PNAS Nexus

PNAS Nexus, 2022, **1**, 1–14
https://doi.org/10.1093/pnasnexus/pgac068
Advance access publication date: 4 July 2022
Research Report

Crop genetic diversity uncovers metabolites, elements, and gene networks predicted to be associated with high plant biomass yields in maize

Mohsen Hajheidari[a,1], Nina Gerlach[a], Kristof Dorau[b], M. Amin Omidbakhshfard[c], Lina Pesch[a], Jörg Hofmann[d], Asis Hallab[e], Gabriel Y. Ponce-Soto[e], Anastasiya Kuhalskaya[c], David B. Medeiros[c], Amélia Bourceret[f,2], the RECONSTRUCT Consortium[g], Björn Usadel[e,h], Jochen Mayer[i], Alisdair Fernie[c], Tim Mansfeldt[b], Uwe Sonnewald[d] and Marcel Bucher[a,*]

[a]Institute for Plant Sciences, Cologne Biocenter, Cluster of Excellence on Plant Sciences, University of Cologne, D-50674 Cologne, Germany
[b]Faculty of Mathematics and Natural Sciences, Department of Geosciences, Institute of Geography, University of Cologne, Albertus-Magnus-Platz, D-50923 Köln, Germany
[c]Max Planck Institute of Molecular Plant Physiology, Department of Molecular Physiology, D-14476 Potsdam-Golm, Germany
[d]Division of Biochemistry, Department of Biology, Friedrich-Alexander-University Erlangen-Nürnberg, D-91054 Erlangen, Germany
[e]Bioinformatics (IBG-4), Forschungszentrum Jülich GmbH, D-52425 Jülich, Germany
[f]Department of Plant Microbe Interactions, Max Planck Institute for Plant Breeding Research, D-50829 Cologne, Germany
[g]RECONSTRUCT Consortium: retrieved November 7 2021 from: https://www.pflanzenforschung.de/de/forschung-plant-2030/projekte/45/detail
[h]HHU Düsseldorf, Institute of Biological Data Science, Cluster of Excellence on Plant Sciences, D-40225 Düsseldorf, Germany
[i]Agroscope, Department of Agroecology and Environment, CH-8046 Zurich, Switzerland
*To whom correspondence may be addressed: Email: m.bucher@uni-koeln.de
[1]Present address: Center for Genomics and Systems Biology, Department of Biology, New York University, 12 Waverly Pl, New York, NY 10003, USA.
[2]Present address: Institut de Systématique, Evolution, Biodiversité, Muséum National d'Histoire Naturelle, CNRS, Sorbonne Université, EPHE, Université des Antilles, Paris, France.
Edited By: Karen E. Nelson

Abstract

Rapid population growth and increasing demand for food, feed, and bioenergy in these times of unprecedented climate change require breeding for increased biomass production on the world's croplands. To accelerate breeding programs, knowledge of the relationship between biomass features and underlying gene networks is needed to guide future breeding efforts. To this end, large-scale multi-omics datasets were created with genetically diverse maize lines, all grown in long-term organic and conventional cropping systems. Analysis of the datasets, integrated using regression modeling and network analysis revealed key metabolites, elements, gene transcripts, and gene networks, whose contents during vegetative growth substantially influence the build-up of plant biomass in the reproductive phase. We found that S and P content in the source leaf and P content in the root during the vegetative stage contributed the most to predicting plant performance at the reproductive stage. In agreement with the Gene Ontology enrichment analysis, the *cis*-motifs and identified transcription factors associated with upregulated genes under phosphate deficiency showed great diversity in the molecular response to phosphate deficiency in selected lines. Furthermore, our data demonstrate that genotype-dependent uptake, assimilation, and allocation of essential nutrient elements (especially C and N) during vegetative growth under phosphate starvation plays an important role in determining plant biomass by controlling root traits related to nutrient uptake. These integrative multiomics results revealed key factors underlying maize productivity and open new opportunities for efficient, rapid, and cost-effective plant breeding to increase biomass yield of the cereal crop maize under adverse environmental factors.

Keywords: systems biology, metabolome, ionome, gene networks, biomass

Significance Statement:

The increasing demand for food from a growing population and the effects of human-induced climate change pose a major challenge for sustainable food production and ecosystem conservation. Plant productivity in the form of yield and biomass is the most important and complex trait of crops. It reflects the interaction of the environment with all growth and development processes that take place throughout the life cycle of the plant. System-level studies are able to reveal the overall strategy of the plant in response to environmental conditions, which also includes a number of key nutrients, metabolic, and genetic factors underlying plant productivity. These results open up new possibilities for biomass-oriented breeding of cereal crops.

Competing Interest: The authors declare no competing interest.
Received: January 15, 2022. **Accepted:** June 29, 2022

Introduction

Due to the complexity of biological and physiological processes affecting agronomic traits in crops, various quantitative genetic and systems biology approaches have been used for deeper understanding and prediction of such traits with measurements at early developmental stages that can lead to efficient, rapid, and cost-effective crop improvement. Complex agronomic traits in maize and other crops are controlled by a large number of small-effect genes, and thousands of associations between molecular markers and traits have been reported for many complex traits in different plant species. Quantitative trait loci-based marker-assisted selection has been used in trials to improve agronomic traits using genetic variation (1, 2). Genome-wide association studies have revolutionized the deciphering of genotype–phenotype associations in many species. However, the identified single nucleotide polymorphisms often explain only a small part of the heritability of complex traits. Therefore, our understanding of the genetic architecture of complex agronomic traits such as green biomass yield is still limited (3–6). Systems biology-based approaches integrate several types of large data sets derived from transcriptome, proteome, and metabolome profiles, combined with network analysis and machine learning approaches, to decipher the complexity of plant systems and better predict complex traits. Considering that important metabolic pathways directly affect plant biomass, metabolites have been widely used to predict plant performance (5–7). Throughout their life, plants assimilate elements such as carbon (C), nitrogen (N), phosphorus (P), and sulfur (S), oxygen and hydrogen to synthesize building blocks that determine the metabolite profile, growth, and development of the plant. Therefore, we hypothesized that the integrated analysis of elemental profiles with other omics datasets, and, more broadly, the analysis of more comprehensive datasets of genetically diverse lines growing in different environments, may ultimately lead to the discovery of key factors involved in the control of complex agronomic traits such as green biomass in crops.

Our integrated study of large-scale lipidome, metabolome, ionome, and transcriptome profiling conducted using machine learning approaches, provides evidence that S and P content in the source leaf and P content in the root, in combination with steady-state levels of a number of key metabolites in plant material collected during vegetative growth, are predictors of green biomass built up during the reproductive phase before harvesting. Our data also suggest biological networks with a number of genes whose expression during the vegetative stage could be used to predict plant biomass at the reproductive stage. Furthermore, a comparative analysis of different maize genotypes grown on cropland with mineral soil fertility management shows a large genetic variation in the response to phosphate deficiency in selected maize lines. This suggests that the assimilation and allocation of C and N to the root system during vegetative growth depends on the genotype and significantly influences the final plant biomass.

Results and Discussion

Variance structure of trial fields

Using a systems biology approach, we wanted to understand the contribution of genetic factors in combination with environmental factors to maize growth in the field and under controlled conditions, and therefore, studied four genetically distant maize lines (Dent line B73, Flint lines DK105 and F2, and Iodent line PH207). We also included the *pht1;6* mutant, which has a defect in the mycorrhiza-specific Pi transporter *PHT1;6* gene in the B73 line and, as a result, a drastic reduction in symbiotic uptake of inorganic phosphate (Pi) in soils with low P input to see the impact of the mycorrhizal Pi uptake pathway on factors affecting biomass yield (8, 9).

We planted maize at two geographically different field sites: the DOK site (dynamic, organic, and conventional managements, since 1978) in Therwil in the south of Basel, and the DEMO site (fertilizer demonstration trial, since 1987) in Reckenholz, Zürich in Switzerland. The field design included eight different plots, one plot for each soil management (NK and NPK) at the DEMO site, and three replicate plots for conventional mineral fertilization (CONMIN) and biodynamic mixed farming (BIODYN) soil managements at the DOK site. The soil type at the DEMO site is a Gleyic Cambisol and at the DOK site a Haplic Luvisol according to the IUSS WRB Working Group (10, 11). To show the overall variation between the study sites or between the different plots, respectively, we analyzed 38 physico-chemical and mineralogical parameters of the top soil in the range from a depth of 0 to 20 cm before sowing (Table S1, Supplementary Material). This revealed some local soil heterogeneity, which can be explained by factors outside of the geographical location (Table S1 and Figures S1–S4, Supplementary Material). There were profound differences in texture between the study sites, with DEMO generally having higher sand contents than DOK (Table S1 and Figure S2, Supplementary Material). DOK exhibited plot-level variations within the CONMIN and BIODYN management, and had higher clay contents in plots 90 and 96 than the surrounding plots (Figure S2, Supplementary Material). A total of 78.1% of the between-plots variances were explained by principal components 1 (51.2%) and 2 (26.9%), to which the following parameters contributed most: C_{org} > N > Ca > POXC > Feo > pH (H2O) >C/N > Mg and K > P > CEC > Fe_d > BS > clay > silt, respectively (Figure S1, Supplementary Material).

