Title: Drought-related secondary metabolites of barley (Hordeum vulgare L.) leaves and their metabolomic quantitative trait loci

Author: Anna Piasecka, Aneta Sawikowska, Anetta Kuczyńska, Piotr Ogrodowicz, Krzysztof Mikołajczak, Karolina Krystkowiak, Kornelia Gudyś, Justyna Guzy-Wróbelska

Citation style: Piasecka Anna, Sawikowska Aneta, Kuczyńska Anetta, Piotr Ogrodowicz, Mikołajczak Krzysztof, Krystkowiak Karolina, Gudyś Kornelia, Guzy-Wróbelska Justyna. (2017). Drought-related secondary metabolites of barley (Hordeum vulgare L.) leaves and their metabolomic quantitative trait loci. "Plant Journal" (Vol. 89, iss. 5 (2017), s. 898-913), doi 10.1111/tpj.13430

the plant journal

SEB Society for Experimental Biology

The Plant Journal (2017) **89,** 898–913 doi: 10.1111/tpj.13430

Drought-related secondary metabolites of barley (*Hordeum vulgare* L.) leaves and their metabolomic quantitative trait loci

Anna Piasecka[1,†], Aneta Sawikowska[1,†,‡], Anetta Kuczyńska[1], Piotr Ogrodowicz[1], Krzysztof Mikołajczak[1], Karolina Krystkowiak[1], Kornelia Gudyś[2,§], Justyna Guzy-Wróbelska[2], Paweł Krajewski[1] and Piotr Kachlicki[1,*]

[1]*Institute of Plant Genetics, Polish Academy of Sciences, Strzeszyńska 34 60-479, Poznań, Poland, and*

[2]*Department of Genetics, Faculty of Biology and Environmental Protection, University of Silesia, Jagiellońska 28 40-032, Katowice, Poland*

Received 6 May 2016; revised 9 November 2016; accepted 17 November 2016; published online 23 November 2016.

*For correspondence (e-mail pkac@igr.poznan.pl).

†These authors made an equal contribution.

‡Department of Mathematical and Statistical Methods, Poznań University of Life Sciences, Wojska Polskiego 28, Poznań, Poland.

§Department of Botany and Nature Protection, Faculty of Biology and Environmental Protection, University of Silesia, Jagiellońska 28 40-032, Katowice, Poland

SUMMARY

Determining the role of plant secondary metabolites in stress conditions is problematic due to the diversity of their structures and the complexity of their interdependence with different biological pathways. Correlation of metabolomic data with the genetic background provides essential information about the features of metabolites. LC-MS analysis of leaf metabolites from 100 barley recombinant inbred lines (RILs) revealed that 98 traits among 135 detected phenolic and terpenoid compounds significantly changed their level as a result of drought stress. Metabolites with similar patterns of change were grouped in modules, revealing differences among RILs and parental varieties at early and late stages of drought. The most significant changes in stress were observed for ferulic and sinapic acid derivatives as well as acylated glycosides of flavones. The tendency to accumulate methylated compounds was a major phenomenon in this set of samples. In addition, the polyamine derivatives hordatines as well as terpenoid blumenol C derivatives were observed to be drought related. The correlation of drought-related compounds with molecular marker polymorphisms resulted in the definition of metabolomic quantitative trait loci in the genomic regions of single-nucleotide polymorphism 3101-111 and simple sequence repeat Bmag0692 with multiple linkages to metabolites. The associations pointed to genes related to the defence response and response to cold, heat and oxidative stress, but not to genes related to biosynthesis of the compounds. We postulate that the significant metabolites have a role as antioxidants, regulators of gene expression and modulators of protein function in barley during drought.

Keywords: secondary metabolites, *Hordeum vulgare*, metabolomic quantitative trait loci, drought response, recombinant inbred lines.

INTRODUCTION

Many physiological functions of plants, including their response to environmental stresses, depend to a large extent on secondary metabolites from several structural classes (Cheynier *et al.*, 2013). However, assigning a common function to the entire group of compounds is not possible, since particular metabolites have different physicochemical properties depending on their chemical structure (Mouradov and Spangenberg, 2014). The large number of different structures arise from the various combinations of multiple hydroxyl groups, methyl groups, glycosides and acyl substituents on the basic C6–C3 and C6–C3–C6 skeletons in phenylpropanoids and flavonoids, respectively (Cheynier *et al.*, 2013). These modifications affect physicochemical properties such as the stability, solubility and biological activity of substituted compounds (Plaza *et al.*, 2014).

Phenolic compounds are a significant group of non-enzymatic antioxidants involved in the response to and tolerance of moisture deficit (Nakabayashi *et al.*, 2014). Their high antioxidant activity depends on structural features

such as the *ortho* position of hydroxyl groups in the B ring and, to a lesser extent, in the A ring (Arora *et al.*, 1998). The presence of a larger number of OH groups increases the antioxidant properties of these compounds. On the other hand, their methylation or glycosylation has the opposite effect. Antioxidative phenolic compounds can protect the enzymatic complexes of the cytoplasm and chloroplasts in mesophyll and epidermal cells from the damaging effects of free radicals (Agati *et al.*, 2007). Phenolic compounds found in vacuoles are substrates for peroxidase in the hydrogen peroxide inactivation reaction (Agati *et al.*, 2007).

Phenolic compounds can also act as plant-protective filters against UV-A and UV-B (Cheynier *et al.*, 2013). In moisture deficit, cell wall-bound phenolic compounds absorb most of the high-energy UV radiation and prevent photoinhibitory injuries to the photosynthetic apparatus (Hura *et al.*, 2012). In addition, phenolic compounds can transmit or absorb energy from other molecules, acting as energy dissipaters and outlets for photo-assimilates if there is energetic imbalance under stress conditions (Hernandez and Van Breusegem, 2010).

The literature on phenolic compounds in barley (*Hordeum vulgare* L.) focuses on the structural analysis of selected metabolites with similar physicochemical properties (Ferreres *et al.*, 2007; Kohyama and Ono, 2013). Recently, 152 phenolic secondary metabolites were characterized in spring barley seedlings using LC-MS-based methods (Piasecka *et al.*, 2015). Changes in their content in this species under water deficiency have not previously been reported, apart from the situation under osmotic stress (Ishihara *et al.*, 2002). Thus, the mechanism of action of phenolic compounds in barley under water deficit is still ambiguous.

The boom in high-throughput metabolomic techniques has enabled researchers to perform complex investigations of metabolomic responses to various factors, including stresses (Mouradov and Spangenberg, 2014). The selection of metabolites associated with drought-related genes has been studied in wheat (Hill *et al.*, 2013; Acuña-Galindo *et al.*, 2015) and perennial ryegrass (*Lolium perenne*) in terms of fructan content (Turner *et al.*, 2008).

No investigations of metabolomic quantitative trait loci (mQTLs) in barley have been published so far. The search for associations of molecular markers with mQTLs in other plant species has been based on genetic maps constructed with the use of molecular markers such as simple sequence repeats (SSR) or single nucleotide polymorphisms (SNP) and by application of diversity array technology (DArT) (Matsuda *et al.*, 2012, 2014). The mQTL analysis was further related to phenotypes by Lisec *et al.* (2008).

This study aimed to identify metabolites that show a significant change in their level under a moisture deficit. The impact of genetic background on the identified changes was investigated in a greenhouse experiment with a population of 100 recombinant inbred lines (RILs) using the mQTL approach. A strong focus was put on the development of statistical gene-to-metabolite networks based on sub-groups of secondary metabolites to functionally interpret mQTLs.

RESULTS

Identification of metabolites

The raw chromatographic UV data submitted for joint analysis were exported from the Empower software in the form of a (13 201 retention time points $\times$2 wavelengths $\times$1632 samples) table. Data pre-processing allowed us to distinguish 135 chromatographic peaks. Identification of the metabolites corresponding to these peaks was performed on samples selected as the most representative from their UV profiles, enabling us to assign the structural features to 75 of them (for details of the identified compounds see Table S1 in the Supporting Information). It should be noted that some chromatographic peaks contained several co-eluting compounds (Table S1).

Glycosides and acylated glycosides of flavones created the most numerous groups of metabolites identified, whereas phenylpropenoic acid derivatives were observed in smaller numbers. Three compounds were chlorogenic acids (metabolites 1, 2 and 3; see Table S1 for the names of metabolites referenced hereafter by numbers). Three isomeric sinapoyl-hexoses (metabolites 72, 73 and 74), feruloyl-hexose (20) and caffeoyl-hexose (14) were identified in barley either as *O*-glycosides or esters of the respective hydroxycinnamic acids, as reported previously (Piasecka *et al.*, 2015). Hordatines and their glycosides (23–26), being products of the esterification of polyamine agmatine with phenylpropenoic acids and the subsequent condensation of two units, were found in our study and were previously described as secondary metabolites characteristic of *Hordeum* species (Kohyama and Ono, 2013).

Many of the observed secondary metabolites were flavone glycosides. Among them many compounds were previously described by Piasecka *et al.* (2015). However, some other compounds, characteristic only for currently investigated RILs (3, 7–9, 11–13, 18, 29, 30, 37, 39, 42, 50, 51, 56, 61, 63, 68, 70, 73–75) were also found. Glycoconjugates of flavones – apigenin, luteolin, chrysoeriol and tricin – were identified with glucose, rhamnose and arabinose forming saccharidic substituents established on the basis of previous NMR investigations on barley flavonoids (Ohkawa *et al.*, 1998; Ishihara *et al.*, 2002; Piasecka *et al.*, 2015).

The MS fragmentation revealed that 6-*C*-glucosides of apigenin, luteolin and chrysoeriol (traditionally called isovitexin, isoorientin and isoscoparin, respectively) were most often observed in the studied RILs. Additional 8-*C*-glycosylation of the 6-*C*-glycosides was also observed. Three types of isomeric diglycosides, 6-*C*-[2″-glycosyl]-glucosides, 6-*C*-

[6″-glycosyl]-glucosides and 6-*C*-glucosyl-7-*O*-glycosides, were detected.

The presence of flavones acylated with hydroxycinnamic acids has been well documented in barley (Ohkawa *et al.*, 1998; Ferreres *et al.*, 2007, 2008; Piasecka *et al.*, 2015). Substitutions with *p*-coumaric, caffeic, ferulic, hydroxyferulic and sinapic acids were identified in our study as acylated 7-*O*-glycosides or acylated aglycone moieties.

