

A systems genetics approach reveals environment-dependent associations between SNPs, protein coexpression, and drought-related traits in maize

Mélisande Blein-Nicolas,¹ Sandra Sylvia Negro,¹ Thierry Balliau,¹ Claude Welcker,² Llorenç Cabrera-Bosquet,² Stéphane Dimitri Nicolas,¹ Alain Charcosset,¹ and Michel Zivy¹

¹Université Paris-Saclay, INRAE, CNRS, AgroParisTech, GQE – Le Moulon, 91190, Gif-sur-Yvette, France; ²LEPSE, INRAE, Université Montpellier, SupAgro, F34060 Montpellier, France

The effect of drought on maize yield is of particular concern in the context of climate change and human population growth. However, the complexity of drought-response mechanisms makes the design of new drought-tolerant varieties a difficult task that would greatly benefit from a better understanding of the genotype–phenotype relationship. To provide novel insight into this relationship, we applied a systems genetics approach integrating high-throughput phenotypic, proteomic, and genomic data acquired from 254 maize hybrids grown under two watering conditions. Using association genetics and protein coexpression analysis, we detected more than 22,000 pQTLs across the two conditions and confidently identified 15 loci with potential pleiotropic effects on the proteome. We showed that even mild water deficit induced a profound remodeling of the proteome, which affected the structure of the protein coexpression network, and a reprogramming of the genetic control of the abundance of many proteins, including those involved in stress response. Colocalizations between pQTLs and QTLs for ecophysiological traits, found mostly in the water deficit condition, indicated that this reprogramming may also affect the phenotypic level. Finally, we identified several candidate genes that are potentially responsible for both the coexpression of stress response proteins and the variations of ecophysiological traits under water deficit. Taken together, our findings provide novel insights into the molecular mechanisms of drought tolerance and suggest some pathways for further research and breeding.

[Supplemental material is available for this article.]

Maize is the world's leading crop (Shiferaw et al. 2011) in terms of production. Having a C4 metabolism, it shows high water use efficiency (WUE). However, it is also highly susceptible to water deficit. For example, maize is more affected by drought than either its close relative sorghum (aboveground dry biomass reduced by 47%–51% and 37%–38%, respectively) (Zegada-Lizarazu et al. 2012; Schittenhelm and Schroetter 2014) or wheat, which is a C3 plant (yield reduction associated with a 40% water reduction of 40% and 20.6%, respectively) (Daryanto et al. 2016). Improving maize yield under drought has been an important aim of breeding programs for decades (Campos et al. 2004, 2006; Cooper et al. 2014). However, despite the overall genetic improvement of maize, increases in drought sensitivity have been reported in several regions (Lobell et al. 2014; Meng et al. 2016; Zipper et al. 2016). In addition, severe episodes of drought are projected to become more frequent in the near future because of climate change (Harrison et al. 2014). Therefore, maize productivity under water deficit is of particular concern, and large efforts are still required to design varieties that are able to maintain high yields in drought conditions.

One lever to accelerate crop improvement is to better understand the genetic and molecular bases of drought tolerance. This highly complex trait is associated with a series of mechanisms oc-

curing at different spatial and temporal scales to (1) stabilize the plant's water and carbon status; (2) control the side effects of water deficit including oxidative stress, mineral deficiencies, and reduced photosynthesis; and (3) maintain plant yield (Chaves et al. 2003). At the physiological level, short-term responses include stomata closure, adjustment of osmotic and hydraulic conductance, leaf growth inhibition, and root growth promotion (Tardieu et al. 2018). At the molecular level, complex signaling and regulatory events occur, involving several hormones, of which abscisic acid (ABA) is a key player, and a broad range of transcription factors (Golldack et al. 2014; Osakabe et al. 2014; Tripathi et al. 2014). Molecular responses also include the accumulation of metabolites involved in osmotic adjustment, membrane and protein protection, as well as scavenging of reactive oxygen species, and the expression of drought-responsive proteins such as dehydrins, late embryogenesis abundant (LEA), and heat shock proteins (HSP) (Valliyodan and Nguyen 2006; Seki et al. 2007). All these responses will depend on the drought scenario, the phenological stage, the genetic makeup, and the general environmental conditions (Tardieu et al. 2018). Taken together, the multiplicity and versatility of the mechanisms involved explain the difficulty in selecting for drought tolerance.

© 2020 Blein-Nicolas et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Corresponding authors: michel.zivy@inra.fr, melisande.blein-nicolas@inra.fr

Article published online before print. Article, supplemental material, and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.255224.119>.

A better understanding of the genotype–phenotype relationship will help guide the development of new drought-tolerant varieties. Systems genetics is a recent approach providing improved insight into this relationship by deciphering the biological networks and molecular pathways underlying complex traits and by investigating how these traits are regulated at the genetic and epigenetic levels (Nadeau and Dudley 2011; Civelek and Lusis 2014; Feltus 2014; van der Sijde et al. 2014; Markowitz and Boutros 2015). This approach compares the position of quantitative trait loci (QTLs) underlying phenotypic traits variation to that of QTLs underlying the variation of upstream molecular traits such as transcript expression (eQTLs) or protein abundance (pQTLs). Until recently, this approach had been mostly applied in human and mice (Moreno-Moral and Petretto 2016). In plants, systems genetics studies have been carried out in a few species including wheat (Munkvold et al. 2013), rapeseed (Basnet et al. 2016), eucalyptus (Mizrachi et al. 2017), and maize (Christie et al. 2017; Jiang et al. 2019).

The first studies comparing QTLs and pQTLs used 2D gel proteomics to quantify proteins (de Vienne et al. 1999; Bourgeois et al. 2011). Since then, proteome coverage and data reliability have been widely improved by the use of mass spectrometry (MS)-based proteomics (Wasinger et al. 2013). Despite these technical advancements, the systems genetics studies published so far have preferentially used transcripts rather than proteins as the intermediate level between the genome and phenotypic traits. One reason is that large-scale proteomics experiments remain challenging (Blein-Nicolas et al. 2015) owing to technical constraints (Balliau et al. 2018) and the trade-off between depth of coverage and sample throughput (Keshishian et al. 2017). However, proteins are particularly relevant molecular components for linking genotype to phenotype. Indeed, protein abundance is expected to be more highly related to phenotype than transcript expression because of the buffering of transcriptional variations and the role of post-translational regulations in phenotype construction (Foss et al. 2011; Vogel and Marcotte 2012; Albertin et al. 2013; Battle et al. 2015; Chick et al. 2016).

Here, we aimed to better understand the molecular mechanisms associated with the genetic polymorphisms underlying the variations of ecophysiological traits related to drought tolerance. To this end, we performed a novel systems genetics study in which MS-based proteomics data acquired from 254 maize genotypes grown in two watering conditions were integrated with high-throughput genomic and phenotypic data. First, protein abundance was analyzed using a genome-wide association study (GWAS) and coexpression networks. Then, these data were integrated with ecophysiological phenotypic data from the same experiment (Prado et al. 2018) using a correlation analysis and by searching for QTL/pQTL colocalizations.

Results

Mild water deficit extensively remodels the proteome

Using MS-based proteomics, we analyzed more than 1000 leaf samples taken from 254 genotypes representing the genetic diversity within dent maize and grown in well-watered (WW) and water deficit (WD) conditions. After data filtering, the peptide intensity data set included 977 samples corresponding to 251 genotypes from which we quantified 2055 proteins described in Supplemental Table S1. Among these, 973 were quantified by integration of peptide intensities (XIC-based quantification). The re-

maining 1082, whose peptides had >10% of missing intensity values, were quantified by spectral counting (SC-based quantification). The latter proteins were less abundant (Supplemental Fig. S1A) and less precisely quantified (Supplemental Fig. S1B) than those that could be quantified by XIC.

Heatmap representations of protein abundance showed that two large separate protein clusters were associated with the two watering conditions (Fig. 1A). This indicates that, although moderate, water deficit had extensively remodeled the proteome of most genotypes. Accordingly, 82.4% and 74.5% of proteins from the XIC-based and SC-based sets, respectively, responded significantly to water deficit (Supplemental Table S1; Supplemental Fig. S2A). These included several proteins known to be involved in responses to drought or stress (Shinozaki and Yamaguchi-Shinozaki 2007; Wang et al. 2016) such as the dehydrins DHN1 (also known as RAB17, GRMZM2G079440) and DHN3 (GRMZM2G373522), an ABA-responsive protein (GRMZM2G106622), the LEA protein MAGI2594 (GRMZM2G352415), the HSPs HSP101 (GRMZM2G360681), HSP8 (GRMZM2G080724) and PZA03529 (GRMZM2G112165), the phospholipase D PLD2 (GRMZM2G061969), the glyoxalase I GLX1 (GRMZM2G181192), and the glutathione-S-transferase PCO124824 (GRMZM2G043291). Induced and repressed proteins constituted two highly differentiated populations in terms of function (Fig. 1B). In particular, transcription, translation, energy metabolism and metabolism of cofactors and vitamins were better represented within repressed proteins, whereas carbohydrate and amino acid metabolism and environmental adaptation were better represented within induced proteins.

The global impact of genotypic change on the proteome was less extensive than that of water deficit, because the proteomes of two different genotypes grown in the same watering condition were more similar than the proteomes of a same genotype grown in different conditions (Fig. 1A). However, the maximum amplitudes of abundance variations were similar (Supplemental Fig. S2B). In addition, 94.9% of proteins from the XIC-based set showed significant variation in abundance attributable to genetic variation (Supplemental Table S1). This was confirmed by broad sense heritability, the median of which was 0.47 and 0.46 for WW and WD conditions, respectively (Supplemental Fig. S3A). In contrast, in the SC-based set, only 34.4% of the proteins showed significant variation in abundance attributable to genetic variation with a median broad sense heritability of 0.08 and 0.10 for WW and WD conditions, respectively (Supplemental Fig. S3B).

