Review

Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics

Dimitry A. Chistiakov a,b,⁎, Bart Hellemans b, Filip A.M. Volckaert b

a Department of Pathology and Laboratory Medicine, University of Cincinnati Medical Center, 231 Albert Sabin Way, Cincinnati, OH 45267-0529, USA
b Laboratory of Aquatic Ecology, Katholieke Universiteit Leuven, Ch. de Bériotstraat 32, B-3000 Leuven, Belgium

Received 31 March 2005; received in revised form 17 November 2005; accepted 25 November 2005

Abstract

Microsatellites represent codominant molecular genetic markers, which are ubiquitously distributed within genomes. Due to their high level of polymorphism, relatively small size and rapid detection protocols, these markers are widely used in a variety of fundamental and applied fields of life and medical sciences. In the field of aquaculture, microsatellites represent workhorse markers, which are useful for the characterization of genetic stocks, broodstock selection, constructing dense linkage maps, mapping economically important quantitative traits, identifying genes responsible for these traits and application to marker-assisted breeding programmes. In this review, genomic distribution, function, evolution and practical applications of microsatellites are considered, with special emphasis on fish genetics and aquaculture.

Keywords: Evolution; Functional relevance; Genomic distribution; Microsatellite; SSR; Teleostei

Contents

1. Introduction ....................................................... 2
2. Genomic distribution of microsatellites ......................................... 2
3. Evolution of microsatellites ................................................ 3
4. Function of microsatellites ................................................ 5
  4.1. DNA structure ................................................... 5
  4.2. DNA recombination ................................................ 5
  4.3. DNA replication .................................................. 5
  4.4. Gene expression .................................................. 5
5. Development of type I (coding) and type II (non-coding) markers . ....................... 7

⁎ Corresponding author. Department of Pathology and Laboratory Medicine, University of Cincinnati Medical Center, 231 Albert Sabin Way, Cincinnati, OH 45267-0529, USA. Tel.: +1 513 558 4402; fax: +1 513 558 2141.
E-mail address: dmitry.chistyakov@uc.edu (D.A. Chistiakov).

© 2005 Elsevier B.V. All rights reserved.

0044-8486/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
1. Introduction

Microsatellites, or simple sequence repeats (SSRs), represent a unique type of tandemly repeated genomic sequences, which are abundantly distributed across genomes and demonstrate high levels of allele polymorphism. They are codominant markers of relatively small size, which can be easily amplified with the polymerase chain reaction. These features provide the foundation for their successful application in a wide range of fundamental and applied fields of biology and medicine, including forensics, molecular epidemiology, population and conservation genetics, genetic mapping and genetic dissection of complex traits. In the field of fisheries and aquaculture, microsatellites are useful for the characterization of genetic stocks, broodstock selection, constructing dense linkage maps, mapping economically important quantitative traits and identifying genes responsible for these traits and application in marker-assisted breeding programmes. Although microsatellites are considered selectively neutral markers, they often represent functionally relevant polymorphisms. SSRs contribute to DNA structure, chromatin organization, regulation of DNA recombination, transcription and translation, gene expression and cell cycle dynamics. To date, microsatellites are far from being completely identified and characterised, providing intriguing perspectives for discovery of new properties and characteristics of SSRs, which will help the design of new research fields and the practical use of these markers.

2. Genomic distribution of microsatellites

Microsatellites are stretches of DNA consisting of tandemly repeated short units of 1–6 base pairs (bp) in length. SSRs typically span between twenty and a few hundred bases (Beckmann and Weber, 1992). They are ubiquitous in prokaryotes and eukaryotes, present even in the smallest bacterial genomes (Gur-Arie et al., 2000). The existence of SSRs in eukaryotic genomes has been known since the 1970s (Bruford et al., 1996). Hamada et al. (1982) demonstrated a large number and wide occurrence of these sequences from yeast through to vertebrates. Tautz and Renz (1984) hybridized different microsatellite sequences to genomic DNA from a variety of organisms and reported many types of simple sequences.

The majority of microsatellites (30–67%) found are dinucleotides. In the genome of vertebrates, (AC)$_n$ is the most common dinucleotide motif. It is 2.3-fold more frequent than (AT)$_n$, the second most general type of dinucleotides (Toth et al., 2000). Interestingly, in primates mononucleotide repeats are mostly represented by poly (A/T) tracts, which are the most frequent classes of SSRs (Beckmann and Weber, 1992). In total, higher-order SSR classes (tri-, tetra-, penta-and hexanucleotides) are about 1.5-fold less common in genomic DNA of vertebrates than dinucleotides (Toth et al., 2000). In the genome of Japanese pufferfish or fugu 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). In total, 1.29% of the genome of Japanese pufferfish consists of microsatellites. For the closely related spotted green pufferfish Tetraodon nigroviridis, SSRs cover 3.21% of the genome (Crollius et al., 2000). In fugu, one microsatellite is found every 1.87 kilobases (kb) of DNA, and a CA repeat (the most common type of tandem repeat) occurs every 6.56 kb of DNA. In three-spined stickleback Gasterosteus aculeatus, CA dinucleotides also are the most common type of microsatellites, occurring approximately once every 14 kb (Peichel et al., 2001). For comparison, in the human genome, one microsatellite was found every 1.87 kilobases (kb) of DNA, and a CA repeat (the most common type of tandem repeat) occurs every 6.56 kb of DNA. In three-spined stickleback Gasterosteus aculeatus, CA dinucleotides also are the most common type of microsatellites, occurring approximately once every 14 kb (Peichel et al., 2001). For comparison, in the human genome, one microsatellite was found every 6 kb and one CA repeat occurred every 30 kb of DNA (Beckmann and Weber, 1992).
Microsatellites can be found anywhere in the genome, both in protein-encoding and noncoding DNA (Toth et al., 2000). In eukaryotic organisms, SSRs have been shown to be in excess in noncoding regions compared to a random distribution pattern (Metzgar et al., 2000). They are relatively rare in coding DNA, ranging between 7–10% in higher plants (Wang et al., 1994; Varshney et al., 2002) and between 9–15% in vertebrates (Moran, 1993; Jurka and Pethiyagoda, 1995; Van Lith and Van Zutphen, 1996). Only 11.6% of a total of 6042 microsatellites were found in protein-coding regions in the genome of Japanese pufferfish (Edwards et al., 2000). Analysis of perfect dimeric SSRs in intronic sequences and intergenic regions, respectively, revealed 42- and 30-fold less frequently in exons than in noncoding regions fits the unbiased single-step mutation model (Bell and Jurka, 1997). This model suggests that repeats change length by plus or minus one unit with equal probabilities, and that base substitutions destroy long perfect repeats, producing two shorter tracts of perfect repeats (see later). As a consequence, long dimeric motifs are highly unstable within expressed sequences, while in noncoding regions most dinucleotide repeats can have surprisingly long stretches, probably due to the high tolerance of noncoding DNA to mutations (Dokholyan et al., 2000). The potential size expansion of di- or tetranucleotide microsatellites in untranslated regions (UTRs) and introns could lead to disruption of native protein and/or formation of new genes with frame-shift (Liu et al., 1999a). These patterns suggest that random distribution of such di- and tetranucleotide SSRs are strongly selected against (Bachtrog et al., 1999).

Dinucleotide repeats in 5'- and 3'-UTRs have been described within genes of a variety of fish species, including channel catfish Ictalurus punctatus (Liu et al., 1999a), Atlantic salmon Salmo salar (Grimholt et al., 2002), zebrafish Danio rerio (Gerhard et al., 2000), Japanese flounder Paralichthys olivaceus (Hirono et al., 2000) and Nile tilapia Oreochromis niloticus (Mansour et al., 1998). Dinucleotide SSRs are also found in introns. For example, intronic dinucleotide microsatellites have been detected in the growth hormone gene of Nile tilapia, barramundi Lates calcarifer, Japanese flounder and Japanese pufferfish (Venkatesh and Brenner, 1997).

In contrast to other types of repeat motifs, triplets are found in both coding and non-coding genomic regions with a high frequency (Wren et al., 2000; Morgante et al., 2002). In all vertebrates, (G+C)-rich motifs (e.g., CCG, CAG) are the most common among trinucleotides. These repeats dominate in exons, whereas they are less common in intronic sequences (Toth et al., 2000). In humans, the expansion of trinucleotides, encoding polyproline (CCG)_n, polyarginine (CGG)_n, polyalanine [(GCC)_n and (GCG)_n] and polyglutamine (CAG)_n tracts within exons has been described. Such expansions can lead to various neurodegenerative and neuromuscular disorders, including myotonic dystrophy, fragile X syndrome, Huntington's disease and spinocerebellar ataxia (Jasinska et al., 2003; Brown and Brown, 2004). Naturally occurring triplet repeat instability of transcribed sequences occurs not only in humans, but also in lower vertebrates. For example, length variability of the signal peptide-encoding region of the melanoma receptor tyrosine kinase (XMRK) gene has been described in the swordtail genus Xiphophorus (Schartl et al., 1998). The signal peptide contains a variable number of CTG repeats, which may differ in length even between closely related individuals. The XMRK locus, encoding a sex-linked oncogene, and responsible for melanoma formation, is located in a highly unstable genomic region (Froschauer et al., 2001). Variability in the signal peptide could influence the efficiency of the functional protein export through the cell membrane and therefore affect fitness traits of individuals expressing the melanoma phenotype. In channel catfish, variable (ACC)_n repeats encoding a polythreonine tract have been found within the RAD23B gene, which is important in the nucleotide excision repair system (Liu et al., 2001). A polymorphic (CAA)_n trinucleotide motif encoding a polyglutamine stretch was observed in the sex-linked NROB1 (DAX1) gene of the European sea bassDicentrarchus labrax (Chistiakov and Hellemans, unpublished data).

### 3. Evolution of microsatellites

The key feature of SSRs as molecular markers is their hypermutability and, hence, their hypervariability in species and populations. The microsatellite mutation rate is estimated at 10^{-2}–10^{-6} per locus per generation (Ellegren, 2000), which is several orders of magnitude greater than that of regular nonrepetitive DNA (10^{-9}; Li, 1997). Analysis of (AC)_n microsatellites in five vertebrate classes (mammals, birds, reptiles, amphibians
and fishes) showed that length is a major factor influencing mutation rate. A directional mutation toward an increase in microsatellite length was also observed (Neff and Gross, 2001).

Two models (DNA polymerase slippage and unequal recombination) have been suggested to explain microsatellite generation and evolution. The first model involves transient dissociation of the replicating DNA strands with subsequent reassociation (Schlötterer and Tautz, 1992; Richards and Sutherland, 1994). The slipped structure can be stabilized by hairpin, triplex or quadruplex arrangement of DNA strands (Sinden, 1999). (CCG)_n × (CGC)_n, (AGG)_n × (CCT)_n and some other triplet motifs are shown to have high hairpin-forming potential and, therefore, could form stable alternative structures, which are implicated in the triplet expansion diseases (Usdin, 1998; Sinden, 1999). Since DNA repeat regions represent preferred target sites for mutations during DNA replication, 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 tumour 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).

