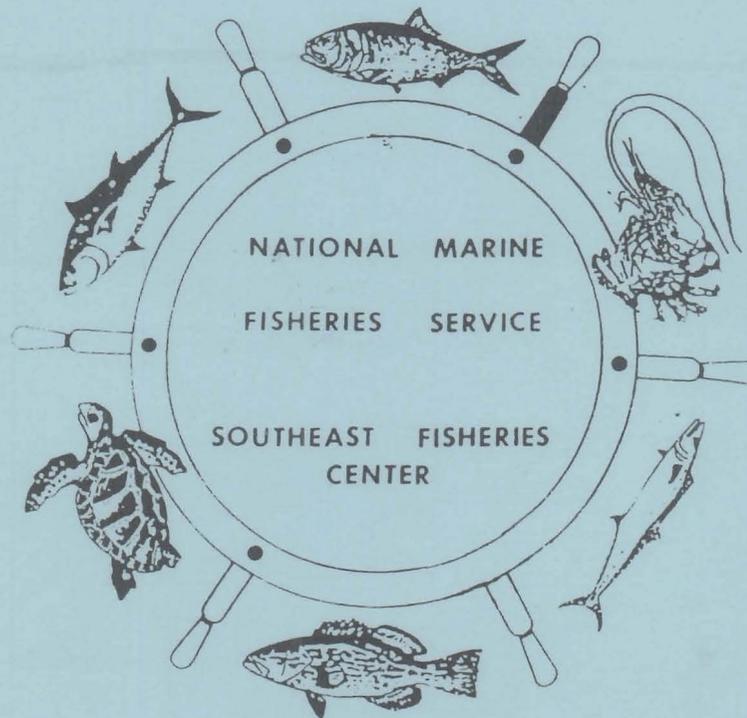


PROCEEDINGS OF THE STOCK IDENTIFICATION WORKSHOP

NOVEMBER 5-7, 1985

PANAMA CITY BEACH, FLORIDA



JULY 1987



U.S. DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration
National Marine Fisheries Service

The Stock Identification Workshop and these Proceedings were supported by the following:

American Institute of Fishery Research Biologists

Cooperative Institute of Marine and Atmospheric Sciences,
University of Miami

Georgia Sea Grant

Florida Sea Grant

Panama City Scientific Society

Panama City Laboratory, Southeast Fisheries Center

NOAA Technical Memorandum NMFS-SEFC-199

**PROCEEDINGS OF THE
STOCK IDENTIFICATION
WORKSHOP**

NOVEMBER 5-7, 1985

PANAMA CITY BEACH, FLORIDA

Editorial Committee

**Herman E. Kumpf, Editor-In-Chief
Rosalie N. Vaught
Churchill B. Grimes
Allyn G. Johnson
Eugene L. Nakamura**



U.S. DEPARTMENT OF COMMERCE

Malcolm Baldrige, Secretary

National Oceanic and Atmospheric Administration

Anthony J. Calio, Administrator

National Marine Fisheries Service

William E. Evans, Assistant Administrator for Fisheries

Technical Memorandums are used for documentation and timely communication of special-purpose information. The Plenary Session papers contained herein have received formal reviews and editing, while the abstracts of Contributed Papers have not received formal reviews.

The National Marine Fisheries Service (NMFS) does not approve, recommend or endorse any proprietary product mentioned in this publication. No reference shall be made to NMFS, or to this publication, in any advertising or sales promotion which would indicate or imply that NMFS approves, recommends or endorses any proprietary product or material mentioned herein, or which has as its purpose an intent to cause directly or indirectly the advertised product to be used or purchased because of this NMFS publication.

1987

PROTEIN ELECTROPHORESIS AND STOCK IDENTIFICATION IN FISHES

FRED M. UTTER

National Marine Fisheries Service
Northwest and Alaska Fisheries Center
2725 Montlake Boulevard East
Seattle, Washington 98112

INTRODUCTION

It is well established in fishery management that harvest regulation should be based on the population dynamics of individual breeding units (summarized in Larkin 1981). Following this principle prevents the inevitable depletion of weaker units that are harvested at an optimal rate for stronger ones. Such management requires an adequate understanding of the genetic population structure within the species that constitute the resource in question.

However, as discussed in Allendorf et al. (1987) there are many reasons why genetic units of fishes are less apparent than those of other groups of intensely managed vertebrates. For instance, the aquatic habitat makes locations and numbers of fishes difficult to identify. Indeterminate growth permits unusually large variations in size among individuals within a breeding unit. Exothermy magnifies the influences of temperature variations. Such attributes have sometimes tended either to mask genetically distinct groups (Shaklee and Tamaru 1981; Grant and Utter 1984) or to give unwarranted taxonomic distinction to apparent differences (Wishard et al. 1984; Hindar et al. 1986). These difficulties in adequately identifying distinct genetic units of fishes have led to a search for purely genetic characters that could serve as markers for genetically distinct groups.

Presently, genetic characters detected by protein electrophoresis are those that are most widely used to identify genetically distinct groups of fishes. Indeed, protein electrophoresis has emerged during the past 20 years from largely a clinical procedure to the primary method for detecting single gene (i.e., Mendelian) variation in all kinds of living organisms (Lewontin and Hubby 1966; Selander 1976; Nevo et al. 1984). A previously unknown (and undetectable) reservoir of Mendelian variation has been amply demonstrated in fishes (see de Ligny 1969, 1972 for early reviews). Genotypic and allelic data from protein coding loci obtained by electrophoretic methods have provided new and valuable insights concerning the population structure of many fish species (Allendorf and Utter 1979; Winans 1980; Shaklee 1983; Ferguson and Mason 1981; Ryman 1983). However, electrophoretic data remain

largely underused among fishery biologists and managers in spite of their potential to identify genetically distinct groups (discussed in Allendorf et al. 1987).

This paper outlines the applications of genotypic and allelic data obtained by electrophoresis in stock identification of fishes. The purpose is to give a basic understanding and appreciation of Mendelian data collected by electrophoresis to potential users having inadequate familiarity with electrophoretic procedures and data. It is intended to provide such workers with a minimal background for applying such data. An overview of underlying molecular and genetic principles is followed by stepwise descriptions of data collection and analyses, with examples from the literature. The citations given are not intended as an exhaustive review of the field, but rather as appropriate examples of the particular situation under consideration. An appendix provides details of calculations carried out in analyses of a hypothetical set of electrophoretic data.

WHY IS INFORMATION ABOUT GENOTYPES AND ALLELES IMPORTANT FOR STUDYING POPULATION STRUCTURES OF FISHES?

This workshop reflects an underlying recognition among fishery biologists of the need to identify distinct breeding units as a basis for proper management of fishery resources. As pointed out above, special attributes of fishes have often resulted in the failure to fulfill this need using criteria that tend to be reliable in other organisms. However, regardless of such attributes, the genetic structure of a species can become apparent when adequate information is available concerning the distribution of genotypes and alleles.

First, what is Mendelian inheritance and variation? Mendel's (1866) classical experiments with peas demonstrated a basis for inheritance that has subsequently proven applicable to most higher organisms including man and most fishes. An individual inherits single, or haploid, sets of genes from each parent. Consequently, each gene occurs in a paired, or diploid, state in most cells. One or the other of each of the paired genes occurs randomly in haploid germ cells (gametes) for transmission to the next generation: such transmission is Mendelian inheritance. Genes occur linearly on bodies called chromosomes. The location of a particular gene on a chromosome is called its locus (plural = loci). Different genes at the same locus are called alleles. Allelic differences are therefore Mendelian variants, and a locus is polymorphic when such variants occur (contrasted with monomorphic in the absence of allelic variation). The diploid allelic constitution of an individual is its genotype. An

individual is homozygous at a locus when both genes are the same, and heterozygous when the respective genes are different alleles.

Next, what can be determined from Mendelian data about genetic population structures that cannot be measured by quantitative genetic characters that typically involve an unknown number of loci plus a variable, and usually substantial, environmental influence? Heritable variations in traits such as size, age at maturity, and time of spawning provide useful information about stock fitnesses in specific environments, as well as guidelines for breeding programs. However, studies of such quantitative traits yield no information on distinct alleles at particular loci. Mendelian data are the fundamental building blocks of empirical population genetics. They provide distinct markers for genetic characterization and monitoring of populations, and for estimating degrees of divergence and evolutionary relationships among populations (Crow and Kimura 1970; Lewontin 1974; Nei 1975; Wright 1978; Hartl 1980; Hedrick 1983).

WHAT IS THE BASIS FOR OBTAINING INFORMATION ABOUT GENOTYPES AND ALLELES FROM ELECTROPHORETIC PATTERNS?

The Relationship of Proteins to Genes

Most genes studied electrophoretically are templates for the synthesis of proteins. The linear arrangement of four different bases (adenine, cytosine, guanine, thymine) in deoxyribonucleic acid (DNA), which is the chemical of the gene, determines the linear arrangement of different amino acids polymerized in polypeptide chains. Active proteins are made up of polypeptide chains (subunits) alone or in aggregate, depending upon the protein. The genetic code relates specific three-base sequences in the DNA to each of the 20 amino acids commonly occurring in nature, thereby assuring that individual protein molecules are direct reflections of individual genes. Further details of this relationship are outlined in Figure 1.

A change in the base sequence (i.e., a mutation) within a segment of DNA encoding a particular subunit (i.e., a locus) may result in a change of amino acids in the subunit. Some amino acids are charged negatively or positively, giving each protein molecule a characteristic net charge. A change of amino acids of different charges can result in a different net charge for the protein. Alleles are DNA segments for the same locus that differ for one or more bases.

Gel electrophoresis separates proteins on the basis of their net charges. Therefore, proteins made up of allelic subunits

having different net charges can be distinguished electrophoretically.

Electrophoresis

The basic procedures of gel electrophoresis are outlined in Figure 2. The process of electrophoresis includes a gel (commonly starch or polyacrylamide) in which introduced solutions of proteins are separated relative to their net charges by passage of a direct electrical current through the gel. Initially, mixtures of proteins are extracted with water (or buffered aqueous solvents) from tissues such as skeletal muscle, heart and liver) or are contained in body fluids such as vitreous humor or blood serum.

Most proteins that are studied by electrophoresis are enzymes, because it is easy to develop specific staining processes for many enzymes. The resolving power of electrophoresis has been substantially increased by applying histochemical staining procedures to visualize activities of specific enzymes (Hunter and Markert 1957). A number of sources give detailed descriptions of many procedures for visualizing enzymatic activities following electrophoresis (e.g., Harris and Hopkinson 1976; Siciliano and Shaw 1976). Each procedure uses a product of the enzyme's specific activity to locate that enzyme precisely in the gel. Specific staining for an enzyme's activity (Figure 2D) permits particular enzymes to be distinguished, one at a time, in the mixture of hundreds of proteins typically found in tissue extracts.

The term isozyme refers to different distinguishable molecules found in the same organism which catalyze the same reaction (Markert and Moller 1959; Shaw 1964; Brewer 1970). The term allozyme is commonly used to refer to allelic variants of the same protein.

