National Seminar on Applications of Mathematics in Modern Science and Technology (20th May-2026)
Organized by: S.R. Government Arts & Science College
Kothagudem, Telangana

International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

Applications of Biostatistics in Modern Botanical Research

D.Ramesh, Assistant Professor
Botany, Government Degree College (A), Paloncha

Abstract- Contemporary botanical research integrates genomic, phenotypic, and environmental data across unprecedented spatial and taxonomic scales. This paper synthesizes recent advances in biostatistical methods applied to plant science, emphasizing their interconnected use rather than isolated application. We systematically review four methodological domains: (1) phylogenetic comparative methods, including relaxed molecular clocks and network inference for divergence timing; (2) genome-wide association studies (GWAS) and quantitative trait locus (QTL) mapping, with innovations in population structure handling, meta-analysis, and machine learning; (3) adaptive evolution inference via genome-environment association (GEA) and quantitative genetic approaches; and (4) spatial statistics and species distribution modeling, including joint and deep-learning-based frameworks. Drawing on 30+ peer-reviewed studies from 2024–2026, we propose an integrative analysis pipeline that bridges these methods. This framework enables researchers to move from pattern description to causal inference in plant evolutionary ecology, biodiversity conservation, and crop breeding.

Keywords: Phylogenetics, GWAS, Adaptive Evolution, Species Distribution Models, Plant Genomics, Structural Equation Modeling

I.INTRODUCTION

The application of bio statistical methods to botanical research has accelerated dramatically over the past decade. High-throughput sequencing has made it routine to obtain genome-wide variation data for hundreds of individuals, while remote sensing and environmental databases provide spatially explicit abiotic and biotic predictors at continental scales.

However, the proliferation of methods has also created fragmentation: researchers working on phylogenetic inference may be unfamiliar with recent advances in GWAS, and vice versa. Yet the questions that drive plant science—how species arise, adapt, and respond to environmental change—demand an integrated statistical approach.

This paper has three objectives. First, we review key bio statistical applications in four interconnected domains. Second, we highlight methodological innovations from 2024–2026 that address longstanding challenges in plant data analysis. Third, we propose a unified analytical framework that connects phylogenetic, genomic, and spatial analyses, enabling researchers to move from correlative to causal inference.

International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

II. PHYLOGENETIC COMPARATIVE METHODS

Phylogenetic methods form the backbone of evolutionary botanical research. Recent advances have addressed two persistent challenges: the integration of genomic data across organizational scales and the calibration of molecular clocks with uncertain fossil records.

Unified Tree-Thinking across Biological Scales

Deng and colleagues (2025) argue that "tree thinking"—once limited to species phylogenetic—now extends to population genomics and cellular biology, revealing the genealogical structure of genetic variation within and across organisms. Ancestral recombination graphs (ARGs) now provide a framework for representing the joint history of recombination and mutation within populations, with computational advances enabling inference from whole-genome data. This unification allows researchers to simultaneously model processes operating at micro evolutionary (population) and macro evolutionary (species) scales, a critical advance for plant speciation genomics.

Divergence Time Estimation with Relaxed Molecular Clocks

The timing of angiosperm diversification remains contested. Clark and Donoghue (2025) employed relaxed molecular clock methods reflecting competing interpretations of the fossil record to show that these methods are capable of recovering an explosive diversification of flowering plants if the fossil record is interpreted confidently. Using a methodological approach that systematically varied fossil calibrations based on their claim to crown-angiosperm affinity, they estimated crown-angiosperm divergence in a Late Jurassic–Early Cretaceous interval, diminishing the "Jurassic gap" between molecular and fossil estimates. This study exemplifies how Bayesian phylogenetic methods require careful treatment of fossil calibration priors, a lesson for all plant divergence time analyses.