Significant gene–environment ($G \times E$) interactions were detected for only four and 12 proteins from the SC-based and XIC-based set, respectively, probably because of a lack of statistical power. These proteins included the LEA protein MAGI 2594 (GRMZM2G352415), HSP18f (GRMZM2G083810), and a COR410 dehydrin (GRMZM2G147014). Although the $G \times E$ interaction was not statistically significant for the dehydrin DHN1 (GRMZM2G079440), this protein was undetectable in the WW condition and more or less expressed, depending on genotype, in the WD condition (Fig. 1C; Supplemental Table S2).

The strength of the genetic control of protein abundance is related to protein function

We performed GWAS for 2501 combinations of protein abundance \times watering condition showing a heritability >0.2. In total, we detected 514,270 significant associations for 2466 (98.6%) combinations of protein abundance \times watering condition involving 1367 proteins. When summarizing associated SNPs into

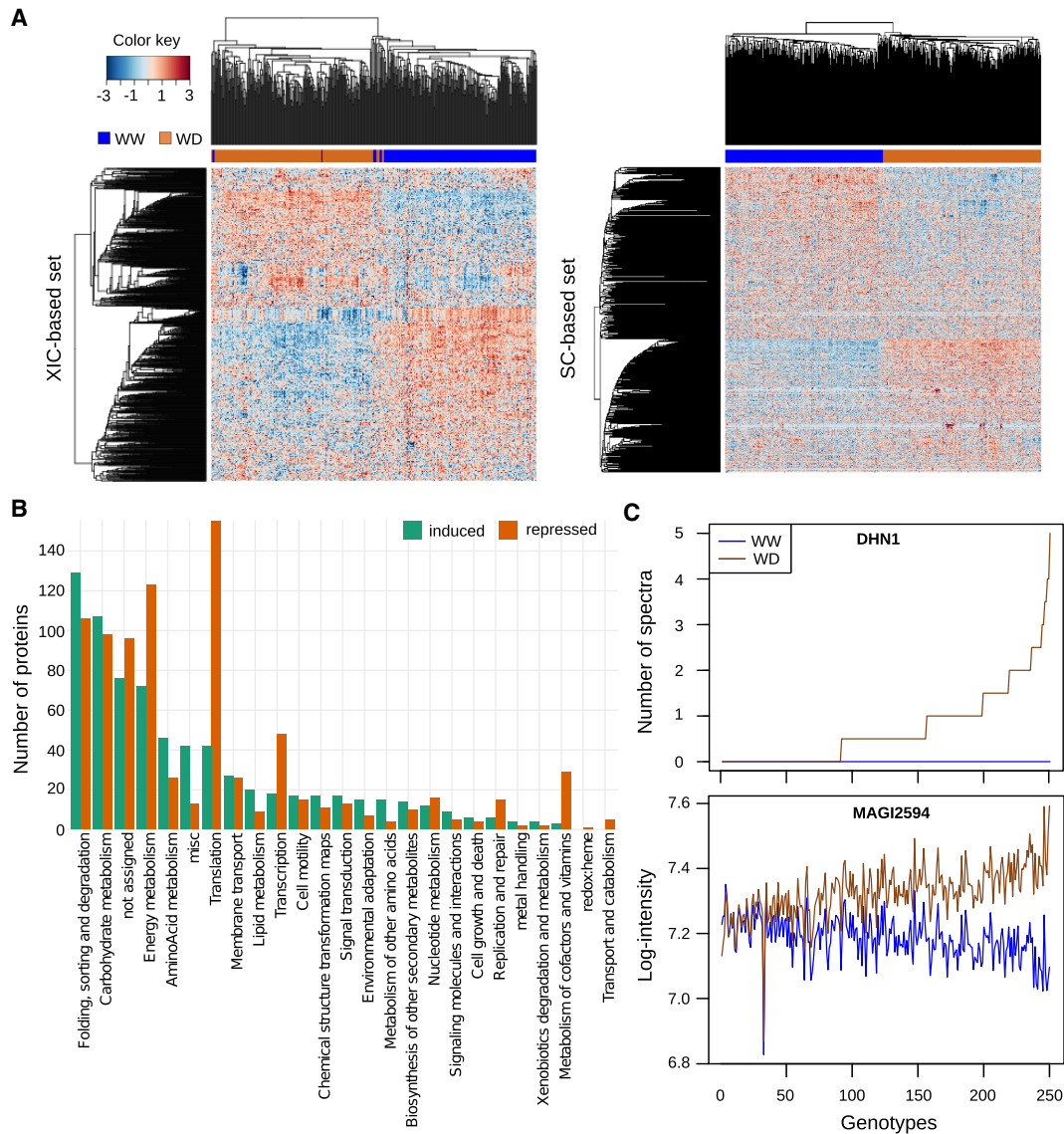


Figure 1. Effect of mild water deficit on the proteome. (A) Heatmap representations of protein abundances estimated for the XIC-based protein set (left) and the SC-based protein set (right). Each line corresponds to a protein and each column to a genotype \times watering condition combination. For each protein, abundance values were scaled and represented by a color code as indicated by the color key bar. Hierarchical clustering of the genotype \times watering condition combinations (top) and of the proteins (left) was built using the $1 - \text{Pearson correlation coefficient}$ as the distance and the unweighted pair group method with arithmetic mean (UPGMA) as the aggregation method. (B) Functions of the induced and repressed proteins under water deficit. (C) Abundance profiles of the DHN1 dehydrin (GRMZM2G079440 quantified based on the number of spectra) and of the MAGI2594 protein (GRMZM2G352415, a LEA protein quantified based on peptide intensity) in the two watering conditions. Genotypes on the x-axis were ordered according to the WD/WW abundance ratio. The lists of genotypes in this order are available in Supplemental Table S2.

pQTLs using classical methods based on genetic distance or linkage disequilibrium (LD), we observed a positive relationship between the number of pQTLs per chromosome and the P -value of the most strongly associated pQTL from the corresponding chromosome (Supplemental Fig. S4A,B). To get rid of this artificial relationship, which could lead to the detection of more than 250 pQTLs on one chromosome, we developed a geometric method based on the P -value signal of SNPs (Supplemental Fig. S4C). This method produced the lowest number of pQTLs per combination of protein abundance \times watering condition (median = 8 vs. 13 for the two other methods) and the lowest maximum number of pQTLs per chromosome (18 vs. 272 and 209 for the methods based

on genetic distance and LD, respectively). Using this geometric method and considering only pQTLs accounting for $>3\%$ of the total variance, we thus detected 22,664 pQTLs accounting for 3%–77.1% of the variance (Supplemental Table S3). Of these, 1113 were local, that is, located $<10^6$ bp from the protein encoding gene, of which 339 were located within the genes. Among distant pQTLs, 80.9% were located on a different chromosome from that of the protein encoding gene. Local pQTLs had stronger effects than distant pQTLs (average $R^2 = 15.3\%$ and 5.2% , respectively) (Supplemental Fig. S5). For 485 proteins, no local pQTL was detected in either condition. This set of proteins was significantly enriched in proteins involved in translation (15.3% vs. 3.8%,

adjusted P -value = 2.3×10^{-10}) and energy metabolism (17.3% vs. 8.5%, adjusted P -value = 8.2×10^{-5}) and depleted in proteins involved in carbohydrate metabolism (9.9% vs. 19.5%, adjusted P -value = 8.2×10^{-5}) compared to the 662 proteins showing a local pQTL in at least one condition. They also showed fewer distant pQTLs and were much less heritable (Supplemental Fig. S6A,B). These results indicate that the strength of the genetic control over protein abundance depends on protein function. This observation is supported by the positive correlation between the mean number of pQTLs and the mean heritability per functional category (Fig. 2).

Identification of loci with potential pleiotropic effects on the proteome

pQTLs were not uniformly distributed in the genome (Fig. 3A,B). Instead, there were genomic regions enriched with pQTLs. We detected 26 and 31 such hotspots that contained at least 19 pQTLs in the WW and WD conditions, respectively (Supplemental Table S4). These hotspots may represent loci with pleiotropic effects on the proteome, that is, loci associated with the abundance variation of several proteins. To refine the detection of such loci, we used a second independent approach based on the search for coexpression QTLs (coQTLs), that is, QTLs associated to the abundance variations of several coexpressed proteins. To do so, we first performed a weighted gene coexpression network analysis (WGCNA) of protein coexpression across the 251 genotypes in the two watering conditions separately (Supplemental Table S5). The two resulting networks differed in the presence of condition-specific modules indicating that water deficit has altered the structure of the protein coexpression network (Fig. 4; Supplemental Fig. S7A–C; Supplemental File S1). For each coexpression module, we then submitted the representative variable, called eigengene according to the WGCNA terminology, to GWAS to identify coQTLs. In total, we detected 176 coQTLs (96 for the 8 WW modules and 80 for the 8 WD modules) (Supplemental Table S3). Fifteen of them colocalized with pQTL hotspots (Supplemental Table S4). Thus, by crossing these results, we confidently identified four loci in the WW condition and 11 loci in the WD condition as having potential pleiotropic effects on the proteome. Of note, the proteins that were associated with hotspots Hs22d and Hs21d and that were also in the modules having coQTLs colocalizing with these hotspots were mainly ribosomal proteins (Supplemental Table S6). This suggests that Hs22d and Hs21d may contain loci involved in ribosome biogenesis.

