Genome-wide unbalanced expression bias and expression level dominance toward *Brassica oleracea* in artificially synthesized intergeneric hybrids of *Raphanobrassica*

Libin Zhang¹, Jianjie He¹, Hongsheng He¹, Jiangsheng Wu² and Maoteng Li¹✉

**Abstract**

*Raphanobrassica* (RrRcCc, 2n = 4x = 36), which is generated by distant hybridization between the maternal parent *Raphanus sativus* (RrRr, 2n = 2x = 18) and the paternal parent *Brassica oleracea* (C°C°, 2n = 2x = 18), displays intermediate silique phenotypes compared to diploid progenitors. However, the hybrid shares much more similarities in silique phenotypes with those of *B. oleracea* than those of *R. sativus*. Strikingly, the silique of *Raphanobrassica* is obviously split into two parts. To investigate the gene expression patterns behind these phenomena, transcriptome analysis was performed on the upper, middle, and lower sections of pods (RCsiu, RCsim, and RCsil), seeds in the upper and lower sections of siliques (RCseu and RCsel), seeds in the upper and lower sections of siliques (RCsiu, RCsim, and RCsil), seeds in the upper and lower sections of siliques (RCseu and RCsel) from *Raphanobrassica*, whole pods (Rsi and Csi) and all seeds in the siliques (Rse and Cse) from *R. sativus* and *B. oleracea*. Transcriptome shock was observed in all five aforementioned tissues of *Raphanobrassica*. Genome-wide unbalanced biased expression and expression level dominance were also discovered, and both of them were toward *B. oleracea* in *Raphanobrassica*, which is consistent with the observed phenotypes. The present results reveal the global gene expression patterns of different sections of siliques of *Raphanobrassica*, pods, and seeds of *B. oleracea* and *R. sativus*, unraveling the tight correlation between global gene expression patterns and phenotypes of the hybrid and its parents.

**Introduction**

Hybridization provides a way of interspecific or inter-generic genome transfer for incorporating preferable traits from the parents to the progeny. Allopolyploid hybrids generated from interspecific or intergeneric hybridization have provided abundant genetic resources for molecular biology research and crop breeding. Many crops, such as wheat¹, cotton², rapeseed³, and Chinese cabbage⁴, are derived from hybridization between different species. However, polyploidization might induce several important changes at the genetic, gene expression, and epigenetic levels⁵–¹⁰. In recent decades, many efforts have been made to produce synthetic allopolyploid species, such as *Arabidopsis¹¹*, *Brassica¹³*, *Oryza¹⁵*, *Nicotiana¹⁶*, *Triticum¹⁷* and *Gossypium¹⁸*. Among them, *Brassica* species are one of the model systems to study crop allopolyploidization¹⁹,²⁰. For example, miRNA expression patterns were explored between *B. napus*, *B. rapa*, and their F1 hybrids²¹. However, the genomic instability and the difficulty of the production of distant hybrids impede a better understanding of global expression patterns and/or regulatory mechanisms of gene expression during hybridization.

Allopolyploids often show genome-wide expression level dominance and/or homoeolog expression bias, namely, the hybrid is more similar and/or biases to one...
parent than the other\textsuperscript{22}. Such expression dominance and/or bias phenomena have been reported in species such as \textit{B. napus}\textsuperscript{23}, \textit{P. hesperium}\textsuperscript{24}, \textit{G. hirsutum}\textsuperscript{25,26}, and \textit{T. turgidum}\textsuperscript{27}. One representative example has been recently reported in newly resynthesized \textit{B. napus} allopolyploids, in which genome-wide expression bias and dominance are both toward the A-genome\textsuperscript{13}. It has also been revealed in which genome-wide expression bias and dominance are both toward the \textit{Aegilops tauschii} (RsRs, maternal parent) and diploid \textit{Aegilops tauschii} \textit{R. sativus}, \textit{C°C°} for \textit{B. oleracea}, and \textit{R'R'C'C°} for \textit{Raphanobrassica}. Siliques of the hybrid exhibit intermediate phenotypes compared to those of the diploid progenitors (Fig. 1b–e and Fig. S1). Specifically, the siliques of the hybrid is obviously split into two parts joined by the valve shoulder (RCsim). The upper section of pods (RCsiu) displays more similarities to those of \textit{R. sativus}, and neither of them possesses the seed-bearing septum (Fig. 1b, c). Instead, seeds in RCsiu and Rsi are enclosed by sponge-like membranes. In contrast, RCsil shares more similarities with the siliques of \textit{C°C°} (Csi), and both harbor a seed-bearing septum (Fig. 1b, c). However, the whole siliques of \textit{R'R'C'C°} (4.65 ± 0.10 cm) are between those of the progenitor \textit{R'R°} (3.33 ± 0.14 cm) and \textit{C°C°} (5.37 ± 0.12 cm), but it is much more similar to that of \textit{C°C°}. A similar phenomenon was found in the phenotype of siliques, with RCsiu (4.83 ± 0.17 mm) and RCsil (4.67 ± 0.21 mm) sharing much more similarities with \textit{C°C°} (3.83 ± 0.17 mm) than \textit{R'R°} (10.00 ± 0.26 mm) (Fig. 1d and Fig. S1). As expected, the number of seeds per siliques of \textit{R'R'C'C°} (10.33 ± 1.21) was between those of \textit{R'R°} (4.17 ± 1.17) and \textit{C°C°} (18.67 ± 1.03), with more seeds in RCsil (7.67 ± 0.82) than in RCsiu (2.67 ± 0.82) (Fig. 1e). In addition, the color of mature seeds, seed area, and thousand seed weight (TSW) of the hybrid were also between those of the two parents but were much more similar to those of \textit{C°C°} (Fig. S1). Of note, few differences in the phenotypes of seed area, TSW, and color of mature seeds between RCseu and RCsel were found.

