in Population Genetics & Fishery Management
N. Ryman & F. Utter, eds. 1987

35

2.
Interpreting Genetic Variation Detected By Electrophoresis

Fred Utter, Paul Aebersold, and Gary Winans

Fishery biologists generally recognize the need to identify differences among stocks and to monitor genetic changes. However, many biologists do not recognize the link between stock differences, protein variation, and genetic variation inherited in a Mendelian manner. This chapter is intended to provide that link.

Outlined are some of the basic genetic principles and procedures underlying the practical application of population genetics to fishery management problems. Some of the information is now fundamental to introductory undergraduate courses, but many fishery biologists completed their training before molecular biology and the molecular basis of hereditary variation were taught at the undergraduate level. However, most of this chapter deals with the understanding and interpretation of genetic data as revealed by electrophoresis. This tool continues to play a major role in understanding the levels and patterns of genetic variability within and among populations.

It was pointed out in Chapter 1 that contributions of individual genes usually cannot be identified from studies of quantitative traits such as length, fecundity, and number of gill rakers. Quantitative characters are therefore excluded in this chapter. The reader is referred to Ferguson (1980) for a general, and in some areas more detailed, examination of many topics approached in this chapter.

Mendel's studies (1866) first identified units of inheritance which were subsequently termed *genes* by Johannsen (1909). Considerable theory concerning the dynamics of Mendelian genes had developed by the early 1930s. A conceptual framework for the interactions of mutation, migration, selection, and drift in the creation, maintenance, and distribution of Mendelian genes in natural populations was established through the writings of Fisher (1930), Haldane (1932), and Wright (1932).

Very little empirical information was available to match this theoretical groundwork for over 30 years. An exception was the detection of sizable levels of deleterious recessive genes detected in *Drosophila* (Dubinin et al. 1937). Also, knowledge of simply inherited human blood groups was accumulating by the 1930s (reviewed in Boyd 1966). Indeed, human blood groups provided early and extensive observations on intraspecific struc-

turing of populations based on information from Mendelian genes (Mourant 1954).

A revolutionary advancement in the ability to identify Mendelian genes arose from two developments in the 1950s. Watson and Crick (1953) deduced the structure of the DNA molecule, which ultimately clarified the direct relationship between genes and proteins. This knowledge was followed by the development of electrophoretic procedures which permitted rapid and reliable identification of protein variations reflecting simple genetic differences (Smithies 1955, Hunter and Markert 1957). The ease with which Mendelian variants could now be detected by electrophoresis (contrasted with the previously great difficulty in detecting such variation) resulted in a proliferation of descriptive studies of Mendelian variants of proteins in many organisms (see deLigny 1969, 1972 for reviews of early electrophoretic studies of fishes). Eventually, the classical studies of Lewontin and Hubby (1966) in *Drosophila pseudoobscura* and of Harris (1966) in man clearly suggested that substantially higher levels of genetic variation exist throughout all classes of organisms than had previously been known.

The simple inheritance observed for blood groups in man (cited above) and other higher vertebrates (e.g., cattle, Stormont et al. 1951; chickens, Briles et al. 1950) suggested the existence of similar Mendelian markers in fishes and led to studies of blood groups in such fishes as tuna (Cushing 1956) and salmonids (Ridgway 1957, Sanders and Wright 1962). The anticipated usefulness of these studies for identifying Mendelian variations was not fulfilled because of technical limitations such as fragility of fish erythrocytes and difficulties in producing and preserving discriminating antisera (reviewed in Hodgins 1972). However, information from electrophoretic studies has subsequently met and surpassed the expectations envisioned from blood groups. Population structures of fish species are being clearly defined with the use of purely genetic data (e.g. Allendorf and Utter 1979, Winans 1980, Shaklee 1983, Ferguson and Mason 1981, Ryman 1983). Using such data, statistical and data processing procedures have been developed for obtaining detailed, accurate, and timely estimates of mixed stock compositions (Grant et al. 1980, Fournier et al. 1984, Milner et al. 1985, Chapter 10).

In this chapter we describe the direct connection between the gene and its expression as a protein molecule, with the intention of making the remaining chapters more accessible to the uninitiated reader. Basic principles and terms are first introduced. An extended section on the most frequently applied technique for studying genetic variation in natural populations—protein electrophoresis—is presented because of the central role this technique has played in revealing Mendelian characters for population genetic studies. Illustrations focus on the complex patterns of the extensively studied salmonids. The chapter closes with observations concerning the limitations of current techniques, which examine only a small portion of the total amount of genetic material.

BASIC PRINCIPLES AND TERMS

Some basic principles and terms concerned with the molecular basis of genetic variation are introduced at this point to clarify their use in this chapter. Most genes in higher diploid organisms such as fishes and man are contained in structures of the cell nucleus called *chromosomes*. A much smaller fraction of genes are found outside the nucleus and include those found in the mitochondria (see Chapters 11 and 12). Chromosomes (and therefore genes) occur in pairs as a consequence of individual sets of chromosomes that are inherited from each parent. These individual sets are transmitted in germ cells called *gametes* (sperm and egg cells in animals). The process of gamete formation (gametogenesis) includes *meiosis* (outlined in any elementary genetics text), which allows chromosomes of each parent to assort independently to each gamete (Mendel's second law). Gametes unite through sexual processes; the egg is fertilized by an individual sperm cell to form a *zygote*. Fertilization, then, results in the pairing of individual sets of chromosomes.

The single-celled zygote soon develops into a highly differentiated collection of tissues and organs performing broadly diverse functions. This differentiation occurs because, although each cell has an identical complement of genes, very few of the total number of genes are active in a particular cell of higher organisms (e.g., see Alberts et al. 1983). Differentiation is the result of the interactions of *regulatory* genes (which determine at what time and in what tissue a particular gene is expressed) and *structural* genes (which contain coded information for proteins that are produced by the organsism). These complex and still poorly understood interactions lie beyond the scope of this chapter (but see McDonald 1983 for a recent review of progress in understanding these interactions). However, the direct relationship between structural genes and their protein products is well understood and is our primary focus.

The condition in which only a single set of chromosomes is present, such as in gametes, is called *haploidy,* while *diploidy* describes the paired chromosome complements in the zygote and subsequently formed tissue cells. Occasionally zygotes are formed with three or four sets of chromosomes *(triploidy* and *tetraploidy)*. The ability to induce the triploid condition is currently receiving considerable attention because of the general sterility of such individuals (see Chapter 13). The tetraploid ancestry of some families of fishes (e.g., Catostomidae and Salmonidae) has resulted in some special evolutionary opportunities for these species (see Ferris and Whitt 1979, Allendorf and Thorgaard 1984). Some consequences of tetraploid ancestry with respect to complexities of electrophoretic expressions are examined later in this chapter.

The location of a gene on a chromosome is called a *locus* (plural *loci*). The paired set of genes inherent in diploidy permits two different forms of a gene for a particular locus to exist in a single individual. Different forms of a gene are called *alleles*. Many alleles may exist for a particular locus in a spe-

cies, but a single diploid individual can carry no more than two alleles at a locus. An individual is *homozygous* at a particular locus if the genes at that locus are identical, and *heterozygous* if they are different. A locus is said to be *monomorphic* if only one allele is known, and *polymorphic* if two or more alleles are known. The set of alleles possessed by an individual at a particular locus (or set of loci) is referred to as the individual's *genotype* at this locus. The *phenotype* is the observed character of an individual, and may be influenced by the environment as well as the genotype.

The fundamental chemical substance of the gene is *deoxyribonucleic acid*, or DNA. DNA is a giant molecule constructed in a so-called double helix like a very long spiral ladder (see Fig. 2.1). The sides of the ladder are alternating sugar (called deoxyribose) and phosphate groups. The rungs of the ladder attached to the sugars are pairs of *bases:* adenine (A), guanine (G), thymine (T), and cytosine (C). These bases always combine as either A – T (or T – A) or C – G (or G – C) because of physical and chemical constraints that preclude other pairings. The genetic information is contained in different sequences of these four bases read from one side of the ladder, the DNA coding strand.

