Indian J. Fish., 59(3) : 21-24, 2012

Rapid and cost effective development of SSR markers using next generation sequencing in Indian major carp, *Labeo rohita* (Hamilton, 1822)

B. P. SAHU, A. PATEL, L. SAHOO, P. DAS, P. K. MEHER AND P. JAYASANKAR
Central Institute of Freshwater Aquaculture, Kausalyaganga, Bhubaneswar - 751 002, Odisha, India
e-mail: pdas77@hotmail.com

ABSTRACT

Development of species specific microsatellite markers using conventional approaches such as partial and enriched genomic library screening are costly and time consuming. Next generation sequencing is a promising methodology for acquisition of abundant microsatellites from the genome. In an effort to enrich the present microsatellite database for marker map generation in *Labeo rohita*, Roche-454 GS-FLX sequencing platform was used and results were compared with the enriched library screening. A library enriched for 'GT' repeat was prepared and screened using Sanger sequencing. Screening of enriched library produced 1840 clones resulting in only 238 sequences having enough flanking region for primer designing. Roche- 454 sequencing of 10 MB of rohu genome generated 387 sequences containing microsatellite repeat with enough flanking region. Enriched library preparation, screening and Sanger sequencing took nearly one and half year, whereas data generation by next generation sequencing took just two months. Moreover equally good number of markers could be generated by high throughput sequencing platform which worked out to be one third cheaper.

Keywords: *Labeo rohita*, Next generation sequencing, Rohu, SSR

Introduction

With the increase in demand for fish as a source of food, there is a need for more efficient and sustainable production systems. DNA marker technology is being efficiently integrated in aquaculture production systems for pedigree information, as tagging individual fish is tedious and it also helps to avoid interference of environmental effects which result from rearing families in separate tanks (Martinez, 2007). Conventional selective breeding programmes have been successful for traits that have high heritability but accuracy of predicting breeding values for traits with low heritability such as disease resistance or carcass quality is low. Thus, breeding programs for traits with low heritability are going to benefit from genome wide selection strategy (Sonesson, 2011). The prerequisite to implement genomic selection include sufficient number of DNA markers in target species. Though an array of DNA markers are being used in fish genetic studies, microsatellite markers are increasingly preferred due to their high polymorphic information content (PIC), co-dominant mode of inheritance, abundance and broad distribution throughout the genome (Liu and Cordes, 2004). Though rohu contribute significantly to the total freshwater aquaculture production of India, only 69 microsatellite loci are so far available from rohu (Das *et al*., 2005; Patel *et al*., 2009). The objective of the present work was to develop and characterise large numbers of microsatellite markers within a short span of time to increase the genomic database of rohu. The conventional microsatellite-enriched library screening method was compared with the advantages offered by next generation sequencing technology.

Materials and methods

Genomic DNA of *L. rohita* was isolated from fresh fin tissue samples using standard phenol-chloroform method (Sambrook and Russell, 2001). The protocol for microsatellite enriched library preparation and screening developed by Bloor *et al*. (2001) was followed with some minor modifications. The genomic DNA (20 µg) was digested in a volume of 150 µl at 37 °C with 20 units of *Sau*3AI (New England Biolabs, USA) in supplier recommended buffer for 3h and 30 min. After digestion, the restriction enzyme was inactivated by incubating the reaction for 20 min at 65 °C. Size fractions of 300-500 bp and 500-800 bp were recovered from a 1% agarose gel using a Qiaquick gel-extraction column (Qiagen Inc., Valencia, CA, USA). For adapter preparation, equal volumes of each oligonucleotide (200 pmol μl^{-1}) (oligo A: 5'-GCGGTACCCGGGAAGCTTGG-3' and oligo B: 5' P-GATCCCAAGCTTCCCGGGTACCGC- 3', where p indicates phosphorylation) were mixed, denatured at 80 °C for 5 min and incubated at room temperature for 1 h to allow them to anneal.

Ligation of adapters with inserts were performed in a total volume of 150 µl containing 50 pmol adapter, 4 µg size selected DNA, 15 µl 10x ligation buffer (New England Biolabs, USA) and 45 Weiss units of T4 DNA ligase (New England Biolabs, USA). The reactions were performed in a thermal cycler (Heto, Denmark) at 16 °C overnight. To confirm that the adapters had ligated to the size selected DNA, PCR was performed in a volume of 25 µl containing 1 µl ligated DNA, 1x *Taq* DNA polymerase reaction buffer, 200 µM dNTPs, 1µl oligo A (10pmol μl^{-1}) and 0.25 units *Taq* polymerase (Bangalore Genei, India) in a thermal cycler with initial denaturation for 5 min at 95 °C followed by 30 cycles at 95 °C for 50 sec, 1 min at 56 °C, 2 min at 72 °C, and a final extension of 10 min at 72 °C. Excess adapter was removed using QIAquick PCR Purification Kit (Qiagen Inc., Valencia, CA, USA).