Genotype is the major factor determining plant biomass, metabolite profile, lipids, and elements

We exploited the heterogeneity of plots and tested the presumed interplant variation based on the dependencies between soil parameters and maize plant genotype affecting physiology and performance. On 240 plants (six replicates per genotype for each plot) at vegetative (V6) or reproductive (R2) stages of plant development, we investigated variation in five agronomic traits such as stem length, total leaf number, shoot fresh weight (SFW), and shoot dry weight (SDW; Table S2, Supplementary Material). Source leaves and root materials were then collected and used for high-throughput profiling of 188 metabolites (Tables S3–S5, Supplementary Material). Of these, 140 annotated lipids (including neutral lipids, phospholipids, and galactolipids) were analyzed by ultraperformance liquid chromatography coupled with Fourier transform mass spectrometry (UPLC-FT-MS; Table S3, Supplementary Material). A total of 26 carboxylic acids and highly phosphorylated intermediates from central metabolic pathways were analyzed using an HPLC system and electrospray ionization (ESI)/MS/MS (Table S4, Supplementary Material). The carbohydrates glucose, fructose, sucrose, and starch were measured with enzymatic assays, and 18 free amino acids with a reversed-phase HPLC system (Table S4, Supplementary Material). The accumulation of a particular element in organs and tissues is a genetically controlled complex process that is crucial for the uptake, binding, transportation, and sequestration of elements. To understand how elemental composition is regulated, profiling and quantification of 26 nutrients and trace elements were performed at the

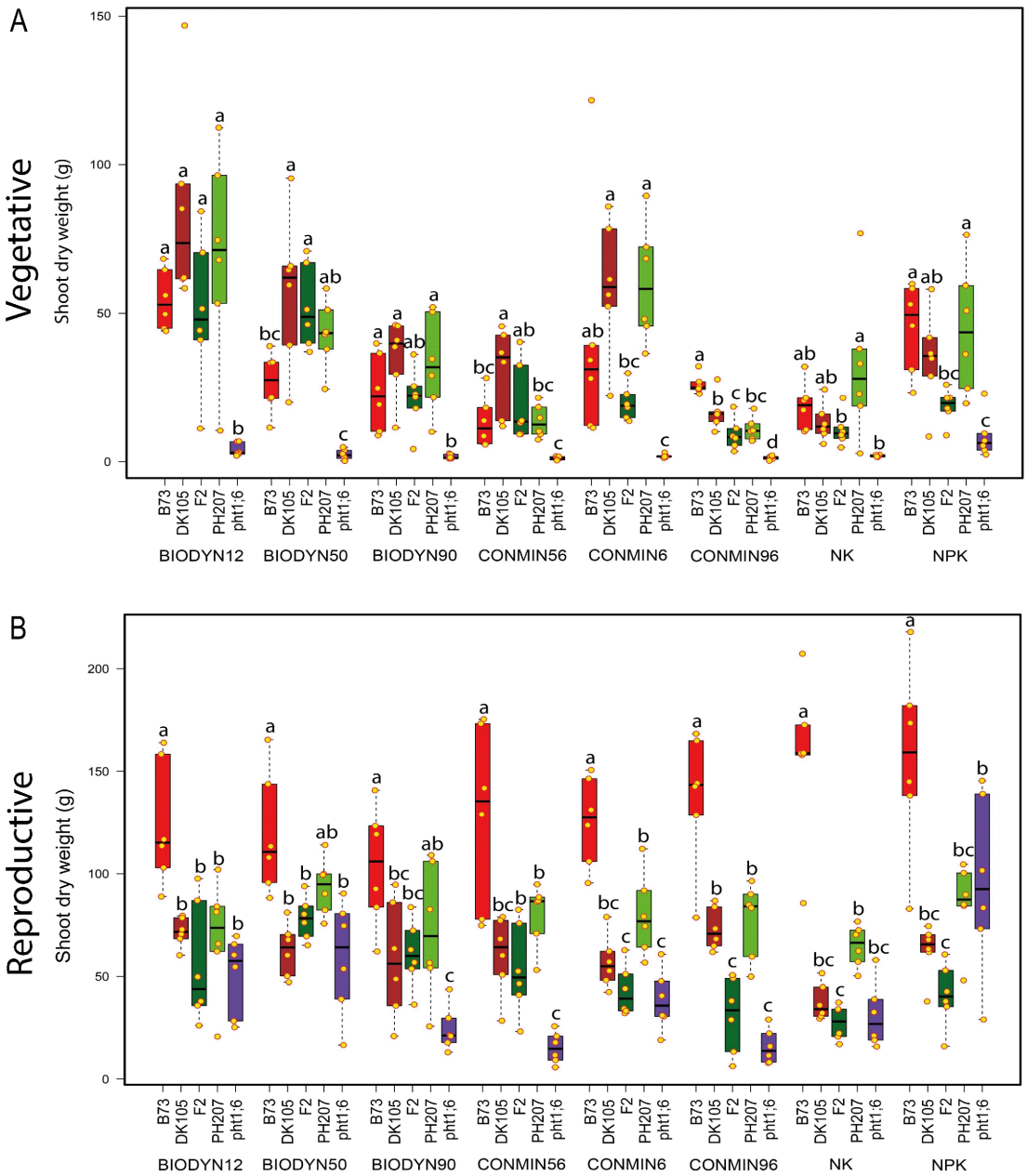

Fig. 1. Comparative analysis of SDW at vegetative and reproductive developmental stages. (A) The median dry weight of shoots of five maize genotypes grown on different plots at vegetative stage ($n = 6$). (B) The median dry weight of shoots of five maize genotypes grown on different plots at reproductive stage ($n = 6$). Significance groups were determined using ANOVA followed by Tukey's HSD test ($P < 0.05$). BIODYN and CONMIN are abbreviations for biodynamic mixed farming and conventional mineral fertilization, respectively. NK and NPK soil management stand for long-term fertilized soils with NK and NPK fertilizers, respectively.

genomic level using inductively coupled plasma mass spectrometry (ICP-MS; Table S5, Supplementary Material).

Permutational multivariate analysis of variance (PERMANOVA) revealed that genotype was a major source of variation in biomass/SDW of maize at vegetative ($P < 0.0001$, 64% of variance, Fig. 1A; Figure S5 and Table S6, Supplementary Material) and reproductive stages of development ($P < 0.0001$, 44% of variance, Fig. 1B; Figure S5 and Table S6, Supplementary Material).

A combination of principal component analysis (PCA), hierarchical clustering, and PERMANOVA showed that mainly genotype and, to a lesser extent, geographical location, followed by soil management, were critical in determining the metabolite profiles of maize (Figures S6–S8 and Table S7, Supplementary Material). PCA analysis with lipids from the source leaves showed that the samples obtained were separable from the DEMO and DOK sites. Hierarchical clustering placed the *pht1;6* mutant and the F2 line in different clusters in the heat map, indicating that the annotated lipid variation in the leaf samples was explained by genotype (PERMANOVA, $P < 0.0001$, 35% of variance), which in combination with site (PERMANOVA, $P < 0.0001$, 12% of variance) and soil management (PERMANOVA, $P < 0.0001$, 3% of variance) mainly determined leaf lipid content and composition. Similarly, to hierarchical clustering, PCA of lipids from root samples separated DEMO and DOK sites into two distinct groups, indicating a significant influence of geographical location (PERMANOVA, $P < 0.0001$, 23% of variance) on lipid content and composition in the root.

PCA analysis of carboxylic acids and highly phosphorylated intermediates from central metabolic pathways grouped source leaves from DEMO and DOK sites into two distinct clusters (Figure S7A, Supplementary Material), which attributed an important role to geographical location (PERMANOVA, $P < 0.0001$, 13% of variance) and soil management (PERMANOVA, $P < 0.01$, 2% of variance) in shaping the content and composition of compounds from the central metabolic pathways in the leaf (Table S7, Supplementary Material). Consistent with hierarchical clustering (Figure S7B, Supplementary Material), statistical analysis (PERMANOVA, $P < 0.0001$, 45% of variance) showed that genotype was the most important factor in shaping the metabolic profile in leaf central metabolism (Table S7, Supplementary Material). The content of pyruvate (Pyr), malate (Mal), phospho*enol*pyruvate (PEP), dihydroxyacetone phosphate (DHAP), 3-phosphoglycerate (3PG), fructose-1,6-bisphosphate (F16BP), ribulose-1,5-bisphosphate (R15BP), UDP-glucose (UDPglc), glucose-6-phosphate (G6P), fructose-6-bisphosphate (F6P), mannose-1-phosphate (M1P), glucose-1-phosphate (G1P), erythrose-4-phosphate (E4P), trehalose-6-phosphate (T6P), and UDP-N-acethylglucosamine (UDPNAG) were low in the *pht1;6* mutant and in DK105 plants, respectively, in all soil management types. This suggested that the Calvin cycle, C4 carbon fixation, glycolysis, and the pentose phosphate pathways are likely to function more efficiently in B73, F2, and PH207 compared to *pht1;6* and DK105. Alternatively, our data suggested that B73, F2, and PH207 became more resilient under conditions in which "metabolic sinks" that utilize sugar phosphate intermediates in anabolic processes downstream of the Calvin cycle and the pentose phosphate pathway become limited. Similarly, slightly increased metabolite pools in *pht1;6* grown on NPK-managed soils, thus providing more available Pi to plants, indicated Pi limitation of anabolism in *pht1;6* compared to the other differentially managed plots. A sharp increase in the citrate and shikimate levels, two stress markers for Pi deficiency and light stress, respectively, in DK105 at both geographical locations and under all soil management regimes indicated that DK105 was either a sensitive genotype or soil management was not optimal for this line (12, 13).

In the case of amino acids, neither maize genotypes nor soil management explained the observed variance in the PCA (Figure S8, Supplementary Material). However, hierarchical clustering presented in a heat map showed that a subset of amino acids, such as the branched-chain nonpolar amino acids isoleucine and valine, nonpolar glycine, polar histidine, and hydrophobic aromatic phenylalanine and tyrosine were high in *pht1;6* across all soil types in the source leaf. Moreover, the content of these amino acids was high in the leaves of B73, PH207, and DK105 plants, respectively, grown on NK-managed soils with a total P content of only 350 mg/kg compared to 670 mg/kg in NPK-managed soils (Figures S4 and S8, Supplementary Material). Overall, these data showed a genotype-specific relationship between the contents of a number of free amino acids in the source leaf in response to Pi limitation.

Multielemental profiling by ICP-MS followed by PCA and multivariate analysis showed that genotype was the main determinant of source leaf (PERMANOVA, $P < 0.0001$, 22% of variance) and root (PERMANOVA, $P < 0.0001$, 34% of variance; Table S7, Supplementary Material) element contents. Moreover, the contents of different elements in the source leaf were only partially influenced by soil management (PERMANOVA, $P < 0.0001$, 8% of variance) on NPK and BIODYN but not on NK or CONMIN plots, respectively (Figure S9 and Table S7, Supplementary Material). Taken together, these data suggest that plant biomass, metabolite profile, lipids, and elemental composition (Figures S5 and S6, Supplementary Material), in the source leaf and root are influenced by genetic input rather than soil management/geographical location (Figure S9, Supplementary Material).

Variance in nutrient levels in maize lines

To investigate the differences in element profiles between maize genotypes, the median element content in the source leaf and root samples of each line was calculated across all plots and analyzed by hierarchical clustering in the heat maps (Figures S10A and B, Supplementary Material). Phosphorus content showed a similar distribution pattern in the source leaf and roots of the different maize lines. Lines B73 and *pht1;6* contained the highest and lowest P contents, respectively (P content B73 > PH207 > F2 > DK105 > *pht1;6*). As already reported by Willmann et al. (9), a dramatic decrease in P content was observed in *pht1;6* plants in all soil management systems (Figures S10C and D, Supplementary Material). This showed that Pi uptake in maize lines grown in agroecosystems was largely dependent on the mycorrhizal Pi transporter PHT1;6 (Table S8, Supplementary Material).