In addition, non-phenolic metabolites belonging to the blumenol terpenoids were also identified in our studies. Blumenols 10-12 and 18 were identified in all samples studied. The fragmentation scheme generated by the ion trap mass spectrometer served to identify particular substituents in these structures. The protonated $[M+H]^+$ ion of compound 10 yielded the release of neutral fragments of 176 amu and subsequently 162 amu for glucuronic acid ($C_6H_8O_6$) and *O*-glucose ($C_6H_{10}O_5$), respectively. The $[M+H\text{-}176\text{-}162]^+$ product ion at $m/z = 211$ corresponded to the blumenol C moiety. The identifications were confirmed by high-resolution mass spectrometry in which the *m*/*z* measured for the $[M+H]^+$ ion was equal to 547.2478 (calculated for $C_{25}H_{39}O_{13}$ 547.2391, error 1.58 p.p.m.). Similarly, the masses of neutral fragments of 176.0321 and 162.0538 as well as the *m*/*z* of the product ion 211.1697 corresponded to the above constituents with an error lower than 2 p.p.m. This enabled us to generate the chemical formula of compound 10 (see Table S1) as blumenol C 2″-*O*-glucuronylglucoside (blumenin) described by Maier *et al.* (1995) and Peipp *et al.* (1997). Compound 11 was identified as an isomer of 10 on the basis of similarities in the fragmentation scheme in the positive ionization (Figure 1a) and they differed in their fragmentation in the negative ion mode. Compound 18 had a very similar fragmentation to that of 10 but the pseudomolecular $[M+H]^+$ ion of compound 18 had an *m*/*z* that was smaller by 2 units, indicating the structure described by Peipp *et al.* (1997) as didehydroblumenol C 2″-*O*-glucuronylglucoside (Figure 1b).

Quantitative changes of metabolite concentrations in drought

The statistical analysis was performed for 98 traits (out of 135) for which a sufficient number of non-zero observations were available in all conditions (single metabolites or groups of co-eluting compounds; detailed results are shown in Table S1) to study their quantitative changes at two time points during application of drought. The principal coordinate analysis showed that the dispersion of lines was the largest at T2 in control conditions (10 days), and the separation between the RILs grown in control and drought conditions was bigger at T2 than at T1 (5 days) (Figure 2).

Comparison of the mean values for parental lines (Table S1) showed that for most of the traits for which the difference between parents was significant, the trait level in the Maresi cultivar was higher than in CamB (e.g. for 49 out of 56 traits discriminating parents under drought at T1) (see also the example chromatograms in Figure S1). The ratio of numbers of traits down-/upregulated by drought changed from 7/42 at T1 to 27/25 at T2 in Maresi, and from 14/23 to 22/30 in CamB. The number of traits with significant mean drought effects increased between T1 and T2 from 61 to 79 for the whole RIL population, and the ratio of the negative to positive effects changed from 14/47 to 6/73, indicating increased accumulation of metabolites in drought over time. The drought effects for metabolites observed at T1 and T2 were correlated ($r = 0.565$, $P < 0.001$) (Figure 3a). Glucosides of hordatine A and C (23), one of sinapic acid glucoside isomers (74) and an unidentified compound (89) revealed the largest negative effects at both T1 and T2. 5-Feruloylquinic acid isomer (3) had an outstanding positive drought effect at T2. A shift from small or negative effects at T1 towards larger, positive effects at T2 was visible, for example for isoorientin 2″-*O*-glucoside (28), isovitexin 4′-*O*-[6″-feruloyl]-glucoside 7-*O*-glucoside (53), isovitexin 7-*O*-[6″-hydroxyferuloyl]-diglucoside (58), isovitexin 7-*O*-diglucoside (63), isovitexin 7-*O*-glucoside (64) and unidentified compounds 99 and 127. It is possible that esterification with ferulic acid and 7-*O*-glycosylation plays crucial role in the accumulation profiles of these metabolites.

A genotype × drought treatment interaction was found for 59 traits at T1 and for 78 at T2 (most, 52, were in common between the time points). The degree of the genotype × drought treatment interaction observed in RILs, measured by the appropriate variance components, was for most traits higher at T2 than at T1 (Figure 3b). The metabolites 2, 11, 37, 73–75 and 89 were characterized by a substantial interaction at T1 only and they were hydroxycinnamic acid derivatives and blumenol C derivatives. On the other hand, interaction for compounds 20, 21, 34, 39, 58, 107, 105, 109, 126 (being ferulic acid derivatives and acylated flavones) was found at T2 only. Compounds with interactions at both T1 and T2 (4, 31, 69, 67) were mostly glucosides and acylated glucosides of apigenin as well as a blumenol derivative (10) and a hexoside of caffeic acid (14).

Heritability coefficients computed under control and drought conditions ranged between 0.96 and 80.72% (in some cases negative variance component estimates did not allow us to estimate heritability; Table S1); the correlation between heritabilities estimated under different conditions was limited (Figure 3c).

The drought effects observed for individual lines at T1 and T2 were visualized using heatmaps and structured by the clustering of RILs and metabolites (Figure 4). The profiles of effects segregated in the population, showing groups of lines particularly different from the parents for specific groups of metabolites (e.g. a group of lines with negative effects for metabolites from cluster 'C' at T1, and a group of lines with positive effects for metabolites from cluster 'A' at T2).

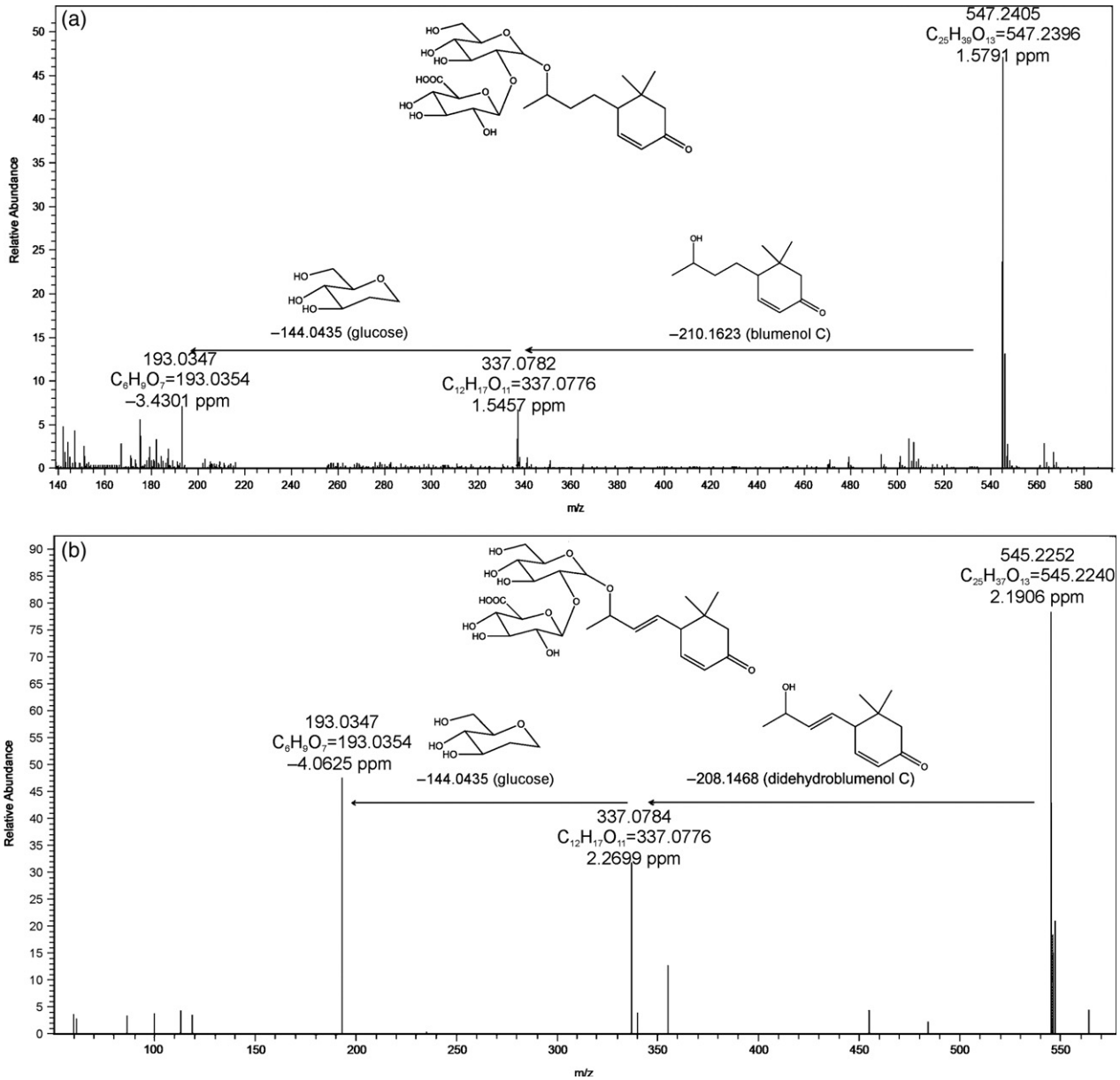

Figure 1. High-resolution mass spectra of (a) blumenol C glucuronylglucoside and (b) didehydroblumenol C glucuronylglucoside.

For the clustering of metabolites, the T1 heatmap allowed the distinction of a large group (group A) of 35 metabolite traits with similar, mostly positive, but relatively small drought effects and a group of 11 traits (group B), with many lines characterized by positive effects, among them glucosides of hordatines (25, 36), which postulates their function in drought response. Group C comprised eight traits with very distinct and mostly negative effects could also be identified, containing, among others, hydroxycinnamic acid derivatives. Two homogeneous sub-groups, each comprising two traits (43 and 66 and 11 and 37), existed within this group. The predominance of 6-*C*-[2″-*O*-glycoside]-glucosides of flavones among the compounds with decreasing concentrations is noticeable.

More positive drought effects were observed at T2. These effects were relatively stronger than those observed at T1 and small groups (pairs) of metabolites with similar drought effects were obtained. A group of seven very distinct compounds with high upregulation in drought could be highlighted. This group comprised isovitexin 7-*O*-[6″-hydroxyferuloyl]-diglucoside (58), isovitexin 2″-*O*-glucoside sinapate (52) and its isomer isovitexin 7-*O*-[6″-sinapoyl]-glucoside (61), 5-feruloylquinic acid (3) with the closest

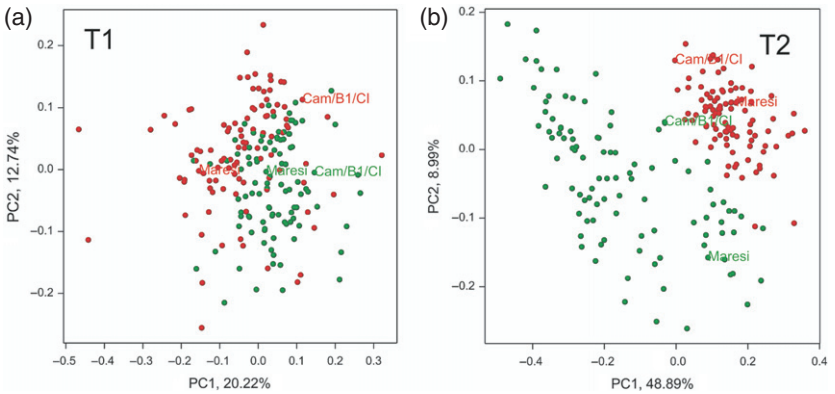

Figure 2. Principal coordinate visualization of recombinant inbred lines based on observations of 98 metabolites in control (green dots) and drought conditions (red dots) at time points T1 and T2.

