Application of a double-enrichment procedure for microsatellite isolation and the use of tailed primers for high throughput genotyping

Fábio Mendonça Diniz1,2, Arati Iyengar2,3, Paulo Sarmanho da Costa Lima1, Norman Maclean2 and Paul Bentzen4

1Embrapa Meio-Norte, Teresina, PI, Brazil.
2School of Biological Sciences, University of Southampton, Southampton, UK.
3Department of Forensic and Investigative Sciences, University of Central Lancashire, Preston, UK.
4Department of Biology, Dalhousie University, Halifax, NS, Canada.

Abstract

The number of microsatellite loci and their allelic diversity contribute to increase accuracy and informativity of genetic estimates, however, the isolation of microsatellite loci is not only laborious but also quite expensive. We used (GATA)n and (GACA)n tetranucleotide probes and single- and double-enrichment hybridization to construct and screen a genomic library with an increased proportion of DNA fragments containing repeat motifs. Repeats were found using both types of hybridization but the double-enrichment procedure recovered sequences of which 100% contained (GATA)n and (GACA)n motifs. Microsatellite loci primers were then designed with an M13R-tail or CAG-tag to produce scorable PCR products with minimal stutter. The approach used in this study suggests that double-enrichment is a worthwhile strategy when isolating repeat motifs from eukaryotic genomes. Moreover, the use of tailed microsatellite primers provides increased resolution for compound microsatellite loci, with a significant decrease in costs.

Key words: double-enrichment, double-hybridization, stutter bands, tailed microsatellite primers.

Received: July 7, 2006; Accepted: October 24, 2006.

Microsatellites, also known as simple sequence repeats (SSRs), are present throughout the eukaryotic genome, often at high concentrations. They are useful markers for a wide range of analyses because in many taxa they show high levels of intraspecific genetic variability. These markers are also easily detectable by the polymerase chain reaction (PCR), producing highly reproducible results when compared with other molecular markers (Tautz et al., 1986; Wright & Bentzen, 1994; O’Connell & Wright, 1997). Applications of microsatellite markers include biomedical diagnosis of diseases (Girardet et al., 2005), genome mapping (Park et al., 2005), parentage analysis and relatedness of individuals or groups (Porta et al., 2006), assessing demographic history (Jacobsen et al., 2005) and examining the genetic structure of subpopulations and populations (Diniz et al., 2004). The information revealed by the genetic data obtained from these markers is a function of the number of loci and allelic diversity, both of which contribute interactively to increase the accuracy and success of genetic estimates and consequently the reliability of inferences (Brookfield & Parkin, 1993; Chakraborty & Jin, 1993; Blouin et al., 1996). However, the isolation of microsatellite loci is not only demanding but also expensive because of the need for numerous polymorphic microsatellites and, consequently, specific fluorescent primers. Many techniques have been applied (Zane et al., 2002; Fujishima-Kanaya et al., 2003; Waldbieser et al., 2003; Chen et al., 2005) but there is no guarantee that a particular method will result in a large number of loci being identified. In addition, current mainstream analysis methods involve the use of fluorescently labeled primers which are expensive, especially when large numbers of loci are required.

In an effort to increase the proportion of genomic DNA fragments containing repeat motifs a simple and inexpensive methodology is described and optimized for the construction of an enriched genomic library. Moreover, one of the primers from every pair designed for the microsatellite flanking sequences were M13-tailed or CAG-tagged to produce microsatellite primer sets that were easier to score and did not require expensive fluorescent labeling.
The spiny lobster *Panulirus argus* was used as a model organism in this study. Specimens were collected along the Brazilian coastline between latitudes 03°43' and 03°54' South. Samples were taken from the fresh tissue of the walking legs of the lobsters and immediately preserved in 20% (v/v) dimethyl sulfoxide (DMSO) in saturated sodium chloride solution (Dawson et al., 1998). Genomic DNA was extracted from the tissue using the phenol/chloroform-isoamyl alcohol protocol described by Sambrook et al. (1989).