The final result of electrophoresis is bands such as those of Figure 2D which identify the locations of various forms of a protein on a gel. The banding pattern of an individual contains information on that individual's genotype with respect to the locus (loci) coding for that particular protein.

Genotypic Interpretations of Gel Banding Patterns

The relationship between DNA base sequences, protein amino acid sequences, and the electrophoretic expression of different genotypes (i.e., the gel phenotype) is most easily illustrated for a monomeric protein. Monomeric proteins are active molecules consisting of only a single protein subunit (i.e., the polypeptide

chain). Let us assume (1) a locus coding for a monomeric protein having two alleles (i.e., a polymorphic locus) designated as A and A', (2) that these alleles produce subunits (the active protein for monomers) that are distinguishable by different electrophoretic mobilities and designated a and a' respectively, and (3) that the a' protein encoded by the A' allele moves slower than the a protein encoded by the A allele.

Three different genotypes are possible at this locus - AA, AA', and A'A'. An individual with the AA homozygous genotype only produces the faster migrating protein form. This form appears at one single location on the gel as a single band. Similarly, a homozygous individual with the A'A' genotype only produces the slower migrating band at a different location on the gel. The heterozygous (AA') genotype produces both protein forms, and an extract from such a fish will express each of these two bands on the gel. Assuming that each allele encodes the production of equal amounts of protein, each band of a heterozygous individual expresses half the production (i.e., dosage) of the single band expressed by a homozygous individual. These patterns of genotypic expression (or phenotypes) of a monomeric protein encoded by a single locus with two alleles are pictured on the top of Figure 3.

Banding patterns on a gel become more complicated when the active protein is a multimer combining two or more protein subunits into an active molecule. Let us assume the above described situation except that the active protein is dimeric, i.e., consisting of two subunits. These electrophoretic expressions are pictured in the middle section of Figure 3. The electrophoretic phenotype for the AA genotype is a single band reflecting identical molecules of a subunits combined in pairs. Similarly, the phenotype A'A' genotype is another single band reflecting paired a'a' subunits at a different location on the gel. However, the phenotype for the AA' genotype consists of three bands reflecting the random combination, in pairs, of the two electrophoretically distinguishable types of subunits. Two of the bands are homomeric combinations of aa and a'a' subunits that are expressions of the respective homozygous genotypes. The third band, not expressed by either of the homozygotes, is a heteromeric band reflecting combinations of a and a' subunits (note that monomers cannot form heteromeric bands because the single subunit is the active protein.) The sum of the intensity of the three bands expressed by heterozygous genotypes is expected to equal the intensity of single banded homozygous expressions, because the same number of subunits are produced by heterozygous or homozygous individuals.

The expected electrophoretic phenotypes for a protein having four subunits (i.e., tetrameric) reflecting a single locus that is polymorphic for two electrophoretically detectable alleles are pictured in the lower portion of Figure 3 and in Figure 4. The

respective phenotypes of homozygous individuals again are single-banded because of the identity of each of the four subunits. The five-banded (heterozygous) phenotype includes three heteromeric bands in addition to the two homomeric bands, and has a total intensity equivalent to the single band of the homozygous expression.

The expected numbers of bands and their relative intensities for individuals heterozygous for protein coding loci can also be predicted from binomial expansion of the two categories of allelic subunits (a and a'). For a dimeric protein this expansion would be

$$(\underline{a} + \underline{a}')^2 = \underline{a}^2 + 2\underline{a}\underline{a}' + \underline{a}'^2.$$

In reference to the left-hand side of the binomial formula, the a and a' represent the actual protein subunits and the exponent (2) represents the number of subunits in the protein, in this case 2 for a dimer. In the expanded right-hand side of the formula, the three terms represent the number of bands and their respective coefficients (1, 2, 1) represent their relative intensities. For a tetramer, the exponent becomes a four. Following expansion then, the relative intensities of 1:4:6:4:1 would be expected from tetramers.

More complicated electrophoretic patterns arise when subunits encoded by two or more loci aggregate to form multimeric proteins. Interpretation of such patterns involve direct extensions of the principles outlined above and are described in sources including Shaw (1964) and Utter et al. (1987).

Strengths and Limitations of Electrophoretic Data for Studying Genetic Population Structures

The principles outlined above for directly obtaining genotypic data from electrophoretic patterns are widely applied and have resulted in electrophoresis being generally recognized as "--the most useful procedure yet devised for revealing genetic variation--" (Hartl 1980). The unmatched power of electrophoresis for detecting allelic variation is enhanced by the volumes of data that can be collected with a given amount of effort. Protein extracts can be prepared with minimal effort. In starch gels (the medium that my colleagues and I have used almost exclusively), many samples can be run on a single gel, and multiple slices of a gel can be stained for different proteins which reflect different loci. For instance, a trained worker can run six gels per day with each gel containing 50 samples for a total of 300 individuals. Data for at least six loci can be obtained from each

individual because each gel can be cut into six or more slices, and each slice can be stained for a different type of protein. Usually data from more than six loci per individual can be obtained because commonly more than one locus encodes for a particular type of protein.

Starch gel is only one of many media used for electrophoresis. Other media including paper, cellulose acetate, agar and acrylamide have also been used. Ferguson (1980) provides a review of the inherent advantages and limitations of many procedures commonly used at present. In addition to starch, acrylamide has been particularly widely used as an electrophoretic medium in population studies. With regard to acrylamide electrophoresis, Ferguson (1980) states the following:

"It is very difficult to evaluate the genetic basis of intra-specific variability from general protein patterns. For most systematic work, staining for specific enzymes is preferable, and reduces the problems of homology and interpretation of variability. In staining for specific enzymes, a maximum resolution technique may not be required, e.g., although starch gel gives poorer resolution than acrylamide for general proteins, enzyme staining is in most cases superior."

Isoelectric focusing is an alternate method to electrophoresis for separating mixtures of proteins in a gel medium. A protein solution is introduced to an acrylamide gel in which a pH gradient has been established through incorporation of synthetic polyamino polycarboxylic acids having a range of isoelectric points. The proteins move in the gel until they have reached the point in the pH gradient equivalent to their own isoelectric points. Dr. Jarle Mork (Biological Station, N-7001 Trondheim, Norway) has contributed the following statement based on starch gel electrophoresis and isoelectric focusing:

"As an analytical tool for multilocus screening of a large number of specimens, starch gel electrophoresis is probably the most efficient method available today. It is generally applicable, is supported by a very broad literature on electrophoretic and staining recipes, and is unmatched in the numbers of samples that can be examined for a given effort because of the capability for obtaining multiple slices from single gel. In special applications its resolving capacity may surpass even that of isoelectric focusing in polyacrylamide gel (IFPAG); proteins with only minor differences in isoelectric points (pI) may be separated by choosing the adequate pH for the gel buffer. In practice, however, there appears to be little need for such micro-adjustments of running conditions to resolve

most isozymes. Thus, although no electrophoretic technique produces protein bands as discrete as those obtained by IFPAG, parallel analysis of tissue enzymes in Atlantic cod showed that all alleles detected by IFPAG were also detected by starch gel electrophoresis without special efforts. IFPAG may have advantages in terms of the simplicity of sample preparation and analytical setup, but appears to be a much harsher procedure than starch gel electrophoresis for many proteins. Enzymes which are readily demonstrated on starch gels may fail to show activity after IFPAG. The reasons for such deactivations are often obscure, but may be sought among some effects inherent with the IFPAG technique. These include ampholyte chelation of metalloproteins (Galante et al. 1975), isoelectric precipitation (Rhigetti and Drysdale 1976), and focusing at unfavorable pH conditions (i.e., proteins with pH in the acidic or basic portion of the gel (cf. Mork and Heggberget 1984, Mork and Haug 1983). In practice therefore, the locus repertory is more restricted in IFPAG than in starch gel electrophoresis. Thus, although IFPAG should be welcomed as a valuable complement to existing techniques, especially when dealing with small amounts of proteins (e.g., planktonic fish eggs, Mork et al. 1983, Mork and Sundnes 1983), it does not replace starch gel electrophoresis in large scale multilocus screening investigations."

There are also limitations to the information about protein coding loci that can be obtained by electrophoresis. The information needed in population genetics relates to base sequences of DNA studied either directly or indirectly. The amino acid substitutions of proteins detected by electrophoretic data are indirect reflections of the actual differences in the base sequences. All base substitutions do not necessarily result in changes of amino acids. Furthermore, all amino acid substitutions do not result in protein changes that are electrophoretically detectable. It has been estimated that only about a third of the amino acid substitutions are detected under the conditions used to collect electrophoretic data in most laboratories (Lewontin 1974). It is apparent then, that electrophoretic identity does not necessarily mean identity of base sequences in DNA. Thus, homozygosity is often a conditional concept with electrophoretic data as it is with many other classes of genetic data where alleles are inferred by phenotypes (Allendorf 1977). Although it is useful to equate electrophoretic and genetic identity, the possibility of unrevealed genetic heterogeneity must be kept in mind.

It must also be kept in mind that even an electrophoretic sample of 100 loci still represents substantially less than 1% of the total number of genes of a particular diploid organism (Nei 1975). Thus, while electrophoretically detected differences among individuals and populations are positive indicators of genetic differences, the absence of differences cannot be equated to genetic identity at the DNA level.

Most allelic differences detected by electrophoresis also appear to have a minimal effect on the fitness of the individual (Kimura 1968; Nei 1983), although exceptions are well documented (Mork et al. 1984; DiMichele and Powers 1982). This apparent neutrality of much of the genetic variation detected by electrophoresis is a disappointment to those who had envisioned electrophoretically detected alleles as "useful genes" for breeding programs assuming that many such genes could be directly related to fitness (Robertson 1972). An investigator should nevertheless continually be aware of the possibility of differential fitness of allelic proteins, but should rigorously pursue alternate explanations when genotypic distributions suggesting selection are encountered. As Ihssen et al. (1981) observed "--a null hypothesis of neutrality rather than selection appears to be the most reasonable expectation as a first approximation of reality--."

However, the general absence of phenotypic effects on fitness of most allelic proteins enhances the value of electrophoretic variation as more or less neutral genetic markers. The primary value of such markers is for inferring the distribution and magnitude of genetic variation resulting from evolutionary processes at the vast remainder of the genome that has not been sampled electrophoretically. In this capacity, electrophoresis appears likely to remain a leading procedure for an extended time because of its capability to generate readily large volumes of reliable genotype and allele frequency data.

HOW ARE SETS OF GENOTYPE AND ALLELE FREQUENCY DATA USED IN STUDIES OF STOCK IDENTIFICATION OF FISHES?

A hypothetical set of electrophoretic phenotypes is used to illustrate the actual application of genotypic and allelic data collected by electrophoresis studies for stock identification in fishes. Actual studies involving similar uses are also referenced at appropriate points throughout this presentation. The set of phenotypes shown in Figure 5 contains data regarding three protein loci for each of three populations of a single species; populations 1 and 2 belong to one subspecies and population 3 belongs to a second.