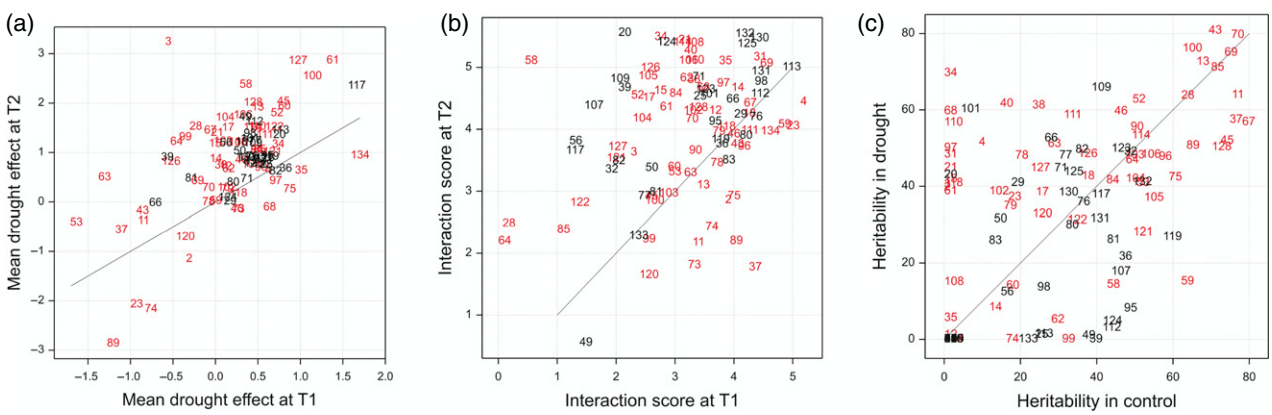

Figure 3. Scatterplots of characteristics of barley recombinant inbred line population for all studied metabolites (represented by numbers, in accordance with Table S1).
(a) Mean drought effects at time points T1 and T2.
(b) Genotype × drought treatment interaction scores (variance component)/(variance component SE) at T1 and T2.
(c) Broad heritability coefficients estimated under control and drought conditions. Continuous lines mark the equality of characteristics. Metabolites for which quantitative trait loci were found at T1 or T2 are marked in red.

relation to isovitexin 7-*O*-[6″-*p*-coumaroyl]-glucoside (60) and two unidentified compounds (117 and 127).

Interestingly, isovitexin 7-*O*-diglucoside (63) was distant from other metabolites at both T1 and T2, and was characterized by the highest variability of effects among genotypes.

Correlation networks

To compare relationships between metabolites under control and drought conditions correlation networks were constructed separately for all four environments, and differential networks comparing correlations in control and drought conditions were made at T1 and T2. The values of connectivity and clustering coefficients for networks were lower in drought than in control conditions, especially so at T2 (Table 1). Within control and drought conditions, the connectivity and clustering increased from T1 to T2.

Modules, which are clusters of correlated metabolites, were distinguished (Table S2a,b). Because of the larger differences between control and drought networks, here we interpret only the situation at T2 (Figure 5). Metabolites in the same module are highly interconnected and modules are visualized as circles. Edges drawn between metabolites show only the highest correlations, and all of them are positive.

The modules 'turquoise', 'blue' and 'brown' in the control network at T2 were the largest and had the highest outside connectivity; therefore all these modules, besides containing correlated metabolites, were also strongly correlated with the other modules (Figure 5a). Since the clustering coefficients for the modules 'turquoise' and 'blue' were higher than that for the 'brown', the neighbourhood of these two modules was more correlated than in case of the 'brown' module (Table S2a). The modules in drought conditions were smaller than in control conditions, and there were no connections between modules (Figure 5b). The module in control conditions at T2, called 'salmon' (Table S2b), consisted of two metabolites, isoorientin 4′,7-di-*O*-glucoside (32) and an unidentified one (85). Conservation of this association was observed, because in drought

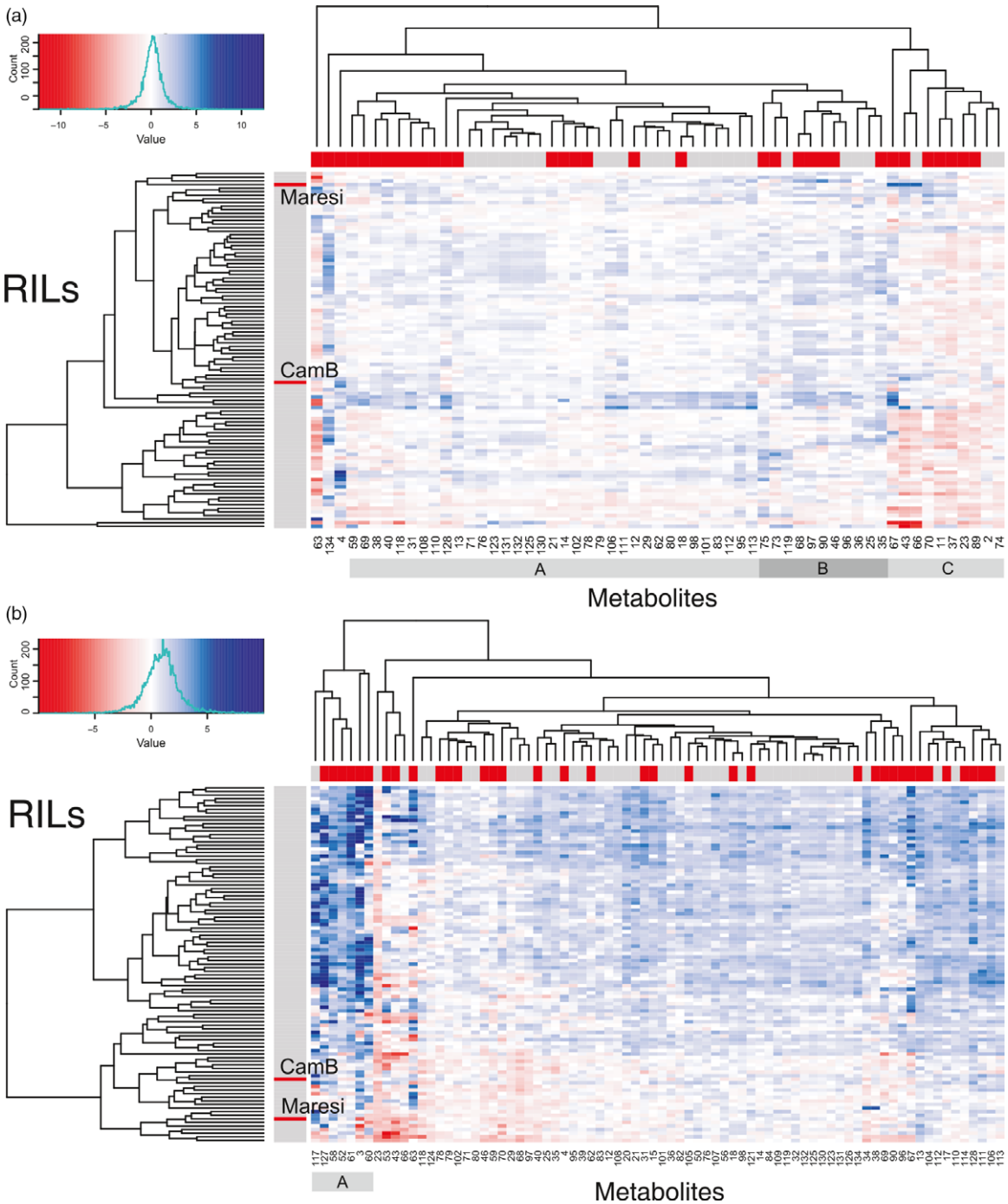

Figure 4. Heatmaps visualizing drought effects for all recombinant inbred lines (RILs) and all metabolites with a significant line × drought treatment interaction. Observations at (a) time point T1 and (b) time point T2. Metabolites for which quantitative trait loci were found at the respective time point and parental lines are marked in red. Letters A, B and C indicate clusters of metabolites discussed in text.

at T2 both metabolites stayed together in one module called 'midnight blue'.

The 'cyan' module in drought at T2 (Table S2a) consisted of two metabolites: isoorientin 2″-*O*-glucoside (28) and isovitexin 7-*O*-glucoside (64): both metabolites stayed together in one module in all networks (Table S2b). However, in all conditions apart from drought at T2 these metabolites were isolated, since their correlations were not sufficiently strong to build an edge in the network. Connectivity for this module increased from 0 in control conditions to 0.46 in stress conditions, which indicated a regulatory connection of these flavone *C*,*O*-diglucosides.

Table 1 Parameters of correlation networks and differential correlation networks for metabolites

Conditions	Number of connected nodes	Connectivity	Clustering coefficient
Correlation networks			
T1 C	44	41.11	0.29
T1 D	24	11.63	0.08
T2 C	68	170.99	0.47
T2 D	30	17.08	0.14
Differential correlation networks			
T1, control versus drought	54	65	0.05
T2, control versus drought	84	567	0.23
Control, T1 versus T2	82	510	0.15

The 'salmon' module in drought at T2 (Table S2b) consisted of feruloyl-hexose (20), feruloyl-quinic acid glucoside (21) and an unknown metabolite (107). All these metabolites were also in one module in control conditions at T2. In drought they belonged to the module with many correlations both inside and outside it, while in control conditions all these metabolites were isolated.

Construction of differential correlation networks allowed us to observe a marked increase in time of the number of significant changes in correlations between control and drought conditions (Figure S2, Table 1). The large number of correlations that changed at T2 was to a great extent due to the fact that the correlation network at T2 C (see Table 1) had a very high connectivity. Nevertheless, considering small differences in connectivity between drought at T1 and T2 (from 11.63 to 17.08, Table 1), a significant increase of connectivity in the differential network at T2 compared with T1 was observed.

Quantitative trait loci analysis and annotation

One hundred and thirty-eight mQTLs found for all 98 traits, listed in Table S3, are shown in the linkage map in Figure 6. They were unevenly distributed among all linkage groups, with 53 mQTLs in 1H.1, 24 in 2H and 17 in 5H.3. Fewer metabolites with corresponding mQTLs, as well as fewer mQTLs, were found at T1 than at T2 (Table 2); for 25 metabolites mQTLs were found at both T1 and at T2. In more than half of cases alleles from the Maresi cultivar provided an increased metabolite content (consistent with comparison of the parental varieties above). The projected two logarithm of the odds (LOD) score intervals for 126 mQTLs were no longer than 30 Mbp and were taken into account in QTL annotation. One hundred and fifteen mQTLs out of 138 (83.3%) contained at least one gene bearing the gene ontology (GO) biological process adnotation to 'response to stress' (GO 06950) or to its subclasses (107 genes in total; total length of intervals approximately 216 Mbp, see Table S4a).

