DOUBLE DIGEST RESTRICTION ASSOCIATED DNA SEQUENCING FOR IDENTIFICATION OF SINGLE NUCLEOTIDE POLYMORPHISMS IN GOATS

Shalu Elizabeth Simon[1], G.Radhika[2], T.V. Aravindakshan[3], Marykutty Thomas[4] and K. Raji[5]
Department of Animal Genetics and Breeding, College of Veterinary and Animal Sciences, Mannuthy, Thrissur- 680 651

Received: 01.09.2018
Accepted:18.09.2018

Abstract

Goat rearing is considered as an Animal husbandry activity which promptly supplements economy of the rural poor. India, with its vast animal genetic resources is blessed with 28 goat breeds which vary in different quantitative traits. Malabari and Attappady Black are two native goat breeds of Kerala, which differ significantly in several economic traits like prolificacy, milk production, milk quality and disease resistance. In this study, a novel reduced representation sequencing technique called double digest restriction associated DNA sequencing (ddRADseq) is applied on Malabari and Attappady Black goats, for identification of genetic variants which can be used for exploring many complex economical traits and to serve as markers for selection. DNA samples were isolated from 10 animals of each breed and were pooled into two samples to represent Malabari and Attappady Black goats. These two samples were digested using SphI and MlucI Restriction enzymes which resulted in 26,77,648 reads in Malabari and 17,40,350 reads in Attappady Black goats. A total of 66299, 28859 and 4266 variants were identified in Malabari in read depths of 2, 5 and 10, respectively. In case of Attappady Black, it was 61241, 21230 and 2251. Variants in read depth 10, were considered as high confidence variants. In Read depth 10, Malabari had 4140 SNPs which included 14 missense SNPs, 24 synonymous SNPs, 58 SNPs in untranslated regions and 4044 SNPs in noncoding regions. In Attappady Black groups, 29 missense SNPs, 30 synonymous SNPs, 82 SNPs in untranslated region and 2051 SNPs in non coding regions were present in 2192 high confidence SNPs discovered.

Keywords: *double digest restriction associated DNA sequencing, Prolificacy, Malabari and Attappady Black goats*

Malabari and Attappady Black are two native goat breeds of Kerala, which differ significantly in several economic traits like prolificacy, milk production, milk quality and disease resistance. Exploring the genetic basis

**Part of MVSc thesis of first author submitted to Kerala Veterinary and Animal Sciences University, Pookode , Wayanad ,Kerala*

1. M.V.Sc. Scholar

2, 4 Assistant Professor

3 Professor and Head

5 Assistant Professor, Department of Physiology, CVAS, Mannuthy

of these differences provide us with a scientific background which could further be utilised for improvement of these traits. DNA sequencing is considered as the main key for opening all those valuable unveiled information in genomes by finding out the orderly arrangement of the four bases namely adenine, guanine, thymine and cytosine. The sequencing costs are decreasing day by day with the advancement of technology but collection, management and storage of whole genome sequencing data of large number of individuals still remain unaffordable for institutions. In such a situation, reduced representation sequencing and its modifications are presumed to be highly demanding, since they survey on a large set of unlinked loci which are representative of the entire genome.

Peterson *et al*. (2012) introduced ddRADseq as a modification of existing protocols of Genome Wide Sampling Sequencing (GWSS) techniques. The protocol was developed by modifying Restriction associated DNA (RAD) sequencing which is considered as one of the successful and cost-effective reduced representation sequencing techniques. ddRADseq protocol was developed by avoiding the random shearing step in RAD sequencing and introducing the use of two restriction enzymes. Since then ddRADseq served as a highly efficient and low cost technology for high density SNP discovery and genotyping, both in model and non -model organisms.

In this study, ddRADseq is applied to highly prolific Malabari and low prolific Attappady Black goats for identification of SNPs which can be used for further studies to explore many complex economical traits and to serve as markers for selection.

Materials and methods

Isolation of DNA from whole blood

DNA was isolated from six ml blood of 10 Malabari and 10 Attappady Black goats from University Goat and Sheep Farm, Mannuthy using standard Phenol chloroform method (Sambrook and Russell, 2001). Isolated 10 DNA samples from each breed were pooled to represent a single sample from each breed. These two samples were subjected to ddRADseq.