In addition to the P content, sulfur (S) content, which is also a vital macronutrient for plant growth and development, was highest in the source leaf and root of the B73 line compared to other genotypes (Figures S10A and B, Supplementary Material). However, the *pht1;6* mutant had the highest S content in the root and showed a strongly reduced S content in the source leaf compared to the corresponding wild type B73 (Figures S10A and B). This confirms a higher remobilization of S from shoot to root under Pi deficiency in mycorrhizal maize plants, similar to mycorrhizal nonhost plants of Arabidopsis (14). However, it cannot be excluded that a disrupted MPU pathway results in reduced translocation of P and S to the shoot and concomitant accumulation of S in the root of *pht1;6* plants.

Potentially toxic elements hyperaccumulate in the roots

In contrast to the levels in the PH207 and DK105 lines, the levels of several potentially toxic metals such as aluminum (Al), cobalt (Co), lead (Pb), lithium (Li), and nickel (Ni) were lowest in the root of B73 and *pht1;6*, respectively, while they reached intermediate levels in F2 (Figures S10B, S11, and Table S5, Supplementary Material). Interestingly, selenium (Se) known for its protective role in metal toxicity showed a similar profile. Thus, considering that plant genotypes with increased tolerance to toxic elements accumulate less of these elements in the root, genotype B73 might be more tolerant to toxic elements (15). When clustering was based on root element content, the *pht1;6* mutant and B73 were grouped together. However, the leaf element content of these two lines was very different, and the potentially toxic elements mentioned above were highly accumulated only in the leaf of the *pht1;6* mutant (Figures S10A, S11, and Table S5, Supplementary Material). Interestingly, the accumulation of potentially toxic elements in the maize genotypes was several times higher in the root (Figure S12, Supplementary Material) than in the source leaf, while source leaf elements critical for plant growth, such as N, P, and K, had higher levels in the source leaf than in the root (Figure S12, Supplementary Material). This suggested that maize plants generally prevent the transport of potentially toxic elements to their above-ground part.

It is known that severe Pi deficiency can lead to acidification of the rhizosphere, resulting in the uptake and transport of a number of potentially toxic elements (16). Previous studies in *Arabidopsis* have shown that Pi deficiency leads to inhibition of primary root

growth due to accumulation and redistribution of Fe in the root (17). In our field study, Pi deficiency led to accumulation of Fe in the source leaf of *pht1;6* (Figures S10A, B, and S13, Supplementary Material), and we did not observe Fe accumulation in the roots of *pht1;6*. In agreement with these data, statistical analysis showed that the main source of Fe variation in the source leaf was PHT1:6 activity (Table S9, Supplementary Material). The genotype effect was high for Fe content in the root with and without *pht1;6* plants (ANOVA, < 0.0001, 57% and 50% of variance, respectively). However, removing of *pht1;6* from the source leaf data set reduced the genotype effect from 45% (ANOVA, $P < 0.0001$) to 4.1% (ANOVA, $P < 0.082$). Overall, these results suggest that genotype is the primary factor determining the content of certain elements in the leaf and root of maize lines and that in maize, unlike Arabidopsis, phosphate deficiency does not lead to the accumulation of Fe in the root.

Sulfur and phosphorus in combination with key metabolites are predictive of plant biomass

To build a statistical model with predictive power for total biomass accumulation, we then analyzed the relationships between SDW at the reproductive stage (SDW-Re) and the metabolites, lipids, and elemental compositions of source leaves and roots at the vegetative stage. Since a normal distribution cannot be assumed for all variables, we used a nonparametric measure of rank correlation.

Central metabolites

The pairwise Spearman's correlation coefficients between the SDW-Re and carbohydrates, major carboxylic acids, and highly phosphorylated intermediates of the central metabolic pathways from the source leaf samples at the vegetative stage showed that most of the central metabolites were strongly positively correlated with SDW-Re (Figure S14A and Table S10, Supplementary Material). The highest correlation was observed for PEP (49%), G6P (43%), and Mal (42%), which are critical for C4 carbon fixation, glycolysis, the pentose phosphate pathway, and starch and lipid biosynthesis. Of the carbohydrates measured at the vegetative stage, glucose and fructose content in the root and glucose content in the source leaf show a significant positive correlation with SDW-Re (Figure S14A and Table S10, Supplementary Material). The sucrose content in the source leaf of B73 is significantly lower and the glucose and fructose content in the source leaf and root significantly higher than in the other genotypes (Figure S15, Supplementary Material). In agreement with these results, we found that the expression pattern of several genes encoding transporters involved in carbon partitioning, such as sugar transporters, sucrose transporters, UDP–glucose transporters, nucleotide/sugar transporter family proteins, SWEET family members, and also sucrose synthases, showed a genotype-dependent pattern in the source leaves (Figure S16, Supplementary Material). Taken together, these results suggest that in the B73 line, sucrose is transported in greater amounts from the leaf to the root, where it is efficiently hydrolysed to glucose and fructose by the activity of invertases and synthases. Similar to the B73 line, the *pht1;6* mutant also transports sucrose in large amounts to the root. However, the mutant is not able to efficiently convert sucrose into hexoses. This is likely due to the strong decreased level of ATP and pyrophosphate (PPi) in the *pht1;6* mutant that are critical for the activity of invertases and sucrose synthases, respectively, and also subsequent ATP/PPi-dependent phosphorylation of hexoses (18, 19). It appears that the *pht1:6* mutant plants attempt to compensate for the low activity of invertases and synthases by increased expression of genes encoding invertases, sucrose synthases, and sugar transporters (Figure S16, Supplementary Material).

Lipids

Correlation studies between lipids from the source leaf and SDW-Re revealed that most of the MGDGs, DGDGs, SQDGs, DAGs, and TAGs were either negatively correlated or not correlated at all (Figures S14B, C, and Table S10, Supplementary Material). In contrast, almost all phospholipids (PCs, PEs, PSs, and PGs) exhibited either positive or no correlations, and phospholipids PC-36–2 (39%), PC-36–3–2 (38%), and PS-40–2–2 (37%) showed the highest correlations with SDW-Re (Figure S14B and Table S10, Supplementary Material). In the root samples, PC-38–2 (29%), PG-34–3 (28%), and PC-34–2 (27%) showed the highest positive correlations with SDW-Re (Table S10, Supplementary Material). Phospholipids are important components of cell membranes and are required for signal transduction cascades (20). They also play an important role in controlling plant growth and development under stress conditions (21). Previous studies have shown that environmental stress leads to extensive remodeling of lipid biosynthesis and consequently to changes in cell membrane composition (21, 22). For example, abiotic stresses lead to a gradual reduction in phospholipid content (23), and under conditions of Pi deficiency, plants replace phospholipids with nonphosphorous glycerolipids, thereby diverting Pi to other vital processes (24). We suggest that a high phospholipid content implies that the plants in question probably received sufficient resources for adequate growth and development and were able to avoid significant stress.

Amino acids

Of the free amino acids in the source leaf, methionine (25%), alanine (17%), and proline (15%) and in the root alanine (30%), glutamate (28%), and serine (27%) showed the highest positive correlations with SDW-Re (Table S10, Supplementary Material). Amino acids are usually multifunctional and do not only serve as building blocks of proteins. Some amino acids, including alanine, the most abundant soluble N containing molecule, are involved in the storage and long-distance transport of N (25). Glutamate is critical for N assimilation into the carbon skeleton, which occurs largely via the glutamine synthetase/glutamate synthase cycle. It is also central to amino acid metabolism and contributes to the regulation of the tricarboxylic acid (TCA) cycle. (26, 27). We have also found negative correlations between several amino acids, including glycine, phenylalanine, isoleucine, histidine, and tyrosine with SDW-Re (Table S10, Supplementary Material).

Potentially toxic elements

The contents of several potentially toxic elements in the source leaf at the vegetative stage, such as chromium (Cr), Se, beryllium (Be), Co, Li, Al, V, Fe, Ni, and cesium (Cs), showed a significant negative correlation with SDW-Re and P content (Figure S14D and Table S10, Supplementary Material), while in the root a considerable accumulation of these elements had no adverse effect. In the source leaf, P (57%), S (53%), Mg (45%), and N (31%) and in the root P (51%), C/N ratio (28%), Mo (15%), and Cu (12%) showed the highest positive correlations with SDW-Re (Table S10, Supplementary Material). The macronutrient potassium (K), an essential element for plant growth, showed a weak but significant negative correlation with SDW-Re (Figure S14D and Table S10, Supplementary Material). K does play an important role in plant growth. However, excess K can also cause stress and growth inhibition (28, 29). The