The patterns of locus 1 typify the expressions of a dimeric protein having two allele variants expressed within a single population (i.e., one heterozygous and two homozygous genotypes). Note however that there are three alleles expressed in the species (based on the total sampling of 150 individuals); alleles A and A' occur in populations 1 and 2, and alleles A' and A'' occur in population 3.

Loci 2 and 3 differ from locus 1 by having no allele variation expressed within any of the populations. However, populations 1 and 2 express the homozygous genotype of a different allele (B) than is expressed in population 3 (B'). Only a single allele (C) is expressed in each of the populations for locus 3. Note that the absence of heterozygous individuals precludes inferring the number of subunits comprising the proteins synthesized by locus 2 and locus 3.

Genetic Characterization of Samples

The first step in analyzing this set of electrophoretic data is to tabulate the individual genotypes expressed for each locus in each population (Table 1). From these genotypic frequencies, the allele frequencies can readily be determined. For instance, in population 1 for locus 1 there are 45 AA and 5 AA' genotypes in this sample of 50 individuals. Each AA individual contains two A alleles and each AA' individual one. There are, then, 100 alleles in this sample of which 95 are A and 5 are A'; and the allele frequencies are 0.95 A and 0.05 A'. The allele frequencies for the other populations and loci are calculated in the same manner. Allele frequencies are fundamentally important genetic characteristics of a particular sample.

It is useful to make further characterizations of samples before comparisons are made between them. A common statistic of genetic variability is the frequency of heterozygotes which can either be estimated directly from counting of heterozygous individuals over all loci examined, or indirectly from allele frequencies (assuming Hardy-Weinberg genotypic proportions - see below); both measures are given in Table 2. Two other estimates also are given in Table 2, the proportion of polymorphic loci and the average number of alleles per locus. The calculations for obtaining these statistics from Tables 1 and 2 are given in the Appendix. Such averages are usually made over substantially more than the three loci used in this example (e.g., 30 or more). Comparisons of these averages within a species should include the same sets of loci.

The Hardy-Weinberg Law (presented in all introductory texts of general and population genetics) is a particularly useful and broadly applied test for the expected distribution of genotypes.

This law predicts that binomial expansion of the allele frequencies of a polymorphic locus establishes the genotypic proportions of that locus under random mating. For a locus with two alleles (A and A' - not to be confused with the specifically designated alleles A and A' in Table 1) having respective frequencies of p(A) and p(A') this expansion is

$$[p(A) + p(A')]^2 = p(A)^2 + 2p(A)p(A') + p(A')^2.$$

In other words, the expected proportion of AA homozygotes in a sample is the frequency of A times itself, $p(A)^2$, etc. The Hardy-Weinberg Law can be extended to more than two alleles (as it is for locus 1 when the genotypes of all three populations are considered jointly - see Appendix).

The Hardy-Weinberg Law provides a valuable first approximation for expected genotypic proportions in samplings of individuals. Genotypes conform to the Hardy-Weinberg Law in large, random mating populations in the absence of migration and mutation, and where the alleles under consideration are not affected by natural selection. Genotype frequencies remain constant over successive generations when these conditions are met, but may deviate from expected Hardy-Weinberg proportions when they are not fulfilled. However, it should also be pointed out that considerable deviation from Hardy-Weinberg conditions may occur within a population but not be detected because of statistical insensitivities of the Hardy-Weinberg Law (Fairbairn and Roff 1980). The genotypic proportions within each population for locus 1 were chosen to conform closely to their binomial expectations; however, such conformance is commonly seen in samples from discrete breeding populations.

A properly genetically characterized sampling of individuals from a population is a valuable genetic snapshot. Data from a sample of at least 50 individuals and 20 loci should include estimates of allele frequencies, heterozygosity, alleles per locus, polymorphic loci, and how closely genotype frequencies conform to Hardy-Weinberg expectations. Characterization may also include equilibrium of between locus genotypes (Hartl 1980). Measurement of so-called "linkage equilibrium" is not included in this review. Adequately characterized samples give insight into the possible influence of migration, genetic drift (chance fluctuations in allele frequencies operating particularly in small populations) or selection on the population at the time of sampling.

In addition, adequate characterizations provide a basis for comparisons with other samplings of the same population. Such genetic monitorings have identified changes that indicate previously unsuspected genetic differences among groups that had

been presumed homogeneous. Allendorf and Ryman (1987) report many such changes in hatchery populations of freshwater and marine species including allelic differences from presumed source populations; allelic differences among year classes; reductions in heterozygosity, number of alleles and polymorphic loci; and deviations from Hardy-Weinberg proportions. These changes were generally interpreted as reflections of insufficient numbers of individuals in establishing and/or perpetuating the hatchery stocks. These findings are not reported as a general indictment of hatcheries. Rather, they point to a need to monitor genetic change in both hatchery and natural populations, and to use the resulting information as guidelines either for continuation of existing procedures or for possible corrective action.

Genetic Comparisons Within Species

The data from Figure 5 and Tables 1 and 2 are now used to look into some of the many procedures that can be used to compare allelic and genotypic data from two or more populations. A contingency test is a simple and effective means to test for differences of genotypic or allelic frequencies among samples. It is obvious from visual examination of Table 1 that population 3 differs substantially from populations 1 and 2 in genotypic and allelic distributions for loci 1 and 2. However, the differences between populations 1 and 2 for locus 1 are less apparent. A contingency table based on the total number of alleles observed in the two samples can be tested with one degree of freedom by either a chi-square or a G statistic derived from a log-likelihood ratio (Zar 1974). The values of both statistics are similar (chi-square = 15.7, G = 16.9 - see Appendix for calculations) and indicate a probability of substantially less than 0.001 that these samples were drawn from the same population. Contingency tests are widely applied as measures of heterogeneity of allele frequencies among samples (Grant et al. 1980; Ryman and Stahl 1981).

A number of measures of genetic similarity or distance have been devised to quantify the amount of genetic differences among groups. The measure that is presently the most widely applied is Nei's genetic distance (D) (Nei 1975). Under certain assumptions, D identifies the average proportion of nucleotide substitutions that have occurred since two groups diverged. Genetic distance is defined as $D = -\ln(I)$ where for a single locus

$$I = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}}$$

Value of x_i and y_i are the frequencies of specific alleles in populations x and y, respectively. Calculations of values of D involving pairwise comparisons of populations 1, 2, and 3 are

given in the Appendix. It is convenient to display such paired genetic comparisons as a matrix when three or more groups are involved. The matrix of D values for populations 1, 2, and 3 measured over all loci is

		Population		
		1	2	3
Population	1	0.000	0.012	1.050
	2		0.000	0.876
	3			0.000

A common procedure for visualizing the patterns of genetic distance or similarity (such as the parameter (I) in Nei's measure of genetic distance) among samples is to carry out a cluster analysis on the paired values of genetic distance or similarity given in the matrix. The simplest and most widely used is the unweighted pair-group method (UPGM, Sneath and Sokal 1973 - see Appendix for computation from present data set) which produces a phenogram or tree visualizing inferred genetic relationships.

Two different values of total average heterozygosity expected under Hardy-Weinberg (i.e., binomial) genotypic proportions were calculated in Table 2 (see Appendix for calculations). The mean average heterozygosity or diversity (symbolized as H(S)) of all populations (i.e., the sum of the average heterozygosity of each population divided by the number of populations) reflects the average amount of allelic variation within populations measured over all loci and populations sampled. The heterozygosity of two populations contributing to the mean average heterozygosity may be the same although different alleles are represented (e.g., populations 2 and 3 for loci 1 and 2 in Table 1). The second value (symbolized as H(T)) is called the total diversity and reflects the amount of allele variation among populations. It differs from the first, however, by being calculated at a particular locus based on the average allele frequencies over all populations sampled (for Table 1, those allele frequencies in Total column). If no differences of allele frequency occur among populations, H(S) and H(T) are the same. However, as differences in allele frequencies increase between two or more samples at one or more loci, H(S) becomes increasingly less than H(T).

This difference is an indicator of the amount of genetic subdivision existing among the populations sampled. Because H(T) is larger than H(S) when there is genetic subdivision, a deficit of heterozygous individuals is expected with admixture of individuals from different subpopulations. Such a deficit of overall heterozygous genotypes is seen in the Total column of

Table 1 where the observed genotypes represent an admixture of 150 individuals from three subpopulations.

The difference between $H(T)$ and $H(S)$ may be expressed as

$$H(T) = H(S) + D(ST)$$

where $D(ST)$ is measure of gene diversity among subpopulations (Nei 1973).

$D(ST)$ may be subdivided by different hierarchical groupings of populations (Chakraborty 1980). Such measurement of genetic differences between populations or groups of populations is called gene diversity analysis. Gene diversity analysis for the data of Table 1 is summarized in Table 3 and Figure 6 (see Appendix for calculations). First $H(S)$ and $H(T)$ are calculated. Next, that portion of $D(ST)$ resulting from averaging allele frequencies between populations within subspecies is determined. The remainder of $D(ST)$ resulting from differences between subspecies is found by subtracting $H(S)$ plus the increase due to populations within subspecies from $H(T)$.

It is convenient to express the partitioning of gene diversity in a subdivided population in relative rather than absolute terms (G values). These coefficients measure the proportion or percentage of the total gene diversity contributed by the different levels of the hierarchy. Comparable values of absolute and relative gene diversities for the data of Table 1 are given in Table 2 and Figure 6A. The coefficient for total population subdivision ($[H(T)-H(S)]/H(T)$) or $G(ST)$ has been recorded from a wide variety of organisms (Hartl 1980) and is equivalent to Wright's fixation index = $F(ST)$ (Wright 1978).

The gene diversity analysis for the data of Table 1 reflects a large amount of population heterogeneity resulting from differences between subspecies. The analysis confirms what is largely apparent from reinspection of Figure 5: only locus 1 contributes to diversity within subspecies, both loci 1 and 2 contribute to the between subspecies diversity and locus 3 proportionately reduces the absolute diversity at all levels but does not affect relative diversity. Differences between subgroups accounts for 71.3% of the total diversity (i.e., $G(ST) = 0.713$) within this hypothetical set of genetic data with by far the largest proportion resulting from differences between subspecies.

These comparative data involving three groups and loci reflect greater genetic heterogeneity than is observed in most data sets reported from conspecific populations. Usually, data are collected from 20 or more loci of which, perhaps, a third are polymorphic. The absence of shared alleles (i.e., fixed

differences) between groups, as seen for locus 2, does not usually occur at the species level, and particularly not at 50% of the polymorphic loci. Consequently, the differences between subspecies are inflated (although differences among cutthroat trout subspecies approach these levels: see Loudenslager and Gall 1980). On the other hand, the comparative data between populations 1 and 2 are more or less typical of different conspecific populations regarding levels of polymorphism and heterozygosity.

