Plant Biotechnology Journal

Plant Biotechnology Journal (2016) **14**, pp. 649–660 doi: 10.1111/pbi.12410

Metabolomic profiling and genomic analysis of wheat aneuploid lines to identify genes controlling biochemical pathways in mature grain

Michael G. Francki[1,2,*], Sarah Hayton[3], Joel P. A. Gummer[3,4,5], Catherine Rawlinson[3,5] and Robert D. Trengove[3,4,5]

[1]*Department of Agriculture and Food Western Australia, Grains Industry, South Perth, WA, Australia*
[2]*State Agricultural Biotechnology Centre, Murdoch University, Murdoch, WA, Australia*
[3]*Separation Science and Metabolomics Laboratory, Research and Development, Murdoch University, Murdoch, WA, Australia*
[4]*School of Veterinary and Life Sciences, Murdoch University, Murdoch, WA, Australia*
[5]*Metabolomics Australia, Murdoch University Node, Murdoch, WA, Australia*

Received 5 March 2015;
revised 1 May 2015;
accepted 5 May 2015.
**Correspondence* (Tel +61 8 93607575;
fax +61 8 93682958;
email michael.francki@agric.wa.gov.au)

Summary

Metabolomics is becoming an increasingly important tool in plant genomics to decipher the function of genes controlling biochemical pathways responsible for trait variation. Although theoretical models can integrate genes and metabolites for trait variation, biological networks require validation using appropriate experimental genetic systems. In this study, we applied an untargeted metabolite analysis to mature grain of wheat homoeologous group 3 ditelosomic lines, selected compounds that showed significant variation between wheat lines Chinese Spring and at least one ditelosomic line, tracked the genes encoding enzymes of their biochemical pathway using the wheat genome survey sequence and determined the genetic components underlying metabolite variation. A total of 412 analytes were resolved in the wheat grain metabolome, and principal component analysis indicated significant differences in metabolite profiles between Chinese Spring and each ditelosomic lines. The grain metabolome identified 55 compounds positively matched against a mass spectral library where the majority showed significant differences between Chinese Spring and at least one ditelosomic line. Trehalose and branched-chain amino acids were selected for detailed investigation, and it was expected that if genes encoding enzymes directly related to their biochemical pathways were located on homoeologous group 3 chromosomes, then corresponding ditelosomic lines would have a significant reduction in metabolites compared with Chinese Spring. Although a proportion showed a reduction, some lines showed significant increases in metabolites, indicating that genes directly and indirectly involved in biosynthetic pathways likely regulate the metabolome. Therefore, this study demonstrated that wheat aneuploid lines are suitable experimental genetic system to validate metabolomics–genomics networks.

Keywords: wheat, metabolomics, genomics, aneuploidy, seed, quality.

Introduction

Detailed knowledge of biological processes can significantly enhance our ability to manipulate desirable phenotypes for crop improvement. Small molecules resulting from metabolism (i.e. metabolites) are an important link between genes and phenotypes as they represent a nearer biological end point to the desired trait than either genes or their encoded protein (Fiehn, 2002; Hall, 2006). The complement of metabolites (metabolome) in any particular plant tissue has the potential to provide a diagnostic and predictive phenotypic trait value. However, changes of the metabolome during plant growth and in response to environmental signals (Hall *et al.*, 2006; Schauer and Fernie, 2006) may resultantly render subsequent comparison of profiles between genotypes unrelated to the genetic differences, but rather reveal metabolites more closely correlated with physiological differences and environmental responses. Therefore, metabolite profiles in mature grain would be most suitable in this regard as there are no further developmental changes within the plant, and association of compounds to phenotypic traits can be made specifically to genotype and environment responses.

Metabolomics is an impartial technology and, when integrated with complementary disciplines, contributes towards the interpretation of interconnecting biological processes associated with phenotypes. Metabolic profiling therefore is becoming an increasingly popular tool in functional genomics (Bino *et al.*, 2004). The co-occurrence of gene transcripts and small-molecule metabolites (associated with the metabolomics discipline) provides a basis for generating data-driven theoretical models of biological networks (Saito and Matsuda, 2010; Saito *et al.*, 2008; Yuan *et al.*, 2008). Although 'guilt by association' of transcripts and metabolites is a widely accepted principle for assuming gene function, the active state of proteins through post-translational modification or the influence of substrates or cofactors by undisclosed interconnecting genes and biological pathways may have a significant influence in metabolite variation (Fridman and Pichersky, 2005; Saito *et al.*, 2008). To this end, a holistic analysis of genes involved in discrete primary and interconnecting secondary pathways and their interplay would significantly

contribute towards understanding the biological networks regulating the plant metabolome and phenotypic variation.

The most comprehensive reconstruction and modelling of a metabolic network based on multi-omics approach was achieved in the filamentous fungus, *Aspergillus niger*, providing a detailed understanding of genes regulating metabolism and new information on physiological traits (Andersen *et al.*, 2008). However, integration of 'omics technologies in crop species is not as well advanced, posing the next major experimental challenge to model metabolic networks that give rise to phenotypes. The genomes of major crop species have been sequenced or are in the process of completion (Feuillet *et al.*, 2011), paving the way to develop the resources needed towards understanding the link between genes of interconnecting biological pathways with phenotypes. Allohexaploid bread wheat ($2n = 6\times = 42$, genomes AABBDD) is one of the more complex crop genomes whereby similar genes on homoeologous chromosomes could pose a significant challenge in reconstructing biological networks. An ordered draft wheat genome sequence has recently been completed with >124 000 gene loci distributed across all chromosomes of the A, B and D subgenomes (International Wheat Genome Sequencing Consortium, 2014) that will assist in identifying genes controlling biological processes and metabolite abundances responsible for phenotypes and trait variation. A preliminary analysis of gene content neither showed a bias in gene composition nor transcription wide global dominance by any particular subgenome but, rather, each had a higher degree of regulatory and transcriptional autonomy (The International Wheat Genome Consortium (IWGSC), 2014). The availability of the draft sequence of the wheat genome enabled the analysis of gene interaction in wheat grain. Expression analysis of a subset of genes confirmed a lack of global dominance from any of the subgenomes during wheat grain development, but rather cell type- and stage-dependent genome dominance, with inter- and intragenomic regulation of gene expression (Pfeifer *et al.*, 2014). Therefore, transcript accumulation is a result of the interplay between subgenomes and amongst individual cell types giving rise to a particular function, confirming the complex regulation of gene expression adding to the multifaceted processes of interacting biological pathways leading to phenotypes in wheat grain.

Data-driven theories of metabolic networks require validation by forward or reverse genetics (Saito *et al.*, 2008), but often the chosen experimental system neither ratifies nor refutes existing hypotheses on key genetic determinants controlling biological processes. Modifying transcript expression using transgenics or mutations, for instance, may not affect metabolite or trait variation if a targeted gene identified from theoretical models is not a rate-limiting step in the primary biochemical pathway or, indeed, other interconnecting biological pathways affect transcriptional regulation or post-translational modification. In this regard, experimental genetic systems capable of simultaneously associating large sets of genes with a metabolite profile are preferred to validate known and discover new interconnecting biological pathways that determine the metabolome. The polyploid nature of bread wheat genome is a particularly unique genetic system as it can tolerate substantial chromosomal aberrations without compromising plant survival, allowing analysis of phenotypic changes caused by multiple gene loss. In particular, aneuploid lines with missing chromosome arms (ditelosomics) or smaller deleted segments have been well characterized (Endo and Gill, 1996) and shown to be useful in identifying genes controlling phenotypes (Erayman *et al.*, 2004). Therefore, metabolic profiling of ditelosomic and deletion lines could provide a powerful genetic system and an appropriate supporting tool to identify genes responsible for primary and interconnecting biological pathways that corroborate networks controlling metabolite accumulation.

Metabolite profiling of mature wheat grain has identified both polar and nonpolar compounds (Bellegia *et al.*, 2013; Lee *et al.*, 2013; Matthews *et al.*, 2012) with significant differences between durum and bread wheat (Matthews *et al.*, 2012). Therefore, it appears that polyploidy has a significant effect on accumulation and composition of metabolites in wheat grain. However, the contribution of the gene content of each chromosome is yet to be realized, although the use of aneuploid lines as a genetic system could greatly enhance our ability to understand and validate interconnecting networks that control the accumulation of compounds of the wheat metabolome. The general aim of this study therefore was to determine the feasibility of using wheat aneuploid lines to identify genes of biological pathways that control the accumulation of metabolites in mature wheat grain. Specifically, this study aimed to (i) develop an untargeted metabolite profile of mature grain of wheat with a full complement of chromosomes and compare metabolite content and composition from selected wheat ditelosomic lines; (ii) identify known biochemical pathways in wheat and ditelosomic lines and interrogate the draft wheat genome sequence to reveal underlying genes that control metabolite accumulation; and (iii) assess the suitability of aneuploid lines as a genetic system to validate genes controlling the abundance and composition of metabolites in mature seed. The study focused on mature grain of Chinese Spring (the genotype from which aneuploid lines were derived) and ditelosomic lines where genes on the short and long arms of A, B and D genomes of homoeologous group 3 chromosomes were deleted, and the resultant effects on seed metabolites. The outcome of this study will determine whether wheat aneuploid lines are appropriate to identify and validate genes controlling metabolite accumulation in mature grain that could be used to manipulate grain quality traits.

