

# Basic Aquaculture Genetics

Jason W. Abernathy,<sup>1</sup> Eric Peatman<sup>1</sup> and Zhanjiang Liu\*

The inheritance of desirable traits in crops and livestock has been the foundation for selective breeding for thousands of years. However, our understanding of genetics, which is the biological basis of heredity and variation among organisms, has deepened dramatically in the last several decades. This new knowledge expands the potential applications of genetics and genome technologies in agricultural practices. In aquaculture settings, current genetic improvement programs focus on selecting superior broodstock, using better breeding practices, increasing sustainability, and minimizing environmental problems. These programs have already led to more efficient, productive and profitable aquaculture systems, but the genomics revolution promises to speed and amplify genetic advances in the near future. This publication describes both basic genetic principles and applied genome technologies that are relevant to today's aquaculture industry.

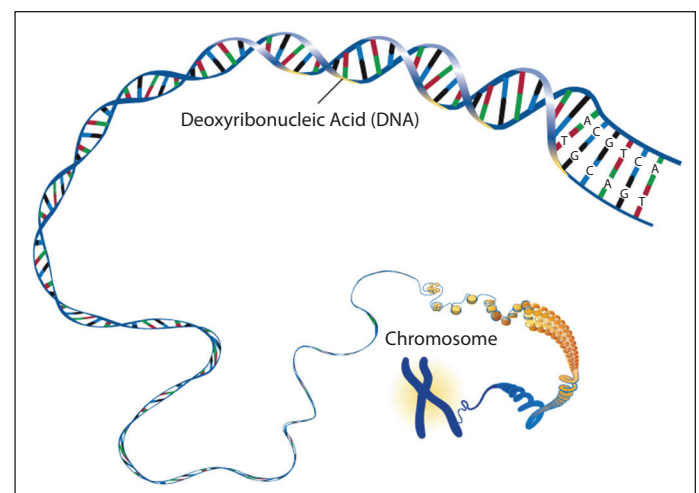
## Introduction to genetics

### *Genes and chromosomes*

The basic unit of inheritance is the gene. Genes contain the biological code for the production of observable traits, or phenotypes, in an organism. Genes are arranged on a large molecule called deoxyribonucleic acid (DNA). DNA consists of subunits called nucleotides (Fig. 1). The four nitrogenous bases that define the four different nucleotides are adenine (A), guanine (G), thymine (T) and cytosine (C). It is the combination of these subunits in a linear arrangement that codes for a

gene. Most of the DNA forms a structure known as the chromosome. Chromosomes are located in all nucleated cells. The total number of chromosomes can differ among species but will generally be constant in the same species. Most aquatic species contain two sets of chromosomes, one inherited from the father and one inherited from the mother. Any organism with two sets of chromosomes is termed diploid, or  $2N$ . We will focus on diploids in our explanations, unless otherwise noted.

The karyotype is the general appearance (size, number, shape) of an organism's chromosomes. There are two types of chromosomes in eukaryotic species (species with nucleated cells): autosomes and sex chromosomes. Autosomes are all chromosomes that are not sex chromosomes. Sex chromosomes usually determine the fish's sex,



**Figure 1.** The physical structure of deoxyribonucleic acid (DNA). DNA is characterized as having a double helix structure with a sugar-phosphate backbone. The nucleotides pair with each other to form the structure. Adenine pairs with thymine, while guanine pairs with cytosine. Courtesy: National Human Genome Research Institute.

<sup>1</sup> The Fish Molecular Genetics and Biotechnology Laboratory, Department of Fisheries and Allied Aquacultures, Program of Cell and Molecular Biosciences, Aquatic Genomics Unit, Auburn University

\*Corresponding author

either male or female. Autosomes are usually designated with a number, while the characteristic notation for sex chromosomes is XX and XY in diploids, where XX are females and XY are males in the XY sex determination system. Many species of fish can have a different sex-determining system. Some may contain only autosomes. Some are hermaphroditic and either change sex as they grow or, in rare cases, possess male and female organs simultaneously. And in some fish species the sex chromosomes have not yet been determined.

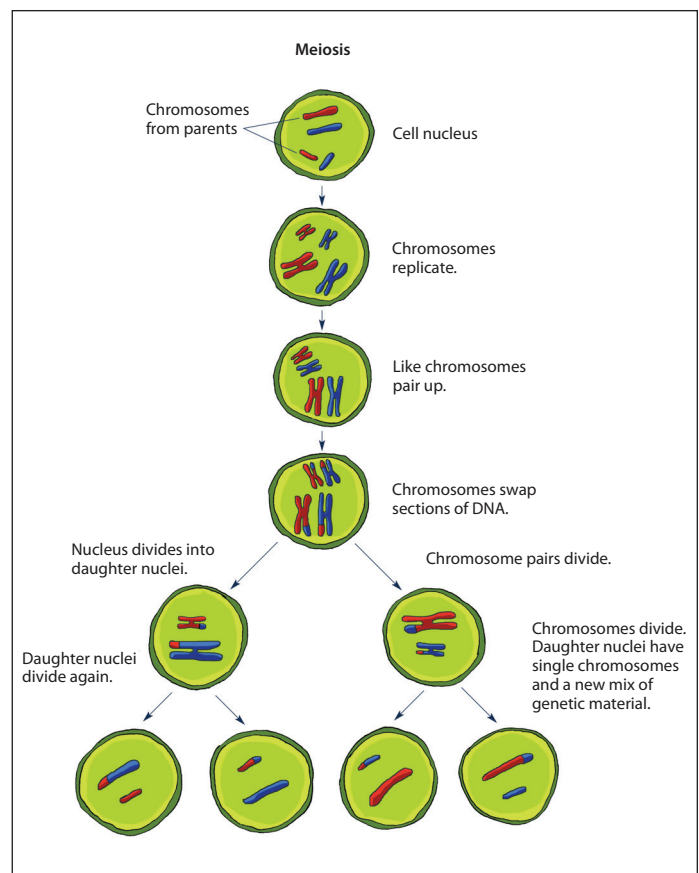
Even though most eukaryotic species are diploid, there are exceptions to the number of chromosome pairs (ploidy) in some aquatic species. Further, the number of chromosome sets can be manipulated in genetic improvement programs. Haploids ( $N$ ) can be created, as well as fish that contain chromosomes from the mother only (gynogens) or from the father only (androgens). Triploids have three sets of chromosomes ( $3N$ ) and tetraploids have four sets of chromosomes ( $4N$ ). In nature, tetraploids are more common than triploids. The channel catfish is diploid ( $2N$ ) and contains 29 pairs of chromosomes ( $2N = 58$ ). The salmonids (salmon and trout) and the common carp are widely thought to be tetraploid ( $4N$ ).

Since chromosomes occur in pairs in diploid organisms, each gene has at least two copies. Each copy of a gene, called an allele, is located at a specific location on each of the sister chromosomes. This location is called the locus of the allele. Two alleles can be identical or can have variations in their DNA sequence. If alleles at a specific locus are identical, the organism is homozygous for that gene. Conversely, if alleles at a specific locus are different, then the organism is heterozygous for that gene. Differences in alleles within an individual can produce genetic variance and thus different phenotypes in a population. The combination of alleles for a given trait is called the genotype.

### Chromosome replication and cell division

There are two forms of cell divisions—mitosis and meiosis. As an organism grows, cells must divide and replicate to increase in number and to replace old or dying cells. Mitosis is the process by which a somatic cell divides to produce two identical daughter cells. Somatic cells are all cells except egg and sperm and the cells that produce them. During mitosis, the entire cell is duplicated, including the chromosomes and all other cellular material. Thus, each daughter cell contains a diploid set of chromosomes.

