www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277- 7695
ISSN (P): 2349-8242
NAAS Rating: 5.03
TPI 2018; 7(10): 131-135

www.thepharmajournal.com
Received: 23-08-2018
Accepted: 24-09-2018

Anu Bosewell
Junior Research Fellow, Department of Veterinary Microbiology, College of Veterinary and Animal Sciences, Mannuthy, Kerala, India

TV Aravindakshan
Professor and Head, Centre for Advanced Studies in Animal Genetics and Breeding, College of Veterinary and Animal Sciences, Mannuthy, Kerala, India

Radhika G
Assistant Professor, Centre for Advanced Studies in Animal Genetics and Breeding, College of Veterinary and Animal Sciences, Mannuthy, Kerala, India

Jinty Sukumaran
Research Assistant, Centre for Advanced Studies in Animal Genetics and Breeding, College of Veterinary and Animal Sciences, Mannuthy, Kerala, India

Correspondence
Anu Bosewell
Junior Research Fellow, Department of Veterinary Microbiology, College of Veterinary and Animal Sciences, Mannuthy, Kerala, India

Whole exome sequencing analysis of native dwarf cattle genetic groups of Kerala by next generation DNA sequencing

Anu Bosewell, TV Aravindakshan, Radhika G and Jinty Sukumaran

Abstract
Indigenous cattle play an important role in the sustenance and welfare of traditional rural farmers. The indicine cattle genetic groups of Kerala, (the southernmost state in India) namely Kasargod dwarf and Vadakara dwarf cattle are believed to possess numerous attributes such as very short stature, disease resistance, heat tolerance, less susceptibility to ticks and lower feed consumption. A detailed genetic study regarding the variations would reveal unique features responsible for its rare phenotypes. The aim of the present study was to perform the whole exome analysis of these native cattle of Kerala using Bovine All Exon capture kit (Agilent Sure Select) followed by sequencing on Illumina HiSeq 2500 platform. We successfully identified a total of 1578008 and 1425816 variants including 1453933 (92.13%) and 1314506 (92.19%) Single Nucleotide Polymorphisms (SNPs) and 124075 (7.86%) and 111310 (7.80%) In Dels (Insertions/Deletions) in Kasargod dwarf and Vadakara dwarf cattle, respectively. Out of the total variants 1245431 and 1134579 were found to be novel in Kasargod dwarf and Vadakara dwarf cattle, respectively. Detailed variant annotation revealed that the majority of the SNPs were located in the intergenic region of the genome. Furthermore the exonic variants were distributed as 63002 and 57516 non synonymous SNPs, 37773 and 35021 missense mutations, 324 and 323 nonsense mutations and 73 and 64 start loss mutations in Kasargod dwarf and Vadakara dwarf cattle, respectively. These findings are particularly interesting because presence of these SNPs might be the probable reason for the peculiar characteristics of these indicine cattle and therefore efforts have to be made to conserve and develop these indigenous cattle resources.

Keywords: Whole exome sequencing, Kasargod dwarf, Vadakara dwarf, Single Nucleotide Polymorphism, Insertion/Deletions

Introduction
India is a vast subcontinent with varied agro climatic conditions and is rich in its animal genetic resources including cattle. Kerala, the southernmost state in India, has a variety of dwarf cattle, referred to as the indigenous cattle of Kerala. In addition to Vechur (only recognised cattle breed from Kerala), there are several genetic groups of cattle in Kerala belonging to different terrains which show unique phenotypic characteristics. These lesser known varieties evolved over the millennia and got adapted to specific agro climatic conditions prevalent in the state. The indigenous Kasargod dwarf originated in the mountain range of Kasargod district and has been considered as small cattle with comparatively better milk production and is well adapted to the hot and humid climate of the region with heat tolerance and disease resistance (figure 1). The rare indigenous Vadakara Dwarf cattle originated from the tribal areas of Chimmini forest, measuring 95- 110 cm in height. A special feature of the cattle is that it is generally hornless and is said to produce milk with medicinal value and withstands high environment temperature (figure 2). Currently, these genetic groups are declining because of extensive crossbreeding practised in the state. Therefore efforts have to be made to study the genome of these livestock for sustainable conservation and management.

