Indian J Dairy Sci 77(1): 76-83
https://doi.org/10.33785/IJDS.2024.v77i01.010

RESEARCH ARTICLE

Genetic blueprinting of novel and performance traits-related SNPs in Indian Gir cattle using latest reference assembly

Nidhi Sukhija[1], Anjali Choudhary[1], Kanaka KK[1], M Joel Devadasan[1], Jayakumar Sivalingam[2#] and Archana Verma[1](✉)

Received: 05 September 2023 / Accepted: 11 October 2023 / Published online: 23 February 2024

Abstract: Among the indicine breeds of cattle, Gir breed thrives in challenging arid habitats, produce more milk with less feed intake and are resistant to tropical diseases. Identification of novel and bias free SNPs are important to include in genetic studies and breeding programs. The current study aimed to explore SNPs in Gir cattle genome followed by annotation of the SNPs to different performance traits. Double-digest restriction site associated DNA sequencing (ddRAD-seq) was done and different bioinformatic tools and software used to mine SNPs and subsequent annotation to different performance traits. A total of 53,243 high quality SNPs were identified using the latest ARS-UCD 1.3 reference genome (GCF_002263795.2). In all, 2,931 SNPs were annotated to 452 candidate genes related to nine different traits of interest such as, milk production, reproduction, adaptation, disease-resistance, growth, carcass, coat thickness, coat colour, and domestication traits. ddRAD-seq offers mining of ascertainment free SNPs in a cost-effective way and hence the SNPs obtained in this studied population can be successfully used for population genetic studies and in breeding programs. The information mined in this study will aid in genetic improvement, designing appropriate breed improvement programmes and leveraging conservation efforts.

Keywords: ddRAD-seq, Gir cattle, Performance traits, SNPs

[1]ICAR-National Dairy Research Institute, Karnal, Haryana, India.

[2]ICAR-National Bureau of Animal Genetics and Resources, Karnal, Haryana, India.

presently at ICAR-Directorate of Poultry Research, Hyderabad 500030, Telangana, India

Archana Verma (✉)
Animal Genetics and Breeding Division, ICAR-National Dairy Research Institute, Karnal 132001, Haryana, India
Email: archana.ndri@gmail.com

Introduction

Agriculture is the basic substance of the Indian rural economy with livestock rearing as its lifeline. In India 53 zebu cattle breeds have been registered, including the major milch purpose breeds *viz*., Sahiwal, Gir, Red Sindhi and Tharparkar cattle breeds (www.nbagr.res.in). The Gir cattle breed is native to Gir forests in Kathiawar region of Gujarat (Gaur et al. 2003). Temperature in their habitat ranges from 7-45°C with an average rainfall of 500-1,500 mm (IMD, 2023). Morphologically, they have leaf-like and the longest ears, sleepy eye appearance, most convex head, largest hump and widest coat colour variation among all Indian cattle breeds (AGRI-IS, 2023). Gir cattle are known for high milk production as compared to other Indian breeds, robustness to stressors as well as tropical diseases and have ~95% A2 allele frequency (Sushil et al. 2018). Gir cattle have acquired cross-country importance shaping Gyr, Girolando and Indubrasil breeds. However, Gir cattle have a lactation yield of about 2,110 kg in India while the one bred overseas, in Brazil has a lactation yield of 3,500 kg (Madalena et al. 2012). Stagnation in productivity and shrinking population size are the main concerns and call for genetic interventions to conserve and bridge the demand-supply gap vis-à-vis growing human population. However, cattle genetics and allied technologies such as SNP arrays and SNP databases developed so far show ascertainment bias (Stothard et al. 2011; Koks et al. 2014; Iqbal et al. 2019) and moreover, exclude discovery of breed-specific or rare variants (Sivalingam et al. 2020; De Donato et al. 2013). Ascertainment bias free markers aid in breed improvement programmes and leveraging conservation efforts. (Jeevan et al. 2019; Kanaka et al. 2023).

Whole Genome Sequencing (WGS) overcomes ascertainment bias but is a costly venture and computationally demanding. Restriction enzyme-based DNA sequencing (RAD-seq) can account for a large number of genetic markers in both referenced and non-referenced organisms and covers up to 40% of the genome (Mishra et al. 2020). Hence, we used double-digest restriction site associated DNA sequencing (ddRAD-seq) (Peterson et al. 2012) which employs rare as well as frequent cutters in this study to omit repetitive and uninformative sequences. Erstwhile, genome-wide identification of SNPs using RAD-seq in seven cattle breeds (Malik et al. 2018), Sahiwal

(Vineeth et al. 2020), Red Sindhi (Iqbal et al. 2019), Vrindavani (Wara et al. 2019) and Tharparkar (Devadasan et al. 2020) cattle breeds have been exercised in the Indian context. However, there is dearth in literature on genome-wide SNPs annotated to candidate genes for different traits in Gir cattle. This catalyzed our study to discover and thereafter, annotate genome-wide SNPs using ddRAD approach with genes related to performance traits in Gir cattle.

Materials and methods

Ethics Statement

The collection of blood samples from animals in this research project was conducted with the highest regard for animal welfare and ethical principles. The procedures were approved in 43rd Institutional Animal Ethics Committee meeting, in compliance with established guidelines and regulations (43-IAEC-18-9).

