The Pharma Innovation Journal 2023; 12(2): 2630-2639

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277-7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2023; 12(2): 2630-2639

www.thepharmajournal.com
Received: 28-12-2022
Accepted: 30-01-2023

Sonali Sonejita Nayak
Division of Animal Genetics, Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, India

Harshit Kumar
Division of Animal Genetics, Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, India

Divya Rajawat
Division of Animal Genetics, Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, India

KA Saravanan
Division of Animal Genetics, Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, India

Anuradha Panwar
Division of Animal Genetics, Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, India

Kanika Ghildiyal
Division of Animal Genetics, Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, India

Nabaneeta Smaraki
Division of Virology, Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, India

Harsh Rajeshbhai Jogi
Division of Virology, Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, India

Anurodh Sharma
Division of Animal Genetics, Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, India

Bharat Bhushan
Division of Animal Genetics, Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, India

Triveni Dutt
Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, India

Dr. Manjit Panigrahi
Division of Animal Genetics, ICAR-Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, India

Corresponding Author:
Dr. Manjit Panigrahi
Division of Animal Genetics, ICAR-Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, India

Coalescence: An anti-clockwise travel

Sonali Sonejita Nayak, Harshit Kumar, Divya Rajawat, KA Saravanan, Anuradha Panwar, Kanika Ghildiyal, Nabaneeta Smaraki, Harsh Rajeshbhai Jogi, Anurodh Sharma, Bharat Bhushan, Triveni Dutt and Manjit Panigrahi

Abstract
Since the early 1980s, the stochastic process known as 'the coalescent' has played a central role in population genetics, and results based on it are now routinely used to analyse DNA sequence polymorphism data. Coalescence theory provides a formal technique to anticipating how diverse biological activities may affect patterns of genomic variation, in addition to establishing the basis for much of modern evolutionary inference. Polymorphisms in current populations are the result of historical occurrences, and population genetic investigations frequently simulate gene lineage branching to anticipate the time to the most recent common ancestor. The emphasis in coalescent thinking is to view populations backward in time, using observable divergence in a population to estimate the time to a most recent common ancestor (MRCA); This ancestor is the point at which gene genealogies come together, or coalesce', in a single biological organism. The main parameters estimated in coalescence analyses are the theta (product of the mutation rate and adequate population size) and coalescence time. The coalescence time refers to the number of generations that have elapsed since the samples had a common ancestor. This approach takes a backward-looking, retrospective look at the genetic process.

Keywords: Coalescence, evolution, population genetics, simulation, wright-fisher model

Introduction
Population genetics originated as an attempt to reconcile Mendelian inheritance with biostatistics models. Fisher, Haldane, and Wright's work established population genetics as a field. Individuals that were more adaptable in our prehistoric age passed their genes on to future generations. The degree of genetic variability between or within livestock species, breeds, or populations is referred to as diversity. The diversity found in nature is the result of evolutionary mechanisms like mutation, gene flow, selection, non-random mating, and genetic drift etc. Natural selection leaves signatures in our genome which can be exploited to discover genes that may be responsible for metabolism variations (Bamshad and Wooding 2003) [3]. Fisher (1930) [19] and Wright (1931) [72] presented a population model that describes the genealogical link between genes. This fundamental reproduction model dynamically follows the evolution of an idealized population and the transmission of genes from one generation to the next. Population biology and genetics are important components of the Theory of Evolution. Motoo Kimura's (1924-1994) contributions later made the theory more robust through the enhanced use of diffusion theory. In 1968, Kimura formulated the neutral hypothesis of molecular evolution. Most of the observed genetic variations are selectively neutral, according to the Neutral Theory. Genetic drift refers to the random variation of allele frequency over time. Furthermore, genetic drift is a process that occurs in the past and continues into the present. Although standard population genetics models of drift use a "forward" approach from the past to the present, we may also study drift from the present back into the past. The coalescent hypothesis is a logical extension of the more traditional Population Genetics concept of neutral evolution, as well as a close approximation to the Wright-Fisher model (for large populations). It was found independently by many researchers in the 1980s, but Kingman is credited with the final formalization. The "backward" method, from the present back into the past, is the foundation of coalescent theory. Following a pair of alleles backward was first proposed by Malécot, which relates the passage of time to their common ancestor. This is the fundamental concept underlying the theory of coalescence. John Kingman developed the mathematical derivation for the coalescent in the early 1980s. Kingman described the coalescent in 1982 (Kingman 1982b) [34] but Hudson (1983) [27] and Tajima (1983) [63] discovered it independently (Donnelly and Tavaré 1995) [14].

For a classical Wright-Fisher model, Kingman explained that this backward process could be well described by a single time-homogeneous Markov chain. The coalescent is a chain of events. Since then, the coalescent model has been applied to a wide range of mutation and demographic models (Donnelly and Tavaré 1995 [14], Donnelly 1999) [15] surprisingly robust to deviations from the norm. Coalescent theory connects population genetic models and molecular data. It explains how demography, recombination, and other factors influence the shape of gene trees and provides statistical tools. Between the early 1970s and the early 1980s, there was a shift in perspective from population genetics to the coalescent approach, which is now the dominant view. A new one starts with a sample and goes back in time (Ewens 1990) [17]. The coalescent theory has evolved from an esoteric problem pursued purely mathematical reasons to an essential conceptual tool used to make testable predictions over the last two decades (Hamilton 2009) [23].

The concept of evolution

The fundamentals of population genetics theory explain how a population evolves under a particular set of circumstances. Evolution is a forward-looking process in which the genetic makeup, allele frequencies, and genotype frequencies change over time. Theodosius Dobzhansky was a noted geneticist who stated that "Nothing in biology does sense apart from in the light of evolution". He wrote an essay on it in 1973, criticizing anti-evolution philosophy and advocating theistic evolution. When evolutionary processes such as natural selection and genetic drift work on variation, some qualities become more frequent or rare within a population (Scott-Phillips 2014) [77], The factors that determine whether a feature is widespread or rare within a population change regularly, resulting in changes in inheritable qualities occurring across succeeding generations. This activity of evolution has resulted in biodiversity at all levels of biological social group, including species, individual creatures, and molecules. Biodiversity is often investigated at three levels namely genetic diversity, species diversity, and ecological diversity. These three layers interact to produce the complexness of life on Earth. The diversity we see today is because of evolutionary forces. Natural selection will only resulted in evolution if a population contains a sufficient amount of genetic variation. Blending inheritance is a popular theory before the discovery of Mendelian genetics. However, with blended inheritance, genetic variation would be swiftly lost, making evolution improbable by natural or artificial selection. The mathematical theory of population genetics, developed by Wright (1931) [72], Fisher (1930) [19], Malecot (1941) [78], and Kimura (1954) [79], states that genomes change by random processes like mutation, drift, migration and selection. The Hardy–Weinberg principle explains how variation is preserved in a population in conjunction with Mendelian heredity.