Although a number of experimental findings argue in favour of the above model, nonreciprocal recombination (gene conversion) also may play a role in genetic instability of some SSRs, including triplet motifs (Jakupciak and Wells, 2000). Gene conversion mechanisms were found to be involved in the differentiation and evolution of paralogous sequences (duplicated loci within species) in members of the family Salmonidae derived from tetraploidization. This observation was taken from the comparative sequence analysis of a simple microsatellite locus Str1INRA, which contains long flanking sequences, including both coding and noncoding regions, in different species of subfamily Salmoninae (Angers et al., 2002). Replication slippage and recombination could interact, affecting stability of microsatellite loci. For example, such “repair-slippage” mechanisms are probably involved in the instability of tri- and dinucleotides in yeast (Richard and Dujon, 1997), polarity of substitutions within repeats in (CA)_n stretches in ovine and bovine genomes (Brohede and Ellegren, 1999), conservation of the dinucleotide microsatellite locus Loc6 in lamniform sharks (Martin et al., 2002) and evolution of intronic microsatellites and some haplotypes within orthologous Rhesus (RH) genes in vertebrates (Okuda and Kajii, 2002).

Slipped-strand mispairing during DNA replication is likely to represent the predominant mutational mechanism for microsatellites (Schlötterer and Tautz, 1992). It results in the nascent strand having a different number of repeats from the template strand once DNA replication is complete. This mutation process allows the same microsatellite allele to arise multiple times, thereby generating size homoplasy. Homoplasy represents similarity of traits or genes for reasons other than common ancestry (e.g., convergent evolution, parallelism, evolutionary reversals, horizontal gene transfer, gene duplications). Homoplasy can violate a basic assumption of the analysis of genetic markers, in which variants of similar phenotype (e.g., base pair size) are assumed to derive from a common ancestor (Sanderson and Hufford, 1996).

Various mutational models that account for this homoplasy have been proposed. The stepwise mutational model (SMM) assumes that all mutational events involve a change in a single repeat only (Kimura and Ohta, 1978; Bell and Jurka, 1997), whereas the two-phase mutational model (TPM) allows a proportion of mutations to involve changes greater than single repeats (Di Rienzo et al., 1994). In contrast, the infinite-allele model (IAM) refuses homoplasy events, suggesting that every mutation results in the creation of a new allele (Kimura and Crow, 1964). Determining which mutational model is most appropriate is important, because estimation of microsatellite-specific genetic distances among populations relies on the underlying assumptions of the chosen model.

During DNA replication, longer stretches of repeated units pose more of a problem to DNA polymerase than do shorter stretches, making longer alleles more prone to slipped-strand mispairing. In addition, larger numbers of repeats provide more opportunities for misalignment during the reannealing of the nascent strand (Eisen, 1999). Therefore, there is thought to be a rough threshold of minimum repeat number below which a microsatellite is not likely to mutate or be variable. Estimations of slippage frequency during polymerase chain reaction (PCR) showed that such a threshold exists and is equal to four and eight repeats.
for (CA/GT)_n dinucleotides and (A/T)_n motifs, respectively (Shinde et al., 2003). Measurements in slime mold Neurospora found that microsatellites are expected to be variable if they have a mean of 5.1 repeats (Dettman and Taylor, 2004).

4. Function of microsatellites

SSRs typically represent selectively neutral DNA markers. However, multiple studies proved the functional relevance of a significant number of SSRs.

4.1. DNA structure

Microsatellites are involved in forming a wide variety of unusual DNA structures with simple and complex loop-folding patterns. Double-stranded alternating purine and pyrimidine sequences such as the (dC−dA)×(dG−dT)_{32} form left-handed Z-DNA structures in vitro (Rich et al., 1984). Several microsatellite sequences, such as (GAA)_n, (AC)_n and composite (GT)_n×(GA)_n simple repeats, also exhibit non-B-DNA structural properties (Epplen et al., 1996).

Telomeric and centromeric chromosome regions have been shown to be rich in long arrays of a variety of mono-, di-, tri-, tetra- and hexanucleotide motifs. Satellite sequences enriched by AT-dinucleotides have been found in the centromeric DNA of various gobid species (Canapa et al., 2002). They are considered important for the control of centromeric chromatin compactness in these fishes. A cryptic RRY(i) microsatellite located close to the centromeric region of an acrocentric chromosome pair was characterized in Atlantic salmon (Martinez et al., 2001). The expanded stretches of a simple repeat sequence (TTAGGG)_n oriented in the 5’ to 3’ direction towards the end of eukaryotic chromosomes constitute a substantial portion of the repetitive DNA in telomeric regions (Henderson, 1995). For example, the telomeric repeat in the Nile tilapia varies in size from 4 to 10 kb (Chew et al., 2002). Telomere-associated repeats can be related to nucleolar organizing regions (NORs), as seen in rainbow trout Oncorhynchus mykiss (Abuin et al., 1996), lake trout Salvelinus namaycush (Reed and Phillips, 1995) and Nile tilapia (Foresti et al., 1993). The (TTAGGG)_n hexamer sequence is recognized by ribonucleoprotein polymerase, a telomerase, which synthesizes telomere repeats onto the chromosome ends to overcome the loss of sequences during DNA replication, whereas other proteins prevent nucleolytic degradation and confer stability of chromosomes (Fang and Sech, 1995; Martins et al., 2004). Therefore, SSRs play an important role in the organization of the chromosome structure.

4.2. DNA recombination

SSRs are considered hot spots for recombination (Jeffreys et al., 1998). Dinucleotide motifs are preferential sites for recombination events due to their high affinity for recombination enzymes (Biet et al., 1999). Some SSR sequences, such as GT, CA, CT, GA and others, may influence recombination directly through their effects on DNA structure (Biet et al., 1999). The orthologous rhesus (Rh) genes, which are responsible for the determination of the blood group in higher vertebrates, were found to contain multiple microsatellites in their introns (Okuda et al., 2000). These SSRs were shown to be associated with the assignment of some Rh phenotypes, and to be involved in the molecular evolution of the human Rh gene family and its orthologs in other eukaryotes and Archaea via replication slippage and recombination (gene conversion) mechanisms (Fujiwara et al., 1999; Okuda and Kajii, 2002).

4.3. DNA replication

SSRs may influence DNA replication. For example, in rat cells, DNA amplification is terminated within a specific fragment which consists of a d(GA)_{27}×d(TC)_{27} tract. This sequence is situated at the end of an amplicon and forms a loop, which serves as a stop signal for DNA polymerase (Li et al., 2002).

Human genes encoding important cell fidelity and growth factors, such as the B-cell leukemia/lymphoma 2 (BCL2)-associated X protein, insulin-like growth factor 2 receptor (IGF2R), breast cancer early onset protein 2 (BRCA2) and transforming growth factor beta 2 (TGF-β2), contain short repeated sequences. MMR deficiency causes frame-shift mutations, resulting in both insertions and deletions of repeat units within these sequences that affect these genes and could therefore initiate tumorigenesis (Johannsdottir et al., 2000). These observations suggest that microsatellites can affect enzymes controlling mutation rate and cell cycles (Chang et al., 2001).

4.4. Gene expression

Numerous data show that SSRs located in promoter regions can influence gene expression. The 5’ upstream region of the insulin gene of Nile tilapia contains a microsatellite close to the same position of a unique
dissociated for a dinucleotide (CA/GT)ₙ microsatellite in
ly influence gene expression level. Such an effect was
al., 1999). SSRs in introns might serve as target
factor-binding sites of promoter regions of many
important genes (Edwards et al., 1998). These SSRs
could affect the sequence of the binding site and there-
fore influence its affinity for the binding of a
corresponding regulatory transcription factor. The tran-
scription-regulating activity of the (CT)ₙ tract in pro-
moters of the HSP26 gene encoding a 26 kDa heat-
shock protein in Drosophila (Sandaltzopoulos et al.,
1995), the gene for glyceraldehyde-3-phosphate dehy-
drogenase in Aspergillus (Punt et al., 1990) and the
piypt1 gene of Phytophthora (Chen and Roxby, 1997)
could serve as examples. The CT-element lying close to
the transcription start point of these genes was shown to
be the target sequence for binding transcription factors
and might be important in determining the frequency of
transcription initiation (Chen and Roxby, 1997).

In many cases, SSR repeat number could significant-
ly influence gene expression level. Such an effect was
described for a dinucleotide (CA/GT)ₙ microsatellite in
the Nile tilapia prolactin 1 (PRL1) promoter (Streelman
and Kocher, 2002). Individuals homozygous for long
microsatellite alleles express less PRL1 in freshwater,
but more in half-seawater than fish with other geno-
types. Interestingly, a similar activity was previously
reported by Naylor and Clark (1990) for TG/CA repeat
sequences in the promoter of the rat prolactin gene. It
suggests the conservation of the regulatory function for
CA/GT microsatellites in the PRL1 promoter over 300
million years of vertebrate evolution.

Transcribed microsatellites located in 5′ untranslated
regions (UTRs) could form specific and unusual DNA
structures. In that case, the length of the repeat region
could affect the translation level from the target mRNA.
The apparent correlation between gene expression and
the number of tandem GAA repeat motifs was observed
for the GAA repeat region which regulates expression
of the M9/pMGA gene family in the avian intracellular
parasite Mycoplasma gallisepticum (Liu et al., 2000).

Intronic SSRs also can affect gene transcription. For
example, such an effect was measured for the tetrameric
microsatellite located in intron 1 of the human tyrosine
hydroxylase gene (Meloni et al., 1998) and the (CA)ₙ
dinucleotide repeat in the first intron of the human
epidermal growth factor receptor gene (Gebhardt et
al., 1999). SSRs in introns might serve as target
sequences for binding a variety of expression-regulating
proteins, as was shown for certain (GT)ₙ or mixed GT/
GA stretches of intronic simple repeats. They have been
preserved in immunologically relevant genes for at least
70 million years and bind nuclear protein regulatory
molecules with high affinities (Epplen et al., 1993;
1996). A number of such SSR repeats could also corre-
late with the strength and affinity of the protein binding
(Winter and Varshavsky, 1989).

Microsatellites situated in the 3′-UTR could affect
gene expression through their influence on the stability
of transcribed products. This role was found for a GA-
rich repetitive DNA segment in the 3′-UTR of the rat
polymeric immunoglobulin receptor gene (Fabregat et
al., 2001) and the chicken elastin gene (Hew et al.,
2000). Long stretches of such polypurine and pyrimi-
dine repeat motifs at 3′-UTRs could destabilize the
structure of the 3′-end of a mRNA molecule and
hence facilitate its availability for degradation by intra-
cellular exonucleases (Wang et al., 2002). Otherwise,
such repeats could affect mRNA stability, representing
binding sites for translation factors, as has been de-
scribed for an AU-rich sequence in the 3′-UTR of
mRNA for human plasminogen activator inhibitor
type 2 (Maurer et al., 1999).

In the European eel Anguilla anguilla, a highly vari-
able region was found at the 3′-UTR of the TSHB gene
encoding thyrotropin β subunit (Pradet-Balade et al.,
1998). The region includes repeat units, which contain
CTG double repeats at their ends. The CTG double
repeats are able to form unusual helix structures within
double-stranded DNA (Chastain and Sinden, 1998).
Such configurations may, therefore, promote genetic
instability and be responsible for heterogeneity in the
number and length of the thyrotropin β mRNA in the
European eel (Pradet-Balade et al., 1997).

