Aquaculture Genome Technologies
Zhanjiang (John) Liu

Chapter 5
Microsatellite Markers and Assessment of Marker Utility

Zhanjiang Liu

Microsatellites consist of multiple copies of tandemly arranged simple sequence repeats (SSR) that range in size from 1–6 base pairs (bp) (e.g., AC, CCA, or GATA) (Tautz 1989). Based on the repeat composition, microsatellites are often classified into mononucleotide microsatellites, dinucleotide microsatellites, trinucleotide microsatellites, tetranucleotide microsatellites, and so on. Microsatellites containing only one type of repeats are called simple microsatellites; microsatellites containing more than one type of repeats are called composite microsatellites. For instance, (CA)15 is a simple microsatellite, but (CA)8(CG)12 is a composite microsatellite. The advantages of microsatellites as molecular markers include its abundance in genomes, even distribution, small locus size facilitating polymerase chain reaction (PCR)-based genotyping, codominant nature of Mendelian inheritance, and high levels of polymorphism.

Abundance of Microsatellites

Microsatellites are highly abundant in various eukaryotic genomes including all aquaculture species studied to date. In most of the vertebrate genomes, microsatellites make up a few percent of the genomes in terms of the involved base pairs, depending on the compactness of the genomes. Generally speaking, more compact genomes tend to contain a smaller proportion of repeats including SSRs, but this generality is not always true. For example, the highly compact genome of the Japanese puffer fish contains 1.29% of microsatellites, but its closely related *Tetraodon nigroviridis* genome contains 3.21% of microsatellites (Grollius et al. 2000). During a genomic sequencing survey of channel catfish, microsatellites were found to represent 2.58% of the catfish genome based on the 11.4 million bps sequenced (Xu et al. 2006). A total of 4,262 microsatellites were found within 11.4 million bps (i.e., one microsatellite exists per 2.67 kilobases [kb] of channel catfish sequences). In fugu, one microsatellite was found every 1.87 kb of DNA. For comparison, in the human genome, one microsatellite was found every 6 kb of DNA (Beckmann and Weber 1992). It is reasonable to predict that in most aquaculture fish species, one microsatellite should exist every 10 kb or less of the genomic sequences, on average.

Dinucleotide repeats are the most abundant forms of microsatellites. For instance, in channel catfish, 67.9% of all microsatellites are present in the form of dinucleotide repeats; 18.5% present as trinucleotide repeats; and 13.5% as tetranucleotide repeats, excluding mononucleotide repeats, which are not nearly useful enough for molecular markers. Generally speaking, dinucleotide microsatellites are the most abundant, followed by tri- or tetra-nucleotide repeats, but in some cases,

tetranucleotide repeats can be more frequent than the trinucleotide repeats. For instance, in the genome of Japanese puffer fish *Takifugu rubripes*, dinucleotide repeats have the highest relative frequency (34%) followed by tetranucleotides (21%), trinucleotides (19%), mononucleotides (16.5%), hexanucleotides (6%), and pentanucleotides (3%) (Edwards et al. 1998).

Of the dinucleotide repeat types, $(CA)_n$, is the most common dinucleotide repeat type, followed by $(AT)_n$, and then $(CT)_n$ (Toth et al. 2000, Xu et al. 2006). $(CG)_n$ type of repeats are relatively rare in the vertebrate genomes. Partially this is because the vertebrate genomes are often A/T-rich. Of the trinucleotide repeats and tetranucleotide repeats, relatively A/T-rich repeat types are generally more abundant than G/C-rich repeat types. For instance, in channel catfish, the most abundant trinucleotide repeat is AAT, followed by AAC, ATC, and ATG. Microsatellites longer than tetranucleotide repeats (penta- and hexanucluotides) are much less abundant and therefore, are less important as molecular markers (Toth et al. 2000). It is important to point out that the definition of microsatellites limiting to repeats of six bases long are quite arbitrary. Technically speaking, repeats with seven bases or longer sequences are also microsatellites, but because they become rarer as the repeats are longer, they are less relevant as molecular markers.

Genomic Distribution of Microsatellites

Microsatellites tend to be, relatively speaking, evenly distributed in the genome on all chromosomes and all regions of the chromosome. They have been found inside gene coding regions (e.g., Liu et al. 2001), introns, and in the nongene sequences (Toth et al. 2000). The best known examples of microsatellites within coding regions are those causing genetic diseases in humans, such as the CAG repeats that encode the polyglutamine tract, resulting in mental retardation. In spite of their wide distribution in genes, microsatellites are predominantly located in noncoding regions (Metzgar et al. 2000). Only about 10–15% of microsatellites reside within coding regions (Moran 1993, van Lith and van Zutphen 1996, Edwards et al. 1998, Serapion et al. 2004). This distribution should be explained by negative selection against frame shift mutations in the translated sequences (Metzgar et al. 2000, Li et al. 2004). Because the majority of microsatellites exist in the form of dinucleotide repeats, any mutation by expansion or shrinking would cause frame shift of the protein encoding open frames if they reside within the coding region. That also explains why the majority of microsatellites residing within coding regions have been found to be trinucleotide repeats, though the presence of dinucleotide repeats and their mutations within the coding regions do occur.

Locus Size of Microsatellites

Most microsatellite loci are relatively small, ranging from a few to a few hundred repeats. The relatively small size of microsatellite loci is important for PCR-facilitated genotyping. Generally speaking, within a certain range, microsatellites containing a larger number of repeats tend to be more polymorphic, though polymorphism has been observed in microsatellites with as few as five repeats (Karsi et al. 2002).

