



# Gene Co-expression Network Analysis Suggests the Existence of Transcriptional Modules Containing a High Proportion of Transcriptionally Differentiated Homoeologs in Hexaploid Wheat

Kotaro Takahagi<sup>1,2,3</sup>, Komaki Inoue<sup>1</sup> and Keiichi Mochida<sup>1,2,3,4\*</sup>

## OPEN ACCESS

### Edited by:

Hans D. Daetwyler,  
La Trobe University, Australia

### Reviewed by:

Divya Mehta,  
Queensland University of Technology,  
Australia

Wilco Ligterink,  
Wageningen University & Research,  
Netherlands

### \*Correspondence:

Keiichi Mochida  
keiichi.mochida@riken.jp

### Specialty section:

This article was submitted to  
Evolutionary and Population Genetics,  
a section of the journal  
Frontiers in Plant Science

**Received:** 31 January 2018

**Accepted:** 23 July 2018

**Published:** 08 August 2018

### Citation:

Takahagi K, Inoue K and Mochida K  
(2018) Gene Co-expression Network  
Analysis Suggests the Existence  
of Transcriptional Modules Containing  
a High Proportion of Transcriptionally  
Differentiated Homoeologs  
in Hexaploid Wheat.  
*Front. Plant Sci.* 9:1163.  
doi: 10.3389/fpls.2018.01163

<sup>1</sup> Bioproductivity Informatics Research Team, RIKEN Center for Sustainable Resource Science, Yokohama, Japan, <sup>2</sup> Graduate School of Nanobioscience, Yokohama City University, Yokohama, Japan, <sup>3</sup> Kihara Institute for Biological Research, Yokohama City University, Yokohama, Japan, <sup>4</sup> Institute of Plant Science and Resources, Okayama University, Kurashiki, Japan

Genome duplications aid in the formation of novel molecular networks through regulatory differentiation of the duplicated genes and facilitate adaptation to environmental change. Hexaploid wheat, *Triticum aestivum*, contains three homoeologous chromosome sets, the A-, B-, and D-subgenomes, which evolved through interspecific hybridization and subsequent whole-genome duplication. The divergent expression patterns of the homoeologs in hexaploid wheat suggest that they have undergone transcriptional and/or functional differentiation during wheat evolution. However, the distribution of transcriptionally differentiated homoeologs in gene regulatory networks and their related biological functions in hexaploid wheat are still largely unexplored. Therefore, we retrieved 727 publicly available wheat RNA-sequencing (RNA-seq) datasets from various tissues, developmental stages, and conditions, and identified 10,415 expressed homoeologous triplets. Examining the co-expression modules in the wheat transcriptome, we found that 66% of the expressed homoeologous triplets possess all three homoeologs grouped in the same co-expression modules. Among these, 15 triplets contain co-expressed homoeologs with differential expression levels between homoeoalleles across  $\geq 95\%$  of the 727 RNA-seq datasets, suggesting a consistent trend of homoeolog expression bias. In addition, we identified 2,831 differentiated homoeologs that showed gene expression patterns that deviated from those of the other two homoeologs. We found that seven co-expression modules contained a high proportion of such differentiated homoeologs, which accounted for  $\geq 20\%$  of the genes in each module. We also found that five of the co-expression modules are abundantly composed of genes involved

in biological processes such as chloroplast biogenesis, RNA metabolism, putative defense response, putative posttranscriptional modification, and lipid metabolism, thereby suggesting that, the differentiated homoeologs might highly contribute to these biological functions in the gene network of hexaploid wheat.

**Keywords:** allopolyploidization, co-expression gene network, hexaploid wheat, homoeolog, transcriptional module

## INTRODUCTION

Interspecific hybridization and polyploidization have played important roles in the evolution and diversification of plants (Soltis and Soltis, 2009; Van de Peer et al., 2009). Allopolyploids are originated from hybridization between different species followed by whole-genome duplication (Ramsey and Schemske, 1998; Comai, 2005). Despite the multiple conditions that need to be met for allopolyploidization to occur, including existing populations of parental lines in the same area, overcoming hybrid incompatibility, gametic non-reduction, and chromosome doubling (Osabe et al., 2012), the occurrence of allopolyploids is widespread in various taxonomic groups in plants (Leitch and Leitch, 2008; Barker et al., 2016). Therefore, it has been hypothesized that allopolyploid species have evolutionary advantages compared to their diploid ancestral species (Wendel, 2000; Doyle et al., 2008).

Improved traits that evolved in allopolyploid plants enhanced their productivity and have contributed to the domestication of many crops (Chen, 2010; Renny-Byfield and Wendel, 2014). For example, the allotetraploid *Arabidopsis suecica* has more vigorous growth and produces more seeds than its ancestral species (Solhaug et al., 2016), whereas the allotetraploid *Coffea arabica* can better adapt to changes in temperature than its diploid ancestors (Combes et al., 2013). In allohexaploid wheat (*Triticum aestivum*), both natural and synthetic plants have higher tolerance to salt stress than their diploid and tetraploid ancestors (Dubcovsky and Dvorak, 2007; Yang et al., 2014). These examples suggest that allopolyploidization often leads to increased productivity through fixation of genomic heterozygosity, which improves environmental fitness and contributes to the habitat expansion of a species.

Allopolyploidization can give rise to transcriptional and/or functional changes in homoeologs (genes that are duplicated due to allopolyploidization) (Mochida et al., 2003; Adams and Wendel, 2005; Moore and Purugganan, 2005). Homoeologs can undergo accelerated evolution due to redundant genetic codes that can evolve new functions without constraints (Kaessmann, 2010; Naseeb et al., 2017). A number of studies have revealed their fates as non-functionalized (loss of function of one of the duplicated genes), subfunctionalized (partitioning of function between duplicated genes), and/or neo-functionalized (diversification of function between the duplicated genes) (Lynch and Conery, 2000; Blanc and Wolfe, 2004; Cusack and Wolfe, 2007). Homoeologs in plants often show different expression patterns across tissues, developmental stages, and conditions, suggesting that they have undergone sub- and/or neofunctionalization (Madlung, 2013). The differential

employment of homoeologs through dynamic transcriptional regulation may contribute to the enhanced evolutionarily adaptability of allopolyploid species.

A number of studies based on homoeolog-specific gene expression analysis have reported the evolutionary fates of homoeologs in various allopolyploid plants (Adams, 2007; Hughes et al., 2014; Takahagi et al., 2018). Transcriptome analysis has revealed that the expression of multiple ribosomal protein-coding homoeologs in *Brassica napus* is tissue-dependent (Whittle and Krochko, 2009). An investigation of the relative levels of allelic and homoeologous gene expression in cotton revealed that subfunctionalized genes are mainly expressed in reproductive tissues, and non-functionalized alleles are typically derived from the A-genome, indicating potential genome-of-origin bias for neofunctionalization (Chaudhary et al., 2009). Differentiation of expression patterns of homoeologs in allopolyploid species might effect changes in their gene regulatory networks owing to transcriptional and/or functional divergence. The evolutionary changes in gene regulatory networks are thought to facilitate responses to developmental programs and environmental cues in allopolyploids (Chen and Ni, 2006).