In brief, siliques phenotypes of siliques structure, siliques length, siliques width, seed area, the color of mature seeds, and TSW of the hybrid are between those of its parents but very similar to those of \textit{C°C°}. Conclusively, phenotypes of the whole siliques of \textit{Raphanobrassica} are much more similar to those of \textit{B. oleracea} than those of \textit{R. sativus}, even though the developing seeds of the hybrid also share slight similarities with those of \textit{R. sativus} (Fig. 1b), suggesting that global gene expression pattern of \textit{Raphanobrassica} siliques is much more similar to its paternal parent than that of its maternal parent.

Global mRNA transcriptome analysis of different sections of siliques

To investigate the gene expression patterns behind these interesting phenomena, we first performed syntenic analysis between the paternal (\textit{C°C°}, assembly BOL) and maternal (\textit{R'R°}, assembly Rs1.0) genomes. It was revealed that the synteny depth between the paternal and maternal genomes is a 3:3 pattern (Fig. 1f), and the genome-wide synteny between these two genomes is of high level (Fig. 1g), indicating that many genes of the two genomes are highly homologous. Subsequently, to unravel the global gene expression patterns, RNA-seq of RCsiu, RCsim,
RCsil, RCsey, RCsel, Rsi, Rse, Csi, and Cse was performed. All of the generated sequencing reads were processed and mapped to the R'R' genome (assembly Rs1.0), C'C' genome (assembly BOL), and integrated genomes of R'R' and C'C' (Table S1). Except for Csi3, we obtained approximately 1.4 × 10⁹ clean reads (5.4 × 10⁷ on average), approximately 1.2 × 10⁹ of which (83.4%) were mapped to the corresponding genomes. Most of the mapped reads were uniquely mapped reads (94.2%), and the vast majority of them were mapped to exons (Table S1). Pearson correlation coefficients (PCCs) between these samples ranged from 0.801–0.996, which indicates the high-quality of the biological replicates (Fig. S2). Twelve genes were randomly selected for further RT–qPCR
analysis, and the results showed the same expression patterns as the mRNA-seq data (Fig. S3).

There were indeed some differences in global gene expression patterns among RCsiu, RCSim, and RCsil, but few were found between those of RCseu and RCSel (Fig. 2a–c). Thus, for convenience, only the transcriptome of RCseu was used for further homoeolog expression bias and expression level dominance analyses in the following text. Moreover, the results also revealed that the global gene expression level distribution of \( R'R'C'C' \) was more similar to that of \( C'C'C' \) than that of \( R'R' \) in all five siliques sections (Fig. 2a and Fig. S4).

In total, we identified 59,018 (41,140 known genes and 17,878 novel gene loci) genes as expressed (empirical cutoff value: FPKM $\geq$ 1) in at least one of twenty-six samples (except for Csi3). Specifically, a total of 28,517, 14,378, 30,084, 17,574, 21,184, 23,899, 23,873, 22,504 and 22,811 expressed genes were identified in Rsi, Csi, Rse, Cse, RCsiu, RCsim, RCsil, RCseu, and RCsel, respectively.
RCsim, RCsil, RCseu and RCsel, respectively (Fig. 2b). The three sections of pods and the seeds in the two sections of \( R'R'C'C' \) siliques have intermediate numbers of expressed genes compared to those of the two progenitors, which is consistent with the intermediate phenotypes of siliques of the hybrid (Fig. 2b). Dramatically, most of the differentially expressed genes (DEGs) in RCsiu, RCseu, and RCsel were downregulated relative to those in the diploid progenitors. However, there were more upregulated DEGs than downregulated DEGs in RCsim and RCsil when compared to Csi (Fig. 2b). Different sets of DEGs were also classified by GO terms. Notably, the upregulated and downregulated DEGs in seeds of \( R'R'C'C' \) relative to the \( R'R' \) genome and the C ‘C’ genome, respectively, were enriched for photosynthesis-related GO terms (Fig. 2d). However, the expression levels of the 20 randomly selected photosynthesis-related DEGs in \( Raphanobrassica \) were comparable between the subgenomes of \( R'R' \) and \( C'C' \) (Fig. 2e).