Fig 1: Kasargod dwarf

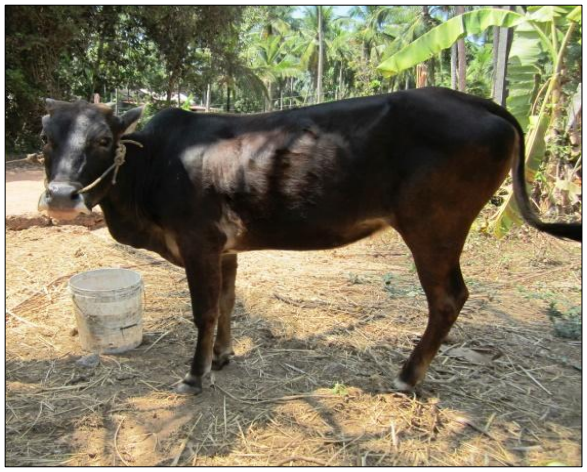

Fig 2: Vadakara dwarf

One of the first mammalian genome sequenced was of bovine because these farm animals are serving as major nutritional sources for human (Lee *et al*., 2013) [10]. After the establishment of bovine reference genome (Elsik *et al*., 2009) [7] a number of cattle breeds were sequenced – Black Angus and Holstein (Stothard *et al*., 2011) [18], Hanwoo, Yanbian (Choi *et al*., 2014; Choi *et al*., 2015) [2, 4], Jeju Hengu and Korean Holstein cattle (Choi *et al*., 2014) [2] and Fleckvieh and Braunvich (Schwarzenbacher *et al*., 2016) [17] leading to the identification of genomic variants. Characterisation of genetic variations in indigenous breeds will be a significant step towards deciphering the mechanism underlying the peculiar traits in those animals.

Recent advances in sequencing technology have enabled to unravel the unique genomic features of ecologically important organism. Discovery and practical use of next generation sequencing (NGS) techniques enabled a drastic increase in availability of sequence data at a significantly low cost (Lee *et al*., 2013) [10]. Although with the emergence of NGS the sequencing cost and sequencing time decreased by several folds, the sequencing of eukaryotic genome is still not affordable. So a more affordable way is to sequence the protein coding regions (exons) which comprises only one to two per cent of the eukaryotic genome depending upon the species (War *et al*., 2015) [19]. Any change in this region will have enormous impact on the phenotype. With the advent of techniques for exon enrichment, it is now possible to sequence the exons genome wide. The genome sequencing studies performed in different cattle breeds mainly focused on identification of breed specific variants (Lee *et al*., 2016) [11]. Therefore the objective of the present study was to perform the whole exome sequencing of the Kasargod dwarf and Vadakara dwarf cattle and the results revealed 1578008 and 1425816 variants of which 1245431 and 1134579 were novel variants in the exome of Kasargod dwarf and Vadakara dwarf cattle, respectively.

Materials and Method

DNA extraction: Genomic DNA was extracted from the whole blood of Kasargod dwarf and Vadakara dwarf cattle by Phenol chloroform method. The quality of DNA were evaluated by Nano Drop TM 1000 spectrophotometer (Thermo Scientific, USA) according to the manufacturer's instruction. The quality of DNA was visually checked on 0.8% agarose gel electrophoresis.

Whole exome sequencing: The exonic regions of Kasargod dwarf and Vadakara dwarf cattle genome were captured using All Exon Bovine (Agilent Sure Select – 54 Mb) capture kit. Briefly, the purified DNA of concentration more than 50ng/µl was randomly sheared and end repaired. The sheared DNA was subjected to hybridization with All Exon Bovine (Agilent Sure Select – 54 Mb SSXT Bovine All Exon. Catalog No.519-5448) capture kit. After hybridization, the hybridized regions were separated using streptavidin-conjugated paramagnetic beads and the unhybridized regions were discarded. Sequences enriched with exonic regions were then recovered by washing the beads and digesting the probes. The fragments were clonally amplified and sequenced using sequencing by synthesis chemistry on Illumina Hi Seq 2500 platform.

