Lawrence Berkeley National Laboratory

Title

Information theory and machine learning illuminate large-scale metabolomic responses of Brachypodium distachyon to environmental change

Permalink

https://escholarship.org/uc/item/4pq369bd

Journal

The Plant Journal, 114(3)

ISSN

0960-7412

Authors

Mahood, Elizabeth H
Bennett, Alexandra A
Komatsu, Karyn
et al.

Publication Date

2023-05-01

DOI

10.1111/tpj.16160

Copyright Information

Peer reviewed

Information theory and machine learning illuminate large-scale metabolomic responses of *Brachypodium distachyon* to environmental change

Elizabeth H. Mahood[1,†] iD, Alexandra A. Bennett[1,‡] iD, Karyn Komatsu[2], Lars H. Kruse[1,¶] iD, Vincent Lau[2], Maryam Rahmati Ishka[1,3] iD, Yulin Jiang[4] iD, Armando Bravo[3,§], Katherine Louie[5,6] iD, Benjamin P. Bowen[5,6] iD, Maria J. Harrison[3], Nicholas J. Provart[2] iD, Olena K. Vatamaniuk[4] iD and Gaurav D. Moghe[1,*] iD

[1]*Plant Biology Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, USA,*
[2]*Department of Cell and Systems Biology, University of Toronto, Toronto, Canada,*
[3]*Boyce Thompson Institute, Ithaca, NY, USA,*
[4]*Soil and Crop Sciences Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, USA,*
[5]*Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA,*
[6]*Lawrence Berkeley National Laboratory, Department of Energy Joint Genome Institute, Berkeley, CA, USA*

SUMMARY

Plant responses to environmental change are mediated via changes in cellular metabolomes. However, <5% of signals obtained from liquid chromatography tandem mass spectrometry (LC-MS/MS) can be identified, limiting our understanding of how metabolomes change under biotic/abiotic stress. To address this challenge, we performed untargeted LC-MS/MS of leaves, roots, and other organs of *Brachypodium distachyon* (Poaceae) under 17 organ–condition combinations, including copper deficiency, heat stress, low phosphate, and arbuscular mycorrhizal symbiosis. We found that both leaf and root metabolomes were significantly affected by the growth medium. Leaf metabolomes were more diverse than root metabolomes, but the latter were more specialized and more responsive to environmental change. We found that 1 week of copper deficiency shielded the root, but not the leaf metabolome, from perturbation due to heat stress. Machine learning (ML)-based analysis annotated approximately 81% of the fragmented peaks versus approximately 6% using spectral matches alone. We performed one of the most extensive validations of ML-based peak annotations in plants using thousands of authentic standards, and analyzed approximately 37% of the annotated peaks based on these assessments. Analyzing responsiveness of each predicted metabolite class to environmental change revealed significant perturbations of glycerophospholipids, sphingolipids, and flavonoids. Co-accumulation analysis further identified condition-specific biomarkers. To make these results accessible, we developed a visualization platform on the Bio-Analytic Resource for Plant Biology website (https://bar.utoronto.ca/efp_brachypodium_metabolites/cgi-bin/efpWeb.cgi), where perturbed metabolite classes can be readily visualized. Overall, our study illustrates how emerging chemoinformatic methods can be applied to reveal novel insights into the dynamic plant metabolome and stress adaptation.

Keywords: computational biology, metabolomics, mass spectrometry, abiotic stress, mycorrhizal symbiosis, Brachypodium.

Linked article: This paper is the subject of a Research Highlight article. To view this Research Highlight article visit https://doi.org/10.1111/tpj.16243.

INTRODUCTION

The central dogma of molecular biology extends from genes to transcripts to proteins. These proteins, however, exert an effect on the phenotype eventually through altering metabolites. Agronomically important traits such as yield, nutritional quality, flavor characteristics, and stress

response are all controlled by underlying metabolic pathways. A revolution in sequencing over the past decade has provided unparalleled insights into the transcriptomic and epigenomic perturbations due to genotypic and environmental changes, yet the global metabolome largely remains a black box, primarily due to our inability to identify compounds from metabolomics data (Chaleckis et al., 2019; Salem et al., 2020). It is estimated that over a million compounds are produced across the plant kingdom (Afendi et al., 2012), with individual plants producing thousands of metabolites (Fernie, 2007). However, <5% of these signals can be annotated using spectral matching (da Silva & Dorrestein, 2015). Thus, patterns of global metabolomic changes still remain unknown despite the importance metabolites have to plant fitness and human society.

To assess metabolomic changes due to genetic variation, developmental progression, and environmental changes, gas chromatography mass spectrometry (GC-MS) and liquid chromatography mass spectrometry (LC-MS) remain the workhorse approaches, with LC-MS typically detecting a much broader set of the metabolome. Although diverse algorithmic innovations have aided in metabolome assessments (Brouard et al., 2016; Dührkop et al., 2019; Schymanski et al., 2017; Tsugawa et al., 2016), LC-MS peaks are primarily annotated using MS/MS spectral matching with entries from public databases (Guijas et al., 2018; Horai et al., 2010; Wang et al., 2016), referred to as 'Level 2' annotations based on Metabolomics Standards Initiative (MSI) guidelines (Sumner et al., 2007). While correct predictions are indeed obtained in this manner, plant-derived compounds are underrepresented in public databases (Fukushima & Kusano, 2013; Shahaf et al., 2016), which results in incorrect hits among the limited numbers of compounds identified. Partly due to this

limitation, many LC-MS-based studies are targeted or semi-targeted, and end up analyzing a small but identifiable portion of the metabolome (Bromke et al., 2015; Itkin et al., 2013; Okazaki et al., 2013; Šimura et al., 2018; Tohge et al., 2016). This strategy produces robust insights, but global shifts in the metabolome and their genetic drivers cannot be assessed via targeted studies. Identifying such patterns can provide novel insights into metabolic plasticity and plant responses to stress conditions, which are important for addressing challenges of agricultural productivity due to climate change, overpopulation, and degrading soil quality.

In recent years, two important resources have emerged for the analysis of global untargeted tandem LC-MS (LC-MS/MS) data. First, the machine learning (ML)-based tool CANOPUS (Dührkop et al., 2021) enables prediction of metabolite structural classes based on the MS/MS spectrum (an MSI Level 3 annotation), providing novel insights into the metabolome composition. For example, even if specific compounds are not identified, recognizing that 'flavonoids' increase in abundance under UV stress provides significant biological insights into the plant's stress response. Second, independent of compound annotation, approaches adapted from information theory can inform about the gross and/or specific shifts in plant metabolomes (Li et al., 2020; Zu et al., 2020). In this study, we combine these two approaches with co-accumulation analysis to illuminate global changes in plant metabolomes under different conditions.

Specifically, we assessed the metabolome of *Brachypodium distachyon* (Brachypodium) under different conditions (Figure 1). Brachypodium is a model C3 grass species in the Poaceae family that shared a common ancestor with rice (*Oryza sativa*) approximately 50 million years ago and Triticeae (wheat [*Triticum aestivum*], barley [*Hordeum*
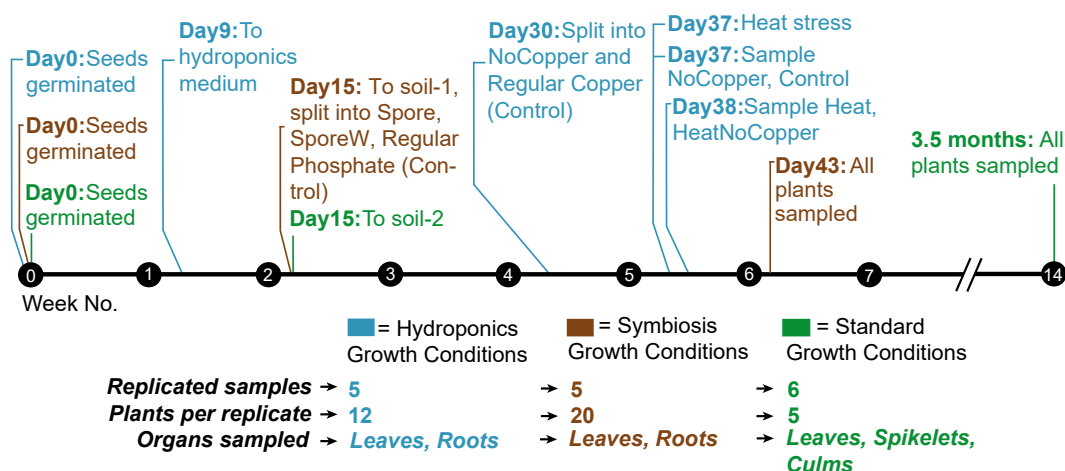
Figure 1. Timeline and schematic of the experimental design. The number of samples, the number of plants per replicate, and the organs sampled for each set of growth conditions are shown, along with the timeline of important events such as treatment induction and harvesting. Divergent growth and stress conditions were chosen to induce variability in metabolic profiles. Days are counted post-germination. Soil-1 and soil-2 refer to different soil mixes. The germination protocol for hydroponics seeds was distinct from the germination protocol for the other growth conditions (see Methods S1).

vulgare]) approximately 35 million years ago (Charles et al., 2009). The short stature of Brachypodium and its fast growth cycle make the species a convenient model for understanding not only Poaceae biology but also for biofuel research (Brkljacic et al., 2011; Douché et al., 2013; Le Bris et al., 2019; Marriott et al., 2014). The main goals of this study were to: (i) assess Brachypodium's metabolomic reconfigurations across different organs and environmental conditions, (ii) identify the metabolite classes most perturbed by different stresses, (iii) discover condition-specific metabolites that may serve as stress biomarkers, and (iv) establish a platform for visualization of the global metabolome changes. Towards these goals, we first performed LC-MS/MS from 17 different organ–condition combinations, including agriculturally relevant conditions such as copper deficiency, heat stress, low phosphate, and arbuscular mycorrhizal symbiosis (AMS). We used CANOPUS and information theory-derived metrics to compare control versus test metabolomes across different organs and to characterize additional metabolome changes through co-accumulation modules and biomarker detection. Finally, these changes were visualized using a novel representation on the Bio-Analytic Resource for Plant Biology (BAR) website (available at: https://bar.utoronto.ca/efp_brachypodium_metabolites/cgi-bin/efpWeb.cgi). Overall, our findings help illuminate a much larger proportion of the metabolome captured by LC-MS methods and provide new insights on metabolic perturbations in Brachypodium under different conditions.

RESULTS

Experimental design and pre-processing of metabolome data

Brachypodium plants were grown to different ages and under different growth conditions with the goal of producing significant metabolome perturbations. Roots, leaves (young and mature combined), and in some cases, culms and spikelets were sampled. Overall, 17 organ–condition combinations were sampled, with plants grown across three major regimens: hydroponics (Hydro), AMS (Sym), and tissue (Tis) (Figure 1, Figure S1). Hydro treatments consisted of regular Cu (Control), Cu deficiency (NoCopper), heat stress (Heat), and heat stress under Cu deficiency (HeatNoCopper). The Sym treatments consisted of plants grown with regular amounts of phosphate fertilization (Control), low phosphate-treated plants inoculated with a wash solution of *Rhizophagus irregularis* spore growth medium (SporeW, not containing any spores, i.e., mock treatment), and low phosphate-treated plants inoculated with *R. irregularis* spores (Spore). The Tis regimen involved growing plants in regular soil until maturity; leaves, culms, and spikelets were sampled in this regime. The effectiveness of the copper deficiency treatment and

presence of colonization were verified through reverse transcriptase-PCR (RT-PCR) of copper deficiency and fungal symbiosis marker genes, respectively (Figure S2). All samples were analyzed via LC-MS/MS in both positive and negative mode to obtain a comprehensive snapshot of their metabolome and to determine common versus different patterns in each mode.

After peak deconvolution and alignment, peak values were filtered using a sequence of steps (Figures S3 and S4). To enable comparisons between different LC-MS runs, we first tested five different data normalization approaches (File S1) and selected variance-stabilized normalization (VSN) as the most appropriate based on its performance (Table S1; File S1). Data imputation was also performed to fill in values lost due to Orbitrap LC-MS detection limits. To ensure that either step does not alter the overall underlying data structure, we first determined the effect of performing imputation before versus after normalization using a dummy dataset where actual peak areas were randomly replaced by zeros. The degree of error in normalization–imputation and imputation–normalization was quantified. Overall, both normalization orders had almost identical errors (Figure S5). Thus, given precedence (Chong et al., 2019; Mock et al., 2018), we first imputed peak areas using k-nearest neighbor and normalized the imputed areas using VSN for further downstream analyses. The final peaks also had very low influence of in-source fragments (Methods S1), and therefore their quality was considered good for downstream analyses (also see Discussion).