Results

Metabolite profiling of mature seed of Chinese Spring and ditelosomic lines

The analyses resolved a total of 412 analytes between the metabolite profiles of mature wheat seed and ditelosomic lines from homoeologous group 3 chromosomes (see Table S1 for metabolite specification). PCA described 53% of the data variance between principal components one and two (Figure 1; 33% and 20% for PC1 and PC2, respectively). The clustering of individual replicates ($n = 4$–5) observed within genotypes, relative to that observed between genotypes, identified distinct groupings correlated to the different genetic backgrounds. Of most interest were those differences in metabolite profiles between the Chinese Spring and ditelosomic lines; however, it was between Chinese Spring, DT3AL, DT3BS and DT3BL that the PCA model described the least variation amongst metabolites (Figure 1). PC1 described the ditelosomic line DT3DS as most distinct from the remaining ditelosomics or Chinese Spring. PC2 described the ditelosomic lines DT3AS and DT3DL to be the next most distinct, describing the difference between these two lines and all others, including Chinese Spring. Overall, the model identified the ditelosomic lines DT3AS, DT3DS and DT3DL as having the largest

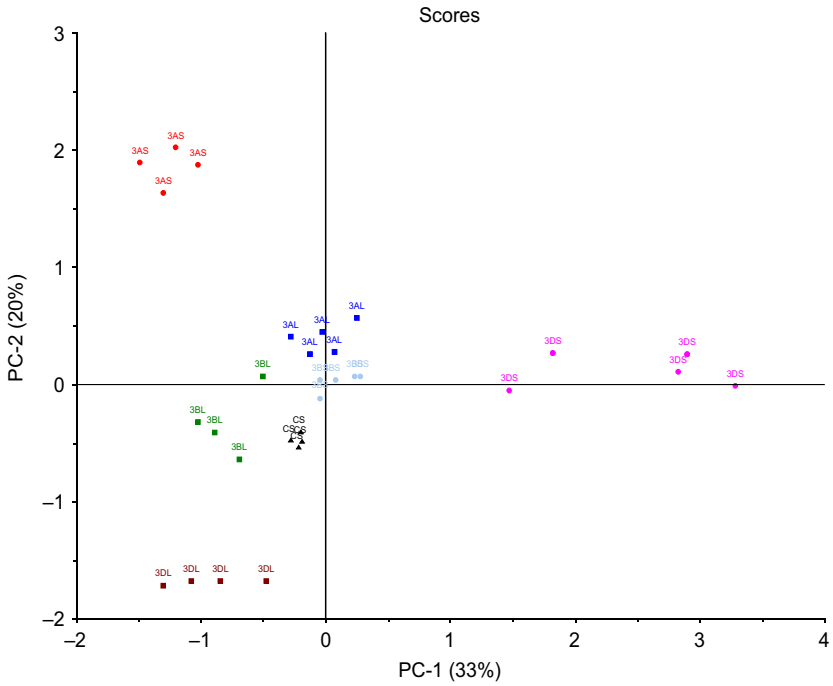

Figure 1 Principal component analysis plot of metabolites for wheat cv Chinese Spring (CS) and genotypes ditelosomic for 3AS, 3AL, 3BS, 3BL, 3DS, 3DL (n = 4–5). The total metabolite variation (53%) represented in the first and second component is 33% and 20%, respectively.

differences between metabolite profiles (Figure 1), indicative of the genes on chromosomes 3AL, 3DL and 3DS, respectively, having a major influence on the variation in polar metabolite composition of mature seed.

The deconvoluted mass spectra from the profiles of Chinese Spring and each ditelosomic lines were compared against the mass spectra and chromatographic features of authentic metabolite standards in combination with the NIST 2011 mass spectral library, to identify individual compounds in mature seed. All identities were determined by a match score of 80% [as scored for uracil (2TMS_18.47_1342)] or above. Succinic acid (2TMS_17.92_1321) and tyrosine (3TMS_30.86_1934) scored the highest match of 96%. The score reflected the relative similarity of ion intensities of the metabolite to that of the library entry. The MS library matches revealed a total of 55 identified metabolite features (Table 1) from the 412 measured analytes (Table S1). Those identified were categorized into one of seven major metabolite classes including amino acids, amino alcohols, fatty acids, nucleosides, organic acids, sugars and sugar alcohols, with the largest number of identified compounds classed as amino acids (Table 1). The comparative metabolite abundances between lines were expressed as a significant ($P < 0.01$) fold change difference to the same compound identified in Chinese Spring. Eighty per cent of the identified metabolites show a significant difference between at least one ditelosomic line and Chinese Spring (Table 1), indicating that genes on the short and long arm of homoeologous group 3 have a major effect on the metabolite profile of mature wheat grain. The significant differences in biochemical profiles between Chinese Spring and ditelosomic lines provided a means to identify putative genes controlling variation in metabolites. Specific metabolites having significant fold -change difference between Chinese Spring and ditelosomic lines or that share common biochemical pathways were selected for further investigation for underlying genes controlling metabolite accumulation.

Metabolite profile and putative genes controlling trehalose accumulation

The MST trehalose (8TMS_42.71_2728) was selected for further analysis as it had the highest fold change difference of all metabolites, with a significant two-fold decrease for DT3BS ($P < 0.01$) and a highly significant 11-fold increase detected for DT3DS ($P < 0.001$) compared to Chinese Spring (Table 1). Consistently, the (non-scaled) chromatograms showed an accumulation of trehalose in DT3DS relative to Chinese Spring and the remaining ditelosomic lines (Figure 2). Trehalose is controlled by a relatively simple biochemical pathway involving three enzymatic steps where UDP-glucose and glucose-6-phosphate are substrates for conversion to trehalose-6-phosphate by trehalose-6-phosphate synthase (TPS) which, in turn, is used to synthesize trehalose through trehalose-6-phosphate phosphatase (TPP) activity (Figure 3). Furthermore, trehalose is converted to form two glucose molecules by trehalase (Figure 3), and putative wheat genes encoding the three enzymes were searched in the wheat genome survey sequence. For comparative purposes, the identification of wheat cDNAs encoding TPS, TPP and trehalase was revealed by TBLASTX analysis using annotated FL-cDNAs from Arabidopsis and rice as query sequences. Full-length wheat cDNA sequences with e-values $<4e^{-105}$ were identified, including two with homology to TPS, one cDNA encoding trehalase and six with significant homology to TPP (Table 2). BLASTN analysis of the draft wheat genome sequence using wheat FL-cDNA as query sequences identified four copies of TPS and three copies of trehalase-related genes on the long arm of homoeologous chromosomes 1 and 5 (Table 2). Genes related to TPP, however, represented a larger multigene family, consisting of at least 13 copies with genes located on homoeologous chromosomes 1, 5 and 6 in addition to copies located on 3AL, 3BL and 3DL (Table 2). Of the ditelosomic lines that lacked the TPP-related sequences, DT3BS showed the expected decrease in metabolite levels (Table 1), indicating that TPP on 3BL may be a