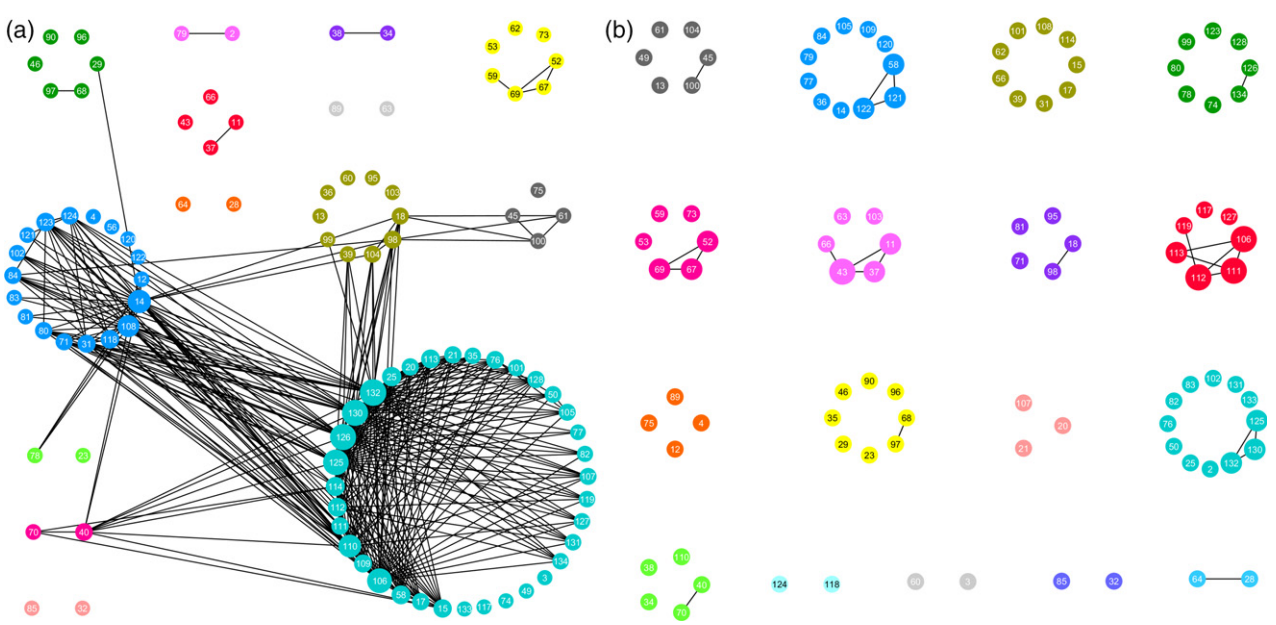

Figure 5. The correlation network of metabolites at time point T2.
The correlation network of metabolites observed at T2 in (a) control conditions and (b) in drought. The size of the circle representing a metabolite is proportional to the number of edges created by this compound. Modules of metabolites are visualized by different colours (colours are assigned randomly, with no correspondence between conditions). Only edges corresponding to elements of the topological overlap matrix greater than 0.2 are shown, both within and between modules. All edges correspond to positive correlations.

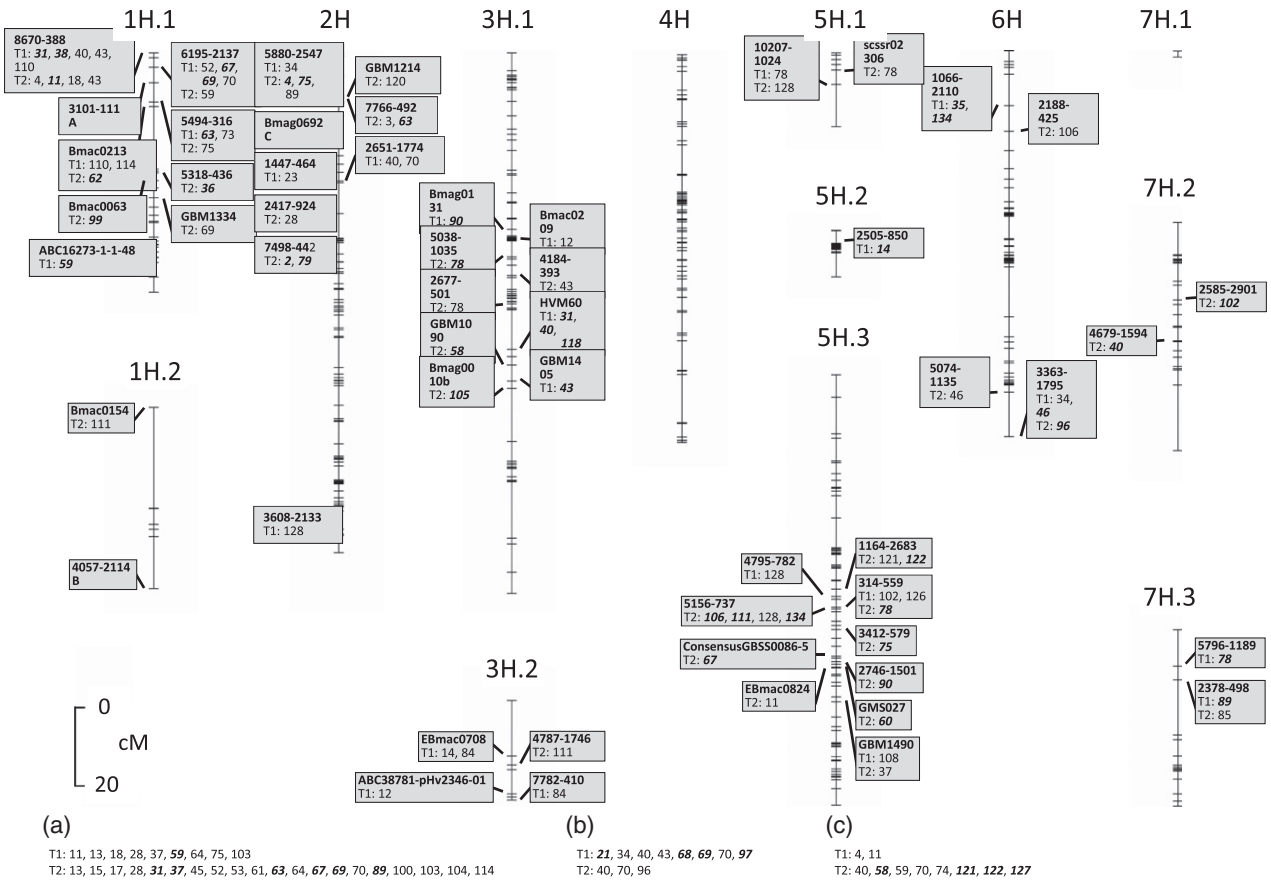

Figure 6. Quantitative trait loci (QTLs) for metabolic traits.
The QTLs for metabolic traits identified in the barley recombinant inbred line population obtained by crossing Maresi and CamB, shown in the genetic map consisting of 13 linkage groups 1H.1, 1H.2, . . ., 7H.3 ascribed to barley chromosomes 1H 7H. Each box contains a marker name and numbers of metabolites with QTLs linked to that marker at time points T1 and T2; metabolites with QTLs at markers 3101-111, 4057-2114 and Bmag0692 are listed below the graph as groups A, B and C, respectively. QTLs expressing interaction with drought treatment (environment) are given in bold italic. The exact positions of all QTLs are given in Table S3.

The QTLs found at T2 could be visualized by a network of metabolites and markers, as shown in Figure S3. In general, the metabolites found to be related in the correlation network analysis (belonging to one module) showed a tendency to be linked to common markers; for example, correlated metabolites (13, 45, 61, 100 and 104) were linked to SNP 3101-111. Some metabolites were linked to isolated markers, but also groups of compounds with a joint system of QTLs were found. Some connections between such groups, for example the link between SNPs 3101-111 and 5880-2547 caused by unidentified metabolite 89, were only significant in the control conditions.

The markers that were found to be linked to the largest number of traits represented mQTL hotspots. The SSR marker Bmag0692 was linked to two and eight metabolites at T1 and T2, respectively. Regions of mQTLs linked to this marker contained 16 genes annotated to stress response, in particular the genes *USP1*, *PRX8* and *MLO3*. Marker 3101-11 (1H.1, 9.27 cM) was linked to nine metabolites at T1 (in one case showing interaction of the genetic effect with drought treatment), and to 20 at T2 (six with interaction); five metabolites (13, 28, 37, 64 and 103) were linked to this marker at both T1 and T2. In most QTLs the Maresi alleles increased the metabolite levels (8 metabolites at T1, 14 at T2). The mQTLs that were found at both T1 and T2 did not express interaction with drought treatment and the sign of their effects was consistent over time. In the mQTL regions marked by SNP 3101-11 there were 12 genes annotated to stress response, in particular *RGH2B* and *CI2D*.

The trait with the strongest mean drought effects in RILs, hordatine A and C glucosides (23), had lower levels in drought at both T1 and T2. At T1 it was linked to SNP 1447-464 (with the additive effect being positive in control and negative in drought conditions) which is located in the region of three oxidative stress-related genes *MLOC_55663, MLOC_66375* and *MLOC_68384*.

A large number of mQTLs (21) was associated with traits in peaks 34, 40, 43 and 70. Only three out of these mQTLs

Table 2 Summary of metabolomic quantitative trait loci (mQTLs) found

Time point	No. of metabolites with mQTLs	No. of mQTLs (per metabolite)	No. of SNPs marking mQTLs (per mQTL)	Fraction of mQTLs interacting with drought treatment	No. of mQTLs with Maresi allele increasing metabolite level (% of all significant)	
					Control	Drought
T1	41	60 (1.46)	28 (0.47)	43.3%	30 (65.2%)	35 (62.5%)
T2	51	78 (1.53)	39 (0.50)	44.9%	42 (65.6%)	38 (57.6%)

expressed interaction with drought treatment. Sixteen mQTLs were associated with intervals containing together 49 stress-related genes. Compounds 40 and 70 were linked at T1 to a common marker, 2651-1774, and at T2 to another, Bmag0692; because of this, at both T1 and T2 two stress-related genes, *USP1* and *PRX8*, were in their mQTL regions. Trait 43 was the only one identified compound whose mQTL interval contained genes annotated specifically to response to water deprivation (GO:0009414); these were: *MLOC_4896* at T1, and *MLOC_43749 (CESA5)* and *MLOC_67_140* at T2.

5-Feruloylquinic acid (3), which was strongly upregulated in RILs in drought at T2, was at T2 assigned a QTL linked to SNP 7766-492 (with the Maresi allele increasing the metabolite level). No stress-related genes were found in the region of this mQTL. On the other hand, for three metabolites (4, 75 and 89), which exhibited 10 mQTLs, all QTL regions contained genes annotated to stress response (31 in total). Four mQTL regions contained genes annotated by GO to phenylpropanoid- or flavonoid-related synthetic and metabolic processes (Table S4b). Variability of a number of metabolites was linked to the polymorphism at SNP 8670-388 that marked an mQTL region at both T1 and T2.

DISCUSSION

Cultivation of a population of barley lines in control and drought conditions allowed us to identify metabolites with an accumulation pattern that was mostly differentiated under two watering regimes. Due to the large amount of plant material (a population of 100 RILs grown in standard conditions and under water deficit, and measured at two time points in three biological replications), we were able to robustly estimate the correlations between quantitative profiles of different metabolites as well as find the cluster structure of the drought effects. Finding those metabolites with a line × drought treatment interaction was especially important, as these are the compounds that in reaction to drought change differently in different lines, and are thus interesting from the point of view of the genetic constitution of the plants. The availability of high-density SSR and SNP marker data allowed us to find the genomic regions responsible for determination of the quantitative levels of a number of traits in the form of mQTLs, taking into account possible QTL × environment interactions.