VSN maximized correlations among replicates while maintaining low correlations between different treatment groups (File S2). The aboveground organs were found to have more peaks as well as a higher total peak abundance than the roots (Figure S6; File S3). The largest numbers of metabolite signals in both organs were observed in Sym samples, indicating that growth media also influenced the Brachypodium metabolome. The high numbers of peaks seen in the Sym Spore root samples likely include metabolites of fungal origin. Correlations between leaf and root and between control and treatment were lower than among replicates (File S2), putatively identifying two other axes of metabolomic divergence between samples, namely organs and conditions. To investigate these further, we first performed a global assessment of similarities and differences between the metabolomes under different conditions.

The root metabolome is less diverse but more specialized and more perturbed than the leaf metabolome

Using the normalized, imputed datasets, we quantified the impact of each stress on the root and leaf metabolomes. Principal component analysis (PCA) identified the organs and the growth media as stronger drivers of metabolic variation in our samples than the stresses. While PC1, explaining 46.88 and 45.6% of the metabolic variation between

samples in positive and negative mode, respectively, was indicative of organ-wise differences, PC2 (12.97 and 13.77%, respectively) revealed a substantial impact of the growth medium (soil type, hydroponics) on both root and leaf metabolomes (Figure S7). PCA as well as hierarchical clustering (Figures S8 and S9) validated close clustering of replicate samples and highlighted a set-wise impact of stresses. For the Hydro set, NoCopper clustered with Control in both leaves and roots, while for the Sym set, SporeW was the more impactful condition for leaves and Spore for the roots. HeatNoCopper clustered closer to Heat than NoCopper in both roots and leaves, indicating that the majority of metabolomic differences in this combined stress was due to heat stress. When PCAs were differentiated by organs (Figure S7b,c,e,f), the effect of different stresses could be observed. Overall, the leaf metabolomes appeared less impacted by the stresses than root metabolomes.

To further quantify the impact of each stress on the overall sampled metabolome, we used two information theory-based measures – diversity (H) and specialization (δ, measuring uniqueness/differentiation) – as well as the Relative Distance Plasticity Index (RDPI, measuring overall perturbation including up- and downregulation; see Experimental Procedures and Li et al., 2020 for explanation of metrics). We first assessed the metabolome differences in non-stress conditions. More peaks as well as more uniformity in the peak areas can increase diversity; thus, given leaves consistently have more peaks than roots, culms, and spikelets (Figure S6), their diversity is the highest (Figure 2a,b; Figure S10a,c). However, roots and spikelets are more metabolically specialized. The degree of specialization and to some extent diversity were clearly dependent on the growth medium and stress (Figure 2a,b; Figure S10). Roots were more specialized in the hydroponic medium (except Sym Spore root) but leaf metabolomes were more specialized in the soil and sand:gravel growth media (Figure S10b,d). Intriguingly, the observation of spikelets having high specialization is congruent with a similar observation in *Nicotiana attenuata* anthers (Li, Heiling, et al., 2016), indicating that the metabolic uniqueness of the reproductive tissues may be a conserved trait across monocots and dicots.

Although differences in specialization and diversity among leaf metabolomes were low, many stresses elicited statistically significant changes (Kolmogorov–Smirnov [KS] test, Table S2). Overall, the stresses appeared to disrupt foliar metabolism far less than that of the roots – especially for leaves from hydroponically grown plants – as indicated by tight clustering of leaf stresses with their controls. In positive mode (Figure 2a), specialization cleanly separated out leaf samples into their growth conditions, but this was not seen in negative mode (Figure 2b), and in both ionization modes, leaf samples had relatively low specialization.

Taken together with the relatively low RDPI values observed for leaf samples (Figure 2c,d), these results indicate that the leaf metabolome is more robust/less responsive to tested environmental changes than the root metabolome.

In contrast, the specialization and RDPI of roots were significantly influenced by stress. In both ionization modes, we found that roots had higher RDPIs (i.e., greater metabolome perturbation) than leaves (except for SporeW, in which leaves had similar RDPIs to roots in negative mode) (Figure 2c,d). Hydro roots had a higher baseline (Control) specialization than Sym (Figure S10b,d), indicating the presence of peaks specific to the hydroponics experimental regimen. However, in both ionization modes, Heat roots and Spore roots had the highest specialization and RDPI. Specialization is a sum of the 'degree of specificity' of each metabolite signal across the different conditions; thus, high specialization in Heat and Spore indicates a greater representation of metabolites that are uniquely changing under these conditions alone. Interestingly, specialization of the HeatNoCopper roots was similar to that of Control roots (Figure 2a,b), while its RDPI was intermediate between those of NoCopper and Heat (Figure 2c,d). These observations suggest that the impact of heat stress on the global root metabolome was less drastic under copper deficiency, which was contradictory to our expectation that HeatNoCopper roots would show a greater perturbation than Heat roots given a combination of two stresses (see Discussion).

To obtain a more granular understanding of the overall induced metabolites, differentially accumulated peaks (DAPs) were estimated in each condition based on false discovery rate (FDR)-corrected *P*-values and fold change criteria (see Methods S1; File S4). The pattern of differential accumulation was similar between positive and negative modes (Figure 2e,f). We found that HeatNoCopper and Heat had a high number of DAPs primarily in the roots (Figure 2e,f; Figures S11–S13) mirroring the RDPI metric. Over 200 metabolites were also perturbed under AMS in positive as well as negative mode; however, many of these metabolites could be of fungal origin. Heat and Spore roots had both the highest numbers of DAPs and unique DAPs, consistent with the finding that they have high RDPI and the highest specialization.

Validation of the structural annotation tool CANOPUS using orthogonal approaches

The above analyses revealed global patterns of change in the Brachypodium metabolome under changing environments. We next sought to understand shifts in specific metabolite classes. While untargeted LC-MS is the method of choice for detecting a diverse range of metabolites, identifying these peaks is a major challenge. We initially performed spectral matching using public databases, yet,
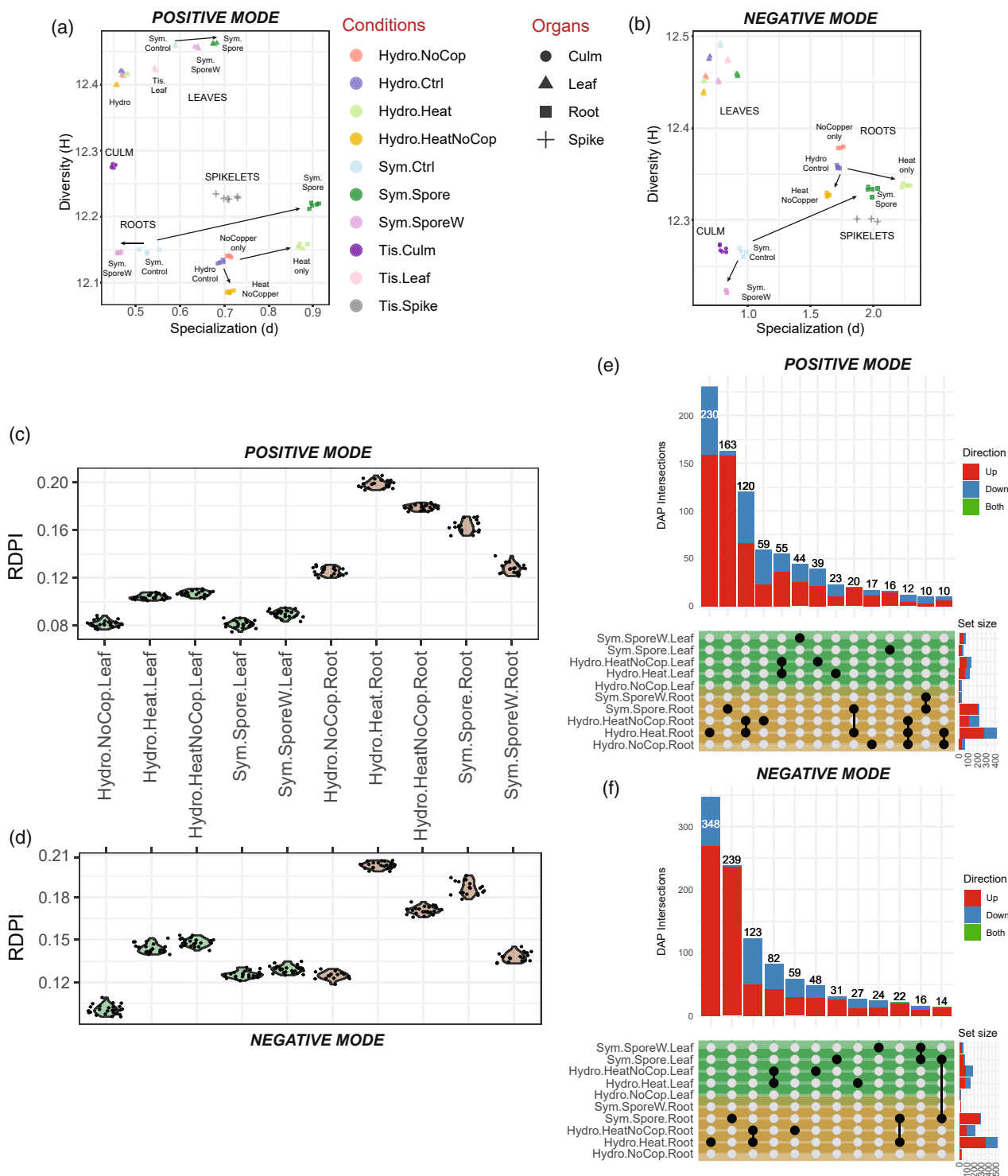
Figure 2. Comparison of metabolomic perturbations among conditions. (a, b) Diversity versus specialization per condition, with organs depicted as different shapes and conditions as different colors. Annotations are added onto these plots for ease of interpretation. (c, d) RDPI per stress condition. (e, f) Upset plots of differentially abundant peaks (DAPs) per stress condition, inclusive of up- and downregulated peaks. Intersections (vertical bars) depict the number of DAPs common between sets of conditions. Only intersections with at least 10 peaks are shown. Bar color represents direction of regulation: up, down, or both (for peaks in intersections of at least two conditions). (a, c, e) Positive mode. (b, d, f) Negative mode.

as expected, this provided structural information for only a small percentage of queried metabolites (6.1% in positive mode, 5.2% in negative mode). Such low degree of identification precludes assessment of how the broader metabolome is changing under the tested conditions. Thus, to annotate a broader proportion of the detected metabolites, we utilized CANOPUS, a deep learning-based method in the SIRIUS4 software which provides database-free predictions of metabolite structural classes corresponding to the MSI Level 3 identifications (Sumner et al., 2007). CANOPUS classifies compounds into the multilabel and hierarchical ChemOnt ontology (Djoumbou Feunang et al., 2016), which is similar to the Gene Ontology (GO) for genes (The Gene Ontology Consortium, 2019). As ChemOnt is multilabel, peaks may receive multiple annotations at each level; however, the classifications we report are of each peak's largest substructure. Of the 3582 and 2996 fragmented peaks in positive and negative mode, 2931 (82%) and 2409 (80%) were annotated by CANOPUS at the Superclass level with posterior probability > 0.5 (Figure S14). Of the 26 Superclasses existing for organic compounds, 14 and 12 were represented in the positive and negative mode data, respectively (Files S5 and S6) before additional filtering, with lipids and lipid-like molecules having the most peaks in both ionization modes.

Given CANOPUS is a relatively new predictive method, we extensively assessed the accuracy of these annotations using four independent approaches: (i) using the aforementioned spectral matches with public databases, (ii) assessing co-clustering in MS/MS molecular networks, (iii) comparing CANOPUS predictions with lipid predictions from MS-DIAL, and (iv) assessing CANOPUS predictions of thousands of authentic standards run on the same instrument.

First, for each compound identified via spectral database hits, we compared their ChemOnt classes to CANOPUS' predictions (File S7a–c). At each level, we calculated misannotations as the percent of peaks identified using spectral matches that were not given the same annotation by CANOPUS. At the Superclass level, we observed good correspondence between CANOPUS classifications and database identifications in both modes. The median CANOPUS misannotation rates at the Class level, when considering correct Classes as determined by ClassyFire, were 54.4 and 28.2% in positive and negative mode (File S7a–c), respectively, indicating that overall CANOPUS predicted Classes well for negative mode only. For example, of the 15 compounds in negative mode that were identified as 'flavonoids' via database searches and also had a CANOPUS Class annotation, 13 (86.7%) were correctly called 'flavonoids' by CANOPUS (File S7). This proportion was only 46.7% in positive mode, although 86.7% were assigned to the correct Superclass ('phenylpropanoids and polyketides'). In positive mode, the most frequently

misannotated Classes were glycerophospholipids (GPs; 65.57% of CANOPUS-predicted GPs were misannotated) and phenols (73.68% misannotated), again most of the misannotations being within the same Superclass (70 and 50%, respectively). The decrease in agreement between positive mode Superclasses and Classes is largely due to the high misclassification rate of GPs and their high presence (24%) in the identified positive mode compounds. However, these results indicated that CANOPUS predictions on negative mode data were more accurate than positive mode data. We further observed that when discrepancies occurred, it was often due to CANOPUS labeling compounds based upon substructures that are present in the molecule but not representative of the whole compound, e.g., labeling 1-palmitoylglycerol as a fatty acyl instead of a glycerolipid, or GPs as fatty acyls/sphingolipids, suggesting that despite misclassification, CANOPUS was identifying common substructures from MS/MS data (File S7a,b).

