
Contents lists available at Sjournals

Scientific Journal of
Biological Sciences

Journal homepage: www.Sjournals.com



Review article

Microsatellite DNA marker in aquatic organisms

A. Moradi*, S. Keyvanshokoh

Department of fishery, Marine Sciences and Technology of Khoramshahr University, Iran.

*Corresponding author; Department of fishery, Marine Sciences and Technology of Khoramshahr University, Iran.

ARTICLE INFO

Article history:

Received 01 September 2013

Accepted 14 September 2013

Available online 28 September 2013

Keywords:

Microsatellite

Aquatic

DNA marker

ABSTRACT

For the last 50 years, attempts have been made to discriminate among populations by using molecular markers. Although some techniques have proved successful in certain circumstances, the consistent trend to newer markers among fishery geneticists highlights the general lack of resolving power observed with older technologies. The last three decade has seen the increasing use of satellite DNA in investigations of genetic variability and divergence. The large interest in microsatellite loci is largely due to the very high levels of variability that have been observed and the ability to investigate this variation using PCR technology. The isolation and application of microsatellites to research fields as diverse as population genetics, parentage analyses and genome mapping are reviewed. Statistical considerations (e.g. appropriate sample sizes, number of loci and the mutation model assumptions on which the estimate is based) have not been considered in detail yet and the problems are often exacerbated in aquatic species, as some species show very large numbers of alleles at microsatellite loci.

© 2013 Sjournals. All rights reserved.

1. Introduction

In an effort to avoid some of the problems associated with minisatellites, research on VNTR loci has turned increasingly to microsatellites. Microsatellites or simple sequence repeats (SSRs) consist of short (1 ± 6 base pair, bp) tandem arrays (Tautz and Renz, 1984; Tautz, 1989). SSRs have generally been investigated by radioactively labelling one of two primers that are complementary to the flanking sequence on either side of the repeat unit array. These primers are used in a PCR reaction to amplify the repeat unit array. Length differences, due to the variable number of repeats among samples, are resolved by running out the amplification products on a sequencing gel. The gel is then dried, exposed to X-ray film and usually developed overnight. Reference to standards (e.g. M13 sequence and samples with known genotype) and/or an allelic ladder commonly determine allele sizes.

Descriptions of protocols used in the isolation and characterization of dinucleotide repeat loci are common (review: O'Reilly and Wright, 1995). Briefly DNA is digested; size fractionated on an agarose gel and a particular size range is excised from the gel (usually around 300 ± 600 bp). The recovered DNA is ligated to a vector, e.g. pUC18. This is used to transform cells and construct a partial genomic library which is screened with a repeat unit probe, e.g. GT (n). The clones that hybridize to the GT (n) probe are sequenced and primers are designed to complement unique sequences on either side of the array. Although the methodology is simple, there are a few key steps that can greatly increase the number of microsatellites isolated from any library. The following list illustrates technical considerations and problems associated with constructing microsatellite libraries.

1. Use good-quality (purified using an organic extraction protocol) DNA when constructing a library.
2. Construct a large library (.20 000 colonies) if possible. Note: Libraries can be made from $\ll 50 \mu\text{g}$ of DNA. However, digesting twice this amount can greatly reduce the time taken to develop primers for microsatellite loci (particularly if the genomic DNA has been digested with only a single enzyme, e.g. Mbol).
3. Check the genomic digest with the individual enzymes to be used for multiple copy elements, e.g. AluI elements, before proceeding with cloning and screening.
4. Keep PCR product size to a minimum: > 220 bp alleles are often fuzzy and faint. This is particularly important if dealing with old or degraded tissue, e.g. otoliths or scales, where products $\gg 120$ bp often fail to amplify.
5. Most applications (excluding mapping) should concentrate on isolating tri- and tetranucleotide primers if possible.
6. Signature (stuttering) patterns and, more seriously, the ability to amplify loci depends on quality of primer synthesis. Thus, when a dependable supplier has been identified, changes should not be made unless this is completely unavoidable.
7. Invest time and resources at the development stage to identify loci that amplify consistently. Avoid the temptation to use the first polymorphic markers identified.

