2
How Genes Vary in Fish Populations

A.J. Gharrett
University of Alaska Fairbanks, School of Fisheries and
Ocean Sciences, Fisheries Division, Juneau, Alaska

This is the second in a series of articles that considers the role genetics plays in conservation and management of our fisheries resources. In the first chapter, we examined the basic genetics of inheritance, Mendel’s laws. In this chapter we look at the genetics of populations, because most of the genetics applications that pertain to conservation and management of fish populations involve the behavior of genes (alleles) in a population and the genetic differences among populations. Of course, the inheritance of alleles and their individual expression follow Mendel’s laws, but now the population is the focus of our interest, rather than progeny from specific crosses. In addition, many of the traits important in survival and adaptation to wild environments, and to improvements in aquacultural applications, result from the combined expression of multiple loci. Traits such as size, fecundity, and thermal tolerance do not result from the expression of single loci, and their study requires some additional approaches.
Gene pool

The term **gene pool** refers to the total aggregation of genes in a population. A gene pool envisions a population as a set of haploid gametes (sperm and eggs) that can unite (fertilization) at random to form diploid individuals. The gene pool concept assumes that random mating occurs in the population.

However, theoretical studies have shown that the random mating assumption does not need to be rigidly adhered to and that a gene pool is a very useful simplification, which accurately describes the genetic composition and genetic dynamics of a population for most purposes. In the gene pool of a population, we look at every individual and count the types of alleles that exist at a locus. For simplicity, we examine one locus at a time. We quantify and describe populations in terms of their allele frequencies, which are the relative proportions of each type of allele at a locus. For example, allele \( a \) may have a frequency of 0.3 and allele \( A \) has a frequency of 0.7. These are referred to as allele frequencies. The allele frequencies at a locus add to one. Because it is impractical to examine every individual in most populations, allele frequencies are usually estimated by taking a sample of individuals from a population and counting the numbers of each type of allele that they carry. Larger samples produce more accurate estimates of allele frequencies.

We will see below that most population genetics processes act by directly altering allele frequencies. For those processes, the gene pool concept works well. We will also see that the gene pool concept may not work simply for quantifying selection, because selection acts on phenotypes and alters phenotypic frequencies, which in turn changes allele frequencies indirectly.

Sources of variation and genetic change in populations

The frequencies of alleles at a locus can be altered by several processes. The common processes, sometimes referred to as “forces,” are **mutation**, **selection**, **gene flow** (= migration), and **random drift**. A simple definition of evolution is a change in allele frequencies in a population. The change can result from the action of any of those forces individually or together.

All variation ultimately arises from mutations

![Mutation](image)

**Gene pool before mutation**  
**Gene pool after mutation**

The paradox is:  
Mutations accumulate very slowly:  
Geologic time, not manager’s time.

**Figure 1.** Mutation alters the DNA sequence in an allele, thereby converting it to a different allele. The mutation rate (\( \mu \)) is often quantified as the proportion of alleles of type \( A \) that are altered to any different allele (\( a \) in the figure) each generation. In this cartoon, the mutation rate is one in 10 (10%). Natural mutation rates are often on the order of one in 10,000 to one in 1,000,000.

**Mutation**

Mutations are changes in the genetic instructions (DNA sequence) that can be passed from generation to generation. Mutation changes an allele to a different, often previously unobserved, allele (Figure 1). Mutation is the ultimate source of genetic variation for every species. Because of the long times that are required for mutations to accumulate, however, the importance of mutation is not always obvious. In fact, all of the diversity that we observe among species and much of the diversity that we see within species derives from mutations that occurred in the past, sometimes the very distant past. It is mind-boggling to realize that time frames that define our conservation and management concerns are mere fractions of a blink of an eye in comparison with the time required for mutations to accumulate and define the genetic nature of species and even populations within a species. Realistically, most mutations are not beneficial and the rare neutral and favorable mutants accumulate too slowly to be of consequence in a manager’s time frame (sidebar 1). Even in the extreme, such as a leak from a nuclear plant, accumulation of deleterious alleles would be a trivial concern relative to damage sustained from irradiation!
Selection