Mapping short reads, identification of SNPs and Indels and their annotation: For each sample, sequence reads were removed if the Phred quality score was less than 30. The remaining reads were mapped against the reference *Bos taurus* genome downloaded from Ensembelftp: //ftp.ensembl.org/pub/release81/gtf/bos_taurus/Bos_taurus.UMD3.1.81.gtf.gz using Burrows – Wheeler Aligner (BWA) Version-0.7.5) programme with default parameters. After mapping, Picard tool (Version-1.100) Sort Sam command was used for sorting alignments. Picard Mark Duplicates was used to exclude duplicate reads. The calling of variants including SNPs and short In Dels were performed using Samtools (Version-0.1.18). The variants were subjected to additional filtering in order to retain good quality variants. Further, after variant calling a detailed functional annotation was carried out.

Results and Discussion

Sequencing and read alignment: The whole exome of Kasargod dwarf and Vadakara dwarf cattle were sequenced using the Illumina Hi Seq 2500 platform. A total of 55946902 and 54023480 reads were obtained in Kasargod dwarf and Vadakara dwarf, respectively, of which 99.9% reads passed the quality control (QC). These resulting reads were mapped to the bovine reference genome UMD3.1 and yielded 2.7 GB of raw sequence data. Approximately, 55065544 and 53130515 of the reads from Kasargod dwarf and Vadakara dwarf respectively, were aligned accurately with reference genome. The raw data quality summary of Kasargod dwarf and Vadakara dwarf are provided in Table 1.

Table 1: Raw read and alignment summary of Kasargod dwarf cattle and Vadakara dwarf cattle

Sample	Read orientation	Raw reads (paired end)	Bases (Gb)	GC%	Total reads	QC passed reads	Aligned read count	Properly paired
Kasargod dwarf	R1	27973451	2.79	44.1	55946902	55946268 (99.99%)	55065544 (98.43%)	96.83%
	R2	27973451	2.79	43.7				
Vadakara dwarf	R1	27011740	2.7	44.01	54023480	54022898 (99.99%)	53130515 (98.35%)	96.44%
	R2	27011740	2.7	43.74				

Li *et al*. (2008) [12] reported that a minimum of 20-30 fold coverage is necessary for all variant detection and a fairly high coverage were obtained for our data. There are several factors such as GC content, sequencing technology, read length, library preparation, structural variants and novel sequences which influence the assembly coverage (Liao *et al*., 2013) [13]. A previous study reported that the coverage obtained by using the Illumina Hiseq 2000 platform was notably high (Choi *et al*., 2014) [2]. Chromosome wise coverage and depth of Kasargod dwarf and Vadakara dwarf cattle are given in Table 2.

Table 2: Chromosome wise coverage and depth of Kasargod dwarf cattle and Vadakara dwarf cattle

Chromosome	Kasargod dwarf		Vadakara dwarf	
	Coverage	Depth	Coverage	Depth
1	20.47	1.75	20.15	1.72
2	22.06	2.02	21.6	1.97
3	24.44	2.35	23.87	2.26
4	21.3	1.79	20.78	1.74
5	24.03	2.22	23.46	2.14
6	20.17	1.74	19.87	1.72
7	23.84	2.11	23.16	2.02
8	22.25	1.78	21.8	1.73
9	19.92	1.63	19.57	1.59
10	23.69	2.3	23.04	2.23
11	25.39	2.16	24.71	2.05
12	19.94	1.52	19.56	1.48
13	26.4	2.01	25.6	1.93
14	22.42	1.69	21.71	1.63
15	23.62	2.11	23.12	2.05
16	24.88	2	24.25	1.92
17	23.66	1.85	23.08	1.77
18	30.85	2.74	29.89	2.59
19	32.85	3.17	31.95	3
20	21.39	1.66	20.92	1.62
21	24.83	1.83	24.15	1.75
22	25.96	2.14	25.08	2.01
23	25.99	2.18	25.2	2.07
24	22.53	1.59	21.87	1.52
25	33.39	2.76	32.31	2.59
26	24.71	2.1	24.12	2.02
27	22.56	1.7	22.08	1.65
28	23.86	1.92	23.26	1.85
29	27.16	2.18	26.2	2.05
X	19.94	1.4	19.76	1.39

SNP and InDel Identification: A total of 1578008 and 1425816 variants were identified throughout the genomes of Kasargod dwarf and Vadakara dwarf cattle, respectively using Samtools. Among them, 1245431 and 1134579 were found in db SNP while the remaining 332577 and 291237 were novel (Table 3). The homozygous to heterozygous ratios of the total SNPs in Kasargod and Vadakara dwarf were 1044295:409638 and 935680:378826, respectively.