Complementarily, Osozawa (2025) constructed a robust, well-calibrated time tree for Spermatophyte, revealing an exponential increase in base substitution rates in recent geologic time, likely linked to Quaternary glacial cycles and C4 grass proliferation. Using BEAST v1.10.4 with alternative dating functions, the study detected both the mid-Cretaceous angiosperm radiation (at the order level) and Quaternary radiations (at the species level), demonstrating that molecular rate heterogeneity can encode biologically meaningful signals of adaptive radiation.

Phylogenetic Networks and Discordance

Beyond strictly bifurcating trees, phylogenetic networks have emerged as essential tools for plant evolution, where hybridization, incomplete lineage sorting, and horizontal gene transfer are common. A 2025 study from the Institute of Botany (CAS) integrated divergence time estimation, gene tree discordance analysis, incomplete lineage sorting assessment, and phylogenetic network inference to resolve relationships among core eudicots. Their finding that nuclear and plastid genomic data generated congruent topologies but different methods recovered different sister lineages underscores the importance of methodological pluralism.

III. GENOME-WIDE ASSOCIATION STUDIES (GWAS) AND QTL MAPPING

GWAS and QTL mapping have become standard approaches for dissecting the genetic architecture of plant traits. Recent innovations address challenges specific to plant data: polyploidy, complex population structure, genotype-by-environment interaction, and the integration of multiple trait measurements.

International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

Handling Population Structure and Interaction Effects

Hamazaki et al. (2025) introduced two novel GWAS models for detecting QTLs interacting with complex population structures, including admixed individuals whose genomes comprise chromosomal regions from different populations. Unlike previous models that require prior information on population structure, these new approaches incorporate interaction terms between SNP/haplotype blocks and genetic background directly into conventional GWAS. Applied to a soybean dataset, one model identified putative associated QTLs that conventional approaches failed to detect. Implemented in the RAINBOWR package (available on CRAN), this method is expected to help uncover complex trait architectures in diverse panels of wild and cultivated plants.

Shang et al. (2025) applied this paradigm to Handroanthus chrysanthus, a landscape tree species introduced to China, genotyping 126 germplasm samples via genotyping-by-sequencing and identifying 124,574 high-quality SNPs. GWAS using a mixed linear model identified 29 significant SNPs associated with four seed morphology traits, with narrow-sense heritability estimates allowing the identification of 68 candidate genes, four of which were validated by qRT-PCR.

Multi-Environment Trials and Meta-Analysis

A significant challenge in plant GWAS is that QTL effects are often environment-dependent. The metaGE method, introduced in PLoS Genetics (2025), provides a flexible meta-analysis approach for jointly analyzing single-environment GWAS from multi-environment trials. metaGE accounts for heterogeneity of QTL effects across environmental conditions and detects QTLs whose allelic effects correlate strongly with environmental cofactors. Applied to Arabidopsis (flowering QTLs modulated by competition) and maize (yield QTLs impacted by drought), the procedure identified known and new QTLs. Notably, metaGE drastically reduces computational burden compared to joint multi-environment analysis and is distributed as an R package.

De Walsche (2025) further developed statistical methods for meta-analysis of GWAS in plant genetics, addressing power limitations inherent in single-study GWAS through latent variable models and meta-analytic techniques. This work is particularly relevant for aggregating results across independent plant studies—a common need in comparative genomics of crop wild relatives.

LASSO-Based Methods for SNP Selection

Puthiyedth et al. (2025) addressed GWAS's limitation in missing important markers by leveraging LASSO-based regression models. Their comparative analysis of Arabidopsis thaliana showed that BIGLASSO aligns strongly with GWAS results, particularly for binary traits, while AUTALASSO complements GWAS for quantitative traits. These methods significantly enhance SNP identification and are made available through an online repository.

Digital Phenotyping and Integration

Zhang et al. (2025) integrated digital phenotyping, GWAS, and transcriptomics to identify a key gene for bud size in tea plant. Digital phenotyping of 280 tea accessions followed by comparative transcriptomics of phenotypic extremes identified CsKNOX6 as a candidate, with functional validation via heterologous transformation in Arabidopsis. This study exemplifies the power of multi-omics integration for trait dissection.

