. G.R. Carvalho and T.J. Pitcher, (Eds.), Molecular Genetics in Fisheries. London: Chapman & Hall: 55-80.

Cavalli-Sforza, L.L. 1998. The DNA revolution in population genetics. Trends in Genetics, 14(2): 60-65.

Cavalli-Sforza, L.L. and Edwards, A.W.F. 1967. Phylogenetic analysis: models and estimation procedures. Evolution, 21: 550-570.

Chakraborty, R., Formage, M., Guequen, R. and Boerwinkle, E. 1991. Population genetics of hypervariable loci: analysis of PCR based VNTR polymorphism within a population. T. Burke, G. Dolf, A.J. Jeffreys and R. Wolff (Eds.), DNA fingerprinting: Approaches and Applications. Basel: Birkhauser Verlag: 127-143.

Chakraborty, R. and Leimar, O. 1987. Genetic variation within a subdivided population. N. Ryman and F.M. Utter (Eds.), Population Genetics and Fishery Management. Washington: University of Washington: 89-120.

Crow, J.F. 1986. Basic concepts in population, quantitive and evolutionary genetics. New York: W.H. Freeman & Company.

Crow, J.F. and Kimura, M. 1970. An Introduction to Population Genetics Theory. London: Harper & Row.

Di Rienzo, A., Peterson, A.C., Garza, J.C., Valdes, A.M., Slatkin, M. and Freimer, N.B. 1994. Mutational processes of simple-sequence repeat loci in human populations. Proceedings of the National Academy of Sciences, 91(8): 3166-3170.

Estoup, A., Presa, P., Krieg, F., Vaiman, D. and Guyomard, R. 1993. $(CT)_n$ and $(GT)_n$ microsatellites: a new class of genetic markers for *Salmo trutta* L. (brown trout). Heredity, 71: 488-496.

Ewens, W.J. 1972. The sampling theory of selectively neutral alleles. Theoretical Poplation Biology 3: 87-112.

Falconer, D.S. 1989. Introduction to Quantitative Genetics. New York: Longman Scientific and Technical.

FAO. 1981. Conservation of the Genetic Resources of Fish: Problems and Recommendations. FAO Fisheries Technical Paper No. 217. Rome: Food and Agriculture Organisation of the United Nations.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution, 39(4): 783-791.

Felsenstein, J. 1993. PHYLIP: Phlogeny Inference Package. Seattle: University of Washington.

Ferguson, A., Taggart, J. B., Prodohl, P. A., McMeel, O., Thompson, C., Stone, C., McGinnity, P. and Hynes, R.A. 1995. The application of molecular markers to the study and conservation of fish populations, with special reference to *Salmo*. Journal of Fish Biology, 47: 103-126.

Ferguson, M. 1995. The role of molecular genetic markers in the management of cultured fish. G.R. Carvalho and T.J. Pitcher (Eds.), Molecular Genetics in Fisheries. London: Chapman & Hall: 81-104.

Gall, G.A.E. 1987. Inbreeding. N. Ryman and F.M. Utter (Eds.), Population Genetics and Fishery Management. Washington: University of Washington: 47-88.

Guo, S.W. and Thompson, E.A. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics, 48: 361-372.

Hauser, L. and Ward, R.D. 1998. Population identification in pelagic fish: the limits of molecular markers. G.R. Carvalho (Ed.), Advances in Molecular Ecology. Oxford: IOS Pres: 191-224.

Hernandez, J.L. and Weir, B.S. 1989. A disequilibrium coefficient approach to Hardy-Weinberg testing. Biometrics, 45: 53-70.

Ihssen, P.E., Booke, H.E., Casselman, J.M., McGlade, J.M., Payne, N.R. and Utter, F.M. 1981. Stock identification: materials and methods. Canadian Journal of Fisheries and Aquatic Science, 38: 1838-1855.

Jarne, P. and Lagoda, P.J.L. 1996. Microsatellites, from molecules to populations and back. Trends in Ecology and Evolution, 11(10): 424-429.

Kimura, M. and Weiss, G.H. 1964. The stepping-stone model of population structure and the decrease of genetic correlation with distance. Genetics, 49: 561-576.

Koray, T. 1993. Su Ürünleri Araştırmalarında Biyometrik Yöntemler. Bornava, İzmir: Ege Üniversitesi Basımevi.

Larkin, P.A. 1981. A perspective on population genetics and salmon management. Canadian Journal of Fisheries and Aquatic Science, 38: 1469-1475.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Research, 27: 209-220.

May, B. and Krueger, C.C. 1990. Use of allozyme data for population analysis. D.H. Whitmore (Ed.), Electrophoretic and Isoelectric Focusing Techniques in Fisheries Management. Boston: CRC Pres: 157-171.

Murphy, R.W., Sites, J.W., Buth, D.G. and Haufler, C.H. 1996. Proteins: Isozyme Electrophoresis. D.M. Hillis, C. Moritz and B.K. Mable (Eds.), Molecular Systematics, 2nd edition. Massachusetts: Sinauer Associates.

Nei, M. 1972. Genetic distance between populations. The American Naturalist, 106: 283-292.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences, 70(12): 3321-3323.

Nei, M. 1976. Mathematical models of speciation and genetic distance. S. Karlin and E. Nevo (Eds.), Population Genetics and Ecology. London: Academic Pres: 723-766.

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

Nei, M. 1987. Genetic distance and molecular phylogeny. N. Ryman and F.M. Utter (Eds.), Population Genetics and Fishery Management. Washington: University of Washington: 193-224.