Hardy-Weinberg Equilibrium

Hardy-Weinberg Law explains the genetic equilibrium in a population. Wilhelm Weinberg (a German physician) & Godfrey Harold Hardy (a British mathematician) postulated the law in 1908, independently. It states that in a sizeable random-mating population, in the absence of selection, mutation, or migration, the gene and genotype frequencies are constant from generation to generation. Furthermore, there is a simple relationship between the gene and genotype frequencies. The assumptions are no selection, no mutation, no migration, random mating between individuals, and a large population. This combined Mendelian genetics with natural selection, which was an important footstep in the development of a cohesive explanation of how evolution works (Larson 2000) [38]. A simple model of populations illustrating the genealogical correlation between genes is that introduced by Fisher (1930) [19] and Wright (1931) [72] called as Wright-Fisher model. This is the basic null model which argues that no change occurs unless evolutionary mechanisms like selection, genetic drift, mutation, and gene flow from other populations are present.

Wright-Fisher Model

It takes a different approach than the Hardy-Weinberg model in that it introduces specific generations and individuals. Here, we need a model of evolution in small populations to think about quantitative changes in such populations. The Wright-Fisher model, named after its creators Sewall Wright and RA. Fisher (Fig. 1). It is one of the simplest models and most beneficial for population genetics. Wright (1931) [72] and Fisher (1930) [19] proposed a simple population model for describing the genealogical relationship between genes (Wright 1931, Fisher 1930, Hein *et al*. 2006) [72, 19, 24].

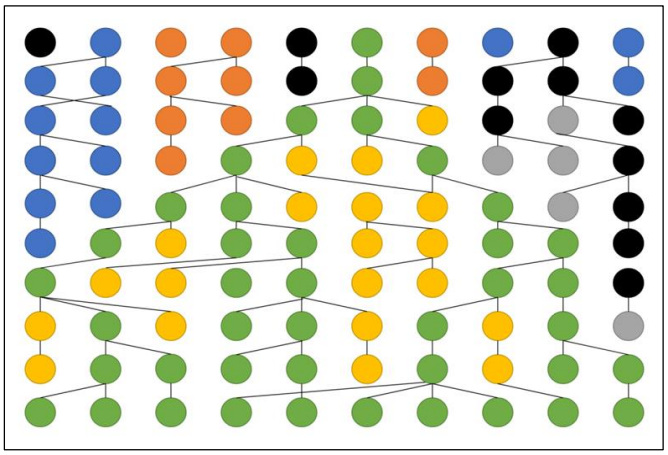

Fig 1: Balls in different strata represent different generations. Each ball represents one individual. Figure depicts the Wright–Fisher model of population size 10 for nine reproduction cycles which equivalent to ten generations. This diagram depicts the genealogical relationships of all genes over the course of ten generations.

Assumptions of wright-fisher model

Generations should be non-overlapping and discrete, constant population size, all individuals are equally fit, the population has no geographical or social structure, genes in the population are not recombining, usually haploid individuals or two sub-populations (male & female) are assumed (Berestycki 2009) [7]. There is no migration, mutation, random mating, or natural selection operating at the locus in question. This model, however, is flawed in that it assumes a constant population size, which is not always the case. A bottom-up analysis of the Wright-Fisher model emphasizing coalescence events yields the coalescent model. Genetic drift, gene flow, recurrent mutation, and natural selection are the primary factors determining allele frequencies. Because of the accessibility of molecular data on all genetic variations, the neutral hypothesis of molecular evolution was developed.

Neutral Theory

Neutral theory was proposed by Kimura (Kimura 1968 [29], Kimura 1977 [30], Kimura 1983 [31]). It suggests that within a population, a considerable majority of variation is neutral. Most of the changes in amino acid sequence or DNA between related species are selectively neutral. It forecasts the amount of variation in a population and the relative rates of non-synonymous and synonymous substitution and other population genetic quantities. In a neutral evolution model, selection does not act on variation at nonsynonymous sites and does not act on variation at synonymous sites. The neutral theory of biodiversity provides a scalable, mechanistic model of biodiversity that includes individual-level dispersal, drift, and speciation dynamics (Thompson *et al.* 2020) [65]. Corresponding to the neutral theory the most of the variations present within a population are selectively neutral, and also most variations in DNA or amino acid sequences over time are also selectively neutral, resulting in many molecular variations between related species. The neutral theory proposes a scalable, mechanical model of biodiversity that takes into account individual-level dispersion, drift, and speciation processes (Thompson *et al.* 2020) [65]. Then, the coalescent theory came as a natural expansion of the more conventional population genetics idea of neutral evolution, as well as a large population idea to the Wright-Fisher model.

Coalescence

In a pedigree, tracing the ancestry history for allele copies gives a means to explain the current patterns in those allele copies. When two homologous DNA molecules merge back into a single DNA molecule in the past, this is a coalescent event. As time passes, the study of random processes in which particles join together to form clusters is known as the coalescent theory (Rosindell and Cornell 2007) [55]. Coalescence is accomplished by tracing lineages backward in time (Rosindell *et al.* 2007) [56]. The coalescence time refers to the number of generations that have elapsed since the samples had a common ancestor. Coalescence explains you take a selection of genes from a population and trace their ancestors back in time when two or more genes in a sample descended from a common ancestor (Wakeley 2020) [69]. With time appropriately scaled and large population size, Kingman (1982b) [34] demonstrated that this backward process is well described for a classical Wright-Fisher model by a single time-homogeneous Markov chain called the coalescent (Fig. 2).

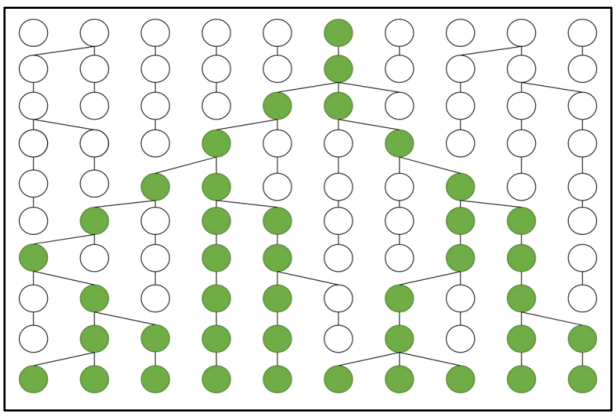

Fig 2: Each ball in the figure represents one individual. Each stratum represents different generations. Branching pattern of Green balls represent a coalescence structure