Two genomic libraries enriched for tetranucleotide repeats were prepared as outlined in Hamilton et al. (1999) and McPherson et al. (2001), with minor modifications (Diniz et al., 2005). The first library was single-enriched for (GACA)₄ and (GATA)₇. The second library was constructed using an additional enrichment after the first hybridization capture (biotin/streptavidin) technique (Kijas et al., 1994) with the previously used tetranucleotide probes. Genomic DNA (~200 ng/μL) of each library was digested separately with HindIII, RsaI, BstU1 and HaeIII at 37 °C overnight. Digests were recovered using PCR Purification columns (Qiagen) according to manufacturer’s instructions, and the terminal ends dephosphorylated using calf intestinal phosphatase. The digested-dephosphorylated genomic DNA was cleaned of all modifying enzymes and then ligated to double-stranded linkers, termed SNX (Sfil, NheI and XmnI) for the restriction sites they contain or form when dimerisation occurs (SNX-F: 5’-CTAAGGCGCTTGCTAGCAGAAGC-3’ and SNX-R: 5’-pGCTTCTGTAGCAAGGCCTTAGAAGAA-3’). The ligation reaction used the enzyme T₄ DNA ligase in the presence of 0.5 ng/μL of each primer of each pair to facilitate fluorescent labeling. Amplifications were carried out in a 50 μL reaction volume containing 20-100 ng DNA, 1 x Thermopol® buffer supplemented with 1.5 mM MgCl₂, 50 μM of each dNTP, 0.5 U DNA polymerase, 5 μM of each primer and the CEQ DTCS - Quick Start kit (Beckman Coulter Inc.) and sequenced on a CEQ 8000XL DNA Analysis System (Beckman Coulter Inc.). Confirmed positives were further sequenced for the opposite strand using the SP6 (5’-CATAACGATTTAGGTGACACTATAG-3’) and M13-R universal primers.

Microsatellite loci primers were designed on the unique flanking regions of each locus using PRIMER 3 (Rozen & Skaltsky, 2000) and an oligonucleotide tail corresponding to the 5’-GGAAACAGCTATGACCATG-3’ of each primer. Each clone was cycle-sequenced (96 °C for 3 min, followed by 40 cycles of 96 °C for 20 s, 52 °C for 20 s and 60 °C for 4 min) in one direction using the M13-F primer and the CEQ DTCS - Quick Start kit (Beckman Coulter Inc.) and sequenced on a CEQ 8000XL DNA Analysis System (Beckman Coulter Inc.). Amplified-enriched DNA from the single- and double-enrichment procedures was ligated into the pDrive cloning vector (Qiagen PCR cloning kit) and used to transform OneShot® competent *Escherichia coli* (Invitrogen) which were plated onto Laura-Bertani agar (LB; Difco) supplemented with 50 μg mL⁻¹ ampicillin. Plasmid inserts were amplified directly from the transformed *E. coli* colonies (PCR screening) using primers (M13-F: 5’-GGAAACAGCTATGACCATG-3’ and M13-R: 5’-GTAAACACGCCTAGAAGAA-3’) that flank the cloning site and size-fractionated on 1.5% (w/v) agarose gels. Recombinant *E. coli* clones giving a positive signal for the insertion of the *P. argus* 500-1000 base-pair (bp) DNA sequence were incubated overnight on LB-ampicillin and the plasmid DNA purified and extracted using the Qiaprep® spin miniprep kit (Qiagen). Each clone was cycle-sequenced (96 °C for 3 min, followed by 40 cycles of 96 °C for 20 s, 52 °C for 20 s and 60 °C for 4 min) in one direction using the M13-F primer and the CEQ DTCS - Quick Start kit (Beckman Coulter Inc.) and sequenced on a CEQ 8000XL DNA Analysis System (Beckman Coulter Inc.). Confirmed positives were further sequenced for the opposite strand using the SP6 (5’-CATAACGATTTAGGTGACACTATAG-3’) and M13-R universal primers.