With regard to the statistical data analysis methods, the algorithms of assessing the variation, correlation and clustering of metabolomic profiles were standard ones that have also been used in other studies (e.g. Schauer *et al.*, 2006; Matsuda *et al.*, 2012). This did not apply to the chromatographic data pre-processing methods and peak identification steps, which, although based on some known procedures, were developed specially for this study. Namely, we developed data processing methods applicable to very large chromatographic data sets. The protocol consisted of several integrated and computationally optimized stages, some performed using publicly available software, some with our own scripts written for known algorithms, such as chromatogram alignment by correlation optimized warping, and others with our own, new algorithms, such as for peak detection and deconvolution. The most time-consuming part was the retention time alignment. There are many known alignment methods, but the most commonly used algorithms are correlation optimized warping (COW) (Nielsen *et al.*, 1998), dynamic time warping (DTW) (Tomasi *et al.*, 2004) and fast Fourier transform cross-correlation (RAFFT) (Wong *et al.*, 2005). DTW uses distances as a measure of the similarity of two peaks and is sensitive to different peak intensities, so, according to Pravdova *et al.* (2002), may lead to unsatisfactory alignment. COW showed very good overall results in a comparison of five frequently used warping algorithms (Jiang *et al.*, 2013), therefore we used the COW algorithm (Nielsen *et al.*, 1998; Pravdova *et al.*, 2002; Tomasi *et al.*, 2004). COW uses the correlation coefficient as the measure of similarity between two chromatograms. The COW method operates on the entire chromatogram; it aligns profiles by matching their shapes and compressing or expanding time intervals, whereas the peak height and area are largely unaffected. It should be noted that our approach to alignment does not require prior chromatographic peak identification, in contrast to several other methods (Daszykowski *et al.*, 2010; Wang *et al.*, 2010; Tomasi *et al.*, 2011; Zhang *et al.*, 2012).

A special algorithm was required for the second step of data processing, peak detection, as the programs commonly used for this in Ultra Performance Liquid Chromatography (UPLC) data, for example Empower, with two algorithms – Traditional Integration and ApexTrack Integration (also see Cardot *et al.*, 1990) – work on single chromatograms and were not applicable to our situation of multiple samples. Instead, we detected peaks within individual chromatograms using profiles of the smoothed second derivative. After peak detection, the problem of peaks that are too long might occur because of the presence of metabolites with very similar (overlapping) retention time intervals. Also, we should assume that in a large experiment, with heterogeneous plant material, different metabolites might be detected in different samples. Again, the existing deconvolution algorithms (e.g. Vivó-Truyols *et al.*, 2005; Romanenko *et al.*, 2006) devised to separate signals for co-eluting compounds could not be applied in our situation due to the large number of samples. According to our proposition peaks that were too long were divided into smaller ones by hierarchical clustering of samples.

The methodology of metabolomic profiling and compound identification enabled annotation of more than a half of chromatographic peaks. The chromatographic data processing allowed detection of 135 peaks that were quantified and submitted to statistical analysis. Out of those, 75 were assigned annotation on the basis of two complementary systems: low-resolution MS with an ion trap analyser and high-resolution MS with an Orbitrap analyser, jointly giving the power of isomer differentiation. Unfortunately, some metabolites had similar physicochemical properties, which caused co-elution of the compounds and annotation of several metabolites to one chromatographic peak. In these cases quantitative description of each compound in the peak was impossible and all compounds present in it were considered as one trait. However, co-eluting compounds usually belonged to common structural groups, for example two glucosides of hordatines were annotated in peak 23; 6,8-di-*C*-glycosides of flavones in peak 4, etc. Thus, quantitative changes of the traits in drought conditions were associated with generic structural features for all co-eluted components in the peak.

In barley, as well as in other plants from the family Poaceae, *C*-glycosides of flavonoids are the most abundant metabolites (Ferreres *et al.*, 2008; Piasecka *et al.*, 2015). Many types of glycosidic substitutions are known in these metabolites (Ferreres *et al.*, 2007, 2008). In our study 7-*O*-glycosides and 6-*C*- and 8-*C*-glycosides as well as di-*O*,*C*-glycosides were detected in isomeric structures. Acylation of glycosides with hydroxycinnamic acids additionally made identification of the metabolites difficult. However, thanks to multiple fragmentation steps in the ion trap multiple-stage mass spectrometry (MS^n) experiments we were able to distinguish isomeric compounds, even when present within the same chromatographic peak (Figure 1).

The statistical analysis showed higher metabolite levels in Maresi than in CamB in all environments, and different metabolomic reprogramming in two parental varieties was observed during stress. Thus, it appears that both parental varieties deploy different antioxidant systems during drought stress. The numbers of metabolites upregulated under drought decreased from T1 to T2 in Maresi and increased in CamB. This would mean a better metabolomic protection of leaves in CamB than in Maresi according to the hypothesis that over-accumulation of particular phenolic structures provides the extended drought tolerance and oxidative properties (Nakabayashi *et al.*, 2014).

For the population of RILs, the experiments showed that the plants reacted to drought conditions during growth with a number of up-regulated metabolites that increased over time. Moreover, differentiation of the reaction of lines also increased over time of the stress application. There were metabolites that had patterns of content change different from the others, and in these cases an exceptional decrease was observed at the beginning of drought followed by an increase at a later stage. The correlations of metabolites were stronger in control conditions than in drought over the whole period of treatment; the metabolites which showed particular changes in correlation with others were also characterized by different reactions of the RILs under different watering conditions. All this showed that the application of stress drastically affected the dynamics of metabolites, especially at T2, and intensive metabolomic reprogramming during plant response to stress was observed.

Glycosides of flavones showed diverse drought effects. Generally, the content of 7-*O*-glycosides and acylated glycosides of flavones increased in drought, while that of specific structures of 6-*C*-[2″-*O*-glycosides]-glucosides of flavones decreased at T1. These compounds were reported to have antioxidative properties in terms of inhibition of malonaldehyde formation from various lipids oxidized by UV light or Fenton's reagent (Nishiyama *et al.*, 1994). Interestingly, isovitexin 7-*O*-diglucoside (63) was distant from the other metabolites at both T1 and T2, and was characterized by the highest variability of effects among genotypes; varied regulation of the compound may indicate a central position of that compound in biosynthesis of flavones during stress and different enzyme activity of glucoside 1,6-glucosyltransferase of flavonoids which catalysed biosynthesis of the compound (Masada *et al.*, 2009).

At T2 hydroxycinnamic esters of flavones were dominantly increased; thus, we postulate an essential role of acylation during the prolonged drought. A similar observation was described for *Arabidopsis thaliana* (Nakabayashi *et al.*, 2014), in which enhancement of drought tolerance

was related to acylated compounds. More specifically, our experiments showed a bigger mean accumulation of hydroxycinnamic acid derivatives in drought. At T1 simple phenylpropanoid derivatives were mainly up-regulated (2, 23, 73 and 74). Previously we demonstrated great variability of hydroxycinnamic acid glycosides and esters in barley (Piasecka *et al.*, 2015). However, only some of these compounds were found in plants of the studied RILs. This group of secondary metabolites plays multiple roles in plants. The concentration of chlorogenic acids can increase drastically during drought in hawthorn (*Crataegus laevigata* and *Crataegus monogyna*) (Kirakosyan *et al.*, 2004), elevated levels of CGAs in transgenic tomato plants were shown to provide increased protection from harmful UV light (Clé *et al.*, 2008) or in response to mechanical damage (Kojima and Kondo, 1985). Meanwhile, the closely related shikimate esters are known to be key intermediates in the synthesis of lignin (Chen and Dixon, 2007), which also plays a crucial role in many environmental stresses; thus, increased level of phenylpropanoids in our experiment may be related to increased biosynthesis of lignin.

It was noteworthy that most metabolites whose contents changed significantly at T2 had methoxyl groups in their structures, either directly on the aglycones or in the acyl moieties. In our experiments contents of metabolites containing ferulic acid were changed by drought to the largest extent and the highest increase was observed for 5-feruloyl-quinic acid at T2. Ferulic acid was previously postulated to be a drought-induced metabolite in wheat (Hura *et al.*, 2009). Similarly, our investigation indicated that ferulic acid and its derivatives were specialized drought-responsive elements. Ferulic acid was also reported to be a compound involved in plant reactions under different stress conditions, for example excess UV radiation (Hura *et al.*, 2012), cold shock in *A. thaliana* (Kaplan *et al.*, 2004) as well as fungal diseases in barley (Hendawey *et al.*, 2014). It was also a factor that increased dehydration tolerance in barley (Li *et al.*, 2013).

The significance of ferulate components in the stress response may be a result of the methoxylation of the ferulate structure. It seems that the most important effect of methoxylation is the shift of the UV absorption maximum towards longer wavelengths. For instance, the maximum of absorption (λ_{max}) is 295 nm for *p*-coumaric and caffeic acids while λ_{max} = 320 nm for ferulic and hydroxyferulic acids and 325 nm for sinapic acid. Similar shifts are observed for apigenin (λ_{max} = 265, 332 nm) and chrysoeriol (λ_{max} = 269, 345 nm). The simultaneous presence of metabolites with and without methoxyl groups extends the range of absorption of the highest energetic UV-B and UV-A radiation and thus may increase the efficiency of UV attenuation and indirectly prevent the formation of free radicals generated by photosystems damaged by drought. The introduction of acyl groups also enhanced physical parameters such as thermostability and light resistivity as well as biological effects such as antioxidant and antimicrobial activity (Ishihara and Nakajima, 2003; Mellou *et al.*, 2005).

Stochmal and Oleszek (2007) showed a significant increase in the absorption of UV-B radiation for acylated flavonoids compared with the respective glycosides. The increased methylation of hydroxycinnamic acids expands the range of absorbed UV light and the connection of these acyls with flavones significantly increases the intensity of absorption. Simultaneously, blocking of the hydroxyl groups of flavonoids by glycosylation suppresses their reactivity and allows these metabolites to accumulate in the vacuole. These facts suggest that phenolic secondary metabolites participate in protection against excessive UV radiation upon photosystem failure related to moisture deficit in barley. This is in agreement with the observed accumulation of flavone as a response of barley to excessive UV radiation (Schmitz-Hoerner and Weissenböck, 2003; Kaspar *et al.*, 2010) and confirms the conclusion of previous studies on the UV-attenuating role of flavonoids under water deficit in wheat (Hura *et al.*, 2009; Neugart *et al.*, 2012).

Hordatines and their glycosides (23–26) are a group of hydroxycinnamic acid amides characteristic of *Hordeum* species (Kohyama and Ono, 2013). Two *p*-coumaroylagmatines form hordatine A, *p*-coumaroyl- and feruloyl-agmatine form hordatine B and two feruloyl-agmatines form hordatine C (Gorzolka *et al.*, 2014; Piasecka *et al.*, 2015). These compounds were regarded to be phytoalexins with an antibiotic effect on pathogens; however, their participation in the drought stress response was not examined previously. In our studies different isomers of this group showed divergent changes in drought. Traits (compound 23) which contained glucosides of hordatines A and C significantly decreased during drought in most of RILs, whereas their isomers in peak 25 and hordatine C (peak 26) increased.

According to our results, non-phenolic blumenol metabolites could also be detected in barley leaves and changes of their contents upon drought conditions were demonstrated. Blumenols were previously identified in roots of Poaceae plants and were known from their mediation in formation of arbuscular mycorrhizas of barley (*H. vulgare* L.), wheat (*Triticum aestivum* L.), rye (*Secale cereale* L.) and oat (*Avena sativa* L.) with *Glomus intraradices* (Maier *et al.*, 1995). However, these compounds were not supposed to act as abiotic stress response factors. The decrease of blumenol C *O*-glucuronylglucoside content at T1 was characteristic for most of RILs. A genotype × drought interaction was also shown for this compound at both T1 and T2; thus some role in stress response might be postulated for this compound.

