A proteomics-based systems genetics approach reveals environment-specific loci modulating protein co-expression and drought-related traits in maize

Running title: Systems genetics of 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], Michel Zivy[1*]

[1] GQE– Le Moulon, INRA, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, 91190, Gif-sur-Yvette, France

[2] LEPSE, INRA, Univ Montpellier, SupAgro, Montpellier, France

* Corresponding authors:

michel.zivy@inra.fr, +33-1-69-33-23-65

melisande.blein-nicolas@inra.fr, +33-1-69-15-68-06

Keywords: proteomics, GWAS, drought, Zea maïs, mass spectrometry, integrative biology

ABSTRACT

The evolution of maize yields under drought is of particular concern in the context of climate change and human population growth. To better understand the mechanisms associated with the genetic polymorphisms underlying the variations of traits related to drought tolerance, we used a systems genetics approach integrating high-throughput phenotypic, proteomics and genomics data acquired on 254 maize hybrids grown under two watering conditions. We show that water deficit, even mild, induced a strong proteome remodeling and a reprogramming of the genetic control of the abundance of many proteins. We identify close co-localizations between QTLs and pQTLs, thus highlighting environment-specific pleiotropic loci associated to the co-expression of drought-responsive proteins and to the variations of phenotypic traits. These findings bring several lines of evidence supporting candidate genes at many loci and provide novel insight into the molecular mechanisms of drought tolerance.

INTRODUCTION

Maize is the main crop worldwide (Shiferaw et al. 2011) in terms of production. Although it exhibits a high water use efficiency thanks to its C4 metabolism, it is also highly sensitive to water deficit. As an illustration, maize is twice as much affected by drought as wheat, with 39.3% and 20.6% respective yield reductions associated with 40% reduction of water (Daryanto et al. 2016). Improving maize yield under drought has been an important goal of breeding programs for several decades (Campos et al. 2004, 2006; Cooper et al. 2014). However, despite the overall genetic progress obtained, increases in drought sensitivity have been reported in several regions (Lobell et al. 2014; Zipper et al. 2016; Meng et al. 2016). In addition, severe drought stress episodes are projected to become more frequent in the near future due to climate change (Harrison et al. 2014). Therefore, the evolution of maize productivity under water deficit is of particular concern and large efforts are still required to design varieties able to maintain high yields in drought conditions.

One lever to accelerate the genetic progress is to better understand the genetic and molecular bases of drought tolerance. This highly complex trait is associated to a series of mechanisms occurring at different spatial and temporal scales to (i) stabilize the plant water and carbon status, (ii) control the side effects of water deficit including oxidative stress, mineral deficiencies and reduced photosynthesis and (iii) maintain the plant yield (Chaves et al. 2003). At the physiological level, short-term responses include stomata closure, osmotic and hydraulic conductance adjustments, 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 or scavenging of reactive oxygen species and the expression of drought-responsive proteins like dehydrins, late embryogenesis abundant (LEA) and heat shock proteins (HSP) (Valliyodan and

Nguyen 2006; Seki et al. 2007). All these responses depend on the drought scenario, the phenological stage, the genetic potential and the surrounding environment (Tardieu et al. 2018). Altogether, the multiplicity and versatility of the mechanisms coming into play explain the difficulty to select for drought tolerance.

Breeding new drought-tolerant varieties would greatly benefit from a better understanding of the genotype-phenotype relationship. Systems genetics is a recent approach allowing to gain better insight into this relationship by deciphering the biological networks and molecular pathways underlying complex traits and by understanding how they are regulated at the genetic and epigenetic levels (Nadeau and Dudley 2011; Civelek and Lusis 2014; Feltus 2014; van der Sijde et al. 2014; Markowetz and Boutros 2015). It consists in comparing the position of quantitative trait loci (QTLs) underlying phenotypic traits variation to that of QTLs underlying the variation of upstream molecular phenotypes such as transcript expressions (eQTLs) or protein abundances (pQTLs). Until now, this approach has been mostly applied in human and animals (Johnson et al. 2015; Williams et al. 2016; Moreno-Moral and Petretto 2016) and to a lesser extent in plants (Moreno-Moral and Petretto 2016; Munkvold et al. 2013; Ogura and Busch 2016; Basnet et al. 2016; Christie et al. 2017; Mizrachi et al. 2017).

The first studies that compared QTLs and pQTLs used 2D gel proteomics to quantify proteins (Bourgeois et al. 2011; de Vienne et al. 1999). 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 progress, the systems genetics studies published so far have preferentially used transcripts rather than proteins as intermediate level between the genome and end-point phenotypes. One reason is that large-scale proteomics experiments remain challenging (Blein-Nicolas et al. 2015) due to technical constraints (Balliau et al. 2018) and to the trade-off between depth of coverage and sample throughput (Keshishian et al. 2017). Yet, proteins are particularly relevant molecular components to link genotype to phenotype. Due to the buffering of

transcriptional variations and to the role of post-translational regulations in phenotypes construction, proteins are indeed expected to be more highly related to end-point phenotypes than transcripts (Foss et al. 2011; Battle et al. 2015; Chick et al. 2016; Albertin et al. 2013; Vogel and Marcotte 2012).

Here, we aimed to better understand the molecular mechanisms associated with the genetic polymorphisms underlying the variations of traits related to drought tolerance. To this end, we performed a unique systems genetics study where MS-based proteomics data acquired for 254 maize genotypes grown in two watering conditions were integrated with high-throughput genomic and phenotypic data. Protein abundances were first analyzed by using genome wide association study (GWAS) and co-expression networks. Then they were integrated with phenotypic data measured for drought-related trait in the same conditions (Prado et al. 2018) through correlation analysis and search for QTL/pQTL co-localizations. We showed that water deficit, even mild, caused a deep proteome remodeling associated to changes in the genetic architecture of protein abundances. Some of these changes could also affect drought-related traits, as indicated by QTL/pQTL co-localizations. These were underlied by exciting candidate genes. In particular, we identified two transcription factors likely involved in the condition-specific co-regulation of drought-responsive proteins as well as in the genetic variations of several phenotypic traits. We also highlighted many cases of QTL/pQTL co-localizations with a reduced number of underlying candidate genes.

RESULTS

Mild water deficit has extensively remodeled the proteome

Using MS-based proteomics, we analyzed more than 1,000 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, our peptide intensity dataset included 977 samples corresponding to 251 genotypes from which we reliably quantified 1,950 proteins. For 977 of them that exhibited too many missing peptide intensity values, quantification was performed based on the number of chromatographic peaks (PC-based set). For the remaining 973 proteins that were on average more abundant (Sup. Figure 1A), quantification was performed based on peptide intensities, that provide a higher precision of quantification than counting data (Sup. Figure 1B, XIC-based set). Functional categories involving highly abundant proteins, such as energy metabolism, were better represented in the XIC-based than in the PC-based set (Sup. Figure 1C).

Heatmap representations of protein abundances show that the two watering conditions were well separated by two large protein clusters (Figure 1A), indicating that, although moderate, water deficit has extensively remodeled the proteome of most genotypes. Accordingly, 82.4% and 71.7% of the proteins of the XIC-based and PC-based sets respectively responded significantly to water deficit (adjusted *P-value* < 0.05, WD/WW ratio > 1.5 or < 0.66, Sup. Table 1). These included several proteins known to be involved in responses to drought or stress (hereafter named drought-responsive proteins) such as dehydrins (GRMZM2G079440, GRMZM2G373522), ABA-responsive protein (GRMZM2G106622), LEA protein (GRMZM2G352415), HSPs (GRMZM2G360681, GRMZM2G080724, GRMZM2G112165), phospholipase D (GRMZM2G061969), glyoxalase I (GRMZM2G181192) or gluthathione-S-transferase (GRMZM2G043291). Induced and repressed proteins constituted two populations highly differentiated in terms of function (Figure 1B). In particular, transcription, translation, energy metabolism and metabolism of cofactors and vitamins

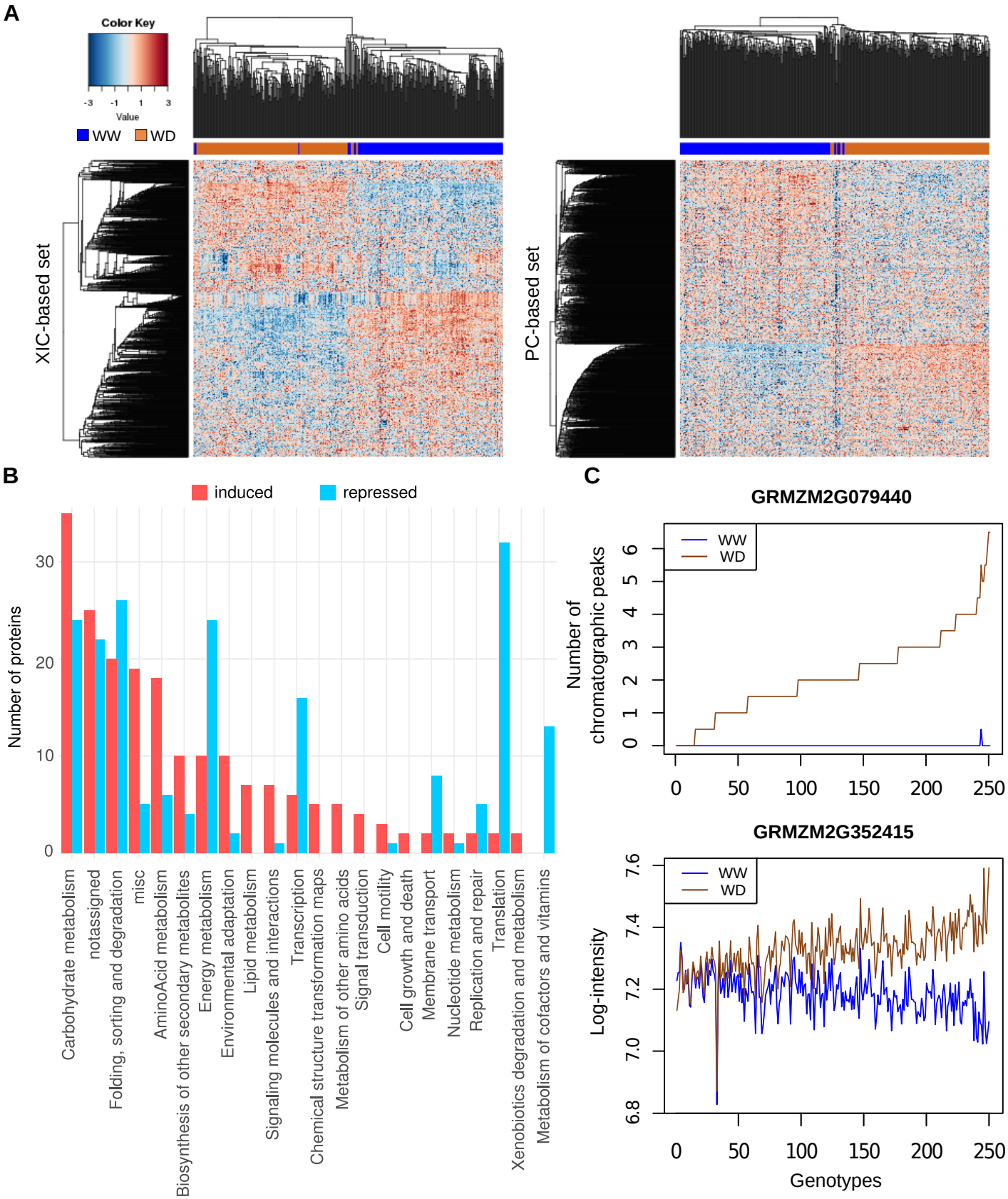

Figure 1. Effect of a mild water deficit on the proteome. (A) Heatmap representations of the abundances estimated for the XIC-based protein set (left) and the PC-based protein set (right). Each line corresponds to a protein and each column to a genotype x watering condition combination. For each protein, abundance values were scaled and represented by a color code as indicated by the color-key bar. Hierarchical clusterings of the genotype x watering condition combinations (top) and of the proteins (left) were built by using the 1-Pearson correlation coefficient as distance and the unweighted pair group method with arithmetic mean (UPGMA) as aggregation method. (B) Functions of the 200 most induced and 200 most repressed proteins under water deficit. (C) Abundance profiles of the RAB17 dehydrin (GRMZM2G079440 quantified based on the number of chromatographic peaks) and of a LEA protein (GRMZM2G352415 quantified based on peptide intensities) in the two watering conditions. Genotypes on the x axis were ordered according to the WD/WW abundance ratio.

were better represented within repressed proteins, while carbohydrate and amino acid metabolisms, environmental adaptation, signaling molecules and interaction were better represented within induced proteins.

The global impact of a genotypic change on the proteome was less extensive than that of water deficit, since the proteomes of two different genotypes grown in the same watering condition were more similar than the proteome of a same genotype grown in different conditions (Figure 1A). However, the maximum amplitudes of abundance variations were similar (Sup. Figure 2). In addition, 94.9% of the proteins from the XIC-based set exhibited significant genetic variations of abundance (adjusted *P-value* < 0.05, fold change > 1.5 or < 0.66). This was confirmed by broad sense heritability, the median of which was 0.47 and 0.46 for the WW and WD conditions, respectively (Sup. Figure 3A). By contrast, in the PC-based set, only 37.4% of the proteins showed significant genetic variations of abundance and the median of broad sense heritability was 0.07 and 0.09 in the WW and WD conditions, respectively (Sup. Figure 3B).

Regarding the genotype x environment (GxE) interactions, several results indicate that they were larger in the PC-based than in the XIC-based set. The correlations of protein abundances between the two watering conditions were lower in the former than in the latter (median of r =0.08 and 0.44, respectively; Sup. Figure 4A). Furthermore, the contribution of the GxE interactions to the total variability of protein abundances was higher in the PC-based than in the XIC-based set (Sup. Figure 4B-D). However, significant GxE interactions were detected for only four and 15 proteins of the PC-based and XIC-based sets, respectively, probably because of a lack of statistical power. These proteins included a LEA protein (GRMZM2G352415) and the ZmRab17 dehydrin (GRMZM2G079440), the latter being undetectable in the WW condition and more or less expressed depending on the genotype in the WD condition (Figure 1C).

Genetic architecture of protein abundance is related to protein function

To identify pQTLs, we submitted the 3,900 molecular phenotypes (1,950 proteins x 2 watering conditions) obtained from the proteomics analysis to GWAS. In total, we detected 583,288 (163,674 in the PC-based set and 419,614 in the XIC-based set) significant associations (*P-value* < 10^{-5}) for 3,759 (96.4%) molecular phenotypes. Associated SNPs explained between 0.3 and 83.2% of variance, with an average at 8.7% (Sup. Figure 5).