Two genomic libraries enriched for tetranucleotide repeats were prepared as outlined in Hamilton et al. (1999) and McPherson et al. (2001), with minor modifications (Diniz et al., 2005). The first library was single-enriched for (GACA)₄ and (GATA)₇. The second library was constructed using an additional enrichment after the first hybridization capture (biotin/streptavidin) technique (Kijas et al., 1994) with the previously used tetranucleotide probes. Genomic DNA (~200 ng/μL) of each library was digested separately with HindIII, RsaI, BstU1 and HaeIII at 37 °C overnight. Digests were recovered using PCR Purification columns (Qiagen) according to manufacturer’s instructions, and the terminal ends dephosphorylated using calf intestinal phosphatase. The digested-dephosphorylated genomic DNA was cleaned of all modifying enzymes and then ligated to double-stranded linkers, termed SNX (Sfil, NheI and XmnI) for the restriction sites they contain or form when dimerisation occurs (SNX-F: 5’-CTAAGGCGCTTGCTAGCAGAAGC-3’ and SNX-R: 5’-pGCTTCTGTAGCAAGGCCTTAGAAGAA-3’). The ligation reaction used the enzyme T₄ DNA ligase in the presence of 0.5 ng/μL of each primer of each pair to facilitate fluorescent labeling. Amplifications were carried out in a 50 μL reaction volume containing 20-100 ng DNA, 1 x Thermopol® buffer supplemented with 1.5 mM MgCl₂, 50 μM of each dNTP, 0.5 U DNA polymerase, 5 μM of each primer and the CEQ DTCS - Quick Start kit (Beckman Coulter Inc.) and sequenced on a CEQ 8000XL DNA Analysis System (Beckman Coulter Inc.). Confirmed positives were further sequenced for the opposite strand using the SP6 (5’-CATAACGATTTAGGTGACACTATAG-3’) and M13-R universal primers.

Microsatellite loci primers were designed on the unique flanking regions of each locus using PRIMER 3 (Rozen & Skaltsky, 2000) and an oligonucleotide tail corresponding to the 5’-GGAAACAGCTATGACCATG-3’ of each primer. Each clone was cycle-sequenced (96 °C for 3 min, followed by 40 cycles of 96 °C for 20 s, 52 °C for 20 s and 60 °C for 4 min) in one direction using the M13-F primer and the CEQ DTCS - Quick Start kit (Beckman Coulter Inc.) and sequenced on a CEQ 8000XL DNA Analysis System (Beckman Coulter Inc.). Confirmed positives were further sequenced for the opposite strand using the SP6 (5’-CATAACGATTTAGGTGACACTATAG-3’) and M13-R universal primers.
HaeIII-digested samples produced mainly high (> 3000 bp) molecular weight fragments with no specific banding pattern and were subjected to further experiments. Minimum over-amplification of ligated DNA and repeat-enriched fragments was warranted when PCR was carried out with 15 cycles.

The libraries enriched for tetranucleotide repeat motifs showed high levels of microsatellite enrichment (Figure 1). Each enriched library had 32 provisional positive clones sequenced. The single-enriched library showed very similar numbers of sequences containing no repeat motifs and sequences recovered in duplicate, except for the (GATA)$_4$ library, which had four times more sequences containing no repeats compared to the library probed with (GATA)$_4$ (Figure 1A). Moreover, this more commonly used sequence enrichment resulted in three sequences being unsuitable for primer design because they either had short or no available flanking regions (Figure 1A). The mean percentage of sequences with repeat motifs from single-enriched libraries was high (67%), ranging from about 9 microsatellite sequences (55%) for (GACA)$_4$ to about 13 microsatellite sequences (80%) for (GATA)$_4$ libraries (Figure 1B). The overall high success rate of the single-enrichment protocol employed here was probably due to the presence of large numbers of these tetranucleotide repeats within the P. argus genome coupled with the effectiveness of using the hybridization capture (biotin/streptavidin) enrichment technique (Kijas et al., 1994). Double-enrichment produced libraries in which all of the sequences contained (GACA)$_n$ and (GATA)$_n$ motifs (16 sequences each). Out of the 32 sequences isolated after double-enrichment, six had limited flanking sequence for primer design or were second copy clones. The high frequency of positive clones found in the recombinant colonies also suggests that PCR screening for fragments of appropriate length (500-1000 bp inserts) considerably increases the chance of detecting clones with repeat arrays and with longer flanking regions for primer design.

Microsatellite sequences from both procedures were characterized by length and type according to Weber (1990) (Figure 1B). Most tandem arrays recovered from the single enrichment procedure were compound (11 sequences) or imperfect repeats (3 sequences). This was even more noticeable in the double-enriched library, where 21 microsatellite sequences were compound repeats and 7 sequences were imperfect repeats. Compound microsatellites were frequently associated with dinucleotides (AC/TG)$_n$ in the case of the GACA-enriched library and with (GA/CT)$_n$ in the case of the GATA-enriched library. The relative frequency of perfect repeat sequences was higher in single-enrichment, especially for the GATA repeat motif. Microsatellite sequences have been deposited in GenBank under accession numbers AY536335 to AY536353 and further characterized (including the primer sequences) by Diniz et al. (2004, 2005).