The fundamental coalescent is predicated on several assumptions, no selection, random mating, and no population structure, including constant population (Hein *et al.* 2006) [24]. These assumptions can be relaxed by including additional parameters such as a population growth term (Slatkin and Hudson 1991) [61], these other parameters are currently not taken into account in current methods for characterizing and detecting recombination (Slatkin and Hudson 1991, Kuhner *et al.* 2000, Fearnhead and Donnelly 2006) [61, 18]. The time to the most recent common ancestor (TMRCA) is often used to refer to the ultimate coalescent time when all DNA copies surveyed today collapse into one ancestral molecule. The coalescent theory uses a retrospective approach to trace alleles of a gene shared by all sampled members of a current population back to one ancestral copy known as the sample's most recent common ancestor (MRCA), (McVean *et al.* 2000) [41]. A sample's genealogy can be described in terms of branch lengths and topology. The measurements of the branches represent the waiting time between subsequent coalescence. Although the coalescence of neutral alleles is easy in ideal populations, it is complicated by biological realities including population structure, population size variations, and non-Poisson variation in reproductive success (Agrawal and Hartfield 2016) [80]. In addition to ecological influences, the selection at one site in the genome can have a significant impact on the coalescence timeframes of connected neutral sites, impacting genomic diversity patterns (Saravanan *et al* 2019 [97]. Coalescence periods can be drastically reduced using selective sweeps (Maynard Smith and Haigh 1974, Kaplan *et al.* 1989) [39, 28].

Felsenstein's bugs in a box

Joe Felsenstein, a population geneticist, has proposed a delightful metaphor for thinking about coalescence as a stochastic process that moves forward in time. Felsenstein imagines a box full of cannibalistic and voracious bugs. The bugs wander around the box at random, and when two bugs contact each other, one eats the other. The procedure is ongoing until there is the last surviving bug in the box. In terms of mathematics, Felsenstein's metaphor of bugs-in-a-box corresponds to the coalescent process for a neutral locus. Felsenstein's bugs-in-a-box analogy is mathematically identical to the coalescent process for a neutral locus but with time running forward instead of backward. According to Felsenstein's metaphor, bugs are copies of genes. When one bug eats another, this is referred to as a coalescent event. When only the last bug is left in the box, the entire population has coalesced.

The Fig: 3 Shows coalescent structures for loci from the very same population assuming pure drift, balancing selection, a selective sweep, and a partial sweep. The points of the tree at the bottom of the graph indicate five sampled alleles, which eventually coalesce into a single lineage as we go back in time. This last coalescent point indicates the MRCA of the sampled alleles. The duration to the MRCA (TMRCA) for balancing selection is longer than for neutral genes, which is longer than for a sweep area. When all alleles are selectively neutral and the effective population size is constant, the conventional coalescent gives the expected branch length. The chance of coalescence changes with changes in the effective population size (population expansion, population bottlenecks). Natural selection influences the likelihood of coalescence by varying the fitness of alleles in each branch.

Natural selection and variations in the effective population size affect the estimated time to coalescence and, as a result, the anticipated branch lengths in a genealogy tree. If the chance of coalescence is greater in the present than in the past, most coalescent events occur near the present and internal branches are long in comparison with external branches.

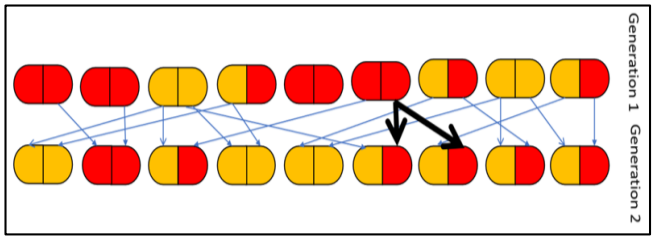

Fig 3: One capsule represents one individual having two alleles at a locus in generation 1(parents) & two (offspring's).Let there are 'N' number of individuals. Probability that an allele in Generation-2 has a parent in Generation-1 = one. Probability that a random allele in Generation-2 has the same parent in Generation-1 = 1/2N. Probability that they don't share common parent = 1-1/2N

Why Coalescence?

Coalescent methods are mainly related to the genealogical tree of the sampled genes when the structure of genetic data is considered. These are useful and attractive because; a) Genealogical methods are more accessible than traditional approaches for quantitative analysis of randomly formed models. b) The application of the coalescent results in highly efficient simulation methods. c) The shape of genetic data reflects the underlying genealogy to a large extent. Hence, an understanding of genealogy improves a qualitative understanding of genetic variation patterns d) it offers inference techniques for genetic data that, for the first time, allow full use of the information possessed by that data (Nordborg 2019) [46].

Coalescence is not only faster than traditional forward-in-time methods but it can also model scenarios that would be virtually impossible to model using the forwards-in-time method, such as scenarios set in almost infinite landscapes (Rosindell and Cornell 2007) [55]. On the one hand, population geneticists have increased their demand for models that incorporate more natural features than Kingman's coalescent allows for (Berestycki 2009) [7].

Coalescence: A Mathematical Approach

The likelihood that the two lineages will select the same parent and coalesce is 1/N, while the possibility that they will choose different parents and remain distinct is 1-1/N. Because generations are self-contained, the probability that they will remain distinct more than t generations in the past is $(1\text{-}1/N)^t$. In a Wright-Fisher model, consider k number copy of genes from N individuals (diploid). How long will it take for all k lineages to unite? For this, we must add the average waiting times for each subsequent coalescent event, beginning with the first when there are k lineages and ending with only two lines.

This gives;

$$\sum_{i=2}^{k} 4N/i(i-1) \; = 4N(1-1/k)$$

As a result, the average coalescent time for k number copy of genes in a large population is about 4N. The expected time for the last event to occur is 2N, which is more than fifty percent of the completely coalescent time for all k lineages. The average time to coalescence for any random gene copy pair in a diploid Wright-Fisher population of size N for a neutral locus is 2N generations.

In a haploid model with 2N genes, what is the distribution of the waiting time until the MRCA of two genes sampled? The likelihood of a coalescence event occurring in a given generation is 1/2N. The likelihood that no coalescence event will occur is (1-1/2N). Because sampling in different generations is independent of one another, the likelihood that two genes find a common ancestor j generation back in time is = $(1\text{-}1/2N)^{j-1}X1/2N$. They chose different ancestors in the first j-1 generations and then the same ancestor in generation j (Fig. 4). For a sample of "n" genes, the likelihood that k genes had k distinct ancestors in the previous generation is

$$\frac{2N-1}{2N} \cdot \frac{2N-2}{2N} \cdots \frac{2N-K+1}{2N} = \prod_{i=1}^{K-1}(1 - i/2N)$$

Expansion of Coalescence

The basic coalescent technology of modeling the genealogical process for a sample of genetic data has been expanded in various ways. Slatkin (1991) [62] considered changes in population size over time (Slatkin and Hudson 1991) [61]; Notohara (1990) [48] provided a general mathematical model of coalescence in a geographically structured population; Kaplan & his co-workers (1988) modeled strong selection, and Krone and Neuhauser (1997) [35] described a framework for the coalescent with weak selection. Only recently has the coalescent approach been formalized for the population structure, with the development of a general model known as the structured coalescent.

