Plant Archives Vol. 25, Supplement 1, 2025 pp. 510-522 e-ISSN:2581-6063 (online), ISSN:0972-5210

Plant Archives

Journal homepage: http://www.plantarchives.org

DOI Url : https://doi.org/10.51470/PLANTARCHIVES.2025.v25.supplement-1.068

ADVANCES IN PHYLOGENETICS: HARNESSING EVOLUTIONARY INSIGHTS FOR CROP IMPROVEMENT

Dhanalakshmi T. N.[1]* Yash Praveen Khot[2] and Shashidhara N.[3]

[1]Department of Genetics and Plant Breeding, College of Agriculture, G.K.V.K., U.A.S., Bengaluru, Karnataka, India

[2]Department of Forestry and Environmental Sciences, College of Agriculture, G.K.V.K., U.A.S., Bengaluru, Karnataka, India

[3]Department of Genetics and Plant Breeding, University of Agricultural Sciences, Dharwad, Karnataka, India

*Corresponding author email: dhanugpb@gmail.com

(Date of Receiving : 04-08-2024; Date of acceptance : 30-09-2024)

ABSTRACT

Phylogenetics serves as a cornerstone in crop improvement, offering invaluable insights into the evolutionary relationships among crop species and guiding breeding strategies for enhanced agricultural productivity and sustainability. This literature comprehensively explores the principles, applications, and challenges of phylogenetics in the context of crop improvement. The literature delves into the construction of phylogenetic trees and the analysis of molecular data to elucidate genetic diversity, trait evolution, and adaptation in crop species. By leveraging phylogenetic approaches, researchers can identify candidate genes associated with agronomic traits, informing breeding efforts aimed at developing improved cultivars with desired characteristics. The implications of phylogenetics for crop improvement are far-reaching. By integrating phylogenetic information into breeding programs, breeders can expedite the development of cultivars with enhanced disease resistance, abiotic stress tolerance, and nutritional quality. Despite its promise, phylogenetics in crop improvement is not without challenges. Issues such as incomplete taxonomic sampling, model assumptions, and phylogenetic uncertainty require careful consideration and methodological refinement. Addressing these challenges and embracing future research directions, including interdisciplinary collaboration and data sharing, will be crucial for advancing phylogenetics in crop improvement and agricultural innovation. Phylogenetics offers a powerful framework for understanding the genetic diversity and evolutionary history of crop species, with profound implications for crop improvement and agricultural sustainability. By harnessing the evolutionary potential of crops and integrating phylogenetic principles into breeding programs, it can address global challenges in agriculture and contribute to a sustainable and resilient food system.

Keywords: Genetic diversity, Agronomic traits, Phylogenetic approaches, Comparative genomics, Phylogenetic trees

Introduction

Phylogenetics, the study of evolutionary relationships among organisms, has emerged as a powerful tool in the field of crop improvement, revolutionizing our understanding of genetic diversity, evolutionary history, and trait evolution in cultivated plants (Hillis *et al*., 1996). By elucidating the evolutionary connections between different crop species and their wild relatives, phylogenetic analyses provide invaluable insights for breeding programs aimed at enhancing crop yield, quality, and resilience to biotic and abiotic stresses.Phylogenetics, a cornerstone of evolutionary biology, encompasses a diverse array of methods and concepts aimed at reconstructing the evolutionary relationships among organisms (Huelsenbeck and Ronquist, 2001). This discipline not only seeks to unravel the branching patterns of life but also elucidates the processes underlying diversification, adaptation, and speciation. In the context of crop improvement, phylogenetics offers a framework for understanding the genetic diversity present within cultivated plants and their wild relatives, thereby informing breeding strategies and conservation efforts (Purugganananand Fuller, 2009).

Recent years have witnessed remarkable progress in phylogenetic methodology, driven in large part by advances in molecular biology and computational techniques (Rannalaand Yang, 2017). The availability of genomic data from an ever-expanding array of species has facilitated the development of sophisticated models for inferring phylogenetic trees and estimating evolutionary parameters. Bayesian inference methods, maximum likelihood estimation, and coalescent-based approaches represent just a few examples of the diverse toolkit employed by phylogeneticists to analyze genetic data (Suchard *et al*., 2018).

In addition to reconstructing evolutionary trees, modern phylogenetics encompasses a range of complementary analyses aimed at elucidating evolutionary processes and patterns. Comparative phylogenetic methods, for instance, allow researchers to test hypotheses about trait evolution, adaptation, and biogeography across multiple lineages (Revell, 2012). By integrating phylogenetic information with ecological and functional data, scientists can gain insights into the factors driving evolutionary change and diversification in plant populations. As we delve deeper into the complexities of phylogenetic inference, it becomes increasingly apparent that a multidisciplinary approach is essential for making sense of the evolutionary history of crops and their wild relatives. The integration of genomics, bioinformatics, statistics, and ecology holds the promise of unlocking new insights into the genetic basis of agronomically important traits and guiding the development of more resilient and sustainable crop varieties (Stamatakis, 2014).

This review aims to provide an overview of the latest models and methodologies in phylogenetics and explore their applications in crop improvement. It will discuss recent studies that have employed phylogenetic analyses to address key challenges in modern agriculture, including the identification of valuable genetic resources, the characterization of genetic diversity, and the development of improved cultivars resistant to pests, diseases, and environmental stresses. By synthesizing recent research findings and highlighting emerging trends in the field, it aims to provide a comprehensive overview of the role of phylogenetics in shaping the future of agriculture.

Importance of Phylogenetics in Crop Improvement

Phylogenetics plays a pivotal role in crop improvement by providing fundamental insights into the evolutionary relationships and genetic diversity of cultivated plants. Understanding the evolutionary history of crop species and their wild relatives is essential for identifying valuable genetic resources, elucidating the genetic basis of agronomically important traits, and guiding breeding efforts to develop improved cultivars with desirable characteristics (Wang *et al*., 2022).

One of the key contributions of phylogenetics to crop improvement is the identification and characterization of crop wild relatives (CWRs), which harbor unique alleles and traits that can be introgressed into cultivated crops to enhance their adaptability, productivity, and resilience to biotic and abiotic stresses (Brozynska *et al*., 2016). Phylogenetic analyses allow researchers to reconstruct the evolutionary relationships between cultivated crops and their wild relatives, facilitating the targeted exploration and utilization of genetic diversity for breeding purposes (Smýkal *et al*., 2015).