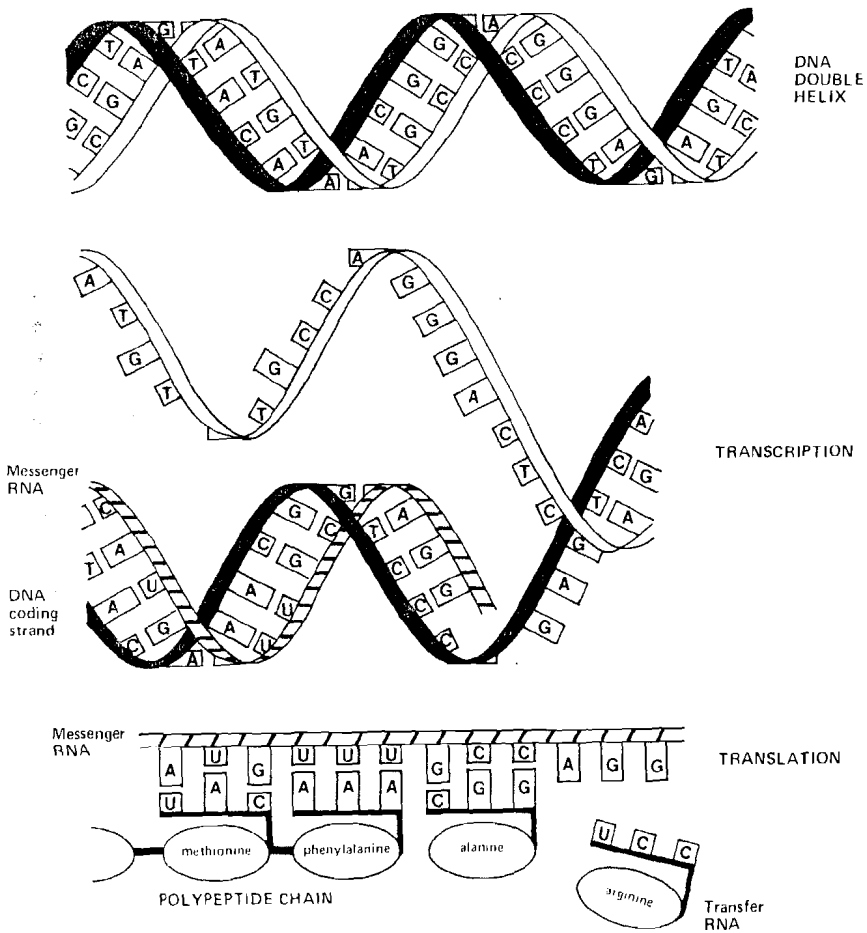

Fig. 2.1 Molecular processes relating base sequences of DNA to amino acid sequences of polypeptide chains (proteins). Messenger RNA, synthesized during transcription, provides a template for the synthesis of the polypeptide chain during translation. The bases in DNA are cytosine (C), guanine (G), adenine (A), and thymine (T). In RNA, the base uracil (U) replaces thymine (T) of DNA.

The base sequence of DNA has a direct linear relationship to the structure of proteins. Proteins are similar to DNA in that they are large molecules made up of different components called *amino acids*. There are 20 common amino acids in nature. Amino acids are connected by peptide bonds in a series to form *polypeptide* chains. Active proteins are made of polypeptide chains alone or in aggregate, depending upon the protein. Each polypeptide chain is called a *subunit*. Most proteins have at least 100 amino acids.

It has been found that different combinations of the four bases read in sequences of three (called *coding triplets* or *codons*) have information (or code) for different amino acids. This information, placed in a line on the DNA molecule, tells the cell which amino acids form a protein molecule and in what order they go. The four bases G, C, A, and T in various combinations of three can be arranged in 64 different ways. Some of the 64 triplets are used to tell the cell where one gene stops and another one starts on the long strands of DNA of a particular chromosome. Others represent a redundancy in the genetic code, more than three times as many combinations as are needed to code for all 20 amino acids; the sequences GGT, GGC, GGA, and GGG, for example, all code for the amino acid glycine.

The uncoding of the segment of triplets in the DNA molecule into the sequence of amino acids in a protein is a two-step process (Fig. 2.1). First, the genetic information from the DNA template is copied, or transcribed, into the nucleotide sequence of a second type of nucleic acid, *ribonucleic acid* (RNA). This RNA is called *messenger* RNA (mRNA) because it carries the coded information of the DNA molecule from the nucleus to the cytoplasm, where protein synthesis occurs. This process of synthesizing mRNA is appropriately called *transcription*. Unlike DNA, the mRNA is single stranded and very small, enabling it to pass from the nucleus to the cytoplasm through small pores in the nuclear membrane. The polypeptide chains are assembled in the cell cytoplasm on structures (organelles) called *ribosomes* through a process called *translation*. This process involves sequential base pairing of triplets on the mRNA with triplets on molecules of *transfer* RNA (tRNA), which, like mRNA, are encoded by a DNA template. Each transfer RNA carries a specific amino acid; there are different tRNAs for each of the 20 amino acids. Coupling between a triplet of mRNA and a specific tRNA occurs through recognition of complementary triplets on the two RNA molecules. Thus, through transcription and translation, the sequences of amino acids in proteins are direct reflections of the base sequences of DNA that constitute the genes. The sequence of DNA

encoding on a single polypeptide chain is currently defined as the *unit gene* (Rieger et al. 1976).

Cell division and chromosome and DNA replication are complicated processes. Mistakes are made occasionally in the formation of gametes. Mispairings of bases can lead to amino acid substitutions in proteins (e.g., a base substitution of TTC for TTA results in an amino acid substitution of phenylalanine for leucine in a polypeptide chain) or to an actual discontinuation of chain building and very likely a nonfunctional protein. Such mispairings are a common source of *mutations* (others include actual structural changes in the chromosomes), which, when passed on to the next generation, are the ultimate origins of all genetic variation.

We can summarize to this point:

- Most genes in most diploid organisms occur in pairs located on individual sets of chromosomes contributed by each parent.
- DNA is the informative chemical substance of the gene.
- Amino acid sequences in proteins directly reflect base sequences in genes.
- Mispairings of bases in DNA are one source of mutations that can result in amino acid alterations in proteins.

GENOTYPIC DATA FROM ELECTROPHORESIS

The Process of Electrophoresis

Five of the 20 common amino acids which make up proteins are charged; the charges of lysine, arginine, and histidine are positive, while those of aspartic acid and glutamic acid are negative. Thus, different proteins tend to have different net electrical charges. Electrophoresis uses this physical chemical property of proteins to separate mixtures of proteins on the basis of charge. If allelic differences (i.e., different forms of a gene) occur at a protein coding locus, the net charge of the protein often changes. Gel electrophoresis makes it possible to identify such allelic differences.

The basic procedures of gel electrophoresis are outlined in Fig. 2.2. The process of electrophoresis includes a gel (commonly starch or polyacrylamide) in which introduced solutions of proteins are separated 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, unless they are already contained in body fluids such as vitreous humor or blood serum. The water soluble protein mixtures are typically introduced to the gel on a piece of filter paper that is saturated with the mixture. Protein mixtures from 50 or more individuals are often introduced to a single gel, although only 10 individuals are pictured in Fig. 2.2.