Table 1 Summary of identified metabolites detected in mature wheat grain

Metabolite	DT3AS	DT3AL	DT3BS	DT3BL	DT3DS	DT3DL
Unidentified pentose sugars						
Unknown aldopentose, 5 TMS, 26.79, 1724	1.15	0.98	0.57**	0.96	3.30**	0.85
Unknown aldopentose, 5 TMS, 26.83, 1724	1.06	0.96	1.05	0.86	4.10**	1.60**
Amino acid						
Aspartic acid, 3 TMS, 22.67, 1516	0.30*	0.66	0.55	0.70	2.40	0.28
GABA, 3 TMS, 22.87, 1527	4.88**	0.62	0.61	0.92	1.41	1.55
Glutamic acid, 3 TMS, 24.79, 1623	0.74	1.25	1.04	1.10	2.26	1.07
Glycine, 2 TMS, 12.72, 1110	0.91	0.77	0.60**	0.66*	0.95	0.84
Glycine, 3 TMS, 17.63, 1308	1.31*	0.75	0.60**	0.79	0.81	0.76*
Isoleucine 1 TMS, 14.36, 1175	0.89	0.76	0.48*	0.54*	0.81	0.46*
Alanine, 2 TMS, 12.11, 1085	1.64**	0.67*	0.65	0.92	1.03	0.83
Glutamic acid, 2 TMS, 22.7, 1519	1.36**	0.81*	0.62**	0.72**	0.92	1.01
Glutamine, 3 TMS, 27.89, 1777	1.75	2.07	1.70	0.85	4.90	5.69
Isoleucine, 2 TMS, 17.32, 1295	1.11	0.61*	0.46**	0.77	1.07	0.41**
Leucine, 1 TMS, 13.94, 1159 (putative)	1.01	0.70	0.53*	0.58*	0.84	0.48*
Methionine, 1 TMS, 20.27, 1416	4.00**	3.02**	1.06	2.41**	0.58	1.57*
Phenylalanine, 1 TMS, 23.37, 1550	1	2.23	0.63	0.71	0.80	0.65
Phenylalanine, 2 TMS, 24.97, 1630	0.92	1.53	0.85	0.86	2.31	0.54
Proline, 2 TMS, 17.43, 1300	2.03**	0.61	0.36*	0.69	0.79	0.95
Proline, × TMS, 24.06, 1585	1.49	0.77	0.42*	0.52	0.85	0.86
Threonine, 3 TMS, 19.59, 1387	1.4	0.86	0.81	1.22	1.96	0.95
Tryptophan, 1 TMS, 34.84, 2172	0.47	1.05	0.84	0.24	3.89	0.01*
Tyrosine, 3 TMS, 30.86, 1934	0.71	0.77	0.77	0.56*	1.61	0.63
Valine, 1 TMS, 12.13, 1085	0.98	0.71	0.47*	0.47*	0.7	0.53*
Serine, 2 TMS, 16.43, 1260	2.32**	1.12	0.92	1.23	0.66	0.91
Serine, 3 TMS, 18.98, 1363	3.11**	0.97	0.86	1.87*	1.21	0.79
Amino alcohol						
Ethanolamine, 3 TMS, 16.6, 1266	1.57**	0.99	1.15	1.16	1.03	0.91
Fatty acid						
Arachidic acid, 1 TMS, 38.86, 2444	1.33	1.09	0.97	0.98	1.06	1.1
Stearic acid, 1 TMS, 35.95, 2244	1.52	1.08	1.06	1.2	0.88	1.07
Nucleoside						
Adenosine, × TMS, 40.95, 2603	0.99	0.62	0.31*	0.68	1.42	0.6
Uracil 2 TMS, 18.47, 1342	1.16	0.46**	0.47**	0.48**	0.62**	0.57**
Uridine, 3 TMS, 38.98, 2462	1.32	0.96	0.64**	0.81	1.52*	0.58**
Organic acids						
Azelaic acid, 2 TMS, 28.33, 1801	0.99	0.87	0.59*	0.59*	0.79	0.54*
Benzoic acid, 1 TMS, 16.26, 1254	0.52	0.85	0.58	0.48	1.24	0.76
Citric acid, 4 TMS, 28.69, 1817	0.59*	0.46**	0.35**	0.30**	0.65*	0.33**
Gluconic acid, 6 TMS, 31.84, 1989	2.50*	1.48*	0.81*	1.05	0.94	1.36
Malonic acid, 2 TMS, 15.13, 1208	1.09	1.06	0.94	1.21	1.31*	1.1
Oxalic acid, 2 TMS, 13.13, 1125	2.87	6.45	0.96	0.8	1.17	0.77
Quinic acid, × TMS, 29.38, 1851	6.76**	1.35*	0.93	2.29**	0.46**	0.89*
Shikimic acid, 28.47, 1809	2.48**	0.98	0.94	1.1	0.55**	0.65
Succinic acid, 2 TMS, 17.92, 1321	0.70**	0.82*	0.71**	0.70**	0.95	0.64**
Fumaric acid, 2 TMS, 18.29, 1357	0.91	1.13	0.66*	0.87	1.07	0.70*
Others						
Tocopherol, 1 TMS, 47.69, 3141	0.96	0.99	0.93	0.82	0.74**	0.9
Squalene, 43.91, 2818	1.2	1.33	1.26	1.17	1.02	1.06
Sugars						
Cellobiose, × TMS, 42.6, 2721	0.64	1.03	0.73	0.8	1.37	1.07
Fructose, 5 TMS, MEOX, 29.60, 1862	2.56**	1.28*	1.38**	1.42**	0.78**	0.82*
Fructose, 5 TMS, MEOX, 29.78, 1871	3.14**	1.47**	1.58**	1.71**	0.74**	0.80*
Ribose, 4 TMS, MEOX, 25.89, 1678	1.68**	1.14	1.15	0.83	1.29	0.74
Glucose, 5 TMS, MEOX, 30.05, 1885	5.19**	1.58**	1.64**	2.24**	0.88	0.94
Mannose, 5 TMS, MEOX, 30.12, 1889	6.86**	1.63*	1.63*	2.39**	0.81	0.99
Trehalose, 8 TMS, 42.71, 2728	1.48	0.93	0.56*	0.67	11.60**	0.91

Table 1 Continued

Metabolite	DT3AS	DT3AL	DT3BS	DT3BL	DT3DS	DT3DL
Xylose, 4 TMS, 25.42, 1656	1.67**	1.14	1.09	1.14	1.59*	0.71*
Stachyose, × TMS, 61.39, 4464	0.82	1.19	1.09	1.05	1.35	0.76
Sucrose, 8 TMS, 41.32, 2630	1.05	1.25	0.93	0.91	1.09	0.98
Sugar alcohols						
Mannitol, 6 TMS, 30.6, 1915	1.46*	1.59**	1.04	1.34	4.96**	1.90**
Myo-inositol, 6 TMS, 33.38, 2081	1.03	0.59**	0.52**	0.60**	1.03	0.46**
Scyllo-inositol, 6 TMS, 32.28, 2020	0.60*	0.8	0.58**	0.75*	1.29*	1.02

The fold change difference in metabolite accumulation compared to Chinese Spring is shown for each ditelosomic (DT) line with significant ($P < 0.01$) and highly significant ($P < 0.001$) differences indicated by * and **, respectively.

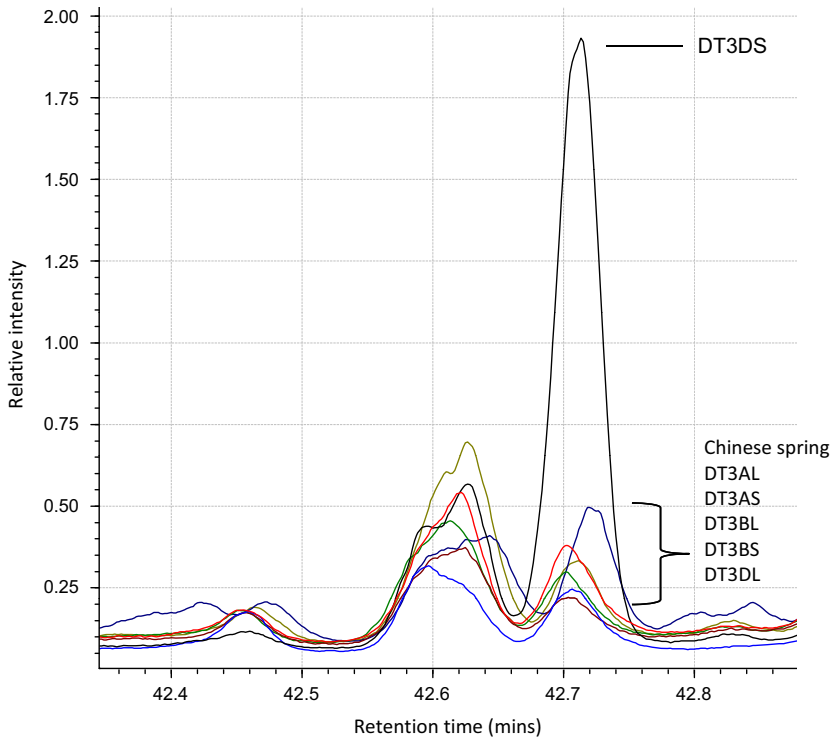

Figure 2 Total ion chromatogram overlays for trehalose 8TMS_42.71_2728 between a representative replicate for Chinese Spring and each ditelosomic line

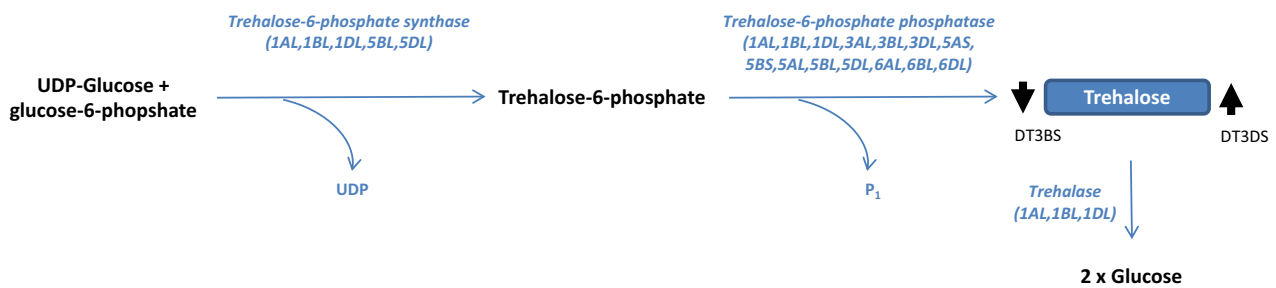

Figure 3 Schematic diagram of the biochemical pathway for trehalose accumulation. Blue box highlights trehalose detected in the untargeted analysis with black arrows indicating increase and decrease in trehalose for DT3BS and DT3DS, respectively. Enzyme names are shown in blue with the chromosomal location of corresponding wheat genes shown in parentheses.

