Chen et al. BMC Genomics (2021) 22:264
https://doi.org/10.1186/s12864-021-07566-6

BMC Genomics

RESEARCH ARTICLE Open Access

Imprints of independent allopolyploid formations on patterns of gene expression in two sibling yarrow species (*Achillea*, Asteraceae)

Duo Chen[1†], Peng-Cheng Yan[2†] and Yan-Ping Guo[1*]

Abstract

Background: Polyploid species often originate recurrently. While this is well known, there is little information on the extent to which distinct allotetraploid species formed from the same parent species differ in gene expression. The tetraploid yarrow species *Achillea alpina* and *A. wilsoniana* arose independently from allopolyploidization between diploid *A. acuminata* and *A. asiatic*a. The genetics and geography of these origins are clear from previous studies, providing a solid basis for comparing gene expression patterns of sibling allopolyploid species that arose independently.

Results: We conducted comparative RNA-sequencing analyses on the two *Achillea* tetraploid species and their diploid progenitors to evaluate: 1) species-specific gene expression and coexpression across the four species; 2) patterns of inheritance of parental gene expression; 3) parental contributions to gene expression in the allotetraploid species, and homeolog expression bias. Diploid *A. asiatica* showed a higher contribution than diploid *A. acuminata* to the transcriptomes of both tetraploids and also greater homeolog bias in these transcriptomes, possibly reflecting a maternal effect. Comparing expressed genes in the two allotetraploids, we found expression of ca. 30% genes were species-specific in each, which were most enriched for GO terms pertaining to "defense response". Despite species-specific and differentially expressed genes between the two allotetraploids, they display similar transcriptome changes in comparison to their diploid progenitors.

Conclusion: Two independently originated *Achillea* allotetraploid species exhibited difference in gene expression, some of which must be related to differential adaptation during their post-speciation evolution. On the other hand, they showed similar expression profiles when compared to their progenitors. This similarity might be expected when pairs of merged diploid genomes in tetraploids are similar, as is the case in these two particular allotetraploids.

Keywords: Allopolyploid speciation, RNA-sequencing, Inheritance of gene expression, Homeolog express bias, *Achillea*

* Correspondence: guoyanping@bnu.edu.cn
†Duo Chen and Peng-Cheng Yan contributed equally to this work.
[1]Key Laboratory of Biodiversity Science and Ecological Engineering of the Ministry of Education, and College of Life Sciences, Beijing Normal University, Beijing, China
Full list of author information is available at the end of the article

Background

Polyploidy is an important mechanism of plant speciation. In an allopolyploid species, the combined effects of two or more diverged subgenomes and their regulatory interactions can lead to a myriad of genetic and epigenetic modifications described as genomic and transcriptomic shock [1–6]. The resulting changes in gene expression may often generate phenotypic variation affecting individual fitness and evolution of allopolyploids [7–13]. Analyses of synthetic polyploid plants have demonstrated that genomic and transcriptomic shock usually occurs immediately after polyploidization [1, 14–17], though changes may also take place during later stages of the evolutionary history of a polyploid species [3, 6, 18, 19].

Polyploid species often consist of lineages that originated independently and recurrently from the same parental species [20, 21]. Such recurrent formation can result in karyotypic, genomic, transcriptomic and phenotypic variation across lineages as demonstrated in recently originated allotetraploid species of *Tragopogon* (Asteraceae) [22–27]. However, whereas different lineages of the same allopolyploid species have been studied in detail, divergent species derived by independent origins from the same parental species have been reported less frequently and studied less [28–30]. Only in the orchid genus, *Dactylorhiza*, has research been conducted on gene expression and epigenetic differences among sibling allotetraploids derived from the same parental species pair. This showed that both kinds of differences occurred and were stable among these allotetraploid species, raising the possibility that they reflect divergent adaptation to the different environmental conditions experienced by the species [28, 29].