The genetic architecture of protein abundance depends on the environment

Of the 11,034 pQTLs detected in the WW condition, only 1124 (10.2%) had a colocalizing pQTL in the WD condition. These pQTLs were generally of strong effect (Supplemental Fig. S8A) and were enriched in local pQTLs (32.6% vs. 4.9% over the entire data set). Although most of the pQTLs that were common to the two conditions had similar effects in both conditions, 75 (6.7%)

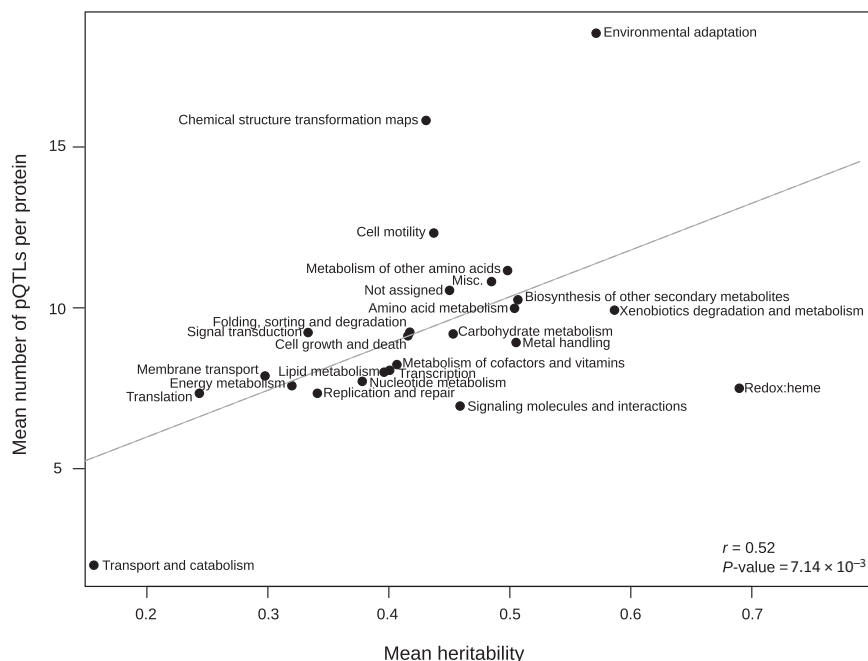


Figure 2. Relationship between the mean number of pQTLs per KEGG category and the mean heritability per KEGG category.

showed contrasted effects (Supplemental Fig. S8B). Half of these pQTLs were local, suggesting that gene promoters may be involved in the $G \times E$ interaction or that the pQTLs that were detected in each condition corresponded to different polymorphic sequences with different effects on protein abundance. These pQTLs were associated with 70 proteins, several of which were stress-responsive like the LEA protein PCO134925a (GRMZM2G045664) or HSPs GRMZM5G813217 and GRMZM2G536644 (Supplemental Table S7). Altogether, these results show that water deficit has altered the genetic architecture of protein abundance.

Identification of loci associated with trait variation at multiple scales

To gain insight into the molecular mechanisms associated with drought tolerance, we searched for colocalizations between the pQTLs, coQTLs, and hotspots detected in our study and the 160 QTLs identified by Prado et al. (2018) on the same plant material. These QTLs were associated with eight ecophysiological traits related to growth and transpiration rate: early leaf area (i.e., before water deficit; LAe), late leaf area (LAL), early biomass (Be), late biomass (Bl), water use (WU), water use efficiency (WUE), stomatal conductance (gs), and transpiration rate (Trate). Robust colocalizations were determined by taking into account the correlation between each trait and protein values.

In total, we identified 68 pairs of SNPs corresponding to QTL/pQTL colocalizations (Fig. 5; Supplemental Table S8). Only one involved a local pQTL. The QTL/pQTL distance was generally <100 kb, with, in 25.7% of cases, the same SNP representing the QTL and the pQTL (Fig. 6A). Most QTL/pQTL colocalizations (98%) were detected in the WD condition, where they corresponded to 39 of the 91 QTLs reported in this condition (Prado et al. 2018). They involved six ecophysiological phenotypic traits (Bl, LAL, WU, WUE, Trate, and gs) and 47 proteins, many of which were

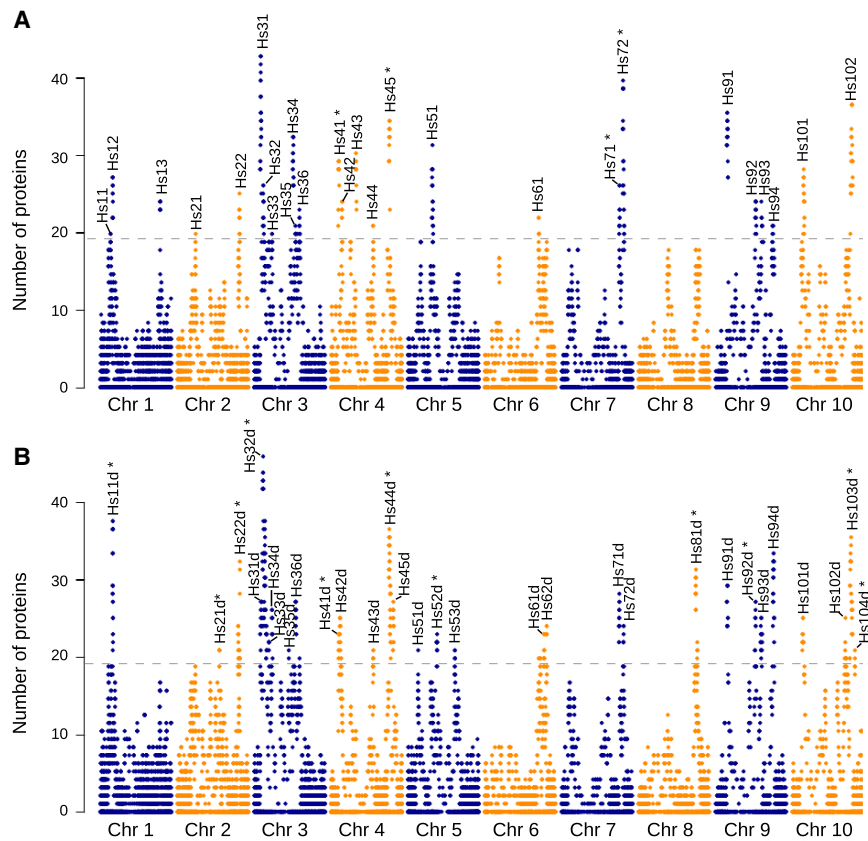


Figure 3. Distribution of pQTLs across the genome. (A) In the well-watered condition. (B) In the water deficit condition. Each point indicates the number of proteins associated with a pQTL located within a given genomic region defined by the linkage disequilibrium interval around an SNP. Dashed horizontal lines indicate the threshold used to detect pQTL hotspots. Names and positions of the pQTL hotspots are indicated above each graph. Asterisks indicate the pQTL hotspots confidently detected as loci with potential pleiotropic effects (for details, see Supplemental Table S4).

stress-responsive (Supplemental Table S9). Twenty-three proteins showed multiple QTL/pQTL colocalizations (Supplemental Table S9).

We further identified 11 pairs of SNPs corresponding to QTL/coQTL colocalizations, all in the WD condition (Supplemental Table S10). They involved three phenotypic traits (WU, Bl, LAI) and two coexpression modules including the WD-specific module (Fig. 5). These two modules were significantly enriched in stress-response proteins and in proteins involved in hormone metabolism and in reactive oxygen species detoxification (Supplemental Table S11). Ten of the 11 QTLs colocalizing with coQTLs also colocalized with pQTLs. The remaining QTL actually also colocalized with pQTLs, but with a low correlation between the phenotypic trait values and the protein abundance levels ($|r_{\text{corrected}}| < 0.23$) (Supplemental Table S10). In contrast, the correlation between trait values and eigengene was much higher ($|r_{\text{corrected}}| = 0.51$), which indicates that proteins were more strongly related to eco-physiological traits when taken collectively through a coexpression module rather than taken individually.

Taken together, these results highlight the presence of loci associated with traits at different biological scales. In the WD condition, several of these loci showed multiple associations both at the proteome and the phenotype level (Fig. 5). On Chromosome 1, a locus spanning 33 kbp contained a QTL for LAI determined by

SNP S1_5382845 as well as a coQTL for the green module and seven pQTLs, all determined by SNP AX-91427638. On Chromosome 5, a locus spanning 1.8 Mb between SNPs AX-91657926 and AX-91658235, contained three QTLs for LAI, Bl, and WU, one coQTLs for the WD-specific module, and six pQTLs. This region also contained hotspot Hs52d. On Chromosome 7, a single SNP (S7_162671160) determined the positions of two QTLs for LAI and WU, two coQTLs, and seven pQTLs. On Chromosome 10, a locus spanning 1.3 Mb between SNPs S10_122802154 and S10_124095144, contained one QTL for LAI, one coQTL, and eight pQTLs. This region also contained hotspot Hs103d. In the WD condition, leaf area (LAI) was repeatedly associated with the green module (on Chromosomes 1, 7, 9, 10) and to proteins belonging to this module. Several of them were detoxification enzymes (that is, a putative polyphenol oxidase, PPO1 GRMZM5G851266; two peroxidases, PRX39 GRMZM2G085967 and GRMZM2G108153; a superoxide dismutase GRMZM2G025992; a glyoxalase GRMZM2G704005).

Identification of candidate genes potentially involved in drought tolerance

Assuming that the genetic polymorphisms associated with protein abundance variations are within genes, we retrieved a list of 1–49 candidate genes

for each of the 69 pairs of SNPs corresponding to a QTL/pQTL or QTL/coQTL colocalization (Supplemental Table S12). Based on gene annotation and the literature, we identified two particularly interesting cases.