A direct current is usually applied for 3–5 hours through the gel. The actual time is determined by such variables as composition of the buffer solution used to make the gel, its ionic strength, and the thickness of the gel. Pro-

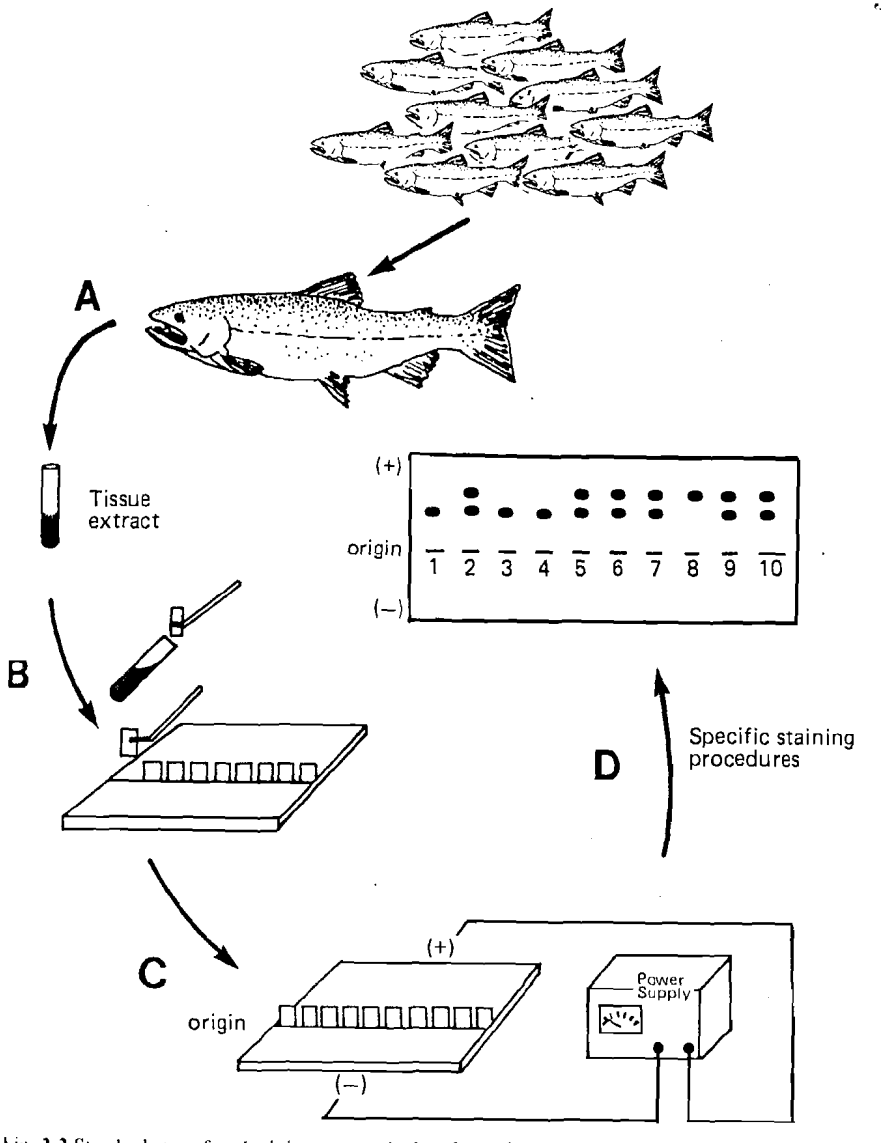

Fig. 2.2 Standard steps for obtaining genotypic data from electrophoresis (modified from Gharrett and Utter 1982). (A) Crude protein is extracted from tissue such as muscle or liver. (B) Extract from each fish is introduced individually to gel by filter paper inserts. (C) Different forms of a particular protein often move different distances from the point of applications when electric current is applied because they do not have identical electrical charges. (D) These forms are readily identified by a specific stain for each protein type. Specificity in staining permits identification of 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.)

teins with a positive net electrical charge move toward the negative (or cathodal) pole and negatively charged proteins move toward the positive (or anodal) pole. The rate of migration is determined by the absolute charge of the protein. A dye solution is added to the sample to mark the progress of electrophoresis. Following electrophoresis, the gel is sliced horizontally into multiple slabs and each slab is stained for the activity of a specific protein.

Most proteins that are studied by electrophoresis are *enzymes,* the catalytic molecules vital to all life, because it is easy to develop 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 precisely locate that enzyme in the gel.

The localization of an enzyme's activity in a gel has been called the "isozyme method." Isozyme refers to different distinguishable molecules found in the same organism which catalyze the same reaction (see Markert and Moller 1959, Shaw 1964, Brewer 1970). *Allozyme* commonly refers to the electrophoretic expression of allelic proteins at a particular locus. The capability to visually localize an enzyme's activity has resulted in the detection of activities of dozens of enzymes reflecting 90 or more loci (e.g., Morizot and Siciliano 1984).

Specific staining for an enzyme's activity permits particular isozymes to be distinguished, one at a time, in a mixture of hundreds of proteins typically found in a tissue extract from an individual fish. The final result of the electrophoretic procedure is bands, such as those in Fig. 2.2, which identify the locations of various forms of a single type of 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.

Expressions of Single Loci

The connection between DNA base sequences, protein amino acid sequences, and the electrophoretic expression of different genotypes is most easily illustrated for a *monomeric* protein. Monomeric proteins are proteins composed of single subunits (i.e., a single polypeptide chain). Let us assume that

- A locus is coded for a monomeric protein having two alleles designated A and A' (i.e., a polymorphic locus);
- These alleles produce subunits (the active protein for monomers), designated a and a' respectively, that are distinguishable by different electrophoretic mobilities; and
- The a' protein encoded by the A' allele migrates more slowly than the a protein encoded by the A allele.

Three different genotypes are possible for an individual at this locus: AA, AA', and $A'A'$. An individual with the AA genotype produces only the faster migrating protein form. This form appears at one location on the gel as a single band. Similarly, an individual with the $A'A'$ genotype produces only the slower migrating form at a different location on the gel. The heterozygous (AA') genotype produces both protein forms and therefore is reflected as two bands on the gel. We assume that each allele results in the production of equal amounts of protein having the same levels of activity. Therefore, each of the two bands of heterozygous individuals is expected to reflect half the amount of protein that is reflected by either of the homozygous types of individuals; that is, each band of a heterozygous individual expresses half the *dosage* of the single band expressed by a homozygous individual. This pattern of genotypic expression of a monomeric protein encoded by a single locus with two alleles is pictured at the top of Fig. 2.3. Commonly studied monomeric proteins include the serum protein transferrin and such enzymes as phosphoglucomutase, mannosephosphate isomerase, and aconitate hydratase.

Banding patterns on a gel become more complicated when the active protein is *multimeric,* i.e., composed of two or more protein subunits. If we extend the above assumptions to a *dimeric* protein (i.e., one consisting of two subunits), the expected banding patterns are those pictured in the middle section of Fig. 2.3. An individual with the genotype AA is expressed as a single

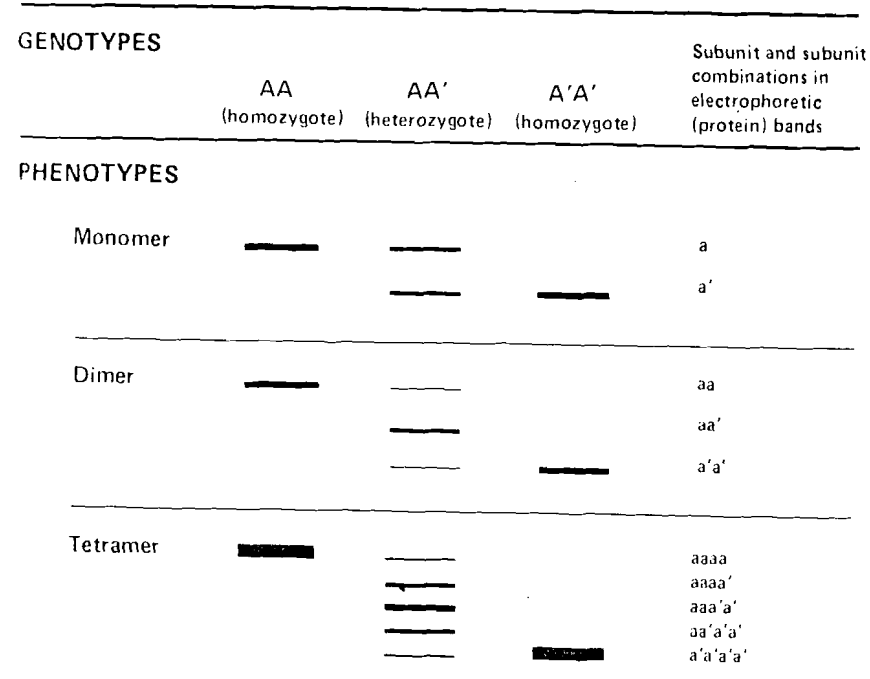

Fig. 2.3 Electrophoretic phenotypes when one locus is expressed. Individuals are homozygous and heterozygous at loci coding for monomeric, dimeric, and tetrameric proteins: the locus is polymorphic, with alleles A and A' resulting in subunits a and a', respectively.

and reflecting identical molecules of *a* subunits combined in pairs. Similarly, the expression of an individual with the *A'A'* genotype is another single band reflecting paired *a'a'* subunits at a different location on the gel. An individual with the *AA'* genotype, however, is expressed by 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. The third middle band 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 both heterozygous and homozygous individuals. Dimeric proteins commonly studied by electrophoresis include the enzymes malate dehydrogenase, isocitrate dehydrogenase, and aspartate aminotransferase.

