Identification and characterization of genes on a single subgenome in the hexaploid wheat (*Triticum aestivum* L.) genotype 'Chinese Spring'

Journal:	<i>Genome</i>
Manuscript ID	gen-2016-0076.R3
Manuscript Type:	Article
Date Submitted by the Author:	28-Jul-2016
Complete List of Authors:	Ma, Jian; Triticeae Research Institute, Sichuan Agricultural University Zheng, Zhi; National Millet Improvement Center, Key Laboratory of Minor Cereal Crops of Hebei Province, Institute of Millet Crops, Hebei Academy of Agriculture and Forestry Sciences Stiller, Jiri; CSIRO, Agriculture Flagship Lan, Xiu-Jin; Sichuan Agricultural University, Triticeae Research Institute Liu, Yaxi; Sichuan Agricultural University Deng, Mei; Sichuan Agricultural University Wang, Penghao; Murdoch University School of Veterinary and Life Sciences Pu, Zhi-En; Sichuan Agricultural University, Triticeae Research Institute Chen, Guangdeng; Sichuan Agricultural University, Triticeae Research Institute Jiang, Qian-Tao; Sichuan Agricultural University, Triticeae Research Institute Wei, Yu-Ming; Sichuan Agricultural University Zheng, You-Liang; Sichuan Agricultural University,
Keyword:	gene loss, evolution, wheat genome, Chinese Spring

SCHOLARONE™
Manuscripts

Identification and characterization of genes on a single subgenome in the hexaploid wheat (*Triticum aestivum* L.) genotype ‘Chinese Spring’

Jian Ma^{1,#,*}, Zhi Zheng^{2,#}, Jiri Stiller³, Xiu-jin Lan¹, Yaxi Liu¹, Mei Deng¹, Penghao Wang⁴, Zhien Pu^{1,5}, Guangdeng Chen⁶, Qian-tao Jiang¹, Yuming Wei¹, You-Liang Zheng¹

¹*Triticeae Research Institute, Sichuan Agricultural University, Wenjiang, Chengdu 611130, China*

²*CSIRO Agriculture, 306 Carmody Road, St Lucia, QLD 4067, Australia*

³*National Foxtail Millet Improvement Centre, Institute of Millet Crops, Hebei Academy of Agricultural and Forestry Sciences, Shijiazhuang, China*

⁴*School of Veterinary and Life Sciences, Murdoch University, 90 South Street, Murdoch WA6150*

⁵*College of Agronomy, Sichuan Agricultural University, Chengdu 611130, P.R.China*

⁶*Institute of Ecological and Environmental Sciences, Sichuan Agricultural University, Wenjiang, Chengdu, 611130, China*

contributed equally to this work.

*Authors for correspondence:

Dr. Jian Ma: jianma@sicau.edu.cn

Abstract

Gene loss during the formation of hexaploid bread wheat has been reported repeatedly. However, our knowledge on genome-wide analysis of the genes present on a single subgenome in bread wheat is still limited. In this study, by analysing the 'Chinese Spring' chromosome arm shotgun sequences together with high-confidence gene models, we detected 433 genes being on a single subgenome (SSGs) only. Greater gene loss was observed in A and D subgenomes compared with B subgenome. More than 79% of the orthologs for these SSGs were detected in diploid and tetraploid relatives of hexaploid wheat as well. Unexpectedly, no bias in expression breadth, or distribution patterns of GO (gene ontology) terms for these genes was detected for all of the high-confidence genes. Further, network and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway analyses indicated that most of these genes were not functionally related to each other. Interestingly, 30.7% of these SSGs are most highly expressed in root, showing biased distribution given the distribution of the whole high-confidence genes. Collectively, these results facilitate our understanding of the loss of the genes that were retained in a single subgenome during the formation of hexaploid wheat.

Keywords gene loss, evolution, wheat genome, Chinese Spring

Introduction

It is widely believed that the allohexaploid bread wheat (*Triticum aestivum* L.; genome AABBDD) has undergone two sequential allopolyploidization events. First, allotetraploidization occurred between two diploid species, *T. urartu* (AA) and an unknown close relative of *Aegilops speltoides* (SS), resulting in the extant tetraploid wheat (*T. turgidum*; AABB) (Petersen et al. 2006; Salamini et al. 2002). Second, allohexaploidization occurred between the tetraploid wheat and goat grass (*A. tauschii*; DD), leading to the formation of bread wheat (Petersen et al. 2006; Salamini et al. 2002).

The formation of hexaploid wheat is a typical process of whole-genome duplication (WGD). It is reported that the duplication of genomes plays a key role during evolution. Gene gain and loss and even chromosomal rearrangements are common (De Smet et al. 2013; Ma et al. 2013; Ma et al. 2014a; Ma et al. 2015; Scannell et al. 2007; Schnable et al. 2012; Zhang 2003). Compared to the appreciated importance of gene gain through duplication, gene loss has only recently attracted attention (De Smet et al. 2013).

The development of sequencing techniques greatly facilitates the generation of genome and transcriptome data from various species. By analysing these publicly-available sequence data, numerous studies have identified lost genes during the evolutionary history of a given species or across different species. For example, the IWGSC (International Wheat Genome Sequencing Consortium) (2014) observed that lost genes during bread wheat evolution were mainly those of expanded families, which is in accordance with previous findings (Brenchley et al. 2012). In contrast, singletons (i.e., genes without paralogous copies within the same genome) did not usually experience gene loss in polyploidization (IWGSC, 2014). However, gene loss on each of the three subgenomes is not clear and functions of the genes that were retained to be on a single subgenome (including genes with paralogous copies within the same subgenome) during polyploidization have not been assessed.

Recently, increasing studies are focused on single-copy genes. For instance, De Smet et al. (2013) identified a large number of genes that are convergently

restored to single-copy status after multiple genome-wide and smaller scale duplication events in the sequenced genomes of 20 flowering plants, consistent with previous observations (Duarte et al. 2010; Paterson et al. 2006). By analysing assembled contigs representing genes for chromosomes 7A, 7B and 7D in bread wheat, Berkman et al. (2013) detected 550 *Brachypodium distachyon* genes with orthologs present on only 2 wheat genomes and 545 *B. distachyon* genes with orthologs present on only a single wheat genome (i.e. either of 7A, or 7B, or 7D).

Nevertheless, to our knowledge, genome-wide analysis of the genes present on a single subgenome in bread wheat is still limited. The newly published wheat chromosome survey sequences (IWGSC, 2014) allow us to systematically analyse specific gene loss, i.e. the genes that were retained to be on a single subgenome during polyploidization. Here, we reported on genome-wide identification and characterization of genes being on a single subgenome in bread wheat by analysing the chromosome shotgun sequence and high-confidence gene models of the reference base of bread wheat, 'Chinese Spring'.

Materials and methods

Data collection

The updated gene models of 'Chinese Spring' ('CS') (version 2.2) were downloaded from the website (https://urgi.versailles.inra.fr/download/iwgsc/Gene_models/) (IWGSC, 2014). To accurately identify genes present on a single subgenome, only gene models (99,386; ta_IWGSC_MIPSV2.2_HighConf_REPR_CDS_2014Jul18) with a home on 'CS' chromosome arm shotgun sequences assembly (and therefore a chromosome arm assignment) and one representative per gene if there were several alternative splice variants for a given gene were retrieved in the present study. As some of these high-confidence gene models were fragmented in the assembly, they could only be partially structurally defined or were classified as gene fragments and pseudogenes (IWGSC, 2014). Further sequencing may improve the coverage and quality of genic sequences and many of them are expected to be merged. Here, to reduce the possibility of

omitting target genes (i.e. single subgenome genes) all the high-confidence gene models (99,386) were used for the further analysis. Thus, the identified genes could be overestimated. The whole genome sequences of *T. turgidum*, *T. monococcum*, *T. urartu*, *A. speltooides*, *A. tauschii*, and *A. sharonensis* were collected from IWGSC (<http://www.wheatgenome.org/>) (IWGSC, 2014). Coding sequences (CDS) of *Arabidopsis* were downloaded from Phytozome (<http://www.phytozome.org/>).