Library preparation work flow

Double digestion of genomic DNA (1µg) was done using *SphI* (rare cutter) and *MlucI (*common cutter) restriction enzymes at 37ºC for 16 hrs. The digested fragments were cleaned using Ampure beads (Agencourt AMPure). These cleaned fragments were ligated with two adapters A and B which are complementary to the restriction sites of the two enzymes *SphI* and *MlucI,* respectively. Besides the complementary sequences, A adapter had an additional 10bp sequence which was TATAGCCT for Malabari and TAATCTTA in Attappady Black group. Ligation of products to A (Barcoded) and B adaptors was done using T4 DNA ligase. Resulting fragments were pooled followed by size selection. Manual size selection was done in which these products were subjected to gel electrophoresis in 2 per cent gels at 85 V for 150 min. A volume of 80 µL of the pooled library was placed in lanes, and DNA in the region of 100–150 bp (based on comparison to a 1-Kbp DNA ladder) was excised from the gel and purified with QiaQuick gelextraction kits (Qiagen).

The resulting libraries were amplified with primers that introduced index sequence of ATTACTCG in Malabari and TCCGGAGA in Attappady Black groups which were read in a separate multiplexing read, as per the standard Illumina multiplexed paired-end sequencing protocol. The quality and concentration of libraries was assessed with a NanoDrop spectrophotometer and quantitative electrophoresis in a Bioanalyzer. The products were again pooled and subjected to Illumina sequencing. A workflow of ddRADseq is given in Fig.1

Bioinformatics Analysis

The samples were demultiplexed to obtain reads for each group and the low quality bases and regions showing base bias at the start or end, were removed from the reads. The Illumina 5′ and 3′ adapter sequences were removed from the reads for final analysis and the processed reads were then aligned to the reference goat genome downloaded from NCBI.

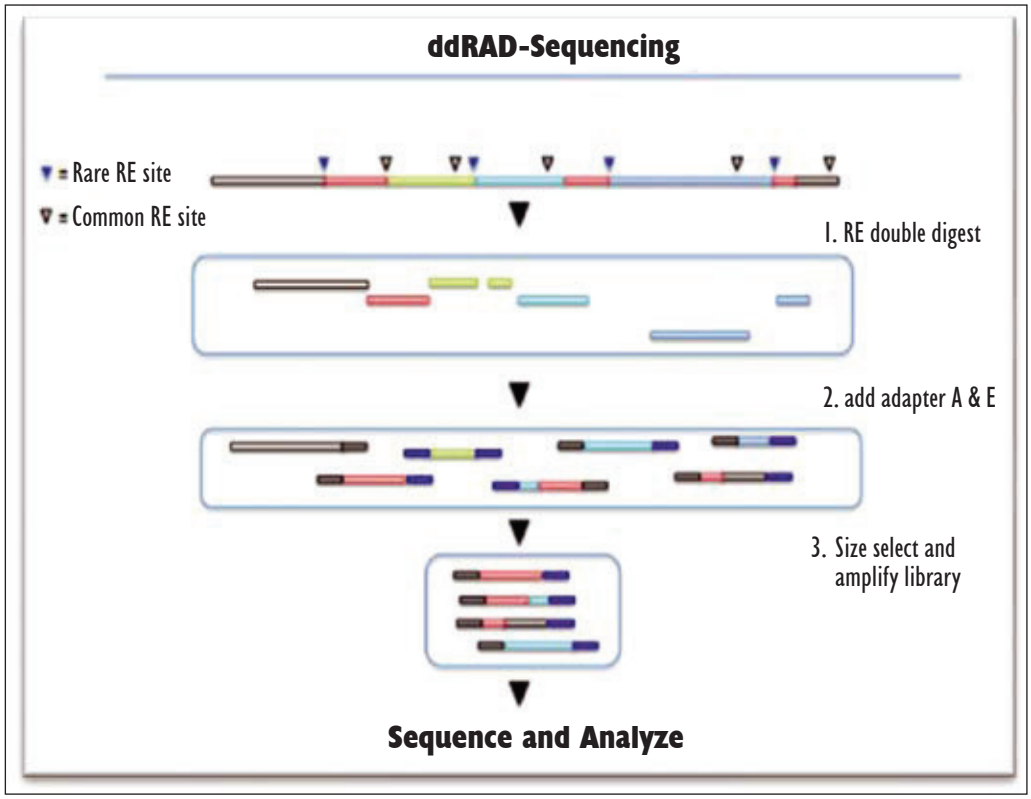

Fig. 1 ddRADseq library
(Peterson *et al*., 2012)

Paired-end alignment was performed using Bowtie2 (Langmead *et al*., 2009) programme with default parameters. Aligned samples and the reference genome sequence were used for variant calling by Samtools programme (Li *et al*., 2009).