Moreover, phylogenetic approaches provide a framework for understanding the genetic architecture of complex traits and predicting the outcomes of hybridization and introgression experiments in crop breeding programs (Tanksley and McCouch, 1997). By integrating phylogenetic information with genomic data, researchers can prioritize candidate genes and genomic regions associated with target traits, accelerating the development of improved cultivars through marker-assisted selection and genomic selection strategies (Elshire *et al*., 2011).

Furthermore, phylogenetic analyses contribute to the conservation and management of plant genetic resources by guiding the establishment of germplasm collections, ex situ conservation strategies, and in situ conservation efforts (Brilhante *et al*., 2023). By identifying genetically diverse and evolutionarily distinct accessions, phylogenetics informs conservation priorities and helps ensure the long-term sustainability of crop genetic diversity for future generations (Dempewolf *et al*., 2017).

Fundamentals of Phylogenetics

Definition and Concept

Phylogenetics, a branch of biology, encompasses the study of the evolutionary relationships among organisms. At its core, phylogenetics seeks to reconstruct the evolutionary history, or phylogeny, of organisms based on shared genetic characteristics, morphological traits, or other observable features (Maddison, 2008). The fundamental concept underlying phylogenetics is the evolutionary tree, also known as a phylogenetic tree or dendrogram, which represents the branching patterns of evolutionary relationships among taxa. The construction of phylogenetic trees involves the analysis of homologous characters traits shared by different species due to common ancestry. These characters can be molecular (e.g., DNA sequences, protein sequences), morphological (e.g., anatomical structures, phenotypic traits), or behavioral (e.g., mating rituals, vocalizations). Phylogenetic methods aim to infer the most likely evolutionary relationships among taxa based on the distribution of these characters across the tree of life.

Recent advancements in sequencing technologies and computational algorithms have revolutionized the field of phylogenetics, enabling researchers to analyze large-scale molecular datasets and reconstruct highly resolved phylogenetic trees (Kumar *et al*., 2017). Phylogenomic approaches, which utilize genome-wide data, have become increasingly prevalent, providing deeper insights into evolutionary patterns and processes across diverse taxa (Hime *et al*., 2021).

Historical Development

The historical development of phylogenetics traces back to the 18thand 19thcenturies when naturalists began to classify and organize organisms based on their shared similarities. However, it was not until the mid-20thcentury that phylogenetics emerged as a formal discipline with the introduction of rigorous mathematical and computational methods for reconstructing evolutionary trees (Hillis *et al*., 1996).One of the seminal contributions to the field of phylogenetics was the development of cladistics by German entomologist Willi Hennig in the 1950s. Cladistics revolutionized the way evolutionary relationships were inferred by emphasizing the importance of shared derived characters, or synapomorphies, in defining evolutionary groups, or clades. Cladistic analysis provided a systematic framework for organizing taxa into nested hierarchies, leading to the concept of monophyly the grouping of organisms that share a common ancestor and all of its descendants.

The introduction of molecular techniques in the latter half of the 20thcentury further revolutionized phylogenetics by enabling the direct comparison of genetic material among organisms. The advent of techniques such as DNA sequencing allowed researchers to analyze molecular data to infer evolutionary relationships and construct phylogenetic trees (Sanger *et al*., 1977). The pioneering work of scientists like Emile Zuckerkandl and Linus Pauling laid the groundwork for molecular phylogenetics, demonstrating the utility of molecular sequences for reconstructing evolutionary histories (Zuckerkandl and Pauling, 1965).

Over the past few decades, phylogenetics has witnessed significant methodological advancements driven by innovations in computational algorithms, statistical models, and data analysis techniques. Phylogenetic inference methods such as maximum likelihood and Bayesian inference have become standard tools for estimating evolutionary trees and assessing the uncertainty associated with inferred relationships (Lartillot and Philippe, 2004). Recent developments in phylogenomics the study of evolutionary relationships using genome-scale data have further expanded the scope and resolution of phylogenetic analyses (Lemmon *et al*., 2012). Phylogenomic approaches leverage whole-genome sequences to infer species trees, resolve deep evolutionary relationships, and study genome evolution across diverse taxa (Brewer *et al*., 2014).

Methods and Approaches

(a) Distance-Based Methods

Distance-based methods represent one of the foundational approaches in phylogenetics, relying on the computation of pairwise genetic distances between taxa to infer evolutionary relationships. These methods are particularly useful when the evolutionary processes are assumed to be relatively simple and when the underlying model of evolution is not well understood (Saitou and Nei, 1987).

One of the most commonly used distance-based methods is the Neighbor-Joining (NJ) algorithm proposed by Naruya Saitou and Masatoshi Nei in 1987. The NJ algorithm constructs a phylogenetic tree by iteratively joining pairs of taxa based on their genetic distances, aiming to minimize the total branch length of the resulting tree (Saitou and Nei, 1987). Despite its simplicity, the Neighbor-Joining algorithm has been widely employed in various phylogenetic studies due to its computational efficiency and ability to produce accurate trees under certain conditions.Another distance-based method is the Unweighted Pair Group Method with Arithmetic Mean (UPGMA), which was introduced by Sokal and Michener in 1958. UPGMA constructs a tree by successively merging taxa into groups based on their pairwise genetic distances, with the resulting tree representing a hierarchical clustering of taxa (Sokal and Michener, 1958). While UPGMA is computationally efficient and easy to implement, it assumes a constant rate of evolution across lineages, which may not always be realistic.

Distance-based methods are also used in constructing phylogenetic networks, which represent reticulate evolutionary events such as hybridization, horizontal gene transfer, and recombination. Methods such as Neighbor-Net, implemented in software packages like SplitsTree, infer phylogenetic networks by visualizing conflicting signal in distance matrices as reticulations or cycles (Huson and Bryant, 2006).Despite their widespread use, distance-based methods have limitations, particularly in handling heterogeneous evolutionary processes and accurately representing the complexities of evolutionary relationships. Consequently, distance-based methods are often supplemented or replaced by model-based approaches, such as maximum likelihood and Bayesian inference, which can accommodate more realistic models of sequence evolution and incorporate additional sources of information.