Hybridization was performed in a total volume of 50 µl containing 3 µg adapter-ligated genomic DNA, 200 pmol of 3' biotinylated $(CA)_{15}$ repeat oligonucleotide and 15 µl of 20x hybridization solution . The mixture was denatured for 10 min at 95 °C and then incubated at 65 °C for 30 min. The hybridization reaction was mixed with 40 µg of streptavidin coated magnetic beads (Invitrogen Dynal AS, Norway). Before use, the Dynabeads were washed according to the instructions provided by the supplier. The mixture was incubated at 68 °C for 2 h with continuous agitation. Beads were washed four times each in 2x SSC and 1x SSC at room temperature. After washing, 25 µl bead suspension and 250 µl of 1x SSC was taken in a eppendorf vial and incubated at 65°C for 10 min. Immediately, supernatant was removed and rinsed with 400 µl of 1x TE followed by removal of the supernatant and washing with 400 µl of 50 mM NaCl. To adjust the beads concentration to 5 µg μl^{-1}, 50 µl of PCR grade water was added to the washed beads..

Microsatellite enriched genomic DNA was amplified in a 50 µl reaction volume containing 8 µl beads suspension, 1x *Taq* polymerase reaction buffer, 200 µM *d*NTPs, 3 µl oligo A (10 pmol μl^{-1}), 0.25 units *Taq* polymerase (Bangalore Genei, India) in a GeneAmp 9700 thermo-cycler (Applied Biosystems, USA) in two steps as follows: initial denaturation at 95 °C (3 min) followed by 9 cycles at 95 °C (30 sec), 65 °C (30 sec), 72 °C (45 sec) then second step : 30 sec at 92 °C followed by 23 cycles at 65 °C (30 sec), 72 °C (55 sec) and a final extension for 30 min at 72 °C. PCR products were purified using the Qiagen PCR purification kit (Qiagen Inc., Valencia, CA, USA). The biotinylated PCR products were ligated to pGEM-T Easy Vector (Promega, USA) at 4 °C overnight according to the instructions provided by supplier. The ligated products were used to transform *Escherichia coli* DH5α competent cells by electroporation, which were then plated on LB plates (containing 50 mg ml^{-1} ampicillin, IPTG and X-gal) at 37 °C to promote selective growth of transformed colonies. All positives (white) colonies were used for sequencing in ABI 3730 automated DNA sequencer (Applied Biosystems, USA).

Electropherograms of all the sequences were checked manually to examine base calling quality. After screening of good sequences, repeat motifs were identified using Web Sat program (Martins *et al*., 2009), setting the value of repeat finding option as di-6, tri-4, tetra-3, penta-3 and hexa-3. All microsatellite containing clones were subjected to redundancy check by DNASTAR and vector/adaptor sequences were chopped off manually. Contigs were generated by SeqMan module of DNAStar (DNAStar, Inc. USA). Each contig sequence was blasted to avoid redundancy with any rohu SSR loci deposited in GenBank earlier. Primers were designed from unique sequences using Batchprimer3 (You *et al*., 2008).

Next generation sequencing (NGS)

Partial genome of rohu was sequenced using GS-FLX 454 Titanium platform (Ecogenics GmBH, Switzerland). Microsatellites were isolated following the protocol adapted from Dubut *et al*. (2010). The selection of sequences for primer design was done with the program QDD (Meglecz *et al*., 2010). Adaptors were removed from sequences. Sequences longer than 80 bp and containing perfect microsatellite motif of at least six repetitions for any microsatellite motif of 2–6 bp were selected for further analyses. Sequence similarities were computed along with the previous sequences present in the database through an 'all against all' BLAST (Altschul *et al*., 1997) in which microsatellite motifs were soft masked. Sequences with more than 95% pairwise similarity along the flanking regions were grouped into contigs. Sequences which had significant BLAST hit with other sequences and an overall similarity among the flanking region below 95% were discarded to avoid potential intragenomic multicopy sequences. All unique sequences (with no BLAST hit to any other '454' reads) and consensus sequences were checked for the presence of short repetitions in the flanking regions. PCR primers were designed using QDD with the following criteria: (a) target microsatellite displayed at least five repetitions, (b) length of PCR product ranging from 80 to 300 bp and (c) annealing temperatures of primer pairs were optimised for falling between 55 °C and 65 °C.

Results and discussion

One enriched (CA repeat) library was prepared for screening microsatellites. Altogether, 1840 recombinant clones were sequenced by Sanger method. Out of this, 122 (6.63%) clones could not be sequenced due to GC reach or long repeat motif that may have caused secondary structure

formation and dislocation of polymerase (Franca *et al*., 2002). Sequence electropherograms of remaining clones (1718) were examined to check quality of base calling and possible manual modification of wrong base. Out of 1718 clones, 1150 (66.93%) clones were microsatellite containing positives, which after redundancy check (872 clones; 50.75% were redundant) produced only 278 (16.18%) unique sequences. BLAST result showed that all sequences were non-identical to that of rohu SSR deposited in NCBI. Out of 278 unique sequences, 238 had enough flanking region for primer design and sequences were deposited in GenBank (Accession number JN581132 to JN581366). Some clones had more than one repeat with enough flanking region so that two primer pairs could be designed from such clones (30 nos.). A total of 272 primer pairs were synthesised at the end of conventional enriched-library screening.

