New Phytologist

Research

Asymmetric epigenetic modification and homoeolog expression bias in the establishment and evolution of allopolyploid *Brassica napus*

Mengdi Li[1], Weiqi Sun[1], Fan Wang[1], Xiaoming Wu[2] and Jianbo Wang[1]

[1]College of Life Sciences, Wuhan University, Wuhan 430072, China; [2]Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture, Oil Crops Research Institute of CAAS, Wuhan 430062, China

Author for correspondence:
Jianbo Wang
Email: jbwang@whu.edu.cn

Received: *4 March 2021*
Accepted: *9 July 2021*

New Phytologist (2021)
doi: 10.1111/nph.17621

Key words: allopolyploidization, asymmetric epigenetic modification, *Brassica napus*, *cis* and *trans* regulatory effects, evolution, homoeolog expression bias.

Summary

- This study explores how allopolyploidization reshapes the biased expression and asymmetric epigenetic modification of homoeologous gene pairs, and examines the regulation types and epigenetic basis of expression bias.
- We analyzed the gene expression and four epigenetic modifications (DNA methylation, H3K4me3, H3K27me3 and H3K27ac) of 29 976 homoeologous gene pairs in resynthesized, natural allopolyploid *Brassica napus* and an *in silico* 'hybrid'.
- We comprehensively elucidated the biased gene expression, asymmetric epigenetic modifications and the generational transmission characteristics of these homoeologous gene pairs in *B. napus*. We analyzed *cis*/*trans* effects and the epigenetic basis of homoeolog expression bias. There was a significant positive correlation between two active histone modifications and biased gene expression.
- We revealed that parental legacy was the dominant principle in the remodeling of homoeolog expression bias and asymmetric epigenetic modifications in *B. napus*, and further clarified that this depends on whether there were differences in the expression/epigenetic modifications of gene pairs in parents/progenitors. The maternal genome was dominant in the homoeolog expression bias of resynthesized *B. napus*, and this phenomenon was attenuated in natural *B. napus*. Furthermore, *cis* rather than *trans* effects were dominant when epigenetic modifications potentially affected biased expression of gene pairs in *B. napus*.

Introduction

Polyploidization has occurred in > 70% of angiosperms over of the course of their evolutionary history (Baduel *et al.*, 2018), and polyploids exhibit an adaptive advantage over diploid progenitors in terms of both phenotype and physiological characteristics (Madlung, 2013). Natural polyploids are formed in two ways, autopolyploidization and allopolyploidization (Leitch & Bennett, 1997). The latter is a powerful evolutionary force in vascular plant diversity and speciation, which consists of hybridization and whole genome duplication (WGD; Soltis *et al.*, 2009; Abbott *et al.*, 2013). Rapid and drastic changes tend to occur in the genome to circumvent incompatibility in allopolyploids, in which two or more divergent genomes abruptly reunite and double in the same cell (Feldman & Levy, 2005). Indeed, genetic and epigenetic changes occur extensively during and/or after allopolyploidization (Jackson & Chen, 2010). All of these changes may enable the successful establishment of nascent allopolyploids, increase ecological diversity, adapt to and expand into new geographic niches, and even alter community structure (Adams & Wendel, 2005; Chen, 2007; Yoo & Wendel, 2014; Segraves, 2017). However, most studies have focused on either natural or resynthesized allopolyploids, rather than on both, and it is necessary to link changes in gene expression that occur immediately after allopolyploidization to those which occur over the course of evolution (Wang *et al.*, 2016).

Allopolyploids often exhibit two or more subgenomes that are asymmetrical in structure and function, which plays an important role in their evolution (Roulin *et al.*, 2013; Li *et al.*, 2015; An *et al.*, 2019). The 'genomic shock' in allopolyploids often leads to subgenome-specific gene expression, which refers to the unequal expression of homoeologs derived from two different parental species (i.e. homoeolog expression bias; Flagel & Wendel, 2010), and this could promote the environmental flexibility and adaptability of allopolyploids (Scarrow *et al.*, 2021). A recent study showed that subgenome-specific selection of defense response genes contributes to the environmental adaptation of allopolyploid *Brassica napus* (Lu *et al.*, 2019). In addition to focusing on the fact of homoeolog expression bias *per se*, the more important issues of generational transmission patterns and the fate of homoeologous gene expression in resynthesized and natural allopolyploids are addressed here.

Gene expression is regulated by *cis* and *trans* regulatory elements, and their divergence is fundamental to phenotypic and

evolutionary diversity in plants (Bao *et al.*, 2019). The concomitant contributions of *cis* and *trans* effects make it difficult to observe or quantify them; however, gene expression analysis of parents and progenies has made it possible to distinguish between them (Xu *et al.*, 2014; Bao *et al.*, 2019). Specifically, homoeologous gene expression differences between parents are caused by different *cis* and *trans* effects, while they are caused by different *cis* effects in progeny (because gene pairs are in the same *trans* regulatory environment). The *cis*/*trans* regulatory patterns underlying the homoeolog expression bias in allopolyploids and their evolutionary divergence are worthy of attention.

Epigenetic modification, mainly including DNA methylation and histone modification (Braszewska-Zalewska *et al.*, 2010), can regulate gene activity in plants (Ding & Chen, 2018). Gene-related DNA methylation can occur in the promoter or gene body in plants (Zhang *et al.*, 2018), and genes are usually expressed constitutively in the latter case (Takuno & Gaut, 2013; Muyle & Gaut, 2019), while the former usually inhibits gene expression but in some cases promotes it (Lang *et al.*, 2017). There are two types of histone modification: those located in euchromatin (promotes gene expression; e.g. H3K4me3 and H3K27ac) and those located in heterochromatin (inhibits gene expression; e.g. H3K27me3 and H3K9me2; Li *et al.*, 2007). So far, the way in which epigenetic modification participates in the regulation of homoeolog expression bias in allopolyploids is still largely unknown. Furthermore, given the extensive epigenetic changes in allopolyploids, another critical issue is the heredity and variation of subgenomic epigenetic modifications during generational transmission in allopolyploids (Lv *et al.*, 2019).

Brassica napus (AACC) was formed by the hybridization and WGD of two diploid progenitors, *Brassica rapa* (AA) and *Brassica oleracea* (CC) *c.* 7500 yr ago, and these three species have been used as a model system for the study of allopolyploidization (Chalhoub *et al.*, 2014). Although the number of studies on changes in gene structure and expression in *B. napus* has been increasing gradually, the expression bias and epigenetic modification bias of homoeologous gene pairs and their evolutionary changes have rarely been reported in *B. napus*. Here, we conducted a comprehensive analysis of homoeolog expression and four epigenetic modifications in natural, resynthesized *B. napus* and *in silico* 'hybrid'. To elucidate the expression and epigenetic remodeling of homoeologs in *B. napus*, we investigated biased expression, asymmetric epigenetic modifications and their generational transmission rules. We also analyzed the *cis*/*trans* effect and the epigenetic bases of homoeolog expression bias, to try to explain the regulation mechanisms in *B. napus*. Our epigenome information is a valuable resource for the study of adaptive mechanisms in allopolyploid *B. napus* and also provides a reference for the study of other allopolyploid plants.

Materials and Methods

Plant materials

Seeds of four plant lines (Supporting Information Fig. S1), including natural *Brassica napus* L. (cv Darmor, AACC, $2n = 4x = 38$), resynthesized *B. napus* L. (HC-2, AACC, $2n = 4x = 38$) and its parents, *Brassica oleracea* L. (cv 3YS013, CC, $2n = 18$, the maternal parent) and *Brassica rapa* L. (cv 9JC002, AA, $2n = 18$, the paternal parent), were provided by the Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, China. Plants were incubated at 23°C under a 16 h : 8 h, light : dark photoperiod. Young leaves of 1-month-old plants were harvested and frozen in liquid nitrogen immediately for further use.

Sequencing and data analysis

Total RNA from 12 samples was extracted using the RNAprep Pure Plant Kit (Tiangen, Beijing, China). A PCR-cDNA sequencing kit (SQK-PCS109; Oxford Nanopore Technologies, Oxford, UK) was used for full-length cDNA library preparation. Sequencing was performed on the PromethION 24 (Oxford Nanopore Technologies). There are two main reasons for us choosing the Oxford Nanopore Technologies (ONT) over the Illumina sequencing platform. First, it is crucial to distinguish homoeologous genes and their expression for our study, and paralogous genes with highly similar sequences can lead to an ambiguous alignment of reads when using short-read sequencing, while ONT long-read sequencing can obtain more uniquely matched reads (Marchet *et al.*, 2019). Second, longer transcripts could be over-represented relative to shorter transcripts in short-read sequencing, and differential gene expression will be biased toward the relatively long transcripts (Oshlack & Wakefield, 2009), but ONT long-read sequencing can effectively avoid such theoretical bias (Dong *et al.*, 2021).

Guppy (v.3.2.10) was used for basecalling and NanoFilt (v.2.5.0) was used for filtering (length > 100, quality > 7) to ensure clean data were obtained. Clean reads were analyzed using Pinfish (v.1.0). Full-length reads from *B. rapa* and *B. olerocea* were mixed in a 1 : 1 ratio to form an *in silico* 'hybrid' (Wang *et al.*, 2016). Minimap2 (v.2.16) was used to map full-length reads to the reference *B. napus* genome v.5 (http://www.genoscope.cns.fr/brassicanapus/data/). Gene expression levels were normalized to TPM (transcripts per million reads).

Genomic DNA was isolated from 12 samples using the cetyl trimethylammonium bromide (CTAB) method. The bisulfite treatment was carried out using the EZ DNA Methylation Gold Kit (Zymo Research, Irvine, CA, USA). Whole-genome bisulfite sequencing (WGBS) libraries were constructed using the Acegen Bisulfite-Seq Library Prep Kit (AG0311; Acegen, Shenzhen, China), and sequencing was performed on an Illumina Hiseq X10 (30-fold sequencing depth). The *in silico* 'hybrid' was constructed by mixing WGBS reads of *B. rapa* and *B. olerocea* in proportion to their genome size, and the data size of the *in silico* 'hybrid' was consistent with that of *B. napus*. Raw reads were filtered using Trimmomatic (v.0.36) and clean reads were mapped to the reference genome using Bsmap (v.2.73).

Chromatin immunoprecipitation (ChIP) experiments were performed on four samples according to standard protocol, and the antibodies used were H3K4me3 (ab8580; Abcam, Cambridge, UK), H3K27me3 (9733; CST, Boston, MA, USA) and

H3K27ac (ab4729; Abcam). The ChIP-seq libraries were constructed using the VAHTS Universal DNA Library Prep Kit for ILLUMINA v.3 (ND607; Vazyme, Nanjing, China), and sequencing was performed on an Illumina Novaseq 6000 (PE150 model). The *in silico* 'hybrid' was constructed by mixing ChIP-seq reads of *B. rapa* and *B. olerocea* at a ratio of 1 : 1. The raw reads were filtered using TRIMMOMATIC (v.0.36), clean reads were mapped to the reference genome using STAR (v.2.5.3a) with default parameters, and the peaks were called using MACS2 (v.2.1.1). The annotation and distribution of peaks were analyzed using BEDTOOLS (v.2.25.0).