The above evidence suggests that SSR variation can
produce either drastic or quantitative variations in gene
expression. Because of genomic overabundance and
high mutability of SSRs, changes in SSR array size
may serve as a rich source of variation in fitness-related
traits in natural populations (Kashi et al., 1997; Streel-
man and Kocher, 2002). Its role may be especially
important for population survival and adaptation to
spatially and temporarily varying environmental condi-
tions (Blankenship et al., 2002). The presence of so-
called contingency loci in many bacterial species may
explain how microsatellite variability influences the
adaptive evolution of microbial pathogens (Metzgar
and Wills, 2000). These loci, containing tandem repeats
within either a coding sequence or a promoter, can be
hypermutable. Altered numbers of repeats thus cause
switches in the reading frame of translation or changes in the level of promoter activity. The contingency loci are specifically associated with genes controlling the antigenic type and/or phase state of pathogenic bacteria (Bayliss et al., 2001). Examples of genes carrying these hypermutable repeats include those encoding surface polysaccharides of Haemophilus influenzae, variable surface lipoproteins of Mycoplasma hyorhinis and opacity proteins of Neisseria sp. (Yoge et al., 1991; Bayliss et al., 2001). The hypervariability of these genes allows clonal lineages to switch rapidly between serotypes and thereby evade clearance by the host immune system, or to switch rapidly between different phenotypic 'phases' required for the invasion of multiple tissue types (Moxon et al., 1994).

5. Development of type I (coding) and type II (non-coding) 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 they commonly are located in noncoding intergenic regions. The fastest and simplest way to detect and characterize a large number of such type II microsatellites lies in the construction of small-insert genomic libraries enriched in arrays of tandem repeats (Zane et al., 2002). The enrichment technique usually includes selective hybridization of fragmented genomic DNA with a tandem repeat-containing oligonucleotide probe and further PCR amplification of the hybridization products. Libraries highly enriched by tandem repeats have been constructed for many organisms, including fishes. The high frequency of tandem repeats in fish genomes provides a good opportunity to obtain libraries significantly enriched in microsatellites. For example, libraries containing 74%, 95% and 96% clones with (CA)n repeats have been developed for the Mediterranean angler fish Lophius sp. (Garoia et al., 2003), gilthead sea bream Sparus aurata (Zane et al., 2002), and Nile tilapia (Carleton et al., 2002), respectively. A library usually contains 1000–4000 recombinant clones. Screening of these clones typically yields 10–15% unique polymorphic SSRs, resulting in the production of 100–500 non-redundant variable microsatellites from a single library (Zane et al., 2002).

In addition, type II markers can be rapidly developed from SSRs isolated previously from closely related species (Bruford et al., 1996). Cross-species amplification of SSRs provides a possibility to superimpose the genetic information or a genetic linkage map from one species to another and therefore to use SSR markers for population genetics, parentage analysis and other applications without having to invest in the isolation of polymorphic microsatellites (Cairney et al., 2000; Leclerc et al., 2000). The presence of highly conserved flanking regions has been reported for some microsatellite loci in cetaceans (Schlöterer et al., 1991), turtles (FitzSimmons et al., 1995) and fishes (Rico et al., 1996; Angers and Bernatchez, 1996), allowing cross-amplification from species that diverged as long ago as 470 million years.

Type I markers are more difficult to develop (Liu et al., 1999b). While non-gene sequences are free to mutate, causing higher levels of polymorphism, sequences within protein-coding regions generally show lower levels of polymorphism because of functional selection pressure. 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 a cDNA clone can be sequenced to produce expressed sequence tags (ESTs). An EST represents a short, usually 200–600 bp-long nucleotide sequence, which represents a uniquely expressed region of the genome. If the EST harbours any polymorphic type I marker [usually SSR and/or single nucleotide polymorphism (SNP)], it can be mapped (see later). cDNA libraries can be routinely sequenced, rapidly producing a bulk of ESTs, with which to organize an EST collection. Such collections provide a robust sequence resource that can be used for gene discovery, genome annotation and comparative genetics (Rudd, 2003; Dunham, 2004; Ng et al., 2005).

EST sequences are archived in a special branch of the GenBank nucleotide database (dbEST) (http://www.ncbi.nlm.nih.gov/dbEST/index.html) (Wheeler et al., 2004). In November 2005, the EST database contained more than 31.3 million sequence entries from around 500 species. The numbers of ESTs for fish species are summarized in Table 1; most of species listed represent model organisms or economically important fishes.

SSRs can be searched for in these EST sequence databases. However, the major drawback for effective and rapid development of type I SSRs is access to sufficient sequence information. As shown in Table 1, for the channel catfish, around 45,000 EST sequences have been developed. This provides a serious source for extracting thousands of sequences containing putative SSRs with the possibility of developing several hundreds of polymorphic microsatellite markers. For example, sequence analysis of 1909 ESTs from a skin cDNA library of Ictalurus punctatus revealed the presence of 89 (4.7% of 1909) microsatellite-containing genes (Karsi et al., 2002). Screening of 1201 ESTs
A Taxonomic assignment was performed through the NCBI taxonomy database (http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html, Wheeler et al., 2004).