Blood collection and DNA isolation

Blood samples, 10 mL each, from seven adult unrelated Gir cows from the Livestock Research Centre at the National Dairy Research Institute in Karnal, Haryana, India (29.704°N Latitude, 76.982°E Longitude) were collected in vacutainer tubes coated with 0.5% EDTA (Ethylene Diamine Tetra Acetic Acid). The genomic DNA was isolated using the phenol-chloroform method as described by Sambrook and Russell (2006).

Library Preparation

The standard RAD protocol (Peterson et al. 2012) was employed for the further sequencing of the DNA after initial evaluation of quality and quantity. The DNA was subjected to double digestion using SphI and MluCI restriction enzymes for constructing the library and thereafter, custom sequencing was done in SciGenom Labs Pvt. Ltd, Cochin, Kerala using Illumina HiSeq 2000 platform.

Bioinformatics analysis

SNPs calling

SNP identification was carried out using the standard procedure (S1). The raw sequencing reads in FastQ format were screened using FastQC *v0.11.9* (Andrews et al. 2010), then adapter sequences and barcodes were removed from the sequence reads using PRINSEQ *v0.20.4* (Schmieder et al. 2011). STACKS *v2.61* (Catchen et al. 2011) was utilized to remove sequences that lacked the restriction enzyme cut site and had a Phred score below 15. The QC-approved reads were then aligned with the Bos taurus (ARS-UCD 1.3) reference genome (RefSeq accession: GCF_002263795.2) using *local very sensitive* mode using Bowtie2 *v2.5.0* (Langmead and Salzberg, 2012). The alignment resulted in SAM (Sequence Alignment Format) files, which were then converted to BAM (Binary Alignment Format) files, sorted, indexed, and merged to produce a single BCF (Binary Call Format) file using BCFtools *v1.16 (Li,* 2011*).* BCF file was converted to VCF (Variant Call Format) equivalent and genome-wide SNPs and InDels were further detected at read depth (RD) thresholds of e" 2, 5, and 10 with a mapping quality (MQ) e" 30 using VCFtoolsv0.1.16 (Danecek et al. *2011).* SNPs were separated from InDels and only those obtained from the aligned reads with e" RD 10 were subjected to further quality control to obtain high-quality SNPs, such as Hardy-Weinberg equilibrium (0.001), Minor Allele Frequency (0.01), and Missing Genotypes (1.0), using VCFtools *v0.1.16* (Danecek et al. 2011).

Annotation

The high-quality SNPs were then annotated using SnpEff *v5.1* (Cingolani et al. 2012). These SNPs located within genes were investigated in the literature to find genes that are linked to performance traits. Graphical maps were generated using MG2C (Chao et al. 2021*)* to show the placement of SNPs on various chromosomes in relation to different traits. Mutations that were potentially harmful with a SIFT (Sorting Intolerant from Tolerant) score of d" 0.05 were recognized and annotated to the relevant protein-coding genes using Variant Effect Predictor, VEP *v108.2* (McLaren et al. 2016). Additionally, new SNPs were discovered using VEP *v108.2.*

Results and Discussion

The objective of this study was genome-wide SNP identification and annotation of SNPs to performance traits in seven Indian Gir cattle using ddRAD-seq. Previous works supporting similar sample size (n=7) include Malik *et al.* (2018) and Devadasan *et al.* (2020). On an average, 1.95 million raw reads were obtained per sample. In all, 13.67 million raw reads were sequenced (Table 1). After de-multiplexing, adapter trimming and quality control of raw reads, 98.18% good quality reads were retained (Table 1). The quality reads were aligned with the latest ARS-UCD 1.3 *Bos* taurus reference genome, with an overall alignment rate of 99.85% (Table 1). The number of sequences in two samples were comparatively lower than the other samples (Table 1) due to differences in read depth because of poor DNA quality similar to Malik *et al.* (2018). STACKS was used for trimming of low quality reads as it checks the mean quality score using sliding windows while PRINSEQ trims across mean values (Surya et al. 2019) (Figure 1).

Overall, 254,128 genome-wide variants, inclusive of 211,560 SNPs were mapped to ARS-UCD 1.3, being the latest and representative assembly. Out of them, a total of 53,243 high quality SNPs were sieved with RD e" 10 and MQ e" 30 for downstream analysis (Figure 2, Table 1, S2). The number of SNPs found are in line with Altmann *et al.* (2012). The number of high quality SNPs is greater than reported 9,638 in Vrindavani crossbred cattle (Wara et al. 2019) and 18,056 in Murrah buffaloes (Jaglan et al. 2023) using