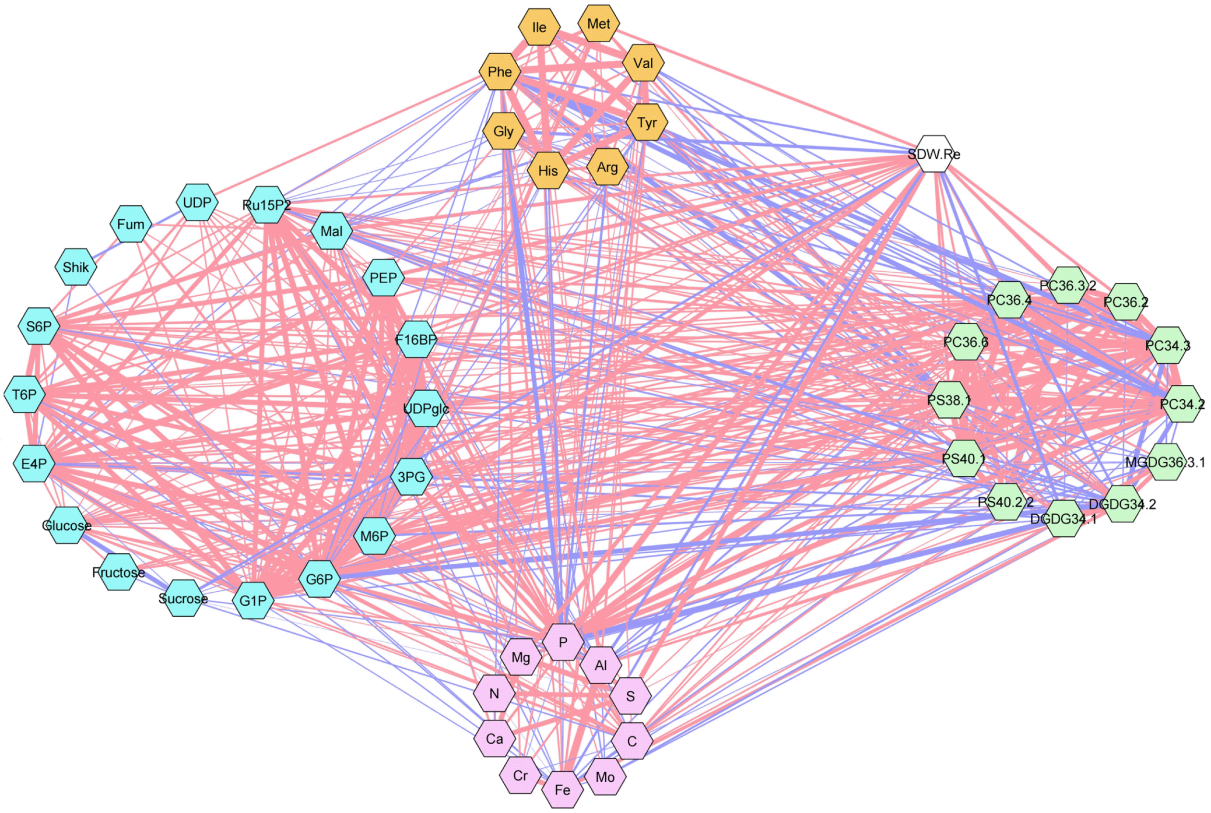

Fig. 2. Pairwise Spearman's rank correlations between dry weight of shoots at reproductive stage and metabolite/lipid/element/amino acid content at vegetative stage from source leaf samples. Network diagram showing some of the most significant correlations identified between dry weight of shoots at the reproductive stage with metabolites, lipids, elements, and amino acids. Positive correlations are represented by red lines and negative correlations by blue lines. The thickness of the lines indicates the strength of the correlations. The turquoise, orange, pink, and green hexagons represent the components of the central metabolites, amino acids, elements and lipids, respectively.

slight negative effect of K on SDW-Re could be due to a moderate excess of available potassium in our experimental fields for the genotypes studied.

Predictors of biomass

Next, we investigated the predictive power of the measured explanatory variables, which consisted of central metabolites, lipids, free amino acids, and elements for SDW-Re, by using a multivariate statistical method, partial least squares regression (PLSR), to identify the most important variables for predicting SDW-Re (30). PLSR is a suitable approach for building predictive models when the explanatory variables are highly correlated with each other (30), as was the case with our datasets (Fig. 2). We considered only those explanatory variables that had significant correlations according to Spearman's rank correlation coefficient and used a leave-one-out cross-validation method to determine the optimal number of latent variables based on the minimum root mean square error of prediction (RMSEP) (31). The PLSR coefficients estimated in the training set were used to predict the SDW-Re in the test set. The model using source leaf data showed a strong relationship between metabolite and element compositions ($R^2 = 0.7382$; Fig. 3A) with SDW-Re. To evaluate the contribution of each predictor of SDW-Re, we examined their regression coefficients. Among the explanatory variables, S (25%), P (10.7%), 3PG (5%), Ala (3%), Mal (2.9%), and PEP (1.5%) had the largest contribution to the PLSR model (Fig. 3C). We also built the PLSR model using the root data sets. The root model also showed strong predictive power of elements and metabolic composition for SDW-Re ($R^2 = 0.6861$; Fig. 3B). The P content in the root exerted the greatest influence (51%) in the PLSR model (Fig. 3D). Since the P content in the *pht1;6* mutant was low compared to other genotypes, we thought that the presence of the *pht1;6* data in our data set might cause a bias towards the P contribution in the root for SDW-Re. Therefore, the *pht1;6* data set was removed and a new model was created based on the dataset of the four lines studied, B73, PH207, F2, and DK105. In this model, P content still contributed by far the most to the prediction of SDW-Re (35%, Fig. 3E). In a further analysis, we removed the data sets from plants grown on NK soil management. In the constructed model, P content was still the most important factor for prediction (42%, Fig. 3F). These results are consistent with the importance of essential macronutrients as key components of a variety of essential biomolecules, including metabolites, and their particular contribution to plant biomass. Taken together, our data suggested that S and P content in the source leaf and P content in the root, which are mainly determined by genotype (Table S9, Supplementary Material), in combination with a number of important metabolites such as 3PG, Ala, Mal, and PEP are key to predicting plant biomass at the reproductive stage.

Genes and gene networks enable the prediction of plant biomass

To identify genes or gene networks critical for plant biomass, we performed Spearman's rank correlation coefficient statistics

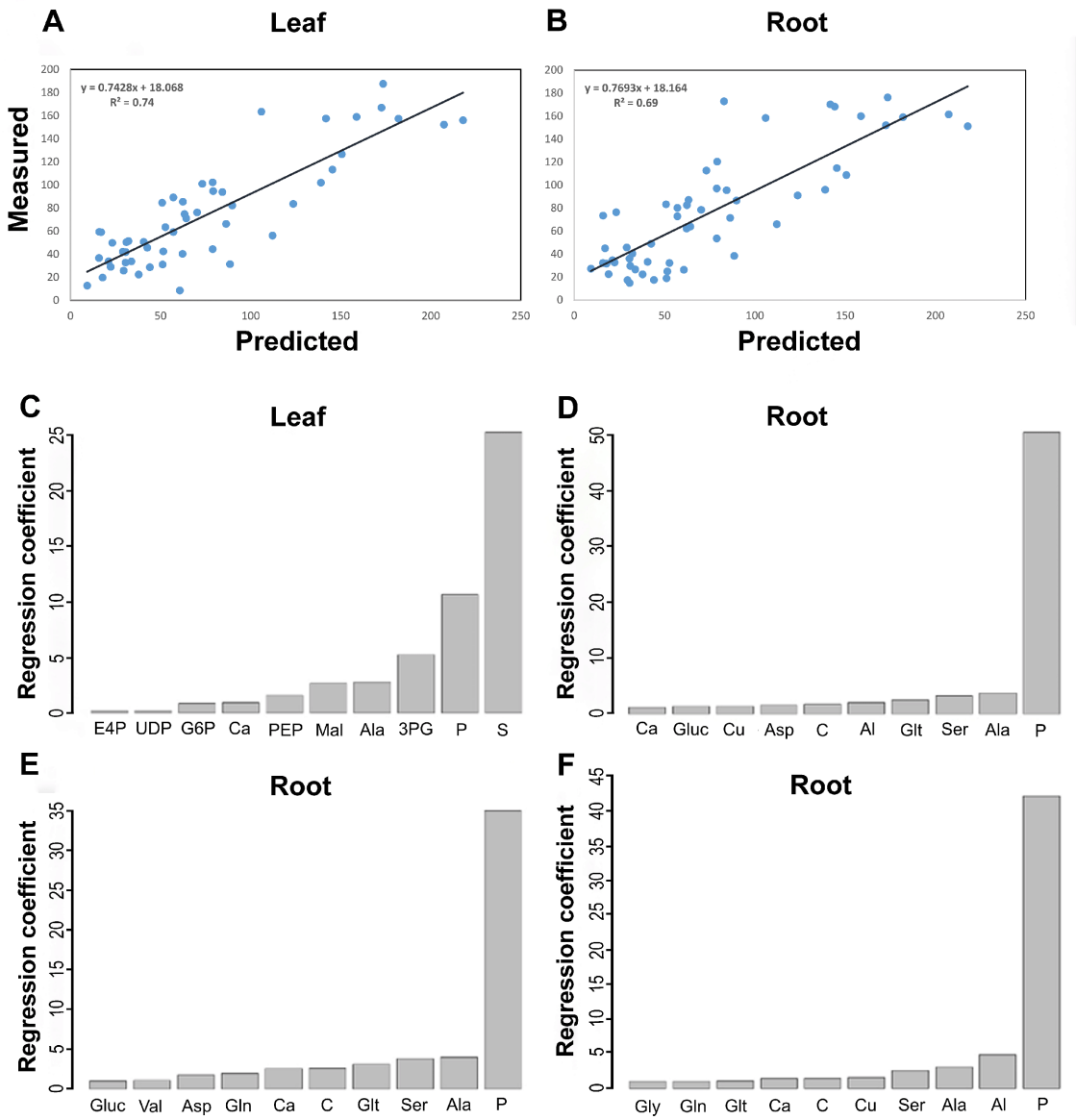

Fig. 3. Key elements S and P are the largest contributors to the PLSR model. (A) PLSR prediction of SDW-Re using metabolite and element compositions from the vegetative stage compared to measured values resulting from source leaf samples of five maize genotypes (B73, PH207, DK105, F2, and the *pht1;6*). (B) PLSR prediction of SDW-Re using metabolite and element compositions from the vegetative stage compared to measured values resulting from the root samples of 5 maize genotypes. (C) Regression coefficients of the top 10 elements and metabolites from the source leaf samples of five maize genotypes that contributed the most contribution to the PLSR model. (D) Regression coefficients of the top 10 elements and metabolites from the root samples of five maize genotypes that contributed the most to the PLSR model. (E) Regression coefficients of the top 10 elements and metabolites from the root samples of four maize genotypes (B73, PH207, DK105, and F2) that contributed the most to the PLSR model. (F) Regression coefficients of the top 10 elements and metabolites from the root samples that contributed the most to the PLSR model, excluding samples grown on NK soil management.

between transcript profiles (gene expression) at vegetative stage and SDW-Re. RNA-seq analysis was performed using 120 transcriptomes (three replicates per genotype for each plot) generated by Illumina paired-end sequencing. Genes that showed significant positive expression correlations with shoot biomass (padj < 0.05) were used as input to the STRING database (32). Networks generated by STRING with default settings were visualized in Cytoscape and partitioned into network modules. We identified seven main modules by analyzing the generated network based on transcript profiles of the source leaves (Fig. 4). Auxin and ABA are plant hormones that play essential roles in plant growth, development, and defense response (33, 34). Components of auxin (TIR1, CULLIN-1, SKP1, RBX1, AUX/IAAs, and ARFs) and ABA signaling pathways (PYLs, PP2Cs, and SNRK2s) were identified as hub genes in module-I and module-II, respectively. Genes encoding ribosomal proteins, RUBISCO, and nitrate reductases are central in module-III. Nitrate reductases and RUBISCO catalyze the first step of nitrogen assimilation and carbon fixation, respectively (35, 36). Ribosomal proteins play a vital role in the cellular process of translation (37). Light receptors such as PHYAs, CRY2, PHOT1, and PHOT2, which are crucial to synchronize plant growth and development through the sensing of light in the environment (38), are at the center of module-IV. Autophagy-related genes, which are critical for recycling of malfunctioning cellular components and nutrient allocation (39, 40), are at the heart of module-V (ATG1a, ATG1t, ATG4b, ATG6, ATG8b, ATG8d, ATG11, ATG13a, ATG13b, and

