REVIEW ARTICLE

The Concept and Application of Simulation in Population Genetics

Sonali Sonejita Nayak[1], Manjit Panigrahi[1]*, Divya Rajawat[1], Kanika Ghildiyal[1], Karan Jain[1], Anurodh Sharma[1], Nabaneeta Smaraki[2], Harsh Rajeshbhai Jogi[2], Bharat Bhushan[1]

Abstract

The process of building a model or copy of a real-world system and assessing its behaviour under various scenarios is referred to as simulation. Simulation in population genetics is the process of simulating a population's genetic makeup and evolutionary history using mathematical models and computer algorithms. It is important in population genetics for a better understanding of the impact of various evolutionary and demographic scenarios on sequence variation and patterns, it allows investigators to better assess and design analytical methods in the study of disease-associated genetic factors. This is an important tool for studying population genetic diversity and how natural selection, genetic drift, mutation, migration, and other evolutionary forces have influenced the genetic makeup of the population. There are three fundamental frameworks for simulation: coalescent, forward, and resampling methods. Numerous simulators that fit under these frameworks can be compared in terms of their evolutionary and demographic scenarios, computing complexity, and particular applications. Population simulation is becoming increasingly important in evolutionary biology, enabling researchers to explore the effects of various genetic models on genetic diversity and DNA sequence patterns.

Key words: Coalescence, Evolution, Genomics, Population Genetics, Simulation.

Ind J Vet Sci and Biotech (2023): 10.48165/ijvsbt.19.6.01

Introduction

In the field of population genetics, simulation involves using mathematical models and computer algorithms to recreate the genetic composition and evolutionary history of a population. This is a crucial tool for researching population genetic diversity, as well as how natural selection, genetic drift, mutation, migration, and other evolutionary forces have affected population genetic makeup. Population simulation is becoming increasingly significant in evolutionary biology, assisting researchers in investigating the effects of various genetic models on genetic diversity and DNA sequence patterns (Calafell *et al.*, 2001). Population size, mutation rate, recombination rate, selection coefficient, migration rate, and other variables are frequently used in population genetics simulation models. Researchers can explore how variations in a population's environment or genetic makeup impact its evolution by changing these parameters. Simulations enable the exploration of situations that are too complex to be solved analytically (Balloux *et al.,* 2001). Researchers can test theories and forecast how populations will evolve over time by using simulation models to create data that can be compared to actual genetic data. These models can also be used to examine how various genetic management techniques, such as gene editing or selective breeding, affect populations. The practice of influencing a system model through imitation is known as simulation. When a system is being simulated, its performance is simulated by creating a huge number of sampling experiments on the system's model without actually viewing the real system. When direct experimentation is not an option, Monte Carlo approaches can be helpful.

[1]Division of Animal Genetics, Indian Veterinary Research Institute, Izatnagar, Bareilly 243122, UP, India

[2]Division of Veterinary Microbiology, Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, 243122, UP, India

Corresponding Author: Dr. Manjit Panigrahi, Senior Scientist, Division of Animal Genetics, Indian Veterinary Research Institute, Izatnagar, Bareilly 243122, UP, India, Email: manjit707@gmail.com, manjit.panigrahi@icar.gov.in.

How to cite this article: Nayak, S. S., Panigrahi, M., Rajawat, D., Ghildiyal, K., Jain, K., Sharma, A., Smaraki, N., & Jogi, H. R. Bhushan, B. (2023). The Concept and Application of Simulation in Population Genetics. Ind J Vet Sci and Biotech. 19(6), 1-7.

Source of support: Nil

Conflict of interest: None

Submitted: 19/07/2023 **Accepted** 29/08/2023 **Published** 10/11/2023

Simulation refers to the process of creating a model or imitation of a real-world system or process and analysing its behaviour under different conditions or scenarios. The purpose of simulation is to gain insights into the functioning of a system and to evaluate the effects of changes to the system before implementing them in the real world. This can be used in a variety of fields, including biology, engineering, physics, economics, and social sciences. For example, engineers may use simulations to test the performance of a new design before building a physical prototype, while economists may use simulations to study the impact of different policy decisions on the economy. Simulation can be either deterministic or stochastic. Deterministic simulations use a set of predetermined rules to determine the behaviour of the system, while stochastic simulations incorporate random elements to account for uncertainties in the system. Monte Carlo simulations, which use random sampling to simulate complex systems, are a common type of stochastic simulation. One advantage of simulations is that they can be used to test hypotheses that may be difficult or impossible to test in the real world. They can also be used to study the behaviour of systems that are too complex or too dangerous to study in the real world. Overall, simulations are a powerful tool for understanding complex systems and predicting their behaviour under different conditions. As computing power continues to advance, simulations are likely to become even more important in fields ranging from science and engineering to finance and public policy.