Hexaploid wheat, *Triticum aestivum*, is a widely cultivated allohexaploid crop ( $2n = 6x = 42$ , AABBDD) that originated from hybridization between the domesticated allotetraploid *Triticum turgidum* ( $2n = 4x = 28$ , AABB) and the diploid goat grass *Aegilops tauschii* ( $2n = 2x = 14$ , DD) approximately 10,000 years ago, followed by genome duplication (Matsuoka, 2011; Feldman and Levy, 2012). Pfeifer et al. (2014) generated a co-expression gene network of hexaploid wheat and examined the contribution of expression of each homoeolog. They found that several network modules exhibit unbalanced homoeolog expression, which might be associated with biological functions and tissue types (Pfeifer et al., 2014). Recently, Tanaka et al. (2016) reported homoeolog-specific regulation of the floral MADS-box genes in wheat, and differential expression patterns of homoeologs were consistently observed in both natural and synthetic allohexaploid wheat varieties (Tanaka et al., 2016). Moreover, Powell et al. (2017) demonstrated that the wheat transcriptome has homoeolog expression bias toward the B- and D-subgenomes in response to pathogen infection (Powell et al., 2017). The divergent expression patterns between homoeologs suggest that they have undergone transcriptional and/or functional differentiation. However, the distribution of transcriptionally differentiated homoeologs in gene regulatory networks and their related biological functions in hexaploid wheat are still largely unexplored.

In this study, to elucidate homoeologous networks in hexaploid wheat and to explore their differentiation, we retrieved

publicly available RNA-sequencing (RNA-seq) datasets from various tissues, developmental stages, and conditions. We categorized hexaploid wheat genes to construct homoeologous groups and identified expressed homoeologous triplets. We also identified differentiated homoeologs that show gene expression patterns that deviate from those of the other two homoeologs. In addition, we explored gene network modules containing a high proportion of differentiated homoeologs in the transcriptome of hexaploid wheat. We assessed enriched functions in the network modules and discussed the evolution of such network modules resulting from transcriptional differentiation of homoeologs in hexaploid wheat.

## MATERIALS AND METHODS

### Data and Data Processing

All publicly available wheat transcriptome sequence datasets were retrieved from the NCBI Sequence Read Archive (April 26, 2017)<sup>1</sup>. To adjust the data format, the datasets were screened according to the following criteria: (1) RNA-seq data strictly (i.e., no EST, FL-cDNA, etc.) from *Triticum aestivum* samples, (2) total number of sequence reads  $\geq 10,000,000$ , and (3) an average sequence read length is 70–1000 bases. The RNA-seq datasets presenting the following characteristics were also removed from analyses, as they were considered inappropriate for gene expression profiling: (1) datasets resulting from pooled samples, taken at different time points, (2) datasets obtained from chromosome deletion and chromosome addition lines, and (3) datasets obtained for poorly described methodologies. RNA-seq reads of the screened datasets were trimmed using Trimmomatic (v.0.32) (Bolger et al., 2014) with the following settings: - thread 1 LEADING: 20 TRAILING: 20 SLIDINGWINDOW:4:15 MINLEN: 50. To obtain high-quality sequence datasets, the trimmed datasets were further screened according to the following criteria: (1)  $\geq 70\%$  of raw reads are maintained after the trimming step and (2) an average sequence read length is 70–1000 bases after trimming. The trimmed reads obtained after the second screening were mapped to the representative cDNA sequences annotated in the genome assembly of Chinese Spring wheat (International Wheat Genome Sequencing Consortium, 2014) downloaded from the Ensembl (v.35)<sup>2</sup> using the BWA program (v.0.7.8) (Li and Durbin, 2009) with its mem command. To use datasets with high-quality alignments of the reads, those that were not uniquely mapped and/or not paired mapped were removed from the read alignment datasets using custom Perl scripts. In total, 727 read alignment datasets (**Supplementary Table S1**), for which  $\geq 50\%$  of raw reads remained after the read removal step, were subjected to further analysis. The reads per million mapped reads (RPM) values were calculated for all genes in the 727 read alignment datasets. Genes with an RPM  $\geq 3$  in at least eight datasets ( $\geq 1\%$  of the 727 RNA-seq datasets) were identified as significantly expressed genes.

<sup>1</sup><https://www.ncbi.nlm.nih.gov/sra>

<sup>2</sup>[http://plants.ensembl.org/Triticum\\_aestivum/Info/Index](http://plants.ensembl.org/Triticum_aestivum/Info/Index)

### Identification of Homoeologous Groups

To identify homoeologous groups, representative protein sequences of the A-, B-, and D-subgenomes annotated in the genome assembly of Chinese Spring wheat (International Wheat Genome Sequencing Consortium, 2014) downloaded from Ensembl (v.35)<sup>2</sup> were compared against each other using BLASTP (v.2.6) (McGinnis and Madden, 2004), applying an *e*-value cut-off of  $1e-5$  and a sequence identity cut-off of 90%. Sets of three homoeologs that were reciprocal best hits in all pairwise comparisons were identified as homoeologous triplets (ABD type in **Figure 1B**). Sets of two homoeologs with reciprocal best hits for two subgenomes and without hits for the other subgenome were identified as homoeologous doublets (AB, AD, and BD types in **Figure 1B**). Genes without hits in any of the other two subgenomes were identified as subgenome-unique genes (A, B, and D types in **Figure 1B**).

### t-distributed Stochastic Neighbor Embedding (t-SNE) Analysis

To summarize expression patterns of the genes with an RPM  $\geq 3$  in a range of 1–7 datasets (spatiotemporally expressed genes), t-SNE analysis was performed using the Rtsne package (v.0.13)<sup>3</sup> in R (v.3.4.3). The number of iterations was set at 10,000, and parameter theta was set at 0.0.

