Microsatellite DNA markers, a fisheries perspective Part 1: The nature of microsatellites

Sekar, M., Suresh, E., Kumar, N.S., Nayak, S.K. and Balakrishna, C.

Central Institute of Fisheries Education, Versova, Andheri (W), Mumbai 400 061, India, e-mail: sekarrajaqua@rediffmail.com

Many new classes of Polymerase Chain Reaction (PCR) -based genetic markers have been developed over the last decade. Of these, microsatellite markers are widely regarded as one of the most useful identified so far[1,2,3]. The significance of microsatellite markers derives from their abundance in the genome, single locus nature, simplicity of assay, high levels of allelic diversity, mendelian inheritance, co-dominance and selective neutrality. Microsatellites are defined as tandem arrays of short DNA sequence motifs of 1-6 base pairs in length. This structure has lead to a number of alternative names such as short or simple tandem repeats (STRs), simple sequence repeats (SSRs), simple sequence length polymorphisms (SSLPs) and variable number tandem repeats (VNTRs)[4]. These are represented as small arrays of repeats with lengths ranging from somewhere around a dozen up to a few hundred base pairs, with many being less than 100 base pairs in total length. Microsatellites are abundantly distributed across the genome, demonstrate high levels of allele polymorphism and can easily be amplified with PCR. These features provide the underlying basis for their successful application in a wide range of fundamental and applied fields of biology and medicine. In the field of fisheries and aquaculture, microsatellites are useful for characterisation 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[5].

Genomic distribution

Microsatellites are ubiquitous in prokaryotes and eukaryotes, present even in the smallest bacterial genomes. The existence of microsatellites in eukaryotic genomes has been known since 1970[6]. In vertebrates these microsatellite loci can be found both in protein coding and non coding regions[7]. However, it has been demonstrated that the presence of microsatellite repeats is much higher in non coding DNA sequences[8]. Microsatellites can be described based on the number of nucleotides in the repeat motifs, with terms such as dinucleotide, trinucleotide and tetranucleotide.

Microsatellites can also be described by the term perfect, imperfect and compounds, which refer to an uninterrupted stretch of identical repeats, a repeat sequences in which there are one or more interruptions and repeat made up of two or more adjacent tandem repeats respectively. The majority of microsatellites (30-67%) found in the genome are dinucleotides. In the genome of vertebrates, the repeat $(AC)_n$ is most common, followed by $(AT)_n$. In total, higher order microsatellite classes such as tri-, tetra-, penta- and hexanucleotides are about 1.5 fold less common in the genomic DNA of vertebrates than dinucleotides[7]. Di and tetranucleotide motifs are mostly clustered in non coding regions. In vertebrates, they are distributed 42- and 30- fold less frequently in exons than in intronic sequences and intergenic regions, respectively. The dimeric microsatellite motifs present within expressed sequences are highly unstable, while in non-coding regions most dinucleotide repeats present as long stretches, probably due to the high tolerance of non-coding DNA to mutations[9].

Dinucleotide microsatellites are found in both exons and introns of a variety of fish species. Intronic dinucleotide microsatellites have been detected in the growth hormone gene of Nile tilapia, barramundi, Japanese flounder, and Japanese puffer fish[10]. Dinucleotide repeats have been described within genes of variety of fish species, including channel catfish[11], Atlantic salmon[12] and zebrafish[10]. Trinucleotide motifs are found in both coding and noncoding genomic regions with a high frequency[13]. In all vertebrates, (G+C) rich motifs are the most common among trinucleotides. These repeats are dominated in exons, whereas they are less common in intronic sequences[7].

Evolution and mutation models of microsatellites

Two models such as DNA polymerase slippage and unequal recombination over in meiosis have been suggested to explain microsatellite generation and evolution. Of these, strand-slippage replication appears to be the predominant mode at microsatellites[14]. Strand-slippage is speculated to occur primarily during lagging strand synthesis[4]. For example, it may involve the slippage of the newly synthesized DNA strand upon dissociation of a polymerase complex. This slippage creates a transient bulge which upon DNA repair would be either removed or lead to the elongation of the repeat[4]. Alternatively, the formation of a transient bulge in the template strand may lead to the shortening of the repeat. Slipped strands mispairing during DNA replication likely to represent the predominant mutational mechanism for microsatellites[15]. Non reciprocal recombination (gene conversion) also plays in genetic instability of some microsatellites including triplet modifs[16]. Gene conversion mechanism were found to be involved in the differentiation and evolution of paralogous sequences (duplicated loci within the species) in members of the family salmonidae, which was derived by tetraploidization.