This process differs from meiosis, which is associated with gamete (egg and sperm) formation (Fig. 2). The eggs and sperm undergo cell division and replication, but produce cells that contain a haploid ( $N$ ) gamete, containing



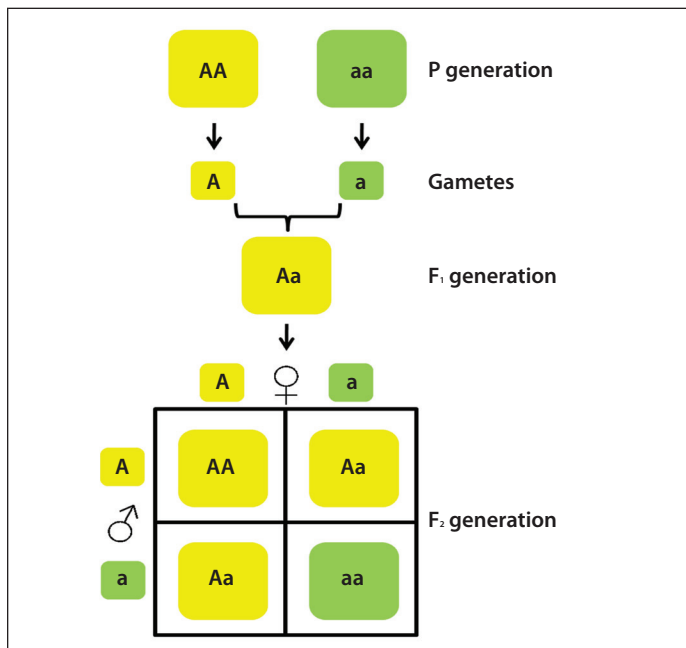
**Figure 2.** The basic principle of meiosis. Courtesy: National Institute of General Medical Sciences.

only one set of each chromosome pair. The process of creating haploid gametes is critical to reproduction in diploid organisms. When an egg ( $N$ ) is fertilized by the sperm ( $N$ ), the correct number of chromosomes ( $2N$ ) is recovered in the zygote. The process of sperm or egg (gamete) generation is called gametogenesis. More specifically, the production of sperm cells is referred to as spermatogenesis, while the production of egg cells is called oogenesis.

In meiosis, at least three important processes occur to produce genetic variability in the sperm and egg: crossing over, segregation, and independent assortment. First, the chromosomes in the gametocyte (precursor cell to an egg or sperm) must undergo replication. After chromosome pairs replicate in diploids, the homologous chromosomes form pairs in bundles of four, called tetrads. At this point, the chromosomes that form tetrads can wrap around each other. The chromosomes may then break apart and pieces from different chromosomes can rejoin. Therefore, genetic material has been swapped from one chromosome to another. A portion of a chromosome that originated from the mother is exchanged between the homologous chromosome from the father, and vice versa. This process of recombination or physical exchange between homolo-

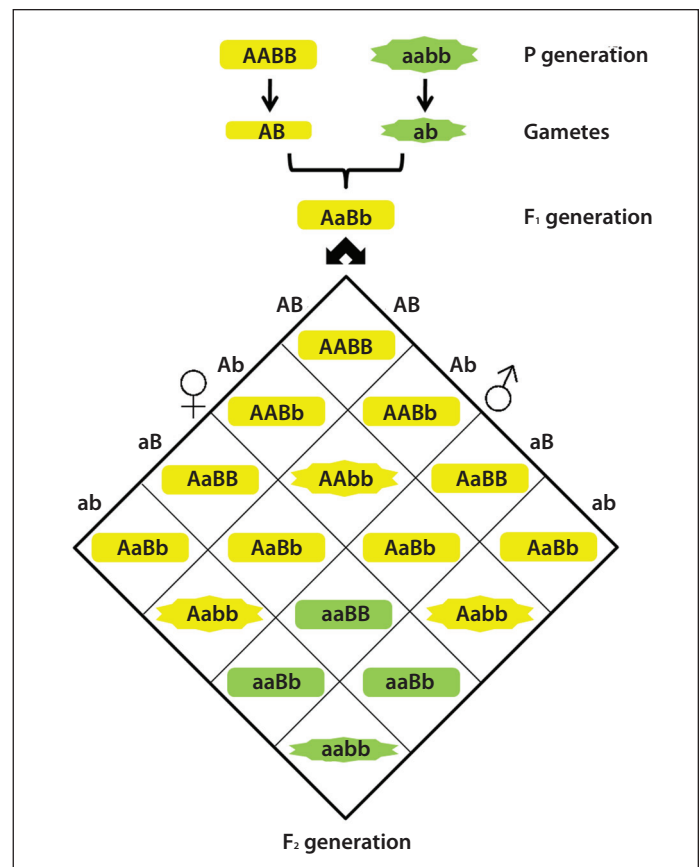
gous chromosomes is called crossing over. This occurs in the cell stage known as the primary gametocyte (primary spermatocyte in males, primary oocyte in females).

The other processes that increase genetic variation include segregation and independent assortment. After chromosome duplication and crossing over in the primary gametocyte, the chromosome complement must be reduced from the diploid state ( $2N$ ) to the haploid state ( $N$ ) in a cell division step for both the egg and sperm. Each chromosome pair will separate, going from a primary gametocyte to a secondary gametocyte. Each secondary gametocyte will contain a single replicated chromosome from each homologous pair. In spermatogenesis, all chromosome pairs separate and one of the chromosomes of each pair goes to a secondary spermatocyte. In oogenesis, all chromosome pairs separate and one of the chromosomes of each pair goes to a secondary oocyte or a polar body. The process by which the duplicated chromosomes separate and form secondary gametocytes is the basis for Mendel's law of segregation (Fig. 3).



**Figure 3.** Mendel's law of segregation. This is the classic example of independent segregation of alleles, using a single observable trait. Gregor Mendel used the observable trait of seed coloration in his plant experiments. In this example, the dominant allele is yellow seed color (A), whereas the recessive allele is green seed color (a). One homozygous dominant (AA) and one homozygous recessive (aa) parent were crossed to produce  $F_1$  progeny that are all heterozygous dominant (Aa) for the trait (and all yellow seeds). A cross between two  $F_1$  progeny (Aa x Aa) produces  $F_2$  progeny with three genotypes (AA, Aa and aa), and two phenotypes (yellow or green seeds). Seed coloration is characterized as either homozygous dominant (AA) yellow, heterozygous dominant (Aa) yellow, or homozygous recessive (aa) green.

Mendel's law of independent assortment states that each gamete receives a random mixture of alleles, one at each locus, from each parent (Fig. 4). The processes of segregation and independent assortment are random in that each secondary gametocyte or polar body receives a random mixture of chromosomes from the mother and the father. The reduction of chromosomes from the primary gametocyte to the secondary gametocyte reduces the number of chromosomes from the diploid complement to the haploid complement. This process ensures that when a sperm ( $N$ ) fertilizes an egg ( $N$ ), the correct complement of chromosomes ( $2N$ ) will be created in the zygote (fertilized egg). The exception to independent assortment of chro-



**Figure 4.** Mendel's law of independent assortment of alleles of different genes. Gregor Mendel also observed the trait producing seed shape, either smooth or rough seeds, along with seed color. In this example, the dominant alleles are yellow (A) and smooth (B) seeds, whereas the recessive alleles are green (a) and rough (b) seeds. The parental cross produces  $F_1$  progeny that are all heterozygous dominant (AaBb) for the traits, and thus yellow smooth seeds. The  $F_2$  generation, produced by a cross of  $F_1$  generation (AaBb x AaBb), yielded progeny with various genotypes and phenotypes. Two new phenotypes appear in the  $F_2$  generation: yellow rough seeds and green smooth seeds. Since the two genes were segregating and assorting independently of each other during meiosis, multiple combinations of alleles (and traits) were produced. Note that independent assortment of genes may not hold true for all combinations of alleles (such as with linked genes).

---

mosomes (and alleles) is linkage. If the genes that control a trait are located on the same chromosome and are close together, they are said to be physically linked. In this case, these genes may be inherited together.