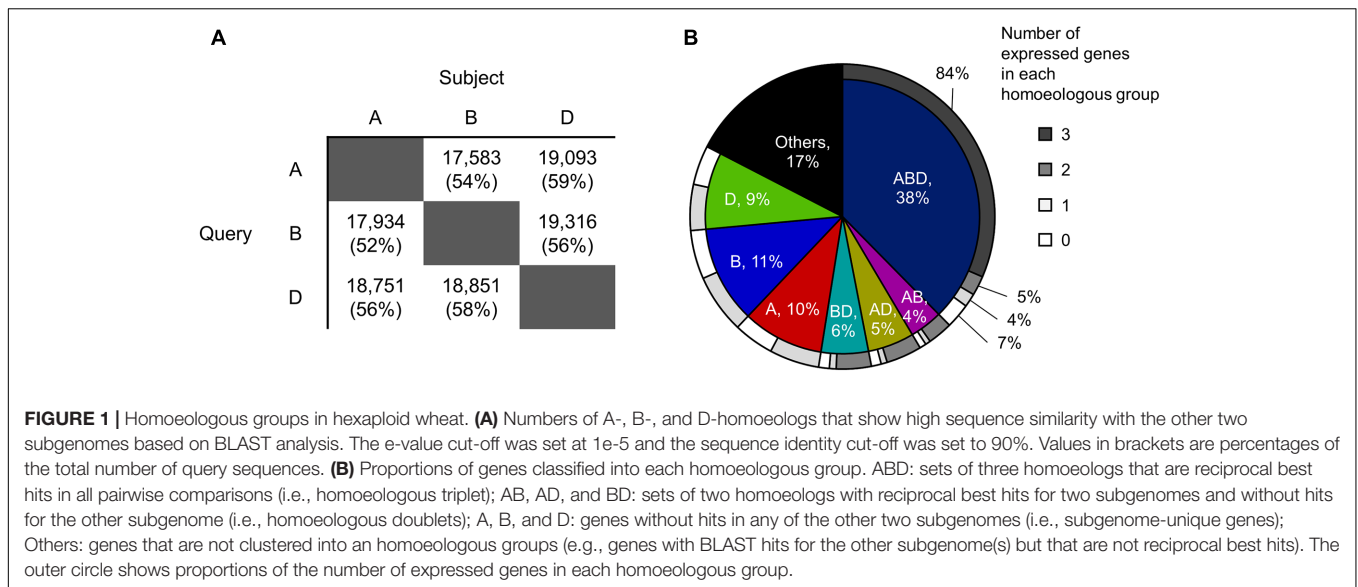
### Co-expression Network Analysis

To compute co-expression modules of homoeologs, WGCNA analysis (Langfelder and Horvath, 2008) was performed based on the normalized RPM using the one-step automatic network construction method with the following parameters: power = 9, networkType = “signed”, TOMType = “unsigned”, minModuleSize = 30, reassignThreshold = 0, mergeCutHeight = 0.25, numericLabels = TRUE, pamRespectsDendro = FALSE. A soft-thresholding power was selected by evaluating the scale-free topology model fit.

### Identification of Differentially Expressed Genes

For identification of the homoeologous triplets containing co-expressed homoeologs with differential expression levels between homoeoalleles, the gene expression fold changes between homoeologs across the 727 RNA-seq datasets were calculated based on RPM. Pairs of homoeologs with a fold change  $\geq 3$  and RPM  $\geq 3$  for at least one of the homoeologs were identified as differentially expressed homoeologs. For the examination of expression bias between homoeologs in the homoeologous triplets, reads used for RPM calculation in a series of RNA-seq datasets (SRR1542404-SRR1542417) (Liu et al., 2015) were subjected to differential gene expression analysis performed by using the edgeR package (v.3.20.9) (Robinson et al., 2010) in R (v.3.4.3). Pairs of homoeologs with a false discovery rate (FDR)  $\leq 0.001$  and RPM (average of 2 biological replicates in the

<sup>3</sup><https://github.com/jkrijthe/Rtsne>



RNA-seq datasets)  $\geq 3$  for at least one of the homoeologs were identified as significantly differentially expressed homoeologs.

## Gene Ontology (GO) Enrichment Analysis

The closest homologs of wheat genes in *Arabidopsis* and rice were identified by BLASTP (v.2.6) (McGinnis and Madden, 2004) searches, applying an *e*-value threshold of  $\leq 1e-5$ . GO terms of the best-hit genes in *Arabidopsis* and rice were used as the customized annotations for wheat genes. To reduce bias, GO terms that were assigned to more than 5,000 wheat genes were excluded. Enriched GO terms were identified for selected genes using BLAST2GO (v.4.1.9) (Conesa et al., 2005) with the customized annotations of wheat genes. For the estimation of the enriched GO terms of genes that are spatiotemporally expressed (representing genes with an RPM  $\geq 3$  in less than 1% (eight datasets) of the 727 RNA-seq datasets) or non-significantly expressed (representing genes with an RPM  $< 3$  in all of the 727 RNA-seq datasets), all of the annotated genes in the Chinese Spring wheat chromosomes were used as a reference set. For estimation of the enriched GO terms of the other sets of genes, those in the expressed homoeologous triplets were used as a reference set. The significance threshold was set at FDR  $\leq 0.001$ . The enriched GO terms were summarized based on their semantic similarities using the web-based tool REVIGO<sup>4</sup> (Supek et al., 2011).

## RESULTS

### Homoeologous Triplets in Hexaploid Wheat

To explore the distribution of transcriptionally differentiated homoeologs in gene regulatory networks and their related biological functions in hexaploid wheat, we identified expressed

homoeologous triplets using publicly available RNA-seq datasets. We gathered 727 RNA-seq datasets from hexaploid wheat composed of as many as 517 biosamples relating to various tissues, developmental stages, and conditions, which enabled us to comprehensively explore functional differentiation of transcription regulatory networks in hexaploid wheat (Supplementary Table S1). We mapped the quality-checked reads of the RNA-seq datasets to the set of representative cDNA sequences annotated in the genome assembly of Chinese Spring wheat. Using a threshold of RPM  $\geq 3$  in at least eight datasets ( $\geq 1\%$  of the 727 RNA-seq datasets), we found that 73,329 genes (74% of the 99,308 genes corresponding to the representative cDNA sequences assigned to each chromosome) are significantly expressed in hexaploid wheat. To construct putative homoeologous groups, and estimate the number of expressed homoeologs from each homoeoloci, we clustered all the 99,308 genes into 49,710 gene groups based on sequence similarity, using a reciprocal BLAST homology search (Figure 1A). Approximately 38% of the genes were classified into gene groups composed of three homoeologs, one from each subgenome (homoeologous triplets, ABD type in Figure 1B), in which 84% of the triplets (10,415 triplets) contained three homoeologs significantly expressed in the RNA-seq datasets (expressed homoeologous triplets; Figure 1B). We also observed that 31,738 genes (39% of 82,012 genes assigned into each of the homoeologous groups) are expressed from one or two homoeologous loci on the subgenomes, which suggests that approximately 40% of the homoeologous groups contain homoeologs rarely expressed or silenced in the wheat transcriptome (Figure 1B).

### Spatiotemporally Expressed Genes in Wheat

To characterize the genes found in the wheat transcriptome that are rarely expressed or silenced, we investigated the

<sup>4</sup><http://revigo.irb.hr/>



chromosomal distribution and function of these genes. Using the threshold to identify significantly expressed genes, we classified 25,979 genes as rarely expressed or silenced, which suggested a transcriptional sign of non-functionalization or acceleration of spatiotemporal transcriptional regulation. To further investigate the functional properties of such genes, we assessed their chromosomal distribution; however, no biased distribution of these genes was found across the 21 wheat chromosomes (**Figure 2A**). We found that 44% of the 25,979 genes were expressed in at least one RNA-seq dataset with an RPM  $\geq 3$ , whereas the remaining 56% genes showed an RPM  $< 3$  in all of the RNA-seq datasets, suggesting spatiotemporal expression and insignificant expression, respectively (**Figure 2B**). To summarize the expression patterns of the spatiotemporally expressed genes across the 727 RNA-seq datasets, we clustered and visualized the expression profiles of these genes using the t-SNE algorithm, and detected several clusters corresponding to the RNA-seq datasets from particular tissues, such as roots, stem, and anthers (**Figure 2C**), suggesting their tissue-specific expressions. To assess gene functions over-represented in the spatiotemporally or non-significantly expressed genes, we performed GO enrichment analysis, and found some enriched GO terms related to the response to abiotic stresses, metabolism, and organ development (**Figure 2D**).