Model Vs. Simulation

A model, which is a replica of a system of interest in the real world, is used to describe the dynamics of population genetics through mathematical equations. It is utilised to forecast changes in genetic variation over time or to calculate figures like mutation rates or effective population size. Modeling is the process of creating a model; a model is a representation of the creation and operation of a system. A simulation of a system is the execution of a system model (Maria, 1997). Running or using a scenario within that model leads to simulation. The simulation is helpful to test ideas on the influence of various genetic and demographic factors on population evolution. In population genetics research, modelling and simulation are complementary techniques. Models can be used to inform the design of simulations, and simulations can be used to test the predictions of models (Zeigler *et al.,* 2000).

Importance of simulation

In contrast to "surface learning," which merely involves memorising; simulations have the potential to engage us in "deep learning" that strengthens understanding. It enables scientists to test theories about how different evolutionary processes, such as natural selection, genetic drift, migration, and mutation affect genetic variation. It enables the investigation of intricate scenarios that would be challenging or impossible to investigate in naturally occurring populations, such as the impact of numerous interrelated influences on genetic diversity. To plan and improve genetic investigations, simulation can be used for tasks like calculating the statistical test's power or assessing the effectiveness of sampling techniques. This is used to solve mathematical puzzles that cannot be solved directly since doing so would be too expensive or time-consuming. Harris and Nielsen (2016) used SLiM to simulate ten thousand human exomes in order to investigate the impact of genetic load and Neanderthal introgression on human genetic diversity. Sanjak *et al.* (2017) used fwdpp to simulate a series of quantitative trait models under mutation-selection balance with population sizes of 2×10^4 diploids in stable populations and populations growing up to around 5×10^5 individuals, using the output to investigate the relationship between the genotype/phenotype model and GWAS outcomes.

Types of Simulations

Coalescent Simulations

Coalescent analysis follows a sample of DNA sequences' lineage backward in time until the sequences coalesce, or have a common ancestor (Nayak *et al.,* 2023a). In a coalescent simulation, scientists begin with a set of DNA sequences sampled from a population and then mimic the random events, such as mutation, recombination, and genetic drift that have an impact on the evolution of those sequences. The coalescent approach is an efficient framework for population genetic simulation. It allows us to simulate the genetic ancestry of a sample from an idealised population model, without explicitly representing the population in memory or walking through the generations (Baumdicker *et al.,* 2022). The simulation mimics the backward-in-time coalescent process, whereby two sequences combine to form a single ancestor. Many factors, including population size, mutation rate, and migration rate, have an impact on the coalescent process. By adjusting these parameters, researchers can explore how various variables affect the sample's genealogy and the genetic diversity of the population. Coalescent simulations are helpful for creating data that can be compared to actual genetic data as well as for testing ideas about the evolutionary history of populations. Coalescent theory had three transformative effects on population genetics: first, it provided researchers with better conceptual tools for describing gene trees, bringing within-population trees into sharper focus; second, it produced analytical methods for estimating parameters of interest from genetic data; and finally, it provided a computationally feasible method for producing computer simulations of population genetics processes (Kelleher *et al.,* 2018). The genetic diversity of human populations, the impact of natural selection on DNA sequences, and the genetic makeup of populations of animals and plants have all been studied using these techniques. Hudson (1983) separately calculated the coalescent in order to efficiently simulate data, and then used these simulations to characterise an analytically difficult distribution.

Coalescent simulation applications that are often used include Multiple Sequential Markov Coalescent (MS), Site Frequency Spectrum Coalescent Simulation (SFS CODE), and Do It Yourself Approximate Bayesian Computation (DIYABC). These popular programs models the evolution of populations under various demographic and genetic conditions (Hudson, 2002). SFS CODE simulates DNA sequences and family trees for sizable sample sizes and intricate demographic scenarios (Eldon *et al.,* 2015). DIYABC is a user-friendly tool that performs statistical inference on demographic parameters while simulating genealogies and genetic data (Beaumont, 2010).

Forward-Time Simulations

If evolutionary processes themselves, rather than their outcomes, are of interest (Calafell *et al.,* 2001), or if population-level features are addressed, forward simulations have a broader applicability range (Balloux and Goudet, 2002). Forward-time simulations are also more adaptable in the sense that any genetic or environmental component can be applied to a population that is evolving in the future (Peng and Kimmel, 2005). A forward simulation begins with a certain generation and allele frequencies at one or more loci. The simulator simulates allele frequencies forward in time until a given ending generation is reached.