For practical applications, microsatellite loci must be amplified using PCR. For best separations of related alleles that often differ from one another by as few as one repeat unit, it is desirable to have small PCR amplicons, most often within 200 bp. However, due to the repetitive nature of microsatellites, their flanking sequences can be a quite simple sequence, prohibiting design of PCR primers for the amplification of microsatellite loci within a small size limit. Consideration should be given regarding whether the quality of PCR primers can be sacrificed to a certain extent to reduce the amplicon size.

Polymorphism of Microsatellites

Microsatellites are highly polymorphic as a result of their hypermutability and thereby cause the accumulation of various forms in the population of a given species. Microsatellite polymorphism is based on size differences due to varying numbers of repeat units contained by alleles at a given locus (Figure 5.1). Microsatellite mutation rates have been reported as high as 10^{-2} per generation (Weber and Wong 1993, Crawford and Cuthbertson 1996, Ellegren 2000), which are several orders of magnitude greater than that of nonrepetitive DNA (10^{9}) (Li 1997).

Two mechanisms have been suggested to explain the hypermutability of microsatellites. (For a recent review, see Chistiakov et al. 2006.) The first involves polymerase slippage during DNA replication, resulting in differences in the number of repeat units (Levinson and Gutman 1987, Tautz et al. 1989). Transient dissociation of the replicating DNA strands followed by subsequent reassociation (Schlötterer et al. 1991, Richards and Sutherland 1994) would lead to slippage of the two strands, leading to the change of repeat numbers in the newly replicated DNA. Direct studies of human families have shown that new microsatellite mutations usually differed from the parental allele by only one or two repeats (Weber and Wong 1993), favoring a stepwise mutation model (see review by Estoup and Cornuet 1999) due to polymerase slippage. Microsatellite stability is controlled at multiple steps *in vivo* through the DNA mismatch repair (MMR) system, as shown for *Escherichia coli*, yeast, and humans (Sia et al. 1997). MMR proteins are found in a wide variety of taxa and are responsible for the correction of replication mistakes and suppression of the recombination between diverged sequences (Kolodner and Marsischky 1999). If the MMR system is defective, coding sequences with tandem repeats become subject to mutations, for example in tumor tissues (Sia et al. 1997). High-frequency microsatellite instability, therefore, plays a pivotal role in carcinogenesis (Atkin 2001). Both minor and major MMR genes contain short $(A)_n$ tracts in their coding regions, which are highly vulnerable to spontaneous deletion or insertion mutations, that could result in the inactivation of the MMR gene and hence cause MMR deficiency (Chang et al. 2001).

The second mechanism involves nonreciprocal recombination within the SSRs, leading to production of significantly larger and smaller alleles (Jakupciak and Wells 2000). In a few fish species, we have observed alleles with very large differences in repeat numbers, predictive of a infinite allele model (Balloux and Lugon-Moulin 2002), as well as alleles with differences of just one repeat unit, characteristic of the stepwise mutation model. Regardless of specific mechanisms, changes in numbers of repeat units can result in a large number of alleles at each microsatellite locus in a population. For evolutionary studies in populations, however, most often the stepwise

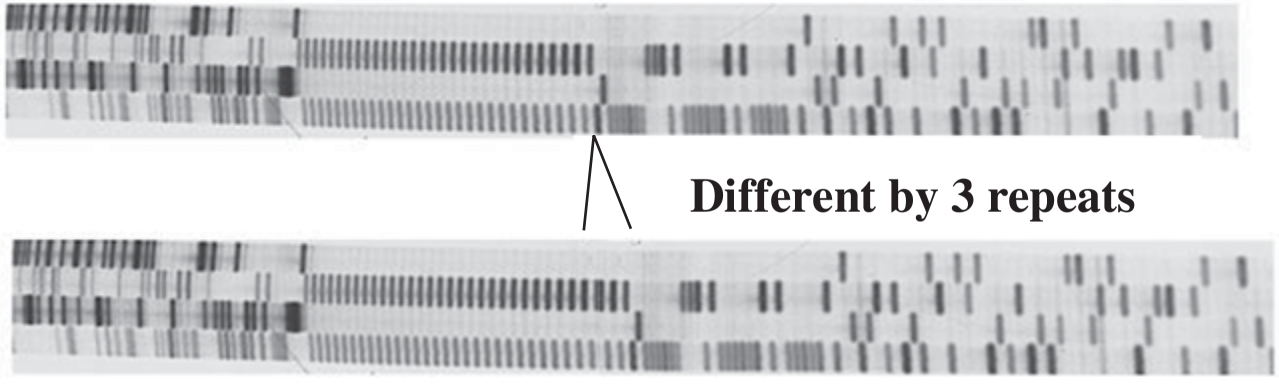

Figure 5.1. Microsatellite polymorphism is caused by the difference in repeat numbers. Note that the sequence of the lower fish harbors three more repeat units leading to a length difference of 6 bp. Figure was adopted from Liu and Cordes (2004b).

mutation model is assumed, right or wrong, as individuals with a similar number of repeats are regarded to be more closely related than those with a large difference in their numbers of repeats.

Inheritance of Microsatellites

Microsatellites are inherited in a Mendelian fashion as codominant markers. This is one of the strengths of microsatellite markers in addition to their abundance, even genomic distribution, small locus size, and high polymorphism.

Genotyping of microsatellite markers is usually straightforward. However, due to the presence of null alleles (alleles that cannot be amplified using the primers designed), complications do exist. As a result, caution should be exercised to assure the patterns of microsatellite genotypes fit the genetic model under application (Figure 5.2).