## Expression Bias Between Homoeologs in Hexaploid Wheat

To examine expression bias between homoeologs in the expressed homoeologous triplets, we computed co-expressed homoeologs and differentially expressed homoeologs based on the 727 RNA-seq datasets. For identification of the co-expressed homoeologs, we applied the WGCNA algorithm, and identified 22 co-expression modules. The results of WGCNA analysis indicated that 66% of the expressed homoeologous triplets possess all three homoeologs grouped in the same co-expression modules (co-expressed triplets, ABD type in **Figure 3A**). For 27% of the triplets, two out of three homoeologs were grouped in the same co-expression modules (AB-D, AD-B, and BD-A types in **Figure 3A**), whereas for the remaining 5% of the triplets, all three homoeologs were assigned to different modules (A-B-D type in **Figure 3A**). To further identify homoeologs that are co-expressed while differentially expressed (representing similar co-expression patterns across the 727 RNA-seq datasets and differential expression levels between homoeoalleles), we identified differentially expressed homoeologs (fold-change  $\geq 3$ ) in the co-expressed triplets, and found that at least 258 triplets contained co-expressed homoeologs with differential expression levels between homoeoalleles across  $\geq 50\%$  of the 727 RNA-seq datasets (**Figure 3B**). We also found that 15 co-expressed triplets contained such homoeologs observed in  $\geq 95\%$  of the datasets, suggesting a consistent trend of homoeolog expression bias (**Figures 3B,C**). On the basis of our GO enrichment analysis of these genes, we observed several over-represented functions, such as biotin metabolism, protein modifications, and response to gibberellin stimulus (**Figure 3D**). Moreover, to illuminate homoeolog-specific expression patterns relative to particular tissue type

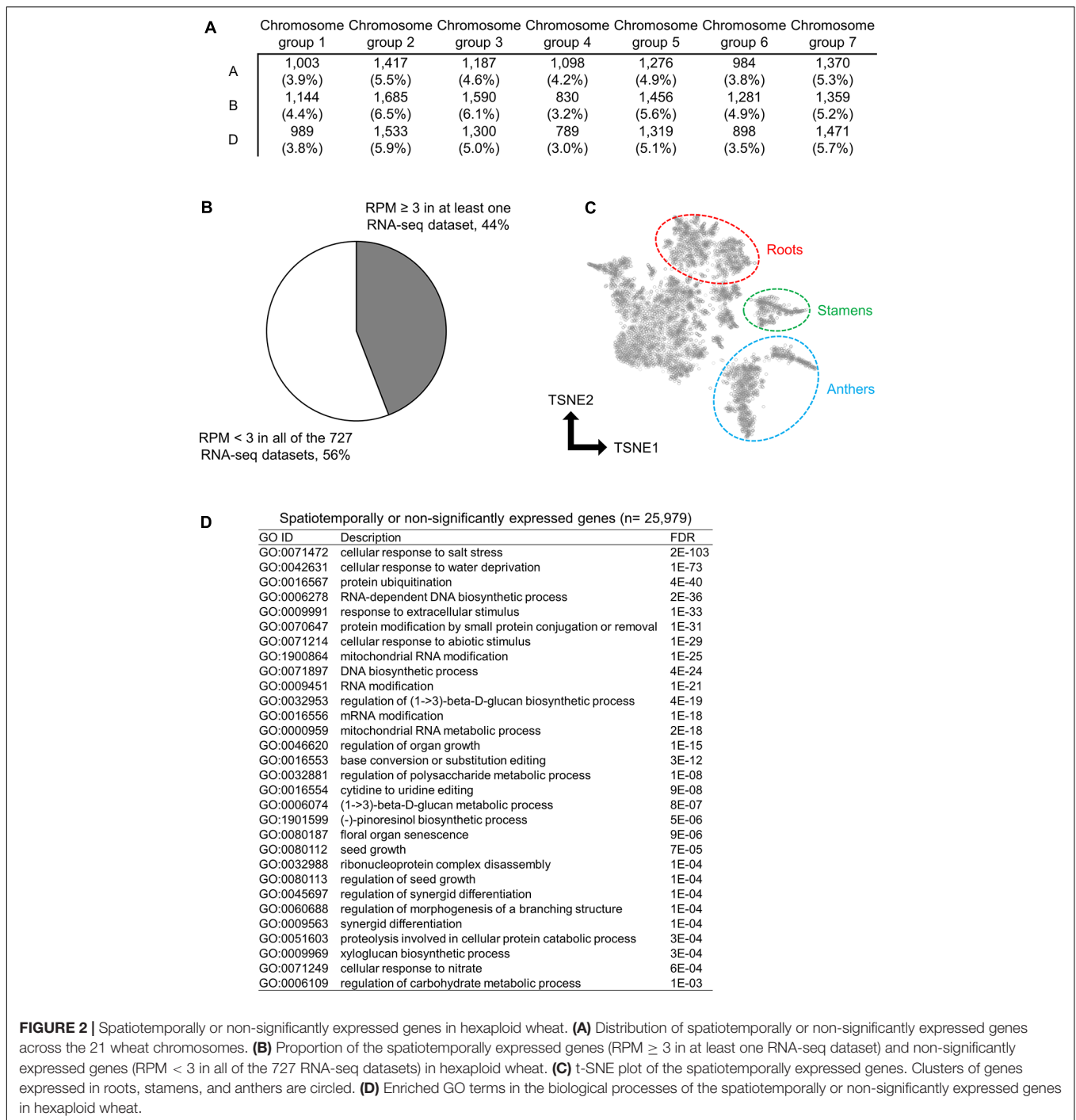
that are supported statistically, we examined the expression bias between homoeologs in the homoeologous triplets in a series of RNA-seq datasets related to multiple abiotic stress conditions such as drought, heat, and combined heat and drought (SRR1542404-SRR1542417) (Liu et al., 2015), and found that an increased number of homoeologous triplets contained differentially expressed homoeologs (FDR  $\leq 0.001$ ) in response to the drought and heat stress conditions, thereby suggesting the differentiation of transcriptional responsiveness between homoeologs to environmental stresses (**Supplementary Table S2**).

## Transcriptional Modules Containing a Number of Differentiated Homoeologs

We constructed co-expression gene networks based on the 727 RNA-seq datasets, and thus found that differentiated homoeologs were unevenly distributed in each of the co-expression modules and that several modules contained high proportions of differentiated homoeologs. On the basis of co-expression modules established from our WGCNA analysis, we identified 2,831 homoeologous triplets containing one homoeolog for which the expression pattern deviated from those of the other two homoeologs, which consisted of 9, 10, and 8% of differentiated homoeologs located in A-, B-, and D-subgenomes, respectively (BD-A, AD-B, and AB-D types, respectively, in **Figure 3A**). We also found that such differentiated homoeologs accounted for approximately 9% of all genes used for the WGCNA analysis (10,415 homoeologous triplets; 31,245 genes), whereas seven co-expression modules contained a high proportion of differentiated homoeologs, accounting for  $\geq 20\%$  of the genes in each module (**Figure 4A**). To estimate enriched biological functions for the genes within the co-expression modules containing a number of differentiated homoeologs, we performed GO enrichment analysis, and found that five of the co-expression modules are abundant in genes involved in biological processes such as chloroplast biogenesis (module 7; **Figure 4B**), RNA metabolism (module 8; **Figure 4C**), putative defense response (module 10; **Figure 4D**), putative posttranscriptional modification (module 15; **Figure 4E**), and lipid metabolism (module 18; **Figure 4F**). These findings suggest that differentiated homoeologs might highly contribute to these biological functions in the gene network of hexaploid wheat.

