



ISSN: 2456-2912

VET 2024; 9(2): 234-237

© 2024 VET

[www.veterinarypaper.com](http://www.veterinarypaper.com)

Received: 26-12-2023

Accepted: 30-01-2024

#### Ravina

<sup>1</sup> Animal Genetics and Breeding Division, ICAR-National Dairy Research Institute, Karnal, Haryana, India

<sup>2</sup> ICAR-National Bureau of Animal Genetic Resources, Karnal, Haryana, India

#### Pallavi Rathi

<sup>1</sup> Animal Genetics and Breeding Division, ICAR-National Dairy Research Institute, Karnal, Haryana, India

<sup>2</sup> ICAR-National Bureau of Animal Genetic Resources, Karnal, Haryana, India

#### Rangasai Chandra Goli

<sup>1</sup> Animal Genetics and Breeding Division, ICAR-National Dairy Research Institute, Karnal, Haryana, India

#### Kiyevi G Chishi

<sup>1</sup> Animal Genetics and Breeding Division, ICAR-National Dairy Research Institute, Karnal, Haryana, India

<sup>2</sup> ICAR-National Bureau of Animal Genetic Resources, Karnal, Haryana, India

#### Chandana Sree Chinnareddyvari

<sup>1</sup> Animal Genetics and Breeding Division, ICAR-National Dairy Research Institute, Karnal, Haryana, India

<sup>2</sup> ICAR-National Bureau of Animal Genetic Resources, Karnal, Haryana, India

#### Dharamshaw CA

<sup>1</sup> Animal Genetics and Breeding Division, ICAR-National Dairy Research Institute, Karnal, Haryana, India

<sup>2</sup> ICAR-National Bureau of Animal Genetic Resources, Karnal, Haryana, India

#### Himanshu Mehta

B.V.Sc & A.H., M.V.Sc. (Animal Genetics and Breeding), Veterinary Officer, Civil Veterinary Hospital Moonak, Department of Animal Husbandry, Punjab, India

#### Gaurav Patel

<sup>1</sup> Animal Genetics and Breeding Division, ICAR-National Dairy Research Institute, Karnal, Haryana, India

<sup>2</sup> ICAR-National Bureau of Animal Genetic Resources, Karnal, Haryana, India

#### Ayushi Singh

Division of Animal Genetics, ICAR-IVRI, Izatnagar, Bareilly, Uttar Pradesh, India

#### Corresponding Author:

##### Kiyevi G Chishi

<sup>1</sup> Animal Genetics and Breeding Division, ICAR-National Dairy Research Institute, Karnal, Haryana, India

<sup>2</sup> ICAR-National Bureau of Animal Genetic Resources, Karnal, Haryana, India

## Unraveling genomic variants and selective sweeps in livestock: A comprehensive analysis and methodological overview

Ravina, Pallavi Rathi, Rangasai Chandra Goli, Kiyevi G Chishi, Chandana Sree Chinnareddyvari, Dharamshaw CA, Himanshu Mehta, Gaurav Patel and Ayushi Singh

DOI: <https://doi.org/10.22271/veterinary.2024.v9.i2d.1212>

### Abstract

This article delves into the exploration of genomic variants and selective sweeps in livestock, leveraging advanced technologies and bioinformatics tools to deepen our understanding. Through the utilization of next-generation sequencing (NGS), the identification of single nucleotide polymorphisms (SNPs) and insertions/deletions (Indels) is elucidated, underscoring their pivotal role in genetic analyses and breeding programs. Particularly, the spotlight is on whole-genome sequencing in goats, accompanied by an exhaustive overview of bioinformatics tools, accentuating their indispensable contribution to such endeavors. Moving forward, the discussion shifts to the profound impact of animal domestication on genetic diversity, shedding light on the detection of selection signatures and the classification of selective sweeps. This segment of the article highlights the various methodologies employed to identify selection signatures, ranging from population genetics approaches to landscape genomics analyses. By employing these diverse approaches, a nuanced understanding of livestock genetics and the intricate dynamics of selection for breeding and conservation purposes is achieved, providing valuable insights for future research and practical applications in livestock management.