Coalescence in simulation algorithm

Simulation has frequently been used to compare extensive hypotheses relating to genetics and population genetic management at a low cost (Krone and Neuhauser 1997) [35]. Simulated data is beneficial for defining the effect of current selection and management practices over time, which is frequently not the case with accurate data due to time and cost constraints (Table 1). Simulation is also helpful in optimizing marker panel construction in terms of SNP uniformity across the genome, the impact of including preselected candidate causative mutations, and the proportion of individual to genotype for a given marker density in a population (Daetwyler *et al.* 2013) [13]. Simulation processes in population genetics are essential for better understanding the effect of several evolutionary and demographic circumstances on sequence patterns and variation and allowing researchers to evaluate better and create analytical methods for the study of disease-related genetic factors (Howard *et al.* 2017) [26]. Mathematical or computer simulations are used to study theoretical models.

A backward-in-time coalescence model entails selecting an individual to be 'un-born' and locating its parent through an 'un-dispersal' event (reproduction is modeled asexually so that each individual has only one parent), (Thompson *et al.* 2020) [65]. There are three types of genetic data simulation algorithms available: backward time, forward time, and resampling approach. The retrospective approach, also called coalescent simulation, begins with the witnessed sample in

the current generation and works backward-that is, starting with a population of individuals, this approach first traces all alleles to a singly inherited allele (Howard *et al.* 2017) [26] Howard. A coalescent framework easily incorporates population structure, demography, and recombination.

Coalescent simulations focus on constructing genealogies rooted in the MRCA under numerous evolutionary and demographic scenarios, ignoring individuals who are not linked to the MRCA (Howard *et al.* 2017) [26]. Geno-Diver is a combined coalescence simulator capable of simulating complex traits with a quantitative component and implementing multiple selections and mating strategies based on genomic information.

Table 1: Description of various simulation tools and softwares available for coalescence targeted simulation algorithms.

Software	Developers	Software	Developers
Ms	Hudson, 2002	Mlcoalsim	Ramos-Onsins and Mitchell-Olds, 2007
SNPsim	Posada and Wiuf, 2003	msHOT	Hellenthal & Stephens, 2007
SARG	Nordborg and Innan, 2003	GENOME	Liang *et al*., 2007
simcoal2	Laval and Excoffier, 2004	IBDSim	Leblois *et al*., 2008
GeneArtisan	Wang and Rannale, 2005	MaCS	Chen *et al*., 2009
CoaSim	Mailund *et al*., 2005	Msms	Ewing & Hermisson., 2010
Cosi	Schaffner *et al*., 2005	DendroPy	Sukumaran, J. and Mark T. Holder, 2010
FastCoal	Marjoram and Wall, 2006	DIYABC	Cornuet *et al*.,2014
GeneRecon	Mailund *et al*., 2006	msprime	Kelleher *et al*., 2016
Recodon	Arenas and Posada, 2007	BEAST	Suchard *et al*, 2018
DIYABC v2.0	Cornuet *et al*., 2014	msprime	Kelleher & Lohse, 2020
Discoal	Kern & Schrider, 2016	Coala	Staab & Metzler, 2016
Phylodyn	Karcher *et al*., 2017	skelesim	Parobek *et al*., 2017

Applications of Coalescence

The main parameters estimated in coalescence studies are the coalescence time and theta (Sigwart 2009). Whether coalescence has occurred in the recent past or the remote past can be determined in a rapidly growing population. Generally, we think if n is large, two individuals chosen at random from the n^{th} generation are unlikely to be closely related in a rapidly growing population, but this is not the case. The use of Y-STR profiles for forensic purposes is gaining popularity (Walsh *et al.* 2008) [71] used a coalescent approach to investigate the effect of Y-STR matching on autosomal matching, and Buckleton and Myers (2014) [81] expanded on their work. Ewing & Hermisson (2010) [82] implemented a coalescent simulation program (msms) for a structured population selected at a single diploid locus. It includes the functionality of the simulator ms to model population structure and demography but adds a model for deme- and time-dependent selection using forward simulations. It is used to investigate both hard and soft selective sweeps in structured populations and the genetic footprint of local adaptation. Walsh *et al.* (2008) [71] described a method for adjusting autosomal coancestry values to account for the presence of a Y chromosome match (Athreya 2012) [2]. Bonuso *et al.* (2009) [9] studied the factors that shape the genetic structure of T. bronchi by 61 demonstrative specimens from all over Italy (Walsh *et al.* 2008) [71]. Tuber bronchi is a significantly cherished truffle traded in Italian markets. They used phylogenetic and coalescent analyses using four loci, revealing two genetically isolated sympatrically distributed groups despite morphological differences. It is challenging to comprehend historical relationships and population evolution without employing various methods, such as coalescent theory (Kingman 1982a, 1982b; Hudson 1983) [34, 27]. It is likely to calculate previous mutation and gene flow events using this and mutation models (Bonuso *et al.* 2009) [9].

Their results showed that the possibility that European HG, South-F, Central-F, and Gurgy were sampled from a single panmictic population could not be rejected. They looked at female effective population size parameter combinations during Europe's colonization 45000 years ago, and the most recent Neolithic samples studied in this study 5900 years ago. They identified conditions under which population panmixia between hunter-gatherers/Early-Middle Neolithic farmers and Gurgy cannot be neglected. Models based on neutral theory have already been applied to a variety of systems, including river systems (Beerli and Felsenstein 1999) [5], coral reefs (Beerli 1998), birds (Dornelas *et al.* 2006, Volkov *et al.* 2007, Graves and Rahbek 2005) [6, 16, 68, 21] and tropical trees (Halley and Iwasa 2011, Hamilton 2009) [23]. Thompson, Chisholm & Rosindell (2020) [65] presented software packages for performing spatially explicit neutral simulations straightforward and effective. Portela *et al.* used coalescent theory to estimate adequate population size and gene flow to characterize the genetic variability of *A. aculeata* populations. The coalescent-based findings contributed to their understanding of *A. aculeate* population genetics and suggested that some traditional assessment methods may not be effective at characterizing historical evolutionary processes. Coalescent methods are best for studying the combined effects of demographics (seed bank, migration, extinction, and recolonization), mutation, recombination, and selection on genetic variation. The coalescing is the essence that it stimulates the genealogical history of a set of genetic data. As a result, history makes predictions about possible patterns of variation among the sample's members.