QTL Mapping for Agricultural Traits

Dynamic QTL mapping in foxtail millet revealed the genetic architecture of stem diameter across developmental stages. Using two RIL populations across five stages and two environments, the study identified 26 unconditional and 16 conditional QTLs, with qSD-9-1 consistently detected across

International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

populations and stages. Such dynamic approaches capture developmental regulation of agronomic traits.

For perennial crops, Lamoumni et al. (2025) mapped flowering date QTLs in olive, constructing high-density parental maps with >10,000 SNPs and identifying 18 significant QTLs. Candidate genes within key QTLs included WRKY71, RLT3, and FT-interacting proteins—showing convergence with flowering regulators known from model systems. Similarly, Jähne et al. (2025) identified a cold tolerance-specific QTL on soybean chromosome 11 and demonstrated that genomic prediction incorporating known QTLs as fixed effects is promising for breeding under controlled climate conditions.

IV. ADAPTIVE EVOLUTION AND GENOME-ENVIRONMENT ASSOCIATION

Understanding how plants adapt to environmental variation is central to evolutionary botany. Recent studies have moved from correlative GEA to experimental validation and trait network analysis.

Experimental Validation of GEA Candidates
Luo et al. (2025) tested 42 GEA-identified genes from Arabidopsis using T-DNA knockout lines under drought. Sixteen genes had significant effects on local adaptation traits or performance responses; notably, WRKY38 knockout lines showed decreased stomatal conductance and specific leaf area under drought, indicating adaptive drought avoidance. This study demonstrates that while only a minority of GEA candidates exhibit genotype-by-environment interactions, the approach successfully identifies genes contributing to local adaptation.

Quantitative Genetic Parameters for Evolutionary Rescue
So et al. (2025) assessed the adaptive capacity of a Québec wild mustard population under climate warming by growing 7,000 pedigreed plants under ambient and heated (+4°C) temperatures. While mean fecundity increased under heating and no significant negative cross-environment genetic correlations were detected, additive genetic variance for fitness did not increase significantly, suggesting limited evolutionary rescue potential. This study provides a rigorous biostatistical framework for predicting population persistence under climate change.

Spectral Network Analysis of Coordinated Trait Adaptation
Ray et al. (2025) combined hyperspectral reflectance data, inverse modeling, and network analysis to investigate population-level adaptation in Streptanthus tortuosus. Using partial least square discriminant analysis for population discrimination, inverse PROSPECT modeling to estimate leaf biochemical traits, and canonical correlation analysis to examine trait-climate relationships, they developed a spectral network approach treating wavelength correlations as biologically meaningful trait networks. Populations showed distinct heritable spectral signatures, with more variable environments displaying greater spectral modularity. Critically, trait-climate correlations shifted between historical (1900–1994) and recent (1995–2024) periods, consistent with ongoing climate adaptation.

Epistasis via Community Ecology Methods
Madrigal-Roca et al. (2025) applied co-occurrence tests from community ecology to identify positive and negative epistasis among 64 inversions segregating in 1,373 F2 monkeyflower plants. The centered Jaccard/Tanimoto index and affinity score described how inversions interact to generate epistasis for plant survival. Network analysis extended the traditional pairwise scope to identify third- and fourth-order interactions. This cross-disciplinary transfer of statistical methods opens new avenues for analyzing structural variant interactions in plant genomes.

International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

V. POPULATION GENOMICS AND CONSERVATION BIOSTATISTICS

Conservation genomics increasingly relies on biostatistical methods that integrate population genetics, demographic reconstruction, and spatial modeling.