Biased expression analysis of gene pairs

To identify homologous gene pairs in the subgenomes of *B. napus*, we used the script for the best-hit method from our previous study (Li *et al.*, 2020). In this study, only unique mapped reads were used for further analysis. Differential expression analysis was performed using DESEQ2 (Love *et al.*, 2014), and the gene pairs were divided into A or C homoeolog biased expression gene pairs (A-/C-BEGs; fold change > 2, false discovery rate (FDR) < 0.001) and nonbiased expression gene pairs (nBEGs). The relative A-homoeolog expression ratio (A%) for each expressed gene pair detected was calculated and was visualized using GGPLOT2 (Villanueva & Chen, 2019). Extremely biased expression gene pairs (ex-BEGs) were defined as gene pairs with A% > 95% or < 5%, and they were divided into two groups: ex-BEG expression patterns in progenies that were inherited from parents/progenitors (group I), and those generated *de novo* (group II). For the generational transmission analysis, gene pairs in progenies were classified into nine patterns belonging to three groups, according to the intrinsic relative orthologous expression in the parents/progenitors, with reference to previous studies (Li *et al.*, 2020). Gene ontology (GO) enrichment analysis was performed using GOSEQ (v.1.10.0), in which a hypergeometric test was used. GO terms with FDR < 0.05 were defined as significantly enriched.

Characterization of *cis*/*trans* regulation effects

Cis and *trans* effects were distinguished using the read counts of *in silico* 'hybrid' (A_C), resynthesized and natural *B. napus* (RAC and NAC, respectively) from RNA sequencing according to previous studies (Xu *et al.*, 2014; Wang *et al.*, 2016; Bao *et al.*, 2019). *Cis* and *trans* effects co-regulated the expression divergence of homoeologous gene pairs in progenitors/parents (represented by A_C); thus, these two effects together were measured by the $\log_2$ ratios of the read count divergences, such that $A = \log_2(A/C)$. *Cis* effects were measured by the read count divergences of homoeologous gene pairs in progenies (RAC/NAC; $B = \log_2(A_n/C_n)$), because these gene pairs were in the common *trans* environment. Therefore, *trans* effects were derived by subtracting the expression divergences of gene pairs in progenies from those of progenitors/parents (A–B). The gene pairs were divided into seven regulatory categories according to the statistical results of A vs C ($A = /\neq 0$), A_n vs C_n ($B = /\neq 0$), and A vs B ($A = /\neq B$). Significant expression divergences of gene pairs in progenitors/parents ($A \neq 0$) and *cis* effects ($B \neq 0$) were determined using DESEQ2, as discussed in the sub-section 'Biased expression analysis of gene pairs', above, and the *trans* effects ($A \neq B$) were tested using Student's *t*-test, followed by the adjustment of *P*-values using FDR (FDR < 0.05; Benjamini & Hochberg, 1995). The line graph and boxplot were drawn using GGPLOT2. Venn diagrams were generated using the online tool VENNY (https://bioinfogp.cnb.csic.es/tools/venny/index.html).

Differential epigenetic modification analysis of gene pairs

The DNA methylation levels of the promoters and gene bodies of 29 976 gene pairs were calculated. Differential DNA methylation levels between subgenomes of *B. napus* were interrogated using Student's *t*-test (*P* value < 0.001). Differential histone modifications were identified using the EDGER package (fold change > 2, FDR < 0.001; Robinson *et al.*, 2010). All correlation scatter plots were plotted using GGPLOT2.

The statistical tests

The statistical significance of each comparison in this study was tested in R (R Foundation for Statistical Computing, Vienna, Austria; https://www.r-project.org), and a variety of statistical tests were employed, including the exact binominal test, Chi-squared test, Kolmogorov–Smirnov test (K-S test), Student's *t*-test, Wilcoxon rank sum test and Pearson's product-moment correlation.

Results

Homoeolog expression bias analysis

Twelve RNA-seq libraries were built and sequenced, including three replicates per leaf sample of the natural *B. napus* (NAC), the resynthesized *B. napus* (RAC) and their two diploid progenitors/parents, *B. oleracea* (C) and *B. rapa* (A; Fig. S1; Table S1). Resynthesized *B. napus* was obtained from diploid *B. oleracea* and *B. rapa* by hybridization and genome doubling. To evaluate the accumulated orthologous gene expression divergence of the two parents, an *in silico* 'hybrid' (A_C) was constructed. To investigate the extent of homoeolog expression bias, the 29 976 homoeologous gene pairs (see 'Biased expression analysis of gene pairs' sub-section in the Materials and Methods section) in three genotypes (A_C, RAC and NAC) were divided into A or C homoeolog biased expression gene pairs (A-/C-BEGs) and nonbiased expression gene pairs (nBEGs; Table 1; Fig. 1a–c). We found that

Table 1 The number of biased/unbiased expression gene pairs in the parental mix (A_C), resynthesized *Brassica napus* (RAC) and natural *B. napus* (NAC).

Type	A_C (%)	RAC (%)	NAC (%)
A bias	737 (3.0)	558 (2.3)	988 (4.2)
C bias	996 (4.1)	668 (2.7)	1015 (4.3)
Unbiased	22 540 (92.9)	23 094 (95.0)	21 648 (91.5)
Total	24 273 (100)	24 320 (100)	23 651 (100)

the total number of BEGs was lowest in RAC and highest in NAC, and the number of C-BEGs was significantly higher than that of A-BEGs in both A_C (exact binominal test, $P < 0.001$) and RAC ($P < 0.01$), but not significantly in NAC ($P > 0.05$). The BEGs/nBEGs were cross-analyzed (Fig. 1a–c) to show whether BEGs were specific to each genotype, and it was found that the majority of BEGs were exclusive to a given genotype. Most nBEGs were identified in two or all three genotypes. A total of 390 gene pairs showed the same expression bias in all genotypes, and the unique BEGs in all genotypes were highly variable. Furthermore, we want to know whether common BEGs would maintain their biased expression during evolution. We quantitatively examined the correlation between common BEGs in every genotype pairing and found that they were highly positively correlated (Figs 1d–f, 2e), suggesting that the biased expression of these gene pairs was highly inherited. Gene ontology (GO) enrichment analysis (Dataset S1) showed that BEGs in NAC were enriched in many novel GO terms, such as mitochondrial membrane and ion channel activity related terms. Moreover, functions of some A-/C-BEGs might have changed during evolution, for example, disulfide oxidoreductase activity related GO terms were mainly enriched by C-BEGs in A_C, but by A-BEGs in NAC. These results suggest that the A subgenome seems to play an increasingly important role in the evolution of *B. napus*.

To explore how parental orthologs would be expressed when they become homoeologs in *B. napus*, we calculated A% values (see 'Biased expression analysis of gene pairs' sub-section in the Materials and Methods section) for each expressed gene pair (Fig. 1g). The spectra for RAC/NAC were significantly reduced compared to that of A_C (K-S test, $P < 2.2 \times 10^{-16}$), and the spectrum for NAC was significantly reduced compared to that of RAC (K-S test, $P < 2.2 \times 10^{-16}$), suggesting that homoeologous gene expression gradually become more aggregated from parents to *B. napus*. We further investigated whether spectral variations would affect subgenome expression dominance. We found that the A subgenome dominance of A_C was not maintained in RAC, while the C subgenome dominance of RAC was maintained in NAC (Fig. 1g). It appears that the overall subgenome dominance and biased expression results in A_C did not match, but in fact this is understandable because the number of BEGs was quite small compared to the total number of gene pairs, and these two different results directly confirmed the large spectrum observed for A_C, which might impact the biased expression of genes in *B. napus*.

To further study the homoeologous gene expression, patterns of ex-BEGs were investigated based on A% data. All ex-BEGs were divided into two groups (Fig. 1h): ex-BEG expression patterns in progenies inherited from parents/progenitors (group I) and those generated *de novo* (group II). We made the following observations: the number of ex-BEGs was lower in group I than group II; there was a similar number of A and C homoeolog ex-BEGs (A-/C-ex-BEGs) in group I, while the number of C-ex-BEGs was greater than the number of A-ex-BEGs in group II; a novel extreme bias was rapidly established in RAC, which was absent in NAC, but novel ex-BEGs were present in NAC as a result of evolution. These results imply that the C subgenome dominance is likely to have been determined by novel ex-BEGs in resynthesized and natural *B. napus*. Gene ontology enrichment analysis (Dataset S2) showed that for group I, the co-enriched GO terms in NAC and A_C were mainly related to gene expression regulation, biosynthesis and metabolism; for group II, the enriched terms in RAC included 'chromatin remodeling/modification/organization' and 'respiratory chain complex', and in NAC included 'enzyme activity' and 'response to auxin', etc.

Asymmetric epigenetic modification analysis

DNA methylation levels (CG, CHG and CHH) in promoters and gene bodies of 29 976 gene pairs in A_C, RAC and NAC (Fig. S2) were assessed using WGBS (Table S2). The promoter region was defined as the sequence from 2 kb upstream to 500 bp downstream of the transcription start site of the genes (Notes S1), and removal of the length possibly involved in the gene body did not affect its relative DNA methylation level (Table S3). We performed differential methylation level (DML) analysis of all gene pairs (Student's *t*-test, $P < 0.001$), and found that the number of gene pairs with DMLs in the promoter was higher than the number with DMLs in the gene body; in addition, we found that the number of gene pairs with C subgenome biased methylation levels in the promoter was significantly higher than the number with A subgenome biased methylation levels (Figs 2a, S3). To further explore the evolutionary changes of gene pairs with DMLs, cross-analysis was performed (Fig. 2b). The gene pairs with DMLs in all three genotypes were divided into seven groups, and the counts of BEGs and nBEGs in each group were used to create a heatmap (Fig. 2c). Group 1 always had the highest number of gene pairs, indicating that many gene pairs with DMLs in parents/progenitors were eliminated in resynthesized/natural *B. napus* either in the promoter or gene body after allopolyploidization. The numbers for groups 2 and 3 were higher than those for groups 4–7, indicating that many new gene pairs with DMLs were generated *de novo* in RAC/NAC. The number of gene pairs with DMLs in group 5 was always higher than that in group 6, suggesting that in natural *B. napus*, the number of gene pairs with DMLs restored from progenitors was greater than that inherited from the resynthesized *B. napus*. The restoration phenomenon was most obvious in the CG context of promoters (45.4%) and least obvious in the CHG context of gene bodies (12.6%). For gene pairs with inherited DMLs from parents/progenitors, it was highly conserved (Figs S4, 2e).