Coalescence and signatures of selection

For combining mathematical and stochastic theory with polymorphism data, the Kingman coalescent (Kingman 1982b) [34], Wright-Fisher (Fisher 1930, Wright 1931) [19, 72], and Moran (1958) models have been instrumental. The Kingman model enables us to interpret observable genetic diversity using a sample of people's population genealogy (called n-coalescent) based on present and past DNA polymorphisms, allowing for model-based inferential analysis. Neutral random processes, such as historical demographic expansion, can generate comparable patterns of nucleotidic diversity in the genome as those coming from natural selection, such as positive selection (Kingman 1982b, Tajima 1989) [34, 63]. However, demographic events are

thought to affect the entire genome, whereas selection may affect only a few sites. All genomic studies that use statistical methods relying on polymorphism data to identify natural selection like Tajima's D, Fay and Wu's H, and McDonald-Kreitman etc (Kumar *et al.,* 2021; Saravanan *et al* 2020a; Panigrahi *et al.,* 2022; Rajawat *et al.*, 2022a; Kumar *et al.* 2022, Rajawat *et al.*, 2022b) [37, 98, 50, 53, 94, 5]; Kumar *et al* or infer past demography and/or selection, rely heavily on predictions from theoretical population genetics, sometimes more particularly from the coalescent theory (Tellier and Lemaire 2014) [64].

Selection and its effect

Selection has been broadly classified into two types i.e., artificial selection and natural selection. Natural selection is caused by the environment in which people with diverse genotypes have a varying ability to contribute to the gene pool of the following generations. Natural selection can be categorized into positive selection, purifying selection, and balancing selection. Artificial selection is a human-mediated process in which the next generation's gene pool is influenced not only by fitness components but also by features chosen by people. Natural and artificial selection signatures left on the genome may be traced back and assist to comprehend the evolutionary processes influencing the genome (Rothammer *et al.* 2013, Gouveia *et al.* 2014) [57, 20]. However, when a variant (either a newly arose variant or a standing variant) provides a fitness advantage to the carrier individual in comparison to other members of the population, the carrier is more likely to thrive and leave more offspring than non-carriers, causing the frequency of the beneficial allele to increase in the population (Utsunomiya *et al.*, 2020) [66]. This leads additional related neutral variations to be carried together with the chosen variant, resulting in a selective sweep (Biswas and Akey 2006, Vitti *et al.* 2013, Gouveia *et al.* 2014, Utsunomiya *et al.* 2020) [8, 67, 20, 66]. It causes decrease in variations surrounding it, resulting in "selective signature" areas. The unique footprints due to the selection, which are found in the genome of individuals are called as selection signatures. The findings of selection signatures has the potential to reveal the identities of genes and mutations linked to critical phenotypic features in livestock species. It is also highly important to evaluate a population's levels of genetic variation, because genetic diversity is the raw material required for breeding. Due to its ability to detect genomic information connected to economic features, scanning genomic footprints has been a popular notion in recent years (Saravanan *et al*. 2021) [59]. Reference genome assembly data can be used for a variety of genomic applications, including admixture analysis (Pal *et al*., 2022) [96], breed-specific SNP panels (Kumar *et al*., 2019; Kumar *et al*., 2021b; Kumar *et al*., 2021c) [90, 91, 92], copy number variations (Kumar *et al*., 2021d) [93], rare SNP Several SNP chips have been created in recent years using various reference assemblies (Panigrahi *et al*., 2022) [50]. These SNP BeadChips have numerous uses in the field of genomics, including the investigation of various diversity parameters and the haplotype block structures of different crossbred cattle (Chhotaray *et al*., 2021b; Saravanan *et al*., 2020b; Saravanan *et al*., 2021) [88, 99, 59]. There are many ongoing works on genomics like identifying genomic breed composition (Ahmad *et al*., 2020; Saravanan *et al*. 2022a, Saravanan *et al*. 2022b; Kaisa *et al*., 2020) [85, 100, 101, 89], ancestry informative markers detection (Chhotaray *et al*. 2020) [86], and Genome-wide association study (Chhotaray *et al*., 2021a) [87].

Methods of detection of selection signatures

Advances in high-throughput sequencing and SNP genotyping technologies have greatly expanded our ability to identify selection signatures. Various statistical methods have been developed for the detection of selection signatures using the SNP data in livestock species. These methods are based on the neutral theory of molecular evolution and can be broadly classified into two groups: intra-population statistics and inter-population statistics. Intra-population statistics search for footprints of selection by comparing genomic data within populations. This group includes three primary methods based on the site frequency spectrum, LD, and reduced local variability. Tajima's D test is based on differences in the shape of lineages. The likelihood of coalescence is constant per lineage throughout time in the typical coalescent theory of genealogical branching. All genomics studies that use statistical methods for detecting natural selection approach that relies on polymorphism data (Tajima's D, Fay and Wu's H, McDonaldKreitman) or infer past demography and/or selection rely heavily on theoretical population genetics predictions, sometimes more specifically from the coalescent model. This theoretical framework, as well as the established statistical inference tools, have been widely applied to study species and population demography (Nelson *et al.* 2012) [43]. Wright (1943) [43] proposed the standard measure of population structure, F_{ST}, as the correlation between alleles in different individuals from the same subpopulation. The mean coalescence periods of alleles within subpopulations and alleles taken at random from the overall metapopulation are closely connected to F_{ST} values.

Saravanan *et al.* (2020a) have evaluated the Tharparkar cow breed's genome-wide evaluation using haplotype block structure, genetic diversity, and linkage disequilibrium. Saravanan *et al.* (2021) [59] studied and compared selection signals in Tharparkar cattle to animal genomes, in pigs Mehrotra *et al.* 2021a [95], Kumar *et al.* (2021) [37] identified Tharparkar cow breed-specific SNPs using an ovine 50K array. Using a publicly accessible collection of ROH islands, Gorssen *et al.* (2021) [83] investigated selection signatures in a variety of cattle and domestic animals. Illa *et al.* (2021) [84] discovered selection signatures in Sahiwal cattle, identifying genomic areas and candidate genes that regulate milk content and coat colour features. Saravanan *et al.* (2021) [59] studied the genetic diversity and selection signatures in three Indian sheep breeds (Changthangi, Deccani, Garole).