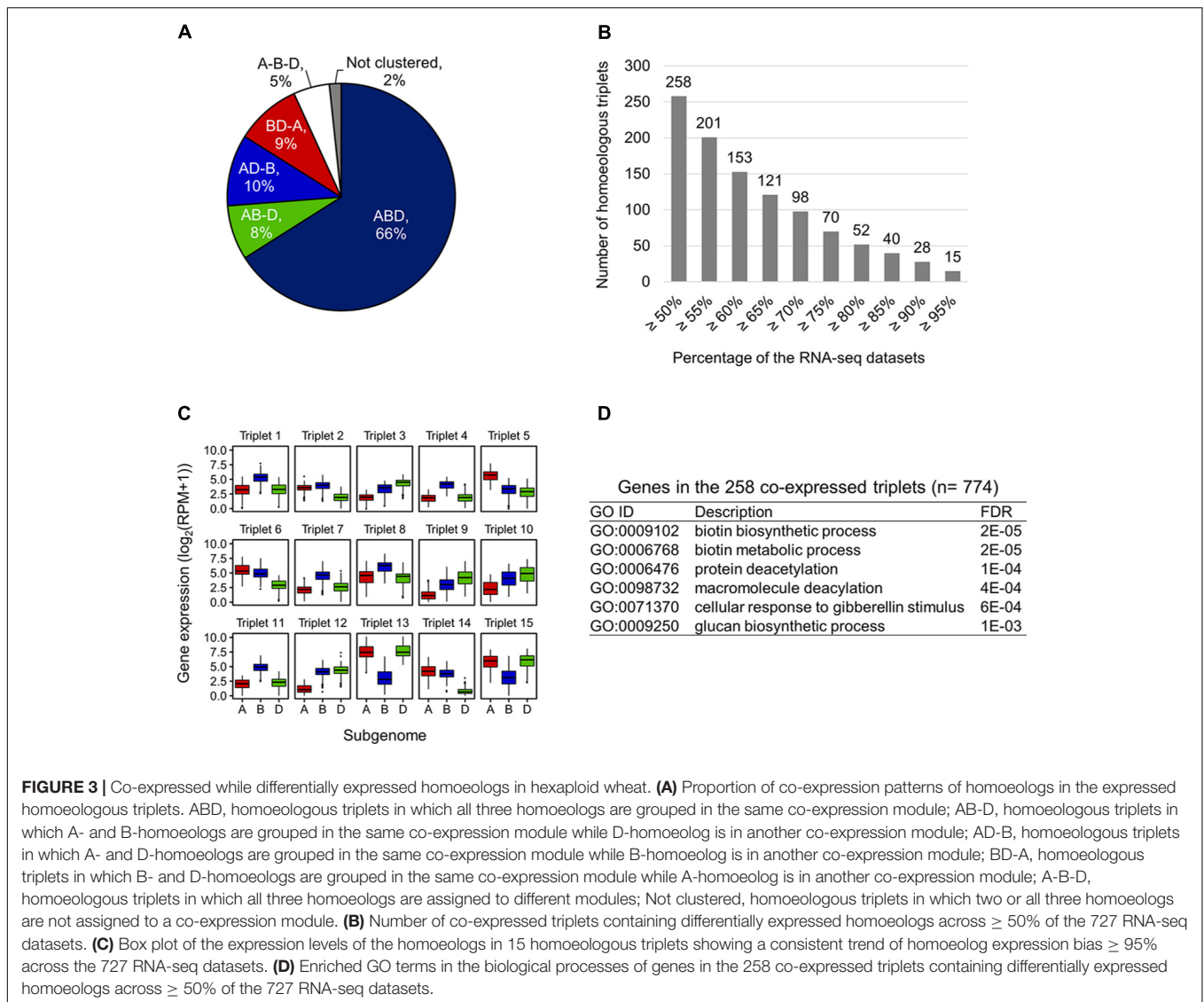
## DISCUSSION

Through our homoeologous gene expression analysis of hexaploid wheat based on a number of RNA-seq datasets, we demonstrated a landscape of transcriptional differentiation among homoeologs. Our comprehensive list of genes that were significantly expressed from one or two homoeologous loci enabled us to identify those genes that may have undergone transcriptional suppression or be directed to spatiotemporal expression. Leach et al. (2014) reported that 55% of genes in hexaploid wheat are expressed from one or two homoeologous loci on the subgenomes in root and shoot tissues (Leach et al., 2014). Using the RNA-seq datasets of 90 wheat lines,



Wang et al. (2017) found that approximately 60% of wheat genes are expressed from one or two homoeologous loci in reproductive tissues (Wang et al., 2017). Our findings based on more comprehensive transcriptome datasets showed that, compared with previous observations, a smaller number of genes ( $\sim$ 40% of genes assigned into each of the homoeologous groups) are expressed from one or two homoeologous loci (Figure 1). These observations suggest that approximately 15–20% of wheat genes, including the silenced loci considered