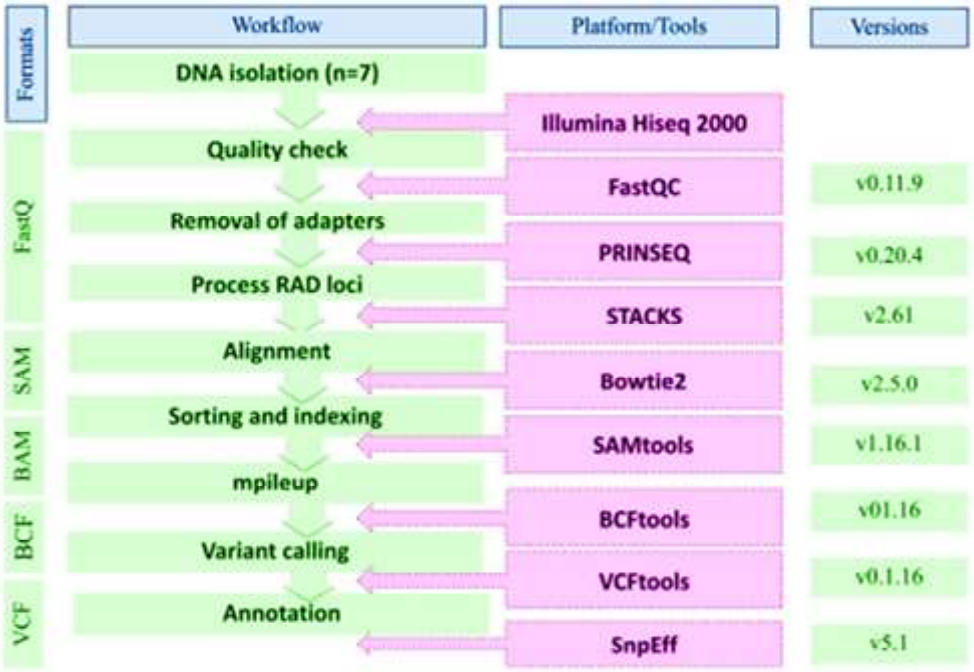

Fig. 1 Schematic representation of steps and tools followed in this annotation

Table 1: Number of raw, processed and aligned reads of Gir Cattle. AR - Alignment Rate, UAR - Unique Alignment Rate. Reads retained = 249,139; Reads removed = 1.82%; Reads retained = 98.18%

No.	Pre-processing of reads			*Bos taurus* (ARS-UCD 1.3)			
	ID	Raw reads	QC-passed reads	No. of aligned reads	Total aligned reads (in %)	AR (%)	UAR (%)
1	G-1	765,970	754,342	743,780	98.60	99.85	17.42
2	G-2	3,310,806	3,223,269	3,180,160	98.66	99.77	21.55
3	G-3	3,305,652	3,249,045	3,197,532	98.41	99.92	21.82
4	G-4	1,283,522	1,264,553	1,246,962	98.61	99.89	23.03
5	G-5	840,808	829,157	818,436	98.71	99.72	22.45
6	G-6	541,836	533,368	525,862	98.59	99.95	22.42
7	G-7	3,626,672	3,572,393	3,536,918	99.01	99.84	22.24
Average		1,953,609	1,918,018	1,892,807	99	99.85	22

the same QC parameters except missing genotypes > 0.8 and > 0.65. The higher number of SNPs despite stringent filtering implies higher polymorphism in Gir cattle. In concordance with Sivalingam et al. (2020) and Kumar et al. (2020), G/A and C/T substitutions were higher than other base changes. Frequent C > T mutations may be due to the deamination of methyl cytosines in CpG dinucleotides (Jaglan et al. 2023). Evidently, transitions were 2.6' more likely than transversions in line with targeted sequencing approaches (Keller et al. 2007; Patel et al. 2017; Ba et al. 2017; Kraus et al. 2011). Average inter-SNP distance in Gir was 49 kb, which was similar to 45 Kb in Indian cattle breeds identified by genotyping by synthesis method (Malik et al. 2018). Maximum numbers of annotated SNPs were discovered in the transcript region, followed by intron and intergenic regions (Table 2), which was consistent with prior reports on SNPs annotation in riverine buffalo (Surya et al. 2019) and Tharparkar cattle (Devadasan et al. 2020). Based on effects by functional class, SNPs within the coding regions were classified as missense (347), silent (627) and nonsense (5).

In contrast to the existing variants, 3,824 (7.2%) SNPs were reported as novel (S). A total of 27 missense variants having SIFT scores d" 0.05 were adjudged deleterious. The deleterious SNPs were annotated to 21 protein-coding genes (S2). On an average, Ts/Tv (Transition and Transversion) ratio of the annotated SNPs was 2.7603 per sample (Table 3). The frequency of G/A and C/T transition substitutions were markedly higher than other substitution types (Figure 3). Sizeable differences were seen in between the number of SNPs in different traits, with maximum number of genes (n=176) and SNPs (n=1,191) traced for the traits related to milk quality and quantity, are in line with Mishra et al. (2020). The deleterious SNPs enumerated (n=27 SNPs) are more than those found in Sahiwal cattle, i.e., 18

Table 2: Region-wise distribution of SNPs

Region	Count	Percent
Downstream	13,962	5.80
Exon	1,361	0.57
Intergenic	29,168	12.12
Intron	90,249	37.51
Splice_site_acceptor	1	0.00
Splice_site_region	123	0.05
Transcript	90,984	37.82
Upstream	13,427	5.58
UTR_3_prime	1,104	0.46
UTR_5_prime	219	0.09

Table 3: Nucleotide base changes in identified SNPs

Nucleotide Change	Count
Transitions	207,345
Transversions	75,116
Ts/Tv ratio	2.7603

deleterious SNPs (Surati et al. 2023), but less than those reported by Kour et al. (2022) in Indian yak (n=166 SNPs). This is suggestive of lesser burden of deleterious variants in Gir cattle due to adaptive selection over many generations. A total of 3,824 novel SNPs were identified, justifying the usage of ddRAD-seq over SNP chips. The novel variants are much lower than 10,519 SNPs reported by Devadasan et al. (2020) may be due to stringent QC applied in this study.