The secondary gametocytes consist of either two secondary spermatocytes (males) or a secondary oocyte and a polar body (females). Each secondary oocyte normally produces only one haploid egg; the rest are polar bodies. In females, after segregation and independent assortment, chromosomes at one end are pinched off with a little surrounding cytoplasm (yolk). This forms the first polar body. Polar body cells are non-functional and are a byproduct of oogenesis. Since the egg is the major site of nourishment to the embryo, a high concentration of cytoplasm is necessary and hence the unequal division of cytoplasm in the oocyte versus the polar body. The first polar body may or may not divide again to produce two small haploid cells. The other daughter cell is the secondary oocyte. The secondary oocyte and first polar body produce one mature egg and one secondary polar body. Secondary oocytes are not stimulated to egg production until after they have contacted the sperm. In males, the two secondary spermatocytes produce four sperm cells. All sperm cells have the haploid number of chromosomes that have been independently assorted and contain equal amounts of cytoplasm.

### **Development**

When an egg and sperm are fused through fertilization, a diploid zygote is produced. This is a single cell; mitosis of the zygote and subsequent daughter cells is responsible for the growth and development of the fish. Fish eggs are mostly composed of yolky cytoplasm. In general, after fertilization cell division begins in a thin, yolk-free region of the egg called the blastodisc. This initial cell division is known as the cleavage phase and the initial cells are termed blastomeres. Blastomere divisions are rapid and cells build upon one another to constitute the blastoderm. These early cells mix and can give rise to a variety of cell and tissue types. After a period of rapid cell cleavage, cell divisions begin to slow and cell movement begins. Blastomeres begin to cluster and segregate throughout the embryo. After the blastoderm has filled about half the yolk, or sooner depending on the species, germ layers begin to form through the process of gastrulation, or cell restructuring. These layers will develop into the tissue and organ systems of the fish. From a genetic standpoint, zygotic gene expression generally begins during the cell division stages.

### **Sex determination**

Sex genes are the main determinants of gender. The sex chromosomes of different fish species may or may

not be morphologically distinct. Many sex chromosomes resemble autosomes, and/or some sex-determining genes may be located on autosomes. In these cases, sex-linked traits must be studied to determine the sex chromosomes or sex genes that are located on autosomes.

The most common sex-determining system is the XY diploid system. This system is the most common in known fish species as well as in humans. Channel catfish have the XY system. This system was so named because the sex chromosomes in humans resemble a Y or an X. In the XY system, the sex chromosomes in females are identical (XX), while those in males are a mix (XY). The Y chromosome is found only in males. Since the pairs in females are the same, these chromosomes are termed homogametic, while in the male they are called heterogametic.

All eggs contain only a single X chromosome. In males, half the sperm population contains a single X chromosome and half contains a single Y chromosome. In the XY system, the heterogametic sex (male) is the one responsible for sex determination in the population. If an egg (X) is fertilized with a sperm carrying the X chromosome, the sex of the offspring will be female (XX). If an egg (X) is fertilized with a sperm carrying the Y chromosome, the offspring will be male (XY). Again, the resulting offspring will be diploid. It is likely that only a portion of each sex chromosome is ultimately responsible for sex determination. Other traits (genes) may be located on the XY chromosomes in many different aquaculture species.

There are several other sex-determining systems in fish besides the XY system. One is the WZ system. It works exactly the opposite to the XY system; a homogametic fish is a male while a heterogametic fish is a female. Blue tilapia is thought to have the WZ sex-determining system. The female blue tilapia contains the different set of sex chromosomes (WZ), while the male blue tilapia contains the same set of sex chromosomes (ZZ). Hence, the W chromosome is the sex-determining chromosome. The sperm (Z) that fertilizes an egg of one chromosome type (Z) will produce males (ZZ), while a sperm (Z) that fertilizes an egg of another chromosome type (W) will produce females (WZ). In other words, sex is controlled by different types of eggs, not different types of sperms as in the XY sex-determination system.

Another system is the ZO sex-determining system. The dwarf gourami fish and a sole fish have been identified as having the ZO system. Here, the females are ZO and the males are ZZ. In these species, the females are the sex-determining species and are the heterogametic sex. The female produces haploid eggs with either the Z chromosome or no sex chromosome at all. If an egg with the Z chromosome (Z) is fertilized by a sperm (Z), then the

---

offspring will be male (ZZ). If an egg has no sex chromosome (O), the resultant fertilization will produce a female offspring (ZO).

The XO system also exists in some species of fish, including a sunfish species. In this system, the males are heterogametic (XO) and determine the sex of the offspring. In a way, the XO sex-determination system is similar to the XY sex-determination system except that in the XO system the males contain just one set of sex chromosomes (X).

Several other sex-determining systems have been discovered in fish. Many of these systems are more complex than the XY system and its variants. These systems have multiple sex chromosomes. One example of a multiplex system is the WXY sex-determining system, common to the platyfish used in aquaria. Here, both males and females can be either homogametic or heterogametic and determine the sex of the offspring. In WXY, the W chromosome acts as a modifying chromosome that can block the male-determining ability of the Y chromosome. Therefore, the XY and YY offspring are males and the XX, WX and WY offspring are females. Other multiplex systems include the  $X_1X_1X_2X_2/X_1X_2Y$ ,  $ZZ/ZW_1W_2$ , and  $XY_1Y_2/XX$  systems. Finally, sex-determining genes also can be located on autosomes. Some aquaculture species do not have any morphologically distinct sex chromosomes. In those species, sex is determined by a number of female or male genes located on specific autosomes.

One important concept to introduce is genetic sex versus phenotypic sex. Fish, being lower vertebrates, have much flexibility in sex organ development, so a phenotypic sex may not reflect the chromosome composition of the fish. A number of factors can influence expression of the sex phenotype. Some of the most common factors are temperature, salinity, population density in tanks, pressure, hormone treatments, radiation and photoperiods. Sex modifications may be induced or may occur naturally in some populations. For instance, hormone treatment (estrogen) of developing males in some fish species can produce phenotypic adult females even though the fish contain the XY chromosomes. Conversely, hormonal treatment can lead to the development of all males even though the fish have female genetic makeup (XX chromosomes). For instance, a common practice in aquaculture is treating tilapia with 17 $\alpha$ -methyltestosterone (MT) to induce all-male populations. As male tilapia grow much faster than females, such hormonal treatment leads to increased yields. In such cases, however, the fish may harbor XY chromosomes (genetic males and phenotypic males as well), or they may harbor XX chromosomes (genetic females, but phenotypic males).

## **Qualitative genetic traits**

Phenotypes can fall under the category of qualitative traits, or traits that can be simply described as one or the other. A trait is qualitative if it can be sorted into one of at least two categories. These traits are not measured over a range, as is the case with quantitative traits. An example of a qualitative trait would be fish pigmentation; the fish is either normally pigmented or albino. Qualitative traits are often the simplest to characterize because they are likely to be controlled by only one or a few genes, unlike quantitative traits. Since an aquaculture species can be categorized by these traits, the population can be described by the ratios of its members with these traits. The number of fish with each trait are simply added up and described, such as 3:1, 1:1, and so on. For example, a population of 20 fish has either a fan tail or a round tail, a qualitative trait. If 15 fish have fan tails and five fish have round tails, the ratio of fan tails to round tails is 3:1. Described another way, one-quarter of the fish population has round tails and three-quarters of the population has fan tails. Qualitative phenotypes are the same in different environments.

