International Journal of Science, Environment and Technology, Vol. 5, No 1, 2016, 229 – 233

ISSN 2278-3687 (O)
2277-663X (P)

NGS APPROACH TO MICROSATELLITE MARKER ISOLATION: AN OVERVIEW

Ilaya Bharathi D.*, Saxena V.K. and Raj Manohar G.*
*College of Poultry Production and Management, Mathigiri, Hosur
Tamilnadu Veterinary and Animal Sciences University

Abstract: Microsatellite markers also known as the Simple sequence repeat (SSR) or Short tandem repeat (STR) they are stretches of DNA, consisting of tandem repeating mono-, di-, tri-, tetra- and penta- nucleotide units, which are arranged throughout the genome of most eukaryotic species. Various DNA markers such as Minisatellites, DNA fingerprinting, Random Amplified Polymorphic DNA markers (RAPD) and microsatellites have been used for estimating genetic variation in chicken lines. Among them microsatellite markers are less laborious and more efficient to estimate genetic variation than other DNA markers.
Keywords: Microsatellite markers, Next generation sequencing technology (NGST).

INTRODUCTION

Microsatellites or simple sequence repeats (SSRs) are tandemly repeated motifs of 1 to 6 bases found in all prokaryotic and eukaryotic genomes present in both coding and noncoding regions usually characterized by a high degree of length polymorphism.

Microsatellites have been classified according to the type of repeat sequence as perfect or imperfect. A common characteristic of imperfect repeats is that there is no more equivalency between fragment length and amplicon sequence: several sequences can correspond to a given length variant. Using microsatellite markers population structures and differences, genetic drift, genetic bottlenecks and even the date of a last common ancestor can be predicted based on microsatellite polymorphism (Moxon E R and Wills C., 1999).

Microsatellite markers belong to a class of highly mutable genomic sequences known as variable number of tandem repeat (VNTR) elements that show extensive levels of intra-specific polymorphisms in both eukaryotic and prokaryotic genomes. The minimum number of repeat motifs used to flag a sequence as containing a microsatellite was 12 repeats for mononucleotide motifs and 5 repeats for the remaining repeat classes (di-, tri-, tetra-, penta-, and hexanucleotides). Attributes of microsatellites as genetic markers as quoted by Mittal N and Dubey AK (2009) are:

- Locus-specific in nature; in contrast to multi-locus markers such as minisatellites or

*Received Jan 20, 2016 * Published Feb 2, 2016 * www.ijset.net*

Random Amplified Polymorphic DNA (RAPDs).

- Co-dominant transmission and therefore the heterozygotes can be distinguished from homozygotes, in contrast to Random amplified polymorphic DNA (RAPD); Amplified fragment length polymorphism (AFLP) which are binary in nature.
- Highly polymorphic and hyper-variable.
- High information content and provides considerable pattern.
- Relative abundance with uniform genome coverage.
- Higher mutation rate than standard sequences (up to 0.001gamates/generation).
- High probability of back mutation.

CONVENTIONAL METHODS

Conventional microsatellite isolation involves screening several thousands of clones through colony hybridization with repeat containing probes (Rassmann *et al.* 1991). The number of positive clones that can be isolated by means of this method ranges from 12% to less than 0.04%. Conventional methods are less useful when dealing with taxa with a very low frequency of microsatellites such as birds or when a large number of microsatellites is required when constructing a genetic map (Cooper *et al* 1999). Under such scenarios, the Simple Hybrid Capture Method is the most popular choice because it does not require any prior sequence data. The repetitive sequences are captured using a biotinylated probe and selectively attached to magnetic beads. After several washing steps, the captured DNA fragments are eluted, amplified, and cloned to produce a clone library enriched for the target sequences which can be sequenced and PCR primers can be developed and validated (Zane *et al.,* 2002).