Identification of single subgenome genes

The 99,386 high-confidence gene models were blasted against A- B- and D-subgenome of 'CS' shotgun sequences using the BLAST+ blastn algorithm, respectively. To avoid missing target genes, a loose criterion: an E-value threshold of 10^{-5} (this value was applied in all subsequent BLAST analyses) was used only. Comparison of the blast results identified a total of 433 genes on a single subgenome only (i.e. single subgenome genes, SSGs). As some of the genes may be fragments and the sequence assembly is not perfect, these identified genes could very likely comprise those that map on several different contigs from a same subgenome. It is also possible that a given gene model could find its paralogous copies in a given subgenome. As these genes were identified on the same subgenome, we here regard them as single subgenome genes as well. Thus, the number of SSGs could be likely overestimated.

Detection of SSGs in diploid and tetraploid wheat

To detect the presence of SSGs in the relatives of wheat, the 433 single subgenome genes were further blasted against the diploid and tetraploid relatives of wheat (*T. turgidum*, *T. monococcum*, *T. urartu*, *A. speltooides*, *A. tauschii*, and *A. sharonensis*). As described in the previous section, the SSGs could most likely comprise their paralogous copies using the present method. To detect possible ortholog of the SSGs, a loose criterion was adopted: a gene was regarded as absent in a given species if no hits were detected. Thus, it is very likely to detect homology rather than orthology of the SSGs in the diploid and tetraploid relatives as well. In addition, the imperfect contigs

assembly of the diploid and tetraploid prevents accurately estimating the ortholog of the SSGs.

Functional annotation of SSGs

Function of the SSGs and the 99,386 high-confidence gene models were annotated by the Blast2GO program (version 2.7.2) (Conesa et al. 2005). E-value threshold of 10^{-6} was used when executing steps of 'Blast'; the default parameters for 'Annotation' were set as following: "Annotation rule cut-off: 55", "E-Value: 1.0E-6", "Hit-HSP overlap: 0", "GO weight: 5" in the Blast2GO program. To summarize the new functional information, a GO-Slim (goslim_generic.obo) analysis was performed. Gene ontology (GO) functional classification of all genes was further performed by choosing Ontology Level 2 for biological process, cellular component and molecular function in Blast2GO. KEGG (Kyoto Encyclopedia of Genes and Genomes) analysis was performed using the online KEGG Automatic Annotation Server (KAAS) (<http://www.genome.jp/kegg/kaas/>) (Moriya et al. 2007). The bi-directional best hit (BBH) method was used to obtain KEGG ortholog assignments. Organism list with all the species in planta were selected.

Network analysis

The identified SSGs and gene models from each of the wheat chromosomes were compared with the *A. thaliana* gene sequences using BLASTN (Altschul et al. 1997) with an E-value cut-off of 10^{-5} . The list of names for the *A. thaliana* genes with top hits against the SSGs and genes on each chromosome were extracted. The names against SSGs were then loaded into STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) (Franceschini et al. 2013) to identify the representation of networked genes with parameters of 'highest confidence: 0.900 (the range of confidence score is 0-1) and custom limit: 0'. Due to the limitation of processing large number of genes in web-based tool, the protein network data (scored links between proteins) was downloaded from STRING 9.1 (<http://string-db.org/>) for analysing networks of genes from each chromosome. Compared with the proteins with confidence score being higher than 0.9, the genes with a link to any of other genes on a given chromosome were recorded for further analysis.

Expression analysis

The public transcriptome data from five tissues (root, leaf, grain, stem and spike) of 'Chinese Spring' bread wheat at three developmental stages (IWGSC, 2014) were used to analyse the expression profiling. For the detected SSGs, Reads Per Kilobase Million (RPKM) was used as expression values. The method of RNA-sequence analysis was adopted from Ma et al. (2014b). Expression breadth is defined as the number of conditions in which a given gene is expressed. Thus, the expression breadth in this study ranges from 0 to 15 as there are five tissues and three developmental stages. To identify the tissue in which a given gene is mainly expressed, the sum of the expression values from the three developmental stages for this given tissue was used to represent the expression value of this tissue.

Results

Distribution of single subgenome genes across genomes

Of the 99,386 high-confidence genes, 433 (0.4%) single subgenome genes were detected (Table S1). Of them, 76 were located on chromosomes belonging to the A subgenome only, 226 on B subgenome only, and 131 on D subgenome only (Table 1). Among the seven homoeologous groups, group 3 contained most genes of these genes (142), and group 2 least (21). No significant correlation between gene amount and chromosome size for any subgenome was detected (Fig. S1).

The presence of the orthologs of single subgenome genes in the evolutionary relatives

A majority of the orthologs of these SSGs identified being on a single subgenome were detected in diploid and tetraploid relatives of hexaploid wheat. For example, 80% and 83% of those A-only genes were detected as well in diploid *A. speltooides* (SS) and *A. tauschii* (DD), respectively (Table 2). Eight-five percent and 84% of those B-only genes were detected in the diploid *T. urartu* (AA) and *A. tauschii* (DD), respectively. Seventy-nine percent and 89% of those D-only genes were detected in *T. urartu* (AA) and *A. speltooides* (SS),

respectively; and ~80% of those D-only genes were detected in tetraploid *T. Turgidum* (AABB) (Table 2).

Expression profiling of the single subgenome genes

The number of genes expressed in 15 conditions accounts for most of those A (21.05%) and B-only (19.47%) genes (Fig. 1). For the D-only genes, the unexpressed genes (18.32%) rank first, which was followed by those expressed in 15 conditions. When compared with the expression breadth for the whole high-confidence genes, a similar trend was observed. The genes expressed in 15 conditions rank first, followed by the unexpressed genes (Fig. 1).

Interestingly, the genes with most expression in root account for most in those A (32.9%), B (31.9%), and D-only (27.5%) genes, which were followed by those in grain for A and D-only genes (A, 25.0%; D: 19.8%) and in spike for B-only genes (B, 24.3%), and least in stem (A, 7.9%; B, 11.9%; D: 2.3%) (Fig. 2). The distribution pattern for the whole high-confidence genes is likely different in that the genes with the most expression in spike (28.2%) account for most, which was followed by those in root (23.6%), grain (15.3%) and leaf (13.3%) (Fig. 2).

Distribution patterns of GO terms between the single subgenome genes and the whole high-confidence genes

Of the 433 SSGs, 248 (57%) genes were successfully annotated and mapped to one or more of the three organizing principles of GO classifications: biological process, molecular function and cellular component (Table S1). In the biological process class, the most abundant groups represented were metabolic process and cellular process ranging from 25% to 29% for the A, B or D-only genes (Fig. 3). For molecular function, catalytic activity (48% to 53%) was the most represented GO term followed by binding (33% to 38%) for A, B, or D-only genes (Fig. S2). Within cellular component, major categories were cell (37% to 44%) and organelle (31% to 38%; Fig. S3).

Significant correlation ($R > 0.94$, $P < 0.01$) was detected between distribution patterns of the whole high-confidence genes and each of A, B and D-only

genes for biological process, molecular function and cellular component (Fig. 3, Fig. S2 and Fig. S3). These results indicated that the distribution patterns of GO terms for SSGs are not significantly different from those of the whole high-confidence genes. Further, enrichment analysis (Fisher's Exact Test) did not detect any GO term that was significantly different between SSGs and the whole genes.

Functional relationships between the single subgenome genes

Of the 433 SSGs, only 44 were identified to be orthologous to 42 unique genes of *Arabidopsis*. Network analysis identified 7 genes only in a given network and the majority are not functionally related to each other (Fig. 4). For a given chromosome, genes detected outside of networks were 2.5 times more than those within networks (Table S2). The numbers of genes within in networks for a given group were not significant. For example, 27%, 28%, and 26% of genes from 1A, 1B, and 1D, respectively, were identified to be involved into networks (Table S2 and Fig. S4).

KEGG pathway analysis assigned 43 of the 433 SSGs to KO (KEGG orthology). These 43 genes were involved in 29 pathways with most in starch and sucrose metabolism (6) followed by glycerolipid metabolism (4) (Table 3). Pathway analysis again revealed that the majority of these genes are not likely functionally related.