2. Population genetics

Prudent management decisions should be based on an understanding of the stock structure of the biological resource in question. However, delimiting the putative boundaries of such stocks in marine/aquatic environments is a complex and difficult task (see Carvalho and Hauser, 1994, for an excellent discussion of this area). Genetic markers have been used to define the underlying stock structure of many fish species, in particular salmonids. The very high levels of variability associated with microsatellites, the speed of processing and the potential to isolate large numbers of loci provides a marker system capable of detecting differences among closely related populations.

One of the first species to be investigated for within and among population variability using microsatellites was rainbow trout (Nielsen et al., 1994). This survey revealed similar patterns of differentiation for mtDNA and the microsatellite locus employed. Populations of potamodromous rainbow trout from Lake Ontario have also been investigated using microsatellite loci and mtDNA (Dueck, 1994; O'Connell et al., 1996a). A comparison of marker sets revealed that the number of mtDNA haplotypes was similar to the number of alleles observed at microsatellite loci, although the single and

combined microsatellite loci data revealed significantly higher levels of differentiation. The higher level of genetic divergence observed for microsatellite data in this case is consistent with sex-biased dispersal and/or the reproductive success of precocious male parr (as mtDNA is inherited primarily through the female line), although such a result most likely reflects a random or chance event. An example of a random event would be where founding individuals in a new population shared allele frequencies at some loci with an adjacent population. Although the allele frequencies at the majority of loci would differ between populations, this would not be apparent if only a few loci were investigated. This type of random event can usually be excluded by sampling several loci. Studies using more than one data set to describe population structure have the potential to describe population interactions more fully than would be possible by using a single marker set. However, it is important to consider the levels of variability of the marker systems employed. If mtDNA is less variable than microsatellites, an increased level of microsatellite DNA structuring, relative to mtDNA structuring, could simply reflect the increased mutation rate of the former and not represent a demographic phenomenon. Caution also has to be exercised when drawing conclusions about differences between mtDNA and nuclear DNA data because random events cannot be eliminated as a possible major determinant of the mtDNA patterns of variability.

Atlantic salmon from sites within the river Dee (Wales) have also been studied using microsatellites (O'Connell et al., 1996b). Previous surveys on the same fish using allozyme variability suggested that the populations were largely isolated from each other (O'Connell, unpublished data). The microsatellite data showed an F_{ST} value (an index of differentiation) twice that of the value generated from the allozyme survey. This is not surprising given that the allozymes identified only 9% of the total number of alleles revealed at the microsatellite loci. Microsatellite data were also more accurate than the allozyme data in determining how precisely individuals from each population could be assigned to their population of origin (using the computer package SPAM 1.01 ± Anon., 1995) although both data sets performed poorly. One potential problem with using microsatellite loci in GSI (genetic stock identification ± determining individual stock proportions in a mixed sample) is the very high number of alleles detected. This phenomenon, also observed with single-locus minisatellite probes, can confound the analysis although binning or grouping together of alleles that fall within a particular size range can eliminate or reduce the problem (discussed in detail later). The very high levels of variability also suggest that some microsatellites will not be useful tools for comparing the levels of genetic diversity among populations. Comparisons of allelic diversity will prove more informative than estimates of heterozygosity, although the very high levels of allelic diversity in some cold-water fishes (Brooker et al., 1994; McConnell et al., 1995a) suggest that microsatellites may be insensitive to all but the most severe population bottlenecks.

3. Parentage and kinship analysis

The highly variable nature of microsatellite loci makes these markers particularly suited for the investigation of kinship relationships and paternity analysis. Such analyses can be useful for the management of captive populations and in understanding mating patterns in the wild. Herbinger et al. (1996b) investigated cod populations from the North-west Atlantic at six microsatellite loci. The objectives of the study were to determine if natural selection could be observed directly in a cohort of larvae and to estimate a minimum effective population size for the spawning aggregation that produced the larval cohort. Kinship relationships among individuals within the study population were assessed using a likelihood ratio method based on the observed and expected number of shared alleles in an idealized situation (details: Thompson, 1991; Herbinger et al., 1996b). This analysis permitted the authors to conclude that there was no evidence of family structure within the larval cohort. This was the first time that family structure in a natural assemblage of larvae from a marine animal in the wild had been determined. This finding is important as it has been proposed that some marine animals may have an effective population size considerably smaller than the actual population size, with relatively few crosses succeeding through sweepstakes evolution (Hedgecock, 1994). The role of natural selection, however, could not be assessed as sibship was not detected in this study, although the inbreeding effective population size of the cohort investigated (in sensu Hartl and Clark, 1989) was calculated to be 2800 individuals. In the past, to estimate effective population size, researchers had to make unrealistic assumptions regarding the data, e.g. no selection, no migration, randomly mating populations, and/or

required temporal data (discussion: Waples, 1989; Bartley et al., 1992). Thus, the estimates were likely to be subject to greater errors than the allele-sharing likely ratio method of Herbinger et al. (1996b). The importance of the effective population size N_e to many aspects of population and conservation biology suggests that this analytical method, used in conjunction with microsatellite data, will prove invaluable to the conservation management of fishes.