We also used MS/MS molecular networking to cluster compounds with similar fragmentation patterns. We then mapped identifications and CANOPUS Superclasses onto this network (Figure 3, Files S8 and S9). Some CANOPUS Superclasses tended to form tight subnetworks, e.g., 236 out of the 240 CANOPUS-annotated GPs in the negative mode network were clustered together (File S9), along with all the database-identified GPs. In the positive mode network, we observed two clusters for GPs – one for peaks identified as glycerophosphocholines/glycerophosphoserines and another for peaks identified as glycerophosphoethanolamines (subnetworks 1 and 2, respectively, in Figure 3, File S8). For other subnetworks (3, 4, and 5; Figure 3), there was good agreement between CANOPUS and identified compound class predictions (File S8).

As a third validation strategy, we compared CANOPUS results to predictions of the MS-DIAL lipidomics analysis pipeline (Tsugawa et al., 2020). We limited this analysis to only metabolites that were (i) predicted by CANOPUS to be in the 'lipids and lipid-like molecules' Superclass and (ii) given a lipid class by MS-DIAL – in order to limit the false positive annotations that may arise when considering a 'metabolomics' experiment as 'lipidomics'. Overall, the Classes of 75.7 and 48.1% of the high confidence MS-DIAL annotated lipids were correctly predicted in negative and positive mode, respectively (File S7d–f), supporting the above results. There was greater variation at the individual Class level – likely due to insufficient sample size – with GPs and sphingolipids having high rates of correct classification in negative mode (File S7f).

Finally, we also assessed whether CANOPUS predicts correct Classes of a large set of authentic standards (3885 in positive mode, 2743 in negative mode) that were run on the same instrument as our dataset, but were not used for training CANOPUS (Methods S1). As the true Classes of
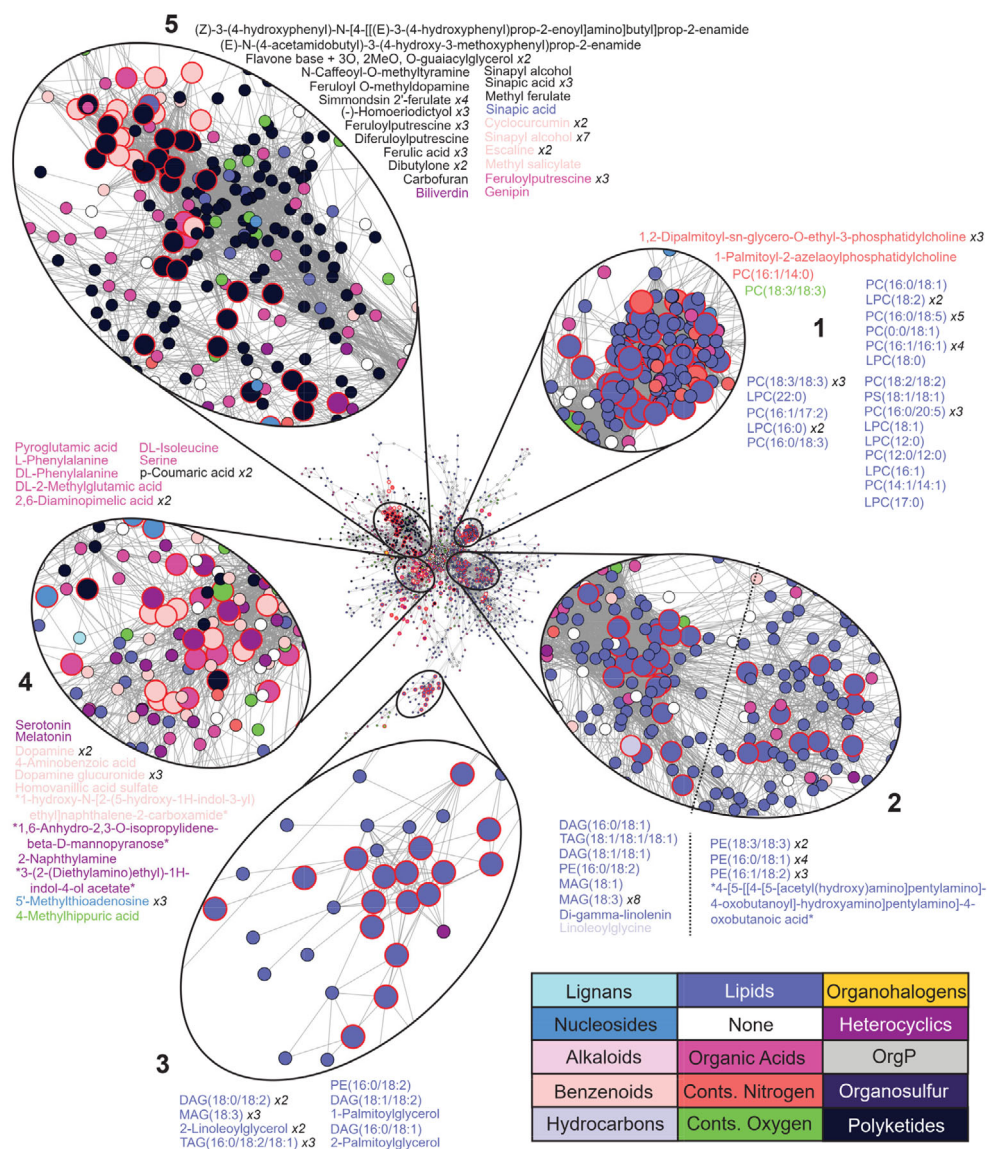
Figure 3. Molecular networking of peaks in positive mode. Network nodes represent peaks detected in positive mode (in any condition/organ), and edges connect nodes that have a pairwise cosine score of >0.7. Large nodes with a red border signify identified peaks. Nodes and identifications (text) are colored with their CANOPUS-annotated Superclass. The number of times each identification occurs in a subnetwork is indicated in italics. Asterisks (*) denote an identification spanning multiple lines. The dashed line in subnetwork 2 separates the majority-glycerolipid section of the subnetwork from the majority-phosphoethanolamine section. Superclass names in legend are shortened for representational purposes. OrgP, organophosphates; Polyketides, phenylpropanoids and polyketides; PC, phosphocholine; L, lyso; DAG, diacylglycerol; TAG, triacylglycerol; PE, phosphoethanolamine; MAG, monoacylglycerol.

these standards are definitively known, we calculated the precision, recall, and F1 score of CANOPUS predictions (see Methods S1). In both modes, on average, only approximately one third of classes with at least five instances had F1 scores of ≥0.6 (File S7f–i). In both modes, Classes such as flavonoids, prenol lipids, steroids, purine and pyrimidine nucleotides, phenylpropanoid acids, and pyrimidine nucleosides passed the F1 score filter. Among well-represented classes (≥10 instances), (i) we find flavonoids to have very similar scores across both ionization modes and (ii) we find several classes (including steroids,

pyridines, and prenol lipids, among others) to have large performance increases (F1 ≥ 0.15 higher) in positive mode (File S7i,j). Finally, we identified 11 classes (including Phenols, Benzene and derivatives, and coumarins and derivatives) that achieved high F1 scores (>0.7) in CANOPUS' training yet received low F1 scores (<0.5) in both ion modes in our analysis, highlighting the importance of this validation with the gold standard dataset.

These results suggest that the accuracy of CANOPUS predictions is dependent on the mode of ionization, compound class, and perhaps specific instrumentation. Thus,

for downstream analyses, we eliminated Classes deemed 'bad' (File S7i,j), i.e., with at least five standards and F1 scores of <0.6. Classes deemed 'Unknown' had less than five authentic standard instances; however, some of them, i.e., GPs and sphingolipids, showed good validation performance using MS-DIAL lipidomics and/or MS/MS network analysis. Due to this reason, 'Unknown' Classes were included in later analyses. Of the 2763 (positive) and 2314 (negative) peaks annotated at the Class level in our dataset, these criteria led us to using only 1353 (49%) and 1107 (48%), respectively, of the peaks for final analyses, which corresponds to approximately 38% of the fragmented peaks in each mode. We primarily show results of negative mode below due to its greater conformity across all four validation approaches, but also analyze filtered CANOPUS annotations of positive mode data in the Supplementary Material.

Compound class annotation reveals an important role of lipids in the induced stress response

Using filtered CANOPUS annotations, we sought to determine (i) how different chemical classes were perturbed under the applied stresses and (ii) whether the relevance of a class to a stress or organ could be quantified. The RDPI metric summarizes both up- and downregulation of all metabolites in a given class, and thus is a useful metric to assess a Class' overall perturbation in a given stress (Figures S15 and S16; File S10). The RDPI distributions of flavonoids were similar to those of the overall metabolome – with roots appearing more inducible than leaves and with Heat, HeatNoCopper, and Spore treatments eliciting the largest metabolome changes. However, some Classes – including lignans and lipids such as GPs, sphingolipids, and steroids – deviated from this overall trend.

Although the RDPI is a useful metric for quantifying gross metabolomic changes, information on whether peaks are upregulated or depleted under stress conditions is lost. Another issue is that our criteria for calling DAPs are stringent, thus high RDPI does not necessarily translate to more DAPs. Lastly, the RDPI metric for a Class with 1000 metabolites versus 10 metabolites can appear the same, confounding the true extent of a metabolite Class' importance in a condition. To address these issues, we identified Classes that were, on average, highly accumulated or depleted in a stress (see Experimental Procedures), and plotted the abundance changes of individual peaks in those Classes (Figure 4, and Figure S17a). Many Classes had expected changes in abundance, which corroborates this methodology. For example, in spore-treated samples, GPs decreased (leaves) while prenol lipids and sphingolipids generally increased (roots) (Figure 4), consistent with their importance in membrane remodeling and signaling during interactions between plants and arbuscular mycorrhizal fungi (Macabuhay et al., 2022; Wewer et al., 2014). In

Cu-deficient plants, GPs showed up- and downregulation in roots and leaves, respectively, while sphingolipids were upregulated in the roots. An increase in root sphingolipids was also seen in heat-stressed roots. A previous study showed that perturbation of sphingolipid biosynthesis in the roots influences the leaf ionome including Cu (Chao et al., 2011), and thus, sphingolipids may play consequential roles in both Cu deficiency and heat stress.

Other Classes showed unexpected changes. Although flavonoids are antioxidants, more flavonoid peaks were depleted rather than upregulated in the roots under multiple stresses (Figure 4a). Multiple Classes possess outliers present on both sides of the distribution, e.g., sphingolipids in Spore leaves, suggesting that peaks within the same structural Class are not necessarily co-regulated. For each condition, we identified Classes that were enriched among the stress-responsive metabolites (Fisher exact test, FDR-adjusted $P < 0.05$). Only GPs and sphingolipids were enriched in any condition (Figure 4a), suggesting that the lipidome is the most stress-responsive portion of the metabolome, possibly resulting from changes in cellular membranes and signaling pathways.

Finally, we assessed if the differing impacts of heat stress versus HeatNoCopper stress on the root metabolome were fueled by differential regulation of a particular Class. When considering only Classes with good performance as validated by Joint Genome Institute (JGI) standards above, we find similar proportions of Classes in the up- and downregulated peaks under each stress (Figure S18). However, Heat appeared to affect a structurally broader portion of the metabolome than HeatNoCopper, as determined by mapping DAPs onto an MS/MS network (Figure 4b, Figures S19 and S20). Many of the heat-specific DAPs were tightly clustered but did not receive CANOPUS annotations, suggesting that specific yet unknown routes of metabolism may be perturbed differently under these stresses.