The banding patterns of a protein having four subunits (a *tetramer)* are pictured in the lower portion of Fig. 2.3. Again, we assume a single locus polymorphic for two electrophoretically detectable alleles. The respective homozygous expressions are single-banded because of the identity of each of the four subunits. The heterozygous individual has five bands, representing random combinations of two allelic subunits in aggregates of four. The five bands include three heteromeric bands in addition to the two homomeric bands; again, their combined intensity is equivalent to the single band of the homozyous expressions. Commonly studied tetrameric proteins include lactate dehydrogenase (see Fig. 2.4), iditol dehydrogenase, and malate dehydrogenase (NADP dependent).

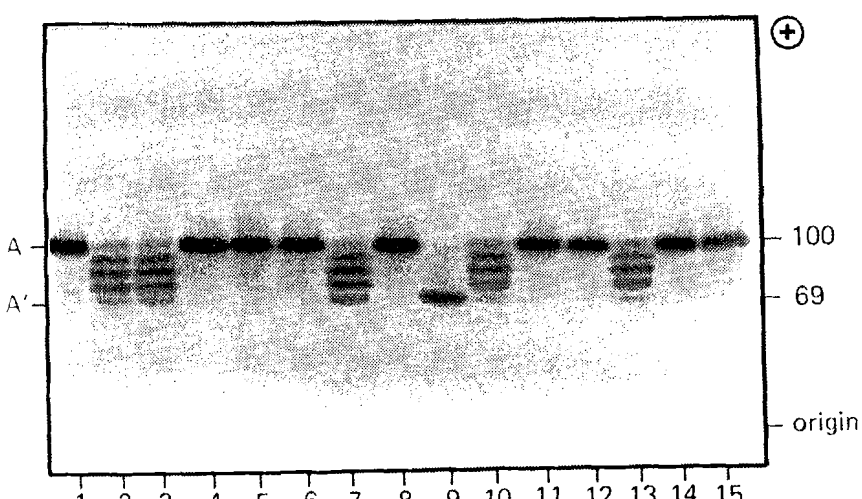

Fig. 2.4 Phenotypes of a two-allele polymorphism for a locus encoding a tetrameric enzyme, lactate dehydrogenase, from livers of 15 rainbow trout. Note five-banded expressions of heterozygous individuals (2, 3, 7, 10, 13).

The expected number of bands and their relative intensities for individuals heterozygous at a particular locus can be predicted assuming that subunits combine randomly following their synthesis. The basis for these expectations can be demonstrated through the randomness of flipping a coin, whose sides represent the allelic subunits *a* and *a'*. The coin is flipped in repeated series, with the number of flips in each series representing the number of subunits in the protein (one flip for a monomer, two flips in a series for a dimer, and four flips in a series for a tetramer). The sequential outcome of each series is recorded before going on to the next series.

For a dimeric protein, four combinations are possible when flipping the coin in a series of two: *aa, aa',a'a,* and *a'a'*. There is an equal probability of getting any of the four types. The sequences *aa'* and *a'a* represent identical dimeric proteins which would form a single electrophoretic band; therefore, their probabilities can be pooled. Thus, the expression of the dimeric heterozygote is three-banded, with the combined *aa'-a'a* band at twice the intensity of the respective *aa* and *a'a'* bands. This coin flip analogy can be applied to proteins with other subunit structures.

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 the expression would be

$$(a + a')^2 = a^2 + 2aa' + a'^2 .$$

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 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 4. Following expansion, then, the relative intensities of 1:4:6:4:1 would be expected from tetramers. Proteins are sometimes encountered whose subunit structures are something other than monomeric, dimeric, or tetrameric (e.g., the enzyme purine nucleoside phosphorylase has three subunits). The expected numbers and relative intensities of electrophoretic bands can be predicted for them in the same manner if the subunit structure of the protein is known.

The expected banding patterns are idealized configurations. It is important to recognize that some deviations from the expected numbers and relative intensities of bands are frequently seen. There are both genetic and nongenetic reasons for these deviations, some of which are discussed later in this chapter. However, understanding the basis of these idealized configurations is essential for properly interpreting the genotypic basis of electrophoretic patterns.

The electrophoretic banding patterns such as those pictured in Fig. 2.2 and Fig. 2.3 are phenotypes. They are expressions of the genotypes (i.e., the actual alleles) with possible—usually minimal—influences of the in vitro en-

vironment of the protein. Thus, genotypes can usually be deduced directly from such phenotypes when the subunit composition of the protein is known. For instance, consider that a group of individuals are subjected to electrophoresis and the resulting gel is stained for lactate dehydrogenase activity. Banding patterns would be observed like those in the lower portion of Fig. 2.3 and in Fig. 2.4. It is safe to assume that such phenotypic patterns reflect the respective homozygous and heterozygous genotypes of an *LDH* (lactate dehydrogenase) locus of that species because the phenotypes conform to the expected numbers and intensities of bands for a tetrameric protein.

Expressions of Additional Loci

More complicated electrophoretic patterns than those depicted in Fig. 2.3 frequently occur when the same type of protein is encoded by two or more loci. These complications include additional protein bands arising from combinations of subunits, encoded by different loci and having different electrophoretic mobilities, or electrophoretic patterns resulting from two (or more) loci whose protein bands have the same or overlapping mobilities.

The latter patterns are particularly frequent in salmonids because of their tetraploid ancestry. Salmonids have about 50% more protein loci expressed than teleosts of diploid ancestry (see Allendorf and Thorgaard 1984). The salmonids have undergone the most intensive electrophoretic examination of any group of fishes, and the complexities of their electrophoretic patterns have often been confusing; an examination of these complexities is therefore warranted.

Let us extend the assumptions of the genotypic expressions of Fig. 2.3 to include a second locus. Every individual is homozygous for the *B* allele at this locus (i.e., a monomorphic locus) which encodes electrophoretically identical *b* subunits. Both loci are expressed at equal levels. Homomeric bands of *b* subunits have electrophoretic mobilities that are distinct from those of the *a* or *a′* subunits encoded by the *A* or *A′* alleles of the first locus. The phenotypes of individuals homozygous at both loci (columns 1 and 3 in Fig. 2.5) resemble the phenotypes of heterozygous individuals in the single locus expression of Fig. 2.3. This resemblence is due to the similarity between the expression of two alleles of a single locus in Fig. 2.3 and of two loci in Fig. 2.5. In monomeric proteins, a single band is expressed for each allele (in this case the *A* and *B* alleles contrasted with the *A* and *A′* alleles in the single locus case). Multimeric proteins express the additional bands of random interactions of the individual subunits (in this case *a* and *b* subunits contrasted with the *a* and *a′* subunits in the single locus case). The expected relative intensities of the bands of dimeric and tetrameric phenotypes of homzygous individuals at two loci are also the same as those of the phenotypes of heterozygous individuals when only a single locus is expressed.