Levels of electrophoretically detected heterogeneity nevertheless vary markedly among species of fishes. Some differences appear to be generally dictated by life history variables. Gyllensten (1985) has compared gene diversities among seven marine, four anadromous, and ten freshwater species of teleosts. A higher average heterozygosity occurs in marine species than in freshwater species, and the fraction of gene diversity attributed to localities increased from marine to anadromous to freshwater species. These observations are consistent with the greater opportunity for gene flow in the marine environment. Local stock discrimination based on genetic isolation may therefore be expected to be less distinct in marine species. However, genetic heterogeneity among populations is not uncommon in marine species when sampling has involved broad geographic areas (e.g., Pacific herring, Grant and Utter 1984; yellowfin sole, Grant et al. 1983; Pacific cod, Grant et al. 1982; Atlantic cod, Mork et al. 1985; milkfish, Winans 1980; Pacific pollock, Iwata 1975; Pacific hake, Utter and Hodgins 1971). Such heterogeneity typically separates different major groups whose existences were often unsuspected prior to genetic investigations. Similar major heterogeneity within marine species undoubtedly remains to be identified on the basis of distinctive allele frequencies.

Two groups having identical allele frequencies at all detectable loci can be made distinguishable provided at least one of the groups is artificially cultured, because allele frequencies of cultured populations can be readily changed (i.e., genetically marked) when parental genotypes are known at polymorphic loci. This process has been effectively used both in short term experimental studies (Reisenbichler and McIntyre 1977; Schroder 1982) as well as in more extensive and long term investigations involving entire segments of established populations (Seeb et al. 1986; Gharrett 1985). The ease and permanence of genetic marking provides a valuable tool for measuring relative reproductive successes and rates of migration between marked and unmarked populations over multiple generations, as well as for identifying origins of individuals or stocks. Guidelines for establishing marked populations include minimizing inbreeding and involving alleles most likely to be neutral to natural selection (Allendorf and Utter 1979; Gharrett 1985).

Knowing the pattern of statistically significant differences in allele frequencies among stocks can be used to estimate compositions of population mixtures. Such use is increasing in the management of Pacific salmon which are usually harvested from complex population mixtures (Grant et al. 1980; Miller et al. 1983; Fournier et al. 1984; Beacham et al. 1985; Milner et al. 1985). Procedures for obtaining estimates (e.g., by maximum likelihood) require (1) that detectable differences in allele frequencies exist among populations potentially contributing to a particular mixture, (2) that sets of allelic data are available for representative groups of such populations, and (3) that sufficient sampling of individuals has been obtained from the mixture for adequate precision of estimates. When these conditions are met, the results to date have provided more detailed estimates than had previously been obtainable within reasonable time intervals, and at comparable or lower costs to other procedures presently used to estimate mixed stock compositions.

Genetic Comparisons Among Species

The primary focus of this paper is on differences within species. However, concerns for stock identification extend between species when individuals of different species cannot be readily distinguished. Electrophoretic data are particularly useful in such circumstances.

Interspecific allelic variations differ from intraspecific variation through fixed allelic differences commonly occurring at one or more loci. Such fixed differences between species are preserved by the absence of gene flow. The proportion of fixed differences generally increases as comparisons among taxonomic levels become higher. A review of many published values of genetic distances of fish at different taxonomic levels (Shaklee et al. 1982) reported average D values at the level of population, species and genus to be 0.05, 0.30 and 0.90, respectively. The relationship between genetic distance and taxonomic level has made allelic data a particularly valuable tool in systematic studies (Awise 1974; Buth 1984).

These qualitative genetic distinctions occurring between species usually preclude the requirement for detailed statistical analyses of allelic data to identify the presence of different species in a sample of a reasonable number of individuals. The presence of only homozygous individuals for different alleles at one or more loci is usually a clear indication of more than one species. This characteristic has proven useful in detecting previously unrecognized sibling species (Shaklee and Tamaru 1981) and in identifying individuals (e.g., eggs, juveniles) when species origins are uncertain (Allendorf and Utter 1979; Mork et

al. 1985). Because of qualitative species differences, we have responded to numerous requests from enforcement personnel for species identifications from fragments of muscle to resolve forensic cases. Shaklee (1983) gives a useful overview of applying qualitative species differences in fishery problems.

Fixed allelic differences between species also usually permit immediate recognition of species hybrids. First generation hybrids are characterized by heterozygous expression of alleles that are fixed in the respective parental species. Consequently, an individual heterozygous at all fixed loci that distinguish two species is readily identifiable as a first generation hybrid between these species. Identification of hybrids is less clear when subsequent hybrid generations or backcrossings are involved because distinguishing alleles may be homozygous. In such cases, as well as in instances where some common alleles are shared by the parent species, a hybrid index may be used to estimate the probability of an individual's hybrid origin (Campton and Utter 1985; Campton 1987).

CONCLUDING REMARKS

The presented material has established (1) that sets of Mendelian data are necessary for properly identifying and measuring genetic population units, (2) that such data were not generally available prior to the development of electrophoretic methods, and (3) that electrophoresis continues to be the primary procedure used for obtaining Mendelian data. It is understandable that new insights into genetic structures of fish species resulted from electrophoretic studies as mentioned in the Introduction. Indeed, such insights tend to be the rule rather than the exception when adequate sets of electrophoretic data are collected on a formerly unstudied species, or group of populations within a species. This point is apparent from a listing of some modified assumptions resulting from electrophoretic studies (Table 4, taken from Allendorf et al. 1987).

The insights gained from such data coupled with their relative ease of collection clearly indicate that any investigation concerned with understanding genetic structures within species as well as among closely related species should include collection of adequate sets of electrophoretic data. These data will not universally provide the biological and management insights that are sought or needed. Biological reality often fails to conform to political boundaries. Limited gene flow among partially reproductively isolated population units requiring separate management may prevent genetic divergence from being detected electrophoretically (although, as mentioned above, this limitation need not apply to cultured populations). Nevertheless, such data are a necessary starting point to determine what additional procedures may be needed to define adequately the

population units in question, and often will provide sufficient information to preclude the use of other procedures.

LITERATURE CITED

ALLENDORF, F.W.

1977. Electromorphs or alleles? *Genetics* 87:821-822

ALLENDORF, F.W., and N. RYMAN.

1987. Genetic management of hatchery stocks. In N. Ryman and F.M. Utter (editors), *Population genetics and its application to fisheries management*, p. 141-159. Univ. Wash. Press, Seattle.

ALLENDORF, F.W., N. RYMAN, AND F.M. UTTER.

1987. Genetics and fishery management: past, present, and future. In *Population genetics and its application to fisheries management*, p. 1-19. Univ. Wash. Press, Seattle.

ALLENDORF, F.W., and F.M. UTTER.

1979. Population genetics. In W.S. Hoar, D.J. Randall and J.R. Brett (editors), *Fish physiology*, Vol. 8, p. 407-454. Academic Press, N.Y.

AVISE, J.C.

1974. Systematic value of electrophoretic data. *Syst. Zool.* 23:465-481.

AVISE, J.C., E. BERMINGHAM, L.G. KESSLER, and N.C. SAUNDERS.

1984. Characterization of mitochondrial DNA variability in a hybrid swarm between subspecies of bluegill sunfish (*Lepomis macrochirus*). *Evolution* 38:931-941.

AVISE, J.C., J.J. SMITH, and F.J. AYALA.

1975. Adaptive differentiation with little genic change between two native California minnows. *Evolution* 29:411-426.

BEACHAM, T.D., R.E. WITHLER, and A.P. GOULD.

1985. Biochemical genetic stock identification of chum salmon (*Oncorhynchus keta*) in southern British Columbia. *Can. J. Fish. Aquat. Sci.* 42:437-448.

BERMINGHAM, E., and J.C. AVISE.

1984. Genetics and zoogeography of freshwater fishes in the southeastern United States: restriction analysis of mitochondrial DNA. *Genetics* 107:s10-s11.

BREWER, G.J.

1970. An introduction to isozyme techniques. Academic Press, N.Y., 186 p.

- BUSACK, C.A., and G.A.E. GALL.
1981. Introgressive hybridization in populations of Paiute cutthroat trout (Salmo clarki seleniris). Can. J. Fish. Aquat. Sci. 38:939-951.
- BUTH, D.G.
1984. The application of electrophoretic data in systematic studies. Annu. Rev. Ecol. Syst. 15:501-522.
- CAMPTON, D.E.
1987. Natural hybridization and introgression in fishes: methods of detection and genetic interpretations. In N. Ryman and F.M. Utter (editors), Population genetics and its application to fishery management, p. 193-223. Univ. Wash. Press, Seattle.
- CAMPTON, D.E., and F.M. UTTER.
1985. Natural hybridization between steelhead trout (Salmo gairdneri) and coastal cutthroat trout (Salmo clarki clarki) in two Puget Sound Streams. Can. J. Fish. Aquat. Sci. 42:110-119.
- CAMPTON, D.E., and F.M. UTTER.
In press. Genetic structure of anadromous cutthroat trout (Salmo clarki clarki) populations in two Puget Sound regions: evidence for restricted gene flow. Can. J. Fish. Aquat. Sci.
- CHAKRABORTY, R.
1980. Gene-diversity analysis in nested subdivided populations. Genetics 96:721-726.
- CHAKRABORTY, R., and O. LEIMAR
1987. Genetic variation within a subdivided population. In N. Ryman and F.M. Utter (editors), Population genetics and its application to fishery management, p. 89-120. Univ. Wash. Press, Seattle.
- CROW, J.F., and M. KIMURA.
1970. An introduction to population genetics theory. Harper & Row, Inc., N.Y., 589 p.
- DiMICHELE, L., and D.A. POWERS.
1982. LDH-B genotype-specific hatching times of Fundulus heteroclitus embryos. Nature 296:563-564.
- FAIRBAIRN, D.J., and D.A. ROFF.
1980. Testing genetic models of isozyme variability without breeding data: can we depend on the chi-square? Can. J. Fish. Aquat. Sci. 37:1149-1159.

- FERGUSON, A.
1980. Biochemical systematics and evolution. Blackie and Son, Ltd., Glasgow. 194 p.
- FERGUSON, A., and F.M. MASON.
1981. Allozyme evidence for reproductively isolated sympatric populations of brown trout Salmo trutta L. in Lough Melvin, Ireland. J. Fish. Biol. 18:629-642.
- FOURNIER, D.A., T.D. BEACHAM, B.E. RIDDELL, and C.A. BUSACK.
1984. Estimating stock composition in mixed stock fisheries using morphometric, meristic, and electrophoretic characteristics. Can. J. Fish. Aquat. Sci. 41:400-408.
- GALANTE, E., T. CARAVAGGION, and P.G. RIGHETTI.
1975. In P.G. Righetti (editor), Progress in isoelectric focusing and isotachopheresis, p. 3-12. ASP Biol. Med. Press B.V., Amsterdam.
- GHARRETT, A.J.
1985. Genetic interaction of Auke Creek hatchery pink salmon with natural spawning stocks in Auke Creek. Alaska Sea Grant Rep. 85-9, Univ. Alaska, Juneau, 40 p.
- GHARRETT, A.J., and F.M. UTTER.
1982. Scientists detect genetic differences. Sea Grant Today 12(2):3-4.
- GRANT, W.S., R. BAKKALA, F.M. UTTER, D.J. TEEL, and T. KOBAYASHI.
1983. Biochemical genetic population structure of yellowfin sole, Limanda aspera, of the North Pacific Ocean and Bering Sea. Fish. Bull., U.S., 81:667-677.
- GRANT, W.S., G.B. MILNER, P. KRASNOWSKI, and F.M. UTTER.
1980. Use of biochemical genetic variants for identification of sockeye salmon (Oncorhynchus nerka) stocks in Cook Inlet, Alaska. Can. J. Fish. Aquat. Sci. 37:1236-1247.
- GRANT, W.S., and F.M. UTTER.
1984. Biochemical population genetics of Pacific herring (Clupea pallasii). Can. J. Fish. Aquat. Sci. 41:856-864.
- GRANT, W.S., C.I. ZHANG, and T. KOBAYASHI.
1982. Population structure of Pacific cod (Gadus macrocephalus). Section II, Report to RACE Division, Northwest and Alaska Fisheries Center, NMFS, Seattle, WA.