Co-accumulated peaks have diverse structural classes, and peaks within a class rarely co-accumulate

As many classes showed broad changes in response to a stress, we next assessed the diversity of structural classes among groups of correlated peaks as determined using weighted gene coexpression network analysis (WGCNA) (Langfelder & Horvath, 2008) (Figure 5a and Figures S21 and S22). WGCNA provides a complimentary approach to assign functional hypotheses to metabolite classes under stresses, as it simultaneously assesses all conditions and classes. We found that most co-accumulation modules contained peaks with high abundance in roots and low in leaves, or vice versa, again highlighting organs as primary drivers for metabolic diversity. One module ('cyan') identified 16 peaks specifically accumulated in Sym Spore roots (Figure S22), four of which were annotated as
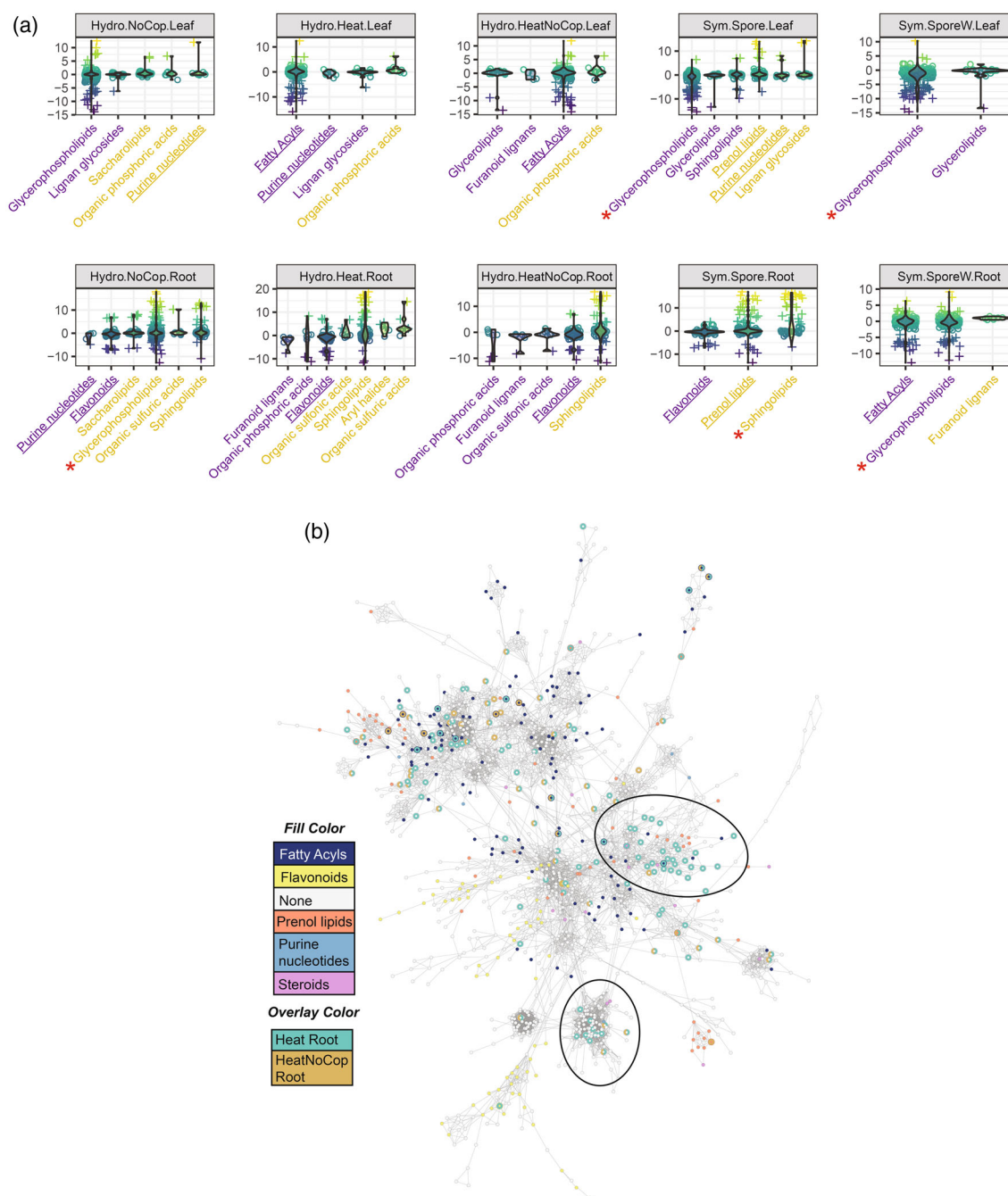
Figure 4. Charting stress-induced shifts of molecular classes, negative mode. (a) Abundance changes of peaks in response to stress. Each stress depicts Classes that were the most upregulated (Class name in yellow) or downregulated (Class name in purple), on average. For a Class to be plotted, its average value must be greater than the 70th percentile (yellow Classes) or lower than the 30th percentile (purple Classes) of all stress-induced peak area changes. Individual metabolites are plotted as circles, outliers are shown as +. Red asterisks (*) denote enrichment of a Class among the stress-induced metabolites in a condition. Classes that validated well with authentic standard data are underlined and those that did not validate well are not shown (see Methods S1). (b) MS/MS networking of all peaks in negative mode. Each node represents a peak, node fill color represents CANOPUS classifications. Only Classes that validated well with authentic standard data *and* only the largest connected network are shown. The full network is shown in Figure S20. Large nodes are peaks that are differentially accumulated in the roots under heat stress (teal overlay), HeatNoCopper stress (gold overlay), or both (both colors). Edges were drawn between nodes if their pairwise cosine similarity was >0.7. Ellipses mark network regions perturbed primarily under heat stress but not as much under HeatNoCopper stress.

sphingolipids, again independently suggesting the importance of sphingolipids in AMS. Other modules contained peaks with more varied accumulation patterns. For

example, the 'turquoise' module identified peaks that were either specifically accumulated in hydroponics roots or excluded from them (Figure 5a). The 'gray60' module
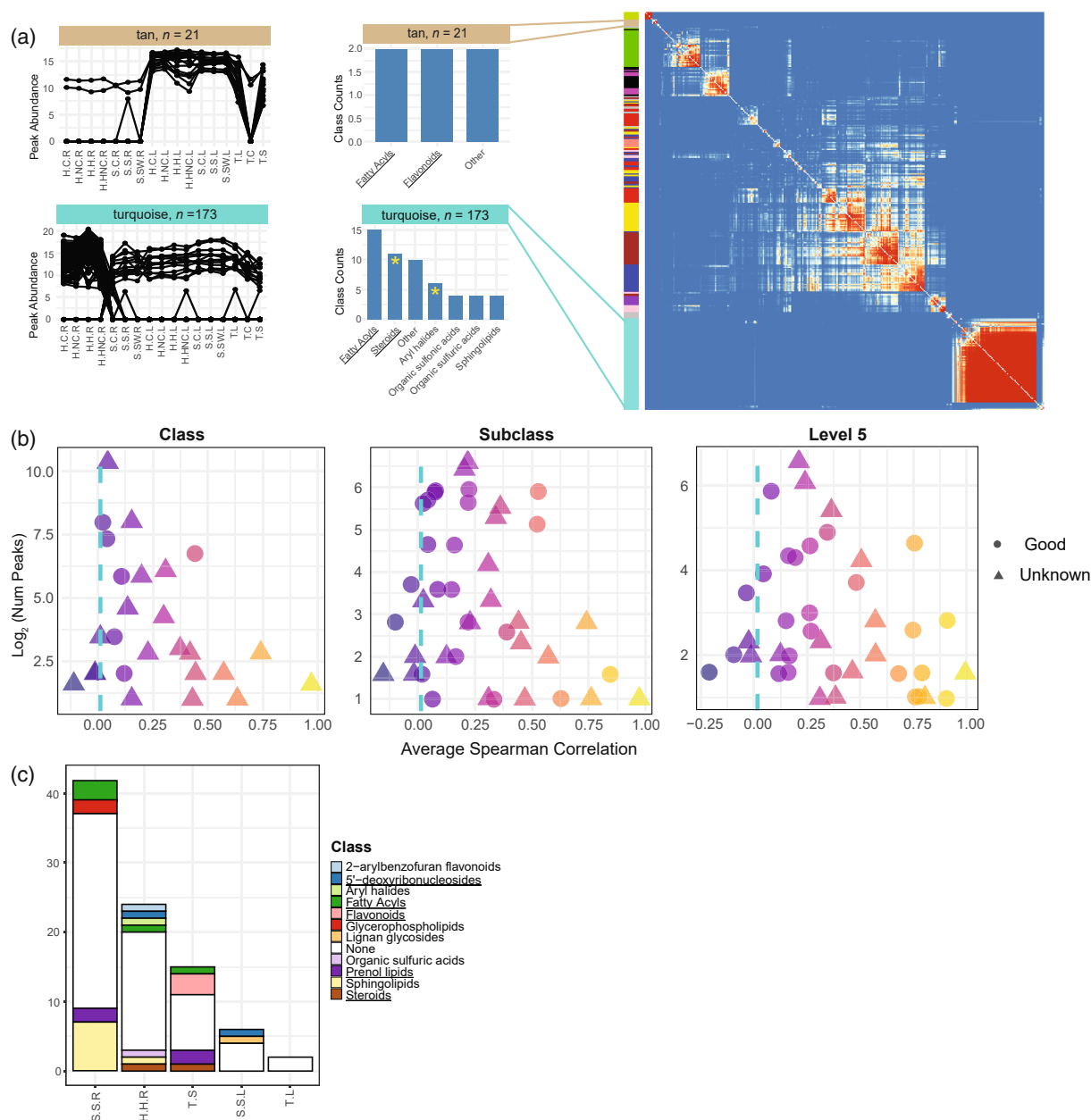
Figure 5. Characterizing metabolite co-abundance, negative mode. (a) WGCNA topography overlap matrix, depicting correlations among peaks placed into significant modules. Normalized abundance patterns and CANOPUS Class distributions are plotted for selected modules (Class 'None' not shown). An asterisk (*) denotes Classes that were significantly enriched in a module (Fisher's exact test, FDR-adjusted *P* < 0.05, count in module at least five). Condition names in the abundance pattern plots (left) are abbreviated such that only the capital letters of the full names (seen in Figure 4) are shown. (b) Average pairwise Spearman correlation among peaks in the same CANOPUS Class, Subclass, or Level 5. The blue line shows average correlation among randomly selected peaks. Good Classes are those that validated well against authentic standards, and Good Subclasses/Level 5 are those contained within Good Classes. (c) Counts of biomarkers found in each stress/tissue, colored by CANOPUS Class. Condition name abbreviations are as in (a). (a–c) Classes that validated well with authentic standard data are underlined (in a, c) or plotted as circles (in b), and those that did not validate well are not shown. Classes (and their subsequent Subclass and Level 5 annotations) unable to be validated by authentic standards data are plotted in (b) as triangles (see Methods S1).

(Figure S22) grouped peaks abundant in leaves but excluded from all roots except those experiencing AMS. These may represent foliar metabolites that undergo transport to the roots and play a role in symbiosis. A more detailed analysis of these peaks can reveal novel insights into the biochemistry of Brachypodium abiotic and biotic responses.

A majority of WGCNA modules contained multiple Classes, and four out of 18 modules were enriched in at least one Class (Fisher's exact test, FDR-adjusted

$P < 0.05$). Some metabolite classes, such as flavonoids (negative mode) and prenol lipids/fatty acyls (positive mode), were enriched in multiple modules with differing abundance patterns (Figures S21 and S22). Flavonoids were enriched in modules with higher accumulation in leaves than roots despite flavonoids being overall highly perturbed (had high RDPIs) in roots but not in leaves (Figures S15 and S16). These results point to differing regulation of individual metabolite classes in roots versus leaves. Additionally, of the seven Classes enriched across all modules, in either positive or negative mode, six were lipids, further highlighting their functional relevance.

To determine if 'Class' is too broad a level for co-regulation and if more evidence for co-regulation is found at the 'Subclass' or 'Level 5' level, the average pairwise Spearman correlation among accumulations of peaks in the same Class, Subclass, or Level 5 category (Figure 5b and Figure S17c) was compared to the average correlation among randomly drawn peaks. At each hierarchy level, a minority of classes had average correlation ≥ 0.5 (15, 19, and 32%, respectively), and most classes had correlation close to random. Notably, at each level of the hierarchy, several classes were unusually large, with >60 members, raising the possibility of low structural similarity within each class. Thus, we asked whether class size and structural similarity within class contribute to average class correlation (referred to as co-abundance, see Methods S1 for calculation). Correlations between average class co-abundance and average class cosine score were usually positive (Figures S23 and S24), suggesting greater structural similarity within a class translates to greater co-abundance. Correlations between class size and co-abundance/cosine score were 0 or negative, highlighting the importance of more specific class definitions. We note that overall, these metrics explained only a very low (approximately 10% or lower) proportion of variance.

Taken together, these results indicate that while some classes (e.g., flavonoids and their subclasses) may represent groups of co-regulated peaks, this is likely not the case for most classes. This may reflect the specificity of underlying metabolic and regulatory pathways, which may significantly increase concentrations of specific individual metabolites of a structural class. These results also suggest that utilizing the multilabel nature of the chemical ontology could be a better approach for finding peaks belonging to coordinated routes of metabolism rather than using single classes.