Most people who have been exposed to Darwin’s ideas on natural selection think of selection as an important force, if not the governing force. Note, however, that the genetic variation must exist (mutation) before selection can play a role. Selection is any force that alters the genetic composition of a population by differentially influencing reproduction or survival. Warm water could reduce the fertility of some individuals more than others. In that instance, temperature would be the selective force. The elements of the natural environment, such as stream temperatures or flows, can change according to year-to-year fluctuations or global changes; as the environment fluctuates or changes, the way in which populations respond genetically also change.

Selection can occur either inadvertently or purposefully in culture situations. Artificial selection—breeding—is the primary reason that agriculture developed so rapidly during the last century. Strong selection (which causes large changes in allele frequencies) can change allele frequencies in very short times. In contrast, weak selection (only small changes in allele frequencies) may be scarcely noticeable, and even modest levels of selection can be very difficult to detect in a manager’s time frame. When one thinks about selection, it is important to realize that an individual either contributes to the next generation or it doesn’t (e.g., genetic life or death). Evolutionary or genetic fitness (more commonly referred to as just plain fitness) is a measure of an individual’s contribution of genes to the next generation. When we say individual, we actually mean the particular phenotype. If the biggest, toughest, meanest son-of-a-gun around has an accident that makes him sterile before puberty, his fitness is zero. Consequently, it is populations and not individuals that evolve; populations not individuals experience changes in their allele frequencies in response to selection pressures.
What Does Genetics Have to Do with It?

Selection acts on phenotypes and in some instances it will favor or select against both AA and Aa genotypes (A is dominant and both genotypes have the same phenotype relative to fitness), as compared to the aa genotype (Figure 2). The connection between altering phenotypic frequencies (which are based on genotypes of two alleles) and allelic frequencies is not simple, and the gene pool concept does not directly apply. One of the reasons low levels of some genetic diseases, like phenylketonuria (PKU), persist in the human population is that the recessive alleles responsible for them get weeded out only in homozygotes (aa). Heterozygous individuals (Aa) pass both alleles to the next generation because their phenotype is unaffected, even though they carry the a allele.

A particular allele may behave differently in different environments. But it is very difficult to measure the effects of different segments of the life history of a fish or population, because we cannot usually monitor them throughout their lives. For example, Pacific salmon experience a diversity of environments: in the gravel as embryos and newly hatched individuals, possibly in freshwater streams or lakes as young juveniles, in estuaries as young fish, and on the high seas as growing fish. However, we only see them early in their lives and not again until they return to spawn.

Gene flow

Gene flow or migration describes the movement of genes from one population to another (Figure 3). In salmon, gene flow results from strays that do not home accurately, whether by accident or design. However, only the strays that successfully produce offspring in the new stream contribute to that population. What this means is that just because straying has been documented does not mean that gene flow has occurred because genes from the stray may not have been passed to the progeny. There is evidence that in some instances the genes of translocated fish do not enter the recipient gene pool in substantial numbers, but it would be folly to count on such failures. The effects of gene flow can accrue very rapidly. If translocated fish can successfully interbreed with native fish, under some situations
their genes can swamp the native population in a very short time. Is this an issue? Stay tuned for other chapters.