GYLLENSTEN, U.

1985. The genetic structure of fish: differences in the intra-specific distribution of biochemical genetic variation between marine, anadromous, and freshwater species. *J. Fish Biol.* 26:691-699.

HARRIS, H., and D.A. HOPKINSON.

1976. Handbook of enzyme electrophoresis in human genetics. American Elsevier, N.Y., 306 p.

HARTL, D.L.

1980. Principles of population genetics. Sinauer Assoc., Sunderland, MA, 488 p.

HEDRICK, P.W.

1983. Genetics of populations. Science Books Int., Boston, MA, 629 p.

HINDAR, K., N. RYMAN, and G. STÅHL.

1986. Genetic differentiation among local populations and morphotypes of Arctic char, Salvelinus alpinus. *Biol. J. Linn. Soc.* 27:269-285.

HUNTER, R.L., and C.L. MARKERT.

1957. Histochemical demonstration of enzymes separated by zone electrophoresis in starch gels. *Science* 125:1294-1295.

IHSSEN, P.E., H.E. BOOKE, J.M. CASSELMAN, J.M. McGLADE, N.R. PAYNE, and F.M. UTTER.

1981. Stock identification: materials and methods. *Can. J. Fish. Aquat. Sci.* 38:1838-1855.

IWATA, M.

1975. Genetic identification of walleye pollock (Theragra chalcogramma) populations on the basis of tetrazolium oxidase polymorphism. *Comp. Biochem. Physiol.* 50B:197-201.

KIMURA, M.

1968. Evolutionary rate at the molecular level. *Nature* 217:624-626.

KORNFIELD, I., K.F. BELAND, J.R. MORING, and F.W. KIRCHEIS.

1981. Genetic similarity among endemic Arctic char (Salvelinus alpinus) and implications for their management. *Can. J. Fish. Aquat. Sci.* 38:32-39.

KORNFIELD, I., D.C. SMITH, P.S. GAGNON, and J.N. TAYLOR.

1982. The cichlid fish of Cuatro Ciénegas, Mexico: direct evidence of conspecificity among distinct trophic morphs. *Evolution* 36:658-664.

- LARKIN, P.A.
1981. A perspective on population genetics and salmon management. *Can. J. Fish. Aquat. Sci.* 38:1469-1475.
- LEWONTIN, R.C.
1974. The genetic basis of evolutionary change. Columbia Univ. Press, N.Y. 346 p.
- LEWONTIN, R.C., and J.L. HUBBY.
1966. A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of Drosophila pseudoobscura. *Genetics* 54:595-609.
- LIGNY, W. de.
1969. Serological and biochemical studies on fish populations. *Oceanogr. Mar. Biol. Annu. Rev.* 7:411-513.
- LIGNY, W. de.
1972. Blood groups and biochemical polymorphisms in fish. In G. Kovacs and M. Papp (editors), 12th Europ. Conf. Anim. Blood Groups Biochem. Polymorph., p. 55-56. W. Junk, The Hague.
- LOUDENSLAGER, E.J., and G.A.E. GALL.
1980. Geographic patterns of protein variation and subspeciation in cutthroat trout, Salmo clarki. *Syst. Zool.* 29:27-42.
- LOUDENSLAGER, E.J., and R.M. KITCHIN.
1979. Genetic similarity of two forms of cutthroat trout, Salmo clarki, in Wyoming. *Copeia* 1979:673-678.
- MENDEL, G.
1866. Versuche uber Pflanzenhybriden. *Verh. Naturforsch. Ver. Brunn* 4:3-47.
- MARKERT, C.L., and F. MOLLER.
1959. Multiple forms of enzymes: tissue, ontogenetic, and species specific patterns. *Proc. Natl. Acad. Sci. U.S.* 45:753-763.
- MILLER, M., P. PATTILLO, G. MILNER, and D. TEEL.
1983. Analysis of chinook stock composition in the May 1982 troll fishery off the Washington coast: an application of the genetic stock identification method. Wash. Dep. Fish. Tech. Rep. No. 74, 27 p.
- MILNER, G.B., D.J. TEEL, F.M. UTTER, and G.A. WINANS.
1985. A genetic method of stock identification in mixed populations of Pacific salmon, Oncorhynchus spp. *Mar. Fish. Rev.* 47:1-8.

- MORK, J., R. GISKEODEGARD, and G. SUNDNES.
 1984. The haemoglobin polymorphism in Atlantic cod (Gadus morhua L.): genotypic differences in somatic growth and in maturing age in natural populations. Flodevigen Rapportser 1:721-732.
- MORK, J., and T. HAUG.
 1983. Genetic variation in halibut Hippoglossus hippoglossus (L.) from Norwegian waters. Hereditas 98:167-174.
- MORK, J., and T. HEGGBERGET.
 1984. Eggs of Atlantic salmon (Salmo salar L.) and trout (S. trutta L.): identification by phosphoglucoisomerase zymograms. Fish. Manage. 15(2):59-65.
- MORK, J., N. RYMAN, G. STÅHL, F. UTTER, and G. SUNDNES.
 1985. Genetic variation in Atlantic cod (Gadus morhus) throughout its range. Can. J. Fish. Aquat. Sci. 42:1580-1587.
- MORK, J., P. SOLEMDAL, and G. SUNDNES.
 1983. Identification of marine fish eggs; a biochemical genetics approach. Can. J. Fish. Aquat. Sci. 40:361-369.
- MORK, J., and G. SUNDNES.
 1983. Population genetic studies in fish may start at the egg stage: examples from gadoid species in Norwegian waters. Sarsia 68:171-175.
- NEI, M.
 1973. Analysis of gene diversity in subdivided populations. Proc Natl. Acad. Sci. USA. 70:3321-3323.
- NEI, M.
 1975. Molecular population genetics and evolution. North-Holland Publ. Co., Amsterdam, 288 p.
- NEI, M.
 1983. Genetic polymorphism and the role of mutation in evolution. In M. Nei and R. Koehn (editors), Evolution of genes and proteins, p. 165-190. Sinauer Assoc., Sunderland, MA.
- NEVO, E., A. BEILES, and R. BEN-SHLOMO.
 1984. The evolutionary significance of genetic diversity: ecological, demographic and life history correlates. In G. S. Mani (editor), Evolutionary dynamics of genetic diversity, p. 13-213. Lect. Notes Biomath., 53 p. Springer-Verlag, Berlin.

- PHELPS, S.R., and F.W. ALLENDORF.
1983. Genetic identity of pallid and shovelnose sturgeon (Scaphirynchus albus and S. platyrhynchus). *Copeia* 1983:696-700.
- REISENBICHLER, R.R., and J.D. MCINTYRE.
1977. Genetic differences in growth and survival of juvenile hatchery and wild steelhead trout, Salmo gairdneri. *J. Fish Res. Board Can.* 34:123-128.
- RIGHETTI, P.G., and J.W. DRYSDALE.
1976. Isoelectric focusing. In T.S. Work and E. Work (editors), *Laboratory techniques in biochemistry and molecular biology*, p. 341-590. North-Holland Publ. Co., Amsterdam.
- ROBERTSON, F.W.
1972. Value and limitations of research in protein polymorphism. In G. Kovacs and M. Papp (editors), 12th Eur. Conf. on Anim. Blood Groups and Biochem. Polymorph., p. 41-54. W. Junk, The Hague.
- RYMAN, N.
1983. Patterns of distribution of biochemical genetic variation in salmonids: differences between species. *Aquaculture* 33:1-21.
- RYMAN, N., F.W. ALLENDORF, and G. STÅHL.
1979. Reproductive isolation with little genetic divergence in sympatric populations of brown trout (Salmo trutta). *Genetics* 92:247-262.
- RYMAN, N., U. LAGERCRANTZ, L. ANDERSSON, R. CHAKRABORTY, and R. ROSENBERG.
1984. Lack of correspondence between genetic and morphologic variability patterns in Atlantic herring (Clupea harengus). *Heredity* 53:687-704.
- RYMAN, N., and G. STÅHL.
1981. Genetic perspectives of the identification and conservation of Scandinavian stocks of fish. *Can. J. Fish. Aquat. Sci.* 38:1562-1575.
- SAGE, R.D., and R.K. SELANDER.
1975. Trophic radiation through polymorphism in cichlid fishes. *Proc. Natl. Acad. Sci. USA* 72:4669-4673.
- SCHRODER, S.
1982. The influence of intrasexual competition on the distribution of chum salmon in an experimental stream. In E.L. Brannon and E.O. Salo (editors), *Salmon and trout migratory behavior symposium*, p. 275-285. Univ. Wash. Coll. Fish.

- SEEB, L.W.
1986. Biochemical systematics and evolution of the scorpaenid genus Sebastes. Ph.D. Thesis, Univ. Wash., 176 p.
- SEEB, J.E., L.W. SEEB, and F.M. UTTER.
1986. Use of genetic marks to assess stock dynamics and management programs for chum salmon. Trans. Am. Fish. Soc. 115:448-454.
- SELANDER, R.K.
1976. Genic variation in natural populations. In F. Ayala (editor), Molecular evolution, p. 21-45. Sinaur Assoc., Sunderland, MA.
- SHAKLEE, J.B.
1983. The utilization of isozymes as gene markers in fisheries management and conservation. In M.C. Rattazzi, J.G. Scandalios, and G.S. Whitt (editors), Isozymes: current topics in biological and medical research, Vol. 11, p. 213-247. Alan R. Liss, Inc., N.Y.
- SHAKLEE, J.B., and C.S. TAMARU.
1981. Biochemical and morphological evolution of Hawaiian bonefishes (Albula). Syst. Zool. 30:125-146.
- SHAKLEE, J.B., C.S. TAMARU, and R.S. WAPLES.
1982. Speciation and evolution of marine fishes studied by the electrophoretic analysis of proteins. Pac. Sci. 36:141-157.
- SHAW, C.R.
1964. The use of genetic variation in the analysis of isozyme structure. In Subunit structure of proteins, biochemical and genetic aspects, p. 117-130. Brookhaven Symp. Biol. No. 17. Brookhaven National Lab., N.Y.
- SICILIANO, M.J., and C.R. SHAW.
1976. Separation and visualization of enzymes on gels. In I. Smith (editor), Chromatographic and electrophoretic techniques, Vol. 2, p. 185-209. Heinemann, London.
- SNEATH, P.H.A., and R.R. SOKAL.
1973. Numerical taxonomy. W.H. Freeman, San Francisco, CA, 573 p.
- STÄHL, G.
1983. Differences in the amount and distribution of genetic variation between natural populations and hatchery stocks of Atlantic salmon. Aquaculture 33:23-32.