(b) Character-based Methods

Character-based methods, also known as parsimony methods, are widely used in phylogenetics for reconstructing evolutionary trees based on discrete character data, such as the presence or absence of morphological traits or molecular sequence substitutions. These methods rely on the principle of parsimony, which posits that the most likely tree is the one that requires the fewest evolutionary changes, or character-state transitions, to explain the observed data (Felsenstein, 2004).

One of the most commonly used character-based methods is the maximum parsimony (MP) criterion, which seeks to identify the tree topology that minimizes the total number of character-state changes required to explain the observed data. MP algorithms search through the space of possible tree topologies and assign branch lengths that minimize the number of evolutionary events needed to reconcile the observed character data with the proposed phylogeny (Felsenstein, 2004).Another character-based method is the maximum compatibility (MC) criterion, which aims to find the tree that maximizes the number of compatible characters, or character-state patterns, observed across taxa (Kluge and Farris, 1969). Unlike MP, which considers all possible character-state changes equally likely, MC assigns greater weight to characters that are consistent with the inferred phylogeny and penalizes conflicting characters.

Character-based methods have several advantages, including computational efficiency, simplicity, and intuitive interpretation of results. These methods are particularly useful for analyzing morphological data or datasets with a limited number of characters, where the assumption of parsimony provides a straightforward criterion for tree inference.However, character-based methods also have limitations, such as their sensitivity to homoplasyconvergent or parallel evolution of traits—and their inability to account for the stochastic nature of evolutionary processes. Furthermore, character-based methods may struggle to accurately resolve phylogenetic relationships in cases of rapid or recent diversification, where multiple equally parsimonious trees exist.

Despite these limitations, character-based methods remain valuable tools in phylogenetic analysis, particularly when combined with other approaches such as model-based methods or Bayesian inference. By integrating information from multiple sources and employing robust statistical frameworks, researchers can obtain more accurate and reliable estimates of evolutionary relationships among taxa.

(c) Model-based Methods

Model-based methods represent a powerful approach in phylogenetics for inferring evolutionary trees by explicitly modeling the evolutionary process. These methods rely on probabilistic models of sequence evolution to estimate the likelihood of different tree topologies given the observed data. By incorporating models of nucleotide or amino acid substitution, as well as parameters such as branch lengths and substitution rates, model-based approaches aim to capture the complexities of molecular evolution and improve the accuracy of phylogenetic inference.

One of the most widely used model-based methods is maximum likelihood (ML) estimation, which seeks to find the tree topology and associated parameter values that maximize the likelihood of observing the sequence data under the assumed evolutionary model (Felsenstein, 2004). ML methods optimize the likelihood function using numerical optimization algorithms, such as the hill-climbing or Newton-Raphson methods, to search the tree space and identify the most likely tree topology and branch lengths.Another popular model-based approach is Bayesian inference, which employs Markov chain Monte Carlo (MCMC) methods to sample from the posterior distribution of trees and model parameters. Bayesian inference integrates prior knowledge or assumptions about the evolutionary process with the likelihood of the observed data to estimate the posterior probability distribution of trees. By sampling from the posterior distribution, Bayesian methods provide estimates of tree topologies, branch lengths, and other parameters along with measures of uncertainty.

Model-based methods also allow for the incorporation of more complex evolutionary models, such as those accounting for heterogeneity in substitution rates among sites (Lartillotand Philippe, 2004) or among lineages (Yang, 1994). These models, known as site-heterogeneous or branch-heterogeneous models, provide a more realistic representation of molecular evolution and can improve the accuracy of phylogenetic inference, particularly for datasets with heterogeneous substitution patterns.Recent advancements in model-based phylogenetic methods have focused on developing more sophisticated models of sequence evolution, improving computational efficiency, and accommodating larger datasets (Lemmon *et al.*, 2012). Phylogenomic analyses, which leverage genome-scale data, have also spurred the development of new model-based approaches tailored to handle the complexities of large genomic datasets.

Steps for Construction of Phylogenetic Trees

Phylogenetic tree construction involves several steps aimed at inferring the evolutionary relationships among taxa based on genetic, morphological, or other types of data. While the specific methodologies and algorithms may vary depending on the dataset and research objectives, the following steps outline a general framework for phylogenetic analysis

Data Collection and Alignment

The first step in phylogenetic tree construction is to gather relevant data, such as DNA or protein sequences, morphological characters, or behavioral traits, from the taxa of interest. For molecular data, sequences are typically obtained from genomic or transcriptomic data sources and aligned to ensure homology among characters (Katoh *et al.*, 2019).

Model Selection

Once the data are collected and aligned, researchers must choose an appropriate evolutionary model that describes the process of sequence evolution. Model selection involves assessing the fit of different substitution models to the dataset and selecting the model that best describes the evolutionary process (Posada and Crandall, 2001).

Phylogenetic Inference

The next step is to infer the phylogenetic tree topology that best explains the observed data. This can be achieved using various methods, including distance-based methods (e.g., neighbor-joining), character-based methods (e.g., maximum parsimony), and model-based methods (e.g., maximum likelihood, Bayesian inference) (Felsenstein, 2004).

Tree Estimation

Once the phylogenetic tree topology is inferred, branch lengths and other parameters of the tree must be estimated. In model-based methods, branch lengths represent the expected number of substitutions per site along each branch of the tree, while other parameters (e.g., substitution rates, nucleotide frequencies) may also be estimated.

Tree Evaluation

After the phylogenetic tree is estimated, it is essential to assess its reliability and robustness. This can be done using statistical measures such as bootstrap support values, which quantify the degree of support for individual branches of the tree based on resampling of the original data (Felsenstein, 2004).

Tree Interpretation

Finally, the inferred phylogenetic tree is interpreted in the context of the research question or hypothesis under investigation. This may involve comparing the tree topology to existing hypotheses of evolutionary relationships, identifying patterns of divergence and speciation, and inferring ancestral states of traits or characters.

Applications of Phylogenetics

Phylogenetics, the study of evolutionary relationships among organisms, has diverse applications across various fields of biology, ecology, and beyond. By elucidating the evolutionary history and relatedness of organisms, phylogenetics provides valuable insights into a wide range of biological phenomena and processes. Some of the key applications of phylogenetics include:

(a) Understanding Biodiversity

Phylogenetics plays a crucial role in understanding the patterns and processes of biodiversity across different taxa and ecosystems. By reconstructing phylogenetic trees, researchers can identify evolutionary lineages, estimate species richness, and assess the distribution of genetic diversity within and among populations (Webb *et al.*, 2002).