Together, the sharp alteration in the global gene expression patterns of pods and seeds of \( R'R'C'C' \) relative to those of the parental genomes indicates a “transcriptome shock” in the hybrid genome. The intermediate numbers of expressed genes of pods and seeds of the hybrid between those of the progenitors are also consistent with the intermediate phenotypes of pods and seeds of the hybrid. However, the genomic regulation mechanisms underlying these phenomena remain unclear.

**Analysis of expression profile differences between the upper and lower sections of siliques of \( R'R'C'C' \)**

Although there are few differences in the expression profile between RCseu and RCsel, we want to know what the specific differences are, as well as those between RCsiu and RCsil. KEGG terms of upregulated DEGs of RCseu relative to RCsel were significantly enriched for ‘plant hormone signal transduction’ and ‘MAPK signaling pathway-plant’ (Fig. 3a). Detailed analysis revealed that two of these DEGs (three unique genes in total) negatively regulate ABA signaling, and one positively regulates...
cytokinin signaling. Suppression of AHG3 (gene ID: 106295669), a negative regulator of ABA signaling, could significantly accelerate fruit ripening in tomato34. However, KEGG analysis of downregulated DEGs of RCseu relative to RCsel was significantly enriched for ‘Protein processing in endoplasmic reticulum’ (Fig. 3b). Interestingly, seven homologous gene pairs were detected within the downregulated DEGs of RCseu relative to RCsel, and three pairs (gene ID: 108812972–106312198, 108817415–106299879, and 108857941–106295243) were identified as molecular chaperones that play critical roles in ER stress35. A previous study revealed that these molecular chaperones might be required for desiccation tolerance in Arabidopsis36. These results indicate that the ripening of RCseu might lag behind that of RCsel, which is in line with the phenotype in which the color of RCsel is deeper than that of RCseu (Fig. 1b). Furthermore, Venn diagram analysis revealed a high proportion of coexpressed genes between RCseu and RCsel (20,845, 95.6% in RCseu and 95.9% in RCsel), and the two tissues shared more coexpressed genes with Rse than Cse (Fig. 3c). In addition, the difference in the expression level of specifically expressed genes in RCseu and RCsel was minimal, and the top eight highly specifically expressed genes in RCsel were identified as novel gene loci (Fig. S5). Together, these results indicate that there are subtle, but nearly no, differences in expression profiles between RCseu and RCsel, which is in line with the phenotypes.

In contrast, the percentages of coexpressed genes between RCsil (18,884, 86.7% of total expressed genes) and RCsiu (18,884, 92.5% of total expressed genes) were smaller than those of RCseu and RCsel, and RCsil conferred more specific expressed genes than RCsiu (Fig. 3f). In addition, upregulated and downregulated DEGs of RCsiu relative to RCsel were enriched for different KEGG terms, for example, ‘Ribosome’ and ‘Plant pathogen interaction’, respectively (Fig. 3d, e). Intriguingly, we noticed that the pod wall of RCsiu is thicker but softer than that of RCsil. The GO term of upregulated DEGs of RCsil relative to those of RCsiu was significantly enriched for “cell wall”, thus indicating that RCsil is much more lignified than RCsiu (Fig. S6). We next identified 14 upregulated homoelogous gene pairs in RCsil enriched for the “cell wall”, which might explain the more rigid pod wall of RCsil than that of RCsiu (Fig. S6). Furthermore, pod shattering was clearly observed in RCsil, but not RCsiu, during harvest. A previous study showed that the IND gene is required for pod shattering37 and that the RPL gene could prevent replum cells from developing into valve margins38. Analysis of the expression level of the two aforementioned genes revealed that RPL is significantly expressed in Rsi, and the expression level in RCsiu is higher than that of RCsil. However, IND is mainly expressed in Csi and RCsil, which might explain the different phenotypes in pod shattering between RCsiu and RCsil (Fig. S6). Nevertheless, more transcriptome data in the early developmental stages of siliques are required to study these differences. Together, these results elucidate the expression profile differences between the upper and lower sections of siliques of the hybrid.