Analysis of specific metabolites and biomarker detection

Our dataset provides a unique opportunity to analyze the accumulation patterns of known metabolites, as well as find biomarkers, i.e., peaks that accumulate highly (not necessarily specifically; see Methods S1) in one condition/organ. We selected salicylic acid (SA), abscisic acid (ABA), and naringenin for analysis as they were identified by GNPS with match scores of ≥ 0.89 (the Classes of ABA and naringenin were additionally correctly annotated by CANOPUS), and may be of relevance in the studied conditions. We further validated these identifications by uncovering their major fragments from the literature and public repositories and checking for matching fragments in our queries (Table S3). SA is known to accumulate in roots under AMS (Zhang et al., 2013) and, in some species, under heat (Hara et al., 2012). We found that SA accumulated (but not significantly increased) in AMS roots, and was mildly but significantly increased in Heat roots (*t*-test, $P < 0.05$) (Figure S25). In contrast, ABA levels highly increased in AMS roots and in Heat and HeatNoCopper leaves (*t*-test, $P < 0.05$). Finally, for naringenin, mean decreases were observed in roots for all conditions (significant decreases seen in Heat and AMS; *t*-test, $P < 0.05$), corroborating our observations of decreases in the broader flavonoid Class.

We also found that the numbers of biomarkers detected in each condition resembled the overall RDPI distribution – roots typically have more biomarkers than their foliar counterparts, and Spore roots and Heat roots have the highest numbers of biomarkers (Figure 5c and Figure S17d; File S11). A large proportion of biomarkers were lipids (70% of annotated biomarkers in negative mode and 80% in positive mode). We found 11-carboxyblumenol C glucoside to be a foliar biomarker for AMS, corroborating previously published data (Wang, Schäfer, et al., 2018) (Figure S26a), and discovered other peaks that shared fragments (Figure S26b). We also detected a peak specific to Spore leaf and not present in AMS roots, which shares no fragment peaks with blumenol C and was classified by CANOPUS as a 5′-deoxyribonucleoside (Figure S26c), suggesting that AMS induces other foliar-specific routes of metabolism.

Visualizing metabolite class importance using the BAR platform

In order to make the data described herein more easily accessible to the scientific community, these data were integrated into the BAR website as a novel electronic Fluorescent Pictograph (eFP) browser (available for testing at: https://bar.utoronto.ca/efp_brachypodium_metabolites/cgi-bin/efpWeb.cgi). CANOPUS Classes surviving validation experiments with at least five members in both positive *and* negative modes were included in this eFP browser. This eFP browser has two viewing options. With the Relative viewing option, the changes of metabolite Class levels across conditions can be readily observed (Figure 6) as the average change in normalized peak area under a condition. With the Absolute viewing option, the average normalized peak areas are plotted per organ and condition. Besides showing how the Class changes in abundance across

conditions, the Absolute view option also provides information about which ionization mode best illustrates changes experienced in that Class. Notably, for some Classes (e.g., furanoid lignans) we observe different changes in abundance across ionization modes. While this may be due to CANOPUS peak misannotations, especially for positive mode, it may also reflect different subclasses being detected in different ionization modes. This finding has implications for targeted comparative metabolomics studies, as results obtained in one ionization mode may not necessarily hold in the other. By establishing our eFP browser, we seek to enable the community to draw further conclusions from our existing results and facilitate the design of future comparative metabolomics and downstream validation studies.

DISCUSSION

While recent improvements in LC-MS hardware have generated impressive advancements in metabolite detection, associating the thousands of metabolites detected in each species with biological processes remains an open challenge (Chaleckis et al., 2019). In this study, using three complementary approaches – information theory, ML-based analysis, and co-accumulation clustering – for LC-MS data, we performed a more comprehensive analysis of metabolome perturbations of *B. distachyon* under different environmental conditions. We performed extensive quality control steps to account for peaks with poor peak shapes, noise, low intensity peaks, in-source fragments, and adducts; however, it is possible that some of the remaining peaks also include cluster ions, unexpected adducts, solvent-derived ions, and other in-source fragments. It is also possible that these factors may affect CANOPUS predictions more in one mode versus another. Nonetheless, we do not expect these unusual forms to affect one sample substantially more than the other and we primarily base our inferences using high-confidence CANOPUS predictions from negative mode; therefore, the overall inferences obtained in this study are likely to be robust.

When applying information theory-based measures to the global metabolome, we found that roots are, on the whole, more stress-responsive than leaves, despite leaves having a more expansive and complex metabolome. The finding that leaves have consistently more peaks than roots may be due to biological or technical/processing reasons, as root harvesting required a washing/drying step to remove the attached soil particles, which may have also removed epidermal metabolites. While the increased number of peaks in foliar samples directly contributes to their increased diversity, the finding that leaf metabolomes are less perturbed than roots under the studied stresses is intriguing. While this may partly be a consequence of the experimental design, previous studies have also found roots to be more impacted than leaves under a variety of stresses, including heat (Giri et al., 2017) and salinity (López-Cristoffanini et al., 2021). Notably, drought stress – not included in our study – appears to be an exception in which leaves are more impacted than roots (Gargallo-Garriga et al., 2014), indicating that the greater metabolic plasticity of the roots is not universal. These results may again be due to technical considerations, as peaks with $m/z > 800$ were not detected, thereby excluding cuticular waxes, which are stress-responsive (Baker, 1974; Kan et al., 2022; Wang, Tian, et al., 2018). Additionally, highly polar and highly non-polar compounds were excluded from our data. Both roots and leaves contain such compounds, and therefore it is unclear how results would differ with these compounds included.

Our analyses revealed that the combined HeatNoCopper stress was less disruptive to the root metabolome than the Heat stress alone, suggesting that 1 week of Cu deficiency primed the roots for subsequent protection against heat stress. Another interpretation is that critical heat response mechanisms were not activated in the roots after a week of Cu deficiency, which could therefore have other long-term impacts. Since the recovery of these plants after stress was not studied, it is not possible to ascertain which interpretation is correct. However, these results reveal an intriguing interplay between heat stress and Cu deficiency. In Arabidopsis, such an interplay is suggested through shared aspects of heat and Cu deficiency responses. For example, Cu deficiency triggers accumulation of ferric superoxide dismutase 1 to account for reduced activity of Cu/Zn superoxide dismutases (Abdel-Ghany & Pilon, 2008). This shift may help protect the roots against reactive oxygen species produced during later heat shock and provide resilience to photosynthesis, which requires Cu. Recent evidence has also suggested that SPL7, a master regulator of the response to Cu deficiency (Yamasaki et al., 2009), may upregulate miR156 under Cu deficiency (Perea-García et al., 2021). In Arabidopsis, miR156 is induced after an initial heat stress event and provides heat shock memory, as plants lacking miR156 showed decreased growth and survival after subsequent heat events (Stief et al., 2014). As miR156 is also induced in wheat after heat stress (Xin et al., 2010) and several miRNAs are known to have different induction patterns in different tissues (Sunkar et al., 2012), we hypothesize that miR156 upregulation under Cu deficiency helps prime Brachypodium roots for heat stress. Another candidate is miR398a, which we find to be upregulated under Cu deficiency (Figure S2) and which is also involved in the heat stress response (Schulten & Krämer, 2018). We also found that sphingolipids, which are associated with both temperature response and ionome regulation (Chao et al., 2011; Huby et al., 2020; Sun et al., 2022), are upregulated in NoCopper, Heat, and HeatNoCopper roots (Figure 4). This finding suggests a role for Cu deficiency-induced sphingolipids in protecting
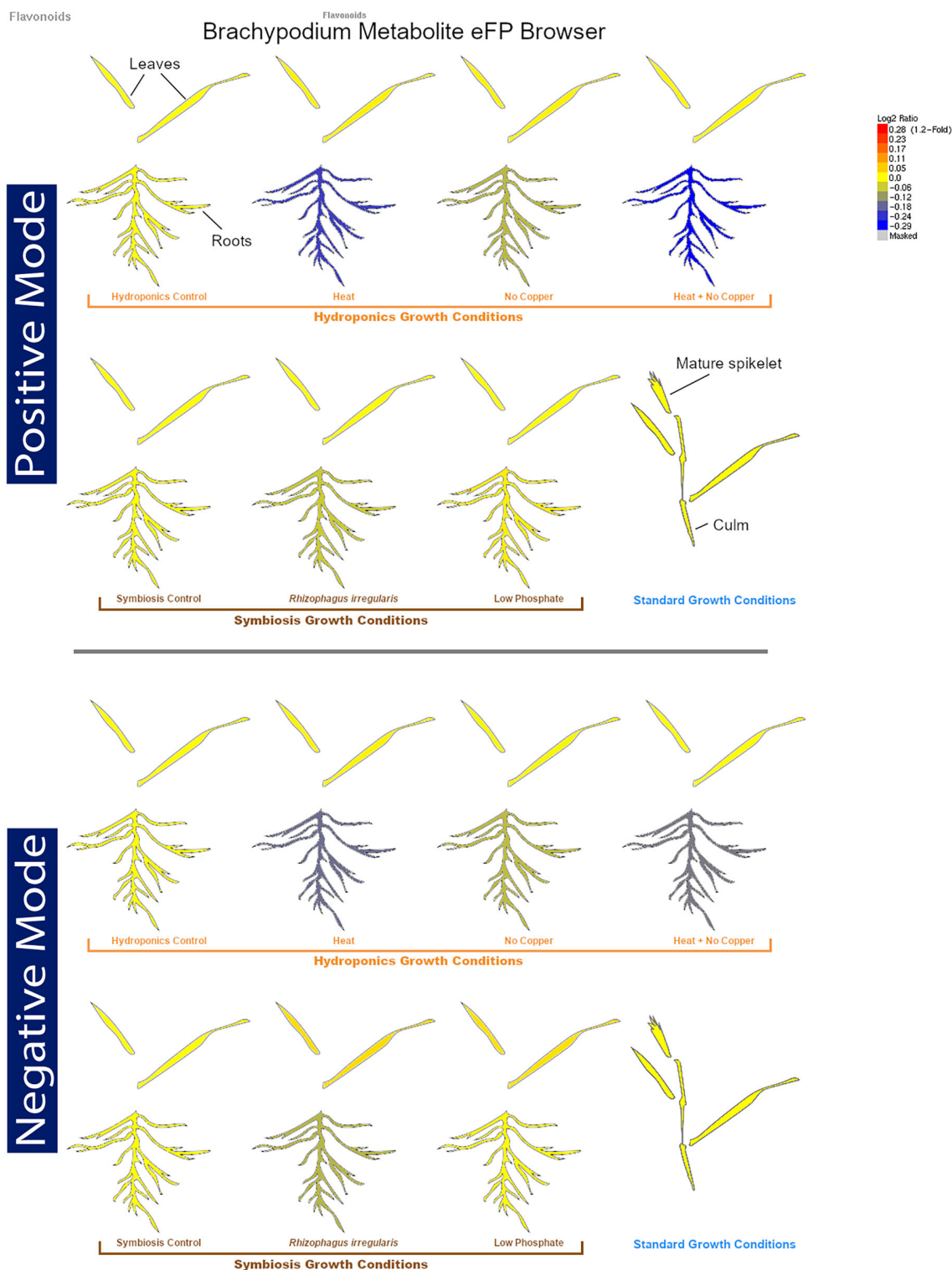
Figure 6. Visualizing stress-induced changes in class abundance. In relative mode of the eFP browser (shown here for flavonoids), the log$_2$(fold change) values in average Class abundance between a condition and its control are plotted. The consistent decreases among stressed roots are again seen.

Brachypodium roots against the impact of transitional heat stress. Future mechanistic studies can help test these hypotheses.

We combined CANOPUS – a tool for structurally annotating peaks – with information theory-based and related measures to analyze more specific metabolome

perturbations. While ML-based tools for peak identification are likely the next frontier in mass spectrometry data analyses, crucial third-party validations of the tools are lacking. Our study provides one of the most extensive, multipronged validations of CANOPUS on plant datasets, revealing different factors that may affect machine learning performance. After validation, we used CANOPUS to uncover extensive metabolic modulations comprising lipidomic perturbations and alterations of phenylpropanoid pathway products such as lignans and flavonoids. We find lipids, on the whole, to be highly stress-responsive, with glycerolipids, GPs, sphingolipids, and fatty acyls having high perturbations under several conditions, yet a limitation of this methodology is that it cannot detect changes in lipid acyl chain composition. However, these Class-level perturbations may be a result of changes in membrane composition (known to occur under heat [Higashi & Saito, 2019] and low P stress [Nakamura, 2013]) and/or production of lipid signaling molecules, such as oxylipins (Ali & Baek, 2020) and sphingolipids (Berkey et al., 2012). Under AMS specifically, certain fatty acyls, sphingolipids, and GPs are known to be produced (Bravo et al., 2017; Moore et al., 2021; Wewer et al., 2014), and while this is indeed reflected in our data (File S10) we also found that prenol lipids were highly altered under AMS, suggesting that AMS has wide-reaching effects on the Brachypodium lipidome. We unexpectedly found that several flavonoids decreased in the roots in response to all conditions except low P – a finding also observed by a focused assessment of naringenin. The validity of CANOPUS flavonoid predictions is not only confirmed in this study but was also previously confirmed in sweet potato (*Ipomoea batatas*) flavonoids and anthocyanins via comparison with MS/MS networking (Bennett et al., 2021). In general, flavonoids are known to accumulate under several stresses (Ferdinando et al., 2012), yet the wholescale labeling of all flavonoids as antioxidants has been questioned (Agati et al., 2020). Previous studies have additionally found disordered regulation of flavonoid biosynthesis, at the level of either individual flavonoids/flavonoid biosynthetic genes (Wu et al., 2020) or post-transcriptional regulation of flavonoid biosynthesis (Cui et al., 2019). These observations reveal a need for a deeper investigation of roles of specific flavonoids and/or their metabolic reprogramming under stress. We further found that WGCNA, a tool commonly used in RNA-seq studies, is effective at uncovering peaks with similar abundance patterns, which are potentially in the same routes of metabolism. Our study was also able to detect biomarkers, which can reveal novel insights into condition-specific activations of metabolic pathways.