Two active histone modification markers (H3K4me3 and H3K27ac) and one repressive marker (H3K27me3) were detected by ChIP-seq (Table S4). Analysis of the difference in histone modification between homoeologs revealed that the number of gene pairs with differential histone modifications (DHMs) with a C-homoeolog bias was always more than the number with an A-homoeolog bias, except for the differential H3K4me3 modification in NAC (Fig. 2d; Table S5), indicating that the number of C-homoeolog with these three histone modifications was slightly higher than that of A-homoeolog in *B. napus*. Similar cross-analysis was performed for gene pairs with DHMs (Fig. S5). Correlation analysis of shared gene pairs with DHMs showed that they were highly positively correlated (Figs S6, 2e).

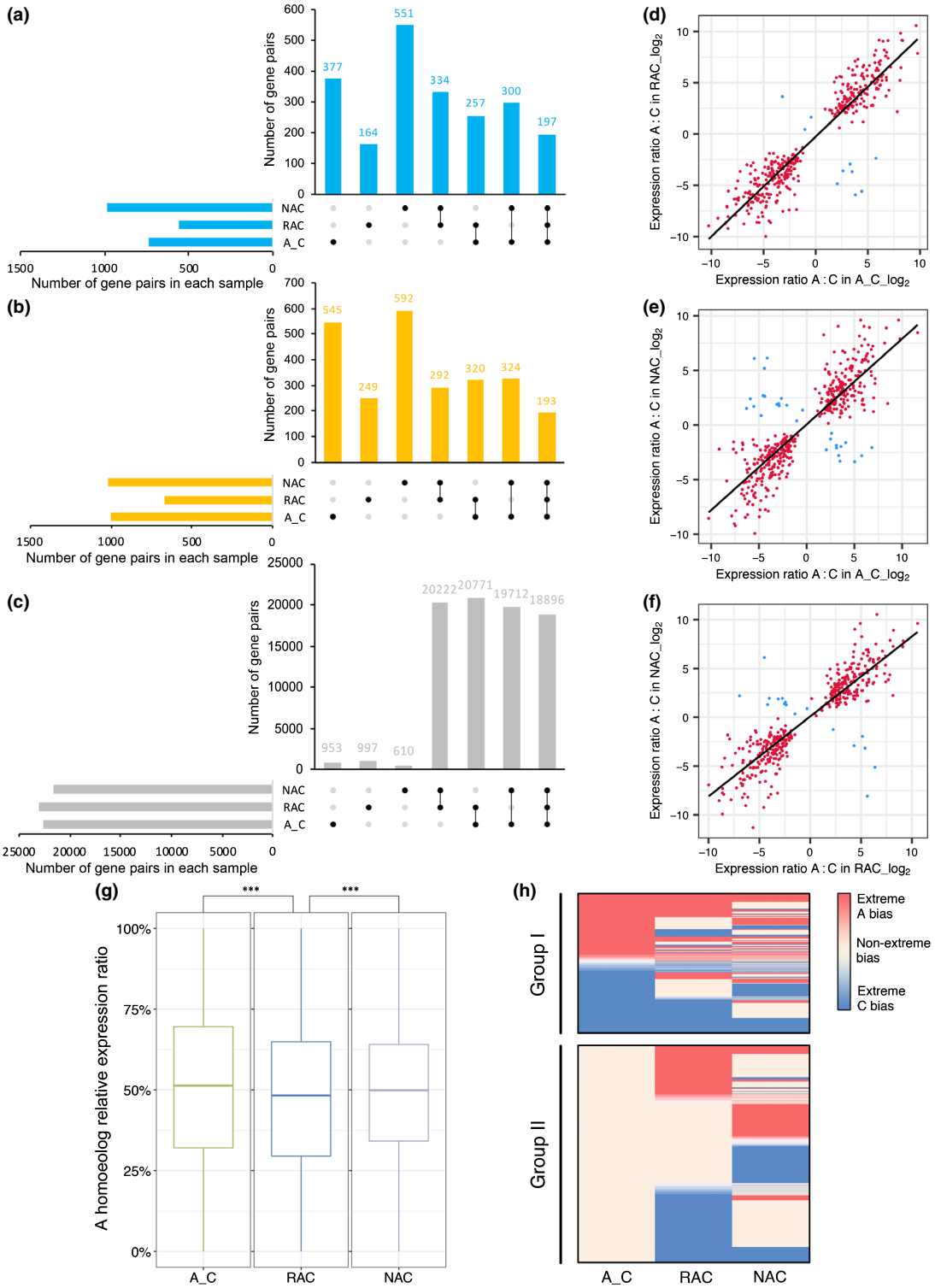

Fig. 1 Homoeolog expression bias of homoeologous gene pairs. (a) The number of gene pairs showing A homoeolog biased expression. (b) The number of gene pairs showing C homoeolog biased expression. (c) The number of gene pairs showing nonbiased expression. In (a–c), black dot(s) at the bottom of each vertical bar indicate that corresponding gene pairs were identified in a given genotype(s), the lines connecting genotypes indicate that they share these gene pairs, and the total number of biased/unbiased gene pairs in a given genotype is shown by the horizontal bars on the left. (d) Correlation between the common biased expression gene pairs (BEGs) in A_C and RAC (Pearson's $r = 0.92$, $P < 2.2 \times 10^{-16}$). (e) Correlation between the common BEGs in A_C and NAC (Pearson's $r = 0.84$, $P < 2.2 \times 10^{-16}$). (f) Correlation between the common BEGs in RAC and NAC (Pearson's $r = 0.88$, $P < 2.2 \times 10^{-16}$). In (d–f), red plots represented the positively correlated BEGs, and blue plots represented the negatively correlated BEGs. (g) The boxplots of the relative A-homoeolog expression ratio of gene pairs in three genotypes (K-S test; ***, $P < 2.2 \times 10^{-16}$). (h) Two groups of extremely biased expression gene pairs (defined as A% > 95% or < 5%). Group I represents the gene pairs with the inherited extremely biased expression from the parents/progenitors; group II represents the gene pairs with novel extremely biased expression. A_C, parental mix; NAC, natural *Brassica napus*; RAC, resynthesized *B. napus*.

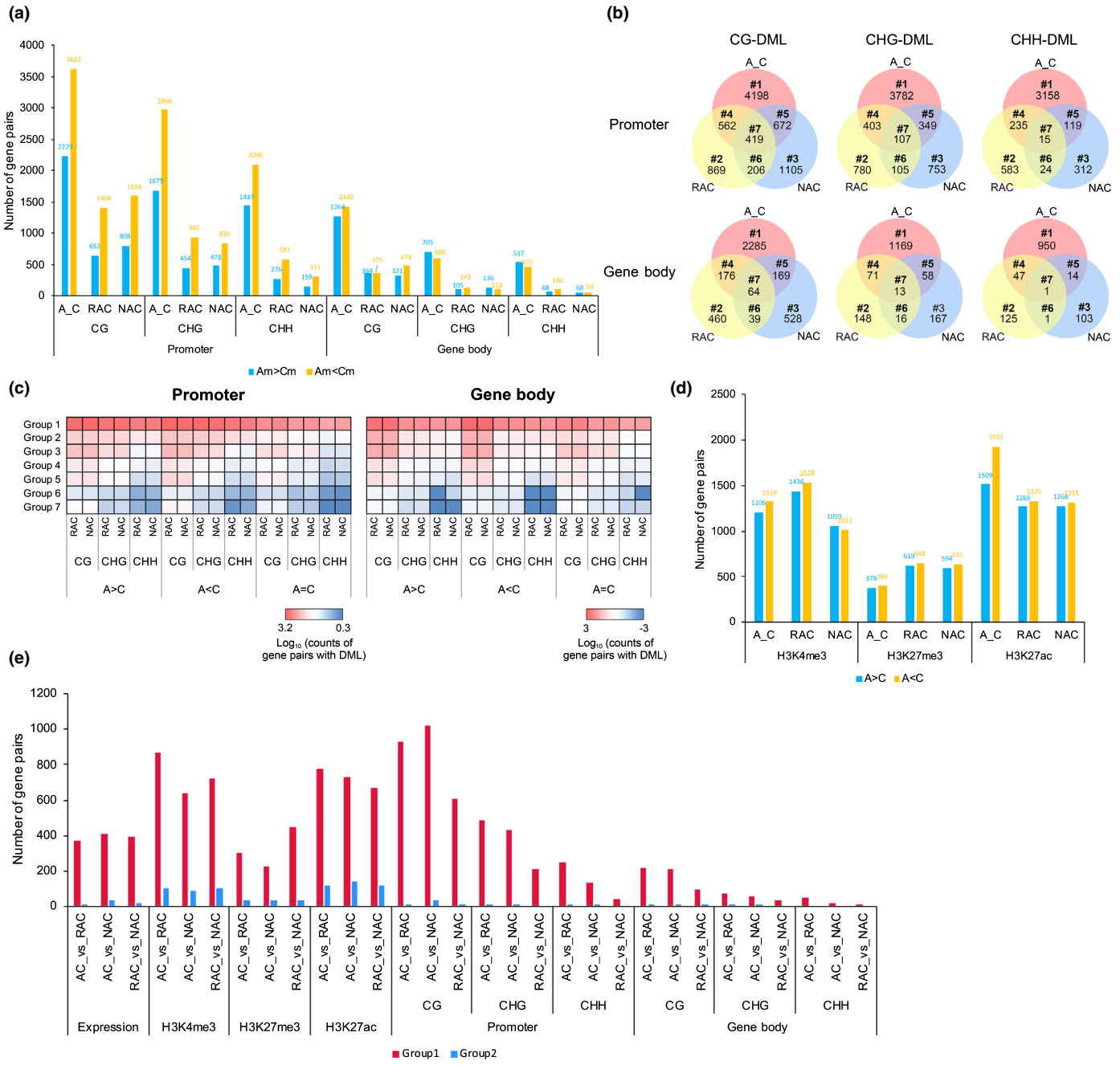

Fig. 2 Biased epigenetic modifications of homoeologous gene pairs. (a) The number of gene pairs with differential methylation levels (DML) in three sequence contexts (CG, CHG and CHH) in the promoter and gene body. Blue and yellow bars represent the numbers of gene pairs with A-homoeolog and C-homoeolog biased DMLs, respectively. (b) Comparison of gene pairs with DMLs in all three genotypes. (c) Counts of gene pairs with DMLs for the seven groups shown in the Venn diagrams in (b). A > C, C > A and A = C represent the statistical expression relationships of the A- and C-homoeologs, respectively. (d) Numbers of gene pairs with differential histone modifications (DHMs). Blue and yellow bars represent the numbers of gene pairs with A- and C-homoeolog biased DHMs, respectively. (e) Numbers of red (group 1) and blue/grey plots (group 2) in Figs 1d–f, S4, S6. A_C, parental mix; NAC, natural *Brassica napus*; RAC, resynthesized *B. napus*.

