

From Traits to Networks: Decoding Wheat's Drought Adaptation Through Multimodal Imaging, Omics, and AI

Submitted by

Mirza Mohammed Shoaib

Master of Agriculture Studies, University of Queensland, Australia

Masters of Botany, National University, Bangladesh

A thesis submitted in total fulfilment of the requirements for the degree of the
requirements for the degree of
Doctor of Philosophy

School of Applied Systems Biology
Faculty of Science, Health and Engineering
La Trobe University
Bundoora, Victoria 3086
Australia

August 2025

Table of Contents

| | |
|--|-----|
| List of Abbreviations..... | 3 |
| Abstract..... | 4 |
| Statement of Authorship | 5 |
| Funding Acknowledgement | 6 |
| Acknowledgement..... | 6 |
| Chapter 1: A Systems Biology Approach to Decoding Plant Drought Resilience..... | 8 |
| Chapter 2. Roots' Drought Adaptive Traits in Crop Improvement..... | 64 |
| Chapter 3. AutoML: A Next-Generation Tool for Mining Multivariate Plant Trait..... | 86 |
| Chapter 4. Seeing the Unseen: Algorithmic Root Traits (ART)..... | 90 |
| Chapter 5. Tissue-Specific Metabolomic Networks Orchestrate Osmotic Stress Adaptation..... | 107 |
| Chapter 6. MOFA+ Transformer: The Ultimate Integration Framework..... | 139 |
| Chapter 7: General Discussion – A New Paradigm for Decoding Plant Resilience..... | 182 |
| Appendix 1: Chapter 4 Supplementary Material..... | 223 |
| Appendix 2: Chapter 5 Supplementary Material..... | 246 |
| Appendix 3: Chapter 6 Supplementary Material..... | 267 |

List of Abbreviations

A

- **ABA** - Abscisic Acid
- **ART** - Algorithmic Root Trait
- **AutoML** - Automated Machine Learning

C

- **CCA** - Canonical Correlation Analysis
- **CV** - Coefficient of Variation

D

- **DAS** - Days After Sowing

F

- **FDR** - False Discovery Rate

L

- **LC-MS** - Liquid Chromatography-Mass Spectrometry
- **LF** - Latent Factor

M

- **ML** - Machine Learning
- **MOFA+** - Multi-Omics Factor Analysis+

P

- **PCA** - Principal Component Analysis
- **PLS** - Partial Least Squares
- **PLS-DA** - Partial Least Squares Discriminant Analysis

R

- **ROC AUC** - Receiver Operating Characteristic Area Under the Curve
- **RWC** - Relative Water Content

S

- **S2M** - Spectral-to-Molecular
- **SHAP** - SHapley Additive exPlanations
- **SSI** - Stress Susceptibility Index

T

- **TRT** - Traditional Root Trait

V

- **VIP** - Variable Importance in Projection

W

- **WGCNA** - Weighted Gene Co-expression Network Analysis

Abstract

Enhancing food security in a changing climate is constrained by a critical gap between our capacity to generate biological data and our ability to extract mechanistic insight for complex traits like drought tolerance in wheat, a crop that sustains a third of humanity. This thesis introduces an integrated, interpretable AI framework to decode the multi-scale architecture of plant resilience by solving four interconnected analytical bottlenecks in plant systems biology: analytical complexity, phenotyping, mechanistic understanding, and systems integration.

To systematically address these bottlenecks, the AI framework was deployed on a multi-modal, time-series dataset from contrasting wheat genotypes under osmotic stress, progressively revealing the multi-scale architecture of resilience. The Algorithmic Root Trait (ART) method resolved the phenotyping bottleneck by discovering latent root architectural features invisible to human perception, increasing drought tolerance classification accuracy from 85.6% to 96.3%, with a 5.8 \times increase in information density. Network architecture analysis addressed the mechanistic understanding bottleneck, revealing that resilience emerges from distinct tissue-specific metabolic architectures: tolerant genotypes form ~40% denser leaf networks for rapid coordination and modular root networks for localised adaptation, both governed by strategic temporal coordination patterns. The MOFA+ Transformer approach overcame the systems integration bottleneck by quantifying that in tolerant genotypes, coordination strength in leaf tissue was 355% stronger, while root networks were more targeted, and these networks were established earlier. The framework's power was validated on an independent single-cell dataset, where it discovered a statistically significant ($p = 0.0099$) non-linear link between cellular phenotype and the key stress-related gene lncRNA NEAT1, a connection not detected by standard correlation analysis.

Together, these results suggest heritable plant resilience is an emergent property of the precision, timing, and architecture of multi-scale coordination, rather than the inherent superiority of individual components. More broadly, the findings of this coordination-centric AI framework provide a new suite of mechanistically-grounded biomarkers—such as 'coordination strength' and algorithmic architectural traits—to accelerate precision breeding. This research delivers not only a unified model of drought adaptation in wheat but also a validated, open-source suite of analytical tools to accelerate the development of climate-resilient crops.

Statement of Authorship

Except where otherwise acknowledged in the text, this thesis contains no material that has been published elsewhere or extracted in whole or in part from a thesis accepted for the award of any other degree or diploma.

No other person's work has been used without due acknowledgment in the main text of the thesis. This thesis has not been submitted for the award of any other degree or diploma at any other tertiary institution.

The contributions of co-authors are specified in the preface of each relevant chapter. Statements of collaboration have been approved by all co-authors, and this approval has been verified by the Principal Supervisor at the beginning of each chapter.

Mirza Mohammed Shoaib

1st August 2025

Funding Acknowledgement

This research was funded by the Department of Energy, Environment and Climate Action (Agriculture Victoria), Australia.

Acknowledgement

I would like to express my sincere gratitude to the many individuals who have supported me throughout my PhD journey. This work would not have been possible without them.

First and foremost, I dedicate this thesis to the memory of my late father. I have no doubt he would have been incredibly proud to see this work completed. His unwavering support and the countless sacrifices he made for our family laid the foundation for all my achievements. This work stands as a testament to his enduring legacy and love.

My deepest gratitude extends to the three most important people in my life: my mother, my beloved wife, and my wonderful daughter. Your constant love, patience, and encouragement have been my anchor throughout this demanding journey. The sacrifices you have all made, often putting my research ahead of your own comfort and time, have not gone unnoticed. This achievement is as much yours as it is mine, and it was only possible with you by my side.

I extend my sincere appreciation to my supervisors, Dr Matthew Hayden and Dr Surya Kant, for their continuous support and expert guidance. Your mentorship was instrumental in my development as a researcher.

I gratefully acknowledge the funding for this research provided by Agriculture Victoria Research.

Several individuals were crucial to the success of my experimental work. I owe a particular debt of gratitude to Emily L. Thoday-Kennedy, who was a key partner in this research. Emily, your deep knowledge of wheat, assistance with every experimental setup, and the personal time you so generously gave were invaluable.

I am incredibly grateful to Dr Adam Dimech for his significant efforts in optimising our image capture and analysis systems at a critical time. I also thank Edmond J. Breen for his essential guidance and for sharing his deep knowledge in this area.

My thanks go to Dr Glenn J. Fitzgerald for providing access to crucial equipment and facilities, and for his helpful technical advice.

Thank you to Dr Denise M. Barbulescu for her kind words, advice, and for the crucial introduction to Dr Priyanka Reddy and Dr Simone J. Rochfort.

This PhD would not have been possible without the vision of Dr Priyanka Reddy, which shaped my key experiments, and her significant support in analysing the LCMS data. I am also deeply thankful to Doris Ram, who dedicated countless hours to helping me execute these experiments.

I am also very grateful to Dr Simone J. Rochfort for her invaluable guidance during the preparation of my manuscripts, and to William Paul Stratmann for his assistance with the experimental setup.

My warm thanks extend to Professor Caixian Tang. His wisdom helped to shift my perspective during a difficult period and guided me to continue with my PhD. I am sincerely grateful for his mentorship.

I would also like to thank Nuseed, and particularly Peter Flett, for the opportunity of casual employment, which provided vital support for my studies after my scholarship funding concluded.

This acknowledgement would not be complete without thanking Kendra L. Whiteman, who was always there to offer support when needed.

Finally, to any friends, colleagues, or mentors whom I have not been able to mention by name, I offer my heartfelt thanks and sincere apologies. Your support was deeply appreciated and has not been forgotten.

Chapter 1: A Systems Biology Approach to Decoding Plant Drought Resilience

1.1 The Global Imperative: Food Security in a Changing Climate

The 21st century is defined by the collision of two powerful and opposing forces: a relentlessly growing global population and the escalating impacts of a changing climate [1-3]. This convergence has created an unprecedented challenge for global agriculture, transforming the pursuit of food security from a logistical problem into a scientific grand challenge [4, 5]. As climatic conditions become more volatile and extreme, the stability of our food systems is increasingly under threat, demanding a paradigm shift in how we breed, grow, and manage our most critical crops [6, 7].

1.1.1 The Escalating Impact of Drought on Global Agriculture

Climate change has fundamentally altered global precipitation patterns, with drought emerging as the most formidable and widespread threat to agricultural productivity [3, 8]. Anthropogenic influences have demonstrably increased drought frequency, duration, and intensity across major agricultural regions [8], causing cascading effects that extend far beyond immediate crop losses to include long-term soil degradation and ecosystem disruption [3, 9]. Drought now affects more agricultural land than any other climatic extreme, with annual economic losses estimated in the hundreds of billions globally [3]. The physiological consequence for plants is osmotic stress, a fundamental cellular challenge that disrupts nutrient uptake, impairs photosynthesis, and ultimately compromises plant growth and survival [10]. Projections from the Intergovernmental Panel on Climate Change (IPCC) indicate this threat will only intensify, even as agricultural water demand is expected to increase by 50% by 2030 amid declining freshwater resources [3, 11].

1.1.2 Wheat (*Triticum aestivum* L.): A Cornerstone of Global Food Security Under Pressure

At the heart of this global challenge lies wheat, a crop that provides over 20% of the world's caloric intake and is a nutritional cornerstone for more than 2.5 billion people [12]. Its cultivation spans more land area than any other commercial crop, making it uniquely exposed to the geographically diverse impacts of climate change. However, wheat systems are particularly vulnerable to drought, with water limitation identified as the primary yield-limiting factor across 60% of the global wheat-growing area [13]. The immediate consequences are severe; with international wheat trade exceeding 180 million tonnes annually [12], localised drought-induced yield reductions can trigger immediate global supply shocks with disproportionate impacts on food-insecure populations [14]. Critically, climate change is already reshaping wheat variety performance rankings across different environments [15], creating unusual conditions that exceed the adaptive capacity of many current elite cultivars [7].

1.1.3 The Need for Climate-Resilient Cultivars

In the face of these compounding pressures, enhancing the intrinsic resilience of the wheat plant itself is the most sustainable and scalable solution [16]. Yet, traditional breeding for drought tolerance has yielded limited success, largely because resilience is not a single trait but an emergent property of multiple physiological, morphological, and biochemical adaptations orchestrated across time and space [17, 18]. Critical adaptations—including root architecture [19, 20], osmotic adjustment [10], antioxidant systems [18], and phenological plasticity [7]—are governed by complex inheritance patterns that have been difficult to select for directly [21].

Recent advances in systems biology and high-throughput phenotyping have identified numerous candidate genes and regulatory networks [22-24]. However, translating this knowledge into improved cultivars is hindered by a critical mismatch: the 10–15 year cycle of conventional breeding is too slow to keep pace with rapid climate change [25, 26], and existing analytical frameworks are unable to effectively integrate the vast phenotypic, genomic, and environmental datasets now being

generated [17, 27]. This analytical gap—between the generation of complex biological data and the extraction of actionable, mechanistic insights—defines the central challenge that this thesis aims to resolve. Overcoming this bottleneck requires a new generation of computational tools capable of accelerating trait discovery and deployment, a prerequisite for fast-forwarding the development of climate-resilient crops [22, 28].

1.2 The Biological Foundation: The Complexity of Drought Adaptation

1.2.1 The Root System: A Plant's Primary Interface with Drought

The imperative to develop climate-resilient crops necessitates a profound understanding of the biological mechanisms governing plant adaptation. At the forefront of this process is the root system, a sophisticated biological interface that orchestrates a multi-layered response to water deficit [29-31]. A comprehensive review of these multifaceted root adaptations is presented in Chapter 2; this section synthesises the key principles to establish the essential analytical challenge this thesis seeks to overcome.

A plant's drought response machinery operates across interconnected physical, biochemical, and symbiotic scales. Architecturally, it deploys both constitutive traits, like steep rooting angles for subsoil access [19, 20], and plastic responses like hydropatterning [32, 33]. The efficacy of this architecture is underpinned by precise anatomical and hydraulic engineering, where traits such as cortical aerenchyma formation reduce metabolic costs [34, 35] while aquaporin expression and suberin deposition dynamically regulate water transport to prevent hydraulic failure [36-38]. Complementing this internal control is a sophisticated biochemical and microbial strategy. Through exudation, roots release metabolites that retain soil moisture and shape a beneficial rhizosphere microbiome [39-41], which functions as a direct extension of the plant's adaptive capacity by enhancing nutrient uptake and stress tolerance [42, 43].

This web of interactions highlights the ultimate challenge: drought adaptation emerges not from any single trait but from their dynamic, system-level integration, where architectural investments come at a metabolic cost and hydraulic properties must converge with spatial deployment for optimal function [44, 45]. This biological reality—a multi-scale, interconnected, and highly plastic system—creates emergent properties that cannot be predicted by studying individual components in isolation. It is precisely this complexity that traditional, reductionist phenotyping methods fail to capture [46, 47]. They can measure a part but not the pattern, a component but not the coordination. This gap between biological reality and our analytical capacity creates the fundamental bottlenecks in phenotyping and mechanistic understanding that this thesis systematically addresses.

1.2.2 Beyond Roots: The Whole-Plant Coordination Challenge

While the root system forms the primary interface with drought, a plant's resilience is not determined by its roots alone. Survival and productivity emerge from a highly coordinated, whole-plant strategy, where roots and shoots engage in a continuous, bidirectional dialogue to balance water acquisition with water loss, and energy investment with metabolic maintenance [48, 49]. This is not a simple stimulus-response pathway but a multi-modal, bidirectional communication system employing a symphony of hydraulic, chemical, electrical, and molecular signals that operate across multiple temporal scales [48, 49]. Understanding this intricate communication network is fundamental, as it reveals that true drought tolerance is an emergent property of systemic coordination, not just the sum of individual component responses [18, 50]. The very complexity of this dialogue defines the analytical bottlenecks that this thesis seeks to overcome.

1.2.2.1 Root-Shoot Communication: The Orchestrated Dialogue of Survival

The dialogue between root and shoot systems represents one of the most sophisticated communication networks in biology, enabling plants to function as unified organisms despite their spatially distributed architecture [48, 49]. Under drought stress, this communication network becomes the central nervous system of

plant survival, orchestrating rapid emergency responses while simultaneously coordinating long-term adaptive strategies [18, 32]. Recent advances in plant systems biology have revealed that this root-shoot dialogue is fundamentally asymmetric, with different tissues deploying specialised communication strategies and network architectures that reflect their distinct functional roles [51, 52]. Understanding these communication protocols is essential for decoding the systems-level organisation of drought tolerance and reveals the profound analytical challenges that limit current approaches to crop improvement [53].

1.2.2.1.1 The Upward Stream: Multi-Modal Emergency Reporting from the Rhizosphere

Upon detecting a water deficit, roots initiate an immediate, multi-modal "emergency report" that propagates through multiple signalling pathways with distinct kinetics and molecular mechanisms [48, 54]. This upward signalling cascade represents one of the fastest and most comprehensive information transfer systems in plant biology, enabling shoot tissues to receive real-time updates on belowground water status [48].

Hydraulic Signalling: The Physical Foundation

The fastest component of this emergency response is hydraulic signalling, where changes in soil water potential create negative pressure waves that propagate through the xylem network almost instantaneously [55, 56]. This physical signal represents the plant's most immediate drought sensor, capable of transmitting information about water deficit across meter-scale distances within minutes [57]. Advanced understanding of root hydraulic conductivity reveals that this hydraulic network is actively regulated by specialised transcription factors such as *XND1*, which controls xylem differentiation and thus modulates the efficiency of hydraulic signal transmission [58]. Moreover, aquaporins (water channel proteins) dynamically regulate radial water transport, providing fine-tuned control over hydraulic signal intensity and enabling tissue-specific modulation of water flow [38, 59].

Electrical Signalling: Rapid Action Potential Networks

Complementing hydraulic signals, electrical signalling provides another rapid-response communication channel through waves of ion flux that propagate systemically throughout the plant [60]. These bioelectrical signals, analogous to action potentials in animal nervous systems, can influence stomatal behaviour within minutes of stress perception and represent a critical component of the plant's emergency response system [61, 62]. Recent research has demonstrated that root-apex proton fluxes serve as key regulators of these electrical signals, positioning root tips as sophisticated electrochemical sensors that continuously monitor soil stress conditions [60].

Chemical Signalling: The Hormonal Command Centre

Following these rapid physical alerts, a more sustained and specific chemical message system is activated, centered around the master regulatory hormone abscisic acid (ABA) [18, 24]. Upon drought perception, roots synthesise and release ABA, which is transported via the xylem sap to aerial tissues where it initiates a comprehensive stress response program [63]. The ABA signalling pathway represents one of the most thoroughly characterised stress communication systems, involving the PYR/PYL/RCAR receptor complex in guard cells, which triggers stomatal closure and activates extensive transcriptional reprogramming through the SnRK2-AREB/ABF regulatory cascade [64, 65]. However, xylem sap pH changes, which become more alkaline under water stress, act synergistically with ABA to enhance stomatal closure, while decreased root-to-shoot transport of cytokinins serves as an additional growth-limiting signal [63, 66].

Molecular Signalling: Long-Distance RNA and Peptide Networks

The most recently discovered component of root-shoot communication involves mobile molecular signals, including regulatory non-coding RNAs and stress-related peptides that provide highly specific, context-dependent information transfer [48, 67]. Long non-coding RNAs such as *ELENA1* demonstrate the capacity for systemic movement and tissue-specific regulation of stress responses, representing a novel layer of communication that enables fine-tuned coordination between distant tissues [67, 68]. These molecular signals appear particularly important for conveying

information about the specific nature and intensity of stress conditions, enabling context-appropriate responses that simple hormonal signals cannot achieve [48].

1.2.2.1.2 The Downward Stream: Strategic Resource Allocation and Metabolic Coordination

In response to these upward stress signals, shoots make critical resource allocation decisions that determine plant survival [69, 70]. This downward signalling stream represents a sophisticated resource management system that must balance immediate survival needs with long-term growth requirements [71].

Carbon Allocation and Root-Shoot Ratio Optimisation

Perhaps the most critical shoot-to-root signal involves the strategic redirection of photosynthetic carbon toward root growth and function [69, 72]. Recent molecular insights reveal that ABA-activated SnRK2 kinases phosphorylate SWEET sucrose transporters create a molecular switch that enhances sugar transport to roots and provides the energy necessary for continued water exploration under stress [69]. This process demonstrates the remarkable precision of plant resource allocation, where chemical signals from stressed roots directly trigger enhanced carbon investment in belowground exploration [70, 71].

Metabolic Network Coordination

Beyond simple carbon allocation, emerging evidence from time-series metabolomics shows that shoots coordinate complex metabolic reprogramming in roots [51, 52]. Drought-responsive metabolites such as raffinose, galactinol, and trehalose show tissue-specific accumulation patterns that are coordinated across the plant, with roots accumulating protective osmolytes while shoots maintain different metabolic profiles optimised for photosynthetic protection [73, 74]. This metabolic coordination suggests the existence of sophisticated signalling networks that enable tissue-specific optimisation within a coherent whole-plant strategy [52]. The breakdown in this coordination in susceptible genotypes highlights its importance for overall resilience [75].

1.2.2.2 Temporal Dynamics of Adaptation Strategies

The intricate dialogue between root and shoot is not static; it is a dynamic process that unfolds as a sophisticated temporal performance, where true resilience is characterised not just by beneficial traits, but by the speed, sequence, and precision of their coordinated activation [18, 48, 76]. Traditional "snapshot" omics studies, which capture single moments of the stress response, fundamentally miss this critical temporal dimension, providing parts lists without the assembly instructions or operational timelines [76].

The plant stress response typically follows a multi-phasic trajectory that distinguishes tolerant from susceptible genotypes at each critical transition [24, 49]. The initial phase is an "emergency" response dominated by rapid, system-wide signals. Hydraulic pressure waves propagate almost instantaneously through xylem networks, alerting shoots to water deficits within minutes [55, 56], while electrical signals and the initial wave of the hormone abscisic acid (ABA) trigger rapid stomatal closure to prevent catastrophic water loss [18, 60]. This immediate, protective phase is crucial for short-term survival.

However, if stress persists, successful adaptation requires a sophisticated transition from immediate defense to a sustained 'acclimation' phase involving profound metabolic reprogramming and strategic resource reallocation, such as shifting carbon allocation to promote root growth [49, 70-72]. It is the efficiency and timing of this transition that fundamentally separates tolerant genotypes from susceptible ones. This aligns with the 'early response hypothesis,' which posits that the capacity for rapid and proactive adjustment is a primary determinant of overall resilience [77, 78]. Tolerant plants deploy these proactive strategies, establishing critical coordination networks at quantifiably earlier stress stages, creating an 'early warning system' that enables managed responses rather than reactive crisis management [67].

This thesis reveals that the tolerant genotype orchestrates a remarkable two-act performance to manage prolonged stress. Initially, in Act I, the plant operates like a

highly centralized command system. The metabolic activities of the roots and leaves are tightly synchronized, showing strong overall correlation ($r \approx 0.55$), ensuring a unified, whole-plant emergency response. Then, in Act II, it executes a strategic pivot. As the stress continues, the plant grants its tissues greater autonomy, allowing them to specialize. This is reflected in a deliberate decoupling of their metabolic states, where the overall correlation between root and leaf metabolism strategically declines by nearly 40% (to $r \approx 0.35$) [79]. This shift allows the roots to focus independently on water foraging while the leaves optimize their distinct task of preserving photosynthetic machinery.

This dynamic pattern—from unified command to specialized delegation—is a hallmark of sophisticated adaptation and is entirely absent in the susceptible genotype, which fails to execute this temporal program effectively. Capturing and quantifying these dynamic patterns—the temporal choreography of resilience—represents a profound analytical challenge and a central motivation for the novel frameworks presented in this work [76, 80].

1.2.2.3 The Complexity of Systems-Level Interactions

The individual adaptive mechanisms outlined—architectural deployment, hydraulic engineering, multi-modal communication, and temporal dynamics—are not isolated modules. They are inextricably linked components of a single, integrated, and highly complex system [49, 50, 81]. This creates a cascade of interdependencies and trade-offs where the behaviour of the whole plant is an emergent property that cannot be predicted by studying its parts in isolation [82]. A prevalent challenge in systems biology is the reduction of high-dimensional data to simple lists of "significant features" without uncovering the principles governing their interaction—providing parts lists but not architectural blueprints [83, 84]. True mechanistic insight requires understanding the system's design, because function is fundamentally an emergent property of structure [82, 85].

This systems-level complexity manifests through **multiple organisational layers** that form an integrated hierarchy. At the foundation lies the **Physical Layer**, where

sophisticated spatial deployment strategies—such as the algorithmically-discovered patterns of dense root cluster positioning as detailed in Chapter 4—optimise water acquisition under heterogeneous soil conditions [19, 86]. However, there are clear trade-offs; for instance, the development of root cortical aerenchyma reduces metabolic construction costs but can compromise colonisation by beneficial mycorrhizal fungi [87, 88]. This physical architecture provides the spatial framework upon which more sophisticated strategies operate.

Building upon this foundation is the **Metabolic Layer**, characterised by striking tissue-specific network architectures that reflect evolved specialisation [52]. As shown in Chapter 5, network analysis reveals that drought-tolerant genotypes deploy fundamentally asymmetric strategies: leaf networks exhibit ~40% higher density and elevated transitivity, creating integrated architectures for rapid coordination, while root networks demonstrate higher modularity and greater fragmentation, enabling flexible, localised adaptation. Critically, architectural and hydraulic traits must functionally converge for optimal performance: a deep rooting architecture is only effective when supported by a hydraulic system capable of efficiently transporting water without cavitation [30, 58].

The most dynamic component is the **Coordination Layer**, encompassing the real-time communication protocols that execute metabolic strategies across tissues and time. This dialogue operates through multiple signalling modalities—hydraulic pressure waves, ABA-mediated chemical signals, electrical action potentials, and mobile RNA networks [56, 67]. As quantified by the MOFA+ Transformer in Chapter 6, tolerant genotypes demonstrate markedly stronger and more precise cross-modal coordination, with specific molecular hubs orchestrating tissue-appropriate responses through quantifiable, directed regulatory relationships [80].

The analysis of susceptible genotypes reinforces this systems view, revealing that resilience breakdown occurs not through the catastrophic failure of a single component, but through a loss of coordination and compromised integration across these layers [81]. This leads to the central hypothesis underpinning this thesis: that the true, heritable drought resilience emerges from the coordination capacity of the

entire integrated system, rather than from the absolute superiority of its individual components [49, 81].

It is precisely this staggering biological complexity—the multi-scale, interconnected, and dynamic nature of drought adaptation—that has created the profound analytical bottlenecks hindering modern crop improvement [53]. Traditional, reductionist approaches are fundamentally ill-equipped to measure, model, or interpret such a system [52, 84]. This chasm between biological reality and our analytical power represents the "grand challenge" this thesis confronts. The subsequent sections will deconstruct this challenge into four specifics, interconnected analytical bottlenecks that must be solved to unlock the next generation of mechanistic insights required for breeding climate-resilient crops [76, 80].

1.3 The Analytical Frontier: Four Bottlenecks Hindering Biological Discovery

The profound biological complexity outlined in the previous sections presents an equally profound analytical challenge. The 21st century has witnessed an exponential expansion in our capacity to generate biological data through high-throughput phenotyping, next-generation sequencing, hyperspectral imaging, and multi-omics platforms [22, 23, 89]. This technological revolution has created a fundamental paradox: while our data archives have swelled to unprecedented scales, our ability to extract actionable, mechanistic insight from them has not kept pace [53, 80, 90]. The primary constraint in modern biological discovery is no longer the generation of data, but its interpretation.

This analytical chasm has created a series of critical bottlenecks that systematically impede progress in decoding complex traits like drought resilience. These are not independent obstacles but an interconnected cascade of limitations, where failure at one stage compromises the potential of all subsequent analyses [80, 84]. This thesis was conceived as a systematic effort to solve four specific, hierarchically organised analytical bottlenecks that currently constrain plant systems biology:

The Analytical Complexity Bottleneck arises from the sheer scale and complexity of modern datasets that render traditional manual analysis obsolete, demanding a transition toward automated, intelligent frameworks [91, 92].

The Phenotyping Bottleneck confronts the fundamental limitations of human visual perception in capturing the subtle, non-obvious architectural patterns that distinguish adaptive strategies [93-95].

The Mechanistic Understanding Bottleneck challenges the prevalent reductionist tendency to catalogue lists of "significant" features without uncovering the organisational principles governing their interaction [82, 83, 85].

The Systems Integration Bottleneck tackles the inability of standard correlation-based approaches to capture the directionality, feature-specificity, and temporal dynamics that define true biological coordination [76, 96].

These bottlenecks represent more than technical hurdles; they constitute a paradigmatic barrier preventing the transition from descriptive to predictive plant biology [27, 80]. The systematic solution of these analytical roadblocks through the development of an integrated computational ecosystem forms the central mission of this thesis.

1.3.1 The Analytical Complexity Bottleneck: From Manual Analysis to Automated Discovery

The first and most foundational bottleneck arises from the sheer scale, velocity, and complexity of modern biological datasets, which have fundamentally overwhelmed traditional analytical paradigms. High-throughput technologies now routinely generate data of unprecedented dimensionality and heterogeneity—genomic studies with millions of variants, metabolomic experiments detecting thousands of compounds, and hyperspectral imagery capturing thousands of wavelengths across temporal series [22, 23, 97]. This data deluge presents the infamous "curse of

"dimensionality," where the number of measured variables (p) vastly exceeds the number of biological samples (n), making conventional statistical models unstable, prone to overfitting, and statistically invalid [23, 92, 98].

In this new data-rich landscape, traditional analytical workflows that rely on manual feature engineering, iterative model selection, and hypothesis testing by human experts have become subjective, error-prone, and critically, unscalable [91, 99]. The manual analytical process introduces systematic biases where researchers may unconsciously gravitate toward familiar patterns, creating a "confirmation bias" that privileges known mechanisms over the discovery of genuinely novel biological insights [80, 100]. This cognitive and perceptual limitation is particularly damaging in complex trait research, where the most biologically informative patterns may be subtle, non-linear, or embedded in high-order interactions—precisely the types of relationships that human analysts are ill-equipped to detect [101, 102].

Furthermore, this reliance on ad-hoc manual analysis contributes directly to the reproducibility crisis in the biological sciences. The lack of standardised, automated, and transparent analytical workflows leads to immense variability between studies, preventing effective meta-analyses and hindering the validation of findings across different laboratories and research programs [80, 84]. This fragmentation is especially detrimental in crop improvement, which demands robust, transferable analytical protocols that can be applied consistently across diverse genetic backgrounds and fluctuating environmental conditions [27, 53].

Automated Machine Learning (AutoML) emerges as the paradigmatic solution to this analytical complexity bottleneck [92, 102, 103]. AutoML represents a fundamental shift by automating the entire machine learning (ML) pipeline—from data preprocessing and feature selection through algorithm selection, hyperparameter optimisation, and model validation [99]. This automation democratises access to advanced data science, empowering domain experts like plant scientists to conduct sophisticated analyses without requiring deep programming or data science expertise [22, 103].

The power of AutoML lies not merely in its efficiency, but in its capacity to systematically and objectively explore a vast model and feature space that would be impossible for a human to investigate manually [92, 104]. By rigorously testing thousands of analytical strategies in parallel, AutoML can uncover the subtle, non-obvious patterns and high-order feature interactions that are critical for understanding complex traits like drought tolerance but would likely be missed by traditional manual approaches [23].

Modern AutoML platforms provide unprecedented analytical comprehensiveness, systematically evaluating diverse feature selection methods, model architectures, and hyperparameter configurations through exhaustive search procedures that ensure optimal analytical strategies are identified rather than assumed [92, 102]. This methodical exploration enables the discovery of unexpected biological relationships and non-linear patterns that emerge only when the full complexity of biological datasets is properly leveraged [80].

Overcoming this initial bottleneck is the essential prerequisite for all subsequent biological discovery. Without robust, scalable, and reproducible tools to manage and mine these immense datasets, the deeper biological questions of phenotyping accuracy, mechanistic understanding, and systems integration remain fundamentally inaccessible [80]. Chapter 3 of this thesis introduces and advocates for the paradigmatic solution offered by AutoML, which automates the entire ML pipeline—from feature selection to model validation—thereby democratising access to advanced data science and empowering domain experts like plant scientists [22, 103].

However, its true power lies not merely in efficiency, but in its capacity for objective discovery. By systematically and objectively exploring a vast model and feature space that is inaccessible to human-led exploration, AutoML can uncover the subtle, non-obvious patterns and high-order feature interactions that are critical for understanding complex traits [23, 92, 99]. Therefore, establishing the principle of automated discovery in Chapter 3 provides the conceptual and methodological

launchpad for the bespoke experimental frameworks developed in Chapters 4, 5, and 6.

1.3. 2 The Phenotyping Bottleneck: From Human Perception to Algorithmic Vision

The sophisticated biological machinery of drought adaptation outlined in the previous sections creates an equally sophisticated challenge for phenotypic measurement. Our ability to quantify biological form is fundamentally constrained by the limits of human visual perception [94], creating a critical bottleneck that prevents us from capturing the full architectural complexity that underpins plant resilience. This phenotyping bottleneck represents more than a technical limitation—it constitutes a paradigmatic barrier that systematically blinds us to the subtle, non-obvious spatial patterns and latent organisational features that distinguish adaptive strategies from maladaptive ones [80, 93, 94].

1.3.2.1 *The Current State: Automation of Human-Defined Traits*

Contemporary root phenotyping represents a remarkable technological achievement in automating the extraction of predefined morphological characteristics. High-throughput platforms employing rhizotrons, minirhizotrons, and digital imaging systems have revolutionised our capacity to generate phenotypic data at unprecedented scales [93, 105, 106]. Software tools such as RhizoVision Explorer [107], WinRHIZO [108], RootNav [109], DIRT [110], and ImageJ [111] can rapidly extract dozens of Traditional Root Traits (TRTs) from segmented images, quantifying established geometric descriptors like total length, surface area, diameter distributions, and branching angles [107, 111].

These traditional approaches have provided valuable insights into architectural patterns and enabled large-scale comparative studies across genotypes and environments [112, 113]. The automation of measurement has dramatically improved throughput, consistency, and objectivity compared to manual assessment methods, establishing the foundation for quantitative root genetics and breeding

programs [114, 115]. Recent advances have further enhanced these capabilities through deep learning-based segmentation [116, 117], automated landmark detection [118], and multi-spectral imaging approaches [119, 120].

1.3.2.2 The Fundamental Limitation: The Perceptual Ceiling

However, this remarkable technological progress conceals a fundamental constraint: these automated methods perpetuate and scale the same perceptual limitations that have historically constrained phenotyping [80, 94]. Traditional root traits represent human-conceived, geometrically-defined descriptors that capture only those architectural features that are intuitively apparent to visual inspection [121]. The traits we measure—length, width, angle, surface area—are those we can conceptualise and define, creating what the Interface Theory of Perception describes as a systematic bias toward simplified representations that may not correspond to the underlying biological reality [94].

This perceptual ceiling manifests in several critical ways. First, traditional geometric descriptors fail to capture the emergent, systems-level properties that arise from spatial organisation [80]. While we can quantify individual root segments, we lack effective methods to characterise the higher-order architectural patterns—the spatial coordination, density clustering, and resource allocation strategies—that represent the true adaptive responses to environmental stress [122-124]. Second, the reliance on predefined trait definitions creates an analytical blind spot for novel, non-obvious features that may be critical for adaptation but do not conform to established morphological categories [100, 125].

Perhaps most critically, this approach assumes that the traits visible and interpretable to human perception are those most relevant for biological function [94]. This assumption becomes particularly problematic in the context of stress adaptation, where subtle spatial patterns, density variations, and organisational strategies may be more informative than gross morphological measurements [46, 95, 126, 127]. The complex environmental challenges faced by roots—navigating variable moisture gradients, mechanical impedance, and nutrient patches—

demand sophisticated architectural solutions that may manifest as subtle pixel-level patterns completely invisible to conventional geometric analysis [122, 123].

1.3.2.3 The Biological Reality: Hidden Architectural Intelligence

The biological foundations established in Chapter 2 reveal that successful drought adaptation depends on architectural sophistication that transcends simple geometric descriptions. Root systems deploy strategic biomass allocation patterns, creating dense exploration zones in regions of resource availability while minimising investment in unproductive soil volumes [19, 128]. These adaptive strategies manifest as complex spatial arrangements, density clustering patterns, and hierarchical organisational structures that represent emergent properties of the entire system rather than characteristics of individual components [82, 129].

Recent advances in computer vision and ML demonstrate that digital images contain vastly more information than can be extracted through human-defined geometric descriptors [125, 130]. Unsupervised learning algorithms can identify spatial patterns, density relationships, and organisational features that are completely invisible to human perception but may represent critical adaptive mechanisms [125, 127, 131]. This raises a profound question: if our measurement methods are constrained by human perceptual limitations, how much biologically relevant information are we systematically missing?

1.3.2.4 The Scale of the Problem: Quantifying the Information Loss

The magnitude of this phenotyping bottleneck becomes apparent when considering the information density of biological images. A single high-resolution root system image contains millions of pixels, each representing spatial, intensity, and contextual information that could potentially reveal adaptive architectural patterns [130]. Traditional geometric analysis reduces this rich information landscape to a handful of predefined measurements, representing a massive compression of potentially relevant phenotypic data [80].

Moreover, the complexity of root-environment interactions suggests that adaptive strategies may manifest as subtle, non-linear spatial patterns that emerge only when the full pixel-level complexity is properly analysed [125]. The dense clustering of fine roots in soil patches with optimal moisture conditions, the strategic spacing of lateral roots to optimise exploration efficiency, and the temporal coordination of growth patterns may all represent critical adaptive features that are completely invisible to traditional geometric analysis [30, 128, 132].

1.3.2.5 The Technological Opportunity: Algorithmic Vision for Biological Discovery

The convergence of advances in unsupervised ML, computer vision, and high-throughput imaging creates an unprecedented opportunity to transcend these perceptual limitations [98]. Algorithmic approaches can systematically explore the full spatial complexity of biological images, identifying and quantifying latent features that represent genuine biological patterns rather than human-imposed geometric simplifications [125, 127].

Ensemble methods employing multiple unsupervised learning algorithms can capture complementary aspects of spatial organisation—from density-based clustering to geometric partitioning—providing comprehensive characterisation of architectural patterns [133]. Unlike supervised approaches that require predefined labels, unsupervised methods can discover novel organisational features directly from pixel distributions, enabling true phenotypic discovery rather than automated measurement of known traits [134, 135]

Critically, these algorithmic approaches can maintain interpretability, producing features that can be linked to established biological processes while revealing hidden spatial patterns [127]. This represents a fundamental shift from automating human-defined measurements to achieving genuine computational discovery of biologically relevant architectural features [80, 125].

1.3.2.6 The Imperative for Change: Enabling Discovery-Driven Phenotyping

The phenotyping bottleneck represents more than a technical challenge—it constitutes a fundamental constraint on our ability to understand and improve plant adaptation to climate stress [53, 93]. Traditional approaches, no matter how highly automated, can only discover what we already expect to find [94]. In an era where climate change demands rapid identification of novel adaptive mechanisms, this perceptual constraint becomes a critical limitation for crop improvement [27, 95].

The experimental work presented in this thesis directly addresses this bottleneck through the development of the Algorithmic Root Trait (ART) framework—a computational approach that systematically transcends human perceptual limitations to achieve true phenotypic discovery. By employing ensemble methods with nine distinct unsupervised learning algorithms, ART can identify and quantify dense root clustering patterns, spatial organisation features, and resource allocation strategies that are completely invisible to traditional geometric analysis but represent critical adaptive mechanisms for drought tolerance.

This transition from human-defined measurement to algorithmic discovery represents the essential foundation for unlocking the next generation of crop improvement strategies [22, 125]. Only by transcending our perceptual limitations can we hope to capture the full architectural sophistication that enables plant resilience in an increasingly challenging climate.

1.3.3 The Mechanistic Understanding Bottleneck: From Components to Network Architecture

Even with superior methods for measuring physical traits, a more profound analytical barrier persists: the prevalent reductionist tendency to distill high-dimensional data into simplified lists of ‘significant’ features—the typical output of differential gene expression or GWAS—thereby failing to uncover the principles governing their interaction. This common approach provides what Fernie and Stitt [83] aptly describe as a “parts list” but not an architectural blueprint. While identifying which genes are upregulated or which metabolites accumulate is a necessary first step, it

cannot explain how these components are organised into a functional, resilient system. This bottleneck represents a conceptual failure to move beyond component-level thinking, obscuring the organisational logic that underpins system behaviour and preventing true mechanistic discovery [81, 129].

Biological function, particularly for a complex trait like drought resilience, is an emergent property that arises from the system's underlying architecture [82, 129]. As established by foundational work in network biology, properties like stability, robustness, and efficiency are emergent phenomena, dictated not by individual components in isolation, but by the network's overall architecture—its topology, modularity, and connectivity. Such design principles, which define the system's organization, are fundamentally invisible when analysis is limited to a mere inventory of its parts [85]. The resilience of these systems depends more on their structure and the strategic placement of highly connected ‘hub’ nodes—a hallmark of biological networks—than on the individual properties of their myriad other components [136-138]. A central tenet of modern systems biology is that function is fundamentally an emergent property of structure [82, 85, 129], yet this principle is systematically overlooked by analytical approaches that treat biological features as independent variables.

This mechanistic understanding bottleneck manifests across multiple critical dimensions. Structurally, conventional analyses fragment interconnected biological networks into isolated components, losing the emergent properties that arise from their architecture [83, 85]. Temporally, static ‘snapshot’ analyses fail to capture the dynamic reorganisation of these networks under stress, missing the critical transitions and coordination patterns that distinguish adaptive from maladaptive responses [80, 82]. Functionally, the focus on individual features obscures the systems-level principles—such as modularity, redundancy, and hierarchical organisation—that determine robustness and evolvability [82, 139].

In the context of drought adaptation, this means that identifying a set of accumulated osmolytes is an insufficient endpoint for analysis [17, 24]. The critical question is not just what changes, but how these changes are organised. Do tolerant genotypes

succeed because they possess quantitatively "better" components, or because they arrange those components into a qualitatively "smarter," more efficient network? Answering this fundamental question requires analytical tools capable of decoding organisational strategies rather than simply cataloguing molecular inventories [80, 85].

The consequences of this bottleneck severely constrain the development of predictive models and rational design strategies for crop improvement [27, 53]. Engineering superior drought tolerance requires understanding not just which traits matter, but how they must be coordinated across the whole plant [50, 81]. A deep root system is only beneficial when coupled with appropriate hydraulic conductivity to move water efficiently; enhanced osmotic adjustment is only effective when integrated with compatible metabolic networks; and stress-responsive gene expression is only advantageous when properly timed and tissue-targeted [18, 30, 58]. Without mechanistic insight into these coordination principles, breeding efforts remain largely empirical and are thus too slow to meet the demands of a changing climate.

Network science provides the essential analytical lens to move beyond component cataloguing and toward a holistic understanding of these biological blueprints [81, 85]. By leveraging advances in interpretable ML, it is now possible to decode the topological principles governing biological function, revealing how modular organisation, hierarchical structure, and dynamic reconfiguration enable adaptive responses [82, 129, 140].

This thesis tackles this bottleneck in Chapter 5 by applying network science to decode the tissue-specific metabolic networks that underpin the stress response in wheat. It tests the central hypothesis that tolerant and susceptible genotypes are distinguished not merely by their metabolic constituents, but by the fundamental architectural principles of their metabolic networks. This approach reveals the evolved, tissue-specific network designs—such as the balance between integrated versus modular architectures—that enable coordinated, whole-plant strategies [50, 83]. By focusing on the system's design, we can begin to understand not just what

molecules are involved, but how they are organised to confer resilience at a systems level, thus providing the mechanistic insight that a simple list of features cannot.

1.3.4 The Systems Integration Bottleneck: From Correlation to Coordination

The fourth and arguably most fundamental bottleneck in contemporary plant systems biology lies in our analytical inability to capture the dynamic, directed communication networks that orchestrate biological resilience [80, 84]. While modern high-throughput technologies generate vast multi-modal datasets spanning genomics, transcriptomics, metabolomics, and phenomics, existing integration approaches remain trapped in a paradigm of static correlation analysis that fundamentally misrepresents how biological systems actually function [84, 141]. This systems integration bottleneck prevents us from answering the most crucial question in stress biology: how do plants coordinate responses across multiple biological scales and temporal windows to achieve adaptive resilience?

1.3.4.1 *The Fundamental Limitations of Correlation-Based Integration*

Current multi-omics integration methods, including canonical correlation analysis, partial least squares regression, and even advanced frameworks like DIABLO and mixOmics, share a critical conceptual flaw: they reduce complex biological coordination to shared variance patterns or co-occurring feature sets [141]. These approaches can identify that spectral reflectance changes correlate with metabolite abundance shifts, but they cannot determine which specific spectral bands drive specific metabolic responses, when these interactions occur, or in which direction the influence flows [81, 84]. This limitation is not merely technical but fundamentally paradigmatic—correlation-based methods provide a static snapshot of what appears connected but offer no insight into the dynamic communication protocols that govern adaptive responses [80].

The biological reality is far more sophisticated. Plant stress responses involve hierarchical cascades of coordinated changes where physiological alterations in one tissue trigger specific biochemical responses in another, mediated through directed

signalling networks that operate across multiple temporal scale [18, 48, 49]. For instance, early spectral changes reflecting altered chlorophyll fluorescence in leaves must rapidly communicate with root metabolic networks to coordinate water acquisition strategies, while simultaneously modulating shoot metabolic pathways to optimise resource allocation [142, 143]. These inter-organ communication networks involve specific molecular hubs, definable temporal sequences, and measurable interaction strengths—none of which can be captured by traditional correlation analysis approaches [48, 144, 145].

1.3.4.2 The Directionality Problem: From Association to Causation

Perhaps the most damaging limitation of current integration approaches is their inability to resolve directional relationships between biological modalities. Standard correlation coefficients, canonical variates, and shared latent factors are inherently symmetric measures—they can tell us that two variables change together but provide no information about which variable influences which [80, 84]. This symmetry assumption fundamentally contradicts biological reality, where information flow is inherently directional: environmental stress signals propagate from sensors to responders, gene expression changes drive metabolic flux alterations, and physiological status modulates biochemical machinery activation [67, 146].

This directionality problem has profound implications for mechanistic understanding and practical applications. Without knowing the direction of influence, we cannot distinguish between primary drivers and secondary responses, cannot identify intervention points for crop improvement, and cannot predict system behaviour under novel conditions [27, 53]. Consider drought adaptation: while we may observe that certain spectral bands correlate with specific metabolites in tolerant genotypes, correlation analysis cannot determine whether spectral changes drive metabolic reprogramming or metabolic changes alter spectral signatures—a distinction that is crucial for developing targeted breeding strategies [115, 147].

Recent advances in causal inference and directed graphical models in other fields have begun to address this limitation, but these approaches have seen limited

application in plant biology due to their computational complexity and requirement for specialised statistical expertise [148]. Moreover, most existing causal inference methods assume linear relationships and static network structures, assumptions that are frequently violated in biological systems characterised by non-linear interactions and context-dependent network topology [82, 83].

1.3.4.3 The Feature-Specificity Gap: From Global Patterns to Molecular Mechanisms

Current integration methods excel at identifying broad patterns of shared variation but fall critically short of revealing feature-specific interactions—the precise molecular relationships that drive biological function [84]. Methods like JIVE, iCluster, and MOFA+ can effectively decompose multi-omics variation into interpretable latent factors, but they provide no information about which specific features from different modalities interact with each other [141]. This limitation creates a mechanistic void where we understand that different data types are broadly coordinated but cannot identify the specific molecular players or precise interaction networks responsible for that coordination [80].

This feature-specificity gap is particularly problematic in plant stress biology, where adaptive responses often involve precise molecular switches rather than broad transcriptional or metabolic reprogramming [24, 64]. For example, drought tolerance may depend on the specific coordination between particular wavelengths of chlorophyll fluorescence (indicating photosystem efficiency) and targeted metabolites involved in osmotic adjustment [147, 149]. Understanding such fine-grained molecular coordination requires analytical frameworks capable of quantifying pairwise relationships between individual features across modalities—a capability that is largely absent from existing integration approaches [84].

1.3.4.4 The Temporal Dynamics Problem: Static Snapshots of Dynamic Processes

Perhaps the most critical limitation of current multi-omics integration approaches is their temporal myopia—the reduction of inherently dynamic biological processes to static snapshots of association [80, 84]. Most integration methods analyse

correlation or covariance structures at individual time points or, at best, treat temporal information as a nuisance variable to be controlled rather than a fundamental dimension to be explored [76]. This temporal blindness fundamentally misrepresents biological reality, where the timing of coordination is often more important than the magnitude of association [76, 146].

Stress adaptation is fundamentally a temporal phenomenon involving precisely orchestrated sequences of physiological and biochemical changes [24, 49]. The speed of response initiation, the temporal coordination between different biological compartments, and the duration of sustained responses are all critical determinants of adaptive success [18, 96]. For instance, drought-tolerant genotypes may not necessarily show stronger overall correlation between physiological and metabolic responses, but rather earlier establishment of coordination networks, more rapid response scaling, or better temporal synchronisation between root and shoot responses [115, 150].

The analytical frameworks developed to capture such temporal coordination dynamics must move beyond simple time-series correlation to quantify how relationships between specific features evolve over time, when coordination networks become established, and how the strength and architecture of inter-modal communication changes in response to environmental challenges [80]. Such capabilities require sophisticated approaches that can simultaneously handle high-dimensional feature spaces, heterogeneous data types, and complex temporal dependencies—requirements that exceed the scope of existing integration methods [84].

1.3.4.5 The Biological Imperative: Understanding Coordination as the Foundation of Resilience

The systems integration bottleneck is not merely a technical limitation but a fundamental barrier to understanding how biological resilience emerges from coordinated network function [80, 82]. Resilience is not a property of individual components but an emergent property of coordinated communication between components across multiple scales of biological organisation [83]. A plant's ability

to survive drought stress depends not on the isolated function of root hydraulic conductivity, leaf osmotic adjustment, or metabolic reprogramming, but on the precise temporal coordination of these responses and their dynamic integration into a coherent adaptive strategy [49, 50, 96].

Understanding this coordination requires analytical frameworks capable of quantifying the communication protocols that orchestrate biological responses—frameworks that can identify which specific features communicate with which others, when these communications are established, how strong these interactions are, and how they change over time [80, 84]. Only by developing such capabilities can we move beyond descriptive correlation analysis toward predictive understanding of how biological systems coordinate complex responses and how we might engineer improved coordination for enhanced stress tolerance [27, 53].

1.3.4.6 The Path Forward: From Static Correlation to Dynamic Coordination

Overcoming the systems integration bottleneck requires a paradigmatic shift from static, correlation-based integration toward dynamic, coordination-focused analysis [80]. This transformation demands analytical frameworks that can simultaneously address four critical capabilities: (i) quantifying directed relationships between specific features across modalities, (ii) capturing temporal dynamics of coordination networks, (iii) providing mechanistic interpretability for biological hypothesis generation, and (iv) robustly handling the scale and heterogeneity of modern multi-omics datasets [84].

Such frameworks must integrate the robustness of traditional statistical methods with the flexibility of modern ML approaches, while maintaining the interpretability essential for biological discovery [80, 101]. They must be capable of learning non-linear relationships, handling missing data, accommodating batch effects, and scaling to high-dimensional feature spaces, while producing interpretable outputs that can guide mechanistic hypothesis generation and practical intervention strategies [84].

The experimental chapters of this thesis systematically dismantle these bottlenecks by introducing a suite of purpose-built computational frameworks. While Chapters 4 and 5 introduce tools to solve the phenotyping and mechanistic gaps respectively, the final and most profound challenge—the systems integration bottleneck—is addressed through the development of the MOFA+ Transformer framework. This culminating methodology is an interpretable deep learning approach that combines robust variance decomposition with attention-based mechanisms. It is uniquely designed to quantify the directed, feature-specific, and temporally dynamic relationships between physiological and biochemical responses, moving far beyond simple correlation. By moving beyond correlation toward coordination quantification, this framework provides a new paradigm for decoding the communication networks that orchestrate biological resilience and establishes the analytical foundation required to engineer enhanced stress tolerance in crops facing an increasingly challenging climate [7, 27, 53].

1.4 Methodological Philosophy: Interpretable AI for Biological Discovery

The analytical bottlenecks outlined in the preceding sections demand more than incremental improvements; they require a fundamental paradigmatic shift in how we conceptualise and conduct biological discovery [80]. This thesis is grounded in the conviction that interpretable AI represents the most promising pathway to bridge the chasm between massive data generation and true mechanistic understanding [22, 151]. The central tenet of this methodological philosophy is that computational tools in biology must serve as instruments of discovery—akin to computational microscopes—rather than opaque endpoints of analysis. The frameworks developed herein therefore reject the false dichotomy between predictive power and interpretability, and are founded on three core principles: the necessity of moving beyond "black box" models, the integration of data across multiple biological scales, and the ultimate service of computational innovation to biological understanding.

1.4.1 Beyond the Black Box: The Need for Mechanistic Insight

The recent explosion of ML has delivered remarkable predictive power, yet this success often comes at the cost of transparency, creating "black box" models where the logic connecting input to output is obscured [151, 152]. This opacity represents a fundamental barrier to biological discovery. While a black box may correctly predict *that* a genotype is drought-tolerant, it often fails to explain *why*, a critical omission that renders it a limited tool for scientific advancement [80, 83].

For applied fields like crop improvement, this limitation is a direct obstacle to progress. Agricultural scientists and breeders require not just predictions, but actionable explanations they can trust and build upon [27]. A black box model that achieves high accuracy by learning from ephemeral artefacts in the data—such as lighting variations or non-heritable physiological states—provides a prediction that is scientifically sterile and practically useless for developing new cultivars [100]. Actionable knowledge demands mechanistic insight: the identification of specific, quantifiable, and biologically plausible features that can be targeted for improvement [27, 53].

This thesis therefore moves beyond post-hoc explanations and instead builds frameworks that are interpretable by design [80, 151]. This principle is woven through every experimental chapter to ensure computational outputs translate directly into testable biological hypotheses:

The Algorithmic Root Traits (ARTs) developed in Chapter 4 are not abstract features from a hidden layer; they are quantifiable metrics like HDBSCAN_density_points and FCM_centre_y that link directly to established biological strategies of biomass allocation and rooting depth [19, 70].

The Network Analysis in Chapter 5 intentionally moves beyond a simple list of significant metabolites—a reductionist "parts list"—to decode the network's architectural blueprint, revealing the system's design principles [52, 85].

The MOFA+ Transformer in Chapter 6 uses its attention mechanism not just to improve prediction, but to explicitly quantify the directed influence of a specific physiological sensor (e.g., spectral band W_565) on a specific biochemical effector (e.g., metabolite N_1909), transforming a black box into a transparent communication schematic [80].

Ultimately, this work is motivated by the conviction that predictive accuracy without explanatory power is of limited use in biological science. A model that achieves high performance but remains an inscrutable "black box" offers no pathway to new knowledge. In contrast, a model that illuminates a novel, testable mechanism provides true scientific value, even if its predictive accuracy is more modest. The goal is therefore to create AI that amplifies, rather than replaces, biological inquiry, ensuring that every computational output serves as a stepping stone toward deeper mechanistic understanding.

1.4.2 Multi-Scale Integration: From Traits to Networks

The principle of interpretable AI is incomplete without a commitment to multi-scale integration. Biological resilience is not an isolated trait but an emergent property of a complex, hierarchical system where function arises from the coordinated interplay of parts across scales—from molecules to tissues to the whole organism [18, 49, 50, 81]. A purely reductionist approach is limited to cataloguing the system's components, missing the architectural logic of how they are assembled and coordinated to produce a functional whole. True mechanistic understanding, as argued by pioneers in network biology, requires decoding the system's underlying design principles, as function is an emergent property of structure [82, 85, 129].

This thesis operationalises this philosophy by deconstructing and then reintegrating the layers of adaptation, moving systematically from the visible to the organisational to the dynamic. This multi-scale perspective is built upon three interconnected layers of investigation:

The Physical Layer: An Architectural Foundation. Resilience begins with the plant's physical scaffold for resource acquisition. Capturing this layer requires moving beyond the limits of human perception, as traditional phenotyping is blind to the latent spatial patterns that distinguish adaptive strategies [94, 100, 127].

The Metabolic Layer: An Organisational Blueprint. Built upon the physical foundation is the plant's metabolic organisation—its strategic blueprint for managing stress. This layer is defined not by the mere presence of metabolites, but by their organisation into functional networks with distinct architectural principles that govern the stress response [52, 83, 85].

The Coordination Layer: Communication Protocols for Real-Time Adaptation. The final and most dynamic layer consists of the communication protocols that execute the metabolic strategy. This requires moving beyond static correlation to quantify the directed, feature-specific, and temporal dialogue between a plant's physiological status and its biochemical state [80, 96].

The analytical journey of this thesis is deliberately structured to mirror this biological hierarchy, with each experimental chapter designed to solve a specific bottleneck associated with each layer. First, **Chapter 4** addresses the Phenotyping Bottleneck, using the ART framework to provide a superior, interpretable set of traits for the **Physical Layer**. Next, **Chapter 5** tackles the Mechanistic Understanding Bottleneck, using network science to decode the architectural blueprints of the **Metabolic Layer**. Finally, **Chapter 6** resolves the Systems Integration Bottleneck, introducing the MOFA+ Transformer to quantify the dynamic, cross-modal communication protocols of the **Coordination Layer**.

This journey—from traits, to networks, to integrated systems—represents a holistic approach that recognises function as an emergent property of hierarchical organisation. By systematically building from component measurement through network organisation to dynamic coordination, this approach enables the transition from descriptive to predictive biology, a critical requirement for developing climate-resilient crops [27, 53, 81].

1.4.3 Computational Innovation in Service of Biological Understanding

These principles of interpretability and multi-scale integration culminate in a philosophy where computational tools must be developed in direct service of biological discovery [80]. This work therefore rejects the passive application of generic ML models, instead championing the purpose-driven engineering of novel frameworks designed to overcome the specific analytical bottlenecks that constrain our understanding of complex living systems [22, 90].

Accordingly, this thesis details the deliberate construction of an integrated methodological ecosystem where each component was engineered as a direct response to the failure of existing tools. The analytical journey begins by establishing automation (AutoML, Chapter 3) as a prerequisite for managing modern data scales [99]. From this foundation, the ART framework (Chapter 4) was built to solve the Phenotyping Bottleneck, engineered to algorithmically discover latent spatial patterns invisible to human perception [95, 127]. To move beyond the reductionist "parts list" of high-dimensional analysis [83], the Network Analysis Pipeline (Chapter 5) was developed to solve the Mechanistic Understanding Bottleneck by decoding the system's underlying architectural blueprints [82, 85]. Finally, the MOFA+ Transformer (Chapter 6) was engineered to solve the Systems Integration Bottleneck, providing an interpretable framework to quantify the directed, dynamic communication protocols that orchestrate adaptation across biological scales [76, 80].

This symbiotic relationship—where biological complexity drives computational innovation, which in turn unlocks deeper biological insight—is the engine of this thesis. Its ultimate objective is not simply to build predictive models, but to generate actionable, mechanistic knowledge that can advance our fundamental understanding of plant resilience and guide the development of climate-resilient crops [27, 53]. The frameworks developed herein are the tangible embodiment of this philosophy: a unified platform to decode complexity and transform data into discovery.

1.5 Thesis Aim and Research Questions

1.5.1 Overarching Aim

To systematically address the analytical bottlenecks constraining plant systems biology by developing an integrated suite of interpretable computational frameworks that progress from algorithmic phenotype discovery to network-based mechanistic understanding to dynamic cross-modal coordination, thereby establishing a unified, multi-scale model that transcends traditional limitations and delivers both predictive power and biological insight in wheat drought tolerance.

1.5.2 Research Questions

To realise the overarching aim of this thesis, the analytical challenges are deconstructed into three targeted research questions. Each question is designed to systematically solve one of the major bottlenecks outlined in Section 1.3, progressing logically from refining the measurement of physical components, to decoding their organisational principles, and finally to quantifying their dynamic, cross-modal integration. Together, they form a cohesive investigative strategy to build a multi-scale, mechanistic model of drought resilience. This progression systematically builds from phenotypic discovery to mechanistic understanding to systems integration, with each question providing essential foundations for the next.

Research Question 1 (The Phenotyping Gap | Addresses Bottleneck 1.3.2): Can the computational extraction of latent architectural traits from root images provide a novel source of information that is complementary to traditional metrics, and does their integration lead to more robust and insightful predictive models?

This question directly addresses the **Phenotyping Bottleneck (Section 1.3.2)**, which postulates that our ability to measure biological form is constrained by the limits of human perception [94]. While high-throughput imaging platforms can automate the extraction of traditional root traits (TRTs) like length and diameter, these metrics are

fundamentally human-conceived and may fail to capture the subtle, non-obvious architectural patterns that are critical for adaptation [46, 95, 153]. Drought resilience in roots often manifests as strategic biomass allocation and complex spatial arrangements—emergent properties that are poorly described by simple geometric descriptors [19]. This research question tests the hypothesis that unsupervised ML can function as a form of "algorithmic vision" to discover and quantify latent, yet biologically interpretable, architectural features directly from pixel data [23, 127]. The goal is not to replace traditional metrics, which provide a valuable foundational assessment, but to investigate whether these algorithmically-derived traits offer a complementary and novel layer of information. It seeks to determine if the synergistic integration of both traditional and algorithmic traits can yield a more comprehensive understanding of root architecture and, consequently, more robust and insightful predictive models than either approach can achieve in isolation.

Research Question 2 (The Mechanistic Basis | Addresses Bottleneck 1.3.3): What are the underlying architectural principles and temporal dynamics of tissue-specific metabolic networks that drive the systemic stress response in wheat?

This question addresses the **Mechanistic Understanding Bottleneck (Section 1.3.3)**, which challenges the reductionist tendency in systems biology to produce lists of significant features without revealing the principles governing their interaction [83]. True mechanistic insight requires moving beyond a "parts list" to understand the system's architectural blueprint, as function is an emergent property of network topology [82, 85, 129]. Given that roots and shoots perform distinct roles in drought adaptation—water acquisition versus photosynthetic maintenance—it is hypothesised that they employ fundamentally different, tissue-specific metabolic network designs (e.g., integrated vs. modular) and temporal strategies [50, 52]. This research question employs network science to decode these organisational principles, seeking to understand not just *what* molecules are involved in the stress response, but *how* they are organised to confer resilience at a systems level.

Research Question 3 (The Dynamic Integration Gap | Addresses Bottleneck 1.3.4): What time-resolved coordination patterns between physiological and

biochemical responses distinguish drought tolerant and susceptible genotypes, and can interpretable deep learning reveal these cross-modal relationships?

This final and most integrative question targets the Systems Integration Bottleneck (**Section 1.3.4**). Biological resilience is not a static state but a dynamic process orchestrated through continuous, multi-modal communication between a plant's physiological status and its biochemical machinery [18, 48]. Current integration methods, largely based on static correlation, are ill-equipped to capture the directionality, feature-specificity, and temporal dynamics of this dialogue [84, 141, 154]. This RQ tests the central hypothesis that superior drought tolerance emerges from superior coordination capacity—quantifiably stronger, earlier, and more precisely directed communication networks [77, 78]. To overcome the limitations of existing tools, this inquiry necessitates the development of a novel, interpretable deep learning framework. Unlike black-box approaches, this framework must provide transparent, mechanistic insights into which specific physiological signals drive which biochemical responses, capable of moving beyond correlation to quantify the directed, time-resolved coordination between physiological sensing and metabolic response, thereby decoding the communication protocols of resilience [80].

1.6 Thesis Outline and Analytical Progression

This thesis is structured as a logical and hierarchical response to the grand challenge of decoding plant resilience. The investigation systematically progresses from establishing the biological and computational foundations to developing and applying a series of increasingly sophisticated analytical frameworks.

1.6.1 A Journey Through the Chapters

The thesis begins in **Chapter 2** with a comprehensive review of the multifaceted nature of root adaptation to drought. It moves beyond a singular focus on architecture to synthesise the critical, yet often-overlooked, contributions of structural, anatomical, hydraulic, and biochemical traits, as well as their dynamic

interaction with the soil microbiome. This chapter establishes the deep biological complexity of the system and provides the essential foundation for understanding why traditional analytical approaches have proven insufficient for decoding drought resilience [95].

Chapter 3 presents a perspective on AutoML, establishing the high-level computational philosophy that a new generation of automated tools is required to tackle the scale of modern biological data. This chapter directly addresses the overarching **Analytical Complexity Bottleneck** by demonstrating how automation can democratise advanced analytics and provide the scalable, reproducible frameworks essential for modern plant systems biology [99, 102].

The experimental work commences in **Chapter 4**, which introduces the Algorithmic Root Trait (ART) framework—a bespoke computational tool developed to solve the **Phenotyping Bottleneck**. Through an ensemble of nine unsupervised learning algorithms, this chapter demonstrates how algorithmic vision can transcend human perceptual limitations to discover and extract biologically meaningful patterns that are invisible to conventional measurement paradigms [94, 127].

Chapter 5 addresses the **Mechanistic Understanding Bottleneck** by using network science to decode the architectural principles of tissue-specific metabolic networks. This chapter moves beyond the traditional reductionist approach of cataloguing significant metabolites to reveal the organisational blueprints—topology, connectivity, and modularity—that determine how metabolic systems function under stress [83, 85].

Chapter 6 tackles the final and most integrative challenge, the **Systems Integration Bottleneck**, by introducing the MOFA+ Transformer. This interpretable deep learning framework quantifies the dynamic coordination between physiological and metabolic data streams, transcending static correlation to decode the directed, feature-specific communication protocols that orchestrate adaptive responses across biological scales and temporal windows [80, 84].

Finally, **Chapter 7**, the General Discussion, synthesises the key findings from the preceding experimental chapters. It revisits the core research questions, presenting a unified, multi-scale model of drought resilience that connects the discoveries from the ART, metabolic network, and MOFA+ Transformer analyses, before discussing the broader implications for plant systems biology and future crop improvement strategies [27, 53].

1.6.2 A Synergistic Methodological Ecosystem

This thesis does not present a series of disconnected solutions, but rather the development of an integrated methodological ecosystem designed to tackle biological complexity through coordinated computational innovation. The analytical journey is deliberately designed to build in sophistication, progressing from a broad call for automation (AutoML), to a targeted solution for a specific data type (ART), to the deep analysis of molecular network architecture, and culminating in a framework that integrates these different scales of biological information (MOFA+ Transformer).

This approach champions interpretable AI, where each framework is designed to amplify human insight by revealing patterns and relationships that are impossible to detect through traditional means, rather than replacing biological expertise with opaque "black-box" automation [151, 152]. The integration of these frameworks creates emergent analytical capabilities that exceed the sum of their individual contributions. The ART-derived phenotypes provide novel inputs for network analysis, the network architectural principles inform the interpretation of cross-modal coordination patterns, and the MOFA+ Transformer reveals how these features and organisational strategies ultimately translate into systemic resilience. This methodological synergy exemplifies the potential for computational approaches to transform plant systems biology from a descriptive to a predictive science [80, 90]. The analytical progression detailed above, from identifying the grand challenge to the development of a synergistic computational ecosystem, is visually summarised in Figure 1. This schematic provides a conceptual roadmap for the thesis, illustrating how each chapter systematically addresses a core analytical bottleneck to build towards a unified, multi-scale understanding of plant resilience.

Thesis Analytical Workflow

A roadmap from the grand challenge to integrated, multi-scale insight

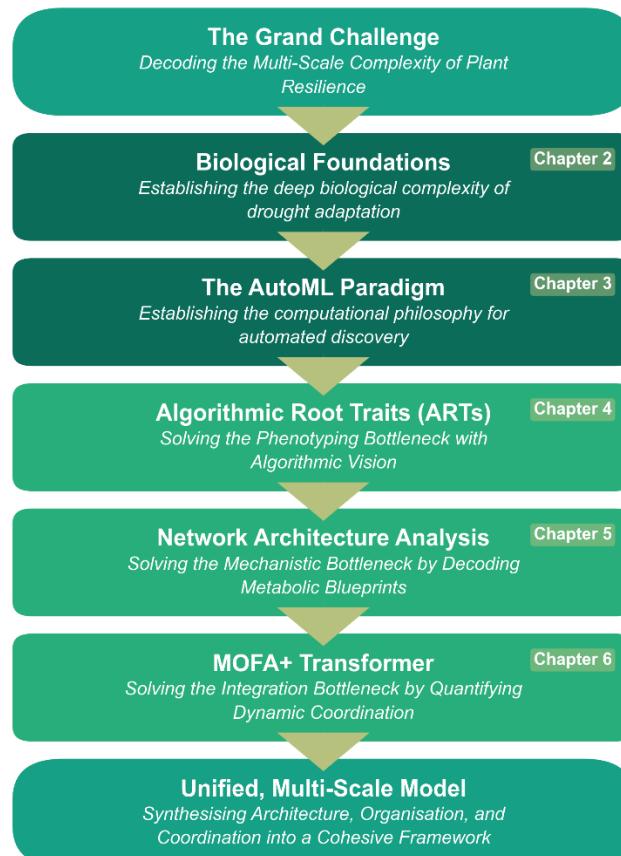


Fig. 1. Thesis Analytical Workflow. This flowchart provides a roadmap of the thesis, illustrating the progression from the overarching scientific challenge to the development of a unified, multi-scale model of plant resilience. The workflow begins by establishing The Grand Challenge and the necessary Biological and Computational Foundations (Chapters 1-3). It then details the development of the core methodological contributions designed to solve specific analytical bottlenecks: the Algorithmic Root Traits (ARTs) framework (Chapter 4), Network Architecture Analysis (Chapter 5), and the MOFA+ Transformer (Chapter 6). The thesis culminates in the synthesis of these findings into a Unified, Multi-Scale Model (Chapter 7), representing an integrated understanding of drought adaptation.

1.7 Research Outcomes and Significance

This thesis was designed to make significant and distinct contributions to both fundamental plant systems biology and applied crop improvement. By systematically developing and applying an integrated computational ecosystem, this work moved beyond incremental advances to establish new analytical paradigms and deliver unprecedented mechanistic insights. The significance of this research is

articulated across three key domains: methodological innovation, mechanistic discovery, and applied impact for future agriculture.

1.7.1 Methodological Innovations for Plant Systems Biology

This research delivered a suite of novel, open-source computational frameworks (ART, MOFA+ Transformer) that were purpose-built to solve long-standing analytical bottlenecks in plant science. The methodological contributions are threefold:

A New Class of Algorithmic Phenotypes: This research introduced Algorithmic Root Traits (ARTs), demonstrating that computational discovery can yield phenotypes with superior predictive power and information density compared to traditional, human-defined methods. The research established a new paradigm for image-based phenotyping, moving from the automation of known measurements to the de novo discovery of biologically informative features directly from sensor data [22, 127].

A Shift from Components to Architecture: The network analysis pipeline developed provides the tools to move beyond a reductionist "parts list" of significant genes or metabolites [83]. It empowers researchers to decode the architectural blueprints governing system behaviour, providing quantitative evidence that biological function is an emergent property of network organisation [82, 85].

A Leap from Correlation to Interpretable Coordination: This research pioneered the use of interpretable deep learning (MOFA+ Transformer) to quantify the directed, dynamic, and feature-specific relationships in multi-omics data. It transforms systems biology from a correlative exercise, typified by methods like WGCNA [154], toward the discovery of mechanistic coordination, providing a clear path from high-dimensional data to testable, systems-level hypotheses [80].

1.7.2 Mechanistic Insights into Drought Adaptation

The application of these novel frameworks yielded profound biological insights, providing a new, systems-level understanding of plant resilience.

A Unified, Multi-Scale Model of Resilience: This research provided one of the first systems-level models of drought resilience in wheat that quantitatively integrates architectural, physiological, and metabolic data across a temporal scale. It revealed that resilience is not determined by any single trait but emerges from coordinated optimisation across hierarchical layers: the Physical Architecture, the Metabolic Organisation, and the Coordination Protocols. [18, 81].

Discovery of Tissue-Specific Network Blueprints: This research was designed to uncover the fundamental architectural principles of tissue-specific metabolic networks. It demonstrated that leaves and roots employ distinct, complementary strategies to manage osmotic stress. It was hypothesised and subsequently shown that leaf networks exhibit densely integrated architectures for rapid coordination, while root networks are modular and fragmented for flexible, localised adaptation—a key organisational principle previously unquantified in wheat [50, 52].

Quantification of Coordination as a Key Trait: This research established that the *timing* and *strength* of network coordination are critical, measurable components of drought tolerance. It provided quantitative evidence that tolerant genotypes exhibit superior coordination capacity, establishing critical communication networks at earlier stress stages. This finding reframes resilience not as a static property, but as a dynamic capability [24, 76].

1.7.3 Applied Implications for Crop Improvement

The methodological and biological breakthroughs of this thesis are expected to have direct and actionable implications for accelerating the development of climate-resilient crops.

Novel Biomarkers for Precision Breeding: This research delivered a new class of mechanistically grounded and highly predictive biomarkers. This work established that specific ARTs and network properties like "coordination strength" are stable, genotype-specific metrics that directly quantify adaptive strategies. This robust

differentiation between genotypes provides a strong foundation for future genetic studies to investigate their heritability. If confirmed as heritable, these traits would represent excellent candidates for integration into modern genomic prediction models to enhance selection accuracy and accelerate the development of crops with 'designer roots' tailored for water-limited environments [17, 27, 155].

Solving the Phenotyping and Analysis Bottleneck: The automated, high-throughput nature of the ART and MOFA+ Transformer frameworks provides scalable analytical pipelines designed to directly address the primary rate-limiting step in modern breeding [23, 93]. Their demonstrated stability across experimental conditions is essential for deployment in multi-environment trials, accelerating modern strategies like 'speed breeding' and enabling the vision of AI-driven smart agriculture [22, 25].

Enabling Climate-Smart Agriculture: By providing mechanistic insights into water-use efficiency, this research supports the development of cultivars that maintain productivity with fewer inputs [156, 157]. By providing insights into the root architectures and coordination strategies associated with efficient water use, this research could help identify traits that enhance plant performance. In turn, such traits have the potential to contribute to more sustainable agricultural practices by improving soil health and promoting carbon sequestration [158, 159].

In summary, this thesis aimed to offer a significant contribution to global scientific knowledge by providing a new perspective on biological resilience, a validated suite of computational tools to decode it, and a clear pathway for translating these discoveries into the climate-resilient crops of the future.

References

- [1] N. Alexandratos and J. Bruinsma, "World agriculture towards 2030/2050: the 2012 revision," 2012.
- [2] U. Desa, "United nations department of economic and social affairs, population division. world population prospects: The 2015 revision, key findings and advance tables," *Online Edition UN DESA, New York*, 2015.
- [3] P. R. Shukla *et al.*, "Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems," 2019.
- [4] J. A. Foley *et al.*, "Solutions for a cultivated planet," *Nature*, vol. 478, no. 7369, pp. 337-342, 2011, doi: 10.1038/nature10452.
- [5] V. Masson-Delmotte *et al.*, "Global warming of 1.5 C," *An IPCC Special Report on the impacts of global warming of*, vol. 1, no. 5, 2018.
- [6] R. M. Rivero, R. Mittler, E. Blumwald, and S. I. Zandalinas, "Developing climate-resilient crops: Improving plant tolerance to stress combination," *The Plant Journal*, 2021, doi: 10.1111/tpj.15483.
- [7] J. Salse, R. L. Barnard, C. Veneault-Fourrey, and H. Rouached, "Strategies for breeding crops for future environments," *Trends in Plant Science*, vol. 29, no. 3, pp. 303-318, 2024, doi: 10.1016/j.tplants.2023.08.007.
- [8] F. Chiang, O. Mazdiyasni, and A. AghaKouchak, "Evidence of anthropogenic impacts on global drought frequency, duration, and intensity," *Nature Communications*, vol. 12, no. 1, p. 2754, 2021/05/12 2021, doi: 10.1038/s41467-021-22314-w.
- [9] J. Sanderman, T. Hengl, and G. J. Fiske, "Soil carbon debt of 12,000 years of human land use," *Proceedings of the National Academy of Sciences*, vol. 114, no. 36, pp. 9575-9580, 2017, doi: 10.1073/pnas.1706103114.
- [10] H. Chen and J.-G. Jiang, "Osmotic adjustment and plant adaptation to environmental changes related to drought and salinity," *Environmental Reviews*, vol. 18, no. NA, pp. 309-319, 2010.

- [11] A. Boretti and L. Rosa, "Reassessing the projections of the World Water Development Report," *npj Clean Water*, vol. 2, no. 1, 2019, doi: 10.1038/s41545-019-0039-9.
- [12] O. Erenstein, M. Jaleta, K. A. Mottaleb, K. Sonder, J. Donovan, and H.-J. Braun, "Global trends in wheat production, consumption and trade," in *Wheat improvement: food security in a changing climate*, M. P. Reynolds and H.-J. Braun Eds. Cham: Springer International Publishing, 2022, pp. 47-66.
- [13] H. Webber *et al.*, "Diverging importance of drought stress for maize and winter wheat in Europe," *Nature Communications*, vol. 9, no. 1, 2018, doi: 10.1038/s41467-018-06525-2.
- [14] K. S. Nelson and E. K. Burchfield, "Landscape complexity and US crop production," *Nature Food*, vol. 2, no. 5, pp. 330-338, 2021/05/01 2021, doi: 10.1038/s43016-021-00281-1.
- [15] W. Xiong *et al.*, "Increased ranking change in wheat breeding under climate change," *Nature Plants*, vol. 7, no. 9, pp. 1207-1212, Sep 2021, doi: 10.1038/s41477-021-00988-w.
- [16] H. Zhang, Y. Li, and J.-K. Zhu, "Developing naturally stress-resistant crops for a sustainable agriculture," in *Nature plants* vol. 4, 2018/11/28 ed, 2018, pp. 989-996.
- [17] A. Sallam, A. M. Alqudah, M. F. Dawood, P. S. Baenziger, and A. Börner, "Drought stress tolerance in wheat and barley: advances in physiology, breeding and genetics research," *International journal of molecular sciences*, vol. 20, no. 13, p. 3137, Jun 27 2019, doi: 10.3390/ijms20133137.
- [18] J.-K. Zhu, "Abiotic Stress Signaling and Responses in Plants," *Cell*, vol. 167, no. 2, pp. 313-324, 2016, doi: 10.1016/j.cell.2016.08.029.
- [19] J. P. Lynch, "Steep, cheap and deep: an ideotype to optimize water and N acquisition by maize root systems," *Annals of Botany*, vol. 112, no. 2, pp. 347-357, 2013, doi: 10.1093/aob/mcs293.
- [20] Y. Uga *et al.*, "Control of root system architecture by DEEPER ROOTING 1 increases rice yield under drought conditions," *Nature Genetics*, vol. 45, no. 9, pp. 1097-1102, Sep 2013, doi: 10.1038/ng.2725.
- [21] R. R. Mir, M. Zaman-Allah, N. Sreenivasulu, R. Trethowan, and R. K. Varshney, "Integrated genomics, physiology and breeding approaches for improving

- drought tolerance in crops," *Theoretical and Applied Genetics*, vol. 125, no. 4, pp. 625-645, 2012/08/01 2012, doi: 10.1007/s00122-012-1904-9.
- [22] A. L. Harfouche, F. Nakhle, A. H. Harfouche, O. G. Sardella, E. Dart, and D. Jacobson, "A primer on artificial intelligence in plant digital phenomics: embarking on the data to insights journey," *Trends in Plant Science*, 2022.
- [23] T. Gill, S. K. Gill, D. K. Saini, Y. Chopra, J. P. de Koff, and K. S. Sandhu, "A Comprehensive Review of High Throughput Phenotyping and Machine Learning for Plant Stress Phenotyping," *Phenomics*, pp. 1-28, 2022.
- [24] K. Shinozaki and K. Yamaguchi-Shinozaki, "Gene networks involved in drought stress response and tolerance," (in eng), *J Exp Bot*, vol. 58, no. 2, pp. 221-7, 2007, doi: 10.1093/jxb/erl164.
- [25] S. Ghosh *et al.*, "Speed breeding in growth chambers and glasshouses for crop breeding and model plant research," *Nature Protocols*, vol. 13, no. 12, pp. 2944-2963, 2018, doi: 10.1038/s41596-018-0072-z.
- [26] Y. Xu *et al.*, "Feeding the world using speed breeding technology," *Trends in Plant Science*, 2022.
- [27] Y. Xu *et al.*, "Smart breeding driven by big data, artificial intelligence, and integrated genomic-enviromic prediction," *Molecular Plant*, vol. 15, no. 11, pp. 1664-1695, 2022.
- [28] R. K. Varshney *et al.*, "Fast-forward breeding for a food-secure world," *Trends in Genetics*, 2021, doi: 10.1016/j.tig.2021.08.002.
- [29] J. P. Lynch, "Harnessing root architecture to address global challenges," *The Plant Journal*, vol. 109, no. 2, pp. 415-431, 2022, doi: 10.1111/tpj.15560.
- [30] C. Maurel and P. Nacry, "Root architecture and hydraulics converge for acclimation to changing water availability," *Nature Plants*, vol. 6, no. 7, pp. 744-749, 2020, doi: 10.1038/s41477-020-0684-5.
- [31] L. Mommer, P. Hinsinger, C. Prigent-Combaret, and E. J. W. Visser, "Advances in the rhizosphere: stretching the interface of life," *Plant and Soil*, vol. 407, no. 1-2, pp. 1-8, 2016, doi: 10.1007/s11104-016-3040-9.
- [32] R. Antoni, D. Dietrich, M. J. Bennett, and P. L. Rodriguez, "Hydrotropism: Analysis of the Root Response to a Moisture Gradient," Springer New York, 2016, pp. 3-9.

- [33] N. E. Robbins and J. R. Dinneny, "Growth is required for perception of water availability to pattern root branches in plants," *Proceedings of the National Academy of Sciences*, vol. 115, no. 4, pp. E822-E831, 2018.
- [34] R. E. Jaramillo, E. A. Nord, J. G. Chimungu, K. M. Brown, and J. P. Lynch, "Root cortical burden influences drought tolerance in maize," *Annals of Botany*, vol. 112, no. 2, pp. 429-437, 2013, doi: 10.1093/aob/mct069.
- [35] P. Saengwilai, E. A. Nord, J. G. Chimungu, K. M. Brown, and J. P. Lynch, "Root Cortical Aerenchyma Enhances Nitrogen Acquisition from Low-Nitrogen Soils in Maize," *Plant Physiology*, vol. 166, no. 2, pp. 726-735, 2014, doi: 10.1104/pp.114.241711.
- [36] R. Aroca, R. Porcel, and J. M. Ruiz-Lozano, "How does arbuscular mycorrhizal symbiosis regulate root hydraulic properties and plasma membrane aquaporins in Phaseolus vulgaris under drought, cold or salinity stresses?," *New Phytologist*, vol. 173, no. 4, pp. 808-816, 2007, doi: 10.1111/j.1469-8137.2006.01961.x.
- [37] M. Calvo-Polanco *et al.*, "Physiological roles of Casparyan strips and suberin in the transport of water and solutes," *New Phytologist*, vol. 232, no. 6, pp. 2295-2307, 2021, doi: 10.1111/nph.17765.
- [38] C. Maurel, Y. Boursiac, D.-T. Luu, V. Santoni, Z. Shahzad, and L. Verdoucq, "Aquaporins in plants," *Physiological reviews*, vol. 95, no. 4, pp. 1321-1358, 2015.
- [39] S. A. Rolfe, J. Griffiths, and J. Ton, "Crying out for help with root exudates: adaptive mechanisms by which stressed plants assemble health-promoting soil microbiomes," *Current Opinion in Microbiology*, vol. 49, pp. 73-82, Jun 2019, doi: 10.1016/j.mib.2019.10.003.
- [40] M. Knott, M. Ani, E. Kroener, and D. Diehl, "Effect of environmental conditions on physical properties of maize root mucilage," 2022.
- [41] P. Mathur and S. Roy, "Insights into the plant responses to drought and decoding the potential of root associated microbiome for inducing drought tolerance," *Physiologia Plantarum*, vol. 172, no. 2, pp. 1016-1029, 2021, doi: 10.1111/ppl.13338.

- [42] J. Banda *et al.*, "Lateral Root Formation in Arabidopsis: A Well-Ordered LRexit," *Trends in Plant Science*, vol. 24, no. 9, pp. 826-839, 2019, doi: 10.1016/j.tplants.2019.06.015.
- [43] Y. Sher *et al.*, "Microbial extracellular polysaccharide production and aggregate stability controlled by switchgrass (*Panicum virgatum*) root biomass and soil water potential," *Soil Biology and Biochemistry*, vol. 143, p. 107742, 2020.
- [44] J. P. Lynch, "Root phenotypes for improved nutrient capture: an underexploited opportunity for global agriculture," *New Phytologist*, vol. 223, no. 2, pp. 548-564, 2019, doi: 10.1111/nph.15738.
- [45] J. P. Lynch, S. J. Mooney, C. F. Strock, and H. M. Schneider, "Future roots for future soils," *Plant, Cell & Environment*, vol. 45, no. 3, pp. 620-636, Mar 2022, doi: 10.1111/pce.14213.
- [46] J. A. Atkinson, M. P. Pound, M. J. Bennett, and D. M. Wells, "Uncovering the hidden half of plants using new advances in root phenotyping," *Current Opinion in Biotechnology*, vol. 55, pp. 1-8, 2019, doi: 10.1016/j.copbio.2018.06.002.
- [47] M. L. McCormack *et al.*, "Building a better foundation: improving root-trait measurements to understand and model plant and ecosystem processes," *New Phytologist*, vol. 215, no. 1, pp. 27-37, 2017, doi: 10.1111/nph.14459.
- [48] H. Li, C. Testerink, and Y. Zhang, "How roots and shoots communicate through stressful times," *Trends in Plant Science*, vol. 26, no. 9, pp. 940-952, 2021, doi: 10.1016/j.tplants.2021.03.005.
- [49] H. Zhang, J. Zhu, Z. Gong, and J.-K. Zhu, "Abiotic stress responses in plants," *Nature Reviews Genetics*, 2021/09/24 2021, doi: 10.1038/s41576-021-00413-0.
- [50] X. Min, X. Lin, B. Ndayambaza, Y. Wang, and W. Liu, "Coordinated mechanisms of leaves and roots in response to drought stress underlying full-length transcriptome profiling in *Vicia sativa* L," *BMC Plant Biology*, vol. 20, no. 1, 2020, doi: 10.1186/s12870-020-02358-8.
- [51] P. Filippou *et al.*, "Systems biology reveals key tissue-specific metabolic and transcriptional signatures involved in the response of plant genotypes to salt

- stress," *Computational and Structural Biotechnology Journal*, vol. 19, pp. 2133-2147, 2021, doi: 10.1016/j.csbj.2021.04.018.
- [52] A. Gargallo-Garriga *et al.*, "Opposite metabolic responses of shoots and roots to drought," *Scientific Reports*, vol. 4, no. 1, p. 6829, 2014/10/29 2014, doi: 10.1038/srep06829.
- [53] M. Reynolds *et al.*, "Addressing Research Bottlenecks to Crop Productivity," *Trends in Plant Science*, vol. 26, no. 6, pp. 607-630, 2021, doi: 10.1016/j.tplants.2021.03.011.
- [54] Z. Shahzad *et al.*, "A Potassium-Dependent Oxygen Sensing Pathway Regulates Plant Root Hydraulics," *Cell*, vol. 167, no. 1, pp. 87-98.e14, 2016, doi: 10.1016/j.cell.2016.08.068.
- [55] A. Carminati and M. Javaux, "Soil Rather Than Xylem Vulnerability Controls Stomatal Response to Drought," *Trends in Plant Science*, vol. 25, no. 9, pp. 868-880, 2020, doi: 10.1016/j.tplants.2020.04.003.
- [56] E. Steudle and C. A. Peterson, "How does water get through roots?," *Journal of Experimental Botany*, vol. 49, no. 322, pp. 775-788, 1998, doi: 10.1093/jxb/49.322.775.
- [57] J. Zhang, X. Zhang, and J. Liang, "Exudation rate and hydraulic conductivity of maize roots are enhanced by soil drying and abscisic acid treatment," *New Phytologist*, vol. 131, no. 3, pp. 329-336, 1995, doi: 10.1111/j.1469-8137.1995.tb03068.x.
- [58] N. Tang, Z. Shahzad, F. Lonjon, O. Loudet, F. Vailleau, and C. Maurel, "Natural variation at XND1 impacts root hydraulics and trade-off for stress responses in Arabidopsis," *Nature Communications*, vol. 9, no. 1, p. 3884, Sep 24 2018, doi: 10.1038/s41467-018-06430-8.
- [59] M. A. Zwieniecki and L. Boersma, "A technique to measure root tip hydraulic conductivity and root water potential simultaneously," *Journal of experimental botany*, vol. 48, no. 2, pp. 333-336, 1997.
- [60] W. Siao, D. Coskun, F. Baluška, H. J. Kronzucker, and W. Xu, "Root-Apex Proton Fluxes at the Centre of Soil-Stress Acclimation," *Trends in Plant Science*, vol. 25, no. 8, pp. 794-804, 2020, doi: 10.1016/j.tplants.2020.03.002.

- [61] H. J. Lee and P. J. Seo, "Ca(2+)-Mediating Initial Responses to Environmental Stresses," *Trends in Plant Science*, vol. 26, no. 8, pp. 849-870, Aug 2021, doi: 10.1016/j.tplants.2021.02.007.
- [62] S. I. Zandalinas and R. Mittler, "Plant responses to multifactorial stress combination," *New Phytologist*, 2022.
- [63] M. M. Chaves, J. Flexas, and C. Pinheiro, "Photosynthesis under drought and salt stress: regulation mechanisms from whole plant to cell," *Annals of Botany*, vol. 103, no. 4, pp. 551-560, 2008, doi: 10.1093/aob/mcn125.
- [64] K. A. H. Hewage, J. F. Yang, D. Wang, G. F. Hao, G. F. Yang, and J. K. Zhu, "Chemical Manipulation of Abscisic Acid Signaling: A New Approach to Abiotic and Biotic Stress Management in Agriculture," *Advanced Science*, vol. 7, no. 18, p. 2001265, 2020, doi: 10.1002/advs.202001265.
- [65] F. Chaumont and S. D. Tyerman, "Aquaporins: highly regulated channels controlling plant water relations," *Plant physiology*, vol. 164, no. 4, pp. 1600-1618, 2014.
- [66] A. R. Fernie and T. Tohge, "The genetics of plant metabolism," *Annual Review of Genetics*, vol. 51, no. 1, pp. 287-310, 2017.
- [67] Y. Yu, Y. Zhang, X. Chen, and Y. Chen, "Plant Noncoding RNAs: Hidden Players in Development and Stress Responses," *Annual Review of Cell and Developmental Biology*, vol. 35, no. 1, pp. 407-431, 2019, doi: 10.1146/annurev-cellbio-100818-125218.
- [68] S. L. H. Cheng, H. Xu, J. H. T. Ng, and N. H. Chua, "Systemic movement of long non-coding RNA ELENA1 attenuates leaf senescence under nitrogen deficiency," *Nature Plants*, vol. 9, no. 10, pp. 1598-1606, Oct 2023, doi: 10.1038/s41477-023-01521-x.
- [69] Q. Chen *et al.*, "Phosphorylation of SWEET sucrose transporters regulates plant root:shoot ratio under drought," *Nature Plants*, vol. 8, no. 1, pp. 68-77, Jan 2022, doi: 10.1038/s41477-021-01040-7.
- [70] A. Eziz, Z. Yan, D. Tian, W. Han, Z. Tang, and J. Fang, "Drought effect on plant biomass allocation: A meta-analysis," *Ecology and Evolution*, vol. 7, no. 24, pp. 11002-11010, 2017, doi: 10.1002/ece3.3630.
- [71] H. Bacher, Y. Sharaby, H. Walia, and Z. Peleg, "Modifying root-to-shoot ratio improves root water influxes in wheat under drought stress," *Journal of*

- Experimental Botany*, vol. 73, no. 5, pp. 1643-1654, Mar 2 2021, doi: 10.1093/jxb/erab500.
- [72] H. Poorter, K. J. Niklas, P. B. Reich, J. Oleksyn, P. Poot, and L. Mommer, "Biomass allocation to leaves, stems and roots: meta-analyses of interspecific variation and environmental control," *New Phytologist*, vol. 193, no. 1, pp. 30-50, 2012, doi: 10.1111/j.1469-8137.2011.03952.x.
- [73] T. Obata and A. R. Fernie, "The use of metabolomics to dissect plant responses to abiotic stresses," (in eng), *Cell Mol Life Sci*, vol. 69, no. 19, pp. 3225-43, Oct 2012, doi: 10.1007/s00018-012-1091-5.
- [74] M. J. Paul, L. F. Primavesi, D. Jhurreea, and Y. Zhang, "Trehalose metabolism and signaling," *Annual Review of Plant Biology*, vol. 59, pp. 417-441, 2008.
- [75] Z. Kang *et al.*, "Comparative metabolomic profiling in the roots and leaves in contrasting genotypes reveals complex mechanisms involved in post-anthesis drought tolerance in wheat," *PLOS ONE*, vol. 14, no. 3, p. e0213502, 2019, doi: 10.1371/journal.pone.0213502.
- [76] Z. Bar-Joseph, A. Gitter, and I. Simon, "Studying and modelling dynamic biological processes using time-series gene expression data," *Nature Reviews Genetics*, vol. 13, no. 8, pp. 552-564, 2012/08/01 2012, doi: 10.1038/nrg3244.
- [77] A. Skirycz *et al.*, "Survival and growth of Arabidopsis plants given limited water are not equal," *Nature Biotechnology*, vol. 29, no. 3, pp. 212-214, 2011/03/01 2011, doi: 10.1038/nbt.1800.
- [78] D. Todaka, K. Shinozaki, and K. Yamaguchi-Shinozaki, "Recent advances in the dissection of drought-stress regulatory networks and strategies for development of drought-tolerant transgenic rice plants," *Frontiers in plant science*, vol. 6, p. 84, 2015.
- [79] Z. Miao, Z. Han, T. Zhang, S. Chen, and C. Ma, "A systems approach to a spatio-temporal understanding of the drought stress response in maize," *Scientific Reports*, vol. 7, no. 1, p. 6590, 2017/07/26 2017, doi: 10.1038/s41598-017-06929-y.
- [80] D. M. Camacho, K. M. Collins, R. K. Powers, J. C. Costello, and J. J. Collins, "Next-Generation Machine Learning for Biological Networks," *Cell*, vol. 173, no. 7, pp. 1581-1592, 2018/06/14/ 2018, doi: <https://doi.org/10.1016/j.cell.2018.05.015>.

- [81] A. R. Joyce and B. Ø. Palsson, "The model organism as a system: integrating 'omics' data sets," *Nature Reviews Molecular Cell Biology*, vol. 7, no. 3, pp. 198-210, 2006/03/01 2006, doi: 10.1038/nrm1857.
- [82] A.-L. Barabási and Z. N. Oltvai, "Network biology: understanding the cell's functional organization," *Nature Reviews Genetics*, vol. 5, no. 2, pp. 101-113, 2004, doi: 10.1038/nrg1272.
- [83] A. R. Fernie and M. Stitt, "On the Discordance of Metabolomics with Proteomics and Transcriptomics: Coping with Increasing Complexity in Logic, Chemistry, and Network Interactions Scientific Correspondence," *Plant Physiology*, vol. 158, no. 3, pp. 1139-1145, 2012, doi: 10.1104/pp.112.193235.
- [84] I. Subramanian, S. Verma, S. Kumar, A. Jere, and K. Anamika, "Multi-omics Data Integration, Interpretation, and Its Application," *Bioinformatics and Biology Insights*, vol. 14, p. 1177932219899051, 2020, doi: 10.1177/1177932219899051.
- [85] D. Toubiana, A. R. Fernie, Z. Nikoloski, and A. Fait, "Network analysis: tackling complex data to study plant metabolism," *Trends in biotechnology*, vol. 31, no. 1, pp. 29-36, 2013.
- [86] A. P. Wasson *et al.*, "Traits and selection strategies to improve root systems and water uptake in water-limited wheat crops," *Journal of Experimental Botany*, vol. 63, no. 9, pp. 3485-3498, 2012, doi: 10.1093/jxb/ers111.
- [87] X. Li, A. Marquardt, and A. Wasson, "Carbon budgeting belowground," *New Phytologist*, 2021, doi: 10.1111/nph.17520.
- [88] L. M. York and J. P. Lynch, "Intensive field phenotyping of maize (*Zea mays*L.) root crowns identifies phenes and phene integration associated with plant growth and nitrogen acquisition," *Journal of Experimental Botany*, vol. 66, no. 18, pp. 5493-5505, 2015, doi: 10.1093/jxb/erv241.
- [89] G. J. Patti, O. Yanes, and G. Siuzdak, "Metabolomics: the apogee of the omics trilogy," *Nature Reviews Molecular Cell Biology*, vol. 13, no. 4, pp. 263-269, 2012/04/01 2012, doi: 10.1038/nrm3314.
- [90] N. Sapoval *et al.*, "Current progress and open challenges for applying deep learning across the biosciences," *Nature Communications*, vol. 13, no. 1, 2022, doi: 10.1038/s41467-022-29268-7.

- [91] X. He, K. Zhao, and X. Chu, "AutoML: A survey of the state-of-the-art," *Knowledge-Based Systems*, vol. 212, p. 106622, 2021, doi: 10.1016/j.knosys.2020.106622.
- [92] S. K. Karmaker, M. M. Hassan, M. J. Smith, L. Xu, C. Zhai, and K. Veeramachaneni, "AutoML to Date and Beyond: Challenges and Opportunities," *ACM Computing Surveys*, vol. 54, no. 8, pp. 1-36, 2022, doi: 10.1145/3470918.
- [93] J. L. Araus and J. E. Cairns, "Field high-throughput phenotyping: the new crop breeding frontier," *Trends in plant science*, vol. 19, no. 1, pp. 52-61, 2014.
- [94] D. D. Hoffman, M. Singh, and C. Prakash, "The Interface Theory of Perception," *Psychonomic Bulletin & Review*, vol. 22, no. 6, pp. 1480-1506, 2015/12/01 2015, doi: 10.3758/s13423-015-0890-8.
- [95] S. R. Tracy, K. A. Nagel, J. A. Postma, H. Fassbender, A. Wasson, and M. Watt, "Crop Improvement from Phenotyping Roots: Highlights Reveal Expanding Opportunities," *Trends in Plant Science*, vol. 25, no. 1, pp. 105-118, 2020, doi: 10.1016/j.tplants.2019.10.015.
- [96] H. L. S. Alves, C. C. Matiolli, R. C. Soares, M. C. Almadanim, M. M. Oliveira, and I. A. Abreu, "Carbon/nitrogen metabolism and stress response networks – calcium-dependent protein kinases as the missing link?," *Journal of Experimental Botany*, vol. 72, no. 12, pp. 4190-4201, 2021, doi: 10.1093/jxb/erab136.
- [97] H. Tao *et al.*, "Proximal and remote sensing in plant phenomics: Twenty years of progress, challenges and perspectives," *Plant Communications*, 2022.
- [98] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Computers and electronics in agriculture*, vol. 147, pp. 70-90, 2018, doi: 10.1016/j.compag.2018.02.016.
- [99] M. Shoaib, N. Sharma, L. Kotthoff, M. Lindauer, and S. Kant, "AutoML: advanced tool for mining multivariate plant traits," *Trends in Plant Science*, 2023, doi: 10.1016/j.tplants.2023.09.008.
- [100] I. Misra, C. Lawrence Zitnick, M. Mitchell, and R. Girshick, "Seeing through the human reporting bias: Visual classifiers from noisy human-centric labels," presented at the Proceedings of the IEEE conference on computer vision and

- pattern recognition, Las Vegas, NV, USA, 2016. [Online]. Available: <https://doi.org/10.48550/arXiv.1512.06974>.
- [101] M. W. Libbrecht and W. S. Noble, "Machine learning applications in genetics and genomics," *Nature Reviews Genetics*, vol. 16, no. 6, pp. 321-332, 2015/06/01 2015, doi: 10.1038/nrg3920.
- [102] Q. Yao *et al.*, "Taking Human out of Learning Applications: A Survey on Automated Machine Learning," *arXiv pre-print server*, 2019-12-16 2019, doi: None
arxiv:1810.13306.
- [103] F. Hutter, L. Kotthoff, and J. Vanschoren, *Automated machine learning: methods, systems, challenges*. Springer Nature, 2019.
- [104] L. Ferreira, A. Pilastri, C. M. Martins, P. M. Pires, and P. Cortez, "A Comparison of AutoML Tools for Machine Learning, Deep Learning and XGBoost," in *2021 International Joint Conference on Neural Networks (IJCNN)*, Shenzhen, China, 18-22 July 2021 2021, pp. 1-8, doi: 10.1109/IJCNN52387.2021.9534091.
- [105] M. G. Huck and H. M. Taylor, "The rhizotron as a tool for root research," *Advances in agronomy*, vol. 35, pp. 1-35, 1982, doi: 10.1016/S0065-2113(08)60320-X.
- [106] P. K. R. Maddikunta *et al.*, "Unmanned aerial vehicles in smart agriculture: Applications, requirements, and challenges," *IEEE Sensors Journal*, vol. 21, no. 16, pp. 17608-17619, 2021.
- [107] A. Seethepalli, K. Dhakal, M. Griffiths, H. Guo, G. T. Freschet, and L. M. York, "RhizoVision Explorer: open-source software for root image analysis and measurement standardization," *AoB PLANTS*, vol. 13, no. 6, 2021, doi: 10.1093/aobpla/plab056.
- [108] *WinRHIZO: Root Measurement and Analysis Software*. (2024). Québec, QC, 2001. [Online]. Available: https://regent.qc.ca/assets/winrhizo_software.html
- [109] R. Yasrab, J. A. Atkinson, D. M. Wells, A. P. French, T. P. Pridmore, and M. P. Pound, "RootNav 2.0: Deep learning for automatic navigation of complex plant root architectures," *GigaScience*, vol. 8, no. 11, Nov 1 2019, doi: 10.1093/gigascience/giz123.