Table 3: Statistics of db SNP filtered variants in Kasargod dwarf cattle and Vadakara dwarf cattle

Samples	Kasargod dwarf	Vadakara dwarf
With db SNP	1245431	1134579
Without db SNP	332577	291237
Total	1578008	1425816

Eck *et al*. (2009) [6] opined that the low proportion of heterozygous SNPs is mainly due to the stringent SNP calling requirements. The transition to transversion (TS/TV) ratio were also computed to evaluate the SNP quality and are the indicators of possible random sequence errors (Kraus *et al*., 2011) [9]. The TS/TV ratios for Kasargod dwarf and for Vadakara dwarf were 2.60 and 2.59 respectively. The ratio were higher than the previous reports (Choi *et al*., 2014; Choi *et al*., 2013; Kawahara-Miki *et al*., 2011; Stothard *et al*., 2011) [2, 3, 8, 18], indicating high quality of SNPs obtained.

A total 124075 and 111310 of In Dels were identified in Kasargod dwarf and Vadakara dwarf cattle, respectively. Variants identified in Kasargod dwarf and Vadakara dwarf cattle were summarized and provided in Table 4.

Table 4: Variant calling summary of Kasargod dwarf cattle and Vadakara dwarf cattle

Sample Name	Kasargod dwarf	Vadakara dwarf
Total variants	1578008	1425816
Total SNPs	1453933 (92.13%)	1314506 (92.19%)
Total indels	124075 (7.86%)	111310 (7.80%)
Total homozygous SNP	1044295 (71.82%)	935680 (71.18%)
Total heterozygous SNP	409638 (28.17%)	378826 (28.81%)
Total transition type SNP	1050156 (72.22%)	948750 (72.17%)
Total transversion type SNP	403777 (27.77%)	365756 (27.69%)
T_S/T_V	2.60	2.59

From a total of 1578008 variants (1453933 SNPs and 124073 InDels) identified 761463 SNPs and 84076 InDels were having a Q score ≥ 20 in Kasargod dwarf while in Vadakara dwarf from a total of 1425816 variants (1314506 SNPs and 111310 InDels) 690160 SNPs and 75420 InDels were having a Q score ≥20. When depth was analysed 555541 SNPs and 50972 InDels in Ksargod dwarf and 507796 SNPs and 45083 InDels in Vadakara dwarf were with a depth > 5. Bodi *et al.* (2013) reported that the high quality variants had a depth of ≥5 and a Q score ≥20. The quality score and read depth of all identified variants are provided in Table 5 and 6.

Table 5: Quality score and read depth distribution of identified variants in Kasargod dwarf cattle

Quality score	Total SNPs	Total indels	Read depth	Total SNPs	Total indels
0-10	381890	22327	0-5	898392	73103
10-20	310580	17672	5-10	149300	14176
20-30	103551	22,798	10-15	80,236	7,443
30-40	111,937	7,773	15-20	56,064	5,209
40-50	91,934	6,331	20-25	42,738	3,803
50-60	47,231	4,880	25-30	33,782	3,045
60-70	47,207	4,192	≥30	82,323	7,219
70-80	39,830	4,584	≥50	111,076	10,073
80-90	31,706	4,696	-	-	-
90-100	28,692	2,460	-	-	-
≥100	259,353	26,358	-	-	-
Total	1,453,933	124,075	-	1,453,933	124,075

Table 6: Quality score and read depth distribution of identified variants in Vadakara dwarf cattle

Quality score	Total SNPs	Total indels	Read depth	Total SNPs	Total indels
0-10	345,948	20,114	0-5	806,710	66,227
10-20	278,398	15,776	5-10	138,232	12,899
20-30	93,650	20,684	10-15	74,352	6,806
30-40	100,464	6,770	15-20	52,003	4,737
40-50	83,662	5,594	20-25	38,908	3,336
50-60	43,581	4,414	25-30	30,831	2,684
60-70	42,683	3,743	≥30	74,418	6,246
70-80	35,835	4,233	≥50	99,045	8,373
80-90	29,056	4,147	-	-	-
90-100	26,260	2,059	-	-	-
≥100	234,962	23,774	-	-	-
Total	1,314,506	111,310	-	1,314,506	111,310