To summarize associated SNPs into pQTLs, we developed a geometric method based on the *P-value* signal of SNPs. Compared to classical methods based on the genetic distance or on linkage disequilibrium (LD), it allowed to detect the lowest number of pQTLs per molecular phenotype (median at 8, 10 and 7, respectively) and the lowest maximum number of pQTLs per chromosome (57, 209 and 15, respectively). It also produced the lowest correlation between the number of pQTLs per chromosome and the *P-value* of the most strongly associated pQTL on the corresponding chromosome (Sup. Figure 6). In total, we thus detected 29,004 pQTLs (16,911 and 12,093 in the XIC-based and PC-based sets, respectively; sup. Table 2). The median number of pQTLs per molecular phenotype was 8 and 5 for the XIC-based and PC-based sets, respectively (Sup. Figure 7A). 1,385 (4.8%) pQTLs were local, *i.e.* located at less than 10^6 bp from the protein encoding gene, of which 442 were located within the genes. Among the distant pQTLs, 81.7% were located on a chromosome different than that of the protein encoding gene. Local pQTLs had stronger effects than distant pQTLs (average *R2*=19.3% and 5.5%, respectively; Sup. Figure 7B). For 992 proteins, no local pQTL was detected in any of the conditions. Compared to the 816 proteins showing a local pQTL in at least one condition, these proteins were significantly enriched in proteins involved in translation (13.6% *vs* 4.2%, adjusted *P-value* = $4.8e^{-11}$) and energy metabolism (11.2% *vs* 6.9%, adjusted *P-value* = 0.016) and depleted in proteins involved in carbohydrate metabolism (10.5% *vs* 18.1%, adjusted *P-value* = $4.2e^{-05}$). They also exhibited less distant pQTLs (Sup. Figure 8 A, B) and were much less heritable (Sup Figure 8 C, D). These results

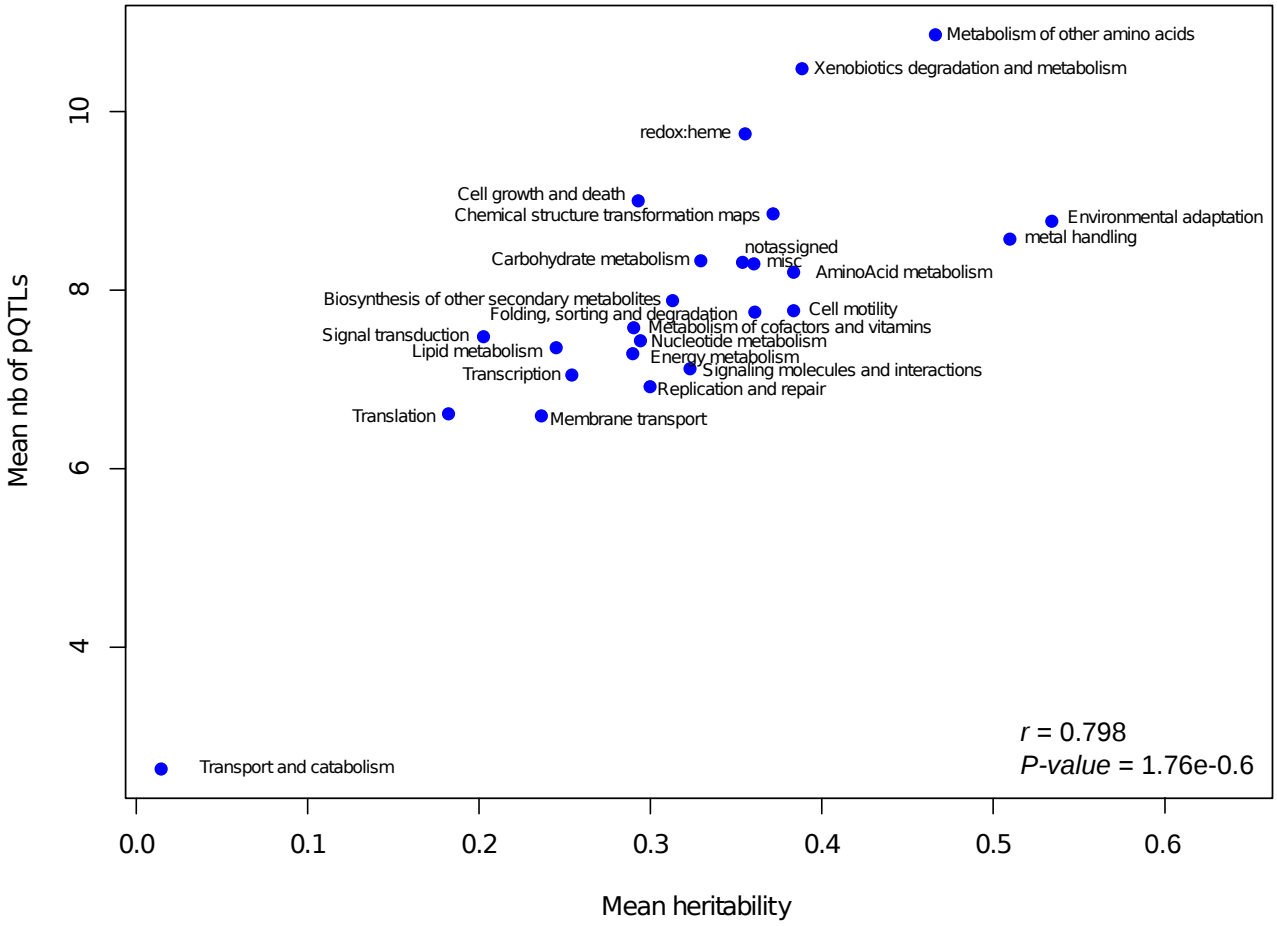

Figure 2. Relationship between the mean number of pQTLs per KEGG category and the mean heritability per KEGG category.

indicate that genetic regulation of protein abundances 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 (Figure 2).

Identification of genomic regions with pleiotropic effects on the proteome

pQTLs were not uniformly distributed in the genome (Figure 3). Instead, there were genomic regions enriched with pQTLs. We detected 25 and 23 of such hotspots that contained more than 20 pQTLs in the WW and WD conditions, respectively (Sup. Table 3). For 15 and 13 of them respectively, the abundances of the associated proteins were more strongly correlated than expected by chance (*P-value* < 0.01; sup. Figure 9, Sup. Table 3), which indicates an enrichment in co-expressed proteins. These hotspots might thus represent regulatory loci with pleiotropic effects on the proteome.

To complement these results, we performed a weighted gene co-expression network analysis (WGCNA) of protein co-expression across the 251 genotypes in the two watering conditions separately. Co-expression networks in the WW and WD conditions captured 1,671 proteins in 11 modules and 1,578 proteins in 14 modules, respectively (Figure 4, Sup. Table 4). Each module contained 31 to 377 proteins. The WD network was well structured according to protein function with many modules showing significant functional enrichments (Figure 4, Sup. Table 5). Because most modules were well preserved between conditions (Sup. Figure 10A, B), we could build a consensus network capturing 698 proteins and composed of 15 modules containing 14 to 197 proteins (Figure 4, Sup. Figure 10C; Sup. Table 4). Consensus modules were more particularly enriched in photosynthesis proteins, ribosomal proteins and in proteins involved in mitochondrial electron transport, ATP synthesis and in the tetrapyrrole pathway (Figure 4, Sup. Table 5). Two modules were condition-specific (Figure 4, Sup. Figure 10). Note that the module specific of the WD condition and was significantly enriched for drought-responsive proteins (Sup. Table 5).

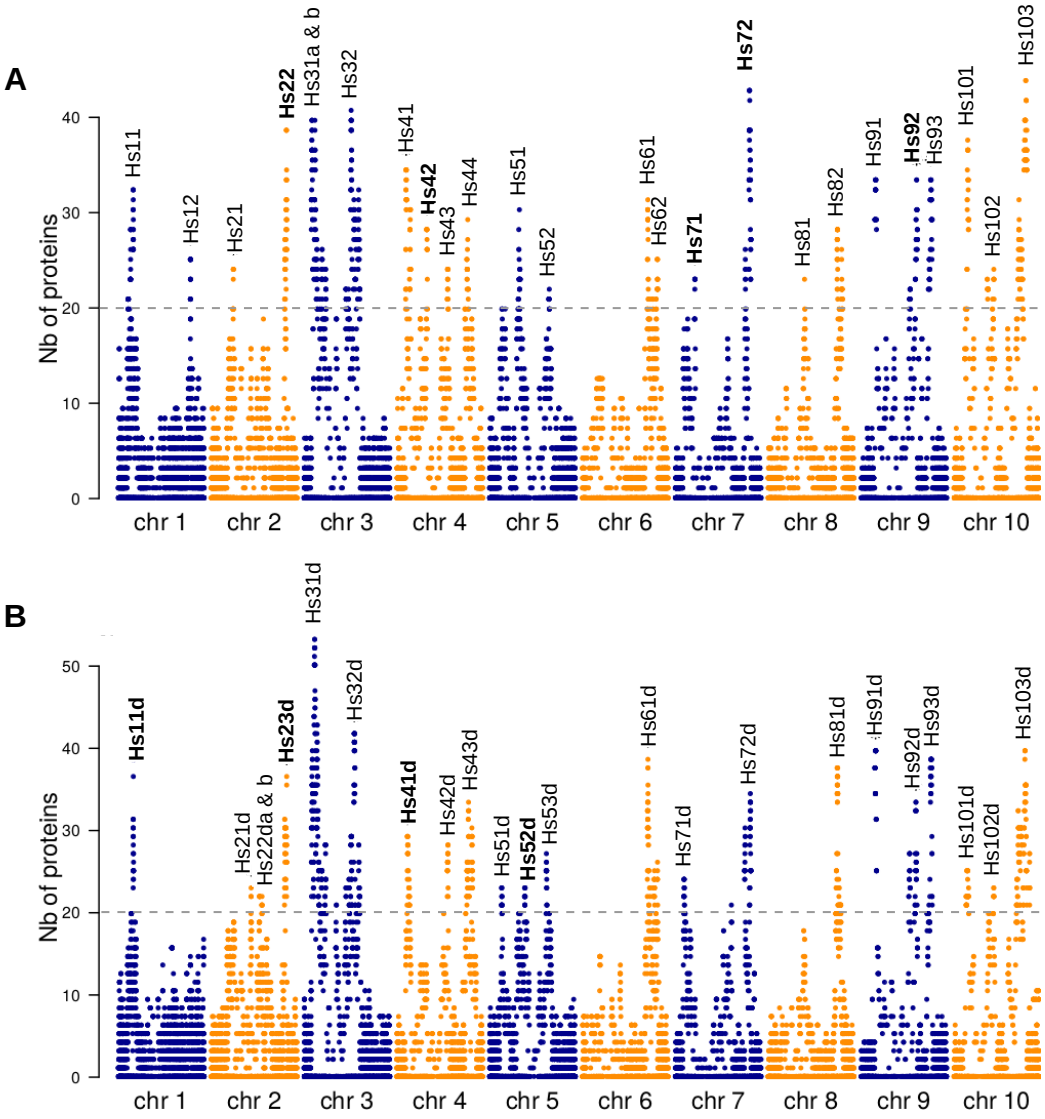

Figure 3. Distribution of pQTLs along the genome. (A) In the well-watered condition. (B) In the water deficit condition. Each point indicates the number of proteins controlled by a pQTL located in a given genomic region defined by the linkage disequilibrium interval around a SNP. Dashed horizontal lines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

Protein abundances within each module were summarized into variables called eigengenes that were considered as molecular phenotypes and submitted to GWAS. In total, we detected 2,859 significant associations (-log10(*P-value*) > 5) for 55 eigengenes (11, 14 and 30 for WW, WD and consensus modules, respectively). Significantly associated SNPs were summarized into 369 co-expression QTLs (coQTLs, Sup. Table 2). When a coQTL co-localized with a pQTL hotspot, we checked whether the proteins associated to the hotspot were significantly enriched for proteins belonging to the module controlled by the coQTL (*P-value* < 0.01). Thus crossing the results, we confirmed nine hotspots as potential pleiotropic loci. Five of them were in the WW condition (Hs22, Hs42, Hs71, Hs72 and Hs92) and four were in the WD condition (Hs11d, Hs23d, Hs41d and Hs52d; Sup. Table 3). Remarkably, several of these hotspots were associated to proteins exhibiting consistent functions. For example, hotspots Hs42 and Hs23d were mainly associated to ribosomal proteins (Sup. Table 6) and hotspots Hs71 and Hs72 were associated to proteins involved in energy metabolism and more particularly ATP synthesis and photosynthesis.

The genetic architecture of protein abundances depends on the environment

Of the 14,432 pQTLs detected in the WW condition, only 1,212 (8.4%) had a co-localizing pQTL in the WD condition. These pQTLs were generally of strong effects (Sup. Figure 11A) and were enriched for local pQTLs (37.7% *vs* 4.8% in the whole dataset). Interestingly, while most of the pQTLs shared across conditions had similar effects in the two conditions, 80 of them (6.6%) exhibited contrasted effects (Sup. Figure 11B). Half of these pQTLs were local, suggesting that gene promoters may be involved in the GxE interaction or that the pQTLs detected in each condition corresponded to different polymorphic sequences with different effects on protein abundances. These pQTLs were associated to 75 proteins, several of which were drought-responsive (Sup. Table 7).