The most cost-effective microsatellite-development method is to use publicly available genetic/genomic information on the species of interest. Microsatellites derived from publicly available sequences may be genomic and gene microsatellites. Genomic are those STR markers present at the non-coding sequences (introns or intergenic spaces) and are more polymorphic and selectively neutral whereas Gene microsatellites are from coding sequences mainly obtained from EST (Expressed Sequence Tag) analysis (Blair *et al.,* 2003: Ellison *et al.,* 2010). Microsatellites from very closely related species can also be utilized for studies on the species of interest. Inoue-Murayama *et al.* (2001) tested chicken microsatellite markers to see if they would be suitable as genetic linkage markers Japanese quail.

Conventional laboratory methods involving cloning, cDNA library construction and Sanger sequencing remain costly and time-consuming regardless of whether genomic or EST

sequences are used for SSR detection. Currently, next-generation sequencing techniques have become popularly used to identify sequences harboring SSR motifs in non-model species. Improvements in methods utilizing biotin-based enrichment in SSR motifs, are also published in combination with next-generation sequencing. Next-generation sequencing also provides information on SSR polymorphism, in particular if more than one genotype is sequenced.

NGS APPROACHES TO SSRS DISCOVERY AND DEVELOPMENT

Next generation sequencing (NGS) allows massive data collection from DNA sequencing at relatively low-cost with high performance. NGSTs require highly sophisticated pre-sequencing and post sequencing (Hui, 2012). The pre-sequencing phase involves two main steps: target enrichment and library constriction. Target DNA enrichment can be acquired either by amplification and sequencing or via hybridization and selection methods.

NGS library preparation can be accomplished using the following procedures: (1) the enriched DNA is fragmented by physical methods into preferable lengths of (150-500) base pairs, (2) small DNA fragments are ligated to adaptor primers, and (3) clonal amplification of the NGS library (by emulsion bead PCR or cluster amplification). Lengths of reads are typically 100 bases in length and each read contains a single sequence clone. The billions of reads can then be assembled as longer sequences referred to as contigs can be screened to identify SSR motifs using relevant bioinformatics tools (Hui 2012). Several recent studies have demonstrated the efficient use of NGSTs for large-scale discovery of SSR loci in non-model species (Zalapa et al. 2012; Franchini, Van der Merwe and Roodt-Wilding 2011; Hou et al. 2011; Lim et al. 2012). Short reads compared with the Sanger approach, abundant repeat regions which are difficult to assemble, sequencing errors and artifacts arising from homopolymers are common limitations with NGSTs. Also construction of DNA library and computation difficulties attributed to the fast rate of generating raw data are callenges to be addressed (Mardis 2008; Shendure & Ji, 2008; Stein 2010).

The major obstacle in using microsatellite markers is the need for isolation of a very high number of microsatellites in order to identify loci that are informative as well as easy to screen. The reason for this is that the degree of length heterogeneity, and therefore the informative value of a particular microsatellite, seems to depend strongly on the number of repeats and whether the repeat is perfect or degenerated. In general, moderately to high levels of allelic variation can be expected only when the microsatellite contains more than 15 repeat units. In particular, for organisms in which hypervariable markers are not as ubiquitous as in mammals ***(e.g.,*** avians), obtaining informative microsatellites may be a time-consuming

process. Hybridisation capture of microsatellite rich sequence is a rapid approach for isolation of microsatellites and other tandem repeated sequences as described by Unn Hilde Refseth *et al.,* 1997 in Atlantic salmon and by Gibbs *et al.,*(1997) in chicken.

Unn Hilde Refseth *et al.,*(1997) reported 35-fold enrichment of AC/GT microsatellites using an $(AC)_{10}$, oligo probe. In addition, four out of five sequences captured by a $(CAG)_{10}$, oligo probe contained one or several CAG repeat arrays. Gibbs *et al.,* 1997 reported that the total number of microsatellite loci is considered to be at least 10-fold lower in avian species than in mammalian species. Therefore, efficient large-scale cloning of chicken microsatellites, as required for the construction of a high-resolution linkage map, is facilitated by the construction of libraries using an enrichment strategy. In their study, a plasmid library enriched for tandem repeats was constructed from chicken genomic DNA by hybridization selection. Using this technique the proportion of recombinant clones that cross-hybridized to probes containing simple tandem repeats was raised to 16%, compared with < 0·1% in a non-enriched library.