Results and discussion

Digestion of the whole genome with restriction enzymes *SphI* and *MlucI,* adapter ligation and size selection (100-150bp) followed by Illumina sequencing produced 26,77,648 reads in Malabari and 17,40,350 reads in Attappady Black. The present study was first of its kind to perform ddRADseq in goat genome and the successful generation of optimum reads for SNP discovery in the present study proved the utility of the enzymes *SphI* and *MlucI* for reduction of complexity of goat genome followed by successful SNP discovery. A summary of reads obtained is given in table 1 and 2. The two enzyme digestion for reduced representation sequencing implied the use of a rare cutter and a common cutter for fragmentation of the genome (Poland *et al*., 2012). Peterson *et al*. (2012) assured that the use of two restriction enzymes resulted in five-fold reduction in the library construction cost and also skipped the random shearing step which reduced unnecessary DNA loss. Selection of enzymes and their combinations needed utmost care considering various factors like the structure and size of genome, coverage needed, purpose of experiment etc. (Mora-Marquez *et al*., 2017). Kess *et al.* (2015) opined that well designed experimental plan is necessary for the success of ddRAD experiments so an *insilico* analysis of the fragments produced by enzymes was done with NEB cutter before the start of work.

Size selection was considered as another valuable step which made the difference between Genotyping by sequencing (GBS) and ddRADseq. Peterson *et al*. (2012) stated that size selection window and restriction enzyme combination are two parameters which affected SNP discovery. Size selection window

varied according to the genome and purpose of work. Kai *et al*. (2014) used a range similar to current study in Japanese eel for construction of genetic map, while Kess *et al*. (2015) used a wider size selection window 150-500 bp, for parentage analysis in *Littorina saxatilis.* However a smaller size selection window used in this study provided us with numerous high confidence SNPs which proved its success for discovery of SNPs in goat genome with ddRADseq. Selection of specified fragment increased the accuracy, but caused a reduction in potential for marker discovery (Yang *et al.,* 2010).

Among total reads, those which had two RAD tags were 26,48,770 in Malabari and 17,09,750 reads in Attappady Black groups. Only these reads were selected for further downstream analysis. The stringent selection of reads reduced the number of reads generated for SNP discovery relatively lesser than other reduced representation techniques. Kai *et al*. (2004) applied ddRADseq in Japanese eel and generated 6.7 million reads from which only around two lakh reads were selected for further sequencing. However, they opined that the disadvantage of lower number of reads produced, was outlooked by the robustness and efficiency of SNPs discovered across groups.

Among 26,48,770 reads in Malabari, 25,30,368 reads were aligned to the goat reference genome, whereas 16,69,056 reads from 17,40,350 reads were aligned for Attappady Black group. Alignment per cent was 95.53 in Malabari and 97.62 per cent in Attappady Black goats. Among total reads aligned, 98.77 per cent reads of Malabari and 98.52 per cent reads of Attappady Black were aligned to unique regions in the genome. This showed the efficiency of both ddRADseq and the tool Bowtie which was used for alignment of the reads Details of alignment data are given in Table 3.

Variant calling was done using Samtools version 0.1.18. Variants were identified from mainly three read depth namely read depth 2, 5 and 10. A summary of SNPs and INDELs discovered in three read depths are summarised in Table 4

Although all the variants discovered are useful markers, SNPs in read depth 10 is considered as high confidence SNPs. In read depth 10, ddRADseq identified 4140 SNPs and

126 INDELS in Malabari and 2192 SNPs and 59 INDELS in Attappady Black goats. These high confidence SNPs included

Table 1. Reads summary of ddRADseq library in Malabari and Attappady Black goats

Group	Total reads	Number of bases (Mb)	GC content (%)
Malabari	2677648	267.76	46.88
Attappady Black	1740350	174.04	47.26

Table 2. RAD tag summary of ddRADseq in Malabari and Attappady Black goats

Groups	Total Reads	Reads with two RAD tags	Reads with R1 RAD tags	Reads with R2 RAD tags	Reads with no RAD tags
Malabari	2677648	2648770	6876	7511	104
Attappady Black	1740350	1709750	6562	8574	328

Table 3. Alignment data summary of total reads obtained by ddRADseq to goat reference genome

Group	RAW Data	Passed Data	Alignment Data	Alignment percent	Uniquely Aligned Read	% of Uniquely Alignment
Malabari	2677648	2648768	2530368	95.53	2499272	98.77
Attappady Black	1740350	1709748	1669056	97.62	1644411	98.52