Discussion

Taking advantage of the recently released 'Chinese Spring' chromosome arm shotgun sequences together with high-confidence gene models, we carried out a systematic assessment of the genes retained on a single subgenome. A total of 433 genes being on a single subgenome only were detected. It is noteworthy that the possible mis-annotated gene models analysed in the present study and imperfect 'CS' sequence assembly could lead to overestimation of the identified SSGs. Although a loose criterion (an E-value threshold of 10^{-5}) was adopted for detecting possible target genes, it is still possible that some potential SSGs could be missed out. This is based on the fact that even if only a small part of the query gene is aligned, this might still

result in an E-value less than 10^{-5} . Thus, a potential SSG could be mistaken as a homolog given that it is aligned to the other subgenome(s). Greater gene loss was observed in A and D subgenomes compared with B subgenome. More than 79% of the orthologs for these SSGs were detected in the diploid and tetraploid relatives of hexaploid wheat. Unexpectedly, no bias in expression breadth, or distribution patterns of Go terms for these genes was detected compared with those for all of the high-confidence genes. Further, network and KEGG pathway analyses indicated that most of these SSGs were not functionally related. Interestingly, 30.7% of these SSGs are most highly expressed in root, followed by spike (22.6%) and grain (21.9%) showing biased distribution compared to the expression patterns of the whole high-confidence genes. Together, these results improve our understanding of the loss of the genes that were kept in a single subgenome.

Most recently, the IWGSC observed that gene loss in the D subgenome appeared slightly lower than A and B subgenomes based on comparisons between hexaploid, tetraploid and diploid genomes (IWGSC, 2014). This findings was in accordance with previous observation by Berkman et al. (2013) who detected greater gene loss in 7A and 7B genomes than 7D genome by analysing against assembled contigs representing genes for chromosomes 7A, 7B and 7D in bread wheat. In our present study, greater gene loss was observed in A and D subgenomes compared with B subgenome. The reason for this difference is probably that the analysis performed by the IWGSC was focused on all the gene loss events including those identified in this study, thus hiding the real picture of gene loss for the single subgenome genes. Despite using strict criteria in this study, we indeed detected similar gene loss events. For example, more single subgenome genes were detected in D subgenome compared with A and B subgenomes in chromosomes 4, 6 and 7 (Table 1). Both the results presented here and those from previous studies showed differential gene loss between A, B and D subgenomes, further supporting the proposed theories for selective gene retention and loss after polyploidization (Bekaert et al. 2011; De Smet et al. 2013).

The fact that a majority of the SSGs were detected in diploid and tetraploid relatives of the hexaploid wheat indicated that most of SSGs were lost during

the formation of hexaploid wheat. It is widely accepted that gene loss occurred in A and B genomes following tetraploidization, with succeeding loss from all three genomes following the formation of the hexaploid wheat (Berkman et al. 2013). It is interesting that a few of the SSGs could not detect their orthologs in diploid and tetraploid relatives. We could not rule out the possibility that it is the imperfect sequence assembly that prevents such of these SSGs not to find blast hits against assembly contigs. It is also possible that most if not all of these SSGs are likely newly created genes during polyploidization.

Numerous studies observed that genes that experienced loss during evolution were not to be a random fraction of the genome. For example, De Smet et al. (2013) found that a large number of single-copy genes (duplication-resistant genes) detected in the sequenced genomes of 20 flowering plants are often involved in essential housekeeping functions. Another example is about the differential gene retention between 7A, 7B and 7D genomes. Berkman et al. (2013) concluded that 7A and 7B genomes have preferentially kept genes participating in basic cellular processes, while 7D genome has mainly retained a greater proportion of genes relating to specialized process. However, in this study, the loss of the SSGs likely occurred randomly. The SSGs are mainly expressed in more than 13 conditions, similar to the expression pattern of housekeeping genes and they as well showed clear differential GO classifications in biological process, molecular function and cellular component. Nevertheless, the analysis of the whole high-confidence genes revealed that similar proportion of genes is expressed in more than 13 conditions, and SSGs are proportionally distributed among the GO terms. In addition, the network and pathway analyses failed to detect significant functionally relatedness between SSGs.

Intriguingly, 30.7% of the SSGs are most highly expressed in root, followed by spike (22.6%) and grain (21.9%). This is different from the patterns of the whole high confidence genes (Fig. 2). This differential distribution allows us to speculate that the SSGs are likely preferentially expressed in root. During formation of hexaploid wheat, the genes with expression enriched in root are prone to be lost. It is likely that these SSGs are sufficient to support the growth and development of the plant without the copies from other

subgenomes. That is gene loss of these SSGs from other subgenomes is tolerated for the plant.

Acknowledgements

This publication is based upon work supported by the projects from the National Natural Science Foundation of China (31601292,31570335). We appreciate the anonymous referees for critical reading of the manuscript.

Reference

- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**(17): 3389-3402. doi: 10.1093/nar/25.17.3389.
- Bekaert, M., Edger, P.P., Pires, J.C., and Conant, G.C. 2011. Two-phase resolution of polyploidy in the Arabidopsis metabolic network gives rise to relative and absolute dosage constraints. *The Plant Cell* **23**(5): 1719-1728. doi: 10.1105/tpc.110.081281.
- Berkman, P.J., Visendi, P., Lee, H.C., Stiller, J., Manoli, S., Lorenc, M.T., Lai, K., Batley, J., Fleury, D., and Šimková, H. 2013. Dispersion and domestication shaped the genome of bread wheat. *Plant Biotechnol. J.* **11**(5): 564-571. doi: 10.1111/pbi.12044.
- Brenchley, R., Spannagl, M., Pfeifer, M., Barker, G.L., D'Amore, R., Allen, A.M., McKenzie, N., Kramer, M., Kerhornou, A., and Bolser, D. 2012. Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature* **491**(7426): 705-710. doi: 10.1038/nature11650.
- Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., and Robles, M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**(18): 3674-3676. doi: 10.1093/bioinformatics/bti610.
- De Smet, R., Adams, K.L., Vandepoele, K., Van Montagu, M.C., Maere, S., and Van de Peer, Y. 2013. Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. *Proc. Natl. Acad. Sci. USA* **110**(8): 2898-2903. doi: 10.1073/pnas.1300127110.
- Duarte, J.M., Wall, P.K., Edger, P.P., Landherr, L.L., Ma, H., Pires, J.C., and Leebens-Mack, J. 2010. Identification of shared single copy nuclear genes in Arabidopsis, Populus, Vitis and Oryza and their phylogenetic utility across various taxonomic levels. *BMC Evol. Biol.* **10**(1): 1. doi: 10.1186/1471-2148-10-61.
- Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., and von Mering, C. 2013. STRING v9. 1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* **41**(D1): D808-D815. doi: 10.1093/nar/gks1094.
- International Wheat Genome Sequencing Consortium (IWGSC). 2014. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* **345**:1251788. doi: 10.1126/science.1251788.
- Ma, J., Stiller, J., Berkman, P.J., Wei, Y., Rogers, J., Feuillet, C., Dolezel, J., Mayer, K.F., Eversole, K., and Zheng, Y.-L. 2013. Sequence-based analysis of translocations and inversions in bread wheat (*Triticum aestivum* L.). *PloS one* **8**(11): e79329. doi: 10.1371/journal.pone.0079329.
- Ma, J., Stiller, J., Wei, Y., Zheng, Y.-L., Devos, K.M., Doležel, J., and Liu, C. 2014a. Extensive Pericentric Rearrangements in the Bread Wheat (*Triticum aestivum* L.) Genotype