A total of 2,931 SNPs were annotated to 452 genes belonging to different traits in the current study (Table 4, Figure 4(i and ii). Overlapping genes between different traits are given in Figure 5. Devadasan et al. (2020) reported 2,871 SNPs in 383 genes despite greater number of SNPs (n=87,047) derived without quality control. This reflects genomic richness of Gir cattle as compared to Tharparkar cattle in terms of polymorphisms. Some of the important genes annotated are discussed here. Pertinent to milk composition, SNPs were mapped in candidate genes, *COL22A1*

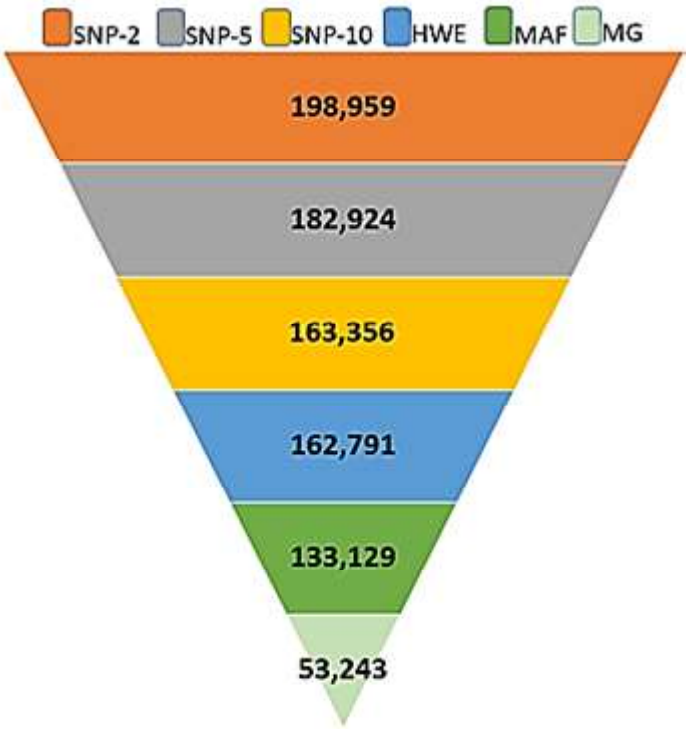

Fig. 2 SNPs retained after applying quality control. SNP-2: read depth 2, SNP-5: read depth 5, SNP-10: read depth 10, HWE-Hardy-Weinberg equilibrium, MAF- Minor Allele Frequencies, MG- Missing genotype

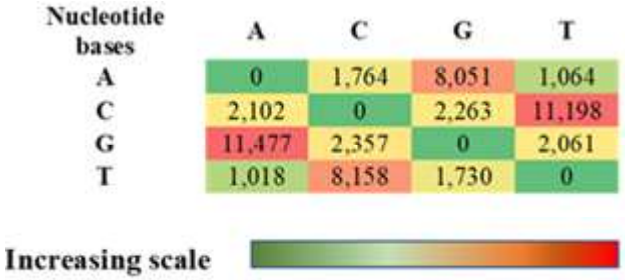

Fig.3 Nucleotide base changes in identified SNPs

Table 4: Number of polymorphisms and genes associated with different categories of traits

Traits	SNPs	Genes	Genes with the highest polymorphisms
Production	1,191	176	*TG* (n=46), *RORA* (n=38)
Reproduction	729	72	*RBFOX1* (n=153), *RBMS3* (n=36)
Disease	615	84	*PLXNA4* (n=69), *PRKG1* (n=30)
Adaptation	380	68	*CTNNA2* (n=41), *PRKCB* (n=22)
Growth	82	17	*PCNX2* (n=22), *LCORL*(n=17)
Carcass	922	101	*RBFOX1* (n=153), *GALNTL6* (n=51)
Coat colour	102	25	*LCORL* (n=17), *DOCK8* (n=9)
Coat thickness	35	10	*TCF7L1* (n=9), *CUX1* (n=9)
Domestication	183	18	*ASTN2* (n=54), *LINGO2* (n=30)
Total	4,239	571	
After removing overlaps*	2,931	452	

* Total number of SNPs and genes may not be equal to the arithmetic sum due to the presence of overlaps