- TURNER, B.J.
1974. Genetic divergence of Death Valley pupfish species: biochemical versus morphological evidence. *Evolution* 28:281-294.
- TURNER, B.J., and D.J. GROSSE.
1980. Trophic differentiation in Illyodon, a genus of stream-dwelling goodeid fishes: speciation versus ecological polymorphism. *Evolution* 34:259-270.
- UTTER, F.M., P. AEBERSOLD, and G. WINANS.
1987. Interpreting genetic variation detected by electrophoresis. In N. Ryman and F. Utter (editors), *Population genetics and its application to fisheries management*, p. 21-45. Univ. Wash. Press, Seattle.
- UTTER, F.M., and H.O. HODGINS.
1971. Biochemical polymorphisms in the Pacific hake (Merluccius productus). *Rapp. P.-v. Réun. Cons. Int. Explor. Mer* 161:87-89.
- UTTER, F., G. MILNER, D. TEEL, and G. STAHL.
In prep. Genetic population structure of North American stocks of chinook salmon.
- WESTRHEIM, S.J., and H. TSUYUKI.
1967. Sebastes reedi, a new scorpaenid fish in the northeast Pacific Ocean. *J. Fish. Res. Board Can.* 24:1945-1954.
- WINANS, G.A.
1980. Geographic variation in the milkfish Chanos chanos. I. Biochemical evidence. *Evolution* 34:558-574.
- WISHARD, L.N., J.E. SEEB, F.M. UTTER, and D. STEFAN.
1984. A genetic investigation of suspected redband trout populations. *Copeia* 1984:120-132.
- WRIGHT, S.
1978. Evolution and the genetics of populations. Vol. IV: Variability within and among natural populations. Univ. Chicago Press, Chicago, IL.
- ZAR, J.H.
1974. Biostatistical analysis. Prentice-Hall, Inc., Englewood Cliffs, NJ, 620 p.

Table 1. Genotypes, allele frequencies and heterozygosities from electrophoretic patterns of Figure 5. Parenthetical figures are expected proportions under binomial expansion of allele frequencies (Hardy-Weinberg equilibrium).

		POPULATION			
		1	2	3	TOTAL
LOCUS 1	<u>Genotype</u>				
	AA	45 (45.1)	28 (28.1)	0	73 (48.2)
	AA'	5 (4.8)	19 (18.8)	0	24 (59.5)
	A'A'	0 (0.1)	3 (3.1)	28 (28.1)	31 (18.4)
	AA''	0	0	0	0 (14.1)
	A'A''	0	0	19 (18.8)	19 (8.7)
	A''A''	0	0	3 (3.1)	3 (1.0)
	<u>Allele frequency</u>				
	A	0.95	0.75	0	0.567
	A'	0.05	0.25	0.75	0.350
	A''	0	0	0.25	0.083
	<u>Heterozygosity</u>	0.10 (0.09)	0.38 (0.375)	0.38 (0.375)	
LOCUS 2	<u>Genotype</u>				
	BB	50	50	0	100 (66.7)
	BB'	0	0	0	0 (66.6)
	B'B'	0	0	50	50 (16.7)
	<u>Allele frequency</u>				
	B	1.0	1.0	0	0.667
B'	0	0	1.0	0.333	
<u>Heterozygosity</u>	0 (0)	0 (0)	0 (0)		
LOCUS 3	<u>Genotype</u>				
	CC	50	50	50	150
	<u>Allele frequency</u>				
	C	1.0	1.0	1.0	1.0
<u>Heterozygosity</u>	0 (0)	0 (0)	0 (0)		

Table 2. Average alleles per locus, proportion of loci polymorphic and average heterozygosity per locus from data of Table 1. Parenthetical heterozygosities are those expected under Hardy-Weinberg equilibrium.

	POPULATION			
	1	2	3	TOTAL
Average no. of alleles per locus	1.3	1.3	1.3	2.0
Proportion of loci polymorphic	0.333	0.333	0.333	0.667
Average heterozygosity per locus	0.033 (0.032)	0.127 (0.125)	0.127 (0.125)	
	Mean heterozygosity (H(S))		0.096	(0.095)
	Total diversity (H(T))		(0.331)	

Table 3. Components of gene diversity from allele frequencies of Table 1.

Locus	ABSOLUTE GENE DIVERSITY			
	Total	Within populations	Between populations (within subspecies)	Between subspecies
1	0.550	0.282	0.013	0.255
2	0.444	0	0	0.444
3	0	0	0	0
Average	0.331	0.095	0.004	0.232

RELATIVE GENE DIVERSITY				
Average	1.000	0.287	0.012	0.701

Table 4. Examples of biochemical genetic studies modifying previous assumptions of the genetic structure of fish species (from Allendorf et al. 1987).

	RELATIONSHIP INDICATED BY BIOCHEMICAL GENETIC DATA	REFERENCE
A. <u>Identification of previously unrecognized systematic groups at the:</u>		
Intra-specific level	Major groups of rainbow trout corresponding to geographic region (coastal-inland) rather than drainage or life history pattern.	Allendorf & Utter 1979
	Major population units of Pacific herring on each side of the Alaska peninsula	Grant & Utter 1984
	Sharp discontinuity of populations east and west of the Apalachicola River in the southeastern U.S.A. of several freshwater species.	Birmingham & Avise 1984; Avise et al. 1984
	Reproductively isolated sympatric populations of brown trout.	Ryman et al. 1979; Ferguson & Mason 1981
Inter-specific level	Identification of previously unrecognized species of rockfish	Westrheim & Tsuyuki 1967; Seeb 1986
	Identification of previously unrecognized species of bonefish.	Shaklee & Tamaru 1981
B. <u>Inconsistencies with previous assumptions of genetic divergence based on:</u>		
Residency vs. anadromy	Conspecificity of anadromous and landlocked forms of char of eastern North America.	Kornfield et al. 1981

Table 4. Continued

	Lack of genetic divergence between anadromous and resident populations in rainbow trout, Atlantic salmon, and brown trout (less than 0.2% and 0.5% of the total gene diversity in Atlantic salmon and brown trout, respectively).	Allendorf & Utter 1979; Ryman 1983; Ståhl 1983
Time of spawning	No apparent genetic divergence between fall and spring spawning Atlantic herring.	Ryman et al. 1984
	Major groups of chinook salmon corresponding to geographic region rather than time of spawning.	Utter et al. in prep.
Morphology	Little genetic divergence among morphologically distinct forms of cutthroat trout.	Busack & Gall 1981; Loudenslager & Kitchin 1979
	Little genetic divergence among morphologically distinct species of pupfish.	Turner 1974
	Little genetic differentiation between minnow species from two genera.	Avise et al. 1975
	Conspecificity (and local random breeding) of distinct morphological types of <u>Illyodon</u> previously considered separate species.	Turner and Grosse 1980
	Lack of apparent genetic divergence between arid adapted (redband) and anadromous (steelhead) populations of rainbow trout.	Wishard et al. 1984
	Lack of genetic divergence between two sturgeon species with overlapping geographic distribution.	Phelps & Allendorf 1983
	Conspecificity of sympatric but trophically specialized forms of Mexican cichlids.	Kornfield et al. 1982; Sage & Selander 1975

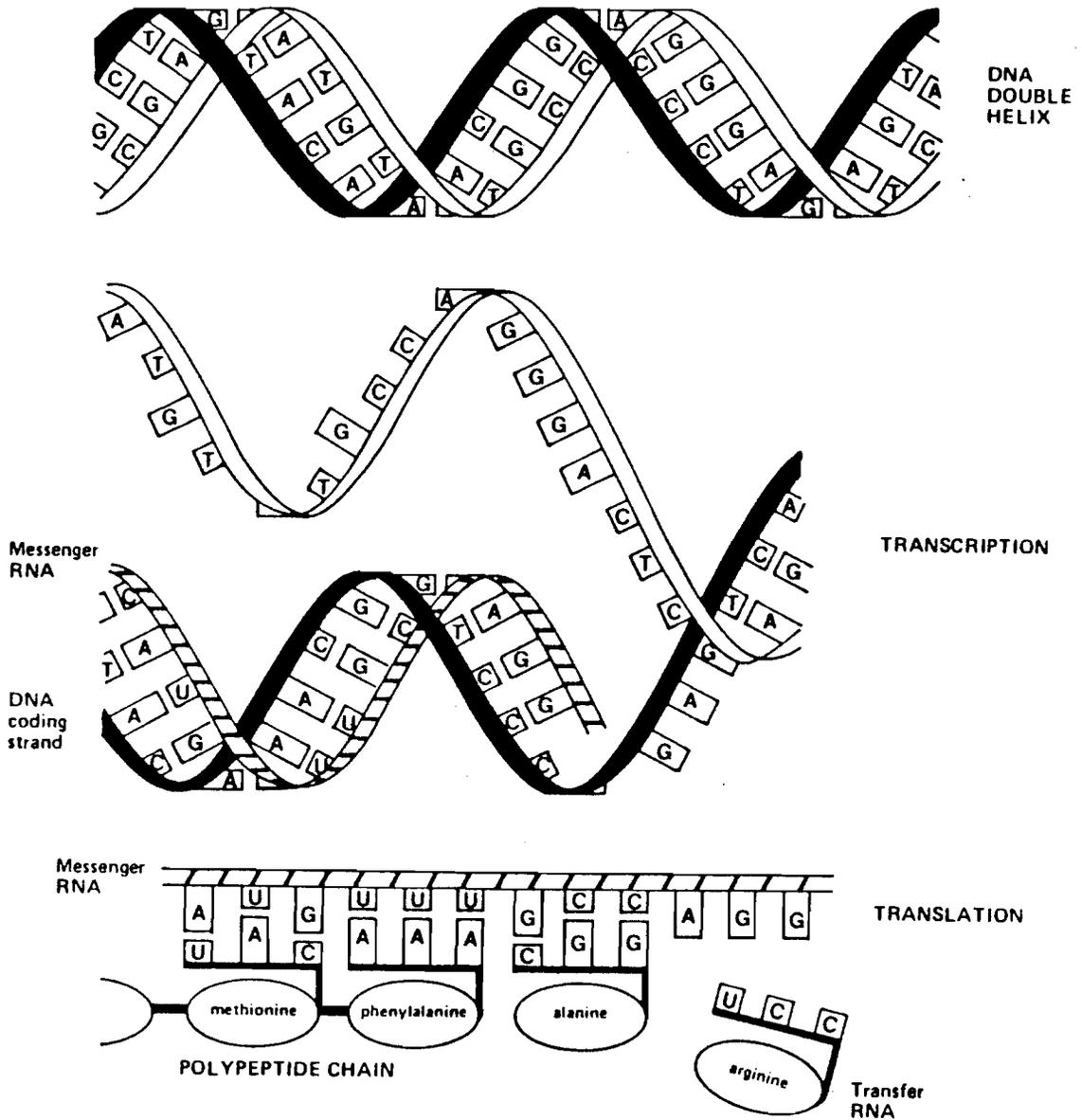


Figure 1. An outline of molecular processes relating base sequences of DNA to amino acid sequences of polypeptide chains (proteins). Messenger RNA is synthesized during transcription, and provides a template for the synthesis of the polypeptide. From Utter et al. (1987).