Conclusion

The coalescent theory is an exciting approach that has transformed our understanding of evolutionary processes (Chave and Leigh 2002) [11]. Because of its nearby association with genetic data samples, the coalescent course provides a natural framework for inferring population structure and history. The ease and effectiveness of the coalescent method under neutrality, which stems from the fact that a sample's genealogy can be modelled without reference to the rest of the population; the coalescent has become a potent inferential tool as the population grows (Condit *et al.* 2002) [12]. The relationship between population history and gene trees is complex, and conclusions can only be drawn after carefully examining the processes that generate gene trees. One of the

most important insights gained from coalescent theory is that the same population history can result in different gene trees if repeated. A very different historical scenario may occasionally produce surprising similarities in gene trees (Nielsen 2012) [44]. The coalescent theory examines drift in reverse, tracking alleles back in time to see which have a common ancestor in a previous generation (Nielsen 2012) [44]. Over time, all alleles in the present-day coalesce into fewer and fewer ancestral alleles until only the last one is left. The coalescent theory allows for many valuable inferences about the most recent common ancestor (MRCA) as well as demographic history (Relethford 2012) [54]. Due to high rates of coalescence observed during the selection phase, positive selection produces an excess of low and high-frequency variations. Coalescent theory not only allows us to infer past demography and the action of selection in the genome from polymorphism data, but it also provides a theoretical framework for developing new statistical methods and a renewed understanding of genome data in a variety of bacteria, fungi, viruses, marine organisms, and plant species with unusual life cycles (Tellier and Lemaire 2014) [64].

Ethics approval and consent to participate
Not Applicable.

Consent for publication
Not Applicable.

Availability of data and material
Not Applicable.

Funding
Not Applicable.

Authors' contributions
Sonali Sonejita Nayak wrote, reviewed, and approved the manuscript with inputs from all co-authors.

Declaration of Competing Interest
There are no conflicts of interest to declare.

Acknowledgments
All the authors of the manuscript wish to thank the Director and Joint Director (Research), ICAR-Indian Veterinary Research Institute (IVRI), Izatnagar, an institute under the Indian Council of Agricultural Research (ICAR), Department of Agricultural Research and Education (DARE), Government of India for their support and providing infrastructural facility to review this study.

References
1. Anon. Human Population Genetics and Genomics. Elsevier; c2019.
2. Athreya KB. Coalescence in the recent past in rapidly growing populations. Stochastic Processes and Their Applications. 2012;122:3757-3766.
3. Bamshad M, Wooding S P. Signatures of natural selection in the human genome. Nature Reviews Genetics. 2003;4(2):99-110.
4. Beerli P, Felsenstein J. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. Proceedings of the National Academy of Sciences of the U S A. 2001;98(8):4563-4568.
5. Beerli P, Felsenstein J. Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. Genetics. 1999;152(2):763-773.
6. Beerli P. Estimation of migration rates and population sizes in geographically structured populations. Advances in Molecular Ecology, NATO Science Series A: Life Sciences. 1998;306:39-53.
7. Berestycki N. Recent Progress in Coalescent Theory (Ensaios Mathematicos 16). Sociedade Brasileira de Matemática, Rio de Janeiro; c2009.
8. Biswas S, Akey J M. Genomic insights into positive selection. Trends in Genetics. 2006;22(8):437-446.
9. Bonuso E, Zambonelli A, Bergemann SE, Iotti M, Garbelotto M. Multilocus phylogenetic and coalescent analyses identify two cryptic species in the Italian bianchetto truffle, Tuber borchii Vittad. Conservation Genetics. 2009;11:1453-1466.
10. Charlesworth B. A hitch-hiking guide to the genome: A commentary on the hitch-hiking effect of a favourable gene' by John Maynard Smith and John Haigh. Genetics Research. 2007;89:389-390.
11. Chave J, Leigh E G Jr. A spatially explicit neutral model of beta-diversity in tropical forests. Theoretical Population Biology. 2002;62(2):153-168.
12. Condit R, Pitman N, Leigh EG Jr. Beta-diversity in tropical forest trees. Science. 2002;295(5555):666-669.
13. Daetwyler HD, Calus MP, Pong-Wong R, De Los Campos G, Hickey J M. Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. Genetics. 2013;193(2):347-365.
14. Donnelly P, Tavaré S. Coalescents and genealogical structure under neutrality. Annual Review of Genetics. 1995;29:401-421.
15. Donnelly P. The coalescent and microsatellite variability. DB Goldstein and C. Schlotter, eds. Microsatellites: evolution and applications; c1999. p. 116-128.
16. Dornelas M, Connolly SR, Hughes TP. Coral reef diversity refutes the neutral theory of biodiversity. Nature. 2006;440(7080):80-82.
17. Ewens WJ. Population genetics theory—the past and the future. Mathematical and Statistical Developments of Evolutionary Theory; c1990. p. 177-227.
18. Fearnhead P and Donnelly P. Estimating recombination rates from population genetic data. Genetics. 2006;159(3):1299-1318.
19. Fisher R A. The genetical theory of natural selection. Oxford: Clarendon; c1930.
20. Gouveia JJ de S, Silva MVGB da, Paiva SR, Oliveira SMP. Identification of selection signatures in livestock species. Genetics and Molecular Biology. 2014;37:330-342.
21. Graves GR, Rahbek C. Source pool geometry and the assembly of continental avifaunas. Proceedings of the National Academy of Sciences of the USA. 2005;102(22):7871-7876.
22. Halley JM, Iwasa Y. Neutral theory as a predictor of avifaunal extinctions after habitat loss. Proceedings of the National Academy of Sciences of the USA. 2011;108(6):2316-2321.
23. Hamilton M. Population Genetics. Wiley–Blackwell, Chichester; c2009.