First, on Chromosome 7, the SNP S7_162671160 was located in *aas8* (also known as *gh3.8*, GRMZM2G053338), which was the only candidate gene. *aas8* is involved in indole-3-acetyl-amide conjugate biosynthesis. In agreement with the role of this gene in drought response (Feng et al. 2015), S7_162671160 was associated with the WD-specific module, WU and LAI, and five stress-response proteins: endochitinase CTA1 (GRMZM2G051943), beta-D-glucanase ENG1 (GRMZM2G073079), peroxidase PRX39 (GRMZM2G085967), polyphenol oxidase PPO1 (GRMZM5G851266), and phospholipase D PLD2 (GRMZM2G061969).

Second, 14 candidate genes were identified in the region of Chromosome 5 covered by hotspot Hs52d, of which two could be associated with the expression variation of a high number of genes. One is a squamosa promoter-binding gene (*sbp1*, GRMZM2G111136) that is inducible by various abiotic stresses including drought (Mao et al. 2016). The other, a C2C2-CO-like transcription factor (*col18*, GRMZM2G148772), was found to be significantly induced by drought and salinity stress in B73 leaves (Forestan et al. 2016). Hotspot Hs52d covered a region of ~4 Mb in which we detected 26 pQTLs (many of which were located

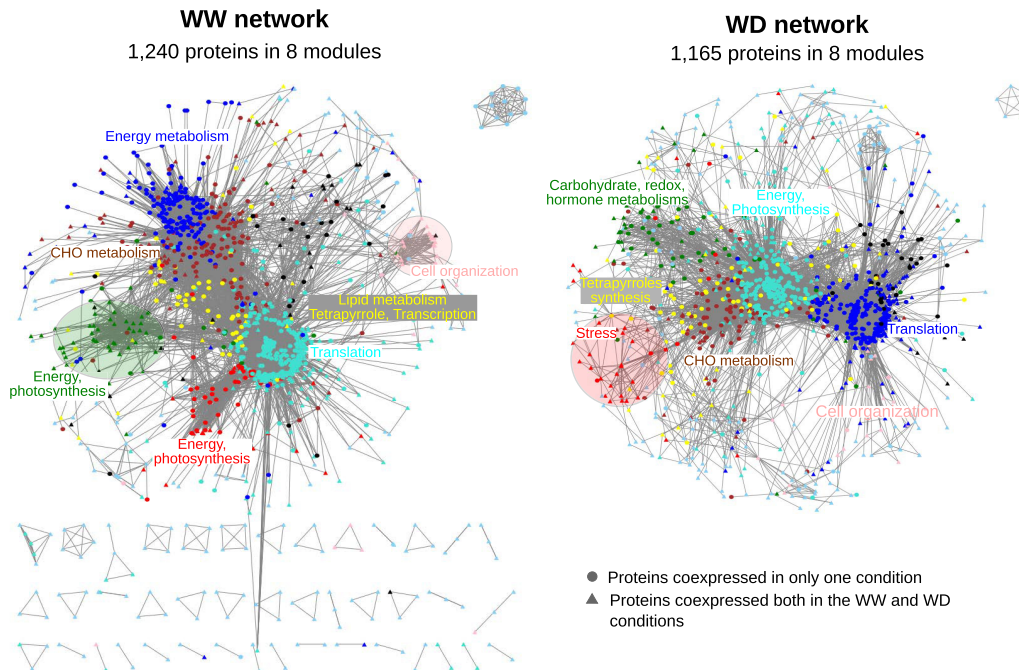


Figure 4. Graphical representation of the coexpression networks resulting from the WGCNA analysis. Only proteins with an adjacency >0.02 are shown. The two views were created by Cytoscape v3.5.1 using an unweighted, spring-embedded layout (cytoscape files are available in Supplemental File S1). The colors displayed on each network represent the different modules identified by WGCNA. Functional enrichments of modules are indicated with corresponding colors. Condition-specific modules are circled. Each module contains 35–471 proteins.

between *sbp1* and *col18*), two coQTLs, and four QTLs (Fig. 6B). A single SNP, AX-91658235, located only 1 kbp from *col18*, determined the position of two QTLs, two pQTLs, and one coQTL. Furthermore, SNP S5_88793314, located within the coding sequence of *sbp1*, determined the position of a QTL and a pQTL. Based on these results, we can hypothesize that hotspot Hs52d may correspond to two *trans*-acting factors for which *sbp1* and *col18* represent good candidates.

Discussion

To better understand the molecular mechanisms associated with the genetic polymorphisms underlying the variations in ecophysiological traits related to drought tolerance, we used a proteomics-based systems genetics approach that allowed us to map 22,664 pQTLs at high-resolution. By relating pQTLs to protein functions and heritability, we showed that the level of genetic control over protein abundances depends on protein function. For instance, proteins involved in translation and energy metabolism showed few pQTLs, with a lack of local pQTLs and low heritability. As these two functional categories mainly contain ancient and evolutionarily conserved proteins (Goldman et al. 2010; Nelson and Junge 2015), our results suggest that evolutionarily ancient proteins have more constrained expressions and fewer associated pQTLs (Popadin et al. 2014; Zhang and Yang 2015; Mähler et al. 2017). They also support the recent hypothesis of Mähler et al. (2017) that, for genes experiencing reduced rates of molecular evolution, purifying selection on individual SNPs is associated with stabilizing selection on gene expression.

pQTLs were found throughout the genome but some of them clustered into hotspots, suggesting the presence of loci with pleiotropic effects on the proteome. The detection of QTL hotspots is

highly dependent on the number of traits studied, the mapping resolution, and the method used to cluster QTLs. This may explain why previous studies have reported hotspots ranging from hundreds of eQTLs (Orozco et al. 2012; Munkvold et al. 2013; Christie et al. 2017), to only a few tens of eQTL or pQTLs (Foss et al. 2011; Ghazalpour et al. 2011; Albert et al. 2014), or even no hotspot at all (Mähler et al. 2017). In our study, false hotspot detection was limited by having a high mapping resolution and by using a pQTL clustering method that takes into account LD variations across the genome (Negro et al. 2019). Based on colocalization with coQTLs, we ultimately cross-validated 15 condition-specific hotspots, suggesting that loci with pleiotropic effects on the proteome can interact with the environment.

By analyzing a diversity panel of 254 genotypes, we showed that many small changes in protein abundance, detected as significant because they occurred in a high number of genotypes, contributed to extensively remodel the proteome in water deficit conditions. In total, ~75% of quantified proteins responded significantly to environmental change. Up- and down-regulated proteins were well differentiated in terms of function and indicated that the photosynthetic, transcriptional, and translational machineries were slowed down while stress responses and signalization mechanisms were activated. All these changes showed that plants clearly perceived a lack of water and presented a coordinated proteome response to water deficit.

Changes in abundance occurring in response to water deficit were associated with changes in the structure of the coexpression network. Indeed, we identified condition-specific modules, one of which, in the WD condition, was significantly enriched for stress-response proteins. Similarly, Munkvold et al. (2013) observed condition-specific modules related to biological processes in response to particular environmental conditions. Such modules suggest

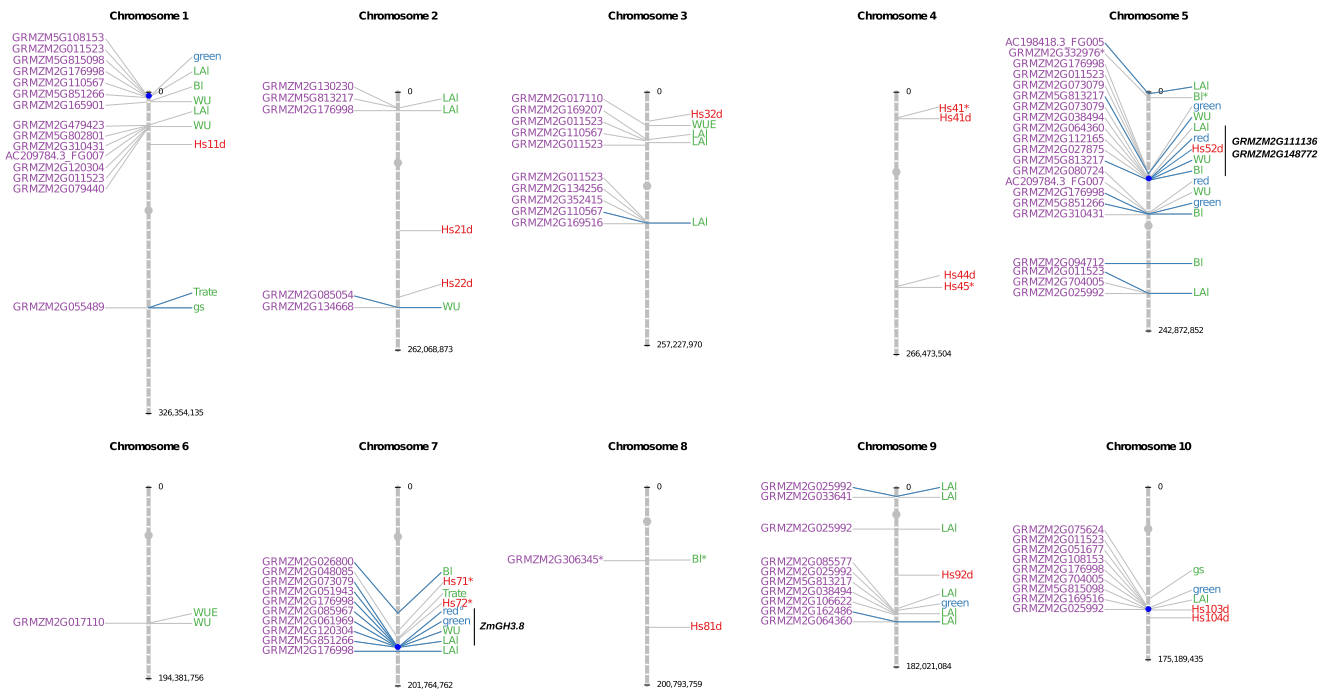


Figure 5. Genomic position of the colocalizing pQTLs, coQTLs, and QTLs. The positions of the 15 pQTL hotspots confidently identified as loci with potential pleiotropic effects are indicated, as well as the positions of the most promising candidate genes. Chromosomes are segmented into 10-Mb bins. Gray dots represent the centromeres and blue dots indicate the position of genomic regions showing evidences for pleiotropy at both the proteome and phenotype level. Blue lines indicate colocalizations with QTLs that are determined by a same SNP. (°) WD-specific module; (*) colocalization found in the WW condition.