### Table 1

<table>
<thead>
<tr>
<th>Species*</th>
<th>Family*</th>
<th># of ESTsb</th>
<th># of SSRsc</th>
<th>Other sequencesc</th>
<th>Totald</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Danio rerio</em></td>
<td>Cyprinidae</td>
<td>689,581</td>
<td>6055</td>
<td>362,754</td>
<td>1,052,335</td>
</tr>
<tr>
<td><em>Oncorhynchus mykiss</em></td>
<td>Salmonidae</td>
<td>239,327</td>
<td>968</td>
<td>2972</td>
<td>243,267</td>
</tr>
<tr>
<td><em>Oryzias latipes</em> (Japanese medaka)</td>
<td>Adrianichthyidae</td>
<td>221,546</td>
<td>2</td>
<td>117,865</td>
<td>339,413</td>
</tr>
<tr>
<td><em>Gasterosteus aculeatus</em> (three-spined stickleback)</td>
<td>Gasterosteidae</td>
<td>170,994</td>
<td>140</td>
<td>10,223</td>
<td>181,357</td>
</tr>
<tr>
<td><em>Salmo salar</em> (Atlantic salmon)</td>
<td>Salmonidae</td>
<td>113,002</td>
<td>1696</td>
<td>1434</td>
<td>116,132</td>
</tr>
<tr>
<td><em>Fundulus heteroclitus</em> (mummichog)</td>
<td>Funditidiae</td>
<td>55,116</td>
<td>83</td>
<td>494</td>
<td>55,693</td>
</tr>
<tr>
<td><em>Ictalurus punctatus</em> (channel catfish)</td>
<td>Ictaluridae</td>
<td>44,476</td>
<td>3921</td>
<td>1614</td>
<td>50,011</td>
</tr>
<tr>
<td><em>Haplochromis chilotes</em> (cichlid fish)</td>
<td>Cichlidae</td>
<td>35,747</td>
<td>–</td>
<td>1</td>
<td>35,748</td>
</tr>
<tr>
<td><em>Takifugu rubripes</em> (Japanese pufferfish)</td>
<td>Tetraodontidae</td>
<td>25,850</td>
<td>430</td>
<td>66,805</td>
<td>93,085</td>
</tr>
<tr>
<td><em>Epiplatys burgeri</em> (inshore hagfish)</td>
<td>Myxiniidae</td>
<td>23,884</td>
<td>–</td>
<td>147</td>
<td>24,031</td>
</tr>
<tr>
<td><em>Haplochromis sp.</em> (red tail sheller)</td>
<td>Cichlidae</td>
<td>14,073</td>
<td>–</td>
<td>–</td>
<td>14,073</td>
</tr>
<tr>
<td><em>Leucoraja erinacea</em> (little skate)</td>
<td>Rajidiae</td>
<td>11,260</td>
<td>–</td>
<td>609</td>
<td>11,869</td>
</tr>
<tr>
<td><em>Squalus acanthias</em> (piked dogfish)</td>
<td>Scualidiae</td>
<td>10,878</td>
<td>8</td>
<td>134</td>
<td>11,020</td>
</tr>
<tr>
<td><em>Petromyzon marinus</em> (sea lamprey)</td>
<td>Petromyzontidae</td>
<td>10,617</td>
<td>11</td>
<td>550</td>
<td>11,178</td>
</tr>
<tr>
<td><em>Cypinus carpio</em> (common carp)</td>
<td>Cyprinidae</td>
<td>10,612</td>
<td>29</td>
<td>1614</td>
<td>12,255</td>
</tr>
<tr>
<td><em>Ictalurus furcatus</em> (blue catfish)</td>
<td>Ictaluridae</td>
<td>10,524</td>
<td>–</td>
<td>25</td>
<td>10,549</td>
</tr>
<tr>
<td><em>Platichthys flesus</em> (European flounder)</td>
<td>Pleuronectidae</td>
<td>5,573</td>
<td>29</td>
<td>225</td>
<td>616</td>
</tr>
<tr>
<td><em>Paralichthys olivaceus</em> (Japanese flounder)</td>
<td>Paralichthyidae</td>
<td>3874</td>
<td>87</td>
<td>368</td>
<td>4329</td>
</tr>
<tr>
<td><em>Astatoctilapia harti</em> (cichlid fish)</td>
<td>Cichlidae</td>
<td>3670</td>
<td>–</td>
<td>90</td>
<td>3760</td>
</tr>
<tr>
<td><em>Hippoglossus hippocoglossus</em> (Atlantic halibut)</td>
<td>Pleuronectidae</td>
<td>3146</td>
<td>29</td>
<td>180</td>
<td>3355</td>
</tr>
<tr>
<td><em>Oncorhynchus tschawytscha</em> (chinook salmon)</td>
<td>Salmonidae</td>
<td>2301</td>
<td>78</td>
<td>591</td>
<td>2970</td>
</tr>
<tr>
<td><em>Carassius auratus</em> (goldfish)</td>
<td>Cyprinidae</td>
<td>2080</td>
<td>21</td>
<td>789</td>
<td>2890</td>
</tr>
<tr>
<td><em>Dicentrarchus labrax</em> (European sea bass)</td>
<td>Moronidiae</td>
<td>1770</td>
<td>208</td>
<td>286</td>
<td>1264</td>
</tr>
<tr>
<td><em>Coregonus clupeaformis</em> (lake whitefish)</td>
<td>Salmonidae</td>
<td>1691</td>
<td>31</td>
<td>13</td>
<td>1735</td>
</tr>
<tr>
<td><em>Pseudopleuronectes americanus</em> (winter flounder)</td>
<td>Pleuronectidae</td>
<td>1663</td>
<td>–</td>
<td>185</td>
<td>1668</td>
</tr>
<tr>
<td><em>Osmerus mordax</em> (rainbow smelt)</td>
<td>Salmonidae</td>
<td>1587</td>
<td>–</td>
<td>15</td>
<td>1600</td>
</tr>
<tr>
<td><em>Sparus aurata</em> (gilthead seabream)</td>
<td>Sparidiae</td>
<td>1518</td>
<td>26</td>
<td>487</td>
<td>2007</td>
</tr>
<tr>
<td><em>Gadus morhua</em> (Atlantic cod)</td>
<td>Gadidiae</td>
<td>1422</td>
<td>12</td>
<td>1599</td>
<td>2954</td>
</tr>
<tr>
<td><em>Perca flavilatilis</em> (European perch)</td>
<td>Percidiae</td>
<td>1097</td>
<td>1</td>
<td>97</td>
<td>1195</td>
</tr>
<tr>
<td><em>Epinephelus coioides</em> (orange-spotted grouper)</td>
<td>Serranidiae</td>
<td>1007</td>
<td>1</td>
<td>167</td>
<td>1193</td>
</tr>
<tr>
<td><em>Xiphophorus maculatus</em> (Southern platyfish)</td>
<td>Poeciliidae</td>
<td>847</td>
<td>266</td>
<td>167</td>
<td>1280</td>
</tr>
<tr>
<td><em>Oncorhynchus nerka</em> (sockeye salmon)</td>
<td>Salmonidiae</td>
<td>664</td>
<td>43</td>
<td>186</td>
<td>893</td>
</tr>
<tr>
<td><em>Opsanus beta</em> (gulf toadfish)</td>
<td>Batrachoidei</td>
<td>619</td>
<td>–</td>
<td>21</td>
<td>640</td>
</tr>
<tr>
<td><em>Seriola quinqueraulata</em> (five-ray yellowtail)</td>
<td>Carangidiae</td>
<td>558</td>
<td>5</td>
<td>178</td>
<td>741</td>
</tr>
<tr>
<td><em>Ctenopharyngodon idella</em> (grass carp)</td>
<td>Cyprinidiae</td>
<td>531</td>
<td>1</td>
<td>186</td>
<td>718</td>
</tr>
<tr>
<td><em>Oreochromis niloticus</em> (Nile tilapia)</td>
<td>Cichlidiae</td>
<td>294</td>
<td>235</td>
<td>3,543</td>
<td>4,072</td>
</tr>
<tr>
<td><em>Anguilla japonica</em> (Japanese eel)</td>
<td>Anguillidiae</td>
<td>196</td>
<td>37</td>
<td>323</td>
<td>556</td>
</tr>
<tr>
<td><em>Gillichthys mirabilis</em> (long-jawed mudsucker)</td>
<td>Gobiidiae</td>
<td>109</td>
<td>–</td>
<td>1</td>
<td>110</td>
</tr>
<tr>
<td><em>Tetradon flavilatilis</em> (green pufferfish)</td>
<td>Tetraodontidae</td>
<td>99</td>
<td>–</td>
<td>89</td>
<td>188</td>
</tr>
<tr>
<td><em>Austrodrinus limnaeus</em> (annual killifish)</td>
<td>Ruvilidiae</td>
<td>91</td>
<td>–</td>
<td>16</td>
<td>107</td>
</tr>
<tr>
<td><em>Salvelinus alpinus</em> (Arctic char)</td>
<td>Salmonidiae</td>
<td>63</td>
<td>12</td>
<td>220</td>
<td>295</td>
</tr>
<tr>
<td><em>Gillichthysseta</em> (goby)</td>
<td>Gobiidiae</td>
<td>62</td>
<td>–</td>
<td>59</td>
<td>121</td>
</tr>
<tr>
<td><em>Torpedo marmorata</em> (spotted ray)</td>
<td>Torpedinidiae</td>
<td>41</td>
<td>–</td>
<td>123</td>
<td>164</td>
</tr>
<tr>
<td><em>Siniperca chuatsi</em> (Chinese perch)</td>
<td>Percichthyidae</td>
<td>32</td>
<td>–</td>
<td>109</td>
<td>141</td>
</tr>
<tr>
<td><em>Chilomycterus plagiosom</em> (white-spotted bambooshark)</td>
<td>Hemiscylliidae</td>
<td>17</td>
<td>–</td>
<td>9</td>
<td>26</td>
</tr>
<tr>
<td><em>Poecilia reticulata</em> (guppy)</td>
<td>Poeciliidae</td>
<td>15</td>
<td>137</td>
<td>339</td>
<td>491</td>
</tr>
<tr>
<td><em>Cyclopterus lumpus</em> (lumpsucker)</td>
<td>Cyclopteridiae</td>
<td>12</td>
<td>–</td>
<td>17</td>
<td>29</td>
</tr>
<tr>
<td><em>Paramisgurnus dabyunianus</em> (Chinese loach)</td>
<td>Cobitidiae</td>
<td>11</td>
<td>1</td>
<td>24</td>
<td>36</td>
</tr>
<tr>
<td><em>Xiphophorus maculatus x Xiphophorus helleri</em> (Southern platyfish x green swordtail)</td>
<td>Poeciliidae</td>
<td>6</td>
<td>–</td>
<td>–</td>
<td>6</td>
</tr>
<tr>
<td><em>Pterophthalmus modestus</em> (shuttles hoppfish)</td>
<td>Gobiidiae</td>
<td>4</td>
<td>–</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td><em>Epiplatys cirratus</em> (New Zealand hagfish)</td>
<td>Myxiniidae</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td><em>Labeo rohita</em> (rohu)</td>
<td>Cyprinidiae</td>
<td>2</td>
<td>21</td>
<td>55</td>
<td>88</td>
</tr>
<tr>
<td><em>Catla catla</em> (Indian major carp)</td>
<td>Cyprinidiae</td>
<td>1</td>
<td>19</td>
<td>14</td>
<td>34</td>
</tr>
<tr>
<td><em>Geotria australis</em> (pouched lamprey)</td>
<td>Petromyzontidiae</td>
<td>1</td>
<td>–</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td><em>Lethenteron japonicum</em> (Japanese lamprey)</td>
<td>Petromyzontidiae</td>
<td>1</td>
<td>–</td>
<td>87</td>
<td>88</td>
</tr>
<tr>
<td><em>Oncorhynchus keta</em> (chum salmon)</td>
<td>Salmonidae</td>
<td>1</td>
<td>16</td>
<td>335</td>
<td>351</td>
</tr>
</tbody>
</table>
from a channel catfish brain cDNA library yielded 88 (7.3%) clones with microsatellites (Liu et al., 2001). Most brain microsatellites represented dinucleotides located in 3′-UTRs. However, 12 (14%) SSRs were tetrameric, whereas 20 (23%) clones contained trinucleotide repeats. A recent bioinformatic analysis of 43,033 ESTs from channel catfish revealed 4855 ESTs (11.3%) containing microsatellites (Serapion et al., 2004). 1312 of these ESTs fell into 569 contigs (e.g., locations within a chromosome map, where contiguous DNA segments overlap) and the remaining 3534 ESTs were singletons (e.g., unmapped to contigs, with unknown position in a chromosome map). A total of 4103 unique microsatellite-containing genes were identified. The dinucleotide CA/TG and GA/TC pairs were the most abundant among EST-derived microsatellites (Serapion et al., 2004).

A typical strategy for the development of EST-derived microsatellite markers (data mining) includes preliminary analysis of EST sequences from the DNA database to remove poly(A) and poly(T) stretches. These mononucleotide repeats are very common in ESTs developed from the 3′-ends of cDNA clones and correspond to the poly(A)-tails in eukaryotic mRNA. Sequences are further screened for putative SSRs. This supports identification of all SSR-containing EST sequences. Following the identification of microsatellite-containing ESTs, flanking primers should be designed to amplify a microsatellite. In order to hypothesize about putative functions of SSR-containing genes, these sequences are needed for comparison to the database of amino acid sequences (Kantety et al., 2002; Thiel et al., 2003). UniProt/Swiss-Prot is an annotated protein sequence database (http://www.ebi.ac.uk/swissprot/) which is extremely helpful for these purposes.

6. Applications of microsatellites

SSRs are often highly polymorphic due to variation in the number of repeats (Amos and Pemberton, 1992). They can be simply and rapidly detected by the polymerase chain reaction (PCR) using two unique oligonucleotide primers that flank the microsatellite and hence define the microsatellite locus. Because of their multiallelic nature, codominant inheritance, small length, extensive genome coverage and relative abundance, microsatellites have been successfully applied in a wide variety of research fields and practical disciplines (Powel et al., 1996).

6.1. Genetic mapping

Genetic mapping represents one of the major research fields in which microsatellite markers have been applied. SSRs remain the markers of choice for the construction of linkage maps, because they are highly polymorphic (and highly informative) and require a small amount of DNA for each test (Fig. 1). Methods for microsatellite detection can be readily automated. A disadvantage of microsatellites is that they are mostly anonymous DNA fragments (Cullis, 2002). However, type II (noncoding) microsatellites are very helpful for building a dense linkage map framework into which type I (coding) markers can then be incorporated.

Compared to type II markers, mapping type I markers directly shows the location of genes within the linkage map. The coding markers often represent genetic variations associated with interesting or economically significant phenotypes. Therefore, enrichment of the linkage map by type I loci greatly benefits the mapping and characterization of genes responsible for medically, agriculturally and evolutionarily important complex traits. This also provides a good opportunity for marker-assisted selection (MAS) in commercially significant species (Poompuang and Hallerman, 1997; Waldbieser and Wolters, 1999). For economically important fishes, EST-derived microsatellites have been isolated and used for genetic mapping in Atlantic salmon (Koop and Davidson, 2005), channel catfish (Liu et al., 2001; Waldbieser et al., 2001; Karsi et al., 2002), European sea bass (Chistiakov, pers. data), Nile tilapia (Cnaani et al., 2002), rainbow trout (Sakamoto et al., 2000) and zebrafish (Knapik et al., 1998).

Linkage maps are known as recombination maps and define the order and distance of loci along a chromosome on the basis of inheritance in families or mapping populations. During meiosis, one random copy of each chromosome pair is passed on to the gamete. Therefore, grandparental copies of genes located on different chromosomes are inherited independently, whereas genes on...

Notes to Table 1:

a-Taxonomic assignment was performed through the NCBI taxonomy database (http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html, Wheeler et al., 2004).
b-DabData are taken from the dbEST database (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html).
the same chromosome are not inherited independently, but are passed on together or “linked”. Only genes located next to each other are tightly linked. Crossing-over results from physical exchange of chromosome segments between two homologous chromosomes of meiosis. Recombination results in the exchange of grandparental alleles of genes further apart on that chromosome (Hartl and Jones, 2001).

During linkage map construction, co-segregating markers are placed into linkage groups, and the proportion of recombinants detected between linked markers is used as a measure of distance between them. Genetic distance is usually measured in centimorgans (cM), where 1 cM is equivalent to 1% recombination between markers. Gene mapping algorithms analyse the co-segregation of markers in the families and assemble the markers into linkage groups, followed by selecting the most likely order of markers within the same linkage group (Hartl and Jones, 2001).