To shed further light on how gene expression might differ between allopolyploid species that originated independently from the same progenitor species, we focus here on two allotetraploid yarrow species, *Achillea alpina* L. and *A. wilsoniana* Heimerl ex Hand. -Mazz., and their parental species, *A. acuminata* (Ledeb.) Sch. -Bip. and *A. asiatica* Serg. (Asteraceae). In China, these tetraploid species have different distributions, with *A. alpina* occurring in the northeast and *A. wilsoniana* in the southwest of the country [30, 31]. Our previous research indicated that the two tetraploids originated independently 35–80 kya following hybridization between their diploid parents during the megainterstadial before the Last Glacial Maximum. Two independent contacts between the parental species were involved, possibly in deglaciated habitats located near refugia present in the mountains of northeast China and relatively southwestern in the Qinling Mountains, respectively [30]. According to plastid sequencing data, *A. asiatica* mostly likely acted as the maternal parent of both tetraploids [32, 33].

To investigate transcriptome changes occurred during allopolyploidization and the following long-term evolution, it is not only necessary to check specific and coexpressed genes among progeny and progenitor species, but also to examine total and relative expression levels of homeologous genes in allotetraploids. Relative expression levels of homeologs may reflect preexisting parental relative levels (parental legacy) or originate following allopolyploidy with one homeolog preferentially expressed relative to the other (expression bias) [34–37]. *Achillea alpina* and *A. wilsoniana* are ideal for such analysis for the following reasons. First, their parental-offspring relationships are clear and simple (no complicated reticulate relationships are involved according to previous studies). Second, the parental species are extant, making it feasible to compare data from allopolyploids with that of their progenitors. Third, the progenitor species, *A. acuminata* and *A. asiatica*, show high levels of genomic sequence divergence [32, 38], while each allopolyploid species maintains both parental genomes intact, having experienced only low levels of homeologous recombination [30, 32, 33]. For these reasons, it is easy to distinguish homeologous genes from each other in the allopolyploid transcriptome, and to measure parental contributions and homeolog expression bias.

In this study, we screened the transcriptome profiles of the two *Achillea* allotetraploid species and their diploid progenitor species by means of whole transcriptome sequencing. By a comparative analysis of these transcriptomes, we examined first the inheritance of parental gene expression, and second relative parental contributions and homeolog expression bias. From our results, we ask whether parental effects which are frequently found in plant hybrid/allopolyploid transcriptomes, are apparent in the present polyploid system. Furthermore and most importantly, we question to what extent inherited patterns of gene expression are similar in different allopolyploids derived from the same parental species, and how significant evolutionary factors, e.g. natural selection and/or genetic drift, have influenced divergent gene expression profiles of the two independently evolved tetraploid species.

Results

Transcriptome profiles

Approximately 34–49 million 100 bp paired-end raw reads were generated for a library of each of the studied *Achillea* species. After removing adapter sequences and filtering out reads with low quality, 93.2–96.4% of clean reads were obtained (Table S1). The initial transcripts were assembled and filtered to 51,414–88,150 unigenes across the studied species, with the N50 length of unigenes always longer than the average length of unigenes in each sample (Table 1). The proportion of unigenes with complete or partial ORFs was 63–71%. These unigenes were used for subsequent gene expression analysis

Table 1 Information of unigenes in the present RNA-Seq data

	acuARX	acuQL	asi	alp	wil
Number of assembled transcripts by Trinity	177,816	180,194	272,030	282,619	300,158
Number of unigenes	51,414	55,391	59,600	81,143	88,150
Average length of unigenes (bp)	1230.20	1243.32	1074.86	976.40	1011.16
N50 length of unigenes (bp)	1678	1687	1544	1386	1432
Number of lncRNAs	5794	5794	7604	11,409	13,748
Number of unigenes with no ORF	9229	11,094	10,342	15,078	18,603
Number of unigenes with complete ORF	21,801	24,123	20,997	24,412	27,984
Number of unigenes with partial ORF	14,590	14,380	20,657	30,244	27,815

Abbreviation of accession names: acuARX for Arxan population of *A. acuminata*; acuQL for Qinling population of *A. acuminata*; alp for *A. alpina*; asi for *A. asiatica*; wil for *A. wilsoniana*

(Table 1). The FPKM values of unigenes showed that data correlation among biological replicates of the same tissue/organ of a species/population was higher than among different tissues/organs, indicating that experimental sampling was repeatable and reliable (Fig. S1).