that, under environmental perturbation, sets of genes or proteins are collectively mobilized by condition-specific factors allowing plant cells to adapt. The WD-specific module was associated with several QTL/coQTL colocalizations, and its eigengene was highly correlated with biomass, water use, and leaf area. Although the approach used here is correlative, these results suggest that, under water deficit, stress-response proteins contribute to phenotypic responses, which is consistent with the fact that many QTL/pQTL colocalizations involved these types of proteins. One coQTL for the WD-specific module was located in a region of Chromosome 5 that also cumulated several QTLs, pQTLs, and the hotspot Hs52d. This indicates that the coexpression observed for stress response proteins may be driven by condition-specific factors, the pleiotropic effects of which resonate across all layers of biological complexity up to phenotype. Altogether, these results suggest that an eigengene may be considered a more integrated molecular trait than protein abundance and can help decipher the genotype-phenotype relationship by bridging the gap between the proteomic and phenotypic level.

Linking phenotypic variation to proteome variation revealed many QTL/pQTL colocalizations for which, using high mapping resolution, we identified a limited number of candidate genes.

Only two of the 69 QTLs detected in the WW condition, versus 39 of the 91 in the WD condition, colocalized with pQTLs. This difference could be explained by the hypothesis that under non-stress conditions, phenotypic variations are driven by many low contribution proteins, whose abundance is probably controlled by low effect genetic polymorphisms, whereas under water stress, phenotypic variations are mainly driven by stress response proteins under the genetic control of condition-specific factors. In agreement with this hypothesis, we robustly identified two geno-

mic regions that could correspond to such factors. The first is located on Chromosome 7, where we identified *aas8* as the sole candidate gene underlying two QTLs (for leaf area and water use), seven pQTLs, of which five were associated with proteins involved in stress responses, and two coQTLs, one of which was associated with the WD-specific module. In maize shoots, Feng et al. (2015) showed that the expression of *aas8* was induced by auxin and reduced under polyethylene glycol treatment. The second region is located on Chromosome 5, in the region of the Hs52d hotspot, where we identified *sbp1* (GRMZM2G111136) and *col18* (GRMZM2G148772) as candidate genes underlying four QTLs, six pQTLs, and one coQTLs. These two transcription factors have been previously shown to be induced by drought in maize (Forestan et al. 2016; Mao et al. 2016). In addition, SBP genes constitute a functionally diverse family of transcription factors involved in plant growth and development (Preston and Hileman 2013). Because of their potential implication in $G \times E$ interactions and because of their roles both in plant growth and development and in drought response, *aas8*, *sbp1*, and *col18* represent promising candidates for drought tolerance breeding.

To conclude, our systems genetics approach, which incorporates MS-based proteomics data, has yielded several new results regarding the drought response in maize. First, we point out that the strength of the genetic control over protein abundance is related to protein function and also probably to the evolutionary constraints on protein expression. Then, we show that even mild water deficit strongly remodels the proteome and induces a reprogramming of the genetic control of the abundance of many proteins including those involved in stress responses. QTL/pQTL colocalizations are mostly found in the WD condition indicating that this reprogramming also affects the phenotypic level. Finally, we identify

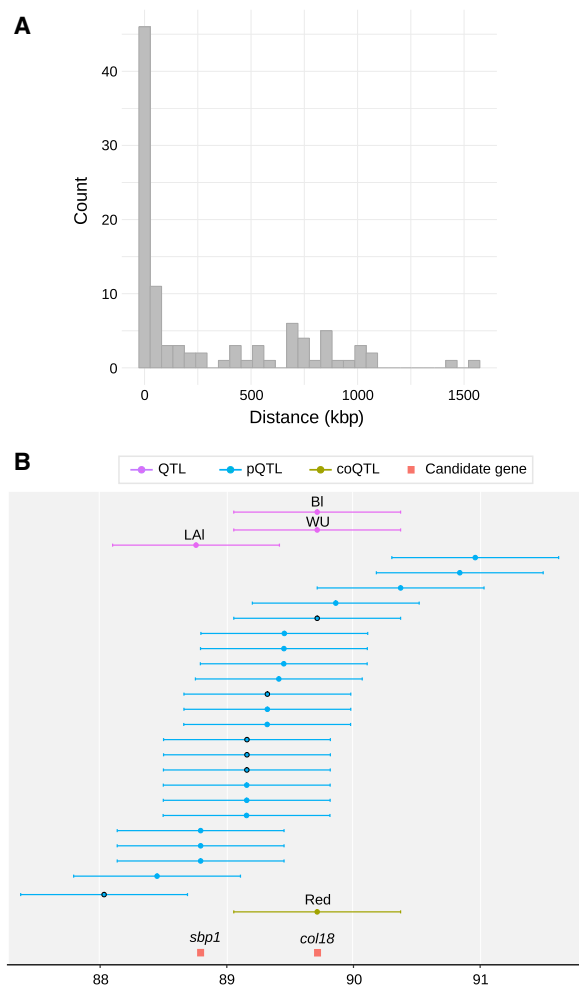


Figure 6. Identification of genomic regions associated with trait variations at multiple scales. (A) Distribution of the distances between colocalizing QTLs and pQTLs. (B) Detailed view of the QTL, pQTL, and coQTL detected in the region covered by the Hs52d hotspot on Chromosome 5. Dots represent the SNPs determining the position of the QTLs, and horizontal bars represent the linkage disequilibrium-based window around each SNP. Black circles indicate the pQTLs that colocalize with QTLs or coQTLs with a high correlation between protein abundance and the phenotypic trait value or the module eigengene. The position of two transcription factors, *sbp1* (GRMZM2G111136) and *col18* (GRMZM2G148772) that represent promising candidate genes, are indicated.

candidate genes that are potentially responsible for both the coexpression of stress-response proteins and the variation of ecophysiological traits under water deficit. Taken together, our findings provide novel insights into the molecular mechanisms of drought tolerance and suggest some pathways for further research and breeding. Our study also shows that proteomics has now reached enough maturity to be fully exploited in systems studies necessitating large-scale experiments.

Methods

Plant material and experiment

Plant material and growth conditions are described in full details in Prado et al. (2018) and in the Supplemental Methods. In brief, a diversity panel of maize hybrids was obtained by crossing a com-

mon flint parent (UH007, paternal parent) with 254 dent lines. Two levels of soil water content were applied: well-watered (soil water potential of -0.05 MPa) and water deficit (soil water potential of -0.45 MPa). Hybrids were replicated three times in each watering condition. Leaf sampling was performed at the preflowering stage in two replicates per hybrid and water condition.

Protein extraction and digestion

Protein extraction and digestion procedures are described in full detail in the Supplemental Methods. In brief, proteins were extracted from frozen ground leaf samples using a standard protocol for protein precipitation with trichloroacetic acid and acetone solution. Tryptic digestion was performed after solubilization, reduction, and alkylation of the proteins. The resulting peptides were desalted by solid phase extraction using polymeric C18 columns.

LC-MS/MS analyses

Samples were analyzed by LC-MS/MS in batches of 96. Analyses were performed using a NanoLC-Ultra System (nano2DUltra, Eksigent) connected to a Q-Exactive mass spectrometer (Thermo Electron). A 400-ng protein digest was loaded at $7.5 \mu\text{L}\cdot\text{min}^{-1}$ on a Biosphere C18 pre-column (0.3×5 mm, 100 \AA , $5 \mu\text{m}$; Nanoseparation) and desalted with 0.1% formic acid and 2% ACN. After 3 min, the pre-column was connected to a Biosphere C18 nanocolumn (0.075×150 mm, 100 \AA , $3 \mu\text{m}$, Nanoseparation). Buffers were 0.1% formic acid in water (A) and 0.1% formic acid and 100% ACN (B). Peptides were separated using a linear gradient from 5% to 35% buffer B for 40 min at $300 \text{ nL}\cdot\text{min}^{-1}$. One run took 60 min, including the regeneration step at 95% buffer B and the equilibration step at 95% buffer A. Ionization was performed with a 1.4-kV spray voltage applied to an uncoated capillary probe ($10 \mu\text{m}$ tip inner diameter; New Objective). Peptide ions were analyzed using Xcalibur 2.2 (Thermo Electron) in a data-dependent acquisition mode as described in the Supplemental Methods.

Peptide and protein identification

Peptide identification was performed using the MaizeSequence genome database (Release 5a, 136,770 entries, <https://ftp.maizgdb.org/MaizeGDB/FTP/>) supplemented with 1821 French maize inbred line F2 sequences with present/absent variants (PAVs) (Darracq et al. 2018) and a custom database containing standard contaminants. Database searches were performed using X! Tandem (Craig and Beavis 2004) (version 2015.04.01.1), and protein inference was performed using a homemade C++ version of X! TandemPipeline (Langella et al. 2017) specifically designed to handle hundreds of MS run files (Supplemental Code). Parameters for peptide identification and protein inference are described in the Supplemental Methods. The false discovery rate (FDR) was estimated at 0.06% for peptides and 0.04% for proteins.