- [110] A. Das *et al.*, "Digital imaging of root traits (DIRT): a high-throughput computing and collaboration platform for field-based root phenomics," *Plant Methods*, vol. 11, no. 1, p. 51, 2015, doi: 10.1186/s13007-015-0093-3.
- [111] R. Tajima and Y. Kato, "Comparison of threshold algorithms for automatic image processing of rice roots using freeware ImageJ," *Field Crops Research*, vol. 121, no. 3, pp. 460-463, 2011, doi: 10.1016/j.fcr.2011.01.015.
- [112] J. Semedo, F. Reboreda, F. Lidon, J. Coutinho, B. Maçãs, and P. Scotti-Campos, "Phenotypic Diversity of Seminal Root Traits in Bread Wheat Germplasm from Different Origins. Plants 2022, 11, 2842," ed: s Note: MDPI stays neutral with regard to jurisdictional claims in published ..., 2022.
- [113] K. W. Shamuyarira, H. Shimelis, S. Figlan, and V. Chaplot, "Path Coefficient and Principal Component Analyses for Biomass Allocation, Drought Tolerance and Carbon Sequestration Potential in Wheat," *Plants*, vol. 11, no. 11, p. 1407, 2022. [Online]. Available: <https://www.mdpi.com/2223-7747/11/11/1407>.
- [114] E. S. Ober *et al.*, "Wheat root systems as a breeding target for climate resilience," *Theoretical and Applied Genetics*, vol. 134, no. 6, pp. 1645-1662, 2021, doi: 10.1007/s00122-021-03819-w.
- [115] A. Ranjan, R. Sinha, S. L. Singla-Pareek, A. Pareek, and A. K. Singh, "Shaping the Root System Architecture in Plants for Adaptation to Drought Stress," *Physiologia Plantarum*, vol. 174, no. 2, p. e13651, Feb 16 2022, doi: 10.1111/ppl.13651.
- [116] K. Seidenthal, K. Panjvani, R. Chandnani, L. Kochian, and M. Eramian, "Iterative image segmentation of plant roots for high-throughput phenotyping," *Scientific Reports*, vol. 12, no. 1, 2022, doi: 10.1038/s41598-022-19754-9.
- [117] C. Shen, L. Liu, L. Zhu, J. Kang, N. Wang, and L. Shao, "High-Throughput in situ Root Image Segmentation Based on the Improved DeepLabv3+ Method," *Frontiers in Plant Science*, vol. 11, 2020.
- [118] N. B. C. Serre and M. Fendrych, "ACORBA: Automated workflow to measure *Arabidopsis thaliana* root tip angle dynamics," *Quantitative Plant Biology*, vol. 3, 2022, doi: 10.1017/qpb.2022.4.

- [119] D. Sun, K. Robbins, N. Morales, Q. Shu, and H. Cen, "Advances in optical phenotyping of cereal crops," *Trends in plant science*, vol. 27, no. 2, pp. 191-208, 2022.
- [120] D. Sun, Y. Xu, and H. Cen, "Optical sensors: deciphering plant phenomics in breeding factories," *Trends in plant science*, vol. 27, no. 2, pp. 209-210, 2022.
- [121] I. Misra, C. L. Zitnick, M. Mitchell, and R. B. Girshick, "Seeing through the Human Reporting Bias: Visual Classifiers from Noisy Human-Centric Labels," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2930-2939, 2015.
- [122] P. J. White, T. S. George, P. J. Gregory, A. G. Bengough, P. D. Hallett, and B. M. McKenzie, "Matching roots to their environment," *Annals of botany*, vol. 112, no. 2, pp. 207-222, Jul 2013, doi: 10.1093/aob/mct123.
- [123] A. G. Jacobsen, G. Jervis, J. Xu, J. F. Topping, and K. Lindsey, "Root growth responses to mechanical impedance are regulated by a network of ROS, ethylene and auxin signalling in *Arabidopsis*," *New Phytologist*, vol. 231, no. 1, pp. 225-242, Jul 2021, doi: 10.1111/nph.17180.
- [124] Y. Uga, "Challenges to design-oriented breeding of root system architecture adapted to climate change," *Breeding science*, vol. 71, no. 1, pp. 3-12, Feb 2021, doi: 10.1270/jsbbs.20118.
- [125] B. Schölkopf *et al.*, "Toward causal representation learning," *Proceedings of the IEEE*, vol. 109, no. 5, pp. 612-634, 2021, doi: 10.48550/arXiv.2102.11107.
- [126] H. M. Schneider and J. P. Lynch, "Should Root Plasticity Be a Crop Breeding Target?," (in English), *Frontiers in Plant Science*, Review vol. 11, 2020-May-15 2020, doi: 10.3389/fpls.2020.00546.
- [127] M. Li, M. H. Frank, V. Coneva, W. Mio, D. H. Chitwood, and C. N. Topp, "The persistent homology mathematical framework provides enhanced genotype-to-phenotype associations for plant morphology," *Plant physiology*, vol. 177, no. 4, pp. 1382-1395, Aug 2018, doi: 10.1104/pp.18.00104.
- [128] A. Hodge, "The plastic plant: root responses to heterogeneous supplies of nutrients," *New Phytologist*, vol. 162, no. 1, pp. 9-24, 2004, doi: 10.1111/j.1469-8137.2004.01015.x.

- [129] U. Alon, "Network motifs: theory and experimental approaches," *Nature Reviews Genetics*, vol. 8, no. 6, pp. 450-461, 2007/06/01 2007, doi: 10.1038/nrg2102.
- [130] Z. Huang, Z. Zeng, B. Liu, D. Fu, and J. Fu, "Pixel-bert: Aligning image pixels with text by deep multi-modal transformers," *arXiv pre-print server*, 2020, doi: 10.48550/arXiv.2004.00849.
- [131] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *the Journal of machine Learning research*, vol. 12, pp. 2825-2830, 2011, doi: 10.48550/arXiv.1201.0490.
- [132] J. P. Lynch, "Rightsizing root phenotypes for drought resistance," *Journal of Experimental Botany*, vol. 69, no. 13, pp. 3279-3292, 2018, doi: 10.1093/jxb/ery048.
- [133] V. Bolón-Canedo and A. Alonso-Betanzos, "Ensembles for feature selection: A review and future trends," *Information fusion*, vol. 52, pp. 1-12, 2019, doi: 10.1016/j.inffus.2018.11.008.
- [134] T. Hastie, "The elements of statistical learning: data mining, inference, and prediction," ed: Springer, 2009.
- [135] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651-666, 2010/06/01/ 2010, doi: <https://doi.org/10.1016/j.patrec.2009.09.011>.
- [136] A.-L. Barabási and R. Albert, "Emergence of Scaling in Random Networks," *Science*, vol. 286, no. 5439, pp. 509-512, 1999, doi: doi:10.1126/science.286.5439.509.
- [137] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences*, vol. 99, no. 12, pp. 7821-7826, 2002, doi: doi:10.1073/pnas.122653799.
- [138] R. Albert, H. Jeong, and A.-L. Barabási, "Error and attack tolerance of complex networks," *Nature*, vol. 406, no. 6794, pp. 378-382, 2000/07/01 2000, doi: 10.1038/35019019.
- [139] A. Küken, D. Langary, and Z. Nikoloski, "The hidden simplicity of metabolic networks is revealed by multireaction dependencies," *Science Advances*, vol. 8, no. 13, p. eabl6962, 2022.

- [140] R. Argelaguet *et al.*, "Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets," *Molecular Systems Biology*, vol. 14, no. 6, p. e8124, 2018, doi: <https://doi.org/10.15252/msb.20178124>.
- [141] F. Rohart, B. Gautier, A. Singh, and K.-A. Lê Cao, "mixOmics: An R package for 'omics feature selection and multiple data integration," *PLoS computational biology*, vol. 13, no. 11, p. e1005752, 2017.
- [142] C. De La Fuente Cantó *et al.*, "An extended root phenotype: the rhizosphere, its formation and impacts on plant fitness," *The Plant Journal*, vol. 103, no. 3, pp. 951-964, 2020, doi: 10.1111/tpj.14781.
- [143] H. Fromm, "Root Plasticity in the Pursuit of Water," *Plants*, vol. 8, no. 7, p. 236, 2019, doi: 10.3390/plants8070236.
- [144] T. Galindo-Castañeda, M. Hartmann, and J. P. Lynch, "Location: root architecture structures rhizosphere microbial associations," *Journal of Experimental Botany*, p. erad421, 2023.
- [145] P. Yu *et al.*, "Plant flavones enrich rhizosphere Oxalobacteraceae to improve maize performance under nitrogen deprivation," *Nature Plants*, vol. 7, no. 4, pp. 481-499, Apr 2021, doi: 10.1038/s41477-021-00897-y.
- [146] J. R. Dinneny, "Developmental Responses to Water and Salinity in Root Systems," *Annual Review of Cell and Developmental Biology*, vol. 35, no. 1, pp. 239-257, 2019, doi: 10.1146/annurev-cellbio-100617-062949.
- [147] A. C. Burnett, S. P. Serbin, K. J. Davidson, K. S. Ely, and A. Rogers, "Detection of the metabolic response to drought stress using hyperspectral reflectance," *Journal of Experimental Botany*, vol. 72, no. 18, pp. 6474-6489, 2021, doi: 10.1093/jxb/erab255.
- [148] L. Cantini *et al.*, "Benchmarking joint multi-omics dimensionality reduction approaches for the study of cancer," *Nature Communications*, vol. 12, no. 1, p. 124, 2021/01/05 2021, doi: 10.1038/s41467-020-20430-7.
- [149] J. You *et al.*, "Transcriptomic and metabolomic profiling of drought-tolerant and susceptible sesame genotypes in response to drought stress," *BMC Plant Biology*, vol. 19, no. 1, 2019, doi: 10.1186/s12870-019-1880-1.

- [150] H. Ye *et al.*, "Genetic diversity of root system architecture in response to drought stress in grain legumes," *Journal of Experimental Botany*, vol. 69, no. 13, pp. 3267-3277, 2018, doi: 10.1093/jxb/ery082.
- [151] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature Machine Intelligence*, vol. 1, no. 5, pp. 206-215, 2019/05/01 2019, doi: 10.1038/s42256-019-0048-x.
- [152] D. Castelvecchi, "Can we open the black box of AI?," *Nature News*, vol. 538, no. 7623, p. 20, Oct 6 2016, doi: 10.1038/538020a.
- [153] J. Ubbens, M. Cieslak, P. Prusinkiewicz, I. Parkin, J. Ebersbach, and I. Stavness, "Latent space phenotyping: automatic image-based phenotyping for treatment studies," *Plant Phenomics*, vol. 2020, p. 5801869, 2020, doi: 10.34133/2020/5801869.
- [154] B. Zhang and S. Horvath, "A general framework for weighted gene co-expression network analysis," (in eng), *Stat Appl Genet Mol Biol*, vol. 4, p. Article17, 2005, doi: 10.2202/1544-6115.1128.
- [155] K. P. Voss-Fels, R. J. Snowdon, and L. T. Hickey, "Designer Roots for Future Crops," *Trends in Plant Science*, vol. 23, no. 11, pp. 957-960, 2018, doi: 10.1016/j.tplants.2018.08.004.
- [156] A. Blum, "Drought resistance, water-use efficiency, and yield potential—are they compatible, dissonant, or mutually exclusive?," *Australian Journal of Agricultural Research*, vol. 56, no. 11, pp. 1159-1168, 2005.
- [157] Z. Zhang and X. Peng, "Bio-tillage: A new perspective for sustainable agriculture," *Soil and Tillage Research*, vol. 206, p. 104844, 2021.
- [158] C. Jansson *et al.*, "Crops for Carbon Farming," (in English), *Frontiers in Plant Science*, Perspective vol. 12, no. 938, 2021-June-04 2021, doi: 10.3389/fpls.2021.636709.
- [159] P. Panchal, C. Preece, J. Peñuelas, and J. Giri, "Soil carbon sequestration by root exudates," *Trends in Plant Science*, vol. 27, no. 8, pp. 749-757, 2022/08/01/ 2022, doi: <https://doi.org/10.1016/j.tplants.2022.04.009>.

Chapter 2. Roots' Drought Adaptive Traits in Crop Improvement

Establishing the Biological Foundation and Analytical Imperative

While drought represents one of agriculture's greatest challenges, our understanding of the biological mechanisms underlying plant adaptation remains fragmented, largely due to methodological limitations in quantifying complex, system-wide traits. This chapter establishes a comprehensive biological foundation for the research undertaken in this thesis by reviewing the architectural, structural, hydraulic, and plastic traits that enable root systems—the hidden half of the plant—to survive water limitation.

From the biomechanics of taproot penetration to the biochemistry of rhizosheath formation, this review maps the sophisticated adaptive machinery evolved to combat drought. Yet, in doing so, it reveals a critical gap: our conceptual understanding has outpaced our analytical capabilities. Traditional phenotyping, constrained by subjectivity and scale, fails to capture the intricate architectural and functional complexity of roots with the precision demanded by modern crop improvement.

This biological complexity directly gives rise to the four analytical bottlenecks identified in Chapter 1 (Sections 1.3.1-1.3.4): the inability of manual methods to handle data complexity, the limitations of human perception in phenotyping, the challenge of integrating across biological scales, and the difficulty of extracting mechanistic insights from correlative data. The biological foundation established here highlights the urgency for the technological expedition that follows—a systematic response to each analytical bottleneck through targeted computational innovation.

Statement of contributions of joint authorship

Mirza Shoaib conducted the literature review, generated the figures, and led the manuscript writing and response to reviewers. Bikram P. Banerjee and Surya Kant contributed to manuscript editing.

Statement from the co-author confirming the PhD candidate's contributions to the authorship

As a co-author of the manuscript:

M. Shoaib, B. P. Banerjee, M. Hayden, and S. Kant, "Roots' drought adaptive traits in crop improvement," Plants, vol. 11, no. 17, p. 2256, 2022.

I confirm that Mirza Shoaib made the following key contributions:

- Conducted the literature review
- Led the writing of the manuscript
- Prepared the response to reviewers



Dr Surya Kant

29 July 2025

Review

Roots' Drought Adaptive Traits in Crop Improvement

Mirza Shoaib ^{1,2}, Bikram P. Banerjee ¹ , Matthew Hayden ^{2,3} and Surya Kant ^{1,2,3,*} ¹ Agriculture Victoria, Grains Innovation Park, 110 Natimuk Road, Horsham, VIC 3400, Australia² School of Applied Systems Biology, La Trobe University, Bundoora, Melbourne, VIC 3083, Australia³ Agriculture Victoria, AgriBio, Centre for AgriBioscience, 5 Ring Road, Bundoora, Melbourne, VIC 3083, Australia* Correspondence: surya.kant@agriculture.vic.gov.au

Abstract: Drought is one of the biggest concerns in agriculture due to the projected reduction of global freshwater supply with a concurrent increase in global food demand. Roots can significantly contribute to improving drought adaptation and productivity. Plants increase water uptake by adjusting root architecture and cooperating with symbiotic soil microbes. Thus, emphasis has been given to root architectural responses and root–microbe relationships in drought-resilient crop development. However, root responses to drought adaptation are continuous and complex processes and involve additional root traits and interactions among themselves. This review comprehensively compiles and discusses several of these root traits such as structural, physiological, molecular, hydraulic, anatomical, and plasticity, which are important to consider together, with architectural changes, when developing drought resilient crop varieties. In addition, it describes the significance of root contribution in improving soil structure and water holding capacity and its implication on long-term resilience to drought. In addition, various drought adaptive root ideotypes of monocot and dicot crops are compared and proposed for given agroclimatic conditions. Overall, this review provides a broader perspective of understanding root structural, physiological, and molecular regulators, and describes the considerations for simultaneously integrating multiple traits for drought tolerance and crop improvement, under specific growing environments.



Citation: Shoaib, M.; Banerjee, B.P.; Hayden, M.; Kant, S. Roots' Drought Adaptive Traits in Crop Improvement. *Plants* **2022**, *11*, 2256. <https://doi.org/10.3390/plants11172256>

Academic Editor: Patrizia Trifilo

Received: 9 August 2022

Accepted: 26 August 2022

Published: 30 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In response to drought, roots adjust their traits, improving plant adaptation, survival, and yield. Among these traits, root system architecture (RSA) is essential in increasing water uptake [1,2], therefore, much of the research has focused on understanding RSA [3]. Phenotyping systems, such as X-ray computed tomography, magnetic resonance imaging, ground-penetrating radar, shovelingomics, rhizotrons, and transparent soils, were developed to study RSA [2,4]. These phenotyping systems identified several root architectural traits that increased water uptake and drought resistance and were utilized in developing drought-resilient plants. [5–9]. Plants also invest a large portion of their photosynthetic carbon (C) as exudates to build root–microbe symbiosis for drought adaptation [1,2,10]. These microbes influence plants' water uptake by extending the water harnessing area [11], changing root hydraulics [12], producing drought stress-reducing antioxidants and influencing many other physiological activities. Thus, beneficial microbial inoculants are widely used in agriculture to improve productivity. Similarly, a crucial drought response is root hydraulics, and plants regulate it for optimal water use. However, rarely it is considered together with water-efficient root architectural traits while proposing drought adaptive ideotype. The effect of interaction among multiple architectural traits and the contribution of their plasticity is also seldomly explored. A further significant response to drought is root exudation and the root-symbiotic microbes' contribution in improving soil structure,

water holding capacity, and storage and nutrient availability. However, these root soil improvement traits are yet to be fully explored for incorporation into drought-resilient crop improvement. Here, we highlight some of these significant and diverse root traits, including architecture, anatomy, hydraulics, plasticity, root-shoot interaction, exudates and interaction with microbes, and emphasize the importance of simultaneous integration of multiple traits for drought tolerance and crop improvement.

2. Root System Architecture and Its Interaction with the Shoot in Response to Drought

To anchor and forage through the soil for water and nutrient uptake, roots dynamically modify their spatial distribution by changing their length, depth, number, angle, diameter, density, and biomass—this spatial distribution and shape is referred as root system architecture (RSA) [2,13]. The functionality of the RSA relies on a dynamic interaction response with its heterogeneous growing environment [13], therefore, RSA has a crucial role in plant adaptation to drought [14]. Although some individual root traits have significant impact on water harnessing in drought, the efficiency and function of those traits are influenced by RSA, for example, root hairs are crucial for water uptake [15], but RSA regulates the efficiency of root hairs. Root hairs in shallow RSAs can access water from surface soil, but root hairs in deep RSAs can continue to access water from deep soil when the surface soil is dry. Roots change their RSA in response to drought, with the adaptation type varying based on the environment; for example, long and metabolically efficient RSA can access water from deep soil during drought [16]. However, in areas where water is mostly available in surface soil due to short periodic rainfall, a shallow RSA is more efficient [17]. Even within the same environment, RSA could change due to heterogeneous water distribution in soil. Roots can preferentially increase their branching towards the water, known as hydropatterning [18] and suppress lateral root formation in soil air spaces, known as xerobranching [19]. Several genes modulate RSA through gravitropic and radial growth during drought, allowing them to form shallow versus deep roots [20]. EXO70A3 regulates the auxin pathway by acting on the distribution of the PIN4 (auxin efflux carrier), resulting in the regulation of gravitropic growth (rooting depth) in RSA [21]. Whereas CKX2 catalyzes the degradation of cytokinin at the upper lateral root flank, suppressing organ growth towards gravity, and allowing radial expansion forming shallow RSA [22].

During drought, root and shoot crosstalk and influence each other's growth. A plant's ability to sense drought and its intensity depends on soil moisture, temperature, radiation, and other environmental factors [23]. When roots sense water stress, it triggers changes in xylem hydraulics, mobile peptides, reactive oxygen species (ROS), and Ca^{2+} signaling, affecting shoot activities, such as stomatal closure to conserve water [24]. The plant then prioritizes its resource allocation and invests more in roots during drought [25,26], affecting the root–shoot ratio [27]. Investing in roots allows the plant to increase water uptake and maintain root water influx under drought conditions, supporting uninterrupted plant growth [24]. In white clover (*Trifolium repens*), the genotype with increased root weight in proportion to total plant weight had improved growth, survival, and seed yield during drought [28]. Chen et al. [27] reported how phosphorylation of sucrose transporters mediates the root–shoot ratio during drought, thus suggesting a strategy for developing drought-resistant crops. They reported that drought stress increases abscisic acid (ABA) levels which activates the SnRK2 protein kinases, phosphorylating the sucrose transporters SWEET11 and 12. This phosphorylation enhances the oligomerization and sucrose transport activity of SWEETs, causing elevated sucrose contents in roots, improving root growth and biomass, root–shoot ratio, and thus drought adaptation.

3. Roots' Structural Response to Drought

Roots adapt their structure in response to drought to increase penetration, distribution, and contact with the soil for improved water and nutrient uptake [6,15]. These structural adaptations ensure necessary nutrition and water acquisition, maintaining plant physiological activities and productivity during drought. Several studies which reported correlations

between root structural traits and plant performance under drought are summarized and presented in Table 1 and Figure 1a,b. Identifying these drought-responsive structural traits in different crops will facilitate plant breeders to utilize these traits for screening drought tolerant genotypes.

Table 1. Root structural traits and their adaptive response to drought.

| Structural Root Traits | Drought Adaptive Responses | Crop | Reference |
|--|--|---|---|
| Taproot diameter | Large taproot diameter genotypes had increased yield and drought resistance. | White clover (<i>Trifolium repens</i> L.), Soybean (<i>Glycine max</i> L.), Chickpea (<i>Cicer arietinum</i> L.) | Caradus and Woodfield [28], Fenta et al. [29], Rabbi et al. [7] |
| Taproot length | Long taproot genotypes yielded higher. | Soybean (<i>Glycine max</i> L.) | Jumrani and Bhatia [30] |
| Root hair | Reduced root hair genotype had lower water absorption and decreased drought resistance. | Arabidopsis (<i>Arabidopsis thaliana</i> L.) | Tanaka et al. [31] |
| Root hair production time | Drought-resistant genotypes had faster root hair production. | Barley (<i>Hordeum vulgare</i> L.) | Carter et al. [32] |
| Root hair length and number | Longer and higher root hair genotypes had less negative leaf water potential and improved water status under drought. | Barley (<i>Hordeum vulgare</i> L.) | Marin et al. [33] |
| Rhizosheath size | Large rhizosheath genotypes were drought resistant. Longer and denser root hairs contributed to larger rhizosheath formation. | Barley (<i>Hordeum vulgare</i> L.), Lotus (<i>Lotus japonicus</i> L.), and Maize (<i>Zea mays</i> L.) | Liu et al. [34], Rabbi et al. [7] |
| Root growth angle and rooting depth | Narrow root angles had downward root growth resulting in deep rooting and better yield under drought. | Rice (<i>Oryza sativa</i> L.), Soybean (<i>Glycine max</i> L.) | Uga et al. [5], Gobu et al. [35], Fenta et al. [29] |
| Seminal and nodal root angle | Steeper seminal and nodal root angle genotypes had a higher yield. | Maize (<i>Zea mays</i> L.) | Ali et al. [36] |
| Tap and lateral root branching intensity | Drought-resistance genotypes had more tap and lateral root branches. | Soybean (<i>Glycine max</i> L.) | Fenta et al. [29] |
| Number of crown root | Low crown root number genotypes had better water status and yield. | Maize (<i>Zea mays</i> L.) | Gao and Lynch [8] |
| Quantity of fine-diameter roots | Drought-resistant genotypes had substantial amounts of small-diameter roots in deep soil. | Wheat (<i>Triticum aestivum</i>) | Becker et al. [25] |
| Lateral root branching density | Genotypes with fewer but longer lateral roots had better water status, biomass, and yield. | Maize (<i>Zea mays</i> L.) | Zhan et al. [6] |
| Root length, branching rate and surface area | Drought-resistant genotypes had increased root length, branching rate, larger root surface, and decreased coarse to fine root ratio. | Oat (<i>Avena sativa</i> L.) | Canales et al. [37] |
| Root volume and dry matter | Drought-resistant genotypes had larger root volumes and more root dry weight. | Sorghum (<i>Sorghum bicolor</i> L. Moench) | Kiran et al. [9] |

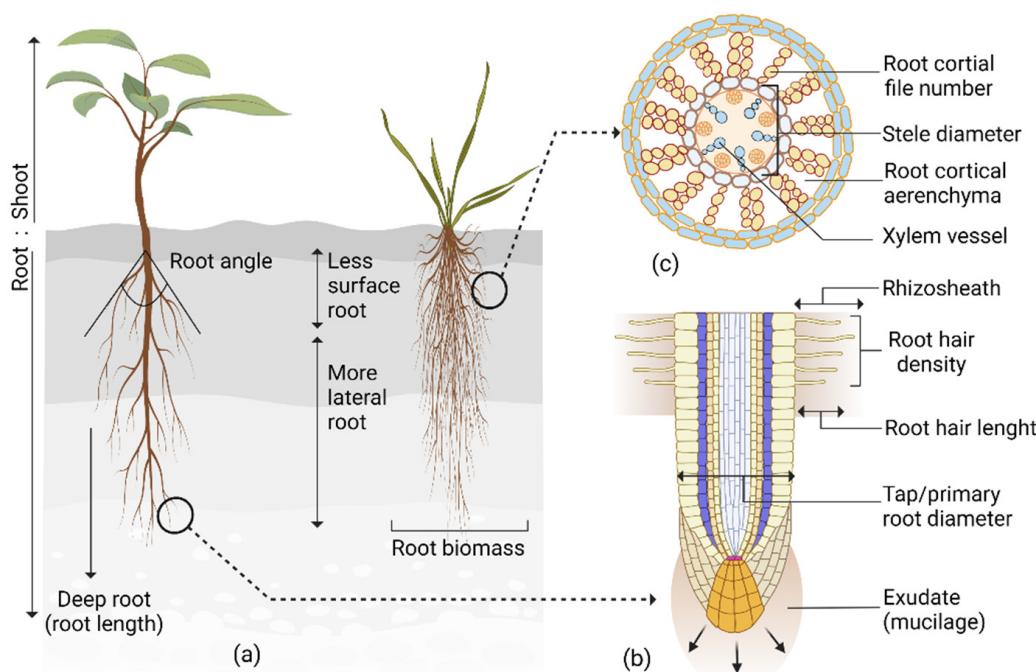


Figure 1. Drought adaptive root traits. (a) Changes in root angle, length, and biomass; the ratio with the shoot, and increased lateral branching facilitate plant adaptation to drought; (b) Root hair length and density, rhizosheath size, taproot diameter, and exudates are crucial drought-responsive traits. (c) Plants adapt their anatomical traits such as root cortical file number, cortical aerenchyma, stele diameter, and xylem vessel in response to drought.

Root drought-responsive structural traits are not static but fluctuate readily in conjunction with the environment, management practice, soil microbes, and genotype. Moreover, different root traits interact, sometimes synergistically or antagonistically, affecting drought adaptation. Thus, careful consideration is needed when planning drought-resistant crops based on structural traits only. However, crop-specific drought-responsive structural traits (Table 1) are easy to phenotype and implement for screening genotypes. Therefore, it has importance in terms of ease of applicability and achieving fast results in developing drought-resistant crops.

4. Root Anatomical Responses to Drought

Similar to architectural traits, roots also adapt their anatomical characteristics in response to drought. Roots increase penetration in soil, reduce metabolic cost, regulate hydraulic conductivity, and facilitate microbial symbiosis to increase resource acquisition [29–33]. Some of these key components are described below.

4.1. Anatomical Adaptation in Reducing Metabolic Cost

Plants invest photosynthates in establishing root systems for exploring water and nutrients needed for metabolic processes. During drought, the expenditure of photosynthates is high, firstly to invest more in root growth to increase water uptake, and secondly to increase respiration to maintain roots in drying soil, thus, compromising plant productivity [34,35]. Anatomical adaptation during drought reduces metabolic costs and allows plants to distribute resources for further resource acquisition, growth, and essential physiological activities, resulting in improved yield. Chimungu et al. [36] reported cortical tissue with fewer but larger cell sizes, and low cortical cell file numbers in maize reduced metabolic cost by decreasing respiration rates during drought. Large cortical genotypes had increased root growth and water acquisition, deeper roots, better stomatal conductance

and leaf CO₂ assimilation, and greater shoot biomass and grain yield. Colombi et al. [34] reported a similar result in wheat, where a genotype with a large root cortical cell diameter significantly reduced root penetration metabolic cost. During drought, root cortical tissue lysed and creates an intercellular vacuum space, cortical aerenchyma [37] (Figure 1c), the formation of which can reduce metabolic costs. This mostly occurs in older roots that no longer take up water efficiently [37], thus having minimal negative impact on water uptake. Instead, it reduces soil exploration metabolic costs [35,38], permits root growth, and improves soil resource acquisition in dry conditions [35]. Under water stress, maize genotypes with high root cortical aerenchyma had 30% more shoot biomass [39].

4.2. Anatomical Response Improving Root Penetration

Water deficit in the soil increases mechanical impedance, thus restricting root penetration into deep soil, which hinders resource capture, consequently reducing crop productivity [40]. Anatomical adaptation facilitates improving root penetration in drying soil. For example, smaller outer cortical cells stabilize the root against ovalization, prevent collapsing, and allow the root to penetrate soil. In contrast, large mesodermis cortical cells and thick axial roots with more aerenchyma reduce the metabolic cost of soil exploration and allow root growth in hard soils [41]. In deep rooting maize genotypes, the roots generated from node three had a reduced cell file number and increased middle cortical area; while the roots generated from node four had increased aerenchyma [38]. Maize and wheat genotypes with multiseriate cortical sclerenchyma had a 22% increase in deep soil penetration and a 49% increase in shoot biomass compared to genotypes that lacked it [33]. These genotypes also had small cells with thick walls in outer cortical tissue, increased root lignin concentration, tensile strength, and root tip bending force.

4.3. Anatomical Attributes Facilitating Microbial Symbiosis

Roots absorb water and nutrients through the root epidermis, hairs, and avascular mycorrhizal (AM) hyphae. The AM colonize in root cortical cells and extend their hyphae into the soil, sometimes expanding soil volume exploration at least 15 cm beyond the root surface [42]. The AM receives organic C from the root and, in return, delivers nutrients to the root [43]. This mutual relationship increases drought resistance and reduces yield loss [44–47]. Root anatomical traits play an essential role in microbial symbiosis. In maize, larger root diameters and larger aerenchyma lacunae increase mycorrhizal colonization, whereas increased aerenchyma and decreased living cortical area reduce mycorrhizal colonization [31]. Cortex thickness is also crucial for AM colonization in woody species [48,49]. Dreyer et al. [50] reported that a continuous sclerenchymatic ring in the outer cortex and aerenchyma in the inner cortex decrease AM colonization in three palm species. They suggested that the sclerenchymatic ring acts as a physical barrier preventing penetration of AM fungal, while the empty aerenchyma area reduces the available tissue in the root for AM colonization.

4.4. Anatomical Adaptation in Regulating Water Transport

Root anatomical traits have a substantial influence on water uptake. Through concentric layers of root cells, water first enters from the soil to the root stele, then into the xylem, and finally into the shoot [2]. Transportation in the xylem is vertical, as casparyan strips limit radial movement, with the size and number of xylem vessels affecting the water transportation rate [51,52]. During root development, two types of xylem vessels are formed, the narrow protoxylem vessels and larger metaxylem vessels, through which the majority of the water is transported [51]. Prince et al. [30] reported soybean genotypes with large xylem vessels, increased xylem diameter, and metaxylem numbers performed well in water-limited environments. Large xylem vessels in olive (*Olea europaea*) increased root conductivity during drought stress, allowing deep rooting and extended water acquisition [53].

Phenotyping root anatomical traits is still low throughout, therefore, less research has focused on drought adaptive root anatomical traits. Further research in this area will improve phenotypic efficiency, accelerate the discovery of additional anatomical traits that assist directly in drought adaptation or indirectly through the facilitation of AM colonization, and ultimately contribute to developing drought-resilient crops.

5. Root Hydraulics

Root water permeability, known as root hydraulic conductivity, plays a crucial role in water uptake from soil [2,39,54,55]. Improved root hydraulic conductivity decreases the need for plants to invest in more root growth for soil exploration to increase water uptake, therefore, reducing the metabolic costs and stabilizing physiological processes under drought [30,37]. Along with root architecture, root hydraulic properties are key to predicting and interpreting plant transpiration activity during drought [56]. In water-limited conditions, root hydraulic conductivity positively correlated with dry biomass in rice and *Arabidopsis* [57,58]. Root hydraulics changes due to temperature [59], water availability [57,58,60], plant genotype [61,62], root anatomical traits [30,63], biochemical traits [55], and soil microbes [12,64].

Transcription factor *XND1* influences root hydraulic conductivity by negatively regulating xylem differentiation (Figure 2). Loss of function of *XND1* increased root hydraulic conductivity and shoot weight, therefore, increasing drought tolerance, with opposite effects seen from the overexpression of *XND1*, which negatively regulated drought stress. *XND1* also interacts with pathogenic bacteria (*Ralstonia solanacearum*) to influence xylem formation, as the presence of the bacteria enhance the expression of *XND1*, resulting in the reduction of root hydraulic conductivity. *XND1*, on the other hand, restricts bacterial growth and thus its pathogenicity [2,57]. In apple (*Malus domestica*), *MdMYB46* was reported to regulate root xylem vessel formation, thus root hydraulic conductivity and tolerance to drought [65] (Figure 2).

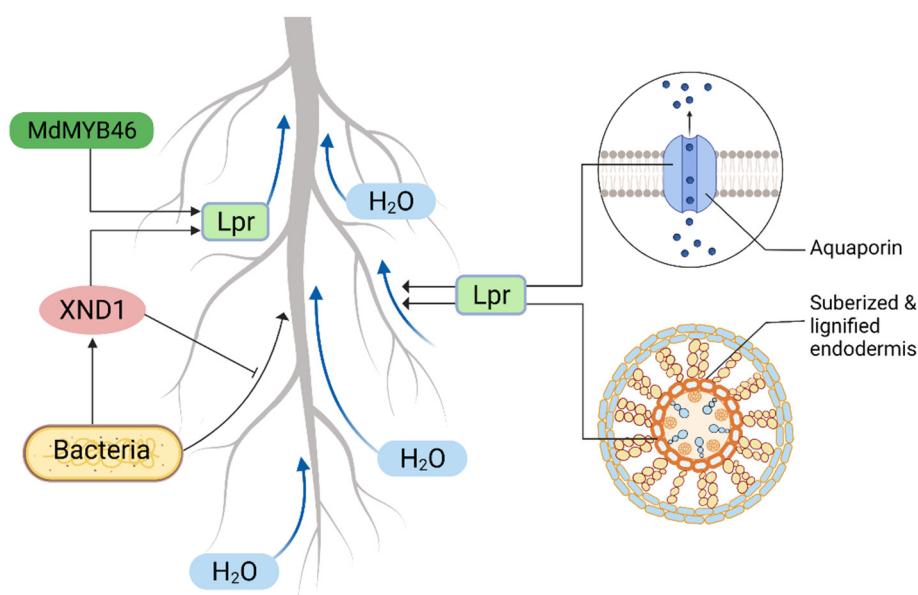


Figure 2. Molecular and cellular regulators of root hydraulics. *XND1* reduced root hydraulics (Lpr) by inhibiting xylem formation. Bacteria increase *XND1* activity and *XND1* reduces the pathogenicity of the bacteria; thus, Lpr is regulated without being affected by bacterial wilt. *MdMYB46* influences Lpr by modifying root xylem vessel formation. Aquaporins control Lpr by regulating radial water transport. Suberin reduces Lpr and lignin indirectly reduces Lpr by facilitating xylem vessel development.

Aquaporins (water channel proteins) also control root hydraulic conductivity by facilitating water diffusion across cell membranes, thus regulating radial water transport [66] (Figure 2). Root hydraulic conductivity can be increased or decreased by regulating plasma aquaporins [67,68], which can also promote the emergence of lateral roots by increasing directional water flow in the root lateral primordia [69]. During water stress, aquaporins are possibly involved in circadian oscillations of root hydraulic conductivity, facilitating root water uptake and promoting growth and photosynthesis [70].

Roots form a thicker suberized endodermis in response to drought [60]. Calvo-Polanco et al. [71] studied a collection of *Arabidopsis* mutants defective in suberin deposition and found mutants with enhanced root hydraulic conductivity, suggesting suberin reduces root hydraulic conductivity. Drought stress also deposits lignin in the root endodermis, indirectly reducing root hydraulic conductivity by contributing to change in xylem vessels [72] (Figure 2).

Most research has been focused on RSA, its adaptive response and role of individual RSA elements and the hormonal regulation of those elements in response to drought. However, Vadez [55] gave examples where root RSA and its element adaptive response do not fully explain plant drought adaptation and suggested root hydraulics to be considered along with RSA. Currently, several methods are available to measure root hydraulic conductivity [60,66,73–75]. Heymans et al. [76] proposed an inexpensive high-throughput method with some optimization suggestions. Therefore, it is feasible to phenotype root hydraulic conductivity with architectural traits.

6. Interaction among Drought Response Root Traits

Plants respond to various environmental conditions and stresses in the field with multiple root traits simultaneously, dynamically, and through the interaction of several traits [13]. These trait interactions sometimes have synergistic or antagonistic effects on drought adaptation [16,41]. Ajmera et al. [77] reported root trait synergism among nodal root angle, smaller diameter nodal roots, nodal root number, and L-type and S-type lateral branching densities that improve yield in dry and low nitrogen (N) conditions. Root anatomical and architectural traits can also interact synergistically, such as in common bean (*Phaseolus vulgaris*), where shallow roots with low axial conductance metaxylem improved water status by reducing water uptake [52]. This interaction of root anatomical and architectural traits likely reflected an adaptation for water uptake to their native environments. Small cells in the outer cortical region and stele diameter size can predict root penetration ability into deep soil [41]. However, these anatomical traits possibly have synergy with root growth angle and root hairs, as root angle can influence rooting depth [5], and root hairs provide anchorage to the growing root tip to penetrate drying soils [78]. Sometimes traits can interact antagonistically, for instance, forming aerenchyma tissue in the root reduces metabolic costs, improving drought adaptation [35], but this increase in aerenchyma tissue also decreases the colonization of drought-reducing root-inhabiting mycorrhiza [31,79–81]. The increased understanding of how plants prioritize root traits and their interaction with other environmental factors will improve the planning of ideotypes and selection of genotypes for drought-prone areas.

7. Root Plasticity in Drought Adaptation

The adaptive abilities of roots to diverse environments, known as root plasticity, are crucial along with RSA for breeding climate resilience crops [18]. In response to drought, the *DRO1* (DEEPER ROOTING 1) gene changes rooting depth by changing root angle and increasing gravitropism, thus increasing yield [5,82,83]. The *EXOCYST70A3* gene also alters root gravitropic responses, resulting in greater rooting depth, by acting on PIN4 protein distribution, influencing auxin transport [21]. During drought, where water is available deep in the soil, fewer but longer lateral roots in conjunction with deep rooting reduces root respiration and improve biomass and yield [6]. Conversely, in low periodic rainfall areas where water is available mainly in surface areas, more lateral branching

towards the surface improves water harnessing efficiency during drought [84]. Plasticity in lateral root formation towards available water i.e., hydropatterning, is therefore crucial in drought adaptation. In *Arabidopsis*, *LBD/ASL* genes are responsible for lateral root formation with AUXIN RESPONSE FACTORs, *ARF7* and *ARF19*, directly regulating these genes [85,86]. Orosa-Puente et al. [87] found roots preferential lateral root branching towards water depends on auxin response factor *ARF7*, inducing differential expression of *LBD16* in the lateral root, with *ARF7* expression itself regulated by the small ubiquitin-like modifier (SUMO) proteins. *Arabidopsis* plants lacking the SUMO protein had defective hydropatterning similar to *arf7* mutants [18,87]. In wheat, the lateral root number is regulated by *TaLBD16-D* [88], while in rice (*Oryza sativa L.*), auxin transporter, *OsAUX1* controls lateral root initiation [89] (Figure 3).

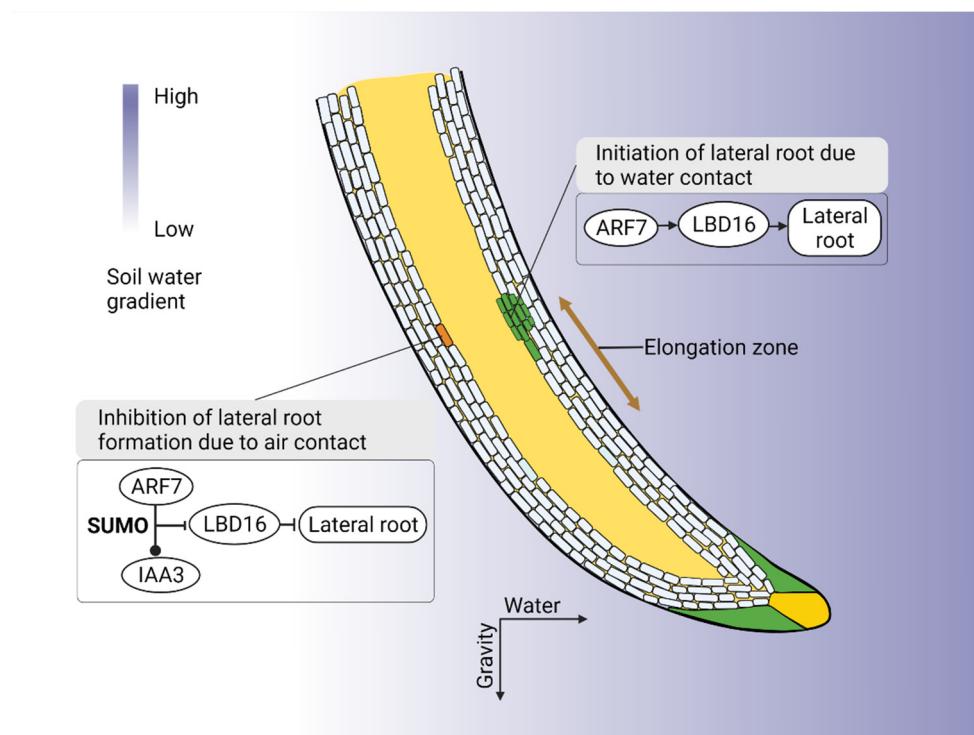


Figure 3. Molecular mechanism of root hydrotropism and hydropatterning. *ARF7* is crucial for preferential lateral root branching towards water (hydropatterning). *ARF7* regulates *LBD16* which initiates lateral root formation towards the water. On the dry side of the soil, SUMO protein enables *ARF7* to increase auxin repressor protein *IAA3*. *IAA3* represses *LBD16*; thus, no lateral root formation [18,87]. In response to a water gradient, ABA mediates differential growth response. In the root elongation zone cortex, ABA signaling kinase *SnRK2.2* and *MIZ1* are expressed that inhibit hydrotropism by preventing cell elongation.

Hydrotropism (the growth of plant roots towards moisture) also plays a key role in drought adaptation. In response to heterogeneous low water potentials, root tips overcome gravity-driven growth and increase branching towards higher water regions, i.e., hydropatterning [90]. When emerged lateral roots are surrounded by dry soil, hydrotropic growth assists roots to maneuver toward the water [18]. Thus, hydrotropism and hydropatterning interplay to adapt plants in a drought. Hydrotropism is dependent on ABA and independent of auxin redistribution, however, its molecular and cellular basis remains unclear [90,91]. Earlier, Kobayashi et al. [92] suggested the hydrotropism response is independent of gravitropism. Recently Dietrich et al. [91] showed the meristem at the root cap does not direct hydrotropism; rather, it depends on the root elongation zone's *SnRK2.2* (ABA signaling kinase) and *MIZ1*. They suggested that the root elongation zone

influences hydrotropism by sensing a water potential gradient and undergoing differential growth (Figure 3).

8. Root Exudates and Microbial Symbiosis in Drought Adaptation

Plants release exudates and root litter into the soil, which microbes feed on and decompose, enhancing nutrient availability in the soil and promoting root growth. Through this symbiotic relationship, root exudates and root-associated microbes can improve plant resilience and performance under drought [93,94]. Out of these microbes, AM and rhizobacteria are major contributors to drought resistance. Here, we highlight a few examples of their contribution to plant drought adaptation.

8.1. Root Exudate Role in Drought Adaptation

Plants change their exudate composition during drought, impacting the plant itself, neighboring plants, soil properties, and organisms [95]. Exudates also serve as a chemoattractant and nutrition source for many drought-reducing microorganisms. For instance, during drought, plants increase their mucilage production, resulting in a larger rhizosphere [96]. Root mucilage plays an important role in retaining water; thus, large rhizosphere formation increases the plant's adaptation to drought [96–98]. Plants preferentially recruit root colonizing drought-reducing microbes by adjusting the composition of root exudates [99–102]. For instance, during drought, plants produce Glycerol-3-phosphate, which increases the abundance of drought stress-reducing monoderm bacteria [103], while maize increases organic acid exudation, particularly malic acid [104], attracting *Bacillus subtilis* known to aid plant drought resistance [105]. Naylor et al. [106] found different plants prefer a specific drought-stress reducing Actinobacteria community, suggesting the plants favored a particular group of microbes for drought adaptation. During drought, exudates including organic acids and mucilage improve nutrition mobilization [107,108], as well as facilitate the belowground N cycle, influencing iron uptake and indirectly assisting in drought adaptation [10] (Figure 4a).

8.2. Contribution of Avascular Mycorrhizal Symbiosis in Drought Response

Avascular mycorrhiza can increase drought resistance and reduce drought susceptibility [109] by improving water and P uptake and enhancing photosynthetic performance [79]. Additionally, through extraradical hyphae, AM extends the soil volume connection beyond the root zone, thus, increasing water uptake [11,110]. AM can occasionally modulate lateral root formation such as in lemon (*Citrus limon* L.) [111] and maize [112]. Symbiosis of AM can influence root hydraulic properties and enhance plant drought resistance by controlling the *PIP* gene that regulates plasma-membrane proteins (PIPs). The downregulation of PIPs reduces water loss during drought [113]. Aroca et al. [12] found AM symbiosis in *Phaseolus vulgaris* was strongly correlated with the regulation of PIP2 protein and reduced hydraulic conductivity during drought. In *Ulmus americana*, AM increased apoplastic water transport and root hydraulic conductivity [64]. During drought, ROS production causes oxidative damage, degrading pigments, proteins, lipids, carbohydrates, and DNA [114]. However, the enzymatic components—superoxide dismutase, catalase, ascorbate peroxidase, glutathione reductase—and non-enzymatic components—cysteine, glutathione, and ascorbic acid—of exudates work as antioxidants and prevent this oxidative damage [114]. AM can increase antioxidant enzyme activities, thus reducing oxidative stress during drought [81], such as inducing more hydrogen peroxide effluxes of the taproot and lateral root of oranges (*Poncirus trifoliata* L.) [80]. In addition, AM contributes to improved soil C [115], which affects water storage, thus indirectly contributing to increased drought resilience (Figure 4b).

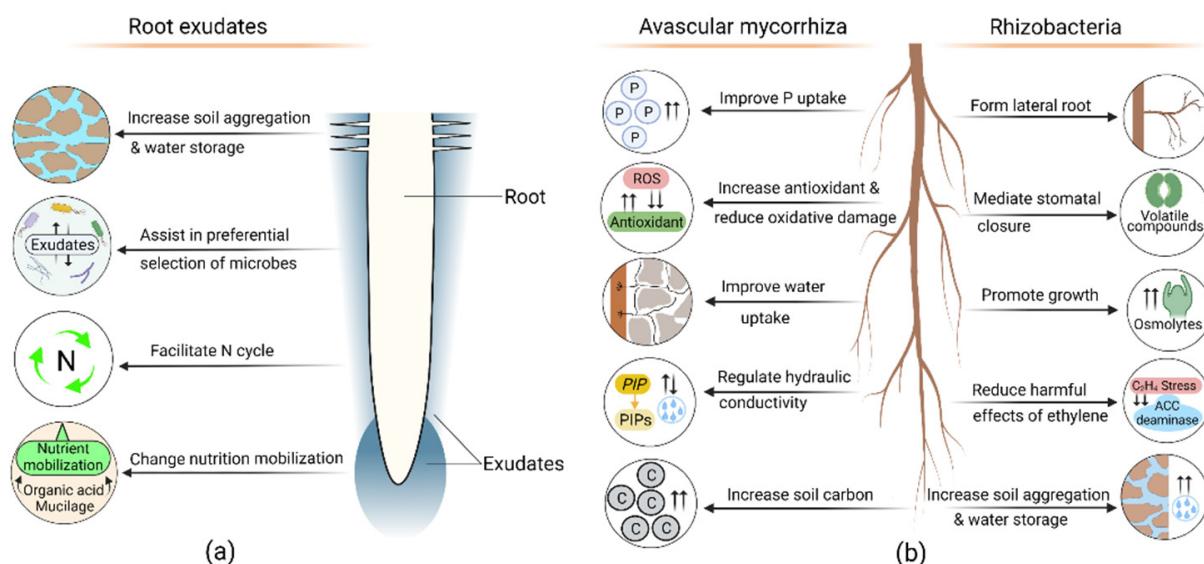


Figure 4. Beneficial roles of root exudates, avascular mycorrhiza (AM), and rhizobacteria in plant drought adaptation. **(a)** Exudates effects soil aggregation, water holding capacity, and nutrient mobilization. The plant preferentially selects microbes through exudation, which assists in drought adaptation. Exudates also influence the soil nitrogen (N) cycle. **(b)** AM increases phosphorus (P) and water uptake, affecting root hydraulic conductivity and reducing drought stress by producing antioxidants. AM also increases soil carbon. Rhizobacteria release exopolysaccharides, volatile compounds, osmolytes, ACC-deaminase, and phytohormones. These compounds increase soil aggregation, lateral root formation, and plant growth; mediate stomatal closure, reduce ethylene's harmful effect, and ultimately increase drought resistance.

8.3. Role of Rhizobacteria in Drought Adaptation

Rhizobacteria release phytohormone, for example, IAA, which promotes lateral root formation [116]. Rhizobacteria inoculated grapevines had increased IAA content and drought resistance [117]. Rhizobacterial exopolysaccharides mediate soil water content, increase soil aggregation, and form a protective capsule around soil aggregates [118], creating an area where water remains longer than the surrounding area [119]. Due to this increased soil aggregation and water content permeability, plants have improved water and nutrient uptake during drought [118]. Some rhizobacteria release volatile compounds that can increase drought resistance in the plant directly and/or indirectly [120], for example, volatile compound acetic acid stimulates bacterial biofilm formation and exopolysaccharides, a major biofilm constituent that increases water retention [120,121]. Similarly, Rhizobacteria, *Pseudomonas chlororaphis* O6, produced volatile metabolites, 2R, 3R-butanediol, mediating stomatal closure [122]. Water deficits increase the concentrations of C-rich compounds that adjust metabolic activity and promote growth, many of which are osmolytes [123,124]. In response to drought, rhizobacteria also produce osmolytes which act as a substitute for plant-produced osmolytes [125]. Staudinger et al. [123] found Rhizobia inoculated *Medicago truncatula* had an enhanced concentration of osmolytes, resulting in a delay in drought-induced leaf senescence. Under stress, ethylene regulates defense and growth responses, reducing root and shoot growth [124,126], with 1-aminocyclopropane-1-carboxylate (ACC) a crucial component of ethylene biosynthesis [126]. Rhizobacterial ACC-deaminase segregates and degrades plant ACC, thus interrupting ethylene biosynthesis and reducing the harmful effect of ethylene in drought [127]. During drought, rhizobacteria also increases activities of the antioxidant enzymes in faba bean (*Vicia faba*) [128] pea (*Pisum sativum* L.) and tomato (*Solanum lycopersicum* L.) [129,130]. Rhizobacteria similarly assist lateral root formation, thus increasing water uptake in maize [131] (Figure 4b).

Although some advances have been made in understanding the role of exudates and microbe symbiosis in plant adaptation to drought under a controlled environment, the study of root exudation in a natural setting is in its infancy [94]. Therefore, it requires a practical standardized root exudates collection method applicable in the field condition [94,132]. To increase the understanding of exudate function and microbial interaction during drought requires how; (i) intensity and composition of exudate changes in response to neighboring plants, (ii) root exudate components interact in the soil, (iii) microbes influence plant exudation, and (iv) microbial mucilage and hyphal exudation influence the overall process.

9. Root Soil Building Attributes in Drought Adaptation

Root and soil interchangeably alter each other, heterogeneous soil can change RSA, similarly root change, and improve soil [133]. Soil structure, nutrient availability, water dynamics, and C storage can be significantly altered by root, improving plant adaptation to drought. Thus, these root soil-building attributes contribute considerably to drought-tolerant crop improvement.

Roots change soil structure through compaction [134], dispersion [7], aggregation and creation of biopores [135], acting as a biological tiller [136]. When the root enters the soil, it first compresses the soil [133,134], then as the root grows, it releases mucilage and exudates, creating a porous rhizosheath [7,137] (Figure 5a). As the root grows further, its elongation zone transits into maturation, producing root hair, releasing exudates, and contributing to soil aggregation [138] (Figure 5c). Similarly, root symbiotic Rhizobacterial exopolysaccharides, AM hyphae, and hyphal exudates increase soil aggregation and change in soil structure [118,139]. Roots also improve soil structure by depositing a substantial portion of photosynthetic C as exudates (Figure 5). Below ground C assimilated by annual crops and grasses accounts for 21% and 33%, respectively [140]. This C increases water sorptivity and storage, and changes nutrition transport, ultimately improving drought adaptation. Moreover, roots represent a large proportion of the world's vegetation C pool [141], contributing more C to the soil. Among these root carbons, deep root carbon is more stable as C decomposition is limited in deep soil [142]. Lignin and suberin in the root cell wall are also key components stabilizing soil C as these degrade slowly [40,142]. Exudate supports rhizobiome communities that release additional C in the soil [143]; thus, roots influence overall C composition, flow, and soil structure near the root zone [144].

The root rhizosphere influences spatial heterogeneities and changes the hydrophilic nature of soil [7,137,145], increasing soil porosity and rendering the rhizosphere anisotropy. It contains a significantly higher water content than bulk soil, with drought-resistant genotypes generating a larger and more porous mass of rhizosheath than drought-sensitive genotypes [7]. Greater mucilage in drought-resistant genotypes may be the main reason for a porous rhizosheath and increased water storage. These changes in the soil are later shaped by rhizosphere microbes [7,145] and create further spatial heterogeneities in the soil, such as Rhizobacterial exopolysaccharides increasing water storage length in the root zone [119]. Root exudates also regulate nutrition flows in soil, with increased exudate concentration blocking soil pores, increasing the friction of immobile water, and creating slow-conducting flow paths, resulting in changes in nutrition transport in the soil [108]. Roots can affect nutrient availability in the soil according to plant needs, for example, organic acids in root exudates act as chelating agents, resulting in increased nutrient availability for ions such as iron and P [146].

Due to current cultivation practices, soil C is reduced in agricultural land, therefore affecting the water storage capacity. Currently, the global C debt in agricultural land for the top two meters of soil is 133 petagram [147]. Drought will continue to worsen due to reduced freshwater supply, soil water absorption and capacity. Roots and their symbiotic microbes offer huge potential for mitigating atmospheric C by sequestering it in the soil. All these roots' direct and indirect soil-building attributes can significantly increase resilience to drought.

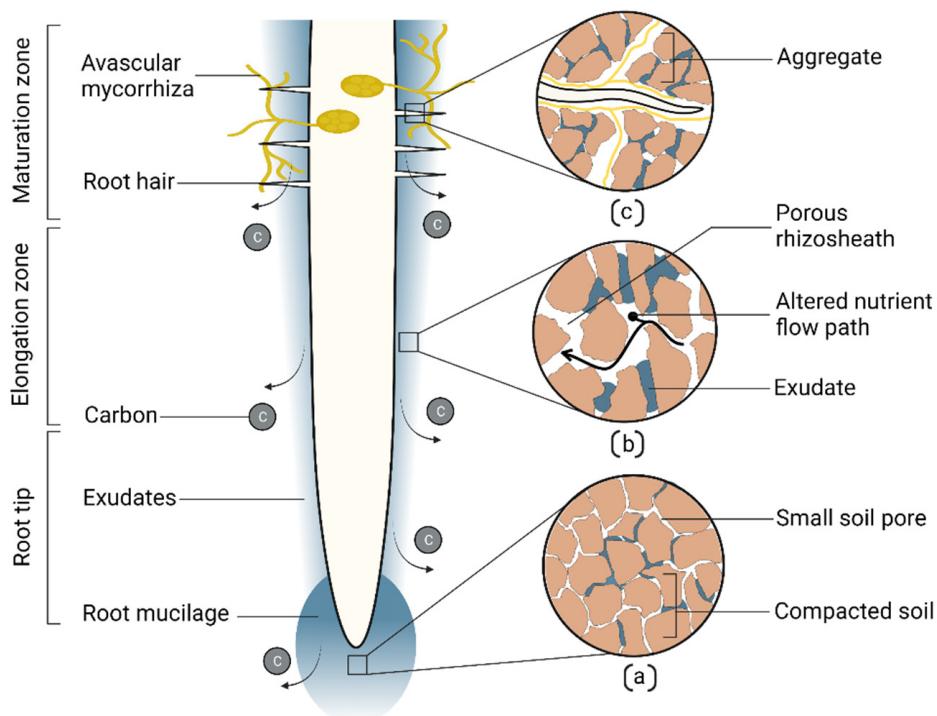


Figure 5. Root soil building attributes. Root exudation releases carbon into the soil. (a) Soil compaction and reduction of pore space in the root tip area. (b) Formation of porous rhizosheath area and altered nutrient flow path due to root mucilage. (c) Soil aggregation by root mucilage, root hair, and avascular mycorrhiza hyphae.

10. Drought Adaptive Root Ideotypes

Proposing an ideotype for drought defines idealized approaches for a breeding program to work towards. A universal ideotype for drought in a broader range of crops might not be applicable for a specific crop because roots of each crop interact and respond differently in response to soil type [148] and growing environment. However, several root drought-resistant traits are shared in a broad range of crops.

Wasson et al. [149] proposed deep root systems, increased root length density in medium and deep soil layers, reduced root length density in the topsoil, increased root hair growth, and increased xylem diameter as an ideal drought-resistant ideotype (Figure 6a). The deep root system would allow access to subsoil moisture in dry conditions but comes at the cost of C, otherwise, the plant would invest more in shoot and reproductive organs. However, it is possible to increase the rooting depth without extra C input by modifying specific root lengths [149]. During anthesis, insufficient root length density at depth can cause a lack of access to water in deep soil. Wasson et al. [149] proposed an ideal ideotype for drought tolerance, with few roots at the surface and more roots at depth, which have a large xylem and increased root hairs. Larger xylem lowers axial resistance, allowing water to pass from the soil into the root easier, and increased root hairs maintain higher hydraulic continuity between roots and soil, improving water acquisition in drying soil. Lynch [16] proposed a “steep, cheap, and deep” ideotype for efficient moisture uptake from the subsoil (Figure 6b), where cheap traits are mostly anatomical traits that reduce the metabolic cost of soil exploration (some of these traits are described in the above sections). Interestingly, Rao et al. [17] argued that the “steep, cheap, and deep” ideotype might not be efficient in low rainfall temperate regions with little water in deep soil, and dicot pulse crops might respond differently to monocot cereals. Thus, their “wide, shallow, and fine” ideotype emphasized dense roots in the upper surface to capture water from low rainfall events before it is lost by evaporation (Figure 6c). These shallow, fine roots might have an advantage in capturing immobile nutrients [84], such as P. How these attributes might

increase drought adaptation need to be investigated to quantify the merits of the “wide, shallow, and fine” ideotype alone. In addition, further research is needed to understand the effect of deep roots on (i) biological tillage [136]; (ii) biopores creation [135]; and (iii) C input in soil, and its impacts on soil water storage, subsequent crops, and long-term drought resilience. Perhaps, the immediate short-term solutions to drought resistance are in the “wide, shallow, and fine” ideotypes and long-term solutions are the deep root ideotypes.

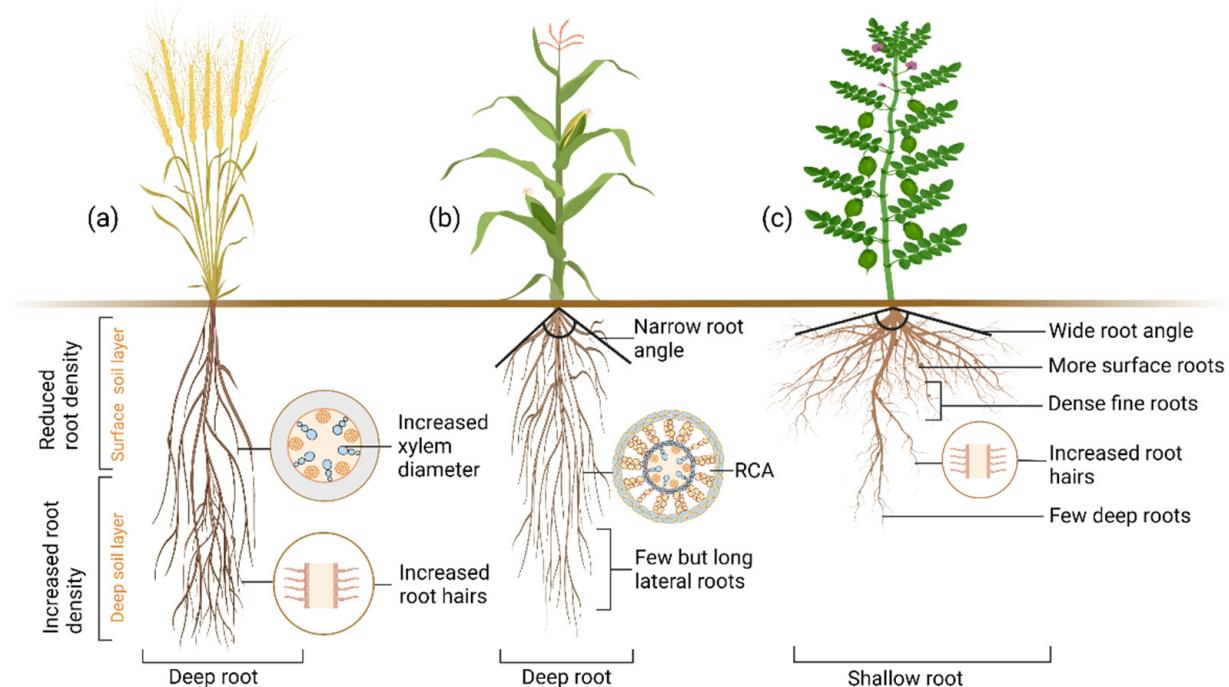


Figure 6. Different drought-adaptive root ideotypes. (a) Deep root, reduced root on soil surface layer, more root in the deep layer, and increased root hair and xylem diameter are crucial drought adaptation response traits (b) Deep root, narrow root angle, few but long lateral roots and more root cortical aerenchyma (RCA) are helpful for drought adaptation. (c) Shallow, wide-angle roots, more dense surface roots but few deep roots, and increased root hairs are the necessary root drought adaptation traits for low rainfall areas.

Contrasting root architectures—deep [16] versus shallow [17]—are proposed for drought-resistant ideotypes for different environments, thus, indicating the vulnerability of the ‘drought ideotype’, which is heavily biased by the plant growing environment. In addition, root architecture and growth do not fully represent root water uptake capability, and more profuse root growth does not necessarily relate to more water extraction [2,55], which also depends on the root hydraulics [2]. Thus, capturing the root’s complete water-harnessing ability and increasing the root drought adaptive ideotype reliability requires the consideration of root hydraulics along with architecture.

The success of a drought-resistant ideotype for a cropping system will depend on genotype, environment, and management practices. Formulating a root ideotype that performs well in broader and varied areas will require incorporating multiple drought response traits. Moreover, developing plants based on a particular root ideotype for a large area will reduce the root diversity and might lack the necessary adaptive ability to cope with climate change. Thus, a root ideotype targeted for a small and specific area might be efficient and facilitate root diversity. When formulating ideotypes, we should also consider incorporating sustainability and soil regeneration along with maximizing productivity. Incorporating root soil-building attributes in the drought ideotype offers a scope of soil regeneration and long-term resilience to drought.

11. Conclusions

Roots are the other half of plants that can significantly contribute to sustain productivity under drought conditions. A range of root attributes, namely RSA, root structural, anatomical, and hydraulic traits are important for drought adaptation. Equally, interactions among these traits, their plasticity and root soil building attributes also affect the overall adaption process. Research has largely been focused on root architectural responses when developing drought-resilient plants. However, root structural, physiological, and cellular regulators, and their interactions with microbes, are also important. Several root structural traits, such as taproot diameter and length, root hair number, density and length, and root angle are phenotyping-friendly, high-throughput, and relatively easy to deploy in plant breeding. Whereas, root anatomical trait screening is still low throughput, requiring further research to find traits that improve root penetration in dry soil, reduce metabolic cost, improve water uptake, and facilitate AM colonization. Consideration also needs to be taken to incorporate root hydraulics and associated molecular regulators such as *XDN1*, *MdMYB46*, and aquaporins (Figure 2) to ensure increased water uptake. Molecular mechanisms for root plasticity, hydropatterning, hydrotropism, and gravitropism are well characterized (Figure 3) to harness this knowledge in drought-resilient crop improvement.

Improvements in root exudate collection and rhizosphere-microbe phenotyping methods will increase understanding of plant–microbes responses during drought. Improving soil C can significantly augment water absorption and storage and offers potential for a long-term solution to continue increasing agricultural production under a limited water supply. Overall, careful selection and incorporation of root architectural, structural, anatomical, hydraulic, molecular, and soil building traits, and consideration of root trait plasticity and interaction are integral in developing desirable ideotypes to sustain agricultural production in drought-prone areas.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We thank Emily Thoday-Kennedy for proof reading of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Den Herder, G.; Van Isterdael, G.; Beeckman, T.; De Smet, I. The roots of a new green revolution. *Trends Plant Sci.* **2010**, *15*, 600–607. [[CrossRef](#)] [[PubMed](#)]
2. Maurel, C.; Nacry, P. Root architecture and hydraulics converge for acclimation to changing water availability. *Nat. Plants* **2020**, *6*, 744–749. [[CrossRef](#)] [[PubMed](#)]
3. Delory, B.M.; Hernandez-Soriano, M.C.; Wacker, T.S.; Dimitrova, A.; Ding, Y.; Greeley, L.A.; Ng, J.L.P.; Mesa-Marín, J.; Xie, L.; Zheng, C. A snapshot of the root phenotyping landscape in 2021. *bioRxiv* **2022**. [[CrossRef](#)]
4. Wasson, A.P.; Nagel, K.A.; Tracy, S.; Watt, M. Beyond Digging: Noninvasive Root and Rhizosphere Phenotyping. *Trends Plant Sci.* **2020**, *25*, 119–120. [[CrossRef](#)]
5. Uga, Y.; Sugimoto, K.; Ogawa, S.; Rane, J.; Ishitani, M.; Hara, N.; Kitomi, Y.; Inukai, Y.; Ono, K.; Kanno, N.; et al. Control of root system architecture by DEEPER ROOTING 1 increases rice yield under drought conditions. *Nat. Genet.* **2013**, *45*, 1097–1102. [[CrossRef](#)] [[PubMed](#)]
6. Zhan, A.; Schneider, H.; Lynch, J.P. Reduced Lateral Root Branching Density Improves Drought Tolerance in Maize. *Plant Physiol.* **2015**, *168*, 1603–1615. [[CrossRef](#)]
7. Rabbi, S.M.F.; Tighe, M.K.; Flavel, R.J.; Kaiser, B.N.; Guppy, C.N.; Zhang, X.; Young, I.M. Plant roots redesign the rhizosphere to alter the three-dimensional physical architecture and water dynamics. *New Phytol.* **2018**, *219*, 542–550. [[CrossRef](#)]
8. Gao, Y.; Lynch, J.P. Reduced crown root number improves water acquisition under water deficit stress in maize (*Zea mays* L.). *J. Exp. Bot.* **2016**, *67*, 4545–4557. [[CrossRef](#)]
9. Kiran, B.O.; Karabhananal, S.S.; Patil, S.B.; Ashwathama, V.H.; M, G.; Sajjanar; Jolli, R.B.; Tonapi, V.A. Phenotyping sorghum for drought-adaptive physiological and root architectural traits under water-limited environments. *Cereal Res. Commun.* **2022**, *50*, 1–9. [[CrossRef](#)]

10. Chai, Y.N.; Schachtman, D.P. Root exudates impact plant performance under abiotic stress. *Trends Plant Sci.* **2022**, *27*, 80–91. [[CrossRef](#)]
11. Thirkell, T.J.; Charters, M.D.; Elliott, A.J.; Sait, S.M.; Field, K.J. Are mycorrhizal fungi our sustainable saviours? Considerations for achieving food security. *J. Ecol.* **2017**, *105*, 921–929. [[CrossRef](#)]
12. Aroca, R.; Porcel, R.; Ruiz-Lozano, J.M. How does arbuscular mycorrhizal symbiosis regulate root hydraulic properties and plasma membrane aquaporins in *Phaseolus vulgaris* under drought, cold or salinity stresses? *New Phytol.* **2007**, *173*, 808–816. [[CrossRef](#)] [[PubMed](#)]
13. Freschet, G.T.; Pagès, L.; Iversen, C.M.; Comas, L.H.; Rewald, B.; Roumet, C.; Klimešová, J.; Zadworny, M.; Poorter, H.; Postma, J.A.; et al. A starting guide to root ecology: Strengthening ecological concepts and standardising root classification, sampling, processing and trait measurements. *New Phytol.* **2021**, *232*, 973–1122. [[CrossRef](#)]
14. Ranjan, A.; Sinha, R.; Singla-Pareek, S.L.; Pareek, A.; Singh, A.K. Shaping the Root System Architecture in Plants for Adaptation to Drought Stress. *Physiol. Plant.* **2022**, *174*, e13651. [[CrossRef](#)]
15. Rongsawat, T.; Peltier, J.B.; Boyer, J.C.; Véry, A.A.; Sentenac, H. Looking for Root Hairs to Overcome Poor Soils. *Trends Plant Sci.* **2021**, *26*, 83–94. [[CrossRef](#)] [[PubMed](#)]
16. Lynch, J.P. Steep, cheap and deep: An ideotype to optimize water and N acquisition by maize root systems. *Ann. Bot.* **2013**, *112*, 347–357. [[CrossRef](#)] [[PubMed](#)]
17. Rao, S.; Armstrong, R.; Silva-Perez, V.; Tefera, A.T.; Rosewarne, G.M. Pulse Root Ideotype for Water Stress in Temperate Cropping System. *Plants* **2021**, *10*, 692. [[CrossRef](#)]
18. Giehl, R.F.H.; Von Wieren, N. Hydropatterning—how roots test the waters. *Science* **2018**, *362*, 1358–1359. [[CrossRef](#)]
19. Orman-Ligeza, B.; Morris, E.C.; Parizot, B.; Lavigne, T.; Babé, A.; Ligeza, A.; Klein, S.; Sturrock, C.; Xuan, W.; Novák, O.; et al. The Xerobranching Response Represses Lateral Root Formation When Roots Are Not in Contact with Water. *Curr. Biol.* **2018**, *28*, 3165–3173.e3165. [[CrossRef](#)]
20. Xiao, G.; Zhang, Y. Adaptive Growth: Shaping Auxin-Mediated Root System Architecture. *Trends Plant Sci.* **2020**, *25*, 121–123. [[CrossRef](#)]
21. Ogura, T.; Goeschl, C.; Filiault, D.; Mirea, M.; Slovak, R.; Wolhrab, B.; Satbhai, S.B.; Busch, W. Root System Depth in *Arabidopsis* Is Shaped by EXOCYST70A3 via the Dynamic Modulation of Auxin Transport. *Cell* **2019**, *178*, 400–412.e416. [[CrossRef](#)] [[PubMed](#)]
22. Waidmann, S.; Ruiz Rosquete, M.; Scholler, M.; Sarkel, E.; Lindner, H.; LaRue, T.; Petrik, I.; Dunser, K.; Martopawiro, S.; Sasidharan, R.; et al. Cytokinin functions as an asymmetric and anti-gravitropic signal in lateral roots. *Nat. Commun.* **2019**, *10*, 3540. [[CrossRef](#)] [[PubMed](#)]
23. Li, H.; Testerink, C.; Zhang, Y. How roots and shoots communicate through stressful times. *Trends Plant Sci.* **2021**, *26*, 940–952. [[CrossRef](#)] [[PubMed](#)]
24. Bacher, H.; Sharaby, Y.; Walia, H.; Peleg, Z. Modifying root-to-shoot ratio improves root water influxes in wheat under drought stress. *J. Exp. Bot.* **2021**, *73*, 1643–1654. [[CrossRef](#)]
25. Becker, S.R.; Byrne, P.F.; Reid, S.D.; Bauerle, W.L.; McKay, J.K.; Haley, S.D. Root traits contributing to drought tolerance of synthetic hexaploid wheat in a greenhouse study. *Euphytica* **2015**, *207*, 213–224. [[CrossRef](#)]
26. Eziz, A.; Yan, Z.; Tian, D.; Han, W.; Tang, Z.; Fang, J. Drought effect on plant biomass allocation: A meta-analysis. *Ecol. Evol.* **2017**, *7*, 11002–11010. [[CrossRef](#)]
27. Chen, Q.; Hu, T.; Li, X.; Song, C.P.; Zhu, J.K.; Chen, L.; Zhao, Y. Phosphorylation of SWEET sucrose transporters regulates plant root:shoot ratio under drought. *Nat. Plants* **2022**, *8*, 68–77. [[CrossRef](#)]
28. Caradus, J.R.; Woodfield, D.R. Genetic control of adaptive root characteristics in white clover. In *Root Demographics and Their Efficiencies in Sustainable Agriculture, Grasslands and Forest Ecosystems*; Springer: Berlin/Heidelberg, Germany, 1998; pp. 651–662.
29. Chimungu, J.G.; Brown, K.M.; Lynch, J.P. Large Root Cortical Cell Size Improves Drought Tolerance in Maize. *Plant Physiol.* **2014**, *166*, 2166–2178. [[CrossRef](#)]
30. Prince, S.J.; Murphy, M.; Mutava, R.N.; Durnell, L.A.; Valliyodan, B.; Shannon, J.G.; Nguyen, H.T. Root xylem plasticity to improve water use and yield in water-stressed soybean. *J. Exp. Bot.* **2017**, *68*, 2027–2036. [[CrossRef](#)]
31. Galindo-Castañeda, T.; Brown, K.M.; Kuldau, G.A.; Roth, G.W.; Wenner, N.G.; Ray, S.; Schneider, H.; Lynch, J.P. Root cortical anatomy is associated with differential pathogenic and symbiotic fungal colonization in maize. *Plant Cell Environ.* **2019**, *42*, 2999–3014. [[CrossRef](#)]
32. Wang, H.; Wang, Z.; Dong, X. Anatomical structures of fine roots of 91 vascular plant species from four groups in a temperate forest in Northeast China. *PLoS ONE* **2019**, *14*, e0215126.
33. Schneider, H.M.; Strock, C.F.; Hanlon, M.T.; Vanhees, D.J.; Perkins, A.C.; Ajmera, I.B.; Sidhu, J.S.; Mooney, S.J.; Brown, K.M.; Lynch, J.P. Multiseriate cortical sclerenchyma enhance root penetration in compacted soils. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2012087118. [[CrossRef](#)] [[PubMed](#)]
34. Colombi, T.; Herrmann, A.M.; Vallenback, P.; Keller, T. Cortical Cell Diameter Is Key To Energy Costs of Root Growth in Wheat. *Plant Physiol.* **2019**, *180*, 2049–2060. [[CrossRef](#)]
35. Lynch, J.P.; Chimungu, J.G.; Brown, K.M. Root anatomical phenes associated with water acquisition from drying soil: Targets for crop improvement. *J. Exp. Bot.* **2014**, *65*, 6155–6166. [[CrossRef](#)] [[PubMed](#)]
36. Chimungu, J.G.; Brown, K.M.; Lynch, J.P. Reduced Root Cortical Cell File Number Improves Drought Tolerance in Maize. *Plant Physiol.* **2014**, *166*, 1943–1955. [[CrossRef](#)]

37. Karlova, R.; Boer, D.; Hayes, S.; Testerink, C. Root plasticity under abiotic stress. *Plant Physiol.* **2021**, *187*, 1057–1070. [[CrossRef](#)]
38. Vanhees, D.J.; Loades, K.W.; Bengough, A.G.; Mooney, S.J.; Lynch, J.P. Root anatomical traits contribute to deeper rooting of maize under compacted field conditions. *J. Exp. Bot.* **2020**, *71*, 4243–4257. [[CrossRef](#)]
39. Zhu, J.; Brown, K.M.; Lynch, J.P. Root cortical aerenchyma improves the drought tolerance of maize (*Zea mays* L.). *Plant Cell Environ.* **2010**, *33*, 740–749. [[CrossRef](#)]
40. Lynch, J.P.; Mooney, S.J.; Strock, C.F.; Schneider, H.M. Future roots for future soils. *Plant Cell Environ.* **2022**, *45*, 620–636. [[CrossRef](#)]
41. Chimungu, J.G.; Loades, K.W.; Lynch, J.P. Root anatomical phenes predict root penetration ability and biomechanical properties in maize (*Zea Mays*). *J. Exp. Bot.* **2015**, *66*, 3151–3162. [[CrossRef](#)]
42. Jansa, J.; Mozafar, A.; Kuhn, G.; Anken, T.; Ruh, R.; Sanders, I.R.; Frossard, E. Soil Tillage Affects the Community Structure of Mycorrhizal Fungi in Maize Roots. *Ecol. Appl.* **2003**, *13*, 1164–1176. [[CrossRef](#)]
43. Smith, S.E.; Smith, F.A. Roles of Arbuscular Mycorrhizas in Plant Nutrition and Growth: New Paradigms from Cellular to Ecosystem Scales. *Annu. Rev. Plant Biol.* **2011**, *62*, 227–250. [[CrossRef](#)] [[PubMed](#)]
44. Chareesri, A.; De Deyn, G.B.; Sergeeva, L.; Polthanee, A.; Kuyper, T.W. Increased arbuscular mycorrhizal fungal colonization reduces yield loss of rice (*Oryza sativa* L.) under drought. *Mycorrhiza* **2020**, *30*, 315–328. [[CrossRef](#)] [[PubMed](#)]
45. Ruiz-Lozano, J.M.; Aroca, R.; Zamarreño, Á.M.; Molina, S.; Andreo-Jiménez, B.; Porcel, R.; García-Mina, J.M.; Ruyter-Spira, C.; López-Ráez, J.A. Arbuscular mycorrhizal symbiosis induces strigolactone biosynthesis under drought and improves drought tolerance in lettuce and tomato. *Plant Cell Environ.* **2016**, *39*, 441–452. [[CrossRef](#)]
46. Kiers, E.T.; Duhamel, M.; Beesetty, Y.; Mensah, J.A.; Franken, O.; Verbruggen, E.; Fellbaum, C.R.; Kowalchuk, G.A.; Hart, M.M.; Bago, A. Reciprocal rewards stabilize cooperation in the mycorrhizal symbiosis. *Science* **2011**, *333*, 880–882. [[CrossRef](#)]
47. Chitarra, W.; Pagliarani, C.; Maserti, B.; Lumini, E.; Siciliano, I.; Cascone, P.; Schubert, A.; Gambino, G.; Balestrini, R.; Guerrieri, E. Insights on the Impact of Arbuscular Mycorrhizal Symbiosis on Tomato Tolerance to Water Stress. *Plant Physiol.* **2016**, *171*, 1009–1023. [[CrossRef](#)]
48. Wang, Y.; Li, Z.; Wang, S.; Wang, W.; Wang, N.; Gu, J. Variations in Arbuscular Mycorrhizal Colonization Associated with Root Diameter and Hypodermis Passages Cells across Temperate and Tropical Woody Species. *Forests* **2022**, *13*, 140. [[CrossRef](#)]
49. Guo, D.; Xia, M.; Wei, X.; Chang, W.; Liu, Y.; Wang, Z. Anatomical traits associated with absorption and mycorrhizal colonization are linked to root branch order in twenty-three Chinese temperate tree species. *New Phytol.* **2008**, *180*, 673–683. [[CrossRef](#)]
50. Dreyer, B.; Morte, A.; López, J.Á.; Honrubia, M. Comparative study of mycorrhizal susceptibility and anatomy of four palm species. *Mycorrhiza* **2010**, *20*, 103–115. [[CrossRef](#)]
51. Lynch, J.P.; Strock, C.F.; Schneider, H.M.; Sidhu, J.S.; Ajmera, I.; Galindo-Castañeda, T.; Klein, S.P.; Hanlon, M.T. Root anatomy and soil resource capture. *Plant Soil* **2021**, *466*, 21–63. [[CrossRef](#)]
52. Strock, C.F.; Burridge, J.D.; Niemiec, M.D.; Brown, K.M.; Lynch, J.P. Root metaxylem and architecture phenotypes integrate to regulate water use under drought stress. *Plant Cell Environ.* **2021**, *44*, 49–67. [[CrossRef](#)] [[PubMed](#)]
53. Tan, J.; Ben-Gal, A.; Shtein, I.; Bustan, A.; Dag, A.; Erel, R. Root structural plasticity enhances salt tolerance in mature olives. *Environ. Exp. Bot.* **2020**, *179*, 104224. [[CrossRef](#)]
54. Shahzad, Z.; Canut, M.; Tournaire-Roux, C.; Martinière, A.; Boursiac, Y.; Loudet, O.; Maurel, C. A Potassium-Dependent Oxygen Sensing Pathway Regulates Plant Root Hydraulics. *Cell* **2016**, *167*, 87–98.e14. [[CrossRef](#)] [[PubMed](#)]
55. Vadez, V. Root hydraulics: The forgotten side of roots in drought adaptation. *Field Crops Res.* **2014**, *165*, 15–24. [[CrossRef](#)]
56. Cai, G.; Ahmed, M.A.; Abdalla, M.; Carminati, A. Root hydraulic phenotypes impacting water uptake in drying soils. *Plant Cell Environ.* **2022**, *45*, 650–663. [[CrossRef](#)]
57. Tang, N.; Shahzad, Z.; Lonjon, F.; Loudet, O.; Vailleau, F.; Maurel, C. Natural variation at XND1 impacts root hydraulics and trade-off for stress responses in *Arabidopsis*. *Nat. Commun.* **2018**, *9*, 3884. [[CrossRef](#)]
58. Matsuo, N.; Ozawa, K.; Mochizuki, T. Genotypic differences in root hydraulic conductance of rice (*Oryza sativa* L.) in response to water regimes. *Plant Soil* **2009**, *316*, 25–34. [[CrossRef](#)]
59. Fennell, A.; Markhart, A.H. Rapid acclimation of root hydraulic conductivity to low temperature. *J. Exp. Bot.* **1998**, *49*, 879–884. [[CrossRef](#)]
60. Lo Gullo, M.A.; Nardini, A.; Salleo, S.; Tyree, M.T. Changes in root hydraulic conductance (KR) of *Olea oleaster* seedlings following drought stress and irrigation. *New Phytol.* **1998**, *140*, 25–31. [[CrossRef](#)]
61. Saliendra, N.Z.; Meinzer, F.C. Genotypic, Developmental and Drought-Induced Differences in Root Hydraulic Conductance of Contrasting Sugarcane Cultivars. *J. Exp. Bot.* **1992**, *43*, 1209–1217. [[CrossRef](#)]
62. Rieger, M.; Litvin, P. Root system hydraulic conductivity in species with contrasting root anatomy. *J. Exp. Bot.* **1999**, *50*, 201–209. [[CrossRef](#)]
63. Fan, M.; Bai, R.; Zhao, X.; Zhang, J. Aerenchyma formed under phosphorus deficiency contributes to the reduced root hydraulic conductivity in maize roots. *J. Integr. Plant Biol.* **2007**, *49*, 598–604. [[CrossRef](#)]
64. Muhsin, T.M.; Zwiazek, J.J. Ectomycorrhizas increase apoplastic water transport and root hydraulic conductivity in *Ulmus americana* seedlings. *New Phytol.* **2002**, *153*, 153–158. [[CrossRef](#)]
65. Geng, D.; Chen, P.; Shen, X.; Zhang, Y.; Li, X.; Jiang, L.; Xie, Y.; Niu, C.; Zhang, J.; Huang, X.; et al. MdMYB88 and MdMYB124 Enhance Drought Tolerance by Modulating Root Vessels and Cell Walls in Apple. *Plant Physiol.* **2018**, *178*, 1296–1309. [[CrossRef](#)] [[PubMed](#)]

66. Javot, H.; Lauvergeat, V.; Santoni, V.; Martin-Laurent, F.; Güçlü, J.; Vinh, J.; Heyes, J.; Franck, K.I.; Schäffner, A.R.; Bouchez, D.; et al. Role of a Single Aquaporin Isoform in Root Water Uptake. *Plant Cell* **2003**, *15*, 509–522. [[CrossRef](#)] [[PubMed](#)]
67. Postaire, O.; Tournaire-Roux, C.; Grondin, A.; Boursiac, Y.; Morillon, R.; Schäffner, A.R.; Maurel, C. A PIP1 Aquaporin Contributes to Hydrostatic Pressure-Induced Water Transport in Both the Root and Rosette of Arabidopsis. *Plant Physiol.* **2009**, *152*, 1418–1430. [[CrossRef](#)]
68. Siefritz, F.; Tyree, M.T.; Lovisolo, C.; Schubert, A.; Kaldenhoff, R. PIP1 Plasma Membrane Aquaporins in Tobacco. *Plant Cell* **2002**, *14*, 869–876. [[CrossRef](#)]
69. Peret, B.; Li, G.; Zhao, J.; Band, L.R.; Voss, U.; Postaire, O.; Luu, D.T.; Da Ines, O.; Casimiro, I.; Lucas, M.; et al. Auxin regulates aquaporin function to facilitate lateral root emergence. *Nat. Cell Biol.* **2012**, *14*, 991–998. [[CrossRef](#)]
70. Caldeira, C.F.; Jeanguenin, L.; Chaumont, F.; Tardieu, F. Circadian rhythms of hydraulic conductance and growth are enhanced by drought and improve plant performance. *Nat. Commun.* **2014**, *5*, 5365. [[CrossRef](#)]
71. Calvo-Polanco, M.; Ribeyre, Z.; Dauzat, M.; Reyt, G.; Hidalgo-Shrestha, C.; Diehl, P.; Frenger, M.; Simonneau, T.; Muller, B.; Salt, D.E.; et al. Physiological roles of Casparyan strips and suberin in the transport of water and solutes. *New Phytol.* **2021**, *232*, 2295–2307. [[CrossRef](#)]
72. Lefebvre, V.; Fortabat, M.-N.; Ducamp, A.; North, H.M.; Maia-Grondard, A.; Trouverie, J.; Boursiac, Y.; Mouille, G.; Durand-Tardif, M. ESKIMO1 Disruption in Arabidopsis Alters Vascular Tissue and Impairs Water Transport. *PLoS ONE* **2011**, *6*, e16645. [[CrossRef](#)] [[PubMed](#)]
73. Zwieniecki, M.A.; Boersma, L. A technique to measure root tip hydraulic conductivity and root water potential simultaneously. *J. Exp. Bot.* **1997**, *48*, 333–336. [[CrossRef](#)]
74. Li, Q.-M.; Liu, B.-B. Comparison of three methods for determination of root hydraulic conductivity of maize (*Zea mays* L.) root system. *Agric. Sci. China* **2010**, *9*, 1438–1447. [[CrossRef](#)]
75. Sutka, M.; Li, G.; Boudet, J.; Boursiac, Y.; Doumas, P.; Maurel, C. Natural variation of root hydraulics in Arabidopsis grown in normal and salt-stressed conditions. *Plant Physiol.* **2011**, *155*, 1264–1276. [[CrossRef](#)] [[PubMed](#)]
76. Heymans, A.; Couvreur, V.; Lobet, G. Combining cross-section images and modeling tools to create high-resolution root system hydraulic atlases in *Zea mays*. *Plant Direct* **2021**, *5*, e334. [[CrossRef](#)]
77. Ajmera, I.; Henry, A.; Radanielson, A.M.; Klein, S.P.; Ianevski, A.; Bennett, M.J.; Band, L.R.; Lynch, J.P. Integrated root phenotypes for improved rice performance under low nitrogen availability. *Plant Cell Environ.* **2022**, *45*, 805–822. [[CrossRef](#)]
78. Haling, R.E.; Brown, L.K.; Bengough, A.G.; Young, I.M.; Hallett, P.D.; White, P.J.; George, T.S. Root hairs improve root penetration, root–soil contact, and phosphorus acquisition in soils of different strength. *J. Exp. Bot.* **2013**, *64*, 3711–3721. [[CrossRef](#)]
79. Li, T.; Lin, G.; Zhang, X.; Chen, Y.; Zhang, S.; Chen, B. Relative importance of an arbuscular mycorrhizal fungus (*Rhizophagus intraradices*) and root hairs in plant drought tolerance. *Mycorrhiza* **2014**, *24*, 595–602. [[CrossRef](#)]
80. Huang, Y.-M.; Zou, Y.-N.; Wu, Q.-S. Alleviation of drought stress by mycorrhizas is related to increased root H₂O₂ efflux in trifoliolate orange. *Sci. Rep.* **2017**, *7*, 42335. [[CrossRef](#)]
81. Li, J.; Meng, B.; Chai, H.; Yang, X.; Song, W.; Li, S.; Lu, A.; Zhang, T.; Sun, W. Arbuscular Mycorrhizal Fungi Alleviate Drought Stress in C3 (*Leymus chinensis*) and C4 (*Hemarthria altissima*) Grasses via Altering Antioxidant Enzyme Activities and Photosynthesis. *Front. Plant Sci.* **2019**, *10*, 499. [[CrossRef](#)]
82. Guseman, J.M.; Webb, K.; Srinivasan, C.; Dardick, C. DRO1 influences root system architecture in Arabidopsis and Prunus species. *Plant J.* **2017**, *89*, 1093–1105. [[CrossRef](#)]
83. Kitomi, Y.; Hanzawa, E.; Kuya, N.; Inoue, H.; Hara, N.; Kawai, S.; Kanno, N.; Endo, M.; Sugimoto, K.; Yamazaki, T.; et al. Root angle modifications by the DRO1 homolog improve rice yields in saline paddy fields. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 21242–21250. [[CrossRef](#)] [[PubMed](#)]
84. Van Der Bom, F.J.T.; Williams, A.; Bell, M.J. Root architecture for improved resource capture: Trade-offs in complex environments. *J. Exp. Bot.* **2020**, *71*, 5752–5763. [[CrossRef](#)] [[PubMed](#)]
85. Okushima, Y.; Fukaki, H.; Onoda, M.; Theologis, A.; Tasaka, M. ARF7 and ARF19 Regulate Lateral Root Formation via Direct Activation of LBD/ASL Genes in Arabidopsis. *Plant Cell* **2007**, *19*, 118–130. [[CrossRef](#)] [[PubMed](#)]
86. Wilmoth, J.C.; Wang, S.; Tiwari, S.B.; Joshi, A.D.; Hagen, G.; Guilfoyle, T.J.; Alonso, J.M.; Ecker, J.R.; Reed, J.W. NPH4/ARF7 and ARF19 promote leaf expansion and auxin-induced lateral root formation. *Plant J.* **2005**, *43*, 118–130. [[CrossRef](#)]
87. Orosa-Puente, B.; Leftley, N.; von Wangenheim, D.; Banda, J.; Srivastava, A.K.; Hill, K.; Truskina, J.; Bhosale, R.; Morris, E.; Srivastava, M. Root branching toward water involves posttranslational modification of transcription factor ARF7. *Science* **2018**, *362*, 1407–1410. [[CrossRef](#)]
88. Wang, H.; Hu, Z.; Huang, K.; Han, Y.; Zhao, A.; Han, H.; Song, L.; Fan, C.; Li, R.; Xin, M.; et al. Three genomes differentially contribute to the seedling lateral root number in allohexaploid wheat: Evidence from phenotype evolution and gene expression. *Plant J.* **2018**, *95*, 976–987. [[CrossRef](#)]
89. Zhao, H.; Ma, T.; Wang, X.; Deng, Y.; Ma, H.; Zhang, R.; Zhao, J. OsAUX1 controls lateral root initiation in rice (*Oryza sativa* L.). *Plant Cell Environ.* **2015**, *38*, 2208–2222. [[CrossRef](#)]
90. Shkolnik, D.; Krieger, G.; Nuriel, R.; Fromm, H. Hydrotropism: Root Bending Does Not Require Auxin Redistribution. *Mol. Plant* **2016**, *9*, 757–759. [[CrossRef](#)]

91. Dietrich, D.; Pang, L.; Kobayashi, A.; Fozard, J.A.; Boudolf, V.; Bhosale, R.; Antoni, R.; Nguyen, T.; Hiratsuka, S.; Fujii, N. Root hydrotropism is controlled via a cortex-specific growth mechanism. *Nat. Plants* **2017**, *3*, 17057. [[CrossRef](#)]
92. Kobayashi, A.; Takahashi, A.; Kakimoto, Y.; Miyazawa, Y.; Fujii, N.; Higashitani, A.; Takahashi, H. A gene essential for hydrotropism in roots. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 4724–4729. [[CrossRef](#)]
93. Vries, F.T.d.; Griffiths, R.I.; Knight, C.G.; Nicolitch, O.; Williams, A. Harnessing rhizosphere microbiomes for drought-resilient crop production. *Science* **2020**, *368*, 270–274. [[CrossRef](#)] [[PubMed](#)]
94. Wen, Z.; White, P.J.; Shen, J.; Lambers, H. Linking root exudation to belowground economic traits for resource acquisition. *New Phytol.* **2022**, *233*, 1620–1635. [[CrossRef](#)] [[PubMed](#)]
95. Gargallo-Garriga, A.; Preece, C.; Sardans, J.; Oravec, M.; Urban, O.; Penuelas, J. Root exudate metabolomes change under drought and show limited capacity for recovery. *Sci. Rep.* **2018**, *8*, 12696. [[CrossRef](#)] [[PubMed](#)]
96. Basirat, M.; Mousavi, S.M.; Abbaszadeh, S.; Ebrahimi, M.; Zarebanadkouki, M. The rhizosheath: A potential root trait helping plants to tolerate drought stress. *Plant Soil* **2019**, *445*, 565–575. [[CrossRef](#)]
97. Etesami, H. Potential advantage of rhizosheath microbiome, in contrast to rhizosphere microbiome, to improve drought tolerance in crops. *Rhizosphere* **2021**, *20*, 100439. [[CrossRef](#)]
98. Young, I.M. Variation in moisture contents between bulk soil and the rhizosheath of wheat (*Triticum aestivum* L. cv. Wembley). *New Phytol.* **1995**, *130*, 135–139. [[CrossRef](#)]
99. Hu, L.; Robert, C.A.M.; Cadot, S.; Zhang, X.; Ye, M.; Li, B.; Manzo, D.; Chervet, N.; Steinger, T.; van der Heijden, M.G.A.; et al. Root exudate metabolites drive plant-soil feedbacks on growth and defense by shaping the rhizosphere microbiota. *Nat. Commun.* **2018**, *9*, 2738. [[CrossRef](#)]
100. Jacoby, R.P.; Koprivova, A.; Kopriva, S. Pinpointing secondary metabolites that shape the composition and function of the plant microbiome. *J. Exp. Bot.* **2021**, *72*, 57–69. [[CrossRef](#)]
101. Huang, A.C.; Jiang, T.; Liu, Y.X.; Bai, Y.C.; Reed, J.; Qu, B.; Goossens, A.; Nutzmann, H.W.; Bai, Y.; Osbourn, A. A specialized metabolic network selectively modulates *Arabidopsis* root microbiota. *Science* **2019**, *364*, eaau6389. [[CrossRef](#)]
102. Zhelnina, K.; Louie, K.B.; Hao, Z.; Mansoori, N.; Da Rocha, U.N.; Shi, S.; Cho, H.; Karaoz, U.; Loqué, D.; Bowen, B.P.; et al. Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly. *Nat. Microbiol.* **2018**, *3*, 470–480. [[CrossRef](#)]
103. Xu, L.; Coleman-Derr, D. Causes and consequences of a conserved bacterial root microbiome response to drought stress. *Curr. Opin. Microbiol.* **2019**, *49*, 1–6. [[CrossRef](#)]
104. Henry, A.; Doucette, W.; Norton, J.; Bugbee, B. Changes in crested wheatgrass root exudation caused by flood, drought, and nutrient stress. *J. Environ. Qual.* **2007**, *36*, 904–912. [[CrossRef](#)] [[PubMed](#)]
105. Williams, A.; Vries, F.T. Plant root exudation under drought: Implications for ecosystem functioning. *New Phytol.* **2020**, *225*, 1899–1905. [[CrossRef](#)] [[PubMed](#)]
106. Naylor, D.; Degraaf, S.; Purdom, E.; Coleman-Derr, D. Drought and host selection influence bacterial community dynamics in the grass root microbiome. *ISME J.* **2017**, *11*, 2691–2704. [[CrossRef](#)] [[PubMed](#)]
107. Jones, D.L.; Darrah, P.R. Role of root derived organic acids in the mobilization of nutrients from the rhizosphere. *Plant Soil* **1994**, *166*, 247–257. [[CrossRef](#)]
108. Paporisch, A.; Bavli, H.; Strickman, R.J.; Neumann, R.B.; Schwartz, N. Root Exudates Alters Nutrient Transport in Soil. *Water Resour. Res.* **2021**, *57*, e2021WR029976. [[CrossRef](#)]
109. Jayne, B.; Quigley, M. Influence of arbuscular mycorrhiza on growth and reproductive response of plants under water deficit: A meta-analysis. *Mycorrhiza* **2014**, *24*, 109–119. [[CrossRef](#)] [[PubMed](#)]
110. Smith, S.E.; Read, D.J. *Mycorrhizal Symbiosis*; Academic Press: Cambridge, MA, USA, 2010.
111. Liu, C.; Guo, X.; Wu, X.; Dai, F.; Wu, Q. The Comprehensive Effects of Rhizophagus intraradices and P on Root System Architecture and P Transportation in *Citrus limon* L.. *Agriculture* **2022**, *12*, 317. [[CrossRef](#)]
112. Paszkowski, U.; Boller, T. The growth defect of *lrl1*, a maize mutant lacking lateral roots, can be complemented by symbiotic fungi or high phosphate nutrition. *Planta* **2002**, *214*, 584–590. [[CrossRef](#)] [[PubMed](#)]
113. Quiroga, G.; Erice, G.; Aroca, R.; Chaumont, F.; Ruiz-Lozano, J.M. Enhanced Drought Stress Tolerance by the Arbuscular Mycorrhizal Symbiosis in a Drought-Sensitive Maize Cultivar Is Related to a Broader and Differential Regulation of Host Plant Aquaporins than in a Drought-Tolerant Cultivar. *Front. Plant Sci.* **2017**, *8*, 1056. [[CrossRef](#)] [[PubMed](#)]
114. Das, K.; Roychoudhury, A. Reactive oxygen species (ROS) and response of antioxidants as ROS-scavengers during environmental stress in plants. *Front. Environ. Sci.* **2014**, *2*, 53. [[CrossRef](#)]
115. Verbruggen, E.; Struyf, E.; Vicca, S. Can arbuscular mycorrhizal fungi speed up carbon sequestration by enhanced weathering? *Plants People Planet* **2021**, *3*, 445–453. [[CrossRef](#)]
116. Dimkpa, C.; Weinand, T.; Asch, F. Plant-rhizobacteria interactions alleviate abiotic stress conditions. *Plant Cell Environ.* **2009**, *32*, 1682–1694. [[CrossRef](#)]
117. Liu, X.; Duan, B.; Li, L.; Chen, G.; Su-Zhou, C.; Li, Y.; Merkeryan, H.; Liu, W. ACC Deaminase-producing PGPRs Improve Drought Stress Tolerance in Grapevine (*Vitis vinifera* L.). *Front. Plant Sci.* **2021**, *12*, 1849. [[CrossRef](#)]
118. Costa, O.Y.A.; Raaijmakers, J.M.; Kuramae, E.E. Microbial Extracellular Polymeric Substances: Ecological Function and Impact on Soil Aggregation. *Front. Microbiol.* **2018**, *9*, 1636. [[CrossRef](#)]

119. Sher, Y.; Baker, N.R.; Herman, D.; Fossum, C.; Hale, L.; Zhang, X.; Nuccio, E.; Saha, M.; Zhou, J.; Pett-Ridge, J. Microbial extracellular polysaccharide production and aggregate stability controlled by switchgrass (*Panicum virgatum*) root biomass and soil water potential. *Soil Biol. Biochem.* **2020**, *143*, 107742. [[CrossRef](#)]
120. Liu, X.M.; Zhang, H. The effects of bacterial volatile emissions on plant abiotic stress tolerance. *Front. Plant Sci.* **2015**, *6*, 774. [[CrossRef](#)]
121. Chen, Y.; Gozzi, K.; Yan, F.; Chai, Y. Acetic acid acts as a volatile signal to stimulate bacterial biofilm formation. *MBio* **2015**, *6*, e00392-15. [[CrossRef](#)]
122. Cho, S.M.; Kang, B.R.; Han, S.H.; Anderson, A.J.; Park, J.-Y.; Lee, Y.-H.; Cho, B.H.; Yang, K.-Y.; Ryu, C.-M.; Kim, Y.C. 2R, 3R-butanediol, a bacterial volatile produced by *Pseudomonas chlororaphis* O6, is involved in induction of systemic tolerance to drought in *Arabidopsis thaliana*. *Mol. Plant-Microbe Interact.* **2008**, *21*, 1067–1075. [[CrossRef](#)]
123. Staudinger, C.; Mehmeti-Tershani, V.; Gil-Quintana, E.; Gonzalez, E.M.; Hofhansl, F.; Bachmann, G.; Wienkoop, S. Evidence for a rhizobia-induced drought stress response strategy in *Medicago truncatula*. *J. Proteom.* **2016**, *136*, 202–213. [[CrossRef](#)]
124. Vurukonda, S.S.; Vardharajula, S.; Shrivastava, M.; Sk, Z.A. Enhancement of drought stress tolerance in crops by plant growth promoting rhizobacteria. *Microbiol. Res.* **2016**, *184*, 13–24. [[CrossRef](#)]
125. Paul, M.J.; Primavesi, L.F.; Jhurreea, D.; Zhang, Y. Trehalose metabolism and signaling. *Annu. Rev. Plant Biol.* **2008**, *59*, 417–441. [[CrossRef](#)]
126. Dubois, M.; Van den Broeck, L.; Inzé, D. The pivotal role of ethylene in plant growth. *Trends Plant Sci.* **2018**, *23*, 311–323. [[CrossRef](#)]
127. Glick, B.R. Modulation of plant ethylene levels by the bacterial enzyme ACC deaminase. *FEMS Microbiol. Lett.* **2005**, *251*, 1–7. [[CrossRef](#)]
128. Mansour, E.; Mahgoub, H.A.M.; Mahgoub, S.A.; El-Sobky, E.E.A.; Abdul-Hamid, M.I.; Kamara, M.M.; AbuQamar, S.F.; El-Tarably, K.A.; Desoky, E.M. Enhancement of drought tolerance in diverse *Vicia faba* cultivars by inoculation with plant growth-promoting rhizobacteria under newly reclaimed soil conditions. *Sci. Rep.* **2021**, *11*, 24142. [[CrossRef](#)]
129. Sutuliené, R.; Rageliené, L.; Samuoliené, G.; Brazaityté, A.; Urbutis, M.; Miliauskiené, J. The Response of Antioxidant System of Drought-Stressed Green Pea (*Pisum sativum L.*) Affected by Watering and Foliar Spray with Silica Nanoparticles. *Horticulturae* **2022**, *8*, 35. [[CrossRef](#)]
130. Mekureyaw, M.F.; Pandey, C.; Hennessy, R.C.; Nicolaisen, M.H.; Liu, F.; Nybroe, O.; Roitsch, T. The cytokinin-producing plant beneficial bacterium *Pseudomonas fluorescens* G20-18 primes tomato (*Solanum lycopersicum*) for enhanced drought stress responses. *J. Plant Physiol.* **2022**, *270*, 153629. [[CrossRef](#)]
131. Yu, P.; He, X.; Baer, M.; Beirinckx, S.; Tian, T.; Moya, Y.A.T.; Zhang, X.; Deichmann, M.; Frey, F.P.; Bresgen, V.; et al. Plant flavones enrich rhizosphere Oxalobacteraceae to improve maize performance under nitrogen deprivation. *Nat. Plants* **2021**, *7*, 481–499. [[CrossRef](#)]
132. Oburger, E.; Jones, D.L. Sampling root exudates—Mission impossible? *Rhizosphere* **2018**, *6*, 116–133. [[CrossRef](#)]
133. Hallett, P.D.; Marin, M.; Bending, G.D.; George, T.S.; Collins, C.D.; Otten, W. Building soil sustainability from root-soil interface traits. *Trends Plant Sci.* **2022**, *27*, 688–698. [[CrossRef](#)] [[PubMed](#)]
134. Lucas, M.; Schlüter, S.; Vogel, H.J.; Vetterlein, D. Roots compact the surrounding soil depending on the structures they encounter. *Sci. Rep.* **2019**, *9*, 16236. [[CrossRef](#)] [[PubMed](#)]
135. Kautz, T. Research on subsoil biopores and their functions in organically managed soils: A review. *Renew. Agric. Food Syst.* **2015**, *30*, 318–327. [[CrossRef](#)]
136. Zhang, Z.; Peng, X. Bio-tillage: A new perspective for sustainable agriculture. *Soil Tillage Res.* **2021**, *206*, 104844. [[CrossRef](#)]
137. Marcacci, K.M.; Warren, J.M.; Perfect, E.; Labbé, J.L. Influence of living grass Roots and endophytic fungal hyphae on soil hydraulic properties. *Rhizosphere* **2022**, *22*, 100510. [[CrossRef](#)]
138. Galloway, A.F.; Pedersen, M.J.; Merry, B.; Marcus, S.E.; Blacker, J.; Benning, L.G.; Field, K.J.; Knox, J.P. Xyloglucan is released by plants and promotes soil particle aggregation. *New Phytol.* **2018**, *217*, 1128–1136. [[CrossRef](#)]
139. Rillig, M.C.; Mumme, D.L. Mycorrhizas and soil structure. *New Phytol.* **2006**, *171*, 41–53. [[CrossRef](#)]
140. Pausch, J.; Kuzyakov, Y. Carbon input by roots into the soil: Quantification of rhizodeposition from root to ecosystem scale. *Glob. Chang. Biol.* **2018**, *24*, 1–12. [[CrossRef](#)]
141. Cabal, C.; Martínez-García, R.; Aguilar, A.d.C.; Valladares, F.; Pacala, S.W. The exploitative segregation of plant roots. *Science* **2020**, *370*, 1197–1199. [[CrossRef](#)]
142. Poirier, V.; Roumet, C.; Munson, A.D. The root of the matter: Linking root traits and soil organic matter stabilization processes. *Soil Biol. Biochem.* **2018**, *120*, 246–259. [[CrossRef](#)]
143. Lange, M.; Eisenhauer, N.; Sierra, C.A.; Bessler, H.; Engels, C.; Griffiths, R.I.; Mellado-Vázquez, P.G.; Malik, A.A.; Roy, J.; Scheu, S.; et al. Plant diversity increases soil microbial activity and soil carbon storage. *Nat. Commun.* **2015**, *6*, 6707. [[CrossRef](#)]
144. Sasse, J.; Martinolia, E.; Northen, T. Feed Your Friends: Do Plant Exudates Shape the Root Microbiome? *Trends Plant Sci.* **2018**, *23*, 25–41. [[CrossRef](#)]
145. Hinsinger, P.; Gobran, G.R.; Gregory, P.J.; Wenzel, W.W. Rhizosphere geometry and heterogeneity arising from root-mediated physical and chemical processes. *New Phytol.* **2005**, *168*, 293–303. [[CrossRef](#)]
146. Goss, M.; Miller, M.; Bailey, L.; Grant, C. Root growth and distribution in relation to nutrient availability and uptake. *Eur. J. Agron.* **1993**, *2*, 57–67. [[CrossRef](#)]

147. Sanderman, J.; Hengl, T.; Fiske, G.J. Soil carbon debt of 12,000 years of human land use. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 9575–9580. [[CrossRef](#)]
148. Carminati, A.; Javaux, M. Soil Rather Than Xylem Vulnerability Controls Stomatal Response to Drought. *Trends Plant Sci.* **2020**, *25*, 868–880. [[CrossRef](#)]
149. Wasson, A.P.; Richards, R.A.; Chatrath, R.; Misra, S.C.; Prasad, S.V.S.; Rebetzke, G.J.; Kirkegaard, J.A.; Christopher, J.; Watt, M. Traits and selection strategies to improve root systems and water uptake in water-limited wheat crops. *J. Exp. Bot.* **2012**, *63*, 3485–3498. [[CrossRef](#)]

Chapter 3. AutoML: A Next-Generation Tool for Mining Multivariate Plant Traits

Addressing the Analytical Complexity Bottleneck Through Automated Discovery

Having established the profound biological complexity of drought adaptation in Chapter 2, this chapter introduces the first component of our computational response: Automated Machine Learning (AutoML). This chapter directly addresses the Analytical Complexity Bottleneck (Section 1.3.1) by demonstrating how automated approaches can handle the scale and complexity of modern biological datasets that overwhelm traditional statistical methods.