- "Chinese Spring" Revealed from Chromosome Shotgun Sequence Data. *Genome Biol. Evol.* **6**(11): 3039-3048. doi: 10.1093/gbe/evu237.
- Ma, J., Stiller, J., Zhao, Q., Feng, Q., Cavanagh, C., Wang, P., Gardiner, D., Choulet, F., Feuillet, C., and Zheng, Y.-L. 2014b. Transcriptome and Allele Specificity Associated with a 3BL Locus for Fusarium Crown Rot Resistance in Bread Wheat. *PloS one* **9**(11): e113309. doi: 10.1371/journal.pone.0113309.
- Ma, J., Stiller, J., Zheng, Z., Wei, Y., Zheng, Y.-L., Yan, G., Doležel, J., and Liu, C. 2015. Putative interchromosomal rearrangements in the hexaploid wheat (*Triticum aestivum* L.) genotype 'Chinese Spring' revealed by gene locations on homoeologous chromosomes. *BMC Evol. Biol.* **15**(1): 37. doi: 10.1186/s12862-015-0313-5.
- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., and Kanehisa, M. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **35**(suppl 2): W182-W185. doi: 10.1093/nar/gkm321.
- Paterson, A.H., Chapman, B.A., Kissinger, J.C., Bowers, J.E., Feltus, F.A., and Estill, J.C. 2006. Many gene and domain families have convergent fates following independent whole-genome duplication events in *Arabidopsis*, *Oryza*, *Saccharomyces* and *Tetraodon*. *Trends Genet.* **22**(11): 597-602. doi: 10.1016/j.tig.2006.09.003.
- Petersen, G., Seberg, O., Yde, M., and Berthelsen, K. 2006. Phylogenetic relationships of *Triticum* and *Aegilops* and evidence for the origin of the A, B, and D genomes of common wheat (*Triticum aestivum*). *Mol. Phylogen. Evol.* **39**(1): 70-82. doi: 10.1016/j.ympev.2006.01.023.
- Salamini, F., Özkan, H., Brandolini, A., Schäfer-Pregl, R., and Martin, W. 2002. Genetics and geography of wild cereal domestication in the Near East. *Nature Reviews Genetics* **3**(6): 429-441. doi: 10.1038/nrg817.
- Scannell, D.R., Frank, A.C., Conant, G.C., Byrne, K.P., Woolfit, M., and Wolfe, K.H. 2007. Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication. *Proc. Natl. Acad. Sci. USA* **104**(20): 8397-8402. doi: 10.1073/pnas.0608218104.
- Schnable, J.C., Freeling, M., and Lyons, E. 2012. Genome-wide analysis of syntenic gene deletion in the grasses. *Genome Biol. Evol.* **4**(3): 265-277. doi: 10.1093/gbe/evs009.
- Zhang, J. 2003. Evolution by gene duplication: an update. *Trends Ecol. Evol.* **18**(6): 292-298. doi: 10.1016/s0169-5347(03)00033-8.

Figure legends:

Fig.1 Distribution of expression breadth for the whole 99,386 high-confidence genes and the single subgenome genes. X axis, gene breadth, i.e. that is, the number of conditions in which the genes was expressed, from 0 to 15. Y axis, A, A-only genes; B, B-only genes; D, D-only genes; All, All genes. Z axis, the percentage of total genes.

Fig. 2 The number of genes expressed most in a given tissue for the whole 99,386 high-confidence genes and the single subgenome genes. A, the whole 99,386 high-confidence genes; B, A-only genes; C, B-only genes; D, D-only genes.

Fig.3 Distribution of genes with a functional annotation (biological process) for the whole 99,386 high-confidence genes and the single subgenome genes(**A**) and the correlation between distribution patterns of the whole high-confidence genes and each of A, B, and D-only genes (**B, C, D**)

Fig.4 Network analysis of single subgenome genes

Draft

Supplementary figure legends:

Fig. S1 Correlation between gene amount and genome size. A, A-only genes; B, B-only genes; C, D-only genes Y axis, genome size (Mb); X axis, gene number

Fig.S2 Distribution of genes with a functional annotation (molecular function) for the whole 99,386 high-confidence genes and the single subgenome genes (**A**) and the correlation between distribution patterns of the whole high-confidence genes and each of A, B, and D-only genes (**B, C, D**)

Fig.S3 Distribution of genes with a functional annotation (cellular component) for the whole 99,386 high-confidence genes and the single subgenome genes (**A**) and the correlation between distribution patterns of the whole high-confidence genes and each of A, B, and D-only genes (**B, C, D**)

Fig.S4 Network analysis of genes from 1A, 1B and 1D chromosomes.

Draft

Table 1. Details of genes with a single subgenome location

Chromosome	1	2	3	4	5	6	7
A only [76]	13	4	5	17	0	10	27
short arm [38]	0	2	2	5	0	5	24
long arm [38]	13	2	3	12	0	5	3
B only [226]	18	12	130	14	29	8	15
short arm [21]	2	7	-	5	1	2	4
long arm [75]	16	5	-	9	28	6	11
D only [131]	13	5	7	20	6	32	48
short arm [35]	1	3	1	11	1	1	17
long arm [96]	12	2	6	9	5	31	31
total	44	21	142	51	35	50	90

Draft

Table 2. Detection of orthologs of single subgenome genes in diploid and tetraploid relatives of wheat

Genome composition	A ^u A ^u	A ^m A ^m	S ^{sh} S ^{sh}	SS	AABB	AABB	DD
Species	<i>T.urartu</i>	<i>T.monococcum</i>	<i>Ae.sharonensis</i>	<i>Ae.speltoides</i>	<i>T.turgidum</i> cv. <i>T.turgidum</i> cv.Cappelli	<i>T.turgidum</i> cv. Strongfield	<i>Ae.tauschii</i>
A only [76]	68	64	66	61	74	71	63
B only [226]	193	194	194	202	217	201	190
D only [131]	104	106	107	116	106	104	121

Draft

Table 3. KEGG analysis of single subgenome genes

Pathway	Gene number
Starch and sucrose metabolism	6
Glycerolipid metabolism	4
alpha-Linolenic acid metabolism	2
Amino sugar and nucleotide sugar metabolism	2
Glyoxylate and dicarboxylate metabolism	2
Linoleic acid metabolism	2
Pentose and glucuronate interconversions	2
Purine metabolism	2
Aminoacyl-tRNA biosynthesis	1
Arginine and proline metabolism	1
Carotenoid biosynthesis	1
Cysteine and methionine metabolism	1
Drug metabolism - other enzymes	1
Flavone and flavonol biosynthesis	1
Flavonoid biosynthesis	1
Galactose metabolism	1
Glycine, serine and threonine metabolism	1
Glycosaminoglycan biosynthesis - chondroitin sulfate / dermatan sulfate	1
Glycosaminoglycan biosynthesis - heparan sulfate / heparin	1
Methane metabolism	1
Other glycan degradation	1
Pantothenate and CoA biosynthesis	1
Photosynthesis	1
Porphyrin and chlorophyll metabolism	1
Pyrimidine metabolism	1
Selenocompound metabolism	1
T cell receptor signaling pathway	1
Thiamine metabolism	1
Various types of N-glycan biosynthesis	1

222x165mm (150 x 150 DPI)

165x148mm (300 x 300 DPI)

453x319mm (100 x 100 DPI)

177x216mm (118 x 78 DPI)

243x167mm (300 x 300 DPI)

464x337mm (100 x 100 DPI)

478x329mm (100 x 100 DPI)

1A

1B

1D

216x182mm (150 x 150 DPI)

Table S1 Details of the 433 single subgenome genes

Gene model	Chromosome	Annotation
Traes_1AL_OA12DA18E.1	1AL	uncharacterized protein loc100846419
Traes_1AL_0B4B6F75A.1	1AL	cell wall invertase
Traes_1AL_17C37FFB0.1	1AL	serine threonine-protein kinase chloroplastic-like
Traes_1AL_269D05D49.1	1AL	
Traes_1AL_51E6DEAC8.1	1AL	uncharacterized protein loc101762814
Traes_1AL_695C34A5A.1	1AL	
Traes_1AL_A1EA76E83.1	1AL	trd1
Traes_1AL_B31DC4745.1	1AL	uncharacterized protein loc100821421
Traes_1AL_BC49C38C1.1	1AL	
Traes_1AL_BE367D600.1	1AL	
Traes_1AL_C65EBCBBA.1	1AL	
Traes_1AL_D9422FB50.1	1AL	
Traes_1AL_F05ABAEDB.1	1AL	ethylene-responsive transcription factor erf091-like
Traes_1BL_1382B02DC.1	1BL	
Traes_1BL_1C9DFF36A.1	1BL	
Traes_1BL_2D221F56E.1	1BL	
Traes_1BL_3CAA69266.1	1BL	
Traes_1BL_4B39056D7.1	1BL	cyclic nucleotide-gated ion channel 4
Traes_1BL_514DCE760.1	1BL	
Traes_1BL_7517181F2.1	1BL	tpa: gid1-like gibberellin receptor
Traes_1BL_84977D514.1	1BL	methionine gamma-lyase-like
Traes_1BL_97C47ECEA.1	1BL	gibberellin 2-oxidase
Traes_1BL_99E050616.1	1BL	
Traes_1BL_A62B4683A.1	1BL	
Traes_1BL_A64CEC09F.1	1BL	
Traes_1BL_A9A3B8388.1	1BL	uncharacterized acetyltransferase at3g50280-like
Traes_1BL_EF484D43E.1	1BL	
Traes_1BL_F247C7924.1	1BL	