A number of factors influence these ratios, including the number of genes needed to produce the phenotype and gene action. The genetic mechanisms described here are typically referred to as classical genetics or Mendelian genetics. Gene action can be characterized as single or multiple genes producing a qualitative phenotype. Classic qualitative traits are dominant or recessive. If a single allele is expressed over the other at the same locus, then the mode of action is termed dominance. Complete dominance mode of action describes the expression of alleles at the same locus where one copy (the dominant allele) masks the effect of the other copy (the recessive allele). The phenotype expressed is termed the dominant phenotype, and the other is the recessive phenotype. When dominant alleles occur, there are three genotypes possible, while only two phenotypes can be produced (Fig. 3). A classic example of complete dominance of a single locus is the gene for albinism in the channel catfish. The dominant genotype that produces normally pigmented catfish can be labeled as AA or Aa, where the capital (A) is the dominant allele and the lower case (a) is the recessive allele. Both AA and Aa fish produce normal pigmentation. Only fish with the complete recessive pigment gene (aa) will be albino. Hence, with dominance there are three combinations of genotypes (AA, Aa, aa), while only two phenotypes are produced—normal (AA or Aa) or albino (aa) coloration. Fish that are homozygous dominant (AA) will obviously be normally pigmented, but so will fish with one dominant (A) and one recessive (a) allele (heterozygous dominant, Aa). With complete dominance, only parents that carry recessive alleles (either Aa or aa) can produce offspring with a recessive trait (Fig. 3).

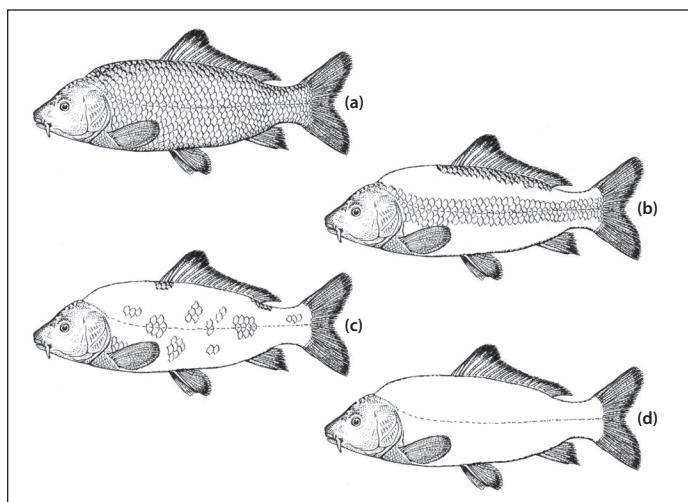
There are two other possibilities concerning qualitative traits—multiple alleles and sex-linked alleles. When alleles at more than one locus control a qualitative phenotype, there are two possibilities of gene action. There are either epistatic effects or non-epistatic effects. Epistatic effects are interactions between genes that can cause modifications or suppressions of phenotypes. Thus, combinations of multiple alleles at different loci can cause different phenotypes than the simple case of either/or patterns.

A classic example of epistasis for a qualitative trait in aquaculture is the scale patterns in the common carp (Fig. 5). Common carp scaling includes wild-scaled, mirror, linear or leather types. Wild-scaled carp have scales all about the fish; mirror carp have scales scattered around the fish; linear carp have scales arranged in a linear array; leather carp have very few scales. These patterns are controlled by genes (S and N) from two loci. One loci (S) determines the degree of scales, either wild-scaled (SS or Ss) or mirror-scaled (ss). The other loci (N) modify these phenotypes in the following manner:

- a. (SS nn or Ss nn); wild-scaled carp
- b. (SS Nn or Ss Nn); linear carp
- c. (ss nn); mirror carp
- d. (ss Nn); leather carp

There is another combination of alleles possible for scale patterns: the homozygous dominant (NN) form of the locus (N). This inheritance is lethal to embryos in the common carp.

Note that the discussions on qualitative traits have concentrated on genes on autosomal chromosomes. However, some phenotypes can be controlled by genes on sex chromosomes as well. The mode of inheritance of a trait may be different when the allele is linked to one of the sex chromosomes.



**Figure 5.** Different phenotypes caused by epistatic effects in common carp: (a) wild-scaled carp, (b) linear carp, (c) mirror carp, and (d) leather carp.

## Quantitative genetic traits

Quantitative traits are phenotypes that have a range of expression. These are traits that can be measured, rather than either/or traits such as normal pigmentation versus albinism. Phenotypes that can be observed and measured may vary in a population. Therefore, genetic programs must be able to properly analyze and understand the traits of interest. Commercially important traits in aquaculture include length, weight, growth rate, feed conversion, oxygen tolerance, percentage of body fat, meat production, disease resistance, and stress resistance, just to name a few. These traits are said to be quantitative because they typically vary among individuals within a population. Quantitative traits are measured using a continuous distribution system and statistics. Such traits are described and reported around their central tendencies, such as average (mean), variance, standard deviation and range. These traits can be controlled by a single gene, but are usually controlled by several to many genes. These traits are also influenced by the environment. Gene expression levels, the environment, and the interaction of the two can play a significant role in the variation of quantitative traits.

Quantitative traits are under a constant variance. Each gene (except linked genes) is segregated and independently assorted. As many quantitative traits are controlled by multiple genes, each of those genes will be segregating and sorting independently as well. Since so many genes are involved and each locus is independently assorted, the potential for a variety of genetic combinations in the offspring is large. Because quantitative traits are also influenced by the environment, the action of both the environment and the multiplex of genes involved produces the distribution of these traits in a population.

## Components of phenotypic expression

In an individual, a quantitative trait is determined by its genes, the environment, and the interactions of the two. This relationship is the sum of any phenotype:

$$P = G + E + GE$$

Where:

P is the phenotypic value for an expressed phenotype

G is the effect of genetics due to a particular phenotype

E is the value for environmental influences

GE is the combined effect of the environment on the genotype of the individual

Genetics and environment interact differently in each individual. An individual may be influenced differently

---

by different environments. Different individuals may respond differently to the same environment.

Importantly, the individual genetic value can be broken down into its principal components. The genetic value (G) can be described as the sum of additive effects (A), dominance effects (D), and epistatic or interaction effects (I) such that:

$$G = A + D + I$$

Additive effects are due to the cumulative effects of alleles at all loci for a given trait. This value is independent of other interactions among alleles or other combinations of alleles, so additive effects are preserved through meiosis and these values are predictably passed from parents to offspring.

Dominance effects are due to the interaction of alleles at each locus. During meiosis, homologous chromosomes are separated and reduced from the diploid to the haploid state. Since gametes are haploid, they can contain no dominance effect; new pairs of alleles form after fertilization to create new and random diploid states. Thus, dominance effects are created in different combinations with each generation.

Epistatic effects are also important in the genetic equation. These effects are caused by the interaction of alleles at different loci. Simply stated, this is the effect of interactions between genes. During meiosis, epistatic effects for most genes are disrupted by segregation and independent assortment (Fig. 4). Therefore, most epistatic effects are recreated with each generation.

Since quantitative traits have a distribution and variance in a population, one way to understand these traits is to analyze their variance and effects. Breeding programs can exploit the genetic variance in a population in order to minimize its effect. The equation for phenotypic effects can be modified from the individual to be relevant to the population. The way to study quantitative traits in a population is to study variance for a trait or traits. Variance is introduced from the environment (VE), from genetics (VG), and from the interaction between genetics and the environment (VGE). Thus, the sum of the phenotypic variance (VP) for any quantitative trait is:

$$V_P = V_G + V_E + V_{GE}$$

Genetic variance is used in breeding programs to assist in the selection of a stock of fish. To study genetic variance, the value must be broken down into its principal components. Genetic variance is the sum of additive genetic variance (VA), dominance genetic variance (VD), and epistatic or interaction genetic variance (VI) such that:

$$V_G = V_A + V_D + V_I$$

The relationships and differences among these genetic variance values, the way they are inherited, and their proportional amounts are important in any breeding program.

### ***Genetic variance and breeding***

Genetic variance is a key component in exploiting the traits of interest in a breeding program. From the equation above, components of genetic variance include epistatic effects, additive effects, and dominance effects. Epistatic effects are rarely used in a breeding program because there so many possible combinations of alleles in this measurement of genetic variance. Most programs focus on additive and dominance genetic variance. The relative contribution of either of these effects on the phenotype of interest determines the best approach to use in a breeding program.