The disadvantage of microsatellites as markers include the requirement for existing molecular genetic information, the large amount of up front work for microsatellite development, and the tedious and labor intensive nature of microsatellite primer design, testing, and optimization of PCR conditions. Each microsatellite locus has to be identified and its flanking region sequenced for the design of PCR primers. Due to polymerase slippage during replication, small size differences between alleles of a given microsatellite locus (as little as 2 bp in a locus comprised of dinucleotide repeats) are possible. Because of this, PCR-amplified microsatellite DNA was traditionally labeled radioactively, separated on a sequencing gel, and then exposed on X-ray film overnight (Sambrook et al. 1989). Significant increases in the number of samples that can be typed in a day have been achieved by using automated fluorescent sequencers coupled with computer imaging systems (O'Reilly and Wright 1995).

The large number of alleles per locus results in the highest heterozygosity and polymorphic information content (PIC) values of any DNA markers. Microsatellites have recently become an extremely popular marker type in a wide variety of genetic investigations, as evidenced by the recent debut of the journal *Molecular Ecology Notes*, dedicated almost entirely to publishing primer and allele frequency data for newly characterized microsatellite loci in a wide range of species. Over the past decade, microsatellite markers have been used extensively in fishery research including studies of genome mapping, parentage, kinships, and stock structure (see O'Connell and Wright 1997 for a review). A cursory online literature search produced more than 500 entries since 1998 involving the use of microsatellites in such studies.

Development of Microsatellite Markers

Technically, the simplest way to identify and characterize a large number of microsatellites is through the construction of microsatellite-enriched small-insert genomic libraries (Orstrander et al. 1992, Lyall et al. 1993, Kijas et al. 1994, Zane et al. 2002). In spite of the variation in techniques for the construction of microsatellite-enriched libraries, the enrichment techniques usually include selective hybridization of fragmented genomic DNA with a tandem repeat-containing oligonucleotide probe and further PCR amplification of the hybridization products. Libraries highly

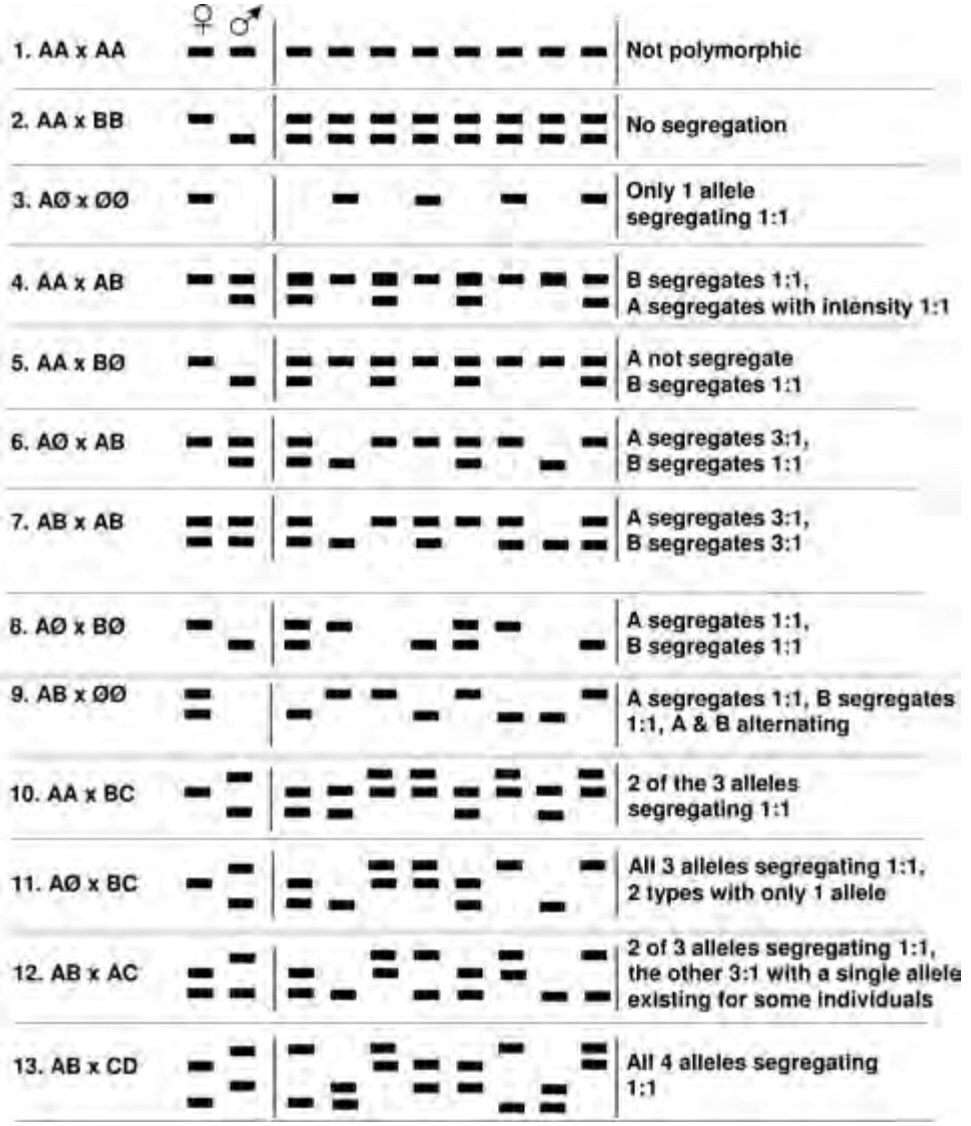

Figure 5.2. Various segregation patterns and the predicted genotypes.

enriched by tandem repeats have been constructed for many organisms, including fish such as *Lophius* sp. (Garoia et al. 2003), gilthead sea bream *Sparus aurata* (Zane et al. 2002), and Nile tilapia (Carleton et al. 2002).