Neigel, J.E. 1997. A comparison of alternative strategies for estimating gene flow from genetic markers. Annual Review of Ecology and Systematics, 28: 105-128.

Nelson, K. and Soule, M. 1987. Genetical conservation of exploited fishes. N. Ryman and F.M. Utter (Eds.), Population Genetics and Fishery Management. Washington: University of Washington: 345-368.

O'Connell, M., Danzmann, R.G., Cornuet, J.M., Wright, J.M. and Ferguson, M.M. 1997. Differentiation of rainbow trout (*Oncorhynchus mykiss*) populations in Lake Ontario and the evaluation of the stepwise mutation and infinite allele mutation models using microsatellite variability. Canadian Journal of Fisheries and Aquatic Science, 54: 1391-1399.

O'Connell, M. and Wright, J.M. 1997. Microsatellite DNA in fishes. Reviews in Fish Biology and Fisheries, 7: 331-363.

Ohta, T. and Kimura, M. 1973. The model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a genetic population. Genetic Research, 22: 201-204.

Ohta, T. 1982. Linkage disequilibrium due to random genetic drift in finite subdivided populations. Proceedings of the National Academy of Sciences, 79: 1940-1944.

Olsen, J.B., Seeb, L.W., Bentzen, P. and Seeb, J.E. 1998. Genetic interpretation of broad-scale microsatellite polymorphism in odd-year pink salmon. Transactions of the American Fisheries Society, 127: 535-550.

Palumbi, S.R. 1996. Nucleic Acids II: The Polymerase Chain Reaction. D.M. Hillis, C. Moritz and B.K. Mable (Eds.), Molecular Systematics 2nd edition. Massachusetts: Sinauer Associates.

Park, L.K. and Moran, P. 1995. Developments in molecular genetic techniques in fisheries. G.R. Carvalho and T.J. Pitcher (Eds.), Molecular Genetics in Fisheries. London: Chapman & Hall.

Raymond, M. and Rousset, F. 1995. GENEPOP: population genetics software for exact tests and ecumenicism. Journal of Heredity, 86: 248-249.

Reynolds, J., Weir, B.S. and Cockerham, C.C. 1983. Estimation of the coancestry coefficient: basis for a short-term genetic distance. Genetics, 105: 767-779.

Rogers, J.S. 1972. Measures of genetic similarity and genetic distance. Studies in Genetics III (pp. 145-153). Texas: University of Texas.

Rousset, F. and Raymond, M. 1995. Testing heterozygote excess and deficiency. Genetics, 140: 1413-1419.

Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics, 145: 1219-1228.

Schlotterer, C. and Pemberton, J. 1994. The use of microsatellites for genetic analysis of natural populations. Molecular Ecology and Evolution: Approaches and Applications, 69: 203-214.

Shriver, M.D., Jin, L., Chakraborty, R. and Boerwinkle, E. 1993. VNTR allele frequency distributions under the stepwise mutation model: a computer simulation approach. Genetics, 134: 983-993.

Slatkin, M. 1985. Rare alleles as indicators of gene flow. Evolution, 39(1): 53-65.

Slatkin, M. 1993. Isolation by distance in equilibrium and non-equilibrium populations. Evolution, 47(1): 264-279.

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

Sokal, R. and Rohlf, F.J. 1981. Biometry. New York: W.H. Freeman & Company.

Swofford, D.L., Olsen, G.J., Waddell, P.J. and Hillis, D.M. 1996. Phylogenetic inference. D.M. Hillis, C. Moritz and B.K. Mable (Eds.), Molecular Systematics, 2nd edition. Massachusetts: Sinauer Associates: 407-514.

Utter, F., Aebersold, P. and Winans, G. 1987. Interpreting genetic variation detected by electrophoresis. N. Ryman and F.M. Utter (Eds.), Population Genetics and Fishery Management. Washington: University of Washington: 21-46.

Utter, F.M. 1991. Biochemical genetics and fishery management: an historical perspective. Journal of Fish Biology, supplement A, 39: 1-20.

Ward, R.D., Woodwark, M. and Skibinski, D.O.F. 1994. A comparison of genetic diversity levels in marine, freshwater and anadromous fishes. J. Fish Biol., 44: 213-232.

Ward, R.D. and Grewe, M. 1995. Appraisal of molecular genetic techniques in fisheries. G.R. Carvalho and T.J. Pitcher (Eds.), Molecular Genetics in Fisheries. London: Chapman & Hall: 29-54.

Weber, J.L. 1990. Informativeness of human $(dC\text{-}dA)_n$. $(dG\text{-}dT)_n$ polymorphisms. Genomics, 7: 524-530.

Weir, B.S. and Cockerham, C.C. 1984. Estimating F-statistics for the analysis of population structure. Evolution, 38(6): 1358-1370.

Weir, B. S. 1990. Intraspecific Differentiation. D. M. Hillis and C. Moritz (Eds.), Molecular Systematics. Massachusetts: Sinauer Associates: 373-410.

Weir, B. S. 1996. Intraspecific differentiation. D. M. Hillis, C. Moritz and B. K. Mable (Eds.), Molecular Systematics, 2nd edition. Massachusetts: Sinauer Associates: 385-406.

Wright, B. S. 1951. The genetical structure of populations. Annual Eugenics, 15: 323-354.

Wright, S. 1969. The Theory of Gene Frequencies. London: The University of Chicago Press.

Yang, R. 1998. Estimating hierarchical F-statistics. Evolution, 52(4): 950-956.