Generational transmission analysis of expression and epigenetic modification patterns

To explore the generational transmission characteristics of gene expression/epigenetic modification patterns, we divided the gene pairs into three groups for RAC/NAC, including nine patterns (Fig. 3a,b). Group A (pattern I–III) consisted of gene pairs for which the expression/histone modification patterns of orthologous genes in parents/progenitors were inherited (i.e. parental legacy) in RAC/NAC; group B (pattern IV–V) consisted of gene pairs for which initial expression/histone modification patterns were eliminated; group C (pattern VI–IX) consisted of gene pairs with biased expression/histone modification patterns which were novel. We found that the proportion of gene pairs belonging to

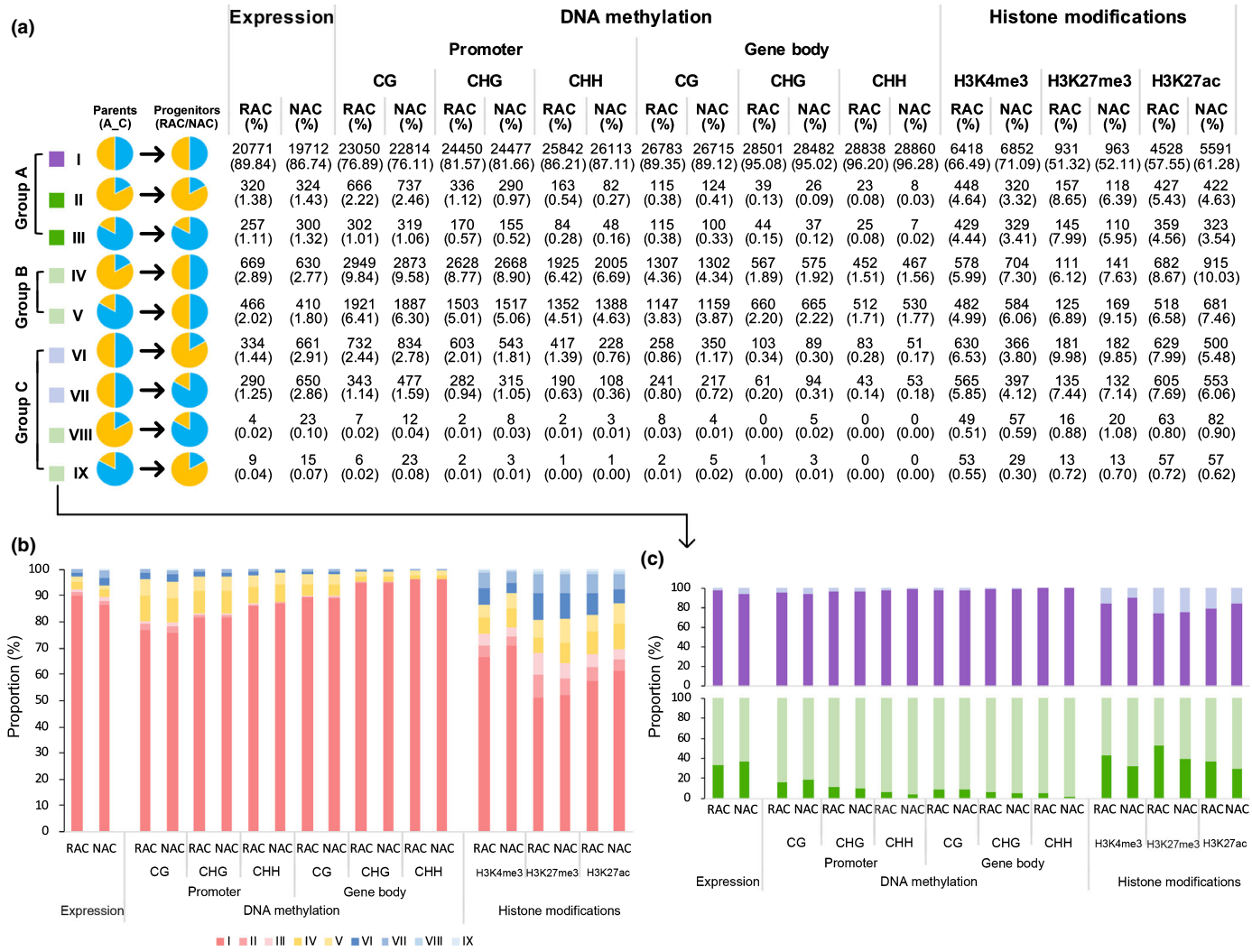

Fig. 3 Generational transmission of expression and epigenetic modification patterns of homoeologous gene pairs. (a) Numbers of homoeologous gene pairs exhibiting different expression and epigenetic modification patterns in RAC/NAC. Blue, A ortholog/homoeolog; yellow, C ortholog/homoeolog. (b) The proportion of gene pairs exhibiting each of nine patterns for all scenarios. (c) The proportions of gene pairs maintaining/changing parental states in RAC/NAC. The relative expression/epigenetic modification states between A ortholog and C ortholog in parents/progenitors were statistically the same (purple), and these states were maintained in the progenies (dark purple) or changed in the progenies (light purple). There were differences in relative expression/epigenetic modification states between A ortholog and C ortholog in parents/progenitors (green), and these states were maintained in the progenies (dark green) or changed in the progenies (light green). NAC, natural *Brassica napus*; RAC, resynthesized *B. napus*.

group A was always the highest, and the proportion of gene pairs represented by patterns VIII and IX was always the lowest (Fig. 3a,b), indicating that parental legacy was overwhelmingly important in the transmission of gene expression and epigenetic modification patterns. The parental legacy of DNA methylation of gene pairs in the gene body (89.9–96.4%) was stronger than that in the promoter (79.6–87.6%). The GO enrichment analysis showed that some genes were enriched in photosynthesis related GO terms in group B, while mitochondria and transmembrane protein activity related GO terms were enriched in group C (FDR < 0.05; Dataset S3).

To explore the transmission characteristics in more detail, we further divided the nine patterns into two categories according to whether there was any difference in relative expression/epigenetic modification state between ortholog A and ortholog C in parents/progenitors (indicated by purple and green squares, respectively, in Fig. 3a). If the state in parents/progenitors was maintained in progenies, it was marked with a dark color in Fig. 3(a); otherwise, it was marked with a light color. We found that when there was no statistical difference in expression/epigenetic modification states of two orthologs in the parents/progenitors, homoeologous gene pairs in progenies tended to maintain this relative state (i.e. parental legacy), and this finding was most pronounced for DNA methylation (93.8–97.1% for gene expression, 94.6–99.6% for DNA methylation and 74.7–90.0% for histone modifications; Fig. 3c). However, when there was a statistical difference, this relative state tended to change, and this finding was most pronounced for DNA methylation (63.3–67.6% for gene expression, 82.0–98.5% for DNA methylation and 46.7%–60.1% for histone modifications; Fig. 3c). There was a

New Phytologist

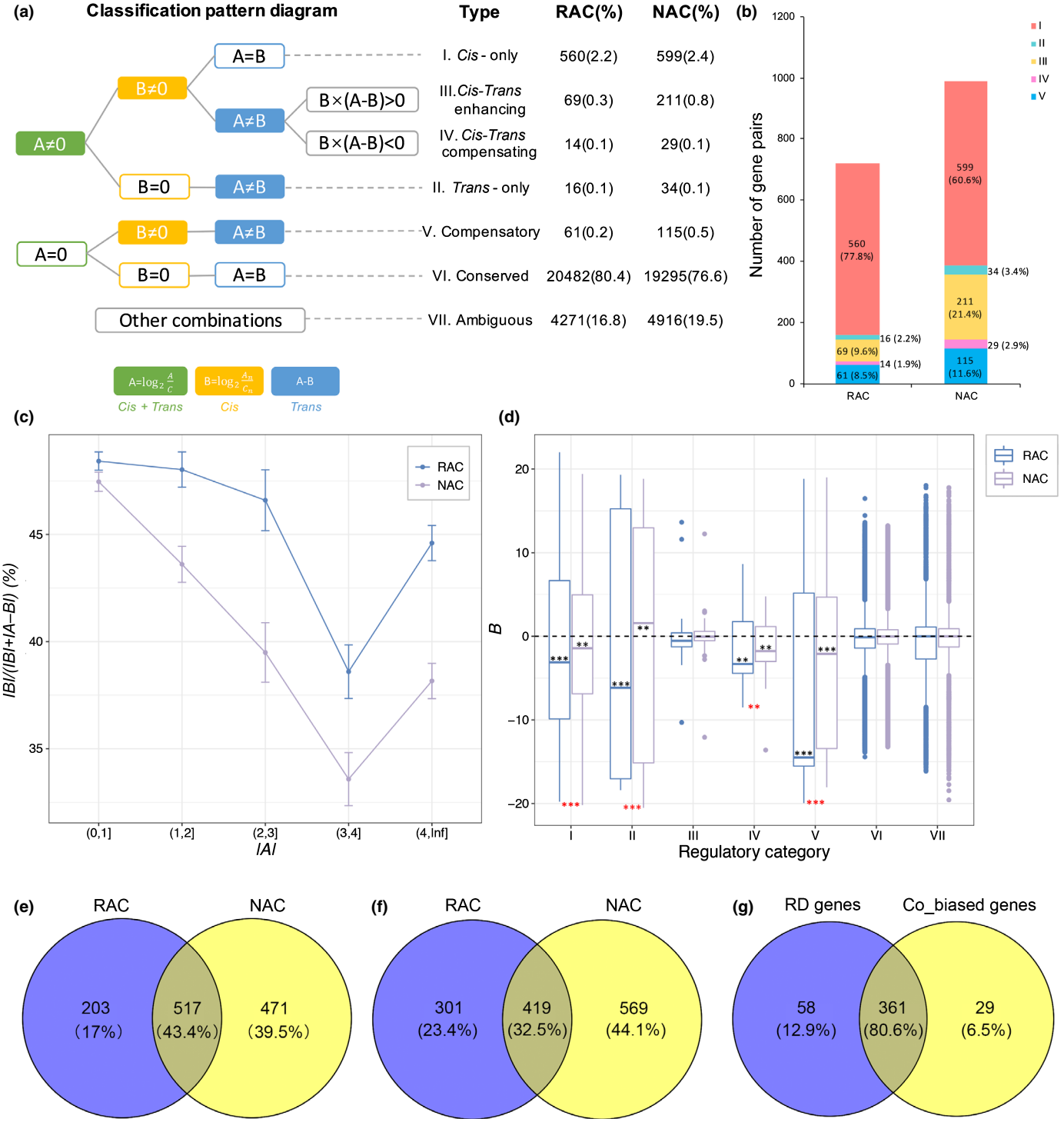

Fig. 4 *Cis*/*trans* regulation of homoeologous gene expression. (a) Gene pairs were assigned to one of seven regulatory (*cis* and/or *trans* effects) categories using the read counts of gene pairs in parents/progenitors (represented by A_C) and progeny (RAC/NAC). *A*, the expression divergence in parents/progenitors; *B*, the expression divergence in progeny (RAC/NAC). See 'Characterization of *cis*/*trans* regulation effects' sub-section in the Materials and Methods section for a detailed description. (b) The numbers and relative percentages for five divergent regulatory categories (I–V) in RAC/NAC. (c) The relative contributions of *cis* effects to the expression divergence of parents/progenitors. The *x*-axis represents the absolute values of the expression divergence of parents/progenitors (|*A*|); the *y*-axis represents the proportion of the total absolute expression divergence due to *cis* effects (|*B*|/(|*B*|+|*A*−*B*|)); error bar, 95% confidence intervals. (d) The expression divergence of progenies (*B*) of every regulatory category. *B* values above or below zero indicate an expression bias toward the A/C subgenome in progenies. Black asterisks indicate significant deviations from zero (Student's *t*-test; **, $P < 0.01$; ***, $P < 0.001$); red asterisks indicate significant differences for the medians of the corresponding categories in RAC and NAC (Student's *t*-test; **, $P < 0.01$; ***, $P < 0.001$). (e) Overlap of gene pairs with divergent regulation (categories I–V) between RAC and NAC. (f) Overlap of gene pairs with the same divergent regulatory category between RAC and NAC. (g) Overlap of 419 regulatory divergence (RD) genes from (f) and 390 biased expression gene pairs (BEGs) with the same biased expression in all genotypes. A_C, parental mix; NAC, natural *Brassica napus*; RAC, resynthesized *B. napus*.

significant difference between these two categories in the proportion of gene pairs maintaining the parental state in RAC/NAC during allopolyploidization and evolution (Chi-squared test, $P < 2.2 \times 10^{-16}$). These results suggested that whether the generational transmission of gene expression/epigenetic modifications was dominated by parental legacy depends on the original relative state of two orthologs in parents/progenitors.