The construction of correlation networks of secondary metabolite profiles is frequently applied in studies of plant

responses to various stresses (Hochberg *et al.*, 2013, 2015; Leisso *et al.*, 2015). Correlation networks were constructed separately for four conditions (two time points, control and drought-treated plants) for profiles obtained in our experiments. The profiles were more correlated in control conditions than in drought. Correlations of several metabolites remained unchanged in drought conditions compared with control conditions. Compounds 52, 53, 59, 67, 69 and 73, being mostly isovitexin glycosides acylated with different cinnamic acid derivatives, formed a separate module in both conditions. On the other hand, large modules formed under control conditions were no longer visible in the drought-treated plants, which indicated drought-related destabilization of the metabolomic correlation. The QTL analysis performed at T1 and T2, using the method that allowed us to find significant QTL × treatment interactions, revealed regions and candidate genes for regulation of the level of a metabolite or groups of metabolites in drought. Some mQTL hotspots were found, the main ones being in the linkage groups 1H.1 (at SNP 3101-111, 7.63 cM) and 2H (SSR Bmag0692, 16.01 cM), with some of the traits determined there being correlated. The existence of regions influencing multiple metabolites has been previously confirmed, for example by Rowe *et al.* (2008) and Matsuda *et al.* (2012). On the other hand groups of metabolites, also of the correlated ones, were separated into regulation groups associated with different genomic regions.

Annotation of the mQTLs that we used in this paper is putative, and its accuracy is highly related to the statistical variability of the mQTL positions and effects, and also the correctness of the linear gene assignment in the current reference genome. To identify genes located in neighbourhood of mQTLs we applied 2-LOD support intervals, which resulted in projections on genomic sequences ranging from 100 to 19 Mbp (projections longer than 30 Mbp were not interpreted). In a similar analysis, Matsuda *et al.* (2012) used 1-LOD support intervals that gave projected intervals of up to about 25 Mbp. We think that our approach is more justified by the statistical properties of estimated QTL positions. However, as very little is known about pathways synthesizing the metabolites that were identified in our study, the likelihood of finding strong enzymatic or regulatory candidates is low, even with quite broad intervals. In the interpretation we concentrated on genes annotated to the biological process 'response to stress'. Taking into account the total length of the barley genomic sequence (about 4 Gbp) and the total number of genes annotated to this GO term (1450), our projected QTL intervals covered 5% of the genome and contained 7% of stress-related genes. Although this cannot be treated as a formal representation test, it shows that the genomic regions involved in regulation of the identified compounds have a role in the reaction of plants to stresses.

The largest number of mQTLs (at least three) was at T1, associated with traits in peaks 40 (a mixture of co-eluting compounds chrysoeriol 7-*O*-diglucoside; isovitexin; isovitexin 2″-O-rhamnoside), 43 (isovitexin 2″-*O*-glucoside isomer 1; isovitexin 2″-*O*-arabinoside), 70 (isovitexin 2″-*O*-rhamnoside) and 34 (isoorientin 7-*O*-[6″-sinapoyl]-glucoside). Most of these metabolites were glycosylated at position 2″-*O*-, which may be an important factor in their transport to vacuoles and storage. These structures were also described as inhibitors of glyoxal formation from the oxidative degradation of three fatty acid esters (Nishiyama *et al.*, 1994). All mQTL regions found for these metabolites contained candidate stress reaction-related genes.

Previous searches for the genetic background of metabolites biosynthesis revealed QTLs related to controlling the levels of phenylpropanoids in *Populus* (Morreel *et al.*, 2006) and amino acids in tomato (Toubiana *et al.*, 2015). As we have pointed out, our analysis of the association of genes with mQTLs was performed on rather wide LOD-supported intervals ranging from 100 to 19 Mbp. Interestingly, almost no putative associations between metabolites and their biosynthetic genes were found. Only mQTLs for several compounds were associated with gene *MLOC_57430* corresponding to an uncharacterized protein connected with flavonoid biosynthetic processes. On the other hand, associations with numerous genes connected to defence response processes, including response to cold, heat and oxidative stresses, were found. However, most of these genes corresponded to unknown proteins according to Ensembl Plants Gene Ontology.

The specific role of the metabolites analysed here in reaction to drought is difficult to predict, with the exception of some cases described in literature. Phenolic secondary metabolites, including flavonoid glycoconjugates, were reported to have a protective effect in chloroplasts, cytoplasm and the nucleus. The relationship of these compounds to chloroplastic metabolism is not yet sufficiently clear; however, their contribution to regulation of the redox state in chloroplasts (Arora *et al.*, 1998) and participation of the chloroplast-located flavonoids in scavenging singlet oxygens from photosystems (Agati *et al.*, 2007) has been reported. In spite of the well-known ability of phenolic compounds to directly scavenge molecular species of active radicals they may be substrates in peroxidase detoxification of H_2O_2 (Agati *et al.*, 2007). Peroxidase activity mainly catalyses the oxidation of flavonoids and hydroxycinnamic acids, such as ferulic and *p*-coumaric acids, in vacuoles and apoplasts.

It should be stressed that the results reported here form part of an investigation performed on the same plant material used to study the biochemical traits, physiology and phenotype of the parental forms and RILs (Filek *et al.*, 2014; Mikołajczak *et al.*, 2016, 2017). More information on the role of phenolic compounds in drought, as well as on

genetic co-regulation of the metabolome and expression of economically important traits related to yield in barley, will be obtained from the joint integrative analysis of all traits measured in the experiments. Such an analysis is in progress and will be the subject of a separate publication.

EXPERIMENTAL PROCEDURES

Plant cultivation

The population of RILs derived from the cross between European and Syrian barley (*Hordeum vulgare* L.) accessions Maresi and Cam/B1/CI08887//CI05761 (hereafter referred to as CamB) that was used in our investigations has been studied for yield-related and other phenotypic traits by Mikołajczak *et al.* (2016, 2017). Maresi is a German semi-dwarf cultivar with the pedigree Cebeco-6801/GB-1605//HA-46459-68, and CamB is a Syrian breeding line adapted to a dry environment. Maresi has been characterized by Mikołajczak *et al.* (2017) as better than CamB for most of the yield-related traits in both control and drought stress conditions. However, the relative decrease of main spike yield characteristics caused by water shortage was lower for CamB than for Maresi.

Plants were grown in pots filled with loam-sandy soil. The control plants (treatment C) were grown at 12% of soil moisture during the entire vegetation season. When the seedlings were at the three-leaf stage (phase 13 in the BBCH scale, reached 16 days after sowing, when the third leaf was observed for at least 50% of plants) the soil moisture was decreased and the moisture level of 6%, achieved within 48 h, was maintained for 10 days (treatment D). Soil moisture was measured by a moisture temperature and salinity of soils (FOM/mts) device and corresponded to 2.2 and 3.2 pF in the control and drought conditions, respectively.

Leaf samples were harvested from plants subjected to drought after 5 and 10 days of moisture deficit (time points T1 and T2). On the same days, leaves from control plants were sampled. All samples were collected in three biological replicates. The harvested material was immediately frozen in liquid nitrogen and stored at −80°C until metabolite extraction and instrumental analysis.

Extraction of phenolic compounds from plant material

Weighted samples of frozen leaves (about 100 mg) were placed in Eppendorf tubes with 1.4 ml of 80% methanol and the internal standard, homogenized using a ball mill (MM 400, Retsch, http://www.retsch.com/) and subsequently placed in an ultrasonic bath for 30 min and centrifuged (11 000 ***g***) for 10 min. Apigenin (5 μl of 1 mg ml^{-1} solution in dimethylsulphoxide) was used as the internal standard in extracts prepared for the UPLC analyses. Samples of the crude extracts were directly subjected to HPLC-MS analysis, and prior to the UPLC analysis 200 μl was diluted with 80% methanol (800 μl). All reagents and solvents for extraction, as well as for UPLC and HPLC-MS analyses described in the "Phenolic compounds profiling and identification" section (methanol, acetonitrile, formic acid, dimethylsulphoxide, polyamide and chloroform), were from Sigma-Aldrich (http://www.sigmaaldrich.com/), and ultrapure water was obtained from a Millipore Direct Q3 device. Standards of flavonoids (apigenin, apigenin-7-*O*-glucoside, apigenin-6-*C*-glucoside, apigenin-8-*C*-glucoside, isorhamnetin, luteolin, luteolin 7-*O*-glucoside, luteolin 4′-*O*-glucoside, luteolin-8-*C*-glucoside, luteolin-6-*C*-glucoside, quercetin 3-*O*-galactoside, quercetin, rutin, isorhamnetin, isorhamnetin 7-*O*-glucoside) were purchased from Extrasynthese (http://www.extrasynthese.com/).

Phenolic compounds profiling and identification

Metabolic profile analyses were performed using two complementary LC-MS systems. The first system, high-performance liquid chromatography coupled with diode-array detection and multiple-stage mass spectrometry (HPLC-DAD-MS^n) consisting of an Agilent 1100 HPLC instrument with a photodiode-array detector (http://www.agilent.com/) and an Esquire 3000 ion trap mass spectrometer (Bruker Daltonics, https://www.bruker.com/) with the XBridge Shield C18 column (150 × 2.1 mm, 3.5 μm particle size, Waters, http://www.waters.com/), the MS^n spectra were recorded separately in the negative and positive ion modes using the previously published approach (Piasecka *et al.*, 2015). The elution was conducted with water containing 0.1% formic acid (solvent A) and acetonitrile (solvent B). The gradient elution was started at 92% of A and linearly changed to 90% of A in 10 min, then to 75% of A in 30 min and to 2% of A over 10 min, followed by the return to the initial conditions and the column was re-equilibrated for 10 min.

The second system consisted of UPLC with a photodiode-array detector PDAeλ (Acquity System; Waters) hyphenated to a high-resolution Q-Exactive hybrid MS/MS quadrupole Orbitrap mass spectrometer (Thermo Scientific, http://www.thermofisher.com/). Chromatographic profiles of metabolites and the quantitative measurements were obtained using water acidified with 0.1% formic acid (solvent A) and acetonitrile (solvent B) with a mobile phase flow of 0.4 ml min^{-1} on a BEH Shield C18 column (2.1 × 150 mm, 1.7 μm particle size; Waters) at 60°C. The injection volume was 2 μl. The gradient elution was started at 99% of A and linearly changed to 80% of A over 2 min, then to 70% of A over 8 min and to 5% of A over 1 min, followed by a return to the initial conditions and re-equilibration for 2 min. UV absorbance was measured in the wavelength range 230–450 nm with a resolution of 2 nm and data were exported with Empower 2 Chromatography Data Software (Waters). Chromatograms drawn at wavelengths of 280 and 330 nm were recorded, used for quantitative measurements of secondary metabolites in samples from the RIL population and subjected to further pre-processing and statistical analysis.