While manual statistical analysis requires deep expertise and iterative optimisation, AutoML democratises advanced data science by automating the entire machine learning pipeline—from feature selection to model deployment. This chapter positions AutoML within the broader methodological philosophy of this thesis: that unlocking the next frontier of biological insight requires a new generation of intelligent, accessible, and scalable analytical tools.

By highlighting AutoML's capacity to automatically navigate the complexities of multivariate plant data, this chapter establishes the foundational principles of automated, scalable, and reproducible analysis. It advocates for a paradigm shift from subjective manual analysis to objective machine-led discovery, thereby forming the conceptual groundwork upon which the bespoke solutions to the subsequent phenotyping, mechanistic, and integration bottlenecks are built.

Statement of contributions of joint authorship

Mirza Shoaib conducted the literature review, generated the figures, and led the manuscript writing and response to reviewers. Neelesh Sharma, Lars Kotthoff, Marius Lindauer and Surya Kant contributed to manuscript editing.

Statement from the co-author confirming the PhD candidate's contributions to the authorship

As a co-author of the manuscript:

M. Shoaib, N. Sharma, L. Kotthoff, M. Lindauer, and S. Kant, "AutoML: advanced tool for mining multivariate plant traits," Trends in Plant Science, 2023, doi: 10.1016/j.tplants.2023.09.008.

I confirm that Mirza Shoaib made the following key contributions:

- Conducted the literature review
- Led the writing of the manuscript
- Generated all figures
- Prepared the response to reviewers



Dr Surya Kant

29 July 2025

AutoML: advanced tool for mining multivariate plant traits

Mirza Shoaib,^{1,2} Neelesh Sharma,^{1,2} Lars Kotthoff,³ Marius Lindauer,⁴ and Surya Kant ^{1,2,5,*}

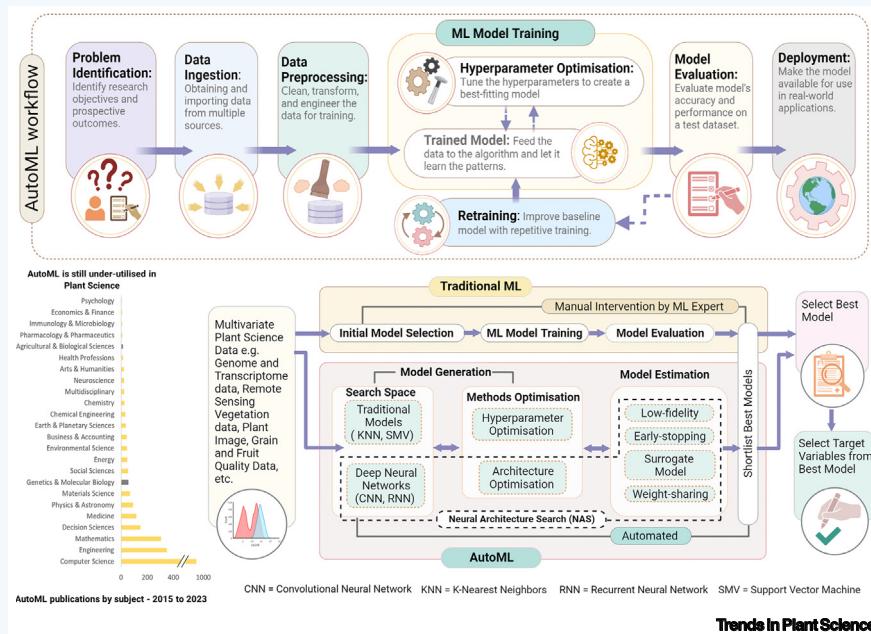
¹Agriculture Victoria, Grains Innovation Park, 110 Natimuk Rd, Horsham, Victoria 3400, Australia

²School of Applied Systems Biology, La Trobe University, Bundoora, Victoria 3083, Australia

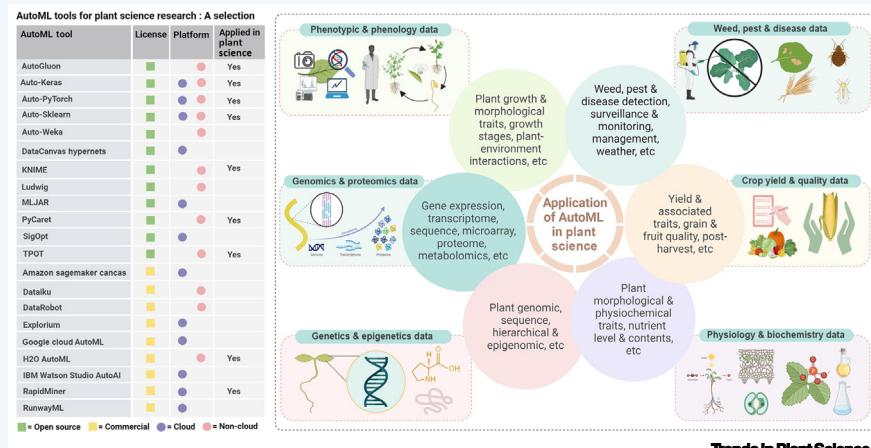
³Department of Electrical Engineering and Computer Science, School of Computing, University of Wyoming, Laramie, WY 82071, USA

⁴Institute of Artificial Intelligence (LUHAI), Leibniz University Hannover, Appelstr. 9A, 30167 Hannover, Germany

⁵Agriculture Victoria, AgriBio, Centre for AgriBioscience, 5 Ring Road, Bundoora, Victoria 3083, Australia



Automated machine learning (AutoML) is an automated version of machine learning (ML) that has the potential to become an integral part of plant science research for dealing with large and complex multivariate datasets. AutoML is rapidly advancing and has the potential to revolutionise the development and deployment of ML as it fully automates the ML process, from selecting the appropriate model and optimising its hyperparameters, to handling data preparation in some instances. This automation saves time and effort, enhances model quality, and increases its usability for novice users such as plant scientists.



AutoML excels in extracting meaningful features from diverse datasets in plant science domains, enhancing understanding of precision agriculture, crop breeding, and disease forecasting; estimating abiotic stresses, molecular genetics, and proteomics analysis; and yield prediction. As AutoML tools evolve, they will become even more powerful and user-friendly, accelerating innovation in plant science.

ADVANTAGES:

AutoML automates feature engineering and scalability, making it easier to extract meaningful information from complex multivariate and large biological datasets.

Efficiency is increased by automating repetitive tasks, reducing errors, and automatically selecting the best features and models for improved accuracy.

Reproducibility is improved by generating a detailed report of the model-building and evaluation process.

ML is democratised by providing a drag-and-drop interface that eliminates the need for programming skills, making ML accessible and usable for beginners.

It integrates seamlessly with existing tools and pipelines, streamlining research workflows and simplifying model building.

CHALLENGES:

Interpreting AutoML, particularly deep learning models, can occasionally present challenges in gaining insights from data, like traditional ML. Therefore, consideration of interpretable AutoML tools may be required in some projects.

In some cases, AutoML training and implementation can be challenged by limited or missing data in plant science.

It lacks the same customisation or flexibility as coding-based approaches, necessitating careful evaluation and domain-specific knowledge.

Most AutoML platforms are versatile and can be used for plant science research. Some are already being used in this field, but others still need to be tested, deployed, and evaluated in the specific plant science domain.

*Correspondence:
surya.kant@agriculture.vic.gov.au (S. Kant).

Trends in Plant Science | Technology of the Month

Declaration of interests

No interests are declared.

Literature

1. Waring, J. et al. (2020) Automated machine learning: review of the state-of-the-art and opportunities for healthcare. *Artif. Intell. Med.* 104, 101822
2. Escalante, H.J. (2021) Automated machine learning—a brief review at the end of the early years. In *Automated Design of Machine Learning and Search Algorithms* (Pillay, N. and Qu, R., eds), pp. 11–28, Springer
3. Hutter, F. et al. (2019) *Automated Machine Learning: Methods, Systems, Challenges*, Springer
4. Korot, E. et al. (2021) Code-free deep learning for multi-modality medical image classification. *Nat. Mach. Intell.* 3, 288–298
5. He, X. et al. (2021) AutoML: a survey of the state-of-the-art. *Knowl. Based Syst.* 212, 106622
6. Nagarajah, T. and Poravi, G. (2019) A review on automated machine learning (AutoML) systems. In *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*, IEEE
7. Truong, A. et al. (2019) Towards automated machine learning: evaluation and comparison of AutoML approaches and tools. In *2019 IEEE 31st International Conference on Tools for Artificial Intelligence (ICTAI)*, IEEE
8. Ferreira, L. et al. (2021) A comparison of AutoML tools for machine learning, deep learning and XGBoost. In *2021 International Joint Conference on Neural Networks*, IEEE
9. Doke, A. and Gaikwad, M. (2021) Survey on automated machine learning (AutoML) and meta learning. In *2021 12th International Conference on Computing and Networking Technologies*, IEEE
10. Wever, M. et al. (2021) AutoML for multi-label classification: overview and empirical evaluation. *IEEE Trans. Pattern Anal. Mach. Intell.* 43, 3037–3054
11. Das, S. and Cakmak, U.M. (2018) *Hands-on Automated Machine Learning: A Beginner's Guide to Building Automated Machine Learning Systems Using AutoML and Python*, Packt Publishing

Chapter 4. Seeing the Unseen: Algorithmic Root Traits (ART)

The First Breakthrough: Solving the Phenotyping Bottleneck Through Algorithmic Vision

While AutoML (Chapter 3) demonstrates the power of automated analysis for complex datasets, this chapter targets the specific phenotyping bottleneck identified in Section 1.3.2—the fundamental limitation of human visual perception in capturing biologically relevant architectural patterns. Traditional trait extraction relies on predefined, visually-derived metrics that are blind to the subtle, non-obvious features that distinguish adaptive strategies.

This chapter introduces the first major methodological contribution of this research: the Algorithmic Root Trait (ART) extraction method—a bespoke computational solution that moves beyond the limits of human perception to achieve true algorithmic discovery.

This work directly addresses Research Question 1: Can the computational extraction of latent architectural traits from root images provide a novel source of information that is complementary to traditional metrics, and can their integration lead to more robust and insightful predictive models?

The results provide a decisive and quantifiable yes. While models based on ARTs alone substantially outperformed traditional methods (96.3% vs. 85.6% accuracy), the framework's true strength lies not in replacement but in synergy. When ARTs and traditional traits were combined, the model achieved the highest classification accuracy of 97.4%, confirming that the two approaches capture complementary facets of root biology. This was achieved with remarkable efficiency, as ARTs delivered 5.8 \times higher information density.

This chapter demonstrates a fundamental principle: for complex biological structures, computation can see what we cannot. Having solved the phenotyping bottleneck by providing a novel and complementary source of information, the stage is set to address the remaining challenges of mechanistic understanding and systems integration.

Statement of contributions of joint authorship

Mirza Shoaib designed and conducted the experiment. Surya Kant guided in refining the experiments. Mirza Shoaib wrote the manuscript and developed the code. Christopher Topp and Matthew J. Hayden contributed to the conceptual ideas of the manuscript. Adam M. Dimech created Fig. S2 and provided guidance in optimising image capture. All reviewed and edited the manuscript.

Statement from the co-author confirming the PhD candidate's contributions to the authorship

As a co-author of the manuscript:

M. Shoaib, A. M. Dimech, S. J. Rochfort, C. Topp, M. J. Hayden, and S. Kant, "Seeing the Unseen: A Novel Approach to Extract Latent Plant Root Traits from Digital Images," Plant Phenomics, p. 100088, 2025/07/09/ 2025, doi: <https://doi.org/10.1016/j.plaphe.2025.100088>.

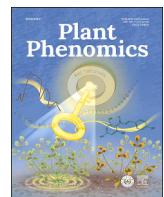
I confirm that Mirza Shoaib has made the following contributions,

- Conducted the literature review
- Designed and conducted the experiments
- Performed all image and data analyses
- Generated all figures, except Figure S2
- Led the writing of the manuscript
- Prepared the response to reviewers



Dr Surya Kant

29 July 2025



Research Article

Seeing the unseen: A novel approach to extract latent plant root traits from digital images



Mirza Shoaib ^{a,b,*}, Adam M. Dimech ^c, Simone J. Rochfort ^{b,c}, Christopher Topp ^d, Matthew J. Hayden ^{b,c}, Surya Kant ^{e,**}

^a Agriculture Victoria, Grains Innovation Park, 110 Natimuk Road, Horsham, Victoria, 3400, Australia

^b School of Applied Systems Biology, La Trobe University, Bundoora, Victoria, 3083, Australia

^c Agriculture Victoria, AgriBio, Centre for AgriBioscience, 5 Ring Road, Bundoora, Victoria, 3083, Australia

^d Donald Danforth Plant Science Center, Saint Louis, Missouri, USA

^e Department of Ecological, Plant and Animal Science, School of Agriculture, Biomedicine & Environment, La Trobe University, Bundoora, Victoria, 3083, Australia

ARTICLE INFO

ABSTRACT

Keywords:

Algorithmic Root Traits (ART)

Root phenotyping

Drought tolerance

Wheat

Latent trait

Machine learning

Root image analysis

A novel approach, the Algorithmic Root Trait (ART) extraction method, identifies and quantifies computationally-derived plant root traits, revealing latent patterns related to dense root clusters in digital images. Using an ensemble of multiple unsupervised machine learning algorithms and a custom algorithm, 27 ARTs were extracted reflecting dense root cluster size and spatial location. These ARTs were then used independently and in combination with Traditional Root Traits (TRTs) to classify wheat genotypes differing in drought tolerance.

ART-based models outperformed TRT-only models in drought classification (e.g., 96.3 % vs. 85.6 % accuracy). Combining ARTs and TRTs further improved accuracy to 97.4 %. Notably, 4 selected ARTs matched the performance of all 23 TRTs, offering $5.8 \times$ higher information density (0.213 vs. 0.037 accuracy/feature). This superiority reflects the ability of ARTs to capture richer, more complex architectural information, evidenced by higher internal variability (35.59 ± 11.41 vs. 28.91 ± 14.28 for TRTs) and distinct data structures in multivariate analyses; PERMANOVA confirmed that ARTs and TRTs provide complementary insights.

Validated through experiments in controlled environments and field conditions with wheat drought-tolerant and susceptible genotypes, ART offers a scalable, customisable toolset for high-throughput phenotyping of plant roots. By bridging conventional, visually derived traits with autonomous computational analyses, this method broadens root phenotyping pipelines and underscores the value of harnessing sensor data that transcends human perception. ART thus emerges as a promising framework for revealing hidden features in plant imaging, with broader applications across plant science to deepen our understanding of crop adaptation and resilience.

1. Introduction

While automated data analysis tools effectively extract valuable root traits [1–5], these typically rely on predefined geometric or morphological characteristics derived from human visual perception. Consequently, subtle or complex patterns within image data, which do not align neatly with established definitions, may remain undetected [6–10]. The identification of latent, complex image-derived traits offers considerable potential for advancing plant phenotyping, particularly for roots, which are inherently concealed and interact dynamically with

their environment [11–13]. In this manuscript, “latent traits” are defined as algorithmically identified, interpretable features—such as root cluster location, size, and density—derived from pixel distributions rather than human visual assessments or abstract neural network representations.

A comprehensive understanding of plant roots, essential for productivity and climate adaptation, is required to facilitate the breeding of climate-resilient crops [14–16]. Critical root attributes such as architecture, structure, anatomy, plasticity, exudation, microbial interactions, and hydraulics significantly influence plant performance

* Corresponding author. Agriculture Victoria, Grains Innovation Park, 110 Natimuk Road, Horsham, Victoria, 3400, Australia.

** Corresponding author.

E-mail addresses: shoaib.mirza@agriculture.vic.gov.au (M. Shoaib), S.Kant@latrobe.edu.au (S. Kant).

[17–23]. Among these, architectural and structural traits have been primary research targets due to their measurability through phenotyping techniques like shovelingomics, rhizotrons, minirhizotrons, and soil coring [24–27]. Software tools such as RhizoVision, WinRHIZO, RootNav, DIRT, and ImageJ typically extract these traits, termed Traditional Root Traits (TRTs) herein, from segmented root images [1–5].

Despite their utility, TRTs may inadequately capture root complexity especially under stress conditions like drought [28]. Roots navigate unique environmental challenges—darkness, variable moisture, mechanical impedance—distinct from aerial tissues, complicating their phenotypic characterisation [11–13].

Alternative approaches have emerged to overcome these limitations. Berrigan et al. [29] adapted SLEAP, a deep-learning animal motion capture system, for root landmark detection without segmentation. Ubbens et al. [30] proposed Latent Space Phenotyping using convolutional neural networks for automatic trait detection from images. Peoples et al. [31] employed Earth Mover's Distance for insights into root system architecture beyond conventional metrics, while Li et al. [32] introduced persistent-homology-based topological methods to quantify morphological variations.

Building on these advances, we introduce the Algorithmic Root Trait (ART) extraction method (see Table S1 for a comparative overview with existing analytical methods). ART systematically integrates multiple unsupervised machine learning (ML) algorithms with a custom-developed algorithm (hereafter, the "Custom" algorithm) to identify and characterise dense root clusters. The novelty of ART lies in harnessing the complementary strengths of these algorithms, uncovering spatial patterns and architectural features often missed by traditional methods.

Unlike methods emphasising global distribution comparisons [31] or abstract topological metrics [32], ART provides directly interpretable localised geometric and density-based traits (e.g., cluster size, centroid coordinates). Additionally, ART autonomously detects emergent root clustering patterns without relying on predefined assumptions or supervised landmark detection approaches [29].

ART's inherent flexibility enables seamless integration with TRTs, facilitating comprehensive analyses of root architecture. Leveraging an algorithmically unconstrained approach, ART extracts latent traits from digital RGB sensor data, significantly enhancing the discovery of nuanced, physiologically relevant traits critical for understanding plant resilience.

To demonstrate ART's practical value, experiments under controlled environment and field conditions were conducted to classify wheat genotypes based on drought tolerance. Wheat, central to global food security [33], faces increasing drought stress, predicted to surpass pathogen threats in its impact on crop productivity [34]. Root density, reflecting plant resource allocation under drought, varies with soil depth and between tolerant and susceptible genotypes [35–37]. Given environmental influences and subjective assessments, accurately identifying dense root areas is challenging.

ART's unsupervised clustering methods were applied to objectively analyse RGB sensor-derived data, identifying dense root clusters to generate informative traits. Integrating these ART-derived traits with TRTs within a supervised ML pipeline improved genotype classification accuracy. Further, combining ARTs with TRTs substantially enhanced predictive performance. Evaluations across controlled environment and variable field conditions confirmed ART's robust performance, underscoring its potential applicability across diverse research and agricultural contexts.

2. Materials and methods

2.1. Experiment conditions and genotypes

A glasshouse rhizotron study was conducted to compare three previously characterised drought-tolerant (DT_1, DT_2, DT_3) and three

drought-susceptible (DS_1, DS_2, DS_3) wheat (*Triticum aestivum* L.) genotypes (Table S2) [38–41]. The experiment utilised 48 rhizotrons (594 cm height × 42 cm width × 6 cm depth; volume: 6.5 L) filled with fertilised coco peat (Table S3), arranged in a completely randomised design with four replicates per genotype per treatment (control and drought). The drought treatment was maintained at 30 % field capacity (FC), while the control was maintained at 60 % FC. Rhizotrons were angled at 45° [38] and wrapped with opaque material to exclude light and mitigate temperature fluctuations. Water was applied by weight to maintain FC.

Three seeds were sown per rhizotron and thinned to a single seedling post-germination. A pilot study determined optimal seed orientation (brush up, crease facing rhizotron wall) for consistent root trait expression (Fig. S1) [39]. Key physiological traits (leaf relative water content, stomatal conductance, tiller number) were recorded at 45 and 49 days after sowing (DAS) to validate genotype drought responses (see 'Drought Tolerance Characterisation' in Result section).

A complementary field experiment was conducted at the research farm of Agriculture Victoria, Horsham (36°43'57"S, 142°06'01"E), planted two contrasting genotypes (DT_1 and DS_3) in a randomised complete block design. Three replicate plots (5 m × 1 m) per genotype (6 plots total) were established, with five rows per plot.

2.2. Imaging setup

A custom glasshouse imaging setup captured rhizotron images (Fig. S2). It featured a blackout curtain, a fixed camera-rhizotron frame, matt black paint, a light blue background (HEX: #0089b6), and four diffused studio lights. Images (n = 960) were taken at 20 time-points (day 5–49) with a Sony camera a7R using Sony FE 35 mm F1.4 GM Lens (Sony Corporation, Tokyo, Japan).

For imaging at field experiment, four weeks after sowing, three minirhizotron tubes (65 mm diameter, 1.2-m long) were installed at 45° within each plot (18 tubes total). These minirhizotron tubes were scanned using an In-Situ Root Imager ICI-600 (CID-Bio-Science, Camas, Washington, United States). Each minirhizotron was scanned three times from top to bottom, resulting in 54 images per imaging day (3 scans × 18 tubes). Seven time-series scans were captured at specific intervals: 147, 151, 157, 163, 170, 177, and 184 DAS, producing a total of 378 images. The field and glasshouse experimental setups are depicted in Fig. S3.

2.3. Image processing pipeline

Python 3.10.2 [40] and R-4.3.2 [41] were used for image pre-processing and analysis. Glasshouse images underwent automated pre-processing: QR code decoding (pyzbar library) for renaming and sample tracking, date suffixing, EXIF-based rotation, and cropping into shoot and root regions (3410 x 5075 pixels for roots) based on fixed coordinates. Field minirhizotron images (three per tube) were merged and oriented, yielding 126 composite images from the original 378.

Images were segmented using RootPainter [42]. To handle distinct imaging environments (glasshouse vs. field), two separate segmentation models were developed. The model was first trained for glasshouse images. This optimised glasshouse model then served as a pre-trained base for fine-tuning with field-specific annotations, reducing training time and adapting to differences in growing media and root morphology. For both datasets, representative images were manually annotated, and models were trained iteratively until the Dice score neared 1 (Fig. S4). The best-performing glasshouse model segmented all glasshouse images, and the fine-tuned field-specific model segmented all field images.

2.4. Traditional Root Traits (TRTs) extraction

TRTs were extracted from segmented binary images (glasshouse and

field) using RhizoVision Explorer [10] (Fig. 1B) with the "broken roots" setting (Table S4) for consistency. This yielded 23 TRTs (listed in Table S5). Field ($n = 126$) and glasshouse ($n = 960$) TRT datasets were merged (total $n = 1086$).

2.5. Defining Algorithmic Root Trait (ART)

TRTs typically quantify pre-defined global or skeleton-based geometric/morphological features [1–3,5]. While valuable, they may not fully capture complex spatial organisation and density variations within the root system. The Algorithmic Root Trait (ART) extraction method, introduced here, employs an ensemble of unsupervised ML algorithms and the Custom algorithm (Table 1) applied post-segmentation.

ARTs are designed to uncover structural and spatial patterns by analysing segmented root pixel distributions, focusing on identifying and quantifying characteristics of algorithmically "discovered dense root clusters". In this context, "latent" refers to cluster characteristics (precise location, boundaries, density) that emerge from data-driven algorithmic analysis rather than being predefined by human visual rules. While the extracted metrics (size, location) are interpretable, the discovery and delineation of clusters themselves are entirely algorithmic.

The novelty of ART lies in its systematic, multi-algorithmic approach to identify dense root clusters and extract interpretable traits: cluster size via [AlgorithmName]_density_points and spatial coordinates via [AlgorithmName]_centre_x/y (Table S5). Unlike methods emphasising global distributions, abstract topological metrics, or supervised landmark detection (Table S1), ART provides localised, interpretable traits from autonomously detected patterns. This approach offers a distinct perspective on root architecture, complementary to global TRT metrics (Fig. 3), as demonstrated by correlation and feature importance analyses (Figs. S5–S8, S14).

2.6. ART extraction

ARTs were extracted from binary segmented images using nine algorithms (eight established unsupervised ML methods plus one custom procedure; Table 1) to identify and characterise the largest dense root cluster in each image. The clustering approach was selected for two key reasons:

Biological rationale: Dense root clusters represent adaptive strategies where plants concentrate biomass in resource-rich zones to optimise water acquisition under drought [35–37,43]. This spatial organisation

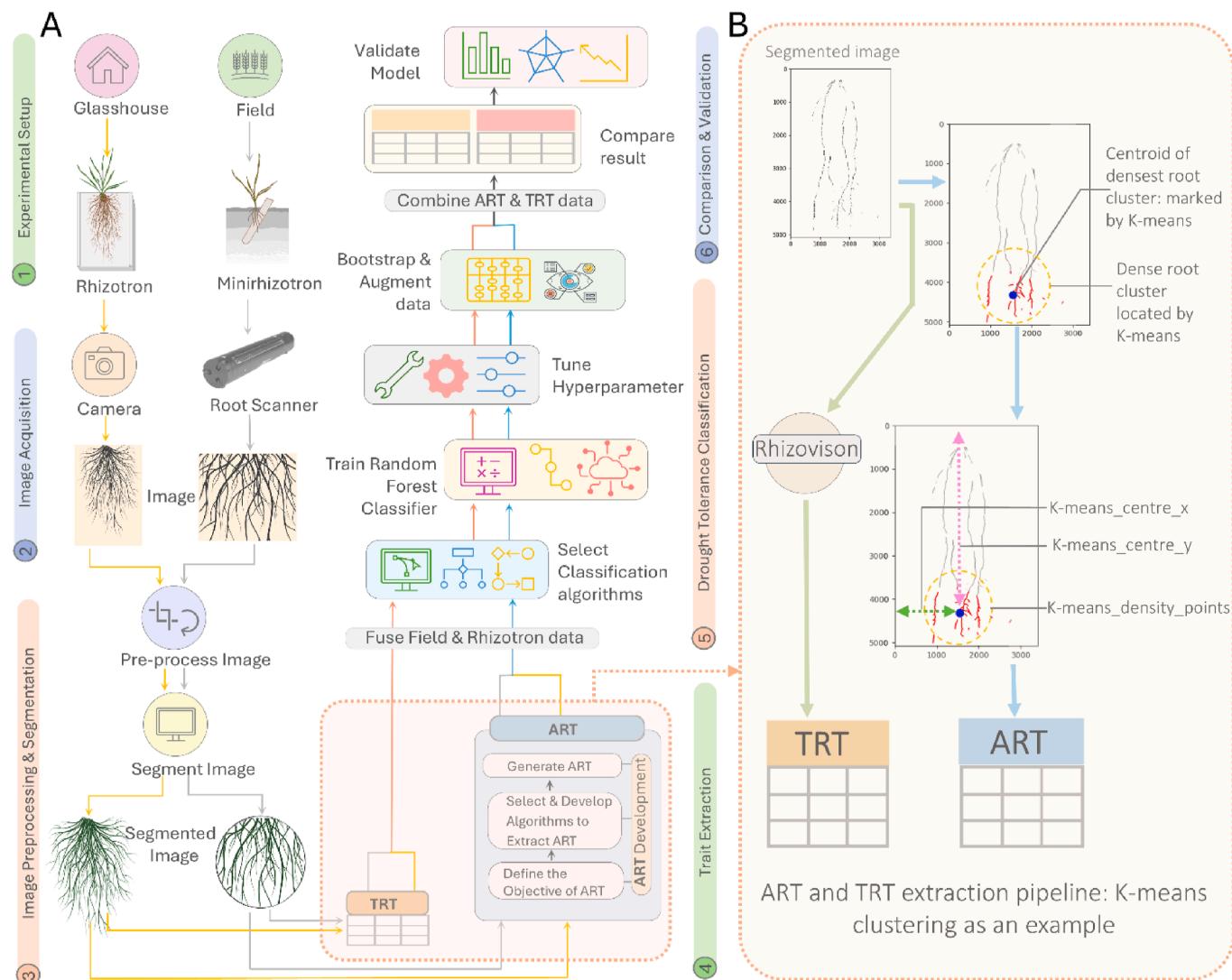


Fig. 1. Experimental setup and workflow. (A) Schematic workflow of ART and TRT extraction and drought tolerance classification. (1) Experimental setup (2) Image Acquisition: Custom imaging setup for glasshouse and C-600 root imager for field. (3) Image preprocessing and segmentation. (4) ART and TRT extraction. (5) Drought tolerance classification with the Random Forest algorithm. (6) Classification result comparison and validation. (B) ART and TRT extraction pipeline: TRT extraction using Rhizovision and ART extraction using K-means Clustering Algorithm.

Table 1

List of unsupervised ML and custom algorithms used in ART extraction.

| Algorithm Category | Algorithm Name | Specific Function | Cluster Shape | Reference |
|-------------------------|--|--|-------------------------------|---|
| Partitioning | Gaussian Mixture Models (GMM) | Soft partitioning | Globular/spherical | Pedregosa et al. [47] |
| Density-based | K-means | Hard partitioning | Globular/spherical | Pedregosa et al. [47] |
| | Density-Based Spatial Clustering of Applications with Noise (DBSCAN) | Density-based clustering | Elongated/irregular | Pedregosa et al. [47] and Ester et al. [48] |
| | Mean-shift | Mode-seeking clustering | Elongated/irregular | Pedregosa et al. [47] and Comaniciu and Meer [49] |
| Hierarchical | Ordering Points to Identify Cluster Structure (OPTICS) | Density-based clustering (extension of DBSCAN) | Elongated/irregular | Pedregosa et al. [47] and Ankerst et al. [50] |
| | Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) | Hierarchical density-based clustering | Elongated/irregular | McInnes et al. [51] |
| Graph-based | Fuzzy C-means (FCM) | Fuzzy clustering | Adaptable to different shapes | Bezdek et al. [52] |
| Superpixel Segmentation | Simple linear iterative clustering (SLIC) | Superpixel generation based on colour and spatial similarity | Irregular | Achanta et al. [53] |
| Custom | Custom | Density estimation and selection of the largest cluster | Density-based | Custom |

reflects phenotypic plasticity consistent with established drought adaptation mechanisms, including strategic positioning for enhanced water uptake efficiency [44–46].

Methodological rationale: Clustering algorithms directly quantify biologically meaningful spatial patterns—cluster size (biomass concentration), centroid coordinates (spatial positioning), and density distributions—that correspond to functional root architecture components known to influence drought tolerance [43,46]. The ensemble approach captures diverse clustering morphologies, from compact aggregations to elongated formations (Table 1), reflecting the architectural diversity of adaptive root responses.

After feasibility screening, algorithms representing diverse clustering approaches were selected: DBSCAN ($\text{epsilon} = 5$, $\text{min_samples} = 10$), Gaussian Mixture Models (GMM; $n_{\text{components}} = 2$), K-means ($n_{\text{clusters}} = 5$), HDBSCAN ($\text{min_cluster_size} = 10$), MeanShift (quantile = 0.1), OPTICS ($\text{min_samples} = 10$), SLIC ($n_{\text{segments}} = 50$), and Fuzzy C-means (FCM). Parameters (Table 1) were optimised in preliminary experiments.

The Custom algorithm identifies dense regions as follows: (1) All root pixels (black pixels) are identified. (2) If > 200 root pixels exist, the 200 most locally dense (pixels surrounded by other root pixels) are selected. (3) A Gaussian kernel density estimate (KDE) weights these selected pixels. (4) Density-weighted coordinates are clustered using K-means (up to 5 clusters, depending on point availability). (5) The largest cluster (by point count) is chosen. (6) Traits extracted are Density_points (cluster size), Density_centre_x, and Density_centre_y (cluster geometric centre).

Each of the nine algorithms extracts three traits from its identified largest cluster: [AlgorithmName]_density_points (size), [AlgorithmName]_centre_x (horizontal position), and [AlgorithmName]_centre_y (vertical position), yielding 27 ARTs (Table S5). These are maintained as distinct features because each algorithm applies different mathematical principles to identify cluster boundaries from the same image data. For example, K-means partitions pixels into globular clusters while HDBSCAN identifies density-connected regions of arbitrary shape, resulting in different cluster definitions and centroid locations from identical inputs (Table 1, Fig. 3). Algorithm-specific naming (e.g., FCM_centre_x vs. DBSCAN_centre_x) preserves information about which computational approach identified each cluster, enabling determination of the most biologically informative clustering perspectives. This methodological diversity ensures comprehensive characterisation of root spatial organisation patterns. All code and data are available for download at <https://github.com/shoibms/ART>.

2.7. Comparative dataset analysis and drought phenotyping

2.7.1. Comparing TRT, ART and combined datasets

Internal variability (mean Euclidean distance), correlation matrices, Euclidean Distance Matrix Heatmaps, Principal Component Analysis (PCA), t-distributed Stochastic Neighbor Embedding (t-SNE), and Multidimensional Scaling (MDS) were used to compare ART, TRT, and combined datasets. PERMANOVA was used to test statistical differences between ART and TRT datasets.

2.7.2. Drought response characterisation

To validate genotype drought responses, key physiological indicators were measured. Leaf Relative Water Content (RWC) was determined at 45 DAS using ~ 100 mg samples from young fully expanded leaves. Fresh weight was recorded, followed by overnight hydration in deionised water at 4°C for turgid weight, and oven-drying at 80°C for 48 h for dry weight. RWC was calculated as $[(\text{FW}-\text{DW})/(\text{TW}-\text{DW})] \times 100\%$ [54]. Stomatal conductance (g_s) was measured on the adaxial surface of the second fully expanded leaf using a portable photosynthesis system (LCpro-SD) between 10:00am-2:30pm to minimise diurnal effects. Measurements were taken at 45 DAS (g_s .1) and 49 DAS (g_s .2) under standardised conditions (PAR 1500 $\mu\text{mol m}^{-2} \text{s}^{-1}$, ambient CO₂). Tiller number was counted at 45 DAS. Effect sizes of these traits were analysed to support their use in genotype classification.

2.7.2.1. Calculation of selected Trait's effects. Effect sizes (Cliff's Delta) between tolerant and susceptible groups were assessed for g_s .1, g_s .2, RWC, and Tiller_no. Normality (Shapiro-Wilk) determined ANOVA (Tukey's HSD for normal RWC) or Kruskal-Wallis (Dunn's test for non-normal g_s .1, g_s .2, Tiller_no).

2.7.2.2. Genotype ranking methodology. To quantitatively assess drought tolerance, genotypes were ranked based on their physiological responses using three complementary approaches. This methodology involved calculating two primary drought response metrics for each genotype (i) and physiological trait (j), from which three final ranking scores were derived.

Variable Definitions:

i = genotype index (from 1 to 6)

j = physiological trait index (g_s .1, g_s .2, RWC, tiller number)

w_j = trait weight (2 for stomatal conductance traits, 1 for RWC and tiller number)

Primary Drought Response Metrics:

To ensure that both metrics were robust to potential outliers in the data, 10 % trimmed means were used for all calculations.

1. Robust Difference (RD):

$$RD_{ij} = |M_{ij,control} - M_{ij,drought}| \quad (1)$$

where $M_{ij,control}$ – $M_{ij,drought}$ are the 10 % trimmed means for trait j of genotype i under control and drought conditions, respectively.

2. Robust Stress Susceptibility Index:

The Stress Susceptibility Index [55] quantifies the relative reduction in performance under stress conditions. A robust version was calculated using trimmed means to maintain consistency with the RD metric:

$$SSI_{ij} = 1 - \left(\frac{M_{ij,drought}}{M_{ij,control}} \right) \quad (2)$$

Final Ranking Scores.

For each metric (RD and SSI), genotypes are ranked from 1 (most tolerant) to 6 (most susceptible) for each of the four traits. The following weighted ranking scores are calculated:

3. RANK₁ (Robust Difference-based Score):

$$RANK_{1,i} = \sum_j [Rank(RD_{ij}) \times w_j] \quad (3)$$

4. RANK₂ (SSI-based Score):

$$RANK_{2,i} = \sum_j [Rank(SSI_{ij}) \times w_j] \quad (4)$$

5. RANK₃ (SSI-based Score):

$$RANK_{3,i} = RANK_{1,i} + RANK_{2,i} \quad (5)$$

Note: The summation \sum_j runs over all four traits. Lower RANK scores indicate higher drought tolerance.

2.7.2.3. Clustering approach. Five unsupervised clustering algorithms (K-means, Agglomerative, Gaussian Mixture, Spectral, and Birch) classified wheat genotypes into “Tolerant” ($T = 1$) and “Susceptible” ($S = 0$) groups under both watering treatments (T0 and T1) using physiological traits (g_s_1, g_s_2, RWC, and tiller number). Data were treatment-segregated and standardised using StandardScaler before generating two clusters per algorithm. Cluster quality was assessed using internal metrics (Silhouette Score, Dunn Index, Calinski-Harabasz Index, and Davies-Bouldin Index) to evaluate separation, and external metrics (Adjusted Rand Index, Normalised Mutual Information, and Fowlkes-Mallows Score) to compare algorithmic assignments against known drought tolerance classifications.

2.7.3. Classification framework

2.7.3.1. Algorithm selection and evaluation. Eight classifiers (AdaBoost, CatBoost, Gradient Boosting, K-Nearest Neighbors, LightGBM, Logistic Regression, Random Forest and Support Vector Machine (SVM)) were evaluated for drought tolerance classification using ART and TRT datasets. Multiple metrics were evaluated including accuracy, precision, recall, F1 score, ROC AUC, and specificity, with priority given to precision and ROC AUC.

2.7.3.2. Development and robust validation of classification models. A multi-stage validation strategy ensured robust evaluation and mitigated overfitting (Fig. S11):

- An independent validation set ($n = 218$: 36 field, 182 glasshouse samples) was set aside at the beginning of the study and kept completely separate from all model development processes to provide an unbiased assessment for final model validation.
- Augmented Development Dataset: Remaining original samples ($n = 868$) were augmented to create a development dataset ($n = 1736$). Augmentation involved bootstrapping original samples with replacement and adding controlled Gaussian noise ($\sigma = 0.02$) to numerical features; thus, the development dataset comprised both original and these newly generated bootstrapped/noised samples.
- Internal Split: This development dataset was split (stratified by drought tolerance) into an internal training set (80 %, $n = 1388$) and an internal test set (20 %, $n = 348$). No data leakage occurred between development and independent validation sets.

The Random Forest classifier was built using a pipeline comprising three sequential stages: (1) data standardisation with StandardScaler, (2) feature selection using SelectFromModel with a Random Forest estimator, and (3) final classification with the main Random Forest model.

Hyperparameter tuning was conducted exclusively on the internal training set ($n = 1388$) using Grid Search CV with 10-fold cross-validation, optimising for precision. This cross-validation showed excellent stability: mean CV% across metrics was 1.95 % (Fig. S12, Table S8). Precision, the optimisation target, had a mean of 0.904 (CV% = 1.876 %, 95 % CI [0.892, 0.916]), indicating robustness against overfitting from augmentation.

The optimal hyperparameters trained the final model on the entire internal training set ($n = 1388$). Performance was evaluated sequentially on the internal test set ($n = 348$) and then on the independent, un-augmented validation set ($n = 218$).

2.7.4. Feature importance and biological interpretation

To assess the biological relevance of ARTs, Pearson correlation analysis was conducted between ART and TRT features (significance at $p < 0.05$). The relative contribution of features to drought tolerance prediction was evaluated using SHAP (SHapley Additive exPlanations) analysis on the Random Forest model trained with the complete feature set. SHAP values, measuring each feature's contribution while accounting for interactions [56], were averaged to determine overall importance (Fig. S16). Based on these results and the algorithmic logic of ART extraction, a conceptual model was developed linking key ARTs to established drought adaptation mechanisms (Fig. S15). This framework helps translate algorithmic outputs into biologically meaningful root architectural strategies. Trait stability was assessed by calculating coefficients of variation for ART and TRT features across experimental conditions (Figs. S9 and S13).

3. Results

3.1. Drought Tolerance Characterisation

3.1.1. Genotype ranking and validation

The genotype ranking based on physiological responses to drought (RANK₁, RANK₂, RANK₃) consistently ranked genotypes (DT₁, DT₂, DT₃) as drought tolerant (i.e., higher ranks) than susceptible genotypes (DS₁, DS₂, DS₃). Minor rank swaps occurred between DS₂ and DS₃, but both consistently ranked as drought-susceptible, confirming their established classifications in the literature [57–60] (Fig. 2A).

3.1.2. Physiological responses to drought

Significant differences (ANOVA/Kruskal-Wallis, $p < 0.0001$) were observed in stomatal conductance (g_s_1, g_s_2), relative water content (RWC), and tiller number (Tiller_no) between drought-tolerant and

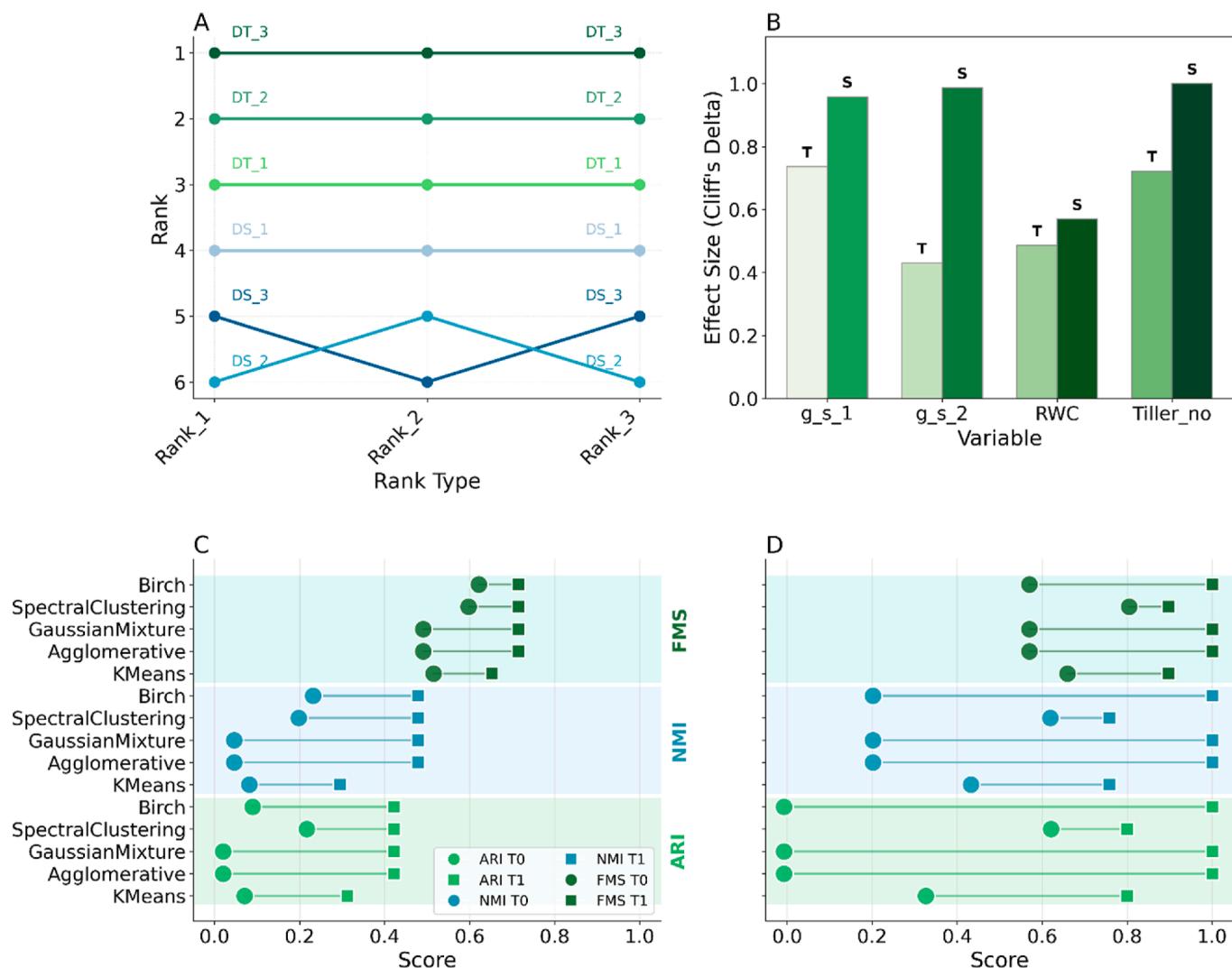


Fig. 2. Drought tolerance characterisation and validation of wheat genotypes. (A) A bump chart illustrates the ranking of drought-tolerant (DT, green hues) and drought-susceptible (DS, blue hues) genotypes based on three physiological ranking methods. DT genotypes consistently rank higher (1–3) than DS genotypes. (B) Bar chart showing the effect size (Cliff's Delta) of drought on key physiological traits. The greater effect on susceptible (S) genotypes compared to tolerant (T) genotypes confirms their higher sensitivity to stress. (C, D) Dumbbell plots comparing external validation scores for five unsupervised clustering algorithms. In these plots, scores under control (T0) conditions are represented by circles, and scores under drought (T1) conditions are represented by squares, as defined in the legend of panel (C). Each line connects the T0 (circle) and T1 (square) scores for a specific algorithm on a given metric (Adjusted Rand Index [ARI], Normalised Mutual Information [NMI], and Fowlkes-Mallows Score [FMS]). The general rightward shift from T0 to T1 indicates that the algorithms achieved better separation of tolerant and susceptible genotypes under drought stress. (D) The separation between tolerant and susceptible groups is further enhanced under drought (T1) after fine-tuning the validation to account for the specific response of genotype DT_3.

susceptible genotypes under stress. Cliff's Delta effect sizes indicated larger negative impacts of drought on the susceptible group for all measured physiological traits, highlighting these traits as robust indicators of drought response (Fig. 2B).

3.1.3. Clustering genotype for drought tolerance

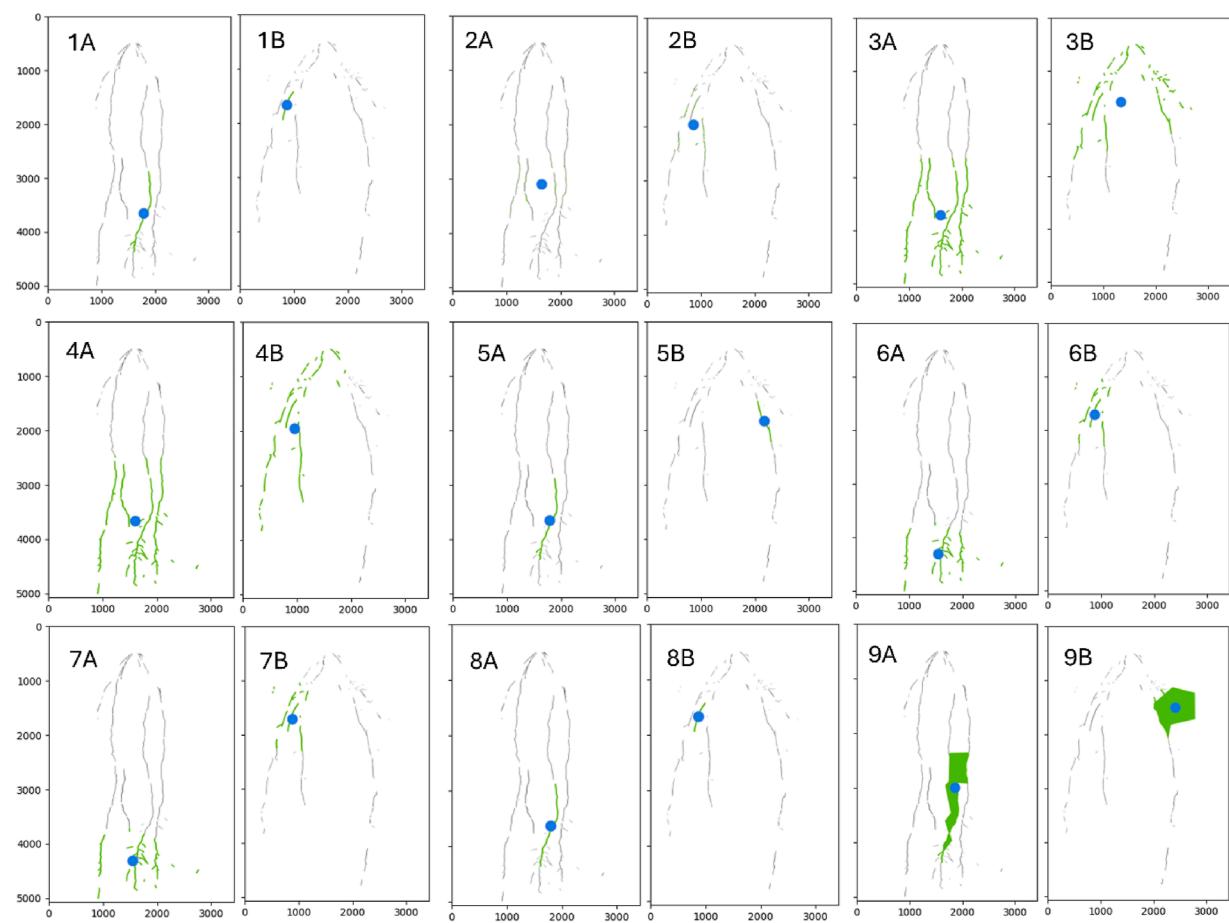
Unsupervised clustering of genotypes based on physiological traits yielded moderate separation between susceptible and tolerant groups under control conditions (T0), with mean internal validation metrics such as Silhouette Score ~ 0.45 , Dunn Index ~ 0.53 , Calinski-Harabasz Index ~ 21.8 , and Davies-Bouldin Index ~ 0.89 (see Table S6 for validation metrics). Cluster separation improved under drought conditions (T1), with mean Silhouette Score increasing to ~ 0.50 , Dunn Index to ~ 1.00 , Calinski-Harabasz Index to ~ 30.4 , and Davies-Bouldin Index decreasing to ~ 0.65 (Table S6). External validation metrics (Adjusted Rand Index, Normalised Mutual Information, Fowlkes-Mallows Score) also showed improved cluster accuracy under T1 (Fig. 2C). This

separation became more pronounced when accounting for the specific response of genotype DT_3, suggesting some intra-group variability within the drought-tolerant classification (Fig. 2D).

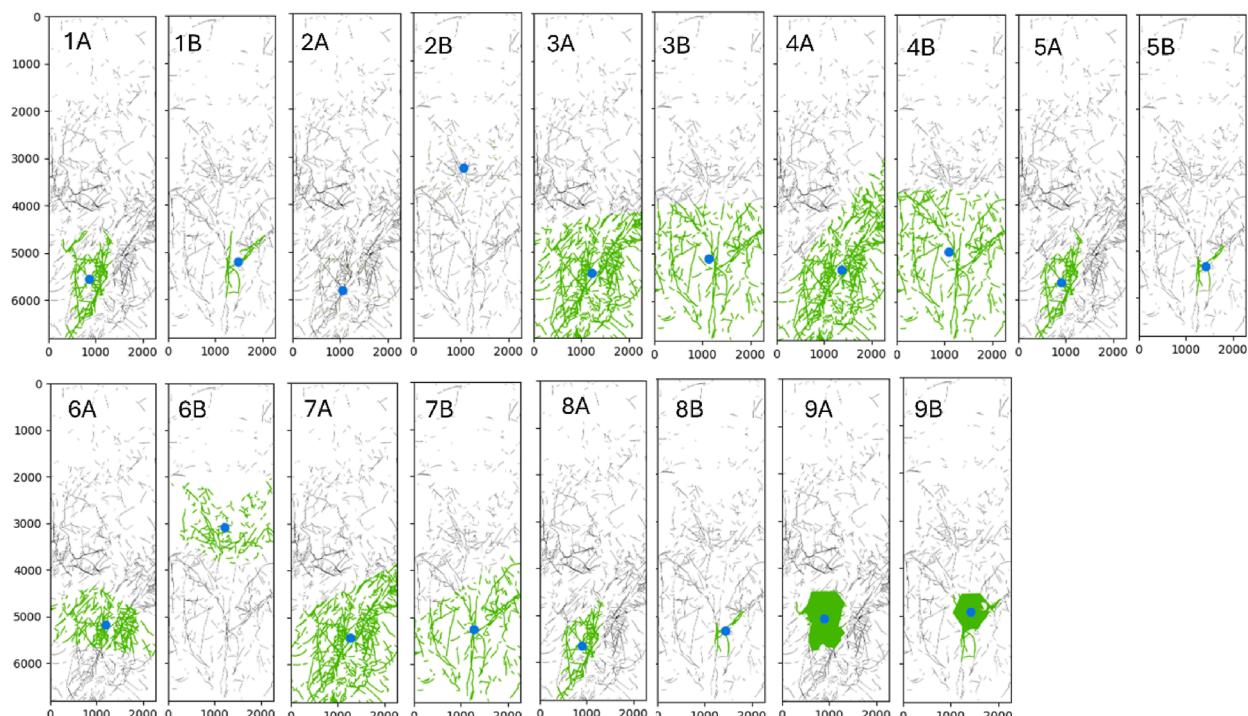
3.2. Algorithmic Root Trait (ART) extraction results

3.2.1. ART extraction process and visualisation

The ART approach, utilising eight unsupervised ML algorithms and the Custom algorithm (Table 1), yielded 27 ARTs (Table S5). Each algorithm generated three ARTs from the largest identified dense root cluster in an image: [AlgorithmName]_density_points (cluster size in pixels), [AlgorithmName]_centre_x (horizontal centroid coordinate), and [AlgorithmName]_centre_y (vertical centroid coordinate). Fig. 3 visually demonstrates how each of the nine algorithms uniquely identifies these dense root clusters and their central locations within root images from both glasshouse (Panel 1, 41 DAS) and field (Panel 2, 147 DAS) experiments for genotypes DT_1 (tolerant) and DS_3 (susceptible).



Panel 1



Panel 2

(caption on next page)

Fig. 3. Visualisation of Algorithmic Root Traits (ART). Panel 1 displays results from glasshouse rhizotron imagery, while **Panel 2** exhibits field minirhizotron imagery. The figures highlight the largest dense root clusters (`_density_points`) in green, with the blue dots indicating the cluster centroids (`_centre_x` and `_centre_y`). Nine algorithms, detailed in [Table 1](#), are represented by numbered overlays, illustrating how each algorithm identifies and locates the largest dense root cluster. The algorithms include: 1. DBSCAN, 2. Custom, 3. FCM, 4. GMM, 5. HDBSCAN, 6. K-means, 7. Mean-shift, 8. OPTICS, and 9. SLIC. In each panel, (A) corresponds to the drought-tolerant genotype DT_1, and (B) corresponds to the drought-susceptible genotype DS_3. For example, Panel 1 (6A) shows the K-means algorithm applied to the drought-tolerant genotype DT_1, and Panel 1 (6B) shows the same algorithm applied to the drought-susceptible genotype DS_3. The green areas represent the largest dense root clusters, and the blue dots indicate their centers as identified by the respective algorithms.

3.3. Comparative analysis: ART vs. TRT

3.3.1. Feature space analysis and multivariate comparisons

The ART dataset exhibited higher internal variability, with a mean Euclidean distance among its features of 35.59 (SD: 11.41), compared to the TRT dataset's 28.91 (SD: 14.28). This suggests ARTs capture more complex and discriminative architectural patterns. Correlation Matrix Heatmaps ([Fig. 4A](#) and B) and Euclidean Distance Matrix Heatmaps ([Fig. 4C](#) and D) visually supported these findings, with ART features generally displaying greater dissimilarity (darker colours in Euclidean distance heatmap). The combined dataset had a mean Euclidean distance of 34.81 (SD: 11.90) ([Fig. 4E](#)). Principal Component Analysis (PCA) plots showed ART variables distributed more broadly across components ([Fig. 4F](#)), unlike TRT variables where variance was more concentrated in the first component ([Fig. 4G](#)). Multidimensional Scaling (MDS) plots mirrored these structural differences ([Fig. 4H](#)). A Permutational Multivariate Analysis of Variance (PERMANOVA) confirmed a significant difference between the ART and TRT datasets (Test Statistic: 5.578, p-value: 0.002).

3.4. Drought tolerance classification performance

3.4.1. Algorithm selection and performance comparison

Among the eight classification algorithms tested, CatBoost and Random Forest generally exhibited superior performance across TRT, ART, and the combined datasets, as shown by the grouped bar chart comparisons ([Fig. 5A–C](#)). The cumulative distribution scores for ROC AUC and precision confirmed that models trained on the combined dataset consistently outperformed those built with either ART or TRT alone ([Fig. 5D](#) and E). Based on its strong and stable performance, Random Forest was selected for all subsequent detailed modeling and analysis. The final tuned Random Forest model's excellent performance on the internal development data ("Model score") and its robust generalisation to the independent validation set ("Validation score") are illustrated in [Fig. 5F](#).

3.5. Model performance and validation

Performance was assessed on internal test sets derived from augmented development datasets.

Classification with TRTs: The optimised Random Forest model trained on the augmented TRT dataset (TRT_A_B) achieved an accuracy of 0.856, precision of 0.860, and ROC AUC of 0.927 on its internal test set ([Table S10](#)). The step-wise improvement from baseline to the final augmented model is visually represented in the bar charts of [Fig. 6A](#).

Classification with ARTs: The model trained on the augmented ART dataset (ART_A_B) yielded superior performance, with an accuracy of 0.963, precision of 0.963, and ROC AUC of 0.997 on its internal test set ([Table S10](#), [Fig. 6A](#)).

Classification with Combined Root Traits (TRT and ART): The Random Forest model trained on the combined, augmented dataset (Combine_A_B) demonstrated the highest performance on its internal test set, with an accuracy of 0.974, precision of 0.975, and ROC AUC of 0.998 ([Table S10](#)). A direct comparison in [Fig. 6B](#) clearly illustrates that the model using the combined dataset outperforms the models trained on TRT or ART alone across most metrics. Confusion matrices for this model on the internal test set showed high true positive and true negative rates ([Fig. 6C](#)).

3.5.1. Independent validation results

The optimal Random Forest model (trained on Combine_A_B) was evaluated on the completely independent, un-augmented validation set ($n = 218$, 50 variables). It achieved a robust performance: accuracy 0.91, precision of 0.93, recall of 0.89, F1 score of 0.91, and ROC AUC of 0.96 ([Fig. 6D](#) and E). Cumulative gains and lift curves further confirmed its effectiveness in distinguishing tolerant and susceptible genotypes ([Fig. 6F](#) and G).

3.6. Biological relevance and feature contribution

3.6.1. ART feature contribution and biological interpretation

3.6.1.1. Contribution of ART in drought tolerance classification. SHAP analysis revealed that ART features accounted for 73.3 % of the combined model's predictive power compared to only 26.7 % for TRTs ([Fig. S14](#)), despite representing just 54 % of total features. Algorithm-level permutation importance confirmed ART dominance, with FCM (0.019), OPTICS (0.013), and HDBSCAN (0.012) showing the strongest individual contributions to model performance ([Table S7](#)).

3.6.1.2. Biological interpretation and mechanism links to drought adaptation. Strong correlations between ART and TRT features ([Fig. 7C](#), [Fig. S6](#)) established the biological validity of algorithmically-derived traits. Notably, 66 out of 129 significant correlations (51 %) exceeded $|r| = 0.7$, demonstrating robust connections between computational and traditional measurements. Root mass clustering ARTs (FCM_density_points, GMM density_points, K-mean_density_points) showed exceptional correlations with TRT volume metrics ($r = 0.93\text{--}0.98$, $p < 0.05$), validating `_density_points` as precise quantifiers of localised root biomass. Vertical position traits (`_centre_y`) correlated positively with depth metrics ($r = 0.60\text{--}0.68$, $p < 0.05$), while density clustering traits (`_density_points`) showed significant negative correlations with diameter metrics ($r = -0.85$ to -0.88 , $p < 0.05$), indicating adaptive trade-offs between root biomass concentration and construction costs.

Feature importance analysis identified FCM_centre_x (SHAP: 0.071) and HDBSCAN_density_points (SHAP: 0.071) as the most influential predictors, with HDBSCAN_centre_y (SHAP: 0.024) also ranking highly ([Fig. S16](#), [Table S9](#)). Complementary analysis using feature-level permutation importance further highlighted the significance of features such as FCM_centre_x, DBSCAN_density_points, and HDBSCAN_density_points ([Fig. 6H](#)). These findings establish a clear mechanistic pathway: ARTs quantify specific architectural adaptations—deeper rooting (`_centre_y`), enhanced lateral exploration (`_centre_x`), and strategic biomass allocation (`_density_points`)—that enable efficient water acquisition during drought stress. This enhanced water uptake maintains favorable plant water status, supporting higher stomatal conductance and ultimately conferring superior drought tolerance. This causal framework explains why ART-based models excel at classifying genotypes initially grouped by physiological drought responses and aligns with established root ideotypes for water-limited environments [[43](#)] ([Fig. 7A](#)).

3.6.2. Model performance and stability assessment

3.6.2.1. Trait stability assessment. ART features demonstrated significantly superior stability across experimental conditions (glasshouse,

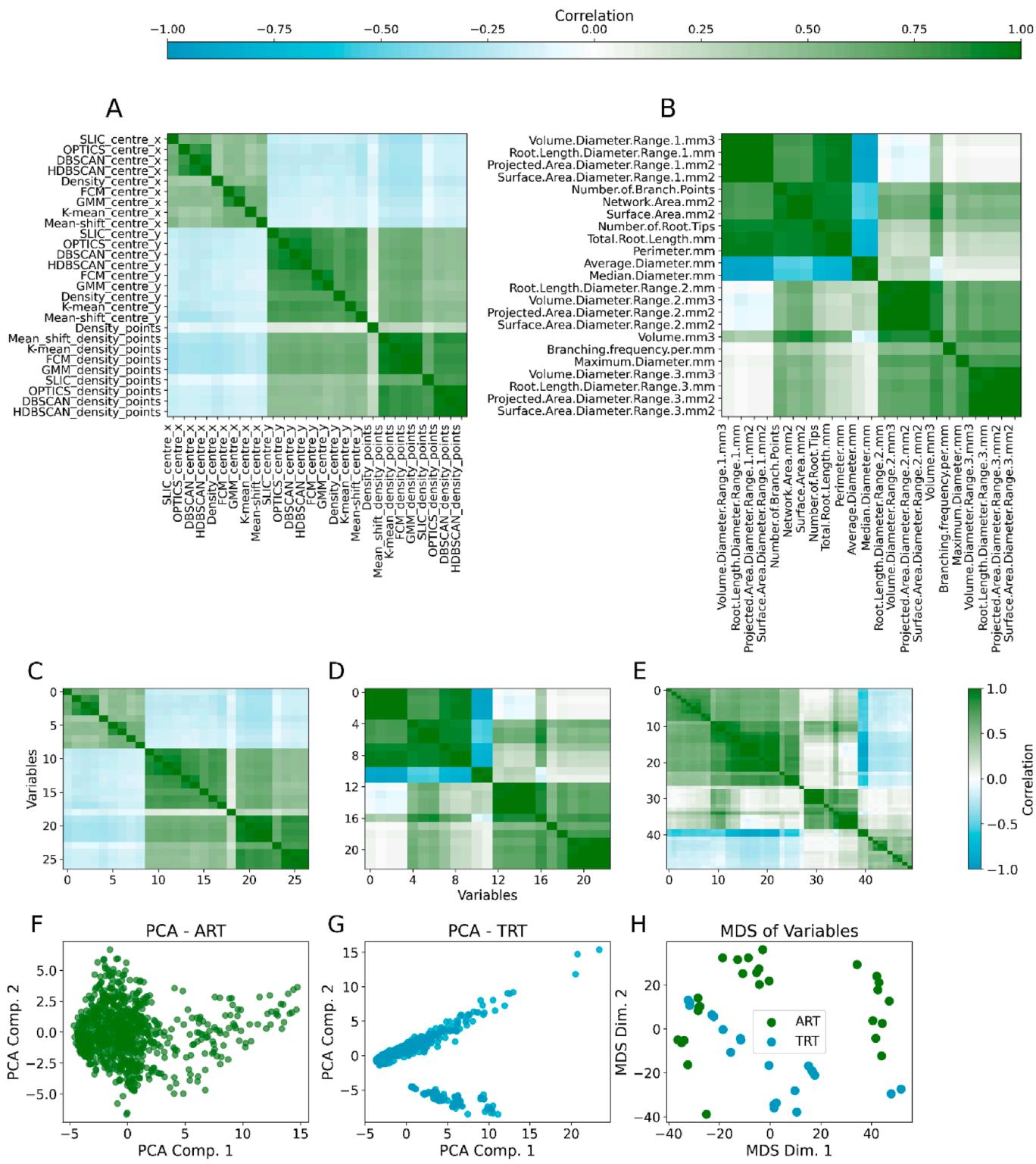


Fig. 4. Comparison between ART and TRT datasets. (A) Correlation matrix heatmap of ARTs. (B) Correlation matrix heatmap of TRTs. (C) Euclidean distance matrix heatmap of ARTs. (D) Euclidean distance matrix heatmap of TRTs. (E) Euclidean distance matrix heatmap for the Combined dataset. (F) PCA plot of ARTs. (G) PCA plot of TRTs. (H) MDS plot comparing ARTs and TRTs.

field, all) compared to TRTs, with a lower mean Coefficient of Variation (CV) (ART mean CV: 0.762 vs. TRT mean CV: 1.361, Welch's *t*-test $p = 0.0137$) (Fig. 7D, Figs. S9 and S13).

3.6.2.2. Model performance advantages and information density of ARTs.

The comparative analysis demonstrates that ARTs significantly

outperform TRTs for drought tolerance classification (Fig. 7B). Models using all 27 ARTs achieved 96.3 % accuracy and 99.7 % ROC AUC, substantially better than models using all 23 TRTs (85.6 % accuracy, 92.7 % ROC AUC)—a 10.7 percentage point improvement. Even just 4 selected ARTs matched the performance of all 23 TRTs, representing 5.8 \times higher information density (0.213 vs. 0.037 accuracy/feature)

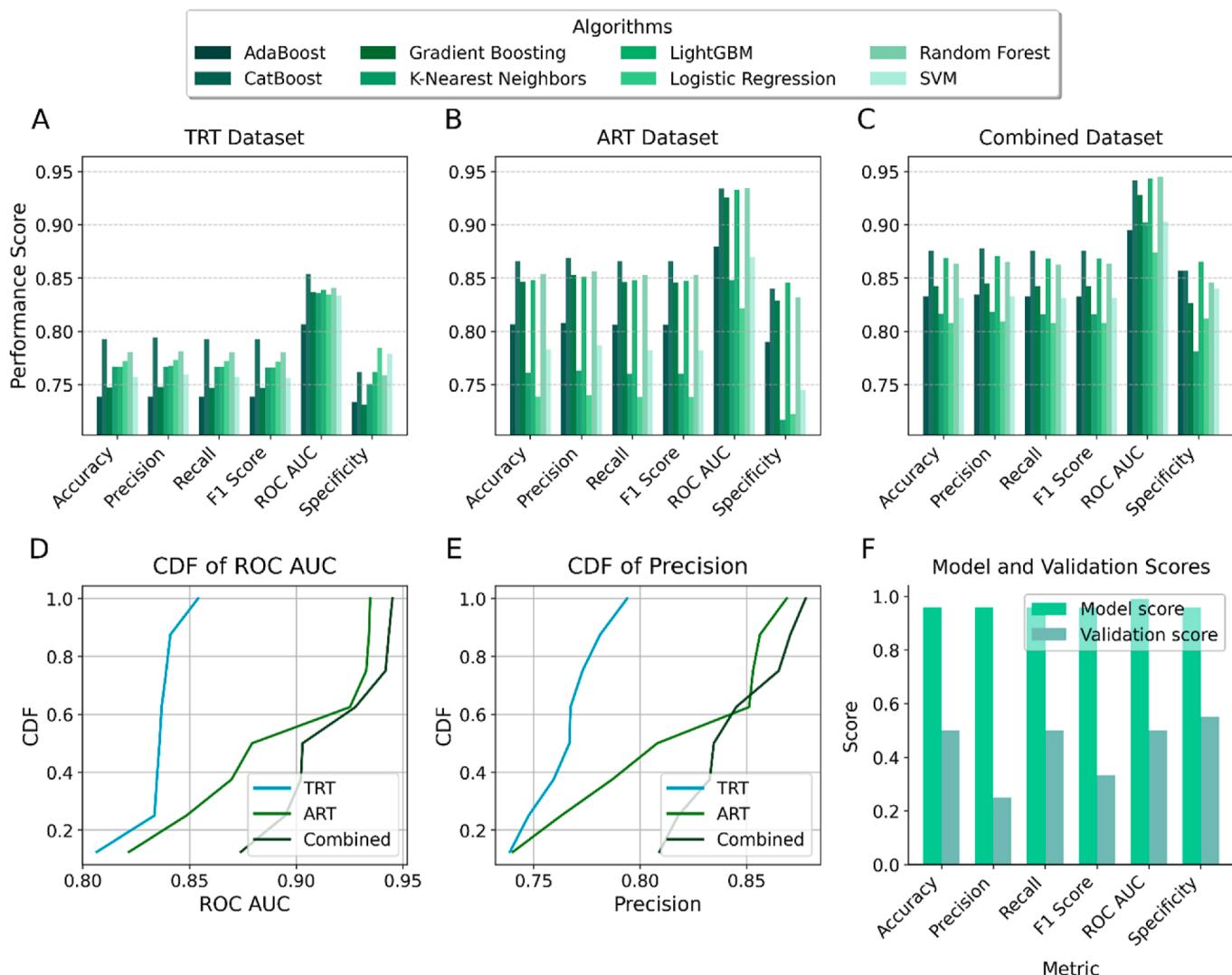


Fig. 5. Comparative performance of classification algorithms and final model validation.

(A–C) Grouped bar charts comparing the performance of eight classification algorithms across six key metrics on the (A) TRT, (B) ART, and (C) combined datasets. For each metric, Random Forest and CatBoost consistently rank among the top performers. (D–E) Cumulative Distribution Function (CDF) plots illustrating the distribution of (D) ROC AUC and (E) Precision scores, showing that models trained on the combined dataset consistently outperform those trained on ART or TRT alone. (F) Performance of the final selected Random Forest model, comparing its scores on the internal development data ("Model score") with its scores on the completely independent validation set ("Validation score"), demonstrating robust generalisation.

(Fig. 7E–Table S10). This efficiency has practical implications for high-throughput phenotyping, where extracting fewer, more informative traits can significantly accelerate screening. The superior performance of ARTs stems from their ability to capture spatial patterns of root organisation that global TRTs miss. ARTs also demonstrated significantly higher stability across environments (mean CV = 0.762 vs. 1.361 for TRTs, $p = 0.0137$), enhancing their reliability for field applications. When combined, ARTs and TRTs showed synergistic effects, with just 8 features (4 ARTs + 4 TRTs) achieving 89.9 % accuracy, further confirming the complementary nature of these trait types and their collective value for drought tolerance screening (Table S10).

3.6.2.3. Cross-validation stability results. The Random Forest model selected via 10-fold cross-validation on the augmented training data demonstrated excellent stability during the model selection process. The cross-validated precision (our optimisation target) averaged 0.904 (95 % CI: [0.892, 0.916]) with minimal variation across folds (CV% = 1.876 %). This stability was consistent across all key performance metrics, with coefficients of variation ranging from 1.242 % (ROC AUC) to 2.591

% (F1 Score), as detailed in Fig. S12 and Table S8. The final model, trained on the augmented training data, achieved excellent performance on the internal test set (accuracy: 0.974, precision: 0.975, F1 score: 0.973, ROC AUC: 0.998) and, importantly, when tested on a completely independent, un-augmented validation set, it achieved an accuracy of 0.913, precision: 0.933, F1 score: 0.911, and ROC AUC: 0.962, confirming its robustness and generalisability.

3.6.3. Comparative validation results

3.6.3.1. Baseline comparisons. To further contextualise the performance of ARTs and TRTs, a simpler baseline was established by training a Random Forest model using only the total pixel count from each segmented root mask as a single feature. This "Total Root Pixels" model, representing the overall scale of the root system, achieved an accuracy of 57.8 % and a ROC AUC of 61.7 % (Table S10). The substantially higher performance of models incorporating multiple TRTs (e.g., "All TRTs" model: 85.6 % accuracy) and particularly ARTs (e.g., "All ARTs" model: 96.3 % accuracy; "Selected 4 ARTs" model: 85.1 % accuracy)

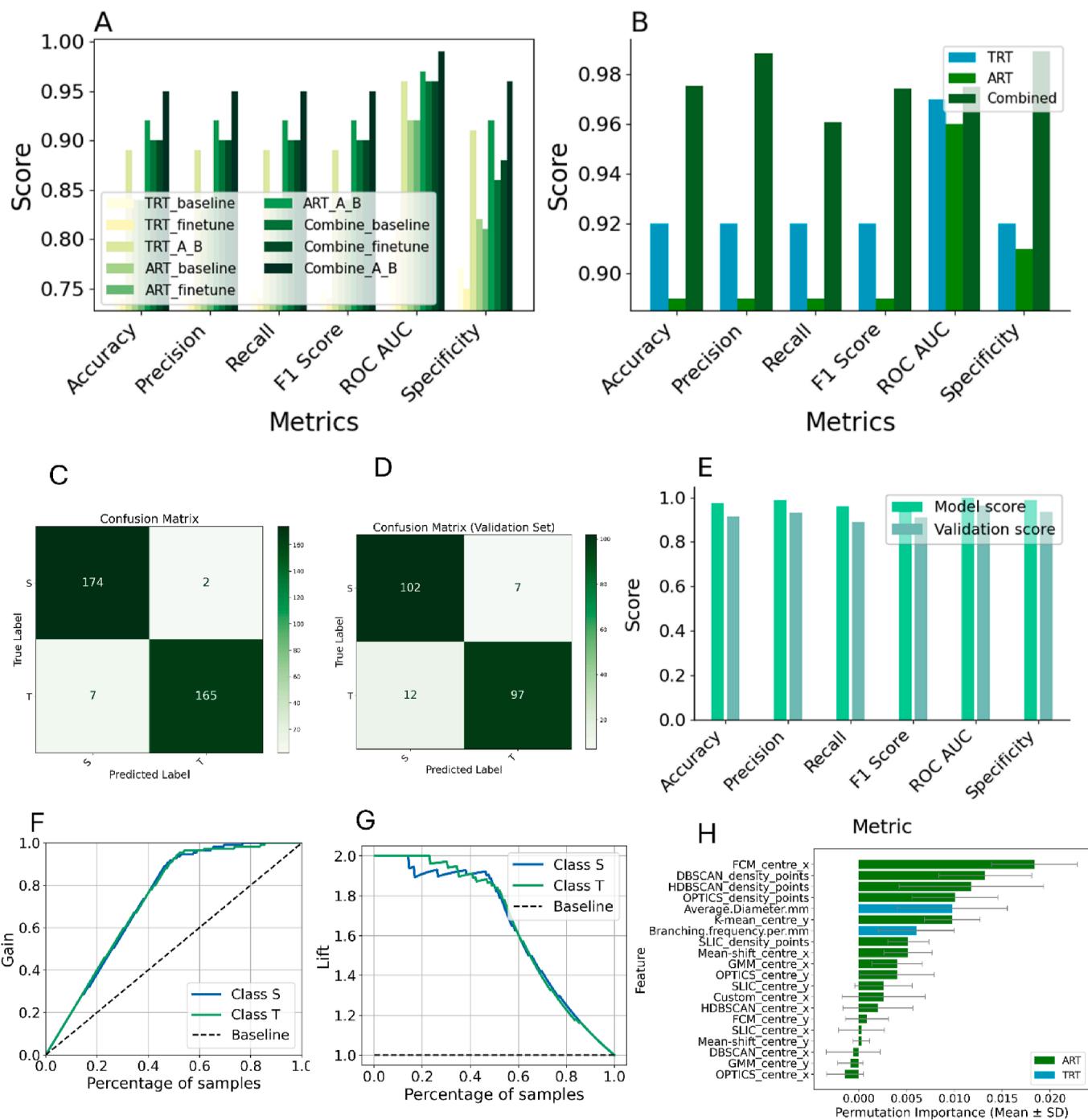


Fig. 6. Performance of Random Forest classifiers in classifying drought tolerance using different datasets.

(A) Grouped bar charts illustrating the step-wise improvement of the Random Forest model's performance. For each data type (TRT, ART, Combine), performance is shown for the baseline, fine-tuned, and augmented (A_B) model versions. (B) Bar chart comparison of the final, optimised models for the TRT, ART, and combined datasets, highlighting the superior performance of the model trained on the combined dataset. (C) Confusion matrix for the best-performing model on the internal test set. (D) Confusion matrix for the best-performing model on the independent validation set. (E) The best-performing model's metric scores during development ("Model score") and on the validation dataset ("Validation score"). (F) Cumulative gain curve for the validation dataset. (G) Lift curve for the validation dataset. (H) Permutation feature importance of the final model.

demonstrates that these more sophisticated trait sets capture critical architectural and distributional information for drought tolerance classification that extends significantly beyond simple root system size.

3.6.3.2. t-SNE comparison results and algorithm contribution. The superior ability of ART features to distinguish between drought-tolerant and susceptible genotypes was visually evident in t-SNE projections

(**Fig. S17**), which showed significantly better group separation compared to TRT features. Specifically, ART features showed a 109 % greater Silhouette Score (0.092 vs. 0.044), 34 % better Davies-Bouldin Score (3.301 vs. 5.004, lower is better), and 113 % greater Calinski-Harabasz Score (61.512 vs. 28.874) compared to TRT features. This pattern of improved separation was consistent across all dimensionality reduction techniques tested (t-SNE and PCA), confirming that ARTs

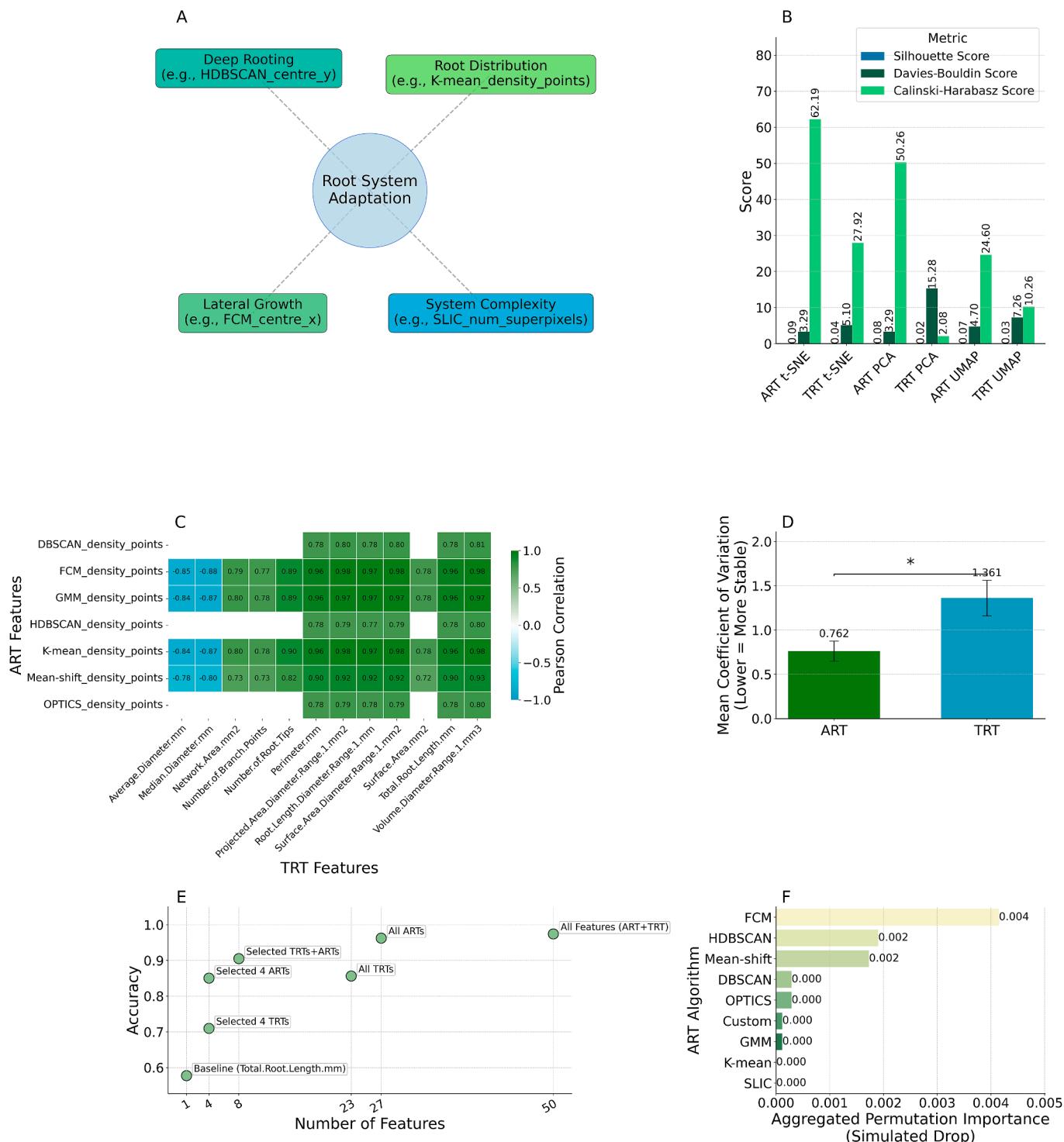


Fig. 7. Biological relevance and performance advantages of ART. (A) Drought adaptation mechanisms captured by ART features, showing connections to established physiological processes; (B) Performance metrics comparison between ART and TRT across multiple evaluation criteria; (C) Significant correlations between ART features and established TRT metrics ($|r| \geq 0.6$, $p < 0.05$); (D) Trait stability comparison showing significantly lower coefficient of variation for ART versus TRT across experimental conditions ($p = 0.0137$); (E) Model accuracy versus feature count, demonstrating the higher information density of ARTs compared to TRTs; (F) Algorithm ablation study results, showing the relative contributions of different algorithms to model performance, with FCM, HDBSCAN, and Mean-shift providing the greatest contributions.

capture distinct information relevant to drought tolerance classification.

The different cluster distributions between ART and TRT feature spaces suggest these approaches capture complementary information about root architecture, explaining why their combination yields superior

classification performance (97.4 % accuracy) compared to either feature set alone. Algorithm ablation study results show the relative contributions of different algorithms to model performance, with FCM, HDBSCAN, and Mean-shift providing the greatest contributions (Fig. 7F).

4. Discussion

This study introduces the ART extraction method, a novel computational framework that significantly advances the capacity to extract nuanced, physiologically relevant information from plant root images. By employing an ensemble of unsupervised ML algorithms and the Custom algorithm applied to segmented binary images, ART uncovers spatially-explicit architectural patterns, particularly the characteristics of dense root clusters. These traits, "latent" in the sense that their specific attributes (e.g., precise location, boundaries, density) emerge from algorithmic analysis rather than being predefined by direct visual rules, proved remarkably effective.

ART-based models substantially outperformed traditional approaches (96.3 % vs. 85.6 % accuracy) while achieving $5.8 \times$ higher information density, demonstrating the significant value of algorithmically-derived traits for drought tolerance classification. This enhanced efficiency and discriminative power has profound implications for high-throughput phenotyping, offering a pathway to accelerate genetic screening for complex stress resilience traits.

The superior predictive power of ARTs stems from their ability to quantify biologically meaningful root system adaptations to water-limited environments, phenomena well-documented in plant physiology [43,45,61,62]. The analyses suggest a clear mechanistic link between these algorithmically-derived traits and drought resilience, as captured by SHAP analysis where ARTs accounted for 73.3 % of the combined model's predictive power (Fig. S14). For instance, vertical distribution traits like HDBSCAN_centre_y (SHAP importance ~ 0.024), which correlated significantly with TRT depth metrics ($r = 0.60\text{--}0.68$, $p < 0.05$), directly quantify deeper rooting—a key component of Lynch's (2013) "Steep, Cheap, and Deep" ideotype. Similarly, lateral exploration traits such as FCM centre x (the most influential single feature, SHAP ~ 0.071), which showed strong negative correlations with TRT average diameter metrics ($r = -0.85$ to -0.88 , $p < 0.05$), likely represent the adaptive trade-off between extensive soil foraging and root construction costs under resource scarcity [63]. Furthermore, _density_points traits, exemplified by HDBSCAN_density_points (second most important feature, SHAP ~ 0.071), quantify strategic root biomass clustering. These correlated strongly with TRT volume/mass metrics ($r = 0.90\text{--}0.98$, $p < 0.05$) and reflect adaptive resource allocation where genotypes concentrate root development in resource-rich zones, indicative of root plasticity [35]. This capacity of ARTs to characterise the precise location and intensity of dense root congregations—putative primary water uptake zones [64]—at a different spatial resolution and perspective than global TRTs, likely explains their dominant predictive contribution. The analyses suggest a plausible biological interpretation: (i) ARTs quantify specific architectural adaptations (deep rooting via _centre_y, lateral exploration via _centre_x, strategic biomass allocation via _density_points); (ii) these adaptations are associated with more efficient water acquisition during drought; (iii) this enhanced water acquisition maintains favorable plant water status; (iv) improved water status supports higher stomatal conductance and leaf relative water content; and (v) these physiological responses confer greater drought tolerance, thereby explaining the superior performance of ART-based models in classifying genotypes initially grouped by these physiological indicators.

Despite the predictive dominance of ARTs, TRTs remain valuable, exhibiting a synergistic relationship when combined. The integrated ART + TRT model achieved the highest classification accuracy (97.4 %), surpassing models based exclusively on ARTs (96.3 %) or TRTs (85.6 %) (Fig. 6B–Table S10). This synergy is further evidenced by the distinct data structures revealed by multivariate analyses (PCA and MDS in Fig. 4F–H; t-SNE in Fig. S17) and PERMANOVA (Test Statistic: 5.578, $p = 0.002$), indicating that ARTs and TRTs capture different, complementary facets of root system architecture. The t-SNE visualisations showed ART features provided substantially better group separation than TRTs (Fig. S17). Even a streamlined model combining just four

selected ARTs with four TRTs achieved comparable performance, demonstrating practical utility for efficient phenotyping.

Beyond its enhanced predictive accuracy, the ART methodology presents significant technical advantages critical for robust scientific inquiry and practical deployment. Notably, ART features demonstrated superior stability across diverse experimental settings (mean CV 0.762 vs 1.361 for TRTs, $p = 0.0137$; Fig. 7D, Fig. S9), a crucial attribute for reliable phenotyping in variable field conditions. The ensemble nature of ART, employing nine distinct algorithms with diverse mathematical principles (Table 1), ensures comprehensive trait capture while yielding interpretable intermediate traits (e.g., FCM_centre_x, HDBSCAN_density_points). While ablation studies confirmed that a core trio of algorithms (FCM, HDBSCAN, and Mean-shift) provided ~ 80 % of the predictive gain (Fig. 7F–Table S7), the full ensemble offers superior robustness, justifying this approach over single-algorithm reliance. This interpretability is pivotal for hypothesis-driven research, offering a clear advantage over "black-box" models by allowing direct investigation of specific root architectural components. Moreover, ART's modular design facilitates customisation and future development: researchers can use the core trio of algorithms for rapid screening, incrementally add others for deeper insights, or integrate new algorithms to target novel root characteristics, ensuring the framework's enduring relevance.

The ART framework signifies a conceptual advancement in image-based phenotyping, shifting beyond predefined geometric descriptors to algorithmically discover and quantify biologically meaningful features. Unlike methods focusing on global distribution comparisons (e.g., Peeples et al. [31]) abstract topology (e.g., Li et al. [32]), supervised landmark detection (e.g., Berrigan et al. [29]), or temporal embeddings (e.g., Ubbens et al. [30]), ART extracts localised, geometric and density traits from algorithmically-identified dense regions (Table S1). This emphasis on explainable features (Fig. 7A and C) distinguishes ART from "black-box" deep learning models where the basis of prediction can be opaque [65]. ART thus bridges traditional trait extraction and machine learning, offering a powerful yet understandable means to leverage unsupervised learning for phenotype discovery. Beyond root phenotyping, this framework could be adapted to quantify plant disease symptoms or other complex visual traits with enhanced objectivity.

While this initial study focused on method development using six wheat genotypes, robustly demonstrating ART's discriminative power based on established drought responses (Fig. 2A and B) [57–60], future work should validate ART across diverse genotypes, species (e.g., maize, rice), and environmental conditions to establish its broader applicability.

The ART framework is inherently extensible, with all code and parameters available on GitHub for replication and adaptation. Future development could involve applying algorithms directly to RGB images to leverage colour and intensity information [66–68], incorporating texture analysis or advanced shape descriptors [69,70], adapting ARTs for regression modeling, or refining parameter optimisation through Monte Carlo simulations [71]. These extensions could potentially reduce dependency on initial segmentation while expanding analytical capabilities.

Crucially, the substantial predictive power of ARTs (73.3 % contribution in combined models, Fig. S14) offers exciting prospects for multi-omics integration. By quantitatively linking these nuanced root phenotypes to genomic, transcriptomic, or metabolomic data, ARTs could facilitate the identification of genetic markers and pathways underpinning drought tolerance and other complex traits.

In conclusion, the ART method provides a novel, objective, scalable, and customisable toolset that augments traditional root phenotyping pipelines. By enabling researchers to delve deeper into the complexities of plant adaptation, this work provides a conceptual blueprint for harnessing latent information within image data. It demonstrates that by relying on human vision alone, we may not truly see what we perceive [6]. By empowering us to 'see the unseen' through algorithmic lenses, this approach prompts a fundamental question for biology: what other

critical patterns lie hidden, just beyond the limits of our perception?