Genotyping 32 P. argus samples at all the microsatellite loci isolated in this study showed clear bands with no shadow bands and minimal stutter (Figure 2). Shadow or stutter bands, occurring during PCR due to replication slippage are not a common difficulty in scoring tetranucleotide loci, but may be a problem for microsatellite loci consisting of compound tetranucleotide loci (i.e. tetra + dinucleotides) producing stutter fragments to levels seen with dinucleotides (Hauge & Litt 1993; O’Reilly et al., 2000). The tailing of one of the microsatellite primers per locus has been found to improve scoring of stuttering fragments from di-
tri- or tetranucleotide microsatellite amplifications (Oetting et al., 1998; Boutin-Ganache et al., 2001). This increases the robustness of scoring, and eliminates the end-labeling of one of the PCR primers for each specific locus, thus reducing the costs associated with labeling primers with fluorescent dye (Yu et al., 2001). However, the inclusion of such tails does increase the chance of formation of hairpins and primer-dimers (Hauswaldt & Glenn, 2003). Therefore, the selection of each tail - primer combination should be tested using a specialized program such as the GENRUNNER program to minimize potential secondary structures. Because the number of primers is increased in the PCR reaction, optimization of each primer set could be more labour-intensive. However, this is a drawback that can easily be overcome, and the benefits of the procedure far outweigh the drawbacks.

Our results show that double-enrichment should certainly be considered in order to overcome low percentages of clones containing tandem repeats. In this study, we were able to increase the number of sequences with repeat sequences to 100%. Even though we did not notice any problem with recombination in the PCR following the hybridization steps, care should be taken to avoid this possible complication. Keeping PCR cycles to a minimum seems to be an effective way to overcome the problem. The use of a double-enrichment protocol and the addition of a sequence tag to the microsatellite amplification primers provide an inexpensive approach for increasing the number of loci that can be screened and decreases the costs of genotyping. We are conducting further work to apply these markers to population discrimination of *P. argus* throughout its range.

### Acknowledgments

We thank Dr. Masayoshi Ogawa, from the Federal University of Ceará, for providing lobster tissue samples and Ian Paterson, from Dalhousie University, Canada, for technical assistance. Thanks are also due to the Brazilian Institute for the Environment and Natural Renewable Resources (IBAMA). FM Diniz was supported by a PhD scholarship from the Brazilian Government (CAPES).