In conclusion, we found that information theory-based metrics and chemical class predictions are effective tools to analyze comparative metabolomics data. Our results reveal a very dynamic plant metabolome influenced by multiple environmental and developmental factors. As more untargeted LC-MS/MS studies are performed, comparative analyses of these datasets may reveal common patterns and the core stress response across groups of plant species. The overall workflow described here can enable a more streamlined analysis of such untargeted datasets. Nonetheless, it is important to carefully validate the ML-based annotations prior to using them for metabolome-wide analyses, given their variable performance based on instrumentation, mode of ionization, and specific metabolite classes. Visualizing metabolomic data using the eFP browser may reveal hidden spatial differences in metabolome perturbations not easily discernible otherwise, and guide the design of targeted studies. For example, this visualization can be a useful tool to identify a better mode of ionization for molecules of interest as well as reveal metabolite classes to be assessed via targeted analyses. Our study shows that data-intensive analytical methods are useful resources for gleaning novel biological insights from untargeted metabolomics studies in plants.

EXPERIMENTAL PROCEDURES

Plant growth conditions and harvesting

Brachypodium distachyon Bd-21 seeds for plants used in the Sym and Tis experiments were sterilized in 10% (v/v) household bleach containing 0.005% (v/v) Tween-20 for 7 min, thoroughly washed five times in sterile water, and germinated in Petri dishes on moist Whatman filter paper in the dark at 4°C for 7 days and at room temperature for an additional 3 days. Germinated seedlings were incubated for an additional 3–5 days under constant light while maintaining constant humidity. The germination protocol for plants used in the Hydro experiment was performed as outlined previously (Sheng et al., 2021). Additional details about plant growth conditions are described in Methods S1. After the growth period, harvesting of all plant material took place between noon and 3:00 pm to maintain circadian profiles of genes and metabolites. All samples were stored at −80°C until further processing. We verified the efficacy of the Cu deficiency and AMS conditions using RT-PCR of previously known condition-specific genes (Rahmati Ishka & Vatamaniuk, 2020) (Methods S1).

Metabolite extraction and sample preparation

All plant material was rough ground over liquid nitrogen using scissors to enable equal and homogenous separation for RNA and metabolite extraction. All samples were further subjected to bead homogenization using a mixer mill (Retsch, Haan, Germany) at 30 bpm with 1-min intervals in 2-ml reaction tubes containing four 2.3-mm chrome steel beads. Ground samples were lyophilized overnight. Sample fresh weights (200 mg leaves, 550 mg roots, 150 mg spikelets and culms) were determined to ensure 50 mg of dry weight for all tissues. Samples were ground again in the bead homogenizer for 10 min and centrifuged at 14 000 *g* for 10 min in order to collect all powdered sample at the bottom of the tube. Metabolites were extracted using a mixture of acetonitrile, isopropanol, and water (ratio of 2:2:1) containing 0.1% (v/v) formic acid and 30 μM of three internal standards (Telmisartan, propyl-4-hydroxy-benzoate, and kanamycin). After solvent addition, samples were vortexed several times over a period of 15 min to

facilitate extraction. After centrifugation for 10 min at 16 000 *g* to remove particulates, the samples were transferred into amber HPLC vials and stored at −80°C until LC-MS/MS analysis. Samples were shipped to the JGI on dry ice for LC-MS/MS analysis, where LC-MS/MS was performed using an Agilent 1290 Infinity LC system (Agilent, Santa Clara, CA, USA) coupled to a Thermo QExactive HF orbitrap mass spectrometer (Thermo Scientific, San Jose, CA, USA). Additional details are provided in Methods S1.

Metabolomic data filtering, normalization, and imputation

All RAW files were converted to mzML format using ProteoWizard v 3.0.7230. TICs were made for all files of a given polarity using XCMS (Mahieu et al., 2016) (Figure S3). All files of a given mode (positive or negative) were then imported into MS-DIAL v4.48 (Tsugawa et al., 2020) for peak deconvolution and alignment. Parameter files for positive and negative mode usage are supplied (File S12). The peak areas of the internal standards Telmisartan and propyl-4-hydroxy-benzoate were manually checked to determine consistency across samples. For each polarity, MS-DIAL outputs a quantitative alignment file, displaying the peak areas of all metabolites in all samples, and a Mascot Generic Format (mgf) file of all fragmented metabolites. Detected metabolites were filtered, imputed, and normalized using a custom R script (developed in R v4.0.4) (R Core Team, 2020), available on GitHub (https://github.com/lizmahood/brachy_metabolomics) as described in Figure S4. Metabolites eluting out at 90 sec or earlier were removed as the total ion current observed at the beginning of runs was high enough that accurate quantification of metabolite values could not be ensured (Figure S3). Imputation was performed with the R package impute and VSN was performed with the R package vsn (Huber et al., 2002). Normalization scheme was chosen using NOREVA (Li, Tang, et al., 2016), followed by identification of differentially accumulated metabolites, both of which are described in more detail in Methods S1.

Peak annotation with CANOPUS

The mgf format MS-DIAL output files were filtered to remove adducts and peaks detected in Blank samples using an in-house python script (https://github.com/lizmahood/brachy_metabolomics). The CANOPUS module (Dührkop et al., 2021) included in the SIRIUS4 v4.9.8 software suite (Dührkop et al., 2019) was used to annotate singly charged peaks with their probable structural classes, as defined in the multilabel ChemOnt ontology (Djoumbou Feunang et al., 2016). The Zodiac module (Ludwig et al., 2020) was additionally used to improve each peak's predicted molecular formula (which CANOPUS uses for annotation). For each compound, CANOPUS predicts the 'Parent Class' – the class of the largest substructure in the molecule – and outputs the probability that the predicted Parent Class is correct, based upon its training data. Other predictions are made at different hierarchies of the ontology (Superclass, Subclass, etc.). Any annotation with prediction probability < 0.5 was not considered in downstream analyses. Additionally, if a classification was discarded for not meeting this probability threshold, each subsequent prediction (at more specific hierarchies) was removed as well, regardless of their prediction probabilities.

Peak identification with GNPS and MS-DIAL

The 'All Public MS/MS' msp files provided by MS-DIAL (http://prime.psc.riken.jp/compms/msdial/main.html#MSP) were used for identification. To remove false positive identifications, we imposed a threshold of >0.8 for both the Dot Product and Reverse Dot Product scores between the query and database match. Feature-based molecular networking through GNPS (Nothias et al., 2020) workflow v28.2 was additionally used for peak identification. Spectral database libraries included those publicly available in GNPS, as well as the NIST 17 library, which was kindly provided by JGI. All parameters for molecular networking were kept at default values except precursor ion mass tolerance (0.01 Da), library search min matched peaks (3), top results to report per query (20), score threshold (0.4), and maximum analog search mass difference (200). We again imposed a threshold of >0.8 for the match score between the query and database match, and only considered the top match per query.

To compare annotations between peak identifications and CANOPUS, InChIs of identified compounds were converted to InChI-Keys through the chembl_ikey python module (https://github.com/mnowotka/chembl_ikey), and structural classifications were obtained with ClassyFire Batch (https://cfb.fiehnlab.ucdavis.edu/).

MS/MS molecular networking

MS-FINDER v3.44 (Tsugawa et al., 2016) was used to perform molecular networking using the filtered mgf files, with the following parameters: mass tolerance 0.01, relative abundance cutoff 5%, MS/MS similarity cutoff 70%, and RT tolerance 100. The Superclass of each peak and the conditions in which each peak was identified as differentially abundant were added to the node file. The edge file and this augmented node file were imported into Cytoscape v.3.8.0 (Su et al., 2014) for figure generation using the Prefuse Force Directed Layout.

Estimating information theory-based measures

The following information theory-based metrics were calculated for our dataset as described previously (Li et al., 2020): Hj (the Shannon entropy/metabolomic profile diversity), Si (metabolomic specificity), and δj (metabolome specialization index). The RDPI, as calculated for all peaks in each stress condition, was also determined as described previously (Valladares et al., 2006). The RDPI calculation was applied to the entire metabolome, and then applied separately to each compound class (for compound classes with at least five peaks classified into them).

The RDPI formula was amended in order to determine if a class is up- or downregulated under a stress. Briefly, for each condition–control pair of samples, a distribution of the abundance changes of all peaks was made, and the mean change in peak abundance was calculated per class. Let $d_{ij \to i'j'}$ represent the peak area changes to all peaks i common to a condition–control pair $(j \to j')$. The mean value of the peak area change for each compound class was computed as $\sum (d_{ij} \to d_{i'j'})/n$, where n is the number of peaks per class. For each condition, these per-class mean values were compared to the overall distribution of $d_{ij \to i'j'}$ across all metabolite peaks to determine the percentile of the per-class value with respect to the peak area changes of all compared metabolites. For the purposes of plotting in Figure 4, the classes with percentiles of >70 (large average increase in abundance) or <30 (large average decrease in abundance) and at least five members were identified, and up to five classes with the highest/lowest percentiles were plotted.

Weighted correlation network analysis construction and module analysis

Using the Weighted Correlation Network Analysis (WGCNA) R package (Langfelder & Horvath, 2008), an unsigned adjacency

network was made from the normalized area of all fragmented peaks. The soft powers were 129 and 131 in positive mode and negative mode, as these were the lowest values achieving an R^2 value of at least 0.8. Hierarchical clustering via the hclust function was performed using method = 'average'. The minimum module size was 10. All peaks that failed to be assigned to a module were discarded, and the remaining peaks were re-clustered into a dendrogram and visualized alongside their topography overlap matrix. The CANOPUS class of all peaks in each significant module was determined. Each class (except 'None', containing unclassified peaks) was analyzed for enrichment in a particular module if there were at least five members in the module. Enrichment was calculated using a Fisher's exact test with all fragmented peaks as the background population.

Visualizing CANOPUS class abundance on the BAR platform

Briefly, an input image was generated representing the experiments described in this paper. The eFP browser code (Winter et al., 2007) was then modified in several ways to be able to display CANOPUS data. First, the color scheme was modified from the default yellow-red color scheme of the original eFP browser to make a visual distinction between the metabolite data being displayed in the modified version and transcript data displayed in the original browser. Second, because CANOPUS data have a lower dispersion, we introduced a possibility of setting a minimum value for the color scale other than zero. Last, CANOPUS classes with at least five members in both positive *and* negative ionization modes were included in this eFP browser, and were databased in such a way that the data from the two modes could be retrieved separately. CANOPUS data may be freely explored at https://bar.utoronto.ca/efp_brachypodium_metabolites/cgi-bin/efpWeb.cgi.

CONFLICTS OF INTEREST

No conflicts of interest exist.

DATA AVAILABILITY STATEMENT

The LC-MS/MS data are deposited on the GNPS website with the accession ID MSV000089340. All code developed for analyses is available on the GitHub repository (https://github.com/lizmahood/brachy_metabolomics) and on the moghelab GitHub page.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

File S1. Raw results of different normalization schemes on our LC-MS/MS data using NOREVA. NEG and POS refer to negative and positive mode data, respectively. Based on comparison of different metrics, VSN was chosen as the normalization approach given it resulted in the highest purity of K-means clusters and the most intergroup separation in the principal component analysis.

File S2. Replicate correlations (determined by Pearson's r) after all filtering and normalization steps.

File S3. Normalized peak area and CANOPUS annotations. All peaks passing the filtering and normalization steps are included in this file. Outlier samples have been detected and removed. High confidence (posterior probability > 50%) CANOPUS annotations for fragmented, singly charged peaks are included. Peaks that received no annotations are labeled as 'None', together with annotations with posterior probability ≤ 50%.