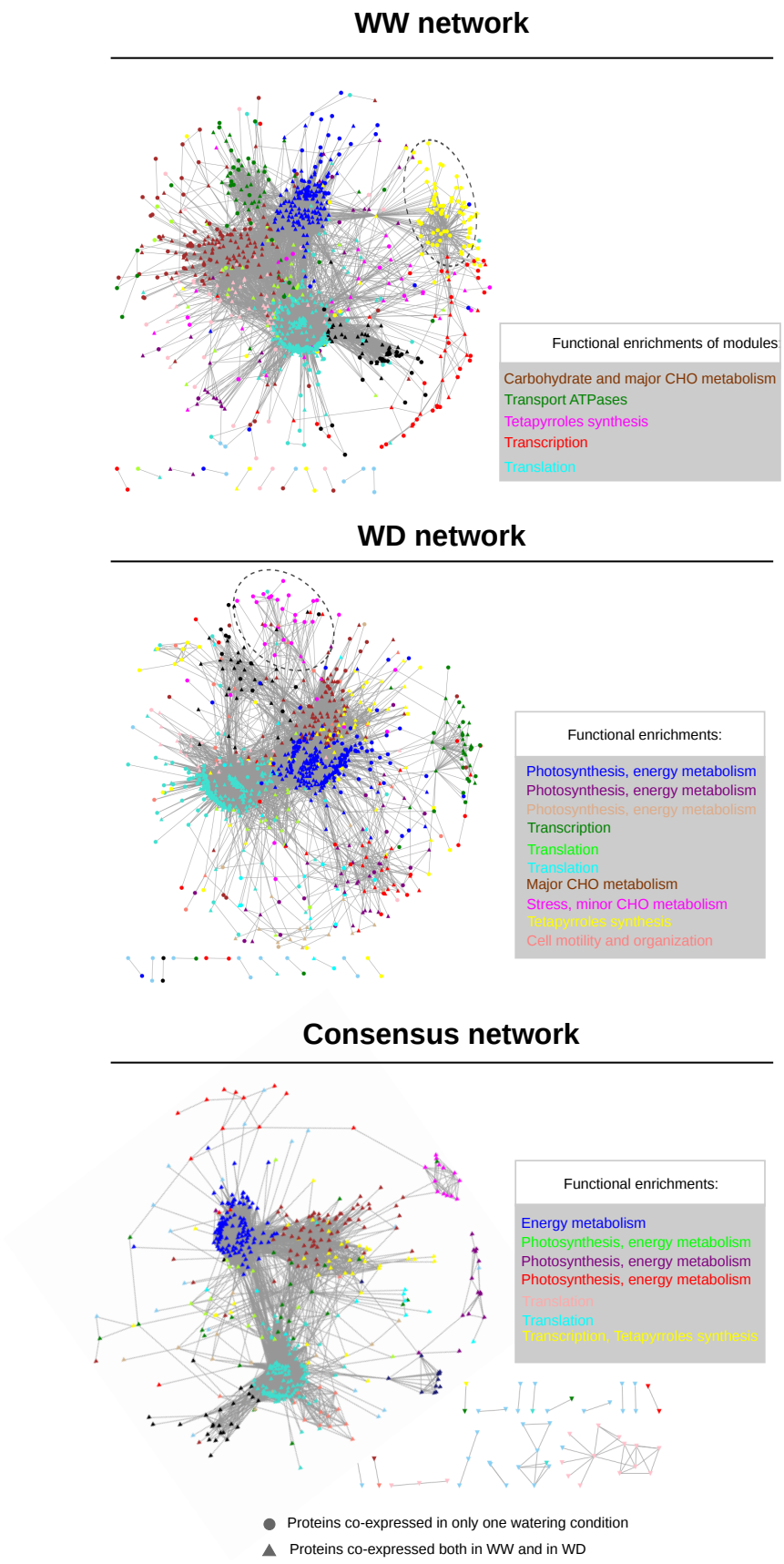

Figure 4. Graphical representation of the co-expression networks resulting from the WGCNA analyses. Only proteins showing adjacencies > 0.02 are shown. The consensus network contains the proteins that were co-expressed in the two watering conditions. The three views were created by Cytoscape v3.5.1 using an unweighted, spring-embedded layout. The colors displayed on each network represent the different modules identified by WGCNA. Functional enrichments of modules are indicated in grey boxes. Condition-specific modules are surrounded by dashed circles.

As observed for individual proteins, eigengenes for consensus modules were poorly to moderately correlated between conditions (*r* between -0.03 to 0.62, Sup. Figure 12) and only one coQTL was shared across conditions. Regarding the nine cross-validated pQTL hotspots, one co-localization was found between Hs22 and Hs23d. However, these two hotspots shared only one associated protein, suggesting that they represented two different loci. Altogether, these results indicate that the positions of genomic regions with pleiotropic effects on the proteome also depended on the environment.

Identification of genomic regions involved in multi-scale genetic control

To gain insight into the molecular mechanisms associated to drought tolerance, we searched for co-localizations between the pQTLs, coQTLs or hotspots detected in our study and the 160 QTLs detected by Alvarez Prado *et al.* (2018) on the same plant material. These QTLs controlled eight phenotypic traits related to growth and transpiration rate in the WW and WW conditions: leaf area early (*i.e.* before water deficit; LAe), leaf area late (LAl), biomass early (Be), biomass late (Bl), water use (WU), water use efficiency (WUE), stomatal conductance (gs) and transpiration rate (Trate). To select for robust co-localizations, we took into account the correlations corrected by the structure and kinship between the trait values and the protein abundances or the module eigengenes ($|r_{corrected}| > 0.3$).

In total, we identified 59 pairs of SNPs corresponding to QTL/pQTL co-localizations (Figure 5, Sup. Table 8). They involved six phenotypic traits (Bl, LAl, WU, WUE, Trate and gs) and 42 proteins, many of which were drought-responsive (Table 1). Most QTL/pQTL co-localizations (96.6%) were detected in the WD condition, where they corresponded to 40 of the 91 QTLs reported in this condition (Prado et al. 2018). All but one involved distant pQTLs. For 12 cases out of 59 (20%), the co-localizing QTL and pQTL were represented by the same SNP. In most of the remaining cases, the QTL/pQTL distance was less than 100 Kb (Figure 6A). Seventeen proteins

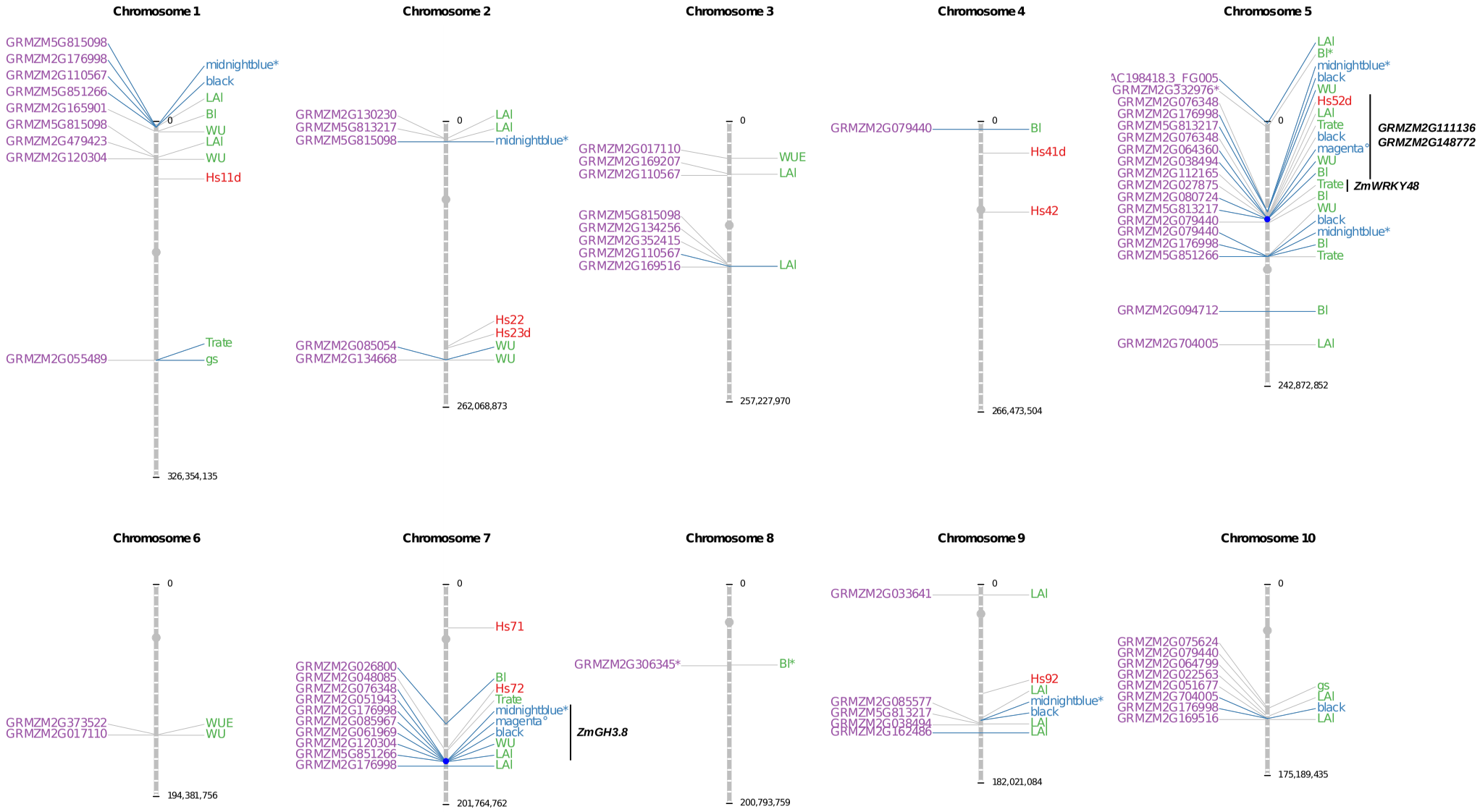

Figure 5. Genomic positions of the co-localizing pQTLs, coQTLs and QTLs. The positions of the nine pQTL hotspots robustly identified as potential loci with pleiotropic effects are indicated as well as the position of the most promising candidate genes. Chromosomes are segmented in 1 Mb bins. Grey dots represent the centromeres and blue dots indicate the position of genomic regions showing evidences for pleiotropy both at the proteome and phenotype level. Blue lines indicate pQTLs, coQTLs and QTLs that are determined by a same SNP.
* consensus module, ° WD-specific module

exhibited multiple QTL/pQTL co-localizations (eight at maximum, Table 1). In particular, the ZmRab17 protein presented seven QTL/pQTLs co-localizations involving four traits (Bl, Trate, WU and gs with respectively 3, 2, 1 and 1 co-localizations; Sup. Table 8). LAl was the trait that exhibited the highest number of co-localizations (44) involving 29 proteins and 16 QTLs (Sup. Table 8).

We further identified 13 pairs of SNPs corresponding to QTL/coQTL co-localizations, all in the WD condition (Sup. Table 9). They involved four phenotypic traits (WU, Bl, LAl and Trate) and three modules including the WD-specific module (Figure 5). Eleven of the 13 QTLs co-localizing with coQTLs also co-localized with pQTLs. The remaining two QTLs (at SNPs S5-88791868 on chromosome 5 for Trate and AX-91801223 on chromosome 9 for LAl) actually also co-localized with pQTLs, but with low correlations between the trait values and the protein abundances ($|r_{corrected}| < 0.06$; Sup. Table 9). By contrast, the correlations between the trait values and the module eigengenes were much higher ($|r_{corrected}| = 0.32$ and 0.50, Sup. Table 9).

Taken together, these results indicate the presence of genomic regions involved in the genetic control of traits at different levels of biological complexity. Some of these regions may control multiple proteins which are more strongly related to end-point phenotypes when taken collectively through co-expression module rather than individually. These results also revealed two genomic regions showing evidences for pleiotropy both at the proteome and phenotype levels (Figure 5). The first, located on chromosome 5 and spanning 1,8 Mb between SNPs AX-91657926 and AX-90612012, contained pQTLs for seven proteins, coQTLs for two modules and QTLs for four phenotypic traits (LAl, BL, WU, and Trate). This region was also covered by hotspot Hs52d. The second is on chromosome 7 where a single SNP (S7_162671160) determined the positions of pQTLs for seven proteins, coQTLs for three modules and QTLs for two phenotypic traits (LAl and WU).

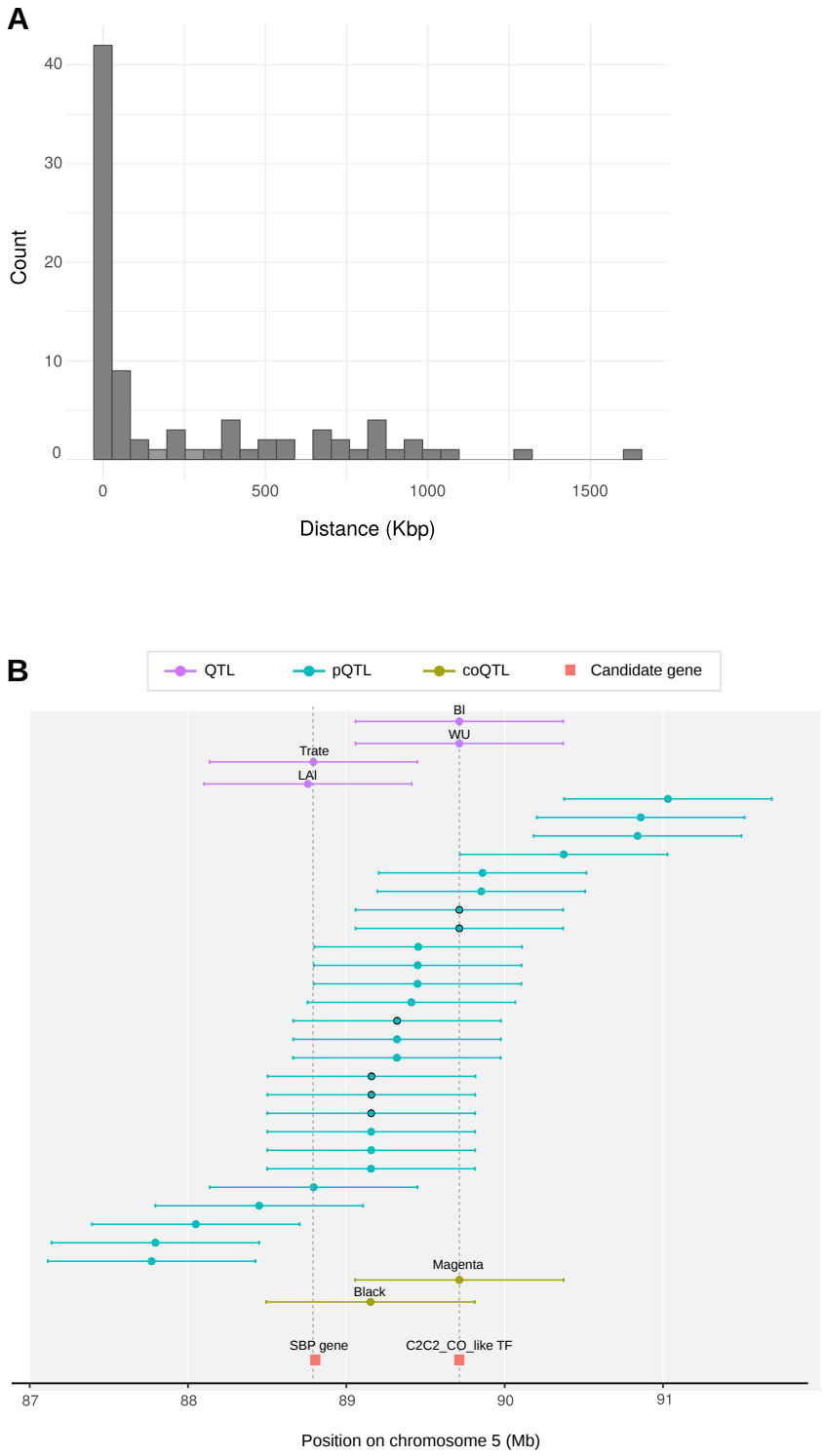

Figure 6. Identification of genomic regions involved in multi-scale genetic control.
(A) Distribution of the distances between co-localizing QTLs and pQTLs. (B) Detailed view of the QTL, pQTL, coQTL detected in the region covered by the hotspot Hs52d on chromosome 5. Dots represent the SNPs determining the position of the QTLs and horizontal bars represent the linkage disequilibrium-based window around each SNP. Black circled dots indicate the pQTLs that co-localize with QTLs or coQTLs with high correlations between the protein abundance and the phenotypic trait value or the module eigengene. Vertical dashed lines indicate the position of SNPs S5_88793314 (on the left) and AX-91658235 (on the right). The position of two transcription factors (a SBP gene, GRMZM2G111136, and a C2C2-CO-like transcription factor, GRMZM2G148772) representing promising candidate genes are indicated.