**Keywords:** Genomic variants, selective sweeps, livestock

### Introduction

In recent years, the advent of next-generation sequencing (NGS) technology has been instrumental in advancing genetic studies of complex traits in domestic animals. Whole-genome sequencing (WGS) approaches have particularly revolutionized these endeavors, offering novel avenues for discerning genomic variants among diverse livestock breeds adapted to various biogeographic regions and production environments (Weldenegodguad *et al.*, 2019) [3]. By comprehensively mapping the entire genome sequence of a species through diverse sequencing and assembly alignment methodologies, WGS facilitates the exploration of genetic diversity, identification of selection signals, and elucidation of the origins and evolutionary trajectories of representative individuals or groups, thus constructing a comprehensive whole-genome genetic variation map (Wang *et al.*, 2016) [2]. The imprints of selection processes, both natural and artificial, have left distinct signatures across the genome. These signatures, termed as selection signatures, manifest as alterations in genetic variation within genomic regions proximal to causal variants, induced by the selective pressures exerted by natural or artificial factors (Nielsen, 2005; Jensen *et al.*, 2016) [14, 18]. These variants often influence multiple traits, ultimately contributing to the distinctive characteristics observed within breeds, encompassing a spectrum of phenotypic features such as size, color, horn morphology, and various production, reproductive, and adaptive traits. The process by which the frequency of a selectively favored variant escalates within a population is known as selective sweep (Maynard and Haigh, 2007) [4]. The emergence of species-specific genomic tools, such as single nucleotide polymorphism (SNP) arrays, has further propelled whole-

genome analyses in livestock species. These tools enable researchers to delve into various facets of genetic diversity, including the elucidation of signatures left by selection processes.

### Genomic Variants

Variations occur both within and between populations, resulting in polymorphism that could be associated with genetic traits or phenotypes, influenced by environmental factors. With the advent of next-generation sequencing (NGS) technologies, the identification of numerous genomic variants, including single nucleotide polymorphisms (SNPs) and insertions and deletions (Indels), has become feasible in livestock (Mullen *et al.*, 2012) [6]. SNPs are classified as variants occurring at the single base pair level, with a frequency of more than 1% in the population. While multiallelic SNPs do exist, the majority are biallelic, characterized by the occurrence of two alternate bases (Wang *et al.*, 1998) [19]. Indel variants entail changes in DNA genomes due to the insertion or deletion of nucleotides within a small length, typically less than 1000 bp (Sehn, 2015) [20]. SNPs represent the most prevalent form of variation in the genome and are extensively utilized to explore genetic disparities among individuals and populations. These SNPs may instigate alterations in the genomic sequence, affecting coding regions (exons), intergenic regions, or noncoding regions (introns) (Dijk *et al.*, 2014; Ahmad *et al.*, 2018) [21, 22].

### Identification of genomic variants using whole genome sequence analysis

With the advancement of high-throughput deep sequencing technology, researchers now have the capability to analyze and sequence the entire genome of animals, thereby detecting genomic variants based on reference genomes, which serve as valuable resources for genetic analyses (Weldenegodguad *et al.*, 2019) [3]. These sequences are currently archived in publicly accessible databases, such as GenBank, established by the National Center for Biotechnology Information (NCBI) (Bai *et al.*, 2012) [8]. Various species have been subject to the identification of variants, such as SNPs and INDELS, through whole-genome sequencing data. For instance, a previous study on goat species reported the detection of millions of variants across eight different goat breeds (Wang *et al.*, 2016) [1]. Although there is limited information available on whole-genome re-sequencing in indigenous goat breeds, Chowdhury *et al.* (2019) [9] documented studies on the Black Bengal goat breed. The recent advancements in sequencing technologies, coupled with their increased cost-effectiveness, have facilitated the identification of high-quality variants.

### Bioinformatic tools for the identification of genomic variants

To ensure the identification of high-quality variants, it is imperative to employ bioinformatics tools that guarantee reliability, accuracy, consistency, efficiency, and cost-effectiveness, given the inherent susceptibility of NGS technology to technological errors (Mielczarek and Szyda, 2016) [10]. A plethora of tools are accessible within the variant calling pipeline, spanning from quality assessment to the identification of pristine, high-quality variants, as elaborated below.

