Integration of AI and Multi-Omics Data in Plant Genetics

Ali Murtaza*[1], Shaiza Rasool[2], Muhammad Awais Tariq[3], Saud Hassan[1] and Muhammad Ahad[1]

1. Department of Plant Breeding and Genetics, University of Agriculture Faisalabad, Pakistan
2. Department of Biotechnology, University of Mainwali, Pakistan
3. Department of Computer Science, COMSATS University, Pakistan

*Corresponding Author: **mralimurtazarai@gmail.com**

ABSTRACT

The growing demand for resilient crops, climate change and the availability of high-throughput sequencing and sensor technologies have created an unprecedented opportunity for data-driven crop improvement. Multi-omics technologies (genomics, transcriptomics, proteomics, metabolomics, epigenomics and phenomics) capture different layers of biological information, yet analysing each layer separately loses the holistic view of how these molecules collectively shape phenotype. Integrating these heterogeneous datasets with artificial intelligence (AI) can reveal complex gene-environment interactions and accelerate trait improvement. This review, written from a plant breeder's perspective, summarizes the current state of AI-assisted multi-omics integration in plant genetics. We describe the omics landscape, discuss machine-learning algorithms and integrative frameworks, review applications in breeding (stress tolerance, disease resistance, yield and quality traits), and examine challenges such as data heterogeneity, model interpretability and equitable data sharing. Finally, we offer recommendations for the next generation of AI-enabled plant breeding programs.

Keywords: Artificial intelligence, Machine learning, Multi-omics integration, Trait prediction, Genotype-by-environment interactions, Digital phenotyping

To cite this article: Murtaza A, S Rasool, MA Tariq, S Hassan & M Ahad. Integration of AI and Multi-Omics Data in Plant Genetics. Biological Times. 2025. December 4(12): 36-38.

Introduction

Traditional plant breeding relies on phenotypic selection and, more recently, marker-assisted selection, which focus on single loci or simple traits. High-throughput sequencing, mass-spectrometry and imaging technologies now produce petabytes of data representing the genome, transcriptome, proteome, metabolome and phenotype of plants. The combination of these so-called "multi-omics" datasets with AI and machine-learning (ML) algorithms promises to revolutionize plant breeding. AI models can uncover hidden patterns, model nonlinear relationships and predict phenotypes from multidimensional data better than traditional statistics. From a breeder's point of view, integrating AI with multi-omics can shorten breeding cycles, increase prediction accuracy and provide insights into trait architecture that would be impossible to obtain through field selection alone [1].

2. Overview of Multi-Omics Technologies and Plant Genetics

Multi-omics refers to the integrative study of several "omic" layers genomics, transcriptomics, proteomics, metabolomics, epigenomics and phenomics.

2.1 Genomics and Pan-Genomics

Genomics forms the foundation of plant breeding. Sequencing technologies now allow assembly of pan-genomes comprising hundreds of accessions, capturing core genes, dispensable genes and rare alleles that contribute to stress adaptation and yield. Pan-genomic approaches can reveal structural variants and presence–absence variation often missed by single reference genomes. High-density markers derived from sequencing are used in genomic selection and genome-wide association studies (GWAS), but their predictive power alone is limited for complex traits [2, 3].

2.2 Transcriptomics

Stress responses involve coordinated transcriptional programs. Single-cell RNA sequencing now allows exploration of cell-type-specific responses; ML algorithms such as SIMLR and neural networks enable clustering, trajectory inference and integration of multimodal data [4]. Integrating transcriptomic data with genomic markers improves prediction accuracy for traits like yield or stress tolerance, as shown in maize and rice genomic-selection models [5].

2.3 Proteomics and Metabolomics

Metabolomics quantifies metabolites reflecting biochemical pathways. Multi-omics studies show that integrating proteomic and metabolomic data with genomic markers enhances trait prediction.

2.4 Multi-Omics Integration Strategies

Horizontal integration merges datasets of the same omics type collected under diverse conditions or populations to increase robustness. Vertical integration combines different omics layers measured in the same samples to elucidate molecular cascades [6]. Multi-Omics toolbox (MOTBX) and Omics Fusion provide user-friendly platforms for breeders to combine omics layers and visualise results [7, 8].