Functional annotation of proteins was based on MapMan mapping (Thimm et al. 2004; Usadel et al. 2009) ([Zm_B73_5b_FGS_cds_2012](https://mapman.gabipd.org/), available at <https://mapman.gabipd.org/>) and on a custom KEGG classification built by manually attributing the MapMan bins to KEGG pathways (C Dillmann, pers. comm.).

Peptide and protein quantification

Peptide quantification was performed using MassChroQ version 2.1.0 (Valot et al. 2011) based on extracted ion chromatograms (XIC) with the parameters described in the Supplemental Methods. Peptide quantification data were filtered to remove genotypes represented by only one or two samples instead of the

expected four, as well as outlier samples for which we suspected technical problems during sample preparation or MS analysis. In the end, the MS data set included 977 samples.

Proteins were quantified from peptides using two complementary methods. First, for XIC-based quantification, proteins were quantified based on peptide intensity data filtered and normalized as described in the *Supplemental Methods*. R script for filtering and normalizing peptide intensity data is available as *Supplemental Code*. We excluded proteins that were quantified by only one peptide. As samples were analyzed by LC-MS/MS in batches over a period of several months, we observed a strong batch effect on normalized peptide intensities. To correct this batch effect, we fitted a linear model to log-transformed intensity data and subtracted the component attributable to batch effects. Then, for each protein, we modeled the peptide data using a mixed-effects model derived from Blein-Nicolas et al. (2012) and described in the *Supplemental Methods*. Protein abundance was subsequently computed as adjusted means from the model's estimates. Second, for spectral counting (SC)-based quantification, proteins that could not be quantified with XIC because their peptides had too many missing intensity values were quantified based on their number of assigned spectra. Proteins with a spectral count less than 2 in any of the samples were discarded. Normalization was then performed as described in the *Supplemental Methods*. As in XIC-based quantification, we corrected the batch effect by fitting a linear model to square-root-transformed and normalized protein abundances. Analysis of variance (ANOVA) was subsequently performed using the mixed-effects model described in the *Supplemental Methods*.

Genome-wide association study

GWAS was performed on protein abundances estimated in each watering condition using the single locus mixed model described in Yu et al. (2006). The variance-covariance matrix was determined as described in Rincent et al. (2014) by a kinship matrix derived from all SNPs except those on the chromosome containing the SNP being tested. SNP effects were estimated by generalized least squares and their significance was tested with an F-statistic. An SNP was considered significantly associated when $-\log_{10}(P\text{-value}) > 5$. A set of 961,971 SNPs obtained from line genotyping using a 50,000 Infinium HD Illumina array (Ganal et al. 2011), a 600,000 Axiom Affymetrix array (Unterseer et al. 2014), and a set of 500,000 SNPs obtained by genotyping-by-sequencing (Negro et al. 2019) were tested. Analyses were performed with FaST-LMM (Lippert et al. 2011) v2.07. Only SNPs with minor allele frequencies $>5\%$ were considered.

Inflation factors were computed as the slopes of the linear regressions on the QQplots between observed $-\log_{10}(P\text{-value})$ and expected $-\log_{10}(P\text{-value})$. Inflation factors were close to 1 (median of 1.08 and 1.06 in the XIC-based and SC-based sets, respectively), indicating low inflation of P -values.

Detection of QTLs from significantly associated SNPs

Three different methods implemented in R (R Core Team 2013) version 3.3.3 were used to summarize significantly associated SNPs into pQTLs. First, in the genetic method, two contiguous SNPs were considered to belong to a same QTL when the genetic distance separating them was <0.1 cM. Second, in the LD-based method, two contiguous SNPs were considered to belong to a same QTL when their LD-based windows (Negro et al. 2019) overlapped. Third, in the geometric method, for each chromosome, we ordered the SNPs according to their physical position. Then, we smoothed the $-\log_{10}(P\text{-value})$ signal by computing the maximum

of the $-\log_{10}(P\text{-value})$ in a sliding window containing N consecutive SNPs. An association peak was detected when the smoothed $-\log_{10}(P\text{-value})$ signal exceeded a max threshold M . Two consecutive peaks were considered to be two different QTLs when the $-\log_{10}(P\text{-value})$ signal separating them dropped below a minimum threshold m . The parameters for QTL detection were fixed empirically at $N=500$, $M=5$, and $m=4$. For the three methods described above, the position of a QTL was determined by the SNP showing the highest $-\log_{10}(P\text{-value})$. A pQTL was considered local if it was within 1 Mb upstream or downstream from the coding sequence of the gene encoding the corresponding protein.

Complementary data analyses

The following complementary data analyses were performed with R (R Core Team 2013) version 3.3.3:

1. Broad sense heritability of protein abundance: For each protein, the broad sense heritability of abundance was computed for each of the two watering conditions from a mixed-effects model as described in the *Supplemental Methods*.
2. Detection of pQTL hotspots: for each SNP position, we counted the number of pQTLs (N) located within its LD-based window (Negro et al. 2019). The threshold used to detect a hotspot was set at the 97% quantile of the distribution of N .
3. Protein coexpression analysis: Protein coexpression analysis was performed using the WGCNA R package (Langfelder and Horvath 2008) with the parameters described in the *Supplemental Methods*. Using a procedure developed to correct the bias owing to population structure and/or relatedness in the LD measure and implemented in the LDcorSV R package (Mangin et al. 2012), we computed pairwise Pearson's correlations corrected by structure and kinship ($|r_{\text{corrected}}|$) and used them as the input similarity matrix. Graphical representations of the resulting networks were performed with Cytoscape (Shannon et al. 2003) v3.5.1 using an unweighted spring-embedded layout.
4. QTL colocalization: We considered QTLs to colocalize when they meet the following two criteria. First, the LD-based windows around the QTLs (Negro et al. 2019) should overlap. Second, the absolute value of the Pearson's correlation of coefficient corrected by structure and kinship (the $|r_{\text{corrected}}|$ mentioned above) between the values of the ecophysiological traits associated with the QTLs should be greater than 0.3. We determined this value empirically in the absence of a statistical test to test the significance of the corrected correlation.
5. Candidate gene identification: For each QTL/pQTL colocalization, gene accessions found within the interval defined by the intersection between the LD-based windows around the QTL and the pQTL were retrieved from the MaizeSequence genome database (Release 5a). Low confidence gene models and transposable elements were not considered.

Data access

The raw MS output files were submitted to PROTiCdb (<http://moulon.inra.fr/protic/amaizing>) (Ferry-Dumazet et al. 2005; Langella et al. 2007, 2013), the MassIVE database (<https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp>) under accession number MSV000085594, and the ProteomeXchange database (<http://www.proteomexchange.org>) under accession number PXD019804. Detailed information on all peptides and proteins identified in the LC-MS/MS runs, as well as peptide intensities and protein abundances obtained for each sample, are also freely available on PROTiCdb at the same URL.

Phenotypic data are available online using the PHIS information system (Neveu et al. 2019) at <http://www.phis.inra.fr/openphis/web/index.php?r=site%2Flogin-as-guest>. From this webpage, data are accessible by clicking on Experimental organization > Project > Systems genetics for maize drought tolerance (Amaizing project). Early leaf area (LAE) was defined at the seven leaves stage, representing 24 d (thermal time in equivalent days at 20°C). Late leaf area (LAL) was defined at the 12 leaves stage, representing 45 d (thermal time in equivalent days at 20°C).

The genotyping data generated in this study were submitted to the European Variation Archive (EVA; <https://www.ebi.ac.uk/eva/>) under accession number PRJEB40124 and are also available at Data INRAE (<https://doi.org/10.15454/GAHEU0>).

Competing interest statement

The authors declare no competing interests.

Acknowledgments

This work was supported by the Agence Nationale de la Recherche project ANT-10-BTBR-01 (Amaizing). Proteomics analyses were performed on the Plateforme d'Analyse Protéomique de Paris Sud Ouest facility (<http://pappso.inra.fr>), which is supported by Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (<http://www.inrae.fr>), the Ile-de-France regional council (<https://www.iledefrance.fr/education-recherche>), Infrastructures en Biologie Santé et Agronomie (<https://www.ibisa.net>), and Saclay Plant Sciences-SPS (ANR-17-EUR-0007). We thank Sylvie Coursol for her critical review of the manuscript; Hélène Corti for her help in sample preparation; Olivier Langella for having specially developed a pipeline to upload the proteomics data on ProticDB; and Delphine Madur, Yannick De Oliveira, and Pierre Montalent for their help in creating the VCF files for genotyping data. We thank people from Laboratoire d'Ecophysiologie des Plantes sous Stress Environnementaux: François Tardieu for his contribution to the coordination of the plant experiment; Benoît Suard, Pauline Sidawi, and Olivier Martin for their technical assistance during the experiment; and Santiago Alvarez Prado for his contribution to plant traits and QTL analysis.

Author contributions: A.C. and M.Z. designed the research; C.W. designed and coordinated the plant experiment in PhenoArch and the genetic analysis of plant traits; L.C.B. performed the plant experiment and analyzed the image-based phenotypic data; M.B.-N. and T.B. performed the proteomics experiments; S.S.N. developed the GWAS pipeline and performed the genotyping quality control; S.D.N. performed the genotyping and estimated local LD; M.B.-N. and M.Z. analyzed the proteomics data; and M.B.-N. performed the systems genetics study and wrote the manuscript. All authors discussed the results and read and approved the final manuscript.