4. Genome mapping

Aquaculture is a vital contributor to the economy of many countries worldwide. However, the vast majority of species and strains reared globally are relatively unimproved for commercially important traits such as growth rate, disease resistance and age of sexual maturation. Thus, the potential for genetic improvement in fish species, compared with domestic livestock, is very high. Most traits of commercial importance are polygenic, i.e. the trait is controlled by more than one locus. These loci are commonly referred to as quantitative trait loci (QTL). The first detection of a QTL was for bunt and rust resistance in wheat. Since this initial description, QTL have been mapped in a wide variety of species including plants (Bradshaw et al., 1995), insects (Coyne and Charlesworth, 1986) and livestock (Andersson et al., 1994). Marker-assisted selection (MAS) can be carried out with an understanding of the linkage relationships between QTLs and markers. This type of selection is particularly useful for characters that are expressed at/after the onset of maturity, because it eliminates the need to rear animals to this late stage. MAS can significantly increase the intensity of selection when complemented with traditional approaches. However, in order to map QTLs, a detailed linkage map is required, with variable markers distributed throughout the genome. Linkage between a marker locus and a segregating QTL allele can only be reliably detected at or below a map distance of 20 centimorgans (cM) (Soller et al., 1976), although the generation of higher-density maps does little to increase the power of QTL detection (Haley and Knott, 1994). Given the tetraploidization, and consequently, the large size of many fish genomes (Allendorf and Thorgaard, 1984), a large number of variable, widely distributed markers are required for many fish genome mapping efforts, e.g. about 200 ± 300 evenly spaced markers are required for salmonids. Poojuang and Hallerman (1996), in an excellent review of mapping and QTL, also reported that the number of QTLs detected is more responsive to the degree of marker polymorphism than to the number of individuals screened. Both these observations suggest that microsatellites will prove to be an ideal tool for genome mapping.

Microsatellites have been quickly adopted for mapping applications in fish including tilapia, zebrafish, Atlantic salmon, rainbow trout and carp. In tilapia, 60 microsatellite loci have been reported from the University of New Hampshire (<http://tilapia.unh.edu> ± Lee and Kocher, 1996). Segregating loci (73% of variable loci) have been mapped to 15 linkage groups. Additional microsatellite loci are also being generated by other groups (Ambali, 1996) and this collaborative approach is the most efficient method for long-term mapping projects. To the best of our knowledge, the first recorded QTL mapped in a fish species is in rainbow trout (Jackson, 1995). In a map generated from allozyme, RAPD and microsatellite data, a significant association between upper temperature tolerance and alleles at two microsatellite loci was observed in more than one family, and the application of these data should prove especially useful in broodstock selection. The testing of loci on more than one family is important because without these data, observations on possible linkage relations are limited. Furthermore, in the salmonids, which are currently 'diploidizing', a process which involves centric and pericentric fusions and fissions, variation in chromosome number has been reported (Allendorf and Thorgaard, 1984).

5. Concluding remarks

A large number of molecular techniques that can identify genetic variation within species are now available. This review has described the application of VNTR DNA and microsatellites in particular to fish studies, although other techniques are available (Greider, 1991; Klein et al., 1993; Carvalho and Hauser, 1994; Ferguson, 1994; Ward and Grewe, 1994; Ferguson et al., 1995). The different molecular techniques now available have particular applications for which they are most powerful. For instance, the rapid development of a framework genetic map may be readily achieved with RAPD markers. An initial survey