Traes_1BL_FE3E68ED7.1	1BL	cation h(+) antiporter 15-like
Traes_1BS_25DF6A270.1	1BS	
Traes_1BS_B61F8895C.1	1BS	
Traes_1DL_106460E68.1	1DL	
Traes_1DL_21D81B0C2.1	1DL	
Traes_1DL_24D22D03F.1	1DL	uncharacterized protein loc100383132
Traes_1DL_3FC530016.1	1DL	
Traes_1DL_49C4E170C1.1	1DL	
Traes_1DL_5353195A7.1	1DL	homeobox-leucine zipper protein roc3
Traes_1DL_5E5E0F267.1	1DL	
Traes_1DL_6E46FF4B9.1	1DL	uncharacterized protein loc101776195
Traes_1DL_76636DBE5.1	1DL	cation h(+) antiporter 15-like
Traes_1DL_794982B23.1	1DL	benzyl alcohol o-benzoyltransferase-like
Traes_1DL_C4B334B3F1.1	1DL	uncharacterized protein loc100277795
Traes_1DL_C7C2228B7.1	1DL	pollen allergen phl p 11
Traes_1DS_6F82D5B34.1	1DS	
Traes_2AL_A8690B42F.1	2AL	mitogen-activated protein kinase kinase kinase a-like
Traes_2AL_DAA33562C.1	2AL	
Traes_2AS_AB8F65F31.1	2AS	ankyrin repeat domain-containing protein 13b-like
Traes_2AS_B7F0117A8.1	2AS	
Traes_2BL_2C7720367.1	2BL	alpha beta fold family expressed
Traes_2BL_41F803764.1	2BL	carotenoid cleavage dioxygenase 7
Traes_2BL_666C4AF9E.1	2BL	
Traes_2BL_D2B962D83.1	2BL	u-box domain-containing protein 19-like
Traes_2BL_EB7A76EA2.1	2BL	
Traes_2BS_19D8CE70D.1	2BS	
Traes_2BS_373595180.1	2BS	coatamer beta subunit family protein
Traes_2BS_488719EBA.1	2BS	uncharacterized protein loc101761302
Traes_2BS_C2CCB6F8E.1	2BS	ankyrin repeat protein family-like protein
Traes_2BS_C527278E9.1	2BS	predicted protein [Hordeum vulgare subsp. vulgare]
Traes_2BS_D1D4A98D2.1	2BS	

Traes_2BS_DC737B3E0. 1	2BS	
Traes_2DL_80FFB7F5C. 1	2DL	
Traes_2DL_E6CD5A5D0. 1	2DL	cysteine histidine-rich c1 domain-containing protein
Traes_2DS_1783BD388. 1	2DS	
Traes_2DS_38022119B. 1	2DS	beta- galactosyltransferase 6
Traes_2DS_B5B2EC8AC. 1	2DS	
Traes_3AL_2AF594324. 1	3AL	
Traes_3AL_7A10DF246. 1	3AL	
Traes_3AL_EE4F2CFBE. 1	3AL	
Traes_3AS_D5CE7702B. 1	3AS	
Traes_3AS_E10DBC85. 1	3AS	
Traes_3B_004D9ADD2. 1	3B	u-box domain-containing protein 16-like
Traes_3B_035C96954. 1	3B	
Traes_3B_03A8319F0. 1	3B	nucleic acid binding protein
Traes_3B_04BBC74A7. 1	3B	hydroxyethylthiazole kinase-like
Traes_3B_05D135D69. 1	3B	
Traes_3B_0791B3EBE. 1	3B	protein
Traes_3B_081CD15BF. 1	3B	universal stress protein a-like
Traes_3B_098D62859. 1	3B	
Traes_3B_0A288E4321. 1	3B	
Traes_3B_0DF56B2BB. 1	3B	l-type lectin-domain containing receptor kinase -like
Traes_3B_1235C7C8A. 1	3B	gdsl esterase lipase at4g10955-like
Traes_3B_16AE3D594. 1	3B	
Traes_3B_1AC570E3A. 1	3B	piriformospora indica-insensitive protein 2-like
Traes_3B_1B6B31737. 1	3B	mitochondrial carnitine acylcarnitine carrier protein cacl
Traes_3B_1DE79A69D. 1	3B	upf0678 fatty acid-binding protein atlg79260-like
Traes_3B_1E3EB01CC. 1	3B	disease resistance protein rpp13-like
Traes_3B_204A47675. 1	3B	uncharacterized protein loc101759345
Traes_3B_20B9ABB22. 1	3B	
Traes_3B_216EC5695. 1	3B	
Traes_3B_22CED8627. 1	3B	transcription initiation factor tfiid subunit 11

Traes_3B_2465D5F97.1	3B	uncharacterized protein loc100831420
Traes_3B_24DF679511.1	3B	lipid-transfer protein dir1-like
Traes_3B_24ED7666D.1	3B	serine-type endopeptidase inhibitor
Traes_3B_25B6CEFD4.1	3B	dna-binding wrky
Traes_3B_29A81F834.1	3B	cysteine proteinase inhibitor 4
Traes_3B_2CE02BF1E.1	3B	disease resistance protein rppl3-like
Traes_3B_2F4607DC5.1	3B	
Traes_3B_34AF28E01.1	3B	probable polyamine oxidase 5-like
Traes_3B_37633E82E.1	3B	transposon protein mutator sub-class
Traes_3B_3AC5C5128.1	3B	gata transcription factor 20-like
Traes_3B_3CFCC10CE.1	3B	crs2-associated factor 2
Traes_3B_3F1BBE2C2.1	3B	viviparous1 protein
Traes_3B_40899AF7E.1	3B	
Traes_3B_42A08E203.1	3B	predicted protein [Hordeum vulgare subsp. vulgare]
Traes_3B_45D5CEEB4.1	3B	nuclease harbil-like
Traes_3B_465485976.1	3B	transcription factor bhlh13-like
Traes_3B_46EA630E4.1	3B	
Traes_3B_489984CE3.1	3B	ribosomal protein s1-like rna-binding domain-containing protein
Traes_3B_4E29C353B.1	3B	transcription factor bhlh13-like
Traes_3B_4F11E8997.1	3B	uncharacterized protein loc100822854
Traes_3B_4F89490D5.1	3B	viviparous1 protein
Traes_3B_514AAB5F3.1	3B	
Traes_3B_5326304FD.1	3B	
Traes_3B_5361D7AFE.1	3B	signal peptidase complex subunit 1
Traes_3B_53E723173.1	3B	uncharacterized protein loc100824905 isoform 1
Traes_3B_547FDBB4D.1	3B	neutral invertase
Traes_3B_557FAA07B.1	3B	
Traes_3B_58309CFB2.1	3B	uridine 5 -monophosphate synthase-like
Traes_3B_59270D484.1	3B	
Traes_3B_609886503.1	3B	
Traes_3B_62FE1820C.1	3B	