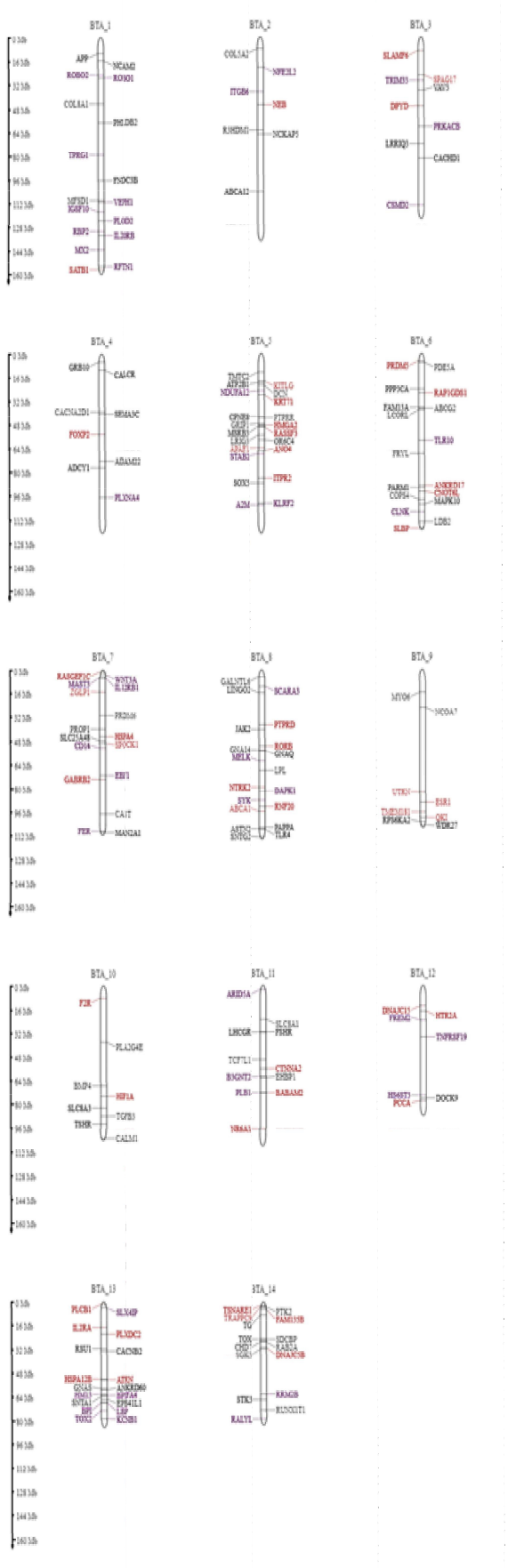

Fig. 4(i) Chromomap displaying genes and their position on respective chromosomes (BTA1 to BTA14). Black-Coat thickness; Blue- Disease; Brown- Growth; Green- Domestication; Orange-Adaptation; Pink- Coat Color; Purple- Carcass; Violet- Reproduction; Red- Production.

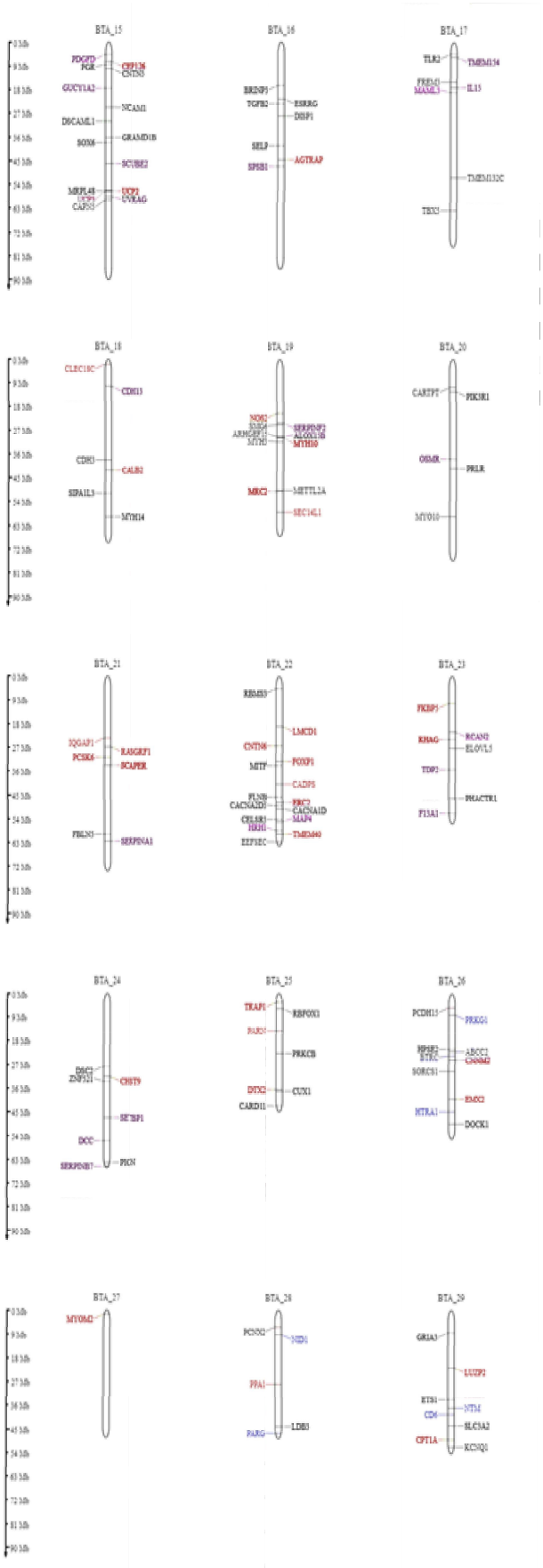

Fig. 4(ii) Chromomap displaying genes and there position on respective chromosomes (BTA 15 to BTA 29). Black-Coat thickness; Blue- Disease; Brown-Growth; Green- Domestication; Orange-Adaptation; Pink- Coat Color; Purple- Carcass; Violet- Reproduction; Red- Production.

and *PTK2* for biological pathways affecting milk fat in dairy cattle (Wang et al. 2013; Buitenhuis et al. 2014). Leptin, a peptide hormone encoded by the *LEP* gene, plays a role in growth, milk production and fertility traits (Clempson et al. 2011). *IGFBP2* gene influences fertility and milk production in cattle (Klein et al. 2006; Pimentel et al. 2011). Dairy cattle's milk fat and protein percentage are found to be influenced by nucleotide base changes in the *LPIN1* gene (Han et al. 2019). The *EPS8* gene is linked to milk output and milk fat percentage (Raven et al. 2014) as it participates in milk fat biosynthesis in the lactating mammary gland by acting through epidermal growth factor (Chatterjee et al. 2009).