rate-limiting step in trehalose accumulation. However, the absence of either TPP or trehalase did not show a similar effect in DT3AS and DT3DS, indicating that TPP genes on 3AL and 3DL may have an alternative role other than trehalose accumulation in mature grain (Figure 3). On the contrary, the 11-fold increase in trehalose for DT3DS was unlikely to be attributed to any gene in the primary biochemical pathway for trehalose, indicating that other unknown genes of intercon-

Table 2 Summary of rice and wheat FL-cDNA sequences with annotation and amino acid identity to enzymes of the trehalose and branched-chain amino acid biosynthetic pathways. The chromosomal locations of wheat cDNA are based on identity with DNA sequences in the survey sequence from International Wheat Genome Sequencing Consortium (IWGSC)

Metabolite	Enzyme	Annotated FL-cDNA*	Wheat FL-cDNA	TBLASTX e-value	Wheat chromosome locations
Trehalose	Trehalose-6-phosphate synthase	Y08568 (*A.t.*)*	FJ167677	e = 0.0	1AL, 1BL, 1DL, 5DL
		AF370287 (*A.t.*)	AK331389	e = 0.0	1AL, 1BL, 1DL, 5DL
		AY063055 (*A.t.*)	FJ167677	e = 0.0	1AL, 1BL, 1DL, 5DL
		AK103775	AK331389	e = 0.0	1AL, 1BL, 1DL, 5DL
			FJ167677	e = 0.0	1AL, 1BL, 1DL, 5DL
			AK331389	e = 0.0	1AL, 1BL, 1DL, 5DL
			FJ167677	e = $2e^{-106}$	1AL, 1BL, 1DL, 5BL
			AK331389	e = $4e^{-105}$	1AL, 1BL, 1DL, 5BL
	Trehalose-6-phosphate phosphatase	AK072132	AK333853	e = 0.0	1AL, 1BL, 1DL, 3AL, 3BL, 3DL, 5AS, 5BS, 5BL, 5DL
			AK334843	e = 0.0	1AL, 1BL, 1DL, 3AL, 3BL, 3DL, 5AS, 5BS, 5BL, 5DL
			FN564426	e = 0.0	1AL, 1BL, 1DL, 3AL, 3BL, 3DL, 5AS, 5BS, 5BL, 5DL
			AK332212	e = 0.0	1AL, 1DL, 3AL, 3BL, 5AL, 5BS, 5BL, 5DL
			AK331757	e = 0.0	1AL, 1BL, 1DL
			BT009244	e = 0.0	6AL, 6BL, 6DL
	Trehalase	BT010732 (*A.t*)	AK331310	e = $2e^{-177}$	1AL, 1BL, 1DL
		AK108163	AK331310	e = 0.0	1AL, 1BL, 1DL
Aspartate	Asparagine synthetase	D83378	AK333183	e = 0.0	1AL,1BL,1DL
			AY621539	e = 0.0	5AL,5BL,5DL
			AK334107	e = 0.0	5AL,5BL,5DL
			BT009245	e = 0.0	5AL,5BL,5DL
			BT009049	e = 0.0	3AS,3DS
Glutamate	Aspartate transaminase	AK069075	AK331565	e = $3e^{-150}$	1AS,1BS,1DS
		AK067732	AK331959	e = $3e^{-150}$	1AS,1BS,1DS
			AK333562	e = $3e^{-127}$	5AS,5BS,5DS
			AK333743	e = $2e^{-124}$	5AL,5L,5DL
		AK068200	BT009428	e = $5e^{-180}$	1AS,1BS,1DS
		AK103586	AK332497	e = 0.0	3AL,3BL,3DL
			BT009009	e = 0.0	3AL,3BL,3DL
			EU346759	e = 0.0	3AL,3BL,3DL
			AK332709	e = $1e^{-144}$	6AL,6BL,6DL
			EU885207	e = $1e^{-136}$	6BL,6DL
			AK333705	e = $7e^{-136}$	6AS,6BS,6DS
Methionine	Aspartate kinase	AK121930	AK333665	e = 0.0	4AL,5BL,5DL
		AK073189	BT009484	e = 0.0	4AL,5BL,5DL
			AK334445	e = 0.0	3AL,3BL,3DL
	Aspartate semialdehyde dehydrogenase	AK060701	BT008970	e = 0.0	5AL,5BL,5DL
			BT009463	e = 0.0	5AL,5BL,5DL
	Homoserine dehydrogenase	AK068391	AK335256	e = 0.0	2AL,2BL,2DL
	Homoserine kinase	AK060519	AK333708	E = $7e^{-136}$	4AS,5AL,5BL,5DL
	Cystathionine-γ-synthase	NM_001071075	AK335253	e = $3e^{-137}$	7AS,7BS,7DS
	Cystathionine-β-lyase	NM_001063486	AK335253	e = 0.0	7AS,7BS,7DS
			BT009509	e = 0.0	7AS,7BS,7DS
	Methionine synthase	AF439723 (Z.m.)	AK335562	e = 0.0	4AL,4DS,4BS,6BS,5DS,5BS,5AS
			BT009353	e = 0.0	4AL,4DS,4BS,6BS,5DS,5BS
			AK335485	e = 0.0	5DS,5BS,5AS,4BS,4DS,4AL,6B
Threonine	Threonine synthase	AK101669	AK330620	e = 0.0	1AL,3AL,3BL,3DL
Isoleucine/Valine	Threonine dehydratase	XM_006650431	tplb0062e09	e = 0.0	4AL,4DS,4BS,5AS,5DS
	Acetolactate synthase	AB049823	AY210406	e = 0.0	6BS,6AL,6BL,6DL
	Ketoacid reductoisomerase	AK072075	BT009123	e = 0.0	1AL
		AK065295	BT009123	e = 0.0	1AL
		AK061892	BT009123	e = 0.0	1AL
	Dihydroxy acid dehydratase	AK102083	AK335234	e = 0.0	7AS,7BS,7DS
	Amino acid aminotransferase	AK120579	AK335425	e = $3e^{-123}$	1AL,1BL,1DL
			AK330986	e = $8e^{-113}$	2AL, 2BL,2DL

Table 2 Continued

Metabolite	Enzyme	Annotated FL-cDNA*	Wheat FL-cDNA	TBLASTX e-value	Wheat chromosome locations
			BT009368	$e = 8e^{-112}$	2AL, 2BL,2DL
		AK108687	AK335425	e = 0.0	1AL,1BL,1DL
			AK330986	$e = 2e^{-176}$	2AL, 2BL,2DL
			BT009368	$e = 2e^{-174}$	2AL, 2BL,2DL
		AK106376	AK330986	e = 0.0	2AL, 2BL,2DL
			BT009368	e = 0.0	2AL, 2BL,2DL
			AK335425	$e = 4e^{-165}$	1AL,1BL,1DL
Alanine	Alanine aminotransferase	AK107237	AK333743	e = 0.0	5AL,5BL,5DL
			AK331959	e = 0.0	1AS,1BS,1DS
			AK331565	e = 0.0	1AS,1BS,1DS
			AK333562	$e = 9e^{-170}$	5AS,5BS,5DS
		AK119373	AK333743	e = 0.0	5AL,5BL,5DL
			AK331959	e = 0.0	1AS,1BS,1DS
			AK331565	e = 0.0	1AS,1BS,1DS
			AK333562	$e = 9e^{-170}$	5AS,5BS,5DS
Leucine	Isopropylmalate synthase	AK066890	AK332549	e = 0.0	5AL,5BL,5DL
		AK243491	AK332549	e = 0.0	5AL,5BL,5DL
	Isopropylmalate isomerase	NM_129871 (A.t.)	BT009140	$e = 6e^{-100}$	6AL,6BL,6DL
	Isopropylmalate dehydrogenase	AK059596	BT009215	e = 0.0	2AL,2BL,2DL,6AL,6BL,6DL
			AK331720	e = 0.0	2AL,2BL,2DL,6AL,6BL,6DL
			BT009114	e = 0.0	2AL,2BL,2DL,6AL,6BL,6DL
			AK331640	e = 0.0	1BL,2AL,2BL,2DL,6DL
		AK120254	AK334888	e = 0.0	2AL,2BL,2DL
			BT009017	$e = 1e^{-126}$	2AL,2BL,2DL
	Amino acid aminotransferase	AK120579	AK335425	$e = 3e^{-123}$	1AL,1BL,1DL
			AK330986	$e = 8e^{-113}$	2AL, 2BL,2DL
			BT009368	$e = 8e^{-112}$	2AL, 2BL,2DL
		AK108687	AK335425	e = 0.0	1AL,1BL,1DL
			AK330986	$e = 2e^{-176}$	2AL, 2BL,2DL
			BT009368	$e = 2e^{-174}$	2AL, 2BL,2DL
		AK106376	AK330986	e = 0.0	2AL, 2BL,2DL
			BT009368	e = 0.0	2AL, 2BL,2DL
			AK335425	$e = 4e^{-165}$	1AL,1BL,1DL

*GenBank Accession Numbers from *Oryza sativa*, *Arabidopsis thaliana* (*A.t.*) or *Zea mays* (*Z.m.*)

necting pathways play a significant role in controlling trehalose in mature grain.