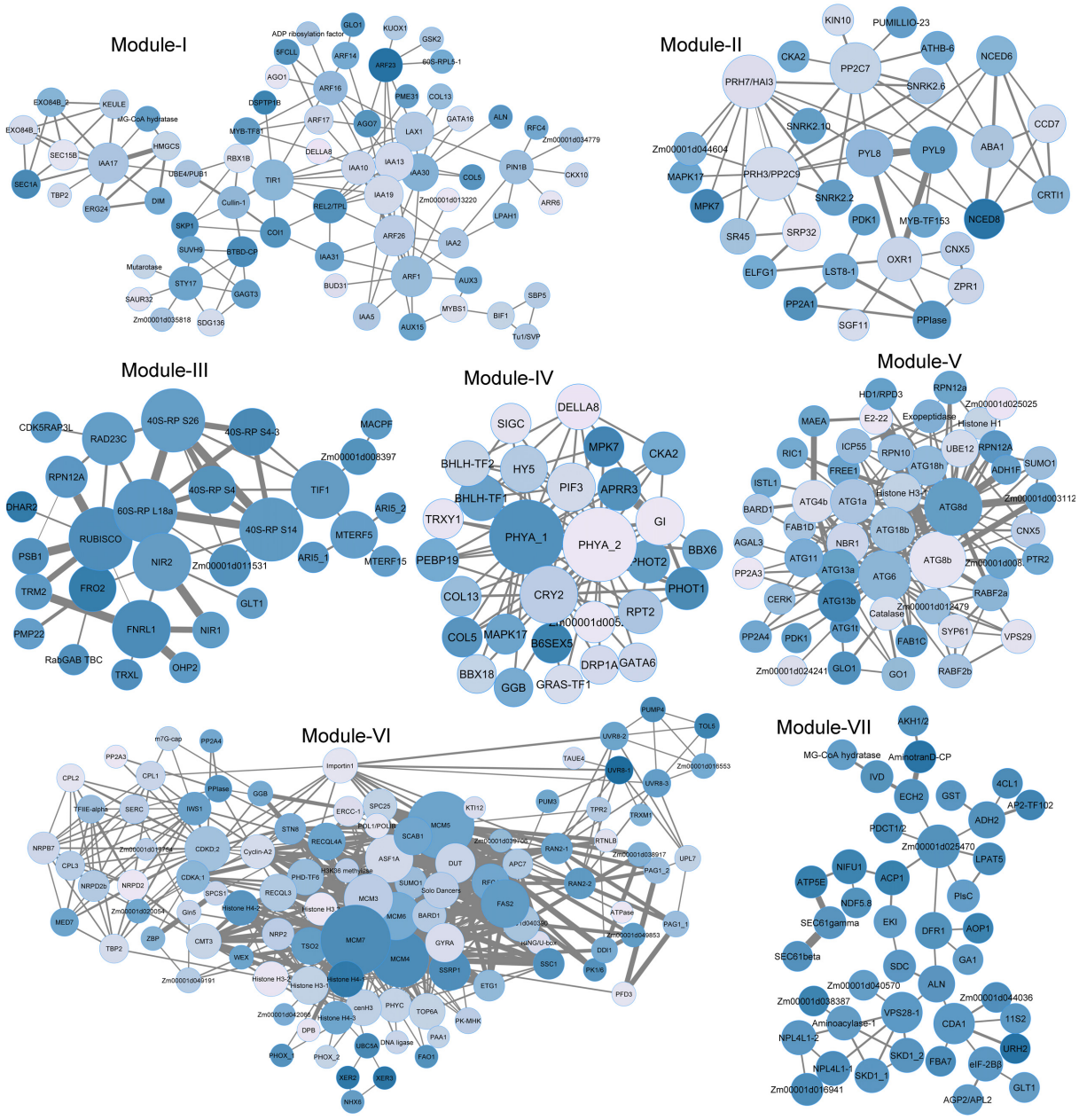

Fig. 4. Modules identified by unweighted network analysis. Genes whose expressions in the source leaf at the vegetative stage showed significant positive correlations (Spearman's rank correlation, *padj* < 0.05) with shoot biomass at the reproductive stage were used as input to the STRING database. The network generated by STRING with default settings was visualized in Cytoscape and seven main network modules was identified. Physical or functional connections are represented by lines. The size of the nodes is proportional to the connectivity of the individual genes within each module. The color intensity of the nodes is proportional to the correlation of gene expression levels with shoot biomass. The edge thickness represents the strength of pairwise coexpression of genes.

ATG18b). Genes involved in DNA replication, RNA transcription, and cell division are central to module-VI. For example, members of the MINICHROMOSOME MAINTENANCE (MCM) complex (MCM3, MCM4, MCM5, MCM6, and MCM7), that are involved in the initiation of DNA replication, are at the center of this module (Fig. 4) (41). UV-B receptors (UVR8s) and cyclin-dependent kinase CDKA;1, which is the master regulator of the cell cycle and controls both G1 to S and G2 to M checkpoints, are present in this module (42, 43). Module-VII contains genes that are critical for phospholipid biosynthesis and the recycling and degradation of plasma membrane proteins (Fig. 4). For example, ethanolamine kinase (EKI) catalyzes the first step of PE biosynthesis, and serine decarboxylase (SDC) catalyzes the biosynthesis of ethanolamine from serine, which is crucial for the synthesis of choline, PEs, and PCs (44, 45). Vacuolar protein sorting-associated protein 28 (VPS28-1) is a component of the ESCRT-I complex (endosomal sorting complex required for transport I), a key regulator of the vesicular trafficking process (46). From the genes presented in the seven modules, based on the connectivity of the genes and the correlation coefficient, we generated a shortlist of candidate marker genes whose expression is likely to be critical for achieving high plant biomass (Table S11, Supplementary Material). In contrast to

the source leaf, we found no modules or pathways when using the root transcriptome data for unweighted correlation network analysis. In addition to unweighted coexpression network analysis, we conducted weighted gene coexpression network analysis (WGCNA) to identify modules (clusters of highly interconnected genes that are represented by colors) during vegetative growth that may be critical for plant biomass formation (47). However, the modules resulting from the source leaf or root samples did not show a strong correlation with SDW-Re (Figure S17, Supplementary Material). The gray module from the source leaf samples containing genes that were not coexpressed with the genes present in other modules showed the highest correlation (47%, $P < 0.0001$) with SDW-Re. This suggested that weighted coexpression network analysis may not be an appropriate approach for identifying gene sets that predict plant biomass. However, some of the modules showed strong correlation coefficients with genotypes. For example, the brown module showed only a strong correlation coefficient (96%) with F2 genotype suggesting that certain signaling pathways or gene networks are strongly associated with the F2 genotype (Figures S18, S19, and Table S12, Supplementary Material).

To identify genes or gene networks that are critical for plant biomass, we conducted an unweighted network analysis, considering physically and/or functionally related genes. With this approach, we risked to losing poorly characterized genes whose expression was crucial for plant biomass. Therefore, we applied PLSR to identify genes whose transcripts predicted SDW-Re with high significance during vegetative growth. For the source leaf samples the genes whose expression had a correlation coefficient of at least 50% with SDW-Re according to the Spearman's rank correlation coefficient were considered. The PLSR model showed strong predictive power of the transcripts for the SDW-Re ($R^2 = 0.6993$; Figure S20, Supplementary Material). Of the top 15 genes with the highest contribution to the prediction of SDW-Re (Table S13, Supplementary Material), only two of them [*PHOTOTROPIN 1*, Zm00001d044599 and *ALANINE AMINOTRANSFERASE* (*AlaAT*) 12, Zm00001d030557] were included in the list of genes generated by the unweighted correlation network analysis (Table S11, Supplementary Material). PHOTOTROPIN 1 is a blue light receptor that promotes growth and development by optimizing photosynthetic light capture under low light conditions (48), and alanine aminotransferases is a N metabolism gene that is essential for the photosynthetic C4 assimilation cycle. Expression of *AlaAT* under the control of a stress-inducible promoter leads to increased biomass in various crops. This increase is probably due to the improvement in nitrogen use efficiency (49, 50). The top gene in the list is a nucleocytoplasmic EULS3-like lectin, which may be involved in ABA-dependent stomatal closure (51) and regulates water transpiration and carbon fixation. Among the root transcripts, a small proportion of genes showed a significant correlation coefficient of more than 50% with the SDW-Re. Therefore, the genes whose expression showed a correlation coefficient of 30% or more with the SDW-Re were considered. The model built with the root transcripts also showed high predictive power ($R^2 = 0.6787$; Figure S20, Supplementary Material). The top 15 genes that contributed the most to the PLSR model are listed in Table S13 (Supplementary Material). The major gene whose transcription performed best in predicting plant biomass is a BURP domain-containing protein (RD22-like gene, Zm00001d031909), which is likely localized in the cell wall matrix and contributes to cell wall loosening by interacting with expansins. Overexpression of an RD22-like gene led to a significant increase in seed mass and plant biomass in cotton and *Arabidopsis* (52, 53).