The highest number of SNPs were found in *RBFOX1* gene (n=153), which regulates expression of large genetic networks during early neuronal development (Bill et al. 2013). The *TG* gene (n=46) related to production, encodes thyroglobulin protein, which serves as substrate for the production of thyroxine and triiodothyronine as well as storage form of thyroid hormone and iodine in their inactive forms (Vassart et al. 1985). *PLXNA4* (n=69) is predicted to enable semaphorin receptor activity, TLR signalling and related to sepsis induced cytokine storming (Wen et al. 2010) and is linked to Alzheimer's disease and Parkinson's disease in humans (Schulte et al. 2013; Jun et al. 2014). Another disease related gene, the *PRKG1* gene (n=30) is a crucial mediator of the nitric oxide/cGMP signalling pathway in different cells (Ørstavik et al. 1992). Highest number of polymorphisms in adaptation related genes were found in a tumor suppressor gene *CTNNA2* (n=41) (Fanjul-Fernández et al. 2013) and in a protein kinase known as *PRKCB* (n=22) has been linked to numerous physiological processes, including activation of B cells, induction of apoptosis, proliferation of endothelial cells, and absorption of intestinal sugars (Philippi et al. 2005). *ASTN* (n=54), belonging to astrotactin gene family controls the movement of *ASTN1* during the migration of glial-guided neurons (Wilson et al. 2010) and *LINGO2* (n=30), predicted to act upstream of or within positive regulation of synapse assembly (Su et al. 2010) displayed highest polymorphism related to domestication. The earlier studies support the SNPs found in the Indian Gir cattle with different performance traits.

Conclusions

This study accounts for the richly polymorphic candidate genes related to varied performance traits. The novel variants reported may provide a sizable addition to Gir genetics, which can be further included in existing SNP chips. Overall, this study offers ascertainment bias free markers which may aid in breed improvement programmes and leveraging conservation efforts. The SNPs mined in the current study may need to be validated in larger herds using association studies.

Authorship Contribution Statement

Archana Verma and Jayakumar Sivalingam conceived and designed the study. Anjali Choudhary, M. Joel Devadasan, conducted experiments. Anjali Choudhary, M. Joel Devadasan, D Ravi Kumar, Vineeth M. R., Surya T., Nidhi Sukhija, Kanaka K. K. conducted bioinformatics and data analysis. Nidhi Sukhija, Kanaka K. K. drafted manuscript. Archana Verma, Jayakumar Sivalingam and Niranjan S. K. provided critical inputs during the data analysis and manuscript preparation.

Acknowledgements

The authors express their gratitude to Director of ICAR-National Bureau of Animal Genetic Resources for the support to carry out the research. They extend their heartfelt thanks to Dr. Parameswari B, Senior Scientist in Plant Pathology and Dr. Neeraj Kulshreshtha, Principal Scientist in Plant Breeding and Head of the Regional Centre of Sugarcane Breeding Institute Regional Centre in Karnal, Haryana, for providing the computational resources needed for this work.

References

AGRI-IS. Animal Genetic Resources of India Information System. Accessed on 07-01-2023 from http://14.139.252.116/agris/breed.aspx

Altmann A, Weber P, Bader D, Preuß M, Binder EB, Müller-Myhsok B (2012) A beginners guide to SNP calling from high-throughput DNA-sequencing data. Human Genet 131**:** 1541-1554

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data

Ba H, Jia B, Wang G, Yang Y, Kedem G, Li C (2017) Genome-wide SNP discovery and analysis of genetic diversity in farmed sika deer (Cervus nippon) in northeast China using double-digest restriction site-associated DNA sequencing. *G3:* Genes Genomes Geneti **7**(9):3169-3176

Bill, B. R., Lowe, J. K., DyBuncio, C. T., & Fogel, B. L. 2013. Orchestration of neurodevelopmental programs by RBFOX1: implications for autism spectrum disorder. *International review of neurobiology*, **113**, 251-267.

Buitenhuis, B., Janss, L. L., Poulsen, N. A., Larsen, L. B., Larsen, M. K., & Sørensen, P. 2014. Genome-wide association and biological pathway analysis for milk-fat composition in Danish Holstein and Danish Jersey cattle. *BMC genomics*, **15**(1), 1-11.

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. 2011. Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes| genomes| genetics*, **1**(3), 171-182.

Chao, J., Li, Z., Sun, Y., Aluko, O. O., Wu, X., Wang, Q., & Liu, G. 2021. MG2C: A user-friendly online tool for drawing genetic maps. *Molecular Horticulture*, **1**(1), 1-4.

Chatterjee, S., Szustakowski, J. D., Nanguneri, N. R., Mickanin, C., Labow, M. A., Nohturfft, A., ... & Sivasankaran, R. 2009. Identification of novel genes and pathways regulating SREBP transcriptional activity. *Plos one*, **4**(4), e5197.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., ... & Ruden, D. M. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *fly*, **6**(2), 80-92.

Clempson, A. M., Pollott, G. E., Brickell, J. S., Bourne, N. E., Munce, N., & Wathes, D. C. 2011. Evidence that leptin genotype is associated with fertility, growth, and milk production in Holstein cows. *Journal of dairy science*, **94**(7), 3618-3628.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... & 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. *Bioinformatics*, **27**(15), 2156-2158.

De Donato, M., Peters, S. O., Mitchell, S. E., Hussain, T., & Imumorin, I. G. 2013. Genotyping-by-sequencing (GBS): a novel, efficient and cost-effective genotyping method for cattle using next-generation sequencing. *PloS one*, **8**(5), e62137.