Variant annotation: Understanding the genomic variation is of primary interest and simplest form of these variations is SNPs. In total 107,172 and 98,143 exonic SNPs were identified including 91,574 and 83,942 in coding region; 2,388 and 2,347 in non-coding region; 426 and 412 in non-coding RNA; 12,784 and 11,442 in untranslated region (UTR) in the genome of Kasargod dwarf and Vadakara dwarf, respectively. Out of the 12,784 and 11,442 SNPs in the UTR, 2,990 and 2,720 were in 5'UTR; 9,794 and 8,722 in 3'UTR of Kasargod dwarf and Vadakara dwarf, respectively. 283 and 259 SNPs in 5'splice site; 417 and 350 in 3' splice site were also identified in Kasargod dwarf and Vadakara dwarf, respectively.

In total 3,373 and 3,135 exonic InDels were found including 1,819 and 1,683 in coding region; 150 and 145 in non-coding region; 38 and 26 in non-coding RNA; 1,366 and 1,281 in untranslated region (UTR) in the genome of Kasargod dwarf and Vadakara dwarf, respectively. Out of the 1,366 and 1,281 SNPs in the UTR, 263 and 247 were in 5'UTR; 1,103 and 1,034 in 3'UTR of Kasargod dwarf and Vadakara dwarf, respectively. 141 and 132 SNPs in 5'splice site; 140 and 138 in 3' splice site were also identified in Kasargod dwarf and Vadakara dwarf, respectively.

Following functional annotation of variants 63,002 and 57,516 non synonymous SNPs in Kasargod and Vadakara dwarf, respectively were identified. The synonymous SNPs does not produce an amino acid change, but may affect the protein folding and function (Parmley and Hurst, 2007) [15]. Of the total variants, 37,773 and 35,021 were predicted to be missense mutation in Kasargod and Vadakara dwarf, respectively. Only 1350 and 1234 of the variants were predicted to have frame shift mutation in Kasargod and Vadakara dwarf, respectively. A 325 and 323 non sense mutations were found in Kasargod and Vadakara dwarf, respectively. The functional annotation also revealed 73 and 64 start loss mutation in Kasargod and Vadakara dwarf, respectively. One of the powerful ways of evaluating the effects of these SNPs in coding region is by focusing on the fraction that alters the encoded amino acid. (Nakken *et al*., 2007) [7].

In indigenous breeds, most of the genetic characteristics are under study. So this information might be useful for understanding and improving them. Therefore detection of genetic variation that contributes to those peculiar characters is important. The genetic variation, particularly the SNPs that result in change in translated amino acid residue is likely to play a major role in functional diversity of coded proteins (Dabhi and Mistry, 2014) [5].

Conclusion

In the present study NGS was successfully employed in identifying SNPs and InDels in Kasargod and Vadakara dwarf cattle. A total of 1,578,008 and 1,425,816 variants were identified in Kasargod dwarf and Vadakara dwarf cattle, respectively. A total of 332,577 and 291,237 variants were novel in Kasargod and Vadakara dwarf cattle, respectively. In depth annotation of SNPs identified 63,002 and 57,516 non synonymous SNPs, 37,773 and 35,021 missense mutations, 324 and 323 nonsense mutation and 73 and 64 start loss mutation in Kasargod dwarf and Vadakara dwarf cattle, respectively. These SNPs and InDels are the genetic parameters which accounts for the unique phenotypic traits in those indigenous breeds. Thus, the results of the present study will help in development of candidate variants of functional significance after validation of these results in larger population.

Acknowledgement

We are thankful to the Kerala Veterinary and Animal Sciences University for providing the facilities for the conduct

of research. We thank SciGenom Labs Pvt Ltd, Kakkanad, Kerala for the sequencing and bioinformatics services.