The expression of a heterozygote when one of two loci is polymorphic for a monomeric protein is again a single band for each allele. However, hetero-

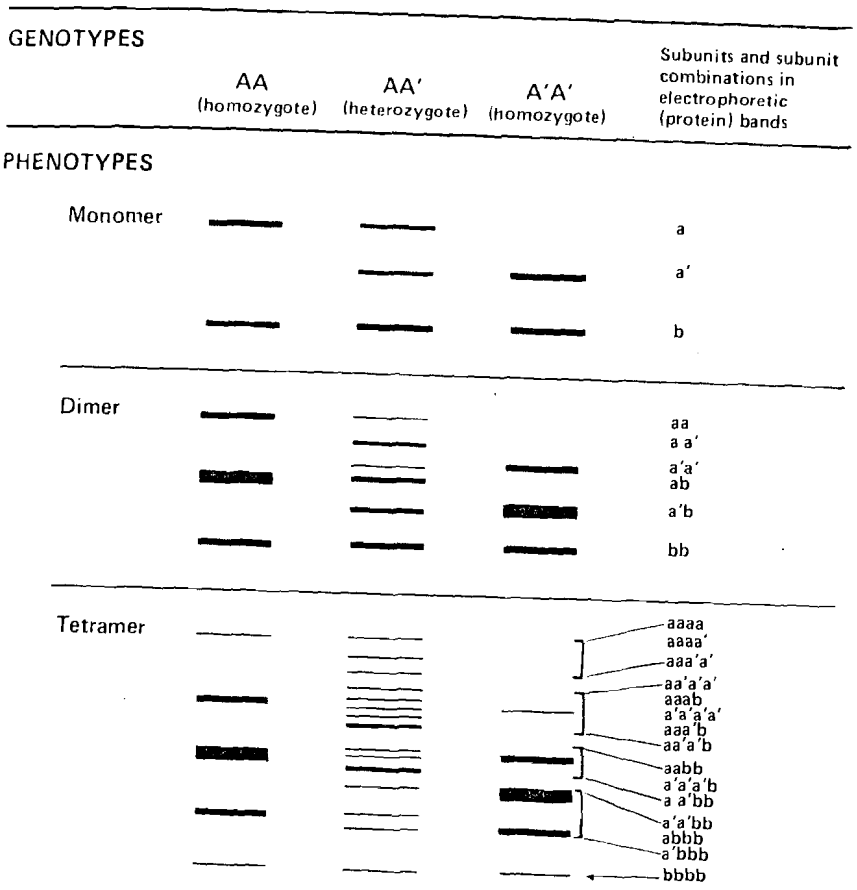

Fig. 2.5 Electrophoretic phenotypes when two loci are expressed. Individuals are homozygous and heterozygous at loci coding for monomeric, dimeric, and tetrameric proteins: one locus is polymorphic (with alleles *A* and *A′* resulting in subunits *a* and *a′*, respectively); and a second is monomorphic, coding for a subunit (*b*) with an electrophoretic mobility that differs from subunits *a* and *a′*

zygote phenotypes of the multimeric proteins in the situations given in Fig. 2.5 are complicated by combinations involving a third electrophoretically distinct subunit. The number of bands involving combinations of the *b* subunit from the second locus with *a* and *a′* subunits from the first locus are readily predicted for heterozygous phenotypic expressions by including this *b* subunit in the binomial expansion. For a dimer, with both doses of the single allele of the monomorphic second locus producing *b* subunits and the respective alleles of the first locus producing *a* and *a′* subunits, the squared expansion becomes

$$(a + a' + 2b)^2 = a^2 + 2aa' + a'^2 + 4ab + 4a'b + 4b^2$$

... five homodimeric and three heterodimeric) electrophoretically distinguishable subunit combinations (as pictured in the central portion of Fig. 2.5). ... binomial power expansion for tetrameric proteins predicts the 15 bands depicted in Fig. 2.5. In practice, some of the predicted bands may not be seen because of overlapping unless the respective homomeric bands are adequately separated by electrophoresis. An actual gel showing a polymorphism for one of two loci encoding a monomeric protein is shown in Fig. 2.6.

Exceptions to Codominant Expression

The phenotypes of Figs. 2.3 and 2.5 are called *codominant* expressions of the respective genotypes because the contributions of all alleles can be identified. Codominant expression is an important attribute of electrophoresis because of the value of genotypic information at individual loci in population genetic studies. However, there are exceptions to codominant electrophoretic expressions that need to be considered.

The occurrence of electrophoretically identical subunits synthesized by two different loci is observed in some fishes (e.g., salmonids). The genetic and evolutionary basis for such *isoloci* (having *isoalleles* giving rise to electrophoretically identical gene products) has been reviewed by Wright et al. (1983) and Allendorf and Thorgaard (1984). One locus or both loci may be variable. In either case the electrophoretic expression of isoloci complicates the determination of genotypes.

A part of this complication is that it is often impossible to assign alleles

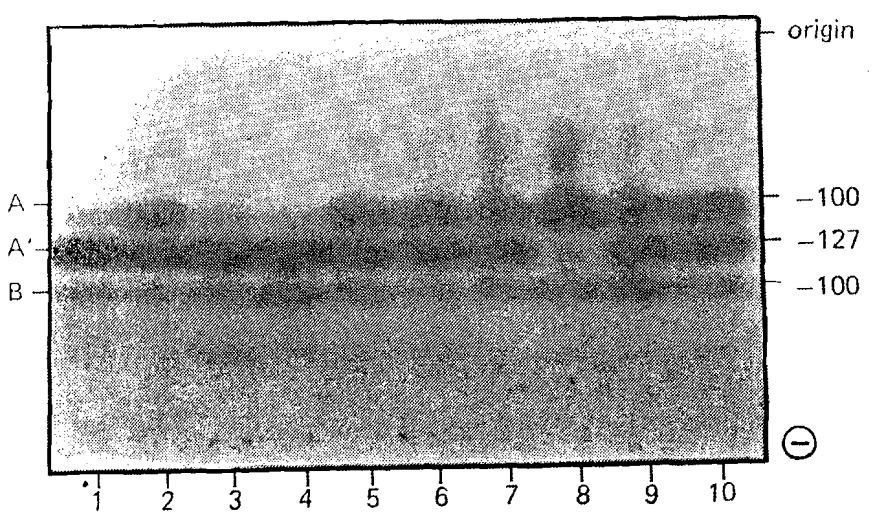

Fig. 2.6 Phenotypes of a two-allele polymorphism for a locus encoding a monomeric enzyme, phosphoglucomutase, from skeletal muscle of 10 sockeye salmon. A second monomorphic locus is expressed cathodal to the bands reflecting the different genotypes of the polymorphic locus. Note that the intensity of expression of the monomorphic locus is less than that of the polymorphic locus, suggesting different levels of synthesis. Note also distinct shadow banding of individuals 1, 3, and 4 that coincides with the mobility of the alternate allelic form.

to specific loci when two (or possibly more) loci code for electrophoretically identical subunits. The problem is apparent from the phenotypes of Fig. 2.7, which pictures the expressions of isoloci where one of the loci is polymorphic and the other monomorphic. The assumptions underlying the phenotypes of Fig. 2.5 are the same for Fig. 2.7 except that the products of the *B* alleles (i.e., *b* subunits) of the monomorphic locus are electrophoretically indistinguishable from those of the *A allele (i.e., a* subunits) of the polymorphic locus. This situation results in the inability to distinguish the contributions of the *a* and *b* subunits from the electrophoretic phenotype. Consequently, the numbers and mobilities of bands expressed by both AA' and $A'A'$ genotypes are the same. Although no *a* subunits are produced by the $A'A'$ genotype, the two doses of *b* subunits having the mobility of the *a* subunit mask this absence. Exactly the same situation would occur if the *B* locus instead of the *A* locus were polymorphic. There is no way to distinguish which locus is polymorphic if both loci are equally expressed in all tissues.

There are ways to deal with this problem. Phenotypes of isoloci are often recorded and analyzed as the summed contribution of four allelic doses for two loci (e.g., May et al. 1979). This procedure gives no information about the diploid genotypes of the individual loci (which has considerable value for determining whether individuals within a sampling represent a single random mating population). If only one locus is polymorphic (as in Fig. 2.7) it is con-

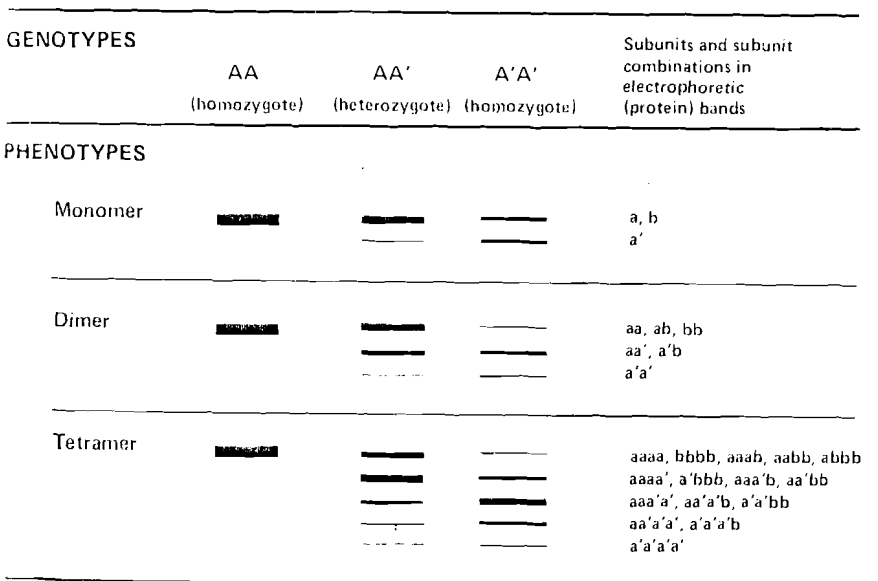

Fig. 2.7 Electrophoretic phenotypes when isoloci are expressed. Individuals are homozygous and heterozygous at loci coding for monomeric, dimeric, and tetrameric proteins: one locus is polymorphic (with alleles *A* and *A'* resulting from subunits *a* and *a'*, respectively); and a second locus is monomorphic, coding for a subunit (*b*) with an electrophoretic mobility identical to that of subunit (*a*).

venient to arbitrarily assign the variation to one of the loci. Such assignment gives information on diploid genotypes within a sampling of individuals but would be misleading in comparisons between samplings if different loci were polymorphic in the sampled groups. A gel showing a polymorphism at one of two isoloci encoding a dimeric protein is seen in Fig. 2.8.