**Genome-wide unbalanced biased expression toward CC in pods and seeds of R′R′C′C′**

Homoeolog expression bias refers to the preferential expression of one homoeolog relative to the other in the hybrids22. To study the homoeolog expression bias in RCsiu, RCsim, RCsil, and RCseu, the genes that were expressed in at least one progenitor and where both homoeologs were expressed in Raphanobrassica were analyzed. First, the BBH method was adopted to identify 23,566 orthologous gene pairs between R′R′ and C′C′ genomes (homoeologous gene pairs of the hybrid subgenomes). Then, a total of 2627, 2939, 2994, and 2594 homoeologous gene pairs were used for expression bias analysis in RCsiu, RCsim, RCsil, and RCseu, respectively. Unbalanced biased expression with a preference toward CC was observed in all four tissues, with a total of 1177 (44.7%), 892 (30.3%), 898 (29.9%), and 1001 (38.6%) homoeologous gene pairs detected in RCsiu, RCsim, RCsil, and RCseu, respectively. Unbalanced biased expression with a preference toward CC was observed in all four tissues, with a total of 1177 (44.7%), 892 (30.3%), 898 (29.9%), and 1001 (38.6%) homoeologous gene pairs detected in RCsiu, RCsim, RCsil, and RCseu, respectively. Unbalanced biased expression with a preference toward CC was observed in all four tissues, with a total of 1177 (44.7%), 892 (30.3%), 898 (29.9%), and 1001 (38.6%) homoeologous gene pairs detected in RCsiu, RCsim, RCsil, and RCseu, respectively. Unbalanced biased expression with a preference toward CC was observed in all four tissues, with a total of 1177 (44.7%), 892 (30.3%), 898 (29.9%), and 1001 (38.6%) homoeologous gene pairs detected in RCsiu, RCsim, RCsil, and RCseu, respectively. Unbalanced biased expression with a preference toward CC was observed in all four tissues, with a total of 1177 (44.7%), 892 (30.3%), 898 (29.9%), and 1001 (38.6%) homoeologous gene pairs detected in RCsiu, RCsim, RCsil, and RCseu, respectively. Unbalanced biased expression with a preference toward CC was observed in all four tissues, with a total of 1177 (44.7%), 892 (30.3%), 898 (29.9%), and 1001 (38.6%) homoeologous gene pairs detected in RCsiu, RCsim, RCsil, and RCseu, respectively. Unbalanced biased expression with a preference toward CC was observed in all four tissues, with a total of 1177 (44.7%), 892 (30.3%), 898 (29.9%), and 1001 (38.6%) homoeologous gene pairs detected in RCsiu, RCsim, RCsil, and RCseu, respectively. Unbalanced biased expression with a preference toward CC was observed in all four tissues, with a total of 1177 (44.7%), 892 (30.3%), 898 (29.9%), and 1001 (38.6%) homoeologous gene pairs detected in RCsiu, RCsim, RCsil, and RCseu, respectively. Unbalanced biased expression with a preference toward CC was observed in all four tissues, with a total of 1177 (44.7%), 892 (30.3%), 898 (29.9%), and 1001 (38.6%) homoeologous gene pairs detected in RCsiu, RCsim, RCsil, and RCseu, respectively. Unbalanced biased expression with a preference toward CC was observed in all four tissues, with a total of 1177 (44.7%), 892 (30.3%), 898 (29.9%), and 1001 (38.6%) homoeologous gene pairs detected in RCsiu, RCsim, RCsil, and RCseu, respectively. Unbalanced biased expression with a preference toward CC was observed in all four tissues, with a total of 1177 (44.7%), 892 (30.3%), 898 (29.9%), and 1001 (38.6%) homoeologous gene pairs detected in RCsiu, RCsim, RCsil, and RCseu, respectively. Unbalanced biased expression with a preference toward CC was observed in all four tissues, with a total of 1177 (44.7%), 892 (30.3%), 898 (29.9%), and 1001 (38.6%) homoeologous gene pairs detected in RCsiu, RCsim, RCsil, and RCseu, respectively. Unbalanced biased expression with a preference toward CC was observed in all four tissues, with a total of 1177 (44.7%), 892 (30.3%), 898 (29.9%), and 1001 (38.6%) homoeologous gene pairs detected in RCsiu, RCsim, RCsil, and RCseu, respectively. Unbalanced biased expression with a preference toward CC was observed in all four tissues, with a total of 1177 (44.7%), 892 (30.3%), 898 (29.9%), and 1001 (38.6%) homoeologous gene pairs detected in RCsiu, RCsim, RCsil, and RCseu, respectively. Unbalanced biased expression with a preference toward CC was observed in all four tissues, with a total of 1177 (44.7%), 892 (30.3%), 898 (29.9%), and 1001 (38.6%) homoeologous gene pairs detected in RCsiu, RCsim, RCsil, and RCseu, respectively.
Fig. 4 Analysis of homoeolog expression bias of homoeologous gene pairs in RrRrCcCr. 