In spite of the simplicity in the construction of microsatellite-enriched libraries and thereby the identification and characterization of microsatellite markers, for large genome projects, the real need of direct microsatellite marker development may not be the wisest approach. This is because microsatellites derived from enriched libraries most often are from anonymous genomic regions and thus are Type II markers. O'Brien (1991) divided molecular markers into Type I markers associated with genes of known functions and Type II markers associated with anonymous genomic

sequences. Microsatellites usually represent Type II markers, since by chance more than 90% of the genome are noncoding regions. Type II markers most often cannot be used for comparative genome studies across a wide spectrum of species. More importantly, microsatellites can be identified and sequenced directly from genome sequence surveys such as BAC end sequencing (see Chapter 15), and from EST analysis from which many microsatellites can be developed into Type I markers (Liu et al. 1999, Serapion et al. 2004).

Type I markers are more difficult to develop (Liu et al. 1999b, Liu and Cordes 2004a). Although nongene sequences are free to mutate, causing higher levels of polymorphism, sequences within protein-coding regions generally show lower levels of polymorphism because of functional restraints. The most effective and rapid way for producing Type I microsatellites is the sequencing of clones from cDNA libraries. Both 5'- and 3'-ends of cDNA clones can be sequenced to produce ESTs (see Chapter 20). Such collections provide a robust sequence resource that can be used for gene discovery and annotation, as well as for the identification of gene-associated microsatellite markers for comparative genetics (Liu 2003, Serapion et al. 2004).

Microsatellites can be searched in EST sequence databases. However, the prerequirement for the development of Type I microsatellites is the access to sufficient sequence information. For instance, in the channel catfish, generation of 45,000 EST sequences allowed the identification of thousands of microsatellites (Serapion et al. 2004). Sequence analysis of 1,909 ESTs from a skin cDNA library of *Ictalurus punctatus* revealed the presence of 89 (4.7% of 1,909) microsatellite-containing genes (Karsi et al. 2002). Screening of 1,201 ESTs from a channel catfish brain cDNA library yielded 88 (7.3%) clones with microsatellites (Ju et al. 2000). It is therefore, reasonable to argue that one does not have to spend resources to intentionally develop microsatellite markers that not only take time and effort but are mostly Type II markers. Instead, identification of microsatellite markers from EST resources should generate mostly Type I markers. Likewise, many microsatellites can be identified from BAC end sequences that fulfill two duties with one action (Xu et al. 2006).

Caution has to be exercised, however, on microsatellites developed from ESTs. First, due to the presence of introns, one has to be careful not to design primers at the exon-intron boundaries. Second, the presence of introns would make allele sizes unpredictable. Finally, many microsatellites exist at the 5'- or 3'-UTR, making flanking sequences not sufficient for the design of PCR primers. Although introns are not a problem for microsatellites derived from BAC end sequencing, sequencing reactions often terminate immediately after the microsatellite repeats, also making flanking sequences not sufficient for the design of PCR primers. Additional sequencing would be required for the development of microsatellite markers from these sequences with SSRs at the ends.

Applications of Microsatellite Markers

Microsatellite markers, one of the many types of molecular markers, are ideal for many types of applications. Microsatellites are the most useful type of molecular markers for genetic linkage mapping, and they are anticipated to remain as the markers of choice for the construction of linkage maps, especially for framework linkage maps. The fundamental reason for this is because of their high polymorphism, high abundance, small

Table 5.1. Some examples of linkage maps constructed with microsatellite markers in aquaculture species.

Species	Common name	References
Salmo trutta	Brown trout	Gharbi et al. 2006
Dicentrarchus labrax	European sea bass	Chistiakov et al. 2005
Oreochromis spp.	Tilapia	Kocher et al. 1998, Agresti et al. 2000, McConnell et al. 2000, Lee et al. 2005
Plecoglossus altivelis	Ayu	Watanabe et al. 2004
Xiphophorus		Walter et al. 2004
Salvelinus alpinus	Arctic char	Woram et al. 2004
Salmo salar	Atlantic salmon	Gilbey et al. 2004, Moen et al. 2004
Oncorhynchus mykiss	Rainbow trout	Sakamoto et al. 2000, Nichols et al. 2003
Ictalurus punctatus	Channel catfish	Waldbieser et al. 2001
Danio rerio	Zebrafish	Knapik et al. 1998, Shimoda et al. 1999
Crassostrea gigas	Pacific oyster	Hubert and Hedgecock 2004
Crassostrea virginica	Eastern oyster	Yu and Guo 2003
Seriola quinqueradiata and *Seriola lalandi*	Yellowtails	Ohara et al. 2005
Cyprinus carpio	Common carp	Sun and Liang 2004
Paralichthys olivaceus	Japanese flounder	Coimbra et al. 2005

locus size, and codominance of inheritance, so all four alleles from a pair of parents can be observed from the gel. In addition, because microsatellite markers are sequence-tagged markers, integration with a physical map is possible.

Some examples of genetic linkage maps constructed using microsatellites are listed in Table 5.1. An excellent review was just published by Chistiakov and others (2005) in which various applications of microsatellites were discussed; interested readers are referred to this review. In Chapter 8, use of microsatellite markers for stock analysis and parentage analysis is described. Chapter 9 describes methodologies involving the use of microsatellites and other markers for population analysis. The use of microsatellite markers for the construction of genetic linkage maps is the subject of Chapter 10. QTL mapping involving the use of microsatellite markers are described in Chapter 11.