Author contributions

MS designed and conducted the experiment. SK guided in refining the experiments. MS wrote the manuscript and developed the code. CT and MH contributed to the conceptual ideas of the manuscript. AD created Fig. S2 and provided guidance in optimising image capture. All reviewed and edited the manuscript.

Funding

This study was funded by Agriculture Victoria Research, Victoria state government, Australia.

Data availability

All code, data and segmented images are available for download at <https://github.com/shoibms/ART>.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was funded by the Agriculture Victoria, Department of Energy, Environment and Climate Action, Australia. The authors wish to extend their sincere thanks to Emily L. Thoday-Kennedy for essential support during the experiments and to Edmond J. Breen for his guidance in optimising image capture.

General: Thank others for any contributions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.plaph.2025.100088>.

References

- [1] A. Seethpalli, K. Dhakal, M. Griffiths, H. Guo, G.T. Freschet, L.M. York, RhizoVision explorer: open-source software for root image analysis and measurement standardization, *AoB PLANTS* 13 (6) (2021), <https://doi.org/10.1093/aobpla/plab056>.
- [2] R. Tajima, Y. Kato, Comparison of threshold algorithms for automatic image processing of rice roots using freeware ImageJ, *Field Crops Res.* 121 (3) (2011) 460–463, <https://doi.org/10.1016/j.fcr.2011.01.015>.
- [3] WinRHIZO, Root Measurement and Analysis Software, 2024. Québec, QC, 2001. [Online], https://regent.qc.ca/assets/winrhizo_software.html.
- [4] R. Yasrab, J.A. Atkinson, D.M. Wells, A.P. French, T.P. Pridmore, M.P. Pound, RootNav 2.0: deep learning for automatic navigation of complex plant root architectures, *GigaScience* 8 (11) (Nov 1 2019), <https://doi.org/10.1093/gigascience/giz123>.
- [5] A. Das, et al., Digital imaging of root traits (DIRT): a high-throughput computing and collaboration platform for field-based root phenomics, *Plant Methods* 11 (1) (2015) 51, <https://doi.org/10.1186/s13007-015-0093-3>.
- [6] D.D. Hoffman, M. Singh, C. Prakash, The interface theory of perception, *Psychon. Bull. Rev.* 22 (6) (2015) 1480–1506, <https://doi.org/10.3758/s13423-015-0890-8>.
- [7] I. Misra, C. Lawrence Zitnick, M. Mitchell, R. Girshick, Seeing through the human reporting bias: visual classifiers from noisy human-centric labels, in: Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016, <https://doi.org/10.48550/arXiv.1512.06974> [Online].
- [8] Z. Huang, Z. Zeng, B. Liu, D. Fu, J. Fu, Pixel-bert: aligning image pixels with text by deep multi-modal transformers, arXiv pre-print server (2020), <https://doi.org/10.48550/arXiv.2004.00849>.
- [9] B. Schölkopf, et al., Toward causal representation learning, *Proc. IEEE* 109 (5) (2021) 612–634, <https://doi.org/10.48550/arXiv.2102.11107>.
- [10] K.N. Kirby, S.M. Kosslyn, Thinking visually, *Mind Lang.* 5 (4) (1990) 324–341, <https://doi.org/10.1111/j.1468-0017.1990.tb00167.x>.
- [11] P.J. White, T.S. George, P.J. Gregory, A.G. Bengough, P.D. Hallett, B.M. McKenzie, Matching roots to their environment, *Ann. Bot.* 112 (2) (Jul 2013) 207–222, <https://doi.org/10.1093/aob/mct123>.
- [12] A.G. Jacobsen, G. Jervis, J. Xu, J.F. Topping, K. Lindsey, Root growth responses to mechanical impedance are regulated by a network of ROS, ethylene and auxin signalling in *Arabidopsis*, *New Phytol.* 231 (1) (Jul 2021) 225–242, <https://doi.org/10.1111/nph.17180>.
- [13] Y. Uga, Challenges to design-oriented breeding of root system architecture adapted to climate change, *Breed. Sci.* 71 (1) (Feb 2021) 3–12, <https://doi.org/10.1270/jbsbs.20118>.
- [14] J.P. Lynch, Root architecture and plant productivity, *Plant Physiol.* 109 (1) (1995) 7–13, <https://doi.org/10.1104/pp.109.1.7>.
- [15] S.J. Tumber-Dávila, H.J. Schenk, E. Du, R.B. Jackson, Plant sizes and shapes above and belowground and their interactions with climate, *New Phytol.* 235 (3) (Aug 2022) 1032–1056, <https://doi.org/10.1111/nph.18031>.
- [16] E.S. Ober, et al., Wheat root systems as a breeding target for climate resilience, *Theor. Appl. Genet.* 134 (6) (2021) 1645–1662, <https://doi.org/10.1007/s00122-021-03819-w>.
- [17] P. Yu, et al., Plant flavones enrich rhizosphere oxalobacteraceae to improve maize performance under nitrogen deprivation, *Nat. Plants* 7 (4) (Apr 2021) 481–499, <https://doi.org/10.1038/s41477-021-00897-y>.
- [18] J.M. Ruiz-Lozano, et al., Arbuscular mycorrhizal symbiosis induces strigolactone biosynthesis under drought and improves drought tolerance in lettuce and tomato, *Plant Cell Environ.* 39 (2) (2016) 441–452, <https://doi.org/10.1111/pce.12631>.
- [19] J.P. Lynch, J.G. Chimgungu, K.M. Brown, Root anatomical phenes associated with water acquisition from drying soil: targets for crop improvement, *J. Exp. Bot.* 65 (21) (2014) 6155–6166, <https://doi.org/10.1093/jxb/eru162>.
- [20] C. Maurel, P. Nacry, Root architecture and hydraulics converge for acclimation to changing water availability, *Nat. Plants* 6 (7) (2020) 744–749, <https://doi.org/10.1038/s41477-020-0684-5>.
- [21] J.P. Lynch, Harnessing root architecture to address global challenges, *Plant J.* 109 (2) (2022) 415–431, <https://doi.org/10.1111/tpj.15560>.
- [22] S.A. Rolfe, J. Griffiths, J. Ton, Crying out for help with root exudates: adaptive mechanisms by which stressed plants assemble health-promoting soil microbiomes, *Curr. Opin. Microbiol.* 49 (Jun 2019) 73–82, <https://doi.org/10.1016/j.mib.2019.10.003>.
- [23] S.J. Prince, et al., Root xylem plasticity to improve water use and yield in water-stressed soybean, *J. Exp. Bot.* 68 (8) (2017) 2027–2036, <https://doi.org/10.1093/jxb/erw472>.
- [24] S. Trachsel, S.M. Kaepller, K.M. Brown, J.P. Lynch, Shovelomics: high throughput phenotyping of maize (zea mays L.) root architecture in the field, *Plant Soil* 341 (1–2) (2011) 75–87, <https://doi.org/10.1007/s11104-010-0623-8>.
- [25] A. Wasson, L. Bischof, A. Zwart, M. Watt, A portable fluorescence spectroscopy imaging system for automated root phenotyping in soil cores in the field, *J. Exp. Bot.* 67 (4) (Feb 2016) 1033–1043, <https://doi.org/10.1093/jxb/erv570>.
- [26] M.G. Huck, H.M. Taylor, The rhizotron as a tool for root research, *Adv. Agron.* 35 (1982) 1–35, [https://doi.org/10.1016/S0065-2113\(08\)60320-X](https://doi.org/10.1016/S0065-2113(08)60320-X).
- [27] F.M. Bauer, et al., Development and validation of a deep learning based automated minirhizotron image analysis pipeline, *Plant Phenomics* 2022 (2022) 9758532, <https://doi.org/10.34133/2022/9758532>.
- [28] J.A. Palta, N.C. Turner, Crop root system traits cannot be seen as a silver bullet delivering drought resistance, *Plant Soil* 439 (1–2) (2019) 31–43, <https://doi.org/10.1007/s11104-018-3864-6>.
- [29] E.M. Berrigan, et al., Fast and efficient root phenotyping via pose estimation, *Plant Phenomics* 6 (2024) 175, <https://doi.org/10.34133/plantphenomics.0175>.
- [30] J. Ubbens, M. Cieslak, P. Prusinkiewicz, I. Parkin, J. Ebersbach, I. Stavness, Latent space phenotyping: automatic image-based phenotyping for treatment studies, *Plant Phenomics* 2020 (2020) 5801869, <https://doi.org/10.34133/2020/5801869>.
- [31] J. Peeples, W. Xu, R. Gloaguen, D. Rowland, A. Zare, Z. Brym, Spatial and texture analysis of root system distribution with Earth mover's distance (STARSEED), *Plant Methods* 19 (1) (Jan 5 2023) 2, <https://doi.org/10.1186/s13007-022-00974-z>.
- [32] M. Li, M.H. Frank, V. Coneva, W. Mio, D.H. Chitwood, C.N. Topp, The persistent homology mathematical framework provides enhanced genotype-to-phenotype associations for plant morphology, *Plant Physiol.* 177 (4) (Aug 2018) 1382–1395, <https://doi.org/10.1104/pp.18.00104>.
- [33] O. Erenstein, M. Jaleta, K.A. Mottaleb, K. Sonder, J. Donovan, H.-J. Braun, Global trends in wheat production, consumption and trade, in: M.P. Reynolds, H.-J. Braun (Eds.), *Wheat Improvement: Food Security in a Changing Climate*, Springer International Publishing, Cham, 2022, pp. 47–66.
- [34] A. Gupta, A. Rico-Medina, A.I. Cano-Delgado, The physiology of plant responses to drought, *Science* 368 (6488) (Apr 17 2020) 266–269, <https://doi.org/10.1126/science.aaz7614>.
- [35] A. Zhan, H. Schneider, J.P. Lynch, Reduced lateral root branching density improves drought tolerance in maize, *Plant Physiol.* 168 (4) (Aug 2015) 1603–1615, <https://doi.org/10.1104/pp.15.00187>.
- [36] A. Odone, O. Popovic, K. Thorup-Kristensen, Deep roots: implications for nitrogen uptake and drought tolerance among winter wheat cultivars, *Plant Soil* 500 (1–2) (2023) 1–20, <https://doi.org/10.1007/s11104-023-06255-5>.
- [37] H. Bacher, Y. Sharaby, H. Walia, Z. Peleg, Modifying root-to-shoot ratio improves root water influxes in wheat under drought stress, *J. Exp. Bot.* 73 (5) (Mar 2 2021) 1643–1654, <https://doi.org/10.1093/jxb/eraab500>.
- [38] T. Bonpart, et al., Affordable and robust phenotyping framework to analyse root system architecture of soil-grown plants, *Plant J.* 103 (6) (Sep 2020) 2330–2343, <https://doi.org/10.1111/tpj.14877>.

- [39] S. Duan, B. Li, H. Gu, H. Jiang, X. Zhang, X. Liu, Root matters: lying seeds flat with the crease Down improves grain yield in winter wheat under drought stress, *Plant Soil* 477 (1–2) (2022) 577–594, <https://doi.org/10.1007/s11104-022-05461-x>.
- [40] Python Software Foundation, Python Language Reference, 2023 [Online], version 3.12.2. <https://docs.python.org/3.12/reference/>.
- [41] R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, 2023. Vienna, Austria. [Online], <https://www.R-project.org/>.
- [42] A.G. Smith, et al., RootPainter: deep learning segmentation of biological images with corrective annotation, *New Phytol.* 236 (2) (Oct 2022) 774–791, <https://doi.org/10.1111/nph.18387>.
- [43] J.P. Lynch, Steep, cheap and deep: an ideotype to optimize water and N acquisition by maize root systems, *Ann. Bot.* 112 (2) (2013) 347–357, <https://doi.org/10.1093/aob/mcs293>.
- [44] P. Lootens, T. Ruttink, A. Rohde, D. Combes, P. Barre, I. Roldán-Ruiz, High-throughput phenotyping of lateral expansion and regrowth of spaced *Lolium* perenne plants using on-field image analysis, *Plant Methods* 12 (2016) 1–14, <https://doi.org/10.1186/s13007-016-0132-8>.
- [45] A. Hodge, The plastic plant: root responses to heterogeneous supplies of nutrients, *New Phytol.* 162 (1) (2004) 9–24, <https://doi.org/10.1111/j.1469-8137.2004.01015.x>.
- [46] A.P. Wasson, et al., Traits and selection strategies to improve root systems and water uptake in water-limited wheat crops, *J. Exp. Bot.* 63 (9) (2012) 3485–3498, <https://doi.org/10.1093/jxb/ers111>.
- [47] F. Pedregosa, et al., Scikit-learn: machine learning in python, *J. Mach. Learn. Res.* 12 (2011) 2825–2830, <https://doi.org/10.48550/arXiv.1201.0490>.
- [48] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, in: Presented at the Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, 1996. Portland, Oregon.
- [49] D. Comaniciu, P. Meer, Mean shift: a robust approach toward feature space analysis, *IEEE Trans. Pattern Anal. Mach. Intell.* 24 (5) (2002) 603–619, <https://doi.org/10.1109/34.1000236>.
- [50] M. Ankerst, M.M. Breunig, H.-P. Kriegel, J. Sander, OPTICS: ordering points to identify the clustering structure, *ACM Sigmod record* 28 (2) (1999) 49–60, <https://doi.org/10.1145/304182.304187>.
- [51] L. McInnes, J. Healy, S. Astels, HdbSCAN: hierarchical density based clustering, *J. Open Source Softw.* 2 (6/22) (2024, 2017), <https://doi.org/10.21105/joss.00205>.
- [52] J.C. Bezdek, R. Ehrlich, W. Full, FCM: the fuzzy c-means clustering algorithm, *Comput. Geosci.* 10 (2–3) (1984) 191–203, [https://doi.org/10.1016/0098-3004\(84\)90020-7](https://doi.org/10.1016/0098-3004(84)90020-7).
- [53] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Süstrunk, SLIC superpixels compared to state-of-the-art superpixel methods, *IEEE Trans. Pattern Anal. Mach. Intell.* 34 (11) (2012) 2274–2282, <https://doi.org/10.1109/TPAMI.2012.120>.
- [54] R.E. Smart, G.E. Bingham, Rapid estimates of relative water content, *Plant Physiol.* 53 (2) (Feb 1974) 258–260, <https://doi.org/10.1104/pp.53.2.258>.
- [55] A. Raman, et al., Drought yield index to select high yielding rice lines under different drought stress severities, *Rice* 5 (1) (Dec 2012) 1–12, <https://doi.org/10.1186/1939-8433-5-31>.
- [56] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, *Adv. Neural Inf. Process. Syst.* 30 (2017).
- [57] S. Bennani, N. Nsarellah, M. Jlibene, W. Tadesse, Efficiency of drought tolerance indices under different stress severities for bread wheat selection, *Aust. J. Crop. Sci.* 11 (4) (2017) 395–405, <https://doi.org/10.21475/ajcs.17.11.04.pne272>.
- [58] A.K. Yadav, A.J. Carroll, G.M. Estavillo, G.J. Rebetzke, B.J. Pogson, Wheat drought tolerance in the field is predicted by amino acid responses to glasshouse-imposed drought, *J. Exp. Bot.* 70 (18) (Sep 24 2019) 4931–4948, <https://doi.org/10.1093/jxb/erz224>.
- [59] H. Hone, et al., Profiling, isolation and characterisation of beneficial microbes from the seed microbiomes of drought tolerant wheat, *Sci. Rep.* 11 (1) (2021/06/07 2021) 11916, <https://doi.org/10.1038/s41598-021-91351-8>.
- [60] G. McDonald, Drought Tolerance of Wheat Varieties, 2023. <https://grdc.com.au/resources-and-publications/grdc-update-papers/tab-content/grdc-update-papers/2016/02/drought-tolerance-fo-wheat-varieties>. (Accessed 5 July 2024).
- [61] X. Li, et al., Deeper roots associated with cooler canopies, higher normalized difference vegetation index, and greater yield in three wheat populations grown on stored soil water, *J. Exp. Bot.* 70 (18) (Sep 24 2019) 4963–4974, <https://doi.org/10.1093/jxb/erz232>.
- [62] Y. Uga, et al., Control of root system architecture by DEEPER ROOTING 1 increases rice yield under drought conditions, *Nat. Genet.* 45 (9) (Sep 2013) 1097–1102, <https://doi.org/10.1038/ng.2725>.
- [63] L.H. Comas, S.R. Becker, V.M. Cruz, P.F. Byrne, D.A. Dierig, Root traits contributing to plant productivity under drought, *Front. Plant Sci.* 4 (Nov 5 2013) 442, <https://doi.org/10.3389/fpls.2013.00442>.
- [64] V. Vadex, Root hydraulics: the forgotten side of roots in drought adaptation, *Field Crops Res.* 165 (2014) 15–24, <https://doi.org/10.1016/j.fcr.2014.03.017>.
- [65] D. Castelvecchi, Can we open the black box of AI? *Nat. News* 538 (7623) (Oct 6 2016) 20, <https://doi.org/10.1038/538020a>.
- [66] A.D. Richardson, Tracking seasonal rhythms of plants in diverse ecosystems with digital camera imagery, *New Phytol.* 222 (4) (Jun 2019) 1742–1750, <https://doi.org/10.1111/nph.15591>.
- [67] J. Johnson, et al., Enhanced field-based detection of potato blight in complex backgrounds using deep learning, *Plant Phenomics* 2021 (2021) 9835724, <https://doi.org/10.34133/2021/9835724>.
- [68] N.-A. Pham, et al., Quantitative image analysis of immunohistochemical stains using a CMYK color model, *Diagn. Pathol.* 2 (Feb 27 2007) 1–10, <https://doi.org/10.1186/1746-1596-2-8>.
- [69] P. Kulkarni, A. Karwande, T. Kolhe, S. Kamble, A. Joshi, M. Wyawahare, Plant disease detection using image processing and machine learning, *arXiv preprint* (2021), <https://doi.org/10.48550/arXiv.2106.10698> arXiv:2106.10698.
- [70] S. Das Choudhury, S. Bashyam, Y. Qiu, A. Samal, T. Awada, Holistic and component plant phenotyping using temporal image sequence, *Plant Methods* 14 (2018) 1–21, <https://doi.org/10.1186/s13007-018-0303-x>.
- [71] S. Shashaani, K. Vahdat, Monte Carlo based machine learning, in: International Conference on Operations Research, Springer International Publishing, in Operations Research Proceedings 2022, Cham, 2022, pp. 625–631, https://doi.org/10.1007/978-3-031-24907-5_75 [Online].

Chapter 5. Tissue-Specific Metabolomic Networks

Orchestrate Osmotic Stress Adaptation

Solving the Mechanistic Understanding Bottleneck Through Network Architecture

Our analytical journey has progressed from automated discovery (AutoML) through algorithmic precision (ART), but these phenotypic observations represent only the external expression of deeper organisational principles. This chapter directly addresses the mechanistic understanding bottleneck (Section 1.3.4) by moving beyond correlative descriptions to reveal the architectural blueprints that govern drought adaptation.

This work targets Research Question 2: What are the underlying architectural principles and temporal dynamics of tissue-specific metabolic networks that drive the systemic stress response in wheat?

Using time-series metabolomics and advanced network science, we demonstrate that drought tolerance is not a simple accumulation of protective compounds, but an emergent property of distinct, tissue-specific network designs.

The architectural discoveries are quantitatively striking: Leaf networks in tolerant genotype were ~40% denser with elevated transitivity (0.740-0.804), engineered for rapid coordination, while root networks exhibited higher modularity (0.213-0.288) and fragmentation (18-21 components), designed for localised adaptation. Temporally, these systems deploy a sophisticated biphasic strategy: initial coordination ($r \approx 0.546$) transitioning to strategic decoupling ($r \approx 0.350$).

Rigorous Bayesian validation confirms these architectures significantly deviate from random expectations ($p < 0.001$), establishing them as evolved strategies rather than stochastic organisation. By revealing the molecular blueprints of adaptation, this chapter establishes that how a plant organises its metabolism is as important as what metabolites it produces. This sets the stage for addressing the final bottleneck: understanding how these architectural principles coordinate with observable physiological responses.

Statement of contributions of joint authorship

Mirza Shoaib designed and conducted the experiment. Surya Kant guided in refining the experiments. Mirza Shoaib wrote the manuscript and developed the code. Simone J Rochfort converted the raw LCMS data. Simone J Rochfort and Priyanka Reddy guided during analysis. Doris Ram helped with LCMS analysis. Matthew J. Hayden, Surya Kant and all reviewed and edited the manuscript.

Statement from the co-author confirming the PhD candidate's contributions to the authorship

As a co-author of the manuscript:

Shoaib M, Rochfort SJ, Reddy P, Ram D, Hayden M, and Kant S. (2025). Tissue-specific Metabolomic Networks Orchestrate Osmotic Stress Adaptation in Wheat. Submitted to Molecular Plant as a Research Perspective, following a positive pre-submission inquiry.

I confirm that Mirza Shoaib has made the following contributions,

- Conducted the literature review
- Designed and conducted the experiments
- Performed all data analyses
- Generated all figures
- Led the writing of the manuscript



Dr Surya Kant

29 July 2025

Tissue-specific Metabolomic Networks Orchestrate Osmotic Stress Adaptation in Wheat

Authors

Mirza Shoaib^{1,2*}, Simone J Rochfort^{2,3}, Priyanka Reddy⁴, Doris Ram³, Matthew J. Hayden^{2,3} and Surya Kant^{5*}

Affiliations

¹Agriculture Victoria, Grains Innovation Park, 110 Natimuk Road, Horsham, Victoria 3400, Australia

²School of Applied Systems Biology, La Trobe University, Bundoora, Victoria 3083, Australia

³Agriculture Victoria, AgriBio, Centre for AgriBioscience, 5 Ring Road, Bundoora, Victoria 3083, Australia

⁴ School of Chemistry, Bio21 Institute of Molecular Science and Biotechnology, The University of Melbourne, Victoria 3010 Australia

⁵Department of Ecological, Plant and Animal Science, School of Agriculture, Biomedicine & Environment, La Trobe University, Bundoora, Victoria 3083, Australia

* Correspondence: shoaib.mirza@agriculture.vic.gov.au and S.Kant@latrobe.edu.au

ORCID

Mirza Shoaib: <https://orcid.org/0000-0002-6773-6050>

Simone J Rochfort: <https://orcid.org/0000-0001-8442-6081>

Priyanka Reddy: <https://orcid.org/0000-0002-1504-1126>

Doris Ram: <https://orcid.org/0000-0003-2717-0424>

Matthew Hayden: <https://orcid.org/0000-0003-3117-5500>

Surya Kant: <https://orcid.org/0000-0001-6178-7036>

networks and understanding how these coordinated networks underpin whole-plant adaptation (Barabási and Oltvai, 2004; Philippou et al., 2021; Gargallo-Garriga *et al.*, 2014; Hsiao and Xu, 2000; Li *et al.*, 2021; Lozano-Elena *et al.*, 2022a; Min et al., 2020; Toubiana *et al.*, 2013). This gap limits our ability to fully grasp how plants maintain homeostasis under water-limited conditions, particularly as drought severity increases due to climate change (Chiang *et al.*, 2021; Fernie and Stitt, 2012; Hodge, 2004; Li *et al.*, 2021; Obata and Fernie, 2012).

Recent advances in metabolomics and systems biology have revealed tissue-specific metabolic reprogramming under osmotic stress, highlighting the dynamic nature of plant responses to abiotic stress (Cardoso *et al.*, 2023; Clark et al., 2020; Fernie and Tohge, 2017; Mintz-Oron *et al.*, 2012; Nakabayashi and Saito, 2015; Obata and Fernie, 2012). However, many studies focus on individual metabolites or pathways, often overlooking the critical role of network architecture in coordinating stress responses (Sévin *et al.*, 2016; Sweetlove and Fernie, 2018; Toubiana *et al.*, 2013). In this context, metabolic network topologies and connectivity emerge as vital determinants of plant adaptation (Sévin *et al.*, 2016; Sweetlove and Fernie, 2018; Toubiana *et al.*, 2013). For example, hub metabolites play pivotal roles in adjusting network function under stress (Cardoso *et al.*, 2023), while trait integration across leaves, stems, and roots is key to optimising resource allocation and growth (Wang *et al.*, 2022). Despite these insights, the specific architectural features, temporal dynamics, and contributions of these networks to drought tolerance remain poorly understood (Fernie and Stitt, 2012; Fernie and Tohge, 2017).

Roots, serving as the primary interface with the soil, require modular and plastic networks to respond to heterogeneous conditions, while leaves are optimised for photosynthetic efficiency through coordinated biochemical networks (Chaves *et al.*, 2008; Hodge, 2004). This functional divergence manifests in opposite metabolic responses to drought stress (Gargallo-Garriga *et al.*, 2014). Yet, the contrasting network strategies of roots and leaves—and how these differ between drought-tolerant and susceptible genotypes—are underexplored. While studies in model species like *Arabidopsis* and maize have demonstrated the importance of network properties during drought stress (Lozano-Elena *et al.*, 2022c; Miao *et al.*, 2017), the tissue-specific architectural principles underlying genotypic differences in stress tolerance remain poorly characterised, particularly in crops such as wheat.

In this study, we combine untargeted metabolomics with network analysis and a time-series approach to uncover a fundamental asymmetry in the tissue-specific molecular architectures underlying osmotic stress resilience in wheat. We selected wheat genotypes with extensively documented drought response differences across multiple independent studies: G1 (Gladius, drought-tolerant) and G2 (DAS5_003811, drought-susceptible) (Bennani et al., 2017; Hone et al., 2021; Shoaib et al., 2025; Yadav et al., 2019). This established phenotypic foundation enables focused mechanistic investigation of tissue-specific network principles. We show that roots and leaves deploy fundamentally different network strategies. In G1, leaves maintain highly integrated networks—with a 40% higher density and significantly greater transitivity—relative to the more modular, fragmented architecture of the roots. Moreover, Temporal analysis reveals a strategic decoupling between leaf and root networks under prolonged stress, with cross-tissue correlation declining from 0.55 to 0.35 as stress progresses, and a biphasic response comprising initial coordinated adjustments followed by tissue-specific adaptations—a pattern consistent with observations in maize (Miao et al., 2017). These findings provide novel insights into the organisational principles governing plant stress responses and establish a framework for understanding and enhancing drought tolerance in crops, with potential applications for improving agricultural resilience under climate change.

RESULTS AND DISCUSSION

Tissue-specific Network Architecture

Our comprehensive network analysis of 2,471 molecular features under osmotic stress revealed a fundamental asymmetry in tissue-specific molecular organisation between roots and leaves of wheat, particularly pronounced in the drought-tolerant genotype (G1) compared to the drought-susceptible genotype (G2) (Fig. 1A, B).

In G1, network analysis of high-confidence features ($n = 964$), identified via partial least squares discriminant analysis (PLS-DA) with variable importance in projection (VIP) scores > 1 (see Supplementary Fig. S1 for the detailed preprocessing workflow and Interactive PLS-DA Loading Plot at <https://github.com/shoaibms/metabo>), revealed striking architectural differences between

tissues that are consistent with observations in other crop species under drought stress (Min *et al.*, 2020). Leaf networks exhibited higher density (0.354) compared to root networks (0.192), and greater transitivity (leaves: 0.740–0.804; roots: 0.686–0.714), indicating that leaf molecular networks are more integrated and facilitate rapid molecular coordination. This architectural divergence parallels observations by Fait *et al.* (Fait *et al.*, 2020). In contrast, root networks displayed higher modularity (roots: 0.213–0.288; leaves: 0.097–0.162) and greater fragmentation (roots: 18–21 components; leaves: 6 components), suggesting a localised and modular organisation optimised for specific environmental adjustments (Fig. 1C).

The organisation of molecular hubs further reinforced the distinct strategies employed by each tissue under osmotic stress. In leaves, where rapid coordination of stress responses is critical, molecular hubs exhibited concentrated high connectivity, with the top 20 hubs averaging 872 connections in G1 (Fig. 1E). These highly connected nodes were predominantly primary metabolites—including carbohydrates and amino acids—reflecting their central role in coordinating rapid metabolic adjustments during stress. (see Supplementary Information 'Metabolite Annotation Methods' section for detailed HMDB and GNPS annotation procedures). In contrast, root hubs displayed lower, more distributed connectivity (top 20 hubs averaging 767 connections in G1), as illustrated by rank-degree decay curves (Fig. 1F), reflecting their role in localised environmental responses. These tissue-specific architectural principles, as identified through Spearman-based analysis, are further corroborated by Bayesian network analysis (Fig. 1D)¹. These architectural differences extend beyond hub organisation. Leaves maintained shorter mean path lengths (2.04–2.17) and higher temporal coherence (0.749–0.782), facilitating efficient signal transmission across the network. Root networks, with their longer path lengths (2.29–2.44) and lower temporal coherence (0.686–0.716), preserved a modular structure that enables independent responses to heterogeneous soil conditions. The drought-tolerant genotype G1 leveraged these tissue-specific architectures effectively, with module-level organisation showing distinct tissue-dependent patterns (Fig. 1G). Analysis of module preservation and network stability demonstrated superior structural robustness in G1 networks compared to G2 (Fig. 1H), showing refined network integration that decreased strategically from 64.2% to 46.5% with convergent response ratios (1.92 to 1.03) (Fig. 1I). In contrast, the susceptible genotype G2 displayed unstable oscillating integration (55.1%, 48.6%, 57.5%), suggesting compromised adaptation. Together, these patterns demonstrate how

¹ Interactive versions of the network analyses are available at: <https://github.com/shoaibms/network>

distinct network architectures - integrated in leaves versus modular in roots - serve complementary functions in osmotic stress adaptation.

Preliminary Bayesian network analyses indicated that these observed molecular architectures deviate significantly from random expectations, suggesting non-random organisation. (For detailed statistical validation and null model comparisons, see the 'Conservation and Validation' section.)

In G2, these architectural distinctions were less pronounced. Leaf and root networks in G2 did not exhibit such significant differences in network density, modularity, or hub connectivity, indicating less specialised molecular organisation compared to G1.

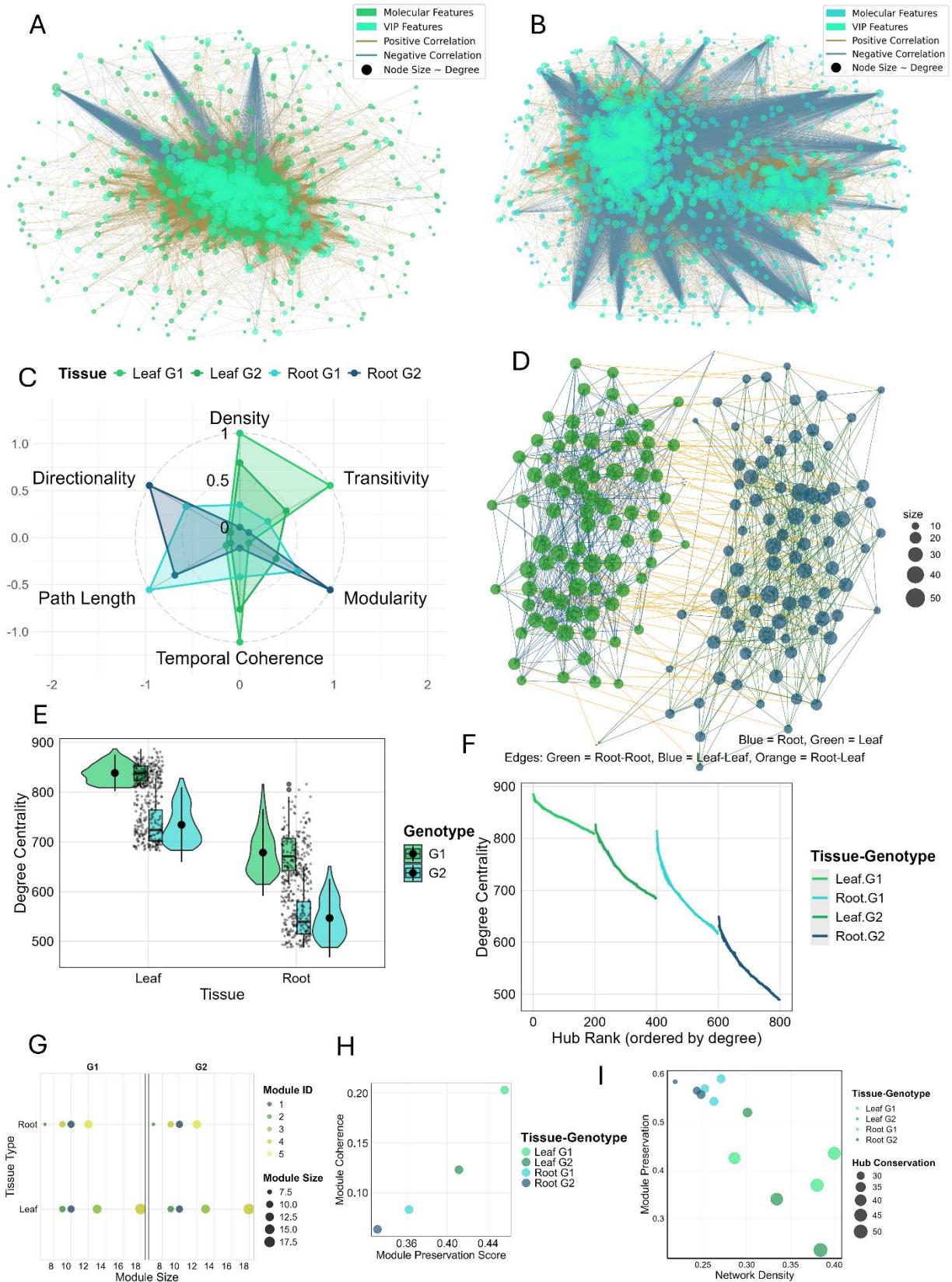


Fig. 1. Tissue-specific molecular network architectures in drought-tolerant wheat under osmotic stress.

Networks were constructed from high-confidence molecular features ($VIP > 1$) identified under osmotic stress conditions, using Spearman correlations ($|r| > 0.7$, $FDR < 0.05$). Nodes represent molecular features, with size scaled by degree centrality; bright light green nodes denote $VIP > 1$ features. Edge colours and thickness indicate correlation direction (gold: positive, blue: negative) and strength, respectively. A, B, Leaf (A) and root (B) molecular networks exhibit distinct organisational patterns. Leaf networks show higher density (0.354) and transitivity (0.740-0.804) compared to root networks (0.192, 0.686-0.714), while root networks display higher modularity (0.213-0.288) and fragmentation (18-21 components) versus leaves (0.097-0.162, 6 components). C, Radar plot comparing network metrics between leaf and root tissues in drought-tolerant (G1) and drought-susceptible (G2) genotypes. D, Bayesian network aligns with Spearman-based analyses. Nodes coloured by tissue origin (blue: root, green: leaf) with cross-tissue connections highlighted in orange ($n=384$). Network topology confirms the fundamental architectural asymmetry between tissues, revealing distinct modular organisation in roots versus integrated connectivity in leaves. E, F, Hub analyses show concentrated distribution in leaves (top 20 hubs) versus distributed patterns in roots. Rank-degree decay curves illustrate these architectural differences through hub organisation patterns. G, Module-level analysis demonstrating tissue-specific organisational strategies. Node sizes reflect module membership represented by distinct colours. Faceted plots by genotype reveal contrasting modular architectures between leaf and root networks, highlighting tissue-specific organisational principles. H, Integrated stability analysis reveals genotype-specific module characteristics. Quantitative assessment of module preservation (x-axis) and coherence (y-axis) demonstrates superior structural integrity in G1 networks (leaf preservation: 0.375, coherence: 0.600; root preservation: 0.357, coherence: 0.497) compared to G2 (leaf: 0.354, 0.482; root: 0.328, 0.451). Node colour indicates tissue-genotype combinations; node position reflects network stability metrics derived from 1,000 permutations. I, Integrated network stability analysis comparing network density, module preservation and hub conservation (node size: 30-50) show distinct tissue strategies, with roots favouring module preservation over density and leaves maintaining a balance between connectivity and modularity.

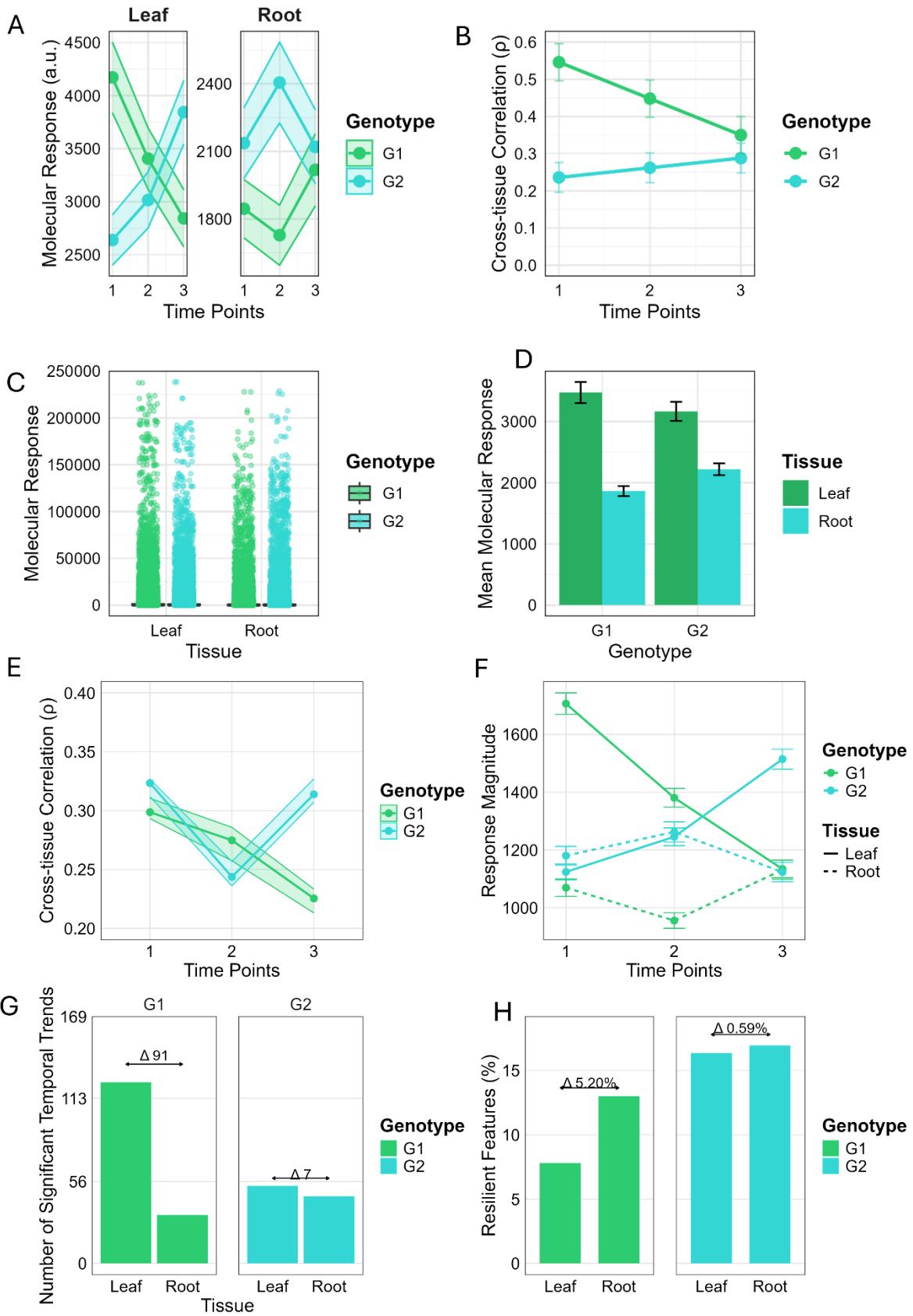
These findings reveal that roots and leaves in the drought-tolerant genotype G1 have evolved distinct molecular network architectures serving complementary functions under osmotic stress, consistent with their contrasting metabolic responses to drought (Gargallo-Garriga *et al.*, 2014). Leaves maintain dense, integrated networks facilitating rapid and coordinated molecular responses essential for photosynthetic adaptation. Roots employ modular, fragmented networks enabling localised molecular adjustments. This architectural asymmetry is leveraged more effectively in G1 than in G2, contributing to enhanced drought tolerance.

Temporal Dynamics and Cross-tissue Coordination

Building on these tissue-specific architectural differences, we next examined how these distinct networks responded dynamically to osmotic stress over time. In G1, the initial strong correlation between leaf and root molecular features ($r = 0.546$ on Time point 1) systematically declined over time ($r = 0.350$ on Time point 3) (Fig. 2A, B), suggesting a strategic decoupling to facilitate tissue-specific adaptations - a phenomenon also observed in transcriptomics analysis in maize under drought stress (Miao *et al.*, 2017). This coordinated response manifested through a precise biphasic strategy in G1, characterised by distinct acute (leaf: 0.46, root: 0.38) and prolonged (leaf: 0.42, root: 0.35) effect sizes, reflecting the tissue-specific temporal progression of drought responses (Fig. 2C) (Gargallo-Garriga *et al.*, 2014). Temporal coherence evolved differently between tissues, with leaves showing strategic attenuation (32.6% to 11.1%) while roots exhibited enhanced coordination (22.0% to 39.6%). This tissue-specific temporal progression (Fig. 2D) was evident in the molecular changes across each time point: leaves exhibited consistent suppression (Time point 1→2: -69.375; Time point 2→3: -79.384), while roots shifted from initial inhibition to stimulation (Time point 1→2: -26.061; Time point 2→3: +26.289). This orchestrated decrease (Fig. 2E) coincided with a strategic redistribution of highly resilient molecular features favouring roots (leaf: 7.80%, root: 13.00%; differential: 5.20%) (Fig. 2F), consistent with the accumulation of osmolytes and secondary metabolites in root tissue under drought stress (see Supplementary Fig. S2-S5 for detailed temporal dynamics of chemical and functional groups). In contrast, G2 displayed weaker and relatively stable leaf-root correlations ($r = 0.236-0.288$) over the same period and a uniform distribution of highly resilient molecular features between leaves and roots (leaf: 16.36%, root: 16.95%; differential: 0.59%), indicating less specialised molecular reprogramming and compromised tissue-specific regulation.

Temporal patterns of molecular responses further distinguished G1 and G2. In G1, leaves exhibited significantly more temporal trends than roots (124 vs. 33; trend_p_adj < 0.05) (Fig. 2G), reflecting extensive molecular reprogramming in the leaves. Despite this asymmetry, G1 maintained precise magnitude control between tissues (leaf-root difference in effect sizes: 0.019; leaves: 0.935 ± 0.289 , roots: 0.916 ± 0.266). Additionally, G1 employed a precise biphasic strategy characterised by strong initial suppression in leaves and calibrated responses in roots (Fig. 2H, I and J), facilitating coordinated adaptation.

In G2, these temporal dynamics were less pronounced. Leaves and roots showed similar numbers of significant temporal trends (leaves: 53; roots: 46), and there was weaker magnitude regulation (leaf–root difference: 0.107; leaves: 0.962 ± 0.280 , roots: 0.855 ± 0.250), suggesting less effective tissue-specific stress responses.



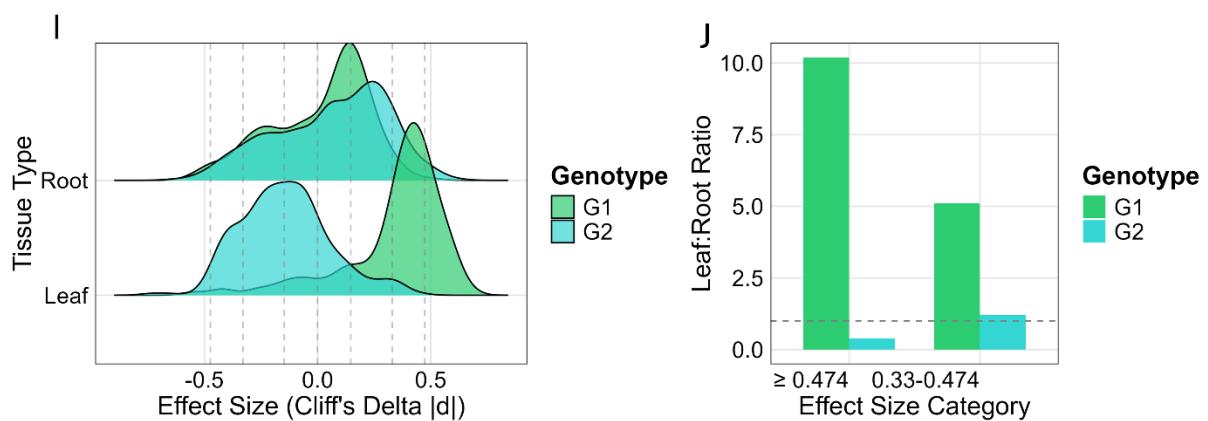


Fig. 2. Temporal dynamics of tissue-specific molecular coordination under osmotic stress.

A, Temporal trajectories of stress-responsive molecular features ($n = 964$) in leaf and root tissues across a 3-time point stress period. Shaded areas represent 95% confidence intervals. B, Cross-tissue correlation analysis revealing genotype-specific coordination strategies. G1 shows strategic decline in correlation ($r = 0.546$ to 0.350) while G2 maintains a lower, stable correlation ($r = 0.236$ - 0.288). C, Distribution of molecular responses across tissues and genotypes. Points represent individual features coloured by response magnitude ($n = 964$). D, Mean molecular response comparison between tissues and genotypes ($n = 7$ biological replicates). G1 demonstrates pronounced tissue differentiation with high leaf response ($3,473 \pm 392$) and restrained root activity ($1,864 \pm 241$), creating a larger leaf-root differential compared to G2 (leaves: $3,166 \pm 475$; roots: $2,220 \pm 377$), indicating less tissue-specific regulation. E, Cross-tissue correlation analysis of stress-responsive features ($VIP > 1$) revealing distinct genotype strategies. G1 shows consistent correlation decline ($p = 0.299$ to 0.225) indicating progressive adaptation; G2 displays oscillating coordination (Time point 1: $p = 0.323$; Time points 2: $p = 0.244$; Time points 3: $p = 0.314$). Error bands: 95% confidence intervals ($n = 1,000$ bootstrap resamples). F, Response magnitude comparison demonstrating G1's biphasic strategy: leaf suppression (Time point 1 \rightarrow 2: -69.375 , Time point 2 \rightarrow 3: -79.384) and dynamic root responses (Time point 1 \rightarrow 2: -26.061 , Time point 2 \rightarrow 3: $+26.289$). Error bars: SE, $n = 7$, biological replicates. G, Temporal trends varied by tissue and genotype. G1 leaves showed significantly more trends (124, trend_p_adj < 0.05) than roots (33, trend_p_adj < 0.05). In contrast, G2 exhibited comparable numbers of temporal trends between leaves (53) and roots (46). H, G1 exhibits strategic tissue specialisation with roots containing 5.2% more resilient features than leaves (roots: 13.0%, leaves: 7.8%), while G2 shows negligible tissue differentiation (roots: 16.95%, leaves: 16.36%, differential: 0.59%), reflecting compromised stress-specific regulation. I, J, Effect size distributions and leaf:root response ratios reinforce these contrasts. G1 maintains controlled tissue differentiation over time, while G2's ratio approaches 1, indicating minimal tissue specialisation. Bootstrap validation ($n = 1,000$) confirms the statistical robustness of these patterns.

Network efficiency metrics underscored the distinctions between G1 and G2. Consistent with their architectural properties (see Network Architecture), leaf networks maintained their characteristic high temporal coherence while demonstrating enhanced response coordination over time and reflecting rapid and coordinated stress responses. Root networks in G1, while exhibiting the

characteristic lower coherence and longer path lengths described earlier, displayed a modular strategy conducive to sustained adaptation. In G2, network properties remained less stable over time, and the weaker tissue-specific differentiation and reduced temporal control reflected disjointed regulation underlying its susceptibility to osmotic stress.

Bayesian network analysis quantitatively validated these architectural distinctions through comprehensive null model comparisons, revealing that tissue-specific organisation significantly exceeded stochastic expectations ($P < 0.001$). This non-random architecture was particularly pronounced in leaf networks relative to roots, as evidenced by comparative network entropy measures and module-level connectivity patterns. Sequential permutation testing further confirmed the robustness of these tissue-specific organisational principles across multiple analytical scales.

These findings establish that drought tolerance in G1 emerges from the precise temporal orchestration of tissue-specific molecular architectures rather than uniform stress responses across tissues. This mirrors observations by Filippou *et al.* (2021), who identified distinct but coordinated tissue-specific metabolic responses to stress—a phenomenon that underscores the importance of specialised, non-uniform mechanisms in stress adaptation. In G1, this evolved complementary strategy integrates rapid, controlled leaf responses (e.g., decreased primary metabolism) with sustained, modular root regulation (e.g., increased metabolic activity), collectively enabling robust whole-plant adaptation to water limitation. Conversely, G2's susceptibility to osmotic stress likely stems from its lack of pronounced tissue specialisation and failure to coordinate precise temporal responses across organs.

Conservation and Validation

To rigorously validate the architectural principles governing tissue-specific molecular responses under osmotic stress, we implemented a multi-tiered validation framework incorporating permutation testing, cross-validation analyses, and Bayesian network modelling. This comprehensive approach confirmed that the observed asymmetry in metabolic network architectures between roots and leaves represents a conserved adaptive mechanism rather than stochastic variation.

Validation of Tissue-Specific Molecular Architectures

We implemented a multi-tiered validation framework to confirm that the previously described tissue-specific architectures represent evolved adaptive strategies rather than stochastic outcomes. In the drought-tolerant genotype (G1), pathway-level analysis revealed significant stress-responsive enrichment (Kruskal-Wallis test, adjusted $P < 0.05$), supporting the architectural distinctions detailed above.

Feature selection via PLS-DA provided additional validation, identifying differential stress responses between tissues (62% root features versus 38% leaf features exceeding VIP > 1). Bayesian network analysis further confirmed the non-random organisation in both tissues (effect sizes: leaves = 25.87, roots = 18.75; $P < 0.001$), with leaf networks showing stronger deviation from null models (493 observed vs 105 expected edges) compared to roots (406 observed vs 106 expected edges).

Coefficient of variation analysis reinforced these tissue-specific patterns, demonstrating higher network integration in leaves ($CV = 0.96$, modularity = 2.65) versus roots ($CV = 0.89$, modularity = 2.06). This architectural distinction was quantitatively supported by differential response distributions: 9.4% of leaf metabolites (52/556) versus 2.5% of root metabolites (23/936) showed significant responses.

These validation analyses collectively confirm that the observed architectural asymmetry represents genuine biological adaptation rather than statistical artifact, with each tissue maintaining distinct organisational principles optimised for their specific roles in stress response.

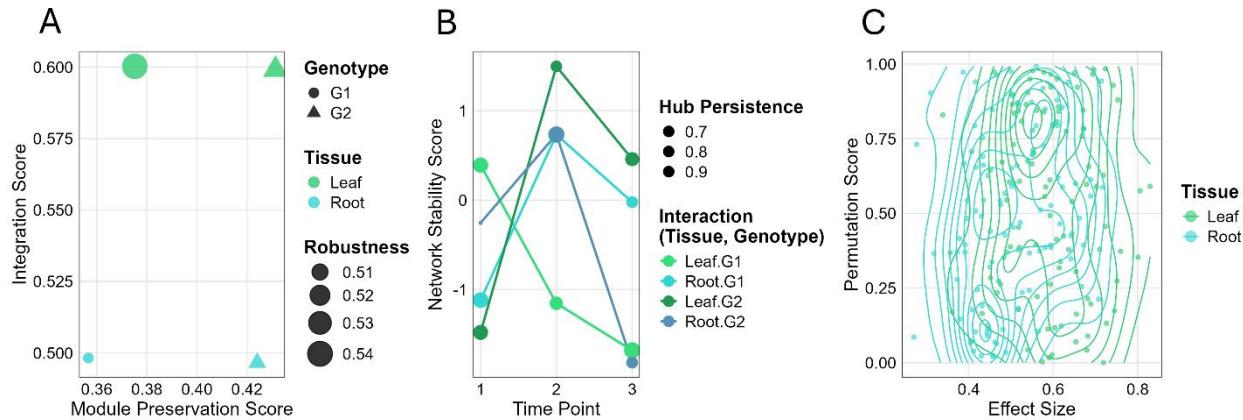


Fig. 3. Multi-level validation of tissue-specific network architectures.

A, Module-level validation through permutation tests ($n = 5,000$, $P < 0.001$) show G1 leaf networks maintain higher integration (0.600) and distinct module organisation (preservation ≈ 0.375) compared to roots (0.497, 0.357), underscoring the stable, adaptive nature of leaf network topologies. Node size (0.504–0.542) indicates network robustness. B, Network stability analysis across time points reveals G1's strategic leaf network remodelling under stress (stability: 0.392 to -1.679; hub persistence: 0.925–0.984), while G2 exhibits oscillating patterns in both tissues, indicating lack of fine-tuning capacity for stress adaptation. C, Permutation-based analyses confirm that observed architectures strongly deviate from random expectations (Benjamini-Hochberg adjusted $P < 0.001$). Leaf networks consistently show stronger effect sizes (Leaf, mean=0.547, CI: 0.521–0.573 vs root, mean=0.396, CI: 0.372–0.420) and tighter clustering compared to roots, highlighting tissue-specific fundamental organisational principles for osmotic stress adaptation.

Statistics: Bayesian analysis with 5,000 bootstrap iterations; FDR correction; sliding window analysis (window=3, step=1).

Network Properties and Temporal Dynamics

Temporal analyses further validated the tissue-specific network architectures by highlighting distinct dynamic patterns under osmotic stress. Leaf networks consistently exhibited higher temporal coherence (0.749–0.782) and shorter mean path lengths (2.04–2.17), reflecting a tightly integrated, rapidly responsive architecture. In contrast, root networks displayed lower coherence (0.686–0.716) and longer path lengths (2.29–2.44), indicative of more modular, locally adaptable structures. These contrasts, derived from correlation-based analyses and Bayesian network modelling, reinforce the notion of complementary tissue-specific strategies.

Quantitative measures further supported these architectural distinctions. Coefficient of variation analysis showed leaves to have higher variability ($CV = 0.96$) and modularity (2.65) compared to roots

(CV = 0.89, modularity = 2.06). False discovery rate (FDR) analysis underscored this differentiation, with a substantially greater proportion of leaf metabolites (9.4%, 52/556) exhibiting significant responses than root metabolites (2.5%, 23/936). Collectively, these results validate that leaves and roots leverage fundamentally different, yet evolutionarily conserved, molecular architectures to achieve robust, tissue-specific resilience under osmotic stress.

Architectural Conservation and Hub Connectivity

Building upon the network density patterns described above, our analysis of hub conservation confirmed tissue-specific organisation. Analysis of hub distribution demonstrated that the previously observed higher leaf network integration manifests through concentrated high-connectivity hubs (top 20 hubs averaging 872.05 for L-G1 and 805.3 for L-G2), while root networks exhibited a more distributed pattern (averaging 767.35 for R-G1 and 620.25 for R-G2).

This architectural distinction was further validated by comprehensive hub analysis (Kruskal-Wallis test, $H = 702.44$, $P = 6.20e-152$), revealing how tissue-specific network properties emerge from fundamentally different hub organisations. The drought-tolerant genotype G1 demonstrated more gradual connectivity decay patterns, particularly in roots, suggesting that enhanced stress resilience emerges not just from overall network structure but from optimised hub distribution patterns.

Bayesian Network Analysis and Structural Validation

Bayesian network analysis provided rigorous statistical validation through comprehensive null model comparisons (Fig. 3C). The non-random organisation of molecular networks was quantitatively validated against null expectations, with leaf networks demonstrating substantially stronger deviation from the null model (effect size: 25.87; observed edges: 493 vs. expected: 105; $P < 0.001$) compared to root networks (effect size: 18.75; observed edges: 406 vs. expected: 106; $P < 0.001$).

Detailed structural analyses incorporating first- and second-order network properties revealed tissue-specific differences in network diameter (leaves: 13; roots: 11) and mean path length (leaves: 6.40; roots: 5.05), while maintaining comparable hierarchical organisation through similar negative assortativity coefficients (roots: -0.102; leaves: -0.120). This conservation of hierarchical properties, combined with tissue-specific architectural differences, validates that the observed asymmetry represents fundamental organisational principles rather than technical artifacts.

The statistical robustness of these architectural distinctions was validated through comprehensive testing (Kruskal-Wallis test, $H = 702.44$, $P = 6.20 \times 10^{-152}$), confirming the non-random organisation of molecular networks across tissues. Multiple model configurations were evaluated, with optimal models selected based on rigorous criteria balancing prediction accuracy and modelled bias. Cross-validation analysis demonstrated high predictive accuracy with tissue-specific mean squared errors (leaf: 0.0622 ± 0.0256 ; root: 0.0268 ± 0.0268). Structural analysis revealed distinct organisational patterns through coefficient of variation (leaf: 0.96, root: 0.89) and modularity metrics (leaf: 2.65, root: 2.06), quantitatively validating the tissue-specific architectural differences.

Importantly, the tissue-specific architectural differences and temporal coordination patterns were consistent across both stress protocols, confirming that our findings reflect fundamental genotype-specific mechanisms rather than protocol-dependent effects.

Conclusion

Our study of contrasting drought-tolerant and -susceptible genotypes provides compelling evidence that drought tolerance in wheat is underpinned by a fundamental architectural asymmetry in tissue-specific metabolic networks. Through integrated metabolomic-network analysis, we demonstrate that drought-tolerant genotypes (G1) employ evolutionarily optimised, tissue-specific architectures: leaves maintain densely integrated networks—exhibiting 40% higher density than roots (0.354 versus 0.192) and higher transitivity (0.740-0.804 versus 0.686-0.714)—with concentrated hub connectivity (top 20 hubs averaging 872.05). In contrast, roots deploy fragmented, modular structures (modularity: 0.213-0.288; 18-21 network components) with distributed hub organisation (averaging 767.35), optimised for localised adaptation to

heterogeneous soil conditions. This architectural distinction was rigorously validated through comprehensive Bayesian network analysis (effect sizes: leaves=25.87, roots=18.75; $P < 0.001$) and permutation testing, confirming these tissue-specific properties represent fundamental organisational principles rather than stochastic variation.

Temporal analysis revealed a sophisticated biphasic adaptation strategy in G1. Initial strong cross-tissue coordination ($r = 0.546$) transitions to strategic decoupling ($r = 0.350$), characterised by precise temporal coherence patterns (leaves: 0.749–0.782; roots: 0.686–0.716) and redistribution of resilient molecular features favouring roots (13.00% versus 7.80% in leaves). In this process, leaves prioritise photosynthetic stability through controlled metabolic suppression (–69% flux change Time Points 1–2), while roots enhance stress sensing and adaptation (+26% Time Points 2–3). G1's controlled network attenuation—from 64.2% to 46.5%—contrasts sharply with the drought-susceptible genotype G2. Consistent with previous metabolomic analyses in drought-sensitive wheat (Kang et al., 2019), G2 shows unstable, oscillating integration (48.6% to 57.5%) and minimal differentiation in stress-persistent features (0.59% divergence). Aligning with previous studies that report minimal tissue differentiation in drought-sensitive wheat genotypes (Kang et al., 2019). This comparison suggests that drought susceptibility stems from compromised network regulation capacity, highlighting network stability as a key determinant of stress resilience.

These findings advance our understanding of plant stress adaptation in three key ways:

1. **Network Architecture Theory:** Establishes that drought tolerance emerges from complementary tissue-specific network architectures—integrated leaf networks for rapid photosynthetic coordination and modular root networks for precise environmental adaptation.
2. **Temporal Control Mechanism:** Identifies strategic network decoupling as essential for sustained adaptation, revealing how initial coordinated responses transition to specialised tissue-specific strategies under prolonged stress.
3. **Potential Breeding Applications:** Suggests that network properties (hub distribution, module stability, network coherence) could potentially serve as novel indicators for drought

tolerance, though field validation and practical implementation methods need to be developed. While these findings are based on two extensively validated genotypes, the consistent architectural patterns and statistical robustness ($n=7$ biological replicates, $P<0.001$) provide compelling evidence for this tissue-specific network hypothesis. Validation across broader genetic backgrounds represents a priority for future research.

While our findings align with recent work in *Arabidopsis* (Lozano-Elena et al., 2022b) highlighting tissue-divergent drought responses, they reveal that network plasticity—manifested through dynamic scaling of network topologies with stress duration—may play a key role in crop resilience. As climate change intensifies agricultural water stress, these architectural insights, supported by putative chemical classification of key network nodes (see Supplementary Fig. S3-S6 for detailed pathway-level analyses), provide a conceptual framework for understanding drought tolerance mechanisms and metabolic resilience under osmotic stress. Future research should focus on: (1) integrating these architectural insights with detailed metabolite identification to understand the functional significance of network differences; (2) validating these network patterns under diverse field conditions and across multiple growing seasons; (3) integrating transcriptomic and proteomic data to elucidate the regulatory mechanisms underlying these network differences; and (4) investigating whether these network properties are heritable and can be meaningfully incorporated into breeding programs. By improving our understanding of how tolerant genotypes organise their molecular responses to stress, this work opens new avenues for research into crop adaptation to water-limited environments—an increasingly important challenge for global agriculture.

MATERIALS AND METHODS

Plant Growth Conditions and Stress Treatments

Genotype selection leveraged extensively documented drought tolerance differences [ART citation], with seven replications used to ensure statistical power within this focused genotypic framework, where genotypes G1 demonstrated superior drought adaptation across multiple metrics including

water retention, gas exchange parameters, and clustering analysis ($p < 0.0001$, Cliff's Delta effect sizes confirming tolerance differences).

Plants were cultivated in an aerated hydroponic system under a 22 h light/2 h dark cycle (22°C during light; 17°C during dark) with 60–70% humidity (Ghosh et al., 2018). The pH was maintained at 6.5 through daily monitoring, and the nutrient solution was replaced every three days during the first four weeks and daily thereafter. Initially, plants of each genotype were grown together in a communal bucket; however, one day prior to treatment, they were transferred to individual buckets to ensure accurate application.

Two osmotic stress protocols were applied ($n=7$ biological replicates each) to capture natural drought variability: Batch 1 received acute stress (0.3M sorbitol, days 38-41) while Batch 2 underwent prolonged stress (0.15M sorbitol, days 37-46) (Supplementary Table S2). Both followed identical pre-treatment protocols, with Time points 1, 2, and 3 corresponding to days 1, 2, and 3 for Batch 1, and days 3, 6, and 9 for Batch 2, respectively. This dual-protocol approach prioritised conserved adaptation signatures transcending specific stress intensities, with primary conclusions validated across both regimes.

Sample preparation

Root and leaf samples were harvested, quenched in liquid Nitrogen and stored at -80°C until further analysis. Samples were freeze-dried using a Christ Alpha 1-4 LDplus (Martin Christ, Germany), with roots in perforated Eppendorf tubes and leaves arranged on racks in opened Ziplock bags. Freeze-dried samples were then ground using a Geno/Grinder® 2010 (Cole-Parmer, Illinois, USA) at 1200 rpm for 4 minutes, with 30-second breaks each minute. For untargeted analysis, 4 ceramic balls (3.5-4.1 mm) were used per sample, while targeted analysis employed 6 balls (two 4.1-4.8 mm and four 2.8-3.2 mm). After grinding, 20 mg of material was extracted with 1 mL of 80% methanol, shaken for 5 minutes, sonicated for 10 minutes, shaken again for 10 minutes, and incubated at room temperature for one hour. The extracts were then centrifuged at 12,300 rpm for 10 minutes, and 100 µL of the supernatant was aliquoted into HPLC vials (Rochfort et al., 2008; Roessner et al., 2001).

LC-MS Analysis

Metabolite separation and analysis were performed using a Vanquish UHPLC system coupled with a Q Exactive Plus Orbitrap mass spectrometer (Thermo Scientific). Chromatographic separation occurred on a C18 column (2.1 × 100 mm, 1.7 µm) at 30°C, using water (A) and acetonitrile (B), both with 0.1% formic acid, as mobile phases (Rochfort *et al.*, 2008). The gradient (0.3 mL/min) progressed from 2% to 100% B over 11 minutes, held for 4 minutes, then returned to initial conditions (5-minute re-equilibration). Mass spectrometric detection employed heated electrospray ionisation in positive and negative modes (spray voltages: 3600 V and 3300 V; capillary temperature: 300°C; sheath gas: 28 arbitrary units; S-lens RF: 64). Full MS scans (m/z 100-1500) were acquired at 70,000 FWHM resolution (m/z 200), with an injection volume of 5 µL. Internal standards were included in each sample to ensure accurate quantification of metabolites and monitor system stability. Additionally, pooled QC samples were regularly inserted into the analytical run to assess and correct for technical variability and ensure consistent performance across the samples (Dunn *et al.*, 2011; Gika *et al.*, 2007).

LC-MS Data Processing and Quality Control

Initial Data Processing

Raw LC-MS data were processed using Genedata Expressionist® Refiner MS 18.0.1. The workflow comprised data import, chromatogram extraction, chemical noise subtraction (via a moving average algorithm), retention time alignment (using a pairwise alignment-based tree method), peak detection (employing a curvature-based algorithm with subsequent refinement and consistency filtering), and isotope clustering (using a peptide isotope shaping method). Separate mode-specific workflows were applied for positive and negative ionisation. Detailed parameters for each step are provided in Supplementary Table S3.

Data Quality Control and Feature Selection

Data preprocessing followed a rigorous multi-step workflow to ensure data quality and reliability (Supplementary Fig. S1). Initial data filtering excluded columns with fewer than three replicates,

reducing molecular features from 4,255 to 1,789 in negative mode and from 3,199 to 1,350 in positive mode. Despite this reduction, missing values persisted, a common issue in LCMS studies (Karpievitch et al., 2012; Kokla et al., 2019). Missing values, which followed a Missing at Random (MAR) pattern, were imputed using a Random Forest approach; imputation quality was validated with Earth Mover's Distance and Hellinger Distance metrics. Outlier detection employed Isolation Forest algorithm, followed by asinh transformation which effectively reduced data variability (CV from 0.876 to 0.206). The final dataset comprised 2,471 molecular features (1,398 from negative mode, 1,073 from positive mode), with tissue-specific distributions detailed in Supplementary Information.

Network Analysis and Statistical Methods

Initial Statistical Analysis

The analytical framework employed complementary approaches addressing distinct biological questions (Supplementary Fig. S7). Spearman correlations identified co-regulated metabolite pairs, network topology quantified architectural principles, and Bayesian structure learning inferred directional dependencies. This layered strategy prevented single-method bias while providing cross-validation of core findings through independent statistical approaches.

Statistical analysis comprised six complementary approaches: (1) tissue comparison (Mann-Whitney U tests), (2) genotype comparison (Cliff's Delta effect sizes), (3) treatment effects (Wilcoxon/Mann-Whitney tests), (4) temporal dynamics (Friedman tests), (5) integrated responses (rank-based analysis), and (6) resilience indices (bootstrap validation, n=5,000). Multiple testing correction via Benjamini-Hochberg FDR ($\alpha=0.05$). Relationship stability was assessed through bootstrap correlation analysis (5,000 iterations) with 95% confidence intervals. Effect sizes were calculated using Cliff's Delta, and statistical power was evaluated through non-parametric simulations (5,000 iterations, $\alpha = 0.05$).

For genotype-specific analyses, tissues were evaluated separately using a comprehensive pipeline. Temporal patterns were assessed using Friedman tests, with Kendall's W employed for effect size estimation. Multiple testing corrections were implemented using both Benjamini-Hochberg FDR and

Bonferroni FWER methods ($\alpha = 0.05$), with significance determined by FDR-adjusted P-values < 0.05 . Tissue-specific responses were evaluated through Mann-Whitney U tests, with P-values adjusted within each tissue group to control for multiple comparisons. Tissue structure analysis employed coefficient of variation metrics and modularity assessment.

Feature Selection

Feature selection used PLS-DA with nested cross-validation (outer: 10-fold, inner: 5-fold) and VIP scoring (threshold > 1.0), implemented via custom ScalerPLSPipeline in Python 3.10. First, Partial Least Squares Discriminant Analysis (PLS-DA) was performed separately for leaf and root tissues using a nested cross-validation framework (outer CV: 10-fold; inner CV: 5-fold) with stratified sampling. Model optimisation employed grid search across scaling methods (Standard, MinMax, Robust) and components (1-10), using mean squared error as the optimisation metric. A custom pipeline incorporating ScalerPLSPipeline was implemented to ensure consistent preprocessing across validation folds. Features were ranked using Variable Importance in Projection (VIP) scores, with scores > 1.0 considered significant, following the established threshold criterion in metabolomics study (Chong and Jun, 2005).

Feature significance was further validated through non-parametric Mann-Whitney U tests comparing metabolite distributions between stressed and control conditions along with detailed metabolite characterisation (see Supplementary Information 'Metabolite Analysis Results' section for comprehensive temporal and pathway-level analyses). Multiple testing correction employed both Bonferroni correction for family-wise error rate control and Benjamini-Hochberg procedure for FDR control ($\alpha = 0.05$). The final feature set comprised molecular features that exceeded the VIP threshold and passed both multiple testing corrections. These validated features were annotated with tissue-specific identifiers and compiled into a unified dataset for subsequent network analyses.

Network Construction and Topology Analysis

Network architecture was characterised through key topological metrics including network density, transitivity, modularity, component analysis, and mean path length calculations (see Supplementary Table S1 for detailed definitions and biological interpretations of all network metrics used). Networks

were constructed using Spearman correlations ($|r| > 0.7$, FDR < 0.05) via NetworkX v2.8.4. Community detection used Louvain algorithm; Bayesian validation performed with bnlearn R package (5,000 bootstrap iterations) (Toubiana and Maruenda, 2021). Network architecture was characterised through five key topological metrics: (1) network density, calculated as the ratio of observed to possible edges; (2) transitivity, quantifying local connectivity clustering; (3) modularity, evaluated using the Louvain community detection algorithm; (4) component analysis to assess network fragmentation; and (5) mean path length calculations for connected components.

Hub identification employed degree centrality analysis with distribution patterns analysed through power-law fitting. Network efficiency was assessed through shortest path length calculations between node pairs and temporal coherence evaluation. For disconnected networks, metrics were calculated for the largest connected component.

Statistical validation incorporated three approaches: (1) permutation testing (5,000 randomisations) with Benjamini-Hochberg correction; (2) hub connectivity pattern analysis using LOWESS smoothing (fraction=0.3) (Clauset et al., 2009; Cleveland, 1979); and (3) tissue-specific network comparisons using non-parametric tests. Module-level organisation was evaluated through coefficient of variation and modularity metrics.

Temporal and Cross-tissue Network Dynamics

Network dynamics were analysed across three dimensions: temporal evolution, cross-tissue coordination, and stability. Temporal patterns were assessed through non-parametric tests across sequential time points and sliding window analysis (window=3, step=1) (Khan et al., 2010), with network coherence quantified via Kendall's W coefficient. Cross-tissue relationships were evaluated through correlation analysis with bootstrap validation ($n=5,000$), stability indices, and information flow calculations. Network stability was measured using moving windows to track hub persistence and module preservation, validated through permutation testing ($n=10,000$) with FDR correction. For fragmented networks, analyses focused on the largest connected component with size normalisation.

Bayesian Network Validation

Network structure was inferred using hill-climbing algorithms with directed acyclic graph constraints. Structure learning employed a non-parametric bootstrap approach (5,000 iterations) with edge screening threshold set at 0.5. Limiting the maximum number of parents per node is a recognised strategy to manage computational complexity in Bayesian network structure learning (Campos and Ji, 2011). In this study, a threshold of 3 parents per node was set to balance model complexity and computational feasibility.

Validation incorporated three sequential steps: (1) null model comparisons through permutation testing while preserving tissue-specific temporal structure; (2) module preservation analysis using standardised preservation statistics and density scores; and (3) cross-tissue comparisons using empirically derived p-values from bootstrap distributions.

Network confidence was assessed through dual validation: bootstrap analysis for edge reliability followed by permutation testing (5,000 iterations) for stability assessment. Community organisation was determined using the Louvain algorithm with multi-level optimisation. For disconnected components, analyses were performed on the largest subgraph with appropriate size normalisation.

Software and Statistical Environment

All analyses were performed in Python 3.8 (using pandas, scikit-learn, NetworkX) and R 4.1.0 (using missForest, bnlearn, igraph). The complete preprocessing pipeline, custom analysis scripts, and data are publicly available on GitHub at: <https://github.com/shoaibms/metabo>.

FUNDING

This work was supported by Agriculture Victoria, an agency of the Department of Energy, Environment and Climate Action (DEECA) within the Victorian State Government, Australia.

ACKNOWLEDGEMENT

The authors thank Emily L. Thoday-Kennedy for her essential technical support during the experimental phase of this project.

AUTHOR CONTRIBUTIONS

M.S. conceived and designed the study, conducted all experiments, performed data analysis, developed computational methods, and wrote the manuscript. S.J.R. and P.R. contributed to LC-MS methodology, metabolite analysis protocols, and data interpretation. D.R. contributed to sample processing. S.K. provided guidance on experimental design and supervised the project. All authors reviewed and approved the final manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

DATA AVAILABILITY

The processed metabolomics data, analysis scripts, and figures generated during this study are publicly available in a GitHub repository (<https://github.com/shoaibms/metabo>). The raw mass spectrometry data have been deposited in the EMBL-EBI MetaboLights repository under accession number MTBLSXXXX.

REFERENCES

- Albert, R., Jeong, H., and Barabási, A.-L.** (2000). Error and attack tolerance of complex networks. *Nature* **406**:378-382. 10.1038/35019019.

- Alon, U.** (2007). Network motifs: theory and experimental approaches. *Nature Reviews Genetics* **8**:450-461. 10.1038/nrg2102.
- Barabási, A.-L., and Albert, R.** (1999). Emergence of Scaling in Random Networks. *Science* **286**:509-512. doi:10.1126/science.286.5439.509.
- Barabási, A.-L., and Oltvai, Z.N.** (2004). Network biology: understanding the cell's functional organization. *Nature Reviews Genetics* **5**:101-113. 10.1038/nrg1272.
- Bennani, S., Nsarellah, N., Jlibene, M., and Tadesse, W.** (2017). Efficiency of drought tolerance indices under different stress severities for bread wheat selection. *Australian Journal of Crop Science* **11**:395-405. 10.21475/ajcs.17.11.04.pne272.
- Campos, C.P.d., and Ji, Q.** (2011). Efficient Structure Learning of Bayesian Networks using Constraints. *J. Mach. Learn. Res.* **12**:663–689.
- Cardoso, L.L., Freire, F.B.S., and Daloso, D.M.** (2023). Plant metabolic networks under stress: a multi-species/stress condition meta-analysis. *Journal of Soil Science and Plant Nutrition* **23**:4-21. <https://doi.org/10.1007/s42729-022-01032-2>.
- Chaves, M.M., Flexas, J., and Pinheiro, C.** (2008). Photosynthesis under drought and salt stress: regulation mechanisms from whole plant to cell. *Annals of Botany* **103**:551-560. 10.1093/aob/mcn125.
- Chen, H., and Jiang, J.-G.** (2010). Osmotic adjustment and plant adaptation to environmental changes related to drought and salinity. *Environmental Reviews* **18**:309-319.
- Chiang, F., Mazdiyasni, O., and AghaKouchak, A.** (2021). Evidence of anthropogenic impacts on global drought frequency, duration, and intensity. *Nature Communications* **12**:2754. 10.1038/s41467-021-22314-w.
- Chong, I.-G., and Jun, C.-H.** (2005). Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems* **78**:103-112. <https://doi.org/10.1016/j.chemolab.2004.12.011>.
- Clark, T.J., Guo, L., Morgan, J., and Schwender, J.** (2020). Modeling plant metabolism: from network reconstruction to mechanistic models. *Annual Review of Plant Biology* **71**:303-326.
- Clauset, A., Shalizi, C.R., and Newman, M.E.** (2009). Power-law distributions in empirical data. *SIAM review* **51**:661-703.

Cleveland, W.S. (1979). Robust locally weighted regression and smoothing scatterplots. Journal of the American statistical association **74**:829-836.

Dunn, W.B., Broadhurst, D., Begley, P., Zelena, E., Francis-McIntyre, S., Anderson, N., Brown, M., Knowles, J.D., Halsall, A., Haselden, J.N., et al. (2011). Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. Nature Protocols **6**:1060-1083. 10.1038/nprot.2011.335.

Erdos, P., and Rényi, A. (1960). On the evolution of random graphs. Publ. math. inst. hung. acad. sci **5**:17-60.

Fait, A., Batushansky, A., Shrestha, V., Yobi, A., and Angelovici, R. (2020). Can metabolic tightening and expansion of co-expression network play a role in stress response and tolerance? Plant Science **293**:110409. 10.1016/j.plantsci.2020.110409.

Fernie, A.R., and Stitt, M. (2012). On the Discordance of Metabolomics with Proteomics and Transcriptomics: Coping with Increasing Complexity in Logic, Chemistry, and Network Interactions Scientific Correspondence. Plant Physiology **158**:1139-1145. 10.1104/pp.112.193235.

Fernie, A.R., and Tohge, T. (2017). The genetics of plant metabolism. Annual Review of Genetics **51**:287-310.

Filippou, P., Zarza, X., Antoniou, C., Obata, T., Villarroel, C.A., Ganopoulos, I., Harokopos, V., Gohari, G., Aidinis, V., Madesis, P., et al. (2021). Systems biology reveals key tissue-specific metabolic and transcriptional signatures involved in the response of plant genotypes to salt stress. Computational and Structural Biotechnology Journal **19**:2133-2147. 10.1016/j.csbj.2021.04.018.

Fortunato, S. (2010). Community detection in graphs. Physics reports **486**:75-174.

Gargallo-Garriga, A., Sardans, J., Pérez-Trujillo, M., Rivas-Ubach, A., Oravec, M., Vecerova, K., Urban, O., Jentsch, A., Kreyling, J., Beierkuhnlein, C., et al. (2014). Opposite metabolic responses of shoots and roots to drought. Scientific Reports **4**:6829. 10.1038/srep06829.

- Ghosh, S., Watson, A., Gonzalez-Navarro, O.E., Ramirez-Gonzalez, R.H., Yanes, L., Mendoza-Suárez, M., Simmonds, J., Wells, R., Rayner, T., Green, P., et al.** (2018). Speed breeding in growth chambers and glasshouses for crop breeding and model plant research. *Nature Protocols* **13**:2944-2963. 10.1038/s41596-018-0072-z.
- Gika, H.G., Theodoridis, G.A., Wingate, J.E., and Wilson, I.D.** (2007). Within-day reproducibility of an HPLC-MS-based method for metabonomic analysis: application to human urine. *J Proteome Res* **6**:3291-3303. 10.1021/pr070183p.
- Girvan, M., and Newman, M.E.J.** (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences* **99**:7821-7826. doi:10.1073/pnas.122653799.
- Grafahrend-Belau, E., Junker, B.H., Klukas, C., Koschützki, D., Schreiber, F., and Schwöbbermeyer, H.** (2009). Topology of plant metabolic networks. *Plant metabolic networks*:173-209.
- Hodge, A.** (2004). The plastic plant: root responses to heterogeneous supplies of nutrients. *New Phytologist* **162**:9-24. 10.1111/j.1469-8137.2004.01015.x.
- Hone, H., Mann, R., Yang, G., Kaur, J., Tannenbaum, I., Li, T., Spangenberg, G., and Sawbridge, T.** (2021). Profiling, isolation and characterisation of beneficial microbes from the seed microbiomes of drought tolerant wheat. *Scientific Reports* **11**:11916. 10.1038/s41598-021-91351-8.
- Hsiao, T.C., and Xu, L.K.** (2000). Sensitivity of growth of roots versus leaves to water stress: biophysical analysis and relation to water transport. *Journal of experimental botany* **51**:1595-1616.
- Jeong, H., Mason, S.P., Barabási, A.-L., and Oltvai, Z.N.** (2001). Lethality and centrality in protein networks. *Nature* **411**:41-42.
- Kang, Z., Babar, M.A., Khan, N., Guo, J., Khan, J., Islam, S., Shrestha, S., and Shahi, D.** (2019). Comparative metabolomic profiling in the roots and leaves in contrasting genotypes reveals complex mechanisms involved in post-anthesis drought tolerance in wheat. *PLOS ONE* **14**:e0213502. 10.1371/journal.pone.0213502.

Karpievitch, Y.V., Dabney, A.R., and Smith, R.D. (2012). Normalization and missing value imputation for label-free LC-MS analysis. *BMC Bioinformatics* **13**:S5. 10.1186/1471-2105-13-s16-s5.

Khan, M.S., Coenen, F., Reid, D., Patel, R., and Archer, L. (2010). A sliding windows based dual support framework for discovering emerging trends from temporal data. *Knowledge-Based Systems* **23**:316-322. 10.1016/j.knosys.2009.11.005.

Kokla, M., Virtanen, J., Kolehmainen, M., Paananen, J., and Hanhineva, K. (2019). Random forest-based imputation outperforms other methods for imputing LC-MS metabolomics data: a comparative study. *BMC Bioinformatics* **20**10.1186/s12859-019-3110-0.

Langfelder, P., Luo, R., Oldham, M.C., and Horvath, S. (2011). Is my network module preserved and reproducible? *PLoS computational biology* **7**:e1001057.

Latora, V., and Marchiori, M. (2001). Efficient behavior of small-world networks. *Physical review letters* **87**:198701.

Li, H., Testerink, C., and Zhang, Y. (2021). How roots and shoots communicate through stressful times. *Trends in Plant Science* **26**:940-952. 10.1016/j.tplants.2021.03.005.

Lozano-Elena, F., Fàbregas, N., Coleto-Alcudia, V., and Caño-Delgado, A.I. (2022a). Analysis of metabolic dynamics during drought stress in *Arabidopsis* plants. *Scientific Data* **9**:90. 10.1038/s41597-022-01161-4.

Lozano-Elena, F., Fàbregas, N., Coleto-Alcudia, V., and Caño-Delgado, A.I. (2022b). Analysis of metabolic dynamics during drought stress in *Arabidopsis* plants. *Scientific Data* **9**10.1038/s41597-022-01161-4.

Lozano-Elena, F., Fàbregas, N., Coleto-Alcudia, V., and Caño-Delgado, A.I. (2022c). Analysis of metabolic dynamics during drought stress in *Arabidopsis* plants. *Sci Data* **9**:90. 10.1038/s41597-022-01161-4.

Miao, Z., Han, Z., Zhang, T., Chen, S., and Ma, C. (2017). A systems approach to a spatio-temporal understanding of the drought stress response in maize. *Scientific Reports* **7**:6590. 10.1038/s41598-017-06929-y.

- Min, X., Lin, X., Ndayambaza, B., Wang, Y., and Liu, W.** (2020). Coordinated mechanisms of leaves and roots in response to drought stress underlying full-length transcriptome profiling in *Vicia sativa* L. *BMC Plant Biology* **2010**.1186/s12870-020-02358-8.
- Mintz-Oron, S., Meir, S., Malitsky, S., Ruppin, E., Aharoni, A., and Shlomi, T.** (2012). Reconstruction of *Arabidopsis* metabolic network models accounting for subcellular compartmentalization and tissue-specificity. *Proc Natl Acad Sci U S A* **109**:339-344. 10.1073/pnas.1100358109.
- Munns, R., and Tester, M.** (2008). Mechanisms of salinity tolerance. *Annu. Rev. Plant Biol.* **59**:651-681.
- Nakabayashi, R., and Saito, K.** (2015). Integrated metabolomics for abiotic stress responses in plants. *Curr Opin Plant Biol* **24**:10-16. 10.1016/j.pbi.2015.01.003.
- Newman, M.E.** (2002). Assortative mixing in networks. *Physical review letters* **89**:208701.
- Newman, M.E. (2010). Networks: an introduction. Oxford university press.
- Newman, M.E.J.** (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences* **103**:8577-8582. doi:10.1073/pnas.0601602103.
- Obata, T., and Fernie, A.R.** (2012). The use of metabolomics to dissect plant responses to abiotic stresses. *Cell Mol Life Sci* **69**:3225-3243. 10.1007/s00018-012-1091-5.
- Rochfort, S.J., Trenerry, V.C., Imsic, M., Panozzo, J., and Jones, R.** (2008). Class targeted metabolomics: ESI ion trap screening methods for glucosinolates based on MS_n fragmentation. *Phytochemistry* **69**:1671-1679.
- Roessner, U., Luedemann, A., Brust, D., Fiehn, O., Linke, T., Willmitzer, L., and Fernie, A.** (2001). Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell* **13**:11-29. 10.1105/tpc.13.1.11.
- Sévin, D.C., Stählin, J.N., Pollak, G.R., Kuehne, A., and Sauer, U.** (2016). Global Metabolic Responses to Salt Stress in Fifteen Species. *PLOS ONE* **11**:e0148888. 10.1371/journal.pone.0148888.
- Shoaib, M., Dimech, A.M., Rochfort, S.J., Topp, C., Hayden, M.J., and Kant, S.** (2025). Seeing the Unseen: A Novel Approach to Extract Latent Plant Root Traits from Digital Images. *Plant Phenomics*:100088. <https://doi.org/10.1016/j.plaphe.2025.100088>.

Chapter 6. MOFA+ Transformer: The Ultimate Integration Framework

The Pinnacle: Solving the Systems Integration Bottleneck Through Cross-Modal Coordination

Our expedition has systematically addressed three of the four analytical bottlenecks: automation (AutoML), phenotyping precision (ART), and mechanistic understanding (networks). This final chapter confronts the systems integration bottleneck (Section 1.3.3)—the challenge of quantifying how different biological layers communicate and coordinate their responses across tissues and time.

The final frontier is to understand how these systems communicate. Real-time adaptation emerges from the dynamic dialogue between a plant's physiological status and its biochemical state. This chapter introduces our most sophisticated analytical achievement—the MOFA+ Transformer—a framework designed to listen to, and quantify, that dialogue.

This work directly addresses Research Question 3: What time-resolved coordination patterns distinguish tolerant and susceptible genotypes, and can interpretable deep learning reveal these cross-modal relationships?

The MOFA+ Transformer transcends simple correlation by integrating hyperspectral and metabolomic data to quantify directed, feature-specific relationships through a cross-modal attention mechanism.

The framework reveals the temporal choreography of resilience: tolerant genotypes exhibited 355% greater Coordination Strength in leaf tissue, establishing critical networks at demonstrably earlier stress stages. It identified molecular command centres—MF N_1909 in leaves and MF N_0512 in roots—that translate physiological signals into metabolic action. Crucially, the framework also established that features driving biological variance were largely distinct from those driving predictive accuracy, a key methodological insight.

The framework's power and generalisability were confirmed beyond plant science. Applied to an independent human single-cell dataset, it discovered and statistically

validated a novel, non-linear link between cellular autofluorescence and the critical stress-related gene lncRNA NEAT1 ($p=0.0099$)—a connection that was not captured by standard correlation analysis. This demonstrates its broad capacity to generate mechanistic hypotheses in diverse biological systems.

The MOFA+ Transformer completes our analytical ecosystem, having systematically solved all four analytical bottlenecks identified in Chapter 1. It represents the pinnacle of our journey from traits to networks, from roots to shoots, and offers a new conceptual framework for understanding resilience in complex biological systems, one driven by integrated computational innovation.

Statement of contributions of joint authorship

Mirza Shoaib designed and conducted the experiment. Surya Kant guided in refining the experiments. Mirza Shoaib wrote the manuscript and developed the code. Priyanka Reddy guided during analysis. Fitzgerald G, Matthew J. Hayden, Surya Kant and all reviewed and edited the manuscript.

Statement from the co-author confirming the PhD candidate's contributions to the authorship

As a co-author of the manuscript:

Shoaib M, Reddy P, Fitzgerald G, Hayden M and Kant S (2025). MOFA+ Transformer: An Interpretable Deep Learning Framework for Dynamic, Feature-Specific Multi-Omics Integration. Submitted to Nature Methods, following a positive pre-submission inquiry.

I confirm that Mirza Shoaib has made the following contributions,

- Conducted the literature review
- Designed and conducted the experiments
- Performed all data analyses
- Generated all figures
- Led the writing of the manuscript



Dr Surya Kant

29 July 2025

MOFA+ Transformer: An Interpretable Deep Learning Framework for Dynamic, Feature-Specific Multi-Omics Integration

Authors

Mirza Shoaib^{1,2*}, Priyanka Reddy³, Glenn J Fitzgerald^{1,4}, Matthew J. Hayden^{2,5} and Surya Kant^{6*}

Affiliations

¹Agriculture Victoria, Grains Innovation Park, 110 Natimuk Road, Horsham, Victoria 3400, Australia

²School of Applied Systems Biology, La Trobe University, Bundoora, Victoria 3083, Australia

³ School of Chemistry, Bio21 Institute of Molecular Science and Biotechnology, The University of Melbourne, Victoria 3010 Australia

⁴Faculty of Science, The University of Melbourne, Parkville, VIC 3010, Australia

⁵Agriculture Victoria, AgriBio, Centre for AgriBioscience, 5 Ring Road, Bundoora, Victoria 3083, Australia

⁶Department of Ecological, Plant and Animal Science, School of Agriculture, Biomedicine & Environment, La Trobe University, Bundoora, Victoria 3083, Australia

* Correspondence: shoaib.mirza@agriculture.vic.gov.au and S.Kant@latrobe.edu.au

ORCID

Mirza Shoaib: <https://orcid.org/0000-0002-6773-6050>

Priyanka Reddy: <https://orcid.org/0000-0002-1504-1126>

Glenn J Fitzgerald: <https://orcid.org/0000-0001-6972-4443>

Matthew Hayden: <https://orcid.org/0000-0003-3117-5500>

Surya Kant: <https://orcid.org/0000-0001-6178-7036>

Abstract

Resolving the feature-specific and dynamic interactions that drive biological processes remains a key challenge in multi-omics analysis. Here, we present MOFA+ Transformer, a framework that synergises unsupervised variance decomposition (MOFA+) with a Transformer's cross-modal attention to quantify directed, feature-pair relationships and their temporal dynamics. Applied to plant drought stress, MOFA+ Transformer reveals that tolerance is an emergent property of the timing and architecture of spectral-metabolite coordination networks. The framework quantifies these dynamics, uncovering a strong predictive relationship between cross-modal attention and biological outcomes, a mechanistic insight not readily apparent using conventional methods. The framework also establishes that features driving variance are distinct from those driving prediction, providing complementary biological insights. We demonstrate its generalisability by discovering and validating ($p = 0.0099$) a non-linear link between cellular autofluorescence and the lncRNA NEAT1 in a single-cell dataset, an association not detected by standard correlation analyses. MOFA+ Transformer addresses the gap between correlational multi-omics and mechanistic understanding by delivering an interpretable, predictive framework that quantifies the temporal dynamics of feature-specific interactions. Ultimately, the framework provides an evidence-based path from high-dimensional data to specific, testable mechanistic hypotheses.

Main

The proliferation of multi-omics data promises to reveal how molecular networks orchestrate biological responses^{1,2}. Yet, extracting predictive, mechanistic insights from these vast and heterogeneous datasets remains a primary analytical challenge^{1,3}. A

suite of computational frameworks has been developed to address this complexity, including specialised multi-omics integration tools^{3,4}, but a systematic comparison reveals they share critical limitations that prevent deeper biological discovery (Table 1)^{5,6}. While existing methods can identify which features co-vary, they largely fail to resolve the directional, feature-specific, and dynamic interactions that govern complex processes like adaptation, development, and disease^{7,8}. By providing static snapshots, they overlook the crucial temporal coordination—the when and how of cross-modal communication, leaving a fundamental gap between data generation and the discovery of manipulable biological mechanisms^{9,10}.

Table 1. Comparison of multi-omics integration frameworks. MOFA+ Transformer (to our knowledge) is among the few frameworks that combines all four essential capabilities for mechanistic discovery: identifying shared variation, quantifying predictive associations, providing feature-pair granularity, and modeling temporal dynamics. This unique synergy enables the shift from descriptive analysis to the generation of predictive, testable hypotheses.

| | Identifies Shared Variation Patterns | Quantifies Predictive Associations | Provides Feature-Pair Granularity | Models Temporal Dynamics | Robustly Handles Data Heterogeneity | Generates Interpretable, Mechanistic Hypotheses |
|--|--------------------------------------|------------------------------------|-----------------------------------|--------------------------|-------------------------------------|---|
| MOFA+ Transformer ^{11,12} | ✓ | ✓ | ✓ | ✓ | ✓ | High |
| Correlation Analysis ¹³⁻¹⁵ | ✓ | x | ✓ | x | Limited | Low |
| PLS/CCA ^{16,17} | ✓ | x | x | x | Limited | Low |
| DIABLO (mixOmics) ^{6,18} | ✓ | x | x | Limited | Moderate | Moderate |
| Standard ML (e.g., RF, SVM) ^{19,20} | Implicitly | x | x | Limited | Moderate | Low |
| JIVE / iCluster ^{6,21,22} | ✓ | x | x | x | Moderate | Low |

Abbreviations: CCA, Canonical Correlation Analysis; DIABLO, Data Integration Analysis for Biomarker discovery using a Latent cOmponent method; JIVE, Joint and Individual Variation Explained; ML, Machine Learning; PLS, Partial Least Squares; RF, Random Forest; SVM, Support Vector Machine.

To address this gap, we developed MOFA+ Transformer, an interpretable deep learning framework that synergises unsupervised variance decomposition (MOFA+)^{11,12} with a Transformer's cross-modal attention mechanism^{23,24}. This design moves beyond simple

co-variation by quantifying directed, feature-pair relationships and their temporal dynamics. The success of Transformer architectures in biological sequence analysis^{25,26} provides strong precedent for their application to multi-omics integration challenges. Our framework thereby provides an evidence-based path from high-dimensional data to testable, mechanistic hypotheses, avoiding the 'black box' problem common to many machine learning applications^{4,20}.

We applied MOFA+ Transformer to dissect the multi-omics response to osmotic stress in wheat, a key component of drought tolerance^{27,28}. The analysis revealed that resilience is not merely a function of molecular abundance but emerges from the specific timing and architecture of coordination networks linking plant physiology and metabolism—a dynamic that is not captured by conventional methods^{29,30}. We further establish that features driving biological variance are distinct from those driving predictive accuracy, a critical distinction that allows for the simultaneous discovery of mechanistic drivers and robust biomarkers. The framework's broad applicability is validated on an independent single-cell dataset, where it uncovers a non-linear link between cellular autofluorescence and the long non-coding RNA (lncRNA) NEAT1³¹—an association not detected by conventional analyses.

By delivering interpretable, dynamic, and quantifiable biological knowledge, MOFA+ Transformer significantly advances multi-omics analysis. It provides a powerful framework for generating testable hypotheses and discovering actionable coordination patterns critical for systems biology, precision agriculture, and biotechnology.

Results

A Robust and Predictive Foundation for Mechanistic Discovery

To deconstruct the complex interplay between plant physiology and metabolism under osmotic stress, we first applied MOFA+ to four integrated data modalities (spectral and molecular features from both leaf and root tissues) derived from 336 distinct biological samples, computationally augmented to 1,344 samples, a process validated to

preserve data structure while enhancing statistical power. The framework successfully decomposed this high-dimensional data into 11 active latent factors (LFs) that capture distinct and interpretable sources of biological variation across the omics layers (Fig. 1a). This integrated latent space provides a structured link between physiological (spectral) and molecular (metabolomic) data (Fig. S1). The features driving these factors demonstrated exceptional stability, with an average selection frequency of 96.0% across 100 bootstrap iterations, confirming the reliability of the variance decomposition.

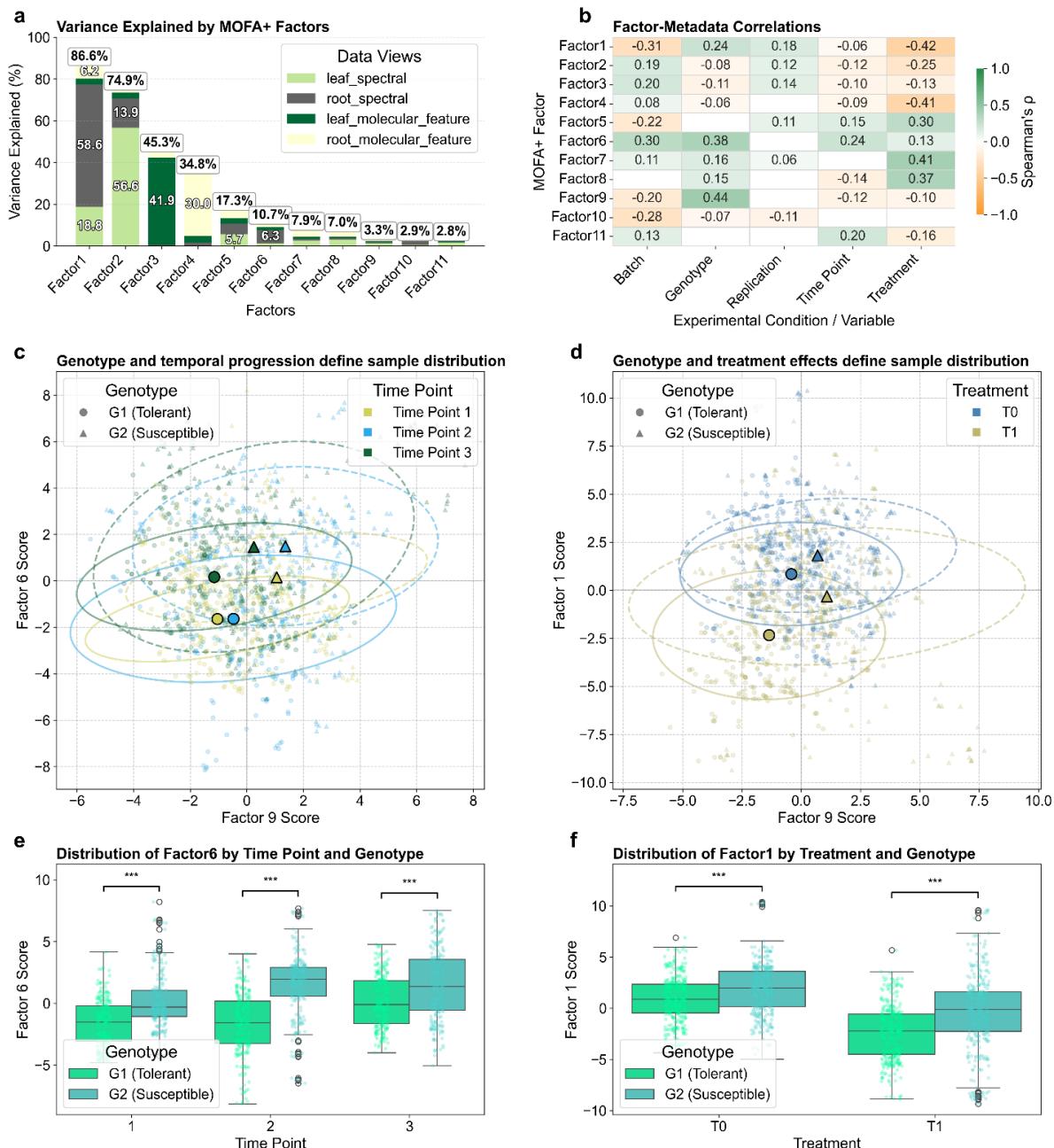


Fig. 1. MOFA+ Decomposes Multi-Omics Variance and Identifies Major Axes of Biological Variation. (a) Variance explained per factor across the four omics views, quantified by the percentage of total variance captured. Total variance explained by each factor is noted above the bars. (b) Biological annotation of latent factors (LFs) via Spearman's correlation with experimental metadata. Only statistically significant associations ($FDR < 0.05$) are shown, with cells coloured by the correlation coefficient (ρ). (c-d) Visualisation of sample distribution based on latent factor scores. Each point represents a sample. Shapes denote genotype, and colours represent either Time Point (c) or Treatment (d). Dashed ellipses represent 95% confidence intervals for each group, with larger solid markers indicating the group centroid. (c) Plot of scores for Factor 9 (associated with Genotype) versus Factor 6 (associated with Temporal progression). (d) Plot of scores for Factor 9 versus Factor 1 (associated with the Treatment/Batch effect). (e-f) Quantification of factor score distributions across conditions. Boxplots showing the distribution of Factor 6 scores across the three experimental time points (e) and Factor 1 scores across treatment conditions (f), separated by genotype. **Significance bars indicate statistically significant differences between genotypes within each condition (Mann-Whitney U test; $*p < 0.001$).