Specifically expressed and coexpressed genes among each allotetraploid species and its diploid progenitors

As shown in the Venn diagrams (Fig. 1), there were 23,614 (29.1%) and 27,535 (31.2%) genes showing species-specific expression in the allotetraploids *A. alpina* and *A. wilsoniana*, respectively, equating to higher proportions than in the diploid parental species (20–25%) and indicating rather high amounts of novel gene expression in both allotetraploids. The numbers of genes expressed in both parents, but not detected in the allotetraploid transcriptome, were 2150/2137 and 2320/2217 in *A. alpina* and *A. wilsoniana*, respectively, suggesting a relatively low level of gene silencing or loss. With regard to coexpression of genes, 35,286 unigenes (about 43.5% of all unigenes) were coexpressed between *A. alpina* and both diploid species, and 36,385 (about 41.3% of all the unigenes) were coexpressed by *A. wilsoninana* and the two diploids (Fig. 1).

Particularly interesting are the genes of each tetraploid specifically coexpressed with each parental species as this indicates the relative contribution of each parent to the transcriptome of each tetraploid. We found that *A. alpina* specifically coexpressed 9922 unigenes with diploids *A. acuminata*, and 12,321

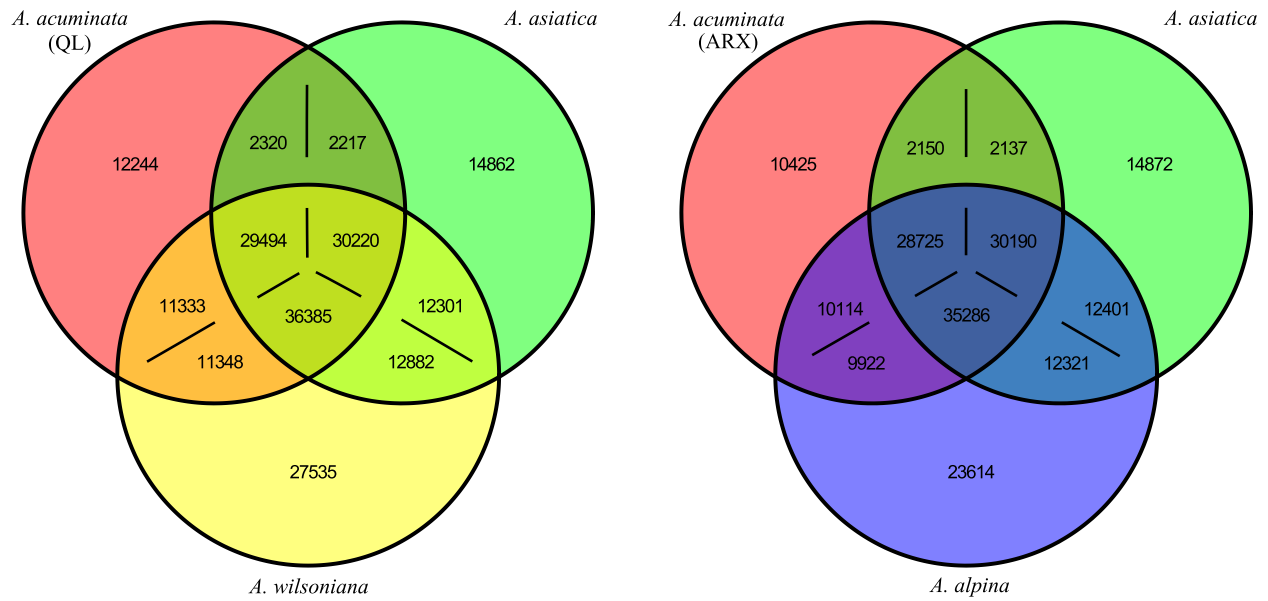

Fig. 1 Venn diagrams showing amounts of coexpressed and specifically expressed genes of the studied allotetraploid species and their diploid progenitors. As the two allopolyploid species originated independently in different regions, and as the diploid *A. acuminata* shows population genetic differentiation, the analysis was conducted separately for each tetraploid species. In the coexpressed gene category, gene number in each species is given (copy-number on some loci may be different among species). Abbreviations: ARX, Arxan Mt.; QL, Qinling Mts