24. Hein J, Schierup MH, Wiuf C. Gene Genealogies, Variation and Evolution: A Primer in Coalescent Theory. Oxford University Press. ISBN: 0-19-852996-1; c2006.
25. Holsinger KE, Weir BS. Genetics in geographically structured populations: Defining, estimating and interpreting F(ST). Nature Reviews Genetics. 2009;10(9):639-650.
26. Howard JT, Tiezzi F, Pryce JE, Maltecca C. Geno-Diver: A combined coalescence and forward-in-time simulator for populations undergoing selection for complex traits. Journal of Animal Breeding and Genetics. 2017;134(6):553-563.
27. Hudson RR. Properties of a neutral allele model with intragenic recombination. Theoretical population biology. 1983;23(2):183-201.
28. Kaplan NL, Hudson RR, Langley CH. The "hitchhiking effect" revisited. Genetics. 1989;123:887–899.
29. Kimura M. Evolutionary rate at the molecular level. Nature. 1968;217(5129): 624-626.
30. Kimura M. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. Nature. 1977;267(5608):275-276.
31. Kimura M. The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge; c1983.
32. Kingman JFC, Koch G, Spizzichino F. Exchangeability and the evolution of large populations. Exchangeability in Probability and Statistics. 97-112.
33. Kingman JFC. On the genealogy of large populations. Journal of Applied Probability. 1982a; 19(A): 27–43.
34. Kingman J F C. The coalescent. Stochastic Processes and Their Applications. 1982b;13:235–248.
35. Krone SM, Neuhauser C. Ancestral Processes with Selection. Theoretical Population Biology. 1997;51(3):210-237.
36. Kuhner MK, Yamato J, Felsenstein J. Maximum likelihood estimation of recombination rates from population data. Genetics. 2000;156(3):1393-1401.
37. Kumar H, Panigrahi M, Saravanan KA, Parida S, Bhushan B, Gaur GK, *et al*. SNPs with intermediate minor allele frequencies facilitate accurate breed assignment of Indian Tharparkar cattle. Gene. 2021a;777:145473. https://doi.org/10.1016/j.gene.2021.145473
38. Larson EJ. Evolution: The remarkable history of a scientific theory (Modern Library Ed.). New York: Modern Library; c2000; 221–243. ISBN 978-0-679-64288-6.
39. Maynard SJ, Haigh J. The hitch-hiking effect of a favourable gene. Genetics Research. 1974;23:23–35.
40. Mayr E. Darwin's influence on modern thought. Scientific American. 2000; 283(1):78-83.
41. McVean G, Awadalla P, Fearnhead P. A coalescent-based method for detecting and estimating recombination from gene sequences. Genetics. 2000;160(3):1231-1241.
42. Muneepeerakul R, Bertuzzo E, Lynch H J, Fagan W F, Rinaldo A and Rodriguez-Iturbe I. Neutral met community models predict fish diversity patterns in Mississippi-Missouri basin. nature. 2000; 453(7192):220-222.
43. Nelson MR, Wegmann D, Ehm MG. An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. Science. 2012;337:100–104.
44. Nielsen R. Molecular signatures of natural selection. Annual Review of Genetics. 2012;39:197-218.
45. Nordborg M, Balding DJ, Bishop M, Cannings C. Coalescent theory. Handbook of Statistical Genetics. 2007;3:843-877.
46. Nordborg M. Coalescent theory. Handbook of Statistical Genomics. 2019;2:145-30.
47. Nordborg M, Balding DJ, Cannings C, Bishop MJW Chichester. Coalescent theory. Handbook of Statistical Genetics; c2001.
48. Notohara M. The coalescent and the genealogical process in geographically structured population. Journal of Mathematical Biology. 1990;29(1):59-75.
49. Ostling A. Ecology: Neutral theory tested by birds. Nature. 1990; 436(7051):635-636.
50. Panigrahi M, Kumar H, Saravanan KA, Rajawat D, Nayak SS, Ghildiyal K, Dutt T. Trajectory of livestock genomics in South Asia: A comprehensive review. Gene; c2022. p. 146808.
51. Qanbari S, Simianer H. Mapping signatures of positive selection in the genome of livestock. Livestock Science. 1990;166:133-143.
52. Rajawat D, Panigrahi M, Kumar H, Nayak SS, Parida S, Bhushan B, *et al*. Identification of important genomic footprints using eight different selection signature statistics in domestic cattle breeds. Gene. 2022(a);816:146165.
53. Rajawat D, Panigrahi M, Kumar H, Nayak SS, Saravanan KA, Bhushan B, *et al*. Revealing genomic footprints of selection for fiber and production traits in three Indian sheep breeds. Journal of Natural Fibers. 2022(b);19(16):14963-14974.
54. Relethford JH. Human population genetics. John Wiley & Sons; c2012. p. 7.
55. Rosindell J, Cornell SJ. Species-area relationships from a spatially explicit neutral model in an infinite landscape. Ecology Letters. 2007;10(7):586-595.
56. Rosindell J, Wong Y, Etienne RS. A coalescence approach to spatial neutral ecology. Ecological Informatics. 2007; 3(3):259-271.
57. Rothammer S, Seichter D, Förster M, Medugorac I. A genome-wide scan for signatures of differential artificial selection in ten cattle breeds. BMC Genomics. 2013;14(908):1-14.
58. Saravanan KA, Panigrahi M, Kumar H, Bhushan B, Dutt T, Mishra BP. Genome-wide analysis of genetic diversity and selection signatures in three Indian sheep breeds. Livestock Science. 2020a;243:104367.
59. Saravanan KA, Panigrahi M, Kumar H, Parida Z, Bhushan B, Gaur GK, *et al* K. Genomic scans for selection signatures revealed candidate genes for adaptation and production traits in a variety of cattle breeds. Genomics. 2021;113:3.
60. Sigwart J. Coalescent Theory. An Introduction, Systematic Biology. 2009;58(1):162-165.
61. Slatkin M, Hudson RR. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics. 1991;129(2):555-562.
62. Slatkin M. Inbreeding coefficients and coalescent times. Genetics Research. 1991;58(2):167-175.
63. Tajima F. Evolutionary relationship of DNA sequences in finite populations. Genetics. 1983;105(2):437-460.
64. Tellier A, Lemaire C. Coalescence 2.0: A multiple branching of recent theoretical developments and their