Assessing the Utility of Markers

Mostly, the number of alleles and allele frequencies measure informativeness of a genetic marker. Specific parameters have been developed through the course of marker development and application in the last two decades. These parameters include two measures of marker informativeness: heterozygosity and polymorphic information content. In addition, the efficiency of marker systems can be evaluated based on the mean number of polymorphic genetic markers per assay. Genotyping errors and multiplex ratio are also used to measure the utility of molecular markers.

Genotyping errors reflect the reproducibility of the marker assay and clarity of the marker genotypes. Multiplex ratio refers to the number of simultaneously assayed loci. In the case of microsatellites, multiplex ratio indicate the number of loci for which PCR primers are compatible for multiplex PCR, and their products are distinct allowing simultaneous analysis on the same gel run.

Heterozygosity (H)

Heterozygosity is a widely used measure of marker informativeness. The informativeness of a genetic marker increases as *H* increases. The heterozygosity of a genetic marker is estimated by the number of alleles and their relative frequencies. Heterozygosity is a function of the individuals and populations sampled. When individuals are sampled from genetically narrow or genetically isolated populations, fewer alleles and a higher frequency of monomorphic loci can be expected than when individuals are sampled from genetically diverse populations. Heterozygosity is defined with the following formula:

$$H = 1 - \sum_{i=1}^{k} Pi^2 \tag{5.1}$$

where P_i is the frequency of the *i*th allele and *k* is the number of alleles (Nei and Roychoudhury 1974, Ott 1992).

Heterozygosity, as its name suggests, is an estimate of probability that a randomly sampled individual is heterozygous when the individuals are sampled from outbred populations. For instance, if a microsatellite has three alleles in the population with a frequency of 10%, 30%, and 60%, H value for this population would be $H = 1-(0.1^2 + 0.3^2 + 0.6^2) = 0.54$. In other words, assuming random mating, the probability of finding a random individual in this population to be heterozygous is 54%.

From the above formula, it is obvious that when the number of alleles is given, the more equal the distribution of all alleles, the greater the H value is; when the allele frequencies are given, the greater the number of alleles, the greater the H value is. For instance, if a population contains two alleles of a microsatellite locus with 10% and 90% frequency each for the two alleles, H value for this microsatellite locus is 0.18. For this microsatellite locus, if the allele frequency of the two alleles is 50% each, then the H value becomes 0.5. If the population has two alleles, then each of the two alleles has a frequency of 50% and the H value would be 0.5. If the population has 10 alleles with 10% each, then the H value would be 0.9.

Polymorphic Information Content

Another measure of the marker informativeness in outbred species is the polymorphic information content (PIC) (Botstein et al. 1980). PIC is defined by the following formula:

$$PIC = 1 - \sum_{i=1}^{k} pi^2 - \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} 2Pi^2 Pj^2 \tag{5.2}$$

In the above example, the population has three alleles with allele frequency of 10%, 30%, and 60%. The PIC value should be calculated as:

$$PIC = 1 - [(0.1)^2 + (0.3)^2 + (0.6)^2] - [2 \times (0.1)^2 \times (0.3)^2 + 2 \times (0.1)^2 \times (0.6)^2 + 2 \times (0.3)^2 \times (0.6)^2] = 0.4662 \quad (5.3)$$

Because of the complexities in deriving PIC, for practical purposes in most cases, H values are used rather than PIC. PIC is always slightly smaller than H, but in most cases is close enough to H so it would not make a major difference. Interested readers are referred to Botstein and others (1980) for more information on PIC.

Mean Heterozygosity

Mean heterozygosity is a useful measurement when dealing with multiple markers or marker types. The mean heterozygosity of n genetic markers is

$$H_{av} = 1 - \sum_{j=1}^{n} H_j / n \quad (5.4)$$

where H_j is the heterozygosity of the jth genetic marker. H_{av} can be estimated with or without monomorphic markers (depending on the context of the analysis). To distinguish between the two cases, let H_{pav} be the mean heterozygosity estimated from polymorphic markers only and H_{Tav} be the mean heterozygosity estimated from polymorphic and monomorphic markers. Powell and others (1996) compared the informativeness and multiplex ratios of RFLP, AFLP, RAPD, and microsatellite markers in soybeans and proposed estimating the mean heterozygosity of genetic markers by summing over monomorphic and polymorphic genetic markers when comparing markers with different multiplex ratios:

$$H_{Tav} = p \sum_{j=1}^{n_p} H_j / n_p \quad (5.5)$$

where H_j is the heterozygosity of the jth polymorphic genetic marker, n_p is the number of polymorphic genetic markers, and p is the percentage of polymorphic markers (number of polymorphic markers over the sum of the polymorphic and monomorphic markers).

The Mean Number of Polymorphic Genetic Markers per Assay

The mean number of polymorphic genetic markers per assay is the product of the mean number of bands per assay and mean heterozygosity:

$$P_{av} = mH_T \quad (5.6)$$

where m is the mean number of bands per assay, and HT is the average heterozygosity. This measurement takes both the heterozygosity and the efficiency of the markers into consideration. For instance, AFLP is much more efficient in producing polymorphic bands than RFLP and it has a greater mean number of polymorphic genetic markers per assay. For example, 10 AFLP primer combinations produce a total of 1,000 bands (100 per AFLP assay on average) and 10 RFLP marker assays produce a total of 12 bands (8 RFLP probes produce 1 band each, and the other 2 RFLP probes produce 2 bands each). If $H_T = 0.2$ for the AFLP markers and $H_T = 0.52$ for the RFLP markers (clearly in this case even though the heterozygosity is lower with AFLP markers than RFLP markers for individual markers), then P_{av} for the AFLP markers is 20, whereas that for the RFLP markers is only 0.624. Thus, the AFLP markers are $20/0.624 = 32.05$ times more informative than the RFLP markers on a per assay basis. Multilocus fingerprinting techniques such as RAPD and AFLP should have a higher mean number of polymorphic genetic markers per assay than single locus marker systems such as RFLP and microsatellites.