The Resampling Method

It uses existing genetic data sets such as HapMap (http://www.hapmap.org/). A simplified version of this method generates samples via bootstrapping from existing data. Several resampling techniques are now being developed to model case-control or affected child trio datasets for power evaluations and explorations of competing genotype-phenotype association methodologies (Wright *et al.,* 2007). Resampling is usually faster than coalescent and forward techniques in terms of computational complexity. This is mostly because resampling does not necessitate modelling a complex evolutionary process, which is typically computationally demanding (Yuan *et al.,* 2012).

Other Simulation Types:

Monte Carlo simulation uses random sampling to model the behaviour of complex systems, where the inputs or parameters of the system are uncertain. It is commonly used in finance and engineering to model risk and uncertainty (Hudson, 2002). Population simulation involves creating a virtual population with known genetic parameters and simulating breeding scenarios to study the effects of selection, genetic drift, mutation, and other factors on the population's genetic diversity and fitness. Genomic simulations focus on specific genes or regions of the genome and model the inheritance of genetic variants and their effects on phenotypic traits. They are used to study the genetic basis of complex traits and to identify genomic regions associated with disease resistance or production traits. Disease simulations model the spread of infectious diseases within a population and the effectiveness of different management strategies, such as vaccination, quarantine, or culling, in controlling the disease. Environmental simulations model the effects of environmental factors, such as temperature, humidity, or feed quality, on animal health and productivity. They are used to optimize environmental conditions for animal welfare and to predict the effects of climate change on livestock production. Multi-trait simulations model the inheritance of multiple traits simultaneously and allow for the study of trade-offs between different traits and the effects of selection on genetic correlations among traits. These tools that allow researchers to study complex biological systems in a controlled, repeatable, and efficient manner, provide insights into the mechanisms of inheritance, genetic variation, and trait evolution in animals.

Simulation In Population Genetics

Simulation in population genetics is used for modeling the genetic changes that occur in populations over time. It allows researchers to test hypotheses about evolutionary processes, such as genetic drift, migration, selection, and mutation, and to generate predictions about the genetic diversity of populations. One common use of simulation in population genetics is to study the effects of natural selection on the distribution of genetic variation in a population. For example, researchers can use simulations to investigate how different types of selection, such as directional, stabilizing, or disruptive selection, affect the frequencies of different alleles and the overall genetic diversity of a population. Another important application is to study the effects of genetic drift, which is the random fluctuation of allele frequencies due to chance events. By simulating populations with different sizes and migration rates, researchers can explore how genetic drift affects the distribution of genetic variation in populations over time. Simulation can also be used to investigate the effects of gene flow, which is the movement of alleles between different populations. By simulating populations with different migration rates and patterns, researchers can study how gene flow affects the genetic diversity and structure of populations. There are numerous current projects in the field of genomics, such as detecting the genetic breed composition (Ahmad *et al.,* 2020; Kaisa *et al.,* 2020; Saravanan *et al.,* 2022a; Saravanan *et al.,* 2022b) genome-wide association research (Chhotaray *et al.,* 2021a; Chhotaray *et al.,* 2021b) and the finding of ancestry informative markers (Chhotaray *et al.,* 2020) are such examples. Additionally, there are numerous researches on selection signatures (Saravanan *et al.,* 2020; Saravanan *et al.,* 2021; Rajawat *et al.,* 2022a; Rajawat *et al.,* 2022b; Rajawat *et al.,* 2023; Nayak *et al.,* 2023b) where the use of simulation can be a possibility.