(a) Homoeolog expression bias analyses of RCsiu, (b) RCsim, (c) RCsil and (d) RCseu. The numbers of homoeologous pairs and their proportions to the total numbers of analyzed pairs are listed. The sizes of the circles in the diploid progenitor RsRs and CoCo indicated the relative expression levels of the orthologs (homoeologs in the hybrid), while relative expression levels of the homoeologs in RrRrCrCr were indicated by the area ratio of the circles. In addition, the numbers in red indicate that the expression patterns of homoeologs in RrRrCrCr maintained parental conditions, while those in blue indicate novel expression bias in the hybrid.
and RCseu share more similarities to *R. sativus* than those of RCsim and RCsil.

**Genome-wide expression level dominance biased toward CC in pods and seeds of *Raphanobrassica***

Expression level dominance (ELD) is another important concept distinct from homoeolog expression bias, which refers to the total expression level of a homoeolog pair in allotetraploids compared to its two progenitors\(^{22}\). To explore the ELD in RCsiu, RCsim, RCsil, and RCseu of *Raphanobrassica*, we only focused on the genes that are expressed in at least one progenitor and where both homoeologs are expressed in *Raphanobrassica*, and these gene pairs were classified into 12 categories by comparing the total expression of the homoeolog pairs in \(R^R R^C C^C\) to those in the progenitors (Fig. 5a).

The results revealed that the numbers of homoeolog pairs displaying CC expression level dominance (ELD-CC, categories II and XI) were much larger than those displaying RR expression level dominance (ELD-RR, categories IV and IX) in RCsiu (621/23.6% vs. 30/1.1%), RCsim (1 169/39.8% vs. 57/1.9%), RCsil (1069/35.7% vs. 79/2.6%), and RCseu (1026/39.6% vs. 45/1.7%) (Fig. 5a). Notably, the homoeolog pairs that display ELD-CC are mainly from category XI, indicating the significant downregulation of \(R^R R^C\) homoeologs compared to those in *R. sativus*. Additionally, 914 (34.8%), 643 (21.9%), 626 (20.9%) and 813 (31.3%) homoeologous gene pairs showed additivity in RCsiu, RCsim, RCsil and RCseu, respectively.
(Fig. 5a). Furthermore, a total of 754 (28.7%), 639 (21.7%), 678 (22.6%), and 482 (18.6%) homoeolog pairs displayed transgressive downregulation (categories III, VII, and X) (Fig. 5a). However, only a total of 269 (10.2%), 408 (13.9%), 467 (15.6%), and 210 (8.1%) homoeolog pairs exhibited equivalent expression (“No change” in Fig. 5a) in RCsiu, RCsim, RCSil, and RCseu, respectively. Moreover, the functional categorizations of different sets of homoeolog pairs showing ELD were also studied in GO terms, with three sets of genes (siu-C-ELD, sil-C-ELD, and seu-C-ELD) enriched for “protein heterodimerization activity” and “rRNA binding” (Table S2). However, genes of sim-C-ELD were enriched for photosynthesis (Table S2).

The relationship between ELD and individual homoeolog expression levels in the RCsiu, RCsim, RCSil, and RCseu of *Raphanobrassica* was also investigated by comparing the individual homoeolog expression levels to those in the diploid progenitors. As expected, the main reason for ELD is that numerous homoeologs or RrRr homoeologs were downregulated compared to those in the diploid progenitors, with 547 (367 + 180, 84.0%), 1 026 (656 + 370, 83.7%), 917 (559 + 358, 79.9%) and 1 026 (670 + 356, 95.8%) homoeolog pairs detected in RCsiu, RCsim, RCSil, and RCseu, respectively (Fig. 5b). Together, these results reveal that the expression level dominance in RCsiu, RCsim, RCSil, and RCseu is toward CC, and the main reason for ELD is due to the downregulated expression of both RrRr and C°C° homoeologs or the downregulated expression of RrRr homoeologs.

**Discussion**

Hybridization between different species promotes genome evolution and the formation of new crops by genetic recombination, which is usually used to generate novel cultivars with novel genetic traits. Hybrids usually display hybrid vigor (heterosis), disease resistance, and some other traits in future generations. For example, the Chinese cabbage *F1* hybrid *Xin No. 3*, a popular leafy crop used for the analysis of mRNA expression profiles. GISH experiments showed that *Raphanobrassica* has the expected chromosome number through many generations of fertility selection. *R. sativus* was reported to resist different diseases of powdery mildew disease, which may provide an important bridge for disease resistance breeding. Moreover, plenty of evidence has shown that *Raphanobrassica* has resistance to clubroot disease, beet cyst nematodes, and effective crossability with *Brassica* species. Therefore, *Raphanobrassica* generated in this work provides a potential new genetic resource for practical use in breeding clubroot-resistant cultivars of *Brassica* crops. In addition, *Raphanobrassica* is superior to its two progenitors in dry weight and crude protein content, which would serve as a useful fodder crop, and the dehiscence characteristic of RCSil makes the siliques easily harvested mechanically. However, how the interaction between the RrRr genome of *R. sativus* and the C°C° genome of *B. oleracea* affects chromosome rearrangement in *Raphanobrassica* remains unclear. Nevertheless, our global gene expression pattern analysis reported here will help the research community broadly understand the consequence of the incorporation of RrRr and C°C° genomes in the silique of *Raphanobrassica*.