Traes_3B_6434A8AA5.1	3B	
Traes_3B_64395F50E.1	3B	
Traes_3B_653BDB9F7.1	3B	xylanase inhibitor
Traes_3B_6594302C6.1	3B	cyclic nucleotide-gated ion channel 4-like
Traes_3B_663B77701.1	3B	glutamate carboxypeptidase 2
Traes_3B_67D605CF81.1	3B	
Traes_3B_67DBA80DE.1	3B	beta-glucan-binding protein
Traes_3B_6C5335E85.1	3B	2-hydroxyacyl- lyase-like
Traes_3B_6D71543F9.1	3B	uncharacterized protein loc100835835
Traes_3B_6F128E9A0.1	3B	
Traes_3B_6FA318BF3.1	3B	
Traes_3B_6FA801BBF.1	3B	
Traes_3B_73B1D08C5.1	3B	uncharacterized protein loc100835227
Traes_3B_744EDC56C.1	3B	uncharacterized protein loc100839687
Traes_3B_769275DE3.1	3B	
Traes_3B_7AC0971AD.1	3B	
Traes_3B_7E85063F2.1	3B	
Traes_3B_7EBCB2ED3.1	3B	pectinesterase pectinesterase inhibitor 18
Traes_3B_7EF47FD09.1	3B	
Traes_3B_856C6F598.1	3B	
Traes_3B_8B9B2BEA1.1	3B	
Traes_3B_8BCC7D2B7.1	3B	
Traes_3B_8C616135A.1	3B	
Traes_3B_8CB8DFA05.1	3B	harpin-induced protein
Traes_3B_921D57F5D.1	3B	
Traes_3B_928A075B3.1	3B	uncharacterized protein loc100836328
Traes_3B_9635D0895.1	3B	pyruvate kinase isozyme chloroplastic
Traes_3B_96705EE0F.1	3B	seed maturation protein pm41
Traes_3B_9E2CA92E7.1	3B	c2h2 zinc finger protein
Traes_3B_9F2AF0F06.1	3B	
Traes_3B_A1A1249AF.1	3B	b3 domain-containing protein

Traes_3B_A5EEE27621.1	3B	grx_a2 - glutaredoxin subgroup iii
Traes_3B_AB2C11A33.1	3B	abc transporter g family member 16-like
Traes_3B_AE68805D8.1	3B	myc2
Traes_3B_AF1377E5A.1	3B	
Traes_3B_B355E7E96.1	3B	squamosa promoter-binding-like protein 2-like
Traes_3B_B3CA55F42.1	3B	
Traes_3B_B6768190E.1	3B	cation h(+) antiporter 15
Traes_3B_BBF8616A5.1	3B	lipase class 3 family
Traes_3B_BEFC0B924.1	3B	
Traes_3B_C13ABF319.1	3B	cysteine proteinase inhibitor 4-like
Traes_3B_C1D8E152A.1	3B	
Traes_3B_C37503560.1	3B	
Traes_3B_C3A2E66D8.1	3B	cbs domain-containing protein cbsx6-like
Traes_3B_C42175777.1	3B	tpa: leucine-rich repeat transmembrane protein kinase family protein
Traes_3B_C47F3F2EB.1	3B	uv-stimulated scaffold protein a homolog
Traes_3B_C4B230D7E.1	3B	
Traes_3B_C58B82890.1	3B	
Traes_3B_C58B828901.1	3B	
Traes_3B_C613F9880.1	3B	
Traes_3B_C9E9632E8.1	3B	cysteine proteinase inhibitor 4
Traes_3B_CAA8F66B9.1	3B	
Traes_3B_CB8CD7F0C.1	3B	
Traes_3B_CC7A20699.1	3B	gdsl esterase lipase at4g10955-like
Traes_3B_D001A64DA.1	3B	
Traes_3B_D00AEDAF6.1	3B	
Traes_3B_D1BF3280B.1	3B	serine threonine-protein kinase bril-like 2-like
Traes_3B_D30684D6C.1	3B	cysteine proteinase
Traes_3B_D6F298575.1	3B	uncharacterized protein loc100845800
Traes_3B_D794E2D14.1	3B	ap2 erebp transcription factor superfamily protein
Traes_3B_D825EC594.1	3B	mitochondrial carnitine acylcarnitine carrier protein cacl-like
Traes_3B_DEB6C5A5C.1	3B	uncharacterized protein loc100824905 isoform 2

Traes_3B_DF2406DC4.1	3B	
Traes_3B_E43E7379E.1	3B	sorting nexin-2-like
Traes_3B_E4896C4E9.1	3B	f-box domain containing protein
Traes_3B_E76AFA5F.1	3B	
Traes_3B_E8ACC946D.1	3B	pentatricopeptide repeat-containing protein at5g66631-like
Traes_3B_E956F4136.1	3B	viviparous1 protein
Traes_3B_EC9D41354.1	3B	dead-box atp-dependent rna helicase 40-like
Traes_3B_ED7524E191.1	3B	skip interacting protein 34
Traes_3B_F0217CAB0.1	3B	
Traes_3B_F14C64C8B.1	3B	
Traes_3B_F3A5AA0D8.1	3B	f-box and wd40 domain
Traes_3B_F46CB2D2E.1	3B	
Traes_3B_F52EDA13C.1	3B	protein kinase-like
Traes_3B_F7FE991F71.1	3B	uncharacterized protein loc100842752
Traes_3B_FAF29A136.1	3B	
Traes_3B_FC2323C30.1	3B	
Traes_3B_FDF72A0C4.1	3B	
Traes_3DL_1B94A9AC3.1	3DL	
Traes_3DL_45B528FFC.1	3DL	
Traes_3DL_B25ABA25A.1	3DL	pentatricopeptide repeat-containing protein at3g21470-like
Traes_3DL_C0B28B652.1	3DL	
Traes_3DL_D6277566A.1	3DL	
Traes_3DL_EA334F981.1	3DL	
Traes_3DS_8AFD0BD51.1	3DS	
Traes_4AL_29D81B290.1	4AL	uncharacterized protein loc100825782
Traes_4AL_2D1E33772.1	4AL	magnesium-protoporphyrin ix methyltransferase
Traes_4AL_30684624B.1	4AL	ribonuclease 3-like
Traes_4AL_351ACDBA9.1	4AL	
Traes_4AL_398B38CAB.1	4AL	helicase-like protein
Traes_4AL_6D83B6137.1	4AL	predicted protein [Hordeum vulgare subsp. vulgare]
Traes_4AL_77913718B.1	4AL	

Traes_4AL_924469221.1	4AL	
Traes_4AL_AB996C292.1	4AL	
Traes_4AL_B23E0F739.1	4AL	
Traes_4AL_D035DB895.1	4AL	
Traes_4AL_F73110922.1	4AL	scarecrow-like protein 28-like
Traes_4AS_0E7A99FE1.1	4AS	sac3 family protein 1-like
Traes_4AS_60A7E57AA.1	4AS	transport inhibitor response 1-like
Traes_4AS_A113A90A5.1	4AS	probable pectinesterase pectinesterase inhibitor 51-like
Traes_4AS_B1C46AD98.1	4AS	
Traes_4AS_D6E6B310D.1	4AS	
Traes_4BL_1B053B606.1	4BL	
Traes_4BL_39D269E13.1	4BL	u3 small nucleolar rna-associated protein 15 homolog
Traes_4BL_3F4480981.1	4BL	hypothetical protein OsJ_35456 [Oryza sativa Japonica Group]
Traes_4BL_4EC4C9D9D.1	4BL	alpha beta fold family expressed
Traes_4BL_754BEE2B7.1	4BL	
Traes_4BL_7E0C13C00.1	4BL	
Traes_4BL_A7420FB2A.1	4BL	probable carboxylesterase 17-like
Traes_4BL_D5362D9D7.1	4BL	type 1 non specific lipid transfer protein precursor
Traes_4BL_ECF34452D.1	4BL	uncharacterized protein loc100821550
Traes_4BS_375EF1D15.1	4BS	squamosa promoter-binding-like protein 2-like
Traes_4BS_467838FA9.1	4BS	uncharacterized protein loc100843204 isoform 1
Traes_4BS_59715D886.1	4BS	
Traes_4BS_5E37B93A8.1	4BS	della protein
Traes_4BS_7D6136FEC.1	4BS	alpha beta fold family expressed
Traes_4DL_122193BA9.1	4DL	
Traes_4DL_47B80EA6B.1	4DL	
Traes_4DL_4D4517045.1	4DL	uncharacterized protein loc100828230
Traes_4DL_892BCAEE9.1	4DL	
Traes_4DL_996028E97.1	4DL	v-type proton atpase 21 kda proteolipid subunit-like
Traes_4DL_DD0565453.1	4DL	
Traes_4DL_E7D476AE1.1	4DL	uncharacterized protein loc101784037