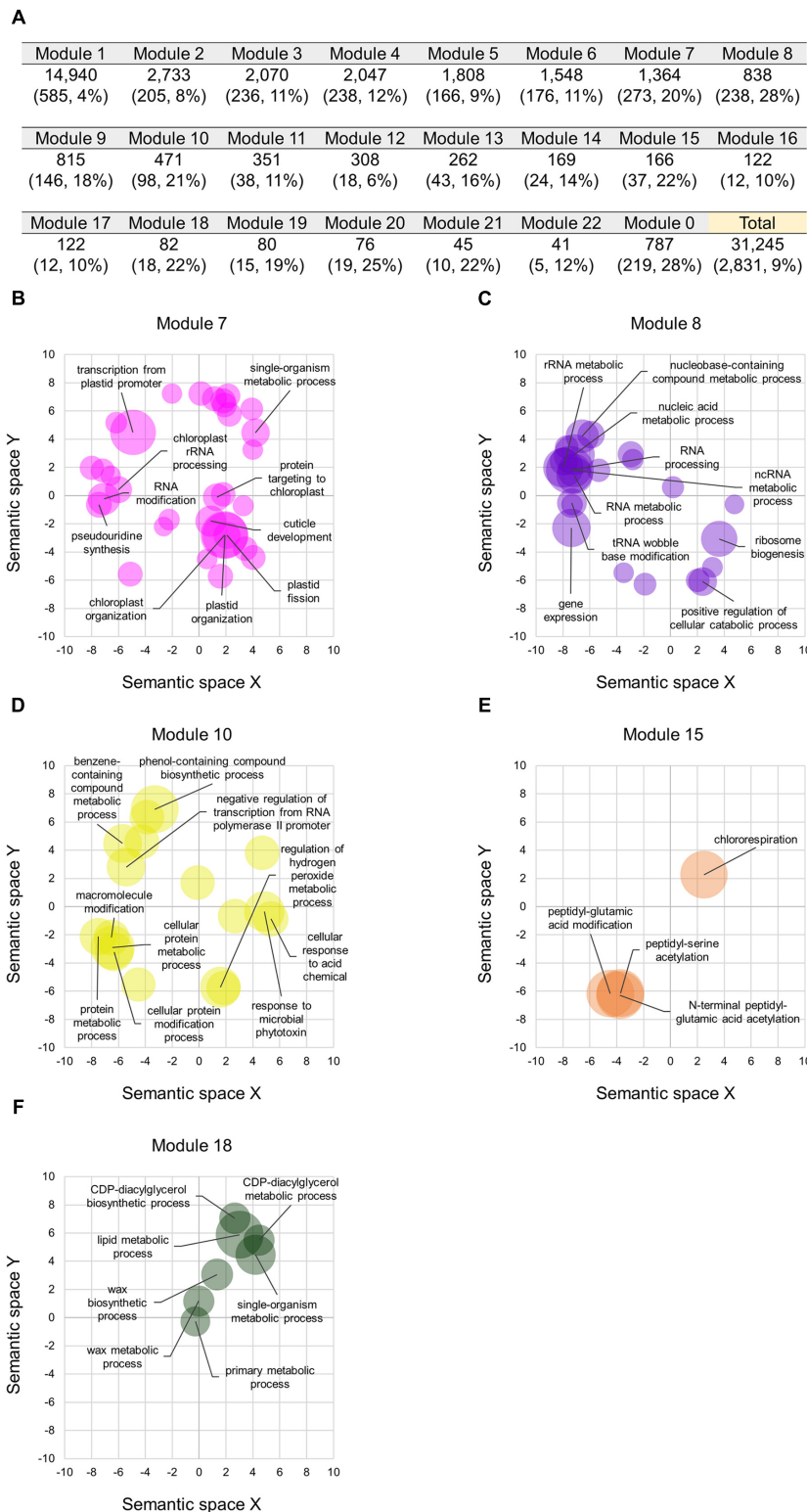
in previous studies, may contain homoeologs that can be expressed in specific tissues, at different developmental stages, or under different conditions. Our list of the spatiotemporally expressed and non-significantly expressed genes represent as many as 44% of those genes expressed (RPM  $\geq$  3) in 1–7 datasets out of the 727 RNA-seq datasets, and suggested that some of these are particularly expressed in specific tissues such as roots, stamens, and anthers (Figures 2B,C). Although we used a threshold of RPM  $\geq$  3 in less than



**FIGURE 3 |** Co-expressed while differentially expressed homoeologs in hexaploid wheat. **(A)** Proportion of co-expression patterns of homoeologs in the expressed homoeologous triplets. ABD, homoeologous triplets in which all three homoeologs are grouped in the same co-expression module; AB-D, homoeologous triplets in which A- and B-homoeologs are grouped in the same co-expression module while D-homoeolog is in another co-expression module; AD-B, homoeologous triplets in which A- and D-homoeologs are grouped in the same co-expression module while B-homoeolog is in another co-expression module; BD-A, homoeologous triplets in which B- and D-homoeologs are grouped in the same co-expression module while A-homoeolog is in another co-expression module; A-B-D, homoeologous triplets in which all three homoeologs are assigned to different modules; Not clustered, homoeologous triplets in which two or all three homoeologs are not assigned to a co-expression module. **(B)** Number of co-expressed triplets containing differentially expressed homoeologs across  $\geq 50\%$  of the 727 RNA-seq datasets. **(C)** Box plot of the expression levels of the homoeologs in 15 homoeologous triplets showing a consistent trend of homoeolog expression bias  $\geq 95\%$  across the 727 RNA-seq datasets. **(D)** Enriched GO terms in the biological processes of genes in the 258 co-expressed triplets containing differentially expressed homoeologs across  $\geq 50\%$  of the 727 RNA-seq datasets.

1% (eight datasets) of the 727 RNA-seq datasets to identify spatiotemporally or non-significantly expressed genes, this threshold depends on the proportion of samples from similar tissues in the dataset, which might present genes specifically expressed in unusually sequenced samples. To further explore spatiotemporally expressed genes, transcriptome datasets obtained from anatomically- or seasonally-distinct samples should be analyzed using emerged technologies such as laser-capture microdissection RNA-seq (LCM RNA-seq) (Zhan et al., 2015) and field transcriptome sequencing (Plessis et al., 2015). These findings may suggest that such genes expressed only from one or two homoeoalleles undergo transcriptional silencing, probably through differentiation of expression patterns and specialization of spatial expression. Consequently, such duplicated genes might be non-functionalized through promoter malfunctions or repression of other transcriptional machineries as a process of functional diploidization (Levy and Feldman, 2002; Rajkov et al., 2014).

Our gene co-expression network analysis enabled us to identify homoeologous triplets containing homoeologs that are co-expressed while differentially expressed (2.5% of the 10,415 expressed homoeologous triplets), as well as differentiated homoeologs that are classified into co-expression modules that differ from the other two homoeologs (27% of the 10,415 expressed homoeologous triplets) (Figures 3A,B). The results of our comprehensive analysis provide evidence that may suggest that most of the differential expression observed between homoeologs represents an alteration of expression patterns in hexaploid wheat. The results of our co-expressed gene network analysis enable us to identify transcriptional modules that contain abundant differentiated homoeologs involved in several particular biological processes, which might have evolved such biological functions in hexaploid wheat through its allopolyploidization (Chen et al., 2007; Feldman and Levy, 2012). Multiple studies have provided evidence to suggest that homoeolog subfunctionalization may be related to enhanced



**FIGURE 4 |** Co-expression modules containing the differentiated homoeologs in hexaploid wheat. **(A)** Number of homoeologs and differentiated homoeologs (numbers with their percentage in brackets) in each of the co-expression modules. The Module 0 represents genes that are not clustered in a co-expression module. Percentages represent proportions of the differentiated homoeologs in each of the co-expression modules. **(B–E)** Enriched GO terms in the biological processes of genes in the co-expression module 7 **(B)**, 8 **(C)**, 10 **(D)**, 15 **(E)**, and 18 **(F)** projected to a 2D semantic space. Circle size represents the  $-\log_{10}$  of FDR values calculated using REVIGO analysis. The top ten enriched GO terms are labeled in the plots.



adaptability to adverse environmental conditions in various allopolyploid species, such as tetraploid cotton, tetraploid coffee, and hexaploid wheat (Liu and Adams, 2007; Hu et al., 2011; de Carvalho et al., 2014; Liu et al., 2015). Consequently, our results suggest that along with other genes, such differentiated homoeologs may have innovated transcriptional networks, which may have contributed to adaptation to environmental change as well as to enhanced productivity during the evolution of hexaploid wheat.