Additive effects are passed down from parent to offspring and dominance effects are created with each generation. If additive effects are large, and thus the variation within a population is large, fish with a desirable trait can be selected and bred. If the additive genetic variance is small, then selecting fish for a specific trait within a population may not produce progeny that express the trait better than the parents. In this case, a recombination of alleles would be the expected approach in the breeding program. More alleles are introduced through an influx of new genetic material, based on trial and error. New strains of fish can be introduced into the breeding program (intraspecific hybridization) or different species bred (interspecific hybridization). A new combination of alleles may produce offspring that harbor the desired combination of phenotypes.

For a selection program to be effective, the amount of additive variance should be determined. The measure of heritability ( $h^2$ ) describes the proportion that additive genetic variance contributes to a phenotype in a population. Heritability describes the percentage of a phenotype that can be inherited in a predicted manner. This value should be reliable since, for a quantitative phenotype, the genotype is not affected by meiosis. These values range from 0 to 1 and are a percentage. A heritability value of 1 suggests that the phenotype observed is 100 percent explained by the additive genetic variance. A heritability value of 0 suggests that additive genetic effects do not contribute to the phenotype. In general, a heritability value greater than 0.2 ( $h^2 = 0.2$ ) suggests that a trait may be reliably exploited in a selection program. If a trait's heritability is lower than  $h^2 = 0.15$ , a selection-based program may not be the answer because dominance genetic variance is more important in this case. Just remember that we are dealing with quantitative traits here. These traits may

---

be affected by the environment and the population, and can vary between generations. Direct selection of traits with low inheritance may be difficult; in such situations, genome-based technologies such as marker-assisted selection may be more applicable.

## Genetic improvement programs

### *Selection*

Selection is a process in which individuals with desired phenotypes for a particular trait are identified and used as future broodstock to produce progeny that are also superior for the trait. Selection programs include mass selection and family selection. In mass selection, the performance of all individuals is compared and selection is based on the performance of each, disregarding the parentage. In family selection, the average performance of families is compared and whole families are selected.

Selection programs are reliable only if the genes responsible for the genetic variation are passed to the offspring. Additive genetic variance is transmitted to offspring in a calculated and reliable manner. Heritability ( $h^2$ ) values must be taken into account. These values are a direct measurement of genetic variance explained by additive genetic effects. The heritability estimates for a trait should be as large as possible to ensure the efficacy of a selection program.

Many traits may be correlated, positively or negatively. For instance, fast growth rate is often correlated with a more efficient feed conversion rate in some species (and the opposite can be true in other cases). Therefore, selection for fast-growing fish also selects for fish with a better feed conversion rate. However, if the traits are negatively correlated, the breeder must be careful because selection for one trait may negatively affect the other. For instance, fast growth may be negatively correlated with reproduction capacity so that selecting for fast-growing fish could reduce reproductive capacity. Therefore, in selection programs for certain traits other important traits must be carefully monitored to determine if there is any correlation among the traits.

### *Hybridization*

If heritability is very low, selection methods will not be the best way to increase performance. To improve a trait in this instance, greater genetic variation for new combinations of alleles is needed. New combinations can be created by mating fish with different genetic histories, a process called hybridization. All offspring are called hybrids. The parents can be of the same species, but different strains (intraspecific hybridization). Or, the parents can be of different species (interspecific hybridization).

Improving a trait using hybridization is done through trial and error. Some of the hybrids may have superior traits and some may not, but hybridization is the best way to increase performance when the heritability for a trait is very low because hybridization creates new combinations of alleles through dominance genetic variance. This process is independent of heritability, so hybridization programs can be used even when heritability is high. Selection and hybridization can both be used to increase fish performance.

### *Intraspecific hybridization*

Intraspecific hybridization can be used to improve fish performance in one of two ways: It can be used to produce a new strain that can undergo selection for a trait, and it can also be terminal with the hybrids being the end product. In general, hybrids have better fitness because of greater genetic variability. This is known as hybrid vigor or enhancement through outbreeding.

To create a new breed, a cross must be made between individuals of two strains with different genetic backgrounds, followed by a selection program to improve performance. A selection program using hybrids usually cannot begin until the second generation of fish ( $F_2$ ) is spawned. This is due largely to the principle of dominance genetic variation; first-generation hybrids ( $F_1$ ) generally do not pass on the hybrid vigor (and superior traits) to all offspring. The hybrids are created using two distinct strains and, therefore, additive genetic effects should be increased. This means that subsequent generations are suitable for use in a selection program.

Another approach is to conduct a selection program within a strain and then use hybridization to try to improve performance. However, not all hybrids will show improved performance. Success depends not only on the strains selected but also on the reciprocal hybridization. This means that a male of one strain crossed with a female of a different strain may produce progeny with different traits, and vice versa. The hybrids that perform the best will be determined by experimentation.

### *Interspecific hybridization*

Fish of different species may be crossed to produce more productive progenies. Interspecific hybridization is usually used to exploit hybrid vigor, or the tendency of a crossbred organism to have qualities superior to those of either parent. However, if progenies are fertile, interspecific hybridization has also been used to improve genetics through introgression, a process by which the genes of one species flow into the gene pool of another. This is achieved by backcrossing an interspecific hybrid with one of its parents. The same principles apply as with intraspe-



---

cific hybridization: Genetic improvement is based on new combinations of alleles.

Interspecific hybrids must be able to produce progeny. Once this is established, a breeding program can be attempted. But many between-species hybrids are sterile, do not reproduce as readily as the parents, or produce progenies that are non-viable or abnormal. Spawning does not occur naturally between many species that can be hybridized, so these species must be artificially spawned.

The process of producing a superior fish involves experimenting with combinations of strains, species and reciprocal crosses.

An example of successful interspecific hybridization is crossing channel catfish and blue catfish. Each of these species has several different superior traits. Channel catfish is best for commercial production because of its growth rate. Blue catfish has a more uniform body shape, yields more fillet, is easier to seine, and is more resistant to certain diseases. A cross between female channel catfish and male blue catfish produces viable offspring. The hybrids have a faster growth rate and are more disease resistant than the parental species. With this knowledge, hybridization programs can produce fish that can be selected for further genetic improvements.

### ***Polyplodization***

Most fish species are diploid; they contain two sets of chromosome pairs ( $2N$ ), one set inherited through the mother and the other set inherited through the father. Polyploids have more than the diploid number of chromosomes. Polyploidy can be induced in fish by using techniques such as temperature variation or pressure applied to the eggs to create triploids ( $3N$ ) or tetraploids ( $4N$ ).

Triploids are created by shocking newly fertilized eggs. The egg does not expel the second polar body when shocked. This creates a fertilized egg with one nuclei from the egg ( $N$ ), one from the sperm ( $N$ ), and one from the second polar body ( $N$ ). During development, the three haploid nuclei will fuse and create a triploid. Triploids also can be created by temperature shocks.

Triploids are created to increase fish growth and to control populations by inducing sterility. Triploids should grow larger because the cells are larger (containing more genetic material and larger nuclei). And since triploids are usually sterile, less energy is needed to produce gametes and this energy may be diverted to growth. Triploids are sterile because the normal  $2N$  number of chromosomes is disrupted, so that segregation and independent assortment are disrupted. This makes gametogenesis difficult because the chromosomes cannot be divided equally. Many triploids also have abnormal gonads, which makes

reproduction difficult.

In aquaculture, triploid grass carp are often grown in ponds along with other species to control grasses and weeds. While there is limited use for triploids in large scale aquaculture, using the process for species with extremely high fecundity, such as oysters, has shown a good level of success. Farmers need large numbers of eggs since many eggs do not survive handling and shocking.

Another way to produce triploids is to mate a diploid with a created tetraploid. This method may increase viability, but requires a tetraploid population.

Tetraploids have four sets of chromosomes ( $4N$ ). They can be created by shocking a zygote when it is undergoing mitosis. Shock should be applied after the chromosomes have replicated and as the nucleus is about to divide into two. The shock prevents the nucleus and cell from dividing so that it retains four chromosomes.

One reason to produce tetraploids is to create triploids, as mentioned above. Triploids can be created more efficiently by mating diploids with tetraploids than they can with shock treatments, because shocking causes significant losses. Many tetraploids can produce viable offspring, so that once a population of tetraploids is created, they could be propagated without creating a new population every time.