(b) Conservation Biology

Phylogenetics informs conservation efforts by identifying evolutionarily distinct and genetically diverse species or populations that are priorities for conservation. Phylogenetic diversity metrics, such as phylogenetic endemism and evolutionary distinctiveness, help guide conservation planning and prioritize areas for protection (Isaac *et al*., 2007).

(c) Evolutionary Biology

Phylogenetics provides a framework for studying evolutionary processes, including speciation, adaptation, and trait evolution. Comparative phylogenetic analyses enable researchers to test hypotheses about the drivers of evolutionary change, such as natural selection, genetic drift, and hybridization (Losos, 2011).

(d) Disease Ecology and Epidemiology

Phylogenetics is increasingly used to study the transmission dynamics and evolutionary history of pathogens, including viruses, bacteria, and parasites. Phylogenetic analyses of pathogen genomes help track the spread of infectious diseases, identify reservoir hosts, and inform public health interventions.

(e) Biogeography and Historical Biogeography

Phylogenetics contributes to our understanding of the historical processes that have shaped the distribution and diversity of organisms across geographic regions. By reconstructing ancestral migration routes and colonization events, phylogenetic biogeography provides insights into the factors driving species distributions and community assembly (Reeand Smith, 2008).

(f) Crop Improvement and Agriculture

Phylogenetics informs crop improvement efforts by identifying genetic resources and wild relatives that can be used to enhance crop productivity, resilience, and adaptability. By elucidating the evolutionary relationships among crop species, phylogenetics guides breeding programs aimed at developing improved cultivars with desirable traits (Brozynska *et al*., 2016).

(g) Forensic Science

Phylogenetics has applications in forensic science for identifying and tracking the sources of biological evidence, such as DNA samples from crime scenes or archaeological remains. Phylogenetic analyses can help determine relationships among individuals or populations and provide evidence for forensic investigations (Phillips and de la Puente, 2021).

Applications of Phylogenetics in Crop Improvement

Genetic Diversity Analysis

Phylogenetics plays a crucial role in crop improvement by facilitating the analysis of genetic diversity within cultivated crops and their wild relatives. Understanding the genetic diversity present in crop germplasm is essential for breeding programs aimed at developing improved cultivars with desirable traits, such as high yield, disease resistance, and abiotic stress tolerance (Brozynska *et al*., 2016). Phylogenetic analysis allows researchers to reconstruct the evolutionary relationships among different accessions or varieties of a crop species, providing insights into the genetic structure and diversity of cultivated populations. By comparing the genetic relatedness among individuals or populations, phylogenetics helps identify genetically distinct groups, assess the extent of genetic variation, and prioritize germplasm for breeding purposes (Vigouroux *et al*., 2008).Recent studies have utilized phylogenetic approaches to analyze genetic diversity in various crop species, such as rice, maize, and wheat. For example, Qiu *et al*. (2017) conducted a phylogenetic analysis of rice landraces from China and identified distinct genetic clusters corresponding to different geographic regions, highlighting the importance of local adaptation in shaping genetic diversity. Similarly, Mammadov *et al*. (2018) used phylogenetic methods to assess the genetic diversity of maize inbred lines and identify elite germplasm with desirable agronomic traits.

Phylogenetic analysis also facilitates the identification of genetic resources and wild relatives that harbor valuable traits for crop improvement. By reconstructing the evolutionary relationships between cultivated crops and their wild relatives, researchers can identify candidate genes associated with traits of interest, such as disease resistance, drought tolerance, or nutritional quality (Brozynska *et al*., 2016). This information guides the introgression of beneficial alleles from wild relatives into cultivated crops through breeding strategies such as marker-assisted selection or genomic selection (Doebley *et al*., 2006).

Germplasm characterization

Germplasm characterization involves the assessment of genetic diversity within a collection of plant materials, such as landraces, cultivars, or wild relatives, known as germplasm. Phylogenetic analysis plays a crucial role in germplasm characterization by elucidating the evolutionary relationships among different accessions and populations. Phylogenetic methods, including neighbor-joining, maximum likelihood, and Bayesian inference, are used to construct phylogenetic trees based on molecular markers or morphological traits. These trees provide insights into the genetic structure and diversity of germplasm collections, revealing patterns of relatedness among accessions and identifying genetically distinct groups (Vigouroux *et al*., 2008).

Germplasm characterization enables breeders to select diverse parental lines for hybridization and breeding programs, leading to the development of improved cultivars with enhanced genetic variability and adaptability to changing environmental conditions (Tanksley and McCouch, 1997).

Identification of genetic resources for breeding

Phylogenetic analysis aids in the identification of genetic resources and wild relatives that harbor valuable traits for crop improvement. By reconstructing the evolutionary relationships between cultivated crops and their wild relatives, researchers can pinpoint candidate genes associated with desirable agronomic traits, such as disease resistance, abiotic stress tolerance, or nutritional quality.This information guides the introgression of beneficial alleles from wild relatives into cultivated crops through breeding strategies such as marker-assisted selection or genomic selection (Brozynska *et al*., 2016). By harnessing the genetic diversity present in wild germplasm, breeders can develop cultivars with improved yield potential, quality, and resilience to biotic and abiotic stresses.

Conservation of genetic diversity

Phylogenetic analysis contributes to the conservation of genetic diversity by guiding the establishment of germplasm collections and conservation strategies for endangered or underutilized crop species and their wild relatives. By identifying genetically diverse and evolutionarily distinct accessions, phylogenetics informs conservation priorities and helps ensure the long-term sustainability of crop genetic resources for future generations (Maxted *et al*., 2011). Conservation efforts based on phylogenetic principles aim to preserve the genetic diversity of crops and their wild relatives, safeguarding against genetic erosion and promoting agricultural resilience in the face of environmental challenges.

Evolutionary Relationships and Taxonomy

Species delineation

Phylogenetics plays a pivotal role in species delineation by providing a framework for elucidating the evolutionary relationships among organisms and defining taxonomic units based on shared ancestry. Traditional species concepts, such as the Biological Species Concept (Mayr, 1942), define species as groups of interbreeding individuals reproductively isolated from other such groups. However, in practice, identifying species boundaries can be challenging, particularly in cases of cryptic species or hybridization events.