Devadasan, M. J., Kumar, D. R., Vineeth, M. R., Choudhary, A., Surya, T., Niranjan, S. K., ... & Sivalingam, J. 2020. Reduced representation approach for identification of genome-wide SNPs and their annotation for economically important traits in Indian Tharparkar cattle. *3 Biotech*, **10**, 1-6.

Fanjul-Fernández, M., Quesada, V., Cabanillas, R., Cadiñanos, J., Fontanil, T., Obaya, Á., ... & López-Otín, C. 2013. Cell–cell adhesion genes CTNNA2 and CTNNA3 are tumour suppressors frequently mutated in laryngeal carcinomas. *Nature communications*, **4**(1), 2531.

Gaur, G. K., Kaushik, S. N., & Garg, R. C. 2003. The Gir cattle breed of India-characteristics and present status. *Animal Genetic Resources/ Resources génétiques animales/Recursos genéticos animales*, **33**, 21-29.

Han, B., Yuan, Y., Liang, R., Li, Y., Liu, L., & Sun, D. 2019. Genetic effects of LPIN1 polymorphisms on milk production traits in dairy cattle. *Genes*, **10**(4), 265.

IMD, 2023.India Meteorological Department,met center ahmedabad. Retrieved January 7, 2023, from https://mausam.imd.gov.in/ahmedabad/

Iqbal, N., Liu, X., Yang, T., Huang, Z., Hanif, Q., Asif, M., ... & Mansoor, S. 2019. Genomic variants identified from whole-genome resequencing of indicine cattle breeds from Pakistan. *PLoS One*, **14**(4), e0215065.

Jaglan, K., Ravikumar, D., Sukhija, N., George, L., Alex, R., Vohra, V., & Verma, A. 2023. Genomic clues of association between clinical mastitis and SNPs identified by ddRAD sequencing in Murrah buffaloes. *Animal Biotechnology*, 1–9. https://doi.org/10.1080/10495398.2023.2165937

Jeevan, C., Kanaka, K. K., Chethan Raj, R., & Harshitha, B. 2019. Prioritization of breeds for the conservation of animal genetic resources. *Journal of Entomology and Zoology Studies,* **7**(1): 587-590

Jun, G., Asai, H., Zeldich, E., Drapeau, E., Chen, C., Chung, J., ... & Farrer, L. A. 2014. PLXNA 4 is associated with Alzheimer disease and modulates tau phosphorylation. *Annals of neurology*, **76**(3), 379-392.

Kanaka, K. K., Sukhija, N., Goli, R. C., Singh, S., Ganguly, I., Dixit, S. P., ... & Malik, A. A. 2023. On the concepts and measures of diversity in the genomics era. Current Plant Biology, **33**, 100278.

Keller, I., Bensasson, D., & Nichols, R. A. 2007. Transition-transversion bias is not universal: a counter example from grasshopper pseudogenes. *PLoS genetics*, **3**(2), e22.

Klein, C., Bauersachs, S., Ulbrich, S. E., Einspanier, R., Meyer, H. H., Schmidt, S. E., ... & Wolf, E. 2006. Monozygotic twin model reveals novel embryo-induced transcriptome changes of bovine endometrium in the preattachment period. *Biology of reproduction*, **74**(2), 253-264.

Koks, S., Reimann, E., Lilleoja, R., Lättekivi, F., Salumets, A., Reemann, P., & Jaakma, Ü. 2014. Sequencing and annotated analysis of full genome of Holstein breed bull. *Mammalian genome*, **25**, 363-373.

Kour, A., Niranjan, S.K., Malayaperumal, M., Surati, U., Pukhrambam, M., Sivalingam, J., Kumar, A., & Sarkar, M. 2022. Genomic Diversity Profiling and Breed-Specific Evolutionary Signatures of Selection in Arunachali Yak. *Genes,* **13**.

Kraus, R. H., Kerstens, H. H., Van Hooft, P., Crooijmans, R. P., Van Der Poel, J. J., Elmberg, J., ... & Groenen, M. A. 2011. Genome wide SNP discovery, analysis and evaluation in mallard (Anas platyrhynchos). *Bmc Genomics*, **12**(1), 1-11.

Kumar, D. R., Devadasan, M. J., Surya, T., Vineeth, M. R., Choudhary, A., Sivalingam, J., ... & Verma, A. 2020. Genomic diversity and selection sweeps identified in Indian swamp buffaloes reveals it's uniqueness with riverine buffaloes. *Genomics*, **112**(3), 2385-2392.

Langmead, B., & Salzberg, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nature methods*, **9**(4), 357-359.

Li, H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, **27**(21), 2987-2993.

Madalena, F. E., Peixoto, M. G. C. D., & Gibson, J. 2012. Dairy cattle genetics and its applications in Brazil. *Livestock Research for Rural Development*, **24**(6), 1-49.

Malik, A. A., Sharma, R., Ahlawat, S., Deb, R., Negi, M. S., & Tripathi, S. B. 2018. Analysis of genetic relatedness among Indian cattle (Bos indicus) using genotyping by sequencing markers. *Animal genetics*, **49**(3), 242-245.

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R., Thormann, A., ... & Cunningham, F. 2016. The ensembl variant effect predictor. *Genome biology*, **17**(1), 1-14.

Mishra, D. C., Sikka, P., Yadav, S., Bhati, J., Paul, S. S., Jerome, A., ... & Chaturvedi, K. K. 2020. Identification and characterization of trait-specific SNPs using ddRAD sequencing in water buffalo. *Genomics*, **112**(5), 3571-3578.