The large number of RNA-seq datasets analyzed in the current study allowed integrating the transcriptional properties of each homoeologous triplet into a dataset (**Supplementary Table S3**), thereby providing a useful information resource for understanding the evolution and function of duplicated genes in hexaploid wheat. Moreover, our analyses using the datasets enabled us to demonstrate the presence of co-expression modules containing a high proportion of differentiated homoeologs in hexaploid wheat, which in turn allowed us to dissect its complex transcriptome derived from duplicated genomes. The considerable recent advances in whole-genome assembly in Triticeae species, including hexaploid wheat and its ancestors (Ling et al., 2013; Mochida and Shinozaki, 2013; International Wheat Genome Sequencing Consortium, 2014; Luo et al., 2017), provide us with an opportunity to further explore sub-/neofunctionalized homoeologs and elucidate the diploidization process that occurred during the evolution of hexaploid wheat after allopolyploidization. Such analysis will enable us to identify genes and transcriptional modules that may be associated with adaptive traits in hexaploid wheat. Such genes and transcriptional modules might also prove useful in enhancing the adaptation of staple crops to counter the potentially

adverse impacts of global climate changes and improve their productivity.

## AUTHOR CONTRIBUTIONS

KT and KM designed the work. KT and KI performed the bioinformatics analysis. KT and KM wrote the manuscript.

## FUNDING

This work was partially supported by grants-in-aid for Young Scientists (A) (Grant No. 26712003 to KM) from the Japan Society for the Promotion of Science (JSPS) and by funds to KM from the CREST (JPMJCR16O4) of the Japan Science and Technology Agency (JST).

## ACKNOWLEDGMENTS

The authors thank Shun Takaya for his help in data analysis. This work was supported by RIKEN Junior Research Associate Program.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.01163/full#supplementary-material>

## REFERENCES

- Adams, K. L. (2007). Evolution of duplicate gene expression in polyploid and hybrid plants. *J. Hered.* 98, 136–141. doi: 10.1093/jhered/esl061
- Adams, K. L., and Wendel, J. F. (2005). Novel patterns of gene expression in polyploid plants. *Trends Genet.* 21, 539–543. doi: 10.1016/j.tig.2005.07.009
- Barker, M. S., Arrigo, N., Baniaga, A. E., Li, Z., and Levin, D. A. (2016). On the relative abundance of autopolyploids and allopolyploids. *New Phytol.* 210, 391–398. doi: 10.1111/nph.13698
- Blanc, G., and Wolfe, K. H. (2004). Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. *Plant Cell* 16, 1679–1691. doi: 10.1105/tpc.021410
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Chaudhary, B., Flagel, L., Stupar, R. M., Udall, J. A., Verma, N., Springer, N. M., et al. (2009). Reciprocal silencing, transcriptional bias and functional divergence of homeologs in polyploid cotton (gossypium). *Genetics* 182, 503–517. doi: 10.1534/genetics.109.102608
- Chen, Z. J. (2010). Molecular mechanisms of polyploidy and hybrid vigor. *Trends Plant Sci.* 15, 57–71. doi: 10.1016/j.tplants.2009.12.003
- Chen, Z. J., and Ni, Z. (2006). Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. *Bioessays* 28, 240–252. doi: 10.1002/bies.20374
- Chen, Z. J., Ha, M., and Soltis, D. (2007). Polyploidy: genome obesity and its consequences. *New Phytol.* 174, 717–720. doi: 10.1111/j.1469-8137.2007.02084.x
- Comai, L. (2005). The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* 6, 836–846. doi: 10.1038/nrg1711
- Combes, M. C., Dereeper, A., Severac, D., Bertrand, B., and Lashermes, P. (2013). Contribution of subgenomes to the transcriptome and their intertwined regulation in the allopolyploid *Coffea arabica* grown at contrasted temperatures. *New Phytol.* 200, 251–260. doi: 10.1111/nph.12371
- Conesa, A., Gotz, S., Garcia-Gomez, J. M., Terol, J., Talon, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676. doi: 10.1093/bioinformatics/bti610
- Cusack, B. P., and Wolfe, K. H. (2007). When gene marriages don't work out: divorce by subfunctionalization. *Trends Genet.* 23, 270–272. doi: 10.1016/j.tig.2007.03.010
- de Carvalho, K., Petkowicz, C. L., Nagashima, G. T., Bessalho Filho, J. C., Vieira, L. G., Pereira, L. F., et al. (2014). Homeologous genes involved in mannitol synthesis reveal unequal contributions in response to abiotic stress in *Coffea arabica*. *Mol. Genet. Genomics* 289, 951–963. doi: 10.1007/s00438-014-0864-y
- Doyle, J. J., Flagel, L. E., Paterson, A. H., Rapp, R. A., Soltis, D. E., Soltis, P. S., et al. (2008). Evolutionary genetics of genome merger and doubling in plants. *Annu. Rev. Genet.* 42, 443–461. doi: 10.1146/annurev.genet.42.110807.091524
- Dubcovsky, J., and Dvorak, J. (2007). Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science* 316, 1862–1866. doi: 10.1126/science.1143986
- Feldman, M., and Levy, A. A. (2012). Genome evolution due to allopolyploidization in wheat. *Genetics* 192, 763–774. doi: 10.1534/genetics.112.146316
- Hu, Z., Yu, Y., Wang, R., Yao, Y., Peng, H., Ni, Z., et al. (2011). Expression divergence of TaMBD2 homoeologous genes encoding methyl CpG-binding