Applications

Simulation used to retrace a population's or species' evolutionary development to determine the connections between various populations and the period at which they separated. It also investigates how migration and gene flow affect genetic diversity and population structure. Simulation is used to investigate the long-term effects of drift on allele frequency. We can study how natural selection affects populations. Researchers can investigate how selection impacts the frequency of alleles and the total genetic diversity of populations by simulating populations under various selection regimes. We can go through genetic make-up of populations in relation to historical occurrences like population bottlenecks or founder effects. Simulation is an important tool used in animal genetics to understand and predict the inheritance of genetic traits and the effects of breeding programs. Animal geneticists use computer simulations to model the genetic and environmental factors that affect animal traits and to make predictions about the outcomes of breeding programs. Animal genetic simulations typically involve creating a virtual population of animals with known genetic parameters, such as allelic frequencies, heritability, and selection coefficients. The virtual population is then subjected to various breeding scenarios, such as selective breeding or inbreeding, to study the effects of these practices on the genetic diversity, health, and productivity of the population. Simulation models can also be used to test the effectiveness of different breeding strategies and to optimize breeding programs for specific goals, such as increasing milk production or improving disease resistance. They can help animal breeders make decisions about which animals to select for breeding, how to manage genetic diversity, and how to avoid negative genetic consequences, such as inbreeding depression. In addition to breeding, animal genetic simulations can also be used to study the genetics of complex traits. Overall, simulations play a critical role in animal genetics research, providing a powerful tool for predicting the outcomes of breeding programs, studying genetic and environmental factors that influence animal traits, and developing strategies for improving animal health and productivity. Today, genomics is ruling the world. Here, we would like to highlight a few significant works in the field of genomics. In addition to admixture analysis (Pal *et al.,* 2022) breed-specific SNP panels (Kumar *et al.,* 2019; Kumar *et al.,* 2021a; Kumar *et al.,* 2021b) copy number variations (Kumar *et al.,* 2021b; Kumar *et al.,* 2023) and rare SNP panels, reference genome assembly data can be used for genomics studies. Recent years have seen the development of several SNP chips employing different reference assemblies (Panigrahi *et al.,* 2022). The research of various diversity factors and the haplotype block architectures of diverse crossbred cattle are just two of the many uses for these SNP BeadChips in the field of genomics (Chhotaray *et al.,* 2021a; Saravanan *et al.,* 2020; Saravanan *et al.,* 2021).

Software used in Simulation

QuantiNemo

Using different population growth models (continuous population, logistically regulated, exponential, etc.) and dispersal models (1 D stepping stone, 2 D stepping stone, island, etc.), QuantiNemo enables the simulation of realistic population dynamics (Neuenschwander *et al.,* 2019). It is a stochastic simulation program for quantitative population genetics. It was created to look into how selection, mutation, recombination, and drift affected neutral markers and quantitative traits in migratory-connected, organized populations in diverse habitats (Neuenschwander *et al.,* 2019). One unique feature is the ability to switch between a population-based faster mode and an individual-based full-featured mode (Neuenschwander *et al.,* 2019).

QuantiNemo 2

For instance, Kanitz *et al.* (2018) recreated the global spread of humanity from East Africa. This requires running simulations for up to five million people over 132 000 years with a wide range of different parameters (migration rates, population growth rates, etc.). Only coalescence can carry out such massive simulations (Neuenschwander *et al.,* 2019).

AdmixSim 2

It is an individual programme that may successfully and adaptably model population genomics data in the context of intricate evolutionary scenarios. Based on the extended Wright-Fisher model, AdmixSim 2 incorporates a number of widely used evolutionary parameters to account for gene flow, natural selection, recombination, and mutation, allowing users to freely create and simulate any intricate scenario involving population mixing. The population size throughout generations can vary and be determined by users, in contrast to the strict WF model (Zhang *et al.,* 2021). Data of dioecious or monoecious populations, autosomes, or sex chromosomes can be simulated using AdmixSim 2 (Zhang *et al.,* 2021). The expanded WF model is the foundation of the individual-based, forward-time simulator known as AdmixSim 2. It can be used to effectively simulate one or more admixed populations under intricate demographic circumstances. AdmixSim 2 is very adaptable and can mimic admixture, recombination, mutation, and natural selection using combinations of different parameter values. Moreover, mixing situations for non-human animals and sex chromosomes can be simulated using AdmixSim 2 (Zhang *et al.,* 2021).

Genie

It is a browser-based instructional application that shows population-genetic principles such as genetic drift, population isolation, gene flow, and genetic mutation. Genie (https://cartwrig.ht/apps/genie/) is a JavaScript-based web-based stochastic simulation app (Castillo *et al.,* 2022).

Stdpopsim

It includes a catalogue of popular population genetic models for six species: *Homo sapiens, Pongo abelii, Canis familiaris, Drosophila melanogaster, Arabidopsis thaliana,* and *Escherichia coli.* The physical organisation of the genome, inferred genetic maps, population-level parameters, and demographic models are all meticulously controlled. Both coalescent and forward simulations are supported by the simulation engines, which have strict quality control for dependable execution (Tang, 2020).

Msprime 1.0

The msprime tree sequence data structure provides insight into not only genetic variation but also the ancestral ancestry that caused that variation. Recent advances in approaches for inferring genetic ancestry in recombining organisms (Kelleher *et al.,* 2019; Speidel *et al.,* 2019, Speidel *et al.,* 2021; Schaefer *et al.,* 2021; Wohns *et al.,* 2022) have allowed for the first time to estimate such ancestry from real data at scale (Harris, 2019; Tang, 2019).

SimuPOP

It is a Python-based forward-time population genetics simulation environment. Python is an 'interpreted, interactive, object-oriented, and extendable' language. SimuPOP is made up of many Python objects and functions, including populations, mating schemes, operators (objects that change populations), and simulators that help to coordinate the evolutionary processes (Peng and Kimmerl, 2005).