Examples of microsatellite-based linkage maps for some important teleost fish species are listed in Table 2. Consolidated linkage maps have been published for fishes, such as Arctic charr Salvelinus alpinus (Woram et al., 2004), Atlantic salmon (Moen et al., 2004a, b), rainbow trout (Sakamoto et al., 2000; Nichols et al., 2003a, b), Xiphophorus sp. (Walter et al., 2004), zebrafish (Woods et al., 2000), Japanese flounder (Coimbra et al., 2003) and Nile tilapia (Kocher et al., 1998; Agresti et al., 2000) (Table 2). These maps comprise different types of markers, with a major contribution of SSRs and amplified fragment length polymorphisms (AFLPs) (Vos et al., 1995) (Table 2). SSRs and AFLPs represent DNA markers, which are extremely useful for constructing a primary framework map that could be further enriched with type I markers (ESTs, SNPs, genes). Zebrafish and rainbow trout linkage maps can serve as examples of such “map evolution”. For example, Woods et al. (2000) used 616 of 2000 microsatellites mapped by Shimoda et al. (1999) to build a linkage map additionally enriched with 1503 coding markers.

The number of linkage groups in a linkage map does not always match the haploid number of chromosomes (Table 2). It may be due to the structure of the population used for construction of the map and the low genome coverage of the markers mapped.

For some aquaculture species, such as rainbow trout (Sakamoto et al., 2000), zebrafish (Singer et al., 2002), Japanese flounder (Coimbra et al., 2003), tilapia (Agresti et al., 2000), Arctic charr (Woram et al., 2004) and European sea bass (Chistiakov et al., 2005), sex-specific maps have been developed (Table 2). Linkage map length differs between sexes. In species with the XY sex determination system, the female map is usually longer than the male map because of higher recombination rates in females compared to males. In zebrafish and rainbow trout, the male recombination rate close to the centromere is greatly reduced compared to the female (Sakamoto et al., 2000; Singer et al., 2002). The molecular basis of suppression in recombination remains unclear. Lindahl (1991) proposed some general explanations, while Sakamoto et al. (2000)
focused their suggestions for salmonids on the model of tetravalent formation, which hinders crossovers between homologous chromosomes via structural constraints in distal regions in males.

Fishes have some of the most complex sex determination systems known in the animal kingdom (Schartl, 2004). Identification of sex-determining loci is hindered in fish species due to the absence of heteromorphic sex chromosomes (Traut and Winking, 2001), variability of genetic sex determination (Völff and Schartl, 2001) and ability to switch sex depending on the environmental conditions (Baroiller and D’Cotta, 2001). However, applying microsatellites provides a good opportunity to find a sex-determining locus due to

Table 2
Microsatellite-based-linkage maps of some economically important marine species

<table>
<thead>
<tr>
<th>Species</th>
<th>Linkage map</th>
<th>Number of markers mapped</th>
<th>Number of linkage groups</th>
<th>Length, cM</th>
<th>Diploid chromosome number (2n)*</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arctic char</td>
<td>(Salvelinus alpinus)</td>
<td>327 (184 SSRs+129 AFLPs+13 ESTs+phenotypic marker Sex)</td>
<td>46</td>
<td>3900 (male)</td>
<td>80</td>
<td>Woram et al., 2004</td>
</tr>
<tr>
<td>Atlantic salmon</td>
<td>(Salmo salar)</td>
<td>Male—251 (31 SSRs+215 AFLPs)</td>
<td>31</td>
<td>103</td>
<td>60</td>
<td>Moen et al., 2004a</td>
</tr>
<tr>
<td>Channel catfish</td>
<td>(Ictalurus punctatus)</td>
<td>263 (19 type 1 SSRs+243 type II SSRs+1 EST)</td>
<td>32</td>
<td>1958</td>
<td>56</td>
<td>Waldbieser et al., 2001</td>
</tr>
<tr>
<td>Yellowtail (Seriola sp.)</td>
<td>Male (S. lalandi): 175 SSRs</td>
<td>21</td>
<td>548.3</td>
<td>48</td>
<td>Ohara et al., 2005</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female (S. quinqueradiata): 122 SSRs</td>
<td>25</td>
<td>473.3</td>
<td>48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Three-spined stickleback</td>
<td>(Gasterosteus aculeatus)</td>
<td>227 SSRs</td>
<td>26</td>
<td>886</td>
<td>42</td>
<td>Peichel et al., 2001</td>
</tr>
<tr>
<td>Nile tilapia</td>
<td>(Oreochromis niloticus)</td>
<td>162 (59 STSs+103 AFLPs)</td>
<td>30</td>
<td>704</td>
<td>44</td>
<td>Kocher et al., 1998</td>
</tr>
<tr>
<td></td>
<td>Male (O. aureus × O. niloticus F1): 214 (60 SSRs+154 AFLPs)</td>
<td>24</td>
<td>1632</td>
<td>44</td>
<td>Agresti et al., 2000</td>
<td></td>
</tr>
<tr>
<td>Hybrid O. aureus × O. niloticus</td>
<td>162 SSRs+21 genes</td>
<td>24</td>
<td>1311</td>
<td>44</td>
<td>Lee et al., 2005</td>
<td></td>
</tr>
<tr>
<td>Rainbow trout</td>
<td>(Oncorhynchus mykiss)</td>
<td>208 (191 SSRs+3 RAPD+7 genes+7 allozymes)</td>
<td>29</td>
<td>463.2 (male)</td>
<td>60</td>
<td>Sakamoto et al., 2000</td>
</tr>
<tr>
<td></td>
<td>Male: 1359 (973 AFLPs+226 SSRs+72 VNTRs+38 SINE markers+29 genes+12 minisatellites+5 RAPD+4 allozymes)</td>
<td>40 (30 major)</td>
<td>4590</td>
<td>44</td>
<td>Nichols et al., 2003a</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female (O. mossambicus): 14 (13 SSRs+42 AFLPs)</td>
<td>16</td>
<td>514</td>
<td>44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>European sea bass</td>
<td>(Dicentrarchus labrax)</td>
<td>162 SSRs</td>
<td>25</td>
<td>815</td>
<td>48</td>
<td>Chistiakov et al., 2005</td>
</tr>
<tr>
<td></td>
<td>Male: 100 SSRs</td>
<td>26</td>
<td>609.1</td>
<td>48</td>
<td>Inami et al., 2005</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female: 89 SSRs</td>
<td>28 (female)</td>
<td>708.5 (female)</td>
<td>48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Platyfish (Xiphophorus sp.)</td>
<td>290 (256 SSRs+22 allozymes+11 genes)</td>
<td>24</td>
<td>2178</td>
<td>48</td>
<td>Walter et al., 2004</td>
<td></td>
</tr>
<tr>
<td>Zebrafish (Danio rerio)</td>
<td>102 SSRs</td>
<td>25</td>
<td>1320</td>
<td>50</td>
<td>Knapik et al., 1996</td>
<td></td>
</tr>
<tr>
<td></td>
<td>705 SSRs</td>
<td>25</td>
<td>2350</td>
<td>50</td>
<td>Knapik et al., 1998</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2119 (616 SSRs+1503 genes and ESTs)</td>
<td>25</td>
<td>2400</td>
<td>50</td>
<td>Woods et al., 2000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2000 SSRs</td>
<td>25</td>
<td>2295</td>
<td>50</td>
<td>Shioda et al., 1999</td>
<td></td>
</tr>
<tr>
<td></td>
<td>141 SSRs</td>
<td>25</td>
<td>2582.7 (female)</td>
<td>50</td>
<td>Singer et al., 2002</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Male—231 (82 SSRs+149 AFLPs)</td>
<td>25</td>
<td>942.5 (male)</td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japanese flounder</td>
<td>(Paralichthys olivaceus)</td>
<td>304 (101 SSRs+203 AFLPs)</td>
<td>27</td>
<td>1635.6 (sex-averaged)</td>
<td>48</td>
<td>Shimoda et al., 2003</td>
</tr>
<tr>
<td></td>
<td>Female—304 (101 SSRs+203 AFLPs)</td>
<td>27</td>
<td>741.1</td>
<td>48</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Data are taken from the Animal Genome Size Database (http://www.genomesize.com/fish.htm). Abbreviations: AFLP, amplified fragment length polymorphism; VNTR, variable number of tandem repeats; STS, sequence tagged site; RAPD, randomly amplified polymorphic DNA; IRS, internal repetitive sequence.
specific features in the heterogametic sex such as an obvious reduction in recombination between markers linked to the sex-determination region in male compared to female meioses (Naruse et al., 2000) and the consistently heterozygous status of males for unique alleles in this region (Peichel et al., 2004). Using this approach, a sex-determining locus has been found in medaka Oryzias latipes (Naruse et al., 2000) and three-spined stickleback (Peichel et al., 2004), species without distinct sex chromosomes.

6.2. Individual DNA identification and parentage assignment

Microsatellites represent codominant single-locus DNA markers. For each SSR, a progeny inherits one allele from the male parent and another allele from the female parent. This simple inheritance pattern can explain the extreme popularity of polymorphic SSR loci in paternity testing. Using a panel of several microsatellite loci, a unique combined SSR genotype profile can be produced for each individual tested. The genotype profile is highly discriminating, which suggests that a random individual would have a low probability of matching a given genotype.

Microsatellites are extensively exploited for paternity and relatedness analysis of natural populations, hatchery broodstocks and trade control of fish products, including those from aquaculture (Liu and Cordes, 2004). Appropriate mathematical tools are available to evaluate genetic relatedness and inheritance in these systems (Luikart and Englund, 1999; Blouin, 2003; Jones and Ardren, 2003). A suitable methodology should be chosen for accurate and correct analysis of genotyping data to reconstruct parentage and pedigree structure in wild populations and broodstocks. Existing analytical packages combine several different approaches in parentage analysis such as merging likelihood techniques with tools assessing statistical confidence in parental assignments that provides a significant power in parentage reconstruction (Jones and Ardren, 2003).

An example of successful application of microsatellite markers in relatedness testing was described by Herbinger et al. (1995), who analyzed a rainbow trout broodstock in a small hatchery in Canada. Using only four of five microsatellite markers, they were able to match 91% of offspring to one or two parental couples of 100 possible parental pairs (ten sires × ten dams) and, in addition, to estimate parental effects on progeny growth and survival. Applications of SSRs have been reported to determine paternity and reproductive contribution in wild and farmed populations of various economically significant species such as bluegill sunfish Lepomis macrochirus (Neff, 2001), red sea bream Pagrus major (Doyle et al., 2001), turbot Scophthalmus maximus (Castro et al., 2004), chinook salmon Oncorhynchus tschawytscha and rainbow trout (Bentzen et al., 2001), Atlantic salmon (Norris et al., 2000; King et al., 2001), European sea bass and gilthead sea bream (Kotoulas and Tzigenopoulos, pers. comm.).

Due to the small size of SSRs, they are relatively stable in degraded DNA (Schneider et al., 2004). This is one reason why polymorphic SSRs are widely used in forensic science for individual DNA identification. In addition, they show a high degree of allelic variability, and hence uniqueness. An interesting example of the application of microsatellites to resolve a case of fishing tournament fraud in Finland was reported by Primmer et al. (2000). Genotyping data provided a highly significant power for excluding the possibility \( P>0.9999 \) of a 5.5 kg Atlantic salmon originating from the fishing competition location, Lake Saimaa (south east Finland). Assignment of the suspect fish to neighbouring natural samples was done with three methods: using a Bayesian-based approach, based on reference population allele frequencies and based on genetic distance, all using the software GeneClass (Cornuet et al., 1999). When presented with this evidence, the offender confessed to purchasing the salmon at a local fish shop, and criminal charges were made. A similar strategy could be also used, for example, in cases of illegal poaching, fraud with food authenticity and the mix-up of commercial fish catches (Fig. 3), in order to assign or exclude individuals from originating from a claimed population (Poetsch et al., 2000).