Analysis of growth phenotype and transcriptome reveals beneficial root traits and gene networks in response to phosphate deficiency in B73

Transcriptional response to Pi deficiency

The machine learning approach has shown that P content in the source leaf and root during the vegetative stage is a crucial factor for predicting plant biomass at the reproductive stage. Thus, to explore the physiological/biological basis of the high predictive power of P content, we conducted a comparative transcriptome analysis of maize plants grown in plots with NK and NPK soil management at the DEMO site. Apart from some biological processes, such as the cellular response to Pi starvation, which were upregulated in all genotypes, we observed a very different response to Pi starvation between genotypes at the source leaf and root levels on NK soil (Figures S21–S23 and Table S14, Supplementary Material). Surprisingly, we found a unique response in the root of the genotype B73 to Pi starvation compared to PH207, F2, and DK105. Signaling pathways involved in processes such as root growth and development, root system development, root hair development and elongation, root morphogenesis, cell growth, cell wall organization or biogenesis, regulation of auxin-mediated signaling, and response to auxin and auxin transport were upregulated only in the root of B73 (Figure S23 and Table S14, Supplementary Material). Although Pi deficiency in B73 resulted in a significant reduction in growth at the vegetative stage, B73 was the only genotype that did not show a significant reduction in SDW at the reproductive stage under Pi deficiency condition and had the highest SDW-Re across all soil management treatments at both geographical locations (DEMO and DOK; Fig. 1A and B; Figure S5, Supplementary Material). Consistent with these results, the SDW of the *pht1;6* mutant grown in the NPK soil management plot was lower than that of the corresponding wild type B73 and other genotypes at the vegetative stage, while its SDW-Re was similar to that of lines PH207 and DK105 and significantly higher than that of the F2 line. This indicated that *pht1;6* was able to catch up in growth during the late developmental phase. The heat map of the genes involved in the Pi starvation response showed that this response was significantly more pronounced in *pht1;6* than in the other maize genotypes studied (Figure S24A, Supplementary Material). Comparative transcript profiling showed that several Pi transporters, SPX domain-containing proteins, INDUCED BY PHOSPHATE STARVATION 1 (IPS1), phosphatases, RNases, β-glucosidase, and PEPCases were strongly upregulated in the *pht1;6* mutant compared to wild type B73 (when grown on NPK-managed soils) to increase Pi uptake and Pi utilization efficiency.

Chromatin-based regulation of root development

Within the upregulated genes in the root of *pht1;6*, GO-terms related to histone H3 lysine 9 (H3K9) methylation, DNA methylation, chromatin silencing, cell division, DNA replication, root development, root hair development, and root hair elongation were significantly enriched (Fig. 5A; Tables S15 and S16, Supplementary Material). Network analysis of the differentially expressed genes showed that epigenetic silencing, DNA replication, cell division, and root/root hair development and growth pathways were strongly linked (Fig. 5A). During cell division, in addition to DNA sequence, the restoration of epigenetic information in daughter cells is critical for maintaining genome stability and controlling the transcriptional status, which in turn is important for maintaining the identity of a particular cell type or its regulated differentiation (54). However, after DNA replication, the establishment of silencing histone methylation marks is a gradual and

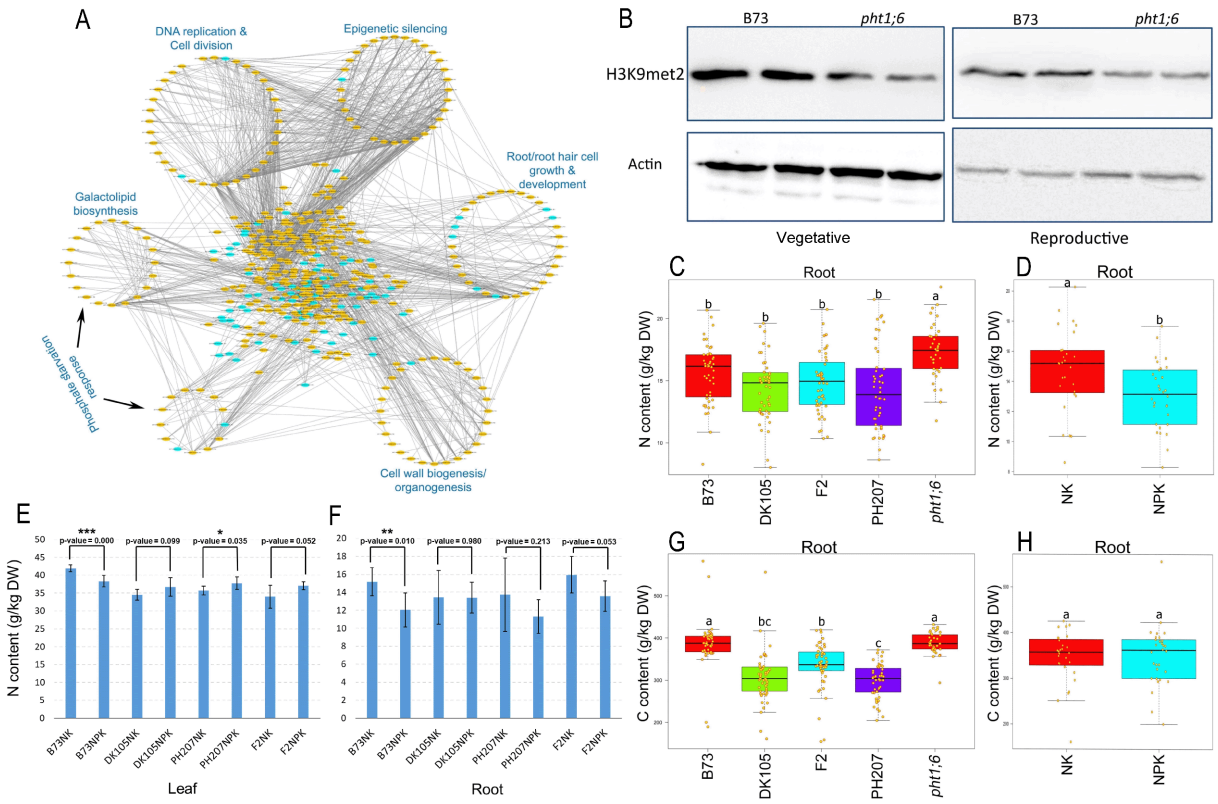

Fig. 5. Pi deficiency promotes root growth and development by promoting DNA replication and cell division. (A) Epigenetic silencing pathways are closely linked with DNA replication, cell division, and root development pathways. Differentially expressed genes ($P < 0.05$ and at least two-fold up- or downregulation) identified by comparative transcript profiling of the *pht1;6* mutant and wild-type B73 grown on NPK soils were subjected to the STRING database. The generated network was visualized using Cytoscape. The upregulated and downregulated differentially expressed genes are represented by yellow and blue colors, respectively. (B) Quantitative analysis of H3K9met2 in two biological replicates at two developmental stages. Equal loading of protein samples is indicated by immunoblotting with antiactin antibody. (C) The median N content of maize lines from root samples grown in all soil management types ($n = 48$). (D) The median N content of maize genotypes from the root samples grown on NK and NPK soil ($n = 30$). (E) The average N content of the maize lines from the source leaf samples grown on NK and NPK soils. The error bars indicate ± 1 SD ($n = 6$). A two-tailed Student's t test was used to calculate significance. (F) The mean N content of maize lines from root samples grown on NK and NPK soils. Error bars indicate ± 1 SD ($n = 6$). A two-tailed Student's t test was used to calculate significance. (G) The median C content of maize lines from root samples grown in all soil treatments ($n = 48$). (H) The median C content of the maize genotypes from the root samples grown in NK and NPK soils ($n = 30$). The significance of the groups was determined using ANOVA followed by Tukey's HSD test ($P < 0.05$).

slow process, and the deposition of new naïve histones with the old recycled histones leads to a twofold dilution of the parental post-translational modification (55–57). Therefore, we hypothesized that if Pi deficiency in the *pht1;6* mutant grown on NPK-managed soil leads to increased DNA replication, cell division, and root growth, the level of histone H3K9 dimethylation (H3K9me2)-dependent silencing should show a significant reduction in *pht1;6* compared with B73. We, therefore, performed immunoblotting analysis to examine H3K9me2 levels in the roots of *pht1;6* and B73 lines grown with NPK. H3K9me2 levels were significantly reduced in *pht1;6* (Fig. 5B). In agreement with the above results, we found that at the physiological level N allocation to the root during vegetative growth was dependent on genotype and Pi deficiency (Figs 5C–F; Figure S24B, Supplementary Material). Similar to *pht1;6*, genotype, B73 assimilated and/or allocated more N to the root under Pi deficiency (Fig. 5F). In contrast to N, the allocation of C to the root during vegetative growth was only dependent on the genotype and was not affected by Pi deficiency. Genotype B73 and *pht1;6* allocated more C to the root than lines F2, PH207, and DK105, respectively (Figs 5G and H; Figure S24C, Supplementary Material). In a complementary independent experiment, we grew the maize genotypes in a hydroponic system with sufficient Pi (SP). After 1 week, the endosperms, which are an important source of P during seed germination (58), were removed from the seedlings. Plants, 12-day-old, were transferred to low-Pi (LP) or SP nutrient media for 8 days. In agreement with the above results, B73 was the only genotype that showed a significant increase in root fresh weight under LP conditions compared to SP conditions (Figures S25 and S26, Supplementary Material) highlighting its high growth potential already at young developmental age.

Transcriptional network underlying biomass

To gain insight into the transcriptional network underlying green biomass formation in the field, we aimed to identify enriched transcription factor (TF) motifs associated with upregulated genes under Pi deficiency. To this end, we examined 2 kb promoter regions of upregulated genes using a combination of unsupervised Multiple Em for Motif Elicitation (MEME), supervised Motif Comparison Tool (TOMTOM), and Analysis of Motif Enrichment (AME) (59, 60). In the source leaf, four significantly enriched TF motifs were unique to the B73 genotype grown on NK (Fig. 6A; Table S17, Supplementary Material). These included the nitrate-responsive *cis*-element (NRE) bound by the family of Nin-like proteins (NLP), which are important regulators of nitrogen signaling, transport, and assimilation (61), an AT-rich Interaction Domain (ARID), a MYB and an AT-hook-type motif. In addition to the enrichment