- domain proteins in wheat (*Triticum aestivum* L.). *Gene* 471, 13–18. doi: 10.1016/j.gene.2010.10.001
- Hughes, T. E., Langdale, J. A., and Kelly, S. (2014). The impact of widespread regulatory neofunctionalization on homeolog gene evolution following whole-genome duplication in maize. *Genome Res.* 24, 1348–1355. doi: 10.1101/gr.172684.114
- International Wheat Genome Sequencing Consortium (2014). A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* 345:1251788. doi: 10.1126/science.1251788
- Kaessmann, H. (2010). Origins, evolution, and phenotypic impact of new genes. *Genome Res.* 20, 1313–1326. doi: 10.1101/gr.101386.109
- Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. doi: 10.1186/1471-2105-9-559
- Leach, L. J., Belfield, E. J., Jiang, C., Brown, C., Mithani, A., and Harberd, N. P. (2014). Patterns of homoeologous gene expression shown by RNA sequencing in hexaploid bread wheat. *BMC Genomics* 15:276. doi: 10.1186/1471-2164-15-276
- Leitch, A. R., and Leitch, I. J. (2008). Perspective – Genomic plasticity and the diversity of polyploid plants. *Science* 320, 481–483. doi: 10.1126/science.1153585
- Levy, A. A., and Feldman, M. (2002). The impact of polyploidy on grass genome evolution. *Plant Physiol.* 130, 1587–1593. doi: 10.1104/pp.015727
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Ling, H. Q., Zhao, S., Liu, D., Wang, J., Sun, H., Zhang, C., et al. (2013). Draft genome of the wheat A-genome progenitor *Triticum urartu*. *Nature* 496, 87–90. doi: 10.1038/nature11997
- Liu, Z., and Adams, K. L. (2007). Expression partitioning between genes duplicated by polyploidy under abiotic stress and during organ development. *Curr. Biol.* 17, 1669–1674. doi: 10.1016/j.cub.2007.08.030
- Liu, Z., Xin, M., Qin, J., Peng, H., Ni, Z., Yao, Y., et al. (2015). Temporal transcriptome profiling reveals expression partitioning of homeologous genes contributing to heat and drought acclimation in wheat (*Triticum aestivum* L.). *BMC Plant Biol.* 15:152. doi: 10.1186/s12870-015-0511-8
- Luo, M. C., Gu, Y. Q., Pui, D., Wang, H., Twardziok, S. O., Deal, K. R., et al. (2017). Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. *Nature* 551, 498–502. doi: 10.1038/nature24486
- Lynch, M., and Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1155. doi: 10.1126/science.290.5494.1151
- Madlung, A. (2013). Polyploidy and its effect on evolutionary success: old questions revisited with new tools. *Heredity (Edinb)* 110, 99–104. doi: 10.1038/hdy.2012.79
- Matsuoka, Y. (2011). Evolution of polyploid triticum wheats under cultivation: the role of domestication, natural hybridization and allopolyploid speciation in their diversification. *Plant Cell Physiol.* 52, 750–764. doi: 10.1093/pcp/pcr018
- McGinnis, S., and Madden, T. L. (2004). BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res.* 32, W20–W25. doi: 10.1093/nar/gkh435
- Mochida, K., and Shinozaki, K. (2013). Unlocking triticeae genomics to sustainably feed the future. *Plant Cell Physiol.* 54, 1931–1950. doi: 10.1093/pcp/pct163
- Mochida, K., Yamazaki, Y., and Ogihara, Y. (2003). Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags. *Mol. Genet. Genomics* 270, 371–377. doi: 10.1007/s00438-003-0939-7
- Moore, R. C., and Purugganan, M. D. (2005). The evolutionary dynamics of plant duplicate genes. *Curr. Opin. Plant Biol.* 8, 122–128. doi: 10.1016/j.pbi.2004.12.001
- Naseeb, S., Ames, R. M., Delneri, D., and Lovell, S. C. (2017). Rapid functional and evolutionary changes follow gene duplication in yeast. *Proc. Biol. Sci.* 284:20171393. doi: 10.1098/rspb.2017.1393
- Osabe, K., Kawanabe, T., Sasaki, T., Ishikawa, R., Okazaki, K., Dennis, E. S., et al. (2012). Multiple mechanisms and challenges for the application of allopolyploidy in plants. *Int. J. Mol. Sci.* 13, 8696–8721. doi: 10.3390/ijms13078696
- Pfeifer, M., Kugler, K. G., Sandve, S. R., Zhan, B., Rudi, H., Hvidsten, T. R., et al. (2014). Genome interplay in the grain transcriptome of hexaploid bread wheat. *Science* 345:1250091. doi: 10.1126/science.1250091
- Plessis, A., Hafemeister, C., Wilkins, O., Gonzaga, Z. J., Meyer, R. S., Pires, I., et al. (2015). Multiple abiotic stimuli are integrated in the regulation of rice gene expression under field conditions. *Elife* 4:e08411. doi: 10.7554/eLife.08411
- Powell, J. J., Fitzgerald, T. L., Stiller, J., Berkman, P. J., Gardiner, D. M., Manners, J. M., et al. (2017). The defence-associated transcriptome of hexaploid wheat displays homoeolog expression and induction bias. *Plant Biotechnol. J.* 15, 533–543. doi: 10.1111/pbi.12651
- Rajkov, J., Shao, Z., and Berrebi, P. (2014). Evolution of polyploidy and functional diploidization in sturgeons: microsatellite analysis in 10 sturgeon species. *J. Hered.* 105, 521–531. doi: 10.1093/jhered/esu027
- Ramsey, J., and Schemske, D. W. (1998). Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu. Rev. Ecol. Syst.* 29, 467–501. doi: 10.1146/annurev.ecolsys.29.1.467
- Renny-Byfield, S., and Wendel, J. F. (2014). Doubling down on genomes: polyploidy and crop plants. *Am. J. Bot.* 101, 1711–1725. doi: 10.3732/ajb.1400119
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616
- Solhaug, E. M., Ihinger, J., Jost, M., Gamboa, V., Marchant, B., Bradford, D., et al. (2016). Environmental regulation of heterosis in the allopolyploid *Arabidopsis suecica*. *Plant Physiol.* 170, 2251–2263. doi: 10.1104/pp.16.00052
- Soltis, P. S., and Soltis, D. E. (2009). The role of hybridization in plant speciation. *Annu. Rev. Plant Biol.* 60, 561–588. doi: 10.1146/annurev.arplant.043008.092039
- Supek, F., Bosnjak, M., Skunca, N., and Smuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6:e21800. doi: 10.1371/journal.pone.0021800
- Takahagi, K., Inoue, K., Shimizu, M., Uehara-Yamaguchi, Y., Onda, Y., and Mochida, K. (2018). Homoeolog-specific activation of genes for heat acclimation in the allopolyploid grass *Brachypodium hybridum*. *Gigascience* 7:giy020. doi: 10.1093/gigascience/giy020
- Tanaka, M., Tanaka, H., Shitsukawa, N., Kitagawa, S., Takumi, S., and Murai, K. (2016). Homoeologous copy-specific expression patterns of MADS-box genes for floral formation in allopolyploid wheat. *Genes Genet. Syst.* 90, 217–229. doi: 10.1266/ggs.15-00029
- Van de Peer, Y., Maere, S., and Meyer, A. (2009). The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* 10, 725–732. doi: 10.1038/nrg2600
- Wang, Y., Yu, H., Tian, C., Sajjad, M., Gao, C., Tong, Y., et al. (2017). Transcriptome association identifies regulators of wheat spike architecture. *Plant Physiol.* 175, 746–757. doi: 10.1104/pp.17.00694
- Wendel, J. F. (2000). Genome evolution in polyploids. *Plant Mol. Biol.* 42, 225–249. doi: 10.1023/A:1006392424384
- Whittle, C. A., and Krochko, J. E. (2009). Transcript profiling provides evidence of functional divergence and expression networks among ribosomal protein gene paralogs in *Brassica napus*. *Plant Cell* 21, 2203–2219. doi: 10.1105/tpc.109.068411
- Yang, C., Zhao, L., Zhang, H., Yang, Z., Wang, H., Wen, S., et al. (2014). Evolution of physiological responses to salt stress in hexaploid wheat. *Proc. Natl. Acad. Sci. U.S.A.* 111, 11882–11887. doi: 10.1073/pnas.1412839111
- Zhan, J., Thakare, D., Ma, C., Lloyd, A., Nixon, N. M., Arakaki, A. M., et al. (2015). RNA sequencing of laser-capture microdissected compartments of the maize kernel identifies regulatory modules associated with endosperm cell differentiation. *Plant Cell* 27, 513–531. doi: 10.1105/tpc.114.135657

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Takahagi, Inoue and Mochida. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.