**Random drift**

Random drift is the term used to describe how the gene pool of a population changes from generation to generation. As its name indicates, random drift is a random process. Random drift occurs because the genes that are passed between generations are a subset of the parental gene pool—winners of the genetics lottery—which is often sampled nearly at random; and only the complete set of alleles in the population represents the gene pool without error. The gametes that successfully produce the next generation can be viewed as a sample of alleles from the gene pool. Because the sample of alleles does not include every gene in the gene pool, the offspring may not perfectly represent the gene pool of the previous generation. For example, if we flip a fair coin, we expect 50% of the tosses to be heads and 50% to be tails. If we tossed the coin 1,000 times, on average we would see close to 500 heads, but if we tossed it only 10 times, we might reasonably expect to see any number from 3 to 7 heads. The same process occurs in the reproduction process, during which alleles are sampled from the population. Because samples may not perfectly represent their source, the allele frequencies in small populations vary randomly between generations. Over a sufficiently long period, small populations may actually lose genetic variability as a result of repeatedly occurring random drift. A short-term dip in population size can alter allele frequencies for a long time (in a manager’s time frame) even if the population size expands. These dips in number are referred to as **bottlenecks** (Figure 4).

If a very small number of individuals initiates a new population, for example a hatchery, the allele frequencies may differ substantially from the donor population. This is referred to as a **founder effect**. In Lancaster County, Pennsylvania, the Amish, a religious sect that colonized North America more than 200 years ago, have a high incidence of a recessive disorder called six-fingered dwarfism. The persistence results because there is little gene flow into the population.

**Biochemical genetic markers versus quantitative or metric traits**

The focus of most population genetics studies is the genetic variation that occurs within and between populations. In fact, the biological function of loci at which the variation exists is often irrelevant to
What Does Genetics Have to Do with It?

Possible consequences of genetic drift

Bottleneck

Recovery

No change

Rare type

Figure 4. Random drift generates random fluctuations in allele frequencies. In this figure, the possible effects of a one-time bottleneck are illustrated. The direction of change is unpredictable. However, the average magnitude of change is related to the number of individuals that exist when the population is at its minimum size. Severe population declines usually cause more severe changes than moderate declines. Loss of relatively uncommon alleles is symptomatic of population declines, whether abrupt and severe or mild but chronic.
geneticists because they are using the differences observed as indices or markers to evaluate genetic structure. In fact, the variation at many of the loci that geneticists use is not significantly influenced by natural selection. Such variation is referred to as **neutral** variation. Most population geneticists study biochemical traits that can be easily detected. Alleles of these loci in aggregate often serve as **markers** (like natural tattoos) that can be used to identify populations. Be careful, though—geneticists have not yet discovered markers that can unequivocally track the natal origin of each single fish.

Although several different methods are used to obtain data, most of the methods directly or indirectly detect differences in the DNA sequence (the actual instructions in the blueprints) of target genes or of the protein products that those genes specify. What makes them target genes? Either through happenstance or deliberate searching, geneticists have learned that the target genes have more than one detectable form (i.e., allele). Recall that the ABO blood group has three alleles (A, B, and o), and some loci may have hundreds of alleles. Of course each individual can carry only two of the many available alleles if they are diploid, but the population can have many. For example at locus A, if Jim has an A₁A₂ genotype, Mary has an A₃A₄ genotype, and Lou has an A₁A₁ genotype, we can count five different alleles (A₁, A₂, A₃, A₄, and A₅) at locus A, which are found in just those three individuals. Note that because they are diploid, each individual carries exactly two alleles. The rest of the population may possess even more alleles. Genes that are targeted for study and are in the nucleus ordinarily obey Mendel’s laws and may be referred to as simple Mendelian traits. We will consider these molecular tools in a later chapter.

Even though allele frequencies at most of the loci that are mined for population genetics data are not influenced substantially by natural selection, deleterious genes do exist. Two examples in humans are deleterious genes that cause phenylketonuria (PKU—the inability to metabolize the amino acid phenylalanine) and sickle cell anemia. We see in Figure 2 that afflictions resulting from recessive alleles may occur if an individual carries two abnormal copies of a gene. For these recessive traits, afflicted individuals can have either one or two normal copies of the instructions, but only individuals homozygous for the defective gene are affected. The normal phenotype is dominant. It is interesting to note that the severity of PKU can be reduced by altering the environment, in this case by eliminating phenylalanine from the diet.