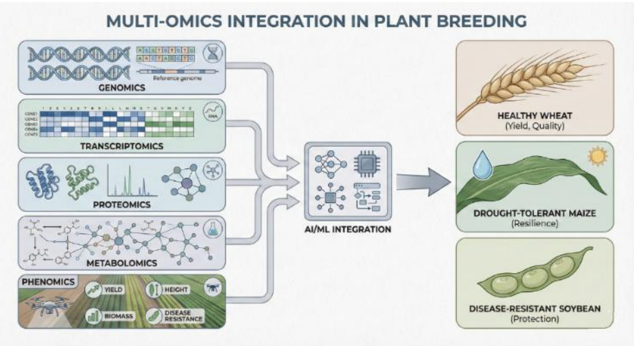


Figure 1: Conceptual framework for AI-driven multi-omics integration in plant genetics

3. Artificial Intelligence in Plant Genetics and Breeding

Gene discovery: Integrating multi-omics data with ML improves the mapping of causal genes. Population-scale multi-omics analyses treat transcripts, proteins and metabolites as molecular traits (mTraits) and high-throughput phenotypic features as imaging traits (iTraits). Dimensionality-reduction algorithms (e.g., non-negative matrix factorization) and clustering (e.g., DBSCAN) are used to create "pseudo-genotype" indices that summarise genomic variation and relate it to mTraits. These strategies enhance GWAS resolution and help fine-map alleles affecting complex traits [9].

Genotype-to-phenotype prediction: Genomic selection models use high-density markers to predict breeding values, but their accuracy suffers when traits have low heritability or strong G×E interactions. ML methods can model nonlinear relationships and integrate multiple data types. For example, integrating SNP genotypes with transcriptomic markers using random forests improved drought-tolerance prediction in maize (predictive $R^2 = 0.72$). In legumes, combining genomic, epigenomic, proteomic and metabolomic data with ML captured nonlinear relationships better than linear models and improved disease-resistance predictions. Automated feature selection using algorithms like Light Gradient Boosting Machine (LightGBM) can identify trait-associated SNPs consistent with GWAS peaks [10, 11].

Phenotyping and stress detection: Computer vision coupled with deep learning accurately classifies stress symptoms from images, quantifies leaf area and estimates biomass. Support-vector machines and neural networks have identified drought- and salt-resistance genes in Arabidopsis, maize and rice. AI also assists in reconstructing gene regulatory networks from time-series expression data and predicting long non-coding RNAs associated with abiotic stress [12, 13].

4. Integrating Multi-Omics Data and AI

4.1 Data Integration Frameworks

Successful integration requires careful preprocessing normalisation, batch correction and feature selection to handle different scales and missing data. Graph-based approaches build networks linking genes, metabolites and proteins; network inference algorithms such as iDREM use hidden Markov models to reconstruct dynamic regulatory events from time-series multi-omics data. Modern integrative frameworks combine these methods with AI to address high dimensionality, heterogeneity and nonlinearity [14, 15].

4.2 AI Algorithms for Multi-Omics Integration

Random forest (RF): RF is an ensemble of decision trees well suited for high-dimensional multi-omics data because it can model nonlinear interactions and does not require normally distributed features. RF models integrating genomic, transcriptomic and metabolomic data have predicted potato quality traits and stress tolerance in maize.

Support-vector machines (SVM): SVM algorithms classify complex patterns by maximizing the margin between classes. Variants of SVM have been used to identify genes associated with drought and salt resistance and to classify abiotic vs. biotic stress responses in rice. Kernel functions allow SVM to model nonlinear relationships in multi-omics data [16].

Dimensionality-reduction (DR) techniques: Principal component analysis (PCA), non-negative matrix factorization (NMF), t-distributed stochastic neighbour embedding (t-SNE) and Uniform Manifold Approximation and Projection (UMAP) reduce dimensionality and visualise high-dimensional data. For example, the Multi-Omics Data Association Studies (MODAS) toolbox uses NMF and clustering to summarise millions of SNPs into thousands of genomic blocks. These methods alleviate the "curse of dimensionality" and facilitate integrative analysis [17].

4.3 Tools and Platforms

An expanding ecosystem of software supports AI-based multi-omics integration. mixOmics and its DIABLO module implement supervised and unsupervised multivariate analyses and produce relevance networks. MOFA and MOFA+ use matrix factorisation to identify latent factors driving variation across omics layers, allowing missing data. iOmicsPASS combines multi-omics data with phenotype information to prioritise biomarkers and modules. MOTBX, Omics Fusion and other cloud-based platforms offer breeders user-friendly interfaces to integrate data and visualise results. These tools typically incorporate ML algorithms and are essential for non-expert breeders to exploit multi-omics data [6].

5. Applications in Plant Breeding

5.1 Stress and Disease Resistance

Abiotic stresses such as drought, heat and salinity threaten crop yields. AI-assisted multi-omics techniques have been used to identify stress-response genes and biomarkers. Integrating genomics, transcriptomics and metabolomics with ML allows accurate prediction of stress tolerance and rapid identification of candidate genes. In legumes, combining genomic, transcriptomic, epigenomic, proteomic and metabolomic data with ML improved predictions of disease resistance and has the potential to shorten breeding cycles [18].