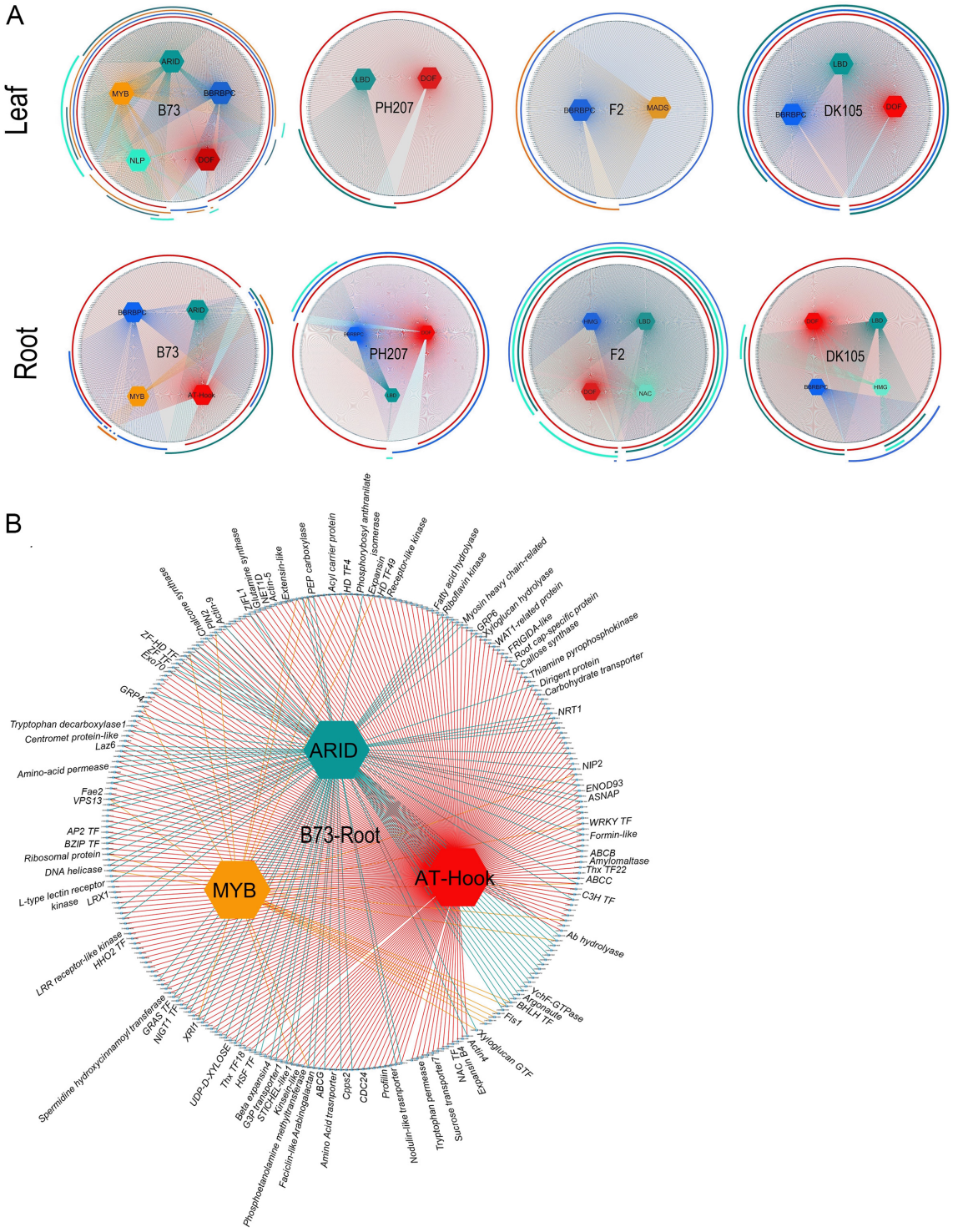

Fig. 6. TF motifs significantly enriched in upregulated genes under phosphate deficiency. (A) Regulatory networks for upregulated genes in the source leaf and root. Hexagons (nodes) represent TFs. Colored lines (edges) represent the interaction of TF motifs present on promoters with TFs. Light blue rectangles (nodes) represent genes. Outer circles show genes grouped with shared motifs. (B) Regulatory network of upregulated genes in the B73 root containing MYB, ARID, and/or AT-hook TF motifs significantly enriched only in the B73 line.

of the NRE, the NLP6-TF was upregulated in the source leaf of B73 under Pi starvation. The enriched TF motifs were probably key to the unique response of the B73 line to Pi deficiency. In the root of B73 plants grown on NK, the NRE was not significantly enriched in upregulated genes, in contrast to the source leaf samples. However, analysis of the upregulated genes with unique enriched TF motifs (Fig. 6B; Table S17, Supplementary Material) showed that several important components of N-signaling such as NRT1 transporter, glutamine synthase, NIGT1.4/HRS1 and NIGT1.2/HHO2 (26, 62), as well as many genes that are required for root growth and development, including carbohydrate transporters, contain AT-hook, ARID, and/or MYB motifs in their

promoter sequence (Fig. 6B). The enriched MYB binding sites are targets for the TFs LHY, CCA1, and RVE in the source leaf and for Blind-like protein RAX in the root. While the first three TFs are components of the central oscillator in the circadian system, RAX regulator was shown to initiate lateral bud formation and thus shoot branching (63). Plants with an environmentally adapted clock period have been shown to contain more chlorophyll, fix more carbon and grow faster than plants with an environmentally deviated circadian period (64, 65). The role and contribution of the correct setting of the circadian clock and the regulation of root branching to biomass growth in B73 require further research. Taken together, our data suggest that, in addition to the allocation of C to the root, higher uptake and assimilation of N in the B73 line leads to greater growth of roots and root hairs under Pi deficiency, which increases the Pi absorption area at the root–soil interface and plant growth. Furthermore, the expression of NLP and NIGT1 proteins in B73 suggests that there is a regulatory framework for the balanced uptake and utilization of N and P in soils managed with NK, which may differ in leaf and roots (62). Thus, in addition to the P content in the root, the C and N allocation to this underground organ during vegetative growth are decisive factors for the nutrient-dependent growth potential and the final biomass of maize.

Concluding Remarks

The way we currently produce food destroys habitats worldwide and contributes to climate change, species extinction, and the occurrence of pandemics (66–68). Therefore, it is desirable to advance crop improvement with efficient breeding methods based on robust biological markers, also in the face of climate change and the accompanying negatively altered environmental conditions. Our data strongly suggest that plant analysis at the system level makes it possible to identify genes, metabolites, and elements as markers that explain biomass variations in the field and allow prediction of the performance of a plant genotype on a given cropping area. We have shown that while key metabolites are important factors in predicting biomass, genetically programmed uptake, and organ-specific distribution of essential (C, N, P, and S) and potentially toxic elements play an equally important role in determining plant biomass. In addition, we have pinpointed a number of genes that shape molecular processes through mathematical modeling and/or from identified gene networks. The expression levels of these genes, therefore, provide information on how efficiently a maize genotype performs under field conditions at the cellular and physiological level. They can be used as markers for predicting biomass in breeding programs that go hand in hand with context-appropriate soil fertility management.

Materials and Methods

All the materials and methods are detailed in the Supplementary Material Appendix: experimental setup; analysis of soil properties; analysis of elemental composition by ICP-MS; metabolite profiling using mass spectrometry; RNA-seq and differential expression analysis; PCA and heatmap hierarchical clustering; monotonic modeling of the relation between shoot dry matter and metabolites, sugars, lipids, and amino acids; biological network analyses; multivariate analysis; identification of enriched regulatory motifs; and statistical methods.

Acknowledgments

We thank Johan Six (ETH Zurich), Paul Schulze-Lefert (PSL, Max Planck Institute for Plant Breeding Research Cologne, Germany (MPI Cologne)), and Stijn Spaepen (MPI Cologne) for insightful discussions during the RECONSTRUCT project. We thank Janine Altmüller from the Cologne Center for Genomics (CCG) for the DNA sequencing and Sina Reichardt, Stephanie Krey, Izabela Fabiańska, Huanhuan Tai, and Sebastian Koschel (University of Cologne, Germany) for assistance with the field and/or laboratory work. We thank René Flisch for access to the field trials, Juliane Hirte and the staff of the Agroscope Research Station (Therwil and Reckenholz) for assistance with the harvesting work in the field. We thank Zoran Nikoloski (University of Potsdam, Germany) for his thorough review of the manuscript.

Supplementary Material

Supplementary material is available at *PNAS Nexus* online.

Funding

This study was supported by the German Federal Ministry of Education and Research (Plant Breeding Research for the Bioeconomy, BMBF) to M.B., U.S., P.S.L., B.U., and A.F. under the RECONSTRUCT project (Unravelling the contribution of soil biodiversity to maize growth and fitness through combined omics-based in silico modeling and reconstruction biology; 031B0200 (A–E).

Authors' Contributions

M.B., A.F., T.M., and U.S. wrote the RECONSTRUCT project. A.B., M.B., K.D., N.G., M.H., and A.O. designed the experiments with support of A.F., T.M., U.S., and the RECONSTRUCT project consortium. J.M. managed the field experiment. A.B., K.D., J.H., A.O., and L.P. collected soil and root material, and measured plant performance parameters. G.P., M.H., and A.H. performed the RNA-seq analysis and K.D. the soil sampling and physico-chemical measurements. A.O., A.K., and D.M. provided lipid profiles, J.H. provided central metabolites, sugars, and free amino acid profiles. N.G. determined the ionome profiles and conducted microscopic mycorrhizal colonization analyses. A.H. and B.U. established, A.H., B.U., and N.G. maintained the RECONSTRUCT Database. L.P. generated Figure S1 (Supplementary Material). K.D. prepared Figures S2–S4 (Supplementary Material). M.H. analyzed the data and created the rest of the figures. M.H. performed the immunoblotting and hydroponic experiment. M.H. and M.B. wrote the manuscript.

Data Availability

The study data are included in the article and/or the Supplementary Material Appendix. The raw RNA-seq data are available at the European Nucleotide Archive under the project name PRJEB45810 (ENA: PRJEB45810) (ENA, https://www.ebi.ac.uk/ena/data/view/PRJEB45810).

References

1. Bernardo R. 2008. Molecular markers and selection for complex traits in plants: learning from the last 20 years. Crop Sci. 48:1649–1664.