Phylogenetic analysis offers a powerful approach to species delineation by examining patterns of genetic divergence and evolutionary history. Molecular phylogenetic methods, including DNA sequencing and phylogenetic tree construction, enable researchers to assess the genetic distinctiveness of populations and infer the presence of independently evolving lineages (Sites and Marshall, 2004). Recent advancements in phylogenomics, which utilize genome-wide data to infer species trees, have enhanced our ability to delineate species boundaries and resolve taxonomic uncertainties (Edwards *et al*., 2016). Integrating multiple lines of evidence, such as molecular data, morphological characters, and ecological traits, allows for a comprehensive assessment of species diversity and facilitates the identification of cryptic species or cases of morphological convergence.

Phylogenetic classification

Phylogenetic classification refers to the organization of organisms into hierarchical taxonomic groups based on their evolutionary relationships. Traditional taxonomic systems, such as the Linnaean classification, group organisms into hierarchical categories, including kingdom, phylum, class, order, family, genus, and species. However, the Linnaean system does not explicitly reflect evolutionary history and may lead to artificial groupings based on superficial similarities.

Phylogenetic classification seeks to overcome these limitations by organizing taxa into monophyletic groups—groups that include a common ancestor and all of its descendants (de Queiroz and Gauthier, 1990). Phylogenetic trees serve as the basis for constructing a natural classification system that reflects the evolutionary history of organisms and captures their genetic relationships.Modern phylogenetic classifications integrate molecular phylogenetic data with morphological and ecological information to construct phylogenetically informative taxonomic schemes. These classifications provide a dynamic framework for understanding biodiversity and evolutionary relationships and facilitate comparative studies across different taxa.

Marker-Assisted Selection

Molecular markers for trait mapping

Marker-assisted selection (MAS) is a breeding strategy that utilizes molecular markers to facilitate the selection of individuals with desired traits. Molecular markers are DNA sequences that vary among individuals and can be linked to genes controlling specific traits of interest. By identifying molecular markers associated with target traits through trait mapping, breeders can expedite the selection process and enhance the efficiency of breeding programs.

Various types of molecular markers are used in MAS, including single nucleotide polymorphisms (SNPs), simple sequence repeats (SSRs), and insertion-deletion polymorphisms (InDels). These markers are distributed throughout the genome and can be genotyped using high-throughput sequencing technologies or polymerase chain reaction (PCR)-based assays (Collard and Mackill, 2008). Trait mapping involves identifying quantitative trait loci (QTLs) genomic regions associated with variation in a specific trait by analyzing the segregation of molecular markers in mapping populations (Xu and Crouch, 2008). QTL mapping allows breeders to pinpoint regions of the genome containing genes or regulatory elements that contribute to the expression of target traits, such as disease resistance, yield potential, or abiotic stress tolerance.

Recent advancements in genomics and bioinformatics have accelerated trait mapping efforts by enabling genome-wide association studies (GWAS) and high-resolution mapping of QTLs (Huang and Han, 2014). GWAS leverage natural variation within diverse germplasm collections to identify marker-trait associations across the entire genome, providing insights into the genetic architecture of complex traits and facilitating the development of marker-assisted breeding strategies.

Marker-assisted breeding

Marker-assisted breeding harnesses the information obtained from trait mapping to improve the efficiency and precision of breeding programs. Once molecular markers linked to target traits are identified, breeders can use these markers to select individuals with favorable alleles during the breeding process (Hospital and Charcoss *et, al*., 1997).

Marker-assisted selection allows breeders to screen large populations of segregating individuals or germplasm collections rapidly and accurately, reducing the time and resources required for phenotypic evaluation (Bernardo, 2016). By selecting for marker-trait associations early in the breeding cycle, breeders can accelerate the development of new cultivars with improved agronomic traits and streamline the breeding process.

In addition to enhancing the selection efficiency, MAS enables the introgression of target traits from exotic or wild germplasm into elite breeding lines through marker-assisted backcrossing (MABC) or marker-assisted introgression (MAI) (Hospital and Charcosset, 1997). These approaches allow breeders to transfer beneficial alleles from unadapted or genetically distant sources into elite breeding materials while minimizing the linkage drag associated with traditional breeding methods.

Genome-wide Association Studies (GWAS)

Genome-wide association studies (GWAS) have emerged as a powerful approach for identifying genetic variants associated with complex traits and diseases in diverse populations. GWAS leverage natural genetic variation across the entire genome to detect marker-trait associations, providing insights into the genetic architecture and evolutionary history of traits of interest.

Principles of GWAS

GWAS involve genotyping thousands to millions of single nucleotide polymorphisms (SNPs) or other genetic markers across the genome in large populations of individuals. By correlating genotype data with phenotype data for specific traits or diseases, researchers can identify genomic regions associated with phenotypic variation (Visscher *et al*., 2012). GWAS are based on the principle of linkage disequilibrium (LD), which describes the non-random association of alleles at different loci within a population. LD enables the detection of marker-trait associations by identifying genomic regions where genetic variants are inherited together due to their physical proximity on the same chromosome (Manolio *et al*., 2009).

Integration with phylogenetic approaches

While GWAS traditionally focus on within-species genetic variation, recent studies have integrated phylogenetic approaches to expand the scope of association mapping across diverse taxa and populations. Phylogenetically informed GWAS (phyloGWAS) leverage the evolutionary relationships among individuals or populations to account for shared ancestry and population structure when detecting marker-trait associations (Frichot*et al*., 2013). PhyloGWAS methods incorporate phylogenetic trees or evolutionary models into the association mapping framework to correct for confounding effects of population structure and relatedness (Huang *et al.*, 2019). By accounting for the evolutionary history of populations, phyloGWAS improves the accuracy and robustness of association mapping analyses, particularly in cases of admixed or structured populations.

Applications and advancements

PhyloGWAS has been applied to diverse biological systems, including plants, animals, and human populations, to dissect the genetic basis of complex traits and diseases while accounting for population history and divergence (Yang *et al.*, 2011). By integrating phylogenetic information with GWAS, researchers can identify candidate genes and genetic variants underlying trait variation and evolutionary adaptation across different taxa and environments.

Recent advancements in computational methods and statistical models have further refined phyloGWAS approaches, enabling the analysis of large-scale genomic datasets and the detection of subtle signals of selection and adaptation (Herrera and Shank, 2016). These advancements have expanded the utility of phyloGWAS for understanding the genetic basis of phenotypic diversity and evolutionary processes in natural and domesticated populations.