In addition, microsatellite loci remain relatively stable in bone remnants and dental tissue, providing the basis for the successful application of ancient DNA for molecular analysis (Burger et al., 1999). Successful extraction and amplification of nuclear DNA from the \( \beta \)-globin gene region containing a polymorphic \((AT) \delta (T) \gamma \) microsatellite from 12,000 year-old human bone specimens has been reported (Beraud-Colomb et al., 1995). Application of microsatellites obtained from historical fish scale collections has helped to explain demographic declines in abundance, which resulted in the complete collapse of populations of lake trout in the upper Laurentian Great Lakes of North America during the past 40 years (Guinand et al., 2003). Other cases where DNA from old fish scales was used to characterise historic populations involve northern pike Esox lucius (Miller and Kapuscinski, 1996; Larsen et al., 2005) and brown trout Salmo trutta (Hansen, 2002). Also the membrane lining of fish otoliths contains DNA, which can be
used to genotype historic populations (Atlantic cod *Gadus morhua*: Hutchinson et al., 2003; New Zealand snapper *Pagrus auratus*: Hauser et al., 2002).

Analysis of nuclear microsatellites from ancient human and animal remnants represents an essential step to understand the genetic diversity in current populations and to provide substantial perspectives for the analysis of archeological issues, establishment of ancient baselines, heritable diseases, determination of relatedness and establishment of genealogies in prehistoric populations (Zierdt et al., 1996).

6.3. Phylogeny, population and conservation genetics

The molecular structure and genetic variability of microsatellites is extensively exploited in evolutionary studies of a wide variety of fish species. The vast majority of these studies attempt to infer phylogenetic relationships from microsatellite data at levels below the species level (Goldstein et al., 1999; Heath et al., 2001; Reusch et al., 2001) or for recently diverged species (McCartney et al., 2003; Stamford and Taylor, 2004), using variability within stretches of tandem repeats, which evolve significantly more rapidly than flanking regions. However, the high incidence of homoplasy (e.g., false equality of alleles based on independent mutation to the same size) with increasing evolutionary distance, may undermine the confidence of the inferred phylogenetic hypotheses, compromise the accuracy and limit the depth of phylogenetic inference (Jarne and Lagoda, 1996). Another obvious problem with using SSRs for phylogenetic inference is that primers developed from one taxon may not work well on all the taxa for which genotypes are required. Although cross-species amplification is common, limits on the utility of primers for amplifying homologous loci in divergent taxa are evident. Even when it is possible to amplify something in divergent taxa, the sequences may not be similar enough to permit confident assessment of orthology. Flanking regions of microsatellites have proven their value in establishing phylogenetic relationships between species and families, because they evolve much more slowly than numbers of tandem repeats. For example, a phylogeny of cichlid fishes was studied based on information from DNA sequences of the flanking region of a (CA)_n microsatellite locus *TmoM27*, which showed particular conservation in several lineages of cichlids diverged more than 80–100 million years ago (Zardoya et al., 1996). Analysis revealed that the repeat region was nearly lost in the ancestor to cichlids and then amplified extensively in African taxa (Steelman et al., 1998). Indian and Malagasy cichlids formed a basal, paraphyletic group, while African and Neotropical cichlids were both monophyletic and sister groups (Zardoya et al., 1996; Steelman et al., 1998). The authors suggested that marker *TmoM27* could be widely applied in phylogenetic studies in other perciform fishes.

Phylogeographical applications of microsatellites are eminently suitable, where population structure is observed over a large geographical scale (Koskinen et al., 2002; Gum et al., 2005) (Fig. 2). The latter study on grayling *Thymallus thymallus* shows that there is strong admixture among major lineages in contact zones between drainages zones. Microsatellites are even more revealing over shorter geographical distances, where a few cases of panmixia (Dannewitz et al., 2005) and numerous cases of isolation by distance patterns (Ruzzante et al., 1999; O’Reilly et al., 2004), clinal variation (Nielsen et al., 2004), fragmentation

Fig. 2. Microsatellite genotypes (scored at dinucleotide locus F14) of common soles (*Solea solea*) collected in the North Sea (*n* = 27) and Adriatic Sea (*n* = 29). Homozygotes are indicated with an arrow. A larger number of individuals from the North Sea population are heterozygous because of the higher allelic diversity. At the right are indicated the marker sizes of the standard (M) (in bp).
(Lemaire et al., 2005), hybridisation (Gum et al., 2005) and cryptic speciation (Fillatre et al., 2003) have been identified (see later). In those cases, differences in the microsatellite allelic composition of populations are converted into evolutionary distances. Microsatellite genotypes are particularly helpful to detect structure in closely related populations, regardless of whether they are in evolutionary equilibrium.

The levels of genetic diversity between fish populations revealed with microsatellite markers are much higher than those obtained with phenotypic or allozyme markers (Miller and Kapuscinski, 1996; Shaw et al., 1999; Triantafyllidis et al., 2002; Corujo et al., 2004). Compared with allozyme markers, microsatellites exhibit higher levels of polymorphism and abundance in genomic DNA (Schlötterer, 2000). Most microsatellites are considered neutral markers, whereas allozymes are sometimes the target for natural selection. In that case, no deviation from the expectations under a neutral model is expected by selection acting on the microsatellite itself. However, if a microsatellite locus is linked to a genomic region which is the target for natural selection, then this microsatellite will show deviation from neutral expectation (Schlötterer et al., 1997; Lemaire et al., 2000).

SSR loci are more sensitive than allozymes for the evaluation of the dynamics of populations, including demographic bottlenecks (Spencer et al., 2000; Guinand and Scribner, 2003; Ramstad et al., 2004), population size fluctuations and effective population sizes (Gold et al., 2001; Bérubé et al., 2002; Waples, 2002). Common measures of genetic diversity are heterozygosity (the proportion of heterozygous individuals in the population), allelic diversity (number of alleles at a locus in the population), and the proportion of polymorphic loci (Pujolar et al., 2005). Marked decreases in the observed heterozygosity and reduced number of observed alleles of tested SSRs might be attributed to the action of population genetic bottlenecks. Since they evolve $10^2$–$10^3$ times faster than single-copy nuclear DNA, they are a powerful tool for analyzing recent and contemporary events (Ellegren, 2000). For example, screening of microsatellites linked to the Y chromosome enabled observation of fine genetic structure of human populations as well as directions of migration and timing of post-glacial human expansion in Europe (Rootsi et al., 2004). In salmonids, SSRs have been successfully used for defining temporal intervals and explaining mechanisms of severe decline of populations of brown trout in Denmark (Hansen et al., 2002) and lake trout in the North American Great Lakes (Guinand and Scribner, 2003; Guinand et al., 2003). However, size homoplasy problems in PCR-based microsatellite assays may affect the inference of recent population history (Estoup et al., 2002). Homoplasy may contribute to biased genetic analyses of natural populations and, hence, limit the use of microsatellites for the identification of conservation units.

Genome-wide scans using microsatellite markers could be applied for a search of locus-specific signatures of positive directional selection in natural populations of any species for which a high-density genetic map is available (Storz, 2005). The expansion and fixation of adaptive mutations is associated with joint fixation of linked neutral variants, or genetic hitch-hiking. Genetic hitch-hiking results in a reduced level of polymorphism and a skewed distribution of allele frequencies at linked neutral markers (Kim and Stephan, 2000). The basic strategy of how to use whole-genome screens to detect loci under positive selection, was explained and referred to as hitch-hiking mapping (Harr et al., 2002). For example, a whole-genome screen of DNA polymorphisms was recently performed in humans and found evidence for selective sweeps, or loci which are driven by positive adaptive selection, in non-African populations (Storz et al., 2004).

In selected cases, it is possible to identify candidate genes, which are responsible for divergent selection in natural populations (Hendry et al., 2000; Santiago and Caballero, 2005). A multilocus scan of microsatellite variation in a southeast Asian population of the malaria parasite Plasmodium falciparum identified evidence for a selective sweep within a 100-kb region on chromosome 4 containing the dihydrofolate reductase (dhfr) gene; point mutations in this gene are known to be responsible for resistance to the antimalarial drug pyrimethamine (Nair et al., 2003).

For spatially separated populations that inhabit different environments or sympatric populations that use different ecological niches, it is possible to find chromosome regions conferring adaptive biodiversity by comparing relative levels of differentiation among multiple unlinked loci (Charlesworth et al., 1997). The occurrence of highly polymorphic microsatellites in the untranslated regions of ESTs is a potentially useful source of gene-associated polymorphisms representing genetic signatures of divergent selection. For example, analysis of 95 EST-microsatellites, isolated during a salmonid EST screening project (Rise et al., 2004), in Salmo salar populations inhabiting contrasting natural environments and geographically distinct regions, resulted in the selection of nine loci linked to candidate genes associated with adaptive divergence (Vasemägi et al., 2005).
Conservation and fisheries genetics focus on the effects of inbreeding, demography, contemporary genetic structuring and adaptation on the long-term survival of a species. Stock identification is a big issue (Ferguson et al., 1995), helping wildlife managers to protect biodiversity by identifying series of conservation units such as evolutionarily significant units (ESUs), management units (MUs) and action units (AUs) (Wan et al., 2004).

If populations within species show significant adaptive differentiation to different habitats (ecological niches) or significant genetic differentiation, they may justify management as separate evolutionary lineages termed ESUs (Moritz, 1994). The ESU concept was developed to assign units for protection below the taxonomical level. The identification of an ESU preferably depends on significantly differentiated genetic structure detected by presumably neutral markers. In that case, SSRs represent markers of choice for identifying ESUs. However, to find a true ESU, multiple, preferably different kinds of markers should be exploited, since size homoplasy and null (e.g., non-amplifiable) alleles could affect PCR-based microsatellite analysis (Brown et al., 2005). For example, mitochondrial and microsatellite DNA markers revealed four genetically differentiated lineages of European grayling (Thymallus thymallus) in central and northern Europe, which evolved in geographical isolation during the Pleistocene and could be recognized as the ESUs (Gum et al., 2005). The genotype data should be complemented with ecological and biological (e.g., life-history and behavioural) evidence (Crandall et al., 2000; Fraser and Bernatchez, 2001).

Genetic analyses often reveal differences between sampled populations with substantial but noncomplete phylogenetic separation, which have minor but statistically significant differences in allele frequency of nuclear or mitochondrial loci. These populations are termed MUs (Wan et al., 2004). The MU is considered to be a conservation unit level below that of the ESU, which is based on multiple evidence such as molecular markers, habitat use and adaptive characters. Using an MU is focused on the monitoring of the contemporary population structure and requires defining genetic structure of currently fragmented populations compared to the ESU, which is derived from historical genetic differentiation (Moritz, 1994). Using an MU designation does not imply detailed genetic structure of populations, but indicates the populations to be treated as a unit. AUs display genetic patterns of living populations. Microsatellites could be more successfully applied for identifying MUs and AUs than mitochondrial DNA markers, since mitochondrial DNA has maternal inheritance. Conservation strategies depend on neither paternal nor maternal variation, but focus on using biparental polymorphism of nuclear DNA to reflect characteristics needed to cope with environmental conditions (Zhang and Hewitt, 2003). A high mutation rate of SSR loci also supports use of these markers in the genetic analysis of very recent events in the dynamics of populations, e.g., to MUs and AUs.