Transcriptome shock in the hybrid refers to extensive changes to patterns of parental gene expression. In this study, transcriptome shock was observed in all five tissues of RrRrC°C°. A previous study revealed that transcriptome shock might be caused by interspecific hybridization and subsequently ameliorated by genome duplication. Thus, we speculated that changes in the expression patterns of the diploid F1 hybrid might be more extensive than those of the allotetraploid hybrid when compared to the diploid progenitors. In addition, altered chromatin compaction and histone methylation in the hybrid, as well as cis- and trans-regulatory divergence between progenitor species, may also explain the transcriptome shock observed here. Considering that most DEGs in RrRrC°C° are downregulated relative to those of the maternal parent *R. sativus*, we speculated that chromatin of the RrRr sub-genome is much more compact than that of the C°C° subgenome, and this hypothesis could be confirmed by ATAC-seq and Hi-C in future studies.

Homoeolog expression bias has been broadly studied in the past decade. Interestingly, in our study, the silique of *Raphanobrassica* was split into two parts, and homoeolog expression bias showed a preference for CC. However, more homoeolog gene pairs showed RR-biased expression in RCsiu and RCseu than in RCsim and RCSil. Strikingly, homoeolog expression and induction bias were revealed in *Triticum aestivum* when bread wheat was infected by the fungal pathogen *Fusarium pseudo-graminearum*. Once infected by the aforementioned fungal pathogen, B and D homoeologs exhibit stronger responses to the infection than A homoeologs, which implies distinctions of homoeolog expression bias under different biotic and abiotic stresses in many resynthesized or natural hybrids. Indeed, some recent studies have also revealed that the expression levels of certain homoeologous genes are altered under heat tolerance, water...
stress, and iron deficiency in polyploid wheat\textsuperscript{48–50}. Hence, homoeolog expression bias of pods and seeds of \textit{Raphanobrassica} under different stresses still remains to be studied.

ELD has also been widely explored in many resynthesized hybrids\textsuperscript{13,15,17,51,52}. The parental genome toward which ELD of the synthetic hybrid biases is consistent with the similarity of phenotypes between the hybrid and parents. For example, in our study, ELD of pods and seeds of \(R'R'C'C'\) showing preference toward \(CC\) is consistent with the fact that phenotypes of the whole siliques of \textit{Raphanobrassica} are similar to those of \textit{B. oleracea}. Hence, we could infer the ELD of the resynthesized hybrid by the similarity of the phenotypes between the progenitors and the hybrid. In addition, there is different ELD biased expression in different organs within the same organism. For example, one previous study revealed that the gene in immature leaves of \textit{Raphanobrassica} displays genome-wide ELD bias toward \(RR\), which is different from that of the pods and seeds in the present study\textsuperscript{51}. Intrigued, the ELD of different developmental stages of pods and seeds of \textit{Raphanobrassica} still remains to be explored to comprehensively understand the relationships of global gene expression patterns between the hybrid and its two diploid progenitors.

Collectively, transcriptome shock was observed in the siliques of \textit{Raphanobrassica} due to the incorporation and interaction of the two parental genomes, which might lead to the split of the siliques into two parts. The upper and lower sections of siliques of the hybrid exhibit expression profile differences, and the related genes are involved in different biological processes. Many more genes in the siliques of \textit{Raphanobrassica} exhibit homoeolog expression bias and expression level dominance toward \textit{B. oleracea} than those toward \textit{R. sativus}. Nonetheless, \(RCSi\) and \(RCSe\) have more RR-biased expression genes than RCsi and RCse, which might explain why the phenotypes of \(RCSi\) and \(RCSe\) of the hybrid share more similarities to \textit{R. sativus} than those of RCsi and RCse.

In conclusion, our mRNA transcriptome profiles are consistent with the observed phenotypes and unravel the tight correlation between the phenotypes and global gene expression patterns in the hybrid and its parents. Our results reported here provide a good reference to study plant polyploidy and mine potential candidate genes responsible for specific phenotypes. However, whether the phenotypes are tightly correlated with other multiomics data in the synthesized hybrid and its parents remains to be explored.