Most population geneticists study simple biochemical genetic traits, which are generally not the traits that are important in fish husbandry or, for the most part, to a population adapting to a local environment. In both fish husbandry and adaptation, traits such as growth rate, fecundity, and temperature tolerance are critical. Clearly, such traits are the net result of the expression of alleles at multiple loci as well as the environment in which they are expressed. Unfortunately, it is usually not possible to disentangle the genes for these quantitative or polygenic traits and study them one at a time (see sidebar 2).

**Summary**

The central concept that will be important to us in the following chapters is local adaptation, which is the process by which (the gene pool of) a population evolves to be most productive under the environmental conditions experienced by the population. Different populations (actually their gene pools) can differ because adaptations to different environments (locations) often require different genetic solutions. We briefly described the primary forces that alter the gene pool: mutation, gene flow, selection, and random drift. Finally, we learned that although each gene follows Mendel’s laws, the aggregation of genetic information that is present for many of the traits, particularly the traits that are acted on by local adaptation, is the result of the expression of multiple genes (quantitative traits) and must be dealt with in a different way than simple Mendelian traits.
What Does Genetics Have to Do with It?

Quantitative traits result from the combined action of several to many genes. Such traits are also called polygenic or metric traits. Many of the traits important for local adaptation, such as salinity or temperature tolerance and migration timing, are influenced by the alleles at many different loci. Even though inheritance at each of the contributing loci obeys Mendel’s laws, the influence of individual genes usually cannot be separated from the total. Consequently, we have to resort to statistical methods to dissect the role of genetics in the expression of these complicated traits. Let’s see how that is accomplished.

The problem is that an individual’s genetic instructions are not the only contributor to its phenotype. If we consider a trait such as weight at age, it is clear that the feeding regimen and many other environmental factors play a role in the expression of the trait. In fact we can look at the environment as a black box that filters the genetic instructions (genotype) in producing the phenotype (Figure 5).

This figure shows that the individuals in a population have a distribution of phenotypes (see Figure 5 also), which results from the expression of the genotype in the population’s environmental background. Some of the variation that we see in a phenotype, such as weight, derives from genetic differences between individuals, and some of it is a result of the different environments that they experience. Even identical twins do not experience exactly the same environments. A humped curve is ordinarily what you will see if you plot the abundances of different quantitative traits (such as size) in a population. Most of the individuals are near the middle; extraordinarily large or small (extreme) individuals are relatively rare and at the ends of the distribution. The midpoint of the distribution is the mean. The extent of the spread of phenotypes, which reflects the variation among individuals, is quantified as the variance. We cannot conveniently graph both the mean and variance in the same figure, but we can show the standard deviation, which is the square root of the variance (S.D. = \sqrt{\text{variance}}) (Figure 6).

The goal of quantitative geneticists is to partition (separate into contributing components) the variation of the phenotypic trait. That is, separate the variance of the phenotype into statistical terms (or \( V_p \)) into the genetic variance (\( V_G \)) and environmental variance (\( V_E \)) components:

\[
V_p = V_G + V_E
\]

When we described Mendelian inheritance, we did so in terms of dominant and recessive traits. For many traits, however, both alleles at a locus contribute to a phenotype, although not necessarily equally. We provided carnation color (red, pink, and white) as an example of incomplete dominance. For a quantitative trait, such as size, a homozygote (AA) might have a value of 2 pounds, the other homozygote (aa) might have a value of 1 pound, and the heterozygote (Aa) might be intermediate at 1.5 pounds. In this example, the effects of the alleles are additive; the A allele contributes 1 pound toward the phenotype and the a allele contributes \( \frac{1}{2} \) pound. In addition, alleles at different loci may act synergistically to produce an unexpected result. This is referred to as epistasis. Often the total genetic variation (\( V_G \)) can be further partitioned into one component that quantifies variation due to...
additive effects ($V_A$) and another that quantifies the nonadditive effects ($V_N$), dominance and epistasis (Figure 7).