Variation for branched-chain amino acids and associated genes

Branched-chain amino acids in ditelosomic lines were selected for further analysis because of their variable metabolite profiles (Table 1) and interconnecting biochemical pathways that link genes regulating amino acid accumulation. Aspartate is the precursor for methionine and threonine and isoleucine, whereas valine, alanine and leucine are derived from a common precursor, pyruvate (Figure 4). Although the aspartate-derived amino acids are known to be amenable to the analytical methods, only a putative identity could be given to leucine. The putative leucine (1TMS_13.94_1159) matched RI criteria, and the identifying ions consistent with this metabolite were observed, but co-elution prevented clean deconvolution or background ion subtraction and therefore confident identification. The corresponding genes for the biosynthesis of leucine were no longer investigated in this study. Nevertheless, the abundance of other amino acids detected in ditelosomic lines relative to Chinese Spring showed unambiguous chromatographic resolution and with a significant decrease in aspartate (aspartic acid 2 and 3 TMS; Table 1) for DT3AS only, whereas the remaining showed either an increase or decrease for at least two ditelosomic lines (Figure 4). Threonine (3 TMS, 19.59, 1387) was the exception where no significant difference was detected between any of the ditelosomic lines and Chinese Spring (Figure 4). Genes on homoeologous group 3 chromosomes, therefore, appear to have a significant effect on the accumulation of many aspartate-derived amino acids and were traced to identify those associated with the primary biochemical pathway having potential regulatory roles.

The interconnected biochemical pathway to convert aspartate to methionine, threonine, isoleucine, valine, alanine and leucine involves 18 enzymes (Figure 4). Although not classified as aspartate-derived amino acids, the biosynthesis of glutamine and asparagine involves two additional enzymes, aspartate transaminase and asparagine synthetase, respectively (Figure 4). Therefore, genes encoding 20 enzymes were analysed for their

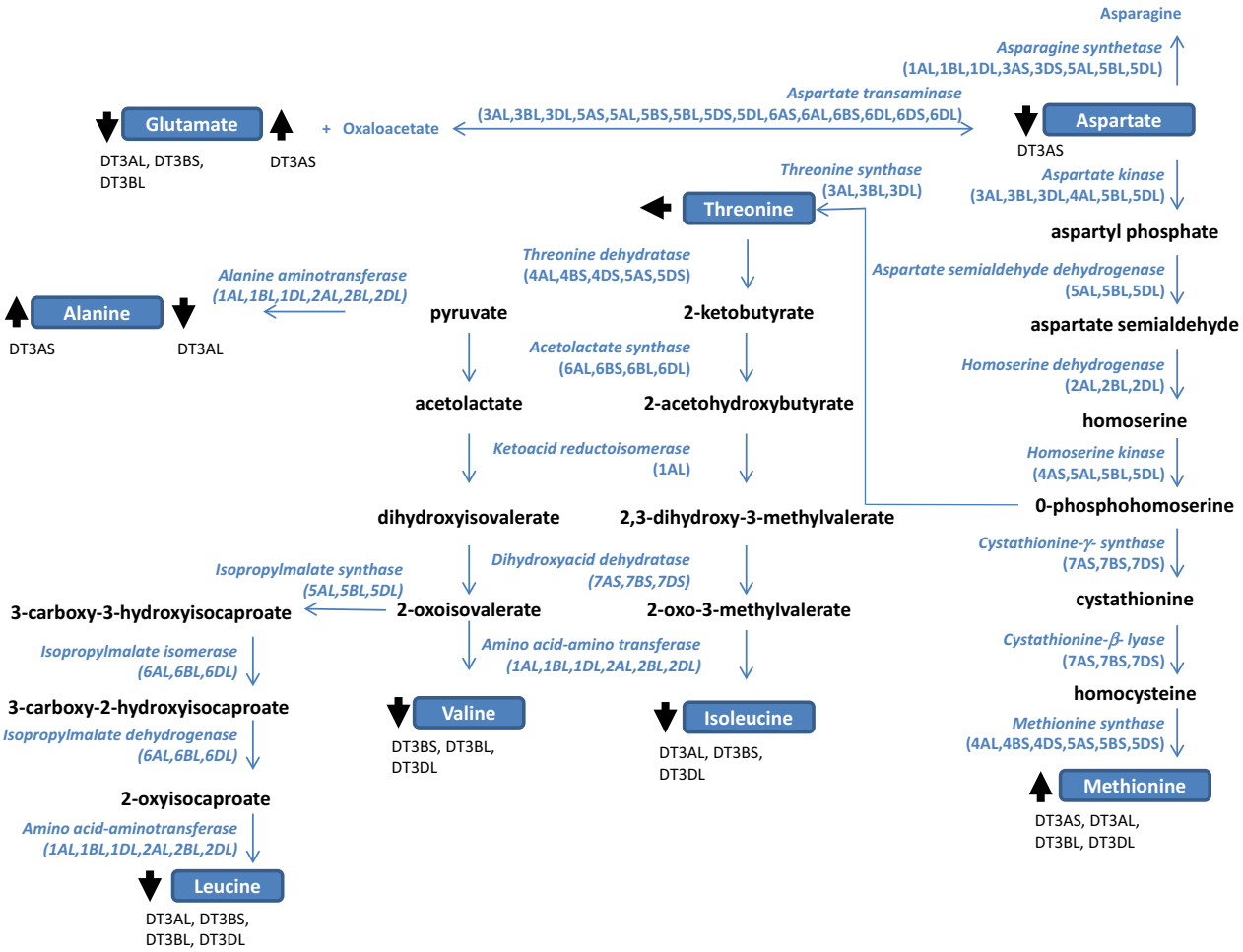

Figure 4 Schematic diagram of the biochemical pathway for the branched-chain amino acids. Blue boxes represent amino acids detected in the untargeted metabolite analysis with black arrows indicating increase or decrease in amino acid for corresponding ditelosomic lines compared with Chinese Spring. Enzymes names are shown in blue with the chromosomal location of their corresponding genes in parentheses.

potential role in controlling amino acid accumulation. The search for annotations from predominantly rice and other plant species identified FL-cDNA for each of the 20 enzymes and was used as query sequences in subsequent TBLASTX analysis to ascertain corresponding wheat FL-cDNA. At least one wheat FL-cDNA was identified for each respective enzyme, but most were represented by several FL-cDNA, indicating that most enzymes were encoded by multigene families and their chromosomal location ascertained by BLASTN search of the draft wheat genome sequence (Table 2 and Figure 4). Interestingly, genes encoding aspartate kinase, an enzyme involved in the first committed step to produce methionine from aspartate and threonine synthase that converts threonine from O-phosphohomoserine, were located on homoeologous group 3 chromosomes. As there is a significant decrease in aspartate for DT3AS (Table 1, Figure 4), the genes controlling its accumulation may either be aspartate transaminase or aspartate kinase located on chromosome 3AL, whereas the deletion of threonine synthase from 3AL, 3BL or 3DL in DT3AS, DT3BS and DT3DS lines, respectively, has no effect on threonine accumulation. The remaining 16 enzymes involved in the biosynthesis of aspartate-derived amino acids were located on chromosomes other than homoeologous group 3 despite increases and decreases in abundance of amino acids in ditelosomic lines relative to Chinese Spring (Table 2 and Figure 4). Therefore, genes from other interacting pathways and located on homoeologous group 3 chromosomes may be responsible for controlling the abundance of methionine, isoleucine, valine, leucine and alanine.

Discussion

This study demonstrated that aneuploid lines were a suitable genetic system to identify changes in metabolite profiles in mature grain when compared with the standard wheat genotype, Chinese Spring. Grain metabolite composition is impacted by different genotypes and genotype-by-environment interaction (Bellegia *et al.*, 2013), whereas differences in extraction processes and analytical instrumentation, even in nontargeted metabolomics, will favour some metabolite classes over others (Khakimov *et al.*, 2013). Nevertheless, grain of the genotypes analysed in this study was comparable; having been grown and stored in the same conditions and metabolites extracted and analysed using the same procedures and instrumentation to minimize any effects caused by the environment or detection methodologies. Additionally, having focussed on the grain rather than plant tissues, metabolism (in terms of normal physiological processes) is assumed to have ceased. Therefore, significant differences in compounds in ditelosomic lines with near identical background to

Chinese Spring were confidently attributed to genes missing from respective chromosome arms.