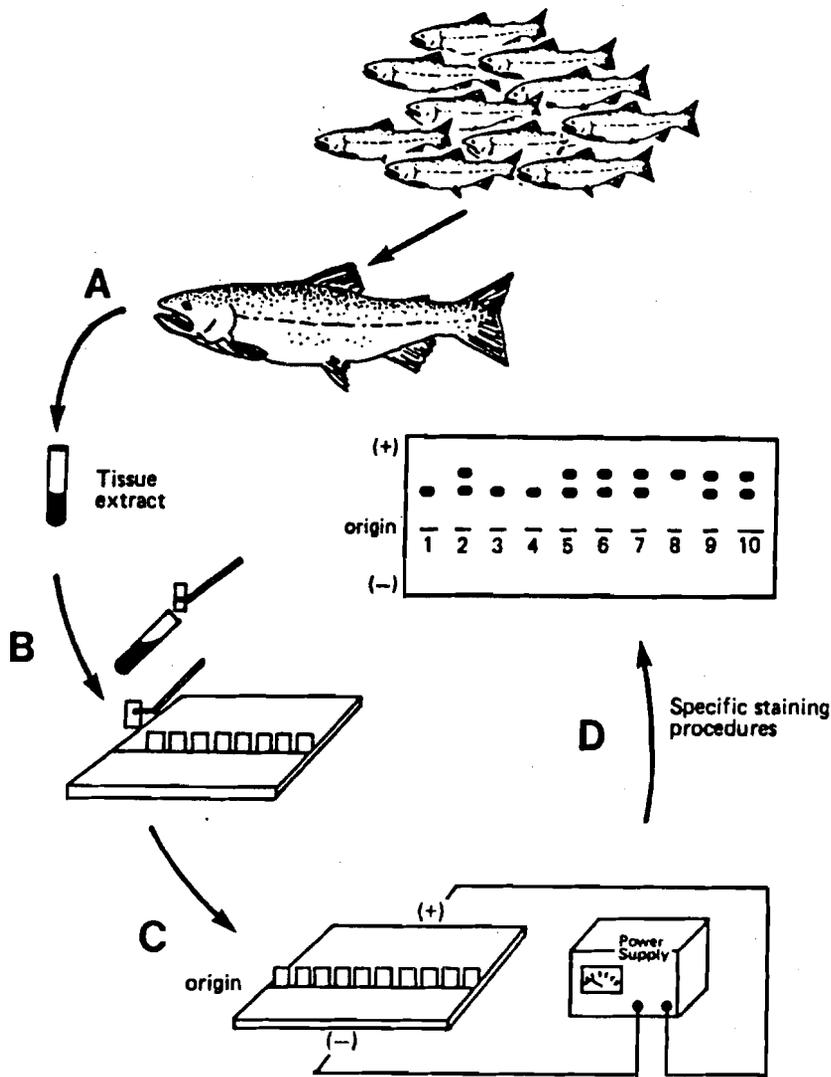


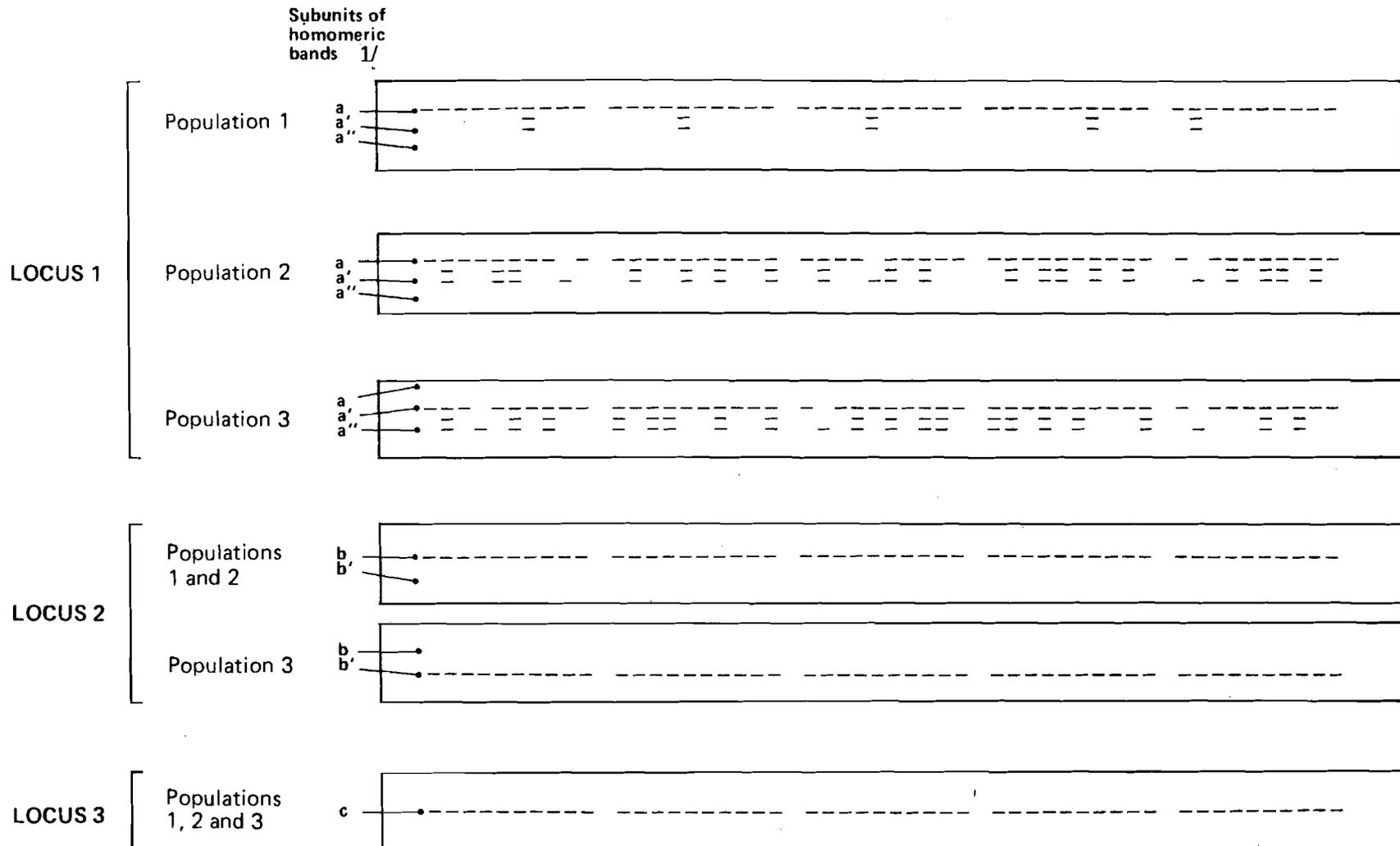
Figure 2. Standard steps for obtaining genotypic data through electrophoresis (modified from Gharrett and Utter 1982). A. Make crude protein extract from a tissue such as muscle or liver. B. Extract from each fish is introduced individually to gel by filter paper inserts. C. Different forms of a particular protein often move different distances from the point of applications when electric current (DC) is applied because of different electrical charges. D. These forms are then readily identified by a specific stain for each protein type. Specificity in staining permits identifying both the activity and the exact location of a particular protein for an individual fish from a complex mixture of proteins in each protein extract (intensities of banding patterns do not reflect differences of gene dosages in this depiction).

GENOTYPES			Subunit and subunit combinations in electrophoretic (protein) bands
AA (homozygote)	AA' (heterozygote)	A'A' (homozygote)	
PHENOTYPES			
Monomer	—	— —	a a'
Dimer	—	— — —	aa aa' a'a'
Tetramer	—	— — — — —	aaaa aaaa' aaa'a' aa'a'a' a'a'a'a'

Figure 3. Electrophoretic phenotypes of two-allele polymorphisms for monomeric, dimeric and tetrameric proteins when one locus is expressed. Allelic genes at the respective loci are designated A and A'; protein subunits synthesized by these alleles are designated a and a'.



Figure 4. Actual gel patterns of single locus polymorphisms. Phenotypes of a tetrameric protein (lactate dehydrogenase) from liver extracts of rainbow trout.



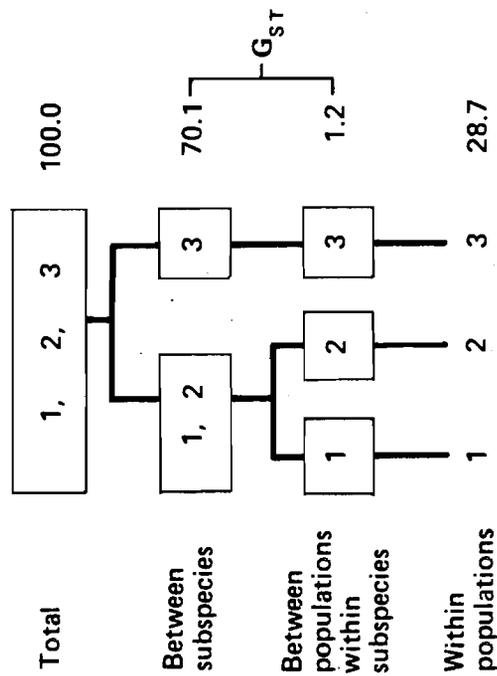
1/ Subunits are the products of A, A', A'' alleles of locus 1;
B, B' alleles of locus 2; and C allele of locus 3.

Figure 5. A hypothetical set of gel phenotypes for three loci from 50 individuals sampled from each of three populations of a species.

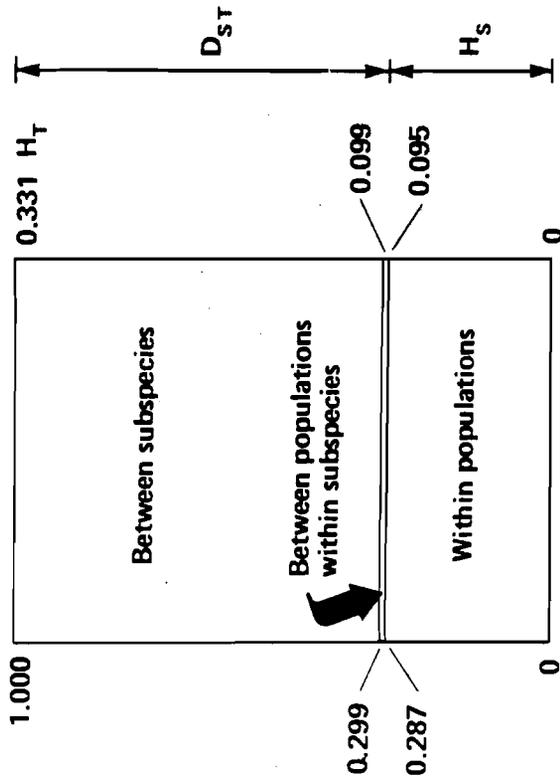
GENE DIVERSITY

Relative Absolute

Source of diversity Percentage of total



A



B

Figure 6A. Hierarchy of gene diversity analysis for data of Figure 5 and Table 1.