NGS methodology generated 3677 contigs that had enough microsatellite repeats but 3290 contigs did not have enough flanking regions for designing primers. Finally, 387 microsatellite containing sequences were used for primer designing and a total of 255 primers (GenBank accession number JQ862039 to JQ862292) were designed so as to have a product size ranging from 98 bp to 350 bp.

Various microsatellite development protocols based on partial genomic library enriched for microsatellite are available in the literature with varying number of microsatellite markers successfully developed. Abdelkrim *et al*. (2009) developed 37 pairs of primers (19.47%) after enriched library screening and sequencing of 190 positive clones in blue duck (*Hymenolaimus malacorhynchos*). In the present study, 272 primers (15.10%) were successfully optimised which is less than the results reported by Abdelkrim *et al*. (2009). This may be due to the presence of higher percentages of redundant clones in the enriched library.

The sequencing of 10 Mb of rohu genome by 454 platforms, lead to the development of 255 microsatellite markers. Compared to the traditional development of microsatellites from a partial genomic library, costs were reduced to about one third and considerable time was saved. Major part of expenditure in traditional method of microsatellite isolation was the cost of sequencing (Sanger method) of large numbers of clones containing genomic DNA insert. The results of the present study clearly indicate that the genomic shotgun sequencing approach using high throughput technology is an effective, economical and fast alternative to traditional protocol for microsatellite isolation.

Acknowledgments

The authors are thankful to the Department of Biotechnology, Govt. of India, New Delhi, for financial help and to the Director, CIFA, Bhubaneswar for providing laboratory facilities to carry out the work.

References

Abdelkrim, J., Robertson, B. C., Stanton, J. A. L. and Gemmell, N. J. 2009. Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. *Biotechniques,* 46: 185-192.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.,* 25: 3389-3402.

Bloor, P. A., Barker, F. S., Watts, P. C., Noyes, H. A. and Kemp, S. J. 2001. *Microsatellite libraries by enrichment.* Protocol available at: http://www. genomics. liv. ac. uk/animal/ research/microsatellite.

Das, P., Barat, A., Meher, P. K., Ray, P. P. and Majumdar, D. 2005. Isolation and characterisation of polymorphic microsatellites in *Labeo rohita* and their cross species amplification in related species. *Mol. Ecol. Notes*, 5: 231-233.

Dubut, V., Grenier, R., Meglecz, E., Chappaz, R., Costedoat, C., Danancher, D., Descloux, S., Malausa, T., Martin, J. F. and Pech, N. 2010. Development of 55 novel polymorphic microsatellite loci for the critically endangered *Zingel asper* (Actinopterygii: Perciformes: Percidae) and cross-species amplification in five other percids. *Eur. J. Wildl. Res.*, 56: 931-938.

Franca, L. T. C., Carrilho, E. and Kist, T. B. L. 2002. A review of DNA sequencing techniques. *Q. Rev. Biophys.,* 35: 169-200.

Liu, Z. J. and Cordes, J. F. 2004. DNA marker technologies and their applications in aquaculture genetics. *Aquaculture*, 238: 1-37.

Martinez, V. 2007. Marker-assisted selection in fish and shellfish breeding schemes. In: Guimaraes E. P. (Ed.), *Marker-assisted selection - Current status and future perspectives in crops, livestock, forestry and fish*. FAO, Rome, p. 329-362.

Martins, W. S., Lucas, D. C. S., de Souza Neves, K. F. and Bertioli, D. J. 2009. WebSat - a web software for microsatellite marker development. *Bioinformation*, 3: 282 pp.

Meglecz, E., Costedoat, C., Dubut, V., Gilles, A., Malausa, T., Pech, N. and Martin, J. F. 2010. QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics*, 26: 403-404.

Patel, A., Das, P., Swain, S. K., Meher, P. K., Jayasankar, P. and Sarangi, N. 2009. Development of 21 new microsatellite markers in *Labeo rohita* (rohu). *Anim. Genet*., 40: 253-254.

Sambrook, J. and Russell, D. W. 2001. *Molecular cloning: a laboratory manual*, 3rd edn., Cold Spring Harbor Laboratory Press, New York, U.S.A.

Sonesson, A. K. 2011. *Genomic selection for aquaculture: principles and procedures. Next generation sequencing and whole genome selection in aquaculture.* Wiley-Blackwell, Oxford, UK. DOI: 10. 1002/9780470958964.ch10.

You, F., Huo, N., Gu, Y., Luo, M., Ma, Y., Hane, D., Lazo, G., Dvorak, J. and Anderson, O. 2008. BatchPrimer3: a high throughput web application for PCR and sequencing primer design. *BMC bioinformatics*, 9: 253.

Date of Receipt : 16.02.2012
Date of Acceptance : 10.08.2012