SIMCOAL

Researchers can examine the effects of demographic events like population bottlenecks and migrations on genetic diversity using the program SIMCOAL, which models the coalescence and migration of individual DNA sequences (Laval and Excoffier, 2004).

Slim

It allows researchers to examine the impact of natural selection, genetic drift, and mutation on genetic diversity and fitness by simulating the evolution of populations with a variety of genetic architectures (Messer, 2013).

Software used for Breeding Aspect of Simulation

Plant breeding programmes (Lin *et al.,* 2016; Gaynor *et al.,* 2017; Gorjanc *et al.,* 2018) and animal breeding programmes (Hayes and Goddard, 2003; Jenko *et al.,* 2015; Johnsson *et al.,* 2019) have both benefited from simulations, also used to address theoretical concepts in quantitative genetics and breeding (Gorjanc *et al.,* 2015). Existing software for breeding programme simulation ranges from cohort-based deterministic simulation based on expected gains, such as ZPLAN+ (Täubert *et al.,* 2010), to applications based on stochastic simulation of single individuals, such as QMSim (Sargolzaei and Schenkel, 2009) and AlphaSim (Faux *et al.,* 2016). Each of these instruments' functionality is greatly dependent on its intended application (Pook *et al.,* 2020). AlphaSimR is a highly adaptable software application that can simulate a variety of plant and animal breeding programmes for diploid and autopolyploid species. This R package is used to perform stochastic simulations of plant and animal breeding plans (Gaynor *et al.,* 2021). MoBPS is a R package that provides a computationally efficient and adaptable framework for simulating complicated breeding plans and comparing their economic and genetic consequences. A Simulation of the effect of gene editing in a cattle breeding programme (Simianer *et al.,* 2018), simulation of a multi-parent advanced generation intercross in maize (Pook *et al.,* 2019), introgression scheme in chicken (Ha *et al.,* 2017), and the generation of a base population with a hard sweep are some of the examples utilizing the above said R package.

Conclusion

In conclusion, simulation is a powerful tool in animal genetics research, providing a means to study genetic and environmental factors that influence animal traits, predict the outcomes of breeding programs, and develop strategies for improving animal health and productivity. Simulations can model complex biological systems in a controlled, repeatable, and efficient manner, allowing researchers to explore hypotheses, test breeding scenarios, and optimize management practices without the need for costly and time-consuming experimental work. As animal genetics research continues to advance, simulation will undoubtedly play an increasingly important role in helping to understand and address the challenges faced by the livestock industry, from improving production efficiency to promoting animal welfare and sustainability. Even while simulation research can never perfectly mirror reality and is based on model assumptions, it has significant advantages and helps the user to derive key conclusions. Overall, population genetic simulation is an effective method for analysing the genetic composition and evolutionary dynamics of populations. It can also be used to guide breeding, biomedical research, and conservation efforts.

References

Ahmad, S.F., Panigrahi, M., Chhotaray, S., Pal, D., Parida, S., Bhushan, B., ... & Singh, R.K. (2020). Revelation of genomic breed composition in a crossbred cattle of India with the help of Bovine50K BeadChip. *Genomics, 112*(2), 1531-1535.

Balloux, F. (2001). EASYPOP (version 1.7): a computer program for population genetics simulations. *Journal of Heredity, 92*(3), 301-302.

Balloux, F., & Goudet, J. (2002). Statistical properties of population differentiation estimators under stepwise mutation in a finite island model. *Molecular Ecology, 11*(4), 771-783.

Baumdicker, F., Bisschop, G., Goldstein, D., Gower, G., Ragsdale, A. P., Tsambos, G., ... & Kelleher, J. (2022). Efficient ancestry and mutation simulation with msprime 1.0. *Genetics, 220*(3), iyab229.

Beaumont, M.A. (2010). Approximate Bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics, 41*(1), 379-406.

Calafell, F., Grigorenko, E.L., Chikanian, A.A., & Kidd, K.K. (2001). 2001Haplotype evolution and linkage disequilibrium: A simulation study. *Human Heredity, 51*(1-2), 85-96.

Castillo, A.I., Roos, B.H., Rosenberg, M.S., Cartwright, R.A., & Wilson, M.A. (2022). Genie: an interactive real-time simulation for teaching genetic drift. *Evolution: Education and Outreach, 15*(1), 1-13.

Chhotaray, S., Panigrahi, M., Bhushan, B., Gaur, G.K., Dutt, T., Mishra, B.P., & Singh, R.K. (2021b). Genome-wide association study reveals genes crucial for coat color production in Vrindavani cattle. *Livestock Science, 247,* 104476.