Methods consistent with untargeted metabolomics, including metabolite isolation and MS acquisition, were used to obtain an unbiased measurement of the metabolites in mature grain. Although classes of compounds identified in this study were similar to those previously reported (Bellegia *et al.*, 2013; and Lee *et al.*, 2013), the specific metabolites identified differed between these studies. For instance, metabolites related to unsaturated fatty acids, fatty alcohols, flavonols, phenolics, phytosterols and vitamins reported in Bellegia *et al.* (2013) and Lee *et al.* (2013) were not detected in this study and can be attributed to differences either in genotypes, environmental effects or disparate metabolite extraction and detection methods. It is interesting to note, however, that no untargeted metabolite profile for wheat grain has reported the detection of compounds associated with carotenoids, which give rise to flour yellowness, important for wheat end products (for review, see Ficco *et al.*, 2014), reflecting variation in compound resolution across studies and the requirement for complementary analyses for greater metabolome coverage (Gummer *et al.*, 2009; Wishart *et al.*, 2009). No single analytical method is capable of resolving the complete metabolome of any given tissue or system, due to the dynamic differences in metabolite chemistries, more specifically chemical structure, polarity, solubility and chromatographic behaviour (Ward *et al.*, 2003). This is a characteristic of small-molecule analysis reflected in metabolomics, more so than other 'omics' disciplines. Therefore, the inclusion of a targeted metabolomics approach, or complementary nontargeted analytical methods, together with different metabolite isolation procedures would provide a more complete measure of the metabolome (Harrigan *et al.*, 2007; Khakimov *et al.*, 2013).

Despite different biological and technical parameters that may affect metabolic profiles, this study used Chinese Spring as a reference genotype for qualitative and quantitative comparison of compounds in ditelosomic lines, to identify chromosome regions affecting metabolite accumulation in wheat seed. Significant differences between Chinese Spring and the ditelosomic lines indicated aneuploid lines are suitable to investigate the underlying genes controlling metabolite variation. It was expected that if genes encoding enzymes directly related to the corresponding biosynthetic pathways were located on homoeologous group 3 chromosomes, then ditelosomic lines would have a significant reduction in corresponding metabolites compared with Chinese Spring. Although a proportion showed a reduction, some ditelosomic lines showed significant increases in specific metabolites, indicating that genes directly and indirectly involved in biosynthetic pathways likely regulate metabolite accumulation.

An example of extreme metabolite variation in this study was in trehalose accumulation. The decrease of trehalose in DT3BS was presumably through the absence of the gene encoding TPP on chromosome 3BL; however, a reduction in trehalose was not detected for DT3AS and DT3DS, indicating that members of the *TPP* gene family on 3AL and 3DL may not serve a similar function to the gene on 3BL. Members of gene families on homoeologous chromosomes have previously been reported to differ in function on the basis of aneuploidy analysis, such as those encoding proteins regulating Na^+/K^+ accumulation (Ariyarathna *et al.*, 2014), and therefore, it is reasonable to assume that *TPP* genes on 3AL and 3DL have alternative functions. Trehalose has been reported to accumulate in roots and shoots of wheat (El-Bashiti *et al.*, 2005), and therefore, *TPP* genes on 3AL and 3DL may function to control trehalose accumulation in specific tissue. However, the contrasting 11-fold accumulation in DT3DS is extraordinary, and it was presumed that genes encoding trehalase that normally reduce trehalose when the enzyme converts it to glucose molecules (Müller *et al.*, 2001) may, indeed, increase trehalose in their absence. Genes related to trehalase were not located on wheat homoeologous group 3 chromosomes and, therefore, assumed that trehalose is regulated by genes of other pathways. Trehalose metabolism in plants is highly regulated through an intricate network of interconnecting pathways involved in post-translational modification, such as AMP-activated protein kinases and Snf-related protein kinases known to affect the altered state of TPS (Halford *et al.*, 2003; Harthill *et al.*, 2006; Martínez-Barajas *et al.*, 2011; Paul *et al.*, 2010; Zhang *et al.*, 2009). Therefore, genes involved in these or similar intricate pathways may be located on 3DL that have an effect on enzyme activity that would normally down-regulate trehalose in mature wheat grain. As trehalose accumulation has been implicated in providing protective mechanisms against stress tolerance in plants (Fernandez *et al.*, 2010; Garg *et al.*, 2002; Iordachescu and Imai, 2008; Penna, 2003), further investigations of interconnecting but yet undefined pathways are certainly warranted to identify key genes on chromosome 3DL that normally inhibit trehalose accumulation. Metabolic profiling of deletion lines (Endo and Gill, 1996) in subsequent studies would assist in ascertaining the key genetic determinants from a smaller pool of candidates on 3DL and develop strategies to manipulate elevated levels of trehalose that may lead to improved stress tolerance during grain filling.

The regulation of aspartate-derived amino acids is of particular interest in this study, not only for their interconnecting biological pathways but their importance in the human diet. Methionine, isoleucine and threonine are essential amino acids, not synthesized in animals, and, therefore, are important for improving the nutritional value of cereal grain (Ufaz and Galili, 2008). Aspartate is the primary amino acid by which these essential amino acids are synthesized, and its accumulation is affected by aspartate kinase for the production of methionine as a substrate for aspartate transaminase to produce glutamate. In this study, the reduction of asparate in DT3AS corresponds to either the loss of aspartate kinase or the aspartate transaminase genes on 3AL, where the latter is complemented by an increase in glutamate in the same ditelosomic line. However, decreases in aspartate or increases in glutamate were not observed for DT3BS and DT3DS, so it is likely that aspartate kinase or aspartate transaminase on chromosome 3BL and 3DL may have a different role other than regulating amino acid accumulation in wheat gain, potentially regulating metabolite accumulation in other tissue. Metabolite profiling of aneuploid lines from other tissue will provide further information on alternative roles of aspartate kinase and aspartate transaminase in regulating amino acid accumulation during plant growth and development.

A notable feature of the branched-chain amino acid pathway is the nonsignificant difference in levels of threonine in ditelosomic lines relative to Chinese Spring despite its biosynthesis from O-phosphohomoserine through threonine synthase; an enzyme encoded by three genes located on the long arm of homoeologous group 3 chromosomes. Therefore, it appears that threonine synthase genes on chromosomes 3AL, 3BL and 3DL encode an active enzyme capable of maintaining threonine homoeostasis during grain development despite the absence of any particular gene in a corresponding ditelosomic line. O-phosphohomoserine

is a common precursor and significant increases in methionine were detected in some ditelosomic lines; however, no genes encoding enzymes controlling methionine or, indeed, accumulation of other amino acids were identified on group 3 chromosomes. Therefore, genes from unidentified interconnecting pathways are likely to control methionine accumulation and the remaining branched-chain amino acids including isoleucine, valine, leucine and alanine. Amino acid accumulation can be affected by other biological processes, including proteins involved in subcellular localization and transport mechanisms, feedback inhibition and activation, post-translation regulation through allosteric regulation of enzymes and transcriptional regulation (Jander and Joshi, 2010; Joshi *et al.*, 2010; Ortiz-Lopez *et al.*, 2000). Indeed, these biological processes integrate in a complex manner to control branched-chain amino acid synthesis whereby some of the underlying but, as yet, unidentified genes may be located on homoeologous group 3 chromosomes. The wheat aneuploid lines would provide an appropriate experimental system in future studies to support the functional analysis of alternative genes involved in the network of numerous biological processes controlling branched-chain amino acid accumulation.

This study has strategically used ditelosomic lines to provide information on genes encoding enzymes of biosynthetic pathways that control metabolite accumulation in a tissue-specific manner. The role of TPP on 3BL reducing trehalose and aspartate kinase on 3AL decreasing aspartate accumulation are good examples on the use of aneuploid lines to discriminate functional roles of genes on homoeologous chromosomes in controlling metabolite accumulation in mature grain. Moreover, the analysis of ditelosomic lines has uncovered a plethora of unidentified biological networks other than genes encoding enzymes of the primary biosynthetic pathway that controls metabolites. The future challenge will be to discover the intricate components of these pathways and their precise role in controlling metabolite accumulation. The completion of the wheat genome sequence including the annotation of pseudomolecules, similar to that for rice and maize (Ouyang *et al.*, 2007; Zhou *et al.*, 2009), coupled with untargeted and targeted metabolite analysis of seeds of wheat aneuploid lines with small deleted regions for all wheat chromosomes (Endo and Gill, 1996) will be a powerful metabolomics–genomics strategy and supporting genetic system to ascertain interconnecting biological networks and underlying genes regulating metabolite and trait variation in wheat grain.

Experimental procedures

Plant material

Seeds of wheat line Chinese Spring and ditelosomic lines, DT3AS, DT3AL, DT3BS, DT3BL, DT3DS, DT3DL, were kindly provided by Dr Jon Raupp, Wheat Genetic and Genomics Resource Center, Kansas State University, Kansas, USA. Seeds were sown in pots and plants grown to maturity in the glasshouse in 2013. Harvested grain from four individual plants of Chinese Spring and each ditelosomic line was pooled and stored at 4 °C in an airtight container with silica gel for 3 months until used for metabolite extraction.