unigenes with *A. asiatica*; while *A. wilsoniana* coexpressed 11,348 and 12,882 unigenes with *A. acuminata* and *A. asiatica*, respectively (Fig. 1). Thus, both tetraploids coexpressed more genes with *A. asiatica* than with *A. acuminata*. Gene Ontology (GO) analysis indicated significant enrichment of these coexpressed genes mostly in terms "response to stress" and "defense response", suggesting that the tetraploid species inherited environmental response genes separately from both progenitors (Fig. 2: A, B; Additional file 6).

Species-specific and coexpressed genes in the two allotetraploid species

Table 2 shows that comparing the expressed genes in the tetraploids, 29.4% genes expressed in *A. alpina* showed species-specific expression and 33.9% genes expressed in *A. wilsoniana* were species-specific. Among the coexpressed genes, 78%–83% were expressed equally in both species, while only about 10% showed up- or down-regulation in one or the other (Table 2). Most enriched GO terms related to biological process (BP) of genes exhibiting species-specific expression pertained to "defense response" in both tetraploids (Fig. 2: C, D; Additional file 7).

In parallel, we found approximately 30% of genes showing population-specific expression in diploid *A. acuminata*; these were most enriched for GO terms pertaining to "defense response" and/or "response to stress" (Fig. 2: E, F; Additional file 7). Moreover, genes coexpressed by each tetraploid with its sympatric *A. acuminata* population were also most enriched for GO terms related to "response to stress" and "defense response" (Fig. 2: A, B; Additional file 6]). These results imply that