Conclusion

Microsatellite markers have been and likely will remain the marker type of choice for genome research because of their high polymorphism, high abundance, small locus size, even genomic distribution, and codominance of inheritance. Microsatellites have the highest heterozygosity among all marker types because of their high number of alleles. Because most RFLP and SNP markers are regarded as biallelic markers, they have a maximal heterozygosity value of 0.5 (when the two alleles have equal allele frequencies). RAPD and AFLP are both biallelic dominant markers, and they can have a maximal heterozygosity of 0.5 as well. Thus, microsatellites are most informative as genetic markers. This feature makes microsatellites the unique marker system for identification of individuals such as parentage analysis, as shown in Chapter 8, as well as the choice of markers for many other types of applications.

The major application of microsatellite markers is for the construction of genetic linkage and QTL maps. This is also because of the high polymorphic rate of microsatellite markers. When a resource family is produced, the male and female fish parents are likely heterozygous in most microsatellite loci. The high polymorphism of microsatellites makes it possible to map many markers using a minimal number of resource families.

There are other reasons for the popularity of microsatellites. One of these is because microsatellites are sequence-tagged markers that allow them to be used as probes for the integration of different maps including genetic linkage and physical maps. Communication using microsatellite markers across laboratories is easy, and use of microsatellite across species borders is sometimes possible if the flanking sequences are conserved (FitzSimmons et al. 1995, Rico et al. 1996, Leclerc et al. 2000, Cairney et al. 2000). As a result, microsatellites can be used also for comparative genome analysis. If microsatellites can be tagged to gene sequences, their potential for use in comparative mapping is greatly enhanced.

Development of microsatellite markers has traditionally been conducted by the development of microsatellite-enriched DNA libraries. However, this may not be the

most optimal situation for genome research. In most cases, EST and BAC end sequence resources are needed earlier or later. Therefore, direct investment into resource development involving EST and BAC end sequencing may prove to be very effective.

References

Agresti J, S Seki, A Cnaani, S Poompuang, EM Hallerman, N Umiel, G Hulata, GAE Gall, and B May. 2000. Breeding new strains of tilapia: development of an artificial center of origin and linkage map based on AFLP and microsatellite loci. Aquaculture, 185, pp. 43–56.

Atkin NB. 2001. Microsatellite instability. Cytogenet Cell Genet, 92, pp. 177–181.

Balloux F and N Lugon-Moulin. 2002. The estimation of population differentiation with microsatellite markers. Mol Ecol, 11, pp. 155–165.

Beckman JS and JL Weber. 1992. Survey of human and rat microsatellites. Genomics, 12, pp. 627–631.

Botstein D, RL White, M Skolnick, and RW Davis. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet, 32, pp. 314–331.

Cairney M, JB Taggart, and B Hoyheim. 2000. Characterization of microsatellite and minisatellite loci in Atlantic salmon (*Salmo salar* L.) and cross-species amplification in other salmonids. Mol Ecol, 9, pp. 2175–2178.

Carleton KL, JT Streelman, BY Lee, N Garnhart, M Kidd, and TD Kocher. 2002. Rapid isolation of CA microsatellites from the tilapia genome. Anim Genet, 33, pp. 140–144.

Chang DK, D Metzgar, C Wills, and CR Boland. 2001. Microsatellites in the eukaryotic DNA mismatch repair genes as modulators of evolutionary mutation rate. Genome Res, 11, pp. 1145–1146.

Chistiakov DA, B Hellemans, CS Haley, AS Law, CS Tsigenopoulos, G Kotoulas, D Bertotto, A Libertini, and FA Volckaert. 2005. A microsatellite linkage map of the European sea bass *Dicentrarchus labrax* L. Genetics, 170, pp. 1821–1826.

Coimbra MRM, K Kobayashi, S Koretsugu, O Hasegawa, E Ohara, A Ozaki, T Sakamoto, K Naruse, and N Okamoto. 2005. A genetic linkage map of the Japanese flounder, *Paralichthys olivaceus*. Aquaculture, 220, pp. 203–218.

Crawford AM and RP Cuthbertson. 1996. Mutations in sheep microsatellites. Genome Res, 6, pp. 876–879.

Crollius HR, O Jaillon, C Dasilva, C Ozouf-Costaz, C Fizames, C Fischer, L Bouneau, A Billault, F Quetier, W Saurin, A Bernot, and J Weissenbach. 2000. Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis. Genome Res, 10, pp. 939–949.

Edwards YJ, G Elgar, MS Clark, and MJ Bishop. 1998. The identification and characterization of microsatellites in the compact genome of the Japanese pufferfish, *Fugu rubripes*: perspectives in functional and comparative genomic analyses. J Mol Biol, 278, pp. 843–854.

Ellegren H. 2000. Microsatellite mutations in the germline: implications for evolutionary inference, Trends Genet, 16, pp. 551–558.

Estoup A and J Cornuet. 1999. Microsatellite evolution: inferences from population data, pp. 49–65 In *Microsatellites: Evolution and Applications*, edited by DB Goldstein and C Schlotterer. Oxford University Press, New York.

FitzSimmons NN, C Moritz, and SS Moore. 1995. Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution. Mol Biol Evol, 12, pp. 432–440.

Garoia F, I Guarniero, and F Tinti. 2003. Polymorphic microsatellites for the Mediterranean angler species (*Lophiidae*). Mol Ecol Notes, 3, pp. 294–296.

