

Computational Approaches for Assessing Genetic Diversity

Karan Sharma, Balkrishna Tiwari, Pravin Rawat, Yamini Sharma*

Genetics and Tree Improvement Division, Indian Council of Forestry Research and Education - Himalayan Forest Research Institute, Conifer Campus, Panthaghathi, Shimla, Himachal Pradesh 171013, India
Corresponding author: yamisharma1996@gmail.com

Received: November, 2025; Accepted: December, 2025; Published: January, 2026

Introduction

Genetic diversity refers to the variation in genetic material within and among populations of species (Hughes et al., 2008; Ennos et al., 2000). It encompasses the differences in DNA sequences, alleles, and genotypes that contribute to individuals' unique traits and characteristics. High genetic diversity within a population is essential for the survival and adaptability of species, as it increases their ability to cope with environmental changes, diseases, and other challenges. Conversely, low genetic diversity can lead to inbreeding and an increased risk of extinction, as populations become more susceptible to genetic disorders and less adaptable to changing conditions. Genetic diversity is a fundamental concept in evolutionary biology, conservation genetics, and biodiversity studies (Frankham et al., 2002). Understanding genetic diversity helps in monitoring the health of populations, designing conservation strategies for endangered species, and studying evolutionary processes. Moreover, it provides insights into the genetic basis of traits that are important for agriculture, medicine, and biotechnology.

Molecular markers are essential tools for assessing genetic diversity in plants. Commonly used markers include Microsatellites (SSRs), which are highly polymorphic and co-dominant; RAPD markers, which are quick and inexpensive but dominant; AFLP, which generates numerous polymorphisms across the genome; SNPs, which are abundant and useful for fine-scale variation studies; ISSRs, which are more reliable than

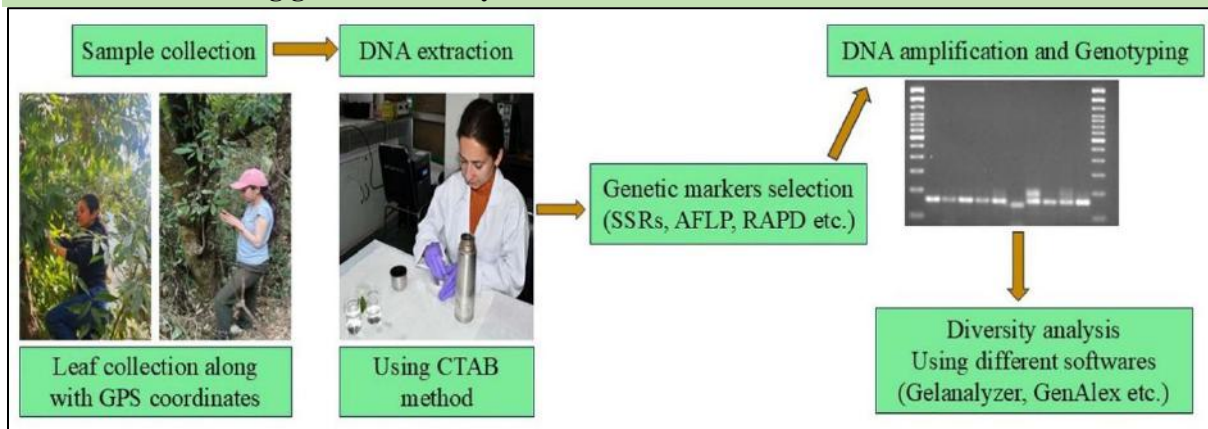
RAPDs and can assess genetic diversity effectively; and EST-SSRs, which focus on functional genomic regions. Additionally, DNA barcoding is used to distinguish species and study biodiversity. The choice of marker depends on the specific objectives, species, and resources.

Software tools for estimating genetic diversity play a crucial role in modern genetic research, allowing scientists to analyse and interpret complex genetic data with accuracy and efficiency. These tools utilize a range of statistical and computational methods to assess the genetic variation within and between populations, helping to quantify parameters such as allele frequencies, genetic distances, and heterozygosity. By integrating data from molecular markers like SSRs, SNPs, or AFLPs, these programs provide insights into the genetic structure of populations, evolutionary processes, and the effects of environmental or human-induced factors on biodiversity. Software such as GenAlex, STRUCTURE, and PAST is widely used in fields such as conservation biology, evolutionary genetics, and plant breeding, enabling researchers to assess the health and adaptability of species, design effective conservation strategies, and study the genetic basis of traits. These tools have become indispensable for understanding the genetic complexity of organisms and managing genetic resources. Therefore, this article provides an overview of various software tools used to estimate genetic diversity, highlighting their features, applications, and significance in genetic

research. These tools are essential for analysing genetic variation within and between populations, enabling researchers to gain

valuable insights into genetic structure, evolutionary processes, and biodiversity conservation.