Unraveling the molecular mechanisms associated to genetic factors underlying the variations of traits related to drought tolerance

Assuming that the genetic determinants controlling the variations of both phenotypic traits and protein abundances (or module eigengenes) were genes, we retrieved the genes underlying the QTL/pQTL (or QTL/coQTL) co-localizations. We thus obtained a list containing one to 45 candidate genes for each of the 64 pairs of SNPs corresponding to QTL/pQTL or QTL/coQTL co-localizations (Sup. Table 10). Based on gene annotation and literature, we identified in most cases interesting candidate genes involved in hormone metabolism, regulation of transcription, signaling, stress or drought response. Three cases especially caught our attention.

First, on chromosome 7, the SNP S7_162671160, mentioned above for its pleiotropy, was located in the *ZmGH3.8* gene (GRMZM2G053338 encoding an auxin-response factor). In agreement with the role of *ZmGH3.8* in drought response (Feng et al. 2015), S7_162671160 was associated to the WD-specific module and to five stress proteins (endochitinase GRMZM2G051943, beta-D-glucanase GRMZM2G073079, peroxidase GRMZM2G085967, polyphenol oxydase GRMZM5G851266 and phospholipase D GRMZM2G061969). These results indicates that *ZmGH3.8* may be the regulator underlying the QTLs, pQTLs and coQTLs located at SNP S7_162671160.

Second, on chromosome 5, we identified a QTL for Trate which co-localized with a pQTL associated to the ZmRab17 dehydrin in the WD condition. Among the six genes underlying both the QTL and the pQTL is the transcription factor *ZmWRKY48* (GRMZM2G120320), which has been shown to be induced by water deficit (Song et al. 2017). This gene is orthologous to *AtWRKY40*, which can bind to a W-box sequence in the promoter of the *AtRab18* gene and repress its expression (Shang et al. 2010). *ZmRab17* is orthologous to *AtRab18* and also contains a W-box sequence in a region upstream of the 3'UTR. Altogether, these results indicate that *ZmWRKY48* probably controls the abundance of ZmRab17.

Third, in the region of chromosome 5 covered by hotspot Hs52d (Figure 5), there were 14 candidate genes, of which two can potentially control a high number of genes. One is a squamosa promoter-binding (SBP) gene (GRMZM2G111136) shown to be induced by various abiotic stresses including drought (Mao et al. 2016). The other, a C2C2-CO-like transcription factor (GRMZM2G148772), was shown to be significantly induced by drought and salinity stress in B73 leaves (Forestan et al. 2016). Hotspot Hs52d covered a region of ca 4 Mb in which we detected 26 pQTLs, many of which were located between the SBP gene and the C2C2-CO-like transcription factor (Figure 6B). In this region, there were also coQTLs for two modules and QTLs for WU, Bl, Trate and LAl. Interestingly, a single SNP, AX-91658235 located only one kbp from the C2C2-CO-like transcription factor, determined the position of two QTLs, two pQTLs and one coQTL. Furthermore, SNP S5_88793314, that fell into the coding sequence of the SBP gene, determined the position of a QTL for Trate and of a pQTL. Based on these results, we can hypothesize that hotspot Hs52d may in fact correspond to two trans-acting regulators for which the SBP gene and the C2C2-CO-like transcription factor represent good candidates.

DISCUSSION

In order to better understand the molecular mechanisms associated with the genetic polymorphisms underlying the variations of traits related to drought tolerance, we analyzed the proteomes of 254 dent maize genotypes grown in two watering conditions. By combining two complementary MS-based quantification methods (XIC and PC-based), we reliably quantified nearly 2,000 proteins in an unprecedented number of samples (>1,000). To our knowledge, this is the best compromise ever obtained in MS-based proteomics between the number of samples analyzed and the number of proteins quantified. Compared to the PC-based set, the proteins of the XIC-based set exhibited higher heritabilities and more genetic variations of their abundances. These discrepancies probably partly arose from the difference of precision in protein quantification.

Based on high-density genotyping data, we subsequently performed GWAS for ca. 4,000 molecular phenotypes, thus mapping 29,004 pQTLs at high-resolution. To achieve this result, we had to summarize associated SNPs into pQTLs. This issue emerged only recently with the advent of high marker densities and no gold standard method is currently available. Here, we developed a geometric method based on the *P-value* signal of SNPs which allowed to take into account that the number of significantly associated SNPs increased with the association strength. To our knowledge, this relationship has never been reported before. We did not go further on this issue which was outside the scope of this study. Nonetheless, our work opens the way towards new methodological developments accounting for the strength of SNP associations for the detection of pQTLs.

Local pQTLs explained a higher proportion of the variance of protein abundance than distant pQTL. They also revealed that the genetic architecture of protein abundances is related to protein function. Notably, proteins involved in translation and in energy synthesis exhibited few associated pQTLs with a lack of local pQTLs. These two functional categories were also particularly well represented in consensus modules, in agreement with previous results showing that

344 genes in consensus modules had few associated eQTLs (Munkvold et al. 2013). As translation and energy metabolism mainly contain ancient and evolutionarily conserved proteins (Goldman et al. 346 2010; Nelson and Junge 2015), our results suggest that the expression of evolutionarily ancient proteins is more constrained with fewer associated pQTLs (Mähler et al. 2017; Popadin et al. 2014; 348 Zhang and Yang 2015). 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 350 associated to stabilizing selection on gene expression.

pQTLs were distributed throughout the genome but some of them clustered into hotspots, 352 suggesting the presence of regulatory loci with large pleiotropic effects on the proteome. The detection of such hotspots is highly dependent on the mapping resolution and on the method used to 354 cluster eQTLs or pQTLs, which may explain the contrasted results reported in the literature. Indeed, while several studies report hotspots associated to hundreds of transcripts (Munkvold et al. 2013; 356 Christie et al. 2017; Orozco et al. 2012), others detected hotspots associated to only a few tens of transcripts or proteins (Foss et al. 2011; Ghazalpour et al. 2011; Albert et al. 2014) or even no 358 hotspot at all (Mähler et al. 2017). In our study, false hotspot detection was limited by the high mapping resolution and by the use of a pQTL clustering method taking into account the variations 360 of LD across the genome (Negro *et al.*, in print). Despite this, we cross-validated only five hotspots in WW and four in WD based on protein co-expression and co-localization with coQTLs. The 362 positions of these hotspots were not shared across conditions. These results indicate that polymorphic loci responsible for the variations in abundance of tens or hundreds of proteins are 364 scarce and can interact with the environment.

By analyzing a diversity panel of 254 genotypes, we showed that many small abundance 366 changes, detected as significant because they occurred in a high number of genotypes, contributed to extensively remodel the proteome under water deficit. In total, approximately 75% of the 368 quantified proteins responded significantly to the environmental change. Up- and down regulated

proteins were well differentiated in terms of function and indicate that the photosynthetic, transcriptional and translational machineries were slowed down while stress response and signalization mechanisms were activated. All these changes show that plants clearly perceived a lack of water and presented a coordinated proteome response to the environmental change.

Interestingly, the abundance changes occurring in response to water deficit were not associated with major changes in the structure of the co-expression network since most co-expression modules were more or less preserved in both conditions. Nonetheless, we identified a WD-specific module that was significantly enriched for drought-responsive proteins. Similarly, Munkvold *et al.* (2013) observed condition-specific modules related to biological processes responsive to particular environmental conditions. Such modules suggest that, under environmental perturbation, sets of genes or proteins are collectively mobilized by condition-specific factors allowing the plant to adapt. In agreement with this hypothesis, the WD-specific module was associated to several QTL/coQTL colocalizations, with high correlations between its eigengene and the phenotypic traits values. In the case of transpiration rate, this correlation was even better than with any of the proteins in the module. This indicates that drought-responsive proteins are major contributors to the phenotypes, an observation reinforced by the fact that many QTL/pQTL co-localizations involved such proteins. Remarkably, 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 co-expression observed for drought-responsive proteins may be driven by condition-specific factors, the pleiotropic effects of which resonate across all layers of biological complexity up to end-point phenotypes.

Linking phenotypic to proteome variations revealed many QTL/pQTL co-localizations for which, thanks to the high mapping resolution, we identified a restrained number of candidate genes. Among these, it is noteworthy to mention Z*mWRKY48* underlying a QTL for transpiration rate and a pQTL for the ZmRab17 protein that is specifically expressed under water deficit. Z*mWRKY48* is a

394 transcription factor known to be induced by water deficit in maize (Song et al. 2017) and its role in the abundance variations of ZmRab17 is strongly supported by the literature data available on
396 *AtWRKY40* and *AtRAB18* in *Arabidopsis* (Shang et al. 2010).

Surprisingly, only two of the 69 QTLs detected in the WW condition were involved in pQTL
398 co-localizations, while in the WD condition, 40 of the 91 QTLs co-localized with at least one pQTL. To explain this discrepancy, we assume that under non-stress condition, phenotypic
400 variations were driven by many low contribution proteins probably controlled by low effect pQTLs while under water stress, phenotypic variations were mainly driven by drought-responsive proteins
402 under the genetic control of condition-specific regulators. In agreement with this hypothesis, we robustly identified two genomic regions that may correspond to such regulators. The first is located
404 on chromosome 7, where we identified the auxin response factor (ARF) *ZmGH3.8* as the unique candidate gene underlying QTLs for leaf area and water use, pQTLs for seven proteins of which
406 five were involved in stress responses and coQTLs for three co-expression modules including the WD-specific module. ARFs play key roles in plant growth and development through the regulation
408 of expression of auxin response genes which can include transcription factors (Li et al. 2016). They are also thought to contribute to drought tolerance (Feng et al. 2015; Zhang et al. 2017). In maize
410 shoots, Feng *et al.* (2015) indeed showed that the expression of *ZmGH3.8* was induced by auxin and reduced under polyethylene glycol treatment. More recently, Zhang *et al.* (2017) identified 13 ARFs
412 as differentially expressed between a drought tolerant and a drought sensitive maize line under different drought scenarios. The second region is located on chromosome 5 where we identified two
414 transcription factors, a SBP gene (GRMZM2G111136) and a C2C2-CO-like gene (GRMZM2G148772), as candidate genes underlying QTLs for key plant growth and transpiration
416 traits, pQTLs for seven proteins and coQTLs for three modules. This region also contained the pQTL hotspot Hs52d. Both GRMZM2G111136 and GRMZM2G148772 were previously shown to
418 be induced by drought in maize (Mao et al. 2016; Forestan 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). Due to their potential implication in the GxE interactions as pleiotropic, condition-specific regulator and because of their roles in both plant growth and development and in drought response, *ZmGH3.8*, the SBP gene and the C2C2-CO-like transcription factor represent particularly promising candidates for drought tolerance breeding.

To conclude, by using a systems genetics approach including MS-based proteomics data, we highlighted several original results. We showed that water deficit, even mild, strongly remodeled the proteome and induced a reprogramming of the genetic control of the abundance of many proteins and notably those involved in drought responses. Furthermore, we point out that the genetic architecture of protein abundances is related to protein function and also probably to the evolutionary constraints on protein expression. Finally, we found QTL/pQTL co-localizations mostly in the WD condition and we identified exciting candidate genes in the vicinity of WD-specific polymorphisms responsible for both the co-expression of drought-responsive proteins and the variations of drought-related traits. This suggests that the reprogramming of the genetic control observed at the proteome level may also affect end-point phenotypes. Taken together, our results demonstrate that proteomics has now reached enough maturity to be fully exploited in systems studies necessitating large-scale experiments. Our findings also provide novel insights into the molecular mechanisms of drought tolerance and highlight some pathways for further research and breeding.

METHODS

Plant material and experiment

A diversity panel of maize hybrids obtained by crossing a common flint parent (UH007) with 254 dent lines was used (Millet et al. 2016). The flint line was used as the paternal parent. The experiment was carried out as in Prado *et al.* (2018). Plant growth conditions (type of soil used, temperature, light, irrigation) are described in full details in the aforementioned publication. Briefly, plants were sown on May 14, 2012 and grown in pots in the phenotyping platform PhenoArch (Cabrera-Bosquet et al. 2016) (https://www6.montpellier.inra.fr/lepse_eng/M3P/PHENOARCH-platform) hosted at the Montpellier Plant Phenotyping Platforms (https://www6.montpellier.inra.fr/lepse_eng/M3P). Two levels of soil water content were applied: well-watered (WW, soil water potential of -0.05 MPa) and water deficit (WD, soil water potential of -0.45 MPa). Hybrids were replicated three times in each of the watering condition.

Sampling was performed at the pre-flowering stage (between June 19 and June 22, 2012) in two replicates per hybrid and water condition. For each sampled plant, 35 to 45 mg of fresh material was taken in a 2 mL tube containing two iron metal beads (5 mm diameter) by punching ten patches (5 mm diameter) in the mature area of the last ligulated leaf. The tubes were frozen in liquid nitrogen immediately after sampling and stored at -80°C until protein extraction.

Protein extraction and digestion

Leaf patches were turn into powder by shaking frozen sample tubes twice at 20 Hz for 20 seconds using TissueLyzer II (Quiagen, Courtaboeuf, France). Beads were removed using a magnet. Proteins were precipitated by incubating the leaf powder in 1.2 ml of an ice-cold solution of acetone containing 10% of trichloroactetic acid and 0.07% β-mercaptoethanol for 1h20 at -20°C. After centrifugation (10 min, 0°C, 14 000 rpm), the supernatants were removed and the protein extracts

were washed by incubation in 1.2 ml of 0.07% β-mercaptoethanol in acetone (1h, -20°C). This step was repeated twice. After the last washing, proteins were dried in a vacuum centrifuge and stored at -80°C until solubilization.

Dried protein pellets were solubilized in 100 µl of a solution containing 6 M of urea, 2 M of thiurea, 10 mM of dithiothreitol (DTT), 30 mM of TrisHCl pH 8.8 and 0.1% of Zwitterionic Acid Labile Surfactant I (ZALS I, Protea Bioscience, Morgantown, USA). Protein powders were mixed in the buffer using a metal spatula before vortexing the tubes for 3 min. Remaining cellular debris were segregated from soluble proteins by centrifugation (12,500 rpm, 25 min, room temperature). Protein concentrations were determined using the PlusOne 2-D Quant kit (GE Healthcare, Little Chalfont, UK) and adjusted to 4 $\mu g.\mu l^{-1}$ prior to digestion.