Ørstavik, S., Sandberg, M., Berube, D., Natarajan, V., Simard, J. S., Walter, U., ... & Jahnsen, T. 1992. Localization of the human gene for the type I cyclic GMP-dependent protein kinase to chromosome 10. *Cytogenetic and Genome Research*, **59**(4), 270-273.

Patel, A. B., Subramanian, R. B., Padh, H., Shah, T. M., Mohapatra, A., Reddy, B., ... & Joshi, C. G. 2017. Identification of single nucleotide polymorphism from Indian Bubalus bubalis through targeted sequence capture. *Current Science*, 1230-1239.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS one*, **7**(5), e37135.

Philippi, A., Roschmann, E., Tores, F., Lindenbaum, P., Benajou, A., Germain-Leclerc, L., ... & Hager, J. 2005. Haplotypes in the gene encoding protein kinase c-beta (PRKCB1) on chromosome 16 are associated with autism. *Molecular psychiatry*, **10**(10), 950-960.

Pimentel, E. C. G., Bauersachs, S., Tietze, M., Simianer, H., Tetens, J., Thaller, G., ... & König, S. 2011. Exploration of relationships between production and fertility traits in dairy cattle via association studies of SNPs within candidate genes derived by expression profiling. *Animal genetics*, **42**(3), 251-262.

Raven, L. A., Cocks, B. G., Goddard, M. E., Pryce, J. E., & Hayes, B. J. 2014. Genetic variants in mammary development, prolactin signalling and involution pathways explain considerable variation in bovine milk production and milk composition. *Genetics Selection Evolution*, **46**(1), 1-13.

Sambrook, J., & Russell, D. W. 2006. Purification of nucleic acids by extraction with phenol: chloroform. *Cold Spring Harbor Protocols*, *2006*(1), pdb-prot4455.

Schmieder, R., & Edwards, R. 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, **27**(6), 863-864.

Schulte, E. C., Stahl, I., Czamara, D., Ellwanger, D. C., Eck, S., Graf, E., ... & Winkelmann, J. 2013. Rare variants in PLXNA4 and Parkinson's disease. *PloS one*, **8**(11), e79145.

Sivalingam, J., Vineeth, M. R., Surya, T., Singh, K., Dixit, S. P., Niranjan, S. K., ... & Ravikumar, D. 2020. Genomic divergence reveals unique populations among Indian Yaks. *Scientific Reports*, **10**(1), 3636.

Stothard, P., Choi, J. W., Basu, U., Sumner-Thomson, J. M., Meng, Y., Liao, X., & Moore, S. S. 2011. Whole genome resequencing of black Angus and Holstein cattle for SNP and CNV discovery. *BMC genomics*, **12**, 1-14.

Su, F. C., Chen, C. M., Chen, Y. C., & Wu, Y. R. 2012. LINGO-2 polymorphism and the risk of Parkinson's disease in Taiwan. *Parkinsonism & Related Disorders*, **18**(5), 609-611.

Surati, U., Verma, A., & Niranjan, S. K. 2023. Genome-wide in silico analysis leads to identification of deleterious L290V mutation in RBBP5 gene in Bos indicus. Animal Biotechnology, 1-9.

Surya, T., Vineeth, M. R., Sivalingam, J., Tantia, M. S., Dixit, S. P., Niranjan, S. K., & Gupta, I. D. 2019. Genomewide identification and annotation of SNPs in Bubalus bubalis. *Genomics*, **111**(6), 1695-1698.

Sushil, K., Anuj, C., Shukla, S. K., & Singh, R. V. 2018. Molecular characterization of A1/A2 beta-casein alleles in Tharparkar and Gir cattle. *Journal of Experimental Zoology, India*, **21**(1), 403-405.

Vassart, G., Bacolla, A., Brocas, H., Christophe, D., De Martynoff, G., Leriche, A., ... & Van Heuverswyn, B. 1985. Structure, expression and regulation of the thyroglobulin gene. *Molecular and cellular endocrinology*, **40**(2-3), 89-97.

Vineeth, M. R., Surya, T., Sivalingam, J., Kumar, A., Niranjan, S. K., Dixit, S. P., ... & Gupta, I. D. 2020. Genome-wide discovery of SNPs in candidate genes related to production and fertility traits in Sahiwal cattle. *Tropical animal health and production*, **52**, 1707-1715.

Wang, H., Jiang, L., Liu, X., Yang, J., Wei, J., Xu, J., ... & Liu, J. F. 2013. A post-GWAS replication study confirming the PTK2 gene associated with milk production traits in Chinese Holstein. *PLoS One*, **8**(12), e83625.

Wara, A. B., Kumar, A., Singh, A., Arthikeyan, A. K., Dutt, T. R. I. V. E. N. I., & Mishra, B. P. 2019. Genome wide association study of test day's and 305 days milk yield in crossbred cattle. *Indian J. Anim. Sci*, **89**, 861-865.

Wen, H., Lei, Y., Eun, S. Y., & P.-Y. Ting, J. 2010. Plexin-A4–semaphorin 3A signaling is required for Toll-like receptor–and sepsis-induced cytokine storm. *Journal of Experimental Medicine*, **207**(13), 2943-2957.

Wilson, P. M., Fryer, R. H., Fang, Y., & Hatten, M. E. 2010. Astn2, a novel member of the astrotactin gene family, regulates the trafficking of ASTN1 during glial-guided neuronal migration. *Journal of Neuroscience*, **30**(25), 8529-8540.