Traes_4DL_FCB128818.1	4DL	mitogen-activated protein kinase kinase kinase a-like
Traes_4DL_FE2566C8B.1	4DL	galactomannan galactosyltransferase 1-like
Traes_4DS_07EEAB06B.1	4DS	v-type proton atpase subunit g1
Traes_4DS_24BB9E06A.1	4DS	ferredoxin--nadp leaf isozyme 2
Traes_4DS_251335DBE.1	4DS	
Traes_4DS_316823E57.1	4DS	protein plant cadmium resistance 2-like
Traes_4DS_5A55A295B.1	4DS	
Traes_4DS_A98066B17.1	4DS	50s ribosomal protein l29
Traes_4DS_AFB7DD996.1	4DS	40s ribosomal protein s30
Traes_4DS_B4D406794.1	4DS	hydroxypyruvate reductase
Traes_4DS_B514BC2D3.1	4DS	peroxisomal -2-hydroxy-acid oxidase glo1
Traes_4DS_B7B909DC0.1	4DS	peroxisomal -2-hydroxy-acid oxidase glo1
Traes_4DS_CC634AFD2.1	4DS	atlg67740 f12a21_i3
Traes_5BL_05AC7CEC0.1	5BL	nodulin-like protein
Traes_5BL_0BFAE33B9.1	5BL	
Traes_5BL_166A442BE.1	5BL	gdsl esterase lipase at5g55050-like
Traes_5BL_1992B070F.1	5BL	
Traes_5BL_25277B66A.1	5BL	uncharacterized protein loc100836395
Traes_5BL_3619A8360.1	5BL	predicted protein [Hordeum vulgare subsp. vulgare]
Traes_5BL_377F4D4C6.1	5BL	
Traes_5BL_3F537AD8F.1	5BL	tpa: transferase
Traes_5BL_43BC70A68.1	5BL	uncharacterized protein loc100828324
Traes_5BL_43BC70A681.1	5BL	uncharacterized protein loc100828324
Traes_5BL_523BB4276.1	5BL	thioredoxin domain-containing protein 9 homolog
Traes_5BL_57554704B.1	5BL	uncharacterized protein loc100835056
Traes_5BL_612C62DE9.1	5BL	pentatricopeptide repeat-containing protein chloroplastic-like
Traes_5BL_6F6DB0191.1	5BL	alpha-1-fucosidase 1-like
Traes_5BL_7035A391D.1	5BL	
Traes_5BL_7FE1F579A.1	5BL	
Traes_5BL_873E079B3.1	5BL	
Traes_5BL_A3467C588.1	5BL	

Traes_5BL_B8F044FF1.1	5BL	lipoxygenase chloroplastic
Traes_5BL_C1F897A4A.1	5BL	ferredoxin
Traes_5BL_C2F35ACFA.1	5BL	wd40-like beta propeller repeat family expressed
Traes_5BL_CD5F2AF30.1	5BL	abc transporter c family member 4-like
Traes_5BL_D18B24652.1	5BL	harpin-induced family
Traes_5BL_DFDF115A6.1	5BL	uncharacterized protein loc100845523
Traes_5BL_E40624C02.1	5BL	u-box domain-containing protein 16-like
Traes_5BL_EE136B1C2.1	5BL	tpa: arm repeat-containing protein containing family protein
Traes_5BL_F4CDCA1B4.1	5BL	uncharacterized protein loc100838527
Traes_5BL_FC28DECD0.1	5BL	
Traes_5BS_E63B64042.1	5BS	
Traes_5DL_OD3054167.1	5DL	disease resistance-responsive (dirigent-like protein) family protein
Traes_5DL_2DC378706.1	5DL	myb-related protein 306-like
Traes_5DL_385138CF0.1	5DL	probable beta- galactosyltransferase 14-like
Traes_5DL_4A1A62C7E.1	5DL	ac090713_7 proteophosphoglycan
Traes_5DL_AAF4C484E.1	5DL	nbs-lrr protein
Traes_5DS_E9025A862.1	5DS	universal stress protein family protein
Traes_6AL_5A0E5217B.1	6AL	
Traes_6AL_5BA58668E1.1	6AL	l-type lectin-domain containing receptor kinase
Traes_6AL_625095415.1	6AL	
Traes_6AL_6C26DE79E.1	6AL	diphthamide synthesis dph2-like
Traes_6AL_92DE09356.1	6AL	udp-glucuronate 4-epimerase 1-like
Traes_6AS_5B414F54C.1	6AS	mitochondrial substrate carrier family protein b-like
Traes_6AS_6D7582A97.1	6AS	lrr receptor-like serine threonine-protein kinase
Traes_6AS_CF0AB9BAB.1	6AS	
Traes_6AS_E06AB7E22.1	6AS	phytosulfokine receptor 1-like
Traes_6AS_E2FA0D8B2.1	6AS	u-box domain-containing protein 4-like isoform 1
Traes_6BL_218BAC3E0.1	6BL	
Traes_6BL_4F6CB9FE9.1	6BL	
Traes_6BL_75986E4E0.1	6BL	
Traes_6BL_8B6B6B2EB.1	6BL	

Traes_6BL_8B6B6B2EB1.1	6BL	
Traes_6BL_B84BC12EC.1	6BL	
Traes_6BS_CA0E2C9F2.1	6BS	
Traes_6BS_F87437638.1	6BS	
Traes_6DL_013E2BAB4.1	6DL	protein tic 20- chloroplastic-like
Traes_6DL_077ADF9F9.1	6DL	triacylglycerol lipase like protein
Traes_6DL_1DD0FF151.1	6DL	lmb1 domain-containing protein 2 homolog a-like
Traes_6DL_2865337731.1	6DL	
Traes_6DL_28B69E568.1	6DL	atp binding
Traes_6DL_2D577F980.1	6DL	probable lrr receptor-like serine threonine-protein kinase at4g36180-like
Traes_6DL_2D910DA74.1	6DL	
Traes_6DL_34C04505F.1	6DL	ethylene receptor-like protein 2
Traes_6DL_44B6FE7CF.1	6DL	ring-h2 finger protein at179-like
Traes_6DL_458CAEF7A.1	6DL	ubiquitin-protein ligase
Traes_6DL_478A5F48C.1	6DL	
Traes_6DL_4C48A6567.1	6DL	
Traes_6DL_4C9665FB5.1	6DL	
Traes_6DL_5BA456CC3.1	6DL	cysteine-rich receptor-like protein kinase 8
Traes_6DL_5DFEB89F6.1	6DL	isoleucyl-trna synthetase
Traes_6DL_5EB76ED40.1	6DL	bzip protein
Traes_6DL_686ED046D.1	6DL	
Traes_6DL_69BE045F9.1	6DL	
Traes_6DL_69D581884.1	6DL	
Traes_6DL_6CE682354.1	6DL	
Traes_6DL_6D896315B.1	6DL	
Traes_6DL_85F5A8FE8.1	6DL	diphthamide biosynthesis protein 1
Traes_6DL_90424C8DB.1	6DL	
Traes_6DL_910A3770D.1	6DL	phosphopantothenate--cysteine ligase 1-like
Traes_6DL_936F83A4F.1	6DL	uncharacterized protein loc100834020 isoform 1
Traes_6DL_B3C85001D.1	6DL	