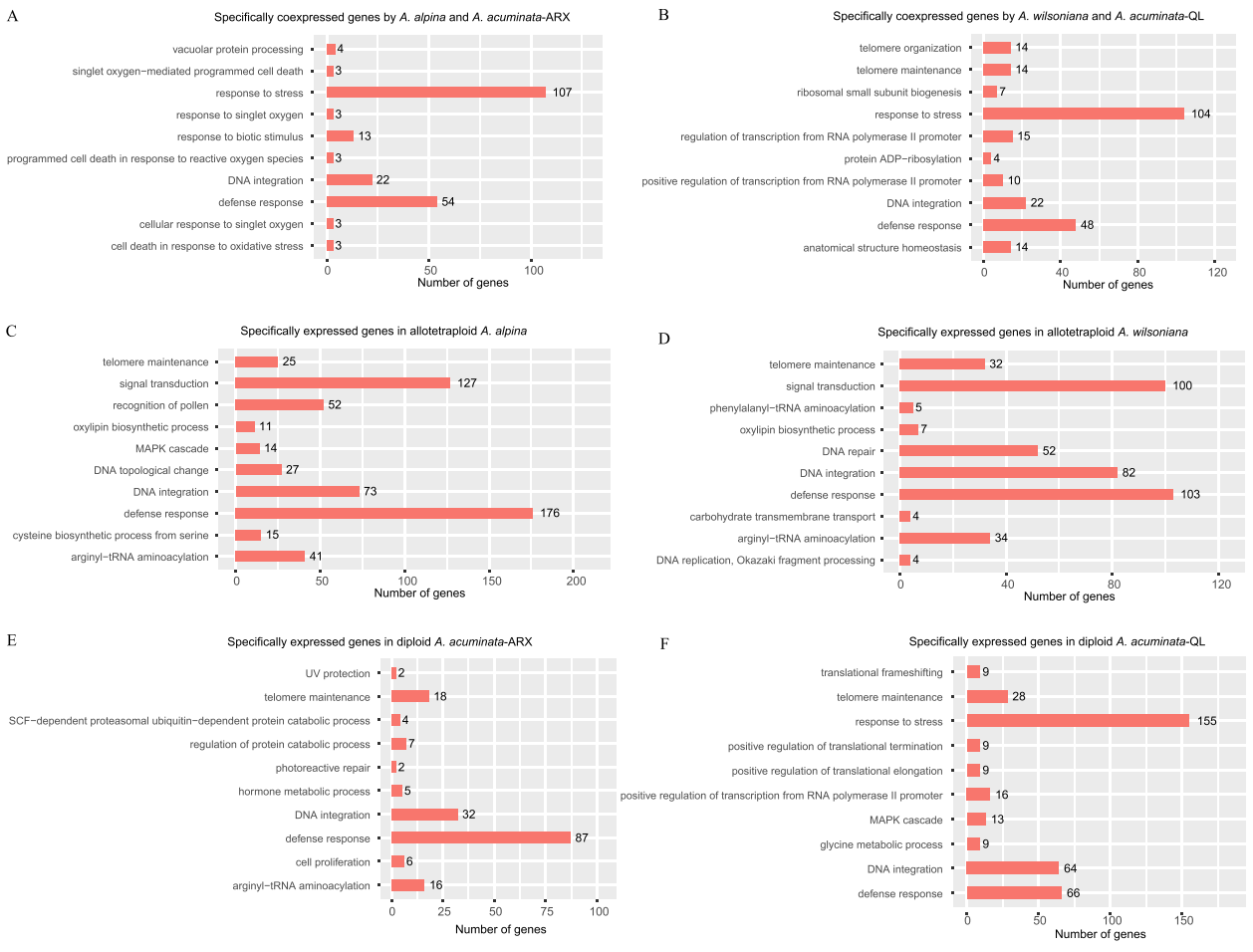

Fig. 2 The top ten most enriched GO terms related to biological process (BP) of specifically coexpressed genes of each allotetraploid species with its sympatric population of *A. acuminata* (**a & b**), species-specific expressed genes in a comparison between the two allotetraploid species (**c & d**), and population-specific expressed genes in a comparison between two populations of diploid *A. acuminata* (**e & f**) (*P*-value < 0.05). These data suggested that the specifically expressed genes were mostly enriched in gene classes pertaining to biological response to environment. The full information of enriched GO terms are listed in Additional Files 6 and 7. Abbreviations: ARX, Arxan population of *A. acuminata*; QL, Qinling population of *A. acuminata*

Table 2 Number of specifically and differentially expressed genes in the two studied allotetraploid species

Specific in *A. alpina*	Specific in *A. wilsoniana*	Expressed in both tetraploids (stem apex)	Expressed in both tetraploids (leaf)
23845 (29.4%)	29881 (33.9%)	4049 (up-regulate in *A. alpina*)	2879 (up-regulate in *A. alpina*)
		8328 (up-regulate in *A. wilsoniana*)	6727 (up-regulate in *A. wilsoniana*)
		44524 (equal expression in both)	47267 (equal expression in both)

the two geographically separated tetraploids may have inherited genes and expression patterns from their sympatric diploid parental species which could be important in local adaptation.