References

1. Bodi K, Perera AG, Adams PS, Bintzler D, Dewar K, Grove DS, *et al*. Comparison of Commercially Available Target Enrichment Methods for Next-Generation Sequencing. Journal of Bimolecular Techniques. 2013; 24:73-86.
2. Choi JW, Choi BH, Lee SH, Lee SS, Kim HC, Yu D *et al.* Whole-Genome Resequencing Analysis of Hanwoo and Yanbian Cattle to Identify Genome-Wide SNPs and Signatures of Selection. Molecules and Cells. 2015; 38(5):466-73.
3. Choi JW, Chung WH, Lee KT, Choi JW, Jung KS, Cho Y *et al*. Whole Genome Resequencing of Heugu (Korean Black Cattle) for the Genome-Wide SNP Discovery. Korean Journal of Food Science and Animal Resources. 2013; 33:715-22.
4. Choi JW, Liao X, Stothard P, Chung WH, Jeon HJ, Miller SP. Whole-Genome Analyses of Korean Native and Holstein Cattle Breeds by Massively Parallel Sequencing. PLoS ONE, 2014; doi:10. 1371/journal. pone.0101127.
5. Dabhi B, Mistry KN. *In silico* analysis of single nucleotide polymorphism in human TNF-α gene. Meta Gene. 2014; 2:586-95.
6. Eck SH, Benet-Pagès A, Flisikowski K, Meitinger T, Fries R, Strom TM. Whole genome sequencing of a single Bos taurus animal for single nucleotide polymorphism discovery. Genome biology, 2009, doi: 10. 1186/gb-2009-10-8-r82.
7. Elsik CG, Tellam RL, Worley KC. The Genome Sequence of Taurine Cattle: A window to ruminant biology and evolution. Science. 2009; 324(5926):522-28.
8. Kawahara-Miki R, Tsuda K, Shiwa Y, Arai-Kichise Y, Matsumoto T, Kanesaki Y, *et al.* Whole-genome resequencing shows numerous genes with nonsynonymous SNPs in the Japanese native cattle Kuchinoshima-Ushi. BMC Genomics. 2011; 12:103-08.
9. Kraus RHS, Kerstens HHD, Hooft PV, Crooijmans RPMA, Poe JJVD, Elmberg J, *et al*. Genome wide SNP discovery, analysis and evaluation in mallard (Anas platyrhynchos). BMC Genomics. 2011; 12:150-54.
10. Lee CY, Chiu YC, Wang LB, Kuo YL, Chuang EY, Lai LC, *et al*. Common applications of next-generation sequencing technologies in genomic research. Translational Cancer Research, 2013, doi: 10. 3978/j.issn.2218-676X.2013.02.09.
11. Lee D, Cho M, Hong W, Lim D, Kim H, Cho Y, *et al*. Evolutionary analyses of Hanwoo (Korean Cattle)-specific –single nucleotide polymorphisms and geens using whole –genome resequencing data of a Hanwoo population. Molecules and Cells. 2016; 39(9):692-98.
12. Li H, Ruan J, Durbin R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Research. 2008; 18:1851-58.
13. Liao X, Peng F, Forni S, McLaren D, Plstow G, Stothard P. Whole genome sequencing of Gir cattle for identifying polymorphisms and loci under selection. Genome. 2013; 56:592-98.
14. Nakken S, Alseth I, Rognes T. Computational prediction of the effects of non-synonymous single nucleotide polymorphisms in human DNA repair genes. Neuroscience. 2007; 145:1273-1279.
15. Parmley JL, Hurst LD. How do synonymous mutations affect fitness? Bio Essays. 2007; 29:515-19.
16. Sambrook J, Russell DW. Molecular cloning: A laboratory manual. Third edition. Cold spring Harbor Laboratory Press. New York, 1886.
17. Schwarzenbacher H, Burgstaller J, Seefried FR, Wurmser C, Hilbe M, Jung S. A missense mutation in TUBD1 is associated with high juvenile mortality in Braunvieh and Fleckvieh cattle. BMC Genomics. 2016; 17:400-05.
18. Stothard P, Choi J, Basu U, Sumner-Thomson JM, Meng Y, Liao X, *et al*. Whole genome resequencing of Black Angus and Holstein cattle for SNP and CNV discovery. BMC Genomics. 2011; 12:559-62.
19. Warr A, Robert C, Hume D, Archibald A, Deeb N, Watson M. Exome Sequencing: Current and Future Perspectives. Genes Genomes Genetics. 2015; 5:1543-49.