To interpret these factors, we assessed their biological relevance via Spearman's rank correlation with experimental metadata. This analysis revealed that MOFA+ effectively disentangled distinct biological signals; for instance, Factor 9 (LF9) was strongly associated with Genotype-specific variance (Spearman's $\rho = 0.44$, $FDR < 1.67e-63$), while Factor 6 (LF6) isolated the temporal progression of the stress response ($\rho = 0.24$, $FDR < 1.39e-18$). Factor 1 (LF1) primarily captured the treatment effect and isolated technical variance from experimental batches (Fig. 1b). The biological significance of these factors was further evidenced by the clear separation of samples in the latent space (Fig. 1c-d) and statistically significant shifts in factor score distributions across conditions (Fig. 1e-f). Together, this analysis establishes that the latent factors identified in Figure 1 serve as robust variables for the primary axes of biological variation.

To provide molecular context for these factors, we examined their underlying feature weights (Supplementary Fig. S1). This analysis revealed that genotype differentiation (LF9) was driven by specific metabolic signatures, while temporal dynamics (LF6) were linked to changes in the visible and red-edge spectral regions, which are known indicators of photosynthetic status. Having established that the MOFA+-derived latent space captures the primary axes of biological variation, we next sought to validate whether these variance-driving features also form a robust basis for predictive

modelling, a critical test of their functional relevance. We used the 519 MOFA+-selected features to train a multi-task Transformer model, benchmarked against Random Forest and K-Nearest Neighbours classifiers. The MOFA+-selected features enabled high predictive accuracy across all models, with the Transformer achieving excellent F1 Macro scores for predicting Genotype, Treatment, and Time Point in both leaf and root tissues (Fig. 2a; Table S3). The model's class-level performance, detailed in the confusion matrices, confirms its ability to reliably distinguish between biological states (Fig. S4).

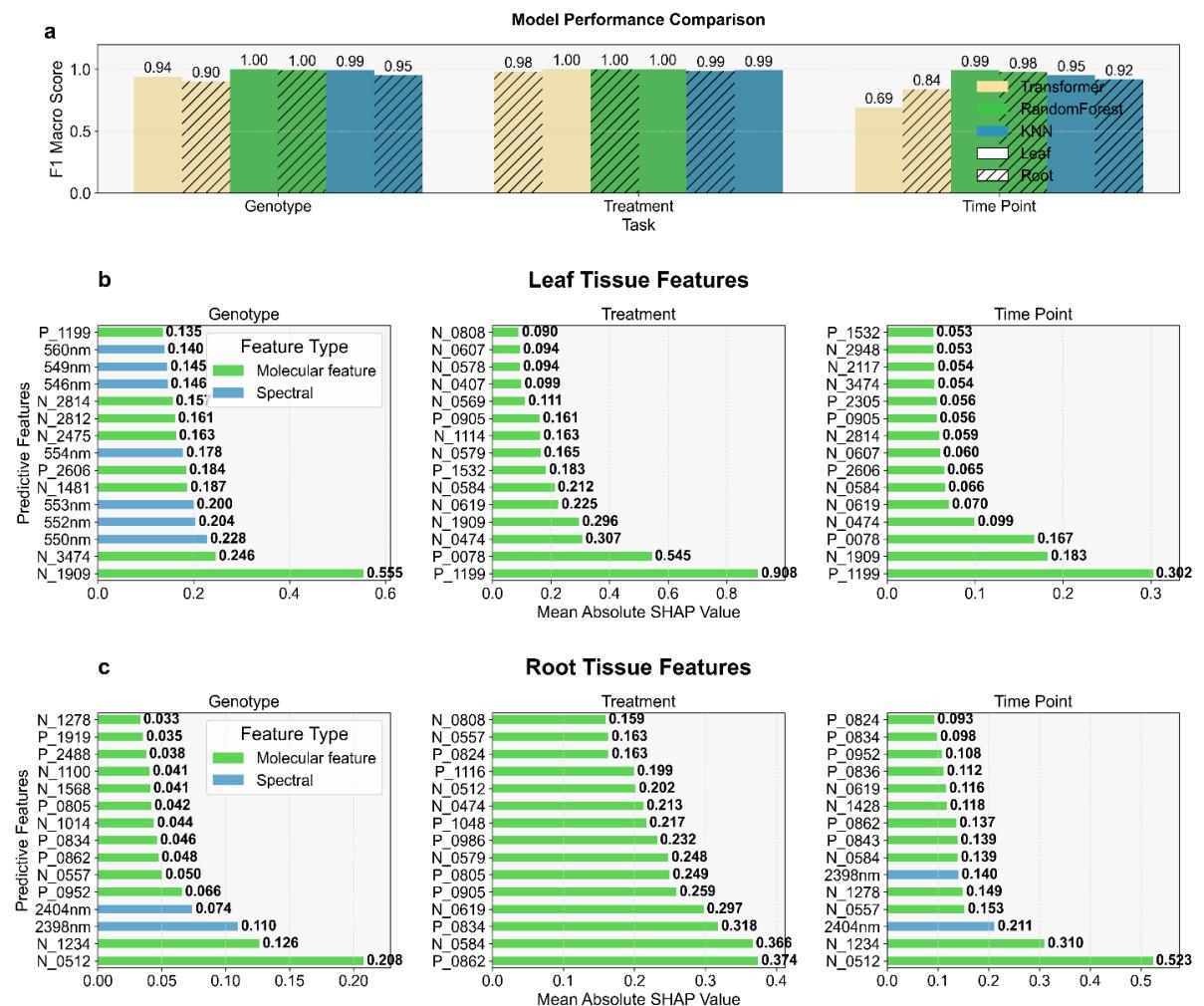


Fig. 2. Predictive validation of MOFA+-selected features and identification of key predictive features via SHAP. (a) Classification performance comparison of the Transformer, Random Forest, and K-Nearest Neighbours (KNN) models on the test set. Bars show the F1 Macro Score for three prediction tasks (Genotype, Treatment, Time Point). Model types are indicated by colour, and tissue data source is distinguished by hatching (no hatch for Leaf, diagonal hatch for Root). Numerical values above each bar indicate the exact F1 score. (b) Top 15 most important features for predicting Genotype, Treatment, and

Time Point in Leaf tissue, as ranked by their mean absolute SHAP value. Bar length corresponds to the feature's contribution to the model's prediction. Bars are coloured by feature type: Molecular features (green) and Spectral features (blue). (c) Top 15 most important features for prediction in Root tissue, displayed as in (b). The analysis consistently identifies molecular features such as N_1909 (Leaf) and N_0512 (Root) as top predictors for genotype.

To move beyond model performance and identify the specific features driving these predictions, we employed Shapley Additive Explanations (SHAP). This analysis consistently highlighted molecular features as the most powerful predictors across most tasks (Fig. 2b-c; Table S4). Crucially, the top predictors were biologically significant molecules that align with the variance patterns identified by MOFA+. For instance, molecular feature N_1909 in leaves and N_0512 in roots emerged as top predictors for genotype (Fig. 2b-c), the same biological axis captured by MOFA+'s Factor 9. This dual validation, demonstrating that our selected features both explain biological variance and power accurate prediction, provides a rigorous foundation for the subsequent mechanistic analysis of the coordination networks that drive stress adaptation.

Revealing the Temporal Architecture of Stress Resilience

Having validated the functional relevance of the MOFA+-selected features, we deployed the Transformer's cross-modal attention mechanism to dissect the coordination networks that link physiology to metabolism. This approach moves beyond simple correlation by learning asymmetric, predictive dependencies (i.e., how spectral features predict molecular responses, rather than simple correlations), quantifying how informative each physiological marker is for predicting a specific biochemical state^{20,23}. Under peak osmotic stress, this analysis revealed qualitatively different network architectures between the tolerant (G1) and susceptible (G2) genotypes (Fig. 3). The tolerant genotype's leaf network was dominated by strong, focused attention radiating

from visible spectral features (e.g., W_565, sensitive to pigment content) to the molecular hub **N_1909** (Fig. 3a). In contrast, its root network centered on interactions between near-infrared features (e.g., W_1100, sensitive to water status and cellular structure) and the molecular hub **N_0512** (Fig. 3c, Table S1). The susceptible genotype's networks were not only weaker but orchestrated by an entirely different set of molecular hubs (Fig. 3b, d). A granular analysis confirmed this differential coordination, revealing a highly structured response where the strength of attention showed a strong linear correlation with genotype-specific fold changes ($r > 0.92$ in leaf) and where the spectral drivers were functionally partitioned by tissue (Supplementary Fig. S2). Quantitatively, the tolerant genotype (G1) exhibited a 355% greater Coordination Strength (average attention across top S2M pairs) in its leaf network, but a 38.9% lower strength in its root network compared to the susceptible genotype (G2), underscoring a highly tissue-specific coordination strategy (Table S9).

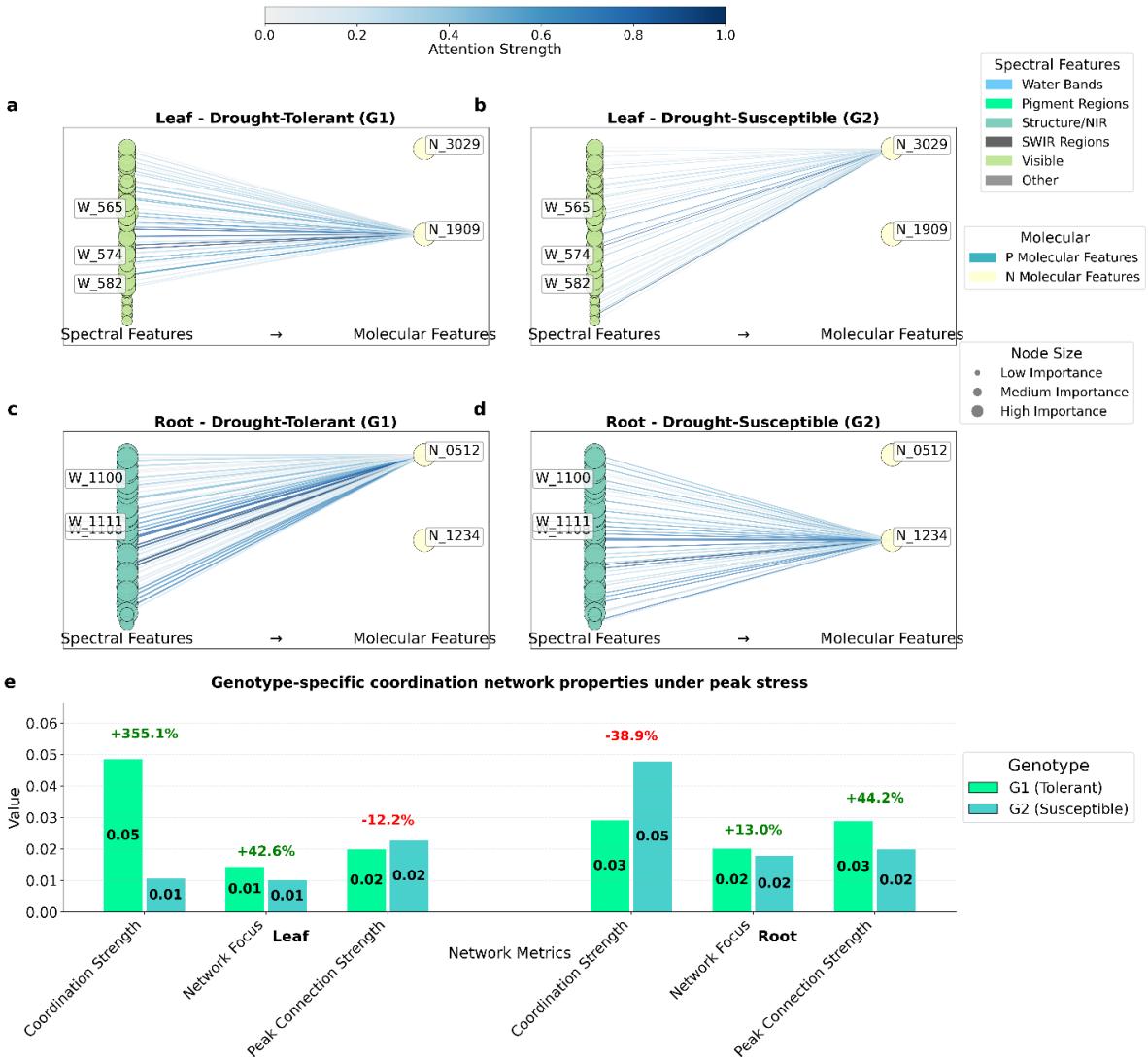


Fig. 3. Genotype-Specific S2M Attention Networks Reveal Differential Coordination Under Peak Osmotic Stress. a-d, Comparison of the top 40 spectral-to-molecular (S2M) attention connections between tolerant (G1; a,c) and susceptible (G2; b,d) genotypes under peak osmotic stress in Leaf (a,b) and Root (c,d) tissues. Edge colour and thickness represent attention strength. Node size reflects a composite importance score derived from total attention and network centrality. Key hub features are labelled. e, Quantitative comparison of network properties. Bar charts show Coordination Strength (mean attention across the top 100 pairs), Network Focus (mean standard deviation of attention scores), and Peak Connection Strength (mean 95th percentile of attention scores). Percentage values indicate the difference of G1 relative to G2.

While these static snapshots reveal architectural differences, a key capability of the framework is its ability to resolve the temporal dynamics of this coordination. We introduce Coordination Advantage Landscapes, a novel methodological output that renders the model's learned attention into a visualisation of how network strategies evolve over the entire stress period (Fig. 4a,b). This technique moves beyond discrete time-point comparisons to quantify the complete temporal strategy of inter-omic communication. It revealed a profound strategic divergence: the tolerant genotype (G1) initiates a powerful, early coordination advantage in its roots that is sustained, while the advantage in its leaves builds progressively, peaking during acclimation. This dynamic orchestration cannot be resolved by conventional, static multi-omics methods and provides direct, mechanistic evidence for the "early response hypothesis," where the timing of the response is as critical as its magnitude^{32,33}.

This macroscopic pattern is quantitatively supported by the dynamics of the top 100 spectral-to-metabolite (S→M) hub pairs (Fig. 4c-f). The tolerant genotype (G1) mounts a powerful and focused response from the outset. Its leaf Coordination Strength is already 2.5-fold higher than G2's at the initial point of stress and intensifies to a peak 3.2-fold higher at the intermediate time point, while the susceptible genotype's network fails to mount a robust response and subsequently collapses (Fig. 4c). This superior coordination is not merely about strength; G1's Network Coordination Focus is significantly higher at the critical initial and late stages, indicating a more efficient and targeted network architecture (Fig. 4d). This efficiency stems from deploying highly stereotyped temporal trajectories for key S→M links, rather than a mixture of dynamic patterns. These uniform strategies—such as a synchronised 'Peak at Time point 2' in leaves—diverged between genotypes in the root, providing strong evidence for a programmed network reconfiguration under stress (Supplementary Fig. S3).

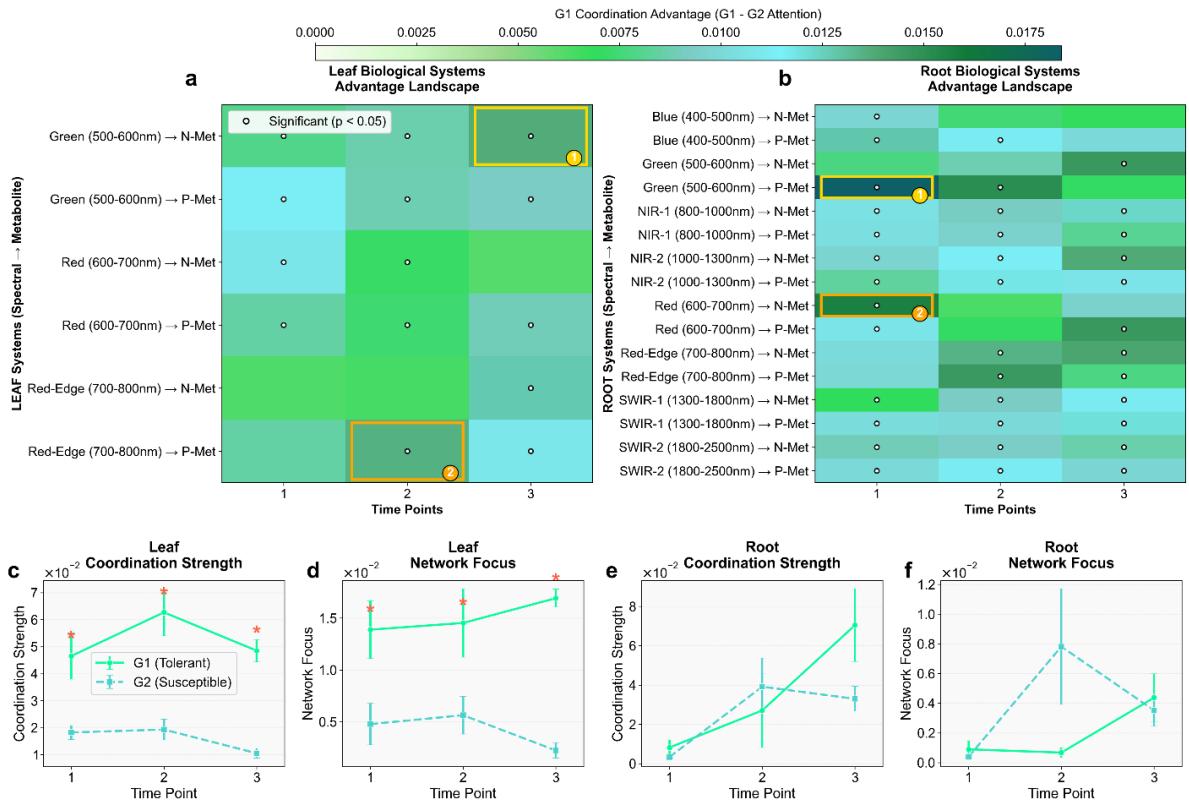


Fig. 4. Drought tolerance is driven by the temporal and tissue-specific architecture of coordination networks. (a, b) Coordination Advantage Landscapes for Leaf (a) and Root (b) tissues under osmotic stress. Heatmaps display the difference in mean Spectral-to-Metabolite (S→M) attention between the tolerant (G1) and susceptible (G2) genotypes, aggregated by biological system pairs. Darker green indicates a stronger coordination advantage in G1. White circles denote statistically significant differences ($p < 0.05$, Mann-Whitney U test), and gold boxes highlight the dominant advantage signatures. (c-f) Temporal dynamics of aggregate network properties, calculated across the top 100 S→M hub pairs. Panels (c, e) plot the overall Coordination Strength (mean S→M attention), while panels (d, f) plot the Network Coordination Focus (standard deviation of S→M attention) for leaf and root tissues, respectively. Trajectories for G1 (Tolerant; solid green line) and G2 (Susceptible; dashed turquoise line) are shown with error bars representing the standard error of the mean (SEM). Asterisks (*) denote a statistically significant difference ($p < 0.05$, Mann-Whitney U test) between genotypes at a given time point. These network-wide metrics provide quantitative, systems-level evidence for the coordination strategies observed in (a, b).

To provide a holistic view of network performance, we integrated these dynamic properties with the predictive power of individual features. The network's coordination focus, and the intensity of its peak connections show that the tolerant genotype (G1) consistently maintains a more structured and powerful communication architecture in its leaves (Fig. 5a, b; Table S2). Critically, the features that act as major network hubs

(high coordination strength) are often the same features identified by SHAP as top predictors of genotype (high predictive power), such as N_1909 in leaves and N_0512 in roots (Fig. 5c). This convergence identifies these features as true 'Key Hubs' of the stress response—molecules that simultaneously drive biological variance, enable accurate prediction, and orchestrate network coordination. By distilling these multi-scale dynamics into a single metric of 'Coordination Efficiency,' we can quantitatively demonstrate that the tolerant genotype's network operates more effectively throughout the stress response (Fig. 5d). This finding reframes resilience not just as the presence of certain molecules, but as a measurable property of the network's timing, architecture, and operational efficiency.

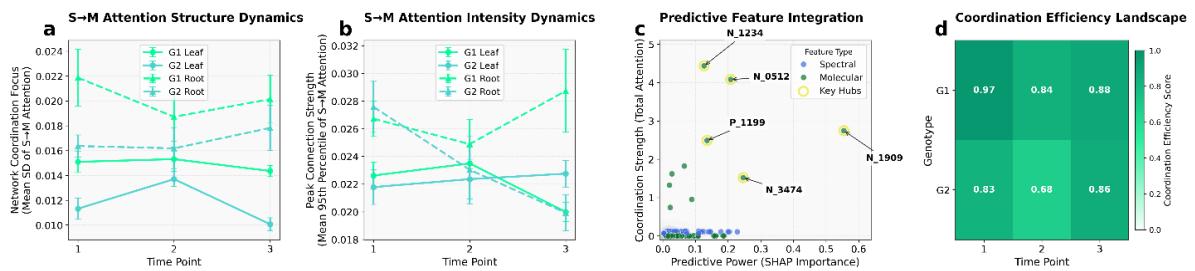


Fig. 5. Multi-scale quantification of coordination network dynamics. (a) S→M Attention Structure Dynamics. Plot of network coordination focus, quantified as the mean standard deviation of S2M attention scores, across the three experimental time points for leaf and root tissues of both genotypes. Error bars represent SEM. (b) S→M Attention Intensity Dynamics. Plot of peak network connection strength, quantified as the mean 95th percentile of S2M attention scores, across time points for both tissues. Error bars represent SEM. (c) Predictive Feature Integration. Scatter plot comparing each feature's coordination strength (y-axis) with its predictive power (x-axis). Points are coloured by feature type (green: molecular; blue: spectral). Key Hubs, features that are important for both network coordination and prediction, are designated with a yellow highlight circle. (d) Coordination Efficiency Landscape. Heatmap of the composite 'Coordination Efficiency' score for each genotype and time point. The score is a holistic measure of network performance, integrating the network's signal-to-noise ratio and directional balance (see Methods).

Uncovering Complementary Layers of Mechanistic Insight

A key innovation of the MOFA+ Transformer framework is its ability to decouple and contrast two distinct modes of feature importance: (1) features that drive biological

variance, identified by MOFA+, and (2) features critical for predictive accuracy, identified by SHAP³⁴. This dual-discovery approach provides a more comprehensive understanding than single-method analyses. A direct comparison revealed that these two feature sets are largely distinct. Across most predictive tasks, the overlap between the top 100 features from MOFA+ and SHAP was minimal, with the Jaccard index for root genotype prediction, for example, being only 0.0050 (Fig. 6a, b). This divergence is further highlighted by the different priorities of each method: MOFA+ tended to prioritise spectral features for explaining variance, while SHAP overwhelmingly favoured molecular features for predictive power (Fig. 6c). The overall correlation between MOFA+ feature weights and SHAP importance scores was correspondingly low (Fig. 6g, h). This demonstrates that feature selection based on variance alone provides an incomplete and potentially misleading view of a system's predictive signatures.

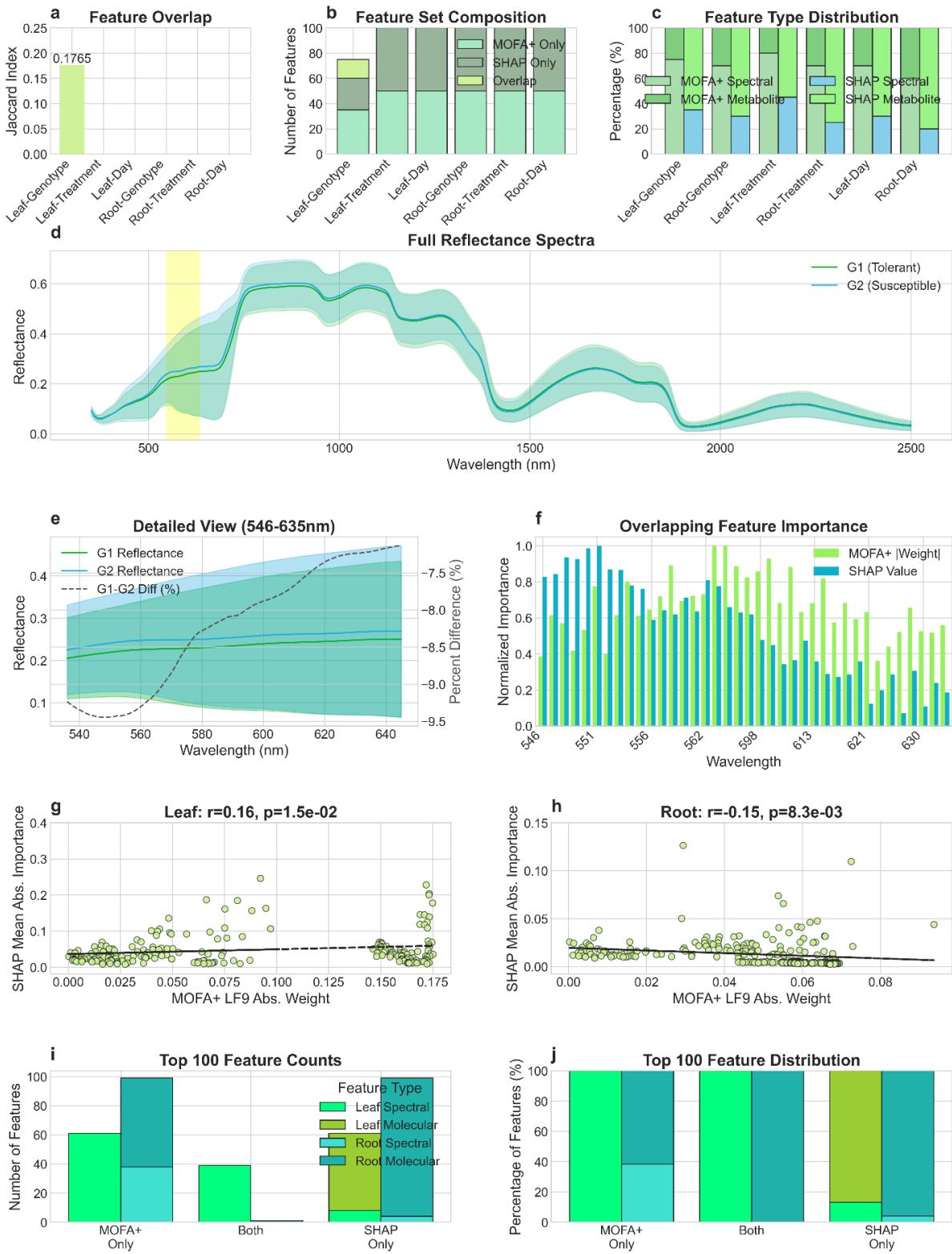


Fig. 6. MOFA+ and SHAP Reveal Complementary Views of Feature Importance. This figure illustrates that features driving biological variance (identified by MOFA+) are largely distinct from those driving predictive accuracy (identified by SHAP), with their convergence highlighting key biomarkers. (a) The Jaccard index quantifies the overlap between the top 100 features from MOFA+ and SHAP, showing it is minimal for most

tasks but moderate for leaf-genotype prediction. (b-c) Stacked bars show the number (b) and percentage (c) of top features identified uniquely by each method versus those in the overlap, highlighting the different feature types prioritised by MOFA+ (more spectral) and SHAP (more molecular). (d-e) The 546–635 nm spectral "sweet spot" (highlighted yellow) where the methods converge is shown in the context of the full leaf reflectance spectra (d) and in a detailed view (e), revealing consistent reflectance differences between genotypes. (f) For features within this sweet spot, their normalised importance scores from MOFA+ (variance-driving weight) and SHAP (predictive value) are comparable. (g-h) Scatter plots showing the low overall correlation between MOFA+ weights and SHAP values for (g) Leaf and (h) Root features, confirming the general divergence in feature prioritisation. (i-j) For the Leaf Genotype task, bar charts of absolute counts (i) and percentage distribution (j) confirm that MOFA+ prioritises spectral features while SHAP strongly prioritises molecular features within the unique and overlapping feature sets.

However, the one exception to this trend proved highly informative. In the task of predicting leaf genotype, we observed a moderate Jaccard overlap of 0.1765 between the top features from MOFA+'s genotype-associated factor (Factor 9) and SHAP (Fig. 6a). Strikingly, these 30 overlapping features were exclusively spectral wavelengths within the 546–635 nm visible range (Fig. 6d, e; Table S5). This 'sweet spot' region, known to be sensitive to leaf pigment composition and photosynthetic health (Gitelson et al., 2003; Sims & Gamon, 2002), was independently identified by MOFA+ as a key driver of biological variance and by SHAP as highly predictive. Within this region, the normalised importance scores from both methods were comparable, confirming the convergence (Fig. 6f). This discovery showcases the framework's unique methodological power to pinpoint features that are simultaneously mechanistically relevant (driving variance) and practically useful as robust biomarkers (powering prediction), an insight that would be missed by either method in isolation (Fig. 6i, j).

To further dissect the architecture of prediction, we used SHAP to explore how feature importance is distributed across different tasks and tissues (Fig. 7). Hierarchical clustering of SHAP values revealed distinct, tissue-specific predictive patterns, with features grouping together based on the tasks they were most influential for (Fig. 7a, b). This analysis reinforced the overall dominance of molecular features for predictive accuracy in both leaf and root tissues (Fig. 7c, d), while also showing that the relative contribution of spectral versus molecular data is highly task-dependent (Fig. 7e, f). For instance, spectral data was significantly more important for genotype prediction in

leaves compared to roots (Fig. 7g). Ultimately, this integrated analysis underscores that the framework's dual-discovery approach provides a more complete and nuanced mechanistic picture, linking features that drive variance, prediction, and, as shown previously, network coordination.

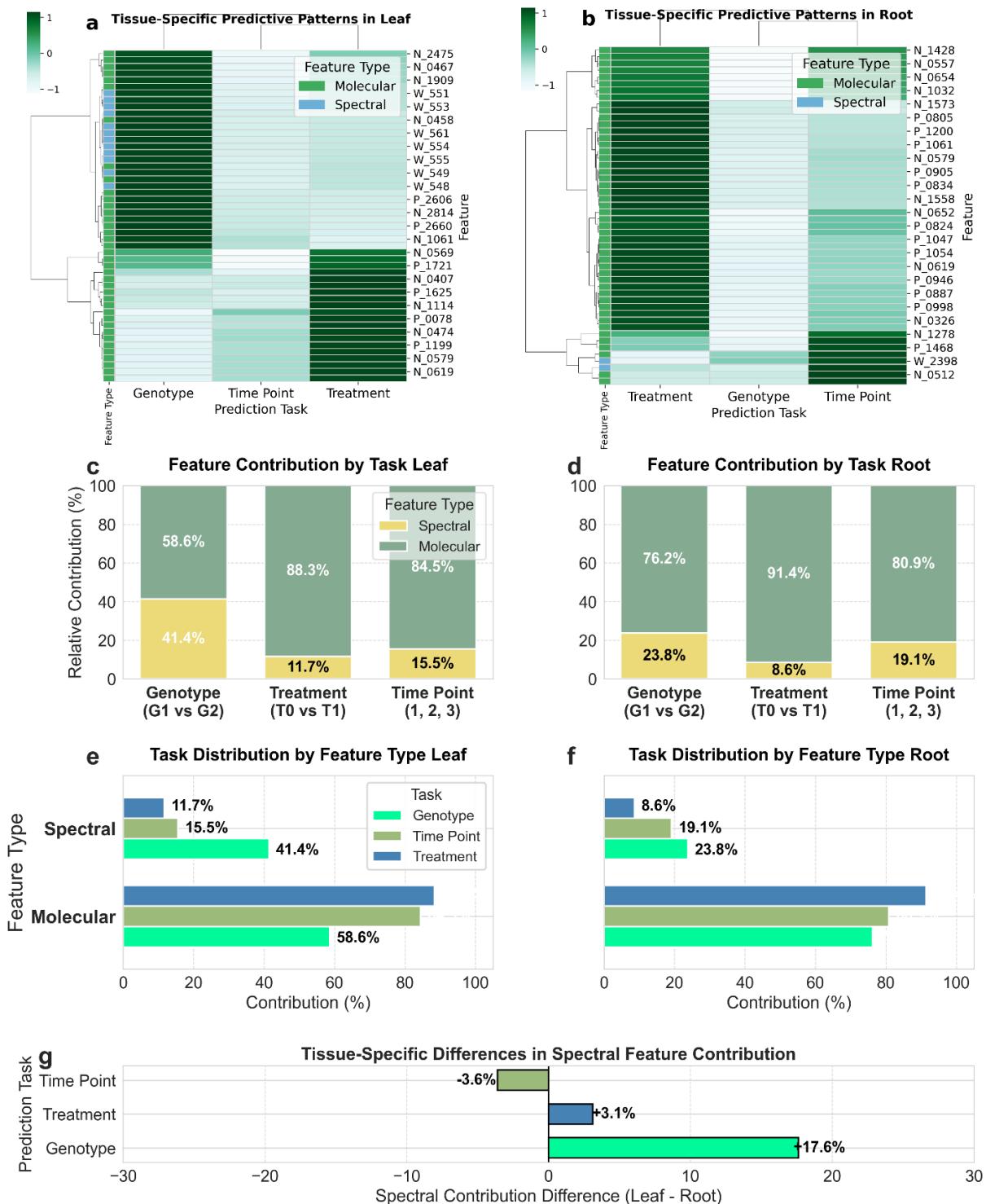


Fig. 7. SHAP Analysis Dissects Tissue-Specific and Task-Dependent Predictive Feature Importance. (a-b) Hierarchical clustering of the top 50 features by their standardised SHAP values reveals tissue-specific predictive patterns for (a) Leaf and (b) Root tissues. Features clustering together are important for the same predictive tasks. (c-d) Bar charts show the aggregated relative contribution (%) of Molecular versus Spectral features within (c) Leaf and (d) Root tissues, demonstrating the overall dominance of molecular features in prediction. (e-f) Horizontal bar charts further break down the feature type contributions for each specific prediction task in (e) Leaf and (f) Root tissues, showing task-specific reliance on different modalities. (g) A bar chart quantifies the difference in spectral feature contribution between leaf and root tissue for each task, highlighting that spectral data is significantly more important for genotype prediction in leaves.

Validating the Framework's Generalisability on a Single-Cell Dataset

To demonstrate the robustness and general applicability of the MOFA+ Transformer framework beyond the initial plant stress model, we applied it to the public HyperSeq dataset, which contains paired single-cell transcriptomics and hyperspectral imaging data from human cells. The MOFA+ stage successfully deconstructed this entirely different multi-modal dataset into interpretable latent factors (Fig. 8a). While Factors 1 and 2 were dominated by the transcriptomic and spectral views, respectively, Factor 3 emerged as a key integration factor, capturing a shared 'Stress and Metabolism' signature. This factor was driven by a distinct set of spectral channels and well-known stress-response genes, including *HSPA6* and *COX6C* (Fig. 8b,c), confirming the framework's ability to identify biologically coherent signatures in diverse data types.

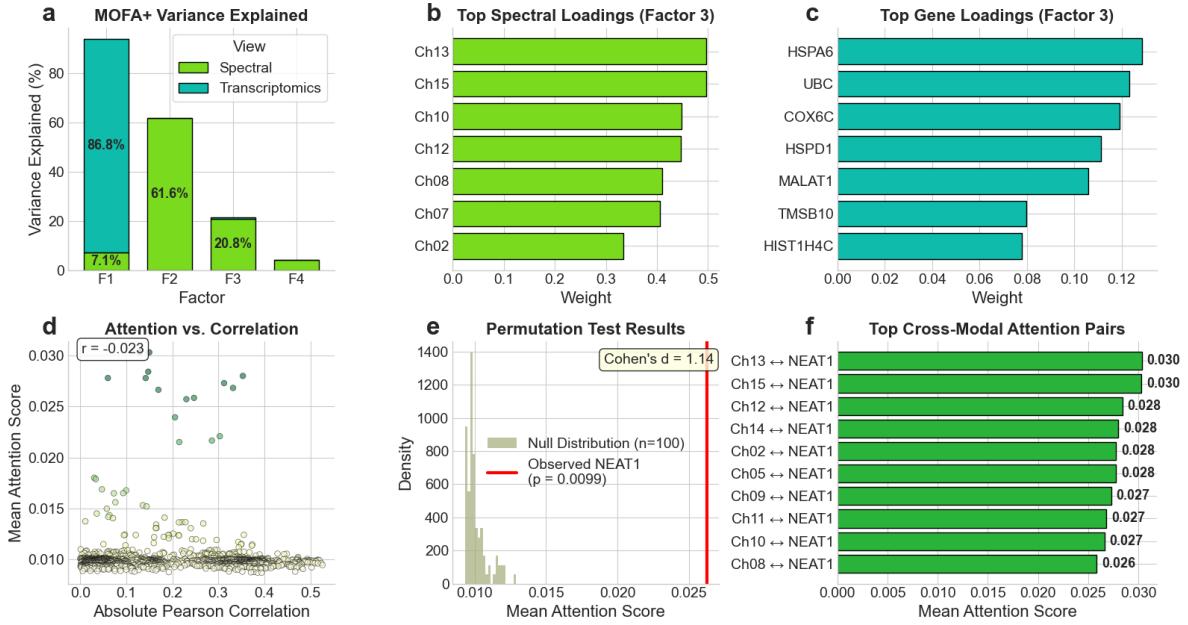


Fig. 8. MOFA+ Transformer discovers and validates novel cross-modal links in an independent single-cell dataset. The framework's generalisability is demonstrated on the public HyperSeq dataset. **a**, MOFA+ successfully deconstructs variance into distinct factors. **b,c**, Top feature loadings for the integration factor (F3) reveal a 'Stress and Metabolism' signature. **d**, Learned attention scores show near-zero correlation with simple Pearson coefficients ($r = -0.023$), demonstrating the model's ability to capture non-linear dependencies. **e,f**, This discovery power is highlighted by a strong, specific link between the cellular spectral phenotype and the lncRNA NEAT1. The top 10 attention pairs converge on NEAT1 (**f**), an association confirmed as highly significant by permutation testing ($n=100$; $p = 0.0099$; Cohen's $d = 1.14$; **e**).

The framework demonstrated genuine discovery power by identifying complex relationships that are invisible to standard linear methods. To test this directly, we compared the Transformer's learned attention scores against Pearson correlation coefficients for the same feature pairs. The analysis revealed a stark divergence, showing no significant linear relationship (Pearson's $r = -0.023$; $p > 0.05$) between the two measures. This result is the critical evidence, confirming the model's capacity to uncover novel, non-linear associations that are not detectable by simple co-expression analysis (Fig. 8d).

The model's most striking demonstration of discovery power was its identification of a novel, biologically profound relationship converging from multiple spectral channels onto the long non-coding RNA (lncRNA) **NEAT1** (Fig. 8f). NEAT1 is a critical architectural

component of paraspeckles, nuclear bodies that form in response to cellular stress and regulate gene expression³⁵. To rigorously validate this data-driven hypothesis, we performed a permutation test by shuffling gene expression values within each sample to generate a null distribution. The analysis confirmed that the intense attention focused on NEAT1 was highly significant and not a product of chance ($p = 0.0099$), with a large effect size (Cohen's $d = 1.14$) (Fig. 8e). This end-to-end validation—from interpretable variance decomposition to the discovery and rigorous statistical validation of a novel, biologically significant link in an entirely independent biological system—establishes the framework as a powerful, generalisable engine for hypothesis generation from complex multi-modal data.

Discussion

The escalating complexity of multi-omics data demands analytical frameworks that move beyond correlative descriptions to uncover the predictive, dynamic, and interpretable mechanisms underlying biological function^{1,6}. Here, we introduce MOFA+ Transformer and demonstrate its ability to move beyond static snapshots of molecular components to reveal a core principle of biological resilience: adaptation to osmotic stress emerges not from the mere presence of specific molecules, but from the precise timing and architecture of the coordination networks that link them^{36,37}. Our application to contrasting wheat genotypes provides novel, mechanistic evidence for the "early response hypothesis"^{38,39}, revealing that tolerance is an actively orchestrated, tissue-specific strategy. The tolerant genotype deploys a strong, early coordination network in its roots, followed by a progressively strengthening network in its leaves (Fig. 4), a dynamic not resolved by conventional methods²⁹. This level of mechanistic insight is achieved by the framework's unique ability to quantify predictive, feature-specific interactions and their temporal dynamics.

The framework's innovation lies in its synergistic architecture, which is designed to overcome the key limitations of existing multi-omics integration approaches (Table 1). It first employs MOFA+ to perform a robust, unsupervised disentanglement of complex variation, separating genotype-specific signatures, temporal progression, and technical effects into stable and biologically interpretable latent factors^{11,12}. While other methods may excel at specific tasks like clustering, this interpretable foundation is critical for structuring heterogeneous data for the subsequent mechanistic inquiry performed by the Transformer. Building on this, the Transformer's cross-modal attention mechanism learns asymmetric, predictive dependencies, moving beyond the simple co-variation captured by correlation-based methods^{14,16}. This design provides a directly interpretable view of feature-pair interactions, thereby avoiding the 'black box' problem common in many machine learning applications and providing an evidence-based route from high-dimensional data to testable hypotheses^{20,40}. The strong linear correlation ($r > 0.92$ in leaves) between the model's internal attention scores and a key biological outcome, the genotype-specific fold change—validates the use of attention as a quantitative proxy for biological effect size, a key test of the model's interpretability (Supplementary Fig. S2c). It confirms that learned attention can serve as a direct, quantitative proxy for biological effect size, bridging the gap between a deep learning mechanism and a tangible biological insight.

Applying MOFA+ Transformer to our osmotic stress dataset validated its discovery power by revealing a key systems-level principle: resilience emerges from a highly dynamic, tissue-specific strategy rather than a simple, uniform advantage. The discovery of a rapid, sustained coordination in roots versus a progressive build-up of coordination in leaves provides a new layer of understanding to plant stress responses^{32,33}. The framework revealed specific principles of this coordination, including tissue-specific spectral partitioning for metabolic sensing—with leaves predominantly utilising visible wavelengths and roots employing near-infrared regions^{41,42}. Furthermore, the method identified genotype-specific network orchestrators, such as N_1909 in leaves and N_0512 in roots for the tolerant genotype. This discovery shifts the focus from "which metabolites are present?" to "which

metabolites serve as network coordinators and how effectively and consistently they do so?"—a systems-level organisational principle underscored by the highly stereotyped temporal trajectories of these networks (Supplementary Fig. S3), with profound implications for breeding strategies^{43,44}. Moreover, the highly structured nature of this coordination, evidenced by the strong linear relationship between attention differences and fold changes (Supplementary Fig. S2c,d), suggests a governing rule in the plant's response. This suggests a highly organised, rather than random, modulation of communication in a consistent manner by its components. Uncovering such quantitative, systems-level organisational principles is a key capability of our framework.

A key methodological advance of our framework is the direct demonstration that features driving biological variance (identified by MOFA+), predictive accuracy (identified by SHAP), and network communication (identified by attention) represent functionally distinct and often non-overlapping sets (Fig. 6). This finding addresses a critical limitation in biology, where a feature's predictive importance in a machine learning model is often conflated with its broader mechanistic role^{20,40}. By distinguishing these contributions, our framework provides a more holistic view. For instance, while metabolomic features were paramount for predictive accuracy (Fig. 7c, d), the convergence of MOFA+ and SHAP on a narrow 'sweet spot' of spectral features (546–635 nm) identified a robust biomarker that is simultaneously a driver of biological variance and highly predictive of genotype (Fig. 6d-f). This spectral region, known to be sensitive to leaf pigment composition^{45,46}, represents an intersection of biological mechanism and practical utility. This capability to pinpoint where different modes of importance converge enables the simultaneous discovery of mechanistic drivers and robust biomarkers.

The capabilities of the MOFA+ Transformer framework extend beyond plant science. While recent large-scale models like scGPT⁴⁷ excel at general-purpose integration, our framework's distinct contribution is its specific focus on providing an interpretable view of the temporal dynamics and feature-specific coordination that generate mechanistic

hypotheses. Its capacity to deliver interpretable, predictive, and dynamic insights from heterogeneous, time-resolved data makes it a powerful tool for any domain requiring mechanistic understanding of complex systems. In clinical research, it could move beyond static biomarkers to reveal how dynamic multi-omics coordination patterns predict patient responses to treatment, a crucial step toward personalised medicine^{2,6}. This end-to-end validation on the HyperSeq single-cell dataset ⁴⁷—from identifying a novel link between a cellular phenotype (autofluorescence) and the stress-related lncRNA NEAT1³⁴ to its statistical confirmation, establishes the framework's broad applicability for hypothesis generation in different biological systems. This principle is broadly applicable to fundamental questions in developmental biology, cancer research, and microbial ecology ⁴⁰. Future methodological developments could incorporate causal inference frameworks or integrate graph neural networks for pathway-aware analysis, further enhancing the framework's discovery capabilities.

While the MOFA+ Transformer framework was developed to overcome key limitations in multi-omics analysis, certain considerations define the scope of this study and offer avenues for future research. The molecular features in this work represent putative identifications under the Metabolomics Standards Initiative⁴⁸, and targeted validation of key network hubs like N_1909 via tandem mass spectrometry is a critical next step for full structural elucidation⁴⁹. Furthermore, while the Transformer's attention mechanism provides an interpretable view of predictive associations, these relationships remain correlational. Establishing direct causality will require moving towards causal inference frameworks and, ultimately, functional validation through genetic or biochemical experiments⁵⁰. Finally, validating these coordination networks under complex field drought conditions and across broader genetic diversity will be essential for translating these findings into agricultural practice⁵¹.

In conclusion, MOFA+ Transformer provides an interpretable deep learning framework capable of quantifying predictive, feature-specific relationships and their temporal dynamics within complex multi-modal data. By revealing that biological resilience is an emergent property of network timing and architecture, our work demonstrates a path to

move beyond descriptive multi-omics to the discovery of functional coordination patterns. This shift from identifying system components to understanding their dynamic orchestration enables deeper mechanistic insight and offers a powerful, generalisable tool to advance hypothesis generation and discovery across the biological sciences.

Methods

The MOFA+ Transformer Framework: Conceptual Overview

The MOFA+ Transformer framework is an interpretable, two-stage deep learning pipeline designed to resolve predictive, feature-specific, and dynamic relationships from heterogeneous multi-omics data. The first stage employs Multi-Omics Factor Analysis+ (MOFA+)^{11,12}, an unsupervised statistical method, to disentangle the sources of variation in the data into a set of biologically interpretable latent factors. This provides a robust, low-dimensional representation and a principled feature selection strategy. The second stage leverages a multi-task Transformer architecture, which uses a cross-modal attention mechanism^{23,24} to learn asymmetric, predictive relationships between specific features from different modalities. This synergistic design moves beyond simple correlation to quantify the architecture and temporal dynamics of inter-omic coordination, providing an evidence-based path to mechanistic hypothesis generation²⁰.

Experimental System for Framework Validation

Plant Material, Growth Conditions, and Stress Application

The framework was validated using two wheat (*Triticum aestivum* L.) genotypes, the tolerant 'Gladius' (G1) and the susceptible 'DAS5_003811' (G2), which have established contrasting responses to drought⁵²⁻⁵⁴. Plants were cultivated in an aerated hydroponic system within a controlled environment (22-h photoperiod at 22°C, 2-h dark period at

17°C, 60–70% humidity). The nutrient solution was maintained at pH 5.5 (full composition in Table S6).

To model ecologically relevant drought responses, osmotic stress was imposed using two complementary protocols: an acute/severe stress (0.3 M sorbitol) and a mild/prolonged stress (0.15 M sorbitol). For downstream analysis, data from both protocols were integrated. The distinct sampling days from each protocol were aligned into three biologically analogous phases: Time Point 1 (Initial Stress Onset), Time Point 2 (Intermediate Acclimation), and Time Point 3 (Late-Stage Adaptation). This abstraction allowed the model to learn robust temporal signatures conserved across different stress intensities.

Sample Collection and Preparation

At each time point, leaf and root tissues were collected separately, snap-frozen in liquid nitrogen, and stored at -80°C. Frozen tissues were lyophilised for 3 days and homogenised to a fine powder using a Geno/Grinder 2010 tissue lyser. Aliquots of 20 mg were used for metabolite extraction in 1 mL of 80% methanol, following a standardised protocol of vortexing, sonication, and incubation^{55,56}.

Multi-Omics Data Generation and Preprocessing

Hyperspectral Reflectance Data Acquisition

Physiological status was captured via spectral reflectance measurements (351–2500 nm) using an ASD FieldSpec 4 spectroradiometer. All measurements were conducted inside a custom-built imaging cube with standardised illumination. Spectra were calibrated against a white Spectralon® reference panel, with each recorded spectrum representing an average of 25 scans. Raw hyperspectral data also underwent a thorough quality assessment. High signal quality (median STD = 0.080, approximate SNR = 2.39; Fig. S8B) and stable spectral shapes confirmed by derivative analysis (Savitzky-Golay

filter; Fig. S8H-I) obviated the need for smoothing or baseline correction. All 336 samples, including 28 potential outliers identified via a multi-method screen using IQR (1.5x multiplier), Modified Z-score (>3.5), and Local Outlier Factor ($n_{neighbors}=20$) (Table S8), were retained to preserve the full spectrum of biological variation for the downstream models. A detailed description of the setup is provided in Supplementary Note 1.

Untargeted LC-MS Molecular Feature Profiling

Biochemical responses were profiled using an untargeted approach on a Vanquish UHPLC system coupled to a Q Exactive Plus Orbitrap Mass Spectrometer. Chromatographic separation was achieved on a C18 reversed-phase column. Data were acquired in both positive and negative electrospray ionisation modes in separate analytical runs. Each detected molecular feature (a putative metabolite) was assigned a unique identifier (e.g., N_1909). System stability was monitored throughout the analytical sequence using internal standards and pooled quality control (QC) samples

^{57,58}.

Data Preprocessing, Quality Control, and Augmentation

Raw LC-MS data were processed using Genedata Expressionist® Refiner MS software (parameters in Table S7). A rigorous, multi-step quality control pipeline was then executed. Missing values were confirmed as Missing at Random (MAR) via logistic regression (Fig. S7A-C); consequently, a Random Forest-based imputation (missForest) was selected for its superior performance over alternatives in key metrics like Earth Mover's Distance and Shannon entropy (Fig. S7D-F). Outlier detection was performed using the Isolation Forest algorithm, and data variability was stabilised via an inverse hyperbolic sine (asinh) transformation, which optimally reduced the coefficient of variation from 0.876 to 0.206 (Fig. S7I-M). Features exhibiting excessive variability (Relative Median Absolute Deviation $> 30\%$) were excluded. To enhance model

robustness, the curated datasets ($n=336$) were computationally augmented to 1,344 samples. A subsequent multi-stage validation confirmed that the augmented data preserved key statistical properties (e.g., Jensen-Shannon divergence ~ 0.11), boosted statistical power (>3.5-fold), and maintained high cross-modal consistency (0.99), as detailed in seven validation reports (Fig. S9-S13). This validation process was performed independently for both leaf and root datasets with comparable outcomes. To avoid redundancy, figures presented here (Fig. S10-S13) show the results from the root dataset as a representative example.

The MOFA+ Transformer Analytical Pipeline

Stage 1: Unsupervised Variance Decomposition with MOFA+

The four pre-processed and augmented data modalities (leaf/root spectral and leaf/root molecular features; $n=1,344$ samples) served as input to the MOFA+ model. Each modality was feature-wise Z-score normalised prior to analysis. Using the mofapy2 Python package¹², we trained the model to identify 20 initial latent factors, with data-driven sparsification enabled via Automatic Relevance Determination (ARD). Factors explaining less than 1% of variance were excluded, resulting in 11 active latent factors (LFs) for downstream analysis. The biological relevance of these factors was established by calculating the Spearman's rank correlation between their scores and the experimental metadata (Genotype, Treatment, Time Point), with significance determined at an FDR < 0.05 . The stability of the feature sets driving these factors was confirmed via 100-iteration bootstrapping, which yielded a 96.0% average selection frequency. Finally, a context-aware feature selection strategy was employed to curate the 519 most biologically relevant features for input into the Transformer model.

Stage 2: Predictive Modelling with a Multi-Task Transformer

A multi-task Transformer model was implemented in PyTorch to learn the predictive relationships within the 519 MOFA+-selected features. Each feature was first embedded

into a 64-dimensional vector using modality-specific linear layers with learnable positional encodings. The core of the architecture consisted of two cross-attention layers, each with four attention heads, enabling the model to learn bidirectional dependencies between modalities (e.g., spectral-to-metabolic and metabolic-to-spectral)^{23,24}. Standard architectural components, including residual connections, layer normalisation, and dropout (rate=0.1), were used to ensure stable training. The final embeddings were passed to three separate multi-layer perceptron (MLP) heads for the parallel classification of Genotype, Treatment, and Time Point.

Model Training and Benchmarking

The data were split using stratified sampling into training (70%), validation (15%), and test (15%) sets. To prevent data leakage, normalisation parameters were learned from the training set only. The model was trained using the Adam optimiser (learning rate = 5e-5, weight decay = 1e-5)⁵⁹ to minimise a combined cross-entropy loss across the three tasks. Training was regulated by an early stopping criterion with a patience of 15 epochs. The performance of the Transformer was benchmarked against two standard classifiers, Random Forest and K-Nearest Neighbours (k=5), using the same training and testing data.

Interpretable Analyses of Model Outputs

To move from prediction to mechanistic insight, we implemented a multi-pronged interpretation strategy. First, we used Shapley Additive Explanations (SHAP) with a GradientExplainer to quantify the predictive importance of each feature for each classification task³⁴. Second, we extracted the raw attention weights from the final cross-attention layer to directly quantify the strength of predictive relationships between specific feature pairs. These weights were then used to construct the Spectral-to-Metabolite (S2M) coordination networks. Third, to analyse the temporal dynamics of these networks, we developed Coordination Advantage Landscapes and calculated

aggregate network properties from the attention scores over time. Finally, we performed a complementarity analysis by directly comparing the top features identified by MOFA+ (variance-driving) and SHAP (prediction-driving) using the Jaccard index, which revealed a key convergence for Leaf-Genotype prediction (Jaccard index = 0.1765) in the 546–635 nm spectral range (Table S5). This integrated approach, detailed below, allowed us to identify features important for variance, prediction, and network coordination.

Framework Generalisability on a Single-Cell Dataset

To test the generalisability of the pipeline, we applied it to the publicly available HyperSeq dataset (GEO: GSE254034), which contains paired single-cell hyperspectral imaging and transcriptomic data⁶⁰. The pre-processed data were analysed using the identical MOFA+ Transformer framework. To validate the discovery of a novel link between spectral features and the lncRNA NEAT1, we performed a permutation test ($n=100$ shuffles) to generate a null distribution of attention scores. The significance of the observed attention was assessed against this null distribution, following standard statistical procedures for Monte Carlo P-value calculation⁶¹.

Statistical Analysis, Visualisation, and Code Availability

All analyses were performed in Python (v3.10.11). Key packages included pandas (v2.2.3), PyTorch (v2.6.0), scikit-learn (v1.6.1), mofapy2 (v0.7.2), and shap (v0.47.1). Visualisations were generated using matplotlib (v3.10.1), seaborn (v0.13.2), and NetworkX (v3.4.2). Detailed descriptions of the quantitative methods used to generate each figure are provided below.

All code used for data preprocessing, analysis, and figure generation is publicly available in a GitHub repository: https://github.com/shoaibms/mofa_transformer. Raw data are available from the corresponding authors upon reasonable request. The processed and augmented data used for all analyses are available in the repository to ensure full reproducibility.

Quantitative Analysis and Visualisation

All statistical tests were two-sided unless otherwise specified. Analyses were performed using Python (v3.10.11), and a complete list of packages and versions is available in the code repository.

Figure 1: MOFA+ Variance Decomposition and Factor Annotation. The MOFA+ model was fitted using the mofapy2 package. To annotate the resulting 11 active factors, we loaded a pre-computed table of Spearman's rank correlations between each factor and the experimental metadata. The most biologically relevant factors were dynamically identified as those with the most significant correlation (lowest FDR-corrected p-value) for Genotype, Time Point, and Treatment/Batch, respectively. Latent space visualisations were generated by plotting the scores of these key factors, with group centroids and 95% confidence ellipses ($n_{std}=1.96$) overlaid. Condition-specific differences in factor score distributions were assessed with a two-sided Mann-Whitney U test.

Figure 2: Predictive Validation and SHAP Analysis. To validate the functional relevance of the 519 MOFA+-selected features, we trained the Transformer, a Random Forest ($n_{estimators}=100$), and a K-Nearest Neighbours ($n_{neighbors}=5$) classifier. Performance was evaluated on a held-out test set (15% of data) using the F1 Macro score, as shown in panel (a). To identify the specific features driving these predictions, we applied a GradientExplainer from the shap library to the trained Transformer. A background dataset of 100 randomly selected training samples established the baseline, and SHAP values were subsequently calculated for 200 samples from the test set. The top 15 features for each task, ranked by their mean absolute SHAP value across all explained instances, were then visualised in panels (b) and (c).

Figure 3: Genotype-Specific S2M Attention Networks. The architecture of predictive spectral-to-molecular (S→M) coordination networks was quantified using the Transformer's attention mechanism. For network visualisation under peak stress, we constructed graphs using the top 40 S→M pairs for each genotype. The layout of these

networks was optimised using hierarchical clustering (method='ward') to group nodes with similar connection patterns. Edge weight was mapped to mean attention strength. Node size was mapped to a composite importance score, calculated as an equal weighting of the node's normalised total attention and its degree centrality. To quantitatively compare network properties, we calculated three metrics based on the top 100 overall S→M hub pairs: Coordination Strength (mean attention), Network Focus (standard deviation of attention), and Peak Connection Strength (95th percentile of attention).

Figure 4: Coordination Advantage Landscapes and Temporal Dynamics. To resolve the temporal evolution of coordination strategies, we developed Coordination Advantage Landscapes. Individual features were first aggregated into functional biological systems by categorising spectral features into canonical wavelength ranges (e.g., 'Blue: 400-500nm') and molecular features by their ionisation mode ('N-Metabolites' or 'P-Metabolites'). The 'Coordination Advantage' plotted in the heatmap was calculated as the difference between the mean S→M attention scores (G1 mean - G2 mean) for each system-pair at each time point under stress. Statistical significance was assessed by performing a two-sided Mann-Whitney U test ($p < 0.05$) on the underlying distributions of all attention scores within each system-pair. Aggregate network dynamics were quantified using the top 100 S→M hub pairs, which were identified based on the global mean attention across all samples and heads. Temporal trajectories represent the mean \pm SEM of per-sample metrics calculated using these hub pairs.

Figure 5: Multi-scale Quantification of Coordination Network Dynamics. To synthesise network performance, we integrated analyses across multiple scales. The background density in the feature integration plot was rendered using a Kernel Density Estimate (KDE). "Key Hubs" were defined quantitatively as features scoring above the 95th percentile for both Predictive Power (SHAP value) and Coordination Strength (total attention). The holistic Coordination Efficiency Score was calculated as the mean of three components: (i) the S→M signal-to-noise ratio (mean/std), (ii) the M→S signal-to-noise ratio, and (iii) a directional balance score ($1 - |\text{Avg_S2M} - \text{Avg_M2S}| / \max(\text{Avg_S2M}, \text{Avg_M2S})$).

Figure 6: MOFA+ and SHAP Feature Importance Complementarity. To investigate the relationship between features driving variance and those driving prediction, we compared the top 100 features from MOFA+ and SHAP. Overlap was quantified using the Jaccard index. The correlation between MOFA+ absolute loading weights and SHAP mean absolute importance was assessed using the Pearson correlation coefficient (scipy.stats.pearsonr).

Figure 7: Dissection of Tissue-Specific and Task-Dependent Predictive Importance. To dissect predictive feature architecture, we performed hierarchical clustering on the top 50 predictive features. The input for clustering was a matrix of SHAP values (features \times tasks) that was row-normalised ($z_score=0$) to cluster features based on their relative importance pattern across tasks. The relative contribution of each data modality to predictive performance was quantified by aggregating the absolute SHAP values by feature type. The tissue-specific difference was calculated as the percentage point difference in spectral contribution (Leaf % - Root %).

Figure 8: Framework Validation on an Independent Single-Cell Dataset. To confirm generalisability, the MOFA+ Transformer pipeline was applied to the HyperSeq dataset. The loadings for the top 7 features of the key integration factor (Factor 3) were visualised. The model's ability to capture non-linear relationships was demonstrated by comparing learned attention scores for the top 10 pairs against their corresponding Pearson correlation coefficients. To statistically validate the discovered link to the lncRNA NEAT1, we performed a permutation test ($n=100$ shuffles). The observed mean attention score for NEAT1 was tested against the generated null distribution to derive an empirical p-value and a Cohen's d effect size.

Acknowledgement

This work was supported by Agriculture Victoria, an agency of the Department of Energy, Environment and Climate Action (DEECA) within the Victorian State Government, Australia.

Authors Contributions

M.S. conceived and designed the study, conducted all experiments, performed data analysis, developed computational methods, and wrote the manuscript. P.R. contributed to LC-MS methodology, metabolite analysis protocols, and data interpretation. S.K. provided guidance on experimental design and supervised the project. All authors reviewed and approved the final manuscript.

Competing Interest

The authors declare no competing interests.

References

- 1 Baysoy, A., Bai, Z., Satija, R. & Fan, R. The technological landscape and applications of single-cell multi-omics. *Nat Rev Mol Cell Biol* **24**, 695-713 (2023). <https://doi.org/10.1038/s41580-023-00615-w>
- 2 Hasin, Y., Seldin, M. & Lusis, A. Multi-omics approaches to disease. *Genome Biology* **18**, 83 (2017). <https://doi.org/10.1186/s13059-017-1215-1>
- 3 Cao, Z.-J. & Gao, G. Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nature Biotechnology* **40**, 1458-1466 (2022). <https://doi.org/10.1038/s41587-022-01284-4>
- 4 Sapoval, N. et al. Current progress and open challenges for applying deep learning across the biosciences. *Nature Communications* **13** (2022). <https://doi.org/10.1038/s41467-022-29268-7>
- 5 Hu, Y. et al. Benchmarking algorithms for single-cell multi-omics prediction and integration. *Nat Methods* **21**, 2182-2194 (2024). <https://doi.org/10.1038/s41592-024-02429-w>
- 6 Cantini, L. et al. Benchmarking joint multi-omics dimensionality reduction approaches for the study of cancer. *Nature Communications* **12**, 124 (2021). <https://doi.org/10.1038/s41467-020-20430-7>
- 7 Ma, A. et al. Single-cell biological network inference using a heterogeneous graph transformer. *Nat Commun* **14**, 964 (2023). <https://doi.org/10.1038/s41467-023-36559-0>
- 8 Schäfer, P. S. L., Dimitrov, D., Villalblanca, E. J. & Saez-Rodriguez, J. Integrating single-cell multi-omics and prior biological knowledge for a functional characterization of the immune system. *Nat Immunol* **25**, 405-417 (2024). <https://doi.org/10.1038/s41590-024-01768-2>
- 9 Yuan, Q. & Duren, Z. Inferring gene regulatory networks from single-cell multiome data using atlas-scale external data. *Nat Biotechnol* **43**, 247-257 (2025). <https://doi.org/10.1038/s41587-024-02182-7>

- 10 Vandereyken, K., Sifrim, A., Thienpont, B. & Voet, T. Methods and applications for single-cell and spatial multi-omics. *Nat Rev Genet* **24**, 494-515 (2023). <https://doi.org/10.1038/s41576-023-00580-2>
- 11 Argelaguet, R. et al. MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biology* **21**, 111 (2020). <https://doi.org/10.1186/s13059-020-02015-1>
- 12 Argelaguet, R. et al. Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. *Molecular Systems Biology* **14**, e8124 (2018). <https://doi.org/https://doi.org/10.15252/msb.20178124>
- 13 Toubiana, D. & Maruenda, H. Guidelines for correlation coefficient threshold settings in metabolite correlation networks exemplified on a potato association panel. *BMC Bioinformatics* **22**, 116 (2021). <https://doi.org/10.1186/s12859-021-03994-z>
- 14 Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* **4**, Article17 (2005). <https://doi.org/10.2202/1544-6115.1128>
- 15 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008). <https://doi.org/10.1186/1471-2105-9-559>
- 16 Rohart, F., Gautier, B., Singh, A. & Lê Cao, K.-A. mixOmics: An R package for ‘omics feature selection and multiple data integration. *PLoS computational biology* **13**, e1005752 (2017).
- 17 Subramanian, I., Verma, S., Kumar, S., Jere, A. & Anamika, K. Multi-omics Data Integration, Interpretation, and Its Application. *Bioinformatics and Biology Insights* **14**, 1177932219899051 (2020). <https://doi.org/10.1177/1177932219899051>
- 18 Singh, A. et al. DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays. *Bioinformatics* **35**, 3055-3062 (2019). <https://doi.org/10.1093/bioinformatics/bty1054>

- 19 Libbrecht, M. W. & Noble, W. S. Machine learning applications in genetics and genomics. *Nature Reviews Genetics* **16**, 321-332 (2015). <https://doi.org/10.1038/nrg3920>
- 20 Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence* **1**, 206-215 (2019). <https://doi.org/10.1038/s42256-019-0048-x>
- 21 Lock, E. F., Hoadley, K. A., Marron, J. S. & Nobel, A. B. JOINT AND INDIVIDUAL VARIATION EXPLAINED (JIVE) FOR INTEGRATED ANALYSIS OF MULTIPLE DATA TYPES. *Ann Appl Stat* **7**, 523-542 (2013). <https://doi.org/10.1214/12-aos597>
- 22 Shen, R., Olshen, A. B. & Ladanyi, M. Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. *Bioinformatics* **25**, 2906-2912 (2009). <https://doi.org/10.1093/bioinformatics/btp543>
- 23 Vaswani, A. et al. Attention is all you need. *Advances in neural information processing systems* **30** (2017).
- 24 Tsai, Y.-H. H. et al. in *Proceedings of the conference. Association for computational linguistics. Meeting*. 6558.
- 25 Jiang, J. et al. Transformer technology in molecular science. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **14**, e1725 (2024).
- 26 Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021). <https://doi.org/10.1038/s41586-021-03819-2>
- 27 Xu, X., Fonseca de Lima, C. F., Vu, L. D. & De Smet, I. When drought meets heat—a plant omics perspective. *Frontiers in Plant Science* **14**, 1250878 (2023).
- 28 Ramírez-González, R. et al. The transcriptional landscape of polyploid wheat. *Science* **361**, eaar6089 (2018).
- 29 Velten, B. & Stegle, O. Principles and challenges of modeling temporal and spatial omics data. *Nature Methods* **20**, 1462-1474 (2023). <https://doi.org/10.1038/s41592-023-01992-y>
- 30 Kaur, S. et al. Temporal ordering of omics and multiomic events inferred from time-series data. *npj Systems Biology and Applications* **6**, 22 (2020). <https://doi.org/10.1038/s41540-020-0141-0>

- 31 Klec, C., Prinz, F. & Pichler, M. Involvement of the long noncoding RNA NEAT1 in carcinogenesis. *Mol Oncol* **13**, 46-60 (2019). <https://doi.org/10.1002/1878-0261.12404>
- 32 Gupta, A., Rico-Medina, A. & Cano-Delgado, A. I. The physiology of plant responses to drought. *Science* **368**, 266-269 (2020). <https://doi.org/10.1126/science.aaz7614>
- 33 Zhang, H., Zhu, J., Gong, Z. & Zhu, J.-K. Abiotic stress responses in plants. *Nature Reviews Genetics* (2021). <https://doi.org/10.1038/s41576-021-00413-0>
- 34 Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. *Advances in neural information processing systems* **30** (2017).
- 35 Hirose, T. et al. NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies. *Mol Biol Cell* **25**, 169-183 (2014). <https://doi.org/10.1091/mbc.E13-09-0558>
- 36 Cramer, G. R., Urano, K., Delrot, S., Pezzotti, M. & Shinozaki, K. Effects of abiotic stress on plants: a systems biology perspective. *BMC Plant Biology* **11**, 163 (2011). <https://doi.org/10.1186/1471-2229-11-163>
- 37 Shinozaki, K. & Yamaguchi-Shinozaki, K. Gene networks involved in drought stress response and tolerance. *J Exp Bot* **58**, 221-227 (2007). <https://doi.org/10.1093/jxb/erl164>
- 38 Skirycz, A. et al. Survival and growth of Arabidopsis plants given limited water are not equal. *Nature Biotechnology* **29**, 212-214 (2011). <https://doi.org/10.1038/nbt.1800>
- 39 Todaka, D., Shinozaki, K. & Yamaguchi-Shinozaki, K. Recent advances in the dissection of drought-stress regulatory networks and strategies for development of drought-tolerant transgenic rice plants. *Frontiers in plant science* **6**, 84 (2015).
- 40 Camacho, D. M., Collins, K. M., Powers, R. K., Costello, J. C. & Collins, J. J. Next-Generation Machine Learning for Biological Networks. *Cell* **173**, 1581-1592 (2018). <https://doi.org/https://doi.org/10.1016/j.cell.2018.05.015>
- 41 Curran, P. J. Remote sensing of foliar chemistry. *Remote sensing of environment* **30**, 271-278 (1989).

- 42 Rewald, B. & Meinen, C. Plant roots and spectroscopic methods – analyzing species, biomass and vitality. *Frontiers in Plant Science* **4** (2013). <https://doi.org/10.3389/fpls.2013.00393>
- 43 Fait, A., Batushansky, A., Shrestha, V., Yobi, A. & Angelovici, R. Can metabolic tightening and expansion of co-expression network play a role in stress response and tolerance? *Plant Science* **293**, 110409 (2020). <https://doi.org/10.1016/j.plantsci.2020.110409>
- 44 Toubiana, D., Fernie, A. R., Nikoloski, Z. & Fait, A. Network analysis: tackling complex data to study plant metabolism. *Trends in biotechnology* **31**, 29-36 (2013).
- 45 Gitelson, A. A., Gritz, Y. & Merzlyak, M. N. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *J Plant Physiol* **160**, 271-282 (2003). <https://doi.org/10.1078/0176-1617-00887>
- 46 Sims, D. A. & Gamon, J. A. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of Environment* **81**, 337-354 (2002). [https://doi.org/https://doi.org/10.1016/S0034-4257\(02\)00010-X](https://doi.org/https://doi.org/10.1016/S0034-4257(02)00010-X)
- 47 Cui, H. et al. scGPT: toward building a foundation model for single-cell multi-omics using generative AI. *Nature Methods* **21**, 1470-1480 (2024). <https://doi.org/10.1038/s41592-024-02201-0>
- 48 Sumner, L. W. et al. Proposed minimum reporting standards for chemical analysis. *Metabolomics* **3**, 211-221 (2007). <https://doi.org/10.1007/s11306-007-0082-2>
- 49 Salek, R. M., Steinbeck, C., Viant, M. R., Goodacre, R. & Dunn, W. B. The role of reporting standards for metabolite annotation and identification in metabolomic studies. *GigaScience* **2** (2013). <https://doi.org/10.1186/2047-217x-2-13>
- 50 Schölkopf, B. et al. Toward causal representation learning. *Proceedings of the IEEE* **109**, 612-634 (2021). <https://doi.org/10.48550/arXiv.2102.11107>
- 51 Tardieu, F., Simonneau, T. & Muller, B. The Physiological Basis of Drought Tolerance in Crop Plants: A Scenario-Dependent Probabilistic Approach. *Annual*

- Review of Plant Biology* **69**, 733-759 (2018). <https://doi.org/10.1146/annurev-arplant-042817-040218>
- 52 Bennani, S., Nsarellah, N., Jlibene, M. & Tadesse, W. Efficiency of drought tolerance indices under different stress severities for bread wheat selection. *Australian Journal of Crop Science* **11**, 395-405 (2017). <https://doi.org/10.21475/ajcs.17.11.04.pne272>
- 53 Hone, H. et al. Profiling, isolation and characterisation of beneficial microbes from the seed microbiomes of drought tolerant wheat. *Scientific Reports* **11**, 11916 (2021). <https://doi.org/10.1038/s41598-021-91351-8>
- 54 Yadav, A. K., Carroll, A. J., Estavillo, G. M., Rebetzke, G. J. & Pogson, B. J. Wheat drought tolerance in the field is predicted by amino acid responses to glasshouse-imposed drought. *Journal of Experimental Botany* **70**, 4931-4948 (2019). <https://doi.org/10.1093/jxb/erz224>
- 55 Roessner, U. et al. Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell* **13**, 11-29 (2001). <https://doi.org/10.1105/tpc.13.1.11>
- 56 Rochfort, S. J., Trenerry, V. C., Imsic, M., Panizzo, J. & Jones, R. Class targeted metabolomics: ESI ion trap screening methods for glucosinolates based on MSn fragmentation. *Phytochemistry* **69**, 1671-1679 (2008).
- 57 Dunn, W. B. et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nature Protocols* **6**, 1060-1083 (2011). <https://doi.org/10.1038/nprot.2011.335>
- 58 Gika, H. G., Theodoridis, G. A., Wingate, J. E. & Wilson, I. D. Within-day reproducibility of an HPLC-MS-based method for metabonomic analysis: application to human urine. *J Proteome Res* **6**, 3291-3303 (2007). <https://doi.org/10.1021/pr070183p>
- 59 Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
- 60 Xie, Y. et al. Integration of hyperspectral imaging and transcriptomics from individual cells with HyperSeq. *bioRxiv*, 2024.2001.2027.577536 (2024).

- 61 North, B. V., Curtis, D. & Sham, P. C. A Note on the Calculation of Empirical
P Values from Monte Carlo Procedures. *The American Journal of
Human Genetics* **71**, 439-441 (2002). <https://doi.org/10.1086/341527>
- 62 Karpievitch, Y. V., Dabney, A. R. & Smith, R. D. Normalization and missing value
imputation for label-free LC-MS analysis. *BMC Bioinformatics* **13**, S5 (2012).
<https://doi.org/10.1186/1471-2105-13-s16-s5>
- 63 Kokla, M., Virtanen, J., Kolehmainen, M., Paananen, J. & Hanhineva, K. Random
forest-based imputation outperforms other methods for imputing LC-MS
metabolomics data: a comparative study. *BMC Bioinformatics* **20** (2019).
<https://doi.org/10.1186/s12859-019-3110-0>
- 64 Wei, R. et al. Missing value imputation approach for mass spectrometry-based
metabolomics data. *Scientific reports* **8**, 663 (2018).

Chapter 7: General Discussion – A New Paradigm for Decoding Plant Resilience

7.1 Recapitulation: The Grand Challenge of Decoding Plant Resilience

The collision of a changing climate with a growing global population has created an unprecedented challenge for agriculture [1, 2]. Drought, in particular, has emerged as a primary and formidable threat to food security, with its escalating frequency and intensity projected to surpass even pathogen pressure in its impact on future crop yields [3, 4]. For wheat (*Triticum aestivum* L.), a crop that sustains a third of humanity, adapting to water limitation is not merely an academic pursuit but a global imperative [5]. Yet, progress is hindered by a critical chasm that has opened between our capacity to generate biological data and our ability to extract systems-level, mechanistic insight [6, 7]. While our knowledge of individual genes and metabolites has expanded exponentially, our understanding of how these components are organised into resilient, coordinated systems—the very architecture of adaptation—has been critically constrained by available analytical tools [6, 8, 9].

This gap between component-level data and systems-level understanding is not abstract; it manifests as a series of concrete analytical bottlenecks that prevent the discovery of the deep, mechanistic insights required for modern crop improvement [10]. Traditional methods, while foundational, have failed to keep pace with the complexity of the data and the systems-level questions we now need to ask [11, 12]. The research undertaken for this thesis was conceived as a direct and systematic response to this grand challenge, designed to solve four specific, interconnected bottlenecks:

- **The Analytical Complexity Bottleneck: From Manual Analysis to Automated Discovery.** High-throughput technologies now generate datasets of a scale and dimensionality that overwhelm traditional statistical methods [13-15]. The sheer volume and heterogeneity of omics and imaging data make

manual feature engineering and model selection subjective, error-prone, and unscalable, demanding a shift toward automated and intelligent analytical frameworks [16, 17].

- **The Phenotyping Bottleneck: From Human Perception to Algorithmic Vision.** Our ability to measure biological form is often constrained by the limits of our own perception. Image-based phenotyping has historically relied on quantifying traits predefined by human visual interpretation—length, angle, diameter—creating a perceptual ceiling that is blind to latent spatial patterns and subtle architectural signatures that may be critical for adaptation [6, 18]. This bottleneck requires moving beyond known measurements to the *de novo* discovery of biologically informative features directly from pixel data.
- **The Mechanistic Understanding Bottleneck: From Components to Network Architecture.** A prevalent challenge in systems biology is the reduction of high-dimensional data to a simple list of "significant features" without uncovering the principles governing their interaction [9]. This component-level view provides a parts list but not the architectural blueprint. True mechanistic insight requires understanding the system's design—the topology, connectivity, and modularity of its underlying networks—because function is an emergent property of structure [6, 19, 20].
- **The Systems Integration Bottleneck: From Correlation to Coordination.** Biological resilience is the product of exquisitely coordinated communication across multiple scales, yet most multi-omics integration methods are limited to identifying static correlations or shared variance [21, 22]. They fail to capture the directionality, feature-specificity, and temporal dynamics of the communication protocols that orchestrate adaptation [6, 12, 21]. This bottleneck prevents us from understanding how a plant truly functions as a cohesive, integrated system.

The mission of this thesis was to systematically solve each of these bottlenecks by developing an integrated computational ecosystem. Comprising AutoML for analytical complexity, the Algorithmic Root Trait (ART) framework for phenotyping,

network science for mechanistic understanding, and the MOFA+ Transformer for systems integration, this work sought to build a unified platform that establishes a new methodological paradigm for decoding the multi-scale complexity of drought adaptation. The frameworks developed herein were created as a contribution to this fundamental shift in biological discovery, where the primary constraint is no longer data generation, but the extraction of actionable, mechanistic insight [6, 7].

7.2 Synthesis of Key Findings: A Direct Response to the Research Questions

This thesis presents three principal findings that directly address the central research questions motivating this work. Each finding stems from the development of a targeted computational framework designed to address a specific analytical bottleneck, providing both methodological innovation and new biological observations.

- **Finding 1: The Algorithmic Root Trait (ART) framework provided a complementary layer of architectural information, enhancing predictive accuracy through the discovery of latent features.** When integrated with traditional metrics, ARTs enabled a more robust classification of drought tolerance. Models combining ARTs and Traditional Root Traits (TRTs) achieved the highest accuracy (97.4%), demonstrating a clear synergistic effect. The novel traits captured by ARTs also offered 5.8 \times greater information density, highlighting their efficiency for high-throughput phenotyping.
- **Finding 2: Network analysis indicated that drought tolerance is associated with distinct, tissue-specific metabolic network architectures and a biphasic temporal coordination strategy.** In the tolerant genotype, leaf networks were observed to be approximately 40% denser than root networks. This architectural asymmetry was governed by a dynamic coordination pattern where initial cross-tissue correlation ($r \approx 0.546$) strategically transitioned to a state of partial decoupling ($r \approx 0.350$) under prolonged stress, a pattern not observed in the susceptible genotype.

- **Finding 3: The MOFA+ Transformer framework quantified temporal coordination patterns, revealing that tolerant genotypes exhibited stronger and earlier cross-modal communication.** The analysis revealed that tolerant genotypes deploy quantifiably stronger physiological-biochemical coordination networks, exhibiting +355% greater average attention across key interaction pairs in leaf tissue. Critically, these networks were established at demonstrably earlier stress stages, a finding that provides quantitative evidence for a proactive, rather than reactive, stress adaptation strategy. The framework's generalisability and power to uncover novel mechanisms were demonstrated on an independent single-cell dataset. There, it moved beyond simple linear associations to discover and statistically validate a novel, non-linear relationship between a cellular spectral phenotype and the critical stress-related long non-coding RNA (lncRNA) NEAT1 ($p=0.0099$), a connection not evident in standard correlation methods. This provides a clear example of the capacity to move beyond static correlation to reveal the dynamic, directed coordination central to complex biological systems—a key challenge across all of systems biology.

7.2.1 Research Question 1: Can the computational extraction of latent architectural traits from root images provides a novel source of information that is complementary to traditional metrics, and does their integration lead to more robust and insightful predictive models?

Answer: Yes—decisively and quantifiably.

The Algorithmic Root Trait (ART) framework moves beyond the geometric simplifications of traditional phenotyping to quantify latent architectural patterns, such as the strategic clustering of root biomass, which are invisible to conventional analysis [18, 23]. The value of this novel information is clear. While models based on 23 Traditional Root Traits (TRTs) achieved a respectable 85.6% accuracy in classifying

drought-tolerant and susceptible genotypes, models using only ARTs reached 96.3% accuracy, highlighting the immense predictive power held within these algorithmically-discovered features.

However, the true strength of the ART framework lies not in replacement but in synergy. When ARTs and TRTs were combined, the model achieved the highest accuracy of 97.4%, confirming that these two approaches capture complementary facets of root biology. This breakthrough was achieved with remarkable efficiency. Just four selected ARTs delivered predictive power equivalent to all 23 TRTs, providing 5.8 \times higher information density (0.213 vs. 0.037 accuracy per feature). This dramatic increase in information content stems from capturing richer, more complex architectural information, a finding underscored by the superior stability of ARTs across diverse experimental conditions (mean CV 0.762 vs. 1.361 for TRTs, $p = 0.0137$). The ART framework thus establishes a new paradigm for image-based phenotyping, proving that computational discovery can yield a complementary set of phenotypes that, when integrated with established metrics, provide a more powerful and holistic understanding of plant adaptation.

7.2.2 Research Question 2: What are the underlying architectural principles and temporal dynamics of tissue-specific metabolic networks that drive the systemic stress response in wheat?

Answer: Our results reveal that drought tolerance in the tested wheat genotypes is strongly associated with the deployment of evolved, tissue-specific network architectures that execute a sophisticated, biphasic temporal coordination strategy.

As detailed in Chapter 5, network analysis of 2,471 molecular features revealed that resilience in this context is an emergent property of network organisation. These architectures, validated by Bayesian analysis ($p < 0.001$) as evolved strategies, are not stochastic but are optimised for distinct functional roles. Leaf networks in

tolerant genotypes are engineered for rapid, robust coordination, exhibiting ~40% greater density and higher transitivity to maintain photosynthetic stability. In contrast, their root networks display a more modular and fragmented architecture, a design that enables flexible, localised adaptation to heterogeneous soil conditions.

The most sophisticated discovery was the identification of a biphasic temporal strategy that distinguishes tolerant from susceptible genotypes. Tolerant plants execute a strategic orchestration where initial strong cross-tissue coordination, reflected in a metabolite correlation of $r \approx 0.546$, transitions to a state of strategic decoupling, with the correlation declining to $r \approx 0.350$ under prolonged stress. This dynamic pattern, which allows for tissue-specific optimisation while maintaining whole-plant coherence, was entirely absent in susceptible genotypes, which showed unstable, oscillating integration. These findings underscore that the architectural organisation of a plant's metabolic response is a critical determinant of resilience, not merely the changes in its individual molecular features.

7.2.3 Research Question 3: What time-resolved coordination patterns between physiological and biochemical responses distinguish drought tolerant and susceptible genotypes, and can interpretable deep learning reveal these cross-modal relationships?

Answer: Our results indicate that tolerant genotypes deploy quantifiably stronger, earlier, and more precisely directed cross-modal communication networks, a strategy that appears central to their enhanced resilience.

The MOFA+ Transformer framework revealed the temporal choreography of resilience by quantifying the dynamic dialogue between a plant's physiological status and its biochemical state. A key finding is that enhanced tolerance is strongly associated with the precision and timing of coordination, rather than being solely dependent on individual molecular or physiological advantages. Tolerant genotypes exhibited 355% greater Coordination Strength (average attention across top S2M

pairs) in leaf tissue alongside a more targeted response in the roots, representing a fundamental difference in coordination capacity. These plants establish critical communication networks at quantifiably earlier stress stages, deploying what can be described as an "early warning system" that enables a proactive, rather than reactive, stress response.

The framework identified key molecular features (MF N_1909 in leaves, MF N_0512 in roots) that functioned as coordination hubs, receiving the strongest cross-modal attention from distinct communication channels (visible vs. NIR spectra). This suggests a functionally specialised sensing system. The framework's power was validated on an independent single-cell dataset, where it discovered and statistically confirmed a novel, non-linear link between a cellular phenotype and the stress-related lncRNA NEAT1 ($p=0.0099$), a connection invisible to standard correlation methods. By integrating variance-driven (MOFA+) and prediction-driven (SHAP) analyses, which showed only 23.46% overlap, this work also revealed that the drivers of biological variance and predictive accuracy are often distinct, providing a novel and powerful dual perspective on feature importance.

7.3 A Multi-Scale Perspective on Drought Resilience

The key findings detailed in Section 7.2, when synthesised, support a coherent, multi-layered model of drought resilience. This model posits that resilience in the studied genotypes is not determined by any single trait but rather emerges from coordinated optimisation across multiple organisational scales—a foundational concept in systems biology when studying complex stress responses [24-26]. This perspective, visually summarised in Figure 7.1, integrates three distinct but interconnected layers that appear to build upon one another: the Physical Architecture, the Metabolic Organisation, and the Coordination Protocols.

The Architecture of Resilience: A Coordination-Centric Framework

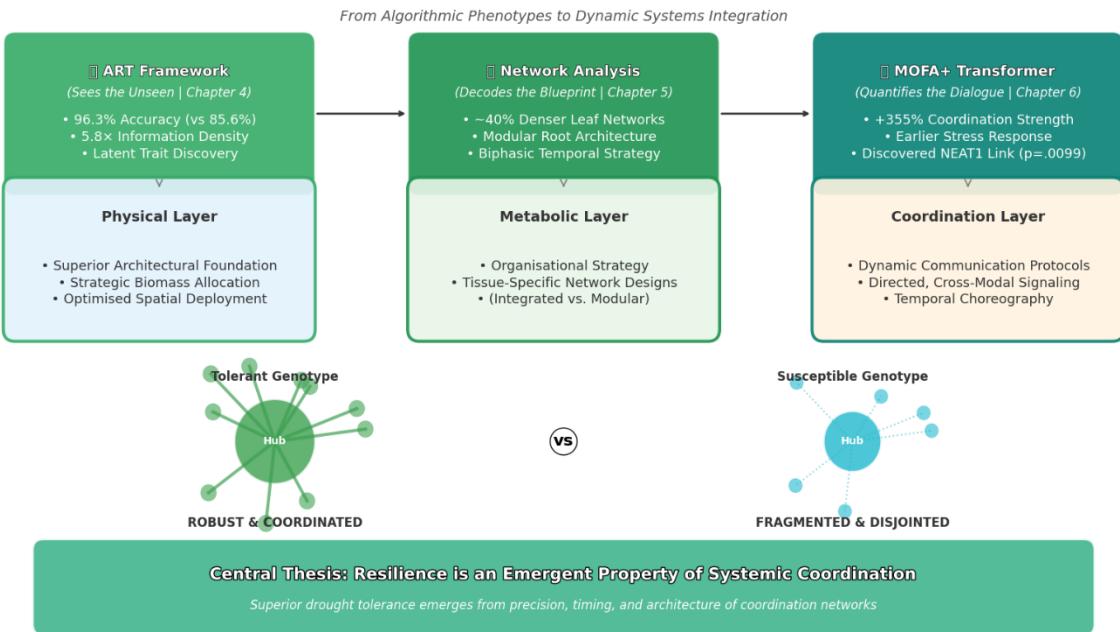


Fig. 7.1. A Unified, Multi-Layer Model of Wheat Drought Tolerance. This figure synthesises the central biological findings of the research thesis into a multi-scale model that contrasts a tolerant and susceptible genotype. The model is structured hierarchically. (Top Row) Three interconnected layers represent the progression of discovery: the Physical Layer (Chapter 4), where Algorithmic Root Traits (ARTs) revealed a superior architectural foundation with high predictive power (e.g., 5.8× information density); the Metabolic Layer (Chapter 5), where network analysis decoded tissue-specific organisational blueprints (e.g., ~40% denser leaf networks); and the Coordination Layer (Chapter 6), where the MOFA+ Transformer quantified the dynamic communication protocols (e.g., +355% stronger coordination strength). (Bottom Row) These layers culminate in distinct systemic outcomes, visually contrasted by a "Robust & Coordinated" network for the tolerant genotype and a "Fragmented & Disjointed" network for the susceptible genotype. The figure encapsulates the thesis's central biological conclusion: that resilience is an emergent property of systemic coordination.

The Physical Layer: An Architectural Foundation for Resource Capture

The architectural sophistication revealed by ART analysis suggests that drought tolerance begins with a superior physical scaffold for resource capture. The computational discovery of latent architectural features reveals that tolerant genotypes deploy strategically positioned dense root clusters to optimise water acquisition—patterns that represent evolved solutions to the fundamental challenge of resource capture under stress [27-29]. These algorithmically-discovered features capture critical dimensions of adaptive root deployment, including strategic

biomass allocation, enhanced lateral exploration, and optimal depth positioning [30]. This physical architecture appears to provide the essential substrate upon which more sophisticated metabolic and coordination strategies can operate [24]. These ARTs provide quantitative evidence for the adaptive root deployment strategies, such as strategic biomass allocation and hydropatterning, detailed in Chapter 2.

The Metabolic Layer: An Organisational Blueprint for Stress Management

Built upon this physical foundation is the plant's metabolic organisation—its strategic blueprint for managing stress. The discovery of a striking architectural asymmetry between tissues reflects an evolved specialisation for different functional roles, a known phenomenon in plant stress response [30-32]. The highly integrated design of leaf networks appears engineered for rapid, robust coordination to protect photosynthesis, while the modular architecture of root networks seems optimised for flexible, localised adaptation to the heterogeneous soil environment. This interpretation aligns with network biology principles where network topology dictates functional capacity [8, 19, 20, 33]. This organisational layer is governed by a sophisticated biphasic timing strategy, where an initial, tightly coupled whole-plant response transitions to decoupled, tissue-specific autonomy. This dynamic pattern, which appears to allow for a shift from an emergency response to a sustained, efficient mode of operation, was absent in susceptible genotypes, highlighting the importance of temporal dynamics in adaptive processes [34]. This discovery of distinct network blueprints provides the mechanistic basis for the functional root-shoot specialisation under drought that was outlined in Chapter 2.

The Coordination Layer: Communication Protocols for Real-Time Adaptation

The final and most dynamic layer consists of the communication protocols that execute the metabolic strategy in real-time. The superior communication protocols decoded by the MOFA+ Transformer reveal how physiological signals direct biochemical responses across tissues and time. This framework's ability to quantify directed, feature-specific relationships addresses a key challenge in systems

biology: moving beyond static snapshots to understand the temporal dynamics of biological processes [6, 12, 34]. The earlier establishment of these coordination networks in tolerant genotypes implies an anticipatory response system that enables proactive rather than reactive stress management, a key feature of the "early response hypothesis" [35, 36]. The temporal choreography revealed by this analysis suggests that coordination timing itself is a critical and measurable component of biological resilience [26, 35]. The quantification of earlier and stronger coordination provides direct evidence for the superior temporal control and communication protocols that, as discussed in Chapter 2, are critical for effective whole-plant resilience.

Systems Integration: A Fully Coordinated System

These three layers—physical, metabolic, and coordinative—appear to function as an integrated hierarchy where each provides the foundation for the next [37-40]. The evidence presented in this thesis suggests a model where superior physical architecture provides the spatial framework for efficient resource acquisition; tissue-specific metabolic organisation provides the strategic blueprint for resource utilisation [8, 20, 41]; and precision coordination protocols provide the dynamic, real-time control necessary for adaptation [24, 26, 42]. This concept of functional convergence—where, for example, architecture must align with hydraulics, and root function with shoot demand—is a recognised principle in plant systems biology [43, 44]. The analysis of susceptible genotypes further supports this systems-level view, suggesting that resilience breakdown occurs not through the failure of a single component, but through a cascade of compromised integration across layers [45-47]. This culminates in the central hypothesis of this thesis: that heritable plant resilience is fundamentally an emergent property of the coordination capacity of the entire system—a quality that is distinct from, and perhaps more critical than, the inherent superiority of its individual molecular components [8, 19, 31, 48-50].

7.4 Methodological Contributions: An Integrated Analytical Ecosystem

The biological perspective presented in Section 7.3 was made possible only through the development of a novel computational platform. This research did not produce a series of disconnected tools, but rather an integrated methodological ecosystem, where each framework was designed to solve a specific analytical bottleneck while synergistically amplifying the capabilities of the others. This ecosystem represents a systematic response to the evolving paradigm in modern biology, where the analytical bottleneck is increasingly defined not by data generation, but by the extraction of mechanistic insight [6, 7] (Fig. 7.2). The four core frameworks—AutoML, ART, the Network Analysis Pipeline, and the MOFA+ Transformer—constitute the pillars of this analytical architecture.

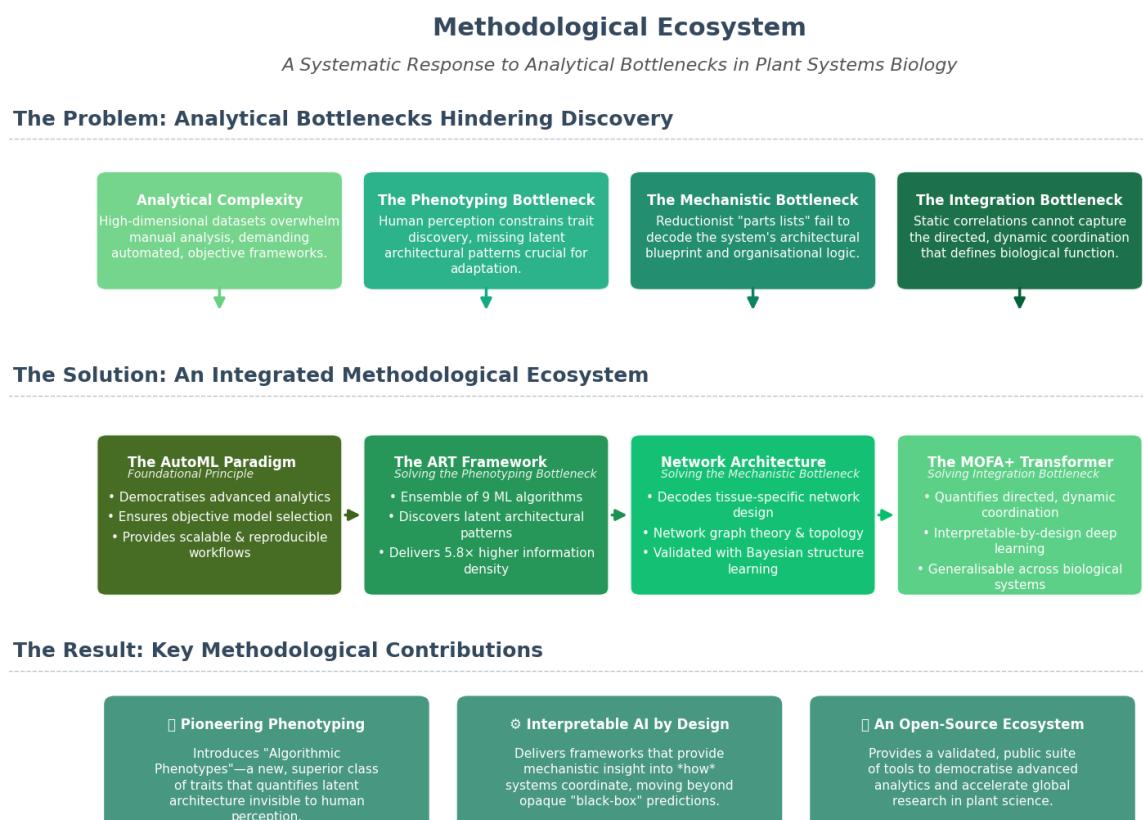


Fig. 7.2. The Methodological Ecosystem: A Systematic Framework for Solving Analytical Bottlenecks in Plant Systems Biology. This schematic illustrates the central argument and structure of the thesis. (Top Panel) The four key analytical bottlenecks that constrain biological discovery are identified as the

primary challenge. (Middle Panel) The thesis's "Solution" is presented as an integrated methodological ecosystem, where each core framework—AutoML, the ART Framework, Network Architecture analysis, and the MOFA+ Transformer—is purpose-built to solve a corresponding bottleneck. This represents the progression from establishing the principle of automation to pioneering new phenotyping methods, decoding mechanistic blueprints, and finally quantifying systemic integration. (Bottom Panel) The "Result" of this ecosystem is a set of key methodological contributions that collectively enable a new paradigm for biological discovery, centered on interpretable AI, pioneering phenotyping, and open-source innovation.

7.4.1 AutoML: Establishing the Principle of Automated Discovery

While not a direct experimental contribution, the perspective on Automated Machine Learning (AutoML) presented in Chapter 3 serves as the philosophical and practical cornerstone of this thesis's methodological ecosystem. Its primary contribution is the introduction and synthesis of the AutoML paradigm for the plant science community, a field where these powerful tools remain underutilised despite their immense potential. This chapter directly addresses the **Analytical Complexity Bottleneck** by demonstrating how the overwhelming scale and multivariate nature of modern biological datasets can be managed through automation [14, 15].

The chapter argued that by automating the entire machine learning pipeline—from feature selection to model optimisation—AutoML democratises access to advanced analytics [51, 52]. This empowers domain experts, such as plant scientists, to lead sophisticated data analysis without requiring deep programming expertise, thereby bridging a critical gap between data generation and insight extraction [16]. By establishing the principles of automated, scalable, and reproducible analysis, this chapter lays the conceptual groundwork for the bespoke experimental frameworks developed to solve the subsequent, more specific bottlenecks in phenotyping, mechanistic understanding, and systems integration.

7.4.2 The ART Framework: Transcending Human Perceptual Limitations

The Algorithmic Root Trait (ART) framework was developed to address the phenotyping bottleneck by introducing the concept of "algorithmic phenotypes," designed to move beyond the constraints of human perceptual limitations [18, 23].

The framework leverages an ensemble of nine distinct unsupervised learning algorithms that work in concert to autonomously discover and quantify latent spatial patterns within root [53-59]. This approach provided a significant uplift in predictive power by capturing novel architectural information missed by traditional methods. ART-based models achieved 96.3% accuracy compared to 85.6% for models based on Traditional Root Traits (TRTs).

The synergistic combination of both trait sets yielded the highest overall performance (97.4%), underscoring that ARTs and TRTs capture complementary aspects of root biology. This finding suggests that while algorithmically-derived traits provide superior predictive power, integrating them with established geometric descriptors offers the most comprehensive view of root system architecture. Critically, unlike opaque "black-box" models, ARTs remain fully interpretable (e.g., `_density_points`, `_centre_y`), directly linking computational outputs to established biological strategies like resource allocation and rooting depth [29, 60]. Furthermore, the framework's superior stability across diverse experimental conditions ($p = 0.0137$) supports its potential for practical deployment in variable field environments.

7.4.3 The Network Analysis Pipeline: Decoding Architectural Blueprints

To address the mechanistic understanding bottleneck, this research developed a comprehensive pipeline that shifts the analytical focus from a simple list of significant metabolites [9] to the organisational principles of the entire metabolic network. This was achieved through an integrated workflow combining robust feature selection with multi-tiered statistical validation, including Bayesian structure learning and permutation testing [61, 62]. This methodological rigour enabled the discovery of the fundamental architectural blueprints of adaptation. It was this pipeline that revealed the tissue-specific network designs—the ~40% denser, integrated architecture of leaves versus the modular structure of roots—and the sophisticated biphasic temporal strategy governing their coordination. By providing the tools to move beyond a component-level view, this pipeline delivers the means to decode a system's underlying design principles, answering not just *what* molecules are involved, but *how* they are organised [19, 20].