File S4. Quantification, metadata, and CANOPUS annotations of differentially accumulated peaks (DAPs). Each peak's Alignment ID corresponds with its Alignment ID in File S3. Average abundance values of 0.01 indicate no detection. 'normalized_fc' shows the fold change calculated using normalized peak abundances (see Experimental Procedures), whereas 'original_fc' shows the fold change of peak abundancies calculated just prior to normalization.

File S5. CANOPUS annotations in positive mode.

File S6. CANOPUS annotations in negative mode.

File S7. Correspondence between peaks identified using spectral matches and their class predictions using CANOPUS. [1]Identified using spectral matches with public repositories. [2]Annotated using CANOPUS. [3]Percentage of the identified and annotated peaks with matching class definitions. Definitions of the identified peaks were obtained using ClassyFire.

File S8. Cytoscape format MS/MS network file containing nodes labeled with their CANOPUS Class and Superclass annotations. Data is for negative mode only.

File S9. Cytoscape format MS/MS network file containing nodes labeled with their CANOPUS Class and Superclass annotations. Data is for positive mode only.

File S10. Average change in peak area and average RDPI per Class, under each condition. The ChangePercentile shows what percent of peaks had a smaller (or more negative) change in peak area

than each Class. This was calculated independently for each condition (see Experimental Procedures for calculation details).

File S11. Condition – enriched metabolites (biomarkers) found in each condition. Alignment ID corresponds with the Alignment ID for each metabolite in File S3. Fold changes of >1000 occur when the respective metabolite is not present in any other condition.

File S12. Parameter settings for MS-DIAL in negative and positive modes.

Figure S1. Experimental design and set-up of the three subexperiments conducted.

Figure S2. Reverse transcriptase (RT) PCR and quantitative RT-PCR for validation of Cu and AMS treatments.

Figure S3. Total ion chromatograms for all samples.

Figure S4. Flow chart of the procedure developed to quantify metabolites.

Figure S5. RMSE between two values and imputed values under two imputation–normalization orders.

Figure S6. Total peak area and peak numbers per sample.

Figure S7. Principal component analysis of all samples surviving the filtering process.

Figure S8. Replicate dendograms. All non-outlier positive samples were clustered using a distance matrix of 1 – pairwise Pearson's correlation. Samples were clustered based on filtered and normalized metabolite values.

Figure S9. Replicate dendograms. All non-outlier negative samples were clustered using a distance matrix of 1 – pairwise Pearson's correlation. Samples were clustered based on filtered and normalized metabolite values.

Figure S10. Diversity and specialization per condition. Bars represent the mean of each metric per condition, and error bars show the standard deviation across all replicates remaining after outlier removal. (a) and (b) show the diversity and specialization in positive mode, while (c) and (d) show them in negative mode.

Figure S11. Full UpSet plots showing intersections of differentially accumulated peaks (DAPs) in all sets of conditions. DAPs include both up- and downregulated peaks. (a) includes peaks detected in positive mode and (b) includes peaks detected in negative mode. The vertical bar charts (showing DAP intersections) show the numbers of DAPs found in each intersection of conditions/organs, as outlined by the combination of dots underneath the bar.

Figure S12. Volcano plots of differentially accumulated peaks in negative mode.

Figure S13. Volcano plots of differentially accumulated peaks in positive mode.

Figure S14. Ratio of compounds annotated at different levels of the ChemOnt ontology. (a) Positive mode. (b) Negative mode. Non-filtered bars represent all compound classifications. Filtered bars represent only those classifications with posterior probabilities of >0.5.

Figure S15. RDPI per CANOPUS Class, positive mode. Each dot represents the RDPI between one replicate in a particular stress and its control. 'None' represents metabolites that received no CANOPUS Class-level annotation. Classes that validated well with authentic standard data are underlined and those that did not validate well are not shown (see Methods S1).

Figure S16. RDPI per CANOPUS Class, negative mode. Each dot represents the RDPI between one replicate in a particular stress and its control. 'None' represents metabolites that received no CANOPUS Class-level annotation. Classes that validated well with authentic standard data are underlined and those that did not validate well are not shown (see Methods S1).

Figure S17. Charting stress-induced shifts of molecular classes.

Figure S18. Distribution of classes among DAPs in roots under heat stress and HeatNoCopper stress. Only Classes that validated well with authentic standards (see Methods S1) are shown. The percentage of peaks in each Class is shown, and the number of peaks in each Class is plotted on/above the bars. (a, b) Positive mode. (c, d) Negative mode. (a, c) Upregulated peaks in each stress. (b, d) Downregulated peaks.

Figure S19. MS/MS networking of peaks in positive mode.

Figure S20. MS/MS networking of peaks in negative mode.

Figure S21. Normalized expression and CANOPUS classifications of peaks in significant WGCNA modules, positive mode.

Figure S22. Normalized expression and CANOPUS classifications of peaks in significant WGCNA modules, negative mode.

Figure S23. Relationship between average Spearman correlation, average cosine score, and number of peaks in each category at the Class, Subclass, and Level 5 levels. All plots are for peaks in positive mode. The 'corr' on each plot is the Pearson correlation between the variable pair. Classes that validated poorly using authentic standards (see Methods S1), as well as the Subclasses and Level 5 annotations within such classes, are not plotted.

Figure S24. Relationship between average Spearman correlation, average cosine score, and number of peaks in each category at the Class, Subclass, and Level 5 levels. All plots are for peaks in negative mode. The 'corr' on each plot is the Pearson correlation between the variable pair. Classes that validated poorly using authentic standards (see Methods S1), as well as the Subclasses and Level 5 annotations within such classes, are not plotted.

Figure S25. Normalized abundance values for salicylic acid, abscisic acid, and naringenin across all organs and conditions. Metabolites were identified through a database search (using GNPS, see Experimental Procedures).

Figure S26. Substructures of foliar biomarkers of symbiosis. (a) 11-Carboxyblumenol C 9-*O*-glucoside was identified to Level 2. Characteristic 'blumenol' fragment peaks are shown in green and orange. (b) Unknown compound with no accumulation in Sym Spore roots yet containing two characteristic 'blumenol' fragment peaks. (c) Unknown compound lacking any 'blumenol' fragment peaks. Asterisks signify foliar symbiosis biomarkers, and each compound's ID corresponds with its Alignment ID in File S3.

Table S1. Coefficients for different normalization methods. Purity of k-means clustering denotes the normalization method's effect on differential metabolic analysis, while the relative weighted consistency reports the method's consistency between identified markers among different datasets.

Table S2. *P*-values of Kolmogorov–Smirnov tests of diversity and specialization distributions. Each Kolmogorov–Smirnov test compared the distributions between a stress and its respective control. Significant *P*-values (<0.05) indicate that the specialization or diversity values between the stress and control arise from different distributions. All *P*-values are two-sided.

Table S3. Comparison of observed MS/MS fragments with database spectra.

Table S4. List of primers used for reverse transcriptase (RT)-PCR and quantitative RT-PCR.

Methods S1. Supplementary methods.

REFERENCES

Abdel-Ghany, S.E. & Pilon, M. (2008) MicroRNA-mediated systemic downregulation of copper protein expression in response to low copper availability in Arabidopsis*. *Journal of Biological Chemistry*, **283**, 15932–15945.

Afendi, F.M., Okada, T., Yamazaki, M., Hirai-Morita, A., Nakamura, Y., Nakamura, K. *et al.* (2012) KNApSAcK family databases: integrated metabolite–plant species databases for multifaceted plant research. *Plant and Cell Physiology*, **53**, e1.

Agati, G., Brunetti, C., Fini, A., Gori, A., Guidi, L., Landi, M. et al. (2020) Are flavonoids effective antioxidants in plants? Twenty years of our investigation. *Antioxidants*, **9**, 1098.

Ali, M.S. & Baek, K.-H. (2020) Jasmonic acid signaling pathway in response to abiotic stresses in plants. *International Journal of Molecular Sciences*, **21**, 621.

Baker, E.A. (1974) The influence of environment on leaf wax development in brassica oleracea var. Gemmifera. *New Phytologist*, **73**, 955–966.

Bennett, A.A., Mahood, E.H., Fan, K. & Moghe, G.D. (2021) Untargeted metabolomics of purple and orange-fleshed sweet potatoes reveals a large structural diversity of anthocyanins and flavonoids. *Scientific Reports*, **11**, 16408.

Berkey, R., Bendigeri, D. & Xiao, S. (2012) Sphingolipids and plant defense/disease: the "death" connection and beyond. *Frontiers in Plant Science*, **3**, 68.

Bravo, A., Brands, M., Wewer, V., Dörmann, P. & Harrison, M.J. (2017) Arbuscular mycorrhiza-specific enzymes FatM and RAM2 fine-tune lipid biosynthesis to promote development of arbuscular mycorrhiza. *New Phytologist*, **214**, 1631–1645.

Brkljacic, J., Grotewold, E., Scholl, R., Mockler, T., Garvin, D.F., Vain, P. *et al.* (2011) Brachypodium as a model for the grasses: today and the future. *Plant Physiology*, **157**, 3–13.

Bromke, M.A., Hochmuth, A., Tohge, T., Fernie, A.R., Giavalisco, P., Burgos, A. *et al.* (2015) Liquid chromatography high-resolution mass spectrometry for fatty acid profiling. *The Plant Journal*, **81**, 529–536.

Brouard, C., Shen, H., Dührkop, K., d'Alché-Buc, F., Böcker, S. & Rousu, J. (2016) Fast metabolite identification with input output kernel regression. *Bioinformatics*, **32**, i28–i36.

Chaleckis, R., Meister, I., Zhang, P. & Wheelock, C.E. (2019) Challenges, progress and promises of metabolite annotation for LC–MS-based metabolomics. *Current Opinion in Biotechnology*, **55**, 44–50.

Chao, D.-Y., Gable, K., Chen, M., Baxter, I., Dietrich, C.R., Cahoon, E.B. *et al.* (2011) Sphingolipids in the root play an important role in regulating the leaf ionome in Arabidopsis thaliana. *The Plant Cell*, **23**, 1061–1081.

Charles, M., Tang, H., Belcram, H., Paterson, A., Gornicki, P. & Chalhoub, B. (2009) Sixty million years in evolution of soft grain trait in grasses: emergence of the softness locus in the common ancestor of Pooideae and Ehrhartoideae, after their divergence from Panicoideae. *Molecular Biology and Evolution*, **26**, 1651–1661.

Chong, J., Wishart, D.S. & Xia, J. (2019) Using MetaboAnalyst 4.0 for comprehensive and integrative metabolomics data analysis. *Current Protocols in Bioinformatics*, **68**, e86.

Cui, L., Guo, F., Zhang, J., Yang, S., Meng, J., Geng, Y. *et al.* (2019) Synergy of arbuscular mycorrhizal symbiosis and exogenous Ca2+ benefits peanut (Arachis hypogaea L.) growth through the shared hormone and flavonoid pathway. *Scientific Reports*, **9**, 16281.

da Silva, R.R. & Dorrestein, P.C. (2015) Illuminating the dark matter in metabolomics. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 12549–12550.

Djoumbou Feunang, Y., Eisner, R., Knox, C., Chepelev, L., Hastings, J., Owen, G. *et al.* (2016) ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. *Journal of Cheminformatics*, **8**, 61.

Douché, T., Clemente, H.S., Burlat, V., Roujol, D., Valot, B., Zivy, M. *et al.* (2013) Brachypodium distachyon as a model plant toward improved biofuel crops: search for secreted proteins involved in biogenesis and disassembly of cell wall polymers. *Proteomics*, **13**, 2438–2454.

Dührkop, K., Fleischauer, M., Ludwig, M., Aksenov, A.A., Melnik, A.V., Meusel, M. *et al.* (2019) SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nature Methods*, **16**, 299–302.

Dührkop, K., Nothias, L.-F., Fleischauer, M., Reher, R., Ludwig, M., Hoffmann, M.A. *et al.* (2021) Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra. *Nature Biotechnology*, **39**, 462–471.

Ferdinando, M.D., Brunetti, C., Fini, A. & Tattini, M. (2012) Flavonoids as antioxidants in plants under abiotic stresses. In: Ahmad, P. & Prasad, M.N.V. (Eds.) *Abiotic stress responses in plants: metabolism, productivity and sustainability*. New York, NY: Springer, pp. 159–179.

Fernie, A.R. (2007) The future of metabolic phytochemistry: larger numbers of metabolites, higher resolution, greater understanding. *Phytochemistry*, **68**, 2861–2880.

Fukushima, A. & Kusano, M. (2013) Recent progress in the development of metabolome databases for plant systems biology. *Frontiers in Plant Science*, **4**, 73.

Gargallo-Garriga, A., Sardans, J., Pérez-Trujillo, M., Rivas-Ubach, A., Oravec, M., Vecerova, K. *et al.* (2014) Opposite metabolic responses of shoots and roots to drought. *Scientific Reports*, **4**, 6829.