References

- Albert FW, Treusch S, Shockley AH, Bloom JS, Kruglyak L. 2014. Genetics of single-cell protein abundance variation in large yeast populations. *Nature* **506**: 494–497. doi:10.1038/nature12904
- Albertin W, Marullo P, Bely M, Aigle M, Bourgeois A, Langella O, Balliau T, Chevret D, Valot B, da Silva T, et al. 2013. Linking post-translational modifications and variation of phenotypic traits. *Mol Cell Proteomics* **12**: 720–735. doi:10.1074/mcp.M112.024349
- Balliau T, Blein-Nicolas M, Zivy M. 2018. Evaluation of optimized tube-gel methods of sample preparation for large-scale plant proteomics. *Proteomes* **6**: 6. doi:10.3390/proteomes6010006
- Basnet RK, Del Carpio DP, Xiao D, Bucher J, Jin M, Boyle K, Fobert P, Visser RGF, Maliepaard C, Bonnema G. 2016. A systems genetics approach identifies gene regulatory networks associated with fatty acid composition in *Brassica rapa* seed. *Plant Physiol* **170**: 568–585. doi:10.1104/pp.15.00853
- Battle A, Khan Z, Wang SH, Mitran A, Ford MJ, Pritchard JK, Gilad Y. 2015. Impact of regulatory variation from RNA to protein. *Science* **347**: 664–667. doi:10.1126/science.1260793
- Blein-Nicolas M, Xu H, de Vienne D, Giraud C, Huet S, Zivy M. 2012. Including shared peptides for estimating protein abundances: a significant improvement for quantitative proteomics. *Proteomics* **12**: 2797–2801. doi:10.1002/pmic.201100660
- Blein-Nicolas M, Albertin W, da Silva T, Valot B, Balliau T, Masneuf-Pomarède I, Bely M, Marullo P, Sicard D, Dillmann C, et al. 2015. A systems approach to elucidate heterosis of protein abundances in yeast. *Mol Cell Proteomics* **14**: 2056–2071. doi:10.1074/mcp.M115.048058
- Bourgeois M, Jacquin F, Casseculle F, Savoie V, Belghazi M, Aubert G, Quillien L, Huart M, Marget P, Bursin J. 2011. A PQL (protein quantity loci) analysis of mature pea seed proteins identifies loci determining seed protein composition. *Proteomics* **11**: 1581–1594. doi:10.1002/pmic.201000687
- Campos H, Cooper M, Habben JE, Edmeades GO, Schussler JR. 2004. Improving drought tolerance in maize: a view from industry. *Field Crops Res* **90**: 19–34. doi:10.1016/j.fcr.2004.07.003
- Campos H, Cooper M, Edmeades GO, Löffler CM, Schussler JR, Ibáñez M. 2006. Changes in drought tolerance in maize associated with fifty years of breeding for yield in the US corn belt. *Maydica* **51**: 369–381.
- Chaves MM, Maroco JP, Pereira JS. 2003. Understanding plant responses to drought—from genes to the whole plant. *Funct Plant Biol* **30**: 239–264. doi:10.1071/FP02076
- Chick JM, Munger SC, Simecek P, Huttlin EL, Choi K, Gatti DM, Raghupathy N, Svenson KL, Churchill GA, Gygi SP. 2016. Defining the consequences of genetic variation on a proteome-wide scale. *Nature* **534**: 500–505. doi:10.1038/nature18270
- Christie N, Myburg AA, Joubert F, Murray SL, Carstens M, Lin YC, Meyer J, Crampton BG, Christensen SA, Ntuli JF, et al. 2017. Systems genetics reveals a transcriptional network associated with susceptibility in the maize–grey leaf spot pathosystem. *Plant J* **89**: 746–763. doi:10.1111/tj.13419
- Civelek M, Lusi AJ. 2014. Systems genetics approaches to understand complex traits. *Nat Rev Genet* **15**: 34–48. doi:10.1038/nrg3575
- Cooper M, Gho C, Leafgren R, Tang T, Messina C. 2014. Breeding drought-tolerant maize hybrids for the US corn-belt: discovery to product. *J Exp Bot* **65**: 6191–6204. doi:10.1093/jxb/eru064
- Craig R, Beavis RC. 2004. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **20**: 1466–1467. doi:10.1093/bioinformatics/bth092
- Darracq A, Vitte C, Nicolas S, Duarte J, Pichon JP, Mary-Huard T, Chevalier C, Bérard A, Le Paslier MC, Rogowsky P, et al. 2018. Sequence analysis of European maize inbred line F2 provides new insights into molecular and chromosomal characteristics of presence/absence variants. *BMC Genomics* **19**: 119. doi:10.1186/s12864-018-4490-7
- Daryanto S, Wang L, Jacinthe PA. 2016. Global synthesis of drought effects on maize and wheat production. *PLoS One* **11**: e0156362. doi:10.1371/journal.pone.0156362
- de Vienne D, Leonardi A, Damerval C, Zivy M. 1999. Genetics of proteome variation for QTL characterization: application to drought-stress responses in maize. *J Exp Bot* **50**: 303–309. doi:10.1093/jxb/50.3.32.303
- Feltus FA. 2014. Systems genetics: a paradigm to improve discovery of candidate genes and mechanisms underlying complex traits. *Plant Sci* **223**: 45–48. doi:10.1016/j.plantsci.2014.03.003
- Feng S, Yue R, Tao S, Yang Y, Zhang L, Xu M, Wang H, Shen C. 2015. Genome-wide identification, expression analysis of auxin-responsive GH3 family genes in maize (*Zea mays* L.) under abiotic stresses. *J Integr Plant Biol* **57**: 783–795. doi:10.1111/jipb.12327
- Ferry-Dumazet H, Houel G, Montalent P, Moreau L, Langella O, Negroni L, Vincent D, Lalanne C, de Daruvar A, Plomion C, et al. 2005. PROTICDB: a web-based application to store, track, query, and compare plant proteome data. *Proteomics* **5**: 2069–2081. doi:10.1002/pmic.200401111
- Forestan C, Aiese Cigliano R, Farinati S, Lunardon A, Sanseverino W, Varotto S. 2016. Stress-induced and epigenetic-mediated maize transcriptome regulation study by means of transcriptome reannotation and differential expression analysis. *Sci Rep* **6**: 30446. doi:10.1038/srep30446
- Foss EJ, Radulovic D, Shaffer SA, Goodlett DR, Kruglyak L, Bedalov A. 2011. Genetic variation shapes protein networks mainly through non-transcriptional mechanisms. *PLoS Biol* **9**: e1001144. doi:10.1371/journal.pbio.1001144
- Ganal MW, Durstewitz G, Polley A, Bérard A, Buckler ES, Charcosset A, Clarke JD, Graner EM, Hansen M, Joets J, et al. 2011. A large maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. *PLoS One* **6**: e28334. doi:10.1371/journal.pone.0028334