### ***Production of gynogens or androgens***

Gynogens are fish that contain chromosomes only from the mother. They are produced by activating oocyte division with irradiated sperm and then restoring diploidy to the developing zygote. Irradiation destroys the DNA in the sperm, but the sperm still can penetrate the egg and induce cell division. After activation of the egg with irradiated sperm, the second polar body normally is extruded, resulting in haploid embryos that eventually die if no additional treatments are given. One way to restore diploidy is to block the extrusion of the second polar body by temperature or pressure shocks. Gynogens so created are called meiogens or meiotic gynogens. Meiotic gynogens are not completely homozygous, even though they contain genetic material only from the mother, because of recombinations between chromosomes in the ovum and in the second polar body. Another way to recover diploidy is to block the first cell cleavage after doubling of the chromosomes. This is done with chemical treatment, temperature shocks, or hydrostatic pressure. Gynogens created this way are called mitogens or mitotic gynogens. Mitotic gynogens also contain genetic material only from the mother and are 100 percent homozygous. Gynogens can be very useful for genetic studies. They are used to reduce genetic variations and to produce all-female populations in the XY sex-determination system. In

---

certain cases, homozygous genetic material can reduce the complexity of a study. For instance, for whole genome sequencing, a completely homozygous DNA template can reduce the complexities caused by DNA sequence variations between the two sets of chromosomes in regular diploid individuals.

Androgens contain chromosomes only from the father and are produced by fertilizing irradiated eggs with regular sperm, followed by doubling of the paternal genome. Fertilizing an irradiated egg with normal sperm produces a haploid zygote. The zygote is shocked after replication, during cleavage, to prevent cell division. The two haploid nuclei fuse together to create a diploid zygote with all-male chromosomal material. This produces two identical copies of haploid male chromosomes. Androgens also can be created by using tetraploid males to fertilize irradiated eggs, which produces diploid offspring.

Androgens are completely homozygous and are often referred to as doubled haploid. Androgens are useful for many genetic studies, including reduction of genetic variation and production of all-male or all-female populations. For instance, in the XY sex-determination system, YY males can be produced by androgenesis. Mating YY males with XX females produces an all-male (XY) population. In some cases, as in tilapia production, all-male populations are desirable because of their higher growth rate.

### **Genetic engineering**

Genetic engineering is the process by which a gene(s) or a functional part of a gene is transferred into an organism. The gene may come from the same species as the recipient or a different species. A transgenic fish is produced upon successful gene transfer. The desirable gene is then propagated in the offspring. A number of processes are involved in genetic engineering. First, a gene of interest is cloned and inserted into a vector, such as a bacterial plasmid DNA. The plasmid is then isolated from the bacteria in large quantities. The gene of interest, or the DNA inserted into the plasmid, is removed from the vector and injected into the fish zygote, where it is expected that the new gene will become part of the host fish DNA. Once the transferred DNA is incorporated into the germ cells, it is inherited as a part of the genome; fish so produced are transgenic fish.

Genetic engineering has the advantage of breaking the species barrier. It avoids epistasis by transferring specific gene(s). Genetic control elements also can be manipulated to allow a gene to be controlled by a different promoter, as desired. Such promoters can be constitutive or inducible. For instance, a growth hormone gene can be placed under the control of a constitutive promoter,

leading to constitutive production of elevated growth hormone, which in turn induces fast growth. An inducible promoter can be used to detect specific contaminants for environmental monitoring. For instance, P450 oxidase promoters are inducible upon exposure to certain contaminants. Transgenic fish with these promoters, along with a marker gene, would express the marker gene when the transgenic fish is exposed to the contaminants.

In spite of these advantages, genetic engineering is an unconventional approach and the uncertainty of its effect on both food safety and ecological safety has generated much public resistance to its application in aquaculture. To date, transgenic fish have had very limited use and limited economic impact. The well-documented commercial application of transgenic fish is the ornamental glowing zebrafish, the GloFish™, created in Singapore and commercially available in the United States.

Although the technology for producing transgenic fish is mature, the production and verification process is long and the cost is still high. Major obstacles for using transgenic fish are social resistance and regulations. Therefore, recent studies have focused on assessing transgenic fish in terms of food and ecosystem safety, and on ways to contain transgenic populations.

### **Molecular genetics and genomics**

Molecular genetics is an emerging field in fish breeding programs. It is the study of genetic material (genotypes) to help determine if fish possess certain traits of interest (phenotypes). One such method of genetic testing that will soon become reality for the aquaculturist is DNA marker-assisted selection. When a certain trait of interest is studied, and a genetic marker found for this trait, a DNA test can determine which fish in the population will be the best to use in a breeding program. Some agricultural programs such as beef and poultry have implemented these technologies in their selection and breeding programs already. As our knowledge of the genetics of aquaculture species increases, genetic testing will become a reality for this industry also.

The entire DNA composition of an organism is called its genome. Genomics is the study of the entire genome or DNA of a species and how genes interact within the whole organism. Whole genome sequences for many important aquaculture species should be known in the very near future because of advances in sequencing technologies, but the complete DNA sequence of most aquaculture species is presently unknown. Genomic programs may still be useful for any breeding program. Maps of useful traits (their position along the chromosomes) will be valuable in the integration of genomic data with traditional selection programs. Long-term goals of a genomics program would be to

identify sets of genes and be able to map multiple production traits to their chromosomes to assist in selection.

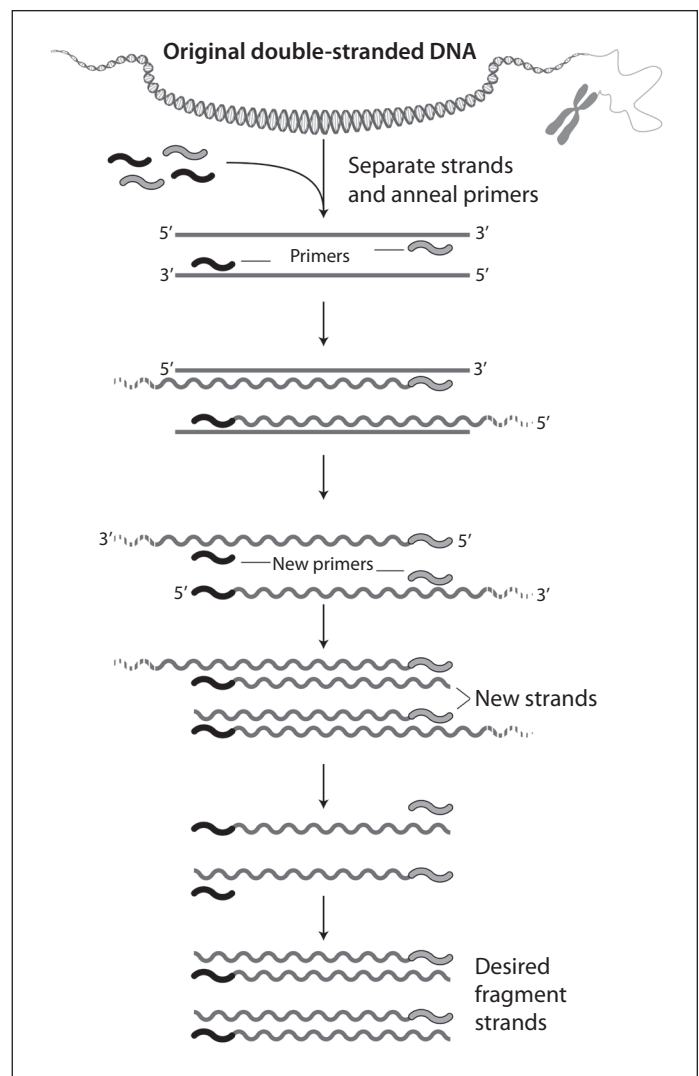
## Genomics concepts and examples of aquaculture genomics research

Genomics is a very active research field. Rather than attempting a thorough review of all the knowledge and progress made, we will discuss some basic concepts of genomics to help readers get to the genomics literature. In aquaculture, major progress has been made in genomics research on many finfish and shellfish species such as Atlantic salmon, rainbow trout, tilapia, carp, striped bass, shrimp, oyster and scallop. We will use research in catfish as examples for convenience.