7.4.4 The MOFA+ Transformer: Interpretable Deep Learning for Systemic Coordination

The MOFA+ Transformer represents the culmination of this thesis's methodological contributions, created to address the systems integration bottleneck. It moves beyond the limitations of traditional correlation-based methods [21] through a novel two-stage architecture that fuses Multi-Omics Factor Analysis+ (MOFA+) for robust variance decomposition with an interpretable Transformer for quantifying directed interactions [63-66]. This hybrid design provides a powerful new ability to quantify the temporal choreography of resilience. It was this framework that revealed the 355% greater Coordination Strength in the leaf tissue of tolerant genotypes and identified the key molecular features functioning as coordination hubs for the stress response. This was complemented by the discovery that the features driving biological variance and predictive accuracy are largely distinct, an insight enabled by integrating MOFA+ with SHAP.

The framework's power and generalisability were demonstrated on an independent public dataset [67], where it discovered and statistically validated a novel, non-linear link between a cellular phenotype and the critical stress-related lncRNA NEAT1 ($p=0.0099$) [68]. By delivering quantifiable, dynamic, and directed relationships, the MOFA+ Transformer helps shift multi-omics analysis from a correlative exercise toward a mechanistic discovery engine, providing a direct response to the need for next-generation tools in biological network analysis [6].

7.5 Implications: New Perspectives for Crop Improvement and Plant Systems Biology

The methodological ecosystem and biological insights developed in this thesis represent more than solutions to specific research problems; they provide a new lens through which to view, measure, and enhance plant resilience. The findings have significant implications, offering new perspectives for both applied crop

improvement and the fundamental study of plant systems biology. As climate change intensifies agricultural challenges worldwide [1, 2], the frameworks established here offer analytical tools and mechanistic insights that can contribute to accelerating the development of the next generation of climate-resilient crops [5, 69, 70].

7.5.1 Applied Implications for Crop Improvement

The urgent need to develop climate-resilient cultivars requires novel strategies that can accelerate breeding cycles and improve selection accuracy, moving beyond traditional, often slow and laborious, selection methods [3, 71, 72]. This research suggests several promising pathways to address this global challenge.

Novel, Mechanistically-Grounded Biomarkers for Precision Breeding

This work identified a new suite of high-value biomarkers that offer a novel approach to a central challenge in crop improvement: moving beyond selection for complex, environmentally-confounded end-point traits like grain yield [3, 73]. The Algorithmic Root Traits (ARTs) and network properties discovered here represent a significant advancement. Specific ARTs like HDBSCAN_density_points and FCM_centre_x are not just predictive; they appear to represent heritable architectural signatures that quantify the very strategies—biomass allocation and spatial deployment—that enable efficient water acquisition, providing a direct, measurable link to the principles of established root ideotypes for water-limited environments [29, 74]. Similarly, the identification of key metabolic network hubs and the quantifiable metric of "coordination strength" define novel, mechanistically-grounded trait classes that capture the dynamic communication capacity of a resilient plant. These traits could be integrated into modern genomic prediction models to enhance selection accuracy and support the development of the 'designer roots' envisioned for future agriculture [72, 74].

Addressing the Phenotyping Bottleneck in Breeding

The frameworks developed here offer a scalable approach to alleviating the primary rate-limiting step in modern breeding: the phenotyping and analysis bottleneck [14, 75, 76]. The automated and high-throughput nature of the ART and MOFA+ Transformer frameworks provides analytical pipelines that enable the analysis of vast germplasm collections with high speed and precision. The demonstrated stability of these frameworks across diverse experimental conditions is essential for practical deployment in multi-environment trials, where genotypic performance must be reliably assessed across fluctuating conditions [77]. This acceleration is valuable for leveraging modern strategies like 'speed breeding' [78, 79] and for advancing the vision of AI-driven smart breeding programs [15, 80].

Enabling Climate-Smart Agriculture

By providing mechanistic insights into water-use efficiency and resilience coordination, this work contributes to the development of cultivars that maintain productivity with reduced external inputs—a cornerstone of climate-smart agriculture [81, 82]. The finding that resilience is associated with optimised coordination rather than simply increased resource consumption informs breeding strategies that enhance a plant's inherent biological efficiency. The architectural strategies inferred from ART and network analyses, such as deeper rooting and optimised biomass allocation, are consistent with traits known to enhance water-use efficiency [81, 83]. Furthermore, these same architectural principles could support sustainable agricultural practices by promoting improved soil carbon sequestration [82, 84-86].

7.5.2 Conceptual Implications for Plant Systems Biology

Beyond its applied utility, this thesis introduces conceptual shifts that redefine how we approach the study of complex biological systems, offering new paradigms for understanding resilience and adaptation.

From Components to Architecture

Much of molecular biology has operated under a reductionist framework, seeking to understand systems by characterising their individual parts—a limitation noted in a foundational review by Fernie and Stitt [9]. The findings in this thesis underscore that for complex traits like drought tolerance, the organisational architecture of a biological system is a critical determinant of its function, alongside the individual molecular components it contains. The discovery of distinct, tissue-specific metabolic network architectures challenges approaches that focus on individual genes or metabolites in isolation. This work offers strong empirical support for foundational network biology principles, which posit that function is an emergent property of topology [19, 20, 33]. In doing so, it highlights the value of shifting the research focus from "which genes are up-regulated?" to "what are the design principles of the responsive network?"

From Static to Dynamic Analysis

Biological resilience is a process, not a state, yet traditional multi-omics studies often provide only snapshots, missing the crucial temporal dimension of adaptation, a challenge highlighted by Bar-Joseph, et al. [34]. The MOFA+ Transformer's ability to quantify the temporal evolution of cross-modal coordination demonstrates that resilience is fundamentally characterised by these dynamic patterns. The discovery of the biphasic coordination strategy and the "early warning networks" in tolerant genotypes establishes that the speed, timing, and precision of biological responses are themselves primary, measurable, and likely heritable components of adaptation. This temporal sophistication challenges static analytical approaches and establishes the need for frameworks that can capture coordination dynamics across multiple time scales [26, 35, 37].

From Correlation to Interpretable Coordination

The systems biology field has been dominated by correlation-based analyses, typified by methods like Weighted Gene Co-expression Network Analysis (WGCNA) [21], which are powerful for pattern discovery but fundamentally limited in their ability to infer mechanism [12]. The MOFA+ Transformer moves plant systems biology beyond identifying co-varying features toward the quantification of directed, feature-specific relationships that enable mechanistic hypothesis generation. This transition is exemplified by the framework's ability to produce precise, testable predictions. By providing a clear path from high-dimensional data to systems-level understanding, this work establishes a new standard for mechanistic discovery that can be applied across biological disciplines [6].

7.5.3 A Practical Roadmap for Implementation: An Integrated Phenotyping and Analysis Platform for Crop Improvement

The methodological ecosystem developed in this thesis offers more than a set of analytical tools; it provides a blueprint for an integrated phenotyping and analysis platform designed to accelerate crop improvement. Translating these advanced frameworks into routine application within breeding programs requires a practical roadmap that addresses key considerations from experimental design to data interpretation and sensor selection [10, 87]. This section outlines a tiered strategy for deploying this ecosystem, providing recommendations on sensing modalities and experimental design to help overcome the persistent phenotyping bottleneck and directly integrate these innovations into modern breeding pipelines [14, 75, 88].

A central challenge in phenotyping is bridging the gap between the precision of controlled environments and the complexity of field conditions, where genotype-by-environment interactions often alter performance rankings [10, 77]. We propose a tiered strategy that strategically allocates resources by beginning with broad, cost-effective screening and progressively focusing more intensive, mechanistic analyses on the most promising candidates. The initial stage would involve high-throughput screening of large germplasm collections in controlled environments, such as glasshouse rhizotrons. Here, low-cost RGB imaging coupled with the ART framework

allows for the rapid, scalable quantification of latent architectural traits. As demonstrated in Chapter 4, this approach maximises throughput and provides high information density ($5.8\times$ greater than traditional metrics), making it an exceptionally efficient primary screen for identifying genotypes with superior root system architecture [89, 90].

From this broad initial screen, a smaller subset of elite candidates would advance to a second tier focused on mechanistic validation. This stage moves beyond what traits are beneficial to understanding why by decoding the underlying principles of resilience. This requires multi-modal sensing, where hyperspectral imaging (350–2500 nm) captures detailed physiological status and untargeted LC-MS metabolomics reveals the plant's biochemical state [91–93]. The MOFA+ Transformer (Chapter 6) is deployed at this stage to dissect the coordination networks, identifying which genotypes exhibit superior, early, and efficient cross-modal communication. The final stage involves testing these mechanistically-understood genotypes in multi-environment field trials. In-situ validation can be achieved using field-deployable sensors, such as minirhizotrons for root analysis with the ART framework [94, 95], alongside drone-based (UAV) hyperspectral and thermal imaging to gather scalable physiological data at the canopy level [96, 97].

The selection of sensing technology is therefore tailored to the goals of each tier. For architectural phenotyping, RGB imaging offers the best balance of cost and scalability in controlled settings, while minirhizotrons are essential for field validation. For the deeper physiological and coordination analysis central to the MOFA+ Transformer, hyperspectral imaging is the gold standard for mechanistic discovery, providing comprehensive data on plant health [91, 98]. For larger-scale field applications, other sensors can be integrated; thermal imaging, for instance, serves as an excellent proxy for stomatal conductance and can be incorporated as a physiological data view [99, 100]. Looking forward, the framework's modularity allows for the integration of 3D imaging modalities, such as using Structure-from-Motion to generate point clouds from UAV imagery. This could add a physical trait view to link coordination dynamics with whole-plant biomass and canopy architecture [90].

By combining this tiered experimental design with targeted sensor deployment, the analytical ecosystem presented in this thesis offers a direct pathway to accelerate selection cycles. Providing high-confidence, mechanistically-grounded biomarkers (e.g., specific ARTs, "coordination strength") early in the breeding pipeline can significantly enhance decision-making, improve genetic gain, and reduce the time and cost required to develop and deliver the next generation of climate-resilient cultivars [71, 78, 79].

7.6 Limitations and Future Directions

While this thesis establishes a novel analytical framework for plant systems biology and delivers significant methodological and biological advances, transparent acknowledgment of current limitations provides essential context for the findings and identifies critical directions for future research. Rather than fundamental flaws, these limitations represent exciting opportunities for expanding the scope and impact of the methodological ecosystem developed here.

7.6.1 Genetic Diversity and Generalisability Constraints

Limitation: The generalisability of the systems-level discoveries in this thesis is constrained by a deliberate experimental design that prioritised mechanistic depth over broad genetic scope for the most intensive analyses. This thesis employed a strategic, two-phase approach:

Phase 1 - Framework Development (Chapter 4): The initial phase utilised six wheat genotypes with documented, contrasting drought responses to establish the ART framework's applicability. This broader genetic sampling, conducted across both controlled glasshouse and variable field experiments, was necessary to validate the framework's robustness and general utility for phenotyping diverse germplasm.

Phase 2 - Systems-Level Analysis (Chapters 5 & 6): The subsequent analysis focused on two representative genotypes (G1: tolerant; G2: susceptible) to enable the intensive, multi-modal data collection required for systems-level investigation. This

narrower focus was necessary to achieve the temporal resolution and data density needed to characterise network architectures and coordination mechanisms.

Impact: While this design enabled deep mechanistic insights, the focused genetic scope for the systems-level work means that the discovered coordination principles and network architectures should be considered well-supported hypotheses requiring validation across broader genetic diversity [3, 101]. The specific molecular hubs and timing patterns identified through MOFA+ Transformer analysis may represent effective strategies within these genotypes. Whether these mechanisms are universal across diverse wheat germplasm remains a key question for future investigation.

Solution Pathway: The established ART framework, validated across multiple genotypes and environments in Phase 1, provides a robust tool for scaling phenotyping to larger genetic panels. The MOFA+ Transformer methodology can now be applied to diversity panels from breeding programs to test the generality of coordination principles and distinguish conserved mechanisms from genotype-specific adaptations [3, 71, 72]. The framework's demonstrated power on an independent single-cell dataset—where it successfully identified a validated, non-linear link to the highly conserved stress-response lncRNA NEAT1 [67, 68]—underscores its potential to uncover significant biological interactions regardless of the specific genotype or organism. Integration with genomic resources will enable the identification of the genetic basis underlying coordination capacity, providing a clear path from the findings of this thesis to applied breeding outcomes [10, 69].

7.6.2 Environmental and Experimental Scope

Limitation: The experimental work was conducted under controlled glasshouse conditions using hydroponic systems, which, while enabling precise control of stress treatments and high-quality data collection, may not fully capture the complexity of field environments where plants experience multiple, interacting stresses [77]. The temporal framework focused on early stress responses (0-7 days), potentially missing longer-term adaptation strategies that develop over weeks or months.

Impact: The coordination patterns and network architectures identified may not translate directly to field conditions where plants face fluctuating water availability, soil heterogeneity, pathogen pressure, and varying climatic conditions [76, 102]. The early-response focus may overlook important adaptive strategies that emerge during prolonged stress or recovery phases, limiting the complete understanding of drought resilience mechanisms [26, 37].

Opportunities for Further Research: The findings from this controlled-environment study provide a strong foundation for targeted field validation. The logical next step is to test the robustness and applicability of these discovered principles under realistic agricultural conditions.

Field-Scale Validation: The analytical frameworks developed here are designed for such a transition. The ART framework, for instance, can be readily applied to lower-resolution field imagery (e.g., from minirhizotrons or soil cores), while the MOFA+ Transformer can integrate field-collected sensor data (e.g., drone-based hyperspectral) with tissue-level omics. This modularity allows for the direct assessment of whether the identified coordination patterns hold true across diverse climatic zones and soil types.

Extended Temporal Analysis: Future studies should also extend the temporal analysis window to capture the full cycle of plant responses, including long-term acclimation and recovery phases. This would be crucial for understanding phenomena like stress memory and the cumulative impact of intermittent drought events [35, 103].

7.6.3 Methodological Extensibility: Integrating New Omics Layers and Causal Inference

Limitation: The current implementation of the framework was designed to establish a proof-of-concept for dynamic integration, focusing on physiological (hyperspectral) and biochemical (metabolomic) data. As such, it does not yet incorporate transcriptomic or proteomic layers. Furthermore, while the Transformer learns directed, predictive relationships, these learned associations are inherently correlational, not causal.

Impact: The absence of transcriptomic and proteomic data constrains a full understanding of the upstream regulatory programs (e.g., transcription factors) and downstream functional execution (e.g., enzyme activity) that govern the observed coordination patterns [35]. Similarly, while the framework moves beyond simple correlation by identifying directed predictive links, these associations cannot, on their own, definitively establish the causal pathways that structure biological networks [6].

Opportunities for Future Development: The MOFA+ Transformer was engineered with extensibility as a core design principle, creating clear opportunities for future expansion.

Multi-Omics Expansion: The framework's modular architecture is designed to accommodate additional data modalities. This flexibility was demonstrated by its successful application to the independent HyperSeq dataset, which seamlessly integrated transcriptomic data instead of metabolomics. Future work will involve incorporating RNA-seq and proteomic datasets from the primary wheat experiment to build a more comprehensive, multi-layered model of regulation, linking gene expression to protein function and metabolic output.

Moving Toward Causal Inference: A key priority for future research is to augment the framework with causal discovery methods. This could involve using the learned attention weights as priors for Bayesian network structure learning or integrating techniques from dynamic causal modelling to move from predictive association toward models that can infer causal influence [6, 7]. This represents a critical step toward creating truly predictive models of biological system behaviour.

7.6.4 Generalisability Across Species and Stress Conditions

While this thesis utilised wheat drought stress as a deep, mechanistically-rich model system to establish foundational principles, the analytical frameworks themselves were engineered for broad utility. Their core strength lies in an ability to integrate heterogeneous data and decode complex coordination patterns—a universal challenge in systems biology. The frameworks are therefore not limited to a single stress or species. The successful application of the MOFA+ Transformer to a human

single-cell dataset (HyperSeq), where it uncovered a validated link to the highly-conserved stress-response gene *NEAT1* [68], strongly supports its versatility. The following examples illustrate how this ecosystem can be adapted to other critical challenges in crop science.

- Heat Stress: The frameworks can be used to decode thermotolerance by integrating thermal imaging data as a primary physiological view [104]. The MOFA+ Transformer could then quantify the coordination between canopy temperature (thermal data) and the molecular heat-shock response (proteomics) or the production of protective metabolites (metabolomics) [105]. This would allow researchers to ask questions like: "Do heat-tolerant genotypes exhibit earlier or stronger coordination between physical cooling mechanisms and the cellular heat-shock protein response?"
- Salinity Stress: To study salinity, ionomics data (e.g., Na^+/K^+ ratios) can be incorporated as a new omics layer within the MOFA+ framework [106]. This would enable the modelling of the critical trade-off between ion toxicity and osmotic adjustment. The framework could reveal how tolerant genotypes coordinate ion exclusion mechanisms in roots with the synthesis of compatible osmolytes in leaves, providing a systems-level view of salt tolerance that transcends single-gene analysis [107, 108].
- Biotic Stresses: The analytical ecosystem is also applicable to understanding plant-pathogen interactions. For a foliar disease, hyperspectral imaging can detect pre-symptomatic stress signatures [66]. The MOFA+ Transformer could then link these early physiological flags to the subsequent activation of specific defense-related metabolic pathways (e.g., phytoalexin production) [109]. For soil-borne pathogens, the ART framework could quantify disease-induced changes in root architecture, capturing how plants adapt their foraging strategy in the presence of root pathogens [110]. In both cases, the frameworks provide tools to decode the temporal dynamics of a successful immune response.

These examples highlight that the methodological ecosystem is stress-agnostic. The path forward involves applying these validated tools to new biological questions,

leveraging their flexibility to integrate the most relevant data modalities for the system under study. To accelerate this, establishing collaborations to test these frameworks across diverse species, environments, and stress types is a key priority.

7.6.5 Technical and Computational Considerations

Limitation: The advanced analytical power of the MOFA+ Transformer framework comes with substantial computational requirements, particularly for the large-scale, augmented datasets used in this thesis. Effective implementation and interpretation currently require access to high-performance computing resources and expertise in computational biology.

Impact: These technical requirements may present a barrier to widespread adoption, particularly in resource-limited research environments where the need for advanced crop improvement tools is often greatest [14, 15]. Without efforts to improve accessibility, the framework's use could be confined to specialised research groups, limiting its potential for broad community-wide impact.

Pathways to Broader Accessibility: A central goal of this work is to ensure these tools are widely available. Several key initiatives are planned to democratise access and facilitate global adoption:

Cloud-Based Implementations and User-Friendly Interfaces: Future development will focus on creating streamlined, cloud-based versions of the analytical pipelines. This will eliminate the need for local high-performance hardware. Paired with the development of user-friendly graphical interfaces, this will significantly lower the technical barrier to entry for framework deployment.

Open-Source Distribution and Documentation: The frameworks are distributed as open-source code, accompanied by comprehensive documentation and step-by-step tutorials. This commitment to open science is intended to facilitate transparent and reproducible research worldwide [15].

Training and Capacity Building: Collaboration with international agricultural research centers will be prioritised to provide training workshops and build local capacity in framework implementation. This will help ensure the tools can be effectively deployed where they are needed most.

7.6.6 Future Research Priorities

The frameworks and insights from this thesis create a launchpad for several critical and exciting avenues of future research. These priorities focus on expanding the biological scope of the findings, integrating them with applied technologies, and moving from predictive association toward causal understanding.

Expanding Biological and Environmental Scope: A primary goal is to test the generalisability of the discovered coordination principles. This involves applying the analytical ecosystem to **broader genetic diversity panels** (e.g., in maize, rice, and barley) and across a range of **abiotic stresses** (e.g., heat, salinity) to distinguish conserved resilience mechanisms from species- or stress-specific adaptations [37, 70]. Furthermore, expanding the analysis to different **developmental stages** and tracking coordination dynamics through **complete stress-and-recovery cycles** will provide more comprehensive models of plant resilience over an entire lifecycle [26, 35].

Integration with Breeding and Precision Agriculture: To accelerate the translation of these findings into practice, future work will focus on integrating the discovered coordination principles with modern breeding technologies. This includes developing **molecular markers for "coordination capacity"** to be used in genomic selection models and identifying key network hubs as potential targets for gene editing. The high-throughput nature of the ART framework makes it ideal for large-scale germplasm screening, while the MOFA+ Transformer can be adapted to analyse field-scale sensor data (e.g., from drones), directly supporting precision agriculture platforms [10, 72].

From Mechanistic Hypothesis to Causal Validation: The interpretable outputs of the MOFA+ Transformer and network analyses generate specific, testable hypotheses about biological coordination. The next crucial step is

functional validation. This involves using reverse genetics (e.g., CRISPR-Cas9) to perturb predicted hub nodes (like MF N_1909) to experimentally confirm their causal role in orchestrating the stress response. Integrating these experimental results with **causal inference models** will be essential for moving from predictive association to true, predictive simulation of biological system behaviour under diverse scenarios [6, 37].

Cross-Disciplinary Applications of Coordination Analysis: The principles of multi-scale coordination are not limited to plants. A final, exciting direction is to explore the application of this analytical ecosystem to other complex biological systems. Fields such as **microbial ecology** (analysing community metabolic networks), **human health** (integrating clinical and multi-omics patient data), and **ecosystem science** could all benefit from tools that quantify dynamic, directed coordination. Such applications would not only demonstrate the broad utility of the framework but also help validate the fundamental principles of biological organisation discovered here [6, 7].

7.7 Conclusion: The Architecture of Resilience

This thesis began by addressing four analytical bottlenecks that have constrained our ability to translate burgeoning biological data into climate-resilient crops. It concludes by offering an integrated computational ecosystem developed to systematically tackle each one, contributing to a framework for biological investigation where artificial intelligence can amplify, rather than replace, biological insight [6]. The journey from automating analysis (AutoML), to seeing the unseen (ART), to decoding network architecture, and finally to quantifying systemic coordination (MOFA+ Transformer) has yielded a unified, multi-scale model of resilience that provides both methodological innovation and new biological understanding.

7.7.1 The Paradigm Revealed: Coordination as the Architecture of Resilience

The central finding of this work establishes a fundamental principle for understanding complex biological traits: biological resilience emerges not from the superiority of individual components, but from the precision, timing, and architecture of their coordination [35, 111]. This coordination-centric paradigm, validated at every scale of this thesis, represents a significant shift from reductionist approaches that catalogue molecular parts [9] toward a systems-level understanding that decodes organisational principles [19, 20].

This coordination framework manifests at multiple scales throughout our discoveries. At the phenotypic level, drought tolerance is associated with the strategic spatial organisation of root biomass rather than simply total root mass [29, 112]. At the metabolic level, adaptation reflects complementary tissue-specific network architectures—integrated leaf networks for rapid response, modular root networks for localised adaptation—rather than merely the accumulation of protective compounds [9]. At the systems level, resilience depends on stronger cross-modal coordination and earlier temporal establishment of communication networks, not just static molecular profiles [26, 37].

The implications extend far beyond drought tolerance in wheat. This framework provides a new lens for understanding biological complexity across disciplines, demonstrating that coordination capacity itself is measurable, heritable, and potentially manipulable [5, 6].

7.7.2 Delivering the Tools for Future Discovery

This work contributes more than a set of findings; it delivers the analytical tools required to ask new kinds of questions. The methodological ecosystem established here—validated through rigorous statistical analysis and confirmed on independent datasets—provides the infrastructure needed to address the escalating challenges facing global agriculture and systems biology [3, 70, 72].

- The **ART framework** provides a transformative new class of algorithmic biomarkers for high-throughput screening, delivering information density that outperforms conventional methods while maintaining complete biological interpretability.
- The **Network Analysis Pipeline** provides the tools to move beyond a simple "parts list" of genes and metabolites to reveal the architectural blueprints governing system behavior, demonstrating how function emerges from organisation [19, 20].
- The **MOFA+ Transformer** provides an unprecedented ability to quantify the directed, dynamic communication between a plant's physiological state and its biochemical machinery, moving systems biology beyond correlation toward mechanistic coordination [6].

Together, this open-source ecosystem transforms how we decode complex biological systems, providing a foundational platform for advancing resilience research and enabling discoveries yet to be envisioned [16, 113].

7.7.3 The Vision for a Resilient Future

The practical implications of this research are both immediate and forward-looking. For crop improvement, the frameworks and biomarkers developed here offer actionable pathways to accelerate breeding programs through more precise, high-throughput screening [10, 72, 74]. The coordination principles discovered provide novel, mechanistically-grounded targets for breeding, while the high-throughput tools enable the pursuit of these targets at scale.

For plant science, this work offers a new conceptual framework for understanding biological complexity, one rooted in the principles of organisation and dynamic coordination rather than static component analysis [9, 111]. As climate change intensifies agricultural stresses worldwide [1, 2], the ability to decode, predict, and ultimately enhance the biological mechanisms that allow life to persist under adversity becomes paramount.

This thesis concludes not as an endpoint, but as a foundation. In a world facing unprecedented agricultural and environmental pressures, it offers a contribution to the global scientific commons—a set of validated tools and a new perspective for understanding the intricate strategies life employs to persist. The analytical barriers that once obscured our view of biological complexity have been systematically addressed. The unseen has been made visible through algorithmic vision. The complex has been rendered quantifiable through network analysis. The communication protocols governing resilience have been decoded.

By learning to quantify biological coordination, we move closer to understanding adaptation itself. From this new vantage point—where coordination capacity can be measured, where network architectures can be decoded, where cross-modal communication can be quantified—the work of building a more resilient agricultural future can proceed with renewed precision and purpose.

The age of coordination-guided crop improvement begins now.

References

- [1] F. Chiang, O. Mazdiyasni, and A. AghaKouchak, "Evidence of anthropogenic impacts on global drought frequency, duration, and intensity," *Nature Communications*, vol. 12, no. 1, p. 2754, 2021/05/12 2021, doi: 10.1038/s41467-021-22314-w.
- [2] P. R. Shukla *et al.*, "Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems," 2019.
- [3] A. Sallam, A. M. Alqudah, M. F. Dawood, P. S. Baenziger, and A. Börner, "Drought stress tolerance in wheat and barley: advances in physiology, breeding and genetics research," *International journal of molecular sciences*, vol. 20, no. 13, p. 3137, Jun 27 2019, doi: 10.3390/ijms20133137.
- [4] O. Erenstein, M. Jaleta, K. A. Mottaleb, K. Sonder, J. Donovan, and H.-J. Braun, "Global trends in wheat production, consumption and trade," in *Wheat*

- improvement: food security in a changing climate*, M. P. Reynolds and H.-J. Braun Eds. Cham: Springer International Publishing, 2022, pp. 47-66.
- [5] H. Zhang, Y. Li, and J.-K. Zhu, "Developing naturally stress-resistant crops for a sustainable agriculture," in *Nature plants* vol. 4, 2018/11/28 ed, 2018, pp. 989-996.
- [6] D. M. Camacho, K. M. Collins, R. K. Powers, J. C. Costello, and J. J. Collins, "Next-Generation Machine Learning for Biological Networks," *Cell*, vol. 173, no. 7, pp. 1581-1592, 2018/06/14/ 2018, doi: <https://doi.org/10.1016/j.cell.2018.05.015>.
- [7] N. Sapoval *et al.*, "Current progress and open challenges for applying deep learning across the biosciences," *Nature Communications*, vol. 13, no. 1, 2022, doi: 10.1038/s41467-022-29268-7.
- [8] U. Alon, "Network motifs: theory and experimental approaches," *Nature Reviews Genetics*, vol. 8, no. 6, pp. 450-461, 2007/06/01 2007, doi: 10.1038/nrg2102.
- [9] A. R. Fernie and M. Stitt, "On the Discordance of Metabolomics with Proteomics and Transcriptomics: Coping with Increasing Complexity in Logic, Chemistry, and Network Interactions Scientific Correspondence," *Plant Physiology*, vol. 158, no. 3, pp. 1139-1145, 2012, doi: 10.1104/pp.112.193235.
- [10] M. Reynolds *et al.*, "Addressing Research Bottlenecks to Crop Productivity," *Trends in Plant Science*, vol. 26, no. 6, pp. 607-630, 2021, doi: 10.1016/j.tplants.2021.03.011.
- [11] H. M. Schneider and J. P. Lynch, "Root Traits for Improving N Acquisition Efficiency," in *Plant Biotechnology*: Springer International Publishing, 2021, ch. Chapter 12, pp. 163-180.
- [12] I. Subramanian, S. Verma, S. Kumar, A. Jere, and K. Anamika, "Multi-omics Data Integration, Interpretation, and Its Application," *Bioinformatics and Biology Insights*, vol. 14, p. 1177932219899051, 2020, doi: 10.1177/1177932219899051.
- [13] H. Tao *et al.*, "Proximal and remote sensing in plant phenomics: Twenty years of progress, challenges and perspectives," *Plant Communications*, 2022.

- [14] T. Gill, S. K. Gill, D. K. Saini, Y. Chopra, J. P. de Koff, and K. S. Sandhu, "A Comprehensive Review of High Throughput Phenotyping and Machine Learning for Plant Stress Phenotyping," *Phenomics*, pp. 1-28, 2022.
- [15] A. L. Harfouche, F. Nakhle, A. H. Harfouche, O. G. Sardella, E. Dart, and D. Jacobson, "A primer on artificial intelligence in plant digital phenomics: embarking on the data to insights journey," *Trends in Plant Science*, 2022.
- [16] M. Shoaib, N. Sharma, L. Kotthoff, M. Lindauer, and S. Kant, "AutoML: advanced tool for mining multivariate plant traits," *Trends in Plant Science*, 2023, doi: 10.1016/j.tplants.2023.09.008.
- [17] X. He, K. Zhao, and X. Chu, "AutoML: A survey of the state-of-the-art," *Knowledge-Based Systems*, vol. 212, p. 106622, 2021, doi: 10.1016/j.knosys.2020.106622.
- [18] D. D. Hoffman, M. Singh, and C. Prakash, "The interface theory of perception," *Psychonomic bulletin & review*, vol. 22, no. 6, pp. 1480-1506, 2015, doi: 10.3758/s13423-015-0890-8.
- [19] A.-L. Barabási and Z. N. Oltvai, "Network biology: understanding the cell's functional organization," *Nature Reviews Genetics*, vol. 5, no. 2, pp. 101-113, 2004, doi: 10.1038/nrg1272.
- [20] D. Toubiana, A. R. Fernie, Z. Nikoloski, and A. Fait, "Network analysis: tackling complex data to study plant metabolism," *Trends in biotechnology*, vol. 31, no. 1, pp. 29-36, 2013.
- [21] B. Zhang and S. Horvath, "A general framework for weighted gene co-expression network analysis," (in eng), *Stat Appl Genet Mol Biol*, vol. 4, p. Article17, 2005, doi: 10.2202/1544-6115.1128.
- [22] F. Rohart, B. Gautier, A. Singh, and K.-A. Lê Cao, "mixOmics: An R package for 'omics feature selection and multiple data integration," *PLoS computational biology*, vol. 13, no. 11, p. e1005752, 2017.
- [23] M. Li, M. H. Frank, V. Coneva, W. Mio, D. H. Chitwood, and C. N. Topp, "The persistent homology mathematical framework provides enhanced genotype-to-phenotype associations for plant morphology," *Plant physiology*, vol. 177, no. 4, pp. 1382-1395, Aug 2018, doi: 10.1104/pp.18.00104.

- [24] H. Li, C. Testerink, and Y. Zhang, "How roots and shoots communicate through stressful times," *Trends in Plant Science*, vol. 26, no. 9, pp. 940-952, 2021, doi: 10.1016/j.tplants.2021.03.005.
- [25] G. R. Cramer, K. Urano, S. Delrot, M. Pezzotti, and K. Shinozaki, "Effects of abiotic stress on plants: a systems biology perspective," *BMC Plant Biology*, vol. 11, no. 1, p. 163, 2011/11/17 2011, doi: 10.1186/1471-2229-11-163.
- [26] J.-K. Zhu, "Abiotic Stress Signaling and Responses in Plants," *Cell*, vol. 167, no. 2, pp. 313-324, 2016, doi: 10.1016/j.cell.2016.08.029.
- [27] A. Wasaya, X. Zhang, Q. Fang, and Z. Yan, "Root Phenotyping for Drought Tolerance: A Review," *Agronomy*, vol. 8, no. 11, p. 241, 2018, doi: 10.3390/agronomy8110241.
- [28] E. S. Ober *et al.*, "Wheat root systems as a breeding target for climate resilience," *Theoretical and Applied Genetics*, vol. 134, no. 6, pp. 1645-1662, 2021, doi: 10.1007/s00122-021-03819-w.
- [29] J. P. Lynch, "Steep, cheap and deep: an ideotype to optimize water and N acquisition by maize root systems," *Annals of Botany*, vol. 112, no. 2, pp. 347-357, 2013, doi: 10.1093/aob/mcs293.
- [30] J. P. Lynch, "Harnessing root architecture to address global challenges," *The Plant Journal*, vol. 109, no. 2, pp. 415-431, 2022, doi: 10.1111/tpj.15560.
- [31] A. Fait, A. Batushansky, V. Shrestha, A. Yobi, and R. Angelovici, "Can metabolic tightening and expansion of co-expression network play a role in stress response and tolerance?," *Plant Science*, vol. 293, p. 110409, 2020/04/01/ 2020, doi: 10.1016/j.plantsci.2020.110409.
- [32] R. F. H. Giehl and N. Von Wieren, "Hydropatterning-how roots test the waters," *Science*, vol. 362, no. 6421, pp. 1358-1359, Dec 21 2018, doi: 10.1126/science.aav9375.
- [33] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences*, vol. 99, no. 12, pp. 7821-7826, 2002, doi: doi:10.1073/pnas.122653799.
- [34] Z. Bar-Joseph, A. Gitter, and I. Simon, "Studying and modelling dynamic biological processes using time-series gene expression data," *Nature Reviews Genetics*, vol. 13, no. 8, pp. 552-564, 2012/08/01 2012, doi: 10.1038/nrg3244.

- [35] K. Shinozaki and K. Yamaguchi-Shinozaki, "Gene networks involved in drought stress response and tolerance," (in eng), *J Exp Bot*, vol. 58, no. 2, pp. 221-7, 2007, doi: 10.1093/jxb/erl164.
- [36] D. Todaka, K. Shinozaki, and K. Yamaguchi-Shinozaki, "Recent advances in the dissection of drought-stress regulatory networks and strategies for development of drought-tolerant transgenic rice plants," *Frontiers in plant science*, vol. 6, p. 84, 2015.
- [37] H. Zhang, J. Zhu, Z. Gong, and J.-K. Zhu, "Abiotic stress responses in plants," *Nature Reviews Genetics*, 2021/09/24 2021, doi: 10.1038/s41576-021-00413-0.
- [38] W. Bechtel, "Hierarchy and levels: analysing networks to study mechanisms in molecular biology," *Philosophical Transactions of the Royal Society B*, vol. 375, no. 1796, p. 20190320, 2020.
- [39] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Barabási, "Hierarchical Organization of Modularity in Metabolic Networks," *Science*, vol. 297, no. 5586, pp. 1551-1555, 2002, doi: doi:10.1126/science.1073374.
- [40] S. Häusler, "Correlations reveal the hierarchical organization of biological networks with latent variables," (in eng), *Commun Biol*, vol. 7, no. 1, p. 678, Jun 3 2024, doi: 10.1038/s42003-024-06342-y.
- [41] A. Gargallo-Garriga *et al.*, "Opposite metabolic responses of shoots and roots to drought," *Scientific Reports*, vol. 4, no. 1, p. 6829, 2014/10/29 2014, doi: 10.1038/srep06829.
- [42] K. Akita and Y. Miyazawa, "The mechanosensitive Ca²⁺ channel, OSCA1.1, modulates root hydrotropic bending in *Arabidopsis thaliana*," *Environmental and Experimental Botany*, vol. 197, p. 104825, 2022, doi: 10.1016/j.envexpbot.2022.104825.
- [43] C. Maurel and P. Nacry, "Root architecture and hydraulics converge for acclimation to changing water availability," *Nature Plants*, vol. 6, no. 7, pp. 744-749, 2020, doi: 10.1038/s41477-020-0684-5.
- [44] F. C. Meinzer, "Functional convergence in plant responses to the environment," *Oecologia*, vol. 134, no. 1, pp. 1-11, 2003/01/01 2003, doi: 10.1007/s00442-002-1088-0.

- [45] R. Albert, H. Jeong, and A.-L. Barabási, "Error and attack tolerance of complex networks," *Nature*, vol. 406, no. 6794, pp. 378-382, 2000/07/01 2000, doi: 10.1038/35019019.
- [46] A. R. Joyce and B. Ø. Palsson, "The model organism as a system: integrating 'omics' data sets," *Nature Reviews Molecular Cell Biology*, vol. 7, no. 3, pp. 198-210, 2006/03/01 2006, doi: 10.1038/nrm1857.
- [47] M. Lucas, L. Laplaze, and M. J. Bennett, "Plant systems biology: network matters," *Plant, cell & environment*, vol. 34, no. 4, pp. 535-553, 2011.
- [48] M. Tomkins, "Towards modelling emergence in plant systems," (in eng), *Quant Plant Biol*, vol. 4, p. e6, 2023, doi: 10.1017/qpb.2023.6.
- [49] H. Kitano, "Systems Biology: A Brief Overview," *Science*, vol. 295, no. 5560, pp. 1662-1664, 2002, doi: doi:10.1126/science.1069492.
- [50] W. Ma, A. Trusina, H. El-Samad, W. A. Lim, and C. Tang, "Defining network topologies that can achieve biochemical adaptation," (in eng), *Cell*, vol. 138, no. 4, pp. 760-73, Aug 21 2009, doi: 10.1016/j.cell.2009.06.013.
- [51] H. J. Escalante, "Automated Machine Learning—A Brief Review at the End of the Early Years," in *Automated Design of Machine Learning and Search Algorithms*, (Natural Computing Series: Springer International Publishing, 2021, ch. Chapter 2, pp. 11-28.
- [52] F. Hutter, L. Kotthoff, and J. Vanschoren, *Automated machine learning: methods, systems, challenges*. Springer Nature, 2019.
- [53] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," presented at the Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, Portland, Oregon, 1996.
- [54] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 5, pp. 603-619, 2002, doi: 10.1109/34.1000236.
- [55] M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander, "OPTICS: Ordering points to identify the clustering structure," *ACM Sigmod record*, vol. 28, no. 2, pp. 49-60, 1999, doi: 10.1145/304182.304187.

- [56] L. McInnes, J. Healy, and S. Astels, "hdbscan: Hierarchical density based clustering," *Journal of Open Source Software*, vol. 2, no. 6/22/2024, 2017, doi: 10.21105/joss.00205.
- [57] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Computers & geosciences*, vol. 10, no. 2-3, pp. 191-203, 1984, doi: 10.1016/0098-3004(84)90020-7.
- [58] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *the Journal of machine Learning research*, vol. 12, pp. 2825-2830, 2011, doi: 10.48550/arXiv.1201.0490.
- [59] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 11, pp. 2274-2282, 2012, doi: 10.1109/TPAMI.2012.120.
- [60] H. Poorter, K. J. Niklas, P. B. Reich, J. Oleksyn, P. Poot, and L. Mommer, "Biomass allocation to leaves, stems and roots: meta-analyses of interspecific variation and environmental control," *New Phytologist*, vol. 193, no. 1, pp. 30-50, 2012, doi: 10.1111/j.1469-8137.2011.03952.x.
- [61] B. V. North, D. Curtis, and P. C. Sham, "A note on the calculation of empirical P values from Monte Carlo procedures," (in eng), *Am J Hum Genet*, vol. 71, no. 2, pp. 439-41, Aug 2002, doi: 10.1086/341527.
- [62] C. P. d. Campos and Q. Ji, "Efficient Structure Learning of Bayesian Networks using Constraints," *J. Mach. Learn. Res.*, vol. 12, no. null, pp. 663–689, 2011.
- [63] Y.-H. H. Tsai, S. Bai, P. P. Liang, J. Z. Kolter, L.-P. Morency, and R. Salakhutdinov, "Multimodal transformer for unaligned multimodal language sequences," in *Proceedings of the conference. Association for computational linguistics. Meeting*, 2019, vol. 2019, p. 6558.
- [64] A. Vaswani *et al.*, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.
- [65] R. Argelaguet *et al.*, "MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data," *Genome Biology*, vol. 21, no. 1, p. 111, 2020/05/11 2020, doi: 10.1186/s13059-020-02015-1.
- [66] R. Argelaguet *et al.*, "Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets," *Molecular Systems*

Biology, vol. 14, no. 6, p. e8124, 2018, doi:
<https://doi.org/10.15252/msb.20178124>.

- [67] Y. Xie *et al.*, "Integration of hyperspectral imaging and transcriptomics from individual cells with HyperSeq," *bioRxiv*, p. 2024.01.27.577536, 2024.
- [68] T. Hirose *et al.*, "NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies," (in eng), *Mol Biol Cell*, vol. 25, no. 1, pp. 169-83, Jan 2014, doi: 10.1091/mbc.E13-09-0558.
- [69] A. L. Harfouche *et al.*, "Accelerating climate resilient plant breeding by applying next-generation artificial intelligence," *Trends in biotechnology*, vol. 37, no. 11, pp. 1217-1235, 2019.
- [70] J. Salse, R. L. Barnard, C. Veneault-Fourrey, and H. Rouached, "Strategies for breeding crops for future environments," *Trends in Plant Science*, vol. 29, no. 3, pp. 303-318, 2024, doi: 10.1016/j.tplants.2023.08.007.
- [71] R. K. Varshney *et al.*, "Fast-forward breeding for a food-secure world," *Trends in Genetics*, 2021, doi: 10.1016/j.tig.2021.08.002.
- [72] Y. Xu *et al.*, "Smart breeding driven by big data, artificial intelligence, and integrated genomic-enviromic prediction," *Molecular Plant*, vol. 15, no. 11, pp. 1664-1695, 2022.
- [73] L. Crespo-Herrera *et al.*, "Genetic gains for grain yield in CIMMYT's semi-arid wheat yield trials grown in suboptimal environments," *Crop Science*, vol. 58, no. 5, pp. 1890-1898, 2018.
- [74] K. P. Voss-Fels, R. J. Snowdon, and L. T. Hickey, "Designer Roots for Future Crops," *Trends in Plant Science*, vol. 23, no. 11, pp. 957-960, 2018, doi: 10.1016/j.tplants.2018.08.004.
- [75] J. L. Araus and J. E. Cairns, "Field high-throughput phenotyping: the new crop breeding frontier," *Trends in plant science*, vol. 19, no. 1, pp. 52-61, 2014.
- [76] S. R. Tracy, K. A. Nagel, J. A. Postma, H. Fassbender, A. Wasson, and M. Watt, "Crop Improvement from Phenotyping Roots: Highlights Reveal Expanding Opportunities," *Trends in Plant Science*, vol. 25, no. 1, pp. 105-118, 2020, doi: 10.1016/j.tplants.2019.10.015.
- [77] W. Xiong *et al.*, "Increased ranking change in wheat breeding under climate change," *Nature Plants*, vol. 7, no. 9, pp. 1207-1212, Sep 2021, doi: 10.1038/s41477-021-00988-w.

- [78] S. Ghosh *et al.*, "Speed breeding in growth chambers and glasshouses for crop breeding and model plant research," *Nature Protocols*, vol. 13, no. 12, pp. 2944-2963, 2018, doi: 10.1038/s41596-018-0072-z.
- [79] Y. Xu *et al.*, "Feeding the world using speed breeding technology," *Trends in Plant Science*, 2022.
- [80] J. Jung, M. Maeda, A. Chang, M. Bhandari, A. Ashapure, and J. Landivar-Bowles, "The potential of remote sensing and artificial intelligence as tools to improve the resilience of agriculture production systems," *Current Opinion in Biotechnology*, vol. 70, pp. 15-22, 2021.
- [81] A. Blum, "Effective use of water (EUW) and not water-use efficiency (WUE) is the target of crop yield improvement under drought stress," *Field Crops Research*, vol. 112, no. 2, pp. 119-123, 2009/06/26/ 2009, doi: <https://doi.org/10.1016/j.fcr.2009.03.009>.
- [82] Z. Zhang and X. Peng, "Bio-tillage: A new perspective for sustainable agriculture," *Soil and Tillage Research*, vol. 206, p. 104844, 2021.
- [83] A. G. Condon, "Breeding for high water-use efficiency," *Journal of Experimental Botany*, vol. 55, no. 407, pp. 2447-2460, 2004, doi: 10.1093/jxb/erh277.
- [84] C. Jansson *et al.*, "Crops for Carbon Farming," (in English), *Frontiers in Plant Science*, Perspective vol. 12, no. 938, 2021-June-04 2021, doi: 10.3389/fpls.2021.636709.
- [85] P. Panchal, C. Preece, J. Peñuelas, and J. Giri, "Soil carbon sequestration by root exudates," *Trends in Plant Science*, vol. 27, no. 8, pp. 749-757, 2022/08/01/ 2022, doi: <https://doi.org/10.1016/j.tplants.2022.04.009>.
- [86] C. Rumpel, J. Lehmann, and A. Chabbi, "'4 per 1,000' initiative will boost soil carbon for climate and food security," *Nature*, vol. 553, no. 7686, pp. 27-27, 2018, doi: 10.1038/d41586-017-09010-w.
- [87] R. R. Mir, M. Zaman-Allah, N. Sreenivasulu, R. Trethowan, and R. K. Varshney, "Integrated genomics, physiology and breeding approaches for improving drought tolerance in crops," *Theoretical and Applied Genetics*, vol. 125, no. 4, pp. 625-645, 2012/08/01 2012, doi: 10.1007/s00122-012-1904-9.

- [88] E. D. Schäfer, M. R. Owen, L. R. Band, E. Farcot, M. J. Bennett, and J. P. Lynch, "Modeling root loss reveals impacts on nutrient uptake and crop development," *Plant Physiology*, 2022, doi: 10.1093/plphys/kiac405.
- [89] J. Casadesús and D. Villegas, "Conventional digital cameras as a tool for assessing leaf area index and biomass for cereal breeding," *Journal of integrative plant biology*, vol. 56, no. 1, pp. 7-14, 2014, doi: 10.1111/jipb.12117.
- [90] N. Lu *et al.*, "Improved estimation of aboveground biomass in wheat from RGB imagery and point cloud data acquired with a low-cost unmanned aerial vehicle system," *Plant Methods*, vol. 15, pp. 1-16, 2019.
- [91] K. Berger *et al.*, "Multi-sensor spectral synergies for crop stress detection and monitoring in the optical domain: A review," *Remote Sensing of Environment*, vol. 280, p. 113198, 2022.
- [92] R. Nakabayashi and K. Saito, "Integrated metabolomics for abiotic stress responses in plants," (in eng), *Curr Opin Plant Biol*, vol. 24, pp. 10-6, Apr 2015, doi: 10.1016/j.pbi.2015.01.003.
- [93] T. Obata and A. R. Fernie, "The use of metabolomics to dissect plant responses to abiotic stresses," (in eng), *Cell Mol Life Sci*, vol. 69, no. 19, pp. 3225-43, Oct 2012, doi: 10.1007/s00018-012-1091-5.
- [94] F. M. Bauer *et al.*, "Development and Validation of a Deep Learning Based Automated Minirhizotron Image Analysis Pipeline," *Plant Phenomics*, vol. 2022, p. 9758532, 2022, doi: 10.34133/2022/9758532.
- [95] S. F. Svane, E. B. Dam, J. M. Carstensen, and K. Thorup-Kristensen, "A multispectral camera system for automated minirhizotron image analysis," *Plant and Soil*, vol. 441, no. 1-2, pp. 657-672, 2019, doi: 10.1007/s11104-019-04132-8.
- [96] P. K. R. Maddikunta *et al.*, "Unmanned aerial vehicles in smart agriculture: Applications, requirements, and challenges," *IEEE Sensors Journal*, vol. 21, no. 16, pp. 17608-17619, 2021.
- [97] W. H. Maes and K. Steppe, "Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture," *Trends in plant science*, vol. 24, no. 2, pp. 152-164, 2019.

- [98] D. Sun, K. Robbins, N. Morales, Q. Shu, and H. Cen, "Advances in optical phenotyping of cereal crops," *Trends in plant science*, vol. 27, no. 2, pp. 191-208, 2022.
- [99] P. M. Correia, J. Cairo Westergaard, A. B. da Silva, T. Roitsch, E. Carmo-Silva, and J. M. da Silva, "High-throughput phenotyping of physiological traits for wheat resilience to high temperature and drought stress," *Journal of Experimental Botany*, 2022.
- [100] V. Sobejano-Paz *et al.*, "Hyperspectral and Thermal Sensing of Stomatal Conductance, Transpiration, and Photosynthesis for Soybean and Maize under Drought," *Remote Sensing*, vol. 12, no. 19, p. 3182, 2020, doi: 10.3390/rs12193182.
- [101] S. Eltaher *et al.*, "Genetic diversity and population structure of F3: 6 Nebraska winter wheat genotypes using genotyping-by-sequencing," *Frontiers in genetics*, vol. 9, p. 76, 2018.
- [102] C. Topp. "Improving soil health." <https://www.danforthcenter.org/our-work/principal-investigators/christopher-topp/> (accessed 27th April, 2022).
- [103] C. Santos-Medellín *et al.*, "Prolonged drought imparts lasting compositional changes to the rice root microbiome," *Nature Plants*, vol. 7, no. 8, pp. 1065-1077, 2021, doi: 10.1038/s41477-021-00967-1.
- [104] B. Bruning, H. Liu, C. Brien, B. Berger, M. Lewis, and T. Garnett, "The Development of Hyperspectral Distribution Maps to Predict the Content and Distribution of Nitrogen and Water in Wheat (*Triticum aestivum*)," (in English), *Frontiers in Plant Science*, Original Research vol. 10, no. 1380, 2019-October-30 2019, doi: 10.3389/fpls.2019.01380.
- [105] S. I. Zandalinas, D. Balfagón, A. Gómez-Cadenas, and R. Mittler, "Plant responses to climate change: metabolic changes under combined abiotic stresses," *Journal of Experimental Botany*, vol. 73, no. 11, pp. 3339-3354, 2022.
- [106] M. Griffiths *et al.*, "A multiple ion-uptake phenotyping platform reveals shared mechanisms affecting nutrient uptake by roots," *Plant Physiology*, vol. 185, no. 3, pp. 781-795, 2021, doi: 10.1093/plphys/kiaa080.
- [107] R. Munns and M. Tester, "Mechanisms of salinity tolerance," *Annu. Rev. Plant Biol.*, vol. 59, no. 1, pp. 651-681, 2008.

- [108] J.-K. Zhu, "Salt and drought stress signal transduction in plants," *Annual review of plant biology*, vol. 53, no. 1, pp. 247-273, 2002.
- [109] U. Baetz and E. Martinoia, "Root exudates: the hidden part of plant defense," *Trends in Plant Science*, vol. 19, no. 2, pp. 90-98, 2014, doi: 10.1016/j.tplants.2013.11.006.
- [110] E. A. Ampt, J. Ruijven, M. P. Zwart, J. M. Raaijmakers, A. J. Termorshuizen, and L. Mommer, "Plant neighbours can make or break the disease transmission chain of a fungal root pathogen," *New Phytologist*, vol. 233, no. 3, pp. 1303-1316, 2022, doi: 10.1111/nph.17866.
- [111] K. Shinozaki and K. Yamaguchi-Shinozaki, "Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signaling pathways," *Current Opinion in Plant Biology*, vol. 3, no. 3, pp. 217-223, 2000/01/01/ 2000, doi: [https://doi.org/10.1016/S1369-5266\(00\)80068-0](https://doi.org/10.1016/S1369-5266(00)80068-0).
- [112] Y. Uga *et al.*, "Control of root system architecture by DEEPER ROOTING 1 increases rice yield under drought conditions," *Nature Genetics*, vol. 45, no. 9, pp. 1097-1102, Sep 2013, doi: 10.1038/ng.2725.
- [113] A. L. Harfouche, F. Nakhle, A. H. Harfouche, O. G. Sardella, E. Dart, and D. Jacobson, "A primer on artificial intelligence in plant digital phenomics: embarking on the data to insights journey," *Trends in Plant Science*, vol. 28, no. 2, pp. 154-184, 2023, doi: 10.1016/j.tplants.2022.08.021.

Appendix 1: Supplementary Material of Chapter 4

SUPPLEMENTARY MATERIALS

Title: Seeing the Unseen: A Novel Approach to Extract Latent Plant Root Traits from Digital Images

Authors

Mirza Shoaib^{1,2*}, Adam M. Dimech³, Simone J Rochfort^{2,3}, Christopher Topp⁴, Matthew J. Hayden^{2,3} and Surya Kant^{5*}

Affiliations

¹Agriculture Victoria, Grains Innovation Park, 110 Natimuk Road, Horsham, Victoria 3400, Australia

²School of Applied Systems Biology, La Trobe University, Bundoora, Victoria 3083, Australia

³Agriculture Victoria, AgriBio, Centre for AgriBioscience, 5 Ring Road, Bundoora, Victoria 3083, Australia

⁴Donald Danforth Plant Science Center, Saint Louis, Missouri, USA

⁵Department of Ecological, Plant and Animal Science, School of Agriculture, Biomedicine & Environment, La Trobe University, Bundoora, Victoria 3083, Australia

* Correspondence: shoaib.mirza@agriculture.vic.gov.au and S.Kant@latrobe.edu.au

Table S1. Comprehensive comparison of ART with STARSEED, Persistent Homology, SLEAP, LSP, and TRTs. A detailed analysis of how ART differs from other prominent root phenotyping methods, comparing core computational philosophy/goals, primary algorithmic engines, input data types, nature of extracted traits/outputs, interpretability approach, custom algorithm/unique implementation details, primary applications demonstrated, and key differentiators from ART's perspective.

| Method/Framework Name (Reference) | Core Computational Philosophy/Goal | Primary Algorithmic Engine(s) | Input Data Type | Nature of Extracted Traits'/Outputs | Interpretability Approach | Custom Algorithm/Unique Implementation Detail | Primary Application Demonstrated | Key Differentiator from ART's Perspective |
|---|---|---|----------------------------|---|---|--|---|---|
| ART (this study) | Unsupervised discovery of significant root density patterns & their spatial locations | Ensemble of unsupervised clustering algs (K-means, DBSCA N, etc.) + Custom "Custom" | Segmented binary image | Geometric attributes of discovered dense root clusters (e.g., centre_y, density_points) | Linking cluster size/location to biological hypotheses (e.g., rooting depth, localised biomass) | "Custom" algorithm for densest cluster ID; Framework combining outputs from 9 diverse algorithmic perspectives | Root drought tolerance classification | Focuses on "what" & "where" dense root regions are via an exploratory, unsupervised ensemble; traits are interpretable spatial/density metrics. Complements & extends TRTs with localised data. |
| STARSEED (Peeples et al., 2023) | Comparing overall root system distributions via optimal transport | Earth Mover's Distance (EMD) | Pre-processed binary image | Scalar EMD score (global difference), Flow matrix (local changes between distribution s) | Visualising flow matrix for local differences ; CH index for group separation | Specific EMD cost function; Grid-based pre-feature extraction (e.g., root pixel %) for EMD input | Sesame root response to genotype/moisture | EMD provides a holistic distance metric between entire distributions; ART identifies specific dense clusters within a distribution. |
| Persistent Homology (Li et al., 2018) | Quantifying shape complexity via topological feature persistence | Persistent Homology (PH) | 2D point clouds (contours) | Persistence barcodes (H_0 for connected component s, H_1 for 1D holes/loops), CV scores from barcode distances | Relating abstract topological features to architectural density for leaf shape; distance function for root branch crossings | Custom PH functions (e.g., Gaussian kernel with PH-derived CV scores) | Tomato leaf shape, serrations, root architecture | PH quantifies abstract topological features (connectivity, holes); ART quantifies geometric features (size, location) of dense pixel clusters. |
| SLEAP for Roots (Berrigan et al., 2024) | Supervised detection & grouping of pre-defined anatomical landmarks | Deep Learning (CNNs for pose estimation) | Raw RGB images | 2D coordinates of pre-defined root landmarks; Derived geometric traits (lengths, angles) | Direct biological meaning of landmarks & derived traits (e.g., primary root length, lateral root angle) | Adaptation of SLEAP (animal pose) to root landmarks; Specific root labelling protocols | Multi-species root topology & trait extraction from landmarks | SLEAP requires supervised labelling of known landmarks; ART is unsupervised, discovering regions of interest (dense clusters) without pre-defined landmark definitions. |
| Latent Space Phenotyping | Learning discriminant | Variational | Raw RGB | Abstract n-dimensiona | Embeddin gs capture | CNN-LSTM | Plant response- | LSP learns an abstract latent |

| | | | | | | | | |
|--------------------------------|---|--|-------------------------|--|---|--|--|---|
| (Ubbens et al., 2020) | tive low-dimension al embeddin gs of dynamic plant response to treatment | Autoencoders (VAEs), CNNs, LSTMs | image sequences | 1 embedding s; Geodesic path length in latent space (response-to-treatment magnitude) | treatment effects; Path length quantifies response. Saliency maps for image region relevance. | encoder for classification, then VAE-like decoder to map latent space for path measurement | to-treatment (drought, N-deficiency) over time | space for overall plant response dynamics; ART extracts concrete, interpretable geometric traits from (typically) static images, focusing on root substructures. |
| Traditional Root Traits (TRTs) | Direct quantification of pre-defined, basic root morphological parameters | Image processing (e.g., thresholding, skeletonisation, pixel counting) | Segmented binary images | Basic geometric measures (e.g., total length, average diameter, area, convex hull, branching angles) | Direct measurement of visually obvious physical properties | Often relies on established algorithms in software like RhizoVision, ImageJ, WinRhizo | General root phenotyping | TRTs provide foundational, often global, metrics. ART aims to discover and quantify localised, potentially latent, density-based traits that complement and enhance TRTs for tasks like classification. |

Table S2. List of genotypes used in the experiment.

| Drought Tolerance | Genotype Name | Abbreviation | Reference |
|-------------------|----------------|--------------|--------------------------|
| Tolerance | Gladius | DT_1 | Bennani et al. [57],[58] |
| | DAS5_005489 | DT_2 | Hone et al. [59] |
| | DAS5_CALINGIRI | DT_3 | Hone et al. [59] |
| Susceptible | Forrest | DS_1 | Glenn and Rebecca [60] |
| | Hartog | DS_2 | Glenn and Rebecca [60] |
| | DAS5_003811 | DS_3 | Hone et al. [59] |

Table S3. Nutrient solution used in the glasshouse experiment.

| No | Nutrient | Concentration |
|----|--------------------------------------|---------------|
| 1 | MgSO ₄ | 4 mM |
| 2 | CaCl ₂ .2H ₂ O | 4 mM |
| 3 | KH ₂ PO ₄ pool | 3 mM |
| | K ₂ HPO ₄ pool | |
| 4 | Fe ⁺ (FeEDTANa Librel) | 400 µM |
| 5 | MnCl ₂ .4H ₂ O | 10 µM |
| | ZnSO ₄ .7H ₂ O | 10 µM |
| | CuSO ₄ | 2 µM |
| | H ₃ BO ₃ | 50 µM |
| | Na ₂ MoO ₄ | 0.2 µM |
| 6 | KNO ₃ | 1M |

Table S4. Rhizovision settings for TRT extraction.

| Parameters | Settings |
|------------|----------|
|------------|----------|

| | |
|---|--------------|
| RhizoVision Explorer Version | 2.0.3 |
| Root type | Broken roots |
| Image Thresholding Level | 255 |
| Invert images | false |
| Keep largest component | true |
| Filter noisy components on background | true |
| Maximum background noisy component size | 0.2 |
| Filter noisy components on foreground | false |
| Maximum foreground noisy component size | 1 |
| Enable edge smoothing | false |
| Edge smoothing threshold | 0 |
| Enable root pruning | true |
| Root pruning threshold | 5 |
| Convert pixels to physical units | true |
| Dots per inch | 600 |
| Pixel to millimeter conversion factor | 0.042333 |
| Diameter Range 1 | 0-0.3 |
| Diameter Range 2 | 0.3-0.6 |
| Diameter Range 3 | 0.6-above |
| Features output file name | features.csv |
| Save segmented images | true |
| Segmented image file name suffix | |
| Save processed feature images | false |
| Processed image file name suffix | |

Table S5. List of ART and TRT variables.

| ART | TRT |
|------------------------|-------------------------------------|
| DBSCAN_density_points | Number.of.Root.Tips |
| DBSCAN_centre_x | Number.of.Branch.Points |
| DBSCAN_centre_y | Total.Root.Length.mm |
| Custom_density_points | Branching.frequency.per.mm |
| Custom_centre_x | Network.Area.mm2 |
| Custom_centre_y | Average.Diameter.mm |
| FCM_density_points | Median.Diameter.mm |
| FCM_centre_x | Maximum.Diameter.mm |
| FCM_centre_y | Perimeter.mm |
| GMM_density_points | Volume.mm3 |
| GMM_centre_x | Surface.Area.mm2 |
| GMM_centre_y | Root.Length.Diameter.Range.1.mm |
| HDBSCAN_density_points | Root.Length.Diameter.Range.2.mm |
| HDBSCAN_centre_x | Root.Length.Diameter.Range.3.mm |
| HDBSCAN_centre_y | Projected.Area.Diameter.Range.1.mm2 |
| K-mean density points | Projected.Area.Diameter.Range.2.mm2 |

| | |
|---------------------------|-------------------------------------|
| K-mean_centre_x | Projected.Area.Diameter.Range.3.mm2 |
| K-mean_centre_y | Surface.Area.Diameter.Range.1.mm2 |
| SLIC_density_points | Surface.Area.Diameter.Range.2.mm2 |
| SLIC_centre_x | Surface.Area.Diameter.Range.3.mm2 |
| SLIC_centre_y | Volume.Diameter.Range.1.mm3 |
| Mean-shift_density_points | Volume.Diameter.Range.2.mm3 |
| Mean-shift_centre_x | Volume.Diameter.Range.3.mm3 |
| Mean-shift_centre_y | |
| OPTICS_density_points | |
| OPTICS_centre_x | |
| OPTICS_centre_y | |

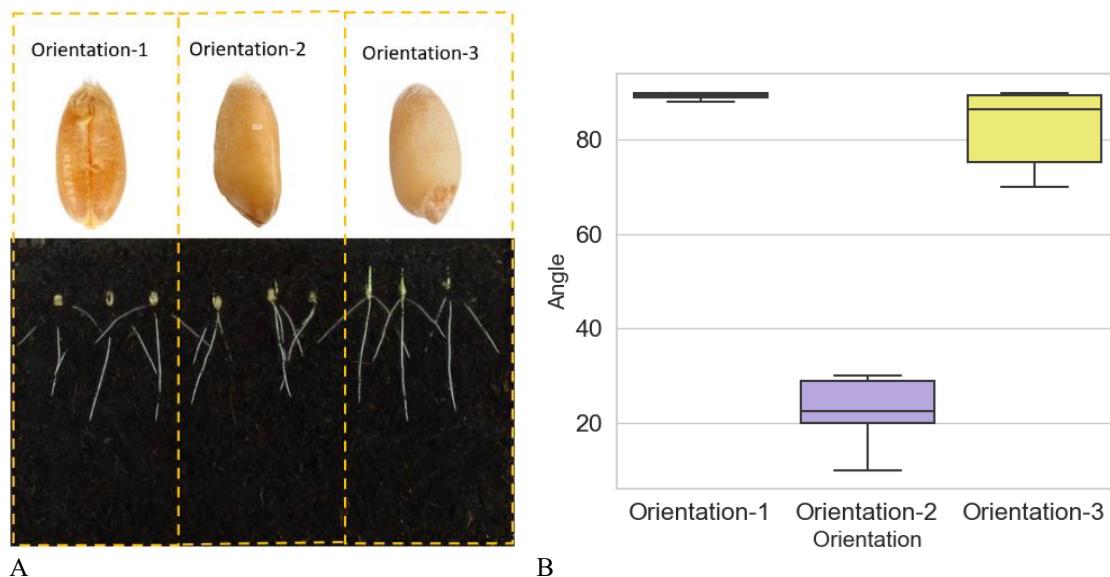


Fig. S1. Effect of seed sowing orientation on root growth. (A) Different seed sowing orientations (Orientation-1, Orientation-2, Orientation-3) and the corresponding root growth patterns for each orientation. (B) Effect of seed sowing orientation on root angle. Orientation-1 = Brush pointed up and crease facing rhizotron wall, Orientation-2 = Brush pointed up and crease facing 90° away from rhizotron wall, Orientation-3 = Brush pointed up and crease facing opposite (180°) to rhizotron wall.

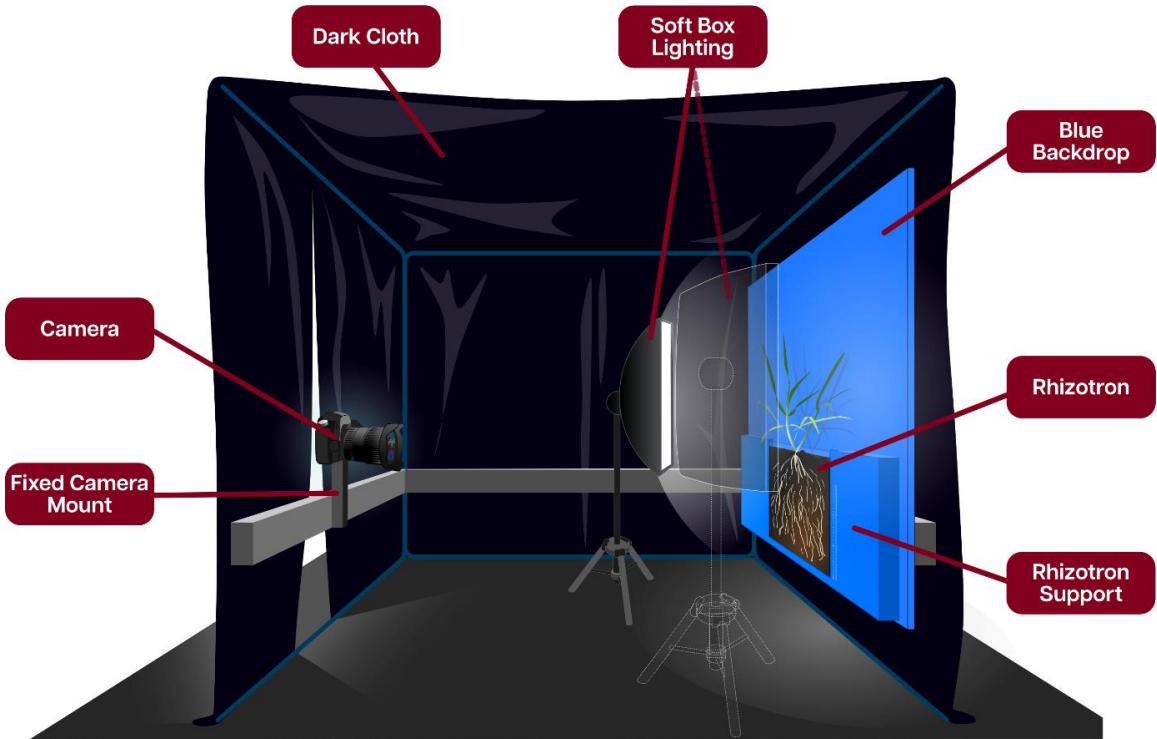


Fig. S2. The imaging set-up used for rhizotron experiments. A custom imaging set-up was built in the glasshouse to capture rhizotron images. To prevent reflectance from the rhizotron, a blackout curtain was used to form a dark tent, and a frame secured the rhizotron and camera at a consistent distance. The set-up was painted matt black, except for the light blue (HEX code: 0089b6) image background, and it was illuminated with four diffused studio lights angled to prevent reflection.

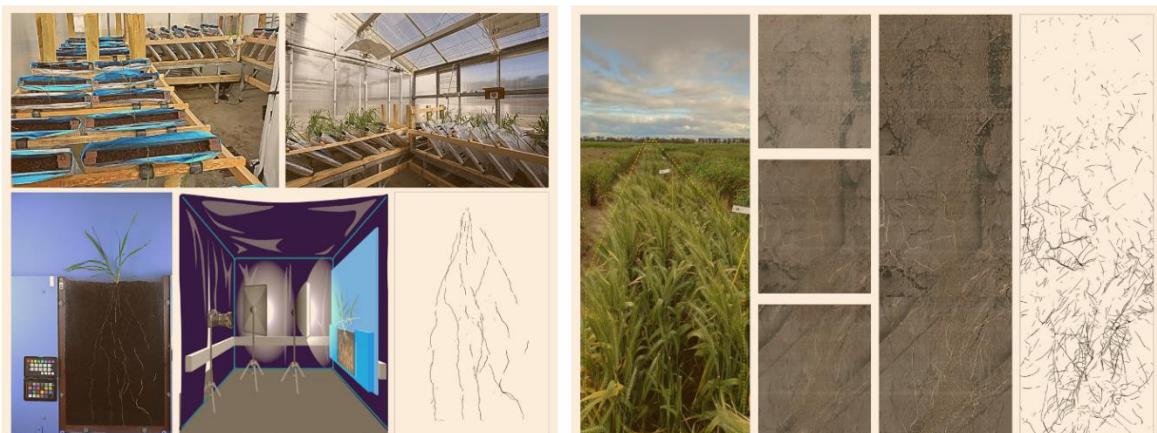
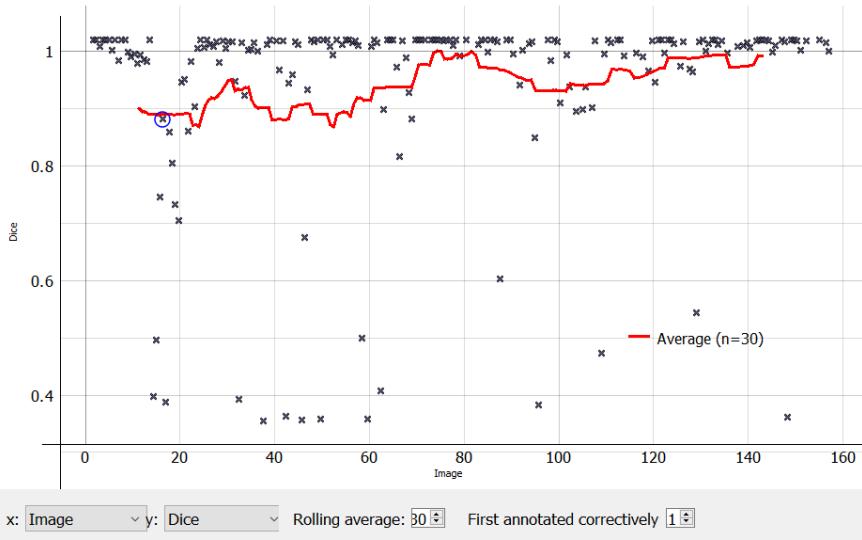
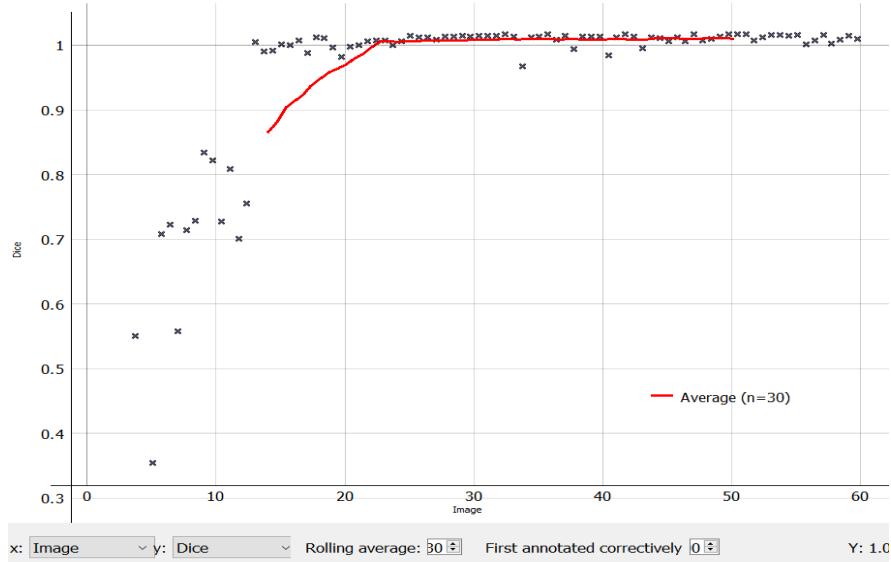


Fig. S3. Experimental set-up. (A) Rhizotron experiment in glasshouse. The experimental set-up included a rhizotron holder, imaging station, rhizotron image, and image after cropping and segmentation. (B) Field experiment. Layout and view of the plots: Three sets of images captured from a single minirhizotron, three images after rotation and switching in the correct order, and a segmented image.



A



B

Fig. S4. Dice score during root segmentation model development with RootPainter. The Dice score, ranging from 0 to 1, indicates the agreement between model predictions and ground truth, with higher scores signifying better agreement. (A) Dice score during segmentation model development for rhizotron images from the glasshouse. (B) Dice score during segmentation model development for minirhizotron images from the field. The trend shows continuous improvement as more images are annotated. Training was stopped when the Dice score approached 1, and the best model was selected.

Algorithm Performance and Biological Relevance

Table S6. Internal validation metrics for clustering algorithms.

| Algorithm | Treatment | Silhouette Score | Dunn Index | Calinski Harabasz Index | Davies Bouldin Index |
|-----------|-----------|------------------|------------|-------------------------|----------------------|
| KMeans | T0 | 0.452 | 0.460 | 22.895 | 0.903 |

| | | | | | |
|-------------------------|----|-------|-------|--------|-------|
| AgglomerativeClustering | T0 | 0.434 | 0.577 | 20.784 | 0.970 |
| GaussianMixture | T0 | 0.434 | 0.577 | 20.784 | 0.970 |
| SpectralClustering | T0 | 0.418 | 0.152 | 20.777 | 0.861 |
| Birch | T0 | 0.410 | 0.976 | 11.518 | 0.738 |
| KMeans | T1 | 0.492 | 0.698 | 30.710 | 0.707 |
| AgglomerativeClustering | T1 | 0.508 | 1.067 | 30.317 | 0.637 |
| GaussianMixture | T1 | 0.508 | 1.067 | 30.317 | 0.637 |
| SpectralClustering | T1 | 0.508 | 1.067 | 30.317 | 0.637 |
| Birch | T1 | 0.508 | 1.067 | 30.317 | 0.637 |

Table S7. Relative contributions of algorithms to model performance. Table showing the permutation importance values for each algorithm, key features with their SHAP importance scores, and hypothesised biological relevance.

| Algorithm | Permutation Importance | Key Features | Top SHAP Score | Hypothesised Biological Relevance |
|-------------|------------------------|----------------------------|----------------|--|
| FCM | 0.019 | FCM_centre_x | 0.071 | Horizontal position of fuzzy root clusters; Lateral exploration for soil volume coverage |
| OPTICS | 0.013 | OPTICS_density_points | 0.06 | Ordered clusters density; Sequential root organisation for uptake |
| HDBSCA N | 0.012 | HDBSCAN_density_poi nts | 0.071 | Hierarchical density clusters; Adaptive clustering in moisture pockets |
| DBSCAN | 0.009 | DBSCAN_density_point s | 0.053 | Root cluster density; Adaptive clustering in moisture pockets |
| SLIC | 0.008 | SLIC_centre_x | 0.026 | Superpixel segmentation; Root cluster density optimisation |
| K-mean | 0.007 | K-mean_centre_x | 0.034 | Globular cluster positioning; Generic lateral root exploration |
| Mean-shift | 0.005 | Mean-shift_centre_x | 0.047 | Mode-seeking clusters; Generic lateral root exploration |
| GMM | 0.003 | GMM_centre_x | 0.037 | Gaussian mixture positioning; Generic lateral root exploration |
| Custom | 0.003 | Custom_centre_x | 0.025 | Density-based selection; Generic lateral root exploration |

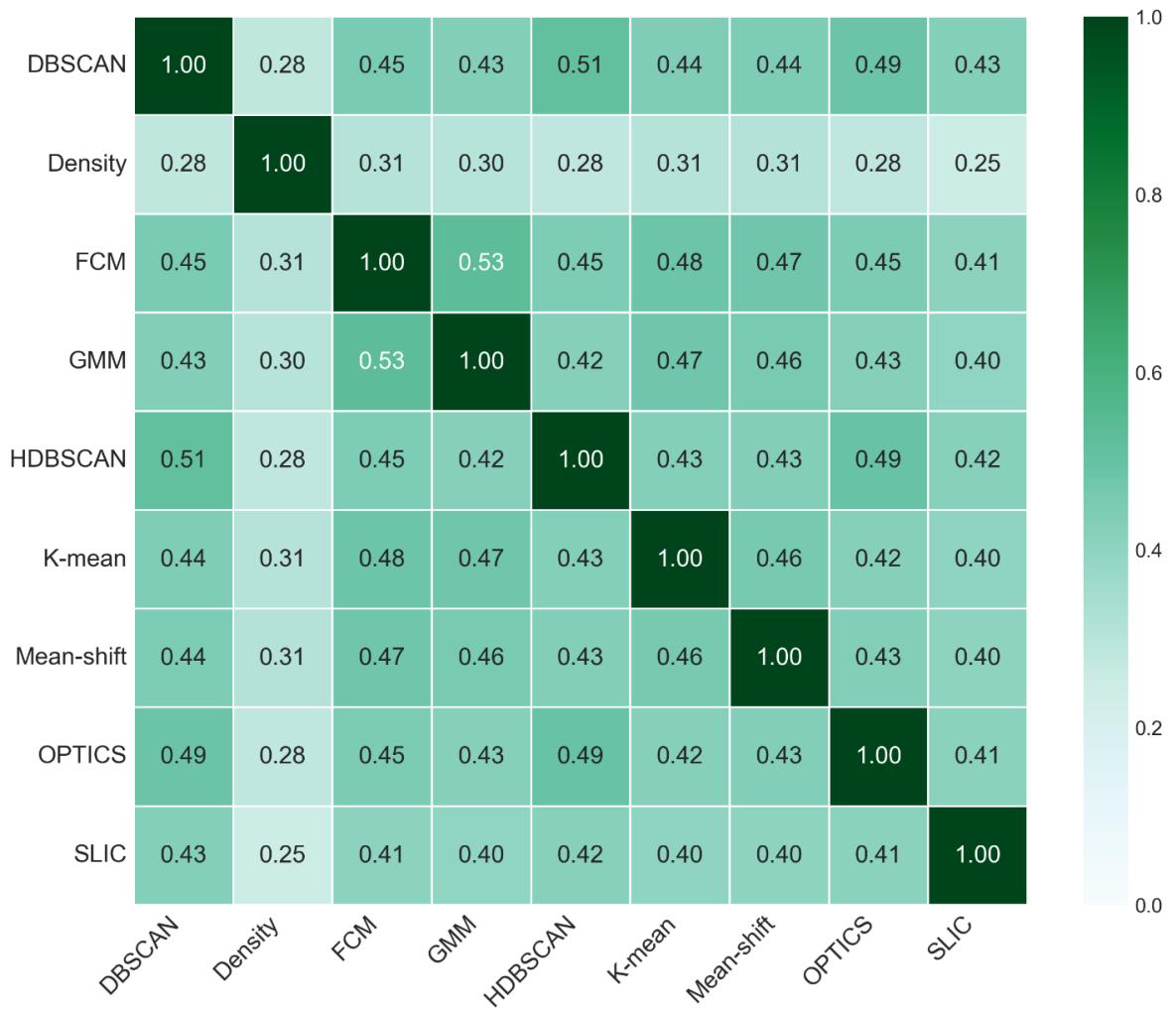


Fig. S5. Algorithm feature correlation matrix. Heatmap of the mean absolute Pearson correlation between features generated by different ART algorithm types. Values closer to 0 (lighter colours) indicate less redundancy and more unique information captured by the respective algorithm pairs.

Biological Correlations and Trait Stability

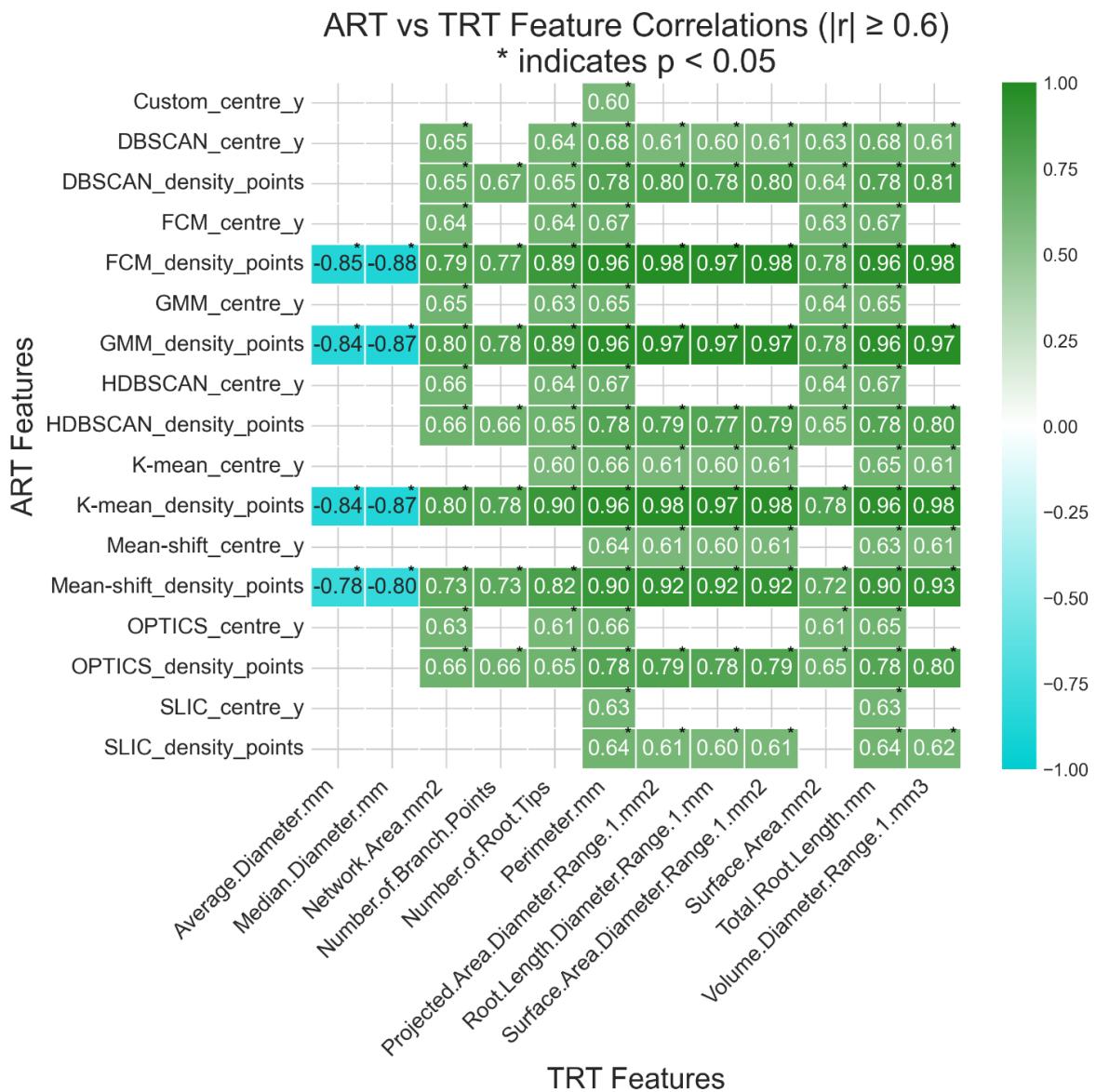


Fig. S6. ART vs TRT Feature Correlations ($|r| \geq 0.6$). Heatmap of significant ($p < 0.05$, indicated by asterisk) Pearson correlations between ART features and TRT features where $|r| \geq 0.6$, showing a broader range of biologically meaningful associations.

ART Features - TRT Features Correlation Network ($|r| \geq 0.7$)

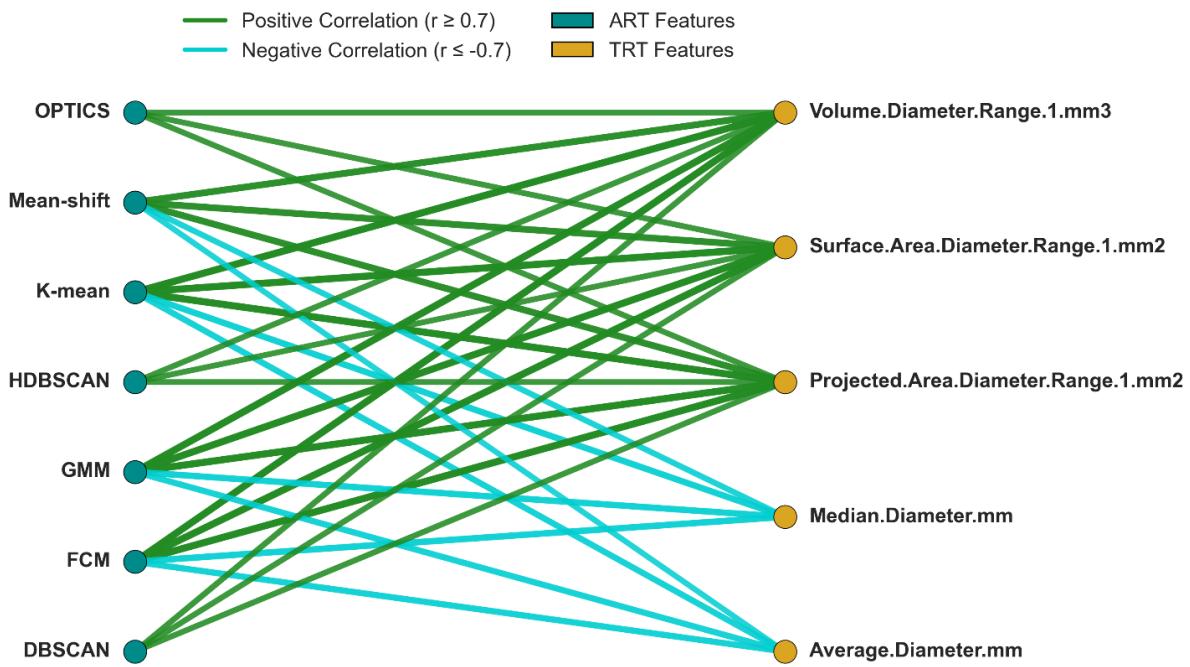


Fig. S7. ART - TRT Feature correlation network. Correlation network visualising significant ($|r| \geq 0.7$, or $|r| \geq 0.6$ if no 0.7 links exist for an algorithm) relationships between ART families (cyan nodes) and TRT features (gold nodes). Green lines denote positive correlations, and cyan lines denote negative correlations. Line thickness is proportional to the absolute correlation strength.

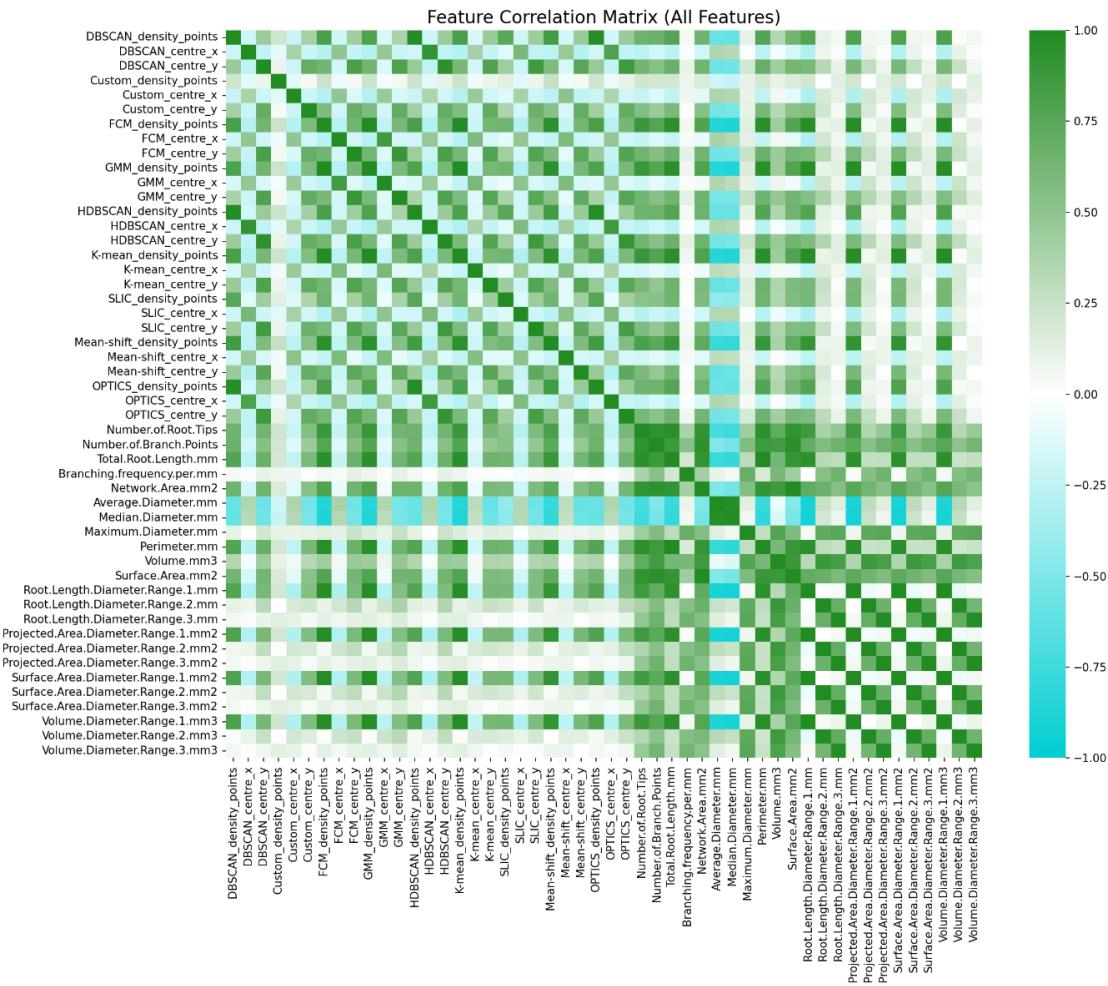


Fig. S8. Feature correlation matrix description: heatmap showing correlations between all features (ART and TRT).

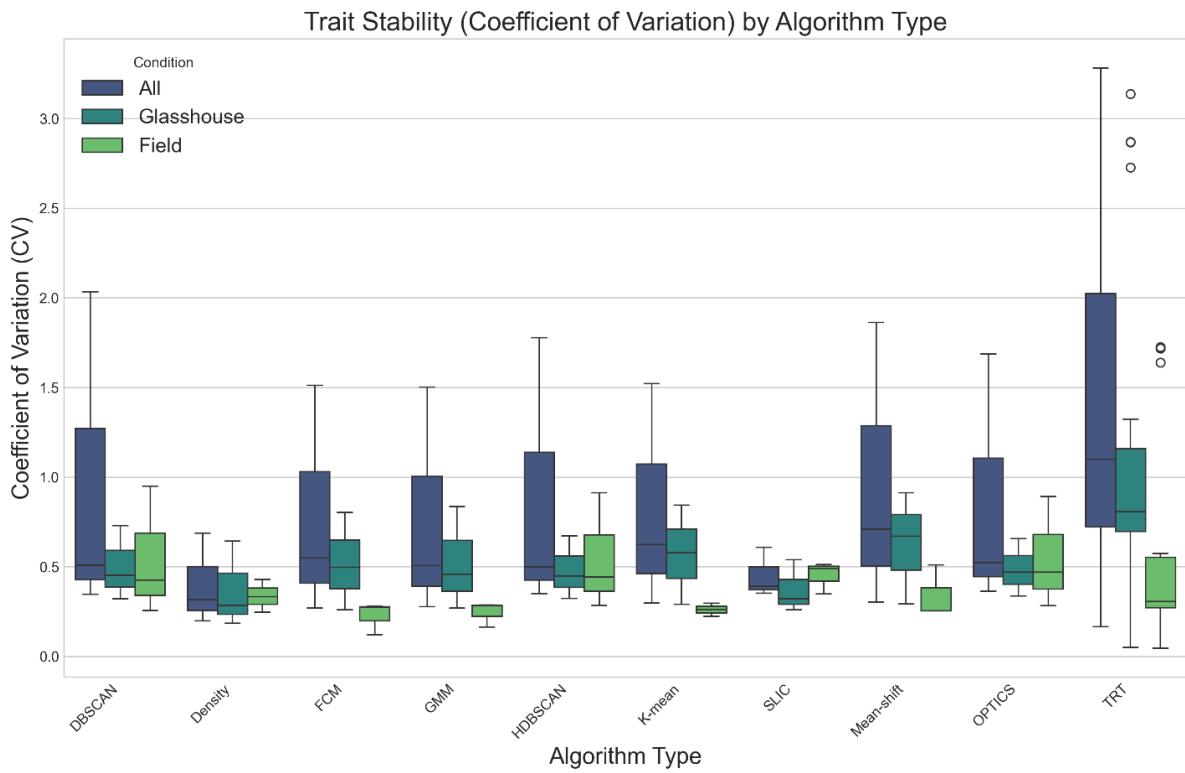


Fig. S9. This boxplot illustrates the Coefficient of Variation (CV) for all 'Algorithm Types' and a combine 'TRT' category for traditional root traits). The data is further categorised by 'Condition' (All, Glasshouse, Field), as indicated by the different coloured boxplots for each algorithm type. The plot aims to compare the trait stability (where a lower CV indicates higher stability or lower relative variability) of features derived from these different algorithms and traditional methods across the specified condition.

Model Development and Validation

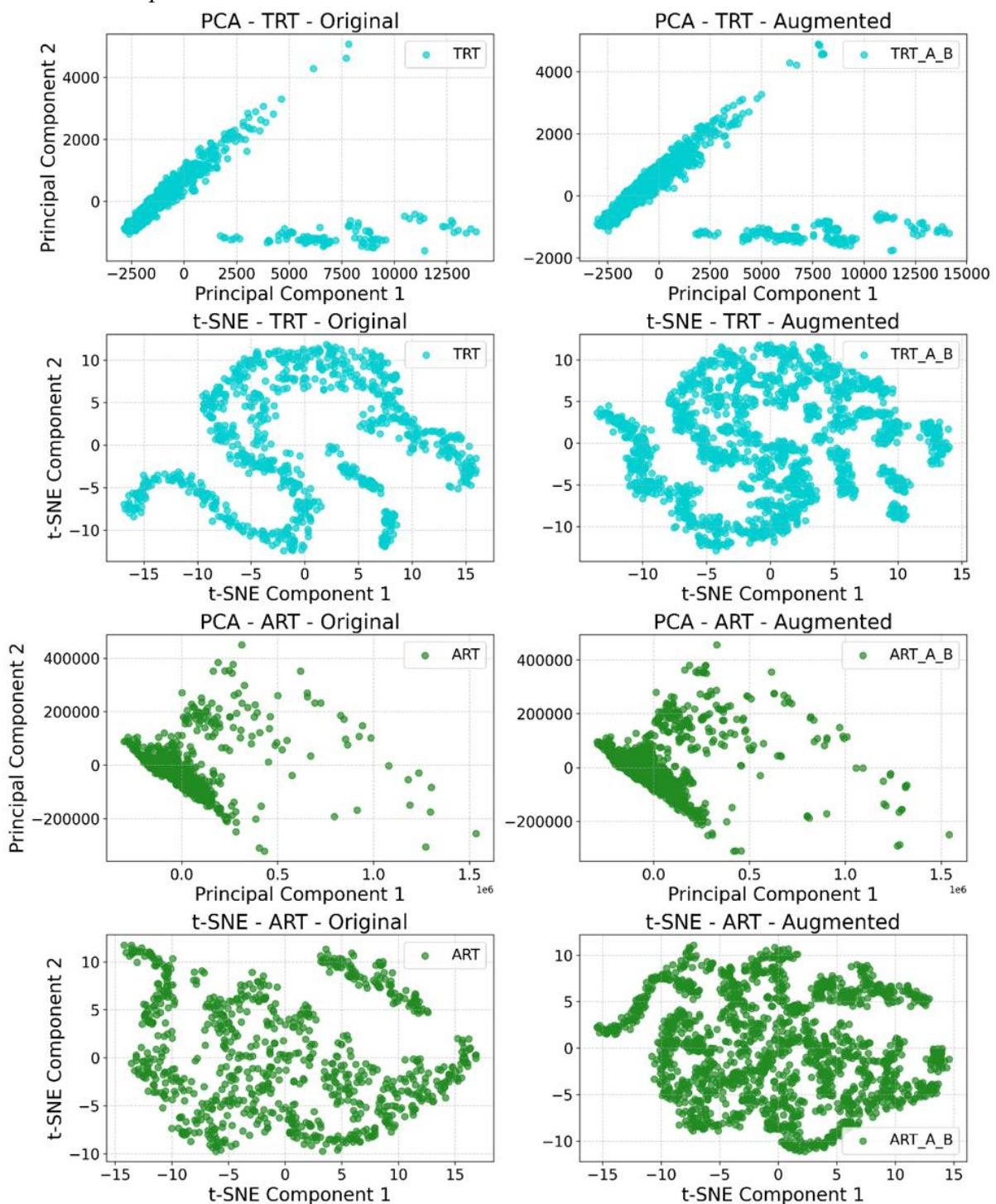


Fig. S10. Comparison of original and augmented datasets using PCA and t-SNE plots. (A) and (B) PCA plot of TRT and augmented and bootstrapped data (TRT_A_B). (C) and (D) PCA plot of Combine and augmented and bootstrapped data (Combine_A_B). (E) and (F) t-SNE plot of TRT and augmented and bootstrapped data (TRT_A_B). (G) and (H) t-SNE plot of Combine and augmented and bootstrapped data (Combine_A_B). Both PCA and t-SNE plots demonstrate increased distribution diversity in augmented datasets.