**Table 1:** Important list of bioinformatics tools for the identification of genomic variants

SL. No	Tool	Functions
1.	Fast QC	Quality control
2.	Trimmomatic	Quality filtering, adapter removal
3.	Fastp	Adapter trimming and quality filtering
4.	Fastx	Trimming the read data
5.	PRINSEQ	Trimming of the read
6.	Cutadapt	Finds and removes adapter sequences
7.	Burrows Wheel Aligner (BWA)	Mapping the sequence against referencegenome
8.	Bowtie 1	
9.	Bowtie 2	
10.	Novoalign	
11.	GEM	
12.	Samtools	Conversion of BAM to SAM, sorting and manipulation of files
13.	Picard tool	Removes duplicates
14.	VCFTools	Filtering the raw data, Manipulation of VCF files, statistics calculation in SNP
15.	GATK	Toolkit used for variant calling
16.	Annovar	Annotation of the variant
	Variant Effect Predictor	
	SnEff	

### Selective sweeps

The process of animal domestication has exerted significant influence on human socioeconomics and cultural identity. As animals underwent domestication, they underwent various changes in physical attributes, behavior, and production traits due to both natural and artificial selection, ultimately leading to the emergence of new breeds (Flori *et al.*, 2009) [11]. These processes of natural and artificial selection deviate from the patterns of molecular variation predicted by the neutral theory, as each type of selection uniquely affects both the selected loci and the neutral loci linked to them (Kreitman, 2000) [12].

When a newly developed allele conferring a selective advantage undergoes positive selection, it tends to increase in frequency within the population and carry linked neutral alleles along with it. This phenomenon is commonly referred to as selective sweeps or hitchhiking effects (Maynard-Smith and Haigh, 1974; Charlesworth, 2007) [4, 23]. Detection of selection signatures has become a primary focus for animal geneticists, as these signatures can unveil genes and advantageous mutations that confer selective advantages within specific livestock populations (Zhao *et al.*, 2015) [13]. The identification of selection signatures has significantly advanced our understanding of the evolutionary processes that shape genetic and genomic variability. It has also contributed

to the development of methods for selection and conservation (Nielsen, 2005) [14]. Maynard-Smith and Haigh proposed three distinct patterns that emerge locally surrounding the location of an advantageous mutation. These include a reduction in segregating sites density in adjacent regions to decrease heterozygosity, a skewness in the site frequency spectrum (SFS) towards extreme frequencies, and an increase in haplotype linkage disequilibrium (LD) around the positively selected allele compared to neutral expectations (Maynard-Smith and Haigh, 1974) [4].

In recent years, the search for genes or genomic regions associated with economically significant traits favored by selection has garnered increased interest among researchers. This heightened interest has been fueled by advancements in next-generation sequencing (NGS) technologies, improved statistical tools, enhanced availability of genomics data, and the proliferation of bioinformatics tools.

### Types of selective sweeps

Selective sweeps come in various types, including hard or soft, complete, or partial, contingent upon the origin, nature, and frequency of mutation.

**Hard/classic selective sweep:** In this scenario, the adaptive allele undergoes an increase in frequency across subsequent generations, eventually achieving fixation within the population. This process leads to a reduction in genetic variation both upstream and downstream of the favorable alleles, termed as a selective sweep (Pritchard *et al.*, 2010) [15].  
**Soft selective sweep:** Soft selective sweeps can be categorized into two types based on the presence of beneficial alleles, namely single-origin and multiple-origin soft sweeps (Hermisson and Pennings, 2017) [16].

In a single-origin soft sweep, the selection process acts upon standing genetic variation, which was previously either neutral or deleterious. The signatures left by selection in such cases tend to be less apparent initially but eventually become fixed due to environmental influences or genetic changes over time (Hermisson and Pennings, 2005) [16]. In contrast, multiple-origin soft sweeps involve the emergence of numerous advantageous mutations at a single locus across various genomic backgrounds. These mutations simultaneously increase in frequency, preventing any one of them from reaching fixation individually during the selective sweep (Saravanan *et al.*, 2020) [17].

### Different approaches used for identification of selection signatures

#### Site Frequency Spectrum

The Site Frequency Spectrum (SFS) constitutes a set of tests reliant on the distribution of allele frequencies within a population (Achaz, 2009; Ronen *et al.*, 2013) [24, 25]. Various methods, including Tajima's D (Tajima, 1989), Fay and Wu's H statistic (Fay and Wu, 2000), and the composite likelihood ratio test (CLR) (Lindsay, 1988), are employed to discern selected regions.

#### Linkage Disequilibrium

Linkage Disequilibrium (LD) methods gauge prolonged homozygous regions characterized by haplotypes with high frequencies engendered by selective sweeps. Techniques such as Long Range Haplotype (LRH), Extended Haplotype Homozygosity (EHH), Relative Extended Haplotype Homozygosity (REHH) (Sabeti *et al.*, 2002) [26], and integrated haplotype score (iHS) (Voight *et al.*, 2006) [27] are utilized for this purpose.