applications. Molecular Ecology. 2014;23(11):2637-2652.
65. Thompson SE, Chisholm RA, Rosindell J. Pycoalescence and coalescence: Packages for simulating spatially explicit neutral models of biodiversity. Methods in Ecology and Evolution. 2020;11(10):1237-1246.
66. Utsunomiya YT, O'Brien AMP, Sonstegard TS, Sölkner J, Garcia JF. Genomic data as the" hitchhiker's guide" to cattle adaptation: tracking the milestones of past selection in the bovine genome. Frontiers in Genetics. 2020;6(36):1-13.
67. Vitti JJ, Grossman SR, Sabeti PC. Detecting natural selection in genomic data. Annual Review of Genetics. 2013;47:97-120.
68. Volkov I, Banavar JR, Hubbell SP, Maritan A. Patterns of relative species abundance in rainforests and coral reefs. Nature. 2007;450(7166):45-49.
69. Wakeley J. Developments in coalescent theory from single loci to chromosomes. Theoretical Population Biology. 2020;133:56-64.
70. Walsh B, Lynch M. Evolution and selection of quantitative traits. Oxford University Press; c2018.
71. Walsh B, Redd AJ, Hammer MF. Joint match probabilities for Y chromosomal and autosomal markers. Forensic Science International. 2008;174(2-3):234-238.
72. Wright S. Evolution in Mendelian Populations. Genetics. 1931;16(2):97-159.
73. Wright S. Isolation by distance. Genetics. 1943;28:114-138.
74. Wright S. The genetically structure of populations. Annals of Eugenics. 1951;15:323-354.
75. Yuan X, Miller DJ, Zhang J, Herrington D, Wang Y. An overview of population genetic data simulation. Journal of Computational Biology. 2021;19(1):42-54.
76. Zhaxybayeva O, Gogarten JP. Cladogenesis, coalescence and the evolution of the three domains of life [Published correction appears in Trends Genet. 2004 Jul;20(7):291]. Trends in Genetics. 2021;20(4):182-18.
77. Scott-Phillips TC, Laland KN, Shuker DM, Dickins TE, West SA. The niche construction perspective: A critical appraisal. Evolution. 2014;68(5):1231-1243.
78. Malécot G. Etude mathematique des populations mendeliennes. Ann. Univ. Lyon. Sci. Sec. 1941;4:45-60.
79. Kimura M. Process leading to quasi-fixation of genes in natural populations due to random fluctuation of selection intensities. Genetics. 1954 May;39(3):280.
80. Agrawal AF, Hartfield M. Coalescence with background and balancing selection in systems with bi-and uniparental reproduction: contrasting partial asexuality and selfing. Genetics. 2016 Jan 1;202(1):313-326.
81. Buckleton J, Myers S. Combining autosomal and Y chromosome match probabilities using coalescent theory. Forensic Science International: Genetics. 2014 Jul 1;11:52-5.
82. Ewing G, Hermisson J. MSMS: A coalescent simulation program including recombination, demographic structure and selection at a single locus. Bioinformatics. 2010 Aug 15;26(16):2064-2065.
83. Gorssen W, Meyermans R, Janssens S, Buys N. A publicly available repository of ROH islands reveals signatures of selection in different livestock and pet species. Genetics Selection Evolution. 2021 Dec;53:1-0.
84. Illa SK, Mukherjee S, Nath S, Mukherjee A. Genome-wide scanning for signatures of selection revealed the putative genomic regions and candidate genes controlling milk composition and coat color traits in Sahiwal cattle. Frontiers in Genetics. 2021 Jul 9;12:699422.
85. Ahmad SF, Panigrahi M, Chhotaray S, Pal D, Parida S, Bhushan B, *et al*. Revelation of genomic breed composition in a crossbred cattle of India with the help of Bovine50K BeadChip. Genomics. 2020;112:1531-1535. https://doi.org/10.1016/j.ygeno.2019.08.025
86. Chhotaray S, Panigrahi M, Pal D, Ahmad SF, Bhushan B, Gaur GK, *et al*. Ancestry informative markers derived from discriminant analysis of principal components provide important insights into the composition of crossbred cattle. Genomics. 2020;112:1726-1733. https://doi.org/10.1016/j.ygeno.2019.10.008
87. Chhotaray S, Panigrahi M, Bhushan B, Gaur GK, Dutt T, Mishra BP, *et al*. Genome-wide association study reveals genes crucial for coat color production in Vrindavani cattle. Livest. Sci. 2021a;247:104476. https://doi.org/10.1016/j.livsci.2021.104476
88. Chhotaray S, Panigrahi M, Pal D, Ahmad SF, Bhanuprakash V, Kumar H, *et al*. Genome-wide estimation of inbreeding coefficient, effective population size and haplotype blocks in Vrindavani crossbred cattle strain of India. Biol. Rhythm Res. 2021b;52(5):666-79. https://doi.org/10.1080/09291016.2019.1600266
89. Kaisa K, Kumar H, Saravanan K, Rajawat D, Bhushan B, Kumar P, *et al*. Concepts of Genomic Selection in Poultry and its Applications. Int. J Livest. Res. 2020;10(10):32-42. https://doi.org/10.5455/ijlr.20200427022022
90. Kumar H, Panigrahi M, Chhotaray S, Pal D, Bhanuprakash V, Saravanan KA, *et al*. Identification of breed-specific SNP panel in nine different cattle genomes. Biomed. Res. 2019;30(1). https://doi.org/10.35841/biomedicalresearch.30-18-1195
91. Kumar H, Panigrahi M, Chhotaray S, Parida S, Chauhan A, Bhushan B, *et al*. Comparative analysis of five different methods to design a breed-specific SNP panel for cattle. Anim. Biotechnol. 2021b;32(1):130-6. https://doi.org/10.1080/10495398.2019.1646266
92. Kumar H, Panigrahi M, Rajawat D, Panwar A, Nayak SS, Kaisa K, *et al*. Selection of breed-specific SNPs in three Indian sheep breeds using ovine 50 K array. Small Rumin. Res. 2021c;205:106545. https://doi.org/10.1016/j.smallrumres.2021.106545
93. Kumar H, Panigrahi M, Saravanan KA, Rajawat D, Parida S, Bhushan B, *et al*. Genome-wide detection of copy number variations in Tharparkar cattle. Anim. Biotechnol; c2021d. p. 1-8. https://doi.org/10.1080/10495398.2021.1942027
94. Kumar H, Panigrahi M, Panwar A, Rajawat D, Nayak SS, Saravanan KA, *et al*. Machine-Learning Prospects for Detecting Selection Signatures Using Population Genomics Data. J Comput. Biol; c2022. https://doi.org/10.1089/cmb.2021.0447
95. Mehrotra A, Bhushan B, Karthikeyan A, Singh A, Panda S, Bhati M, *et al*. Genome-wide SNP data unravel the ancestry and signatures of divergent selection in Ghurrah pigs of India. Livest Sci. 2021a;250:104587. https://doi.org/10.1016/j.livsci.2021.104587.
96. Pal D, Panigrahi M, Chhotaray S, Kumar H, Nayak SS,

Rajawat D, *et al*. Unraveling genetic admixture in the Indian crossbred cattle by different approaches using Bovine 50K BeadChip. Trop. Anim. Health Prod. 2022;54(2):135. https://doi.org/10.1007/s11250-022-03133-7

97. Saravanan KA, Panigrahi M, Kumar H, Bhushan B. Advanced software programs for the analysis of genetic diversity in livestock genomics: A mini review. Biol. Rhythm Res; c2019. p. 1-11. https://doi.org/10.1080/09291016.2019.1642650
98. Saravanan KA, Panigrahi M, Kumar H, Bhushan B, Dutt T, Mishra BP. Selection signatures in livestock genome: A review of concepts, approaches and applications. Livest. Sci. 2020a;241:104257. https://doi.org/10.1016/j.livsci.2020.104257
99. Saravanan KA, Panigrahi M, Kumar H, Parida S, Bhushan B, Gaur GK, *et al*. Genome-wide assessment of genetic diversity, linkage disequilibrium and haplotype block structure in Tharparkar cattle breed of India. Anim. Biotechnol; c2020b. p. 1-15. https://doi.org/10.1080/10495398.2020.1796696
100. Saravanan KA, Panigrahi M, Kumar H, Nayak SS, Rajawat D, Bhushan B, *et al*. Progress and future perspectives of livestock genomics in India: a mini review. Anim. Biotechnol; c2022a. https://doi.org/10.1080/10495398.2022.2056046
101. Saravanan KA, Panigrahi M, Kumar H, Rajawat D, Nayak SS, Bhushan B, *et al*. Role of genomics in combating COVID-19 pandemic. Gene. 2022b;823:146387. https://doi.org/10.1016/j.gene.2022.146387