Gharbi K, A Gautier, RG Danzmann, S Gharbi, T Sakamoto, B Hoyheim, JB Taggart, M Cairney, R Powell, F Krieg, N Okamoto, MM Ferguson, LE Holm, and R Guyomard. 2006.

A linkage map for brown trout (Salmo trutta): chromosome homeologies and comparative genome organization with other salmonid fish. Genetics, 172, pp. 2405–2419.
Gilbey J, E Verspoor, A McLay, D Houlihan. 2004. A microsatellite linkage map for Atlantic salmon (*Salmo salar*). Anim Genet, 35, pp. 98–105.
Hubert S and D Hedgecock. 2004. Linkage maps of microsatellite DNA markers for the Pacific oyster *Crassostrea gigas*. Genetics, 168, pp. 351–62.
Jakupciak JP and RD Wells. 2000. Genetic instabilities of triplet repeat sequences by recombination. IUBMB Life, 50, pp. 355–359.
Ju Z, A Karsi, A Kocabas, A Patterson, P Li, D Cao, R Dunham, and ZJ Liu. 2000. Transcriptome analysis of channel catfish (*Ictalurus punctatus*): genes and expression profile from the brain. Gene, 261, pp. 373–382.
Karsi A, D Cao, P Li, A Patterson, A Kocabas, J Feng, Z Ju, K Mickett, and ZJ Liu. 2002. Transcriptome analysis of channel catfish (*Ictalurus punctatus*): Initial analysis of gene expression and microsatellite-containing cDNAs in the skin. Gene, 285, pp. 157–168.
Kijas JM, JC Fowler, CA Garbett, MR Thomas. 1994. Enrichment of microsatellites from the citrus genome using biotinylated oligonucleotide sequences bound to streptavidin-coated magnetic particles. Biotechniques, 16, pp. 656–660.
Knapik EW, A Goodman, M Ekker, M Chevrette, J Delgado, S Neuhauss, N Shimoda, W Driever, MC Fishman, HJ Jacob. 1998. A microsatellite genetic linkage map for zebrafish (*Danio rerio*). Nat Genet, 18, pp. 338–343.
Kocher TD, WJ Lee, H Sobolewska, D Penman, B McAndrew. 1998. A genetic linkage map of a cichlid fish, the tilapia (*Oreochromis niloticus*). Genetics, 148, pp. 1225–1232.
Kolodner RD, GT Marsischky. 1999. Eukaryotic DNA mismatch repair. Curr Opin Genet Dev, 9, pp. 89–96.
Leclerc D, T Wirth, and L Bernatchez. 2000. Isolation and characterization of microsatellite loci in the yellow perch (*Perca flavescens*), and cross-species amplification within the family Percidae. Mol Ecol, 9, pp. 995–997.
Lee BY, WJ Lee, JT Streelman, KL Carleton, AE Howe, G Hulata, A Slettan, JE Stern, Y Terai, and TD Kocher. 2005. A second-generation genetic linkage map of tilapia (*Oreochromis spp.*). Genetics, 170, pp 237–244.
Levinson G and GA Gutman. 1987. High frequency of short frameshifts in poly-CA/GT tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. Nucl Acids Res, 15, pp. 5323–5338.
Li W-H. 1997. *Molecular Evolution.* Sinauer Associates, Sunderland, MA, pp. 177–213.
Liu ZJ. 2003. A review of catfish genomics: progress and perspectives. Comp Funct Genomics, 4, pp. 259–265.
Liu ZJ and JF Cordes. 2004a. DNA marker technologies and their applications in aquaculture genetics. Aquaculture, 238, pp. 1–37.
Liu ZJ and JF Cordes. 2004b. Erratum to DNA marker technologies and their applications in aquaculture genetics. Aquaculture, 242, pp. 735–736.
Liu ZJ, A Karsi, and RA Dunham. 1999. Development of polymorphic EST markers suitable for genetic linkage mapping of catfish. Mar Biotechnol, 1, pp. 437–447.
Liu ZJ, P Li, A Kocabas, Z Ju, A Karsi, D Cao, and A Patterson 2001. Microsatellite-containing genes from the channel catfish brain: evidence of trinucleotide repeat expansion in the coding region of nucleotide excision repair gene RAD23B. Biochem Biophys Res Comm, 289, pp. 317–324.
Lyall JE, GM Brown, RA Furlong, MA Ferguson-Smith, and NA Affara. 1993. A method for creating chromosome-specific plasmid libraries enriched in clones containing [CA]n microsatellite repeat sequences directly from flow-sorted chromosomes. Nucleic Acids Res. 21, pp. 4641–4642.
McConnell SK, C Beynon, J Leamon, and DO Skibinski. 2000. Microsatellite marker based genetic linkage maps of *Oreochromis aureus* and *O. niloticus* (Cichlidae): extensive linkage group segment homologies revealed. Anim Genet, 31, pp. 214–218.

Metzgar D, J Bytof, and C Wills. 2000. Selection against frameshift mutations limits microsatellite expansion in coding DNA. Genome Res, 10, pp. 72–80.

Moen T, B Hoyheim, H Munck, and L Gomez-Raya. 2004. A linkage map of Atlantic salmon (*Salmo salar*) reveals an uncommonly large difference in recombination rate between the sexes. Anim Genet, 35, pp. 81–92.

Moran C. 1993. Microsatellite repeats in pig (*Sus domestica*) and chicken (*Gallus domesticus*) genomes. J Hered, 84, pp. 274–280.

Nei M and AK Roychoudhury. 1974. Sampling variances of heterozygosity and genetic distance. Genetics, 76, pp. 379–390.