Chhotaray, S., Panigrahi, M., Pal, D., Ahmad, S.F., Bhanuprakash, V., Kumar, H., ... & Singh, R.K. (2021a). Genome-wide estimation of inbreeding coefficient, effective population size and haplotype blocks in Vrindavani crossbred cattle strain of India. *Biological Rhythm Research, 52*(5), 666-679.

Chhotaray, S., Panigrahi, M., Pal, D., Ahmad, S.F., Bhushan, B., Gaur, G.K., ... & Singh, R.K. (2020). Ancestry informative markers derived from discriminant analysis of principal components provide important insights into the composition of crossbred cattle. *Genomics, 112*(2), 1726-1733.

Eldon, B., Birkner, M., Blath, J., & Freund, F. (2015). Can the site-frequency spectrum distinguish exponential population growth from multiple-merger coalescents? *Genetics, 199*(3), 841-856.

Faux, A.M., Gorjanc, G., Gaynor, R.C., Battagin, M., Edwards, S.M., Wilson, D.L., ... & Hickey, J.M. (2016). AlphaSim: software for breeding program simulation. *The Plant Genome, 9*(3), plantgenome2016-02.

Gaynor, R.C., Gorjanc, G., & Hickey, J.M. (2021). AlphaSimR: An R package for breeding program simulations. *G3, 11*(2), jkaa017.

Gaynor, R.C., Gorjanc, G., Bentley, A.R., Ober, E.S., Howell, P., Jackson, R., Mackay, I.J., & Hickey, J.M. (2017). A two-part strategy for using genomic selection to develop inbred lines. *Crop Science, 57*(5), 2372-2386.

Gorjanc, G., Cleveland, M. A., Houston, R. D., & Hickey, J. M. (2015). Potential of genotyping-by-sequencing for genomic selection in livestock populations. *Genetics Selection Evolution, 47,* 1-14.

Gorjanc, G., Gaynor, R. C., & Hickey, J. M. (2018). Optimal cross selection for long-term genetic gain in two-part programs with rapid recurrent genomic selection. *Theoretical and applied genetics, 131,* 1953-1966.ca

Ha, N.-T., Pook, T., Dierks, C., Weigend, S., Preisinger, R. (2017). A simulation approach to optimize breeding programs with application to the introgression of the blue egg color into a high performing layer line. 1*0th European Symposium on Poultry Genetics,* 26-28, June, Saint-Malo, France p. 107.

Harris, K. (2019). From a database of genomes to a forest of evolutionary trees. *Nature Genetics, 51*(9), 1306-1307.

Harris, K., & Nielsen, R. (2016). The genetic cost of Neanderthal introgression. *Genetics, 203*(2), 881-891.

Hayes, B., & Goddard, M.E. (2003). Evaluation of marker assisted selection in pig enterprises. *Livestock Production Science, 81*(2-3), 197-211.

Hudson, R.R. (1983). Testing the constant-rate neutral allele model with protein sequence data. Evolution; *International Journal of Organic Evolution, 37*(1), 203-217.

Hudson, R.R. (2002). Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics, 18*(2), 337-338.

Jenko, J., Gorjanc, G., Cleveland, M.A., Varshney, R.K., Whitelaw, C.B.A., Woolliams, J.A., & Hickey, J.M. (2015). Potential of promotion of alleles by genome editing to improve quantitative traits in livestock breeding programs. *Genetics Selection Evolution, 47*(1), 1-14.

Johnsson, M., Gaynor, R.C., Jenko, J., Gorjanc, G., De Koning, D.J., & Hickey, J.M. (2019). Removal of alleles by genome editing (RAGE) against deleterious load. *Genetics Selection Evolution, 51,* 1-18.

Kaisa, K., Kumar, H., Saravanan, K., Rajawat, D., Bhushan, B., Kumar, P., *et al*. (2020). Concepts of genomic selection in poultry and its applications. *International Journal of Livestock Research, 10*(10), 32-42.

Kanitz, R., Guillot, E. G., Antoniazza, S., Neuenschwander, S., & Goudet, J. (2018). Complex genetic patterns in human arise from a simple range-expansion model over continental landmasses. *PLoS One,* 13(2), e0192460.

Kelleher, J., Thornton, K.R., Ashander, J., & Ralph, P.L. (2018). Efficient pedigree recording for fast population genetics simulation. *PLOS Computational Biology, 14(*11), e1006581.

Kelleher, J., Wong, Y., Wohns, A.W., Fadil, C., Albers, P.K., & McVean, G. (2019). Inferring whole-genome histories in large population datasets. *Nature Genetics, 51*(9), 1330-1338.