### Reduced local variability

Reduced local variability methodologies target the identification of genomic regions exhibiting diminished variation relative to the genome average. Approaches such as runs of homozygosity (ROH) (McQuillan *et al.*, 2008) [28] and pooled heterozygosity (Hp) (Rubin *et al.*, 2010) [29] are employed in this regard.

### Population Differentiation

Population differentiation incorporates two approaches: fixation index (FST) (Wright, 1949) [30] and FLK (Bonhomme *et al.*, 2010) [31], within differentiation-based methodologies.

### Conclusion

The integration of next-generation sequencing (NGS) technology, particularly whole-genome sequencing (WGS), has significantly propelled genetic studies in domestic animals. By leveraging diverse sequencing and assembly alignment methodologies, WGS has enabled the comprehensive mapping of entire genome sequences, facilitating the exploration of genetic diversity and the identification of selection signals. Selection signatures, resulting from natural and artificial selection processes, manifest as alterations in genetic variation within specific genomic regions, ultimately shaping various phenotypic traits observed within breeds. The emergence of species-specific genomic tools, such as single nucleotide polymorphism (SNP) arrays, further enhances the elucidation of genetic disparities and selection signatures. Additionally, the identification of genomic variants, including SNPs and insertions/deletions (Indels), through whole-genome sequencing analysis, offers valuable insights into genetic traits and population diversity. To ensure the accuracy and reliability of variant identification, the utilization of bioinformatics tools within the variant calling pipeline is essential. Overall, the utilization of advanced genomic technologies and bioinformatics tools has revolutionized genetic analyses in livestock, paving the way for comprehensive understanding and practical applications in breeding and conservation efforts.

### References

1. Wang X, Liu J, Zhou G, Guo J, Yan H, Niu Y, *et al.* Whole-genome sequencing of eight goat populations for the detection of selection signatures underlying production and adaptive traits. *Sci. Rep.* 2016;6(1):38932. DOI: 10.1038/srep38932
2. Wang Z, Sun H, Chen Q, Zhang X, Wang Q, Pan Y, *et al.* A genome scan for selection signatures in Taihupig breeds using next-generation sequencing. *Anim.* 2019;13(4):683-693. DOI: 10.1017/S1751731118001714
3. Weldenogodguad M, Popov R, Pokharel K, Ammosov I, Ming Y, Ivanova Z, *et al.* Whole-genome sequencing of three native cattle breeds originating from the northernmost cattle farming regions. *Front Genet.* 2019;9:728. DOI: 10.3389/fgene.2018.00728
4. Maynard J, Haigh J. The hitch-hiking effect of a favourable gene. *Genet Res.* 2007;89(5-6):391-403. DOI: 10.1017/S0016672308009579
5. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297-1303. DOI: 10.1101/gr.107524.110