The situation becomes more complicated if both loci are polymorphic. Different procedures have been used to assign alleles to one or the other locus (e.g., Imhoff et al. 1980, Gall and Bentley 1981). Accurate characterization of individual diploid genotypes is still precluded through these methods.

Unambiguous genotypic information can be obtained for isoloci whose encoded proteins are synthesized at different levels in different tissues (e.g., Allendorf and Thorgaard 1984). However, such differences have not often been found.

Codominant expression of isoloci can be masked even when it is known which of the two isoloci is the polymorphic one (as is assumed in Fig. 2.7). The problem here, as indicated above, is the potential ambiguity of phenotypic expression for heterozygous individuals (the second column of Fig. 2.7 and individuals 4–6 of Fig. 2.8) and individuals homozygous for the *A′* allele (the third column of Fig. 2.7 and individuals 7–9 of Fig. 2.8). The same number of bands is expressed in both instances, although the expected relative intensities of the bands differ between the two genotypic expressions. The gene dosages differ for the production of the respective bands of the heterozygote and the *A′A′* homozygote (3:1 for *AA′BB* versus 2:2 for *A′A′BB*). The heterozygous genotype is expressed by asymmetrical relative intensities of banding. Expected

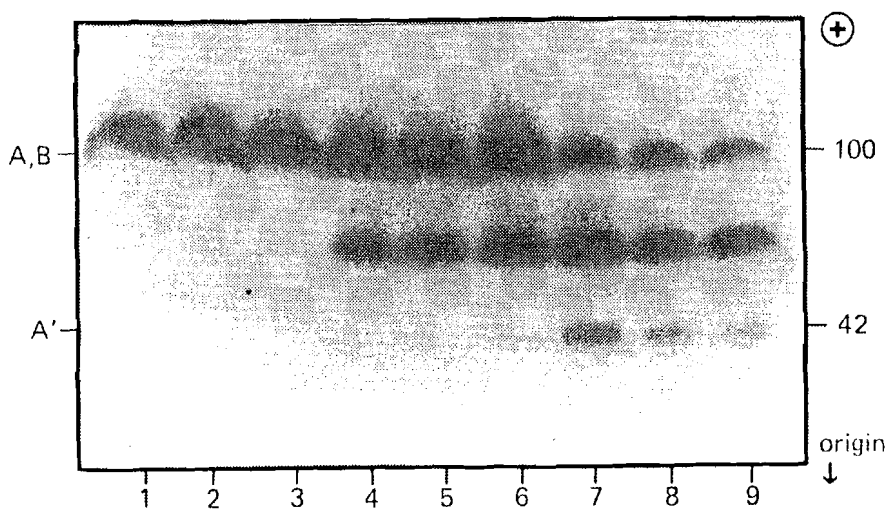

Fig. 2.8 Phenotypes of a two-allele polymorphism for one of two isoloci encoding a dimeric enzyme, isocitrate dehydrogenase, from livers of nine rainbow trout. Note that the same bands are expressed in individuals having single doses (4–6) and double doses (7–9) of the varying allele; only their relative intensities differ.

intensities based on binomial expansions are 3:1, 9:6:1, and 81:108:54:12:1 for the monomeric, dimeric, and tetrameric proteins, respectively. On the other hand, the homozygous *A′A′* genotype has equivalent production of electrophoretically distinguishable subunits; and the respective relative intensities of bands for monomeric, dimeric, and tetrameric proteins are symmetrical 1:1, 1:2:1, and 1:4:6:4:1 ratios.

These differing expected intensities can be used to differentiate between homozygous and heterozygous individuals expressing the same bands. However, in practice, such distinctions are sometimes difficult or impossible to make. Insensitivities of the electrophoretic procedures coupled with possibly different activities or levels of synthesis of allelic products may prevent identifying genotypes among isoallelic phenotypes (e.g., Utter and Hodgins 1972, Allendorf et al. 1976).

Inactive, or *null* alleles, which result in no active protein being detected electrophoretically (e.g., Lim and Bailey 1977), are another category of variants whose genotypes are often difficult or impossible to determine from electrophoretic phenotypes. The expressions of genotypes having no null alleles and those heterozygous for a null allele are usually ambiguous and distinguishable only on the basis of different gene dosages and consequently differing intensities of the same banding patterns.

Heterozygous genotypes for null alleles are particularly difficult to detect when only a single locus is expressed. In such instances, the only clue to the correct genotype is a reduced intensity of the single band. The existence of the null allele is usually verified by the absence of any electrophoretic banding from individuals with homozygous genotypes for the null allele (see Utter et al. 1984). Such quantitative differences, particularly in the absence of individuals that are homozygous for the null allele, usually cannot be identified reliably. Heterozygotes for null alleles are more readily detected for loci encoding multimeric proteins whose expressions include interactions of subunits with those of the same protein encoded by other loci. The reduced synthesis of subunits caused by the null allele results in reduced intensities of multiple bands, providing more visual clues for genotypic recognition. However, correct identification of the heterozygous nature of such individuals by visual observation is still difficult (see Stoneking et al. 1981) and may require that reduced activity be verified by measuring the intensities of the banding patterns or the protein activities (e.g., Allendorf et al. 1984). Such measurements are also usually required to differentiate between a null allele and an allele (isoallele) encoding a homomeric protein of the same mobility as the homomer of the second locus.

Exceptions to the models for molecular aggregation outlined above are also known. A notable example is the electrophoretic expression of the dimeric enzyme creatine kinase extracted from the skeletal muscle of teleost fishes. These banding patterns do not reflect the subunit aggregations expected from a dimeric protein; in fact, they show no heteromeric bands at all. Fisher and

Whitt (1978) and Utter et al. (1979) have provided molecular and genetic interpretations of this anomalous electrophoretic expression.

Genotypic Nomenclature

The material introduced to this point has involved not more than two loci and, in most instances, not more than two alleles at each locus. A particular diploid species has thousands of loci, and a particular locus often has more than two alleles (although an individual of course, can express only one or two of them). The electrophoretic expressions of a monomeric protein in which four alleles are segregating are pictured in Fig. 2.9. The designations for loci and alleles require more explicit designations when multiple loci and alleles are simultaneously considered. A useful convention of allelic nomenclature that has been adopted by many workers is one outlined by Allendorf and Utter (1979), in which loci are identified by italicized lettered abbreviations reflecting the name of the protein (e.g., *LDH* for lactate dehydrogenase). If there is more than one locus coding for that protein in the species in question, a hyphenated number is added to the locus abbreviation (e.g., *LDH-1* and *LDH-2);* the sequence of these numbers starts with the least anodal homomeric band. Alleles are also numerically identified on the basis of electrophoretic mobilities of homomeric proteins. One allele is designated 100—(e.g., *LDH − 2(100)*—and other alleles are designated according to the percent mobility of their homomeric proteins relative to the distance migrated by the homomeric protein of the *100* allele. For instance, an allelic designation of *LDH − 2(80)* would indicate that the homomer encoded by this allele migrates 80% of the distance of the homomeric protein of the *LDH − 2(100)* allele.