Kumar, H., Panigrahi, M., Chhotaray, S., Pal, D.V.B., Ka, S., Shandilya, R., Parida, S., & Bhushan, B. (2019). Identification of breed-specific SNP panel in nine different cattle genomes. *Biomedical Research, 30*(1).

Kumar, H., Panigrahi, M., Chhotaray, S., Parida, S., Chauhan, A., Bhushan, B., Gaur, G.K., Mishra, B.P., & Singh, R.K. (2021a). Comparative analysis of five different methods to design a breed-specific SNP panel for cattle. *Animal Biotechnology, 32*(1), 130-136.

Kumar, H., Panigrahi, M., Rajawat, D., Panwar, A., Nayak, S.S., Kaisa, K., Bhushan, B., & Dutt, T. (2021b). Selection of breed-specific SNPs in three Indian sheep breeds using ovine 50 K array. *Small Ruminant Research, 205,* 106545.

Kumar, H., Panigrahi, M., Saravanan, K.A., Rajawat, D., Parida, S., Bhushan, B., ... & Singh, R.K. (2023). Genome-wide detection of copy number variations in Tharparkar cattle. *Animal Biotechnology, 34*(2), 448-455.

Laval, G., & Excoffier, L. (2004). SIMCOAL 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. *Bioinformatics, 20*(15), 2485-2487.

Lin, Z., Cogan, N.O., Pembleton, L.W., Spangenberg, G.C., Forster, J.W., Hayes, B.J., & Daetwyler, H.D. (2016). Genetic gain and inbreeding from genomic selection in a simulated commercial breeding program for perennial ryegrass. *The Plant Genome, 9*(1), plantgenome 2015-06.

Maria, A. (1997). Introduction to modeling and simulation. In: *Proceedings of the 29th conference on Winter simulation.* 7 - 10, December, 1997, Atlanta Georgia, USA. pp. 7-13.

Messer, P.W. (2013). SLiM: Simulating evolution with selection and linkage. *Genetics, 194*(4), 1037-1039.

Nayak, S.S., Kumar, H., Rajawat, D., Saravanan, K.A., Panwar, A., Ghildiyal, K., ... & Panigrahi, M. (2023a). Coalescence: An anti-clockwise travel. 12, (2) 2630-2639

Nayak, S.S., Panigrahi, M., Kumar, H., Rajawat, D., Sharma, A., Bhushan, B., & Dutt, T. (2023b). Evidence for selective sweeps in the MHC gene repertoire of various cattle breeds. *Animal Biotechnology,* (1-7) https://doi.org/10.1080/10495398.2023.2240387.

Neuenschwander, S., Michaud, F., & Goudet, J. (2019). QuantiNemo 2: a Swiss knife to simulate complex demographic and genetic scenarios, forward and backward in time. *Bioinformatics, 35*(5), 886-888.

Pal, D., Panigrahi, M., Chhotaray, S., Kumar, H., Nayak, S. S., Rajawat, D., Parida, S., Gaur, G. K., Dutt, T., & Bhushan, B. (2022). Unraveling genetic admixture in the Indian crossbred cattle by different approaches using Bovine 50K BeadChip. *Tropical Animal Health and Production, 54*(2), 135.

Panigrahi, M., Kumar, H., Saravanan, K.A., Rajawat, D., Sonejita Nayak, S.S., Ghildiyal, K., Kaisa, K., Parida, S., Bhushan, B., & Dutt, T. (2022). Trajectory of livestock genomics in South Asia: A comprehensive review. *Gene, 843,* 146808.

Peng, B., & Kimmel, M. (2005). simuPOP: a forward-time population genetics simulation environment. *Bioinformatics, 21*(18), 3686-3687.

Pook, T., Schlather, M., & Simianer, H. (2020). MoBPS-modular breeding program simulator. *G3, 10*(6), 1915-1918.

Pook, T., Schlather, M., de Los Campos, G., Mayer, M., Schoen, C. C., & Simianer, H. (2019). HaploBlocker: creation of subgroup-specific haplotype blocks and libraries. *Genetics, 212*(4), 1045-1061.

Rajawat, D., Panigrahi, M., Kumar, H., Nayak, S.S., Parida, S., Bhushan, B., ... & Mishra, B.P. (2022a). Identification of important genomic footprints using eight different selection signature statistics in domestic cattle breeds. *Gene, 816,* 146165.

Rajawat, D., Panigrahi, M., Kumar, H., Nayak, S.S., Saravanan, K.A., Bhushan, B., & Dutt, T. (2022b). Revealing genomic footprints of selection for fiber and production traits in three Indian sheep breeds. *Journal of Natural Fibers, 19*(16), 14963-14974.