REFERENCES

[1] Blair M.W., F. Pedraza, H.F. Buendia, E.Gaitan-Solıs, S.E. Beebe (2003). Development of genome-wide anchored microsatellite map for common bean (Phaseolus vulgaris L.). *Theor. Appl. Genet.* 107: 1362–1374.

[2] Cooper G, Amos W, Bellamy R *et al.* (1999). An empirical exploration of the $(dm)^2$ genetic distance for 213 human microsatellite markers. *American Jour. of Human Genetics,* **65**, 1125-1133.

[3] Ellison CK, Shaw KL (2010). Mining non-model genomic libraries for microsatellites: BAC versus EST libraries and the generation of allelic richness. *BMC Genomics* 11: 428.

[4] Franchini, P., Van der Merwe, M. andRoodt-Wilding R. 2011. Transcriptome characterization of the South African abalone *Haliotismidae* using sequencing-by-synthesis.BMC Research Notes.**4** (59): 1-11.

[5] Gibbs M, D A Dawson, C McCamley, A F Wardle, J A L Armour, T Burke (1997). Chicken microsatellite markers isolated from libraries enriched for simple tandem repeats. *Animal Genetics,* 28, 401–417, 401.

[6] Hou, R., Bao, Z., Wang, S., Su, H., Li, Y., Du, H., Hu, J. and X. Hu. 2011. Transcriptome sequencing and de novo analysis for yesso scallop (*Patinopectenyessoensis*) using 454 GS FLX. PloS One, **6** (6): 21560.

[7] Hui, P. 2012. Next generation sequencing: chemistry, technology and applications. Top.Curr. Chem. **336:** 1-18.

[8] Inoue-Murayama, M., Kayang, B.B., Kimura, K., Ide, H., Nomura, A., Takahashi, H.,Nagamine, Y., Takeda, T., Hanada, H., Tatsuda, K., Tsudzuki, M., Matsuda, Y.,Mizutani, M., Murayamam, Y. and Ito, S. 2001. Chicken microsatellite primers are not efficient markers for Japanese quail. Anim. Genet. **32**: 7-11.

[9] Lim, K.G., Kwoh, C.K., Hsu L.Y. and Wirawan A. 2012. Review of tandem repeat search tools: a systematic approach to evaluating algorithmic performance. Brief Bio. Inform, **14**(1): 67-81.

[10] Mardis, E.R. 2008. Next-generation DNA sequencing methods. Ann. Rev. Genom.Human Genet.**9**: 387-402.

[11] Mittal, N. and Dubey, A.K. 2009. Microsatellite markers- A new practice of DNA basedmarkers in molecular genetics. Pogo Rev. **3**: 235-46.

[12] Moxon, E.R. and Wills, C. 1999. DNA microsatellites: Agents of Evolution?.Scientific American, 72-77.

[13] Rassmann. K., Schlotterer, C. and Tautz, D. 1991. Isolation of simplesequence loci for use in polymerase chain reaction-based DNA fingerprinting.Electrophoresis, **12**: 113-118.

[14] Shendure, J. and Ji, H. 2008. Next-generation DNA sequencing. Nature Biotech. **26**(10): 1135-1145.

[15] Stein, L.D. 2010. The case for cloud computing in genome informatics. Genome Biol. **11** (5):207.

[16] Unn Hilde Refseth, Bjern Magne Fangan and Kjetill Sigurd Jakobsen, 1997. Hybridization capture of microsatellites directly from genomic DNA. Electrophoresis, **18:**1519-1523.

[17] Zalapa, J.E., Cuevas, H., Zhu, H., Steffan, S., Senalik, D., Zeldin, E., McCown, B., Harbut, R. and Simon, P. 2012. Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. Am. J Bot.**99** (2): 193-208.

[18] Zane L, Bargelloni,L. and Patarnello, T. 2002. Strategies for microsatellite isolation: a review. Mol. Ecol.**11**: 1- 16.