Q-Exactive MS operated in Xcalibur version 3.0.63 with the following settings: heated electrospray ionization ion source voltage −3 kV or 3 kV; sheath gas flow 30 L min^{-1}; auxiliary gas flow 13 L min^{-1}; ion source capillary temperature 250°C; auxiliary gas heater temperature 380°C. The tandem mass spectrometry (MS/MS) experiments were performed using a collision energy 15 eV in full MS/data-dependent MS/MS scan modes in both positive and negative ionization.

The application of two LC-MS systems – low-resolution ion trap and high-resolution Orbitrap – and comparison of mass spectra with own library and data bases (KNApSAcK, PubChem) provided means for tentative identification of the analysed compounds (Piasecka *et al.*, 2015) and annotation of these traits to UV signals in chromatographic profiles.

Chromatographic data pre-processing

The raw data were normalized by dividing every single chromatogram value by the mass of the sample. The chromatographic baselines were removed without their explicit estimation using numerical differentiation (Feuerstein *et al.*, 2009). Chromatogram alignment was performed using COW (Nielsen *et al.*, 1998). Following the alignment, peak detection was performed, with 'peaks' interpreted as the retention time intervals in which a metabolite, or group of metabolites, with similar chromatographic properties occurs, using profiles of the second derivative smoothed with

cubic smoothing splines. The inflection points were located to find boundaries for the individual peaks and peaks common for all chromatograms were built. In the case of too long peaks that occurred because of the presence of metabolites with very similar (overlapping) retention time intervals, deconvolution was done by clustering of profiles using the average link algorithm with a correlation-based similarity coefficient (using function *pvclust* in R). Finally, the chromatograms were integrated over peaks to obtain numerical values for all peaks and all samples. For more details of the data pre-processing procedure see Methods S1. Computations were implemented in R scripts (available at https://github.com/asaw/POLAPGEN).

Statistical analysis

Observations equal to zero (below the detection level) were substituted by half of the minimum non-zero observation for each metabolite. Then observations were transformed by $10^9 \log_2(x)$. Statistical analysis was performed separately for T1 and T2 (in Genstat 18 (VSN International, 2015) except for parts referenced below to the R system) only for those traits for which the number of lines with mean values above the detection limit was bigger than 70 in each of four environments (C, D) $\times$ (T1, T2). Analysis of variance was used, separately at T1 and T2, to test the significance of the difference between treatments C and D with respect to metabolite level (fixed effect, $P < 0.05$, Bonferroni correction for the number of metabolites), as well as of the differences between RILs and of the interaction of RILs with treatments (random effects, variance components significant when larger than three times the SE). The influence of drought was measured by the difference in mean values (D − C) indicating positive (up-) or negative (down-) regulation of metabolite accumulation in drought (significance of individual drought effects for lines assessed by a *t*-test at $P < 0.05$). Analysis of variance separately for control and drought was used to estimate broad sense heritability coefficients for metabolomic traits (as the ratio of appropriate variance components). Principal coordinates analysis (with similarity coefficients based on Euclidean distances) was done for all traits and hierarchical bi-clustering of RILs and traits (with a complete link algorithm, using the function *heatmap2* in R) was performed for the subset of traits expressing RIL $\times$ treatment interaction. Correlation networks and differential correlation networks were constructed according to Langfelder and Horvath (2008) (using the *WGCNA* package in R) and visualized by Cytoscape (Shannon *et al.*, 2003) (see also Methods S1).

Metabolomic QTL analysis

Genotyping of parental forms and RILs with respect to SSR and BOPA1 SNP molecular markers and construction of the linkage map (containing 117 SSR and 702 SNP markers, comprising 13 linkage groups attributed to barley chromosomes 1H 7H) was performed as described by Mikołajczak *et al.* (2016). QTL localization was performed for each metabolic trait independently, separately at T1 and T2, with the mixed linear model-based interval linkage mapping method described by Malosetti *et al.* (2013) and implemented in Genstat 18 (VSN International, 2015), using the compound symmetry model of environmental covariation between control and drought conditions, a genome-wide error rate smaller than 0.01 and selection of significant QTL effects in the final model at $P < 0.05$. For each QTL the 2-LOD QTL support interval was computed according to Xu (2010); the mean length of all intervals, *m*, was further used as a parameter in QTL annotation. SNP sequences taken from Close *et al.* (2009; supplementary material file BOPA1 SNP 1471-2164-10-582-S19.xls) were mapped to the barley genome space in Ensembl Plants (v.082214v1; reference repeat masked sequence Hordeum_vulgare.082214v1.28.dna_rm.toplevel.fa, NCBI Blast for Windows with maximum *E*-value = 1×10^{-60}, minimum 95% identity of the SNP sequence) (Table S4). The QTL intervals of length *m* around the closest markers were projected onto the genomic sequence, with the use of SNPs corresponding to the interval boundaries; all genes located in projected intervals were listed and annotated using GO terms.

ACKNOWLEDGEMENTS

The work was supported by the European Regional Development Fund through the Innovative Economy Program for Poland 2007-2013, project POLAPGEN-BD no. WND-POIG.01.03.01-00-101/08. There are no conflicts of interest for any of the authors.

AUTHOR CONTRIBUTIONS

AP and PKa metabolomic assays, metabolite identification, interpretation, manuscript preparation. AS metabolomic data processing, manuscript preparation. PO, KM, AK, KK, KG and JG-W genetic map integration. PKr statistical data analysis, interpretation and manuscript preparation.

SUPPORTING INFORMATION

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

Figure S1. Comparative view of chromatographic UV absorbance profiles of Maresi and CamB recorded at 330 nm.

Figure S2. The differential correlation network for metabolites observed at T1 and at T2.

Figure S3. The network of metabolites linked to molecular simple sequence repeat and single nucleotide polymorphism markers constructed by metabolomic quantitative trait locus analysis at time point T2.

Table S1. Secondary metabolites in barley leaves: identification and statistical characteristics.

Table S2. Parameters and modules of correlation networks constructed for metabolites in control conditions and in drought at time points T1 and T2.

Table S3. Metabolic quantitative trait loci found at two time points T1 and T2 with their statistical characterization.

Table S4. Annotation of genes located in quantitative trait locus regions using Ensembl Plants Gene Ontology information.

Methods S1. Chromatographic data pre-processing; correlation networks.

REFERENCES

Acuña-Galindo, M.A., Mason, R.E., Subramanian, N.K. and Hays, D.B. (2015) Meta-analysis of wheat QTL regions associated with adaptation to drought and heat stress. *Crop Sci.* **55**, 477–492.

Agati, G., Matteini, P., Goti, A. and Tattini, M. (2007) Chloroplast-located flavonoids can scavenge singlet oxygen. *New Phytol.* **174**, 77–89.

Arora, A., Nair, M.G. and Strasburg, G.M. (1998) Structure activity relationship for antioxidant activities of series of flavonoids in liposomal system. *Free Rad. Biol. Med.* **24**, 1355–1363.

Cardot, P.J.P., Trolliard, P. and Tembely, S. (1990) A fully automated chromatographic peak detection and treatment software for multi-user multitask computers. *J. Pharm. Biomed. Anal.* **8**, 755–759.

Chen, F. and Dixon, R.A. (2007) Lignin modification improves fermentable sugar yields for biofuel production. *Nat. Biotechnol.* **25**, 759–761.

Cheynier, V., Comte, G., Davies, K.M., Lattanzio, V. and Martens, S. (2013) Plant Phenolics: Recent advances on their biosynthesis, genetics, and ecophysiology. *Plant Physiol. Biochem.* **72**, 1–20.

Clé, C., Hill, L.M., Niggeweg, R., Martin, C.R., Guisez, Y., Prinsen, E. and Jansen, M.A.K. (2008) Modulation of chlorogenic acid biosynthesis in Solanum lycopersicum; consequences for phenolic accumulation and UV-tolerance. *Phytochemistry*, **69**, 2149–2156.

Close, T.J., Bhat, P.R., Lonardi, S. *et al.* (2009) Development and implementation of high-throughput SNP genotyping in barley. *BMC Genomics*, **10**, 582.

Daszykowski, M., Van der Heyden, Y., Boucon, C. and Walczak, B. (2010) Automated alignment of one-dimensional chromatographic fingerprints. *J. Chromatogr. A*, **1217**, 6127–6133.

Ferreres, F., Gil-Izquierdo, A., Andrade, P.B., Valentão, P. and Tomás-Barberán, F.A. (2007) Characterization of C-glycosyl flavones O-glycosylated by liquid chromatography-tandem mass spectrometry. *J. Chromatogr. A*, **1161**, 214.

Ferreres, F., Andrade, P.B., Valentao, P. and Gil-Izquierdo, A. (2008) Further knowledge on barley (*Hordeum vulgare* L.) leaves O-glycosyl-C-glycosyl flavones by liquid chromatography-UV diode–array detection-electrospray ionisation mass spectrometry. *J. Chromatogr. A*, **1182**, 56–64.

Feuerstein, D., Parker, K.H. and Boutelle, M.G. (2009) Practical methods for noise removal: applications to spikes, nonstationary quasi-periodic noise, and baseline drift. *Anal. Chem.* **8**, 4987–4994.

Filek, M., Łabanowska, M., Kościelniak, J., Biesaga-Kościelniak, J., Kurdziel, M., Szarejko, I. and Hartikainen, H. (2014) Characterization of barley leaf tolerance to drought stress by chlorophyll fluorescence and electron paramagnetic resonance studies. *J. Agric. Crop Sci.* **201**, 228–240.

Gorzolka, K., Bednarz, H. and Niehaus, K. (2014) Detection and localization of novel hordatine-like compounds and glycosylated derivatives of hordatines by imaging mass spectrometry of barley seeds. *Planta*, **239**, 1321–1335.

Hendawey, M.H., Gharib, M.M., Marei, T.A. and Mohammed, A.S. (2014) Implication of phenolic compounds and amino acids in tolerance against net blotch disease in barley (*Hordeum vulgare* L.). *Glob. J. Biotechnol. Biochem.* **9**, 105–129.

Hernandez, I. and Van Breusegem, F. (2010) Opinion on the possible role of flavonoids as energy escape valves: novel tools for nature's Swiss army knife? *Plant Sci.* **179**, 297–301.

Hill, C.B., Taylor, J.D., Edwards, J., Mather, D., Bacic, A., Langridge, P. and Roessner, U. (2013) Whole-genome mapping of agronomic and metabolic traits to identify novel quantitative trait loci in bread wheat grown in a water-limited environment. *Plant Physiol.* **162**, 1266–1281.

Hochberg, U., Degu, A., Toubiana, D., Gendler, T., Nikoloski, Z., Rachmilevitch, S. and Fait, A. (2013) Metabolite profiling and network analysis reveal coordinated changes in grapevine water stress response. *BMC Plant Biol.* **13**, 184.

Hochberg, U., Batushansky, A., Degu, A., Rachmilevitch, S. and Fait, A. (2015) Metabolic and physiological responses of Shiraz and Cabernet Sauvignon (*Vitis vinifera* L.) to near optimal temperatures of 25 and 35 °C. *Int. J. Mol. Sci.* **16**, 24276–24294.

Hura, T., Hura, K. and Grzesiak, S. (2009) Possible contribution of cell-wall-bound ferulic acid in drought resistance and recovery in triticale seedlings. *Plant Physiol.* **166**, 1720–1733.