Traes_6DL_B67849C3E.1	6DL	probable carotenoid cleavage dioxygenase chloroplastic-like
Traes_6DL_D81C60D57.1	6DL	u-box domain-containing protein 35-like
Traes_6DL_E1CEB78EC.1	6DL	
Traes_6DL_E6E7B314B.1	6DL	u-box domain-containing protein 25-like
Traes_6DL_FEB17A5E4.1	6DL	u-box domain-containing protein 26-like
Traes_6DS_7732A9607.1	6DS	
Traes_7AL_OC3473DBB.1	7AL	
Traes_7AL_16CAC66ED.1	7AL	7 protein
Traes_7AL_6D67B292E.1	7AL	
Traes_7AS_01C673190.1	7AS	l-type lectin-domain containing receptor kinase
Traes_7AS_OC4F461A6.1	7AS	
Traes_7AS_ODBFD4657.1	7AS	
Traes_7AS_152E27DE8.1	7AS	
Traes_7AS_1A16D24B7.1	7AS	flavonoid 3 -monooxygenase-like
Traes_7AS_2BAEE42D8.1	7AS	
Traes_7AS_2F023334C.1	7AS	
Traes_7AS_38F4D18AC.1	7AS	mediator of rna polymerase ii transcription subunit 33a-like isoform x5
Traes_7AS_48EC42B53.1	7AS	scarecrow-like protein 28-like
Traes_7AS_4BB18C367.1	7AS	granule-bound starch synthase 1
Traes_7AS_54CBF74FB.1	7AS	starch synthase ii
Traes_7AS_6E668B80.1	7AS	armadillo beta-catenin repeat related protein
Traes_7AS_7499A734B.1	7AS	nadh dehydrogenase complex assembly factor 6-like
Traes_7AS_81E840BCA.1	7AS	probable l-type lectin-domain containing receptor kinase -like
Traes_7AS_830505028.1	7AS	calcium-dependent protein kinase
Traes_7AS_A048C3935.1	7AS	
Traes_7AS_A476464CF1.1	7AS	activator of 90 kda heat shock protein atpase homolog 1-like
Traes_7AS_A820BBAAA.1	7AS	granule-bound starch synthase
Traes_7AS_B02BB100C.1	7AS	
Traes_7AS_B39DF1E58.1	7AS	
Traes_7AS_B9AB23B80.1	7AS	

Traes_7AS_E012CA4D9.1	7AS	ras group-related lrr 5 protein
Traes_7AS_E5659F6DC.1	7AS	gibberellin 20 oxidase 2
Traes_7AS_EC40B9415.1	7AS	lectin-domain containing receptor kinase
Traes_7BL_1DD9E9A8E.1	7BL	
Traes_7BL_537EA1748.1	7BL	disease resistance protein rppl3
Traes_7BL_7D6A00C8C.1	7BL	fertility restorer
Traes_7BL_80C59C507.1	7BL	
Traes_7BL_A6732AAAA.1	7BL	
Traes_7BL_AD93089B3.1	7BL	
Traes_7BL_D5B527888.1	7BL	
Traes_7BL_E0E018397.1	7BL	
Traes_7BL_E8FB13575.1	7BL	
Traes_7BL_EF2EA8F09.1	7BL	
Traes_7BL_FAF92E64F.1	7BL	
Traes_7BS_01835AEAE.1	7BS	probable long-chain-alcohol o-fatty-acyltransferase 4-like
Traes_7BS_751146381.1	7BS	lipxygenase chloroplastic
Traes_7BS_9212EB953.1	7BS	leucine zipper
Traes_7BS_DD21953A5.1	7BS	phosphate transporter
Traes_7DL_03E44563A.1	7DL	high affinity nitrate transporter -like
Traes_7DL_0762CF867.1	7DL	
Traes_7DL_07D295BFC.1	7DL	crinkly4-like receptor protein kinase family protein
Traes_7DL_13621DE2E.1	7DL	
Traes_7DL_2AFCB81A4.1	7DL	serine threonine-protein kinase-like protein ccr4
Traes_7DL_2EA8D5524.1	7DL	serine threonine-protein kinase-like protein ccr4
Traes_7DL_32AC0405A.1	7DL	
Traes_7DL_4F8080979.1	7DL	uncharacterized protein loc100828305
Traes_7DL_51D5AA31C.1	7DL	
Traes_7DL_5414349CB.1	7DL	ptil-like tyrosine-protein kinase at3g15890-like
Traes_7DL_54415F964.1	7DL	mitogen-activated protein kinase kinase kinase 1-like
Traes_7DL_5AA017B3C.1	7DL	protein iq-domain 14-like
Traes_7DL_5F1F26574.1	7DL	

Traes_7DL_62D5D1C30.1	7DL	lrr receptor-like serine threonine-protein kinase
Traes_7DL_7716C6556.1	7DL	
Traes_7DL_7D5DBCCA0.1	7DL	
Traes_7DL_851E1A20D.1	7DL	
Traes_7DL_85BA805E2.1	7DL	aspartyl protease family protein
Traes_7DL_8D5704082.1	7DL	protein transparent testa 1-like
Traes_7DL_9235C4392.1	7DL	
Traes_7DL_929BEA59C.1	7DL	uncharacterized protein loc101786056
Traes_7DL_B0FC8A744.1	7DL	macrophage migration inhibitory factor homolog
Traes_7DL_B8FCF6E36.1	7DL	udp-glycosyltransferase 75d1-like
Traes_7DL_BE13C5308.1	7DL	protein kinase
Traes_7DL_C2D5C1A92.1	7DL	zinc finger ccch domain-containing protein 7
Traes_7DL_C3F07D103.1	7DL	floral organ number partial
Traes_7DL_C4F9BED90.1	7DL	aspartic proteinase nepenthesin-1 precursor
Traes_7DL_CDB7B43A8.1	7DL	
Traes_7DL_CE076898F.1	7DL	probable ubiquitin conjugation factor e4-like
Traes_7DL_DCOE268FC.1	7DL	uncharacterized protein loc101776480
Traes_7DL_EA4E031BE.1	7DL	
Traes_7DS_0F9499EB6.1	7DS	protein
Traes_7DS_1AE62ED43.1	7DS	
Traes_7DS_22CCC8B31.1	7DS	e3 ubiquitin-protein ligase xbat33-like
Traes_7DS_308B45EF2.1	7DS	
Traes_7DS_30BC9884E.1	7DS	rpml-interacting protein 4-like
Traes_7DS_5084227F2.1	7DS	hypothetical protein F775_05705 [Aegilops tauschii]
Traes_7DS_554A7D67C.1	7DS	u-box domain-containing protein 12-like
Traes_7DS_619733045.1	7DS	lrr receptor-like serine threonine-protein kinase
Traes_7DS_8453B74CD.1	7DS	casp-like protein sb07g025950-like
Traes_7DS_943233EC4.1	7DS	
Traes_7DS_B466B7F71.1	7DS	uncharacterized protein isoform 1
Traes_7DS_B82B9C04E.1	7DS	burp domain-containing protein
Traes_7DS_C1DBCA997.1	7DS	uncharacterized protein chloroplastic-like

Traes_7DS_C92CA6864.1	7DS	uncharacterized protein loc100827453
Traes_7DS_C932222B7.1	7DS	atp-dependent zinc metalloprotease ftsh chloroplastic-like
Traes_7DS_D661C8118.1	7DS	
Traes_7DS_EAE15169D.1	7DS	glycoprotein 3-alpha-l-fucosyltransferase

1

2

Draft

Table S2. Gene network analysis of genes from each chromosome and single subgenome genes

Chr.	total # of genes	# of unique Arabidopsis orthologs*	# of orthologs within in networks (score > 0.9)	# of orthologs outside of networks (score > 0.9)	# of orthologs within in networks (score > 0.9)	% orthologs outside of networks (score > 0.9)
1A	4155	1928	511	1417	27%	73%
1B	4178	1906	530	1376	28%	72%
1D	3773	1823	479	1344	26%	74%
2A	5626	2496	702	1794	28%	72%
2B	6196	2622	722	1900	28%	72%
2D	6357	2447	688	1759	28%	72%
3A	4606	2051	533	1518	26%	74%
3B	5862	2388	606	1782	25%	75%
3D	4517	1940	524	1416	27%	73%
4A	4830	2211	651	1560	29%	71%
4B	4467	2038	577	1461	28%	72%
4D	4683	2218	706	1512	32%	68%
5A	4159	2062	587	1475	28%	72%
5B	5762	2544	704	1840	28%	72%
5D	4921	2275	599	1676	26%	74%
6A	4038	1841	479	1362	26%	74%
6B	3746	1631	438	1193	27%	73%
6D	3869	1747	431	1316	25%	75%
7A	4667	1987	553	1434	28%	72%
7B	4015	1716	457	1259	27%	73%
7D	4959	2153	579	1574	27%	73%

3

4 * As several splicing variants belong to a single protein, several splicing variants were
5 merged into a single gene.

6