Cis/*trans* regulation analysis

To explore the *cis*/*trans* regulation mechanism underlying gene expression, gene pairs were divided into seven categories (Fig. 4a). We found that the proportion of gene pairs exhibiting conserved regulation (category VI) was the highest RAC/NAC, suggesting that the expression of the majority of genes was regulated conservatively during evolution. The number of gene pairs belonging to category VI in NAC was significantly lower than in RAC (Chi-squared test, $P < 2.2 \times 10^{-16}$), indicating that the regulation mechanism of most gene pairs has changed from convergent regulation (category VI) to divergent regulation (categories I–V) after evolution. For divergent regulation (Fig. 4b), we found that category I accounted for the largest number of gene pairs, while the number belonging to category II was much lower, indicating that *cis* action tends to play a role independently while *trans* action tends to work with *cis*. When *cis* and *trans* effects work together, the number of gene pairs showing opposing directions (category IV and V) was lower than the number showing the same direction (category III) in RAC, and the opposite phenomenon was observed in NAC. The number of gene pairs belonging to categories II–V was significantly higher in NAC than RAC (Chi-squared test, $P < 0.05$), and category III showed the largest increase, indicating that the conserved regulation of gene pairs (category VI) in RAC was mainly converted to enhancing regulation (category III) when they converted to divergent regulation during evolution.

We investigated the relative contributions of *cis* and *trans* regulation of gene expression in *B. napus*, and found that the degree of *trans* regulation was significantly greater than that of *cis* regulation in RAC and NAC when all gene pairs were taken into account ($|trans| > |cis|$, Wilcoxon rank sum test, $P < 0.001$; Fig. 4c). We then asked whether the degree of expression divergence of gene pairs in parents/progenitors (i.e. *A* values; see 'Characterization of *cis*/*trans* regulation effects' sub-section in the Materials and Methods section) would influence the *cis* or *trans* regulation in *B. napus*. A significant positive correlation was found between *cis*/*trans* effects and *A* values in RAC/NAC, and the correlation between *trans* effects and *A* values (Pearson's $r = 0.53$ in RAC, $r = 0.61$ in NAC, $P < 2.2 \times 10^{-16}$) was higher than that between *cis* effects and *A* values ($r = 0.44$ in RAC, $r = 0.39$ in NAC, $P < 2.2 \times 10^{-16}$). Although the proportion of *cis* effects seems to decrease with the magnitude of parental/progenitors' expression divergence (represented by $|A|$), the correlation coefficients were very weak (Pearson's $r = -0.02$, $P < 0.001$ in RAC; $r = -0.10$, $P < 2.2 \times 10^{-16}$ in RAC; Fig. 4c).

To explore the regulatory categories of homoeolog expression bias in RAC/NAC, we grouped the *B* values into seven categories and generated a boxplot (Fig. 4d). We found that the main mechanism underlying homoeolog expression bias was divergent regulation – specifically, regulation of C-homoeolog expression bias occurred via category I, II, IV and V effects in RAC ($B < 0$, Student's *t*-test, $P < 0.01$; black asterisks in Fig. 4d), while in NAC it occurred via category I, IV and V effects ($B < 0$, Student's *t*-test, $P < 0.01$; black asterisks in Fig. 4d). The reduction of C-homoeolog expression bias was mainly regulated via category I, II, IV and V effects in NAC (the median in RAC < NAC, Student's *t*-test, $P < 0.01$; red asterisks in Fig. 4d). A total of 517 gene pairs were found to show regulatory divergence (RD) in RAC and NAC, among which 419 (81.0%) exhibited consistent regulatory categories (Fig. 4e–f), indicating that regulatory categories of 517 RD gene pairs were highly conserved during evolution. Of 419 RD gene pairs, 364 (86.9%) belong to category I (*cis* only; Fig. S7), indicating that this category was moderately conserved during evolution. We then performed the cross analysis of 419 RD genes and 390 BEGs with the same biased expression in all genotypes identified previously. We found that 361 gene pairs (> 80%) were shared between them (Fig. 4g), and BEGs were significantly prone to showing divergent regulation in RAC and NAC (Chi-squared test, $P < 0.001$, Fig. S8), indicating that gene pairs with co-biased expression in all three genotypes (A_C, RAC and NAC) were mainly regulated by conservative divergent regulation categories in *B. napus*. The GO enrichment analysis (Dataset S4) showed that the divergent regulatory genes in NAC were uniquely enriched in GO terms associated with photosystem/thylakoid, indicating that some genes related to photosynthesis were regulated divergently in natural *B. napus* after evolution.

Asymmetric epigenetic modifications contribute to homoeolog expression bias

We further explored whether homoeolog expression bias was correlated with asymmetric epigenetic modifications. The first question was whether the BEGs were significantly enriched in gene pairs with DMLs/DHMs compared with those without DMLs/DHMs. As expected, more biased expression was detected in gene pairs with DMLs/DHMs than those without DMLs/DHMs in most cases (Fig. 5a,b). Notably, two active histone modifications were highly significantly correlated with the biased expression of gene pairs across all genotypes. The next question was whether the biased expression of gene pairs was positively or negatively correlated with different epigenetic modifications. A positive correlation was defined as gene expression and epigenetic modifications of A-/C-homoeolog biased gene pairs both being larger/smaller than that of C-/A-homoeolog biased gene pairs, and a negative correlation as the opposite (Fig. 5c). We observed only the cases marked with asterisk(s) in Fig. 5b and found that the number of negative correlations was always greater than the number of positive ones for both DNA methylation and the repressive H3K27me3 marker (with one exception), whereas the opposite was observed for both of the two active histone modifications. The results reflected in these data – that DNA methylation and H3K27me3 markers may inhibit gene expression, while H3K4me3 and H3K27ac markers may activate gene expression –

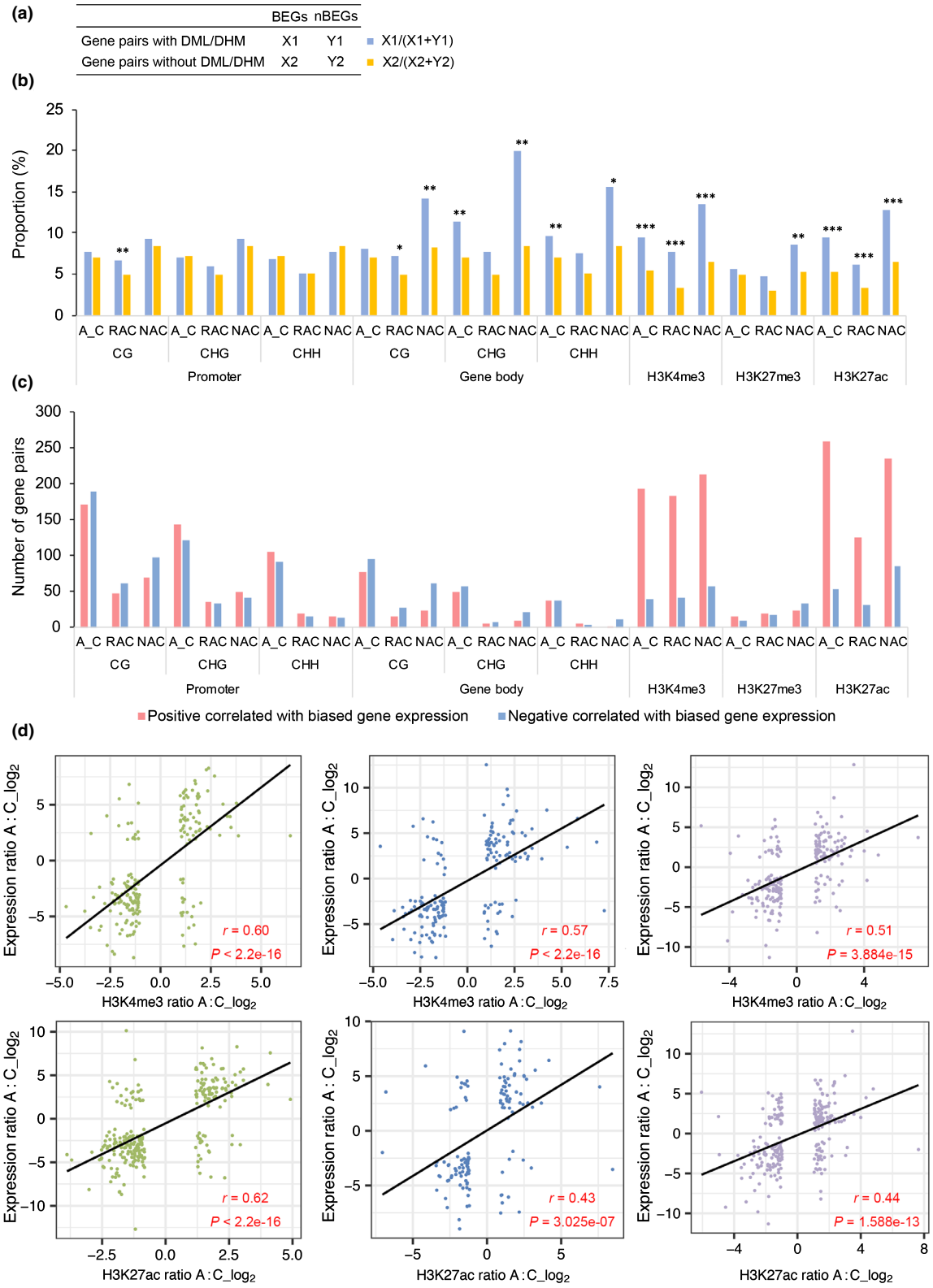

Fig. 5 Correlation analysis for four asymmetric epigenetic modifications and biased expression of homoeologous gene pairs. (a) Cross-table of biased expression gene pairs (BEGs)/nonbiased expression gene pairs (nBEGs) and gene pairs with/without differential methylation level (DMLs)/differential histone modifications (DHMs). (b) Comparison between the percentage of BEGs in gene pairs with DMLs/DHMs and those without DMLs/DHMs. Chi-squared test (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$). Blue bars, X1/(X1+Y1) in (a); yellow bars, X2/(X2+Y2) in (a). (c) Numbers of positive/negative correlation events for all epigenetic modifications. (d) Quantitative correlations between two active asymmetric histone modifications and biased expression in A_C (green), RAC (blue) and NAC (purple). Pearson's *r* and *P*-values are shown in the figure. A_C, parental mix; NAC, natural *Brassica napus*; RAC, resynthesized *B. napus*.

are consistent with the findings of most studies to date. Furthermore, the correlation between epigenetic modifications and homoeolog expression bias was quantitatively analyzed, and we found that H3K4me3 and H3K27ac significantly positively regulated the biased expression of genes in all genotypes (Pearson's $r = 0.51–0.60$, $P < 0.001$ in H3K4me3; $r = 0.43–0.62$, $P < 0.001$ in H3K27ac; Fig. 5d).