There are three types of models of mutation that have been proposed to describe variations at microsatellite loci, which include stepwise mutation model (SMM), K" alleles model and the infinite alleles model (IAM). The SMM model holds that when microsatellites mutate, they only gain or lose one repeat and this implies that two alleles that differ by one repeat are more closely related than alleles that differ by many repeats. The genetic distance statistic that uses this model is called RST, it is generally the preferred model when calculating relatedness between individuals and population sub struc-

turing[17]. The K" Alleles Model holds that a microsatellite can mutate into any one of "K" alleles randomly. The IAM model predicts that mutation will lead only to new allelic states and may involve any number of repeat units, the statistic that uses this model is called FST[18]. More precise estimates of population size and structuring events would be possible assessing which model provides a better fit to microsatellite data. The yearly work on mutation models showed that the SMM more accurately predicted the type of variation commonly observed at microsatellite loci. Studies in human families apprised that most mutations led to new alleles which differed from the parental allele by only one or two repeat units. Thus, it is suggested that SMM provides the best fit to the microsatellite data available at this time. However, in fishes particularly from rainbow trout, it has been observed that the IMA gave a consistently better fit than the predictions of SMM to microsatellite variability data using uninterrupted microsatellite loci[19].

Microsatellite markers

The existence of variability in the microsatellite loci is a common phenomena, this variability consists of length polymorphisms caused by changes in the number of repeat units in the microsatellite array. The most likely change is the gain or loss of single repeat units, which suggests that microsatellite repeats are changing in the stepwise fashion. The observed mutation rates range from 10-3 to 10-6 event per locus per generation[8,20] which is several orders of magnitude greater than that of regular non repetitive DNA[21]. This higher rate of mutation at microsatellite loci causes hyper variable marker locus with many alleles. The allelic state at a particular microsatellite locus can be visualized by the analysis of length variation seen in PCR products. Because the regions flanking the microsatellites are generally conserved within a species, primers in these flanking regions can be used for PCR amplification and the products screened for size variation on polyacrylamide gels[22]. As microsatellite alleles are co-dominant they will be seen as one or two bands after PCR amplification representing homozygous or heterozygous states respectively. Microsatellites are being used as genetic markers in many living organisms, including agriculturally important species and in species of taxonomic interest or conservation concern.

Limitations

Microsatellites have proved to be versatile molecular markers, particularly for population analysis, but they are not without limitations. A major drawback to the use of microsatellite markers is that it is necessary to know the sequence of DNA flanking the various tandem repeats in order to design appropriate PCR primers. Unless the DNA sequence for the particular organism is readily available, it is necessary to create and screen a genomic DNA library in order to locate microsatellite loci[2,22].

Genotyping with microsatellite (especially those with dinucleotide repeats) is often complicated by the presence of stutter bands. The stuttering at dinucleotide microsatellite loci can often occlude adjacent alleles. The stutter bands are caused by polymerase slippage during PCR amplification, which results in secondary products containing one or more repeat units less than the primary allelic band. Stutter bands can sometimes equal the intensity of the primary band, making it difficult to accurately characterize genotypes, particularly in population studies where the accurate scoring is essential for parentage determination and family reconstruction from wild populations[11]. This problem is exacerbated in microsatellite loci that have large microsatellites, as the level of stuttering is generally higher in microsatellites with large repeated arrays. One possible approach to avoid this problem is to select tetranucleotide loci instead of dinucletide. A second method for reducing the potential scoring difficulties is to use dinucleotide loci with a reduced product size (<120)[23].

Functions

Although microsatellites are usually considered to be evolutionarily neutral markers, some microsatellites have been proven to have functional significance by critical test in various biological events in several organisms[21].

Chromatin organisation

Microsatellites play an important role in chromosome organisation as they are capable of forming a wide variety of unusual DNA structures with simple and complex loop folding patterns. For example, a hairpin formed by fragile X repeats (CCC)n and the bipartite triplex formed by (GAA)n / (TTC)n, have shown simple loop folding, and such triplex structures may have important regulatory effects on gene expression[24]. In many species, the centromeric and telomeric region of chromosomes is composed of numerous tandem repeats that affect the organisation of the centromere and telomere. Long microsatellites with mono-, di-, tri-, and tetranucleotide motifs are highly clustered in the centromeric regions of tomato, arabidopsis and sugar beet. In fishes, satellite sequences enriched by AT-dinucleotides have been found in the centromeric DNA of various gobiid species[25].

Regulation of DNA metabolic process

Numerous microsatellite and microsatellite DNA have been proposed as hot spots for recombination[26]. Dinucleotide motifs are preferential sites for recombination events due to their high affinity for recombination enzymes. Some microsatellite sequences, such as GT, CA, CT, GA and others, may influence recombination directly through their effects on DNA structure[28]. Microsatellites may affect DNA replication. In rat cells DNA amplification is arrested within a specific fragment which consists of a $d(GA)^{27}$ $d(TC)^{27}$ tract. It is found at the end of an amplicon, and in conjunction with the inverted repeat, may serve as an arrest site for DNA replication in-vivo[29].

Regulation of gene activity

Microsatellite occurrence in coding regions seems to be limited by non-perturbation of the reading frame. In many cases, microsatellite repeat numbers appears to be a key factor for gene expression and expression level. Some genes can only be expressed at a specific repeat number of microsatellites and some genes can be expressed within a narrow range of microsatellite repeat number, and out of this range gene activity would be turned off[21].

Part 2 of this article "Applications of microsatellite markers in fisheries" will follow in the July-September issue.