Phylogenetic Comparative Methods

Phylogenetic comparative methods (PCMs) are analytical approaches that integrate phylogenetic information into comparative analyses to study evolutionary patterns and processes across multiple species. PCMs enable researchers to investigate the evolution of traits, genes, and molecular sequences in a phylogenetic context, providing insights into evolutionary relationships, adaptation, and diversification.

Comparative genomics

Comparative genomics is a field that compares the genomes of different species to identify similarities and differences in gene content, organization, and evolution. PCMs in comparative genomics utilize phylogenetic trees to infer patterns of genome evolution, including gene gains and losses, gene family expansions and contractions, and genome rearrangements (Wolf *et al.*, 2002). Phylogenetic methods such as ancestral state reconstruction and molecular evolution models allow researchers to trace the evolutionary history of genes and genomic features across the tree of life (Felsenstein, 2004). Comparative genomics provides insights into the genetic basis of phenotypic diversity, evolutionary innovations, and species adaptations, informing our understanding of genome evolution and function.

Comparative transcriptomics

Comparative transcriptomics involves the comparison of gene expression profiles across different species or conditions to identify conserved and divergent patterns of gene regulation and expression. PCMs in comparative transcriptomics integrate phylogenetic information to account for shared ancestry and evolutionary relationships when analyzing gene expression data (Revell, 2012). Phylogenetic comparative methods, such as phylogenetic independent contrasts and Brownian motion models, enable researchers to test hypotheses about the evolution of gene expression traits, including rates of expression divergence, gene co-expression networks, and regulatory changes associated with adaptation and speciation (Gutenkunst *et al.*, 2009). Comparative transcriptomics provides insights into the molecular basis of phenotypic variation and evolutionary responses to environmental challenges.

Comparative proteomics

Comparative proteomics explores the diversity and evolution of protein expression patterns and functions across different species or conditions. PCMs in comparative proteomics utilize phylogenetic trees to analyze protein sequences, structures, and interactions in a phylogenetic context, revealing evolutionary trends and innovations in protein evolution (Liu *et al.*, 2016).

Phylogenetic comparative methods in proteomics allow researchers to identify conserved and lineage-specific protein features, infer ancestral protein states, and investigate the functional consequences of protein evolution (Thomas *et al.*, 2003). Comparative proteomics contributes to our understanding of protein function, adaptation, and disease mechanisms, providing insights into the evolutionary dynamics of molecular traits across diverse organisms.

Limitations of Phylogenetics

While phylogenetics is a powerful tool for understanding evolutionary relationships and processes, it is not without limitations. Some of the key limitations of phylogenetics are explained below

- Phylogenetic analyses often rely on available sequence data from a limited number of species or taxa. Incomplete taxonomic sampling can lead to biased or inaccurate reconstructions of evolutionary relationships, particularly when important lineages or key transitional forms are missing from the analysis.

- Phylogenetic methods are based on mathematical models that make simplifying assumptions about evolutionary processes, such as constant rates of molecular evolution, absence of horizontal gene transfer, and absence of recombination. Violations of these assumptions can lead to inaccurate phylogenetic reconstructions and erroneous conclusions about evolutionary history.
- Homoplasy refers to the independent evolution of similar traits or character states in different lineages, often due to convergent evolution, parallel evolution, or evolutionary reversals. Homoplastic characters can confound phylogenetic analyses by misleadingly grouping unrelated taxa together or obscuring true evolutionary relationships.
- Long-branch attraction occurs when rapidly evolving lineages (long branches) are erroneously attracted to each other in phylogenetic trees, leading to the incorrect inference of a close evolutionary relationship. Long-branch attraction is particularly problematic when analyzing distantly related taxa or when using phylogenetic markers with high substitution rates.
- Hybridization events between different species or horizontal gene transfer between distantly related lineages can result in discordant phylogenetic signals and complicate the inference of evolutionary relationships. Phylogenetic methods may fail to accurately capture the complex evolutionary histories of organisms affected by hybridization or horizontal gene transfer.
- Incomplete lineage sorting occurs when ancestral genetic polymorphisms are retained in descendant populations, leading to incongruences between gene trees and species trees. Incomplete lineage sorting can result in conflicting phylogenetic signals and complicate efforts to reconstruct accurate evolutionary relationships, particularly in rapidly diverging lineages or recent radiations.
- Phylogenetic analyses are sensitive to errors and biases in sequence data, alignment methods, and tree-building algorithms. Poor-quality data, sequence misalignment, and methodological artifacts can introduce noise and uncertainty into phylogenetic reconstructions, affecting the reliability and robustness of inferred evolutionary relationships.

Challenges and Future Directions in Phylogenetics

As phylogenetics continues to advance, several challenges and future directions emerge that shape the field's trajectory and its applications in various scientific disciplines. Addressing these challenges and embracing new opportunities will be crucial for unlocking the full potential of phylogenetic analyses and their contributions to understanding evolutionary biology, biodiversity, and beyond.

Big Data and Computational Complexity

One of the significant challenges in phylogenetics is dealing with the increasing volume and complexity of genomic data. As sequencing technologies continue to produce vast amounts of genetic information, phylogenetic analyses must grapple with the computational demands of processing, analyzing, and interpreting these data. Developing scalable algorithms, efficient computing infrastructure, and novel computational approaches will be essential for handling big data in phylogenetics effectively.

Integrating Multiple Data Types

Phylogenetic analyses often rely on different types of data, including DNA sequences, morphological traits, and ecological characteristics. Integrating multiple data types into phylogenetic reconstructions poses challenges related to data heterogeneity, model selection, and data integration. Future directions in phylogenetics involve developing robust methods for combining disparate data sources and leveraging complementary information to improve the accuracy and resolution of phylogenetic trees.

Addressing Incomplete Lineage Sorting and Gene Tree Discordance

Incomplete lineage sorting and gene tree discordance can complicate phylogenetic inference, particularly in rapidly diverging lineages and recent radiations. Overcoming these challenges requires developing sophisticated models and statistical frameworks that explicitly account for gene tree heterogeneity, population genetic processes, and complex evolutionary scenarios. Integrating coalescent-based approaches and species tree estimation methods can help reconcile conflicting signals and improve the accuracy of phylogenetic reconstructions.