- Ghazalpour A, Bennett B, Petyuk VA, Orozco L, Hagopian R, Mungrue IN, Farber CR, Sinshheimer J, Kang HM, Furlotte N, et al. 2011. Comparative analysis of proteome and transcriptome variation in mouse. *PLoS Genet* **7**: e1001393. doi:10.1371/journal.pgen.1001393
- Goldman AD, Samudrala R, Baross JA. 2010. The evolution and functional repertoire of translation proteins following the origin of life. *Biol Direct* **5**: 15. doi:10.1186/1745-6150-5-15
- Golladack D, Li C, Mohan H, Probst N. 2014. Tolerance to drought and salt stress in plants: unraveling the signaling networks. *Front Plant Sci* **5**: 151. doi:10.3389/fpls.2014.00151
- Harrison MT, Tardieu F, Dong Z, Messina CD, Hammer GL. 2014. Characterizing drought stress and trait influence on maize yield under current and future conditions. *Glob Change Biol* **20**: 867–878. doi:10.1111/gcb.12381
- Jiang LG, Li B, Liu SX, Wang HW, Li CP, Song SH, Beatty M, Zastrow-Hayes G, Yang XH, Qin F, et al. 2019. Characterization of proteome variation during modern maize breeding. *Mol Cell Proteomics* **18**: 263–276. doi:10.1074/mcp.RA118.001021
- Keshishian H, Burgess MW, Specht H, Wallace L, Clauser KR, Gillette MA, Carr SA. 2017. Quantitative, multiplexed workflow for deep analysis of human blood plasma and biomarker discovery by mass spectrometry. *Nat Protoc* **12**: 1683–1701. doi:10.1038/nprot.2017.054
- Langella O, Zivy M, Joets J. 2007. The PROTiCdb database for 2-DE proteomics. *Methods Mol Biol* **355**: 279–303. doi:10.1385/1-59745-227-0:279
- Langella O, Valot B, Jacob D, Balliau T, Flores R, Hoogland C, Joets J, Zivy M. 2013. Management and dissemination of MS proteomic data with PROTiCdb: example of a quantitative comparison between methods of protein extraction. *Proteomics* **13**: 1457–1466. doi:10.1002/pmic.201200564
- Langella O, Valot B, Balliau T, Blein-Nicolas M, Bonhomme L, Zivy M. 2017. X!TandemPipeline: a tool to manage sequence redundancy for protein inference and phosphosite identification. *J Proteome Res* **16**: 494–503. doi:10.1021/acs.jproteome.6b00632
- Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**: 559. doi:10.1186/1471-2105-9-559
- Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D. 2011. FaST linear mixed models for genome-wide association studies. *Nat Methods* **8**: 833–835. doi:10.1038/nmeth.1681
- Lobell DB, Roberts MJ, Schlenker W, Braun N, Little BB, Rejesus RM, Hammer GL. 2014. Greater sensitivity to drought accompanies maize yield increase in the U.S. midwest. *Science* **344**: 516–519. doi:10.1126/science.1251423
- Mähler N, Wang J, Teriebienic BK, Ingvarsson PK, Street NR, Hvidsten TR. 2017. Gene co-expression network connectivity is an important determinant of selective constraint. *PLoS Genet* **13**: e1006402. doi:10.1371/journal.pgen.1006402
- Mangin B, Siberchicot A, Nicolas S, Doligez A, This P, Cierco-Ayrolles C. 2012. Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. *Heredity (Edinb)* **108**: 285–291. doi:10.1038/hdy.2011.73
- Mao HD, Yu LJ, Li ZJ, Yan Y, Han R, Liu H, Ma M. 2016. Genome-wide analysis of the SPL family transcription factors and their responses to abiotic stresses in maize. *Plant Gene* **6**: 1–12. doi:10.1016/j.plgene.2016.03.003
- Markowitz F, Boutros M. 2015. An introduction to systems genetics. In *Systems genetics* (ed. Markowitz F, Boutros M), pp. 1–11. Cambridge University Press, Cambridge, UK. doi:10.1017/CBO9781139012751.001
- Meng Q, Chen X, Lobell DB, Cui Z, Zhang Y, Yang H, Zhang F. 2016. Growing sensitivity of maize to water scarcity under climate change. *Sci Rep* **6**: 19605. doi:10.1038/srep19605
- Mizrachi E, Verbeke L, Christie N, Fierro AC, Mansfield SD, Davis MF, Gjersing E, Tuskan GA, Montagu MV, de Peer YV, et al. 2017. Network-based integration of systems genetics data reveals pathways associated with lignocellulosic biomass accumulation and processing. *Proc Natl Acad Sci* **114**: 1195–1200. doi:10.1073/pnas.1620119114
- Moreno-Moral A, Petretto E. 2016. From integrative genomics to systems genetics in the rat to link genotypes to phenotypes. *Dis Model Mech* **9**: 1097–1110. doi:10.1242/dmm.026104
- Munkvold JD, Laudencia-Chinguanco D, Sorrells ME. 2013. Systems genetics of environmental response in the mature wheat embryo. *Genetics* **194**: 265–277. doi:10.1534/genetics.113.150052
- Nadeau JH, Dudley AM. 2011. Systems genetics. *Science* **331**: 1015–1016. doi:10.1126/science.1203869
- Negro SS, Millet EJ, Madur D, Bauland C, Combes V, Welcker C, Tardieu F, Charcosset A, Nicolas SD. 2019. Genotyping-by-sequencing and SNP arrays are complementary for detecting quantitative trait loci by tagging different haplotypes in association studies. *BMC Plant Biol* **19**: 318. doi:10.1186/s12870-019-1926-4
- Nelson N, Junge W. 2015. Structure and energy transfer in photosystems of oxygenic photosynthesis. *Annu Rev Biochem* **84**: 659–683. doi:10.1146/annurev-biochem-092914-041942
- Neveu P, Tireau A, Hilgert N, Nègre V, Mineau-Cesari J, Bricchet N, Chapuis R, Sanchez I, Pommier C, Charnomordic B, et al. 2019. Dealing with multi-source and multi-scale information in plant phenomics: the ontology-driven phenotyping hybrid information system. *New Phytol* **221**: 588–601. doi:10.1111/nph.15385
- Orozco LD, Bennett BJ, Farber CR, Ghazalpour A, Pan C, Che N, Wen P, Qi HX, Mutukulu A, Siemers N, et al. 2012. Unraveling inflammatory responses using systems genetics and gene-environment interactions in macrophages. *Cell* **151**: 658–670. doi:10.1016/j.cell.2012.08.043
- Osakabe Y, Osakabe K, Shinozaki K, Tran LSP. 2014. Response of plants to water stress. *Front Plant Sci* **5**: 86. doi:10.3389/fpls.2014.00086
- Popadin KY, Gutierrez-Arcelus M, Lappalainen T, Buil A, Steinberg J, Nikolaev SI, Lukowski SW, Bazykin GA, Seplyarskiy VB, Ioannidis P, et al. 2014. Gene age predicts the strength of purifying selection acting on gene expression variation in humans. *Am J Hum Genet* **95**: 660–674. doi:10.1016/j.ajhg.2014.11.003
- Prado SA, Cabrera-Bosquet L, Grau A, Coupel-Ledru A, Millet EJ, Welcker C, Tardieu F. 2018. Phenomics allows identification of genomic regions affecting maize stomatal conductance with conditional effects of water deficit and evaporative demand. *Plant Cell Environ* **41**: 314–326. doi:10.1111/pce.13083
- Preston JC, Hileman LC. 2013. Functional evolution in the plant *SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE (SPL)* gene family. *Front Plant Sci* **4**: 80. doi:10.3389/fpls.2013.00080
- R Core Team. 2013. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. <http://www.R-project.org/>.
- Rincint R, Moreau L, Monod H, Kuhn E, Melchinger AE, Malvar RA, Moreno-Gonzalez J, Nicolas S, Madur D, Combes V, et al. 2014. Recovering power in association mapping panels with variable levels of linkage disequilibrium. *Genetics* **197**: 375–387. doi:10.1534/genet.ics.113.159731
- Schittenhelm S, Schroetter S. 2014. Comparison of drought tolerance of maize, sweet sorghum and sorghum-sudangrass hybrids. *J Agron Crop Sci* **200**: 46–53. doi:10.1111/jac.12039
- Seki M, Umezawa T, Urano K, Shinozaki K. 2007. Regulatory metabolic networks in drought stress responses. *Curr Opin Plant Biol* **10**: 296–302. doi:10.1016/j.pbi.2007.04.014
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**: 2498–2504. doi:10.1101/gr.1239303
- Shiferaw B, Prasanna BM, Hellin J, Bänziger M. 2011. Crops that feed the world 6. Past successes and future challenge to the role played by maize in global food security. *Food Secur* **3**: 307–327. doi:10.1007/s12571-011-0140-5
- Shinozaki K, Yamaguchi-Shinozaki K. 2007. Gene networks involved in drought stress response and tolerance. *J Exp Bot* **58**: 221–227. doi:10.1093/jxb/erl164
- Tardieu F, Simonneau T, Muller B. 2018. The physiological basis of drought tolerance in crop plants: a scenario-dependent probabilistic approach. *Annu Rev Plant Biol* **69**: 733–759. doi:10.1146/annurev-arplant-042817-040218
- Thimm O, Blasing O, Gibon Y, Nagel A, Meyer S, Krüger P, Selbig J, Müller LA, Rhee SY, Stitt M. 2004. Mapman: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J* **37**: 914–939. doi:10.1111/j.1365-313X.2004.02016.x
- Tripathi P, Rabara RC, Rushton PJ. 2014. A systems biology perspective on the role of WRKY transcription factors in drought responses in plants. *Planta* **239**: 255–266. doi:10.1007/s00425-013-1985-y
- Unterseer S, Bauer E, Haberer G, Seidels M, Knaak C, Ouzunova M, Meitinger T, Strom TM, Fries R, Pausch H, et al. 2014. A powerful tool for genome analysis in maize: development and evaluation of the high density 600 k SNP genotyping array. *BMC Genomics* **15**: 823. doi:10.1186/1471-2164-15-823
- Usadel B, Pore F, Nagel A, Lohse M, Czedik-Eysenberg A, Stitt M. 2009. A guide to using MapMan to visualize and compare omics data in plants: a case study in the crop species, maize. *Plant Cell Environ* **32**: 1211–1229. doi:10.1111/j.1365-3040.2009.01978.x
- Valliyodan B, Nguyen HT. 2006. Understanding regulatory networks and engineering for enhanced drought tolerance in plants. *Curr Opin Plant Biol* **9**: 189–195. doi:10.1016/j.pbi.2006.01.019
- Valot B, Langella O, Nano E, Zivy M. 2011. MassChroQ: a versatile tool for mass spectrometry quantification. *Proteomics* **11**: 3572–3577. doi:10.1002/pmic.201100120
- van der Sluis MR, Ng A, Fu J. 2014. Systems genetics: from GWAS to disease pathways. *Biochim Biophys Acta* **1842**: 1903–1909. doi:10.1016/j.bbdis.2014.04.025

- Vogel C, Marcotte EM. 2012. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* **13**: 227–232. doi:10.1038/nrg3185
- Wang X, Cai X, Xu C, Wang Q, Dai S. 2016. Drought-responsive mechanisms in plant leaves revealed by proteomics. *Int J Mol Sci* **17**: 1706. doi:10.3390/ijms17101706
- Wasinger VC, Zeng M, Yau Y. 2013. Current status and advances in quantitative proteomic mass spectrometry. *Int J Proteomics* **2013**: 180605. doi:10.1155/2013/180605
- Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, et al. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* **38**: 203–208. doi:10.1038/ng1702
- Zegada-Lizarazu W, Zatta A, Monti A. 2012. Water uptake efficiency and above- and belowground biomass development of sweet sorghum and maize under different water regimes. *Plant Soil* **351**: 47–60. doi:10.1007/s11104-011-0928-2
- Zhang J, Yang JR. 2015. Determinants of the rate of protein sequence evolution. *Nat Rev Genet* **16**: 409–420. doi:10.1038/nrg3950
- Zipper SC, Qiu J, Kucharik CJ. 2016. Drought effects on US maize and soybean production: spatiotemporal patterns and historical changes. *Environ Res Lett* **11**: 094021. doi:10.1088/1748-9326/11/9/094021

Received July 29, 2019; accepted in revised form September 24, 2020.