To ascertain whether there was a correlation between the occurrence of multiple (concurrent) epigenetic modifications and the biased expression of gene pairs, a hierarchical analysis was performed (Fig. 6a). The gene pairs were clustered according to whether biased expression or asymmetric epigenetic modifications were detected. We found that a total of 517, 340 and 518 gene pairs in A_C, RAC and NAC were involved, among which the proportion of epigenetic modification concurrence (clusters 1–7, 9–11 and 13) was 26.5%, 27.5% and 29.3%, respectively. The concurrence ratio of H3K4me3 and H3K27ac histone modification was the highest (Figs 6a, S9), indicating that the probability of their concurrence in BEGs was greater than that of the other epigenetic modifications. We then carried out quantitative correlation analysis for the representative gene pairs (clusters 8, 11, 12, 14, 15 in Fig. 6a), and generated plot diagrams (Figs 6a, S10, S11). We found that two active histone modifications significantly positively correlated with the biased expression of gene pairs ($r = 0.77$ in A_C, $r = 0.54$ in RAC, $r = 0.47$ in NAC, $P < 0.001$; Fig. 6b), when they were jointly detected in BEGs (clusters 12 and 15 in Fig. 6a). Moreover, He *et al.* (2010) have shown that epigenetic natural variation may influence gene expression variation via *cis*/*trans* effects, so we wanted to identify what was the main regulatory categories of homoeolog expression bias that was correlated with asymmetric epigenetic modifications in our study. We investigated how many gene pairs belonging to each of the seven regulatory categories (Fig. 4a) were present in the representative clusters (clusters 8, 11, 12, 14, 15 in Fig. 6a,c), and found that gene pairs belonging to categories III and VI were not detected, and that category I (*cis*-only) represented the largest numbers of gene pairs in all clusters, if category VII (ambiguous) was not considered, indicating that the *cis* effect is dominant when epigenetic modifications potentially affect the biased expression of gene pairs in *B. napus*.

We then wanted to know what epigenetic modifications primarily occurred in the 361 previously identified RD gene pairs with the same biased expression direction in all three genotypes. We found that only 22.7–33.2% of the 361 gene pairs exhibited differential epigenetic modifications (Fig. 6d), the majority of which were accounted for by H3K4me3 and/or H3K27ac (76.7% to 86.3%, respectively; Fig. 6e). Further analysis revealed that only two categories of regulation were detected, I (*cis*-only) and IV (compensatory), with category I dominating as expected (Fig. 6f). These results showed that two active histone modifications might affect the biased expression of these gene pairs, mainly via *cis* effects.

Discussion

Homoeolog expression bias in allopolyploids has been widely studied, but the underlying epigenetic basis is not well understood. Here, we comprehensively analyzed biased gene expression and four asymmetric epigenetic modifications of homoeologous gene pairs in resynthesized, natural *B. napus*, and an *in silico* 'hybrid'. One question we focused on was how the expression/epigenetic modifications of homoeologous gene pairs were reshaped after allopolyploidization in *B. napus*. We found that parental legacy was the dominant principle in these remodeling processes (Fig. 3a,b), and a similar phenomenon has also been found in other studies (Wang *et al.*, 2016; Bird *et al.*, 2021). Our results further support the idea that subgenome dominance in allopolyploids is primarily inherited from their progenitors, rather than being the outcome of allopolyploidization (Buggs *et al.*, 2014; Bird *et al.*, 2021). The highlight of this study is the demonstration of the possible specific genetic rule behind the parental legacy, that is, whether the parental legacy is dominant depends on whether there are differences in the expression/epigenetic modifications of homoeologous gene pairs in the parents/progenitors. Specifically, when there is no difference between them, parental legacy is dominant; when they differ, they are more likely to be altered in progenies after allopolyploidization (Fig. 3c). Overall, the number of homologous gene pairs with asymmetric expression/epigenetic modifications was significantly lower than the number of those without such modifications in both parents and progenies, leading to the dominance of parental legacy in the remodeling processes. In addition, although the subgenome DNA methylation dominance was inherited mainly from two parents/progenitors (i.e. parental legacy), we found that DNA methylation differences between subgenomes were modified repeatedly during early allopolyploidization and/or the subsequent evolution of *B. napus*. Specifically, we found that the number of gene pairs with asymmetric DNA methylation decreased significantly in *B. napus* after allopolyploidization, which was partially reversible in natural *B. napus* (Fig. 2b). This result indicates that the allopolyploidization might attenuate the differences in DNA methylation between two subgenomes in *B. napus*. Indeed, many studies have found that DNA methylation patterns change repeatedly after hybridization, genome doubling and other evolutionary processes in allopolyploids. For example, a recent study on allohexaploid wheat showed that genome merger and separation lead to dynamic and reversible changes in DNA methylation, which are related to changes in transposable element (TE) activity and gene expression (Yuan *et al.*, 2020). Another study on allotetraploid rice showed that genome merger resulting from hybridization attenuates the initial DNA methylation differences, and the WGD resulted in these differences being re-augmented (Li *et al.*, 2019). The DNA methylation patterns in different generations of resynthesized *B. napus* also changed to different degrees (Bird *et al.*, 2021).

The next question of interest is whether the A or C subgenome regions showed dominant expression in resynthesized and natural *B. napus*. We found that the C subgenome was dominant in the gene expression of resynthesized and natural *B. napus*, and the C genome was the maternal genome in this study. Dominant expression of the maternal genome has also been observed in other resynthesized allopolyploids (Xu *et al.*, 2012; Qi *et al.*, 2018; Wu *et al.*, 2018; Ferreira de Carvalho *et al.*, 2019; Bird

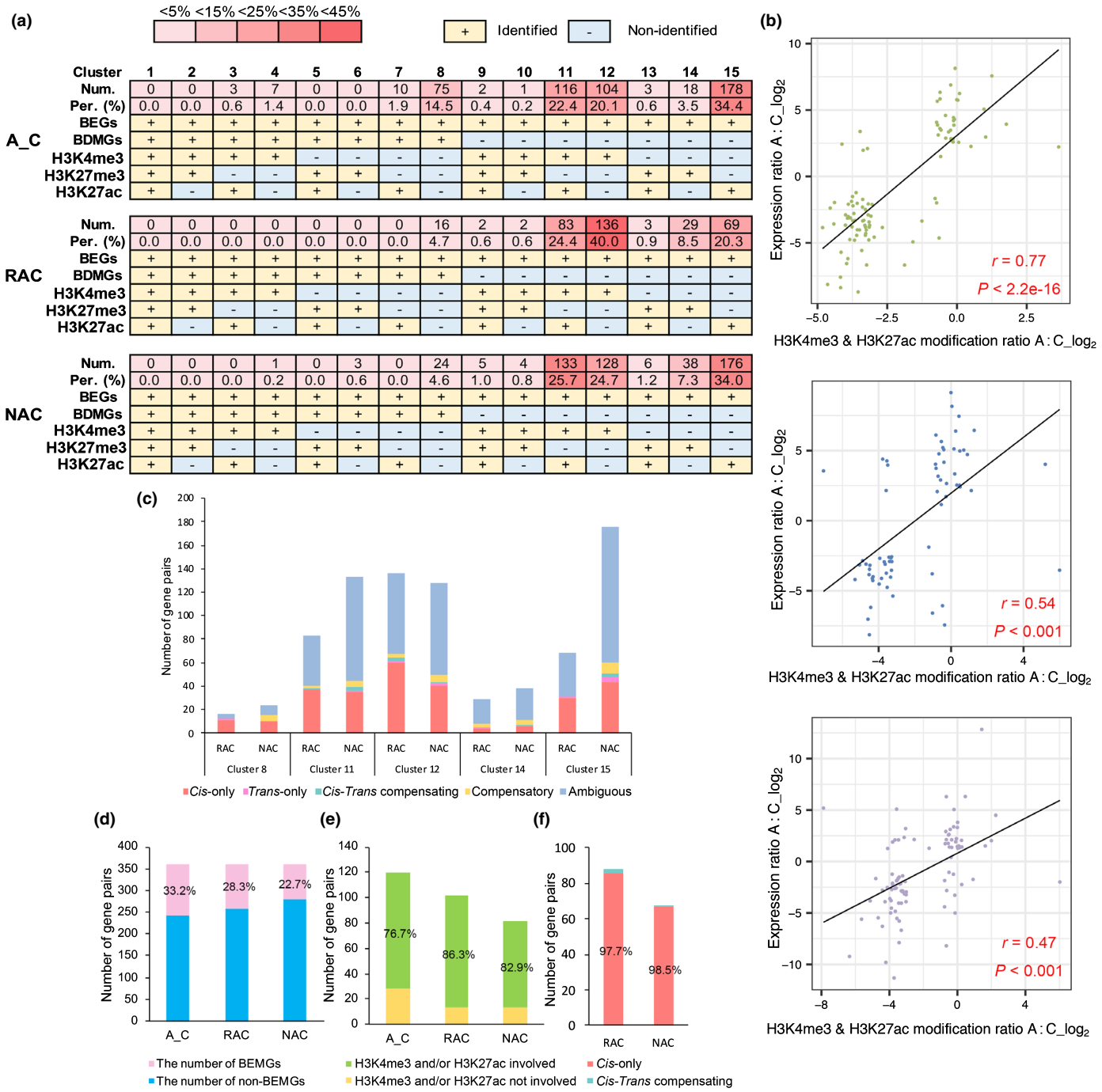

Fig. 6 The relationships between multiple asymmetric epigenetic modifications and the biased expression of homoeologous gene pairs. (a) Permutation table for all four epigenetic modifications and biased expression of gene pairs. BEGs, biased expression gene pairs; BDMGs, biased DNA methylation gene pairs; H3K27ac, biased H3K27ac modification gene pairs; H3K27me3, biased H3K27me3 modification gene pairs; H3K4me3, biased H3K4me3 modification gene pairs; Num., number of genes belonging to each cluster; Per., percentage of genes belonging to each cluster. (b) Correlations between two histone modifications (H3K4me3 and H3K27ac) and the expression of gene pairs in cluster 11 in (a). Green plot, the correlations in A_C; blue plot, the correlations in RAC; purple plot, the correlations in NAC. Pearson's *r* and *P*-values are shown in the figure. (c) Numbers of gene pairs belonging to regulatory categories of the representative clusters in (a). (d) Numbers of biased epigenetic modification gene pairs (BEMGs) and nonBEMGs in 361 regulatory divergence (RD) gene pairs with the same biased expression direction in all three genotypes. (e) The numbers of gene pairs for which two active histone modifications (H3K4me3 and/or H3K27ac) were/were not observed. (f) The numbers of gene pairs belonging to two regulatory categories for the two active histone modifications in (e). A_C, parental mix; NAC, natural *Brassica napus*; RAC, resynthesized *B. napus*.