Phylogenetic Uncertainty and Error Estimation

Assessing phylogenetic uncertainty and quantifying error propagation are critical aspects of robust phylogenetic inference. Future directions in phylogenetics involve developing methods for estimating uncertainty in phylogenetic trees, assessing the reliability of inferred relationships, and quantifying the impact of data quality, model assumptions, and methodological choices on phylogenetic outcomes. Bayesian approaches, bootstrapping techniques, and sensitivity analyses offer promising avenues for addressing phylogenetic uncertainty and error estimation.

Phylogenomics and Comparative Analyses

With the advent of high-throughput sequencing technologies, phylogenomics the study of evolutionary relationships using genome-scale data has emerged as a transformative approach in phylogenetics. Future directions in phylogenomics involve harnessing genomic data to reconstruct robust phylogenetic trees, infer genome-wide evolutionary patterns, and elucidate the genetic basis of phenotypic diversity. Integrating phylogenomic data with comparative analyses, functional genomics, and ecological modeling will advance our understanding of evolutionary processes and their ecological and functional consequences.

Interdisciplinary Collaboration and Data Sharing

Addressing complex biological questions in phylogenetics requires interdisciplinary collaboration and data sharing across scientific disciplines. Collaborative efforts between biologists, statisticians, computer scientists, and other experts are essential for developing innovative methods, advancing theoretical frameworks, and applying phylogenetic analyses to diverse research areas. Embracing open science principles, sharing data, code, and analytical pipelines, and fostering collaborative networks will facilitate knowledge exchange and accelerate scientific discovery in phylogenetics.

Conclusion

In this review, it is explored the fundamental principles, applications, and challenges of phylogenetics in the context of crop improvement. Phylogenetics offers valuable insights into the evolutionary relationships among crop species, facilitating the identification of genetic diversity, trait evolution, and adaptation. By reconstructing phylogenetic trees and analyzing molecular data, researchers can elucidate the genetic basis of agronomic traits, inform breeding strategies, and guide crop improvement efforts.

The integration of phylogenetics into crop improvement programs holds immense promise for enhancing agricultural productivity, sustainability, and resilience. By leveraging phylogenetic approaches, breeders can accelerate the development of improved cultivars with desired traits, such as disease resistance, abiotic stress tolerance, and nutritional quality. Future research directions in phylogenetics for crop improvement involve leveraging genomic resources, integrating multi-omics data, and embracing interdisciplinary collaborations to address key challenges in agriculture and advance breeding methodologies.

As it looks for the future, phylogenetics will continue to play a vital role in shaping the trajectory of crop improvement and agricultural innovation. By harnessing the power of evolutionary biology and genomics, we can unlock the genetic potential of crop species, mitigate the impacts of climate change, and ensure food security for future generations. With continued research efforts and collaborative initiatives, phylogenetics will empower breeders, researchers, and policymakers to address the global challenges facing agriculture and contribute to a sustainable and resilient food system.

References

Bernardo, R. (2016). Bandwagons I, too, have known. *Theoretical and Applied Genetics*, **129**, 2323-2332.

Brewer, M.S., Cotoras, D.D., Croucher, P.J. and Gillespie, R.G. (2014). New sequencing technologies, the development of genomics tools, and their applications in evolutionary arachnology. *The Journal of Arachnology*, **42(1)**, 1-15.

Brilhante, M., Catarino, S., Darbyshire, I., Bandeira, S., Moldão, M., Duarte, M. C. and Romeiras, M.M. (2023). Diversity patterns and conservation of the *Vigna* spp. in Mozambique: A comprehensive study. *Frontiers in Ecology and Evolution*, **10**, 1057785.

Brozynska, M., Furtado, A. and Henry, R.J. (2016). Genomics of crop wild relatives: expanding the gene pool for crop improvement. *Plant Biotechnology Journal*, **14(4)**, 1070-1085.

Collard, B.C. and Mackill, D.J. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **363(1491)**, 557-572.

De Queiroz, K. and Gauthier, J. 1990. Phylogeny as a central principle in taxonomy: phylogenetic definitions of taxon names. *Systematic Zoology*, **39(4)**, 307-322.

Dempewolf, H., Baute, G., Anderson, J., Kilian, B., Smith, C. and Guarino, L. (2017). Past and future use of wild relatives in crop breeding. *Crop Science*, **57(3)**, 1070-1082.

Doebley, J.F., Gaut, B.S. and Smith, B.D. 2006. The molecular genetics of crop domestication. *Cell*, **127(7)**, 1309-1321.

Edwards, S.V., Xi, Z., Janke, A., Faircloth, B.C., McCormack, J.E., Glenn, T.C. and Davis, C.C. (2016). Implementing and testing the multispecies coalescent model: a valuable paradigm for phylogenomics. *Molecular Phylogenetics and Evolution*, **94**, 447-462.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S. and Mitchell, S.E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PloS one, **6(5)**, e19379.

Felsenstein, J. (2004). Inferring phylogenies. *Sunderland,* Mass.: Sinauer. Gene, 1.

Frichot, E., Schoville, S.D., Bouchard, G. and François, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, **30(7)**, 1687-1699.

Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H. and Bustamante, C.D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, **5(10)**, e1000695.

Herrera, S. and Shank, T.M. (2016). RAD sequencing enables unprecedented phylogenetic resolution and objective species delimitation in recalcitrant divergent taxa. *Molecular Phylogenetics and Evolution*, **100**, 70-79.

Hillis, D.M., Moritz, C. and Mable, B.K. (1996). Molecular Systematics, 2nd edn. Sinauer Assoc. Inc., Sunderland, Massachusetts.

Hime, P.M., Lemmon, A.R., Lemmon, E.C.M., Prendini, E., Brown, J.M., Thomson, R.C. and Weisrock, D.W. (2021). Phylogenomics reveals ancient gene tree discordance in the amphibian tree of life. *Systematic Biology*, **70(1)**, 49-66.

Hospital, F. and Charcosset, A. (1997). Marker-assisted introgression of quantitative trait loci. *Genetics*, **147(3)**, 1469-1485.

Huang, W., Massouras, A., Inoue, Y., Peiffer, J., Ràmia, M., Tarone, A. M. and Mackay, T. F. (2014). Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines. *Genome Research*, **24(7)**, 1193-1208.

Huang, X. and Han, B. (2014). Natural variations and genome-wide association studies in crop plants. *Annual Review of Plant Biology*, **65(1)**, 531-551.

Huelsenbeck, J.P., andRonquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics, 17(8), 754-755.