5.2 Yield and Quality Traits

Yield and quality traits (e.g., seed size, nutrient content, tuber texture) are polygenic and influenced by environment. Integrating multi-omics layers enhances predictive accuracy for such traits. In potato, a random-forest model combining genomic, transcriptomic, proteomic and metabolomic data predicted tuber flesh colour, shape and enzymatic discolouration better than single-omics models. In maize and rice, integrating transcriptomic markers with genomic data improved yield predictions. Multi-omics data also enable mapping of metabolic pathways controlling nutritional traits. For example, multi-omics approaches identified candidate genes for kernel size in maize and oat nutritional traits through combined proteomic and metabolomic QTL analysis [19, 20].

5.3 Genotype-by-Environment Interactions and Climate Resilience

Climate change intensifies abiotic and biotic stresses, necessitating crops resilient to variable environments. Multi-omics integration coupled with predictive modelling can capture G×E interactions by incorporating environmental covariates and digital phenotyping. Integrating multi-omics technologies (genomics, transcriptomics, proteomics, metabolomics, phenomics) with machine learning and deep learning has revolutionised our understanding of plant stress networks and enables accurate genotype-to-phenotype predictions [21].

6. Future Perspectives and Recommendations

6.1 Developing Robust AI Models and Expanding Training Datasets

Future models must handle heterogeneous, high-dimensional data, integrate prior biological knowledge and deliver interpretable outputs. Ensemble methods, graph neural networks and transfer learning can improve prediction robustness. Federated learning may allow sharing model parameters without exposing data, addressing privacy concerns. Collecting diverse multi-omics datasets across environments and breeding populations will improve model generalisation and capture G×E interactions. Collaboration between breeders, computational biologists and data scientists is crucial [22].

6.2 Enhancing Multi-Omics Data Quality and Standardization

Standardised protocols for sampling, sequencing and data processing are needed to ensure comparability across studies. FAIR data repositories, integrated omics databases and metadata standards should be adopted. Platforms like MOTBX and Omics Fusion should incorporate quality-control pipelines and promote reproducibility. Investment in high-throughput phenotyping infrastructure will generate consistent iTraits, enabling integration with molecular data [23].

6.3 Training Plant Breeders in Data Science and AI

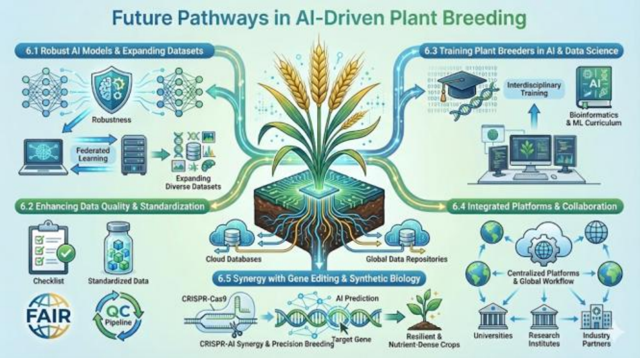
The next generation of plant breeders must be proficient not only in genetics and agronomy but also in statistics, programming and machine learning. Educational programs should integrate computational biology, bioinformatics and data science courses. Interdisciplinary training will empower breeders to design experiments, manage data and interpret AI outputs. Collaborative networks and workshops can facilitate knowledge exchange and bridge the gap between breeders and data scientists [22].

6.4 Integrated Breeding Platforms and Collaborative Networks

Integrated breeding platforms that combine genotypic, phenotypic, environmental and management data, along with AI analytical tools, will streamline breeding workflows. Centralised databases enable real-time data sharing and collaborative analysis across institutions. Breeders should leverage cloud-based infrastructure to scale their analyses and share insights with global partners. Citizen science and participatory breeding programs can contribute phenotypic data and improve model generalisation [24].

6.5 Synergy with Gene Editing and Synthetic Biology

AI-guided multi-omics analysis can identify causal genes and regulatory networks that can be targeted by genome editing tools like CRISPR/Cas. Combining genomic selection, multi-omics integration and gene editing allows breeders to design and introduce beneficial alleles with precision, accelerating the development of climate-resilient and nutrient-dense crops. Synthetic biology may enable the construction of novel metabolic pathways or regulatory circuits predicted by integrative models, although ethical and regulatory frameworks will be needed[15, 25].



7. Conclusion

Integrating AI with multi-omics data ushers in a new era of plant breeding. Multi-omics technologies capture the complexity of plant biology, and AI algorithms can distil this complexity into actionable insights. Evidence from recent studies shows that combining genomic, transcriptomic, proteomic, metabolomic and phenomic data with ML and deep learning improves prediction accuracy for stress tolerance, disease resistance, yield and quality traits. Random-forest and support-vector-machine models have already enhanced trait prediction in potato, maize, rice and legumes. However, challenges remain in data integration, computational resources, interpretability and ethical considerations. Overcoming these barriers will require robust algorithms, standardised data, investment in infrastructure and multidisciplinary training. From a breeder's perspective, the integration of AI and multi-omics data holds immense promise to accelerate genetic gains, adapt crops to changing environments and secure global food supply. Adoption of these technologies, coupled with responsible governance and collaboration, will transform plant breeding from selection to intelligent design.