et al., 2021). Sharbrough *et al.* (2017) proposed that the typical maternal inheritance of cytoplasmic genomes may lead to the dominant expression of these organelle-targeted maternal homoeologous genes. Moreover, Bird *et al.* (2021) found that C-BEGs in resynthesized *B. napus* were highly enriched in organelle functions. Similarly, in our study, the organelle components of

C-BEGs were more highly enriched than those of A-BEGs in resynthesized *B. napus* (Dataset S1). These results provide further evidence that cyto-nuclear interactions may be one of the driving forces behind the subgenome expression dominance in allopolyploids (Sharbrough *et al.*, 2017; Bird *et al.*, 2021). Another important finding of this study is that C subgenome expression dominance is significant in resynthesized *B. napus* but not in natural *B. napus*, and the total number of BEGs in natural *B. napus* was greater than that in resynthesized *B. napus* (Table 1), suggesting that the maternal subgenome expression dominance phenomenon reduced and the total number of BEGs increased in *B. napus* during the subsequent evolutionary processes. At present, the main reasons attributed to subgenome dominance in allopolyploids include the influence of genomic features of progenitors and the 'genomic shock' caused by allopolyploidization, such as TE density and homologous exchanges (HEs) of subgenomes (Sharbrough *et al.*, 2017; Bird *et al.*, 2018; Bird *et al.*, 2021). Homologous exchanges continue after the formation of allopolyploids (Xiong *et al.*, 2011; Chester *et al.*, 2012), and many regions of the C subgenome were replaced by A subgenome regions through HEs in *B. napus* (Chalhoub *et al.*, 2014), which may explain why the C subgenome expression dominance in natural *B. napus* weakened after evolutionary processes in this study.

We also examined epigenetic subgenome dominance and its relationship to subgenome expression dominance in *B. napus*. We found that the C subgenome was significantly DNA methylation dominant in both resynthesized and natural *B. napus* (Figs 2a, Fig. S3), which was consistent with the findings of previous studies (Chalhoub *et al.*, 2014; Bird *et al.*, 2021; Zhang *et al.*, 2021). Chalhoub *et al.* (2014) proposed that the high degree of DNA methylation in the C subgenome might be associated with its high TE density. Moreover, for asymmetric histone modification, the A/C subgenome bias was balanced (i.e. there was no statistically significant difference between the numbers of A-homoeolog biased and C-homoeolog biased gene pairs) in resynthesized and natural *B. napus* (though a slight C-over-A subgenome dominance was observed, except for H3K4me3 in NAC; Table S5). This result differed from those reported in a study by Zhang *et al.* (2021), in which the A subgenome showed histone modification dominance in *B. napus*. There are two possible explanations. First, genetic differences in the plant materials used may be the main reason. Natural and resynthesized *B. napus* were used in our study, while sterile and restorer lines of *B. napus* were used in the study of Zhang *et al.* (2021). The parent/progenitor genome may have a strong influence on the subgenome distribution of epigenetic modifications in allopolyploids, for various reasons including TE density, and TE content has been shown to be highly variable within a single species (Golicz *et al.*, 2016; Anderson *et al.*, 2019). Second, different methods of analysis may also have an impact on results. Zhang *et al.* (2021) examined the subgenome coverage of histone modifications, while we examined histone modifications of homoeologous gene pairs between two subgenomes. We then explored the relationship between epigenetic modifications and subgenome expression dominance, and found that two active markers (H3K4me3 and H3K27ac) frequently appeared together in BEGs, and that they were significantly positively correlated with biased gene expression (Figs 5d, 6b, S10). Chromatin regions containing the H3K4me3 marker are targeted by histone acetyltransferases for histone acetylation (Wang *et al.*, 2009), and these markers might lead to genes being actively expressed. A study on rice also showed that clusters of enhancers enriched by H3K4me3 were simultaneously enriched by different levels of H3K27ac (Sun *et al.*, 2019). In our study, no association was found between H3K4me3 and H3K27me3, which is widely found in animals and has also been identified in rice (He *et al.*, 2010), but not in Arabidopsis (Ha *et al.*, 2011), which may indicate that animals and plants as well as monocotyledons and dicotyledons adopt different strategies for cell differentiation and organ development. For DNA methylation and H3K27me3 markers, we found that the number of negative correlations was always greater than the number of positive ones (with one exception; Fig. 5c) in the BEGs significantly enriched cases shown in Fig. 5(b). Quantitative correlation analysis showed that DNA methylation in the CG context was significantly negatively correlated with biased expression of gene pairs in A_C and NAC when DMLs were detected in BEGs alone (Fig. S10). There was no significant correlation between H3K27me3 and biased expression of gene pairs in RAC and NAC, which was consistent with previous findings in rice (Lv *et al.*, 2019). A previous study proposed that epigenetic variation might influence gene expression variation through *cis*/*trans* effects (He *et al.*, 2010), but no further relevant research has been done. Our integrated analysis revealed that the *cis* effect was dominant when these four epigenetic modifications potentially affect the biased expression of gene pairs in *B. napus*.

A systematic *cis*/*trans* effect analysis has not yet been reported for *B. napus*, but it is one of tools that we can use to understand its phenotypic variation after allopolyploidization and evolution. Our study showed that *cis* and *trans* effects account for 33.6–48.4% and 51.6–66.4% of the total regulatory effects in *B. napus*, respectively (Fig. 4c), and these figures are similar to results reported for maize and cotton (Lemmon *et al.*, 2014; Bao *et al.*, 2019). Our results suggest that these two regulatory effects are almost equally important in the evolution of allopolyploids. With the process of allopolyploidization, a large number of highly similar or different *trans* factors interact with homoeologous *cis* elements, creating multiple evolutionary possibilities for allopolyploids (Bao *et al.*, 2019). A recent study in cotton suggested that the phenomenon of enhanced *trans* effect evolution might be a general feature of polyploids (Bao *et al.*, 2019), and our study supports this conclusion, which helps to explain the evolution of allopolyploids. Another interesting finding in our study is that the main mechanism underlying the homoeolog expression bias in *B. napus* was divergent regulation (Fig. 4d), suggesting that regulatory effects of biased gene expression were more likely to change in allopolyploid *B. napus* during its early establishment and subsequent evolution. A similar phenomenon has been found in cotton (Bao *et al.*, 2019), suggesting that this may be a common feature of allopolyploids, and further study in other allopolyploids is needed.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (31970241).

Author contributions

JW and ML conceived and designed the study. ML performed the bioinformatics analyses. ML wrote the manuscript. XW provided some critical suggestions for this study, and the experimental materials. JW and XW revised the manuscript. ML, WS and FW were responsible for planting materials. All authors read and approved the final manuscript.

ORCID

Mengdi Li https://orcid.org/0000-0001-5526-9670
Jianbo Wang https://orcid.org/0000-0002-3040-5522

Data availability

The data that support the findings of this study are openly available in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database with accession no. SRR13302173-SRR13302184 (RNA-seq), SRR13306925-SRR13306936 (WGBS) and SRR13318007-SRR13318030 (ChIP-seq).

References

Abbott R, Albach D, Ansell S, Arntzen Jw, Baird S, Bierne N, Boughman J, Brelsford A, Buerkle Ca, Buggs R *et al.* 2013. Hybridization and speciation. *Journal of Evolutionary Biology* **26**: 229–246.

Adams KL, Wendel JF. 2005. Polyploidy and genome evolution in plants. *Current Opinion in Plant Biology* **8**: 135–141.

An H, Qi X, Gaynor ML, Hao Y, Gebken SC, Mabry ME, McAlvay AC, Teakle GR, Conant GC, Barker MS *et al.* 2019. Transcriptome and organellar sequencing highlights the complex origin and diversification of allotetraploid *Brassica napus*. *Nature Communications* **10**: 2878.

Anderson SN, Stitzer MC, Brohammer AB, Zhou P, Noshay JM, O'Connor CH, Hirsch CD, Ross-Ibarra J, Hirsch CN, Springer NM. 2019. Transposable elements contribute to dynamic genome content in maize. *The Plant Journal* **100**: 1052–1065.

Baduel P, Bray S, Vallejo-Marin M, Kolar F, Yant L. 2018. The "polyploid hop": shifting challenges and opportunities over the evolutionary lifespan of genome duplications. *Frontiers in Ecology and Evolution* **6**: 117.

Bao Y, Hu G, Grover CE, Conover J, Yuan D, Wendel JF. 2019. Unraveling *cis* and *trans* regulatory evolution during cotton domestication. *Nature Communications* **10**: 5399.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate – a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**: 289–300.

Bird KA, Niederhuth CE, Ou S, Gehan M, Pires JC, Xiong Z, VanBuren R, Edger PP. 2021. Replaying the evolutionary tape to investigate subgenome dominance in allopolyploid *Brassica napus*. *New Phytologist* **230**: 354–371.

Bird KA, VanBuren R, Puzey JR, Edger PP. 2018. The causes and consequences of subgenome dominance in hybrids and recent polyploids. *New Phytologist* **220**: 87–93.

Braszewska-Zalewska A, Bernas T, Maluszynska J. 2010. Epigenetic chromatin modifications in *Brassica* genomes. *Genome* **53**: 203–210.

Buggs RJ, Wendel JF, Doyle JJ, Soltis DE, Soltis PS, Coate JE. 2014. The legacy of diploid progenitors in allopolyploid gene expression patterns. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences* **369**: 20130354.

Chalhoub B, Denoeud F, Liu S, Parkin I, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B *et al.* 2014. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* **345**: 950–953.

Chen ZJ. 2007. Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annual Review of Plant Biology* **58**: 377–406.

Chester M, Gallagher JP, Symonds VV, Cruz da Silva AV, Mavrodiev EV, Leitch AR, Soltis PS, Soltis DE. 2012. Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon miscellus* (Asteraceae). *Proceedings of the National Academy of Sciences, USA* **109**: 1176–1181.