Figure 6B. Histogram comparing relative and absolute measures of gene diversity.

APPENDIX

(For calculations from data of Table 1)

The reader is referred to Ferguson (1980) and Hartl (1980) for details of additional calculations involving Mendelian data collected by electrophoresis for protein-coding loci.

Proportion of loci polymorphic

The proportion of polymorphic loci is the fraction of loci that are polymorphic of the loci examined. It is customary to specify a particular level of polymorphism (e.g., 0.95, 0.99). For the individual populations the proportion is the same, i.e.,

$$1/3 = 0.333.$$

For locus 2, populations 1 and 2 and population 3 are fixed for different alleles. Consequently, the species is considered polymorphic for this locus as well (although no heterozygous individuals were observed) and the proportion of polymorphic loci is

$$2/3 = 0.667.$$

Average number of alleles

The average number of alleles per locus is the number of different detected alleles divided by the number of examined loci. For each population, this number is

$$(2 + 1 + 1)/3 = 1.3.$$

For combined samples, this number is

$$(3 + 2 + 1)/3 = 2.0$$

Hardy-Weinberg proportions

For locus 1 and population 1 the expected number of AA genotypes in the sample of 50 individuals is

$$p(A)^2 \times 50 = (0.95)^2 \times 50 = 45.1.$$

Similarly, the expected numbers of AA' and A'A' genotypes are respectively

$$2p(A)p(A') \times 50 = 2(0.95)(0.05) \times 50 = 4.8$$

and

$$p(A')^2 \times 50 = (0.05)^2 \times 50 = 0.1.$$

If each of the samples of 50 individuals were drawn from the same panmictic population, the expected numbers of the total genotypes should approximate the numbers predicted from the binomial expansion of the mean values of the allele frequencies, i.e.,

$$p(A) = (0.95 + 0.75 + 0)/3 = 0.567,$$

etc. The expected numbers of the six genotypes for locus 1 are

$$AA = p(A)^2 \times 150 = (0.567)^2 \times 150 = 48.2$$

$$AA' = 2p(A)p(A') \times 150 = 2(0.567)(0.350) \times 150 = 59.5$$

$$A'A' = p(A')^2 \times 150 = (0.350)^2 \times 150 = 18.4$$

$$AA'' = 2p(A)p(A'') \times 150 = 2(0.567)(0.083) \times 150 = 14.1$$

$$A'A'' = 2p(A')p(A'') \times 150 = 2(0.350)(0.083) \times 150 = 8.7$$

$$A''A'' = p(A'')^2 \times 150 = (0.083)^2 \times 150 = 1.0.$$

It is clear that Hardy-Weinberg proportions of genotypes are not approximated from comparing these values with the observed totals of Table 1. Less obvious deviations may be tested statistically by chi-square or G-statistic (see contingency test, below). Appropriate degrees of freedom for tests of Hardy-Weinberg proportions are the number of phenotypes minus 1, minus the number of alleles minus 1; with three alleles the degrees of freedom would be

$$(6 - 1) - (3 - 1) = 5 - 2 = 3.$$

Heterozygosity at individual loci is the proportion of observed or expected heterozygotes relative to the total genotypes, e.g.,

$$5/50 = 0.10 \text{ (observed heterozygosity)}$$

$$4.8/50 = 0.09 \text{ (expected heterozygosity)}$$

for locus 1 and population 1.

Average heterozygosity of a population is the sum of the heterozygotes at individual loci divided by the number of loci, e.g., for the observed heterozygotes of locus 1 and population 1

$$0.10/3 = 0.033.$$

The average heterozygosity within populations is the sum of the heterozygosities for each population divided by the number of populations, e.g.,

$$(0.033 + 0.127 + 0.127)/3 = 0.096$$

for the observed average heterozygosity within populations.

Total gene diversity is the expected heterozygosity based on the mean values of allele frequencies averaged over all loci, i.e.,

$$[2(0.567)(0.350) + 2(0.567)(0.083) + 2(0.350)(0.083) + 2(0.667)(0.333)]/3 = 0.331$$

Contingency tests

A contingency table for the observed and expected (in parentheses) numbers of A and A' alleles in populations 1 and 2 is

		Populations		
		1	2	
Alleles	A	95(85)	75(85)	170
	B	5(15)	25(15)	30
	Total	100	100	200

where the expected values are the averages of the two populations for alleles A and A'.

A chi-square test for these data with

$$(\text{number of alleles} - 1)(\text{number of populations} - 1) = 1$$

degrees of freedom is the sum of

$$(O - E)^2 / E$$

where O is the observed and E the expected value, i.e.,

$$(10)^2/85 + (10)^2/85 + (10)^2/15 + (10)^2/15 = 15.68.$$

A G-test for the same data set and degrees of freedom is

$$2[(\text{flnf for the allelic frequencies}) - (\text{flnf for the row and column totals}) + (\text{nlfn})]$$

where f is the number of alleles observed in each cell and n is the total number of alleles, i.e.,

$$2[(95 \ln 95 + 5 \ln 5 + 75 \ln 75 + 25 \ln 25) - (100 \ln 100 + 100 \ln 100 + 170 \ln 170 + 30 \ln 30) + (200 \ln 200)] = 16.9.$$

The distribution of the G-statistic approximates that of chi square and significance levels of both tests can be found in tables of chi square.

Genetic distance

Genetic distance (D) as given in the text is

where

$$I = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}}$$

and x_i and y_i are frequencies of specific alleles in populations x and y.

Calculation of D for populations 1 and 2 is as follows:
In locus 1,

$$I = \frac{(0.95)(0.75) + (0.05)(0.25)}{[(0.95)^2 + (0.05)^2][(0.75)^2 + (0.25)^2]}^{0.5} = 0.964.$$

For both loci 2 and 3

$$I = \frac{(1)(1) + (0)(0)}{[(1)^2 + (1)^2]}^{0.5} = 1$$

The combined value of I for the 3 loci is

$$(1 + 1 + 0.964)/3 = 0.988,$$

and

$$D = -\ln(0.988) = 0.012.$$

Calculation of D for populations 1 and 3 is as follows:

In locus 1

$$I = (0.50)(0.75) / [(95)^2 + (0.05)^2][(0.75)^2 + (0.25)^2]^{0.5} = 0.050$$

In locus 2, I=0 because of no common alleles and a numerator of 0.

In locus 3, I=1 in the same manner as populations 1 and 2, loci 2 and 3.

The mean value for I over all three loci is

$$1.05/3 = 0.350,$$

and

$$D = -\ln 0.350 = 1.05$$

In a similar manner, the I values for populations 2 and 3 are

locus 1 - 0.249
locus 2 - 0
locus 3 - 1
average - 0.416

and

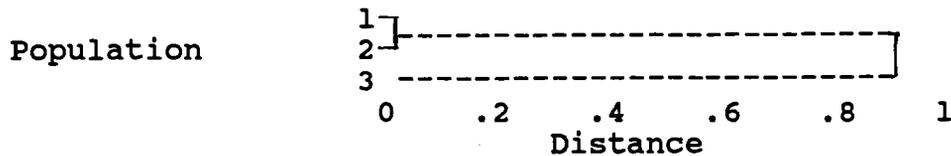
$$D = 0.876.$$

Cluster analysis

The unweighted pair-group method (UPGM) of clustering for constructing a dendrogram starts with a matrix of paired distance (or similarity) values. The first two groups to be clustered are those with the lowest distance (or highest similarity) value. A new matrix is then formed - reduced by one row and column - containing these paired groups as a single member. The recalculated values involving this new member are the means of the values of its two component groups with each of the other members from the original matrix. The process is repeated (in large matrices) through the ultimate averaging of values from two remaining groups. In the present case, populations 1 and 2 join at a D value of 0.012. The new member (comprising populations 1 and 2) joins population 3 at a D value of

$$(1.050 + 0.876)/2 = 0.963.$$

The cluster then, appears as



Gene diversity analysis

The initial calculations for gene diversity analysis have already been outlined under different calculations of expected heterozygosity in the section presenting Hardy-Weinberg calculations. An alternate formula for expected heterozygosity at a locus is

$$(1 - \text{the sum of squared allelic frequencies}).$$

The mean expected heterozygosity of Table 1 is $H(S)$ as outlined in the text; likewise, the total diversity is $H(T)$.

The only additional value that requires calculation in the present hierarchy is the increase in expected heterozygosity from $H(S)$ at the level of populations within subspecies. This value is obtained by pooling the allelic frequencies of the populations within each subspecies (in this instance, populations 1 and 2) and recalculating the mean heterozygosity from $H(S)$ at the level of populations within subspecies. This value is obtained by pooling the allelic frequencies of the populations within each subspecies (in this instance, populations 1 and 2) and recalculating the mean heterozygosity as if the subspecies themselves were the individual populations. This calculation can be weighted assuming the sampling is representative of the actual number of population units existing in the species (Chakraborty 1980); i.e., for locus 1

$$[2(2)(0.85)(0.15) + 2(0.75)(0.25)]/3 = 0.295.$$

Alternatively, the calculation can be unweighted, assuming the sampling is independent of the actual number of units (D. Campton, personal communication; see also Chakraborty and Leimar, 1987); i.e.,

$$[2(0.85)(0.15) + 2(0.75)(0.25)]/2 = 0.315$$

The difference between either of these figures and $H(S)$ is that portion of the structuring that is due to population heterogeneity within subspecies depending on which assumption is most appropriate. Under the first assumption,

$$.295 - .282 = .013.$$

The remainder of the structuring within the species, i.e.,

$$0.268 - 0.013 = 0.255$$

is that portion due to differences between subspecies.

For locus 2, all diversity is due to differences between groups because of fixation of alternate alleles in different subspecies. The absolute gene diversity calculated on the weighted allele frequencies is

$$2(0.667)(0.333) = 0.444.$$

The average absolute gene diversity between subspecies is

$$(0.255 + 0.444)/3 = 0.232.$$

Similarly, the average gene diversity between populations is

$$(0.013)/3 = 0.004.$$

Calculations of coefficients of gene diversity are based on that proportion of H(T) that is due to population heterogeneity at a particular level of the hierarchy: i.e., for the within population component

$$0.095/0.331 = 0.287;$$

the combined diversity from subdivision at both hierarchical levels (G(ST)) is

$$(0.331 - 0.095)/0.331 = 0.713;$$

the coefficient of gene diversity due to subspecies is

$$(0.232)/(0.331) = 0.701;$$

the coefficient of gene diversity due to populations within subspecies is

$$(0.004)/(0.331) = 0.012.$$