Digestion was performed in 0.2 ml strip tubes from 10 µl of diluted proteins. Proteins were incubated one hour at room temperature for reduction by the 10 mM DTT present in the buffer. Thereafter, proteins were alkylated one hour in 40 mM iodoacetamide (room temperature in the dark) and diluted with 50 mM ammonium bicarbonate to decrease total urea and thiourea concentration to 0.77 M. Overnight digestion was performed at 37°C with 1/50 (w/w) trypsin (Promega, Charbonnières-les-Bains, France) and stopped by acidification (1% total volume of trifluoroacetic acid, TFA). The resulting peptides were desalted on solid phase extraction using polymeric C18 columns (strata XL 100 µm, ref 8E-S043-TGB; Phenomenex, Le Pecq, France) as follows. Peptides were first diluted in 3% ACN and 0.06% acetic acid in water (washing buffer) up to a final volume of 500 µl. Then, they were loaded onto cartridges previously conditioned with 500 µl of ACN and rinsed three times with 500 µl of washing buffer. Peptides were rinsed three times with 500 µl of washing buffer and eluted twice by adding 300 µl of 40% ACN and 0.06% acetic acid. To finish, eluted peptides were speed-vac dried and suspended in a solution containing 2% ACN, 0.05% TFA and 0.05 % formic acid.

488

490 **LC-MS/MS analyses**

Samples were analyzed by LC-MS/MS by batches of 96. Analyses were performed using a
492 NanoLC-Ultra System (nano2DUltra, Eksigent, Les Ulis, France) connected to a Q-Exactive mass spectrometer (Thermo Electron, Waltham, MA, USA). A 400 ng of protein digest were loaded onto
494 a Biosphere C18 pre-column (0.3 × 5 mm, 100 Å, 5 µm; Nanoseparation, Nieuwkoop, Netherlands) at 7.5 µl.min^{-1} and desalted with 0.1% formic acid and 2% ACN. After 3 min, the precolumn was
496 connected to a Biosphere C18 nanocolumn (0.075 × 150 mm, 100 Å, 3 µm, Nanoseparation). Buffers were 0.1% formic acid in water (A) and 0.1% formic acid and 100% ACN (B). Peptides
498 were separated using a linear gradient from 5 to 35% buffer B for 40 min at 300 nl.min^{-1}. One run took 60 min, including the regeneration step at 95% buffer B and the equilibration step at 95%
500 buffer A.

Ionization was performed with a 1.4-kV spray voltage applied to an uncoated capillary
502 probe (10 µm tip inner diameter; New Objective, Woburn, MA, USA). Peptide ions were analyzed using Xcalibur 2.2 (Thermo Electron) with the following data-dependent acquisition steps: (1) MS
504 scan (mass-to-charge ratio (m/z) 400 to 1400, 70,000 resolution, profile mode), (2) MS/MS (17,500 resolution, collision energy = 27%, profile mode). Step 2 was repeated for the eight major ions
506 detected in step 1 with a charge of 2 or 3. Dynamic exclusion was set to 40 s. Xcalibur raw datafiles were transformed to mzXML open source format using msconvert software in the ProteoWizard
508 3.0.3706 package (Kessner et al. 2008). During conversion, MS and MS/MS data were centroided.

510 **Peptide and protein identification**

Peptide identifications were performed using the MaizeSequence genome database (Release
512 5a, 136,770 entries, https://ftp.maizegdb.org/MaizeGDB/FTP/) supplemented with 1821 FV2

sequences showing presence/absence variations (Darracq et al. 2018) and with a custom database containing standard contaminants. Database searches were performed using X!Tandem (Craig and Beavis 2004) (version 2015.04.01.1) with the following main parameters. Enzymatic cleavage was declared as a trypsin digestion with one possible misscleavage. Cystein carboxyamidomethylation and methionine oxidation were set to static and possible modifications, respectively. Precursor mass error was set to 10 ppm and fragment mass tolerance was set to 0.02 Da. In refine mode, a second search was performed with the same settings, except that protein N-ter acetylation was added as a potential modification and that the point mutations option was activated to detect possible single amino acid changes in the peptide sequences. Only peptides with an E-value smaller than 0.01 were reported.

Identified proteins were filtered and grouped using a homemade C++ version of X! TandemPipeline (Langella et al. 2017) especially designed to handle hundreds of MS run files. Only the proteins identified with a minimum of two peptides were considered as valid. Protein inference was performed using all samples together. The false discovery rate (FDR) was assessed from searches against a decoy database (using the reversed amino acid sequence for each protein) and was estimated at 0.03% for peptides.

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

Protein quantification

Protein quantification was performed from the peptide data obtained after ion chromatograms extraction and retention time (RT) alignment using MassChroQ software version 2.1.0 (Valot et al. 2011) with the following parameters: "ms2_1" alignment method,

tendency_halfwindow of 10, MS1 smoothing halfwindow of 0, MS2 smoothing halfwindow of 15, "quant1" quantification method, XIC extraction based on max, min and max ppm range of 10, anti-spike half of 5, background half median of 5, background half min max of 20, detection thresholds on min and max at 30 000 and 50 000, respectively, peak post-matching mode, "ni min abundance" of 0.1. Peptide data were filtered to remove the genotypes represented by only one or two samples instead of the expected four as well as outliers samples for which we suspected technical problems during sample preparation or MS analysis. At the end, the MS dataset included 977 samples. Proteins were then quantified using two methods.

XIC-based quantification. Proteins were quantified based on peptide intensity data processed as follows. We first removed the peptides ions showing standard deviations of retention time >15 s, which may arise from mis-identifications. Intensity normalization was subsequently performed to take into account possible global quantitative variations between LC-MS/MS runs. For this, we used a local normalization method described in Millan-Oropeza *et al.* (2017). We then removed the peptides shared between several proteins as well as the peptides for which both the unmodified form and a mass modification corresponding to an amino acid change was detected by the point mutation option of X!Tandem. These mutated peptide forms represent allelic versions of the sequences present in the searched protein database. In heterozygous genotypes, each allelic version produces its own MS signal, so that their measured intensities are not representative of the total protein abundance. We also removed the peptide ions presenting more than 10% missing values and those showing inconsistent intensity profiles. To this end, we computed Pearson correlations between log-intensities averaged across replicates for each pair of peptide ions belonging to the same protein. The peptide ion with the highest number of significant correlations (*P-value* < 0.01 after adjustment for multiple testing (Benjamini and Hochberg 1995)) was chosen as a reference for the protein. The peptide ions showing non-significant correlation to the reference (adjusted *P-value* >= 0.01) or whose coefficients of correlation to the reference were inferior to 0.3

were removed. To finish, we excluded the proteins quantified by only one peptide. As samples were grouped by batches of 96 analyzed by LC-MS/MS over a period of several months, we observed a strong batch effect on peptide normalized intensities. To correct this batch effect, we fitted a linear model to log-transformed intensity data, including only batches, and we subtracted the component due to the batch effects. Then, for each protein, we modeled the peptide data using the following mixed-effects model derived from Blein-Nicolas *et al.* (2012):

$$I'_{ijkl} = \mu + G_j + E_k + (GxE)_{jk} + R_{l(k)} + P_i + \theta_{jkl} + \varepsilon_{ijkl} \quad (1)$$

where I'_{ijkl} is the corrected, normalized log-intensity measured for peptide *i* in genotype *j*, watering condition *k* and replicate *l;*

μ is the mean intensity for a given protein;

G_j is the effect of the genotype *j;*

E_k is the effect of the watering condition *k;*

$(GxE)_{jk}$ is the effect of the genotype *j* x watering condition *k* interaction;

$R_{l(k)}$ is the effect of the replicate *l* nested in the watering condition *k;*

P_i is the effect of the peptide *i;*

$\Theta_{jkl} \sim \mathcal{N}(0, \sigma_\theta^2)$ is the random technical variation due to handling and injection in the mass spectrometer of the sample *jkl*;

$\varepsilon_{ijkl} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ is the residual error.

Model parameters were estimated by maximizing the restricted log-likelihood (REML method) and the differential protein abundance analysis was performed by analysis of variance (ANOVA). The resulting *P-value*s were adjusted for multiple testing by the Benjamini-Hochberg procedure (Benjamini and Hochberg 1995).

To subsequently perform GWAS at the protein level, we estimated protein abundances in each watering condition using the following model:

$$I'_{ijkl} = \mu_k + G_{jk} + R_{lk} + P_{ik} + \theta_{jkl} + \varepsilon_{ijkl} \quad (2)$$

where μ_k is the mean intensity obtained for a given protein in the watering condition *k*;

G_{jk} is the effect of the genotype *j* in the watering condition *k;*

R_{lk} is the effect of the replicate *l* in the watering condition *k;*

P_{ik} is the effect of the peptide *i* in the watering condition *k.*

Protein abundances were computed as adjusted means as follows:

$$A_{jk} = \mu_k + G_{jk} \quad (3)$$

Peak counting (PC)-based quantification. Proteins that could not be quantified based on XIC because their peptides showed to many missing intensity values were quantified based on their number of associated chromatographic peaks (*i.e.* quantified peptide ions). The shared peptides, the peptide ions showing variable RT and the peptides for which a mass modification corresponding to an amino acid change was detected were removed before computing protein abundances as peak numbers. The proteins for which the peak counts < 2 in any of the samples were deleted.

Normalization was then performed as follows:

$$Anorm_{ps} = \frac{A_{ps}}{\sum_{n=1}^{P} A_{ns}} x \frac{\sum_{m=1}^{S} \sum_{n=1}^{P} A_{nm}}{S} \quad (4)$$

where A_{ps} is the abundance of protein *p* in sample *s*;

P is the number of quantified proteins;

S is the number of samples.

As mentioned above, we corrected the batch effect by fitting a linear model to square-root transformed, normalized protein abundances, including only batches, and subtracting the component due to the batch effects. We then used the following model to detect protein abundance changes:

$$A'_{jkl} = \mu + G_j + E_k + (GxE)_{jk} + \alpha_l + \epsilon_{jkl} \quad (5)$$

where A'_{jkl} is the corrected, normalized, square-root transformed abundance obtained for a given protein in genotype *j*, watering condition *k* and replicate *l;*

μ is the mean abundance for the protein;

G_j is the effect of the genotype *j*;

E_k is the effect of the watering condition *k*;

$(GxE)_{jk}$ is the effect of the genotype *j* x watering condition *k* interaction;

$\alpha_l \sim \mathcal{N}(0, \sigma_\alpha^2)$ is the random effect of the replicate *l*;

$\varepsilon_{ijkl} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ is the residual error.

Estimation of the model parameters, differential analysis and *P-value* adjustment were performed as described above. Finally, for GWAS, we estimated the protein abundances separately in each watering condition with a mixed model derived from (5) and including only a fixed effect of the genotype and a random effect of the replicate. Protein abundances were computed as adjusted means as in (3).

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. The SNP effects were estimated by generalized least squares and their significance was tested with an F-statistic. A SNP was considered as significantly associated when -log10(*P-value*) > 5. A set of 961,971 SNPs obtained from lines genotyping using a 50 K Infinium HD Illumina array (Ganal et al. 2011), a 600 K Axiom Affymetrix array (Unterseer et al. 2014) and a set of 500 K SNPs obtained by genotyping by sequencing (Negro *et al.*, in print) was 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(*P-value*) and expected -log(*P-value*). Inflation factors were close to 1 (median of 1.08 and 1.05 in the XIC-based and PC-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 the significantly associated SNPs into pQTLs. *The genetic method:* two contiguous SNPs were considered as belonging to a same QTL when the genetic distance separating them was inferior to 0.1 cM. *The LD-based method:* two contiguous SNPs were considered as belonging to a same QTL when their LD-based windows (Negro *et al.*, in print) overlapped. *The geometric method:* for each chromosome, we ordered the SNPs according to their physical position. Then, we smoothed the -log10(*P-value*) signal by computing the maximum of the -log10(*P-values*) in a sliding window containing *N* consecutive SNPs. An association peak was detected when the smoothed -log10(*P-value*) signal exceeded a max threshold *M*. Two consecutive peaks were considered as two different QTLs when the *P-value* signal separating them went below a min 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 exhibiting the highest -log10(*P-value*). A pQTL was considered as local when located within 1 Mb upstream or downstream 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.

Broad sense heritability of protein abundance.

For each protein, the broad sense heritability of abundance was computed in each of the two watering conditions as follows. For the proteins quantified by the XIC-based approach, abundances were estimated in each sample as adjusted means from (1), by excluding the peptide effect P_{ik} and the random sample effect Θ_{jkl}. For the proteins quantified by the PC-based approach, the corrected, normalized, square-root transformed abundances were used. Protein abundances were then modeled as follows:

$$A_{jkl} = \mu_k + \beta_{jk} + \gamma_{kl} + \epsilon_{jkl} \quad (6)$$

where A_{ijk} is the abundance estimated for a given protein in the genotype *j*, the replicate *i* and the watering condition *k*

$\beta_{jk} \sim \mathcal{N}(0, \sigma_\beta^2)$ is the random effect of the genotype *j* in the watering condition *k*

$\gamma_{kl} \sim \mathcal{N}(0, \sigma_\gamma^2)$ is the random effect of the replicate *l* in the watering condition *k*.

Heritability was subsequently computed as

$$H^2 = \frac{\sigma_\beta^2}{\left(\sigma_\beta^2 + \sigma_\gamma^2 / N\right)} \quad (7)$$

where *N* is the number of replicates.

Protein co-expression analysis

Protein co-expression analysis was performed using the WGCNA R package (Langfelder and Horvath 2008). Using a procedure developed to correct the LD by the structure and/or the relatedness and implemented in the LDcorSV R package (Mangin et al. 2012), we computed pair-wise Pearson's correlations corrected by structure and kinship ($|r_{corrected}|$) and used them as input similarity matrix. The sotfpower parameter was set at 2. Adjacency and topological ovelap matrices were both unsigned. Protein modules were constituted with a minimum module size set at 20 and a control over sensitivity splitting set at 4. The minimum height for merging modules was set at 0.25. The other parameters were left at default values. Graphical representations of the resulting networks

were performed with Cytoscape (Shannon et al. 2003) v3.5.1 using an unweighted spring embedded layout. Module eigengenes were computed as described in the WGCNA R package.

QTL co-localization

Co-localizations between QTLs were detected when they fulfilled two criteria. First, the LD-based windows around the QTLs (Negro *et al.*, in print) should overlap. Second, the absolute value of the Pearson's correlation of coefficient corrected by structure and kinship (the $|r_{corrected}|$ mentioned above) between the values of the phenotypes controlled by the QTLs should be superior to 0.3. We determined this value empirically, in absence of a statistical test to test the significance of the corrected correlation.