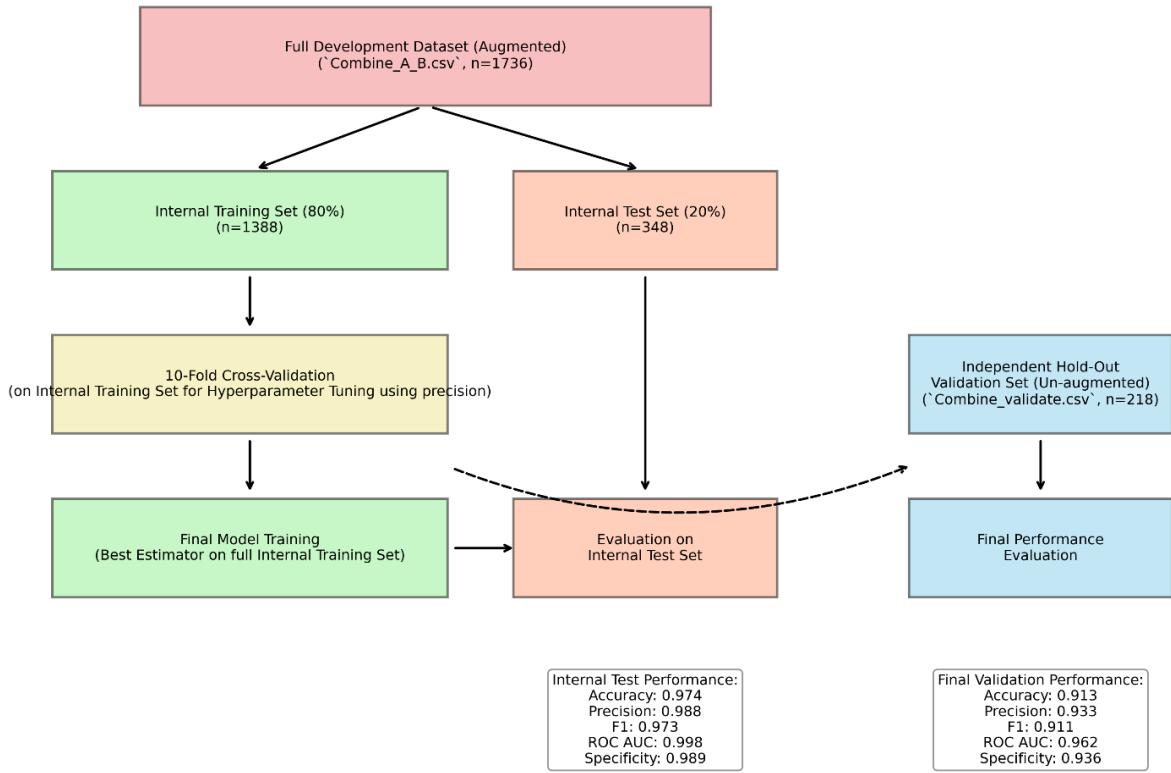


Fig. S11. Validation methodology diagram. Schematic overview of the model training and validation methodology. The full augmented development dataset ($n = 1736$) was split into an internal training set (80%, $n = 1388$) and an internal test set (20%, $n = 348$). Hyperparameters for the Random Forest model was tuned using 10-fold cross-validation on the internal training set (optimising for precision). The final model was trained on the entire internal training set and evaluated on both the internal test set and a completely independent, un-augmented hold-out validation set ($n = 218$). Key performance metrics are displayed for both evaluation stages.

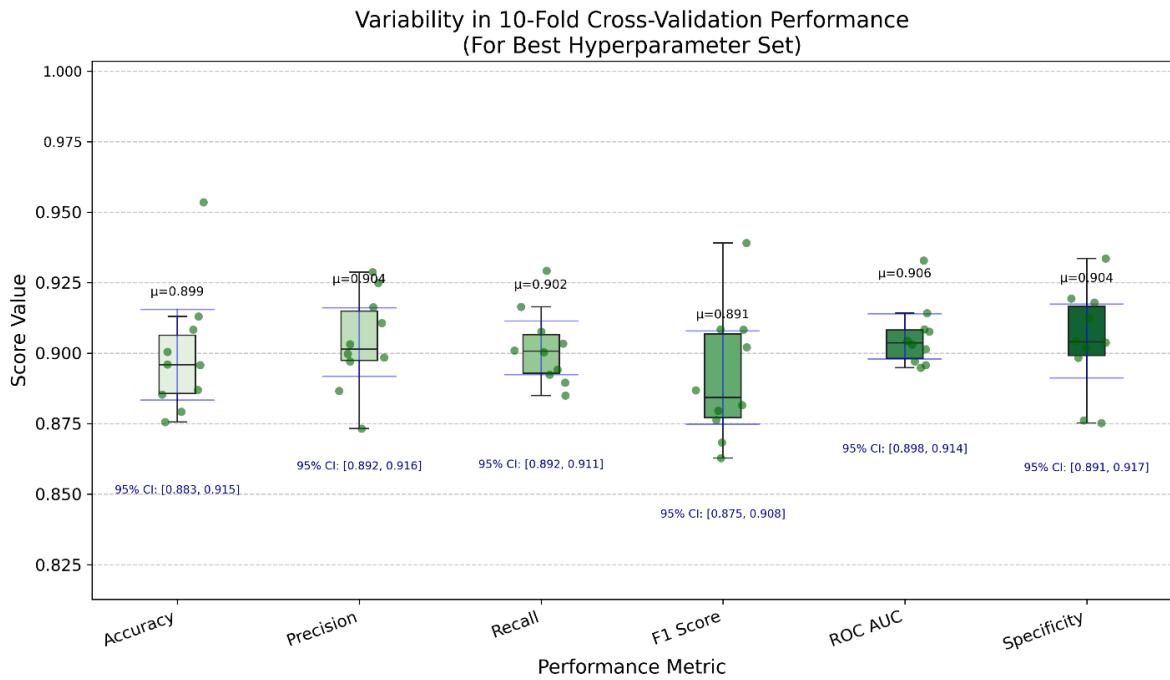


Fig. S12. Cross-validation stability analysis. Variability in 10-fold cross-validation performance for the best hyperparameter set (determined by optimising for precision). Boxplots illustrate the distribution of key performance metrics (Accuracy, Precision, Recall, F1 Score, ROC AUC, and Specificity) across the 10 folds of cross-validation performed on the internal training set. For each metric, the mean value (μ) and 95% confidence interval (CI) are shown. Individual points represent scores from each fold. The low variability (average CV% = 1.95%) demonstrates the model's stability across different data partitions.

Table S8. Cross-Validation stability metrics. Quantitative assessment of model stability across the 10-fold cross-validation for the best hyperparameter set. For each performance metric, the table provides the mean value, standard deviation (Std), coefficient of variation (CV%), and 95% confidence interval width (CI_Width). The consistently low CV% values (1.242%-2.591%) indicate excellent stability regardless of data partitioning, with precision (our optimization target) showing particularly high consistency (CV% = 1.876%).

| Metric | Mean | Std | CV(%) | CI_Lower | CI_Upper | CI_Width | N |
|-------------|----------|----------|----------|----------|----------|----------|----|
| Accuracy | 0.899406 | 0.022474 | 2.498809 | 0.883329 | 0.915483 | 0.032154 | 10 |
| Precision | 0.903876 | 0.016955 | 1.87583 | 0.891747 | 0.916005 | 0.024258 | 10 |
| Recall | 0.901885 | 0.013251 | 1.469292 | 0.892405 | 0.911364 | 0.018959 | 10 |
| F1 Score | 0.891315 | 0.023091 | 2.590689 | 0.874797 | 0.907834 | 0.033037 | 10 |
| ROC | 0.905936 | 0.01125 | 1.241834 | 0.897888 | 0.913984 | 0.016096 | 10 |
| AUC | | | | | | | |
| Specificity | 0.904276 | 0.018284 | 2.021932 | 0.891196 | 0.917355 | 0.026159 | 10 |

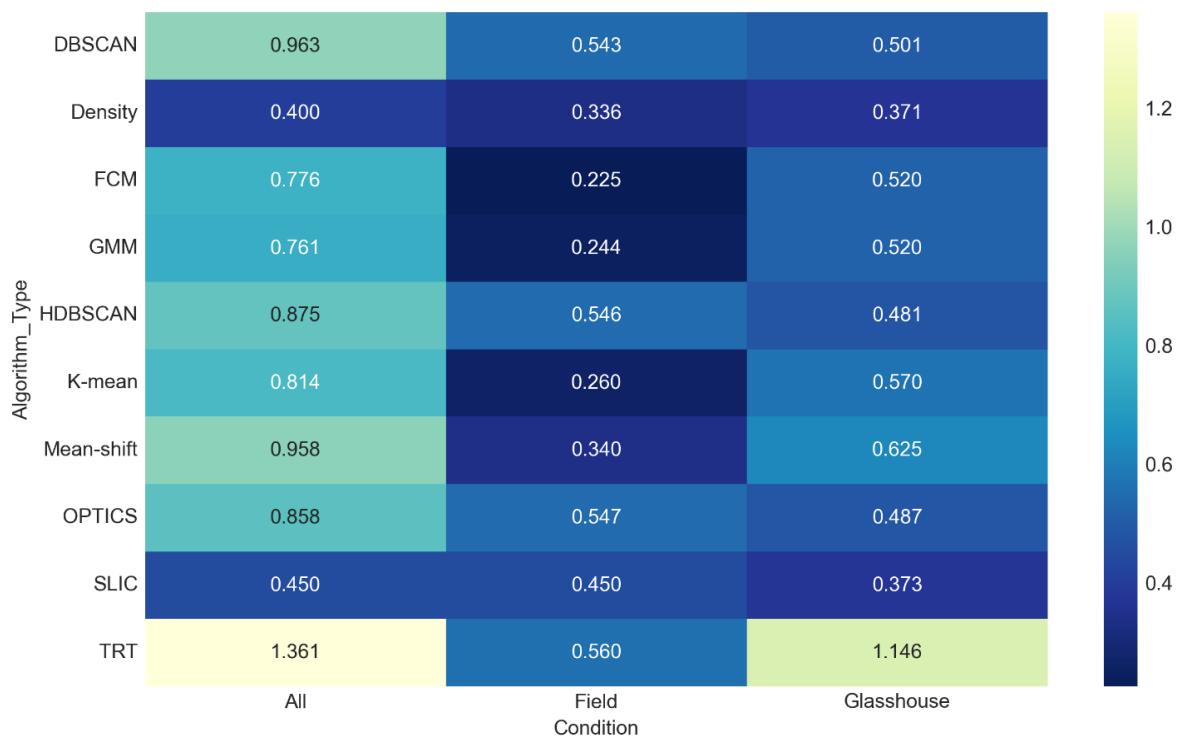


Fig. S13. Mean CV by algorithm type and condition Description: Heatmap of mean coefficient of variation by algorithm type and experimental condition, identifying algorithms with best cross-environment stability.

Feature Importance and Model Performance

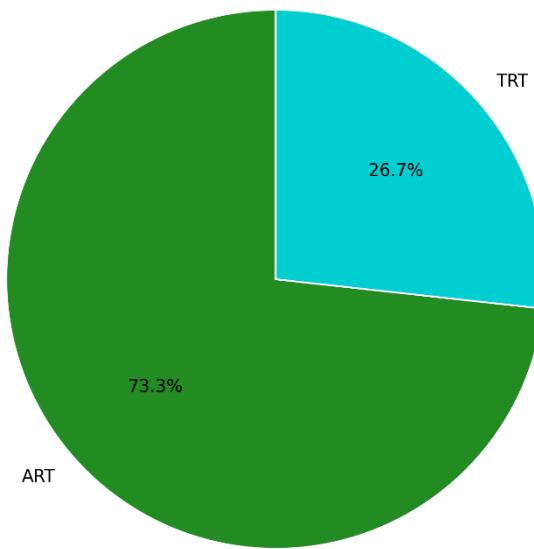


Fig. S14. Feature type contribution. Relative contribution of Algorithmic Root Traits (ARTs) and Traditional Root Traits (TRTs) to model performance based on mean absolute SHAP values from the 'All Features (ART+TRT)' model. Despite representing only 54% of the total features, ARTs contribute 73.3% of the predictive power.

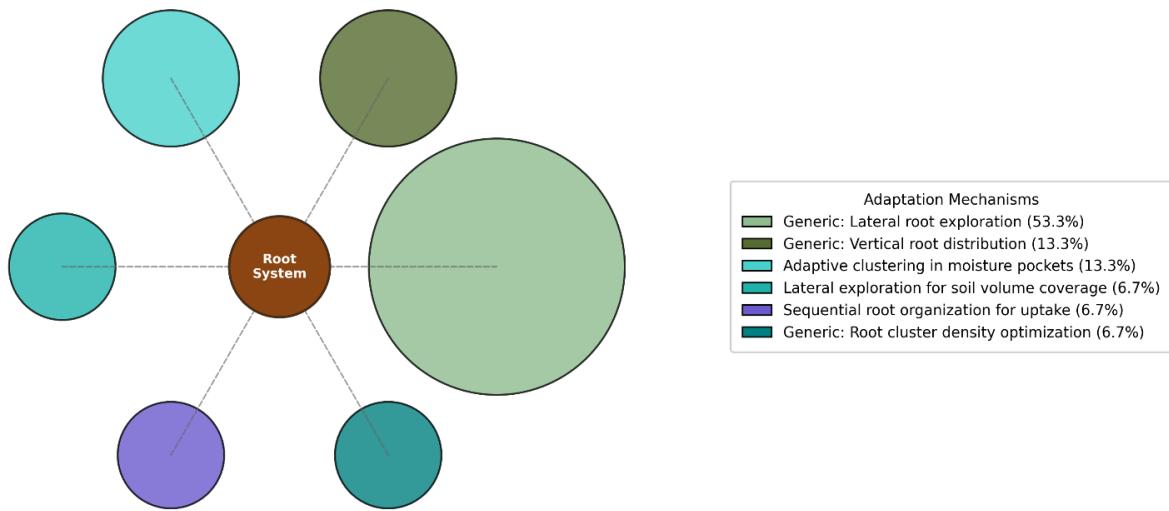


Fig. S15. Hypothesised drought adaptation mechanisms. Conceptual representation of the drought adaptation mechanisms captured by the top ART features from the 'All Features (ART+TRT)' model. The diagram illustrates the relative prominence of different adaptive strategies based on the frequency of top ART features associated with each mechanism. Circle size is proportional to this frequency. Key adaptations highlighted in the legend include "Generic: Lateral root exploration" (53.3%), "Generic: Vertical root distribution" (13.3%), and "Adaptive clustering in moisture pockets" (13.3%).

Note: The "Hypothesised Drought Adaptation Mechanisms" diagram was derived from the Random Forest model trained on the "All Features (ART+TRT)" dataset. The importance of each Algorithmic Root Trait (ART) feature within this model was determined by its mean absolute SHAP (SHapley Additive exPlanations) value. The top 15 ART features, ranked by these SHAP values (see Table S7 for details on these features, their SHAP scores, and their individual hypothesised biological relevance), were selected for this analysis.

Each of these 15 ART features was programmatically assigned a 'Hypothesised Drought Adaptation Mechanism' by the `get_drought_mechanism` function within the `baseline_comparison_shap_6.py` script. This function applies a set of predefined rules to interpret patterns in the ART feature names (e.g., `_centre_y` indicating vertical distribution, `_density_points` suggesting clustering, or specific feature names mapping to more granular mechanisms).

The frequency of occurrence for these programmatically assigned mechanisms was then tallied across the 15 top ART features. These frequencies were used to calculate the percentage contribution for each distinct mechanism category displayed in the diagram's legend, which may involve grouping similar individual mechanisms. The categories and their contributions, accounting for all 15 top ART features, are:

Generic: Lateral root exploration: 53.3% (representing 8 out of 15 features)

Generic: Vertical root distribution: 13.3% (representing 2 out of 15 features)

Adaptive clustering in moisture pockets: 13.3% (representing 2 out of 15 features)

Lateral exploration for soil volume coverage: 6.7% (representing 1 out of 15 features)

Sequential root organisation for uptake: 6.7% (representing 1 out of 15 features)
 Generic: Root cluster density optimisation: 6.7% (representing 1 out of 15 features)

Visually, the diagram connects a central "Root System" node to peripheral circles, each representing one of these mechanism categories. The radius of each mechanism's circle is scaled proportionally to its frequency count ($\text{radius} \propto \text{frequency}^{0.65}$), thereby illustrating the relative prominence of each hypothesised adaptive strategy among the features most influential to the model.

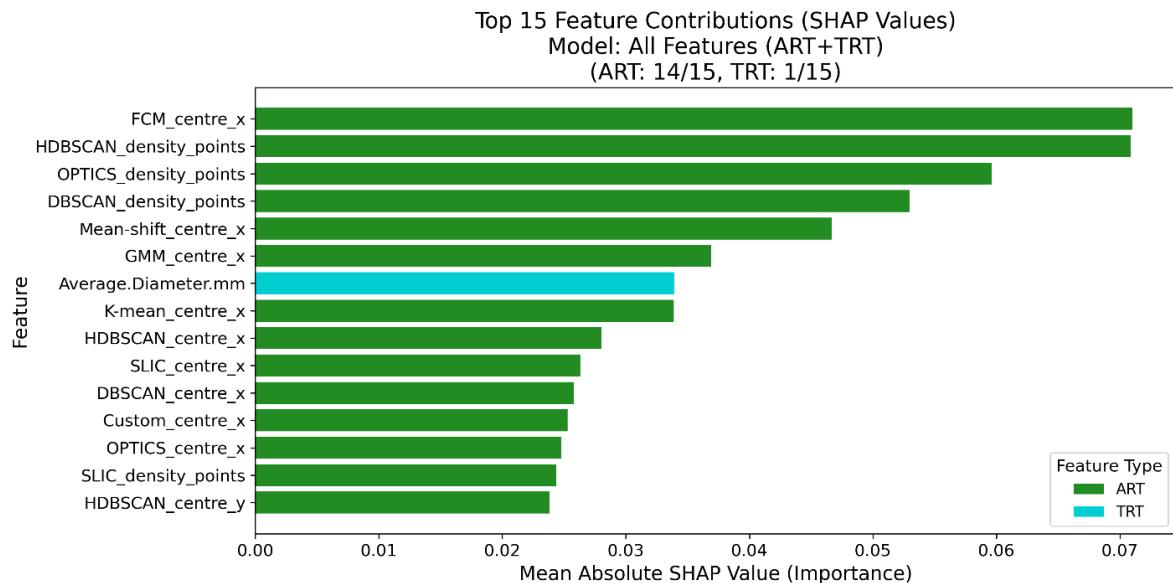


Fig. S16. Top 15 Individual feature contributions. Top 15 most important features from the 'All Features (ART+TRT)' model, ranked by mean absolute SHAP values. ART features dominate the top positions, with *FCM_centre_x* and *HDBSCAN_density_points* being the most impactful overall.

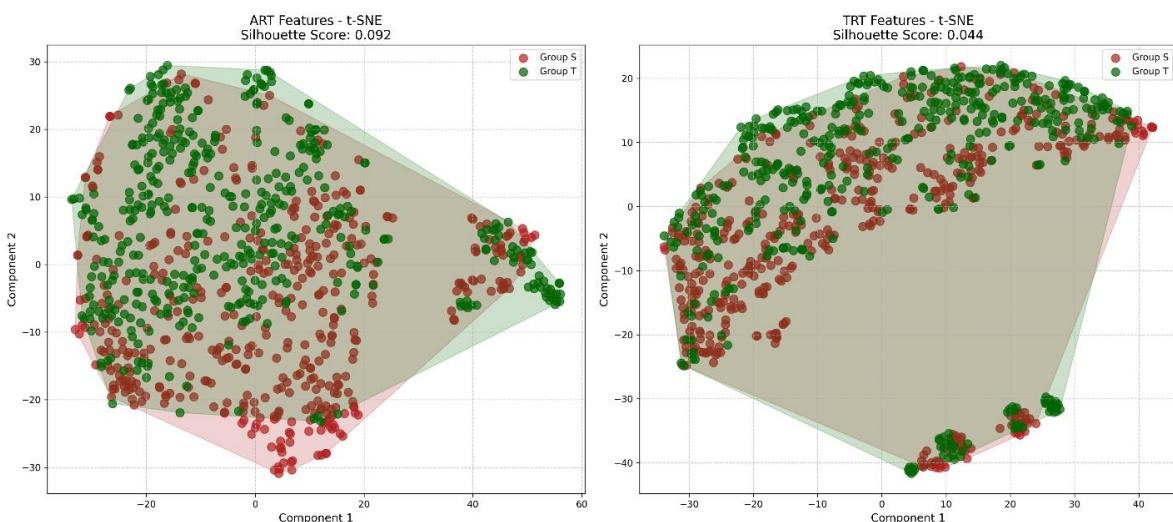


Fig. S17. t-SNE Comparison of ART vs TRT feature spaces. t-SNE projections showing drought-tolerant (green) and drought-susceptible (red) genotypes in (A) ART and (B) TRT feature spaces. Data points include 95% confidence ellipses to highlight group separation. ART features provide substantially superior group separation

(silhouette score: 0.092 vs 0.044).

Table S9. Comprehensive feature importance values across models. This table presents the importance values for all features across different model configurations (All ARTs, All Features (ART+TRT), All TRTs, selected 4 ARTs, Selected 4 TRTs, Selected TRTs+ARTs, and Baseline). This detailed quantitative data supports the feature importance analysis presented in Figures S14 and S16 and demonstrates the consistent dominance of ART features, particularly FCM centre x, HDBSCAN density points, and OPTICS density points across models.

| Feature | Importance | Feature_Type | Model |
|---------------------------|------------|--------------|------------------------|
| FCM_centre_x | 0.081651 | ART | All ARTs |
| DBSCAN_density_points | 0.074981 | ART | All ARTs |
| OPTICS_density_points | 0.073769 | ART | All ARTs |
| HDBSCAN_density_points | 0.072179 | ART | All ARTs |
| Mean-shift_centre_x | 0.053449 | ART | All ARTs |
| GMM_centre_x | 0.052605 | ART | All ARTs |
| K-mean_centre_x | 0.04229 | ART | All ARTs |
| SLIC_centre_x | 0.042076 | ART | All ARTs |
| DBSCAN_centre_y | 0.041266 | ART | All ARTs |
| DBSCAN_centre_x | 0.035813 | ART | All ARTs |
| HDBSCAN_centre_y | 0.034441 | ART | All ARTs |
| OPTICS_centre_x | 0.032205 | ART | All ARTs |
| GMM_centre_y | 0.031736 | ART | All ARTs |
| HDBSCAN_centre_x | 0.03169 | ART | All ARTs |
| FCM_centre_y | 0.030842 | ART | All ARTs |
| OPTICS_centre_y | 0.029639 | ART | All ARTs |
| Custom_centre_x | 0.027288 | ART | All ARTs |
| Mean-shift_centre_y | 0.026375 | ART | All ARTs |
| Mean-shift_density_points | 0.025637 | ART | All ARTs |
| SLIC_density_points | 0.025428 | ART | All ARTs |
| SLIC_centre_y | 0.024351 | ART | All ARTs |
| K-mean_centre_y | 0.021728 | ART | All ARTs |
| K-mean_density_points | 0.020601 | ART | All ARTs |
| FCM_density_points | 0.019961 | ART | All ARTs |
| GMM_density_points | 0.017776 | ART | All ARTs |
| Custom_density_points | 0.016273 | ART | All ARTs |
| Custom_centre_y | 0.013953 | ART | All ARTs |
| FCM_centre_x | 0.070996 | ART | All Features (ART+TRT) |
| HDBSCAN_density_points | 0.07089 | ART | All Features (ART+TRT) |
| OPTICS_density_points | 0.059607 | ART | All Features (ART+TRT) |
| DBSCAN_density_points | 0.052967 | ART | All Features (ART+TRT) |
| Mean-shift_centre_x | 0.04668 | ART | All Features (ART+TRT) |
| GMM_centre_x | 0.036932 | ART | All Features (ART+TRT) |
| Average.Diameter.mm | 0.033912 | TRT | All Features (ART+TRT) |
| K-mean centre x | 0.033888 | ART | All Features (ART+TRT) |

| | | | |
|-------------------------------------|----------|-----|------------------------|
| HDBSCAN_centre_x | 0.028036 | ART | All Features (ART+TRT) |
| SLIC_centre_x | 0.026317 | ART | All Features (ART+TRT) |
| DBSCAN_centre_x | 0.025789 | ART | All Features (ART+TRT) |
| Custom_centre_x | 0.025295 | ART | All Features (ART+TRT) |
| OPTICS_centre_x | 0.024782 | ART | All Features (ART+TRT) |
| SLIC_density_points | 0.024365 | ART | All Features (ART+TRT) |
| HDBSCAN_centre_y | 0.023816 | ART | All Features (ART+TRT) |
| Branching.frequency.per.mm | 0.023256 | TRT | All Features (ART+TRT) |
| DBSCAN_centre_y | 0.022083 | ART | All Features (ART+TRT) |
| GMM_centre_y | 0.021627 | ART | All Features (ART+TRT) |
| SLIC_centre_y | 0.018402 | ART | All Features (ART+TRT) |
| FCM_centre_y | 0.015897 | ART | All Features (ART+TRT) |
| K-mean_centre_y | 0.015859 | ART | All Features (ART+TRT) |
| OPTICS_centre_y | 0.015794 | ART | All Features (ART+TRT) |
| Number.of.Root.Tips | 0.015156 | TRT | All Features (ART+TRT) |
| Mean-shift_centre_y | 0.014961 | ART | All Features (ART+TRT) |
| Maximum.Diameter.mm | 0.01281 | TRT | All Features (ART+TRT) |
| Root.Length.Diameter.Range.1.mm | 0.012557 | TRT | All Features (ART+TRT) |
| Volume.Diameter.Range.3.mm3 | 0.012071 | TRT | All Features (ART+TRT) |
| Surface.Area.Diameter.Range.2.mm2 | 0.011968 | TRT | All Features (ART+TRT) |
| Mean-shift_density_points | 0.011399 | ART | All Features (ART+TRT) |
| Number.of.Branch.Points | 0.010833 | TRT | All Features (ART+TRT) |
| Custom_density_points | 0.010793 | ART | All Features (ART+TRT) |
| Projected.Area.Diameter.Range.2.mm2 | 0.009771 | TRT | All Features (ART+TRT) |
| Surface.Area.Diameter.Range.1.mm2 | 0.009741 | TRT | All Features (ART+TRT) |
| FCM_density_points | 0.009667 | ART | All Features (ART+TRT) |
| Projected.Area.Diameter.Range.1.mm2 | 0.009633 | TRT | All Features (ART+TRT) |
| Root.Length.Diameter.Range.2.mm | 0.009402 | TRT | All Features (ART+TRT) |
| Median.Diameter.mm | 0.009311 | TRT | All Features (ART+TRT) |
| K-mean_density_points | 0.0093 | ART | All Features (ART+TRT) |
| Root.Length.Diameter.Range.3.mm | 0.009218 | TRT | All Features (ART+TRT) |
| Network.Area.mm2 | 0.008909 | TRT | All Features (ART+TRT) |
| Volume.Diameter.Range.2.mm3 | 0.008825 | TRT | All Features (ART+TRT) |
| Surface.Area.Diameter.Range.3.mm2 | 0.008762 | TRT | All Features (ART+TRT) |
| Custom_centre_y | 0.008753 | ART | All Features (ART+TRT) |
| Volume.Diameter.Range.1.mm3 | 0.008732 | TRT | All Features (ART+TRT) |
| Perimeter.mm | 0.008694 | TRT | All Features (ART+TRT) |
| Total.Root.Length.mm | 0.00864 | TRT | All Features (ART+TRT) |
| Volume.mm3 | 0.008469 | TRT | All Features (ART+TRT) |
| Surface.Area.mm2 | 0.008429 | TRT | All Features (ART+TRT) |
| Projected.Area.Diameter.Range.3.mm2 | 0.008286 | TRT | All Features (ART+TRT) |

| | | | |
|-------------------------------------|----------|-----|------------------------|
| GMM_density_points | 0.007721 | ART | All Features (ART+TRT) |
| Average.Diameter.mm | 0.088393 | TRT | All TRTs |
| Branching.frequency.per.mm | 0.084057 | TRT | All TRTs |
| Number.of.Root.Tips | 0.06687 | TRT | All TRTs |
| Root.Length.Diameter.Range.2.mm | 0.049926 | TRT | All TRTs |
| Maximum.Diameter.mm | 0.046677 | TRT | All TRTs |
| Volume.mm3 | 0.045621 | TRT | All TRTs |
| Surface.Area.Diameter.Range.2.mm2 | 0.044908 | TRT | All TRTs |
| Network.Area.mm2 | 0.043871 | TRT | All TRTs |
| Number.of.Branch.Points | 0.041907 | TRT | All TRTs |
| Surface.Area.mm2 | 0.041197 | TRT | All TRTs |
| Volume.Diameter.Range.2.mm3 | 0.04109 | TRT | All TRTs |
| Root.Length.Diameter.Range.1.mm | 0.040422 | TRT | All TRTs |
| Projected.Area.Diameter.Range.2.mm2 | 0.04001 | TRT | All TRTs |
| Projected.Area.Diameter.Range.1.mm2 | 0.03916 | TRT | All TRTs |
| Volume.Diameter.Range.1.mm3 | 0.038128 | TRT | All TRTs |
| Projected.Area.Diameter.Range.3.mm2 | 0.035866 | TRT | All TRTs |
| Surface.Area.Diameter.Range.1.mm2 | 0.033092 | TRT | All TRTs |
| Volume.Diameter.Range.3.mm3 | 0.03198 | TRT | All TRTs |
| Perimeter.mm | 0.031304 | TRT | All TRTs |
| Root.Length.Diameter.Range.3.mm | 0.030753 | TRT | All TRTs |
| Total.Root.Length.mm | 0.029774 | TRT | All TRTs |
| Surface.Area.Diameter.Range.3.mm2 | 0.028475 | TRT | All TRTs |
| Median.Diameter.mm | 0.026519 | TRT | All TRTs |
| | | | Baseline |
| Total.Root.Length.mm | 1 | TRT | (Total.Root.Length.mm) |
| FCM_centre_x | 0.34095 | ART | Selected 4 ARTs |
| HDBSCAN_density_points | 0.23341 | ART | Selected 4 ARTs |
| OPTICS_density_points | 0.227308 | ART | Selected 4 ARTs |
| DBSCAN_density_points | 0.198332 | ART | Selected 4 ARTs |
| Volume.mm3 | 0.266735 | TRT | Selected 4 TRTs |
| Total.Root.Length.mm | 0.25701 | TRT | Selected 4 TRTs |
| Network.Area.mm2 | 0.246895 | TRT | Selected 4 TRTs |
| Surface.Area.mm2 | 0.229359 | TRT | Selected 4 TRTs |
| FCM_centre_x | 0.228256 | ART | Selected TRTs+ARTs |
| HDBSCAN_density_points | 0.15278 | ART | Selected TRTs+ARTs |
| OPTICS_density_points | 0.138376 | ART | Selected TRTs+ARTs |
| DBSCAN_density_points | 0.125975 | ART | Selected TRTs+ARTs |
| Volume.mm3 | 0.092278 | TRT | Selected TRTs+ARTs |
| Total.Root.Length.mm | 0.087952 | TRT | Selected TRTs+ARTs |
| Network.Area.mm2 | 0.087207 | TRT | Selected TRTs+ARTs |
| Surface.Area.mm2 | 0.087177 | TRT | Selected TRTs+ARTs |

Table S10. Comparative model performance metrics. This table presents the quantitative performance metrics (Accuracy, Precision, Recall, F1, ROC_AUC, CV_Accuracy_Mean, CV_Accuracy_Std) for each model configuration tested (Baseline, selected 4 TRTs, Selected 4 ARTs, Selected TRTs+ARTs, All TRTs, All ARTs, All Features). The data numerically demonstrates the superior performance of ART-based models and the synergistic effect when combining ARTs with TRTs.

| Model | Features | Accuracy | Precision | Recall | F1 | ROC_AUC | CV_Accuracy_Mean | CV_Accuracy_Std |
|---|----------|----------|-----------|----------|----------|----------|------------------|-----------------|
| Baseline (Total.Root.Length.mm) | 1 | 0.577586 | 0.577944 | 0.577828 | 0.577499 | 0.61656 | 0.588145 | 0.022623 |
| Selected 4 TRTs | 4 | 0.70977 | 0.70981 | 0.709831 | 0.709768 | 0.76807 | 0.7143 | 0.024365 |
| Selected 4 ARTs | 4 | 0.850575 | 0.850575 | 0.850621 | 0.85057 | 0.924931 | 0.861749 | 0.010493 |
| Selected TRTs+ARTs | 8 | 0.899425 | 0.900199 | 0.899181 | 0.899325 | 0.961912 | 0.910154 | 0.018699 |
| All TRTs | 23 | 0.856322 | 0.860423 | 0.855708 | 0.855745 | 0.926582 | 0.882512 | 0.030886 |
| All ARTs | 27 | 0.962644 | 0.962861 | 0.96254 | 0.962629 | 0.996779 | 0.975821 | 0.019646 |
| All Features (ART+TRT) | 50 | 0.974138 | 0.974675 | 0.973969 | 0.974121 | 0.997605 | 0.974093 | 0.019076 |

Appendix 2: Supplementary Material of Chapter 5

SUPPLEMENTARY INFORMATION

Tissue-specific Metabolomic Networks Orchestrate Osmotic Stress Adaptation in Wheat

General Supplementary Information

This document provides details about methods and findings that support the manuscript "Tissue-specific Metabolomic Networks Orchestrate Osmotic Stress Adaptation in Wheat." The methods section outlines protocols for metabolite identification, including database matching, structure techniques, and pathway mapping. The results section presents analyses of responses to stress, featuring profiling of groups, analyses across genotypes, and visualisations that complement findings. These materials include figures, analyses, and tables that show the approach and provide context for metabolic adaptations in wheat under stress conditions.

The following Supporting Information is available for this article:

Table S1: Network analysis terminology

| Network Metric / Concept | Definition (General Network Science) | Conceptual Calculation / Underlying Logic | General Meaning/Interpretation | Biological Definition & Interpretation | Relevant Reference(s) for Biological |
|--------------------------|--|---|--|---|--------------------------------------|
| Network Density | The proportion of existing connections in a network is relative to the total number of possible connections. | Logic: Measures how “full” or “sparse” a network is. Calculated as $2E / (N(N-1))$ for undirected networks or $E / (N(N-1))$ for directed, where E is edges, N is nodes. Ranges from 0 (no connections) to 1 (fully connected). Most biological networks have low density | Biologically, high density can suggest a system with many interactions, potentially leading to faster signal propagation, robustness to loss of individual connections, or higher functional redundancy. Low density might imply | Barabási and Oltvai (2004); (Erdos and Rényi, 1960; Graharend-Belau et al., 2009) | |

| Network Metric / Concept | Definition (General Network Science) | Conceptual Calculation / Underlying Logic | General Meaning/Interpretation | Biological Definition & Biological Interpretation | Relevant Reference(s) for |
|---|---|--|--|--|--|
| | | (< 0.1) due to metabolic constraints. | specialisation or efficiency with fewer links. | | |
| Transitivity (Global Clustering Coefficient) | A measure of the overall tendency for nodes in a network to cluster together. Specifically, it quantifies the likelihood that two neighbors of a given node are connected. Logic: Assesses the "cliquishness" of a network. Calculated as $3 * (\text{number of triangles}) / (\text{number of connected triplets of nodes})$. A triangle is three nodes all connected to each other. A connected triplet is a node connected to two others. Ranges from 0 to 1. Values > 0.3 typically indicate significant clustering in biological networks. | Logic: Assesses the "cliquishness" of a network. Calculated as $3 * (\text{number of triangles}) / (\text{number of connected triplets of nodes})$. A triangle is three nodes all connected to each other. A connected triplet is a node connected to two others. Ranges from 0 to 1. Values > 0.3 typically indicate significant clustering in biological networks. | High transitivity often indicates the presence of tightly coupled functional modules or pathways where components work closely together. It can contribute to network robustness and efficient local information processing. | Barabási and Oltvai (2004); Watts and Strogatz, 1998 | |
| Modularity (Q) | A measure of the strength of division of a network into modules (also called groups, clusters, or communities). Networks with high modularity have dense connections between the nodes within modules but sparse connections between nodes in different modules. Logic: Quantifies how well a network can be partitioned into non-overlapping communities. Algorithms (e.g., Louvain) optimise Q by comparing the fraction of edges within communities to the expected fraction if edges were random. Higher Q indicates stronger community structure. Ranges from -0.5 to 1. Values > 0.3 generally indicate strong community structure. Biological networks typically show Q = 0.3-0.7. | Logic: Quantifies how well a network can be partitioned into non-overlapping communities. Algorithms (e.g., Louvain) optimise Q by comparing the fraction of edges within communities to the expected fraction if edges were random. Higher Q indicates stronger community structure. Ranges from -0.5 to 1. Values > 0.3 generally indicate strong community structure. Biological networks typically show Q = 0.3-0.7. | Biologically, modularity often reflects functional specialisation. Different modules might correspond to distinct biological pathways, complexes, or processes that can operate independently, allowing for adaptability and robustness. | (Barabási and Oltvai, 2004; Fortunato, 2010; Girvan and Newman, 2002); Newman (2006) | |
| Mean Path Length | The average number of edges along the shortest paths connecting all possible pairs of nodes in a network. | Logic: Measures the typical separation between any two nodes, reflecting the efficiency of information or mass transfer across the network. Calculated by | Short mean path lengths are characteristic of "small-world" networks and imply efficient communication or rapid propagation of signals | Latora and Marchiori, 2001; Watts and Strogatz (1998) | (Barabási and Oltvai, 2004; Latora and Marchiori, 2001; Watts and Strogatz (1998)) |

| Network Metric / Concept | Definition (General Network Science) | Conceptual Calculation / Underlying Logic | General Meaning/Interpretation | Biological Definition & Biological Interpretation | Relevant Reference(s) for |
|--|---|--|--|---|--|
| | | finding all the shortest paths, summing their lengths, and dividing by the total number of pairs. | throughout the system. This can be crucial for coordinated responses. | | |
| Degree Centrality (Node Degree) | The number of direct connections (edges) a node has to other nodes in the network. | Logic: A simple, local measure of a node's direct influence or activity within the network. Count the number of edges incident to a node. | Nodes with high degree (hubs) are often critical for network function and stability. They can act as control points, integrate signals, or be essential components of many pathways. Loss of hubs can be highly disruptive. | | Barabási and Oltvai (2004); (Clauset <i>et al.</i> , 2009; Jeong et al., 2001) |
| Network Components (Connected Components) | Subgraphs in which any two nodes are connected to each other by paths, and which are connected to no additional nodes in the supergraph. A network can consist of one or multiple components. | Logic: Identifies disconnected parts of a network. Algorithms like Breadth-First Search or Depth-First Search can identify all nodes reachable from a starting node, thus defining a component. | The number and size of components indicate the overall cohesiveness of a network. A single large component suggests an integrated system. Multiple components might suggest independent functional units or fragmentation due to stress or missing data. | | (Erdos and Rényi, 1960); Newman (2010) |
| Assortativity (Degree Assortativity) | A measure of the preference for nodes in a network to attach to other nodes that are similar (assortative) or dissimilar (disassortative) in terms of their degree. | Logic: Assesses correlation between the degrees of connected nodes. Calculated as the Pearson correlation coefficient of the degrees at either end of every edge. Positive values indicate assortativity; negative values indicate disassortativity. | Biological networks are often disassortative (hubs connect to low-degree nodes), which can make them robust to random failures but vulnerable to targeted attacks on hubs. It can also reflect hierarchical organisation. | | Newman (2002) |

| Network Metric / Concept | Definition (General Network Science) | Conceptual Calculation / Underlying Logic | General Meaning/Interpretation | Biological Definition & Biological Interpretation | Relevant Reference(s) for |
|--|---|---|---|---|--|
| Network Stability/Robustness | The ability of a network to maintain its essential structural properties despite perturbations (e.g., node/edge removal, changes over time). | Logic: Assessed through various means, e.g., measuring how key metrics (density, path length, size of largest component) change after simulated perturbations, or how module structures persist across different states (module preservation). | High stability/robustness is crucial for biological systems to function reliably in fluctuating environments or under stress. It reflects the system's resilience and capacity to maintain homeostasis or adapt effectively. | | Albert et al. (2000); (Fait et al., 2020) |
| Hub Distribution/Connectivity Pattern | The way high-degree nodes (hubs) are distributed across the network - either concentrated (few very high-degree nodes) or distributed (more evenly spread). | Logic: Analyse the degree distribution curve, often through rank-degree plots or power-law fitting. Concentrated = steep decay; Distributed = gradual decay. Can be quantified through measures like the Gini coefficient of degree distribution. | Concentrated hubs provide centralised control points for rapid coordination (like in leaves for photosynthetic responses). Distributed hubs allow for more flexible, decentralised responses (like roots adapting to soil heterogeneity). | | Barabási and Albert (1999); (Clauset et al., 2009; Jeong et al., 2001) |
| Temporal Coherence | A measure of how consistently connected nodes in a network respond together over time. | Logic: Often calculated as correlation of responses between connected nodes across time points, or using metrics like Kendall's W for concordance. Can also be measured as the stability of correlation patterns over time. | High temporal coherence suggests coordinated responses within functional modules. Low coherence may indicate independent responses or system fragmentation under stress. Essential for understanding dynamic adaptation strategies. | | Fait et al. (2020) |
| Cross-tissue Correlation | The degree of similarity in metabolic responses between different tissues (e.g., leaf vs root) over time. | Logic: Correlation analysis (e.g., Spearman correlation) of metabolite profiles between tissues, often tracked over multiple time points. Can reveal coordination patterns and strategic decoupling. | High cross-tissue correlation suggests coordinated whole-plant responses. Strategic decoupling (decreasing correlation) may indicate tissue-specific adaptation strategies. | | Gargallo-Garriga et al. (2014) |

| Network Metric / Concept | Definition (General Network Science) / Underlying Logic | Conceptual Calculation / General Meaning/Interpretation | Biological Definition & Biological Interpretation | Relevant Reference(s) for |
|----------------------------|---|--|---|--|
| | | | | Critical for understanding organ-level coordination. |
| Module Preservation | The extent to which network modules (communities) maintain their structure across different conditions or time points. | Logic: Statistical measures comparing module membership between conditions, often using preservation statistics like those in WGCNA (Weighted Gene Co-expression Network Analysis). Involves permutation testing and standardised preservation scores. | High module preservation indicates stable functional organisation. Dynamic module reorganisation can reflect adaptive responses to environmental changes. Important for understanding functional stability under stress. | (Fortunato, 2010); Langfelder et al. (2011); Zhang and Horvath, 2005 |
| Network Motifs | Recurring patterns of interconnections that occur in networks at numbers that are significantly higher than those in randomised networks. | Logic: Identified by comparing subgraph frequencies in real networks vs. randomised null models. Common motifs include feedforward loops, feedback loops, and bi-fans. Statistical significance assessed through Z-scores. | Network motifs represent fundamental building blocks of biological networks and often correspond to basic regulatory or metabolic functions. They provide insights into evolutionary constraints and functional organisation. | Alon (2007); (Barabási and Oltvai, 2004) |

Table S2 Nutrition solution used in the glasshouse experiments.

| No | Nutrient | Concentration |
|----|--------------------------------------|---------------|
| 1 | MgSO ₄ | 4 mM |
| 2 | CaCl ₂ .2H ₂ O | 4 mM |
| 3* | KH ₂ PO ₄ pool | 3 mM |
| | K ₂ HPO ₄ pool | |
| 4 | Fe ⁺ (FeEDTA) | 400 µM |
| 5 | MnCl ₂ .4H ₂ O | 10 µM |
| | ZnSO ₄ .7H ₂ O | 10 µM |
| | CuSO ₄ | 2 µM |
| | H ₃ BO ₃ | 50 µM |
| | Na ₂ MoO ₄ | 0.2 µM |
| 6 | KNO ₃ | 1M |

*Pooled to make pH 6

Materials and Methods Details

Data Processing Parameters

Table S3. Genedata setting for quantitation

| Processing Step | Parameter | Positive Mode | Negative Mode |
|--------------------|------------------------|----------------------|----------------------|
| Data Import | RT Range | 0-16 minutes | 0-16 minutes |
| Chromatogram | | | |
| Chemical Noise | Smoothing RT Window | 3 scans | 3 scans |
| Subtraction | Subtraction RT Window | 51 scans | 51 scans |
| | Quantile | 50% | 50% |
| | Intensity Threshold | 8000, 10000 | 3000 |
| RT Alignment | Max RT Shift | 0.4 minutes | 0.2 minutes |
| | Gap Penalty | 1 | 1 |
| Peak Detection | Summation Window | 5 scans | 5 scans |
| | Minimum Peak Size | 3 scans | 4 scans |
| | Maximum Merge Distance | 5 points | 5 points |
| | Merge Strategy | Centers | Boundaries |
| | Smoothing | m/z window: 3 points | m/z window: 3 points |
| | Refinement Threshold | 5% | 5% |
| | Consistency Threshold | 0.6 | 0.6 |
| Isotope Clustering | RT Tolerance | 0.05 minutes | 0.05 minutes |
| | m/z Tolerance | 5 ppm | 5 ppm |
| | Ionization | Protonation | Deprotonation |
| | Charge Range | 1-2 | 1-5 |
| | Distance Measure | Log-Ratio | Log-Ratio |
| | Max Distance | 0.5 | 0.5 |

Detailed Data Preprocessing Workflow

Initial Data Cleaning

We first excluded columns containing fewer than three replicates for each molecular feature. Despite this reduction, missing values persisted, a common issue in LCMS studies (Karpievitch *et al.*, 2012) (Kokla *et al.*, 2019). We assessed the nature of the missing data using Little's MCAR

test and logistic regression to determine whether they were Missing Completely at Random (MCAR), Missing at Random (MAR), or Missing Not at Random (MNAR).

Missing Value Analysis and Imputation

We employed a diverse set of imputation techniques, including machine learning (Wei et al., 2018), neighbour-based methods, and matrix factorisation. In R, we used Random Forest (RF) (maxiter = 5, ntree = 50) and Predictive Mean Matching (PMM), while in Python, we applied Bayesian PCA, k-Nearest Neighbours (KNN), SVD, Gaussian Process Regression (GPR), and Expectation-Maximisation (EM). The performance of these methods was evaluated using Earth Mover's Distance (EMD) and Hellinger Distance, alongside visual comparisons via Q-Q plots, ECDF plots, and KDE plots. Additionally, we calculated richness, Shannon entropy, Simpson's diversity index, and sparsity for datasets imputed with median and RF methods.

Outlier Detection, Removal

Post-imputation, we numerically encoded categorical variables and standardised molecular features. Seven outlier detection techniques were applied: Mahalanobis Distance, Isolation Forest, Elliptic Envelope, Robust PCA, Local Outlier Factor, Z-Score, and IQR. We assessed the effectiveness of these methods by quantifying outliers and adjusting contamination thresholds, with visualisation via PCA and t-SNE plots. Outliers identified were removed, and resulting NaN values were imputed using the validated Random Forest method (missForest package). The 30 most affected variables were plotted with standard deviation error bars to highlight changes after outlier removal and imputation.

Data Transformation and Variable Selection

Following outlier removal, applied various data transformations: Log, Square Root, Box-Cox, Yeo-Johnson, Generalised Hyperbolic Sine (asinh) (using its standard mathematical definition, $\text{asinh}(x) = \ln(x + \sqrt{x^2 + 1})$), Generalised Logarithmic (glog), and Anscombe. We assessed their impact on data variability and distribution using Coefficient of Variation (CV), MA-transform (log ratio M), Relative Standard Deviation (RSD), and Relative Median Absolute Deviation (rMAD). Normality was evaluated using Shapiro-Wilk and Anderson-Darling tests, with density plots

visualising the distribution of test statistics and p-values for each variable and transformation method.

After the asinh transformation, we refined the dataset by excluding variables with an rMAD exceeding 30%, a threshold set to eliminate highly variable metabolites, and this clean data later used for subsequent analysis.

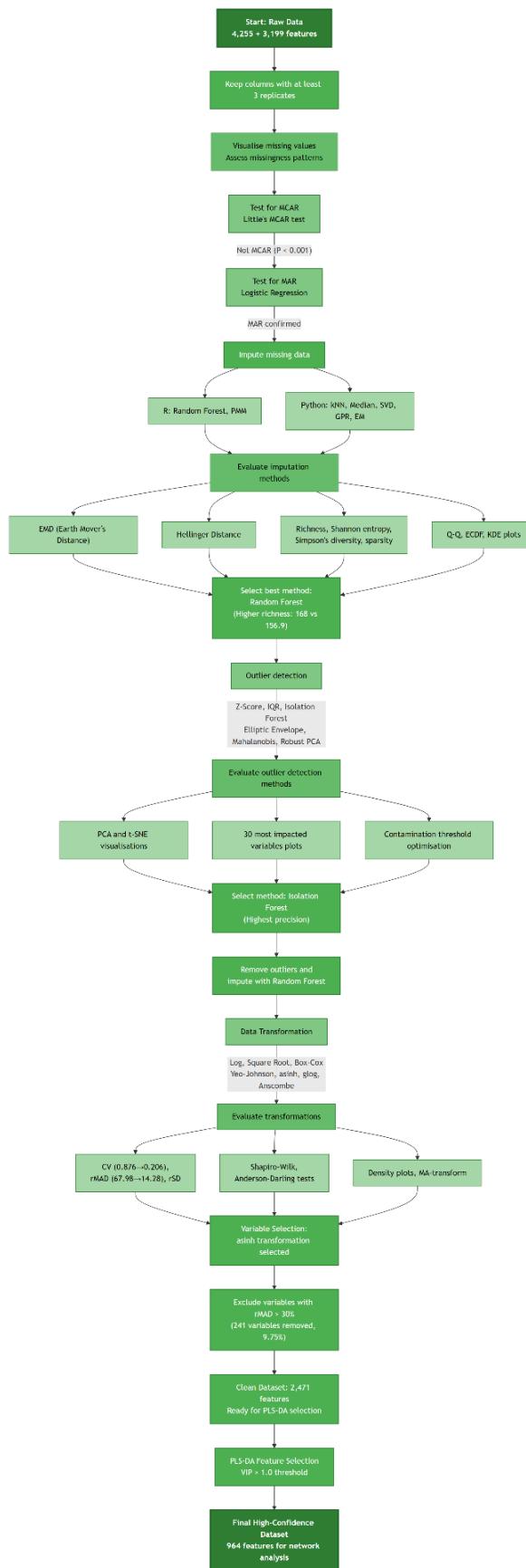


Fig. S1. Data preprocessing steps

Metabolite Annotation Methods

Database Matching (HMDB)

Untargeted metabolomics data was matched against the Human Metabolome Database (HMDB) using a custom Python script. The script first extracted the HMDB data from a compressed XML file, parsing key metabolite properties including chemical identifiers (InChI, InChIKey, SMILES), structural information (formula, super class, sub class), biological context (pathways, tissue locations), and database cross-references (KEGG, PubChem, Metlin IDs). For metabolite matching, the script prioritised matching by exact mass when available, with m/z as a fallback, applied a 10 ppm mass tolerance filter, selected the match with the smallest ppm difference for each cluster, and calculated quality metrics including m/z rank (normalised mass difference) and RT rank (normalised retention time difference). The results were exported as separate CSV files for negative and positive ionisation modes, containing comprehensive metabolite annotations and match quality assessments. The script handled missing values gracefully and included extensive error checking to ensure data integrity.

$$m/z \text{ Rank} = \frac{[\text{Experimental m/z} - \text{HMDB Exact mass}]}{\text{Experimental m/z}}$$

$$RT \text{ Rank} = \frac{[\text{Experimental RT} - \text{HMDB RT}]}{\text{Maximum RT in Dataset}}$$

GNPS-Based Annotation Enhancement

The metabolite matches were further annotated using the GNPS database API. The script queried GNPS using chemical identifiers (InChIKey, SMILES, or PubChem CID) to retrieve additional chemical classifications, pathways, and biospecimen data. The data was then enhanced with tissue-type annotations by mapping metabolite clusters to their corresponding root or leaf tissue origins. Finally, the script integrated statistical metrics from VIP and Mann-Whitney analyses, creating comprehensive annotation files for both ionisation modes. All

results were exported as CSV files with complete metabolite characterisation and tissue distribution information.

Structural Analysis using RDKit

The putatively identified metabolites underwent comprehensive structural characterisation using RDKit through two complementary approaches: structural classification and functional group analysis. In the classification phase, each metabolite's SMILES structure was evaluated against curated SMARTS patterns representing both broad biomolecule classes (like lipids and carbohydrates) and specific plant compounds (such as phytohormones). A confidence scoring system assessed matches based on molecular pattern complexity and atomic substructure matching, with weighted priorities resolving overlapping classifications. The functional group analysis then systematically identified 30 key structural features (including hydroxyls, carboxyls, and aromatic rings) using parallel processing for computational efficiency. The analyses enriched the metabolite annotations with structural classifications, quantitative confidence scores (0-1), and detailed functional group profiles.

Pathway Mapping

We performed metabolic pathway analysis of stress responses using non-parametric statistics. Metabolite changes between control and stress conditions were assessed using Mann-Whitney U tests, with false discovery rate (FDR) correction (Benjamini-Hochberg) applied within each tissue-genotype-time points-batch combination. Significant changes were defined by FDR < 0.05 and minimum three replicates per condition. Median fold changes and 95% confidence intervals were calculated using bootstrap resampling (1000 iterations). These significant metabolite changes were then mapped to KEGG pathways to assess pathway-level responses, considering both the number and magnitude of metabolite changes within each pathway. This approach provided a comprehensive view of metabolic pathway modulation under stress conditions.

Supplementary Results

Data Preprocessing Results

Initial Data Cleaning Outcomes

Initial data filtering reduced the number of molecular features from 4,255 to 1,789 in one dataset and from 3,199 to 1,350 in another, by excluding entries with fewer than three replicates. Little's MCAR test (P-value = 1.0) indicated that missing metabolite values were not Missing Completely at Random. Logistic regression revealed significant associations between missingness probability and observed variables (genotype, treatment group, time points, and replicate; Fig. S2A), suggesting a Missing at Random (MAR) scenario. Heatmaps (Fig. S2B-C) illustrate missing value patterns in leaf and root datasets for negative columns, with missing data in green and non-missing in grey.

Missing Value Patterns and Imputation Performance

The stacked bar plot of Earth Mover's Distance (Fig. S2D) indicated that median imputation performed well overall, closely followed by Random Forest (RF). Hellinger Distance (Fig. S2E) supported these findings. However, RF imputation outperformed median imputation in key metrics, with higher richness (168 vs. 156.9), Shannon entropy (5.124 vs. 4.999), and Simpson's diversity index (0.994 vs. 0.992). RF also demonstrated lower sparsity (0.00595 vs. 0.0204), preserving more variability.

Graphical comparisons, including Q-Q plots (Fig. S2F, left panel), ECDF, and KDE plots (Fig. S2F, middle and right panels), showed that RF imputation more closely aligned with the original data distribution. Based on these visual and statistical evaluations, RF was selected as the optimal imputation method.

Outlier Analysis Results

Isolation Forest emerged as the most effective outlier detection method, as visualised through PCA and t-SNE plots (Fig. S2g, h) and supported by the analysis of the 30 most impacted variables. This method identified outliers with the highest precision, leading to its selection for further analysis.

Transformation Effects

The asinh transformation most effectively reduced data variability, decreasing CV from 0.876 to 0.206 (Fig. S2I), rMAD from 67.98 to 14.28 (Fig. S2J), and RSD from 87.36% to 20.58% (Fig. S2K). MA-transformation plots confirmed tighter data clustering (Fig. S2L, M). Normality improved moderately across transformations, with asinh offering the best balance (Fig. S2N-Q). However, most variables remained non-parametric, guiding subsequent analysis choices.

To further refine the dataset, an rMAD-based selection process removed 241 variables (9.75%) with rMAD >30%, retaining 2,230 variables. This step enhanced reliability by excluding high-variability metabolites, providing a robust foundation for analysing metabolomic responses to osmotic stress.

(In negative mode, leaf data contains 807 molecular features, root data contains 982, with 391 features common to both, resulting in a combined total of 1,398 molecular features. In positive mode, leaf data contains 611 molecular features, root data contains 739, with 277 features common to both, resulting in a combined total of 1,073 molecular features. Overall, the combined dataset includes 2,471 molecular features.)

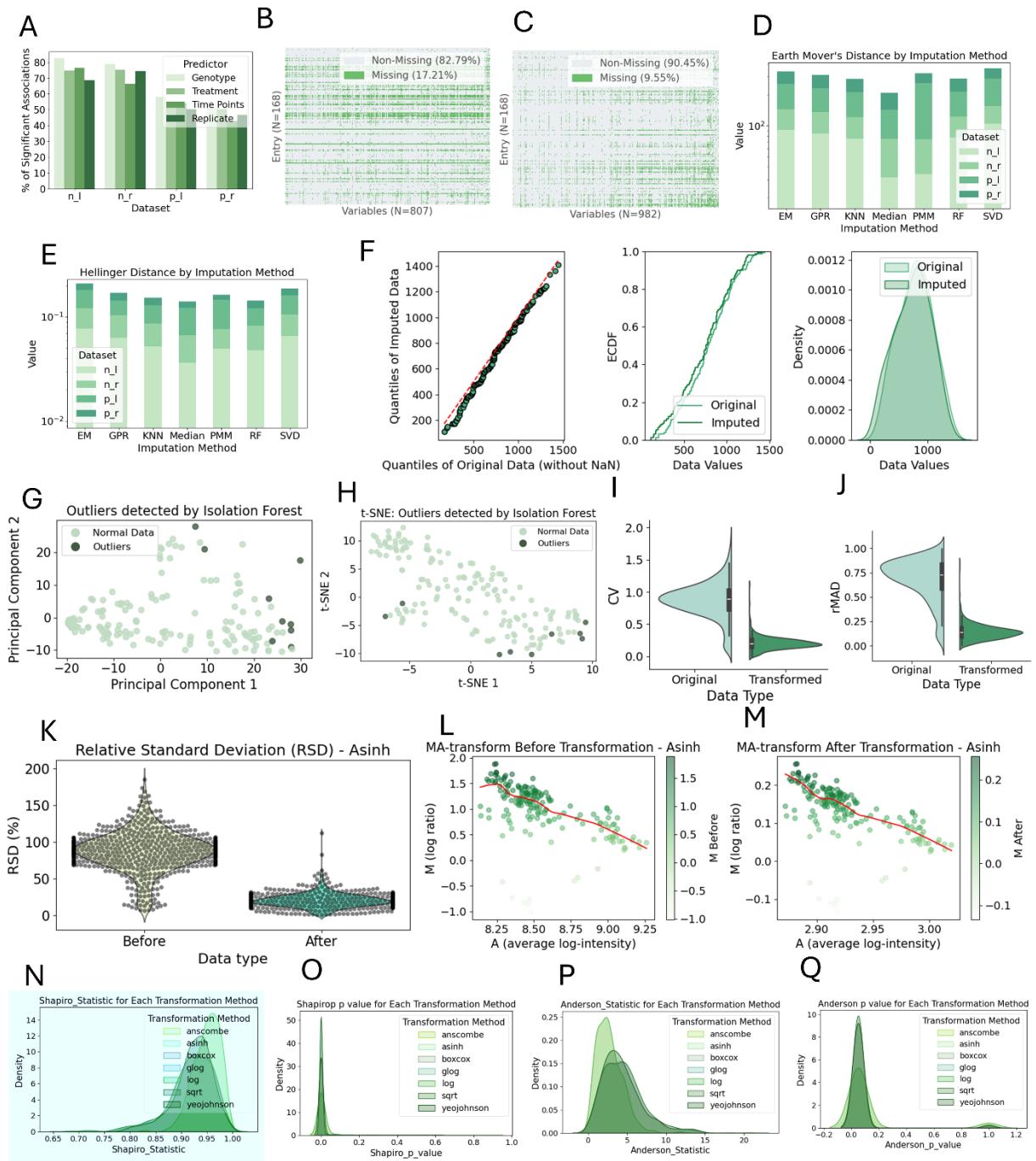


Fig. S2. Data Preprocessing and Evaluation

A, Logistic Regression for Missing Data Analysis. Bar chart illustrating significant predictors (genotype, treatment, time points, replicate) associated with missing data, supporting the MAR scenario across all datasets. B, Missing Value Patterns. Heatmaps depicting missing (green) and non-missing (grey) data distribution in negative mode leaf (n_l) B and root (n_r) C, datasets. D-E, Imputation Method Comparison. Stacked bar plots of Earth Mover's Distance D and Hellinger Distance B for various imputation methods across all datasets, highlighting the superior performance of RF and median imputation. F, Graphical Evaluation of Imputation Methods. Q-Q, ECDF, and KDE plots comparing original and RF-imputed data distributions across all datasets, showing that RF-imputed data closely aligns with the

original distribution. G-H, PCA and t-SNE plots of the n_l dataset. These plots show outliers (dark points) at the periphery of clustered normal data (light green), demonstrating Isolation Forest's effectiveness in identifying anomalies across different dimensions. I-J, Data Transformation Effects. Density plots showing the impact of asinh transformation on Coefficient of Variation (CV) I and Relative Median Absolute Deviation (rMAD) J. K Relative Standard Deviation (RSD) Change. Violin plot demonstrating RSD reduction in the negative mode leaf dataset (n_l). L-M, MA-Transform Plots. Scatter plots illustrating the distribution of the n_l dataset before L and after M asinh transformation. N-Q, Normality Assessment. Density plots of Shapiro-Wilk N,O and Anderson-Darling P,Q test statistics and p-values for various transformation methods on the n_l dataset.

Metabolite Analysis Results

Tissue-specific Temporal Adjustments

Temporal analysis revealed distinct response patterns between leaf and root tissues under osmotic stress. Leaf tissue exhibited rapid metabolic adjustments at the initial timepoint, while root tissue showed a delayed response peaking at the second timepoint. This tissue-specific temporal divergence was evident across multiple chemical classes and functional groups, with leaves showing coordinated early responses and roots displaying more gradual adaptation patterns. Notably, thiol, a key osmotic stress marker, was exclusively detected in root tissue. These tissue-specific metabolic dynamics are comprehensively visualised in Fig S3-S6, which detail the temporal progression of chemical groups, functional responses, and pathway-level changes across tissues and genotypes. The figures illustrate how these coordinated metabolic adjustments contribute to the distinct network architectures observed in the main manuscript, highlighting the sophisticated regulation of both broad chemical classes and specific metabolic pathways during stress adaptation.

Supplementary Figures

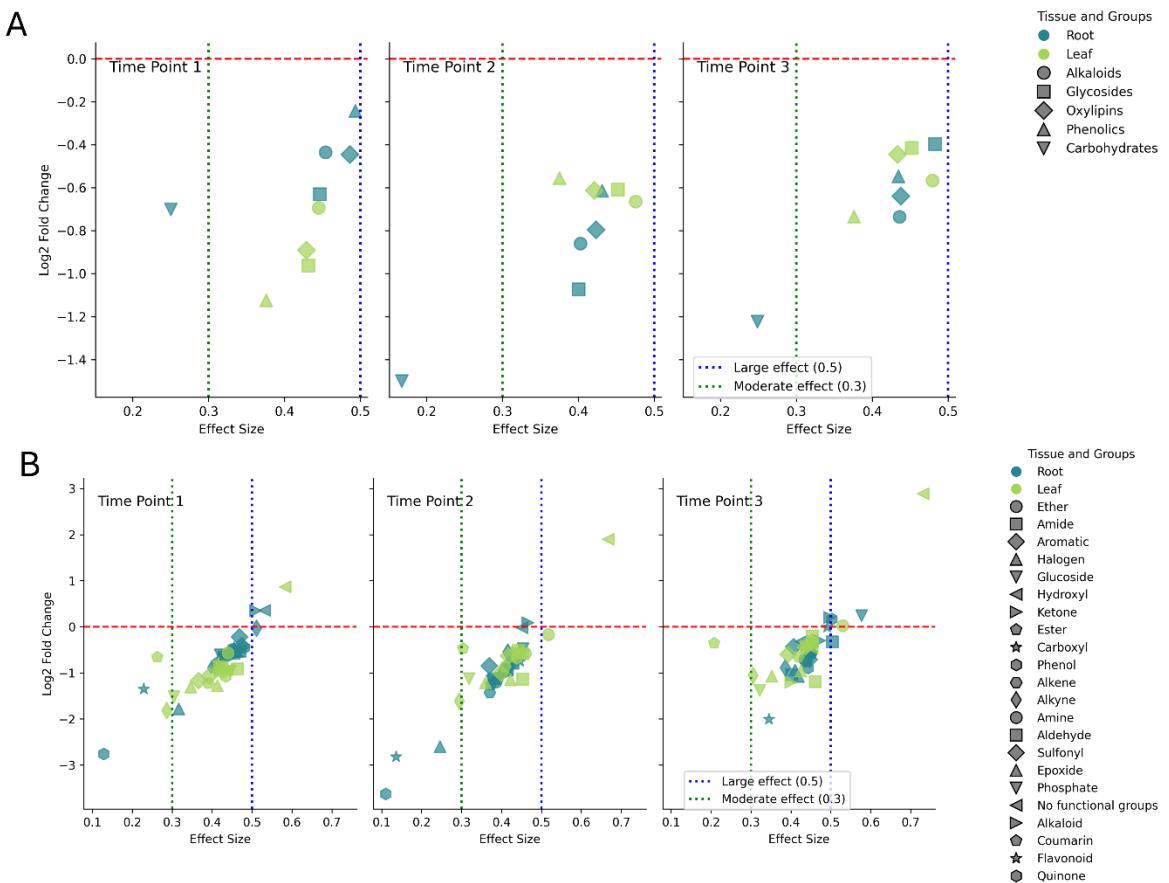


Fig. S3. Temporal dynamics of chemical and functional group responses to osmotic stress across tissues.

The temporal progression reveals tissue-specific patterns of chemical (A) and functional (B) groups. In chemical groups (A), leaf tissue demonstrates a rapid response with maximum decreases in metabolites at Time Point 1, while root tissue shows delayed responses peaking at Time Point 2. Most functional groups (B) mirror this tissue-specific temporal pattern, with leaves showing early responses (Time Point 1) and roots exhibiting delayed but sustained modulation across Time Points 2 and 3. Effect sizes are demarcated by vertical dotted lines indicating moderate (0.3, green) and large (0.5, blue) effects. Log₂ fold changes quantify the magnitude and direction of metabolic adjustments relative to control conditions. This figure captures the distinct temporal orchestration of metabolic responses between tissues under osmotic stress, highlighting the sophisticated regulation of both broad chemical classes and specific functional groups during stress adaptation.

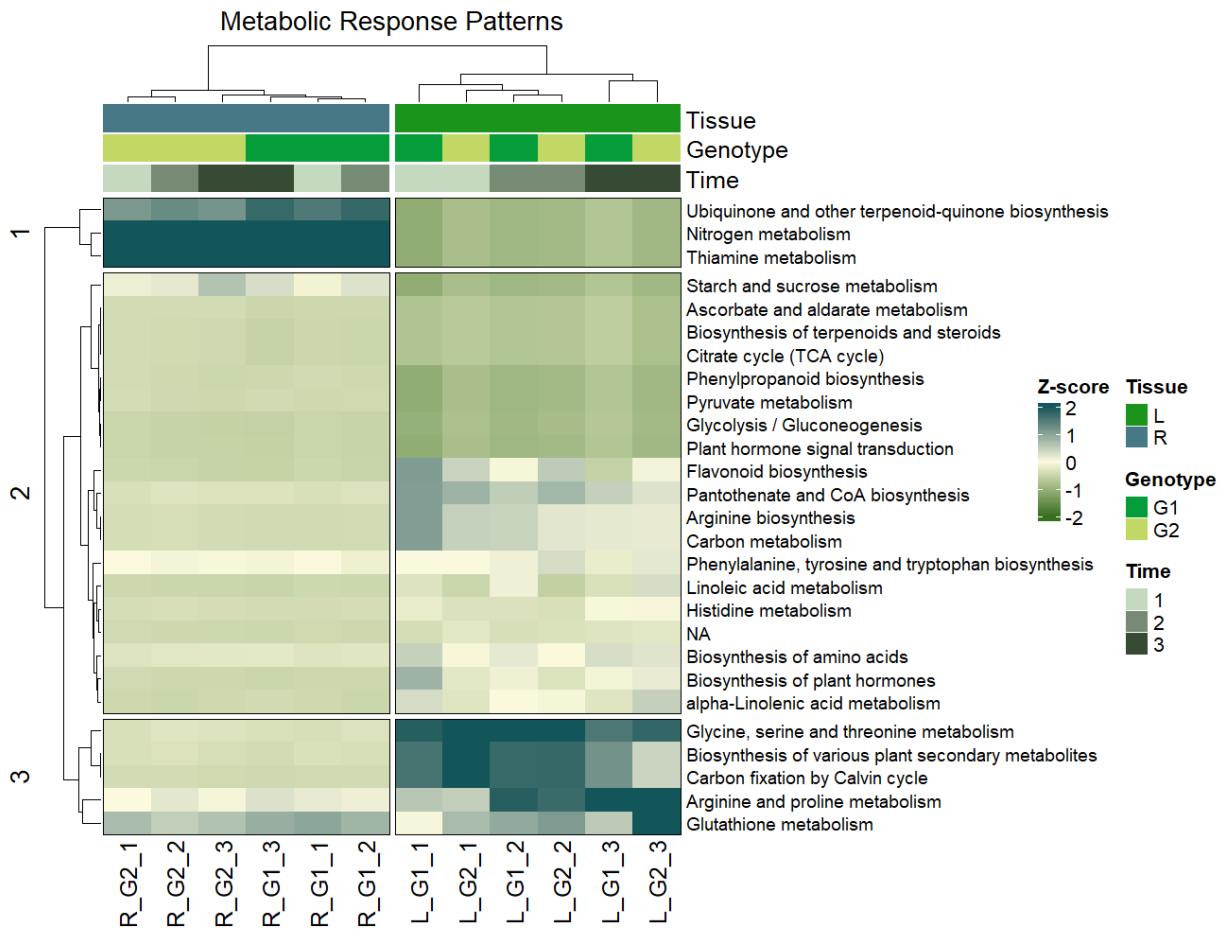


Fig. S4. Tissue and Genotype-Specific Metabolic Pathway Responses under Osmotic Stress in Wheat.

This hierarchically clustered heatmap shows Z-score transformed metabolic pathway responses (blue: upregulation, green: downregulation) in leaf (L) and root (R) tissues of drought-tolerant (G1) and drought-susceptible (G2) genotypes across three time points under osmotic stress. Three clusters emerge: (1) downregulated primary metabolism (e.g., ubiquinone biosynthesis), (2) moderately modulated central carbon and amino acid pathways (e.g., TCA cycle), and (3) differential regulation of stress-associated pathways (e.g., glutathione metabolism, flavonoid biosynthesis). The clustering patterns suggest distinct metabolic responses between tissues and genotypes during osmotic stress adaptation.

Temporal Metabolic Response Patterns
Top Responsive Pathways

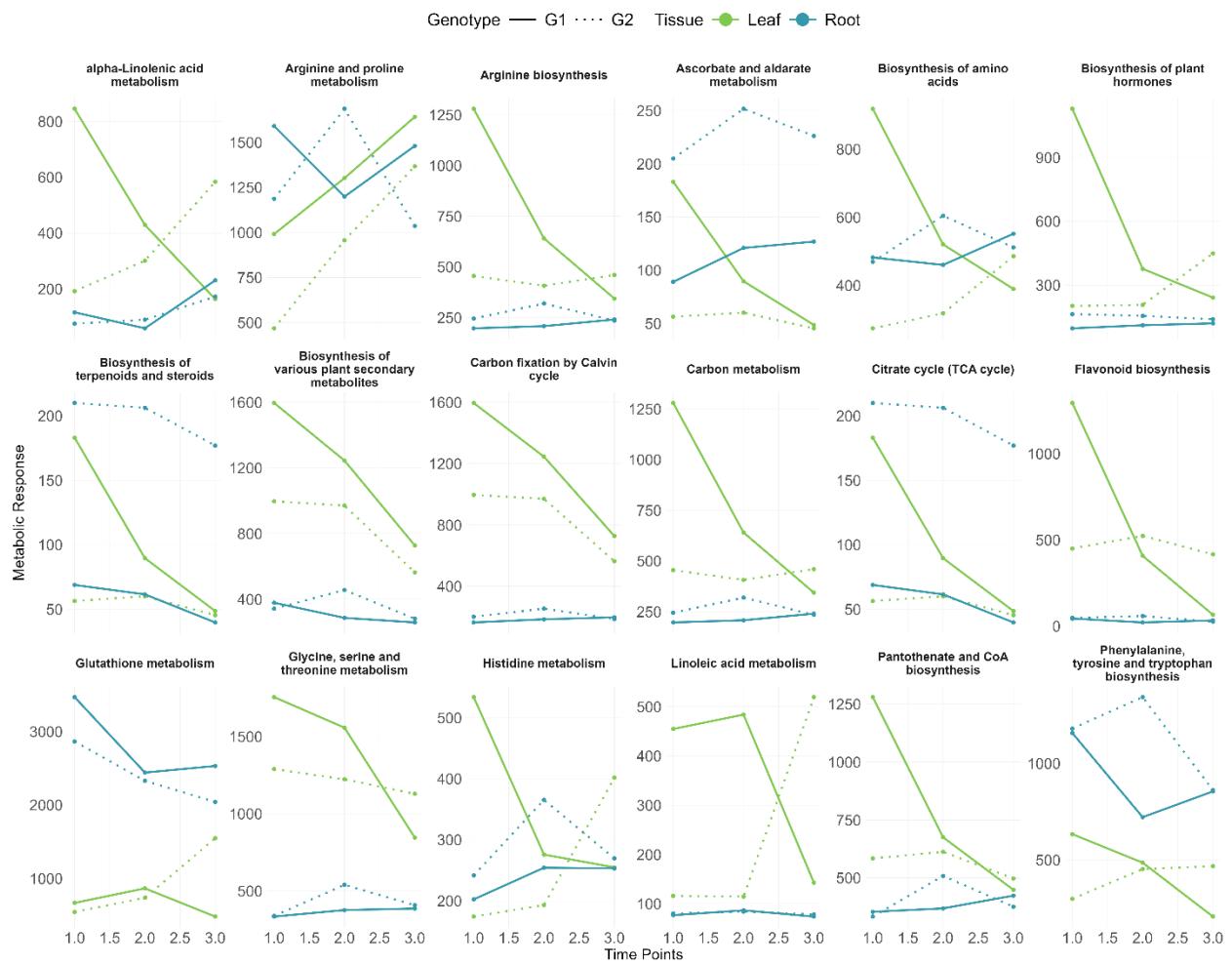


Fig. S5. Temporal Dynamics of Major Metabolic Pathways Under Osmotic Stress

Each subplot illustrates the temporal dynamics of putatively identified metabolic pathways in leaf and root tissues under osmotic stress conditions, differentiated by genotype (G1, G2) and tissue type (leaf, root). The visualisation captures tissue-specific response patterns, with leaves typically showing pronounced metabolic adjustments while roots display more gradual changes over time. This aligns with the observed tissue-specific network architecture differences, particularly in G1, where leaves demonstrate coordinated responses across primary metabolism pathways (amino acid biosynthesis, carbon fixation), while roots exhibit modulated changes in stress-responsive pathways (e.g., glutathione metabolism). These pathway-level responses provide additional context for understanding the tissue-specific molecular strategies employed under osmotic stress conditions.

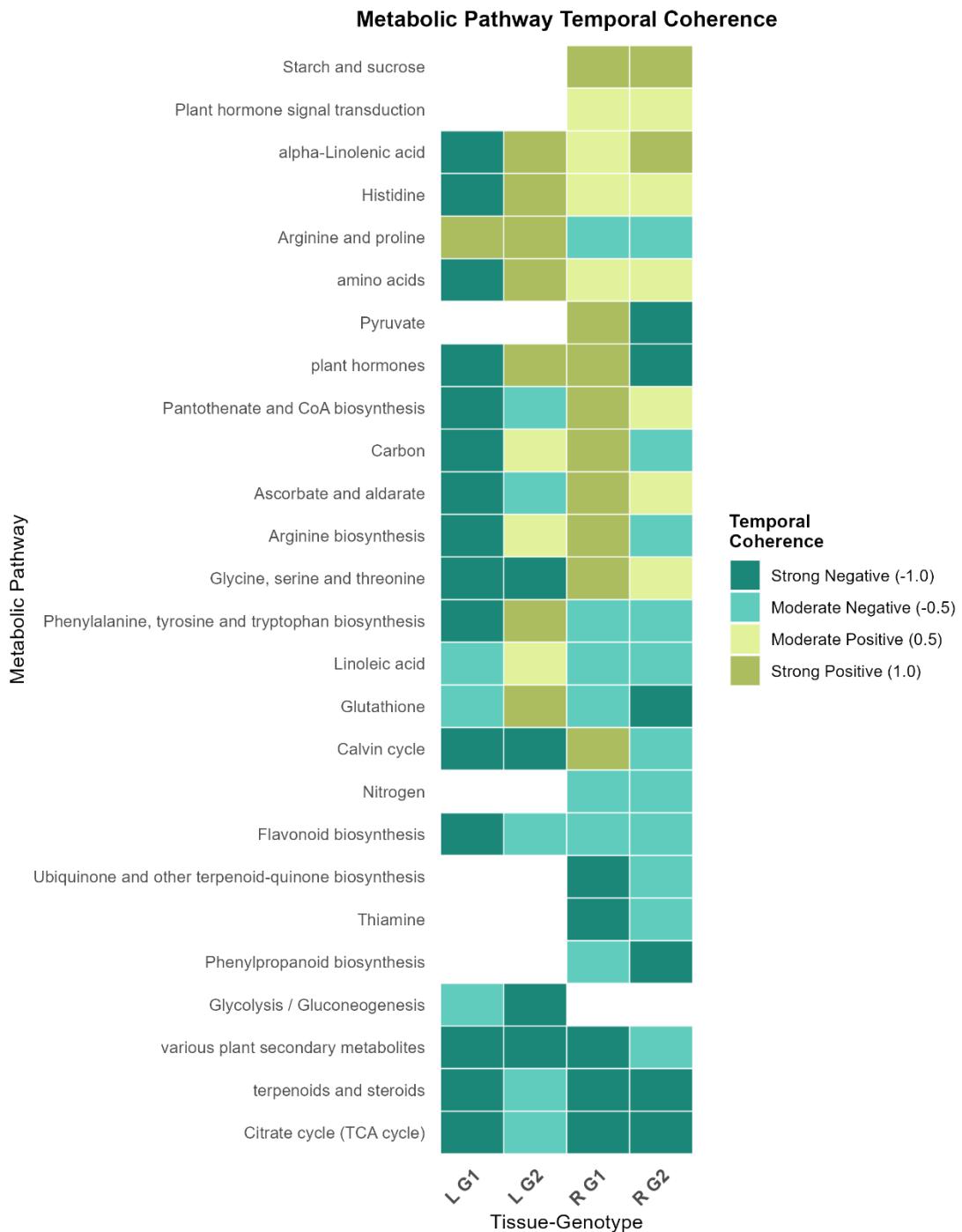


Fig. S6. Tissue-Specific Temporal Coherence of Metabolic Pathways Under Osmotic Stress

Heatmap visualisation reveals the temporal coherence patterns of putatively identified metabolic pathways in leaf (L) and root (R) tissues of drought-tolerant (G1) and drought-susceptible (G2) wheat genotypes under osmotic stress. Temporal coherence scores (-1.0 to +1.0) indicate the consistency and direction of metabolic responses over time, where strong positive values (light green) represent sustained increases and strong negative values (dark green) indicate consistent decreases. The analysis captures tissue-specific regulation of multiple pathways, including

primary metabolism (amino acids, carbon metabolism) and stress-responsive processes (glutathione, flavonoid biosynthesis). These pathway-level temporal coherence patterns provide additional context for understanding the metabolic responses under osmotic stress conditions.

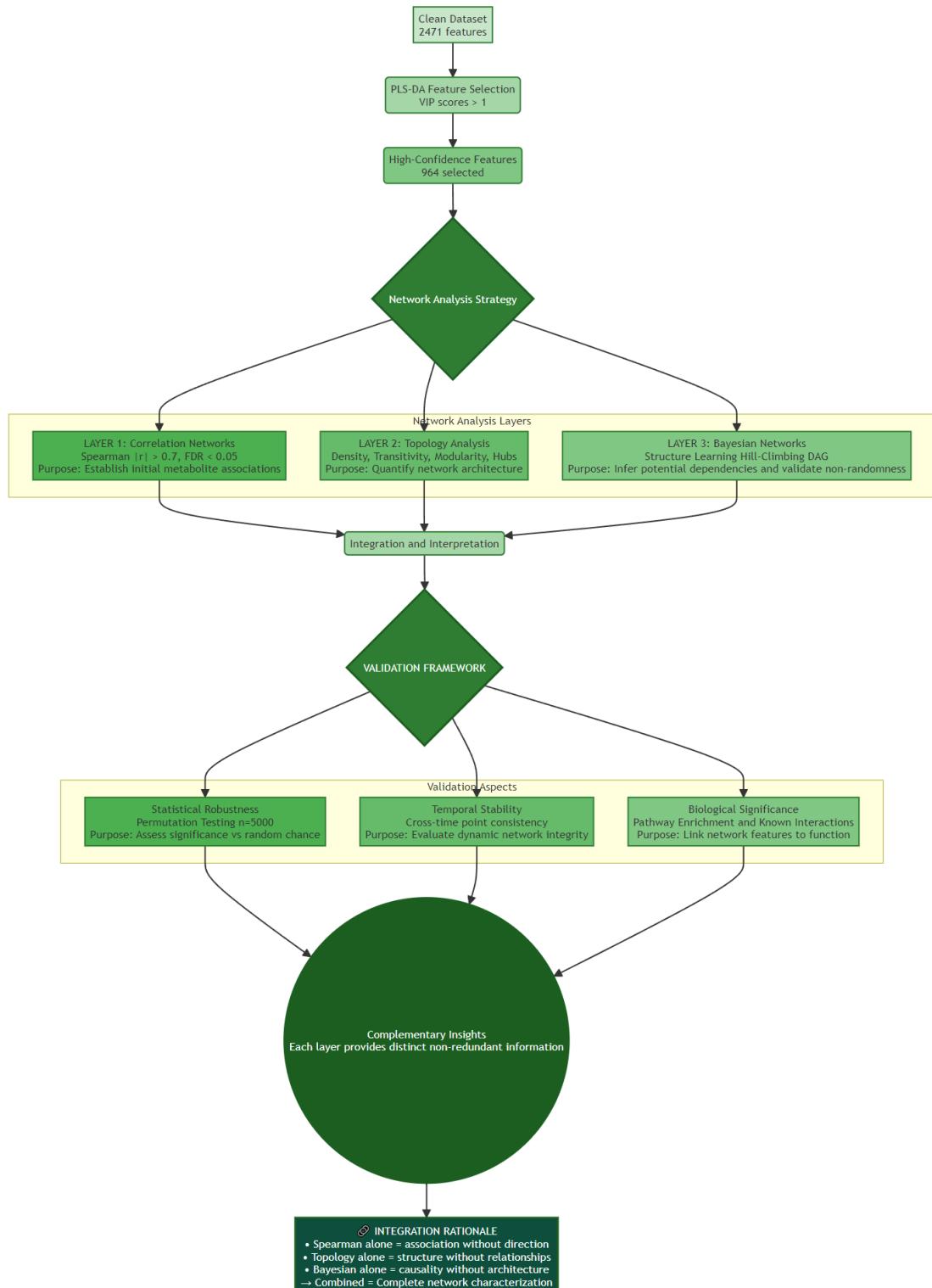


Fig. S7. Multi-layered analytical framework for metabolic network analysis.

Each layer addresses distinct biological questions using complementary statistical approaches. Layer 1 identifies co-regulated metabolite pairs, Layer 2 quantifies architectural principles distinguishing tissue-specific organisation strategies, and Layer 3 validates directional relationships. Statistical robustness is ensured through bootstrap resampling ($n=5,000$) and permutation testing with sample sizes optimised for computational tractability ($n=1,000$ - $10,000$). All analyses achieved $P < 0.001$ significance thresholds.

Appendix 3: Supplementary Material of Chapter 6

Supplementary Information

Guide to Supplementary Information

This document contains supplementary figures, tables, and notes that provide detailed support for the main findings of the manuscript. The materials are organised as follows:

Supplementary Figures S1-S5: These figures provide visualisations of the biological discoveries, including the integrated multi-omics network (Fig. S1), analyses of cross-modal attention dynamics and spectral drivers (Fig. S2), stereotyped temporal patterns (Fig. S3), model performance metrics (Fig. S4), and the identification of robust early biomarkers (Fig. S5).

Supplementary Figures S6-S8: These figures document the quality control and preprocessing pipelines for the untargeted LC-MS (Figs. S6-S7) and hyperspectral reflectance data (Fig. S8), establishing the integrity of the input datasets.

Supplementary Figures S9-S13: This section provides the validation of the data augmentation strategy. It includes a workflow diagram (Fig. S9) and comparisons confirming that the augmented data preserves spectral signatures (Fig. S10), molecular profiles (Fig. S11), and overall data quality, structure, and biological signals (Figs. S12-S13).

Supplementary Tables S1-S9: These tables provide the quantitative data and statistical results that underpin key analyses. This includes lists of top attention pairs (S1), statistical comparisons of network metrics (S2), predictive model performance scores (S3), ranked lists of predictive features from SHAP analysis (S4), details of the MOFA+/SHAP feature overlap (S5), experimental parameters (S6-S7), a list of flagged spectral outliers (S8), and a quantitative comparison of network coordination strength (S9).

Supplementary Note 1: This note contains a detailed description of the experimental and computational methods, including specific protocols for spectral measurements, the LC-MS data preprocessing workflow, and an outline of the data augmentation and validation framework.

Code and Data Availability: All code used for data preprocessing, the MOFA+ Transformer analysis, and figure generation is available in the associated GitHub repository. The processed and augmented datasets used for all analyses are also provided to ensure full reproducibility.

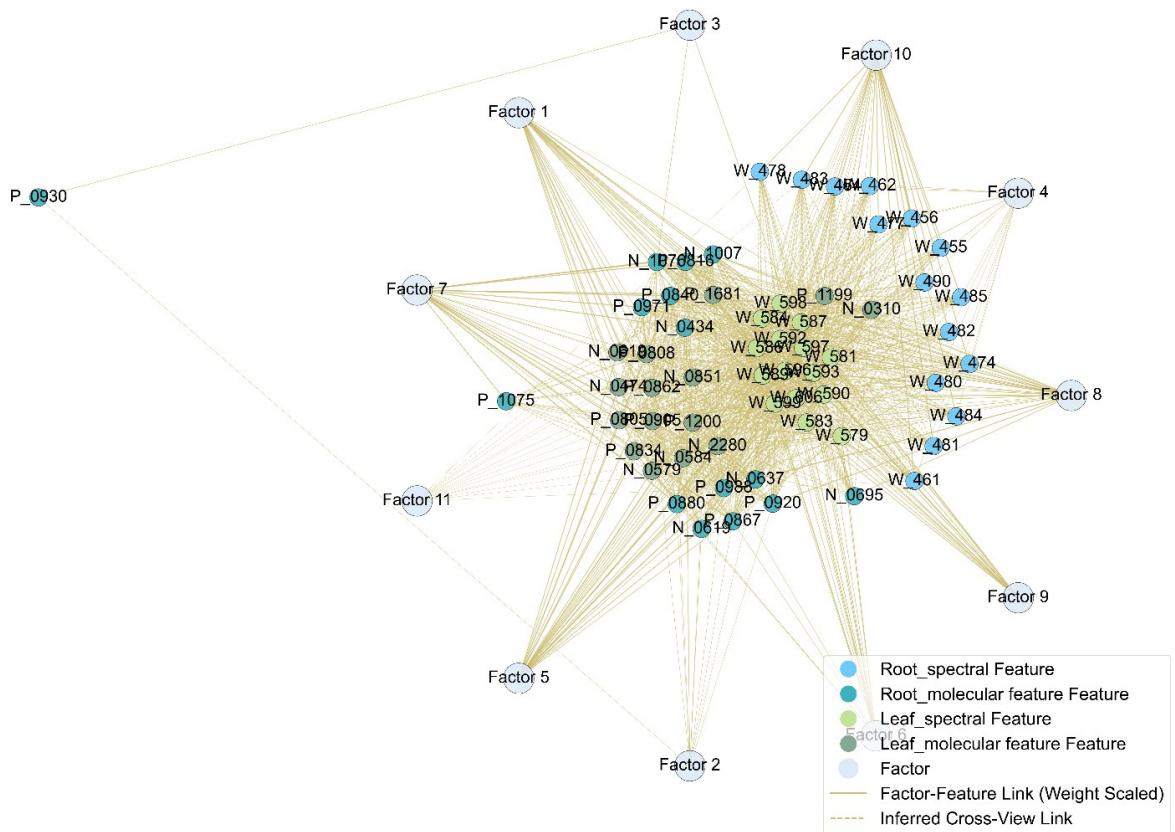


Fig. S1. Cross-View Feature Integration Network (Top Features). This network visualises relationships between selected top features (coloured nodes: Root Spectral Feature, light blue; Root Molecular Feature, dark cyan; Leaf Spectral Feature, light green; Leaf Molecular Feature, dark green) and active latent factors (light grey-blue nodes) from MOFA+. Solid gold lines represent MOFA+ feature weights (loadings), linking features to factors. Dashed gold lines infer cross-view links between features from different modalities that load onto the same factor. Clustering highlights MOFA+'s grouping of features based on shared variance (e.g., genotype or stress response patterns), illustrating the integrated data structure underpinning downstream analyses.

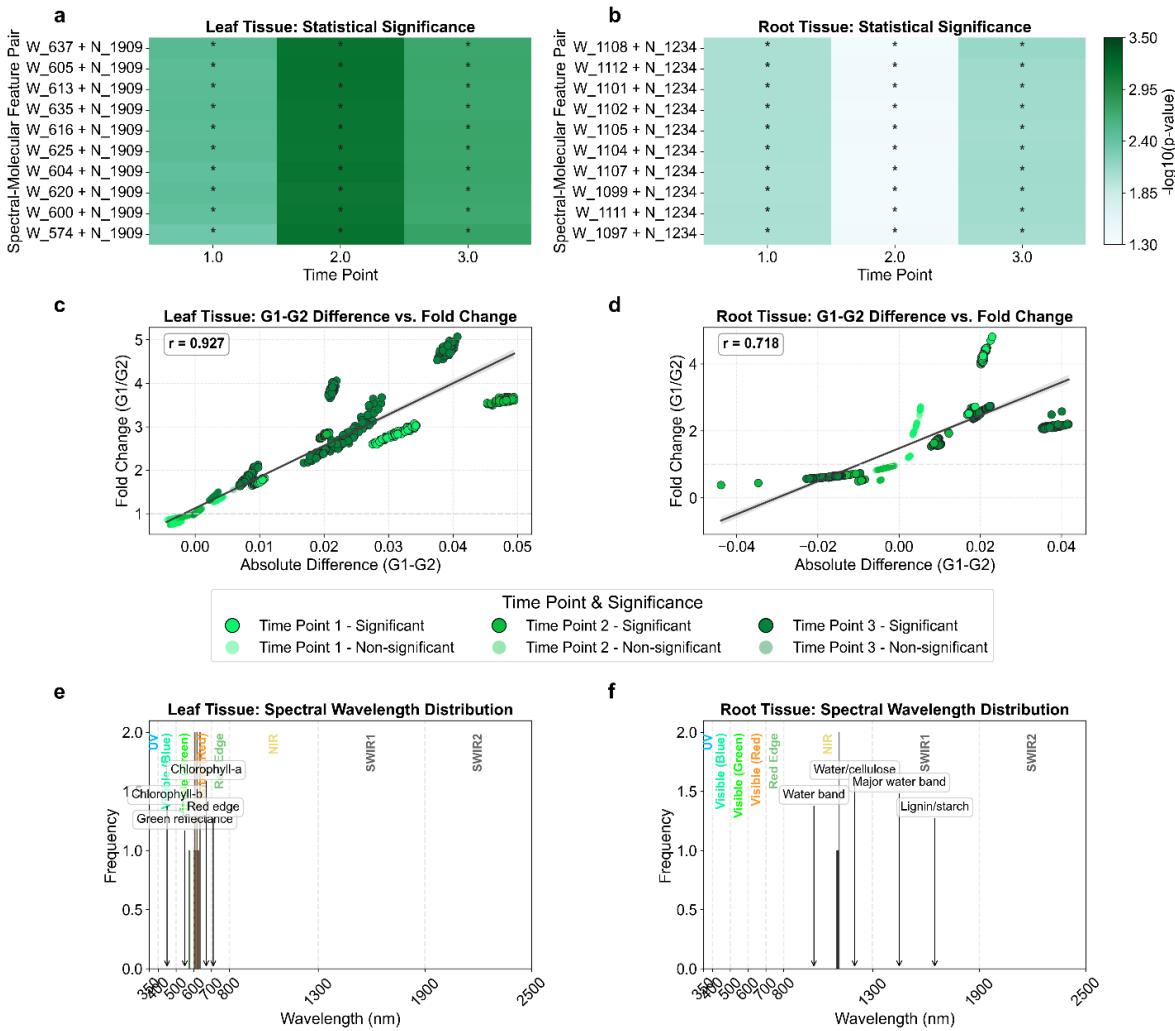


Fig. S2. Genotype-Specific Cross-Modal Attention Differences, Temporal Dynamics, and Spectral Enrichment. Statistical comparison of Spectral-to-Molecular (S2M) attention between tolerant (G1) and susceptible (G2) genotypes under stress reveals differences in interaction strength, timing, and the spectral features involved. (a, b) Heatmaps display the statistical significance (-log₁₀ p-value) of the attention difference (G1 vs. G2) for top-ranked S2M pairs over the three defined stress phases. The prevalence of highly significant differences (darker green cells) and the trend of increasing significance over time highlights a robust and escalating differential response. (c, d) Scatter plots visualise the magnitude of these differences. The y-axis shows the Fold Change (G1/G2) in attention, while the x-axis shows the Absolute Difference. The strong positive correlation reveals a highly structured relationship between the magnitude and relative change in attention. (e, f) Histograms reveal the spectral origins of these important interactions. In leaf tissue (e), attention is predominantly linked to spectral features in the visible spectrum (~500-700 nm), implicating changes in pigment content and photosynthetic status. In

root tissue (f), interactions are dominated by the Near-Infrared (NIR) region (~900-1300 nm), suggesting links to water status and tissue structure.

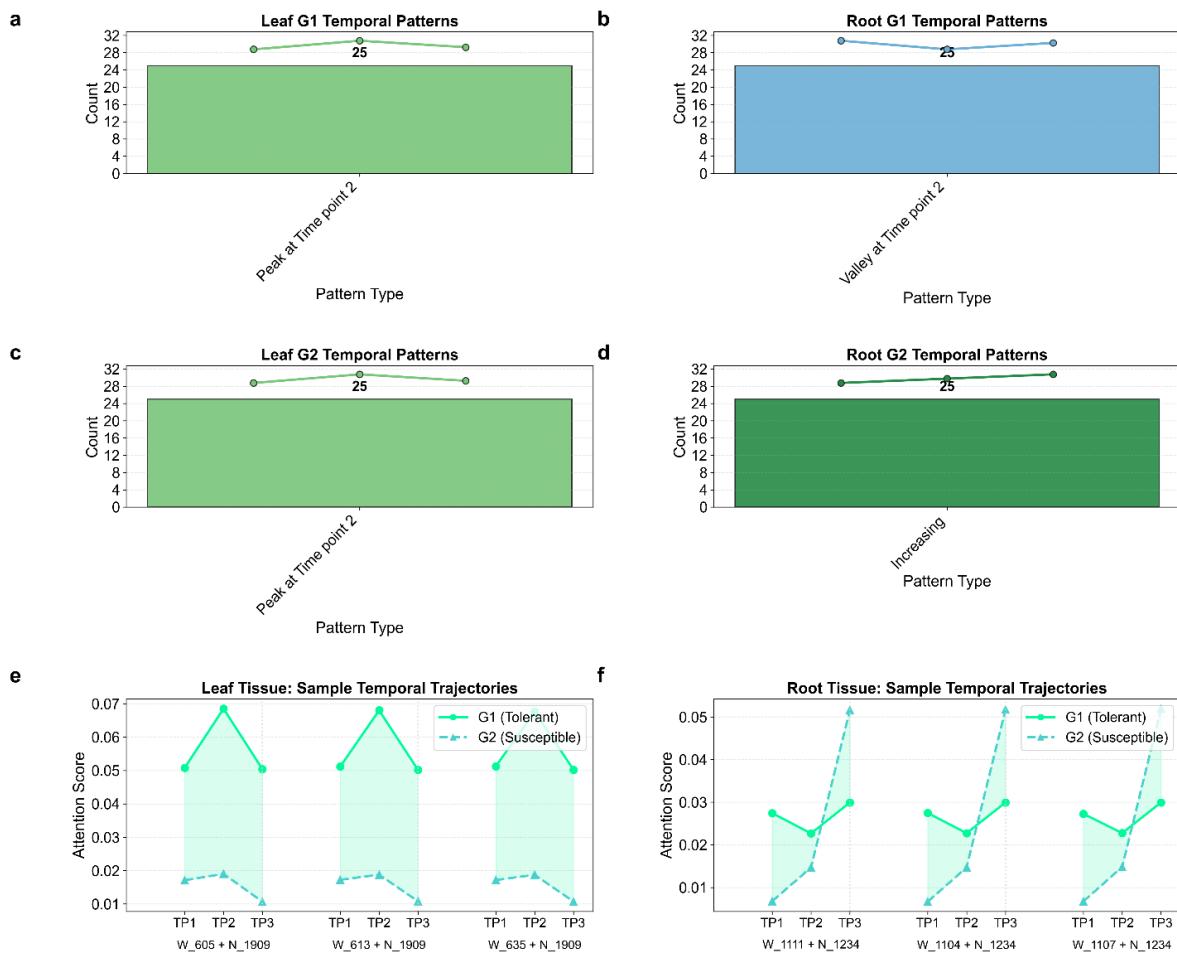


Fig. S3. Genotype-Specific Temporal Patterns in Cross-Modal Attention Reveal a Stereotyped Stress Response. The analysis of S2M attention dynamics over three time points reveals highly consistent, tissue-specific strategies. (a-d) Bar charts summarise the prevalent temporal trajectories for the top 25 S→M pairs. The results show a strikingly uniform temporal signature for each condition, indicating a programmed rather than varied response. Leaf tissue (a, c) is overwhelmingly characterised by a 'Peak at Time point 2' pattern. In root tissue, the tolerant genotype (G1) is dominated by a 'Valley at Time point 2' pattern (b), while the susceptible genotype (G2) shows a consistently 'Increasing' response (d). (e, f) Representative trajectories for the top S→M pairs visualise these dominant patterns, directly comparing G1 (solid line) and G2 (dashed line) in (e) Leaf and (f) Root tissues and illustrating the consistent dynamics within each group.

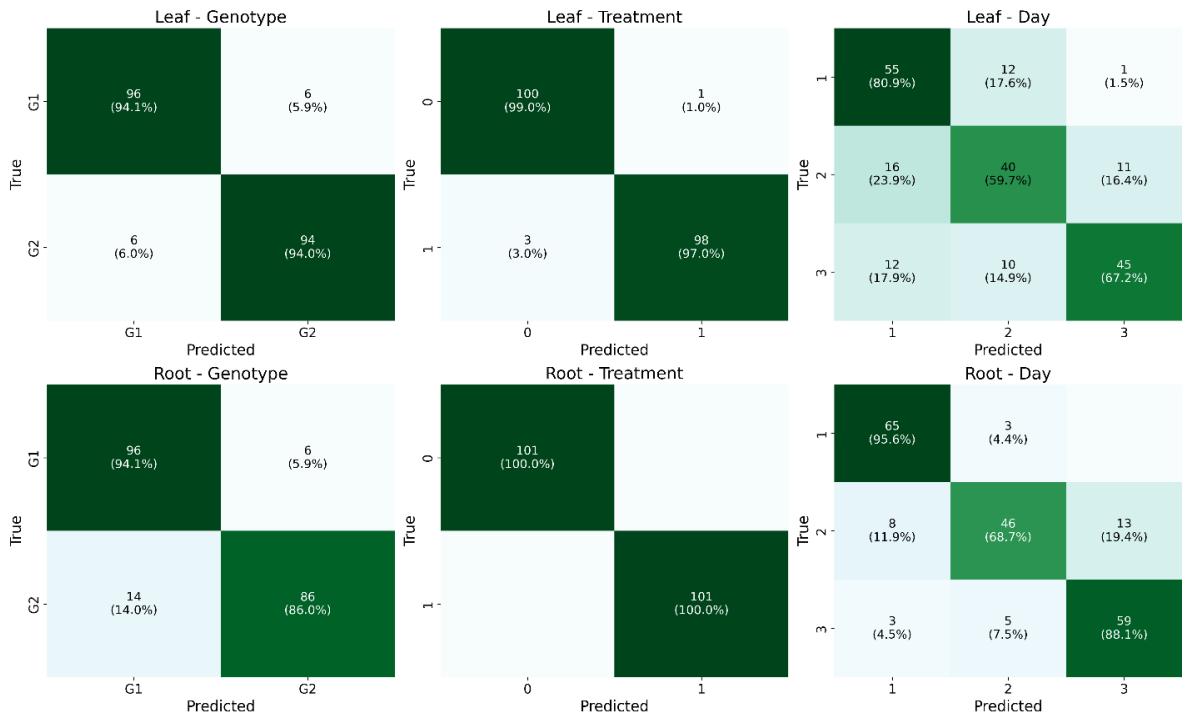


Fig. S4. Confusion Matrices for Classification Tasks by Tissue. Transformer model classification performance on the test set for Genotype, Treatment, and Time Point prediction tasks, shown for Leaf (top) and Root (bottom). Cells display normalised sample counts and percentages (True class vs. Predicted class). The confusion matrices for the test set show high classification accuracy, particularly for high accuracy, supporting metrics in Fig. 5A and Table S3. Matrices reveal patterns like higher G2 misclassification in Root Genotype task and confusion between adjacent time points (e.g., TP1 vs. TP2) in Time Point prediction.