Huson, D. H. and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23(2)**, 254-267.

Isaac, N.J., Turvey, S.T., Collen, B., Waterman, C. and Baillie, J.E. (2007). Mammals on the EDGE: conservation priorities based on threat and phylogeny. *PloS One*, **2(3)**, e296.

Katoh, K., Rozewicki, J. and Yamada, K.D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, **20(4)**, 1160-1166.

Kluge, A.G. and Farris, J.S. (1969). Quantitative Phyletics and the Evolution of Anurans. *Systematic Zoology*, *18(1)*, 1–32.

Kumar, S., Stecher, G., Suleski, M. and Hedges, S.B. (2017). TimeTree: a resource for timelines, timetrees, and divergence times. *Molecular Biology and Evolution*, **34(7)**, 1812-1819.

Lartillot, N. and Philippe, H. (2004). A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Molecular Biology and Evolution*, **21(6)**, 1095-1109.

Lemmon, A.R., Emme, S.A. and Lemmon, E.M. (2012). Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology*, **61(5)**, 727-744.

Liu, Y., Beyer, A. and Aebersold, R. 2016. On the dependency of cellular protein levels on mRNA abundance. *Cell*, **165(3)**, 535-550.

Losos, J.B. (2011). Convergence, adaptation, and constraint. *Evolution*, **65(7)**, 1827-1840.

Maddison, W.P. (2008). Mesquite: a modular system for evolutionary analysis. *Evolution*, **62**, 1103-1118.

Mammadov, J., Buyyarapu, R., Guttikonda, S.K., Parliament, K., Abdurakhmonov, I.Y. and Kumpatla, S.P. (2018). Wild relatives of maize, rice, cotton, and soybean: treasure troves for tolerance to biotic and abiotic stresses. *Frontiers in Plant Science*, **9,** 886.

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J. and Visscher, P.M. (2009). Finding the missing heritability of complex diseases. *Nature*, **461(7265)**, 747-753.

Maxted, N., Castañeda Álvarez, N.P., Vincent, H.A. and Magos Brehm, J. (2011). Gap analysis: a tool for genetic conservation. *Collecting Plant Genetic Diversity: Technical Guidelines*, 1-17.

Mayr, E. (1942). Systematics and the origin of species– Columbia Univ. Press, New York, 99-107.

Phillips, C. and de la Puente, M. (2021). The analysis of ancestry with small-scale forensic panels of genetic markers. *Emerging Topics in Life Sciences*, **5(3)**, 443-453.

Posada, D. and Crandall, K.A. (2001). Selecting the best-fit model of nucleotide substitution. Systematic biology, 50(4), 580-601.

Purugganan, M.D. and Fuller, D.Q. (2009). The nature of selection during plant domestication. *Nature*, **457(7231)**, 843-848.

Qiu, J., Zhou, Y., Mao, L., Ye, C., Wang, W., Zhang, J. and Lu, Y. (2017). Genomic variation associated with local adaptation of weedy rice during de-domestication. *Nature Communications*, **8(1)**, 15323.

Rannala, B. and Yang, Z. (2017). Efficient Bayesian species tree inference under the multispecies coalescent. *Systematic Biology*, **66(5)**, 823-842.

Ree, R. H. and Smith, S.A. (2008). Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology*, **57(1)**, 4-14.

Revell, L.J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, **(2)**, 217-223.

Saitou, N. and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4(4)**, 406-425.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of The National Academy of Sciences*, **74(12)**, 5463-5467.

Sites Jr, J.W. and Marshall, J.C. (2004). Operational criteria for delimiting species. *Annu. Rev. Ecol. Evol. Syst.*, **35(1)**, 199-227.

Smýkal, P., Coyne, C.J., Ambrose, M.J., Maxted, N., Schaefer, H., Blair, M.W. and Varshney, R.K. (2015). Legume crops phylogeny and genetic diversity for science and breeding. *Critical Reviews in Plant Sciences*, **34(1-3)**, 43-104.

Sokal, R.R. and Michener, C.D. (1958). A Statistical Method for Evaluating Systematic Relationships. *University of Kansas Science Bulletin*, **28(2)**, 1409–1438.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30(9)**, 1312-1313.

Suchard, M.A., Lemey, P., Baele, G., Ayres, D.L., Drummond, A.J. and Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution*, **4(1)**, vey016.

Tanksley, S.D. and McCouch, S.R. (1997). Seed banks and molecular maps: unlocking genetic potential from the wild. *Science*, **277(5329)**, 1063-1066.

Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R. and Narechania, A. (2003). PANTHER: a library of protein families and subfamilies indexed by function. *Genome Research*, **13(9)**, 2129-2141.

Vigouroux, Y., Glaubitz, J. C., Matsuoka, Y., Goodman, M.M., Sánchez G,J. and Doebley, J. (2008). Population structure and genetic diversity of New World maize races assessed by DNA microsatellites. *American Journal of Botany*, **95(10)**, 1240-1253.

Visscher, P.M., Brown, M.A., McCarthy, M.I. and Yang, J. (2012). Five years of GWAS discovery. *The American Journal of Human Genetics*, **90(1)**, 7-24.

Wang, J., Zhang, L., Wang, J., Hao, Y., Xiao, Q., Teng, J. and Wang, J. (2022). Conversion between duplicated genes generated by polyploidization contributes to the divergence of poplar and willow. *BMC Plant Biology*, **22(1)**, 298.

Webb, C.O., Ackerly, D.D., McPeek, M.A. and Donoghue, M.J. (2002). Phylogenies and community ecology. *Annual Review of Ecology and Systematics*, **33(1)**, 475-505.

Wolf, Y.I., Rogozin, I.B., Grishin, N.V. and Koonin, E.V. (2002). Genome trees and the tree of life. *TRENDS in Genetics*, **18(9)**, 472-479.

Xu, Y. and Crouch, J.H. (2008). Marker-assisted selection in plant breeding: From publications to practice. *Crop Science*, **48(2)**, 391-407.

Yang, J., Lee, S.H., Goddard, M.E. and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*, **88(1)**, 76-82.

Yang, Z. (2014). Molecular Evolution: A Statistical Approach. *Oxford University Press.*

Zuckerkandl, E. and Pauling, L. (1965). Evolutionary divergence and convergence in proteins. *In Evolving Genes and Proteins* (pp. 97-166). Academic press.