### Molecular markers

A first step toward improving aquaculture programs through molecular genetics is to identify molecular markers. DNA sequences vary within a population; that is, alleles at a given locus may be different within populations. Such differences are termed polymorphisms. Identifying polymorphic markers within a species can have commercial importance. For example, a molecular marker can be identified that differentiates a population of slow-growing fish from a population of fast-growing fish based on allele usage. To identify molecular markers, blood or tissue samples may be collected in the field, while DNA isolation and the molecular techniques must be performed in a laboratory.

There are many techniques for identifying molecular markers. When no genetic information is available for the aquaculture species of interest, random amplified polymorphic DNA (RAPD) and amplified fragment length polymorphism (AFLP) markers can be used. DNA of the aquatic species is isolated and RAPD or AFLP techniques are used to try to find polymorphisms in the DNA. RAPD and AFLP markers are identified with the help of polymerase chain reactions (PCR). PCR uses primers (short sequences of synthesized DNA) that bind to DNA and amplify a stretch of DNA between the primer binding sites (Fig. 6). In RAPD, short random primers are synthesized and PCR with low annealing temperature is performed on the DNA. The reaction is visualized on a gel and polymorphisms may be identified between DNA samples; they are seen as a presence or absence of an amplified product. The basis of polymorphism in RAPD markers is the differences in DNA sequences between samples. If the DNA sequence at the primer binding site(s) and/or the length of the DNA sequence between the primer sites is different between samples, polymorphisms will likely be



**Figure 6.** The polymerase chain reaction (PCR). The DNA strands are separated by heating. Primers anneal to the target (complementary) sequence and the DNA sequence between the primers is replicated. The process is repeated and the number of DNA molecules is doubled with each cycle of PCR to produce many copies of the desired fragment. Courtesy: National Human Genome Research Institute.

observed. The AFLP technique also can identify polymorphic DNA by using PCR. Of the two techniques used to identify molecular markers, AFLP is highly robust and more reliable, but requires more steps than RAPD, as well as some specialized equipment and training. The basis of polymorphism in AFLP is also caused by differences in DNA sequences between samples, observed as the presence or absence of amplified product on a gel. While RAPD and AFLP markers are a quick and economical way to identify polymorphic DNA between samples, they are inherited as dominant markers. Dominant markers, as the name implies, will generate a marker with a single dose of alleles. Thus, with dominant markers, dominant

---

homozygous and heterozygous genotypes are not distinguishable on the basis of the presence or absence of amplified products on a gel. Dominant markers are generally less informative than co-dominant markers. Communications of dominant markers across laboratories can be difficult as well.

In aquaculture species where some genetic information exists, more molecular markers can be identified. Highly robust and informative markers include microsatellite markers and single nucleotide polymorphism (SNP) markers. Microsatellites are stretches of DNA within a genome that contain simple sequence repeats. As we know, DNA sequences are composed of four nucleotides: A, T, C and G. When the combination of nucleotides at a locus is repeated, such as CACACACACACA, a microsatellite exists. Microsatellites are generally highly polymorphic, abundant, distributed throughout the genome, and inherited as co-dominant markers. This makes microsatellites very useful in developing polymorphic markers. In some cases, the DNA sequences differ at the primer binding site(s) used to amplify the locus by PCR. This would lead to the non-amplification of the allele (so-called null allele). The basis of polymorphism between samples for microsatellites is the number of sequence repeats at a locus such as (CA)<sub>8</sub> versus (CA)<sub>10</sub>. Use of microsatellites as molecular markers requires prior DNA sequence information to identify these repeats, as well as some extra cost and training.

SNP markers are co-dominant markers and using them requires some prior DNA sequence information. SNP markers are also very useful and provide allele-specific information. SNPs are defined as a base change at any given position along the DNA chain (e.g., A to G, or C to T). Theoretically, SNPs should have a total of four alleles (A, T, G and C at any position); observations suggest, however, that they most often exist as bi-allelic markers (e.g., the two alleles can be A or G). SNPs can be identified either within an individual (between sister chromosomes in diploid organisms) or between individuals. For example, one allele has the DNA sequence AATAGCTG and another allele has the sequence AATACCTG. In this case, an SNP marker has been identified at that locus.

Identified molecular markers can have several uses in genetic analysis. When a polymorphic marker has been identified between populations selected for important traits, individuals can be selected for likely trait performance based upon their genotype (marker-assisted selection). Further, molecular markers can help identify DNA variation useful for inducing new and favorable traits in a selection program where needed. When using an interspecific hybrid system, molecular markers can help

identify important genes from each species. Molecular markers can also help identify strain and parentage, and thus confer lineage-specific information.

### **Genomic mapping**

Genome mapping techniques include linkage mapping, physical mapping, molecular cytogenetic mapping by fluorescent *in situ* hybridizations, and radiation hybrid mapping. We will not cover molecular cytogenetic mapping and radiation hybrid mapping in this publication.

To use linkage mapping, multiple molecular markers must be identified and a resource family defined. A resource family is a population of individuals (parents and progeny) whose DNA is used for genotyping. Some planning must be done when choosing a resource family for genetic studies. If the resource family is highly informative, a monohybrid cross can be used if parents are true-breeding, or homozygous, for alternate forms of a trait (Fig. 4). True-breeding parents produce offspring with a phenotype of interest, say disease resistance. When a species is mated from true-breeding parents (P generation), the offspring are termed the first filial generation, or F<sub>1</sub> generation. These F<sub>1</sub> progeny are used to create F<sub>2</sub> progeny by self-mating, back-crossing, or hybridization. For linkage mapping studies, the F<sub>2</sub> generation (and beyond) are chosen as resource families. If the F<sub>1</sub> generation is chosen as a resource family, only heterozygous loci of the parents are segregating; the vast majority of homozygous loci of the parents are not segregating. As a result, the F<sub>1</sub> resource family may not be fully informative, as all heterozygous siblings will have genotypes with no allele segregation. For instance, if homozygous P fish (AA) and (aa) are mated all progeny will have the Aa genotype and no allele segregation will be observed. F<sub>2</sub> progeny produced by a backcross of F<sub>1</sub> fish will produce offspring with the genetic diversity to use as a resource family, as Aa x Aa produces progeny with different genotypes (AA, Aa, and aa alleles are possible). As a practical example, F<sub>1</sub> interspecific hybrid catfish (channel catfish x blue catfish) have been created for their superior performance in several commercially important traits. An F<sub>2</sub> generation was created by mating F<sub>1</sub> generation hybrids with channel catfish and is currently being used as a highly informative resource family.

Linkage maps are created using multiple polymorphic molecular markers within a resource family. Remember, recombination of alleles occurs by crossover of homologous chromosomes. However, recombination frequency is not consistent throughout a genome. Genes on different chromosomes segregate completely independently. When genes are close together on the same chromosome, they are physically linked. These genes are expected to have a

---

lower recombination frequency than genes on the same chromosome but far away from each other. In a resource family, when multiple loci are screened by using multiple molecular markers, linkage maps can be created, usually with the aid of software programs. While the creation of linkage maps is time-consuming and can be complex, the process simply involves the reconstruction of chromosomes (by creating linkage groups) using the recombination differences of molecular markers. When recombination frequency between markers is low (near 0 percent), it is expected that the markers are linked and little or no recombination has occurred. Conversely, when recombination frequency between markers is high (approaching 50 percent), it is expected that the markers are independently assorted during meiosis. The creation of linkage groups is very useful in determining the position and order of markers/alleles along chromosomes.