### References

- Blouin MS, Parsons M, Lacaille V and Lotz S (1996) Use of microsatellite loci to classify individuals by relatedness. Mol Ecol 7:393-401.
- Boutin-Ganache I, Raposo M, Raymond M and Deschepper CF (2001) M13-tailed primers improve the readability and usability of microsatellite analyses performed with two different allele-sizing methods. Biotechniques 31:24-28.
- Brookfield JFY and Parkin DT (1993) Use of single-locus DNA probes in the establishment of relatedness in wild populations. Heredity 70:660-663.
- Chakraborty R and Jin L (1993) A unified approach to study hypervariable polymorphism: Statistical considerations of determining relatedness and population distances. In: Pena SDJ, Chakraborty R, Epplen JT and Jeffreys AJ (eds) DNA Fingerprinting: State of the Science. Birkhauser Verlag, Basel, pp. 153-175.
- Chen K, Knorr C, Bornemann-Kolatzki K, Ren J, Huang L, Rohrer GA and Brenig B (2005) Targeted oligonucleotide-mediated microsatellite identification (TOMMI) from large-insert library clones. BMC Genet 6:54.
- Dawson MN, Raskoff KA and Jacobs DK (1998) Field preservation of marine invertebrate tissue for DNA analyses. Mol Mar Biol Biotech 7:145-152.
- DINIZ FM, Maclean N, Ogawa M, Paterson I and Bentzen P (2005) Microsatellites in the overexploited spiny lobster, *Panulirus argus*: Isolation, characterization of loci and potential for intraspecific variability studies. Cons Gen 6:637-641.
- DINIZ FM, Maclean N, Paterson IG and Bentzen P (2004) Polymorphic tetranucleotide microsatellite markers in the Caribbean spiny lobster, *Panulirus argus*. Mol Ecol Notes 4:327-329.
- Fujishima-Kanaya N, Toki D, Suzuki K, Sawazaki T, Hiraiwa H, Iida M, Hayashi T, et al. (2003) Development of 50 gene-associated microsatellite markers using BAC clones and the construction of a linkage map of swine chromosome 4. Anim Genet 34:135-141.
- Girardet A, Moncla A, Hamamah S and Claustres M (2005) Strategies for preimplantation genetic diagnosis of Angelman syndrome caused by mutations in the *UBE3A* gene. Reprod Biomed Online 10:519-526.
- Hamilton MB, Pinceau EL, Di Fiore A and Flesher RC (1999) A universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. Biotechniques 27:500-507.
- Hauge XY and Litt M (1993) A study of the origin of shadow bands seen when typing dinucleotide repeat polymorphisms by PCR. Hum Mol Genet 2:411-415.
- Hauswaldt JS and Glenn TC (2003) Microsatellite DNA loci from Diamondback terrapin (*Malaclemys terrapin*). Mol Ecol Notes 3:174-176.
- Jacobsen BH, Hansen MM and Loescheke V (2005) Microsatellite DNA analysis of northern pike (*Esox lucius*) populations: Insights into the genetic structure and demographic history of a genetically depauperate species. Biol J Linn Soc Lond 84:91-101.
- Kijas JMH, Fowler JCS, Garbett CA and Thomas MR (1994) Enrichment of microsatellites from the citrus genome using biotinylated oligonucleotide sequences bound to streptavidin-coated magnetic particles. Biotechniques 16:657-662.
- McPherson AA, O’Reilly PT, McParland TL, Jones MW and Bentzen P (2001) Isolation of nine novel tetranucleotide microsatellites in the Atlantic herring (*Clupea harengus*). Mol Ecol Notes 1:31-32.
- O’Connell M and Wright JM (1997) Microsatellite DNA in fishes. Rev Fish Biol Fish 7:331-363.
- Oetting WS, Armstrong CM, Ronan SM, Young TL, Sellers TA and King RA (1998) Multiplexed short tandem repeat polymorphisms of the Weber 8A set of markers using tagged primers and infrared fluorescence detection. Electrophoresis 19:3079-3083.
- Oetting WS, Lee HK, Flanders DJ, Wiesner GL, Sellers TA and King RA (1995) Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tagged primers. Genomics 30:450-458.
O’Reilly PT, Canino MF and Bentzen P (2000) Isolation of twenty low stutter di- and tetranucleotide microsatellites for population analyses of walleye pollock (Theragra chalcogramma) and other gadoids. J Fish Biol 56:1074-1086.

Park Y-H, Alabady MS, Ulloa M, Sickler B, Wilkins TA, Yu J, Stelly DM, et al. (2005) Genetic mapping of new cotton fiber loci using EST-derived microsatellites in an interspecific recombinant inbred line cotton population. Mol Genet Genomics 274:428-441.

Porta J, Porta JM, Martínez-Rodríguez G and Alvarez MC (2006) Genetic structure and genetic relatedness of a hatchery stock of Senegal sole (Solea senegalensis) inferred by microsatellites. Aquaculture 251:46-55.

Rozen S and Skaletsky HJ (2000) Primer3 on the WWW for general users and biologist programmers. In: Krawetz S and Misener S (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, pp 365-386.

Sambrook J, Fritsch EF and Maniatis T (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp 6.4-6.11.

Tautz D, Trick M and Dover GA (1986) Cryptic simplicity in DNA is a major source of genetic variation. Nature 322:652-656.

Waldbieser GC, Quiniou SM and Karsi A (2003) Rapid development of gene-tagged microsatellite markers from bacterial artificial chromosome clones using anchored TAA repeat primers. Biotechniques 35:976-979.

Weber JL (1990) Informativeness of human (dC-dA)n(dG-dT) polymorphisms. Genomics 7:524-530.

Wright JM and Bentzen P (1994) Microsatellites: Genetic markers for the future. Rev Fish Biol Fish 4:384-388.

Yu K, Haffner M and Poysa V (2001) Tailed primer base excision sequence scanning (TP-BESS) for detection of single nucleotide polymorphisms (SNPs). Plant Mol Biol Rep 19:49-54.

Zane L, Bargelloni L and Patarnello T (2002) Strategies for microsatellite isolation: A review. Mol Ecol 11:1-16.

Associate Editor: Emmanuel Dias Neto