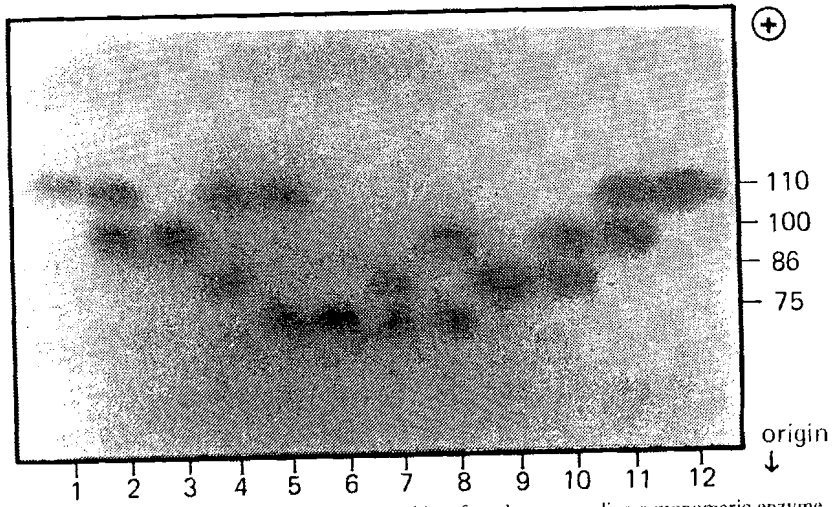

Fig. 2.9 The phenotypes of a four-allele polymorphism for a locus encoding a monomeric enzyme, aconitate hydratase, from livers of 12 chinook salmon.

STRENGTHS AND LIMITATIONS OF ELECTROPHORETIC DATA FOR STUDYING PROTEIN LOCI

The principles outlined above for obtaining genotypic data directly 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 to detect allelic variation is enhanced by the large volumes of data that can be collected with a given amount of effort. Protein extracts can be prepared easily. Many samples can be run on a single gel, and multiple slices can be made from a gel, with each slice stained for different proteins to reveal different loci. For instance, a trained worker can run six gels per day, 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 sliced into six or more slabs and each slab can be stained for a different type of protein. Usually data from more than six loci per individual can be obtained because of the expression of more than one locus for particular types of proteins.

There are limitations to the information that can be obtained by electrophoresis at protein coding loci. 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 electrophoresis are indirect reflections of the actual base substitutions in 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 whose alleles are inferred by phenotypes (see Allendorf 1977). The term *electromorph* (e.g., King and Ohta 1975) has been introduced to deal with this potential genetic heterogeneity within allelic classes. The distinction between alleles that are alike in state and identical by descent is fundamental in measurements of inbreeding (see Chapter 3). Although it is useful to treat homozygous phenotypes as expressions of the same allele, the possibility of unrevealed genetic heterogeneity must be kept in mind.

All electrophoretic expressions described to this point have had a genetic basis. It soon becomes apparent to any person collecting genetic data by electrophoresis that electrophoretic phenotypes can be affected by influences other than the individual genotypes. Electrophoretic expressions of proteins can be strongly modified by length and conditions of storage. A common reflection of such variables is extra bands expressed anodal or cathodal to the primary

band of a particular protein. We have found that collection on dry ice (−80°C), storage at similarly low temperatures, and analysis within a few weeks of collection provide optimal conditions for minimizing the occurrence of these "shadow bands," or "conformational isozymes." These artifacts are more of a problem for some proteins, e.g., glucosephosphate isomerase, adenosine deaminase, and phosphoglucomutase , than for others (see Fig. 2.6). Harris and Hopkinson (1975) describe shadow bands and the use of thiol reagents for altering or reducing these spurious bands.

Shadow bands can present problems in recording genetic information from observed phenotypes on a gel. Several points should be kept in mind for accurate interpretations of gels. If a homomer of one allele is accompanied by a shadow band, homomers of each allele for the same locus will probably be accompanied by its own shadow band, in the same direction (cathodal or anodal) and at the same distance from the homomeric bands. It is therefore understandable why a monomer such as phosphoglucomutase can have heterozygous genotypes from a single locus expressing a phenotype of four bands (e.g., Winans 1980). Often, additional shadow bands occur beyond the initial shadow band, again with the same direction and spacing. They usually occur in progressively reduced intensities forming a serial pattern.

Knowing the subunit structure of the protein and, therefore, the expected number of bands and their relative intensities in the heterozygous phenotypes, helps in accurate gel interpretation. In gels in which more than one locus is expressed for a particular multimeric protein, heteromeric bands and shadow bands make correct gel interpretation even more difficult. Correct gel interpretation may initially be possible only when the frequency of a particular allele is high enough to detect a homozygote for that allele. Alternatively, such homozygous individuals may be generated if inheritance studies are feasible.

Criteria for Allelic Variation: Importance of Inheritance Studies

The ultimate test for an allelic basis of an electrophoretic variant is through inheritance studies. It is highly likely that progeny from crosses between parents expressing the phenotypes shown in Figs. 2.3 and 2.5 will verify the inferred parental genotypes through Mendelian ratios. These ratios are readily calculated from a "Punnett square," which plots the expected genotypic ratios of the progeny from the gametic contribution of each parent. Since two heterozygous parents produce two types of gametes, *A* and *A′*, a Punnett square for a cross of these parents would look like the following:

Gametes of parent 2 \ Gametes of parent 1	A	A′
A	AA	AA′
A′	A′A	A′A′

Four different genotypes are produced in the progeny. However, since *A′A* equals *AA′*, the predicted genotypic ratio among the progeny would be 1*AA*:2*AA′*:1*A′A′*. Expected genotypic ratios from other crosses (e.g., a homozygote and a heterozygote) are calculated similarly with a Punnett square. Genotypic proportions conforming to these expected ratios validate the Mendelian inheritance of the observed variation. Statistically significant deviations from these genotypic ratios suggest, among other thing, nongenetic origins of observed variations.

Understanding the heritable basis of some electrophoretic phenotypes is sometimes impossible without inheritance data, particularly in species, such as salmonids, with a high frequency of isoloci. However, it is frequently impossible or impractical to carry out breeding studies, and alternative criteria for establishing an allelic basis to variability must be met in such cases. A demonstrated genetic basis through inheritance studies for a presumed homologous variant in a closely related species is a strong criterion for an allelic basis of a particular electrophoretic variant. Extensive inheritance studies such as those of May (1980) in salmonids and Kornfield et al. (1981b) in herring provide valuable information to verify the simple inheritance of similar phenotypic variations among closely related species. Another good criterion is consistent expression of a variant electrophoretic pattern in different tissues within an individual where the same locus is expressed. Criteria relating banding patterns to the known subunit structure of a protein have been mentioned above. A Mendelian basis can usually be determined by meeting such criteria (see Allendorf and Utter 1979).

The electrophoretic expressions of some proteins do not conform to the phenotypes that are expected based on the known subunit structure of the protein. For example, heteromeric bands are not observed for the dimeric enzyme creatine kinase in the skeletal muscle of teleost fishes (Ferris and Whitt 1978). These unusual expressions, when coupled with the general absence of homozygotes for infrequent alleles (this absence is explained later), can only be interpreted through inheritance studies. In our own studies, phenotypic expressions of individuals homozygous for an infrequent allele that are generated from crosses of heterozygous parents have clarified the genetic basis for phenotypic variation of muscle creatine kinase (mentioned above) and adenosine deaminase (Kobayashi et al. 1984) in salmonids. The studies of adenosine deaminase also clarified the nature of artifact bands that had previously confounded interpretations of electrophoretic phenotypes. Of course, the Mendelian ratios observed in these studies further confirmed the heritable basis of these phenotypes.

Inheritance studies are needed more often for verification of null and monomeric protein alleles than for variants of multimeric proteins such as those typified in Fig. 2.5. The heteromeric bands of multimeric proteins aid in distinguishing true genetic variation from banding patterns that can arise from nongenetic causes such as storage and bacterial contamination. Our own studies indicate that artifact bands of monomeric proteins are more readily confused with

true allelic variation. Particularly deceptive are the occasional second bands seen above or below a band that appears to be monomorphic for other individuals. In such cases it is uncertain whether the phenotype reflects a heterozygous individual for an infrequent allele or a shadow band. As indicated above, heterozygous individuals for null alleles of loci coding for monomeric proteins or multimeric proteins that are singly expressed in a particular tissue are difficult to distinguish from expressions of genotypes lacking the null allele. The null allele would be expressed only as heterozygotes in most samplings from populations where the allele is infrequent. In such instances, the null homozygotes needed to verify the existence of the null allele would require intentional mating of two heterozygous individuals.