References

1. Bruford, M.V., Wayne, R. K., 1993. Microsatellite and their application to population genetics studies. Current Opinions in Genetics and Development 3: 939-943.
2. Queller, D. C., Strassmann, J. E., Hughes, C. R., 1993 Microsatellite and Kinship. Trends Ecol Evo. 8: 285-289.
3. Sunnucks, P., 2000. Efficient genetic markers for population biology. Trends Ecol Evol. 15: 199-203.
4. Schlötterer, C., Tautz, D., 1992. Slippage synthesis of simple sequence DNA. Nucleic Acids Res. 20: 211–215.
5. Chistiakov, D.A., Hellemans, B., Volckaert, A.M., 2006. Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. Aquaculture 255 (1-4): 1-29.
6. Bruford, M.V., Wayne, R. K., 1993. Microsatellite and their application to population genetics studies. Current Opinions in Genetics and Development 3: 939-943.
7. Toth, G., Gaspari, Z., Jurka, J., 2000. Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res. 10: 967–981.
8. Hancock, J.M., 1999. Microsatellite and other simple sequences in a minimal genome. J Mol. Evol. 41: 1038-1047.
9. Ellegren, H., 2000. Microsatellite mutations in the germline: implications for evolutionary inference. Trends Genet. 16: 551–558.
10. Venkatesh, B. and Brenner, S., 1997. Genomic structure and sequence of the Pufferfish (*Fugu rubripes*) growth hormone-encoding gene: a comparative analysis of teleost growth hormone genes, Gene. 187: 211–215.
11. Liu, Z.J., Cordes, J.F., 2004. DNA marker technologies and their applications in aquaculture genetics. Aquaculture 238: 1–37.
12. Grimholt, U., Drablos, F., Jorgensen, S.M., Hoyheim, B., Stet, R.J., 2002. The major histocompatibility class I locus in Atlantic salmon (*Salmo salar* L.): polymorphism, linkage analysis and protein modelling. Immunogenetics 54: 570–581.
13. Wren, J.D., Forgacs, E., Fondon, J.W., 2000. Repeat polymorphism within gene regions: phenotypic and evolutionary implications. Am. J. Hum. Genet. 67: 345–356.
14. Wolff, R.K., Plaetke, R., Jeffreys, A.J., White, R., 1989. Unequal crossing over between homologous chromosomes is not the major mechanism involved in the generation of new alleles at VNTR loci. Genomics 5: 382-384.
15. Richards, R.I., Sutherland, G.R.., 1994. Simple repeat DNA is not replicated simply. Nat. Genet. 6: 114–116.
16. Jakupciak, J.P., Wells, R.D., 2000. Genetic instabilities of triplet repeat sequences by recombination, UBMB Life 50: 355–359.
17. Moxon, Richard E. and Christopher Wills. 1999. DNA Microsatellites: Agents of Evolution? Scientific American, January: 94-99.
18. Jarne, P. and Pierre J.L. Lagoda. 1996. Microsatellites, form molecules to populations and back. Trends in Evolution and Ecology 11 (10): 424-429.
19. O'Connell, M., Skibinski, D.O.F., Beardmore, J.A., 1996. Allozyme amd mtDNA divergence between Atlantic salmon populations in North Wales. J. Fish. Biol. 48: 1023-1026.
20. Hancock, J.M., 1999. Microsatellite and other simple sequences: genomic context and mutational mechanism. In: Goldstein, D. B and Schlotterer, C., Microsatellites: Evolution and applications. Oxford University press, New York, pp1-9.
21. Li, Y.C., Koroi, A.B., Fahima, T., Beiles, A.A., Evitara. N. 2002. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Molecular Ecology 11: 2453-2465.
22. Strassmann, J. E., Solis, C. R., Peters, J.M., Queller, D. C., 1996. Strategies for finding and using high polymorphic DNA microsatellite loci for studies of genetic relatedness and pedigrees. In: Ferraris, J. D and Palumbi, S. R., Molecular zoology. Wiley-Liss, New York. pp 163-180.
23. O' Reilly, P., Wright, J.M., 1995. The evolving technology of DNA fingerprinting and its application to fisheries and aquaculture. J. Fish Bio. l47: 29-55.
24. Hulata, G., 2001. Genetic manipulations in aquaculture: a review of stock improvement by classical and modern technologies. Genetica 111: 155–173.
25. Canapa, A., Cerioni, P.N., Barucca, M., Olmo, E., Caputo, V. 2002. A centromeric satellite DNA may be involved in heterochromatin compactness in gobiid fishes, Chromosom. Res. 10: 297–304.
26. Jeffreys, A.J, Neil, D.L, Neumann, R. 1998. Repeat instability at human minisatellites arising from meiotic recombination, EMBO J. 17: 4147–4157.
27. Biet, E, Sun, J, Dutreix, M. 1999. Conserved sequence preference in DNA binding among recombination proteins; an affect of SSDNA secondary structure. Nucleic acid research 27: 596-600.
28. Kantety, R.V, La Rota, M., Matthews, D.E., and Sorrells, M.E. 2002. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol. Biol. 48:501-510.
29. Field, D, Wills, C., 1996. Long, polymorphic microsatellite in simple organisms. Proceeding of Royal Society of London (Biological Science) 263: 209-215.