Metabolite extraction

Harvested grain was retrieved from 4 °C storage, and 10–15 seeds per technical replicate (five replicates total) were lyophilized for 16 h in a LABCONCO Freezone 2.5 Plus (Labconco Corp Kansas City, MO). Seeds were ground to a fine powder in a mortar and pestle, chilled with dry ice and 25 mg per replicate transferred to a 2.0 mL tube. Methanol was added to each tube, together with 650 ng $^{13}C_6$-sorbitol (ISTD; in methanol) to a combined volume of 500 μL, and vigorously agitated within a Precellys 24 lysis cryo-mill tissue lyser (Bertin Technologies, Aix-en-Provence, France) at ~5000 g for two subsequent rounds of 20 s. The suspension was agitated in a thermomixer (Eppendorf, South Pacific Pty. Ltd., North Ryde, Australia) at ~1000 g for 15 min at 10 °C and sample particulate collected by centrifuge at 10 000 *g*. The supernatant was transferred to a fresh tube and the extraction repeated with another 500 μL methanol, without any further addition of ISTD. The supernatants were combined and dried in preparation for derivatization by vacuum removal of the organic solvent, followed by drying by lyophilization, as described by Gummer *et al.* (2013). This required concentration of the extract to <100 μL volume in an Eppendorf Concentrator Plus vacuum concentrator (Eppendorf, South Pacific Pty. Ltd., North Ryde, Australia) and the subsequent addition of 300 μL of LC-MS grade water. The sample was then frozen in liquid nitrogen and dried by lyophilization in a LABCONCO Freezone 2.5 Plus (Labconco Corp Kansas City, MO). The dried extracts were stored at −80 °C until metabolite derivatization.

GC-MS analysis of metabolites

The metabolites were derivatized by a combination of methoximation and silylation reactions. To the dried metabolites was added 20 μL of methoxylamine HCl (Sigma-Aldrich, Castle Hill, NSW, Australia) [20 mg/mL in pyridine (UNIVAR)], followed by brief mixing by vortex and incubation at 30 °C for 90 min with agitation at ~800 g in an Eppendorf thermomixer. Fourty microlitres of MSTFA (Sigma-Aldrich, Castle Hill, NSW, Australia) was then added and mixed briefly by vortex before incubation at 37 °C for 30 min with agitation at 300 rpm. The entire volume was then transferred to a 200-μL glass insert within a 2-mL analytical vial and, five microlitres of *n*-alkanes [(C_{10}, C_{12}, C_{15}, C_{19}, C_{22}, C_{28}, C_{32} and C_{36}); Sigma-Aldrich] in hexane [for retention index (RI) calculation] was added and mixed. Samples were loaded on to the GC-MS in a randomized sequence for analysis.

Metabolites were analysed using a Shimadzu QP2010 Ultra GC-MS with AOC-20i Autosampler and injector unit (Shimadzu, Kyoto, Japan) equipped with an Agilent Factor Four fused silica capillary column (VF-5 ms 30 × 0.25 mm × 0.25 μm + 10 m EZ-Guard; Agilent Technologies, Santa Clara, CA). Helium was used as the carrier gas at constant flow. One microlitre of sample was injected into a split/splitless GC inlet, held at 230 °C using a splitless mode of injection, with an initial GC column temperature held at 70 °C. The oven temperature was initially ramped one °C/min for five minutes before a final ramp of 5.6 °C/min to a final temperature of 320 °C. The transfer line and ionization source were held at 280 °C and 230 °C, respectively. The mass spectrometer was set to scan a mass range of 40–600 *m/z* at 10 scans/s using a 70 eV electron beam.

Metabolomics data analyses and metabolite identification

GC-MS data were analysed using Shimadzu GC-MS solution 2.61 (Shimadzu, Kyoto, Japan). A target list of detected analytes was assembled from the collected GC-MS data for development of a processing method. Each detected metabolite was assigned three (unique, where possible) ions: one quantifier and two qualifier ions. The ions were recognized within a five-second retention time (RT) window. Each metabolite entry was checked for the

presence of conflicting ions within the assigned RT deviation. Relative quantitation was determined by calculation and comparison of quantifier ion peak areas. Raw peak areas were normalized to the ISTD ($^{13}C_6$ sorbitol).

Metabolites were identified by mass spectral match to an in-house library, generated from the analysis of authentic metabolite standards. Identification required a minimum forward match percentage of 80% or higher, to be within 5 retention indices (RIs) of the analysed standard compound. Putative identification of metabolites was carried out using the National Institute of Standards and Technology (NIST) 2011 mass spectral library. Metabolites were assigned a mass spectral tag (MST) describing the respective identification and analytical features of the analyte of ontology 'metabolite ID_RT_RI'.

Multivariate and statistical analyses were performed using the Unscrambler X software, version 10.1 (CAMO Software, Oslo, Norway). The ISTD-corrected data matrix was scaled by $\log_{10}(x + 1)$ transformation prior to principal component analysis (PCA), using noniterative partial least squares algorithm, cross-validation with no rotation. Significant differences in metabolite abundance between metabolite profiles were determined using an independent, two-tailed Student's t-test and were deemed to be significant or highly significant when $P \leq 0.01$ or $P \leq 0.001$, respectively.

BLAST similarity searching, gene identification and location in the wheat genome

Full-length (FL-) cDNA from rice (The Rice Full-Length cDNA Consortium, 2013) annotated to encode enzymes of biochemical pathways was retrieved from National Center for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov/) using key word searching. In the event that rice cDNA sequences were not identified, key word searches were extended to identify annotated FL-cDNA from *Zea mays* or *Arabidopsis thaliana*. Annotated cDNA sequences were used as query sequences in TBLASTX search to identify corresponding wheat FL-cDNA from the Chinese Spring collection (Kawaura *et al.*, 2009). Wheat sequences were identified as orthologs of the annotated rice cDNA when e-values of TBLASTX hits were $<6e^{-100}$. The wheat FL-cDNA sequences were used as a query in BLASTN searching against the wheat genome survey sequence (http://wheat-urgi.versailles.inra.fr/Seq-Repository/) and assigned chromosomal location based on 90% sequence identity threshold value.

References

Andersen, M.R., Nielsen, M.L. and Nielsen, J. (2008) Metabolic model integration of the bibliome, genome, metabolome, and reactome of *Aspergillus niger*. *Mol. Syst. Biol.* **4**, 178.

Ariyarathna, H.A.C.K., Ul-Haq, T., Colmer, T.D. and Francki, M.G. (2014) Characterization of the multigene family *TaHKT2;1* in bread wheat and the role of gene members in plant Na^+ and K^+ status. *BMC Plant Biol.* **14**, 159.

Bellegia, R., Platani, C., Nigro, F., De Vita, P., Cattivelli, L. and Papa, R. (2013) Effect of genotype, environment, and genotype-by-environment interaction on metabolite profiling in durum wheat (*Triticum durum* Desf.) grain. *J. Cereal Sci.* **57**, 183–192.

Bino, R.J., Hall, R.D., Fiehn, O., Kopka, J., Saito, K., Draper, J., Nikolau, B.J., Mendes, P., Roessner-Tunali, U., Beale, M.H., Trethewey, R.N., Lange, B.M., Wurtele, E. and Sumner, L.W. (2004) Potential of metabolomics as a functional genomics tool. *Trends Plant Sci.* **9**, 418–425.

El-Bashiti, T., Hamamci, H., Öktem, H. and Yücel, M. (2005) Biochemical analysis of trehalose and its metabolizing enzymes in wheat under abiotic stress conditions. *Plant Sci.* **169**, 47–54.

Endo, T.R. and Gill, B.S. (1996) The deletion stocks of common wheat. *J. Hered.* **87**, 295–307.

Erayman, M., Sandhu, D., Sidhu, D., Dilbirligi, M., Baenziger, P.S. and Gill, K.S. (2004) Demarcating the gene rich regions of the wheat genome. *Nucleic Acids Res.* **32**, 3546–3565.

Fernandez, O., Béthencourt, L., Quero, A., Sangwan, R.S. and Clément, C. (2010) Trehalose and plant stress responses: friend or foe? *Trends Plant Sci.* **15**, 409–417.

Feuillet, C., Leach, J.E., Rogers, J., Schnable, P.S. and Eversole, K. (2011) Crop genome sequencing: lessons and rationales. *Trends Plant Sci.* **16**, 77–88.

Ficco, D.B.M., Mastrangelo, A.M., Trono, D., Borrelli, G.M., De Vita, P., Fares, C., Beleggia, R., Platani, C. and Papa, R. (2014) The colors of durum wheat: a review. *Crop Past. Sci.* **65**, 1–5.