Inheritance data are also necessary to study *linkage* relationships among loci, i.e., the occurrence of two loci on the same chromosome (e.g., May et al. 1979, Morizot and Siciliano 1982). The extensive gene duplications of salmonids have resulted in complicated inheritance patterns that have only recently been clarified through inheritance studies involving allelic variants at many protein loci; they are reviewed in Wright et al. (1983) and Allendorf and Thorgaard (1984).

Describing Populations Through Information from Individual Genotypes

The principles outlined above can now be extended to describe a sample in terms of genotypic and allelic frequencies. Let us use the electrophoretic results pictured in Fig. 2.2 as an example of 10 individuals subjected to electrophoresis and stained for a locus segregating for two alleles in this sample. We arbitrarily designate the allele encoding the more anodal (or faster) protein as A and the other allele A'. Among the 10 individuals, 1 has the genotype AA, 6 have AA', and 3 have $A'A'$. The genotype frequency, then, is 1:6:3.

The allelic frequencies are calculated by counting the number of A and A' alleles. The 10 diploid individuals represent 20 alleles. The 1 AA homozygous individual represents 2 A alleles, and the 6 AA' heterozygous individuals each contribute 1 A allele. Thus, the number of A alleles in this sample is $2 + 6 = 8$; and the frequency of the A allele is $8/20 = 0.40$. Similarly, the frequency of the A' allele is $12/20 = 0.60$.

These genotype frequencies (1:6:3) and allele frequencies ($p = 0.40$, $q = 0.60$) can be compared with genotype and allele frequencies of other samples for this locus. The more distinct the differences between two samples, the greater the "genetic distance" between them. Comparisons among samples are usually based on data from several loci. The amount of genetic distance is then averaged over all loci (see Chapters 4 and 8). Such genotype and allele frequencies, usually calculated for many loci (e.g., 30 or more), are the basic units of information for genetically describing a particular sample of individuals, and for making genetic comparisons between this sample and other samples (see Chapters 5 and 9).

Usually larger sample sizes than 10 individuals (and 20 genes) are required for accurate estimates of allele frequencies of a particular population because of the low precision of estimates from small sample sizes. For reasonable precision of estimates, a minimum of from 50 to 100 individuals is commonly required. For instance, approximate 95% confidence intervals on the estimates of allele frequencies from Fig. 2.2 with 10 samples are ± 0.23. For sample sizes of 50 and 100 the intervals are reduced to ± 0.10 and ± 0.07, respectively.

In addition to permitting genetic comparisons among groups, allele and genotypic data are extremely useful for genetically characterizing a sample of individuals. A common statistic of genetic variability is the frequency of heterozygotes which can be estimated either directly from counting heterozygous individuals over all loci examined or indirectly from allele frequencies (assuming Hardy-Weinberg genotypic proportions; see below). Two other estimates are the percentage of loci that are polymorphic and the average number of alleles per locus.

The Hardy-Weinberg Law (presented in all introductory texts of general and population genetics) is a particularly useful and broadly applied test for the random 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') having respective frequencies of $p(A)$ and $p(A')$ this expansion is

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

where the expected proportions of the homozygous (AA and $A'A'$) and the heterozygous (AA') genotypes are, respectively, the first, third, and second terms of the expanded expression. The Hardy-Weinberg Law can be extended to more than two alleles, and to two or more loci (see Chapters 4 and 7).

Genotypic proportions predicted by the Hardy-Weinberg Law provide a valuable first approximation for expectations in samplings of individuals. For instance, it has been stated earlier in this chapter that homozygous genotypes for a particular allele are not expected to occur in samplings of individuals when the frequency of that allele is low in the sampled population. This expectation becomes apparent from the binomial proportions of an allele occurring at low frequencies. Assume that two alleles (A and A') are present in a population for a protein locus, and 1 heterozygous individual (AA') and 49 homozygous individuals (AA) are seen out of a sample of 50 individuals following electrophoresis and staining for that locus. The frequency of the A' allele $p(A')$ in this sample is $1/100 = 0.01$. Therefore, the probability of an $A'A'$ genotype is

$$p(A')^2 = 0.0001,$$

or 1 in 10,000. Thus the $p(A')$ would have to be about 0.14 before a single $A'A'$ homozygote is expected in a sample of 50 individuals.

Deviations from expected Hardy-Weinberg proportions may result from forces including mutation, migration, selection, and genetic drift (chance fluctuations in allele frequencies operating particularly in small populations). In the absence of such forces, allele and genotype frequencies remain constant over successive generations. Data comparing the observed and expected genotypic frequencies provide valuable insights concerning the operation of such forces within the group of individuals from which a sample was drawn, as do data from repeated samplings of a group at different times.

CONCLUDING OBSERVATIONS

The material presented in this chapter has reviewed the fundamentals of Mendelian genetics that relate to electrophoretic procedures. The collection and interpretation of genotypic data by electrophoresis is currently the primary means for measuring genetic variability within and among species of diploid organisms. Genotypic data from one or more loci are the basis of much of the material presented in subsequent chapters. Among other things, such data permit quantitative estimates of the amount of genetic variation in a sample. Tests and analyses of genotypic distributions in a sample may give information concerning its genetic variation and insight into past and present actions of evolutionary processes (cf. Chapters 4 and 8). Allelic frequencies for many loci estimated from genotypic data collected for two or more populations permit quantitative estimates of the amounts and distributions of genetic variation between and among populations. Questions concerning relative levels of genetic variability and relationships among populations can be answered.

In spite of the unquestionable power of electrophoresis to reveal genetic variation, it must be kept in mind that an electrophoretic sample of 100 loci still represents substantially less than 1% of the total number of genes of a particular diploid organism (Crow 1976). It must also be remembered that electrophoresis detects only a part of the genetic variation of the loci studied. 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.

The differences detected by electrophoresis of proteins encoded by different alleles at the same locus appear to have very little or no effect on the fitness of the individual (see Kimura 1968, Nei 1983). This situation is a disappointment to those who had envisioned electrophoretically detected alleles as "useful genes" for breeding programs and assumed that many such genes could be directly related to fitness (see Robertson 1972). However, the general absence of phenotypic effects on fitness of most allelic proteins enhances the value of this variation as a more or less neutral genetic marker. 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. The value of this information is explored in many of the following chapters of this book.

Despite valuable yet largely preliminary sketches revealed by electrophoresis, considerable genetic variation remains to be detected. New procedures are continually being developed for examining previously unstudied proteins. Previously undetected alleles have also been revealed from proteins that are commonly studied through application of more refined techniques, such as modifications of buffer and gel concentrations and testing for different thermal stabilities of allelic proteins (Singh et al. 1976, Coyne et al. 1978, Coyne 1982).

Procedures for directly examining the genes are emerging as tools for a potentially broader and deeper examination of genetic variation than is possible with methods of protein electrophoresis. Nucleotide sequencing of nuclear genes provides the ultimate information on genetic variation. Methods are tedious and data (see Nei 1983) are as yet sparse and extremely complex. Nevertheless, such information will inevitably become more readily and widely collected.

Restriction enzyme analyses of DNA are seeing accelerated application, particularly for the small but accessible and informative mitochondrial genome. Applications and advantages of these procedures for studies of fish populations are described by Ferris and Berg in Chapter 11 and Gyllensten and Wilson in Chapter 12. Restriction enzyme studies of nuclear DNA that are currently in progress are also yielding promising preliminary results (Bruce J. Turner, University of Virginia, personal communication).

Until recently, studies of fish chromosomes have been largely unproductive in identifying Mendelian variants within species relative to electrophoretic methods (Ihssen et al. 1981). However, new procedures and refined techniques that permit more detailed examinations of chromosome morphology (see Chapter 13) indicate a previously unrecognized potential of cytogenetic studies for identifying Mendelian variations in fish populations.

These procedures appear certain to become and remain valuable tools for fish population genetics. It seems equally certain that electrophoresis will remain a leading procedure because it can readily generate large volumes of reliable genotypic and allele frequency data.