Inheritance patterns of gene expression

Figure 3 shows the numbers and proportions of differentially expressed genes (DEGs) among all expressed genes in the allotetraploids. Most of these genes (71.49% in *A. alpina* and 67.30% in *A. wilsoniana*) were 'conserved', meaning that the total expression of homeologs for a given gene in the allotetraploids was statistically similar to the expression levels of that gene in both parental species.

Altered gene expression in the tetraploids was evidenced by expression inheritance patterns classified into 12 categories. Thus, 5.8 and 5.0% of expressed genes in *A. alpina* and *A. wilsoniana*, respectively, had expression levels intermediate to the parental species (categories I and XII in Fig. 3). Approximately 15% of genes showed "expression-level dominance" (categories II, XI, IV and IX) with both tetraploids exhibiting greater *A. asiatica* expression-level dominance (S-dominance) than *A. acuminata* dominance (C-dominance) (categories IV and IX vs. II and XI). Finally, both tetraploids possessed more transgressively downregulated genes (categories III, VII and X) than transgressively upregulated genes (categories V, VI and VIII).

Relative homeolog contribution and homeolog expression bias

The two allotetraploids displayed a relatively small proportion of silent/lost parental genes. Moreover, they exhibited imbalanced silencing/loss of homeologs between the two parental subgenomes. Silence/loss of genes were more evident for *A. acuminata*-homeologs than for *A. asiatica*-homeologs, implying preferential expression of the *A. asiatica*-subgenome in both tetraploids (Table 3).

The relative homeolog contribution to total expression levels of allotetraploid genes was quantified by Rh [Rh = $\log_2$ (acu-homeolog/asi-homeolog)] (Fig. 4).

Categories	Intermediate		C-expression level dominance		S-expression level dominance		Transgressive down-regulation			Transgressive up-regulation			Conserved	Total
	I	XII	II	XI	IV	IX	III	VII	X	V	VI	VIII		
	S C	S C	S C	S C	S C	S C	S C	S C	S C	S C	S C	S C		
A. alpina (Stem apex)	1165	859	1254	938	1373	1568	121	2569	112	57	33	625	23962	34636
A. alpina (Leaf)	1100	867	1151	760	1146	1389	62	1988	80	55	17	456	25558	34629
Proportion	5.76%		5.92%		7.91%		7.12%			1.79%			71.49%	----
A. wilsoniana (Stem apex)	1003	758	1234	1018	2063	1964	150	2590	184	145	120	1023	22270	34522
A. wilsoniana (Leaf)	972	698	1149	862	1639	1695	110	2145	104	129	43	780	24192	34518
Proportion	4.97%		6.17%		10.66%		7.65%			3.24%			67.30%	----

Fig. 3 Inheritance categories of gene expression of the studied allotetraploid species. The categorization involving 12 states of differential expression (labeled with Roman numeral I–XII) is modified from Rapp et al. (2009) [39]. A cartoon depiction is provided for each of the 12 states, where parental states (S for *A. asiatica*; C for *A. acuminata*) are on the outer edges and the allotetraploid is in the middle. Dots on the same horizontal line indicate statistically equal expression level, whereas dots on higher or lower horizontal lines refer to significantly higher or lower expression level. The 'Intermediate' states, I and XII, indicate gene expression levels in the allopolyploid being significantly different from, but intermediate between the parental levels. The 'conserved' refers to genes with basically equal expression levels among the allotetraploid and both parental species. The number of genes of each category is given, and the percentage of each category group in all expressed genes is provided

Approximately two-thirds of homeolog pairs displayed equal expression of parental copies, and the remaining one-third exhibited different expression levels of parental homeologs. Among the differentially expressed homeologous pairs, more exhibited higher expression of the *A. asiatica* copy than the *A. acuminata* copy.

To determine if the detected differential expression of homeologs is derived from pre-existing differences in parental gene expression levels, or is due to homeolog expression bias, we compared Rh with the relative expression of orthologs between the parental species, Rp [Rp = $\log_2$ (*A. acuminata*/*A. asiatica*)] (Fig. 5). Approximately 79% of homeolog pairs in the tetraploids displayed vertical inheritance of pre-existing parental expression levels, that is, without expression bias. Among the remaining 21% homeolog pairs that displayed parental expression bias, S-bias (bias toward *A. asiatica* copy) was more common than C-bias (bias toward *A. acuminata* copy).