Candidate genes identification

For each QTL/pQTL co-localization, gene accession present in 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 deposited online using PROTICdb (Langella et al. 2007; Ferry-Dumazet et al. 2005; Langella et al. 2013) at the following URL: http://moulon.inra.fr/protic/amaizing. They are currently available with the following username: reviewer and password: reviewer. They will be made freely available after publication. Detailed information on the peptides and proteins identified in all 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 the following URL: http://www.phis.inra.fr/openphis/web/index.php?r=project%2Fview&id=Systems+genetics+for+maize+drought+tolerance+%28Amaizing+project%29. Leaf area early was defined at the seven leaves stage, representing 24 $d_{20°C}$ (thermal time in equivalent days at 20°C). Leaf area late was defined at the 12 leaves stage, representing 45 $d_{20°C}$.

Genotyping data are available at the following URL: https://data.inra.fr/privateurl.xhtml?token=a0dbedde-249e-4808-9f51-51e2004678f7.

ACKNOWLEDGMENTS

This work was supported by the Agence Nationale de la Recherche project ANT-10-BTBR-01 (Amaizing). Proteomics analyses were performed on the PAPPSO platform (http://pappso.inra.fr) which is supported by INRA (http://www.inra.fr), the Ile-de-France regional council (https://www.iledefrance.fr/education-recherche), IBiSA (https://www.ibisa.net) and SPS (https://www6.inra.fr/saclay-plant-sciences). The authors want to thank Sylvie Coursol for her critical reviewing of the manuscript, Hélène Corti for her help in sample preparation and Olivier Langella for having specially developed a pipeline to upload the proteomics data on ProticDB. They are also grateful to people from INRA LEPSE: François Tardieu for its participation to the coordination of the plant experiment; Benoît Suard, Pauline Sidawi and Olivier Martin for their technical assistance during the experiment; Santiago Alvarez Prado for his contribution to plant traits and QTL analysis.

AUTHORS CONTRIBUTION

AC and MZ designed research; CW designed and coordinated the plant experiment in PhenoArch and the genetic analysis of plant traits; LBC performed the plant experiment and analyzed image-based phenotypic data; MBN and TB performed the proteomics experiments; SSN developed the GWAS pipeline and performed the quality control on the genotyping; SDN performed the genotyping and estimated local LD; MBN and MZ analyzed the proteomics data, MBN performed the systems genetics study and wrote the manuscript. All authors discussed the results and read and approved the final manuscript.

DISCLOSURE DECLARATION

The authors declare no competing interest.

738 **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.

Albertin W, Marullo P, Bely M, Aigle M, Bourgais 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.

Balliau T, Blein-Nicolas M, Zivy M. 2018. Evaluation of optimized tube-gel methods of sample preparation for large-scale plant proteomics. *Proteomes* **6**.

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.

Battle A, Khan Z, Wang SH, Mitrano A, Ford MJ, Pritchard JK, Gilad Y. 2015. Impact of regulatory variation from RNA to protein. *Science* **347**: 664–667.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol* **57**: 289–300.

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.

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.

Bourgeois M, Jacquin F, Cassecuelle F, Savois V, Belghazi M, Aubert G, Quillien L, Huart M, Marget P, Burstin J. 2011. A PQL (protein quantity loci) analysis of mature pea seed proteins identifies loci determining seed protein composition. *Proteomics* **11**: 1581–1594.

Cabrera-Bosquet L, Fournier C, Brichet N, Welcker C, Suard B, Tardieu F. 2016. High-throughput estimation of incident light, light interception and radiation-use efficiency of thousands of plants in a phenotyping platform. *New Phytol* **212**: 269–281.

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.

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.

Chaves MM, Maroco JP, Pereira JS. 2003. Understanding plant responses to drought — from genes to the whole plant. *Funct Plant Biol* **30**: 239–264.

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.

Christie N, Myburg AA, Joubert F, Murray SL, Carstens M, Lin Y-C, 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.

Civelek M, Lusis AJ. 2014. Systems genetics approaches to understand complex traits. *Nat Rev Genet* **15**: 34–48.

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.

Craig R, Beavis RC. 2004. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **20**: 1466–1467.

Darracq A, Vitte C, Nicolas S, Duarte J, Pichon J-P, Mary-Huard T, Chevalier C, Bérard A, Le Paslier M-C, 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**.

Daryanto S, Wang L, Jacinthe P-A. 2016. Global synthesis of drought effects on maize and wheat production. *PLoS ONE* **11**.

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.

Feltus FA. 2014. Systems genetics: a paradigm to improve discovery of candidate genes and mechanisms underlying complex traits. *Plant Sci* **223**: 45–48.

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.

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.

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.

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.

Ganal MW, Durstewitz G, Polley A, Bérard A, Buckler ES, Charcosset A, Clarke JD, Graner E-M, 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.

Ghazalpour A, Bennett B, Petyuk VA, Orozco L, Hagopian R, Mungrue IN, Farber CR, Sinsheimer J, Kang HM, Furlotte N, et al. 2011. Comparative analysis of proteome and transcriptome variation in mouse. *PLoS Genet* **7**: e1001393.

Goldman AD, Samudrala R, Baross JA. 2010. The evolution and functional repertoire of translation proteins following the origin of life. *Biol Direct* **5**: 15.

Golldack D, Li C, Mohan H, Probst N. 2014. Tolerance to drought and salt stress in plants: Unraveling the signaling networks. *Front Plant Sci* **5**.

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.

Johnson MR, Behmoaras J, Bottolo L, Krishnan ML, Pernhorst K, Santoscoy PLM, Rossetti T, Speed D, Srivastava PK, Chadeau-Hyam M, et al. 2015. Systems genetics identifies Sestrin 3 as a regulator of a proconvulsant gene network in human epileptic hippocampus. *Nat Commun* **6**: 6031.

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.

Kessner D, Chambers M, Burke R, Agus D, Mallick P. 2008. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* **24**: 2534–2536.

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.

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.

Langella O, Zivy M, Joets J. 2007. The PROTICdb database for 2-DE proteomics. *Methods Mol Biol* **355**: 279–303.

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

Li S-B, Xie Z-Z, Hu C-G, Zhang J-Z. 2016. A review of auxin response factors (ARFs) in plants. *Front Plant Sci* **7**.

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.

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.

Mähler N, Wang J, Terebieniec BK, Ingvarsson PK, Street NR, Hvidsten TR. 2017. Gene co-expression network connectivity is an important determinant of selective constraint. *PLoS Genet* **13**: e1006402.

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* **108**: 285–291.

Mao H-D, Yu L-J, Li Z-J, 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.

Markowetz F, Boutros M. 2015. An introduction to systems genetics. In *Systems Genetics* (eds. F. Markowetz and M. Boutros), pp. 1–11, Cambridge University Press, 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.

Millan-Oropeza A, Henry C, Blein-Nicolas M, Aubert-Frambourg A, Moussa F, Bleton J, Virolle M-J. 2017. Quantitative proteomics analysis confirmed oxidative metabolism predominates in *Streptomyces coelicolor* versus glycolytic metabolism in *Streptomyces lividans*. *J Proteome Res* **16**: 2597–2613.

Millet EJ, Welcker C, Kruijer W, Negro S, Coupel-Ledru A, Nicolas SD, Laborde J, Bauland C, Praud S, Ranc N, et al. 2016. Genome-wide analysis of yield in Europe: Allelic effects vary with drought and heat scenarios. *Plant Physiol* **172**: 749–764.

Mizrachi E, Verbeke L, Christie N, Fierro AC, Mansfield SD, Davis MF, Gjersing E, Tuskan GA, Montagu MV, Peer YV de, 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.

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.

Munkvold JD, Laudencia-Chingcuanco D, Sorrells ME. 2013. Systems genetics of environmental response in the mature wheat embryo. *Genetics* **194**: 265–277.

Nadeau JH, Dudley AM. 2011. Systems genetics. *Science* **331**: 1015–1016.

Nelson N, Junge W. 2015. Structure and energy transfer in photosystems of oxygenic photosynthesis. *Annu Rev Biochem* **84**: 659–683.

Neveu P, Tireau A, Hilgert N, Nègre V, Mineau-Cesari J, Brichet 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.

Ogura T, Busch W. 2016. Genotypes, networks, phenotypes: Moving toward plant systems genetics. *Annu Rev Cell Dev Biol* **32**: 103–126.

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.

Osakabe Y, Osakabe K, Shinozaki K, Tran L-SP. 2014. Response of plants to water stress. *Front Plant Sci* **5**.

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.

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.

Preston JC, Hileman LC. 2013. Functional Evolution in the Plant SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE (SPL) Gene Family. *Front Plant Sci* **4**: 80.

R core team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rincent 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.

Seki M, Umezawa T, Urano K, Shinozaki K. 2007. Regulatory metabolic networks in drought stress responses. *Curr Opin Plant Biol* **10**: 296–302.

Shang Y, Yan L, Liu Z-Q, Cao Z, Mei C, Xin Q, Wu F-Q, Wang X-F, Du S-Y, Jiang T, et al. 2010. The Mg-chelatase H subunit of Arabidopsis antagonizes a group of WRKY transcription repressors to relieve ABA-responsive genes of inhibition. *Plant Cell* **22**: 1909–1935.

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.

Shiferaw B, Prasanna BM, Hellin J, Bänziger M. 2011. Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security. *Food Secur* **3**: 307.

Song K, Kim HC, Shin S, Kim K-H, Moon J-C, Kim JY, Lee B-M. 2017. Transcriptome analysis of flowering time genes under drought stress in maize leaves. *Front Plant Sci* **8**.

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.

Thimm O, Bläsing 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.

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.

Unterseer S, Bauer E, Haberer G, Seidel 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.

Usadel B, Poree 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.

Valliyodan B, Nguyen HT. 2006. Understanding regulatory networks and engineering for enhanced drought tolerance in plants. *Curr Opin Plant Biol* **9**: 189–195.

Valot B, Langella O, Nano E, Zivy M. 2011. MassChroQ : A versatile tool for mass spectrometry quantification. *Proteomics* **11**: 3572–3577.

van der Sijde MR, Ng A, Fu J. 2014. Systems genetics: From GWAS to disease pathways. *Biochim Biophys Acta* **1842**: 1903–1909.

Vogel C, Marcotte EM. 2012. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* **13**: 227–232.

Wasinger VC, Zeng M, Yau Y. 2013. Current Status and Advances in Quantitative Proteomic Mass Spectrometry. *Int J Proteomics* 2013:180605

Williams EG, Wu Y, Jha P, Dubuis S, Blattmann P, Argmann CA, Houten SM, Amariuta T, Wolski W, Zamboni N, et al. 2016. Systems proteomics of liver mitochondria function. *Science* **352**: aad0189.

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.

Zhang J, Yang J-R. 2015. Determinants of the rate of protein sequence evolution. *Nat Rev Genet* **16**: 409–420.

Zhang X, Liu X, Zhang D, Tang H, Sun B, Li C, Hao L, Liu C, Li Y, Shi Y, et al. 2017. Genome-wide identification of gene expression in contrasting maize inbred lines under field drought conditions reveals the significance of transcription factors in drought tolerance. *PLoS ONE* **12**: e0179477.

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.

1.

FIGURE LEGENDS

Figure 1. Effect of a mild water deficit on the proteome. (A) Heatmap representations of the abundances estimated for the XIC-based protein set (left) and the PC-based protein set (right). Each line corresponds to a protein and each column to a genotype x watering condition combination. For each protein, abundance values were scaled and represented by a color code as indicated by the color-key bar. Hierarchical clusterings of the genotype x watering condition combinations (top) and of the proteins (left) were built by using 1-the Pearson correlation coefficient as distance and the unweighted pair group method with arithmetic mean (UPGMA) as aggregation method. (B) Functions of the 200 most induced and 200 most repressed proteins under water deficit. (C) Abundance profiles of the RAB17 dehydrin (GRMZM2G079440 quantified based on the number of chromatographic peaks) and of a LEA protein (GRMZM2G352415 quantified based on peptide intensities) in the two watering conditions. Genotypes on the x axis were ordered according to the WD/WW abundance ratio.

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

Figure 3. Distribution of pQTLs along the genome. (A) In the well-watered condition. (B) In the water-deficit condition. Each point indicates the number of proteins controlled by a pQTL located in a given genomic region defined by the linkage desequilibrium interval around a SNP. Dashed horizontal lines indicate the arbitrary threshold used to detect pQTL hotspots. Names and positions of the pQTL hotspots are indicated above each graph. Names in bold indicate the pQTL hotspots confidently detected as potential pleiotropic loci (see Sup. Table 3 for details).

764 **Figure 4. Graphical representation of the co-expression networks resulting from the WGCNA analyses.** Only proteins showing adjacencies > 0.02 are shown. The consensus network contains the
766 proteins that were co-expressed in the two watering conditions. The three views were created by Cytoscape v3.5.1 using an unweighted, spring-embedded layout. The colors displayed on each
768 network represent the different modules identified by WGCNA. Functional enrichments of modules are indicated in grey boxes. Condition-specific modules are surrounded by dashed circles.

770

Figure 5. Genomic positions of the co-localizing pQTLs, coQTLs and QTLs. The positions of
772 the nine pQTL hotspots robustly identified as potential loci with pleiotropic effects are indicated as well as the position of the most promising candidate genes. Chromosomes are segmented in 1 Mb
774 bins. Grey dots represent the centromeres and blue dots indicate the position of genomic regions showing evidences for pleiotropy both at the proteome and phenotype level. Blue lines indicate
776 pQTLs, coQTLs and QTLs that are determined by a same SNP.
* consensus module, ° WD-specific module

778

Figure 6. Identification of genomic regions involved in multi-scale genetic control.
780 (A) Distribution of the distances between co-localizing QTLs and pQTLs. (B) Detailed view of the QTL, pQTL, coQTL detected in the region covered by the hotspot Hs52d on chromosome 5. Dots
782 represent the SNPs determining the position of the QTLs and horizontal bars represent the linkage disequilibrium-based window around each SNP. Black circled dots indicate the pQTLs that co-
784 localize with QTLs or coQTLs with high correlations between the protein abundance and the phenotypic trait value or the module eigengene. Vertical dashed lines indicate the position of SNPs
786 S5_88793314 (on the left) and AX-91658235 (on the right). The position of two transcription factors (a SBP gene, GRMZM2G111136, and a C2C2-CO-like transcription factor,
788 GRMZM2G148772) representing promising candidate genes are indicated.

TABLES

Table 1. Proteins associated to pQTLs co-localizing with QTLs.