Ding M, Chen ZJ. 2018. Epigenetic perspectives on the evolution and domestication of polyploid plant and crops. *Current Opinion in Plant Biology* **42**: 37–48.

Dong X, Tian L, Gouil Q, Kariyawasam H, Su S, De Paoli-Iseppi R, Prawer Y, Clark MB, Breslin K, Iminitoff M *et al.* 2021. The long and the short of it: unlocking nanopore long-read RNA sequencing data with short-read differential expression analysis tools. *NAR Genomics and Bioinformatics* **3**: lqab028.

Feldman M, Levy AA. 2005. Allopolyploidy – a shaping force in the evolution of wheat genomes. *Cytogenetic and Genome Research* **109**: 250–258.

Ferreira de Carvalho J, Lucas J, Deniot G, Falentin C, Filangi O, Gilet M, Legeai F, Lode M, Morice J, Trotoux G *et al.* 2019. Cytonuclear interactions remain stable during allopolyploid evolution despite repeated whole-genome duplications in *Brassica*. *The Plant Journal* **98**: 434–447.

Flagel LE, Wendel JF. 2010. Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation. *New Phytologist* **186**: 184–193.

Golicz AA, Bayer PE, Barker GC, Edger PP, Kim H, Martinez PA, Chan CKK, Severn-Ellis A, McCombie WR, Parkin IAP *et al.* 2016. The pangenome of an agronomically important crop plant *Brassica oleracea*. *Nature Communications* **7**: 13390.

Ha M, Ng DW, Li WH, Chen ZJ. 2011. Coordinated histone modifications are associated with gene expression variation within and between species. *Genome Research* **21**: 590–598.

He G, Zhu X, Elling AA, Chen L, Wang X, Guo L, Liang M, He H, Zhang H, Chen F *et al.* 2010. Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids. *Plant Cell* **22**: 17–33.

Jackson S, Chen ZJ. 2010. Genomic and expression plasticity of polyploidy. *Current Opinion in Plant Biology* **13**: 153–159.

Lang Z, Wang Y, Tang K, Tang D, Datsenka T, Cheng J, Zhang Y, Handa AK, Zhu JK. 2017. Critical roles of DNA demethylation in the activation of ripening-induced genes and inhibition of ripening-repressed genes in tomato fruit. *Proceedings of the National Academy of Sciences, USA* **114**: E4511–E4519.

Leitch IJ, Bennett MD. 1997. Polyploidy in angiosperms. *Trends in Plant Science* **2**: 470–476.

Lemmon ZH, Bukowski R, Sun Q, Doebley JF. 2014. The role of *cis* regulatory evolution in maize domestication. *PLoS Genetics* **10**: e1004745.

Li B, Carey M, Workman JL. 2007. The role of chromatin during transcription. *Cell* **128**: 707–719.

Li LF, Liu B, Olsen KM, Wendel JF. 2015. A re-evaluation of the homoploid hybrid origin of *Aegilops tauschii*, the donor of the wheat D-subgenome. *New Phytologist* **208**: 4–8.

Li M, Wang R, Wu X, Wang J. 2020. Homoeolog expression bias and expression level dominance (ELD) in four tissues of natural allotetraploid *Brassica napus*. *BMC Genomics* **21**: 330.

Li N, Xu C, Zhang A, Lv R, Meng X, Lin X, Gong L, Wendel JF, Liu B. 2019. DNA methylation repatterning accompanying hybridization, whole genome doubling and homoeolog exchange in nascent segmental rice allotetraploids. *New Phytologist* **223**: 979–992.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**: 550.

Lu K, Wei L, Li X, Wang Y, Wu J, Liu M, Zhang C, Chen Z, Xiao Z, Jian H *et al.* 2019. Whole-genome resequencing reveals *Brassica napus* origin and genetic loci involved in its improvement. *Nature Communications* **10**: 1154.

Lv Z, Zhang W, Wu Y, Huang S, Zhou Y, Zhang A, Deng X, Xu C, Xu Z, Gong L *et al.* 2019. Extensive allele-level remodeling of histone methylation modification in reciprocal F_1 hybrids of rice subspecies. *The Plant Journal* **97**: 571–586.

Madlung A. 2013. Polyploidy and its effect on evolutionary success: old questions revisited with new tools. *Heredity* **110**: 99–104.

Marchet C, Lecompte L, Silva CD, Cruaud C, Aury JM, Nicolas J, Peterlongo P. 2019. *De novo* clustering of long reads by gene from transcriptomics data. *Nucleic Acids Research* **47**: e2.

Muyle A, Gaut BS. 2019. Loss of gene body methylation in *Eutrema salsugineum* is associated with reduced gene expression. *Molecular Biology and Evolution* **36**: 155–158.

Oshlack A, Wakefield MJ. 2009. Transcript length bias in RNA-seq data confounds systems biology. *Biology Direct* **4**: 14.

Qi X, Wang H, Song A, Jiang J, Chen S, Chen F. 2018. Genomic and transcriptomic alterations following intergeneric hybridization and polyploidization in the *Chrysanthemum nankingense* × *Tanacetum vulgare* hybrid and allopolyploid (Asteraceae). *Horticulture Research* **5**: 5.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.

Roulin A, Auer PL, Libault M, Schlueter J, Farmer A, May G, Stacey G, Doerge RW, Jackson SA. 2013. The fate of duplicated genes in a polyploid plant genome. *The Plant Journal* **73**: 143–153.

Scarrow M, Wang Y, Sun G. 2021. Molecular regulatory mechanisms underlying the adaptability of polyploid plants. *Biological Reviews* **96**: 394–407.

Segraves KA. 2017. The effects of genome duplications in a community context. *New Phytologist* **215**: 57–69.

Sharbrough J, Conover JL, Tate JA, Wendel JF, Sloan DB. 2017. Cytonuclear responses to genome doubling. *American Journal of Botany* **104**: 1277–1280.

Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, Depamphilis CW, Wall PK, Soltis PS. 2009. Polyploidy and angiosperm diversification. *American Journal of Botany* **96**: 336–348.

Sun J, He N, Niu L, Huang Y, Shen W, Zhang Y, Li L, Hou C. 2019. Global quantitative mapping of enhancers in rice by STARR-seq. *Genomics, Proteomics & Bioinformatics* **17**: 140–153.

Takuno S, Gaut BS. 2013. Gene body methylation is conserved between plant orthologs and is of evolutionary consequence. *Proceedings of the National Academy of Sciences, USA* **110**: 1797–1802.

Villanueva RAM, Chen ZJ. 2019. ggplot2: elegant graphics for data analysis. *Measurement: Interdisciplinary Research and Perspectives* **17**: 160–167.

Wang X, Zhang H, Li Y, Zhang Z, Li L, Liu B. 2016. Transcriptome asymmetry in synthetic and natural allotetraploid wheats, revealed by RNA-sequencing. *New Phytologist* **209**: 1264–1277.

Wang Z, Zang C, Cui K, Schones DE, Barski A, Peng W, Zhao K. 2009. Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell* **138**: 1019–1031.

Wu J, Lin L, Xu M, Chen P, Liu D, Sun Q, Ran L, Wang Y. 2018. Homoeolog expression bias and expression level dominance in resynthesized allopolyploid *Brassica napus*. *BMC Genomics* **19**: 586.

Xiong Z, Gaeta RT, Pires JC. 2011. Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. *Proceedings of the National Academy of Sciences, USA* **108**: 7908–7913.

Xu C, Bai Y, Lin X, Zhao N, Hu L, Gong Z, Wendel JF, Liu B. 2014. Genome-wide disruption of gene expression in allopolyploids but not hybrids of rice subspecies. *Molecular Biology and Evolution* **31**: 1066–1076.

Xu Y, Zhao Q, Mei S, Wang J. 2012. Genomic and transcriptomic alterations following hybridisation and genome doubling in trigenomic allohexaploid *Brassica carinata* × *Brassica rapa*. *Plant Biology* **14**: 734–744.

Yoo MJ, Wendel JF. 2014. Comparative evolutionary and developmental dynamics of the cotton (*Gossypium hirsutum*) fiber transcriptome. *PLoS Genetics* **10**: e1004073.

Yuan J, Jiao W, Liu Y, Ye W, Wang X, Liu B, Song Q, Chen ZJ. 2020. Dynamic and reversible DNA methylation changes induced by genome separation and merger of polyploid wheat. *BMC Biology* **18**: 171.

Zhang H, Lang Z, Zhu JK. 2018. Dynamics and function of DNA methylation in plants. *Nature Reviews Molecular Cell Biology* **19**: 489–506.

Zhang Q, Guan P, Zhao L, Ma M, Xie L, Li Y, Zheng R, Ouyang W, Wang S, Li H *et al.* 2021. Asymmetric epigenome maps of subgenomes reveal imbalanced transcription and distinct evolutionary trends in *Brassica napus*. *Molecular Plant* **14**: 604–619.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Dataset S1 GO enrichment of biased/nobiased expression gene pairs (different colored backgrounds represent different types, green for biological process, blue for cellular component, and yellow for molecular function).

Dataset S2 GO enrichment of extremely biased expression gene pairs (different colored backgrounds represent different types, green for biological process, blue for cellular component, and yellow for molecular function).

Dataset S3 GO enrichment of gene pairs in three expression inheritance pattern groups (different colored backgrounds represent different types, green for biological process, blue for cellular component, and yellow for molecular function).

Dataset S4 GO enrichment of gene pairs with conserved/divergent regulation (different colored backgrounds represent different types, green for biological process, blue for cellular component, and yellow for molecular function).

Fig. S1 Experimental materials for this study.

Fig. S2 The DNA methylation levels of gene pairs in promoters and gene bodies.

Fig. S3 The number of gene pairs with/without differential methylation levels (DMLs) in promoters and gene bodies.

Fig. S4 Correlations of gene pairs with DMLs in promoters/gene bodies for each genotype pairing.

Fig. S5 Cross-analysis of homoeologous gene pairs with differential histone modifications (DHMs) in three genotypes.

Fig. S6 Correlations of gene pairs with DHMs for each genotype pairing.

Fig. S7 The proportion of each divergent regulation categories in 419 regulatory divergence genes.

Fig. S8 Relationship between biased expression and divergent regulation in resynthesized and natural *Brassica napus* (RAC and NAC, respectively).

Fig. S9 The concurrence ratio of four epigenetic modifications in all genotypes.

Fig. S10 Correlations of DNA methylation and expression of gene pairs in cluster 8.

Fig. S11 Correlations of three histone modifications and expression of gene pairs in clusters 12, 14 and 15.

Notes S1 The promoter region definition for DNA methylation analysis.

Table S1 Summary of RNA-seq data for all leaf samples.

Table S2 Summary of WGBS data for all leaf samples.

Table S3 Correlation between DNA methylation levels in two kinds of promoter region (see Notes S1 for definitions).

Table S4 Summary of chromatin immunoprecipitation sequencing (ChIP-seq) data for all leaf samples.

Table S5 The number of gene pairs with differential histone modifications in three genotypes.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.