Integrating Population Genomics with SDMs for Threatened Species

A 2025 study on Calycanthus chinensis, an endangered shrub in subtropical China, integrated population genomics with species distribution modeling to investigate genetic diversity and habitat suitability changes. Sequencing 75 plastomes and obtaining nuclear genome-wide SNP data, AMOVA revealed that genetic variation occurred mainly within populations, with low nucleotide diversity suggesting genetic erosion. Demographic analysis identified a bottleneck event from 0.4–0.7 Ma associated with the Wangkun glaciation. Ecological niche modeling revealed that both genetic lineages face loss of highly suitable habitats under future climate change, providing evidence-based conservation priorities.

Assessing Rarity in Tropical Trees

Montoya et al. (2025) integrated census data with genomics to assess rarity in the poorly known Neotropical tree Magnolia yantzazana. Despite relatively high nucleotide diversity, analysis revealed loss of heterozygosity, inbreeding (FIS ≥ 0.5), and demographic reconstruction showed population decline since the late Pleistocene with predicted effective population size of ~103. This genomic rarity assessment supports updating the species' conservation status to Critically Endangered—a framework applicable to thousands of data-poor tropical plant species.

New Population Genetic Diversity Estimators

Frey and Heine (2025) introduced SPrUCE (Sigmoid Pi requiring UCEs), a reference-free method estimating nucleotide diversity from aligned ultraconserved element data. SPrUCE is fast, scalable, and effective even with missing data, offering a reliable new option for conservation genetics where reference genomes are unavailable.

VI. SPATIAL STATISTICS AND SPECIES DISTRIBUTION MODELING

Species distribution models (SDMs) are essential for predicting plant responses to environmental change, yet traditional SDMs neglect biotic interactions. Recent advances address this limitation.

Joint Species Distribution Models with Remote Sensing

Hofmeyr et al. (2025) investigated whether Earth observation (EO) data improves predictions of plant community change in the Greater Cape Floristic Region. Using joint species distribution models (JSDMs) fitted with standard static environmental variables plus EO data, EO inclusion increased explanatory power for distribution models by 3% but improved abundance models by up to 30%. Critically, EO variables replaced much of the residual variance previously explained by estimated spatial latent variables, enabling more accurate predictions of composition across the landscape.

Deep Learning for Incomplete Observations

Abdelwahed et al. (2025) proposed CISO (Conditioned on Incomplete Species Observations), a deep learning-based SDM that incorporates incomplete biotic information alongside environmental variables. Using plant data from sPlotOpen (covering thousands of vegetation plots globally), CISO conditions predictions on a flexible number of species observations, accommodating the variability and incompleteness of available biotic data. Results show that including partial biotic information improves predictive performance on spatially separate test sets, and combining observations from multiple datasets further enhances accuracy.

International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

VII. STRUCTURAL EQUATION MODELING FOR TRAIT NETWORKS

Structural equation modeling (SEM) is increasingly used to disentangle direct and indirect effects among plant traits. Two 2025 studies exemplify this trend.

Suela et al. (2025) applied SEM to interpret GWAS results for soybean morphological and yield traits. Using the hill-climbing algorithm (a score-based Bayesian network learning method) to construct phenotypic networks, SEM decomposed SNP effects into direct and indirect components. They identified negative interrelationships between number of grains and hundred-grain weight, and positive relationships between number of pods and grains, and between grain weight and pod thickness. Critically, SEM revealed that many traits showed only indirect SNP effects—information unavailable from standard GWAS. In total, 46 candidate genes were identified, with 11 common to three yield-related traits.

For coffee breeding, a complementary study integrated confirmatory factor analysis, Bayesian networks, SEM, and genome-wide selection using 195 arabica coffee plants genotyped with 21,211 SNPs. CFA established latent variables for vigor, nodes, leaf length, and yield. SEM revealed causal pathways: leaf length influences vigor and node number, which in turn influence yield. Importantly, using latent variable node number to predict yield increased selection gains by 66.35% compared to using yield directly, demonstrating that SEM-informed breeding strategies outperform naive multi-trait selection.