Giri, A., Heckathorn, S., Mishra, S. & Krause, C. (2017) Heat stress decreases levels of nutrient-uptake and -assimilation proteins in tomato roots. *Plants*, **6**, 6.

Guijas, C., Montenegro-Burke, J.R., Domingo-Almenara, X., Palermo, A., Warth, B., Hermann, G. *et al.* (2018) METLIN: a technology platform for identifying knowns and unknowns. *Analytical Chemistry*, **90**, 3156–3164.

Hara, M., Furukawa, J., Sato, A., Mizoguchi, T. & Miura, K. (2012) Abiotic stress and role of salicylic acid in plants. In: Ahmad, P. & Prasad, M.N.V. (Eds.) *Abiotic stress responses in plants*. New York, NY: Springer, pp. 235–251.

Higashi, Y. & Saito, K. (2019) Lipidomic studies of membrane glycerolipids in plant leaves under heat stress. *Progress in Lipid Research*, **75**, 100990.

Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K. *et al.* (2010) MassBank: a public repository for sharing mass spectral data for life sciences. *Journal of Mass Spectrometry*, **45**, 703–714.

Huber, W., von Heydebreck, A., Sültmann, H., Poustka, A. & Vingron, M. (2002) Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*, **18**, S96–S104.

Huby, E., Napier, J.A., Baillieul, F., Michaelson, L.V. & Dhondt-Cordelier, S. (2020) Sphingolipids: towards an integrated view of metabolism during the plant stress response. *New Phytologist*, **225**, 659–670.

Itkin, M., Heinig, U., Tzfadia, O., Bhide, A.J., Shinde, B., Cardenas, P.D. *et al.* (2013) Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. *Science*, **341**, 175–179.

Kan, Y., Mu, X.-R., Zhang, H., Gao, J., Shan, J.-X., Ye, W.-W. *et al.* (2022) TT2 controls rice thermotolerance through SCT1-dependent alteration of wax biosynthesis. *Nature Plants*, **8**, 53–67.

Langfelder, P. & Horvath, S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**, 559.

Le Bris, P., Wang, Y., Barbereau, C., Antelme, S., Cézard, L., Legée, F. *et al.* (2019) Inactivation of LACCASE8 and LACCASE5 genes in Brachypodium distachyon leads to severe decrease in lignin content and high increase in saccharification yield without impacting plant integrity. *Biotechnology for Biofuels*, **12**, 181.

Li, B., Tang, J., Yang, Q., Cui, X., Li, S., Chen, S. *et al.* (2016) Performance evaluation and online realization of data-driven normalization methods used in LC/MS based untargeted metabolomics analysis. *Scientific Reports*, **6**, 38881.

Li, D., Halitschke, R., Baldwin, I.T. & Gaquerel, E. (2020) Information theory tests critical predictions of plant defense theory for specialized metabolism. *Science Advances*, **6**, eaaz0381.

Li, D., Heiling, S., Baldwin, I.T. & Gaquerel, E. (2016) Illuminating a plant's tissue-specific metabolic diversity using computational metabolomics and information theory. *Proceedings of the National Academy of Sciences of the United States of America*, **113**, E7610–E7618.

López-Cristoffanini, C., Bundó, M., Serrat, X., San Segundo, B., López-Carbonell, M. & Nogués, S. (2021) A comprehensive study of the proteins involved in salinity stress response in roots and shoots of the FL478 genotype of rice (Oryza sativa L. ssp. indica). *The Crop Journal*, **9**, 1154–1168.

Ludwig, M., Nothias, L.-F., Dührkop, K., Koester, I., Fleischauer, M., Hoffmann, M.A. *et al.* (2020) Database-independent molecular formula annotation using Gibbs sampling through ZODIAC. *Nature Machine Intelligence*, **2**, 629–641.

Macabuhay, A., Arsova, B., Walker, R., Johnson, A., Watt, M. & Roessner, U. (2022) Modulators or facilitators? Roles of lipids in plant root–microbe interactions. *Trends in Plant Science*, **27**, 180–190.

Mahieu, N.G., Genenbacher, J.L. & Patti, G.J. (2016) A roadmap for the XCMS family of software solutions in metabolomics. *Current Opinion in Chemical Biology*, **30**, 87–93.

Marriott, P.E., Sibout, R., Lapierre, C., Fangel, J.U., Willats, W.G.T., Hofte, H. *et al.* (2014) Range of cell-wall alterations enhance saccharification in Brachypodium distachyon mutants. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 14601–14606.

Mock, A., Warta, R., Dettling, S., Brors, B., Jäger, D. & Herold-Mende, C. (2018) MetaboDiff: an R package for differential metabolomic analysis. *Bioinformatics*, **34**, 3417–3418.

Moore, W.M., Chan, C., Ishikawa, T., Rennie, E.A., Wipf, H.M.-L., Benites, V. *et al.* (2021) Reprogramming sphingolipid glycosylation is required for endosymbiont persistence in Medicago truncatula. *Current Biology*, **31**, 2374–2385.e4.

Nakamura, Y. (2013) Phosphate starvation and membrane lipid remodeling in seed plants. *Progress in Lipid Research*, **52**, 43–50.

Nothias, L.-F., Petras, D., Schmid, R., Dührkop, K., Rainer, J., Sarvepalli, A. *et al.* (2020) Feature-based molecular networking in the GNPS analysis environment. *Nature Methods*, **17**, 905–908.

Okazaki, Y., Otsuki, H., Narisawa, T., Kobayashi, M., Sawai, S., Kamide, Y. *et al.* (2013) A new class of plant lipid is essential for protection against phosphorus depletion. *Nature Communications*, **4**, 1510.

Perea-García, A., Andrés-Bordería, A., Huijser, P. & Peñarrubia, L. (2021) The copper-microRNA pathway is integrated with developmental and environmental stress responses in Arabidopsis thaliana. *International Journal of Molecular Sciences*, **22**, 9547.

R Core Team. (2020) *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rahmati Ishka, M. & Vatamaniuk, O.K. (2020) Copper deficiency alters shoot architecture and reduces fertility of both gynoecium and androecium in Arabidopsis thaliana. *Plant Direct*, **4**, e00288.

Salem, M.A., Perez de Souza, L., Serag, A., Fernie, A.R., Farag, M.A., Ezzat, S.M. *et al.* (2020) Metabolomics in the context of plant natural products research: from sample preparation to metabolite analysis. *Metabolites*, **10**, 37.

Schulten, A. & Krämer, U. (2018) Interactions between copper homeostasis and metabolism in plants. In: Cánovas, F.M., Lüttge, U. & Matyssek, R. (Eds.) *Progress in botany*, Vol. **79**. Cham: Springer International Publishing, pp. 111–146.

Schymanski, E.L., Ruttkies, C., Krauss, M., Brouard, C., Kind, T., Dührkop, K. *et al.* (2017) Critical assessment of small molecule identification 2016: automated methods. *Journal of Cheminformatics*, **9**, 22.

Shahaf, N., Rogachev, I., Heinig, U., Meir, S., Malitsky, S., Battat, M. *et al.* (2016) The WEIZMASS spectral library for high-confidence metabolite identification. *Nature Communications*, **7**, 12423.

Sheng, H., Jiang, Y., Rahmati, M., Chia, J.-C., Dokuchayeva, T., Kavulych, Y. *et al.* (2021) YSL3-mediated copper distribution is required for fertility, seed size and protein accumulation in Brachypodium. *Plant Physiology*, **186**, 655–676.

Šimura, J., Antoniadi, I., Široká, J., Tarkowská, D., Strnad, M., Ljung, K. *et al.* (2018) Plant Hormonomics: multiple phytohormone profiling by targeted metabolomics. *Plant Physiology*, **177**, 476–489.

Stief, A., Altmann, S., Hoffmann, K., Pant, B.D., Scheible, W.-R. & Bäurle, I. (2014) Arabidopsis miR156 regulates tolerance to recurring environmental stress through SPL transcription factors. *The Plant Cell*, **26**, 1792–1807.

Su, G., Morris, J.H., Demchak, B. & Bader, G.D. (2014) Biological network exploration with Cytoscape 3. *Current Protocols in Bioinformatics*, **47**, 8.13.1–8.13.24.

Sumner, L.W., Amberg, A., Barrett, D., Beale, M.H., Beger, R., Daykin, C.A. *et al.* (2007) Proposed minimum reporting standards for chemical analysis chemical analysis working group (CAWG) metabolomics standards initiative (MSI). *Metabolomics*, **3**, 211–221.

Sun, A.-Z., Chen, L.-S., Tang, M., Chen, J.-H., Li, H., Jin, X.-Q. *et al.* (2022) Lipidomic remodeling in Begonia grandis under heat stress. *Frontiers in Plant Science*, **13**, 843942.

Sunkar, R., Li, Y.-F. & Jagadeeswaran, G. (2012) Functions of microRNAs in plant stress responses. *Trends in Plant Science*, **17**, 196–203.

The Gene Ontology Consortium. (2019) The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Research*, **47**, D330–D338.

Tohge, T., Wendenburg, R., Ishihara, H., Nakabayashi, R., Watanabe, M., Sulpice, R. *et al.* (2016) Characterization of a recently evolved flavonol-phenylacyltransferase gene provides signatures of natural light selection in Brassicaceae. *Nature Communications*, **7**, 12399.

Tsugawa, H., Ikeda, K., Takahashi, M., Satoh, A., Mori, Y., Uchino, H. *et al.* (2020) A lipidome atlas in MS-DIAL 4. *Nature Biotechnology*, **38**, 1159–1163.

Tsugawa, H., Kind, T., Nakabayashi, R., Yukihira, D., Tanaka, W., Cajka, T. *et al.* (2016) Hydrogen rearrangement rules: computational MS/MS fragmentation and structure elucidation using MS-FINDER software. *Analytical Chemistry*, **88**, 7946–7958.

Valladares, F., Sanchez-Gomez, D. & Zavala, M.A. (2006) Quantitative estimation of phenotypic plasticity: bridging the gap between the evolutionary concept and its ecological applications. *Journal of Ecology*, **94**, 1103–1116.

Wang, M., Carver, J.J., Phelan, V.V., Sanchez, L.M., Garg, N., Peng, Y. *et al.* (2016) Sharing and community curation of mass spectrometry data with global natural products social molecular networking. *Nature Biotechnology*, **34**, 828–837.

Wang, M., Schäfer, M., Li, D., Halitschke, R., Dong, C., McGale, E. *et al.* (2018) Blumenols as shoot markers of root symbiosis with arbuscular mycorrhizal fungi. *eLife*, **7**, e37093.

Wang, Z., Tian, X., Zhao, Q., Liu, Z., Li, X., Ren, Y. *et al.* (2018) The E3 ligase DROUGHT HYPERSENSITIVE negatively regulates cuticular wax biosynthesis by promoting the degradation of transcription factor ROC4 in rice. *The Plant Cell*, **30**, 228–244.

Wewer, V., Brands, M. & Dörmann, P. (2014) Fatty acid synthesis and lipid metabolism in the obligate biotrophic fungus Rhizophagus irregularis during mycorrhization of Lotus japonicus. *The Plant Journal*, **79**, 398–412.

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G.V. & Provart, N.J. (2007) An "electronic fluorescent pictograph" browser for exploring and analyzing large-scale biological data sets. *PLoS One*, **2**, e718.

Wu, X., Zhang, S., Liu, X., Shang, J., Zhang, A., Zhu, Z. *et al.* (2020) Chalcone synthase (CHS) family members analysis from eggplant (Solanum melongena L.) in the flavonoid biosynthetic pathway and expression patterns in response to heat stress. *PLoS One*, **15**, e0226537.

Xin, M., Wang, Y., Yao, Y., Xie, C., Peng, H., Ni, Z. *et al.* (2010) Diverse set of microRNAs are responsive to powdery mildew infection and heat stress in wheat (Triticum aestivum L.). *BMC Plant Biology*, **10**, 123.

Yamasaki, H., Hayashi, M., Fukazawa, M., Kobayashi, Y. & Shikanai, T. (2009) SQUAMOSA promoter binding protein–Like7 is a central regulator for copper homeostasis in Arabidopsis. *The Plant Cell*, **21**, 347–361.

Zhang, R.-Q., Zhu, H.-H., Zhao, H.-Q. & Yao, Q. (2013) Arbuscular mycorrhizal fungal inoculation increases phenolic synthesis in clover roots via hydrogen peroxide, salicylic acid and nitric oxide signaling pathways. *Journal of Plant Physiology*, **170**, 74–79.

Zu, P., Boege, K., Del-Val, E., Schuman, M.C., Stevenson, P.C., Zaldivar-Riverón, A. *et al.* (2020) Information arms race explains plant-herbivore chemical communication in ecological communities. *Science*, **368**, 1377–1381.