**Materials and methods**

**Plant materials and sampling**

The diploid maternal \textit{R. sativus} (\(2n = 2x = 18\)) and paternal \textit{B. oleracea} var. \textit{alboglabra} (\(2n = 2x = 18\)) were used as parental lines to generate allotetraploid \textit{Raphanobrassica} (\(2n = 4x = 36\)) by crossing, and the hybrid chromosomes were doubled in the \(F_1\) generation by colchicine treatment followed by seed fertility selection\textsuperscript{32}. All of the plant materials were planted in Wuhan, China. Seeds with good fertility were obtained in the \(F_{10}\) generation, and GISH analysis revealed that the hybrid genome contained all 18 chromosomes from \(R'R^4\) and all 18 chromosomes from \(C'C^4\).

**Library construction, Illumina sequencing, read mapping and DEG analysis**

Whole pods (\(Rsi\) and \(Csi\)), seeds (\(Rse\) and \(Cse\)) of siliques of the diploid progenitors, three sections of the whole pods (\(RCsi\), \(RCsm\), and \(RCs\)) and all seeds in the two sections of the siliques (\(RCse\) and \(RCsel\)) of the hybrid, with three biological replicates of each sample, were obtained at 35 DAF for total mRNA isolation and sequencing library construction.

For RNA-seq, RNA library preparation for each sample was performed according to the manuals provided by Illumina. The aforementioned 27 sequencing libraries were sequenced on the Illumina HiSeq 2000 platform. The raw data were filtered by NGS QC Toolkit\textsuperscript{53} to obtain high-quality clean reads. The total clean reads of \textit{R. sativus} and \textit{B. oleracea} obtained after filtration were mapped to the \(R'R^4\) and \(C'C^4\) genomes, respectively, while those of \(R'R'C'C'\) were mapped to the integrated genomes of \(R'R^4\) and \(C'C^4\) using HISAT software (Table S1)\textsuperscript{54}. We also mapped the clean reads of \textit{R. sativus} and \textit{B. oleracea} to the \(C'C^4\) and \(R'R^4\) genomes to examine whether erroneous mapping existed in the hybrid, and some of the randomly selected results were further confirmed by RT–qPCR. The novel gene loci were predicted by StringTie\textsuperscript{55} and annotated by the Pfam database\textsuperscript{56}. The mapping results were also processed by StringTie to obtain FPKM for all 27 samples, and the genes whose FPKM \(\geq 1\) were identified as expressed. Differentially expressed gene analysis between the 27 samples was conducted using the \(R\) package DEseq\textsuperscript{57}, and genes that exhibited a difference of at least twofold with a corrected \(P\) value \(\leq 0.05\) were regarded as significantly differentially expressed.

**GO and KEGG enrichment analysis**

GO and KEGG enrichment analyses for the DEGs were conducted using the \(R\) package clusterProfiler\textsuperscript{58} and the KEGG database\textsuperscript{59}, respectively. The GO and KEGG terms of the three ontologies exhibiting \(p_{adj} \leq 0.05\) were regarded as significantly enriched.

**Annotation of homoeologous gene pairs in \textit{Raphanobrassica}**

For convenience, we considered that most of the homoeologs remained in a one-to-one relationship after
polypliodization within the subgenomes of R'RC'C', although the homoeologous relationships are not necessarily one-to-one. Based on this, we adopted the BBH method to infer homoeologs, although this method has inherent drawbacks. Thus, as shown in Fig. S7, the 43,882 gene models of C'C' and the 48,203 gene models of R'RC' were blasted against each other using the blastn program with a search cutoff E-value of 1e-10 to identify the orthologous gene pairs, which were termed homoeologous gene pairs in the hybrid. Gene pairs with a sequence identity > 90% were regarded as orthologous gene pairs in the two progenitors or homoeologous gene pairs in the hybrid.

Analysis of homoeolog expression bias and expression level dominance

In the process of homoeolog expression bias and ELD analyses, we only paid attention to the genes that were expressed in at least one progenitor and where both homoeologs were expressed in Raphanobrassica. Student’s t-test (P < 0.05) was used to compare the expression level of each homoeologous gene pair in the two progenitors (R'RC' vs. C'C') and the Raphanobrassica hybrid (R'RC' vs. C'C') in the process of homoeolog expression bias according to previously described methods. During the analyses of ELD in different tissues, the collective expression level of a homoeologous gene pair in the Raphanobrassica hybrid was compared to that of the diploid progenitors, that is, (R'RC' + C'C') vs. R'RC' and (R'RC' + C'C') vs. C'C', using Student’s t-test (P < 0.05) as described previously. Furthermore, twelve possible bins were classified according to Yoo et al..