Protein ID	Gene accession	Plaza 4.0 annotation[a]	# of QTL/pQTL co-localizations	Phenotypic trait
a3.a1	GRMZM2G306345	pyruvate orthophosphate dikinase1	1[c]	BI
c111.a2	GRMZM2G085054	Glycosyltransferase benzoxazinone synthesis 8	1	WU
c135.a1	GRMZM2G085577	Alpha-1,4 glucan phosphorylase	1	LAI
c177.a1	GRMZM2G027875	Aminopeptidase M1	1	LAI
c236.a1[b]	GRMZM2G130230	Glucose-6-phosphate dehydrogenase	1	LAI
c263.a1	GRMZM2G162486	**Glutathione S-transferase**	1	LAI
c264.a1	GRMZM2G061969	**Phospholipase D**	1	LAI
c291.a1	GRMZM2G064799	Succinate dehydrogenase 1	1	LAI
c317.a1	GRMZM2G134256	Transaldolase 2	1	LAI
c364.a1[b]	GRMZM2G075624 GRMZM2G108474	Translationally controlled tumor protein	1	gs
c513.a2	GRMZM2G134668	Calnexin	1	WU
c517.a1	AC198418.3	RNA helicase4	1	LAI
c547.a1	GRMZM2G033641	Patellin-1 (SEC14-like protein)	1	LAI
c561.a1[b]	GRMZM2G048085	**Senescence-associated protein DIN1**	1	Trate
c587.a1[b]	GRMZM2G051677	Fructokinase-2	1	LAI
c607.a1[b]	GRMZM2G352415	**Late embryogenesis abundant (LEA) protein**	1	LAI
c739.a1[b]	GRMZM2G085967	**Peroxidase**	1	LAI
c790.a1	GRMZM2G332976	short chain alcohol dehydrogenase1	1[c]	BI
c914.a2	GRMZM2G373522	**Dehydrin**	1	WU
c959.a2[b]	GRMZM2G479423	Aldose reductase	1	LAI
c976.a1	GRMZM2G064360	**Basic endochitinase 1**	1	LAI
d1139.a1[b]	GRMZM2G169207	IMP dehydrogenase	1	LAI
d1261.a1	GRMZM2G026800	Probable plastid lipid-associated protein 10	1	BI
d1513.a1	GRMZM2G094712	Aspartate aminotransferase	1	BI
d1624.a1	GRMZM2G022563	oxoacyl-[acyl-carrier-protein] synthase 3	1	LAI
b33.a2	GRMZM2G112165	**Heat shock protein 90-2**	2	Lal, WU
c178.a1	GRMZM2G055489	Sucrose-phosphatase 1 (ZmSPP1)	2	gs, Trate
c395.a1	GRMZM2G165901	**Glycine-rich RNA-binding, ABA-inducible protein**	2	BI, WU
c490.a1	GRMZM2G038494	Obg-like ATPase 1, GTP-binding protein-related	2	LAI
c997.a1	GRMZM2G051943	**Endochitinase A**	2	Lal, WU
d1066.a1	GRMZM2G704005	**Lactoylglutathione lyase / glyoxalase I family protein**	2	LAI
d1218.a1[b]	GRMZM2G169516	Indole-3-glycerol phosphate synthase	2	LAI
c144.a1[b]	GRMZM2G120304	hydroxyproline-rich glycoprotein family protein	3	Lal, WU
c654.a1	GRMZM2G110567	Protein binding, zinc finger family protein	3	LAI
c778.a1	GRMZM2G017110	**Glutamate decarboxylase**	3	WU, WUE
d1404.a1	GRMZM2G080724	**25.3 kDa heat shock protein**	3	BI, Lal, WU
d1140.a1[b]	GRMZM2G073079 GRMZM2G076348	alpha/beta-Hydrolases superfamily protein	4	Lal, WU
d1512.a1[b]	GRMZM5G815098	Unknow	4	LAI
b94.a2	GRMZM5G851266	Putative polyphenol oxidase family protein	5	BI, Lal, WU
c115.a2	GRMZM5G813217	**Heat shock protein 90-5**	6	BI, Lal, WU
d1161.a1	GRMZM2G079440	**Dehydrin DHN1 (RAB-17 protein)**	7	BI, gs, Trate, WU
c880.a1	GRMZM2G176998	Putative WD40-like beta propeller repeat family protein	8	BI, Lal, WU

a Functions in bold are related to drought- or stress-response.
b Protein ID corresponding to different isofoms or accessions
c QTL/pQTL co-localization observed in the WW condition.

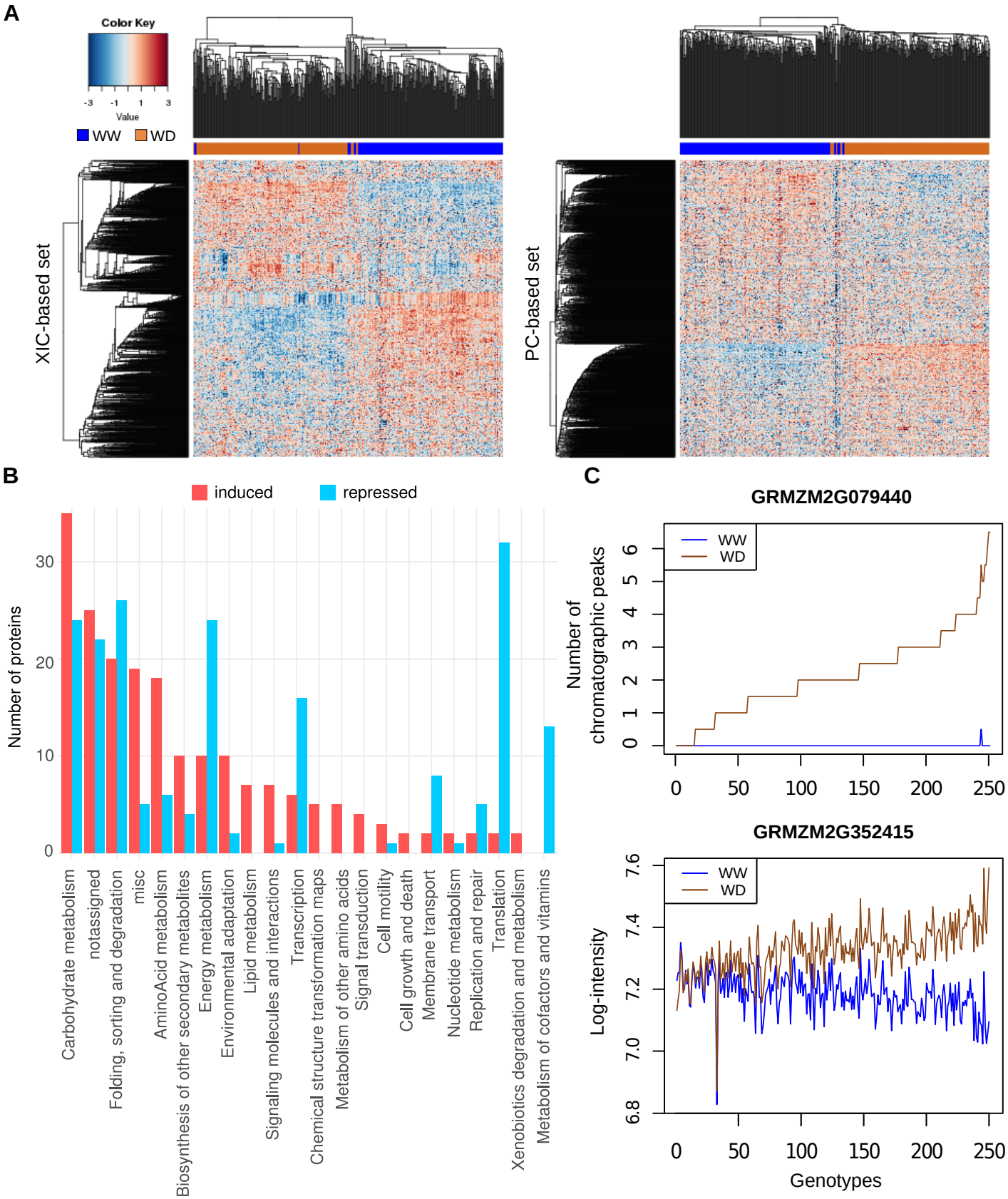

Figure 1. Effect of a mild water deficit on the proteome. (A) Heatmap representations of the abundances estimated for the XIC-based protein set (left) and the PC-based protein set (right). Each line corresponds to a protein and each column to a genotype x watering condition combination. For each protein, abundance values were scaled and represented by a color code as indicated by the color-key bar. Hierarchical clusterings of the genotype x watering condition combinations (top) and of the proteins (left) were built by using the 1-Pearson correlation coefficient as distance and the unweighted pair group method with arithmetic mean (UPGMA) as aggregation method. (B) Functions of the 200 most induced and 200 most repressed proteins under water deficit. (C) Abundance profiles of the RAB17 dehydrin (GRMZM2G079440 quantified based on the number of chromatographic peaks) and of a LEA protein (GRMZM2G352415 quantified based on peptide intensities) in the two watering conditions. Genotypes on the x axis were ordered according to the WD/WW abundance ratio.

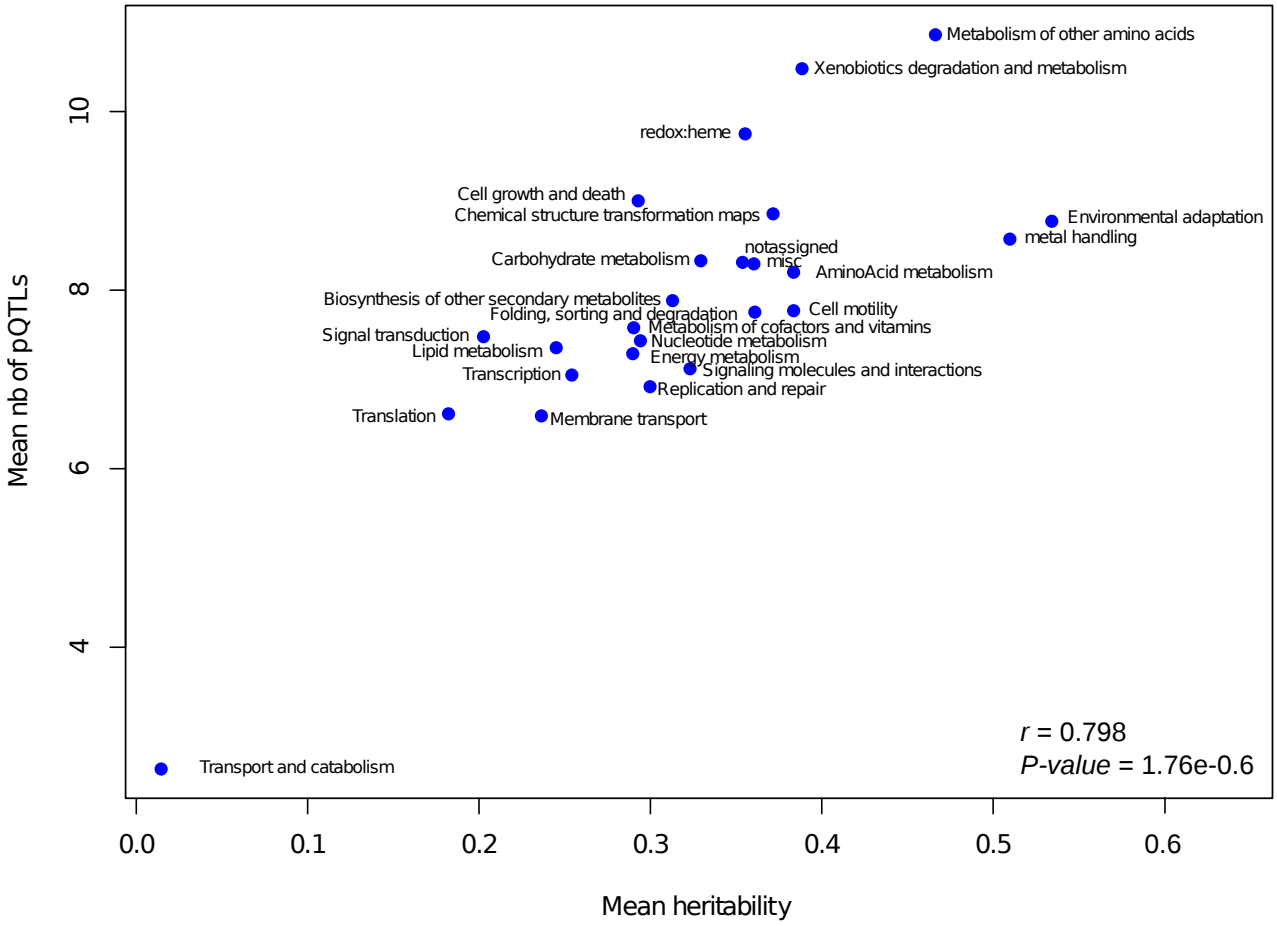

Figure 2. Relationship between the mean number of pQTLs per KEGG category and the mean heritability per KEGG category.

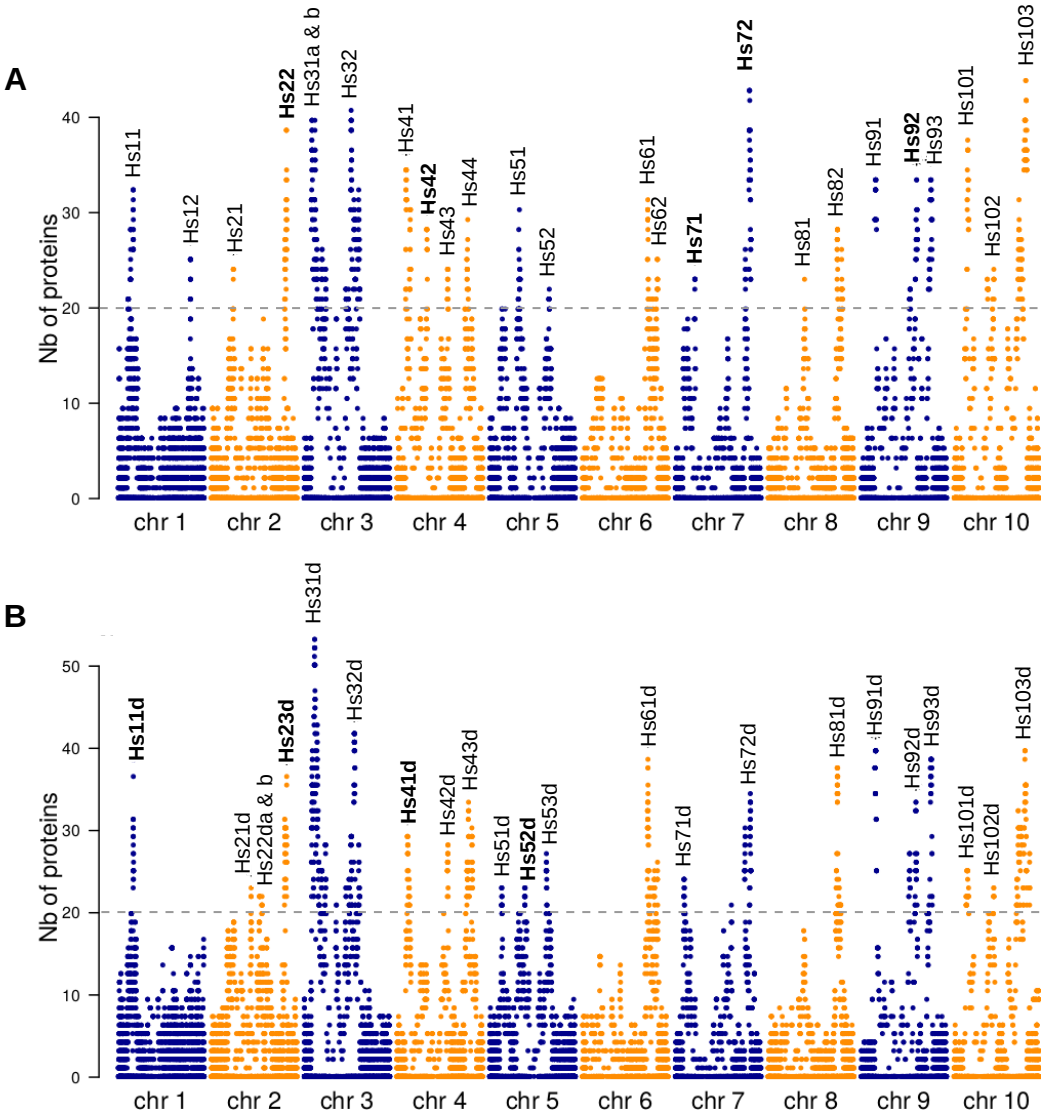

Figure 3. Distribution of pQTLs along the genome. (A) In the well-watered condition. (B) In the water deficit condition. Each point indicates the number of proteins controlled by a pQTL located in a given genomic region defined by the linkage disequilibrium interval around a SNP. Dashed horizontal lines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