2. Xu Y, Crouch JH. 2008. Marker-assisted selection in plant breeding: from publications to practice. Crop Sci. 48. DOI: 10.2135/cropsci2007.04.0191.
3. Yang N, *et al.* 2014. Genome wide association studies using a new nonparametric model reveal the genetic architecture of 17 agronomic traits in an enlarged maize association panel. PLoS Genet. 10:e1004573.
4. Chen H, Hao Z, Zhao Y, Yang R. 2020. A fast-linear mixed model for genome-wide haplotype association analysis: application to agronomic traits in maize. BMC Genomic. 21:151.
5. Meyer RC, *et al.* 2007. The metabolic signature related to high plant growth rate in *Arabidopsis thaliana*. Proc Natl Acad Sci USA. 04:4759–4764.
6. Sulpice R, *et al.* 2009. Starch as a major integrator in the regulation of plant growth. Proc Natl Acad Sci USA. 106:10348–10353.
7. Cañas RA, *et al.* 2017. Exploiting the genetic diversity of maize using a combined metabolomic, enzyme activity profiling, and metabolic modeling approach to link leaf physiology to kernel yield. Plant Cell. 29:919–943.
8. Muraya MM, Schmutzer T, Ulpinnis C, Scholz U, Altmann T. 2015. Targeted sequencing reveals large-scale sequence polymorphism in maize candidate genes for biomass production and composition. PloS ONE. 10:e0132120.
9. Willmann M, *et al.* 2013. Mycorrhizal phosphate uptake pathway in maize: vital for growth and cob development on nutrient poor agricultural and greenhouse soils. Front Plant Sci. 4:533.
10. Mäder P, *et al.* 2002. Soil fertility and biodiversity in organic farming. Science. 296:1694–1697.
11. Wrb. 2014. World reference base for soil resources 2014, update 2015. International soil classification system for naming soils and creating legends for soil maps. World Soil Resources Report No. 106.
12. Obata T, Fernie AR. 2012. The use of metabolomics to dissect plant responses to abiotic stresses. Cell Mol Life Sci. 69:3225–3243.
13. Barnaby JY, *et al.* 2013. Drought responses of foliar metabolites in three maize hybrids differing in water stress tolerance. PLoS ONE. 8:e77145.
14. Rouached H, Secco D, Arpat B, Poirier Y. 2011. The transcription factor PHR1 plays a key role in the regulation of sulfate shoot-to-root flux upon phosphate starvation in *Arabidopsis*. BMC Plant Biol. 11:19.
15. Kochian LV, Hoekenga OA, Piñeros MA. 2004. How do crop plants tolerate acid soils? Mechanisms of aluminum tolerance and phosphorous efficiency. Ann Rev Plant Biol. 55:459–493.
16. Hinsinger P, Plassard C, Tang C, Jaillard B. 2003. Origins of root-mediated pH changes in the rhizosphere and their responses to environmental constraints: a review. Plant and Soil. 248:43–59.
17. Mora-Macías J, *et al.* 2017. Malate-dependent Fe accumulation is a critical checkpoint in the root developmental response to low phosphate. Proc Natl Acad Sci USA. 114:E3563–E3572.
18. Stitt M. 1998. Pyrophosphate as an energy donor in the cytosol of plant cells: an enigmatic alternative to ATP. Botan Acta. 111:167–175.
19. Carstensen A, *et al.* 2018. The impacts of phosphorus deficiency on the photosynthetic electron transport chain. Plant Physiol. 177:271.
20. Nakamura Y. 2017. Plant phospholipid diversity: emerging functions in metabolism and protein–lipid interactions. Trends Plant Sci. 22:1027–1040.
21. Zhang Q, *et al.* 2018. Knock-down of *Arabidopsis* PLC5 reduces primary root growth and secondary root formation while over-expression improves drought tolerance and causes stunted root hair growth. Plant Cell Physiol. 59:2004–2019.
22. Rawat N, Singla-Pareek SL, Pareek A. 2020. Membrane dynamics during individual and combined abiotic stresses in plants and tools to study the same. Physiol Plant. 171. DOI: 10.1111/ppl.13217.
23. Lu S, *et al.* 2019. Genome-wide analysis of phospholipase D gene family and profiling of phospholipids under abiotic stresses in *Brassica napus*. Plant Cell Physiol. 60:1556.
24. Pant BD, *et al.* 2015. The transcription factor PHR1 regulates lipid remodeling and triacylglycerol accumulation in *Arabidopsis thaliana* during phosphorus starvation. J Exp Bot. 66:1907–1918.
25. O'Neill KC, Lee YJ. 2020. Visualizing genotypic and developmental differences of free amino acids in maize roots with mass spectrometry imaging. Front Plant Sci. 11:639.
26. Lam HM, Coschigano KT, Oliveira IC, Melo-Oliveira R, Coruzzi GM. 1996. The molecular-genetics of nitrogen assimilation into amino acids in higher plants. Ann Rev Plant Physiol Plant Mol Biol. 47:569–593.
27. Hijaz F, Killiny N. 2019. Exogenous GABA is quickly metabolized to succinic acid and fed into the plant TCA cycle. Plant Signal Behav. 14:e1573096.
28. Stromberger J, Tsai C, Huber D. 1994. Interactions of potassium with nitrogen and their influence on growth and yield potential in maize. J Plant Nut 17:19–37.
29. Xu X, *et al.* 2020. Effects of potassium levels on plant growth, accumulation and distribution of carbon, and nitrate metabolism in apple dwarf rootstock seedlings. Front Plant Sci. 11:904.
30. Geladi P, Kowalski BR. 1986. Partial least-squares regression: a tutorial. Anal Chim Acta. 185:1–17.
31. Mevik BH, Wehrens R. 2007. The pls package: principal component and partial least squares regression in R. J Stat Softw. 18. DOI: 10.18637/jss.v018.i02.
32. Szklarczyk D, *et al.* 2017. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. 45:D362–D368.
33. Kazan K, Manners JM. 2009. Linking development to defense: auxin in plant-pathogen interactions. Trends Plant Sci. 14:373–382.
34. Raghavendra AS, Gonugunta VK, Christmann A, Grill E. 2010. ABA perception and signalling. Trends Plant Sci. 15:395–401.
35. Erb TJ, Zarzycki J. 2018. A short history of RubisCO: the rise and fall (?) of nature's predominant CO2 fixing enzyme. Curr Opin Biotechnol. 49:100–107.
36. Wang R, *et al.* 2004. Genomic analysis of the nitrate response using a nitrate reductase-null mutant of *Arabidopsis*. Plant Physiol. 136:2512–2522.
37. Muñoz A, Castellano MM. 2012. Regulation of translation initiation under abiotic stress conditions in plants: is it a conserved or not so conserved process among eukaryotes?. Comp Funct Genomics. 2012:406357.
38. Kami C, Lorrain S, Hornitschek P, Fankhauser C. 2010. Light-regulated plant growth and development. Curr Top Dev Biol. 91: 29–66.
39. Michaeli S, Galili G, Genschik P, Fernie AR, Avin-Wittenberg T. 2016. Autophagy in plants - what's new on the menu?. Trends Plant Sci. 21:134–144.
40. Li F, *et al.* 2015. Autophagic recycling plays a central role in maize nitrogen remobilization. Plant Cell. 27:1389–1408.
41. Evrin C, *et al.* 2009. A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. Proc Natl Acad Sci USA. 106:20240–20245.

42. Rizzini L, *et al.* 2011. Perception of UV-B by the *Arabidopsis* UVR8 protein. Science. 332:103–106.
43. Iwakawa H, Shinmyo A, Sekine M. 2006. *Arabidopsis* CDKA;1, a cdc2 homologue, controls proliferation of generative cells in male gametogenesis. Plant J. 45:819–831.
44. Eastmond PJ, *et al.* 2010. Phosphatidic acid phosphohydrolase 1 and 2 regulate phospholipid synthesis at the endoplasmic reticulum in *Arabidopsis*. Plant Cell. 22:2796–2811.
45. Yunus IS, Liu YC, Nakamura Y. 2016. The importance of SERINE DECARBOXYLASE1 (SDC1) and ethanolamine biosynthesis during embryogenesis of *Arabidopsis thaliana*. Plant J. 88:559–569.
46. Gao C, *et al.* 2014. A Unique plant ESCRT component, FREE1, regulates multivesicular body protein sorting and plant growth. Curr Biol. 24:2556–2563.
47. Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 9:559.
48. Takemiya A, Inoue SI, Doi M, Kinoshita T, Shimazaki KI. 2005. Phototropins promote plant growth in response to blue light in low light environments. Plant Cell. 17:1120–1127.
49. Schlüter U, Bräutigam A, Droz JM, Schwender J, Weber APM. 2019. The role of alanine and aspartate aminotransferases in C4 photosynthesis. Plant Biol. 21:64–76.
50. Tiong J, *et al.* 2021. Improving nitrogen use efficiency through overexpression of alanine aminotransferase in rice, wheat, and barley. Front Plant Sci. 12:29.
51. Van Hove J, De Jaeger G, De Winne N, Guisez Y, Van Damme EJM. 2015. The *Arabidopsis* lectin EULS3 is involved in stomatal closure. Plant Sci. 238:312–322.
52. Xu B, *et al.* 2013. A cotton BURP domain protein interacts with α-expansin and their co-expression promotes plant growth and fruit production. Mol Plant. 6:945–958.
53. Park J, Cui Y, Kang BH. 2015. AtPGL3 is an arabidopsis BURP domain protein that is localized to the cell wall and promotes cell enlargement. Front Plant Sci. 6:412.
54. Stewart-Morgan KR, Petryk N, Groth A. 2020. Chromatin replication and epigenetic cell memory. Nat Cell Biol. 22:361–371.
55. Xu M, Wang W, Chen S, Zhu B. 2012. A model for mitotic inheritance of histone lysine methylation. EMBO Rep. 13:60–67.
56. Alabert C, *et al.* 2015. Two distinct modes for propagation of histone PTMs across the cell cycle. Genes Dev. 29:585–590.
57. Stewart-Morgan KR, Reverón-Gómez N, Groth A. 2019. Transcription restart establishes chromatin accessibility after DNA replication. Mol Cell. 75:284–297.e6.
58. Bledsoe SW, *et al.* 2017. The role of Tre6P and SnRK1 in maize early kernel development and events leading to stress-induced kernel abortion. BMC Plant Biol. 17:74.
59. Reynoso MA, *et al.* 2019. Evolutionary flexibility in flooding response circuitry in angiosperms. Science. 365:1291–1295.
60. Bailey TL, *et al.* 2009. MEME Suite: tools for motif discovery and searching. Nucleic Acids Res. 37:W202–208.
61. Gaudinier A, *et al.* 2018. Transcriptional regulation of nitrogen-associated metabolism and growth. Nature. 563:259–264.
62. Maeda Y, *et al.* 2018. A NIGT1-centred transcriptional cascade regulates nitrate signalling and incorporates phosphorus starvation signals in *Arabidopsis*. Nat Commun. 9:1376.
63. Müller D, Schmitz G, Theres K. 2006. Blind homologous R2R3 Myb genes control the pattern of lateral meristem initiation in *Arabidopsis*. Plant cell. 18:586–597.
64. Alabadí D, *et al.* 2001. Reciprocal regulation between TOC1 and LHY/CCA1 within the *Arabidopsis* circadian clock. Science. 293:880–883.
65. Dodd AN, *et al.* 2005. Cell biology: plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage. Science. 309:630–633.
66. Poore J, Nemecek T. 2018. Reducing food's environmental impacts through producers and consumers. Science. 360:987–992.
67. Clark MA, *et al.* 2020. Global food system emissions could preclude achieving the 1.5° and 2°C climate change targets. Science. 370:705–708.
68. Wernecke B, *et al.* 2020. 'Preventing the next pandemic' – a 2020 UNEP frontiers series report on zoonotic diseases with reflections for South Africa. S Afr J Sci. 116. DOI: 10.17159/sajs.2020/8531.