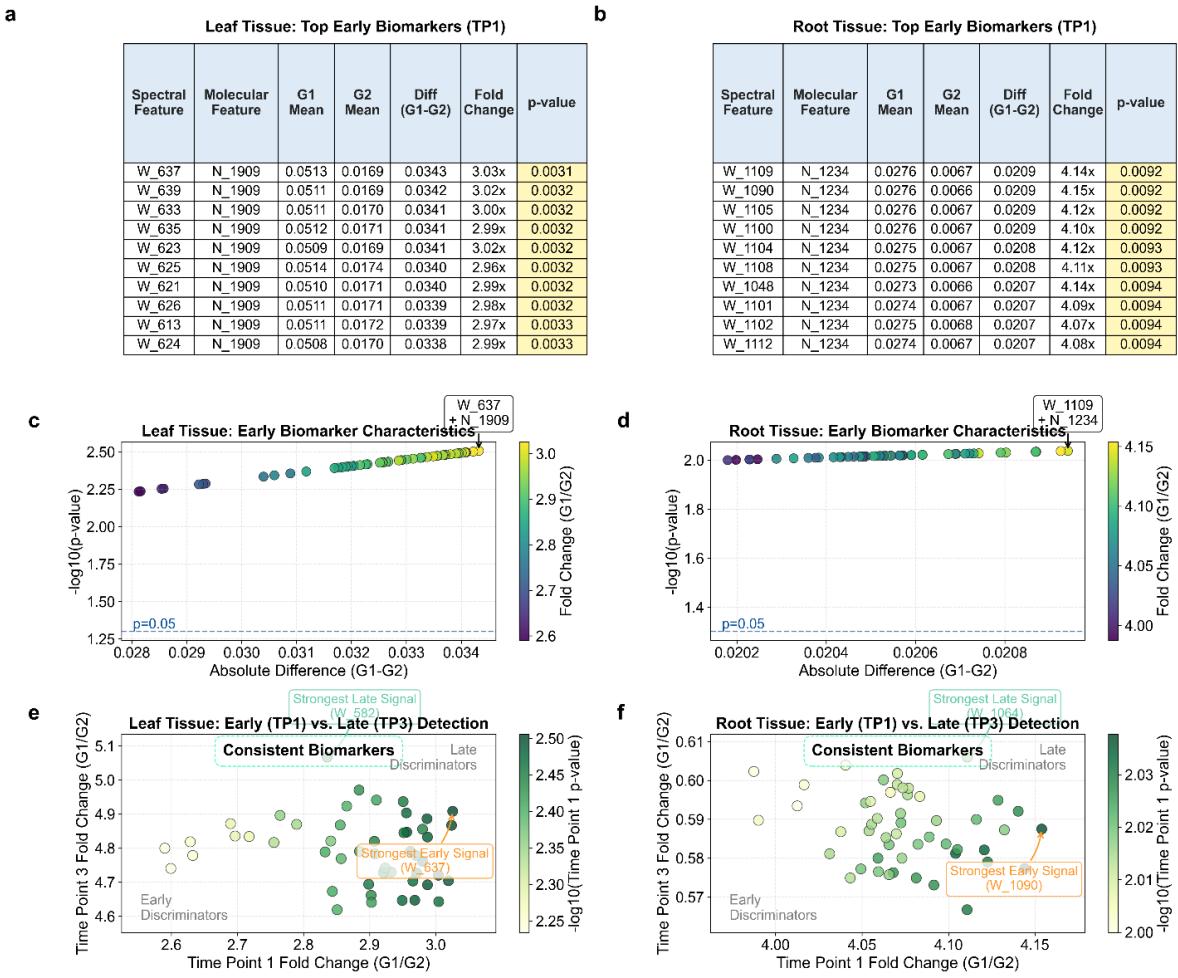


Fig. S5. Identification and Characterisation of Robust Early Biomarkers. The framework successfully identifies S \rightarrow M pairs with highly significant attention differences at the initial stress time point (TP1). (a, b) Tables of the top 10 early biomarkers for Leaf and Root tissues, ranked by statistical significance, show consistently low p-values and high fold changes. (c, d) Volcano-style plots visualise these characteristics, confirming that all top biomarkers are well above the significance threshold ($p=0.05$). (e, f) A comparison of discriminatory power (Fold Change G1/G2) at TP1 versus TP3 reveals the temporal stability of these biomarkers. The tight clustering of points confirms that the vast majority of early biomarkers are also 'Consistent Biomarkers', maintaining their high discriminatory power into the late stress phase. The zoomed-in view highlights the specific features that perform strongest as early versus late signals within this highly stable group.

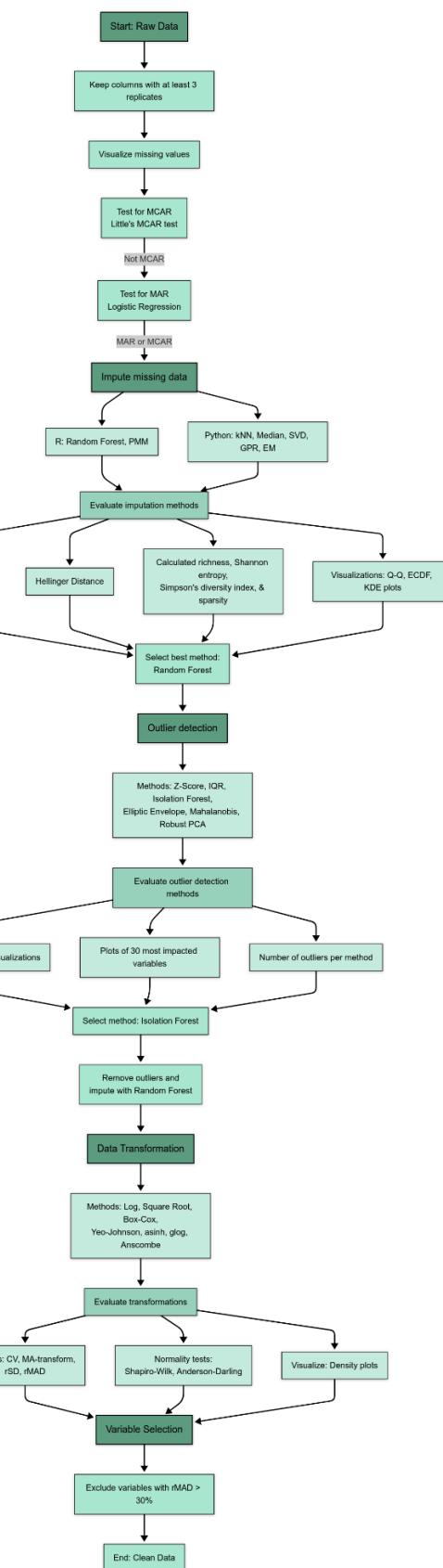


Fig. S6. LCMS Data Preprocessing Workflow

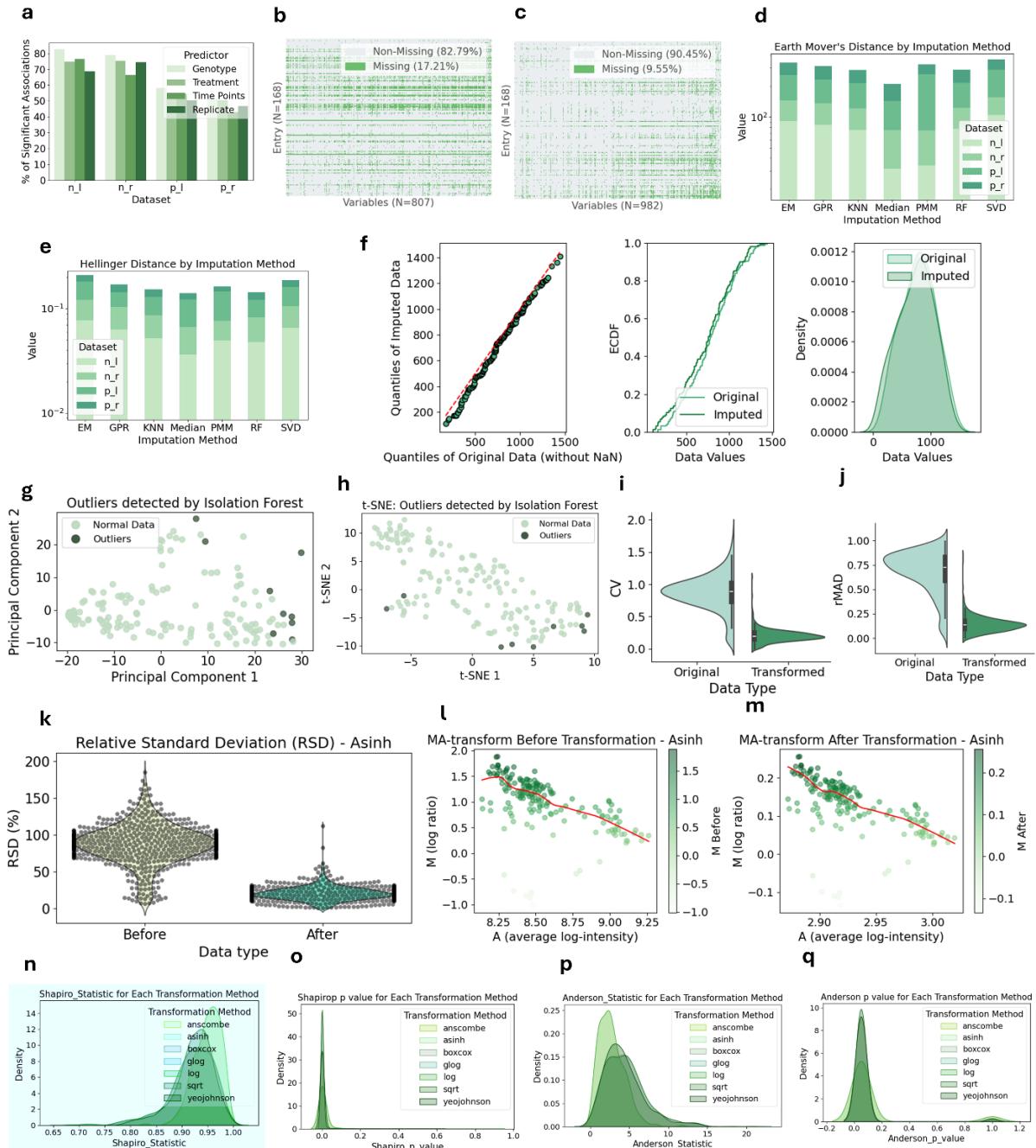


Fig. S7. LCMS Data Preprocessing and Evaluation Plots. (A) Logistic Regression for Missing Data Analysis. Bar chart illustrating significant predictors (genotype, treatment, Time Point, replicate) associated with missing data, supporting the MAR scenario across all datasets. (B-C) Missing Value Patterns. Heatmaps depicting missing (green) and non-missing (grey) data distribution in negative mode leaf (*n_l*) (B) and root (*n_r*) (C) datasets. (D-E) Imputation Method Comparison. Stacked bar plots of Earth Mover's Distance (D) and Hellinger Distance (E) for various imputation methods across all datasets, highlighting the superior performance of RF and median imputation. (F) Graphical

Evaluation of Imputation Methods. Q-Q, ECDF, and KDE plots comparing original and RF-imputed data distributions across all datasets, showing that RF-imputed data closely aligns with the original distribution. (G-H) PCA and t-SNE plots of the *n_l* dataset. These plots show outliers (dark points) at the periphery of clustered normal data (light green), demonstrating Isolation Forest's effectiveness in identifying anomalies across different dimensions. (I-J) Data Transformation Effects. Density plots showing the impact of asinh transformation on Coefficient of Variation (CV) (I) and Relative Median Absolute Deviation (rMAD) (J). (K) Relative Standard Deviation (RSD) Change. Violin plot demonstrating RSD reduction in the negative mode leaf dataset (*n_l*). (L-M) MA-Transform Plots. Scatter plots illustrating the distribution of the *n_l* dataset before (L) and after (M) asinh transformation. (N-Q) Normality Assessment. Density plots of Shapiro-Wilk (N, O) and Anderson-Darling (P, Q) test statistics and p-values for various transformation methods on the *n_l* dataset.

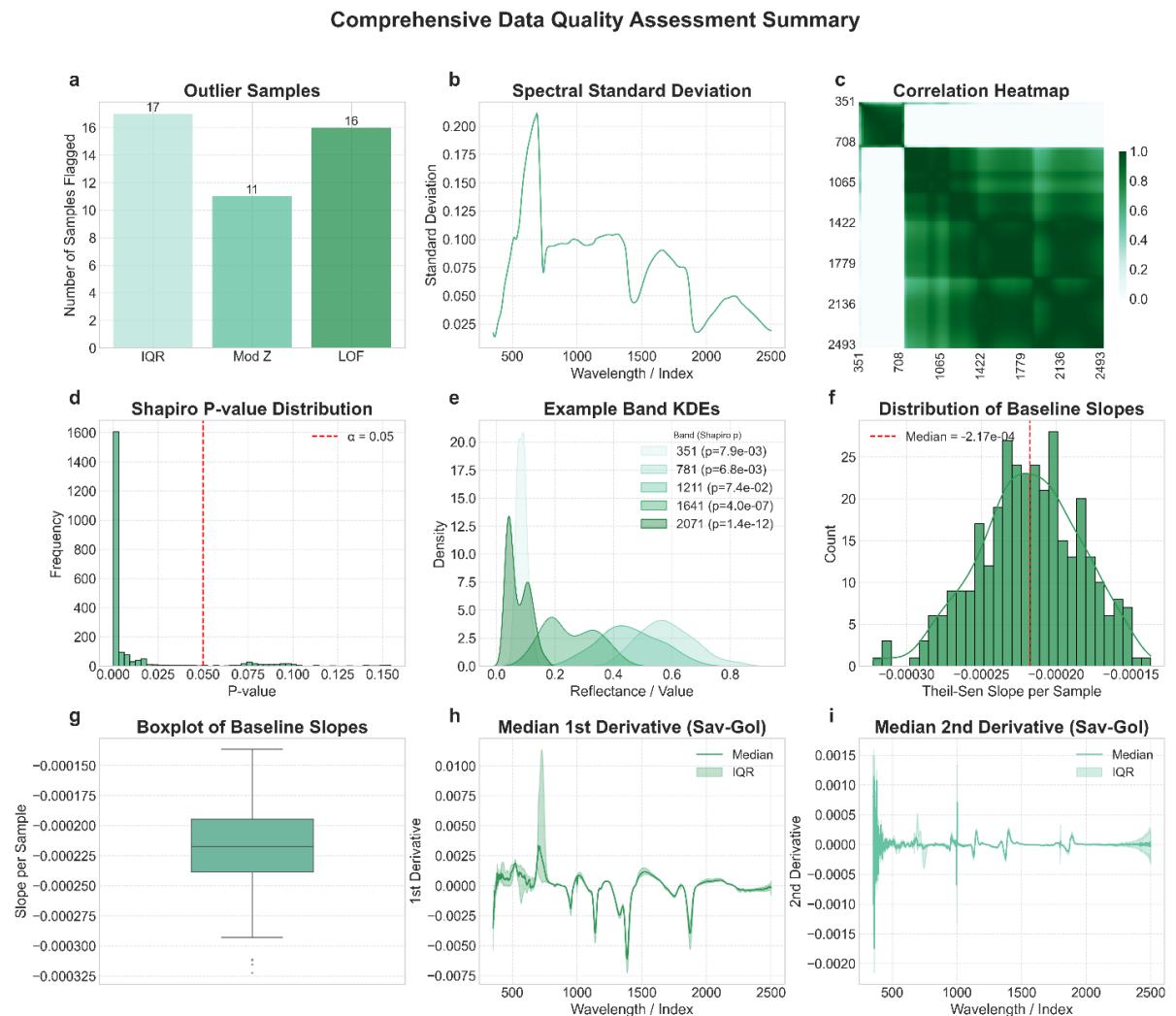


Fig. S8. Comprehensive Hyperspectral Data Quality Assessment Plots. (A) Number of potential outlier samples identified by Interquartile Range (IQR), Modified Z-score (Mod Z), and Local Outlier Factor (LOF) methods. (B) Spectral standard deviation calculated across all samples for each wavelength, indicating regions of higher variability (e.g., water absorption bands around 1450 nm and 1940 nm). (C) Correlation heatmap showing Pearson correlation

coefficients between spectral bands (sampled across ~100 bands for clarity). Darker green indicates higher positive correlation. (D) Distribution of Shapiro-Wilk test *p*-values for normality across all 2150 bands. The red dashed line indicates the significance threshold ($\alpha = 0.05$). The high frequency of *p*-values near zero confirms widespread non-normality. (E) Kernel density estimates (KDEs) illustrating the non-normal distributions of five representative spectral bands. (F) Histogram showing the distribution of Theil-Sen baseline slopes calculated per sample. The red dashed line indicates the median slope (-2.17e-4), revealing a consistent negative trend. (G) Boxplot visualising the distribution, median, interquartile range (IQR), and outliers for the per-sample Theil-Sen baseline slopes. (H) Median first-order Savitzky-Golay derivative spectrum (window=5, polyorder=2) across all samples. The shaded area represents the IQR, indicating low variability in the rate of spectral change. (I) Median second-order Savitzky-Golay derivative spectrum, highlighting spectral curvature and inflection points. The narrow IQR indicates stable spectral shapes.

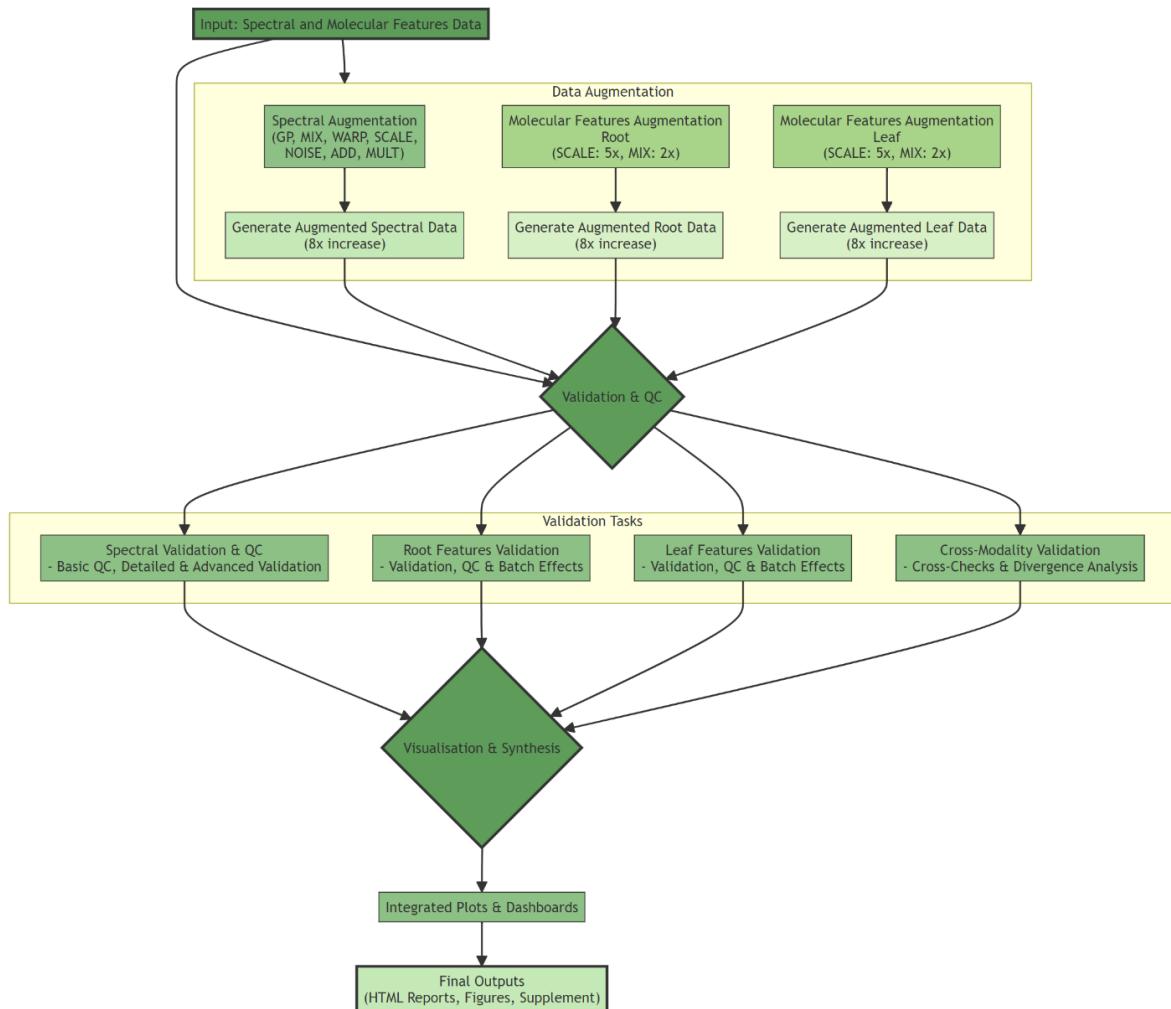


Fig. S9. Workflow diagram illustrating the data augmentation and validation pipeline for spectral and Molecular feature datasets.

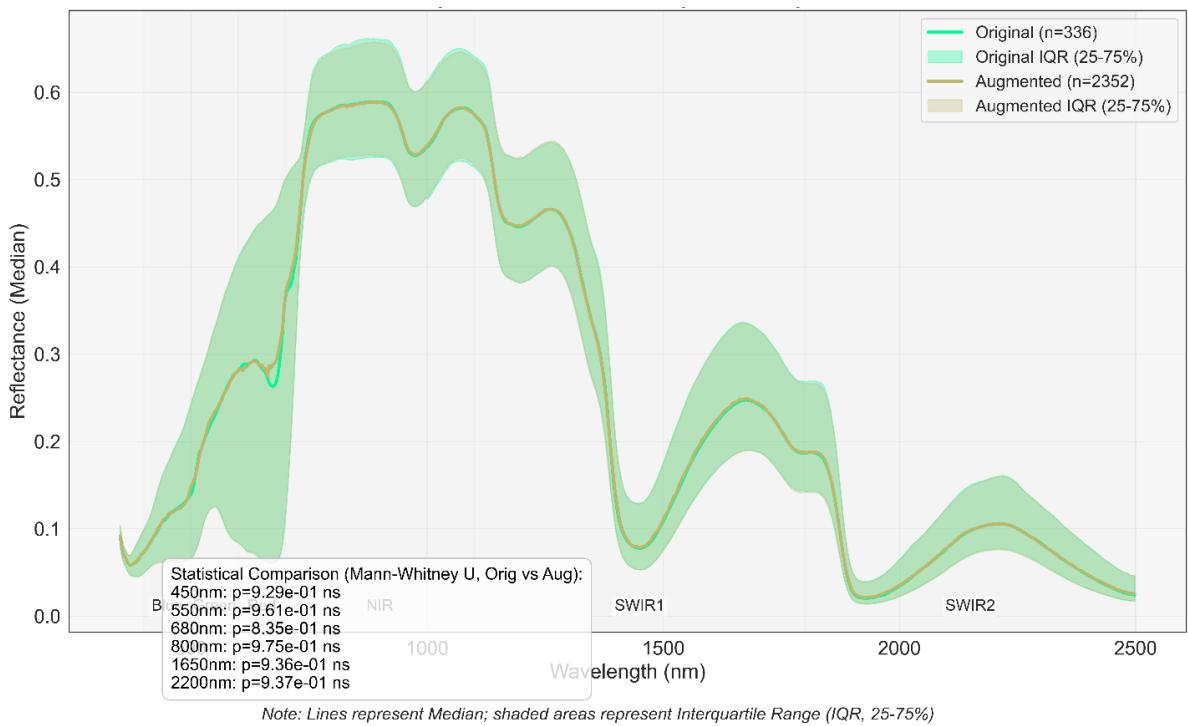


Fig. S10. Spectral Signature Comparison for the Root Dataset. Median spectral reflectance profiles (350-2500 nm) for original ($n=336$) and augmented ($n=2352$) datasets. Shaded areas represent the Interquartile Range (IQR, 25-75%). Statistical comparisons (Mann-Whitney U) at key wavelengths indicate excellent preservation of the overall spectral shape and distribution (all $p > 0.05$).

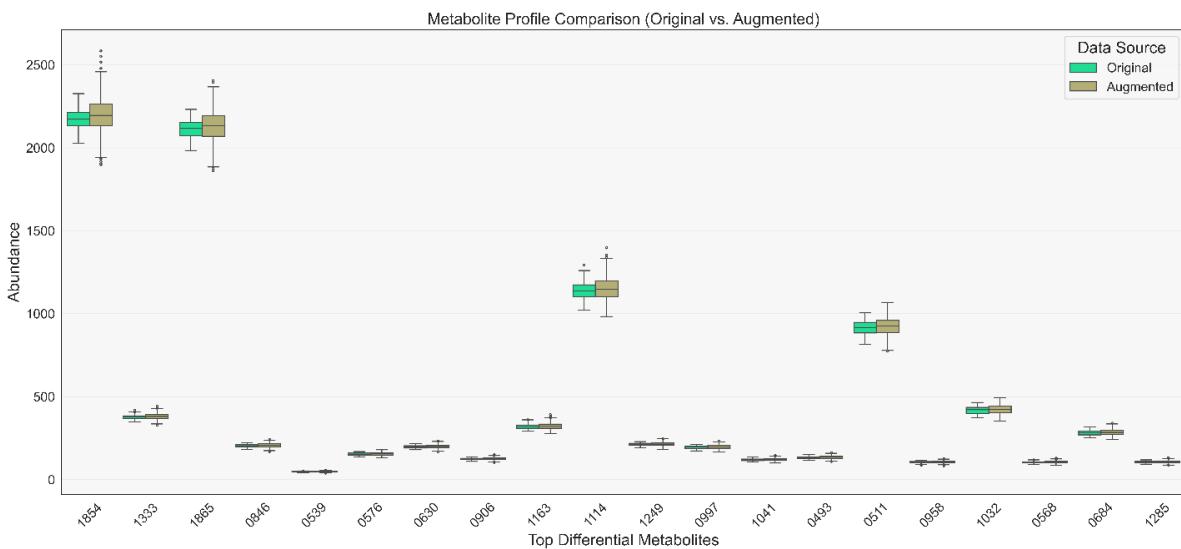
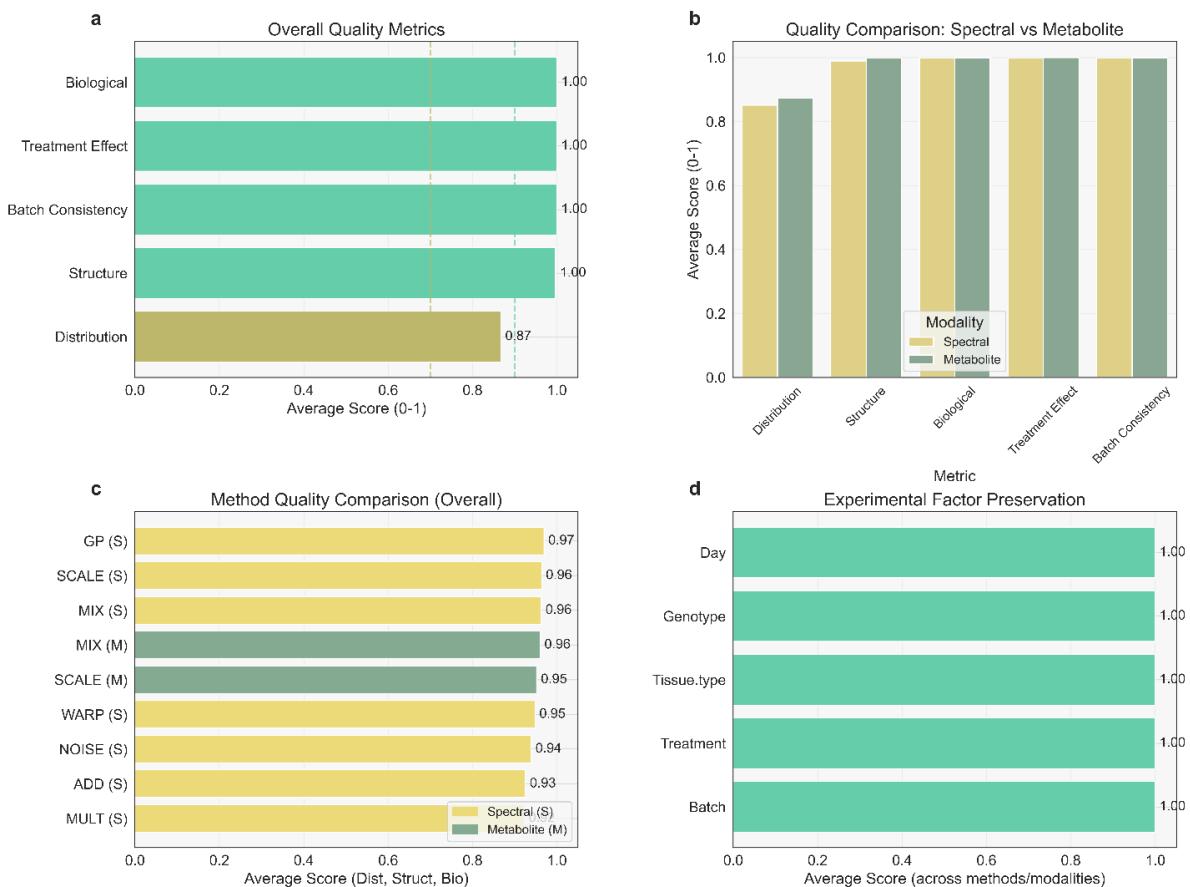


Fig. S11. Molecular Feature Profile Comparison for the Root Dataset. Comparison of abundance distributions for the top 20 most variable molecular features between original and augmented root (or specify leaf/root if different plots) datasets. Boxplots represent IQR, with whiskers extending to 1.5x IQR, demonstrating high similarity in distributions for key features post-augmentation.

Comprehensive Augmentation Quality Assessment



Note: Scores based on non-parametric statistics (JS Div., Spearman Corr., Median Profiles)

Fig. S12. Comprehensive Augmentation Quality Assessment for the Root Dataset. Multi-panel figure summarising key validation metrics. (A) Overall quality scores across five dimensions (Distribution, Structure, Biological Signal, Treatment Effect, Batch Consistency) averaged across modalities and methods. (B) Comparison of average quality scores between spectral and molecular feature data for each metric. (C) Overall quality comparison of individual spectral (S) and molecular feature (M) augmentation methods. (D) Experimental factor preservation scores, averaged across methods and modalities. Scores approaching 1.0 indicate better preservation/quality.

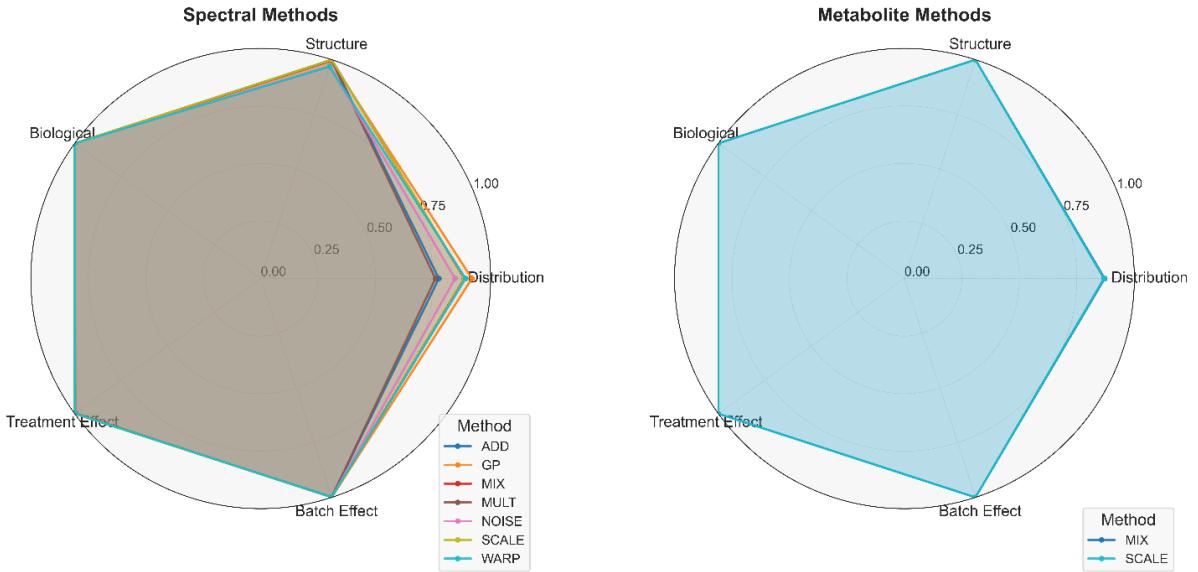


Fig. S13. Augmentation Method Comparison Across Key Metrics for the Root Dataset. Radar charts comparing the performance of spectral methods (left) and molecular feature methods (right) across five validation dimensions (Distribution, Structure, Biological, Treatment Effect, Batch Effect). Scores are normalised (0-1, higher is better), illustrating the relative strengths of each augmentation approach.

Detailed validation outputs are accessible via GitHub at

https://github.com/shoaibms/mofa_transformer/tree/main/html

Supplementary Tables

Table S1: Top 30 Conditional Spectral-to-Molecular (S2M) Attention Pairs under Peak Stress.

| Tissue | Condition | Rank | Spectral_Feature | Molecular_Feature | Mean_Attention |
|--------|-----------|------|------------------|-------------------|----------------|
| Leaf | G1_T1_TP3 | 1 | W_565 | N_1909 | 0.050723 |
| Leaf | G1_T1_TP3 | 2 | W_582 | N_1909 | 0.050689 |
| Leaf | G1_T1_TP3 | 3 | W_574 | N_1909 | 0.050566 |
| Leaf | G1_T1_TP3 | 4 | W_559 | N_1909 | 0.050538 |
| Leaf | G1_T1_TP3 | 5 | W_604 | N_1909 | 0.050423 |
| Leaf | G1_T1_TP3 | 6 | W_605 | N_1909 | 0.050381 |
| Leaf | G1_T1_TP3 | 7 | W_563 | N_1909 | 0.050252 |
| Leaf | G1_T1_TP3 | 8 | W_562 | N_1909 | 0.050235 |
| Leaf | G1_T1_TP3 | 9 | W_635 | N_1909 | 0.050194 |
| Leaf | G1_T1_TP3 | 10 | W_547 | N_1909 | 0.05018 |
| Leaf | G1_T1_TP3 | 11 | W_613 | N_1909 | 0.050159 |

| | | | | | |
|------|-----------|----|-------|--------|----------|
| Leaf | G1_T1_TP3 | 12 | W_616 | N_1909 | 0.050141 |
| Leaf | G1_T1_TP3 | 13 | W_580 | N_1909 | 0.05014 |
| Leaf | G1_T1_TP3 | 14 | W_561 | N_1909 | 0.050119 |
| Leaf | G1_T1_TP3 | 15 | W_577 | N_1909 | 0.05008 |
| Leaf | G1_T1_TP3 | 16 | W_558 | N_1909 | 0.050077 |
| Leaf | G1_T1_TP3 | 17 | W_618 | N_1909 | 0.050035 |
| Leaf | G1_T1_TP3 | 18 | W_599 | N_1909 | 0.049933 |
| Leaf | G1_T1_TP3 | 19 | W_612 | N_1909 | 0.049927 |
| Leaf | G1_T1_TP3 | 20 | W_570 | N_1909 | 0.049925 |
| Leaf | G1_T1_TP3 | 21 | W_564 | N_1909 | 0.049885 |
| Leaf | G1_T1_TP3 | 22 | W_620 | N_1909 | 0.049868 |
| Leaf | G1_T1_TP3 | 23 | W_625 | N_1909 | 0.049854 |
| Leaf | G1_T1_TP3 | 24 | W_556 | N_1909 | 0.049841 |
| Leaf | G1_T1_TP3 | 25 | W_546 | N_1909 | 0.049829 |
| Leaf | G1_T1_TP3 | 26 | W_584 | N_1909 | 0.049815 |
| Leaf | G1_T1_TP3 | 27 | W_592 | N_1909 | 0.049798 |
| Leaf | G1_T1_TP3 | 28 | W_575 | N_1909 | 0.049784 |
| Leaf | G1_T1_TP3 | 29 | W_600 | N_1909 | 0.04972 |
| Leaf | G1_T1_TP3 | 30 | W_593 | N_1909 | 0.049702 |
| Leaf | G2_T1_TP3 | 1 | W_574 | N_3029 | 0.024271 |
| Leaf | G2_T1_TP3 | 2 | W_565 | N_3029 | 0.024247 |
| Leaf | G2_T1_TP3 | 3 | W_559 | N_3029 | 0.024134 |
| Leaf | G2_T1_TP3 | 4 | W_577 | N_3029 | 0.024132 |
| Leaf | G2_T1_TP3 | 5 | W_582 | N_3029 | 0.02413 |
| Leaf | G2_T1_TP3 | 6 | W_563 | N_3029 | 0.024064 |
| Leaf | G2_T1_TP3 | 7 | W_598 | N_3029 | 0.024057 |
| Leaf | G2_T1_TP3 | 8 | W_556 | N_3029 | 0.024052 |
| Leaf | G2_T1_TP3 | 9 | W_625 | N_3029 | 0.024013 |
| Leaf | G2_T1_TP3 | 10 | W_558 | N_3029 | 0.02398 |
| Leaf | G2_T1_TP3 | 11 | W_637 | N_3029 | 0.02398 |
| Leaf | G2_T1_TP3 | 12 | W_611 | N_3029 | 0.023972 |
| Leaf | G2_T1_TP3 | 13 | W_579 | N_3029 | 0.023971 |
| Leaf | G2_T1_TP3 | 14 | W_616 | N_3029 | 0.023964 |
| Leaf | G2_T1_TP3 | 15 | W_580 | N_3029 | 0.023962 |
| Leaf | G2_T1_TP3 | 16 | W_575 | N_3029 | 0.023962 |
| Leaf | G2_T1_TP3 | 17 | W_561 | N_3029 | 0.023958 |
| Leaf | G2_T1_TP3 | 18 | W_600 | N_3029 | 0.023953 |
| Leaf | G2_T1_TP3 | 19 | W_612 | N_3029 | 0.023952 |
| Leaf | G2_T1_TP3 | 20 | W_597 | N_3029 | 0.023952 |
| Leaf | G2_T1_TP3 | 21 | W_564 | N_3029 | 0.023947 |
| Leaf | G2_T1_TP3 | 22 | W_572 | N_3029 | 0.023944 |
| Leaf | G2_T1_TP3 | 23 | W_610 | N_3029 | 0.023936 |
| Leaf | G2_T1_TP3 | 24 | W_599 | N_3029 | 0.023932 |

| | | | | | |
|------|----------------|----|--------|--------|----------|
| Leaf | G2_T1_TP3 | 25 | W_596 | N_3029 | 0.02393 |
| Leaf | G2_T1_TP3 | 26 | W_566 | N_3029 | 0.023922 |
| Leaf | G2_T1_TP3 | 27 | W_562 | N_3029 | 0.023922 |
| Leaf | G2_T1_TP3 | 28 | W_618 | N_3029 | 0.023907 |
| Leaf | G2_T1_TP3 | 29 | W_609 | N_3029 | 0.023903 |
| Leaf | G2_T1_TP3 | 30 | W_570 | N_3029 | 0.023898 |
| Leaf | Overall_T1_TP3 | 1 | W_565 | N_1909 | 0.050723 |
| Leaf | Overall_T1_TP3 | 2 | W_582 | N_1909 | 0.050689 |
| Leaf | Overall_T1_TP3 | 3 | W_574 | N_1909 | 0.050566 |
| Leaf | Overall_T1_TP3 | 4 | W_559 | N_1909 | 0.050538 |
| Leaf | Overall_T1_TP3 | 5 | W_604 | N_1909 | 0.050423 |
| Leaf | Overall_T1_TP3 | 6 | W_605 | N_1909 | 0.050381 |
| Leaf | Overall_T1_TP3 | 7 | W_563 | N_1909 | 0.050252 |
| Leaf | Overall_T1_TP3 | 8 | W_562 | N_1909 | 0.050235 |
| Leaf | Overall_T1_TP3 | 9 | W_635 | N_1909 | 0.050194 |
| Leaf | Overall_T1_TP3 | 10 | W_547 | N_1909 | 0.05018 |
| Leaf | Overall_T1_TP3 | 11 | W_613 | N_1909 | 0.050159 |
| Leaf | Overall_T1_TP3 | 12 | W_616 | N_1909 | 0.050141 |
| Leaf | Overall_T1_TP3 | 13 | W_580 | N_1909 | 0.05014 |
| Leaf | Overall_T1_TP3 | 14 | W_561 | N_1909 | 0.050119 |
| Leaf | Overall_T1_TP3 | 15 | W_577 | N_1909 | 0.05008 |
| Leaf | Overall_T1_TP3 | 16 | W_558 | N_1909 | 0.050077 |
| Leaf | Overall_T1_TP3 | 17 | W_618 | N_1909 | 0.050035 |
| Leaf | Overall_T1_TP3 | 18 | W_599 | N_1909 | 0.049933 |
| Leaf | Overall_T1_TP3 | 19 | W_612 | N_1909 | 0.049927 |
| Leaf | Overall_T1_TP3 | 20 | W_570 | N_1909 | 0.049925 |
| Leaf | Overall_T1_TP3 | 21 | W_564 | N_1909 | 0.049885 |
| Leaf | Overall_T1_TP3 | 22 | W_620 | N_1909 | 0.049868 |
| Leaf | Overall_T1_TP3 | 23 | W_625 | N_1909 | 0.049854 |
| Leaf | Overall_T1_TP3 | 24 | W_556 | N_1909 | 0.049841 |
| Leaf | Overall_T1_TP3 | 25 | W_546 | N_1909 | 0.049829 |
| Leaf | Overall_T1_TP3 | 26 | W_584 | N_1909 | 0.049815 |
| Leaf | Overall_T1_TP3 | 27 | W_592 | N_1909 | 0.049798 |
| Leaf | Overall_T1_TP3 | 28 | W_575 | N_1909 | 0.049784 |
| Leaf | Overall_T1_TP3 | 29 | W_600 | N_1909 | 0.04972 |
| Leaf | Overall_T1_TP3 | 30 | W_593 | N_1909 | 0.049702 |
| Root | G1_T1_TP3 | 1 | W_1100 | N_0512 | 0.076602 |
| Root | G1_T1_TP3 | 2 | W_1111 | N_0512 | 0.076592 |
| Root | G1_T1_TP3 | 3 | W_1105 | N_0512 | 0.076379 |
| Root | G1_T1_TP3 | 4 | W_1109 | N_0512 | 0.07616 |
| Root | G1_T1_TP3 | 5 | W_1108 | N_0512 | 0.076007 |
| Root | G1_T1_TP3 | 6 | W_1104 | N_0512 | 0.075841 |
| Root | G1_T1_TP3 | 7 | W_1112 | N_0512 | 0.075829 |

| | | | | | |
|------|-----------|----|--------|--------|----------|
| Root | G1_T1_TP3 | 8 | W_1101 | N_0512 | 0.075526 |
| Root | G1_T1_TP3 | 9 | W_1107 | N_0512 | 0.0754 |
| Root | G1_T1_TP3 | 10 | W_1106 | N_0512 | 0.075291 |
| Root | G1_T1_TP3 | 11 | W_1102 | N_0512 | 0.075199 |
| Root | G1_T1_TP3 | 12 | W_1099 | N_0512 | 0.075012 |
| Root | G1_T1_TP3 | 13 | W_1092 | N_0512 | 0.07488 |
| Root | G1_T1_TP3 | 14 | W_1090 | N_0512 | 0.074855 |
| Root | G1_T1_TP3 | 15 | W_1096 | N_0512 | 0.074656 |
| Root | G1_T1_TP3 | 16 | W_1098 | N_0512 | 0.07454 |
| Root | G1_T1_TP3 | 17 | W_1097 | N_0512 | 0.074492 |
| Root | G1_T1_TP3 | 18 | W_1103 | N_0512 | 0.074348 |
| Root | G1_T1_TP3 | 19 | W_1095 | N_0512 | 0.074208 |
| Root | G1_T1_TP3 | 20 | W_1091 | N_0512 | 0.074202 |
| Root | G1_T1_TP3 | 21 | W_1094 | N_0512 | 0.074148 |
| Root | G1_T1_TP3 | 22 | W_1082 | N_0512 | 0.074076 |
| Root | G1_T1_TP3 | 23 | W_1088 | N_0512 | 0.073957 |
| Root | G1_T1_TP3 | 24 | W_1048 | N_0512 | 0.073866 |
| Root | G1_T1_TP3 | 25 | W_1089 | N_0512 | 0.073851 |
| Root | G1_T1_TP3 | 26 | W_1084 | N_0512 | 0.073725 |
| Root | G1_T1_TP3 | 27 | W_1071 | N_0512 | 0.073618 |
| Root | G1_T1_TP3 | 28 | W_1093 | N_0512 | 0.073612 |
| Root | G1_T1_TP3 | 29 | W_1083 | N_0512 | 0.073584 |
| Root | G1_T1_TP3 | 30 | W_1086 | N_0512 | 0.073486 |
| Root | G2_T1_TP3 | 1 | W_1108 | N_1234 | 0.052533 |
| Root | G2_T1_TP3 | 2 | W_1107 | N_1234 | 0.051924 |
| Root | G2_T1_TP3 | 3 | W_1100 | N_1234 | 0.051792 |
| Root | G2_T1_TP3 | 4 | W_1112 | N_1234 | 0.051782 |
| Root | G2_T1_TP3 | 5 | W_1105 | N_1234 | 0.051746 |
| Root | G2_T1_TP3 | 6 | W_1102 | N_1234 | 0.051728 |
| Root | G2_T1_TP3 | 7 | W_1104 | N_1234 | 0.05171 |
| Root | G2_T1_TP3 | 8 | W_1111 | N_1234 | 0.051608 |
| Root | G2_T1_TP3 | 9 | W_1109 | N_1234 | 0.051407 |
| Root | G2_T1_TP3 | 10 | W_1099 | N_1234 | 0.051214 |
| Root | G2_T1_TP3 | 11 | W_1101 | N_1234 | 0.05116 |
| Root | G2_T1_TP3 | 12 | W_1090 | N_1234 | 0.050908 |
| Root | G2_T1_TP3 | 13 | W_1094 | N_1234 | 0.05089 |
| Root | G2_T1_TP3 | 14 | W_1106 | N_1234 | 0.050878 |
| Root | G2_T1_TP3 | 15 | W_1097 | N_1234 | 0.050821 |
| Root | G2_T1_TP3 | 16 | W_1103 | N_1234 | 0.050819 |
| Root | G2_T1_TP3 | 17 | W_1098 | N_1234 | 0.050795 |
| Root | G2_T1_TP3 | 18 | W_1096 | N_1234 | 0.050794 |
| Root | G2_T1_TP3 | 19 | W_1084 | N_1234 | 0.050678 |
| Root | G2_T1_TP3 | 20 | W_1092 | N_1234 | 0.050545 |

| | | | | | |
|------|----------------|----|--------|--------|----------|
| Root | G2_T1_TP3 | 21 | W_1086 | N_1234 | 0.050433 |
| Root | G2_T1_TP3 | 22 | W_1083 | N_1234 | 0.050308 |
| Root | G2_T1_TP3 | 23 | W_1082 | N_1234 | 0.0503 |
| Root | G2_T1_TP3 | 24 | W_1095 | N_1234 | 0.05011 |
| Root | G2_T1_TP3 | 25 | W_1089 | N_1234 | 0.049995 |
| Root | G2_T1_TP3 | 26 | W_1091 | N_1234 | 0.049961 |
| Root | G2_T1_TP3 | 27 | W_1093 | N_1234 | 0.049917 |
| Root | G2_T1_TP3 | 28 | W_1071 | N_1234 | 0.049915 |
| Root | G2_T1_TP3 | 29 | W_1069 | N_1234 | 0.049814 |
| Root | G2_T1_TP3 | 30 | W_1079 | N_1234 | 0.049768 |
| Root | Overall_T1_TP3 | 1 | W_1100 | N_0512 | 0.076602 |
| Root | Overall_T1_TP3 | 2 | W_1111 | N_0512 | 0.076592 |
| Root | Overall_T1_TP3 | 3 | W_1105 | N_0512 | 0.076379 |
| Root | Overall_T1_TP3 | 4 | W_1109 | N_0512 | 0.07616 |
| Root | Overall_T1_TP3 | 5 | W_1108 | N_0512 | 0.076007 |
| Root | Overall_T1_TP3 | 6 | W_1104 | N_0512 | 0.075841 |
| Root | Overall_T1_TP3 | 7 | W_1112 | N_0512 | 0.075829 |
| Root | Overall_T1_TP3 | 8 | W_1101 | N_0512 | 0.075526 |
| Root | Overall_T1_TP3 | 9 | W_1107 | N_0512 | 0.0754 |
| Root | Overall_T1_TP3 | 10 | W_1106 | N_0512 | 0.075291 |
| Root | Overall_T1_TP3 | 11 | W_1102 | N_0512 | 0.075199 |
| Root | Overall_T1_TP3 | 12 | W_1099 | N_0512 | 0.075012 |
| Root | Overall_T1_TP3 | 13 | W_1092 | N_0512 | 0.07488 |
| Root | Overall_T1_TP3 | 14 | W_1090 | N_0512 | 0.074855 |
| Root | Overall_T1_TP3 | 15 | W_1096 | N_0512 | 0.074656 |
| Root | Overall_T1_TP3 | 16 | W_1098 | N_0512 | 0.07454 |
| Root | Overall_T1_TP3 | 17 | W_1097 | N_0512 | 0.074492 |
| Root | Overall_T1_TP3 | 18 | W_1103 | N_0512 | 0.074348 |
| Root | Overall_T1_TP3 | 19 | W_1095 | N_0512 | 0.074208 |
| Root | Overall_T1_TP3 | 20 | W_1091 | N_0512 | 0.074202 |
| Root | Overall_T1_TP3 | 21 | W_1094 | N_0512 | 0.074148 |
| Root | Overall_T1_TP3 | 22 | W_1082 | N_0512 | 0.074076 |
| Root | Overall_T1_TP3 | 23 | W_1088 | N_0512 | 0.073957 |
| Root | Overall_T1_TP3 | 24 | W_1048 | N_0512 | 0.073866 |
| Root | Overall_T1_TP3 | 25 | W_1089 | N_0512 | 0.073851 |
| Root | Overall_T1_TP3 | 26 | W_1084 | N_0512 | 0.073725 |
| Root | Overall_T1_TP3 | 27 | W_1071 | N_0512 | 0.073618 |
| Root | Overall_T1_TP3 | 28 | W_1093 | N_0512 | 0.073612 |
| Root | Overall_T1_TP3 | 29 | W_1083 | N_0512 | 0.073584 |
| Root | Overall_T1_TP3 | 30 | W_1086 | N_0512 | 0.073486 |

Table S2: Statistical Comparison of View-Level S2M Attention Metrics Between Conditions. This table presents the results of Mann-Whitney U tests comparing view-level S2M attention statistics (Standard Deviation - StdAttn_S2M; 95th Percentile - P95Attn_S2M) between specified conditions. P-values were adjusted for multiple comparisons using the Benjamini-Hochberg False Discovery Rate (FDR) method. Significance was determined at FDR < 0.05.

| Comparison Description | Tissue | Metric | Group 1 Mean (SD) [n] | Group 2 Mean (SD) [n] | P-value (Raw) | P-value (FDR) | Significant (FDR < 0.05) |
|---|--------|-------------|---------------------------------------|---------------------------------------|---------------|---------------|--------------------------|
| StdAttn_S2M: G1 vs G2 (T1, Time Point 3) | Leaf | StdAttn_S2M | 0.0144 (0.0019) [n=17] | 0.0101 (0.0022) [n=17] | 2.64E-05 | 0.00031 | TRUE |
| StdAttn_S2M: G1 vs G2 (T1, Time Point 3) | Root | StdAttn_S2M | 0.0201 (0.0080) [n=17] | 0.0178 (0.0075) [n=17] | 0.2704 | 0.4897 | FALSE |
| P95Attn_S2M: G1 vs G2 (T1, Time Point 3) | Leaf | P95Attn_S2M | 0.0200 (0.0029) [n=17] | 0.0228 (0.0040) [n=17] | 0.09146 | 0.2195 | FALSE |
| P95Attn_S2M: G1 vs G2 (T1, Time Point 3) | Root | P95Attn_S2M | 0.0287 (0.0123) [n=17] | 0.0199 (0.0053) [n=17] | 0.05375 | 0.1613 | FALSE |
| StdAttn_S2M: G1 Time Point 3 vs Time Point 1 (T1) | Leaf | StdAttn_S2M | 0.0144 (0.0019) [n=17] (Time Point 3) | 0.0151 (0.0036) [n=17] (Time Point 1) | 0.9177 | 0.9177 | FALSE |
| StdAttn_S2M: G1 Time Point 3 vs Time Point 1 (T1) | Root | StdAttn_S2M | 0.0201 (0.0080) [n=17] (Time Point 3) | 0.0219 (0.0094) [n=17] (Time Point 1) | 0.4084 | 0.5383 | FALSE |
| P95Attn_S2M: G1 Time Point 3 vs Time Point 1 (T1) | Leaf | P95Attn_S2M | 0.0200 (0.0029) [n=17] (Time Point 3) | 0.0226 (0.0040) [n=17] (Time Point 1) | 0.04961 | 0.1613 | FALSE |
| P95Attn_S2M: G1 Time Point 3 vs Time Point 1 (T1) | Root | P95Attn_S2M | 0.0287 (0.0123) [n=17] (Time Point 3) | 0.0267 (0.0053) [n=17] (Time Point 1) | 0.4084 | 0.5383 | FALSE |
| StdAttn_S2M: G2 Time Point 3 vs Time Point 1 (T1) | Leaf | StdAttn_S2M | 0.0101 (0.0022) [n=17] (Time Point 3) | 0.0113 (0.0035) [n=17] (Time Point 1) | 0.4486 | 0.5383 | FALSE |
| StdAttn_S2M: G2 Time Point 3 vs Time Point 1 (T1) | Root | StdAttn_S2M | 0.0178 (0.0075) [n=17] (Time Point 3) | 0.0164 (0.0037) [n=17] (Time Point 1) | 0.7829 | 0.8541 | FALSE |
| P95Attn_S2M: G2 Time Point 3 vs Time Point 1 (T1) | Leaf | P95Attn_S2M | 0.0228 (0.0040) [n=17] (Time Point 3) | 0.0218 (0.0052) [n=17] (Time Point 1) | 0.2856 | 0.4897 | FALSE |
| P95Attn_S2M: G2 Time Point 3 vs Time Point 1 (T1) | Root | P95Attn_S2M | 0.0199 (0.0053) [n=17] (Time Point 3) | 0.0276 (0.0076) [n=17] (Time Point 1) | 0.00341 | 0.02049 | TRUE |

SD: Standard Deviation. n: Number of samples per group.

Time Point comparisons show Time Point 3 as Group 1 and Time Point 1 as Group 2.

Table S3: Predictive Model Performance Metrics

| Tissue | Task | RandomForest | KNN (k=5) | Transformer |
|--------|------------|--------------|-----------|-------------|
| Leaf | Genotype | 1 | 0.9901 | 0.9505 |
| Leaf | Treatment | 1 | 0.9851 | 0.9802 |
| Leaf | Time Point | 0.99 | 0.9503 | 0.7559 |
| Root | Genotype | 0.995 | 0.9505 | 0.8096 |
| Root | Treatment | 1 | 0.9901 | 1 |

Table S4: Top 20 Predictive Features Ranked by SHAP Importance

Leaf Tissue

| Rank | Genotype_Feature | Genotype_SHAP | Genotype_FeatureType | Treatment_Feature | Treatment_SHAP | Treatment_FeatureType | Time_Point_Feature | Time_Point_SHAP | Time_Point_FeatureType |
|------|------------------|---------------|----------------------|-------------------|----------------|-----------------------|--------------------|-----------------|------------------------|
| 1 | N_1909 | 0.578962 | Molecular Features | P_1199 | 1.088617 | Molecular Features | P_1199 | 0.311726 | Molecular Features |
| 2 | N_3474 | 0.262881 | Molecular Features | P_0078 | 0.587104 | Molecular Features | N_1909 | 0.200217 | Molecular Features |
| 3 | W_550 | 0.213938 | Spectral | N_0474 | 0.301634 | Molecular Features | P_0078 | 0.171235 | Molecular Features |
| 4 | W_552 | 0.185353 | Spectral | N_1909 | 0.288474 | Molecular Features | N_0474 | 0.097239 | Molecular Features |
| 5 | P_2606 | 0.184694 | Molecular Features | N_0619 | 0.216223 | Molecular Features | P_2606 | 0.071501 | Molecular Features |
| 6 | W_553 | 0.175338 | Spectral | N_0584 | 0.209585 | Molecular Features | N_0584 | 0.066713 | Molecular Features |
| 7 | N_1481 | 0.171813 | Molecular Features | P_1532 | 0.201628 | Molecular Features | N_0619 | 0.066266 | Molecular Features |
| 8 | W_554 | 0.166336 | Spectral | N_0579 | 0.161876 | Molecular Features | N_3474 | 0.066208 | Molecular Features |
| 9 | N_2812 | 0.16231 | Molecular Features | P_0905 | 0.160727 | Molecular Features | N_0607 | 0.065063 | Molecular Features |
| 10 | N_2814 | 0.160455 | Molecular Features | N_1114 | 0.159035 | Molecular Features | N_2814 | 0.062427 | Molecular Features |
| 11 | N_2475 | 0.145299 | Molecular Features | N_0407 | 0.101844 | Molecular Features | P_2305 | 0.057564 | Molecular Features |
| 12 | W_546 | 0.136296 | Spectral | N_0569 | 0.098648 | Molecular Features | N_2948 | 0.055863 | Molecular Features |
| 13 | W_549 | 0.131387 | Spectral | N_0607 | 0.094963 | Molecular Features | N_2117 | 0.055109 | Molecular Features |

| | | | | | | | | | |
|----|--------|--------------|-----------------------|--------|----------|-----------------------|------------|--------------|-----------------------|
| 14 | P_1199 | 0.12885 6 | Molecular Features | N_1800 | 0.091045 | Molecular Features | N_180 0 | 0.054 351 | Molecular Features |
| 15 | W_560 | 0.11636 3 | Spectral | N_0808 | 0.086488 | Molecular Features | N_049 5 | 0.053 136 | Molecular Features |
| 16 | W_557 | 0.11150 7 | Spectral | N_0578 | 0.085821 | Molecular Features | P_090 5 | 0.052 818 | Molecular Features |
| 17 | N_2797 | 0.10733 8 | Molecular Features | P_1625 | 0.083718 | Molecular Features | N_057 9 | 0.052 253 | Molecular Features |
| 18 | N_1061 | 0.10631 1 | Molecular Features | N_0037 | 0.08187 | Molecular Features | N_162 9 | 0.051 408 | Molecular Features |
| 19 | W_551 | 0.10550 2 | Spectral | N_0495 | 0.081684 | Molecular Features | P_153 2 | 0.051 053 | Molecular Features |
| 20 | N_0467 | 0.10532 3 | Molecular Features | P_0930 | 0.08019 | Molecular Features | P_260 5 | 0.050 85 | Molecular Features |

Root Tissue

| Rank | Genotype_Feature | Genotyp e_SHAP | Genotype_Fe atureType | Treatment _Feature | Treatmen t_SHAP | Treatment_Fe atureType | Time Point_F eature | Time Point_ SHAP | Time Point_Feat ureType |
|------|------------------|----------------|-----------------------|--------------------|-----------------|------------------------|---------------------|------------------|-------------------------|
| 1 | N_1234 | 0.158798 | Molecular Features | P_0862 | 0.387749 | Molecular Features | N_1234 41 | 0.3905 | Molecular Features |
| 2 | N_0512 | 0.144171 | Molecular Features | N_0584 | 0.367427 | Molecular Features | N_0512 72 | 0.3693 | Molecular Features |
| 3 | W_2398 | 0.100917 | Spectral | N_0619 | 0.332646 | Molecular Features | W_2404 65 | 0.2062 | Spectral |
| 4 | W_2404 | 0.07583 | Spectral | P_0834 | 0.327219 | Molecular Features | N_0557 98 | 0.1673 | Molecular Features |
| 5 | N_0557 | 0.060491 | Molecular Features | P_0905 | 0.271153 | Molecular Features | W_2398 86 | 0.1458 | Spectral |
| 6 | P_0952 | 0.050167 | Molecular Features | P_0805 | 0.255789 | Molecular Features | N_1278 45 | 0.1414 | Molecular Features |
| 7 | P_0834 | 0.046898 | Molecular Features | N_0579 | 0.243081 | Molecular Features | P_0843 25 | 0.1344 | Molecular Features |
| 8 | P_0862 | 0.046614 | Molecular Features | P_0986 | 0.237489 | Molecular Features | N_0584 74 | 0.1252 | Molecular Features |
| 9 | N_1568 | 0.043444 | Molecular Features | N_0474 | 0.234178 | Molecular Features | P_0862 46 | 0.1244 | Molecular Features |
| 10 | N_1100 | 0.041033 | Molecular Features | P_1048 | 0.22347 | Molecular Features | N_1428 03 | 0.1077 | Molecular Features |
| 11 | P_0805 | 0.039392 | Molecular Features | P_1116 | 0.213175 | Molecular Features | P_0952 98 | 0.1076 | Molecular Features |
| 12 | P_0811 | 0.038139 | Molecular Features | N_0557 | 0.190509 | Molecular Features | P_0824 11 | 0.0970 | Molecular Features |

| | | | | | | | | | |
|----|--------|----------|--------------------|--------|----------|--------------------|--------|--------|--------------------|
| 13 | P_2488 | 0.03783 | Molecular Features | N_0512 | 0.179457 | Molecular Features | P_0834 | 0.0965 | Molecular Features |
| 14 | P_1919 | 0.035454 | Molecular Features | P_0824 | 0.171381 | Molecular Features | N_0619 | 0.0909 | Molecular Features |
| 15 | P_0917 | 0.034342 | Molecular Features | N_0808 | 0.158116 | Molecular Features | N_0654 | 0.0905 | Molecular Features |
| 16 | N_1866 | 0.033171 | Molecular Features | P_0843 | 0.149558 | Molecular Features | P_0876 | 0.0796 | Molecular Features |
| 17 | P_1467 | 0.032425 | Molecular Features | P_0876 | 0.14518 | Molecular Features | P_2488 | 0.0768 | Molecular Features |
| 18 | P_1647 | 0.032309 | Molecular Features | P_1200 | 0.143659 | Molecular Features | N_0808 | 0.0755 | Molecular Features |
| 19 | N_0734 | 0.032284 | Molecular Features | N_1328 | 0.134725 | Molecular Features | N_0579 | 0.0718 | Molecular Features |
| 20 | N_1328 | 0.030757 | Molecular Features | N_0326 | 0.134027 | Molecular Features | N_0326 | 0.0717 | Molecular Features |

Table S5: Overlapping Features (MOFA+ Factor 9 vs. SHAP Genotype)

| Feature | Wavelength | MOFA_Rank | SHAP_Rank | Weight | MeanAbsoluteSha p | p_value | cohen_d |
|---------|------------|-----------|-----------|----------|----------------------|----------|----------|
| W_546 | 546 | 16 | 5 | 0.168901 | 0.136296 | 0.069034 | -0.19903 |
| W_547 | 547 | 18 | 27 | 0.167816 | 0.042817 | 0.069299 | -0.19884 |
| W_548 | 548 | 17 | 10 | 0.168775 | 0.104274 | 0.06981 | -0.19847 |
| W_549 | 549 | 15 | 6 | 0.170953 | 0.131387 | 0.070411 | -0.19804 |
| W_550 | 550 | 13 | 1 | 0.171759 | 0.213938 | 0.071155 | -0.19752 |
| W_551 | 551 | 14 | 9 | 0.171035 | 0.105502 | 0.072107 | -0.19685 |
| W_552 | 552 | 6 | 2 | 0.173247 | 0.185353 | 0.07328 | -0.19605 |
| W_553 | 553 | 8 | 3 | 0.173113 | 0.175338 | 0.074639 | -0.19512 |
| W_554 | 554 | 1 | 4 | 0.17493 | 0.166336 | 0.076266 | -0.19403 |
| W_555 | 555 | 11 | 11 | 0.172381 | 0.100598 | 0.07831 | -0.19269 |
| W_556 | 556 | 3 | 37 | 0.17356 | 0.036698 | 0.080863 | -0.19105 |
| W_557 | 557 | 12 | 8 | 0.172276 | 0.111507 | 0.083795 | -0.18923 |
| W_558 | 558 | 9 | 26 | 0.172741 | 0.044463 | 0.087071 | -0.18725 |
| W_560 | 560 | 4 | 7 | 0.173366 | 0.116363 | 0.094709 | -0.18285 |
| W_561 | 561 | 10 | 12 | 0.172728 | 0.068141 | 0.099084 | -0.18046 |
| W_562 | 562 | 7 | 18 | 0.173189 | 0.056555 | 0.103761 | -0.178 |
| W_564 | 564 | 2 | 13 | 0.174548 | 0.066203 | 0.113774 | -0.17301 |
| W_568 | 568 | 5 | 16 | 0.173304 | 0.060152 | 0.139041 | -0.1618 |
| W_590 | 590 | 19 | 36 | 0.164778 | 0.037147 | 0.253992 | -0.12468 |
| W_597 | 597 | 22 | 32 | 0.159375 | 0.039311 | 0.272357 | -0.11996 |

| | | | | | | | |
|-------|-----|----|----|----------|----------|----------|----------|
| W_598 | 598 | 21 | 38 | 0.160943 | 0.036618 | 0.274696 | -0.11938 |
| W_603 | 603 | 23 | 28 | 0.159278 | 0.042795 | 0.285513 | -0.11672 |
| W_604 | 604 | 20 | 34 | 0.161218 | 0.037339 | 0.288206 | -0.11607 |
| W_609 | 609 | 25 | 35 | 0.155236 | 0.03716 | 0.304179 | -0.11229 |
| W_612 | 612 | 24 | 31 | 0.15561 | 0.04144 | 0.316209 | -0.10952 |
| W_613 | 613 | 27 | 19 | 0.152665 | 0.05607 | 0.320427 | -0.10857 |
| W_615 | 615 | 26 | 25 | 0.15452 | 0.045069 | 0.328239 | -0.10683 |
| W_616 | 616 | 31 | 15 | 0.150604 | 0.060405 | 0.332096 | -0.10598 |
| W_619 | 619 | 29 | 21 | 0.15155 | 0.052511 | 0.34201 | -0.10382 |
| W_620 | 620 | 38 | 30 | 0.149568 | 0.041561 | 0.344813 | -0.10322 |
| W_621 | 621 | 32 | 22 | 0.150422 | 0.050475 | 0.34741 | -0.10267 |
| W_622 | 622 | 28 | 29 | 0.152117 | 0.042663 | 0.349627 | -0.1022 |
| W_623 | 623 | 30 | 33 | 0.151411 | 0.039099 | 0.351569 | -0.10178 |
| W_624 | 624 | 34 | 14 | 0.150261 | 0.065938 | 0.353443 | -0.10139 |
| W_625 | 625 | 37 | 20 | 0.14982 | 0.054346 | 0.354864 | -0.10109 |
| W_630 | 630 | 33 | 17 | 0.150295 | 0.059966 | 0.359593 | -0.1001 |
| W_634 | 634 | 36 | 24 | 0.149934 | 0.047175 | 0.36315 | -0.09936 |
| W_635 | 635 | 35 | 23 | 0.150206 | 0.048022 | 0.364457 | -0.09909 |

Table S6: Nutrient Solution

| No | Nutrient | Concentration |
|----|--------------------------------------|---------------|
| 1 | MgSO ₄ | 4 mM |
| 2 | CaCl ₂ .2H ₂ O | 4 mM |
| 3* | KH ₂ PO ₄ pool | 3 mM |
| | K ₂ HPO ₄ pool | |
| 4 | Fe ⁺ (FeEDTANa Librel) | 400 µM |
| 5 | MnCl ₂ .4H ₂ O | 10 µM |
| | ZnSO ₄ .7H ₂ O | 10 µM |
| | CuSO ₄ | 2 µM |
| | H ₃ BO ₃ | 50 µM |
| | Na ₂ MoO ₄ | 0.2 µM |
| 6 | KNO ₃ | 1M |

*Pooled to make pH 6

Table S7: Genedata Expressionist® Refiner MS Settings

| Processing Step | Parameter | Positive Mode | Negative Mode |
|-----------------|-----------|---------------|---------------|
|-----------------|-----------|---------------|---------------|

| Data Import | RT Range | 0-16 minutes | 0-16 minutes |
|---|------------------------|----------------------|----------------------|
| Chromatogram Chemical Noise Subtraction | Smoothing RT Window | 3 scans | 3 scans |
| | Subtraction RT Window | 51 scans | 51 scans |
| | Quantile | 50% | 50% |
| | Intensity Threshold | 8000, 10000 | 3000 |
| RT Alignment | Max RT Shift | 0.4 minutes | 0.2 minutes |
| | Gap Penalty | 1 | 1 |
| Peak Detection | Summation Window | 5 scans | 5 scans |
| | Minimum Peak Size | 3 scans | 4 scans |
| | Maximum Merge Distance | 5 points | 5 points |
| | Merge Strategy | Centers | Boundaries |
| | Smoothing | m/z window: 3 points | m/z window: 3 points |
| | Refinement Threshold | 5% | 5% |
| | Consistency Threshold | 0.6 | 0.6 |
| Isotope Clustering | RT Tolerance | 0.05 minutes | 0.05 minutes |
| | m/z Tolerance | 5 ppm | 5 ppm |
| | Ionisation | Protonation | Deprotonation |
| | Charge Range | 1-2 | 1-5 |
| | Distance Measure | Log-Ratio | Log-Ratio |
| | Max Distance | 0.5 | 0.5 |

Table S8: List of Unique Spectral Outlier Samples

| Outlier sample index |
|----------------------|
| 28 |
| 31 |
| 65 |
| 82 |
| 89 |
| 105 |
| 125 |
| 134 |
| 148 |
| 159 |
| 183 |
| 203 |
| 205 |
| 208 |
| 215 |

219
221
239
240
242
243
244
246
247
248
249
250
285

Table S9: Comparison of Average Genotype S2M Attention

This table shows the comparison of the average Spectral-to-Molecular (S2M) attention between the tolerant (G1) and susceptible (G2) genotypes under peak stress conditions (Treatment 1, Time Point 3). The average attention was calculated across the top 100 overall S2M pairs identified from the analysis.

| Tissue | N Pairs Compared | Avg G1 S2M Attention | Avg G2 S2M Attention | % Difference (G1 vs G2) |
|--------|------------------|----------------------|----------------------|-------------------------|
| Leaf | 100 | 0.0484 | 0.0106 | 355.10% |
| Root | 100 | 0.0292 | 0.0478 | -38.90% |

Supplementary Note 1: Detailed Experimental and Computational Methods

LCMS Detailed Data Preprocessing Workflow

Initial Data Cleaning

We first excluded columns containing fewer than three replicates for each molecular feature. Despite this reduction, missing values persisted, a common issue in LCMS studies^{62,63}. We assessed the nature of the missing data using Little's MCAR test and logistic regression to determine whether they were Missing Completely at Random (MCAR), Missing at Random (MAR), or Missing Not at Random (MNAR).

Missing Value Analysis and Imputation

We employed a diverse set of imputation techniques, including machine learning⁶⁴, neighbour-based methods, and matrix factorisation. In R, we used Random Forest (RF) (maxiter = 5, ntree = 50) and Predictive Mean Matching (PMM), while in Python, we applied Bayesian PCA, k-Nearest Neighbours (KNN), SVD, Gaussian Process Regression (GPR), and Expectation-Maximisation (EM). The performance of these methods was evaluated Earth Mover's Distance (EMD) and Hellinger Distance, alongside visual comparisons via Q-Q plots, ECDF plots, and KDE plots. Additionally, we calculated richness, Shannon entropy, Simpson's diversity index, and sparsity for datasets imputed with median and RF methods.

Outlier Detection and Removal

Post-imputation, we numerically encoded categorical variables and standardised molecular features. Seven outlier detection techniques were applied: Mahalanobis Distance, Isolation Forest, Elliptic Envelope, Robust PCA, Local Outlier Factor, Z-Score, and IQR. We assessed the effectiveness of these methods by quantifying outliers and adjusting contamination thresholds, with visualisation via PCA and t-SNE plots. Outliers identified were removed, and resulting NaN values were imputed using the validated Random Forest method (missForest package). The 30 most affected variables were plotted with standard deviation error bars to highlight changes after outlier removal and imputation.

Data Transformation and Variable Selection

Following outlier removal, we applied various data transformations: Log, Square Root, Box-Cox, Yeo-Johnson, Generalised Hyperbolic Sine (asinh) (using its standard mathematical definition, $\text{asinh}(x) = \ln(x + \sqrt{x^2 + 1})$), Generalised Logarithmic (glog), and Anscombe. We assessed their impact on data variability and distribution using Coefficient of Variation (CV), MA-transform (log ratio M), Relative Standard Deviation (RSD), and Relative Median Absolute Deviation (rMAD). Normality was evaluated using Shapiro-Wilk and Anderson-Darling tests, with density plots visualising the distribution of test statistics and p-values for each variable and transformation method.

After the asinh transformation, we refined the dataset by excluding variables with an rMAD exceeding 30%, a threshold set to eliminate highly variable molecular feature, and this clean data later used for subsequent analysis.

Spectral Measurement Methods

Spectral Measurement Protocol

Spectral reflectance measurements were obtained using an ASD FieldSpec 4 spectroradiometer (Malvern Panalytical Ltd, UK). Plants were measured inside a custom-built, matte-black imaging cube (90 cm side length) designed to minimise stray light and external interference. Uniform illumination was achieved using two 500 W halogen lamps (Arlec, Australia), and internal temperature was stabilised with a ceiling-mounted exhaust fan (Ventair Pty Ltd, Australia). Spectra were recorded from approximately 2×2 cm areas from the central region of the shoot and root system immediately after subsampling for LC-MS analysis. Optimal lens-to-sample distance was calculated based on the spectroradiometer's field-of-view (FOV) and the target sample dimensions to ensure consistent sampling geometry.

$$D = S / (2 * \tan(FOV/2))$$

Where:

- D: Distance between the spectroradiometer lens and sample surface (cm)
- S: Dimension of the side of the square sample area measured (e.g., 2 cm)
- FOV: Spectroradiometer lens field of view (degrees)

Imaging Cube and Sample Environment

The imaging cube (90 cm per side) was fabricated from MDF boards, with internal and external surfaces receiving four coats of matte-black paint to minimise internal reflections and stray light. A matte-black copy stand provided stable mounting for samples and held the pistol grip bare lens of the ASD FieldSpec 4 spectroradiometer. The cube opening was covered by a black blackout curtain to block external light during measurements. Two Arlec 500 W halogen work lights provided consistent illumination, while a Manrose inline exhaust fan (150 mm diameter) maintained internal temperature stability during measurements. Sample holders were custom-

made from aluminium sheets coated with four layers of matte-black paint to further minimise unwanted reflections.

Spectral Data Quality Assessment and Preprocessing

Data Overview

The initial dataset contained hyperspectral reflectance measurements for 336 plant samples, covering 2151 spectral bands ranging from 351 nm to 2500 nm.

Data Integrity and Outlier Screening

A comprehensive outlier screening was performed using multiple robust statistical methods: the Interquartile Range (IQR, using a 1.5x multiplier), Modified Z-score (threshold > 3.5), and Local Outlier Factor (LOF, n_neighbors=20, contamination='auto'). This multi-method approach aimed to identify diverse potential anomalies. Combining the samples flagged by any of these methods yielded 28 unique potential outlier samples (8.3% of the dataset, Fig. S8A). Following identification, these flagged samples underwent careful review. Considering the study's objective to capture a wide spectrum of osmotic stress responses across different genotypes and conditions, and the potential for these samples to represent genuine biological extremes rather than solely technical errors, the decision was made to retain all 336 samples for subsequent analyses. This strategy preserves the full observed variance and leverages the robustness of the downstream MOFA+ and Transformer analytical framework. A list of the 28 unique sample indices flagged during this screening process is provided for transparency (Table S8).

Signal Quality and Spectral Characteristics

Overall signal quality was assessed. The median standard deviation across all wavelengths was 0.080 (Mean=0.076), indicating generally low noise levels relative to the signal magnitude (Fig. S8B). The median approximate signal-to-noise ratio (Mean/STD per band) was 2.39. Spectral bands exhibited moderate-to-high inter-correlation (Mean Pearson correlation = 0.54), particularly within the visible, NIR, and SWIR regions, reflecting expected plant spectral features (Fig. S8C).

Normality Assessment

Normality of each spectral band was assessed using the Shapiro-Wilk test ($\alpha = 0.05$). Results showed that 90.7% (1950 out of 2150) of the spectral bands significantly deviated from a normal distribution (Fig. S8D). Kernel density estimates for representative bands confirmed non-Gaussian shapes (Fig. S8E). This widespread non-normality justifies the use of analytical methods, such as MOFA and potentially transformations within the deep learning framework, that are robust to distributional assumptions.

Smoothing Decision

Spectral smoothing is often applied to reduce noise but can obscure subtle spectral features relevant to biological stress. Given the observed high signal quality (Median STD = 0.080, Fig. S8B) and the demonstration of stable spectral shapes via derivative analysis (Fig. S8H, I), no smoothing filters were applied to the data to retain the maximum spectral information for the interpretable deep learning model.

Baseline Assessment and Correction Decision

Baseline stability and systematic trends were evaluated using non-parametric methods. Theil-Sen regression estimated the linear slope for each spectrum across the wavelength range. While the median slope was small (-2.17e-4), the distribution was narrow (IQR = 4.38e-5), and 100% of samples exhibited slopes exceeding a robust threshold (1.5 * IQR), indicating a subtle but consistent negative baseline trend across the dataset (Fig. S8F, G). First-derivative analysis (median absolute difference = 4.53e-4, not plotted) showed good point-to-point stability. Based on this assessment and to leverage the model's ability to handle such variations, no baseline correction was applied.

Derivative Analysis (Savitzky-Golay)

To further characterise spectral shape stability, first and second derivatives were computed using a Savitzky-Golay filter (window size=5, polynomial order=2). The median derivative spectra across all samples showed clear and consistent spectral features typical of vegetation, with

narrow Interquartile Range (IQR) bands indicating high reproducibility of spectral shapes (Fig. S8H, I). The median values of the per-sample median derivatives were low (1st: 5.56e-5, 2nd: -2.41e-6) with small IQRs, confirming minimal noise amplification and stable curvature characteristics, further supporting the decision not to apply smoothing.

Final Dataset

This pre-processed spectral dataset, comprising all 336 samples and 2151 features, formed the input for the data augmentation step prior to MOFA+ Transformer analysis.

Data Augmentation and Validation Strategy

Data Augmentation Strategy

To enhance model robustness and statistical power, the spectral and molecular datasets were expanded 8-fold while carefully preserving all experimental conditions and biological relationships. For the hyperspectral dataset (hyper_full_w.csv), augmentation was performed using spectral_augmentation.py, applying seven different techniques: Gaussian Process regression (GP), Spectral Mixup (MIX), Peak-Preserving Warp (WARP), Reflectance Scaling (SCALE), Band-Specific Noise addition (NOISE), Additive Mixup (ADD), and Multiplicative Mixup (MULT). Each augmented sample received a unique identifier indicating the method used (e.g., _GP, _WARP) for complete traceability. The leaf (n_p_l2.csv) and root (n_p_r2.csv) molecular datasets were augmented using molecular_feature_augmentation.py with two primary approaches: scaling existing features (SCALE method, 5 copies) and mixing features between samples (MIX method, 2 copies), with sample names encoding both method and copy number (e.g., _1_SCALE, _2_MIX).

Validation Framework and Results

The augmented datasets underwent extensive validation to ensure biological accuracy and analytical value, a process documented across seven comprehensive quality reports.

The augmented spectral data demonstrated high quality with an overall score of 87.25 out of 100 (SR1.html). Key validation findings included low outlier rates (<5.1%), high signal-to-noise ratios (>56 dB), and strong preservation of important vegetation indicators, including NDVI and Red

Edge Inflection Point. The data maintained excellent statistical similarity to the original measurements (Jensen-Shannon divergence ~ 0.11) and demonstrated outstanding functional equivalence in machine learning tests, with models achieving 108% and 113% of original performance in cross-dataset validation ([SR1.html](#), [SR2.html](#)).

The augmented molecular feature data for both leaf and root tissues also maintained strong quality, scoring 0.86 and 0.87 respectively ([SR3.html](#), [SR4.html](#)). These datasets exhibited acceptable outlier rates (<9%), excellent signal preservation (>97%), and a significant increase in statistical power for detecting biological differences. Notably, the SCALE method improved the ability to distinguish between treatments in leaves and genotypes in roots, while machine learning models retained over 72% of their original performance ([SR3.html](#), [SR4.html](#)).

To ensure suitability for integrated analysis, cross-modal consistency was validated and found to be excellent, with the augmented spectral and molecular datasets showing high compatibility (overall score: 0.99) ([SR5.html](#)). All experimental factors—treatments, genotypes, and batch effects—were perfectly preserved (score = 1.0), confirming that the augmented data maintained coherent biological patterns across different measurement types ([SR5.html](#)). Finally, a comprehensive statistical analysis confirmed that the augmented datasets closely matched their original counterparts in distribution patterns (preservation scores >0.8) while providing a substantial >3.5-fold increase in statistical power. The high consistency between different augmentation methods (uncertainty <0.001) confirmed the reliability of the approach ([SR6.html](#), [SR7.html](#)).

Validation Outcome Summary

This rigorous seven-stage validation process ([SR1.html](#) through [SR7.html](#)) confirmed that the augmentation strategy successfully: (1) preserved essential biological and spectral characteristics; (2) significantly improved statistical power for detecting biological differences; (3) maintained cross-modal consistency; and (4) created high-quality synthetic data suitable for advanced machine learning analysis. This thorough validation ensures confidence in all subsequent results and biological conclusions.

MOFA+ Feature Selection Stability Validation

To quantify feature stability, we performed bootstrap analysis by repeating the MOFA+ analysis 100 times ($n = 100$), each time sampling with replacement from the 1,344 augmented samples while maintaining constant model parameters. For each iteration, features were selected using a two-step criterion: first identifying ‘active’ latent factors ($>1\%$ total variance), then selecting features within these factors with absolute loading weights exceeding 0.1. Each feature’s ‘Selection Frequency’—the proportion of iterations meeting this criterion—provided a quantitative stability measure. The high average selection frequency of 96.0% confirms exceptional stability of the biological signals captured by our framework, indicating strong resilience to data perturbations.

Supplementary Note 2: LC-MS Data Preprocessing Results

Initial Data Cleaning Outcomes

Initial data filtering reduced the number of molecular features from 4,255 to 1,789 in one dataset and from 3,199 to 1,350 in another, by excluding entries with fewer than three replicates. Little’s MCAR test (P -value = 1.0) indicated that missing molecular feature values were not Missing Completely at Random. Logistic regression revealed significant associations between missingness probability and observed variables (genotype, treatment group, Time Point, and replicate; Fig. S7A), suggesting a Missing at Random (MAR) scenario. Heatmaps (Fig. S7B-C) illustrate missing value patterns in leaf and root datasets for negative columns, with missing data in green and non-missing in grey.

Missing Value Patterns and Imputation Performance

The stacked bar plot of Earth Mover’s Distance (Fig. S7D) indicated that median imputation performed well overall, closely followed by Random Forest (RF). Hellinger Distance (Fig. S7E) supported these findings. However, RF imputation outperformed median imputation in key metrics, with higher richness (168 vs. 156.9), Shannon entropy (5.124 vs. 4.999), and Simpson’s diversity index (0.994 vs. 0.992). RF also demonstrated lower sparsity (0.00595 vs. 0.0204), preserving more variability.

Graphical comparisons, including Q-Q plots (Fig. S7F, left panel), ECDF, and KDE plots (Fig. S7F, middle and right panels), showed that RF imputation more closely aligned with the original data distribution. Based on these visual and statistical evaluations, RF was selected as the optimal imputation method.

Outlier Analysis Results

Isolation Forest emerged as the most effective outlier detection method, as visualised through PCA and t-SNE plots (Fig. S7G, H) and supported by the analysis of the 30 most impacted variables. This method identified outliers with the highest precision, leading to its selection for further analysis.

Transformation Effects

The asinh transformation most effectively reduced data variability, decreasing CV from 0.876 to 0.206 (Fig. S7I), rMAD from 67.98 to 14.28 (Fig. S7J), and RSD from 87.36% to 20.58% (Fig. S7K). MA-transformation plots confirmed tighter data clustering (Fig. S7L, M). Normality improved moderately across transformations, with asinh offering the best balance (Fig. S7N-Q). However, most variables remained non-parametric, guiding subsequent analysis choices.

To further refine the dataset, an rMAD-based selection process removed 241 variables (9.75%) with rMAD >30%, retaining 2,230 variables. This step enhanced reliability by excluding high-variability Molecular feature, providing a robust foundation for analysing metabolomic responses to osmotic stress.

(In negative mode, leaf data contains 807 molecular features, root data contains 982, with 391 features common to both, resulting in a combined total of 1,398 molecular features. In positive mode, leaf data contains 611 molecular features, root data contains 739, with 277 features common to both, resulting in a combined total of 1,073 molecular features. Overall, the combined dataset includes 2,471 molecular features.)

References

- 1 Baysoy, A., Bai, Z., Satija, R. & Fan, R. The technological landscape and applications of single-cell multi-omics. *Nat Rev Mol Cell Biol* **24**, 695-713 (2023). <https://doi.org/10.1038/s41580-023-00615-w>
- 2 Hasin, Y., Seldin, M. & Lusis, A. Multi-omics approaches to disease. *Genome Biology* **18**, 83 (2017). <https://doi.org/10.1186/s13059-017-1215-1>
- 3 Cao, Z.-J. & Gao, G. Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nature Biotechnology* **40**, 1458-1466 (2022). <https://doi.org/10.1038/s41587-022-01284-4>
- 4 Sapoval, N. et al. Current progress and open challenges for applying deep learning across the biosciences. *Nature Communications* **13** (2022). <https://doi.org/10.1038/s41467-022-29268-7>
- 5 Hu, Y. et al. Benchmarking algorithms for single-cell multi-omics prediction and integration. *Nat Methods* **21**, 2182-2194 (2024). <https://doi.org/10.1038/s41592-024-02429-w>
- 6 Cantini, L. et al. Benchmarking joint multi-omics dimensionality reduction approaches for the study of cancer. *Nature Communications* **12**, 124 (2021). <https://doi.org/10.1038/s41467-020-20430-7>
- 7 Ma, A. et al. Single-cell biological network inference using a heterogeneous graph transformer. *Nat Commun* **14**, 964 (2023). <https://doi.org/10.1038/s41467-023-36559-0>
- 8 Schäfer, P. S. L., Dimitrov, D., Villalblanca, E. J. & Saez-Rodriguez, J. Integrating single-cell multi-omics and prior biological knowledge for a functional characterization of the immune system. *Nat Immunol* **25**, 405-417 (2024). <https://doi.org/10.1038/s41590-024-01768-2>
- 9 Yuan, Q. & Duren, Z. Inferring gene regulatory networks from single-cell multiome data using atlas-scale external data. *Nat Biotechnol* **43**, 247-257 (2025). <https://doi.org/10.1038/s41587-024-02182-7>

- 10 Vandereyken, K., Sifrim, A., Thienpont, B. & Voet, T. Methods and applications for single-cell and spatial multi-omics. *Nat Rev Genet* **24**, 494-515 (2023). <https://doi.org/10.1038/s41576-023-00580-2>
- 11 Argelaguet, R. et al. MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biology* **21**, 111 (2020). <https://doi.org/10.1186/s13059-020-02015-1>
- 12 Argelaguet, R. et al. Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. *Molecular Systems Biology* **14**, e8124 (2018). <https://doi.org/https://doi.org/10.15252/msb.20178124>
- 13 Toubiana, D. & Maruenda, H. Guidelines for correlation coefficient threshold settings in metabolite correlation networks exemplified on a potato association panel. *BMC Bioinformatics* **22**, 116 (2021). <https://doi.org/10.1186/s12859-021-03994-z>
- 14 Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* **4**, Article17 (2005). <https://doi.org/10.2202/1544-6115.1128>
- 15 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008). <https://doi.org/10.1186/1471-2105-9-559>
- 16 Rohart, F., Gautier, B., Singh, A. & Lê Cao, K.-A. mixOmics: An R package for ‘omics feature selection and multiple data integration. *PLoS computational biology* **13**, e1005752 (2017).
- 17 Subramanian, I., Verma, S., Kumar, S., Jere, A. & Anamika, K. Multi-omics Data Integration, Interpretation, and Its Application. *Bioinformatics and Biology Insights* **14**, 1177932219899051 (2020). <https://doi.org/10.1177/1177932219899051>
- 18 Singh, A. et al. DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays. *Bioinformatics* **35**, 3055-3062 (2019). <https://doi.org/10.1093/bioinformatics/bty1054>

- 19 Libbrecht, M. W. & Noble, W. S. Machine learning applications in genetics and genomics. *Nature Reviews Genetics* **16**, 321-332 (2015). <https://doi.org/10.1038/nrg3920>
- 20 Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence* **1**, 206-215 (2019). <https://doi.org/10.1038/s42256-019-0048-x>
- 21 Lock, E. F., Hoadley, K. A., Marron, J. S. & Nobel, A. B. JOINT AND INDIVIDUAL VARIATION EXPLAINED (JIVE) FOR INTEGRATED ANALYSIS OF MULTIPLE DATA TYPES. *Ann Appl Stat* **7**, 523-542 (2013). <https://doi.org/10.1214/12-aos597>
- 22 Shen, R., Olshen, A. B. & Ladanyi, M. Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. *Bioinformatics* **25**, 2906-2912 (2009). <https://doi.org/10.1093/bioinformatics/btp543>
- 23 Vaswani, A. et al. Attention is all you need. *Advances in neural information processing systems* **30** (2017).
- 24 Tsai, Y.-H. H. et al. in *Proceedings of the conference. Association for computational linguistics. Meeting*. 6558.
- 25 Jiang, J. et al. Transformer technology in molecular science. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **14**, e1725 (2024).
- 26 Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021). <https://doi.org/10.1038/s41586-021-03819-2>
- 27 Xu, X., Fonseca de Lima, C. F., Vu, L. D. & De Smet, I. When drought meets heat—a plant omics perspective. *Frontiers in Plant Science* **14**, 1250878 (2023).
- 28 Ramírez-González, R. et al. The transcriptional landscape of polyploid wheat. *Science* **361**, eaar6089 (2018).
- 29 Velten, B. & Stegle, O. Principles and challenges of modeling temporal and spatial omics data. *Nature Methods* **20**, 1462-1474 (2023). <https://doi.org/10.1038/s41592-023-01992-y>
- 30 Kaur, S. et al. Temporal ordering of omics and multiomic events inferred from time-series data. *npj Systems Biology and Applications* **6**, 22 (2020). <https://doi.org/10.1038/s41540-020-0141-0>

- 31 Klec, C., Prinz, F. & Pichler, M. Involvement of the long noncoding RNA NEAT1 in carcinogenesis. *Mol Oncol* **13**, 46-60 (2019). <https://doi.org/10.1002/1878-0261.12404>
- 32 Gupta, A., Rico-Medina, A. & Cano-Delgado, A. I. The physiology of plant responses to drought. *Science* **368**, 266-269 (2020). <https://doi.org/10.1126/science.aaz7614>
- 33 Zhang, H., Zhu, J., Gong, Z. & Zhu, J.-K. Abiotic stress responses in plants. *Nature Reviews Genetics* (2021). <https://doi.org/10.1038/s41576-021-00413-0>
- 34 Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. *Advances in neural information processing systems* **30** (2017).
- 35 Hirose, T. et al. NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies. *Mol Biol Cell* **25**, 169-183 (2014). <https://doi.org/10.1091/mbc.E13-09-0558>
- 36 Cramer, G. R., Urano, K., Delrot, S., Pezzotti, M. & Shinozaki, K. Effects of abiotic stress on plants: a systems biology perspective. *BMC Plant Biology* **11**, 163 (2011). <https://doi.org/10.1186/1471-2229-11-163>
- 37 Shinozaki, K. & Yamaguchi-Shinozaki, K. Gene networks involved in drought stress response and tolerance. *J Exp Bot* **58**, 221-227 (2007). <https://doi.org/10.1093/jxb/erl164>
- 38 Skirycz, A. et al. Survival and growth of Arabidopsis plants given limited water are not equal. *Nature Biotechnology* **29**, 212-214 (2011). <https://doi.org/10.1038/nbt.1800>
- 39 Todaka, D., Shinozaki, K. & Yamaguchi-Shinozaki, K. Recent advances in the dissection of drought-stress regulatory networks and strategies for development of drought-tolerant transgenic rice plants. *Frontiers in plant science* **6**, 84 (2015).
- 40 Camacho, D. M., Collins, K. M., Powers, R. K., Costello, J. C. & Collins, J. J. Next-Generation Machine Learning for Biological Networks. *Cell* **173**, 1581-1592 (2018). <https://doi.org/https://doi.org/10.1016/j.cell.2018.05.015>
- 41 Curran, P. J. Remote sensing of foliar chemistry. *Remote sensing of environment* **30**, 271-278 (1989).

- 42 Rewald, B. & Meinen, C. Plant roots and spectroscopic methods – analyzing species, biomass and vitality. *Frontiers in Plant Science* **4** (2013). <https://doi.org/10.3389/fpls.2013.00393>
- 43 Fait, A., Batushansky, A., Shrestha, V., Yobi, A. & Angelovici, R. Can metabolic tightening and expansion of co-expression network play a role in stress response and tolerance? *Plant Science* **293**, 110409 (2020). <https://doi.org/10.1016/j.plantsci.2020.110409>
- 44 Toubiana, D., Fernie, A. R., Nikoloski, Z. & Fait, A. Network analysis: tackling complex data to study plant metabolism. *Trends in biotechnology* **31**, 29-36 (2013).
- 45 Gitelson, A. A., Gritz, Y. & Merzlyak, M. N. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *J Plant Physiol* **160**, 271-282 (2003). <https://doi.org/10.1078/0176-1617-00887>
- 46 Sims, D. A. & Gamon, J. A. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of Environment* **81**, 337-354 (2002). [https://doi.org/https://doi.org/10.1016/S0034-4257\(02\)00010-X](https://doi.org/https://doi.org/10.1016/S0034-4257(02)00010-X)
- 47 Cui, H. et al. scGPT: toward building a foundation model for single-cell multi-omics using generative AI. *Nature Methods* **21**, 1470-1480 (2024). <https://doi.org/10.1038/s41592-024-02201-0>
- 48 Sumner, L. W. et al. Proposed minimum reporting standards for chemical analysis. *Metabolomics* **3**, 211-221 (2007). <https://doi.org/10.1007/s11306-007-0082-2>
- 49 Salek, R. M., Steinbeck, C., Viant, M. R., Goodacre, R. & Dunn, W. B. The role of reporting standards for metabolite annotation and identification in metabolomic studies. *GigaScience* **2** (2013). <https://doi.org/10.1186/2047-217x-2-13>
- 50 Schölkopf, B. et al. Toward causal representation learning. *Proceedings of the IEEE* **109**, 612-634 (2021). <https://doi.org/10.48550/arXiv.2102.11107>
- 51 Tardieu, F., Simonneau, T. & Muller, B. The Physiological Basis of Drought Tolerance in Crop Plants: A Scenario-Dependent Probabilistic Approach. *Annual*

- Review of Plant Biology* **69**, 733-759 (2018). <https://doi.org/10.1146/annurev-arplant-042817-040218>
- 52 Bennani, S., Nsarellah, N., Jlibene, M. & Tadesse, W. Efficiency of drought tolerance indices under different stress severities for bread wheat selection. *Australian Journal of Crop Science* **11**, 395-405 (2017). <https://doi.org/10.21475/ajcs.17.11.04.pne272>
- 53 Hone, H. et al. Profiling, isolation and characterisation of beneficial microbes from the seed microbiomes of drought tolerant wheat. *Scientific Reports* **11**, 11916 (2021). <https://doi.org/10.1038/s41598-021-91351-8>
- 54 Yadav, A. K., Carroll, A. J., Estavillo, G. M., Rebetzke, G. J. & Pogson, B. J. Wheat drought tolerance in the field is predicted by amino acid responses to glasshouse-imposed drought. *Journal of Experimental Botany* **70**, 4931-4948 (2019). <https://doi.org/10.1093/jxb/erz224>
- 55 Roessner, U. et al. Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell* **13**, 11-29 (2001). <https://doi.org/10.1105/tpc.13.1.11>
- 56 Rochfort, S. J., Trenerry, V. C., Imsic, M., Panizzo, J. & Jones, R. Class targeted metabolomics: ESI ion trap screening methods for glucosinolates based on MSn fragmentation. *Phytochemistry* **69**, 1671-1679 (2008).
- 57 Dunn, W. B. et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nature Protocols* **6**, 1060-1083 (2011). <https://doi.org/10.1038/nprot.2011.335>
- 58 Gika, H. G., Theodoridis, G. A., Wingate, J. E. & Wilson, I. D. Within-day reproducibility of an HPLC-MS-based method for metabonomic analysis: application to human urine. *J Proteome Res* **6**, 3291-3303 (2007). <https://doi.org/10.1021/pr070183p>
- 59 Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
- 60 Xie, Y. et al. Integration of hyperspectral imaging and transcriptomics from individual cells with HyperSeq. *bioRxiv*, 2024.2001.2027.577536 (2024).

- 61 North, B. V., Curtis, D. & Sham, P. C. A Note on the Calculation of Empirical
P Values from Monte Carlo Procedures. *The American Journal of
Human Genetics* **71**, 439-441 (2002). <https://doi.org/10.1086/341527>
- 62 Karpievitch, Y. V., Dabney, A. R. & Smith, R. D. Normalization and missing value
imputation for label-free LC-MS analysis. *BMC Bioinformatics* **13**, S5 (2012).
<https://doi.org/10.1186/1471-2105-13-s16-s5>
- 63 Kokla, M., Virtanen, J., Kolehmainen, M., Paananen, J. & Hanhineva, K. Random
forest-based imputation outperforms other methods for imputing LC-MS
metabolomics data: a comparative study. *BMC Bioinformatics* **20** (2019).
<https://doi.org/10.1186/s12859-019-3110-0>
- 64 Wei, R. et al. Missing value imputation approach for mass spectrometry-based
metabolomics data. *Scientific reports* **8**, 663 (2018).