VIII. DISCUSSION: TOWARD AN INTEGRATED ANALYTICAL FRAMEWORK

The methods reviewed above are often applied in isolation. However, the most powerful botanical research will integrate these approaches. We propose a three-stage integrated framework:

Stage 1: Evolutionary Context. Use phylogenetic comparative methods (relaxed molecular clocks, ARG inference, and network analysis) to establish the evolutionary relationships and divergence times of study taxa. This provides the historical framework for all downstream analyses.

Stage 2: Genetic Architecture. Conduct GWAS or QTL mapping using methods that account for population structure (RAINBOWR) and environment (metaGE). For multi-trait systems, apply SEM-GWAS to decompose direct and indirect SNP effects. For trait networks, use spectral or Bayesian network approaches.

Stage 3: Spatial-Environmental Inference. Project results onto landscapes using SDMs (JSDMs with EO data, CISO for incomplete biotic data) to predict responses to environmental change. Integrate population genomics with niche modeling for conservation prioritization.

Several 2025 studies have begun such integration. The Calycanthus chinensis conservation study combined population genomics, phylogenomics, and SDMs. The soybean SEM-GWAS integrated network learning and association mapping. The tea bud size study integrated digital phenotyping, GWAS, and transcriptomics. The path forward lies in further methodological integration.

International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

IX. CONCLUSION

Biostatistics has transformed botanical research from descriptive to predictive. Relaxed molecular clocks reveal the tempo of angiosperm evolution. GWAS methods that account for population structure and environment uncover the genetic basis of adaptive traits. Experimental validation of GEA candidates moves from correlation to causation. Joint species distribution models incorporating remote sensing and deep learning enable landscape-scale predictions. Structural equation modeling reveals causal networks among traits, guiding breeding.

The synthesis presented here demonstrates that these methods are not alternatives but complements. Phylogenetic history constrains genetic architecture. Genetic architecture determines adaptive potential. Adaptive potential, projected through spatial environmental models, predicts conservation outcomes. By adopting an integrated biostatistical framework, plant scientists can address the grand challenges of the Anthropocene: feeding a growing population, preserving biodiversity, and understanding the evolutionary processes that sustain life on Earth.

REFERENCES

1. Y. Deng et al. (2025). Tree Thinking in the Genomic Era: Unifying Models Across Cells, Populations, and Species. arXiv:2512.05499. (https: //arxiv.org/abs/2512.05499v1)
2. J.W. Clark & P.C.J. Donoghue (2025). Uncertainty in the timing of diversification of flowering plants rests with equivocal interpretation of their fossil record. Royal Society Open Science, 12(5), 242158.(https:/ /pmc.ncbi.nlm.nih.gov/articles/PMC12115813/)
3. S. Osozawa (2025). Spermatophyta Molecular Clock: Time Drift and Recent Acceleration. Plant-Environment Interactions, 6(5), e70084. (https: //pubmed.ncbi.nlm.nih.gov/40978090/)
4. K. Hamazaki et al. (2025). A novel genome-wide association study method for detecting quantitative trait loci interacting with complex population structures in plant genetics. Genetics, iyaf038. (https:/ /pubmed.ncbi.nlm.nih.gov/40091626/)
5. X. Shang et al. (2025). Population structure, genetic diversity and genome-wide association analysis of the seed morphology traits in Handroanthus chrysanthus. BMC Plant Biology (via AGRIS). (https: //agris.fao.org/search/en/records/690c994ce36ca62843606f78)
6. metaGE authors (2025). metaGE: Investigating genotype × environment interactions through GWAS meta-analysis. PLoS Genetics, 21(1), e1011553. (https:// pmc.ncbi.nlm.nih.gov/articles/PMC11756807/)
7. A. De Walsche (2025). Development of statistical methods for meta-analysis of genome-wide association studies, applications in plant genetics. PhD thesis, Université Paris-Saclay. (https: //theses.hal.science/tel-05073179v1)
8. N. Puthiyedth et al. (2025). Leveraging LASSO-based methodologies for enhanced SNP analysis in plant genomes. Bioinformatics Advances, 5(1), vbaf014. (https:/ /pubmed.ncbi.nlm.nih.gov/40092525/)
9. S. Zhang et al. (2025). Integration of digital phenotyping, GWAS, and transcriptomic analysis revealed a key gene for bud size in tea plant. Horticulture Research, 12(6), uhaf051. (https: //pubmed.ncbi.nlm.nih.gov/40271457/)
10. Dynamic QTL authors (2025). Dynamic QTL mapping reveals the genetic architecture of stem diameter across developmental stages in foxtail millet. Planta, 261(4), 70.
11. O. Lamoumni et al. (2025). Unraveling the genetic basis of full flowering date in olive tree through QTL mapping. HAL Preprint. (https: //hal.inrae.fr/hal-05207778v1)
12. F. Jähne et al. (2025). Cold stress tolerance of soybeans during flowering: QTL mapping and efficient selection strategies. Plant Breeding, 138(6), 708–720. (https: //agris.fao.org/search/es/records/67a09eed20478411b025a1ee)