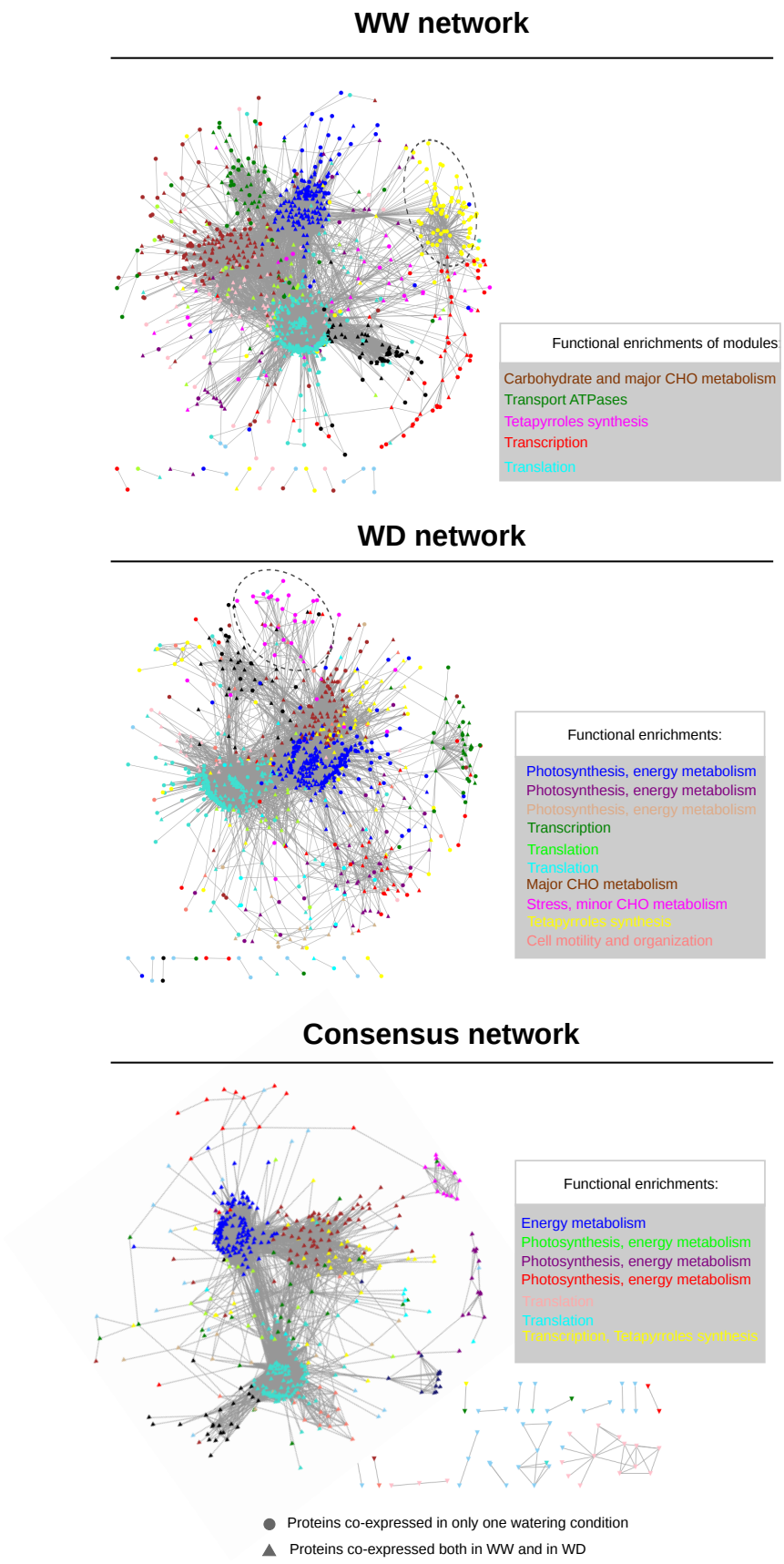

Figure 4. Graphical representation of the co-expression networks resulting from the WGCNA analyses. Only proteins showing adjacencies > 0.02 are shown. The consensus network contains the proteins that were co-expressed in the two watering conditions. The three views were created by Cytoscape v3.5.1 using an unweighted, spring-embedded layout. The colors displayed on each network represent the different modules identified by WGCNA. Functional enrichments of modules are indicated in grey boxes. Condition-specific modules are surrounded by dashed circles.

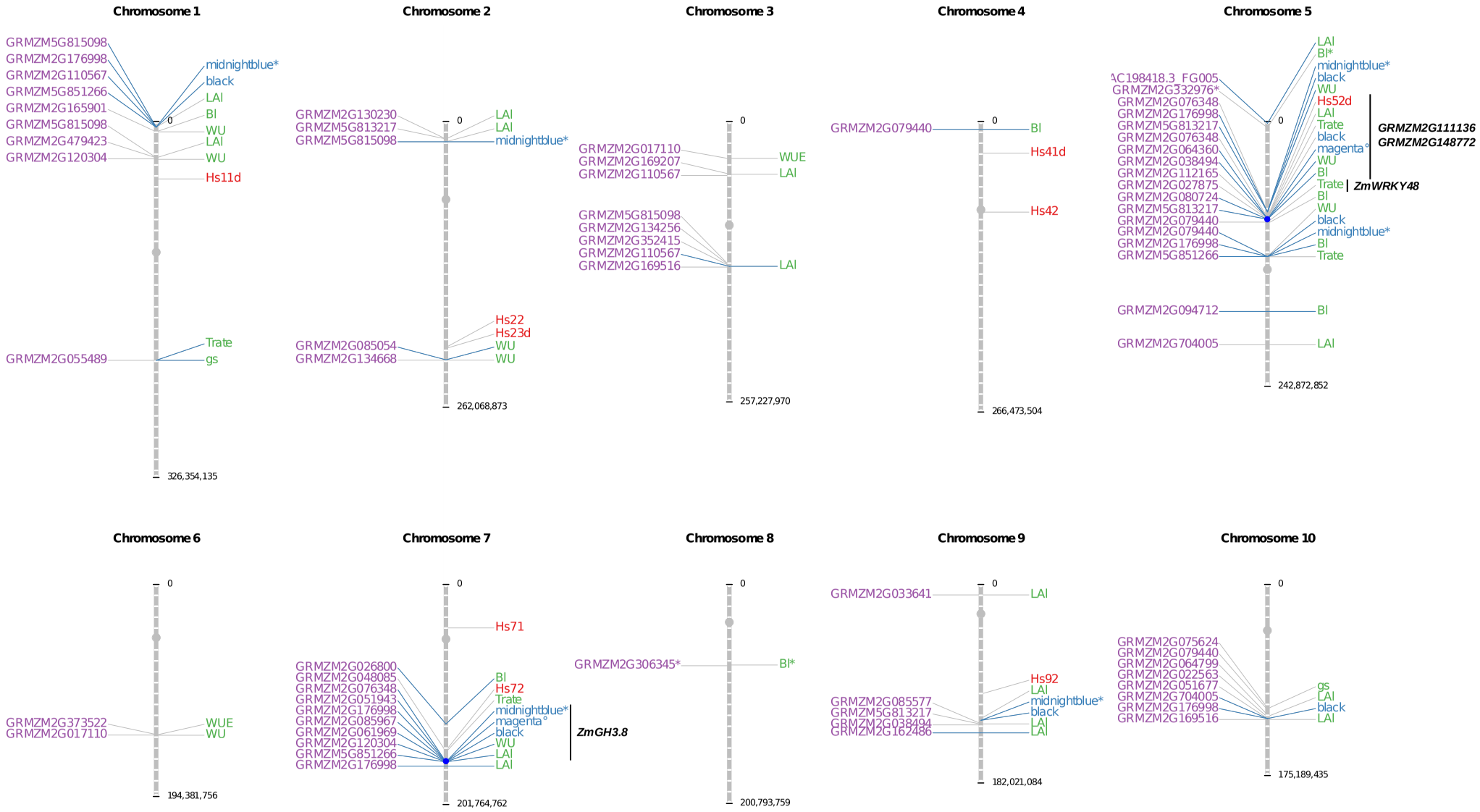

Figure 5. Genomic positions of the co-localizing pQTLs, coQTLs and QTLs. The positions of the nine pQTL hotspots robustly identified as potential loci with pleiotropic effects are indicated as well as the position of the most promising candidate genes. Chromosomes are segmented in 1 Mb bins. Grey dots represent the centromeres and blue dots indicate the position of genomic regions showing evidences for pleiotropy both at the proteome and phenotype level. Blue lines indicate pQTLs, coQTLs and QTLs that are determined by a same SNP.
* consensus module, ° WD-specific module

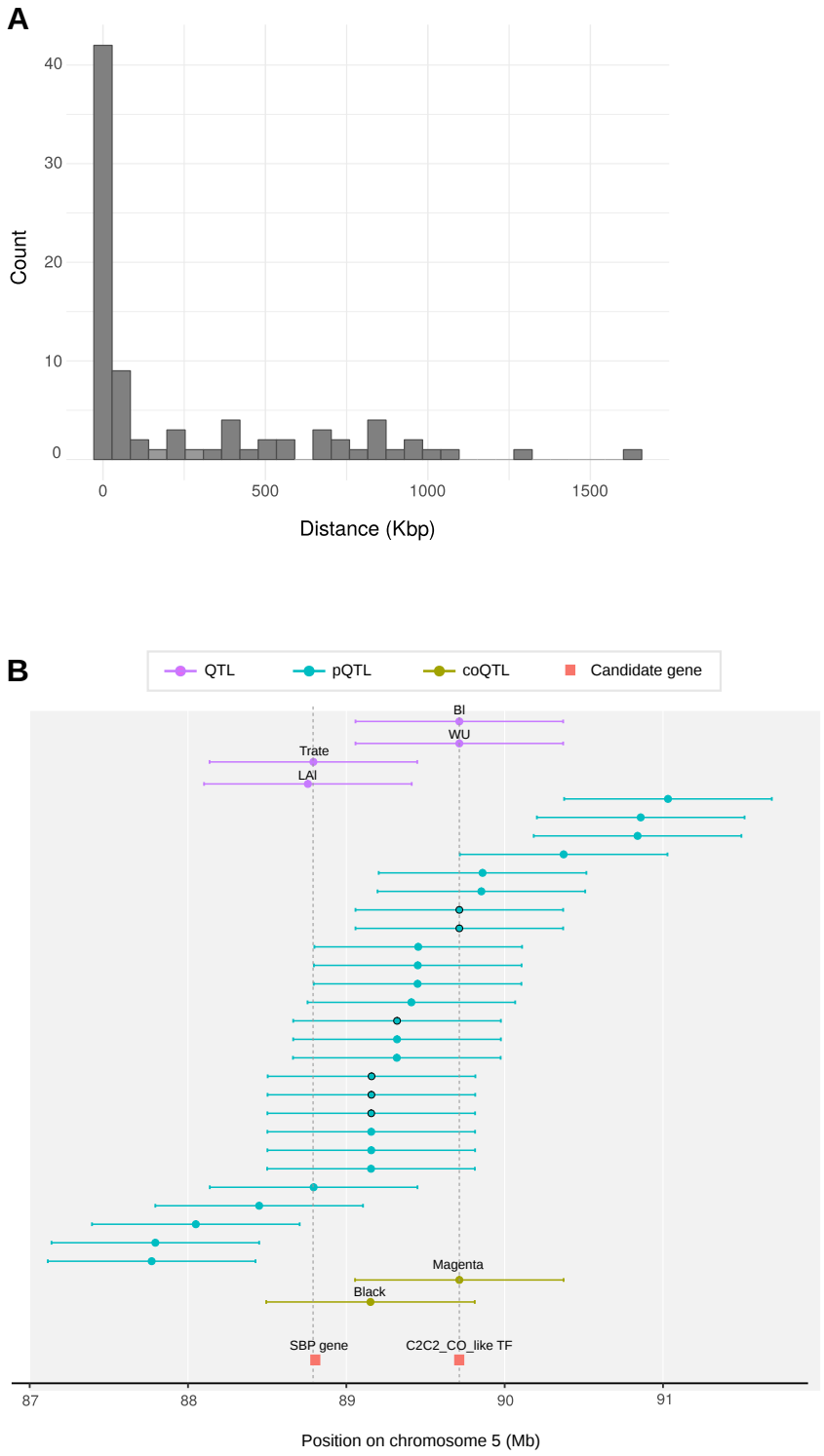

Figure 6. Identification of genomic regions involved in multi-scale genetic control.
(A) Distribution of the distances between co-localizing QTLs and pQTLs. (B) Detailed view of the QTL, pQTL, coQTL detected in the region covered by the hotspot Hs52d on chromosome 5. Dots represent the SNPs determining the position of the QTLs and horizontal bars represent the linkage disequilibrium-based window around each SNP. Black circled dots indicate the pQTLs that co-localize with QTLs or coQTLs with high correlations between the protein abundance and the phenotypic trait value or the module eigengene. Vertical dashed lines indicate the position of SNPs S5_88793314 (on the left) and AX-91658235 (on the right). The position of two transcription factors (a SBP gene, GRMZM2G111136, and a C2C2-CO-like transcription factor, GRMZM2G148772) representing promising candidate genes are indicated.