6. Mullen MP, Creevey CJ, Berry DP, McCabe MS, Magee DA, Howard DJ, *et al.* Polymorphism discovery and allele frequency estimation using high-throughput DNA sequencing of target-enriched pooled DNA samples. *BMC Genomics*. 2012;13(1):16. DOI: 10.1186/1471-2164-13-16
7. Muñoz M, Bozzi R, García F, Núñez Y, Geraci C, Crovetto A, *et al.* Diversity across major and candidate genes in European local pig breeds. *PLoS ONE*. 2018;13(11):e0207475. DOI: 10.1371/journal.pone.0207475
8. Bai Y, Sartor M, Cavalcoli J. Current status and future perspectives for sequencing livestock genomes. *J Anim. Sci. Biotechnol.* 2012;3(1):8. DOI: 10.1186/2049-1891-3-8
9. Chowdhury SMZH, Nazir KHMHNH, Hasan S, Kabir A, Mahmud MM, Robbani M, *et al.* Whole genome analysis of black bengal goat from savar goat farm, bangladesh. *BMC Res Notes*. 2019;12(1):687. DOI: 10.1186/s13104-019-4700-7
10. Mielczarek M, Szyda J. Review of alignment and SNP calling algorithms for next-generation sequencing data. *J Appl Genet*. 2016;57(1):71-79. DOI: 10.1007/s13353-015-0292-7
11. Flori L, Fritz S, Jaffrézic F, Boussaha M, Gut I, Heath S, *et al.* The genome response to artificial selection: A case study in dairy cattle. *PLoS ONE*. 2009;4(8):e6595. DOI: 10.1371/journal.pone.0006595
12. Kreitman M. Methods to detect selection in populations with applications to the human. *Annu. Rev. Genomics Hum Genet*. 2000;1(1):539-559. DOI: 10.1146/annurev.genom.1.1.539
13. Zhao F, McParland S, Kearney F, Du L, Berry DP. Detection of selection signatures in dairy and beef cattle using high-density genomic information. *Genet Sel Evol*. 2015;47(1):49. DOI: 10.1186/s12711-015-0127-3
14. Nielsen R. Molecular signatures of natural selection. *Annu Rev Genet*. 2005;39(1):197-218. DOI: 10.1146/annurev.genet.39.073003.112420
15. Pritchard JK, Pickrell JK, Coop G. The genetics of human adaptation: Hard sweeps, soft sweeps, and polygenic adaptation. *Curr. Biol*. 2010;20(4):R208-R215. DOI: 10.1016/j.cub.2009.11.055
16. Hermisson J, Pennings PS. Soft sweeps and beyond: Understanding the patterns and probabilities of selection footprints under rapid adaptation. *Methods Ecol. Evol*. 2017;8(6):700-716. DOI: 10.1111/2041-210X.12808
17. Saravanan KA, Panigrahi M, Kumar H, Bhushan B, Dutt T, Mishra BP, *et al.* Selection signatures in livestock genome: A review of concepts, approaches and applications. *Livest Sci*. 2020;241:104257. DOI: 10.1016/j.livsci.2020.104257
18. Jensen JS, Cusini M, Gomberg M, Moi H. 2016 European guideline on *Mycoplasma genitalium* infections. *Journal of the European Academy of Dermatology and Venereology*. 2016 Oct;30(10):1650-6.
19. Wang XZ, Harding HP, Zhang Y, Jolicoeur EM, Kuroda M, Ron D, *et al.* Cloning of mammalian Irel reveals diversity in the ER stress responses. *The EMBO journal*. 1998 Oct 1;17(19):5708-5717.
20. Sehn LH, Gascoyne RD. Diffuse large B-cell lymphoma: optimizing outcome in the context of clinical and biologic heterogeneity. *Blood, the Journal of the American Society of Hematology*. 2015 Jan 1;125(1):22-32.
21. Roofthoof DW, Simons SH, Anand KJ, Tibboel D, Van Dijk M. Eight years later, are we still hurting newborn infants?. *Neonatology*. 2014 Feb 4;105(3):218-226.
22. Ahmad MA, Eckert C, Teredesai A. Interpretable machine learning in healthcare. In *Proceedings of the 2018 ACM international conference on bioinformatics, computational biology, and health informatics*; c2018 Aug 15. p. 559-560.
23. De Miguel E, Iribarren I, Chacon E, Ordonez A, Charlesworth S. Risk-based evaluation of the exposure of children to trace elements in playgrounds in Madrid (Spain). *Chemosphere*. 2007 Jan 1;66(3):505-13.
24. Achaz G. Frequency spectrum neutrality tests: one for all and all for one. *Genetics*. 2009 Sep 1;183(1):249-58.
25. Christiansen M, Fagerholt K, Nygreen B, Ronen D. Ship routing and scheduling in the new millennium. *European Journal of Operational Research*. 2013 Aug 1;228(3):467-483.
26. Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, *et al.* Detecting recent positive selection in the human genome from haplotype structure. *Nature*. 2002 Oct 24;419(6909):832-837.
27. Voight BF, Kudravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS biology*. 2006 Mar;4(3):e72.
28. McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, *et al.* Runs of homozygosity in European populations. *The American Journal of Human Genetics*. 2008 Sep 12;83(3):359-372.
29. Rubin RB, Rubin AM, Graham EE, Perse EM, Seibold D. *Communication research measures II: A sourcebook*. Routledge; c2010 Apr 26.
30. Wright S. The genetical structure of populations. *Annals of eugenics*. 1949 Jan;15(1):323-354.
31. Bonhomme M, Chevalet C, Servin B, Boitard S, Abdallah J, Blott S, *et al.* Detecting selection in population trees: the Lewontin and Krakauer test extended. *Genetics*. 2010 Sep 1;186(1):241-262.