Geneticists have coined a term that quantifies the relative importance of heredity in expression of a phenotype in a population, the heritability of the trait. There are two different heritabilities. **Heritability in the broad sense** ($H^2 = V_G/V_P$) estimates the proportional influence of all genetic influences, and **heritability in the narrow sense** ($h^2 = V_A/V_P$) estimates the proportion of a phenotype that is attributable only to additive genetic influences. The latter is used to predict the results of selection in agricultural and aquacultural directed selection programs.

To study quantitative traits, geneticists conduct breeding experiments. Two different approaches may be used. In one approach, individuals are selected for their trait (say the largest 10%) and the trait is measured in their offspring (Figure 8). Of course, it is usually necessary to follow a nonselected control population, which serves as a reference for the response to selection, the name of this approach.

Another approach is to make multiple crosses among individuals that were chosen at random from the population. Heritabilities are deduced from these crosses by measuring the phenotypic similarities that related individuals share with each other but do not share with other members of the populations. The proportions of shared alleles are also factored into this computation. For example, large parents generally produce large offspring and small parents generally produce small offspring.

**Figure 7.** The phenotypic variation ($V_P$) of a quantitative trait in a population is the total of the variation (the statistical variance) that is observed for the trait. $V_P$ is the sum of genetic influences ($V_G$) and environmental influences ($V_E$). $V_G$ can be further separated into genetic contributions that reflect additive effects of alleles ($V_A$ = additive variance) and contributions that result from interactions between genes and loci ($V_N$ = nonadditive variance).

**Figure 8.** Response to selection. Individuals in one tail of a phenotypic distribution are used as breeders—e.g., the largest ones. This tail, which is the parents, has a mean of $M_{Par}$. If their offspring have an average phenotype ($M_{Off}$) that exceeds the population mean ($M_{Pop}$)—in size for this example—there is a genetic component for the phenotype. The bigger the deviation from the population mean, the greater the genetic component. Heritability can be estimated: $h^2 = (M_{Off} - M_{Pop})/(M_{Par} - M_{Pop})$. 
(Figure 9). The trait "size" is heritable and the heritability ($h^2$, strength of genetic contribution) can be quantified in the terms shown in Figure 7. Offspring share half of their alleles with each parent (from the parent's gamete). For many traits, the progeny must be tracked to adulthood or marked by family and recovered at maturity in order to obtain comparable measurements of relatives. Breeding experiments are expensive and usually require dedicated facilities, which make them difficult to conduct on wild populations. However, quantitative traits are as a rule far more important to an individual's and population's success than the vast majority of biochemical genetic traits used by population geneticists to study population structure. Two other problems are that heritabilities are population-specific characteristics and when they are estimated, they are usually estimated for a particular environment or generation. Finally, $h^2$ estimates the relative proportion of additive genetic variation, but it does not reveal the actual allelic composition.

The similarity (statistical correlation) between parents and offspring can be demonstrated by a plot of an offspring's value (e.g., length) versus its parent's value (Figure 9). Estimates of the average similarities can be made with regression analysis, a statistical tool that quantifies the relationships between the sizes of parents and offspring. If a trait is heritable, one would also expect to see a relationship between the sizes of siblings, cousins, half-siblings, and so on for any pair of "blood" relatives. Other statistical methods are used to quantify these relationships and extract information about the heritability of traits.

Figure 9. This is a plot of the length of a son on the y-axis against the length of its father (= sire) on the x-axis. There is a linear relationship, although it is not perfect, which indicates that there is a genetic basis to length. If you were expecting the pitch of the line (slope) to be one (an incremental increase in sire size produces the same incremental increase in the size of its son), you would have been wrong. First, only half of the genetic information for size is inherited from the father, the mother provides the other half; and second, environment also contributes to the variation in size of both fathers and sons.