Fiehn, O. (2002) Metabolomics- the link between genotypes and phenotypes. *Plant Mol. Biol.* **48**, 155–171.

Fridman, E. and Pichersky, E. (2005) Metabolomics, genomics, proteomics, and the identification of enzymes and their substrates and products. *Curr. Opin. Plant Biol.* **8**, 242–248.

Garg, A.K., Kim, J.-K., Owens, T.G., Ranwala, A.P., Choi, Y.D., Kochian, L.D. and Wu, R.J. (2002) Trehalose accumulation in rice plants confers high tolerance levels to abiotic stresses. *Proc. Natl Acad. Sci. USA*, **99**, 15898–15903.

Gummer, J., Banazis, M., Maker, G., Solomon, P., Oliver, R. and Trengove, R. (2009) Use of mass spectrometry for metabolite profiling and metabolomics. *Aust. Biochem.* **40**, 5–16.

Gummer, J.P.A., Trengove, R.D., Oliver, R.P. and Solomon, P.S. (2013) Dissecting the role of G-protein signalling in primary metabolism in the wheat pathogen *Stagonospora nodorum*. *Microbiology*, **159**, 1972–1985.

Halford, N.G., Hey, S., Jhurreea, D., Laurie, S., McKibbin, R.S., Paul, M. and Zhang, Y. (2003) Metabolic signalling and carbon partitioning: role of Snf1-related (SnRK1) protein kinase. *J. Exp. Bot.* **54**, 467–475.

Hall, R.D. (2006) Plant metabolomics: from holistic hope, to hype, to hot topic. *New Phytol.* **169**, 453–468.

Harrigan, G.C., Martino-Catt, S. and Glenn, K.C. (2007) Metabolomics, metabolic diversity and genetic variation in crops. *Metabolomics*, **3**, 259–272.

Harthill, J.E., Meek, S.E.M., Morrice, N., Peggie, M.W., Borch, J., Wong, B.H.C. and MacKintosh, C. (2006) Phosphorylation and 14-3-3 binding of Arabidopsis trehalose-phosphate synthase 5 in response to 2-deoxyglucose. *Plant J.* **47**, 211–223.

Iordachescu, M. and Imai, R. (2008) Trehalose biosynthesis in response to abiotic stresses. *J. Integr. Plant Biol.* **50**, 1223–1229.

Jander, G. and Joshi, V. (2010) Recent progress in deciphering the biosynthesis of aspartate-derived amino acids in plants. *Mol. Plant.* **3**, 54–65.

Joshi, V., Joung, J.-G., Fei, Z. and Jander, G. (2010) Interdependence of threonine, methionine and isoleucine metabolism in plants: accumulation and transcriptional regulation under abiotic stress. *Amino Acids*, **39**, 933–947.

Kawaura, K., Mochida, K., Enju, A., Totoki, Y., Toyoda, A., Sakaki, Y., Kai, C., Kawai, J., Hayashizaki, Y., Seki, M., Shinozaki, K. and Ogihara, Y. (2009) Assessment of adaptive evolution between wheat and rice as deduced from full-length common wheat cDNA sequence data and expression patterns. *BMC Genom.* **10**, 271.

Khakimov, B., Bak, S. and Engelsen, S.B. (2013) High-throughput cereal metabolomics: current analytical technologies, challenges and perspectives. *J. Cereal Sci.* **59**, 393–418.

Lee, D.P., Alexander, D. and Jonnalgadda, S.S. (2013) Diversity of nutrient content in grains – a pilot metabolomics analysis. *J. Nutr. Food Sci.* **3**, 2.

Martínez-Barajas, E., Delatte, T., Schluepmann, H., de Jong, G.J., Somsen, G.W., Nunes, C., Primavesi, L.F., Coello, P., Mitchell, R.A.C. and Paul, M.J. (2011) Wheat grain development is characterized by remarkable trehalose accumulation trehalose 6-phosphate accumulation pregrain filling: tissue distribution and relationship to SNF1-related protein kinase1 activation. *Plant Physiol.* **156**, 373–381.

Matthews, S.B., Santra, M., Mensack, M.M., Wolfe, P., Byrne, P.F. and Thompson, H.J. (2012) Metabolite profiling of a diverse collection of wheat lines using ultraperformance liquid chromatography coupled with time-of-flight mass spectrometry. *PLoS ONE*, **7**, e44179.

Müller, J., Aeschbacher, R.A., Wingler, A., Boller, T. and Wiemken, A. (2001) Trehalose and trehalase in Arabidopsis. *Plant Physiol.* **125**, 1086–1093.

Ortiz-Lopez, A., Chang, H.-C. and Bush, D.R. (2000) Amino acid transporters in plants. *Biochim. Biophys. Acta*, **1465**, 275–280.

Ouyang, S., Zhu, W., Hamilton, J., Lin, H., Campbell, M., Childs, K., Thibaud-Nissen, F., Malek, R.L., Lee, Y., Zheng, L., Orvis, J., Haas, B., Wortman, J. and Buell, C.R. (2007) The TIGR rice genome annotation resource: improvement and new features. *Nucleic Acids Res.* **35**, D883–D887.

Paul, M.J., Jhurreea, D., Zhang, Y., Primavesi, L.F., Delatte, T., Schluepmann, H. and Wingler, A. (2010) Upregulation of biosynthetic processes associated with growth by trehalose 6-phosphate. *Plant Signal. Behav.* **5**, 386–392.

Penna, S. (2003) Building stress tolerance through over-producing trehalose in transgenic plants. *Trends Plant Sci.* **8**, 355–357.

Pfeifer, M., Kugler, K.G., Sandve, S.R., Zhan, B., Rudi, H., Hvidsten, T.R., International Wheat Genome Sequencing Consortium, Mayer, K.F.X. and Olsen, O.-A. (2014) Genome interplay in the grain transcriptome of hexaploid bread wheat. *Science*, **345**, 1250091.

Saito, K. and Matsuda, F. (2010) Metabolomics for functional genomics, systems biology and biotechnology. *Ann. Rev. Plant Biol.* **61**, 463–489.

Saito, K., Hirai, M.Y. and Yonekura-Sakakibara, K. (2008) Decoding genes with co-expression networks and metabolomics- 'majority report by precogs'. *Trends Plant Sci.* **13**, 36–43.

Schauer, N. and Fernie, A.R. (2006) Plant metabolomics: towards biological function and mechanism. *Trends Plant Sci.* **11**, 508–516.

The International Wheat Genome Consortium (IWGSC). (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science*, **345**, 1251788.

The Rice Full-Length cDNA Consortium. (2013) Collection, mapping and annotation of over 28,000 cDNA clones from *japonica* rice. *Science*, **301**, 376–379.

Ufaz, S. and Galili, G. (2008) Improving the content of essential amino acids in crop plants: goals and opportunities. *Plant Physiol.* **147**, 954–961.

Ward, J.L., Harris, C., Lewis, J. and Beale, M.H. (2003) Assessment of ^{1}H NMR spectroscopy and multivariate analysis as a technique for metabolite fingerprinting of *Arabidopsis thaliana*. *Phytochemistry*, **62**, 949–957.

Wishart, D.S., Knox, C., Guo, A.C., Eisner, R., Young, N., Gautam, B., Hau, D.D., Psychogios, N., Dong, E., Bouatra, S., Mandal, R., Sinelnikov, I., Xia, J., Jia, L., Cruz, J.A., Lim, E., Sobsey, C.A., Shrivastava, S., Huang, P., Liu, P., Fang, L., Peng, J., Fradette, R., Cheng, D., Tzur, D., Clements, M., Lewis, A., De Souza, A., Zuniga, A., Dawe, M., Xiong, Y., Clive, D., Greiner, R., Nazyrova, A., Shaykhutdinov, R., Li, L., Vogel, H.J. and Forsythe, I. (2009) HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.* **37**, D603–D610.

Yuan, J.S., Galbraith, D.W., Dai, S.Y., Griffin, P. and Stewart, C.N. (2008) Plant systems biology comes of age. *Trends Plant Sci.* **13**, 165–171.

Zhang, Y., Primavesi, L.F., Jhurreea, D., Andralocj, P.J., Mitchell, R.A.C., Powers, S.J., Schluepmann, H., Delatte, T., Wingler, A. and Paul, M.J. (2009) Inhibition of SNF1-related protein kinase1 activity and regulation of metabolic pathways by trehalose 6-phosphate. *Plant Physiol.* **149**, 1860–1871.

Zhou, S., Wei, F., Nguyen, J., Bechner, M., Potamousis, K., Goldstein, S., Pape, L., Mehan, M.R., Churas, C., Pasternak, S., Forrest, D.K., Wise, R., Ware, D., Wing, R.A., Waterman, M.S., Livny, M. and Schwatrz, D.C. (2009) A single molecule scaffold for the maize genome. *PLoS Genet.* **5**, e1000711.

Supporting information

Additional Supporting information may be found in the online version of this article:

Table S1 Metabolite profile of mature wheat grain from Chinese Spring and ditelosomic lines.