of genetic variation for a species in which no DNA development work has been carried out is most easily realized with an allozyme survey. Nevertheless, the tool with the greatest overall potential to address the most common applications of molecular markers to fish and fisheries-related questions (excluding cross-species phylogenetic applications) is microsatellites. There are, however, a number of potentially serious problems associated with generating and interpreting microsatellite data, including null alleles, poor resolution of larger alleles, designing appropriate sampling regimes, and deciding on which statistics to employ (FST versus RST). Most of these problems can be offset, or at least reduced, by carefully considering the biological questions which are to be addressed, using tetranucleotide loci and keeping the product size to a minimum. Although smaller loci may reveal lower numbers of alleles, they amplify more reliably and are considerably easier to score. As the allele size range in smaller loci is reduced, more than one locus can also be frequently amplified in a single reaction, i.e. multiplexing, which considerably reduces the costs associated with investigating microsatellite variability. Microsatellites are being rapidly characterized for many teleosts, although most of the research emphasis to date has concentrated on the commercially important salmonids and cichlids. However, the ability to amplify presumably homologous loci in related species (McConnell et al., 1995a; Morris et al., 1996; Rico et al., 1996) can help reduce the development costs associated with related, but less commercially important, species. It is to be hoped that this will allow those countries, which often have limited financial resources, to address population and conservation genetics issues. This is important, as these countries are frequently rich in biological diversity but have been unable to investigate genetic variability owing to the high technical demands and running costs associated with the employment of other types of DNA markers.

References

- Tautz, D., Renz, M., 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Res.*, 12, 4127-4138.
- Tautz, D., 1989. Hyper variability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res.*, 17, 6463-6471.
- O'Reilly, P., Wright, J.M., 1995. The evolving technology of DNA fingerprinting and its application to fisheries and aquaculture. *J. Fish. Biol.* 47., 29-55.
- Carvalho, G.R., Hauser, L., 1994. Molecular genetics and the stock concept in fisheries. *Rev. Fish. Biol. Fish.*, 4, 326-350.
- Nielsen, J.L., Gan, C.A., Wright, J.M., Morris, D.B., Thomas, W.K., 1994. Biogeographic distributions of mitochondrial and nuclear markers for southern steelhead. *Mol. Mar. Biol. Biotech.*, 3, 281-293.
- Dueck, L.A., 1994. Population divergence of introduced rainbow trout (*Oncorhynchus mykiss*) in the Lake Ontario watershed, based on the mitochondrial genome. MSc thesis, Univ. of Guelph, Canada., 173 pp.
- O'Connell, M., Danzmann, R.G., Cornuet, J. M., Wright, J.M., Ferguson, M.M., 1996a. Differentiation of rainbow trout populations in Lake Ontario and the evaluation of the stepwise mutation and infinite allele mutation models using microsatellite variability. *Can. J. Fish. Aquat. Sci.*, 54, 122-143.
- O'Connell, M., Dillon, M.C., Wright, J.M., 1996b. Development of microsatellite markers for genetic discrimination of Prince William Sound herring populations. Report to the Alaska Dept of Fish and Game., 31 pp.
- Anon. A., 1995. SPAM 1.01 ± Statistics Program for Analyzing Mixtures. Alaska Dept of Fish and Game, Division of Commercial Fisheries Management and Development, Genetics Lab, 333 Raspberry Rd. Anchorage., AK 99518.
- Brooker, A.L., Cook, D., Bentzen, P., Wright, J.M., Doyle, R.W., 1994. Organization of microsatellites differs between mammals and cold-water teleost fishes. *Can. J. Fish. Aquat. Sci.*, 51, 1959-1966.
- McConnell, S.K., Hamilton, L., Morris, D., Cook, D., Pacquet, D., Bentzen, P., Wright, J.M. 1995a. Isolation of salmonid microsatellite loci and their application to the population genetics of Canadian east coast stocks of Atlantic salmon. *Aquaculture.*, 137, 19-30.
- Herbinger, C.M., Doyle, R.W., Taggart, C.T., Lochmann, S.E., Brooker, A.L., Wright, J.M., Cook, D. 1996b. Family relationships and effective population size in a natural cohort of cod larvae. *Can. J. Fish. Aquat. Sci.*, 54, 11-18.