6.4. Molecular epidemiology and pathology

Genomic instability of microsatellites has been extensively evaluated in the field of carcinogenesis, where chromosomal rearrangements (e.g., translocations, insertions and deletions of genomic regions) occur (Charames and Bapat, 2003). Carcinogenic events often happen within a genomic region harbouring a tumour suppressor gene and hence inactivate the gene (Grady, 2004). Carcinogenic rearrangements are associated with loss of heterozygosity (LOH) in microsatellites located within the affected chromosome region. Thus, detecting microsatellite LOH in tumour tissues contributes not only to molecular diagnosis of cancer, but also points the possible location of a tumour suppressor gene (Presneau et al., 2003).

Zebrafish is a relevant vertebrate system for modeling human cancer, displaying many similarities in tumorigenic pathways, despite the evolutionary divergence of tetrapods and fishes more than 300 million years ago. The zebrafish (whose cell-cycle genes, tumor suppressor genes and oncogenes are homologous to those found in humans and other mammals) has enormous potential as a vertebrate system in which to identify novel molecular pathways of oncogenesis, especially because they are prone to develop tumors (Amatruda et al., 2002; Rubinstein, 2003). Zebrafish also have played an important role as a vertebrate model system for carcinogenesis regarding environmental effects, genetic susceptibility, and environmental–genomic interactions (Amatruda et al., 2002).

The instability of triplet motifs was found in lower vertebrates, including fishes (Schartl et al., 1998; Liu et al., 2001). A variable number of trinucleotide repeats occurred within the coding region of functionally important genes expressed in the brain of adult fishes such as the channel catfish orthologue of the RAD23B gene (Liu et al., 2001) and the zebrafish Clock gene (Saleem et al., 2001). A RAD23B gene product functions in the nucleotide exclusion repair (NER) pathway. NER defects are associated with higher incidence of mutagenesis and carcinogenesis and cause Xeroderma...
pigmentosum, an autosomal recessive disease in humans (Sancar and Hearst, 1993). The *Clock* locus regulating circadian rhythms is highly conserved in various organisms. Alterations of circadian rhythms could be related to a large number of diseases, including psychiatric disorders in humans (King and Takahashi, 2000). The polyglutamine \((\text{CAG})_n\) tract at the *Clock* gene is highly polymorphic in *Drosophila*, ranging from 25 to 33 pure glutamine repeats. In zebrafish, the \((\text{CAG})_n\) stretch includes up to 51 repeat units (Saleem et al., 2001). However, it is shortest and non-polymorphic in human. The lack of polymorphism may indicate that variation at this locus is deleterious to the individual and hence not tolerated. The above examples suggest that expansion of trimeric repeats within exons is tightly regulated during molecular evolution and that regulatory mechanisms preserve coding sequences from the uncontrolled extension of triplet motifs.

Since microsatellite markers are usually selectively neutral and often represent non-functional sequences, they cannot be defined as loci directly responsible for disease phenotype. However, SSR markers, showing linkage and association with disease, can be in strong linkage disequilibrium with other functional genetic variations which truly cause the pathological phenotype. These disease-associated markers typically are represented by single nucleotide polymorphisms, or SNPs. They are often functionally relevant and, therefore, could be responsible for determination of the pathogenic phenotypes (Schork et al., 2000).

Genome-wide scans using SSRs often reveal linkage to susceptibility loci on chromosomes that harbour genes contributing to genetic predisposition/resistance to pathology. Further fine mapping with microsatellite markers located within the region of linkage often results in identifying the susceptibility gene, but not in the detection of the genetic variation, which directly causes clinical outcomes of the disease. In some cases, only when a microsatellite marker has direct functional significance, can it represent a genetic marker for disease (Fornoni et al., 2002; Wang et al., 2002).

Microsatellite-based screening strategies can be used in the fields of veterinary and medical parasitology and for molecular studies of infectious diseases. This includes mapping and further identification of genes responsible for resistance to parasites and pathogens and the identification of genes controlling drug resistance in pathogenic organisms (Gasser, 1999; Naidoo and Chetty, 1998; Behnke et al., 2003; Anderson, 2004). Such approaches have been applied in a variety of domesticated and farmed animals and plants (Naidoo and Chetty, 1998; Yencho et al., 2000). In farmed fish, they have been mostly performed on *Salmonidae*. In rainbow trout, genomic DNA was screened for loci controlling natural killer cell-like activity (Zimmerman et al., 2004) and linked to resistance to infectious hematopoietic necrosis virus (Palti et al., 1999; Khoo et al., 2004), ceratomyxosis (Nichols et al., 2003b) and pancreatic necrosis virus (Ozaki et al., 2001; Gibson, 2002). In Atlantic salmon, polymorphic loci associated with resistance to infectious salmon anemia have been identified (Grimholt et al., 2003; Moen et al., 2004b). For an interspecific tilapia hybrid (*Oreochromis mossambicus × O. aureus*), a genome scan using microsatellite markers found five loci in different linkage groups linked to innate immunity (Cnaani et al., 2004). Identification of six microsatellite loci linked to resistance to red sea bream iridovirus, which causes high mortalities in cultured red sea bream in Japan, have also been reported (Inami et al., 2005).

Transmission of diseases is a major problem in aquaculture production, and determining the genetic architecture of disease resistance traits is of great interest to geneticists working on aquaculture species. Due to the rapid accumulation of genetic and genomic data and availability of high-density microsatellite maps for certain farmed fishes, the number of studies searching for marker loci and genes associated with resistance to pathogens is expected to increase dramatically.

### 6.5. Quantitative trait loci mapping

A quantitative trait is one that has measurable phenotypic variation owing to genetic and/or environmental influences. The variation can be measured numerically (for example, height, size or blood pressure) and quantified. Generally, quantitative traits are complex (multifactorial) and influenced by several polymorphic genes and by environmental conditions. A QTL is a genetic locus (gene), the alleles of which affect phenotypic variation. One or many QTLs can contribute to a trait or a phenotype. When more than one QTL influences a particular trait, each might have a different effect size, and the effects of individual QTLs can vary from strong to weak. The size and nature of these effects also can be affected by the genetic background (the total genotype of the individual), and interactions between QTLs are common (Mackay, 2001).

To date, microsatellite-based strategies (scans across individual chromosomes and a whole genome) represent appropriate techniques to identify QTLs, particularly those that are associated with medically, economically and evolutionarily important complex traits. Due to the genome-wide distribution and high
levels of allelic polymorphism, SSR loci are very helpful in coarse and fine linkage mapping approaches. Coarse mapping resolves detection of a putative QTL in a chromosomal region, usually within a range of 10–30 cM. For a given QTL, the likelihood of success and mapping resolution depends on the number of loci screened and the magnitude of their effect on the trait of interest. Also important are recombination events in the mapping population, the mode of expression of the trait (dominant, recessive or additive), size of the mapping population, and number of genes that define the quantitative trait (Glazier et al., 2002).

Characteristics and types of experimental crosses used to breed a mapping population are also crucial in precise QTL mapping. Typically, approaches that increase the number of breakpoints in a mapping population provide higher mapping resolution, but also require a larger number of animals to achieve significance for a given size of a QTL effect. Lander and Kruglyak (1995) defined threshold values to assign significance of linkage of a marker to a given trait in a genome-wide screen: ‘highly significant’ refers to $P < 4.9 \times 10^{-6}$, ‘significant’ refers to $1.7 \times 10^{-3} \times P \times 4.9 \times 10^{-6}$ and ‘suggestive’ refers to $P \times 1.7 \times 10^{-3}$ after correction for multiple testing. Permutation tests represent a proven technique to calculate threshold values adjusted for multiple testing (Doerge and Churchill, 1996).

All QTLs detected, including those only with suggestive significance, should be designated as proposed by Lander and Kruglyak (1995); it facilitates further confirmation of such loci. It is wise to reconfirm any significant linkages before proceeding to finer mapping, to avoid unexpected effects. The reconfirmation can be done with independent crosses but requires a simple test of the proposed chromosome interval for linkage to the quantitative trait. Another way includes analysis of a congenic strain produced by repeated backcrosses to an inbred strain with selection for a particular marker or chromosomal region, which shows a maximum strength of linkage. This approach was widely applied in QTL mapping in rat and mouse, providing rapid confirmation of loci with modest genetic effect using a limited number of animals (Nadeau et al., 2000).

Fine mapping narrows the putative location of a QTL within the chromosome to the interval of less than 1–5 cM but requires more recombination events to separate the genes that drive expression of the quantitative trait from closely linked markers. If there are several linked QTLs, it is necessary to perform more sensitive phenotyping procedures. Creation of subcongenic strains, subdividing the crucial interval of linkage into several segments, greatly facilitates fine mapping and further discovery of a suitable candidate gene (Nadeau et al., 2000). However, it is difficult to find a candidate gene even when the crucial region was 0.5 cM if no significant sequence information on that region is available.

Generation and analysis of knock-out and knock-in (for a given gene) animals (Hentschel and Bonventre, 2005), transgenesis with bacterial artificial chromosomes carrying the target gene (Yang and Seed, 2003), detection and evaluation of polymorphic markers within the candidate gene (Fanara et al., 2002), in vitro functional studies (Ponsuksili et al., 2005), examination of gene function (Salvi and Tuberosa, 2005) and other techniques can be used for further identification and confirmation of the gene responsible for the quantitative trait. However, identification of genes controlling complex traits often is complicated by the action of non-coding regulatory variants that are problematic for simple functional characterization. Until 2002, a total of 1700 genes responsible for a variety of genetic traits have been detected in various species, and only 31 of them (including 7 human genes) have been defined for complex traits (Peltonen and McKusick, 2001; Glazier et al., 2002). Five complex QTLs have been defined in agricultural plants and only one QTL in farmed animals (Glazier et al., 2002). The latter involves the DGAT1 gene encoding an enzyme which catalyzes the final step in tryglyceride synthesis. Three polymorphisms within the gene are found to cosegregate with milk production, and one of them represents lysine-to-alanine amino acid substitution in codon 232, which underlies variation in milk yield and composition (Coppieri et al., 1998).

For farmed fishes, the first mapping of an economically important QTL was reported in 1998 (Jackson et al., 1998). To date, no QTL gene has been defined, but several microsatellite-based QTL screenings have been performed. Most of these mapping experiments have targeted three salmonid species (Atlantic salmon, rainbow trout and Arctic char). These screenings include searches for QTLs related to temperature tolerance (Jackson et al., 1998; Danzmann et al., 1999; Perry et al., 2001; Cnaani et al., 2003; Somorjai et al., 2003), body weight (Cnaani et al., 2003; O’Malley et al., 2003; Borrell et al., 2004; Reid et al., 2005), body length (Borrell et al., 2004), spawning date (Sakamoto et al., 1999; O’Malley et al., 2003), embryonic development rate (Nichols et al., 2000; Robison et al., 2001) and condition factor (Nakamura et al., 2001; Perry et al., 2003; Reid et al., 2005). In rainbow trout, the sex-linked microsatellite marker OmyFGT19TU showed significant association with fork length (FL) and upper thermal tolerance (UTT), explaining dependence of male advantages in FL and UTT compared to their female
sibs with the origin of the Y chromosome (Perry et al., 2005). A body weight QTL search has been reported for all three salmonids mentioned above (Reid et al., 2005). Several markers linked to feed conversion efficiency have been found in channel catfish (Karsi et al., 2000).