Another useful genomic mapping strategy is physical mapping. Physical maps are created with the help of a DNA library. A library, in terms of genetics, is any collection of DNA fragments that have been inserted into a cloning vector for propagation. For physical mapping, the whole genome of a species can be fragmented and used to construct large-insert DNA libraries. One example of a large-insert DNA library is a bacterial artificial chromosome (BAC) library. A BAC library contains many long pieces (~200,000 nucleotides) of genomic DNA of the species of interest. BAC libraries can be used to create physical maps by a technique called DNA fingerprinting. The genomic DNA contained in a BAC library is isolated and fragmented to create “fingerprints,” or highly specific DNA patterns based on the nucleotide composition of the DNA sequence. Overlapping fingerprints are then used to reconstruct the DNA, with the goal of creating a map spanning the entire genome of the species of interest. Physical maps are used in whole genome sequencing projects and are a useful resource for many other genomic projects, including further marker and gene identification, whole genome comparisons to other highly characterized species (map-rich or model species) such as zebrafish and *Tetraodon*, and in map integration projects.

When linkage and physical maps are created, the maps can be integrated (aligned) together, which is particularly useful in analyzing quantitative trait loci (QTL). QTLs are regions of DNA where a correlation has been identified with a trait(s) of interest. QTLs can be added to genomic maps. By integrating the linkage and physical maps, a QTL identified within a linkage map can be located along the physical map. If the marker corresponds to a known gene, the function of this gene may be determined. If the QTL occurs along a region of DNA where no function can yet be assigned, maps are especially use-

ful in locating specific regions along the chromosome for further study. QTL analysis forms the basis for marker-assisted selection, where loci that correspond to candidate genes or traits can be used to assist classical breeding programs.

Ultimately, the best genome map can be obtained by sequencing the whole genome for the species of interest. Sequencing a whole genome is costly and requires time, effort and specialized equipment. Even so, the whole genome sequence of a species can provide a wealth of information. Most whole genome sequencing projects to date are performed by shotgun sequencing and/or minimal tiling path sequencing. Shotgun sequencing is done by sequencing many random, short (~500 nucleotides) segments of DNA and assembling the sequence with the help of computers to re-construct the chromosomes. With minimal tiling path sequencing, DNA is sequenced in an orderly manner and guides are used to help in the reconstruction process. An example of the minimal tiling path method would be using a BAC library to sequence individual BACs and assembling each BAC one by one. Both strategies are effective, and many projects use a combination of the methods to provide a highly accurate genome sequence. Recent progress in whole genome sequencing includes the use of next-generation (second generation and third generation of sequencers) sequencing technologies. Currently, next-generation sequencing can produce hundreds of thousands to tens of millions of DNA sequences from a sequencing reaction. As more sequencing technologies are developed and they become less costly, whole genome sequencing projects for many aquaculture species are expected.

### ***Gene expression studies***

Other important research is the study of gene expression. The central dogma in molecular biology is the flow of genetic information—DNA to RNA to proteins. A gene (DNA) is transcribed into RNA. There are several types of RNA; messenger RNA (mRNA) translates the genetic information to proteins, the biologically active endpoint of many genes. Many gene expression studies involve the study of mRNA. The level of expression of a given gene has a direct relation to the amount of mRNA present for that gene. When candidate genes have been identified, gene expression studies can help determine where (what cell or tissue) and how much (quantity of mRNA present) of the gene has been expressed. There are many reasons to perform gene expression studies. An example would be determining gene expression in a control versus a treatment group, such as in healthy fish versus diseased fish.

Several genomic tools can help in gene expression studies. The most common technique for studying the

---

expression of individual genes is reverse transcription-PCR (RT-PCR). PCR requires DNA templates, so for gene expression studies mRNA must be converted to complementary DNA, or cDNA, using reverse transcriptase, thus the name RT-PCR. Once the mRNA is converted to cDNA, the PCR process works the same way as regular PCR. The idea is that if the starting material contains more of a specific mRNA, more PCR products will be generated using specific PCR primers than when the starting material contains less of the mRNA. This type of RT-PCR is sometimes referred to as semi-quantitative PCR because the quantification is not always perfect. Quantification of starting mRNA relies on the PCR to be conducted under identical conditions and stays within the log phase of PCR amplification. A better approach is to use real time RT-PCR. It provides a highly accurate assessment of gene expression levels but is costly and requires specific equipment.

Gene expression profiling is used to assess gene expression at the genome scale. When working with a species for which little or no genetic data is available, cDNA libraries are useful in gene discovery projects. Not all genomic DNA corresponds to the gene coding (protein coding) sequence. Therefore, by using a cDNA library for sequencing (instead of genomic DNA), the sequences correspond to gene coding products. Single-run sequencing does not guarantee that the complete mRNA (cDNA) will be sequenced. Usually, only a short fragment of the complete cDNA is sequenced. Single-run sequences of cDNA are called expressed sequence tags (ESTs). Many gene products in an organism can be discovered by generating a large set of ESTs. The frequency of ESTs in a large-scale EST sequencing project is a rough reflection of the gene expression patterns of the organism. However, repeated sequencing of the most abundantly expressed transcripts prevents the rarely expressed genes from being sequenced at all.

To create a highly efficient cDNA library for gene discovery projects, multiple sources (cells and tissues) can be used. To circumvent the problem of repeated sequencing of the most highly expressed genes, various types of DNA libraries can be created, including normalized and subtracted libraries. Normalized and subtracted cDNA libraries are often created in the effort to sequence the greatest number of genes. Since cDNAs in a library are sequenced at random, gene products will be sequenced at a frequency relative to their expression levels represented in the library. A normalized library maximizes gene discovery because all genes represented in the library are at a more equal abundance (in theory, at least). Subtracted libraries are useful when comparing two or more expression groups. Subtracted libraries eliminate the cDNA shared by the groups. The remaining cDNA corresponds

to the genes that are expressed in one group but not the other(s). The traditional Sanger sequencing method has limitations. The highly efficient next generation sequencers have the ability to discover the entire mRNA composition of an organism without any subtraction or normalization, which greatly facilitates genome-scale gene expression studies.

Once cDNA libraries have been made and ESTs generated, these resources can be used to develop microarrays. In the absence of a whole genome sequence in most of the aquaculture species today, microarrays are often created using all available EST sequences for a species of interest. Common ESTs are combined (clustered) to create a unique set of sequences. These sequences are used to synthesize DNA “features.” Features are applied to a media to create an array of sequences. These features can detect the presence and level of expressed genes in a population of cDNA through hybridization. Microarrays are very powerful tools, limited only by the number and quality of EST resources available with which to design features. Many studies involving microarray technology have been and are currently being performed in aquaculture species, and there can be many applications. One general application would be to determine which genes are differentially expressed in a control group versus a treatment group. This is generally performed using a specific tissue or cell type. An example would be to study the expression profile of healthy fish liver tissue (control) versus liver from a diseased fish (treatment) to attempt to discover genes involved in disease resistance or disease susceptibility. In this example, when cDNA is created from both tissues, the microarray can be used to detect the genes that are up-regulated or down-regulated between the tissues. These data are useful in determining a global gene expression profile between treatment groups, and also in further research to produce species with superior performance traits.

## Afterword

An understanding of the principles of genetics is useful in any aquaculture program. Genetics programs can help increase productivity in aquaculture systems, as in using hybridization and selection to produce strains with superior performance traits. Genetics research has improved the quality and production of aquatic species throughout the years, but there is a need for further and faster genetic gains. Much of the progress to date has been made using traditional selection. As molecular genetics and genomics tools and technologies are implemented in aquaculture systems, they will complement and extend classical genetic improvement programs.



---

SRAC fact sheets are reviewed annually by the Publications, Videos and Computer Software Steering Committee. Fact sheets are revised as new knowledge becomes available. Fact sheets that have not been revised are considered to reflect the current state of knowledge.



The work reported in this publication was supported in part by the Southern Regional Aquaculture Center through Grant No. 2007-38500-18470 from the United States Department of Agriculture, National Institute of Food and Agriculture.

---