International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

13. Y. Luo et al. (2025). Experimental Validation of Genome-Environment Associations in Arabidopsis. Molecular Ecology, 34(21), e70129. [https ://pubmed.ncbi.nlm.nih.gov/41054271/](https //pubmed.ncbi.nlm.nih.gov/41054271/)
14. C.P. So et al. (2025). The capacity for adaptation to climate warming in a naturalized annual plant (Brassica rapa). Evolution, qpaf187. (https ://pubmed.ncbi.nlm.nih.gov/40982227/)
15. R. Ray et al. (2025). Spectral network analysis illuminates coordinated trait adaptation across plant populations. bioRxiv. (https: //www.biorxiv.org/content/10.1101/2025.09.18.676927v1)
16. L.J. Madrigal-Roca et al. (2025). Are you with me? Co-occurrence tests from community ecology can identify positive and negative epistasis between inversions in Mimulus guttatus. PLoS ONE, 20(4), e0321253. [(https:/ /pubmed.ncbi.nlm.nih.gov/40294049/)
17. Integrative population genomics study (2025). Calycanthus chinensis conservation. Global Ecology and Conservation, 60, e03614. (https:/ /library.cnu.ac.kr/eds/detail/edsdoj_edsdoj.6374bf8186a64462aad937a3504a2862)
18. S.J. Montoya et al. (2025). Assessing rarity: genomic insights for population assessments and conservation of the most poorly known Neotropical trees. Biological Conservation, 309, 111280.(https:/ / www.sciencedirect.com/science/article/pii/S0006320725003179)
19. S.J. Frey & J. Heine (2025). SPrUCE: Utilizing Ultraconserved Elements of DNA for Population-Level Genetic Diversity Estimation. bioRxiv.(https://www.biorxiv.org/content/10.1101/2025.11.16.664130v1)
20. M. Hofmeyr et al. (2025). Predicting plant community change using satellite remote sensing in the Greater Cape Floristic Region. South African Journal of Botany, 186, 238–250. [(https //www.sciencedirect.com/science/article/pii/S0254629925005058)
21. H.R. Abdelwahed et al. (2025). CISO: Species Distribution Modeling Conditioned on Incomplete Species Observations. arXiv:2508.06704. (https ://arxiv.org/abs/2508.06704)
22. M.M. Suela et al. (2025). Using Structural Equation Models to Interpret Genome-Wide Association Studies for Morphological and Productive Traits in Soybean. Plants, 14(19), 3015. (https:// www.mdpi.com/2223-7747/14/19/3015)
23. Coffee SEM-GWS study (2025). Agronomy, 15(7), 1686. [(https:/ /www.mdpi.com/2073-4395/15/7/1686)