Nichols KM, WP Young, RG Danzmann, BD Robison, C Rexroad, M Noakes, RB Phillips, P Bentzen, I Spies, K Knudsen, FW Allendorf, BM Cunningham, J Brunelli, H Zhang, S Ristow, R Drew, KH Brown, PA Wheeler, and GH Thorgaard. 2003. A consolidated linkage map for rainbow trout (*Oncorhynchus mykiss*). Anim Genet, 34, pp. 102–115.

O'Brien SJ. 1991. Molecular genome mapping lessons and prospects. Curr Opin Genet Dev, 1, pp. 105–111.

O'Connell M and JM Wright. 1997. Microsatellite DNA in fishes. Rev Fish Biol Fish, 7, pp. 331–363.

Ohara E, T Nishimura, Y Nagakura, T Sakamoto, K Mushiake, and N Okamoto. 2005. Genetic linkage maps of two yellowtails (*Seriola quinqueradiata* and *Seriola lalandi*). Aquaculture, 244, pp. 41–48.

O'Reilly P and JM Wright. 1995. The evolving technology of DNA fingerprinting and its application to fisheries and aquaculture. J Fish Biol, 47, pp. 29–55.

Ostrander EA, PM Jong, J Rine, and G Duyk. 1992. Construction of small-insert genomic DNA libraries highly enriched for microsatellite repeat sequences. Proc Natl Acad Sci, USA, 89, pp. 3419–3423.

Ott J. 1992. Strategies for characterizing highly polymorphic markers in human gene mapping. Am J Hum Genet, 51, pp. 283–290.

Powell W, M Morgante, C Andre, M Hanafey, J Vogel, S Tingey, and A Rafalski. 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Molecular Breeding, 2, pp. 225–238.

Rico C, I Rico, and G Hewitt. 1996. 470 million years of conservation of microsatellite loci among fish species, Proc R Soc Lond B Biol Sci, 263, pp. 549–557.

Sakamoto T, RG Danzmann, K Gharbi, P Howard, A Ozaki, SK Khoo, RA Woram, N Okamoto, MM Ferguson, LE Holm, R Guyomard, and B Hoyheim. 2000. A microsatellite linkage map of rainbow trout (*Oncorhynchus mykiss*) characterized by large sex-specific differences in recombination rates. Genetics, 155, 1331–1345.

Sambrook J, EF Fritsch, and T Maniatis. 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Schlötterer C, B Amos, and D Tautz. 1991. Conservation of polymorphic simple sequence loci in cetacean species. Nature, 354, pp. 63–65.

Serapion J, H Kucuktas, J Feng, and Z Liu. 2004. Bioinformatic mining of type I microsatellites from expressed sequence tags of channel catfish (*Ictalurus punctatus*). Mar Biotechnol, 6, pp. 364–377.

Shimoda N, EW Knapik, J Ziniti, C Sim, E Yamada, S Kaplan, D Jackson, F de Sauvage, H Jacob, and MC Fishman. 1999. Zebrafish genetic map with 2,000 microsatellite markers. Genomics, 58, pp. 219–232.

Sia EA, RJ Kokoska, M Dominska, P Greenwell, TD Petes. 1997. Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes. Mol Cell Biol, 17, pp. 2851–2858.

Sun X and L Liang. 2004. A genetic linkage map of common carp (*Cyprinus carpio* L.) and mapping of a locus associated with cold tolerance. Aquaculture, 238, pp. 165–172.

Tautz, D. 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucl Acids Res, 17, pp. 6563–6571.

Toth G, Z Gaspari, and J Jurka. 2000. Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res, 10, pp. 967–981.

Van Lith HA and LF Van Zutphen. 1996. Characterization of rabbit DNA microsatellites extracted from the EMBL nucleotide sequence database. Anim Genet, 27, pp. 387–395.

Waldbieser GC, BG Bosworth, DJ Nonneman, and WR Wolters. 2001. A microsatellite-based genetic linkage map for channel catfish, *Ictalurus punctatus*. Genetics, 158, pp. 727–734.

Walter RB, JD Rains, JE Russell, TM Guerra, C Daniels, DA Johnston, J Kumar, A Wheeler, K Kelnar, VA Khanolkar, EL Williams, JL Hornecker, L Hollek, MM Mamerow, A Pedroza, and S Kazianis. 2004. A microsatellite genetic linkage map for *Xiphophorus*. Genetics, 168, pp. 363–372.

Watanabe T, H Fujita, K Yamasaki, S Seki, and N Taniguchi. 2004. Preliminary study on linkage mapping based on microsatellite DNA and AFLP markers using homozygous clonal fish in ayu (*Plecoglossus altivelis*). Mar Biotechnol, 6, pp. 327–334.

Weber JL and C Wong. 1993. Mutation of human short tandem repeats. Hum Mol Genet, 2, pp. 1123–1128.

Woram RA, C McGowan, JA Stout, K Gharbi, MM Ferguson, B Hoyheim, EA Davidson, WS Davidson, C Rexroad, and RG Danzmann. 2004. A genetic linkage map for Arctic char (*Salvelinus alpinus*): evidence for higher recombination rates and segregation distortion in hybrid versus pure strain mapping parents. Genome, 47, pp. 304–315.

Xu P, S Wang, L Liu, E Peatman, B Somridhivej, J Thimmapuram, G Gong, and ZJ Liu. 2006. Generation of channel catfish BAC end sequences for marker development and assessment of syntenic conservation with other fish species. Anim Genet, 37, pp. 321–326.

Yu Z and X Guo. 2003. Genetic linkage map of the eastern oyster *Crassostrea virginica* Gmelin. Biol Bull, 204, pp. 327–338.

Zane L, L Bargelloni, and T Patarnello. 2002. Strategies for microsatellite isolation: a review. Mol Ecol, 11, pp. 1–16.