References

[1] Mahmood, U., et al., Multi-omics revolution to promote plant breeding efficiency. Frontiers in Plant Science, 2022. 13.

[2] Liu, Y., Construction and Analysis of a Rice Pan-genome Reveals Structural Variation Hotspots Across Subspecies. Rice Genomics and Genetics, 2025.

[3] Jin, S., et al., Structural variation (SV)-based pan-genome and GWAS reveal the impacts of SVs on the speciation and diversification of allotetraploid cottons. Molecular Plant, 2023. 16(4): p. 678-693.

[4] Dylan, M., et al., Deep Learning in Single-cell Analysis. ACM Transactions on Intelligent Systems and Technology, 2022.

[5] Brendel, M., et al., Application of Deep Learning on Single-Cell RNA Sequencing Data Analysis: A Review. Genomics, Proteomics & Bioinformatics, 2022. 20(5): p. 814-835.

[6] Vahabi, N. and G. Michailidis, Unsupervised Multi-Omics Data Integration Methods: A Comprehensive Review. Frontiers in Genetics, 2022. 13.

[7] Huang, W., et al., A Review of Fusion Methods for Omics and Imaging Data. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2023. 20(1): p. 74-93.

[8] Montesinos-López, O.A., et al., A review of multimodal deep learning methods for genomic-enabled prediction in plant breeding. GENETICS, 2024.

[9] Alemu, R., et al., Multi-omics approaches for understanding gene-environment interactions in noncommunicable diseases: techniques, translation, and equity issues. Human Genomics, 2025. 19(1).

[10] Chen, L., Causal Genomics in the Deep Learning Era. 2025.

[11] Nam, Y., et al., Harnessing Artificial Intelligence in Multimodal Omics Data Integration: Paving the Path for the Next Frontier in Precision Medicine. Annual Review of Biomedical Data Science, 2024. 7(1): p. 225-250.

[12] Wen, T., et al., Thermal imaging: The digital eye facilitates high-throughput phenotyping traits of plant growth and stress responses. Science of The Total Environment, 2023. 899: p. 165626.

[13] Singh, A., et al., Challenges and Opportunities in Machine-Augmented Plant Stress Phenotyping. Trends in Plant Science, 2021. 26(1): p. 53-69.

[14] Subramanian, I., et al., Multi-omics Data Integration, Interpretation, and Its Application. Bioinformatics and Biology Insights, 2020. 14: p. 117793221989905.

[15] Mallik, S. and Z. Zhao, Graph- and rule-based learning algorithms: a comprehensive review of their applications for cancer type classification and prognosis using genomic data. Briefings in Bioinformatics, 2019. 21(2): p. 368-394.

[16] Thanmalagan, R.R., et al., Comprehensive meta-analysis and machine learning approaches identified the role of novel drought specific genes in Oryza sativa. Plant Gene, 2022. 32: p. 100382.

[17] Ros, F. and R. Riad, Dimensionality reduction. Unsupervised and Semi-Supervised Learning, 2023: p. 11-25.

[18] Roychowdhury, R., et al., Multi-Omics Pipeline and Omics-Integration Approach to Decipher Plant's Abiotic Stress Tolerance Responses. Genes, 2023. 14(6): p. 1281.

[19] Yang, Y., et al., Applications of Multi-Omics Technologies for Crop Improvement. Frontiers in Plant Science, 2021. 12.

[20] Iqbal, Z., et al., Toward Integrated Multi-Omics Intervention: Rice Trait Improvement and Stress Management. Frontiers in Plant Science, 2021. 12.

[21] Khan, M.K.R., et al., The Intervention of Multi-Omics Approaches for Developing Abiotic Stress Resistance in Cotton Crop Under Climate Change. Sustainable Agriculture in the Era of the OMICs Revolution, 2023: p. 37-82.

[22] Hafner, A., et al., Data Reuse in Agricultural Genomics Research: Present Challenges and Future Solutions. 2024.

[23] Demirbaga, Ü., et al., Cloud Computing for Big Data Analytics. Big Data Analytics, 2024: p. 43-77.

[24] Tiwari, J.K., et al., Germplasm, Breeding, and Genomics in Potato Improvement of Biotic and Abiotic Stresses Tolerance. Frontiers in Plant Science, 2022. 13.

[25] Murmu, S., et al., A review of artificial intelligence-assisted omics techniques in plant defense: current trends and future directions. Frontiers in Plant Science, 2024. 15.