- Thompson, E.A., 1991. Estimation of relationships from genetic data. In Rao, C.R. and Chakraborty, R., eds. *Handbook of Statistics*, Vol. VIII. Amsterdam, Holland: Elsevier, pp. 255-269.
- Hedgecock, D., 1994. Does variance in reproductive success limit effective population sizes of marine organisms In Beaumont, A.R., ed. *Genetics and Evolution of Aquatic Organisms*. London: Chapman and Hall., pp. 122-134.
- Hartl, D.L., Clark, A.G., 1989. *Principles of Population Genetics*. Sunderland, MA: Sinauer Assoc.
- Waples, R.S., 1989. A generalized method for estimating population size from temporal changes in allele frequency. *Genetics.*, 121, 379-391.
- Bartley, D., Bagley, M., Gall, G., Bentley, B., 1992. Use of linkage disequilibrium data to estimate effective population size of hatchery and natural fish populations. *Cons. Biol.*, 6, 365-375.
- Coyne, J.A., Charlesworth, B., 1986. Location of an X-linked factor causing sterility in male hybrids of *Drosophila simulans* and *D. mauritiana*. *Heredity.*, 57, 243-246.
- Andersson, L., Haley, C.S., Ellegren, H., Knott, S.A., Johansson, M., Andersson, K., Andersson- Eklund, L., Edfors-Lilja, I., Fredholm, M., Hansson, I., Hakansson, J., Lundstrom, K., 1994. Genetic mapping of quantitative trait loci for growth and fatness in pigs. *Genetics.*, 263, 1771-1774.
- Soller, M., Brody, T., Genizi, A., 1976. On the power of experimental designs for the detection of linkage between marker loci and quantitative trait loci in crosses between inbred lines. *Theor. Appl. Genetic.*, 47, 35-39.
- Haley, C.S., Knott, S.A., 1994. Interval mapping. In *Proc. 5th World Congress on Genetics Applied to Livestock Production*, Vol. 21. Guelph, Ontario: Dept of Animal and Poultry Sciences, Univ. of Guelph, pp. 25-32.
- Allendorf, F.W., Thorgaard, G.H., 1984. Tetraploidy and the evolution of salmonid fishes. In Turner, B.J., ed. *Evolutionary Genetics of Fishes*. New York: Plenum Press, pp. 1-53.
- Lee, W. J., Kocher, T.D., 1996. Microsatellite DNA markers for genetic mapping in *Oreochromis niloticus*. *J. Fish Biol.*, 49, 169-171.
- Ambali, A.J.D., 1996. The relationship between domestication and genetic diversity of *Oreochromis* species in Malawi: *Oreochromis shiranus shiranus* (Boulenger) and *Oreochromis shiranus chilwae* (Trewavas). PhD thesis, Dalhousie Univ., Canada. 204 pp.
- Jackson, T.R., 1995. Linkage analysis of molecular markers and a search for quantitative trait loci for upper temperature tolerance in rainbow trout (*Oncorhynchus mykiss*). PhD thesis, Univ. of Guelph, Canada. 81 pp.
- Greider, C.W., 1991. Telomeres. *Curr. Opin Cell Biol.*, 3, 444-451.
- Klein, D., Ono, H., O'hUigin, C., Vincek, V., Goldschmidt, T., Klein, J., 1993. Extensive MHS variability in cichlid fishes. *Nature.*, 364, 330-334.
- Carvalho, G.R., Hauser, L., 1994. Molecular genetics and the stock concept in fisheries. *Rev. Fish Biol. Fish.*, 4, 326-350.
- Ferguson, A., Taggart, J.B., Prodohl, P.A., McMeel, O., Thompson, C., Stone, C., McGinnity, P., Hynes, R.A., 1995. The application of molecular markers to the study and conservation of fish populations, with special reference to *Salmo*. *J. Fish. Biol.*, 47, 103-126.
- Ward, R.D., Grewe, P., 1994. Appraisal of molecular genetic techniques in fisheries. *Rev. Fish Biol. Fish.*, 4, 300-325.
- Ferguson, M. 1994. The role of molecular genetic markers in the management of cultured fishes. *Rev. Fish Biol. Fish.*, 4, 351-373.
- Morris, D.B., Richard, K.R., Wright, J.M., 1996. Microsatellites from rainbow trout (*Oncorhynchus mykiss*) and their use for genetic studies of salmonids. *Can. J. Fish. Aquat. Sci.*, 53, 120-126.
- Rico, C., Rico, I., Hewitt, G., 1996. 470 million years of conservation of microsatellite loci among fish species. *Proc. R. Soc. Lond.*, 263, 549-557.