Rajawat, D., Panigrahi, M., Nayak, S. S., Ghildiyal, K., Sharma, A., Kumar, H., ... & Dutt, T. (2023). Uncovering genes underlying coat color variation in indigenous cattle breeds through genome-wide positive selection. *Animal Biotechnology*, (1-14) https://doi.org/10.1080/10495398.2023.2240387.

Rasmussen, M.D., Hubisz, M.J., Gronau, I., & Siepel, A. (2014). Genome-wide inference of ancestral recombination graphs. *PLOS Genetics, 10*(5), e1004342.

Sanjak, J.S., Long, A.D., & Thornton, K.R. (2017). A model of compound heterozygous, loss-of-function alleles is broadly consistent with observations from complex-disease GWAS datasets. *PLOS Genetics, 13*(1), e1006573.

Saravanan, K.A., Panigrahi, M., Kumar, H., Bhushan, B., Dutt, T., & Mishra, B.P. (2020). Selection signatures in livestock genome: A review of concepts, approaches and applications. *Livestock Science, 241,* 104257.

Saravanan, K.A., Panigrahi, M., Kumar, H., Parida, S., Bhushan, B., Gaur, G.K., ... & Singh, R.K. (2022a). Genome-wide assessment of genetic diversity, linkage disequilibrium and haplotype block structure in Tharparkar cattle breed of India. *Animal Biotechnology, 33*(2), 297-311.

Saravanan, K.A., Panigrahi, M., Kumar, H., Rajawat, D., Nayak, S.S., Bhushan, B., & Dutt, T. (2022b). Role of genomics in combating COVID-19 pandemic. *Gene, 823,* 146387.

Saravanan, K.A., Panigrahi, M., Kumar, H., Parida, S., Bhushan, B., Gaur, G.K., Dutt, T., Mishra, B.P., & Singh, R.K. (2021). Genomic scans for selection signatures revealed candidate genes for adaptation and production traits in a variety of cattle breeds. *Genomics, 113*(3), 955-963.

Sargolzaei, M., & Schenkel, F.S. (2009). QMSim: a large-scale genome simulator for livestock. *Bioinformatics, 25*(5), 680-681.

Schaefer, N.K., Shapiro, B., & Green, R.E. (2021). An ancestral recombination graph of human, Neanderthal, and Denisovan genomes. *Science Advances, 7*(29), eabc0776.

Simianer, H., Pook, T., & Schlather, M. (2018). Turning the PAGE–the potential of genome editing in breeding for complex traits revisited. 2018, researchgate.net.

Speidel, L., Cassidy, L., Davies, R.W., Hellenthal, G., Skoglund, P., & Myers, S.R. (2021). Inferring population histories for ancient genomes using genome-wide genealogies. *Molecular Biology and Evolution, 38*(9), 3497-3511.

Speidel, L., Forest, M., Shi, S., & Myers, S.R. (2019). A method for genome-wide genealogy estimation for thousands of samples. *Nature Genetics, 51*(9), 1321-1329.

Tang, L. (2019). Genealogy at the genome scale. *Nature Methods, 16*(11), 1077.

Tang, L. (2020). Standardizing population genetics simulations. *Nature Methods, 17*(9), 876-876.

Täubert, H., Reinhardt, F., & Simianer, H. (2010). ZPLAN+, a new software to evaluate and optimize animal breeding programs. In *Proceedings of the 9th world congress on genetics applied to livestock production* (Vol. 1, No. 6).

Wohns, A.W., Wong, Y., Jeffery, B., Akbari, A., Mallick, S., Pinhasi, R., ... & McVean, G. (2022). A unified genealogy of modern and ancient genomes. *Science, 375*(6583), eabi8264.

Wright, F.A., Huang, H., Guan, X., Gamiel, K., Jeffries, C., Barry, W.T., ... & Zou, F. (2007). Simulating association studies: a data-based resampling method for candidate regions or whole genome scans. *Bioinformatics, 23*(19), 2581-2588.

Yuan, X., Miller, D.J., Zhang, J., Herrington, D., & Wang, Y. (2012). An overview of population genetic data simulation. J*ournal of Computational Biology, 19*(1), 42-54.

Zeigler, B.P., Kim, T.G., & Praehofer, H. (2000). *Theory of modeling and simulation.* Academic Press, Cambridge, Massachusetts, United States.

Zhang, R., Liu, C., Yuan, K., Ni, X., Pan, Y., & Xu, S. (2021). AdmixSim 2: a forward-time simulator for modeling complex population admixture. *BMC bioinformatics, 22,* 1-15.