Process for estimating genetic diversity



Software tools for diversity analysis and population structure analysis

1. **Gel analyzer software (23.1.1)** is used to estimate the molecular weight of bands in gel electrophoresis by comparing them to a known molecular weight ladder. The software helps identify the size of the molecular bands by analyzing their migration through the gel matrix. Typically, smaller molecules travel faster and thus appear lower in the gel, while larger molecules travel slower and remain higher. The software not only estimates molecular weight but also provides precise measurements, visual analysis, and automated band detection, making it an essential tool for quantifying and comparing protein or DNA samples in research and diagnostics.
2. **GenAlex software (6.503)** is a robust and advanced tool designed to estimate genetic diversity by analyzing molecular data across various populations, genotypes, and loci. To use this software, you first need to gather and input data on the molecular weights of the bands obtained from Gel analyzer software. This data should include information from different populations and genotypes across various loci. Once the data is organized into an Excel sheet, the software can analyze it to calculate key genetic diversity parameters, such as allele frequencies, heterozygosity, and genetic differentiation. GenAlex allows for the assessment of genetic variability within and between populations, helping researchers
- to better understand evolutionary processes, gene flow, and the genetic structure of populations. The software also provides various statistical outputs and graphical representations, AMOVA and Principal component analysis (PCoA) making it a valuable tool in population genetics studies.
3. **HP-rare software (1.1)** is used to estimate allelic richness, which is a measure of the number of different alleles present in a population, normalized for sample size. This software is particularly useful in population genetics studies, as it helps quantify genetic diversity by providing insights into how many distinct alleles are present across different loci. HP-rare can analyse data from multiple populations and account for variations in sample sizes, ensuring that comparisons of allelic richness are not biased by unequal sample sizes.
4. **Structure software (2.3.4)** is used to analyse population structure by identifying and estimating the genetic composition of populations, including the number of distinct genetic clusters (or subpopulations) within a dataset. It uses genetic data, typically from multiple loci, to assign individuals to different populations based on their genetic similarities. Structure can reveal the underlying patterns of genetic differentiation among populations and provide insights into gene flow, migration,

and the historical processes that shape population structure. By running simulations for various numbers of populations (K), the software helps researchers determine the most likely number of genetic clusters in the data. It is widely used in conservation genetics, evolutionary biology, and studies of human and plant genetic variation, offering graphical outputs and statistical estimates to support inferences about population connectivity and genetic diversity. Further, the best K value (the optimal number of genetic clusters) is determined using ΔK , which can be calculated through an online tool called **Structure Selector**. This tool is specifically designed to assist in selecting the most appropriate K value by analysing the results of multiple Structure runs. It uses the ΔK method, which evaluates the rate of change in the likelihood of the data between different K values, helping to identify the point where the model best fits the genetic data. Structure Selector provides an easy-to-use interface for visualizing and interpreting the results, making it a valuable resource for researchers working with Structure software. By accurately selecting the optimal K, researchers can more confidently interpret population structure and genetic differentiation in their studies.

5. **PAST (Paleontological Statistics) software (4.0.3)** is a versatile statistical tool widely used for analyzing and clustering genetic diversity data. It provides a range of statistical methods, including clustering algorithms, that help researchers group populations or individuals based on their genetic similarities. Using genetic data such as allele frequencies or molecular markers, PAST can perform various types of clustering, such as hierarchical clustering or principal component analysis (PCoA), to identify patterns of genetic variation. This makes it particularly useful for studying population structure, evolutionary relationships, and genetic

diversity across different groups or species. In addition to clustering, PAST offers a variety of other statistical tests, data visualization options, and graphing tools, making it a comprehensive software for conducting genetic diversity assessments in both ecological and evolutionary research.

6. **Darwin software (6.0.021)** is a tool used for analyzing molecular data and estimating genetic diversity in populations. It is particularly useful for handling large datasets derived from genetic markers such as microsatellites, SNPs, and other molecular techniques. Darwin software allows researchers to calculate various genetic parameters, including allele frequencies, genetic differentiation (F_{st}), and heterozygosity, making it an essential tool for population genetics studies. The software also supports various statistical analyses, such as clustering, principal component analysis (PCoA), and multivariate analysis, to assess genetic structure and relationships among populations. Its user-friendly interface and robust analytical capabilities make it popular in both basic and applied genetic research, including conservation genetics, evolutionary biology, and breeding programs.
7. **PowerMarker (3.2.5)** is a widely used, though older, software tool designed for analyzing genetic data, particularly for assessing Polymorphism Information Content (PIC). PIC is a measure of the informativeness of a genetic marker, reflecting its ability to differentiate between different genotypes within a population. PowerMarker calculates PIC values for various molecular markers, including SSRs (simple sequence repeats), SNPs, and AFLPs, helping researchers assess the diversity and variability of genetic traits. In addition to PIC, PowerMarker can also compute other genetic parameters like allele frequencies, gene diversity, and heterozygosity.

References

1. Ennos, R.A., Worrell, R., Arkle, P., & Malcolm DC. (2000). Genetic diversity and conservation. In: Genetic variation and conservation of British native trees and shrubs current knowledge and policy implications. *Forestry commission technical paper 31. Forestry Commission, Edinburgh.*
2. Frankham, R., Ballou, J.D. & Briscoe, D.A. 2002. Introduction to Conservation Genetics. *Cambridge: Cambridge University Press*, 617.
3. Hughes RA, Brian D, Inouye BD, Marc TJJ, Underwood N, Vellend M (2008) Ecological consequences of genetic diversity. *Ecol Lett*, 11, 609–623.