Collinearity analysis

The protein, CDS, and mRNA FASTA files and GFF files for B. oleracea and R. sativus were retrieved from the NCBI genome database, and only genes of R. sativus from the nine longest scaffolds were used for analysis. Moreover, only the first transcript was used when the gene had more than one transcript. Homologous genes within the self-genome or cross-genome were obtained using BLASTP (blastp–2.9.0+) with a cutoff e-value of 1e-10, and only the top five hits were used for downstream collinearity analysis using the MCSanX package.

RT–qPCR validation

The expression level of mRNA was detected in the nine tissues of R. sativus, B. oleracea, and hybrids using quantitative reverse transcription-PCR (RT–qPCR). For mRNA expression level detection, 1 μg of total RNA was reverse-transcribed using SuperScript III Reverse Transcriptase (Invitrogen) and oligo (dT)18 according to the manufacturer’s protocol. The qPCR experiment was carried out using an ABI 7300 (ABI), and each reaction was performed in triplicate. U6 RNA was set as an internal reference gene for mRNA expression detection. The primers for mRNA RT–qPCR are listed in Table S3.

Acknowledgements

This research was supported by the National Key Research and Development Program of China (2017YFD0101701) and the National Natural Science Foundation of China (Grant No. 31540083).

Author contributions

M.L. conceived the idea, designed the study, and revised the manuscript. L.Z., H.H., and J.W. performed the experiments. J.H. performed the bioinformatic analysis. L.Z. and J.H. wrote the manuscript and contributed equally to this work. All the authors have reviewed and approved the manuscript.

Data availability

Sequencing data of R. sativus, B. oleracea, and the hybrid Raphanobrassica were deposited in the SRA Database in NCBI (Accession numbers: SRR12191747, SRR12191748, SRR12191749, SRR12191750, SRR12191751, SRR12191752, SRR12191753, SRR12191754, SRR12191755, SRR12191756, SRR12191757, SRR12191758, SRR12191759, SRR12191760, SRR12191761, SRR12191762, SRR12191763, SRR12191764, SRR12191765, SRR12191766, SRR12191767, SRR12191768, SRR12191769, SRR12191770, SRR12191771, SRR12191772, SRR12191773).

Conflict of interest

The authors declare no competing interests.

Supplementary information

The online version contains supplementary material available at https://doi.org/10.1038/s41438-021-00672-2.

Received: 29 March 2021 Revised: 27 July 2021 Accepted: 30 July 2021

Published online: 01 December 2021

References

1. Brenchley, R. et al. Analysis of the bread wheat genome using whole-genome shotgun sequencing. Nature 491, 705–710 (2012).
2. Wang, M. et al. Reference genome sequences of two cultivated allotetraploid cottons, Gossypium hirsutum and Gossypium barbadense. Nat. Genet. 51, 224–229 (2019).
3. Chalhoub, B. et al. Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. Science 345, 950–953 (2014).
4. Li, P. et al. Genome-wide analysis of changes in mRNA and target gene expression reveals key roles in heterosis for Chinese cabbage biomass. Hortculture Res. 8, 39 (2021).
5. Jackson, S. & Chen, Z. J. Genomic and expression plasticity of polyploidy. Curr. Opin. Plant Biol. 13, 153–159 (2010).
6. Sosnovska, K. et al. Chromosome instabilities in resynthesized Brassica napus revealed by FISH. J. Appl. Genet. 61, 323–335 (2020).
7. Schiauva, M., Bodrug-Scheper, A., Dohrn, J. C. & Himmelbauer, H. Sub-genome evolution in allotetraploid plants. Plant J. 106, 672–688 (2021).
8. Qiu, T., Liu, Z. & Liu, B. The effects of hybridization and genome doubling in plant evolution via allopolyploidy. Mol. Biol. Rep. 47, 5549–5558 (2020).
9. Glombik, M., Bačovský, V., Hobza, R. & Kopecký, D. Competition of parental genomes in plant hybrids. Front. Plant Sci. 11, 200 (2020).
10. Voshall, A. & Motyama, E. N. Next-generation transcriptome assembly and analysis: Impact of ploidy. Methods: 176, 14–24 (2020).
11. Ni, Z. et al. Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. Nature 457, 327–331 (2009).
12. Letch, A. & Letch, I. Genomic plasticity and the diversity of polyploid plants. Science 320, 481–483 (2008).
13. Wu, J. et al. Homoeolog expression bias and expression level dominance in resynthesized allopolyploid Brassica napus. BMC Genomics 19, 586 (2018).
14. Zhang, L. et al. Integration of metabolome and transcriptome reveals flavonoid accumulation in the intergeneric hybrid between Brassica rapa and Raphanus sativus. Sci. Rep. 9, 18368 (2019).
15. Wu, Y. et al. Transcriptome shock in an interspecific F1 triploid hybrid of Oryza revealed by RNA sequencing. J. Integr. Plant Biol 58, 150–164 (2016).