Hura, T., Hura, K., Dziurka, K., Ostrowska, A., Baczek-Kwinta, R. and Grzesiak, M. (2012) An increase in the content of cell wall-bound phenolics correlates with the productivity of triticale under soil drought. *J. Plant Physiol.* **169**, 1728–1736.

Ishihara, K. and Nakajima, N. (2003) Structural aspects of acylated plant pigments: stabilization of flavonoid glucosides and interpretation of their functions. *J. Mol. Catal. B Enzym.* **23**, 411–417.

Ishihara, A., Ogura, Y., Tebayashi, S. and Iwamura, H. (2002) Jasmonate-induced changes in flavonoid metabolism in barley (*Hordeum vulgare*) leaves. *Biosci. Biotechnol. Biochem.* **66**, 2176–2182.

Jiang, W., Zhang, Z.-M., Yun, Y.-H., Zhan, D.-J., Zheng, Y.-B., Liang, Y.-Z., Yang, Z.Y. and Yu, L. (2013) Comparisons of five algorithms for chromatogram alignment. *Chromatographia*, **76**, 1067–1078.

Kaplan, F., Kopka, J., Haskell, D.W., Zhao, W., Schiller, K.C., Gatzke, N., Sung, D.Y. and Guy, C.L. (2004) Exploring the temperature-stress metabolome of Arabidopsis. *Plant Physiol.* **136**, 4159–4168.

Kaspar, S., Matros, A. and Mock, H.P. (2010) Proteome and flavonoid analysis reveals distinct responses of epidermal tissue and whole leaves upon UV-B radiation of barley (*Hordeum vulgare* L.) seedlings. *J. Proteome Res.* **9**, 2402–2411.

Kirakosyan, A., Kaufman, P., Warber, S., Zick, S., Aaronson, K., Boilin, S., Chul Chang, S. (2004) Applied environmental stresses to enhance the levels of polyphenolics in leaves of hawthorn plants. *Physiol. Plant.* **121**, 182–186.

Kohyama, N. and Ono, H. (2013) Hordatine A ß-D-glucopyranoside from ungerminated barley grains. *J. Agric. Food Chem.* **61**, 1112–1116.

Kojima, M. and Kondo, T. (1985) An enzyme in sweet potato root which catalyzes the conversion of chlorogenic acid, 3-caffeoylquinic acid, to isochlorogenic acid, 3,5-dicaffeoylquinic acid. *Agric. Biol. Chem.* **49**, 2467–2469.

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

Leisso, R.S., Buchanan, D.A., Lee, J. *et al.* (2015) Chilling-related cell damage of apple (Malus × domestica Borkh.) fruit cortical tissue impacts antioxidant, lipid and phenolic metabolism. *Physiol. Plant.* **153**, 204–220.

Li, D.-M., Nie, Y.-X., Zhang, J., Yin, J.S., Li, Q., Wang, X.-J. and Bai, J.-G. (2013) Ferulic acid pretreatment enhances dehydration-stress tolerance of cucumber seedlings. *Biol. Plant.* **57**, 711–717.

Lisec, J., Meyer, R.C., Steinfath, M. *et al.* (2008) Identification of metabolic and biomass QTL in *Arabidopsis thaliana* in a parallel analysis of RIL and IL populations. *Plant J.* **53**, 960–972.

Maier, W., Peipp, H., Schmidt, J., Wray, V. and Strack, D. (1995) Cell levels of a terpenoid glycoside (Blumenin) and wall-bound phenolics in some cereal mycorrhizas. *Plant Physiol.* **109**, 465–470.

Malosetti, M., Ribaut, J.-M. and van Eeuwijk, F.A. (2013) The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. *Front. Physiol.* **4**, 44.

Masada, S., Terasaka, K., Oguchi, Y., Okazaki, S., Mizushima, T. and Mizukami, H. (2009) Functional and structural characterization of a flavonoid glucoside 1,6-glucosyltransferase from *Catharanthus roseus*. *Plant Cell Physiol.* **50**, 1401–1415.

Matsuda, F., Okazaki, Y., Oikawa, A., Kusano, M., Nakabayashi, R., Kikuchi, J., Yonemaru, J., Ebana, K., Yano, M. and Saito, K. (2012) Dissection of genotype–phenotype associations in rice grains using metabolome quantitative trait loci analysis. *Plant J.* **70**, 624–636.

Matsuda, F., Nakabayashi, R., Yang, Z., Okazaki, Y., Yonemaru, J.-I., Ebana, K., Yano, M. and Saito, K. (2014) Metabolome-genome-wide association study dissects genetic architecture for generating natural variation in rice secondary metabolism. *Plant J.* **81**, 13–23.

Mellou, F., Lazari, D., Skaltsa, H., Tselepis, A.D., Kolisis, F.N. and Stamatis, H. (2005) Biocatalytic preparation of acylated derivatives of flavonoid glycosides enhances their antioxidant and antimicrobial activity. *J. Biotechnol.* **116**, 295–304.

Mikołajczak, K., Ogrodowicz, P., Gudyś, K. *et al.* (2016) Quantitative trait loci for yield and yield-related traits in spring barley populations derived from crosses between European and Syrian cultivars. *PLoS ONE*, **11**, e0155938.

Mikołajczak, K., Kuczyńska, A., Krajewski, P. *et al.* (2017) Quantitative trait loci for plant height in Maresi × CamB barley population and their association with yield-related traits under different water regimes. *J. Appl. Genet.* **58**, 23–35.

Morreel, K., Goeminne, G., Storme, V. *et al.* (2006) Genetical metabolomics of flavonoid biosynthesis in *Populus*: a case study. *Plant J.* **47**, 224–237.

Mouradov, A. and Spangenberg, G. (2014) Flavonoids: a metabolic network mediating plants adaptation to their real estate. *Front Plant Sci.* **5**, 620.

Nakabayashi, R., Yonekura-Sakakibara, K., Urano, K. *et al.* (2014) Enhancement of oxidative and drought tolerance in Arabidopsis by overaccumulation of antioxidant flavonoids. *Plant J.* **77**, 367–379.

Neugart, S., Zietz, M., Schreiner, M., Rohn, S., Kroh, L.W. and Krumbein, A. (2012) Structurally different flavanol glycosides and hydroxycinnamic acid derivatives respond differently to moderate UV-B radiation exposure. *Physiol. Plant.* **145**, 582–593.

Nielsen, N.P.V., Carstensen, J.M. and Smedsgaard, J. (1998) Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping. *J. Chromatogr. A*, **805**, 17–35.

Nishiyama, T., Hagiwara, Y., Hagiwara, H. and Shibamoto, T. (1994) Inhibitory effect of 2″-O-glycosyl isovitexin and alpha-tocopherol on genotoxic glyoxal formation in a lipid peroxidation system. *Food Chem. Toxicol.* **32**, 1047–1051.

Ohkawa, M., Kinjo, J., Hagiwara, Y., Hagiwara, H., Ueyama, H., Nakamura, K., Ishikawa, R., Ono, M. and Nohara, T. (1998) Three new anti-oxidative saponarin analogs from young green barley leaves. *Chem. Pharm. Bull.* **46**, 1887–1890.

Peipp, H., Maier, W., Schmidt, J., Wray, V. and Strack, D. (1997) Arbuscular mycorrhizal fungus-induced changes in the accumulation of secondary compounds in barley roots. *Phytochemistry*, **44**, 581–587.

Piasecka, A., Sawikowska, A., Krajewski, P. and Kachlicki, P. (2015) Combined mass spectrometric and chromatographic methods for in-depth analysis of phenolic secondary metabolites in barley leaves. *J. Mass Spectrom.* **50**, 513–532.

Plaza, M., Pozzo, T., Liu, J., Gulshan Ara, K.Z., Turner, C. and Nordberg Karlsson, E. (2014) Substituent effects on *in vitro* antioxidizing properties, stability, and solubility in flavonoids. *J. Agric. Food Chem.* **62**, 3321–3333.

Pravdova, V., Walczak, B. and Massart, D.L. (2002) A comparison of two algorithms for warping of analytical signals. *Anal. Chim. Acta*, **456**, 77–92.

Romanenko, S.V., Stromberg, A.G. and Pushkareva, T.N. (2006) Modeling of analytical peaks: peaks properties and basic peak functions. *Anal. Chim. Acta*, **580**, 99–106.

Rowe, H.C., Hansen, B.G., Halkier, B.A. and Kliebenstein, D.J. (2008) Biochemical networks and epistasis shape the *Arabidopsis thaliana* metabolome. *Plant Cell*, **20**, 1199–1216.

Schauer, N., Semel, Y., Roessner, U. *et al.* (2006) Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nat. Biotechnol.* **24**, 4.

Schmitz-Hoerner, R. and Weissenböck, G. (2003) Contribution of phenolic compounds to the UV-B screening capacity of developing barley primary leaves in relation to DNA damage and repair under elevated UV-B levels. *Phytochemistry*, **64**, 243–255.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504.

Stochmal, A. and Oleszek, W. (2007) Effect of acylation of flavones with hydroxycinnamic acids on their spectral characteristics. *Nat. Prod. Commun.* **2**, 571–574.

Tomasi, G., van den Berg, F. and Andersson, C. (2004) Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. *J. Chemom.* **18**, 231–241.

Tomasi, G., Savorani, F. and Engelsen, S.B. (2011) icoshift: an effective tool for the alignment of chromatographic data. *J. Chromatogr. A*, **1218**, 7832–7840.

Toubiana, D., Batushansky, A., Tzfadia, O., Scossa, F., Khan, A., Barak, S., Zamir, D., Fernie, A.R., Nikoloski, Z. and Fait, A. (2015) Combined correlation-based network and mQTL analyses efficiently identified loci for branched-chain amino acid, serine to threonine, and proline metabolism in tomato seeds. *Plant J.* **81**, 121–133.

Turner, L.B., Cairns, A.J., Armstead, I.P., Thomas, H., Humphreys, M.W. and Humphreys, M.O. (2008) Does fructan have a functional role in physiological traits? Investigation by quantitative trait locus mapping. *New Phytol.* **179**, 765–775.

Vivó-Truyols, G., Torres-Lapasió, J.R., Van Nederkassel, A.M., Van der Heyden, Y. and Massart, D.L. (2005) Automatic program for peak detection and deconvolution of multi-overlapped chromatographic signals: part I: peak detection. *J. Chromatogr. A*, **1096**, 133–145.

VSN International (2015) Genstat for Windows 18th Edition. VSN International, Hemel Hempstead, UK, Genstat.co.uk.

Wang, S.Y., Ho, T.J., Kuo, C.H. and Tseng, Y.J. (2010) Chromaligner: a web server for chromatogram alignment. *Bioinformatics*, **26**, 2338–2339.

Wong, J.W., Durante, C. and Cartwright, H.M. (2005) Application of fast Fourier transform cross-correlation for the alignment of large chromatographic and spectral datasets. *Anal. Chem.* **77**, 5655–5661.

Xu, Y. (2010) *Molecular Plant Breeding*. Wallingford, UK: CAB International.

Zhang, Z.M., Liang, Y.Z., Lu, H.M., Tan, B.B., Xu, X.-N. and Ferro, M. (2012) Multiscale peak alignment for chromatographic datasets. *J. Chromatogr. A*, **1223**, 93–106.