An interspecific hybrid tilapia stock (four-way cross between Oreochromis mossambicus, O. aureus, O. mossambicus and Sarotherodon galilaeus) has been successfully developed for linkage mapping and QTL analysis (Agresti et al., 2000; Cnaani et al., 2003). A search for association with deleterious alleles and sex-ratio distortions revealed three microsatellite loci linked to sex distortion genes in tilapia, and one of the markers was likely to be related to a modifier of these genes (Shirak et al., 2002). A recent complex search for genetic markers is also crucial (Slate, 2005).

QTL mapping in natural populations represents a powerful tool to study the genetic architecture of fitness traits and reproductive isolation. This approach has not been extensively exploited yet since there are no well-developed genetic and genomic tools for most free-living species. The search for QTLs in wild populations requires the design or observation of appropriate crosses to create a suitable mapping population, which contains individuals of measured phenotype and which can be pedigreed. The availability of a genetic map of variable markers is also crucial (Slate, 2005).

Marker-assisted selection is based on the concept that it is possible to infer the presence of a gene from the presence of a marker tightly linked to the gene. For...
this purpose, it is important to have high-density and high-resolution genetic maps, which are saturated by markers in the vicinity of a target locus (gene) that will be selected. The degree of saturation is the proportion of the genome that will be covered by markers at the density such that the maximum separation between markers is no greater than a few centimorgans (usually 1–2 cM), within which linkage of markers and QTLs can be detected. Strategies to find markers tightly linked to the target gene are similar to those that are used for fine QTL mapping. Strategies, such as flanking marker analysis (Dixon et al., 1995) and pooled sample mapping (Churchill et al., 1993), are used to order these markers. When implementing flanking marker analysis, a large segregating population is screened with markers flanking the target interval (usually 5–10 cM) in order to identify individuals with a crossover within that interval. In pooled sample mapping, DNA from individuals from a large segregating population that share a given phenotype is pooled. DNA from each pool is analyzed with markers flanking the target gene. The order of markers then is determined from the proportion of pools that show a crossover with respect to the markers used for analysis.

Once a tight linkage is found between a molecular marker and a gene of interest, the inheritance of the gene can be traced in breeding programs. The availability of a phenotype, which can be clearly identified and quantified, plays an important role in successful MAS programmes. MAS should be performed with strong control of the correlation and matching ratio between the phenotype of interest and genotype information. Breeders select animals or plants carrying beneficial genotypes and alleles of markers that associate with or contribute to a trait of interest. Successful implementation of MAS requires well-developed genomic tools, including optional information on genetic variations relevant to the QTL phenotype, mode of inheritance, interactions with other contributing QTLs and economical magnitude of the QTL studied (Poompuang and Hallerman, 1997). To plan MAS, breeders also should take into account possible interactions between QTLs, which could relate to each other and have overlapping genetic backgrounds. In that case, MAS should preferably represent a complex selection index and take into consideration all economically significant traits that interact.

Production (such as growth rate, stress response and disease resistance) and reproductive (such as sex determination and development rate) traits are extremely important to breeders. Discovery of genes that control these features could greatly benefit MAS in breeding programs. Some marker-assisted breeding approaches have been implemented in agricultural plants, for which significant genetic and genomic knowledge is available (Koebner and Summers, 2002; Dale and Bradshaw, 2003). For example, wheat lines resistant to leaf rust, a world-wide disease caused by *Puccinia recondita f.sp. tritici*, have been selected using molecular markers linked to resistant alleles of genes *Lr9, Lr10, Lr19, Lr21, Lr24, Lr28, Lr29* and *Lr35* through construction and analysis of near-isogenic and single-gene lines. These lines differ in the presence or absence of the target gene and a small region flanking the target gene (Kolmer, 2005). SSR markers located in the vicinity of these genes were converted into sequence-tagged sites, or STS, representing short DNA sequences, which have a single occurrence in the genome and whose location and base sequence are known. The STS tightly linked to each of the leaf rust protective genes are clearly associated with the preferable phenotype and can be rapidly detected by PCR (Singh et al., 2004; Blaszczyk et al., 2004). The *Lr10* and *Lr21* genes have been cloned and sequenced (Feuillet et al., 2003; Huang et al., 2003). Both genes have sequences encoding nucleotide-binding site-leucine-rich regions, which are characteristic of disease resistance genes in plants. Leaf rust infections occur when the resistance gene interacts with a specific antivirulence gene in the rust, while compatible infections occur in the absence of an antivirulence gene. New races of rust develop by mutation and selection of virulence against rust resistance genes in wheat (Kolmer, 2005).

MAS was successfully implemented in farmed animals, such as cattle (Maillard et al., 2003), pig (Rothschild, 2003), sheep (Notter and Cockett, 2005) and chicken (Malek and Lamont, 2003), for which many QTLs have been completed. It can be combined with candidate gene analyses that have identified important chromosomal regions and individual genes associated with traits of economic interest (Rothschild, 2003).

To date, most of genetic improvements in aquaculture have been performed following the use of traditional breeding approaches. For example, the Norwegian selective breeding programme for Atlantic salmon has focused on growth rate, age at sexual maturity and later, disease resistance and carcass quality traits. The selection response for growth rate in salmon is generally much higher than that obtained for traditional livestock animals. Gjerde and Korsvoll (1999) reported the average genetic gain for growth rate of 14% per generation through six generations per selection. These authors also reported that four rounds of selection for delayed sexual maturation have reduced the incidence of salmon becoming sexually mature before they reach market size.
However, traits difficult to measure, such as disease resistance, low-heritability traits, sex-limited traits and traits expressed late in life would benefit from the use of gene technology (Poompuang and Hallerman, 1997).

In fish culture, DNA marker-based techniques have been applied in several cases, for example, in breeding programmes for Atlantic halibut *Hippoglossus hippoglossus* (Jackson et al., 2003), channel catfish (Waldieiser and Wolters, 1999), European sea bass (García de León et al., 1995), Japanese flounder (Hara and Sekino, 2003; Sekino et al., 2003) and salmonids (Herbinger et al., 1995; Fjalestad et al., 2003; Wilson et al., 2003). Apart from facilitating survival under variable environmental conditions, a higher genetic diversity within a progeny array stemming from multiple matings by females might also serve to reduce the potential cost of inbreeding and reduce the deleterious effects of genetic incompatibility between two partners (Jennions and Pettire, 2000).

There are not yet large-scale fish breeding programmes using MAS. However, the current status of genetic information and advanced genomic tools available for rainbow trout, Atlantic salmon and channel catfish provide an opportunity for the successful application of the DNA marker-based technology to selective breeding in these species. Predictions of the benefits of MAS in beef production are that genetic progress may increase by about 11% under specific circumstances (Gomez-Raya and Klemensdal, 1999). Estimates of benefits to aquaculture breeding are yet undefined.

Microsatellite markers are useful in early stages of MAS for the primary selection of parents for further crossing and subsequent genetic characterization of progeny. For this, SSRs linked to the target QTL would be used. Further improvements, such as enrichment of linkage maps with type I markers, construction of high-resolution linkage maps, development of physical maps and their integration with linkage maps, fine QTL mapping using the candidate gene approach, will lead to the replacement of SSRs by other types of genetic markers (ESTs and SNPs) in later stages of marker-assisted breeding programmes. This would lead to even more precise selection by gene-assisted selection (GAS), based on the use of favorable haplotypes and genotypes derived from genes directly contributing to the target trait (Hulata, 2001).

### 7. Prospects

In this review, with special emphasis on fishes, we considered microsatellites as structural genomic components and as genetic markers which have specific evolutionary mechanisms, functions and applications. Microsatellite loci have a high utility for constructing a genetic framework onto which other markers and genes are incorporated using various mapping strategies (linkage mapping, physical mapping and comparative genetic and genomic tools). Using microsatellite markers will greatly benefit the genetic dissection of complex and quantitative traits in order to map, identify and eventually clone and characterize the candidate genes controlling economically important traits. In aquaculture, SSRs represent the markers of choice for genetic monitoring of farmed stocks in view of breeding programs through the analysis of genetic variability and pedigree structure to design beneficial crosses, select genetically improved stocks, minimize inbreeding and increase selection response (Davis and DeNise, 1998; Knibb, 2000).

Large-scale DNA marker-based approaches derived from population and conservation genetics will greatly assist the management of wild fish stocks, such as trout and salmon. As advanced genomic resources become more and more available in model and economically important organisms, the possibility of applying them in related nonmodel species increases. For example, mapping studies in red deer have benefited from resources created for cattle (Slate et al., 2002). Genetic and genomic information from rainbow trout and

---

![Fig. 3. Results of STRUCTURE clustering and assignment analysis (Pritchard et al., 2000). Bar plot of individual assignment scores for eight populations of four morphologically defined eel species (represented by four colours). Each bar represents a single eel. Green: European eel *Anguilla anguilla*; yellow: American eel *A. rostrata*; blue: Japanese eel *A. japonica*; red: giant mottled eel *A. marmorata*. A colour different from the baseline may point to introgression (e.g., between *A. anguilla* and *A. rostrata*) (courtesy of G. Maes). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)](image-url)
Atlantic salmon can be used for comparative mapping in other salmonids and predictive genetic reconstruction of their common ancestor (Rexroad et al., 2005). In these studies, the main strategy lies in the identification of common SSR type I markers with subsequent mapping and annotation of the ESTs harbouring these SSRs to develop comparative anchor-tagged sequences (CATS). CATS primers anneal to conserved exonic regions, enabling the amplification and sequencing of adjacent introns. Since the primers are in conserved regions, CATS loci can be useful for identifying SNPs for comparative genomics projects (O’Brien et al., 1993).

Microsatellite-based techniques will be applied in genome scans and QTL mapping in natural populations to search for the genetic basis of adaptive selection and biodiversity in an increasing number of species (Rogers and Bernatchez, 2005). As our knowledge of the genome expands, we predict that more difficult and formal proofs of QTL identity will become unnecessary. New genetic and experimental tools will allow us to rapidly identify more and more genes, which drive quantitative variation, and then to evaluate their importance.

Acknowledgements

This work was supported by the EU-funded research projects BASSMAP (FP5; Q5RS-2001-01701) and AQUAFIRST (FP6; STREP contract no. 513692). We thank G. Maes for Fig. 3.

References


Maurer, F., Tierney, M., Medcalf, R., 1999. An AU-rich sequence in the 3-UTR of plasminogen activator, inhibitor-type (PAI-2) mRNA promotes PAI-3 mRNA decay and provides a binding site for nuclear HuR. Nucleic Acids Res. 27, 1664–1673.


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