To understand the possible influence of expression bias to the relative contribution of the parental homeologs, we integrated data sets of relative homeolog expression level and homeolog bias (Table S2). Of the homeolog pairs showing equal expression of parental copies, 35% showed expression bias, while the rest simply maintained pre-existing parental expression levels. Of the homeolog pairs with unequal expression levels, most might have resulted from homeolog expression bias. For instance, out of 1396 homeolog pairs showing higher expression level of the *A. asiatica* copy, 1037 (74.3%) displayed expression bias toward the *A. asiatica* copy (Table S2).

Validation of RNA-Seq analysis by RT-qPCR

To validate the analysis and data obtained by RNA-sequencing, differential expression of genes was checked using RT-qPCR assays. Unigenes exhibting different inheritance patterns of gene expression (intermediate expression, *A. acuminata*/*A. asiatica* expression-level dominance, transgressive expression) were randomly chosen for RT-qPCR verifying. For all 10 unigenes tested, expression patterns revealed by qRT-PCR assays were consistent with those evident in the RNA-Seq data (Fig. S2), demonstrating the reliability of data produced by RNA-sequencing.

Discussion

To understand the influence of hybridization and polyploidy on the inheritance of gene expression from parental to allopolyploid species, we conducted a transcriptome analysis on two allotetraploid *Achillea* species that originated independently from the same two parental species. We evaluated RNA-Sequencing data to determine: (*i*) species-specific gene expression and coexpression among both tetraploid and progenitor diploid species; (*ii*) inheritance patterns of parental gene expression; and (*iii*) parental contribution to gene expression level in the tetraploids, and occurrence of homeolog expression bias.

Gene expression profiles in the allotetraploid species with influence of maternal effect

Both hybridization and polyploidization can alter gene expression between progenitors and allopolyploid offspring by affecting the number of expressed genes and their expression levels. In the present analysis only 3.6%–4.7% (Fig. 1) genes expressed in the diploids were not detected in the tetraploid species, suggesting a low level of gene silencing (or loss). On the other hand, each of the tetraploid species possessed a high proportion (approximately 30%) of species-specific expressed genes (23,614 out of 81,143 genes in *A. alpina* and 27,535 out of 88,150 genes in *A. wilsoniana*, Fig. 1), suggesting that hybridization and polyploidy activate some genes not expressed in the diploids.

In hybrid plants, maternal effects may have a strong influence on morphological, life-history and physiological traits, which can be beneficial if the maternal phenotype is linked to increased fitness [40–43]. The present study showed that global gene expression of both *Achillea* tetraploids was frequently more similar to *A. asiatica* than to *A. acuminata*, as reflected by the number of coexpressed genes between species, expression-level dominance, relative homeolog contribution, homeolog-specific expression and homeolog expression bias. This similarity to *A. asiatica* suggests a maternal effect on gene expression with both tetraploids previously shown to have had an *A. asiatica*-like ancestor as their maternal parent [32, 33].

It has been suggested that parental expression-level dominance in allopolyploids mainly results from up- or down-regulation of one of the homeologous copies,

Table 3 Number of silent/lost homeologs in the studied allotetraploids

Samples	Number of silent/lost *A. asiatica*-homeologs (%)	Number of silent/lost *A. acuminata*-homeologs (%)
A. alpina (Stem apex)	362 (3.60%)	539 (5.36%)
A. alpina (Leaf)	311 (3.67%)	479 (5.65%)
A. wilsoniana (Stem apex)	370 (3.50%)	517 (4.89%)
A. wilsoniana (Leaf)	331 (3.99%)	417 (5.02%)

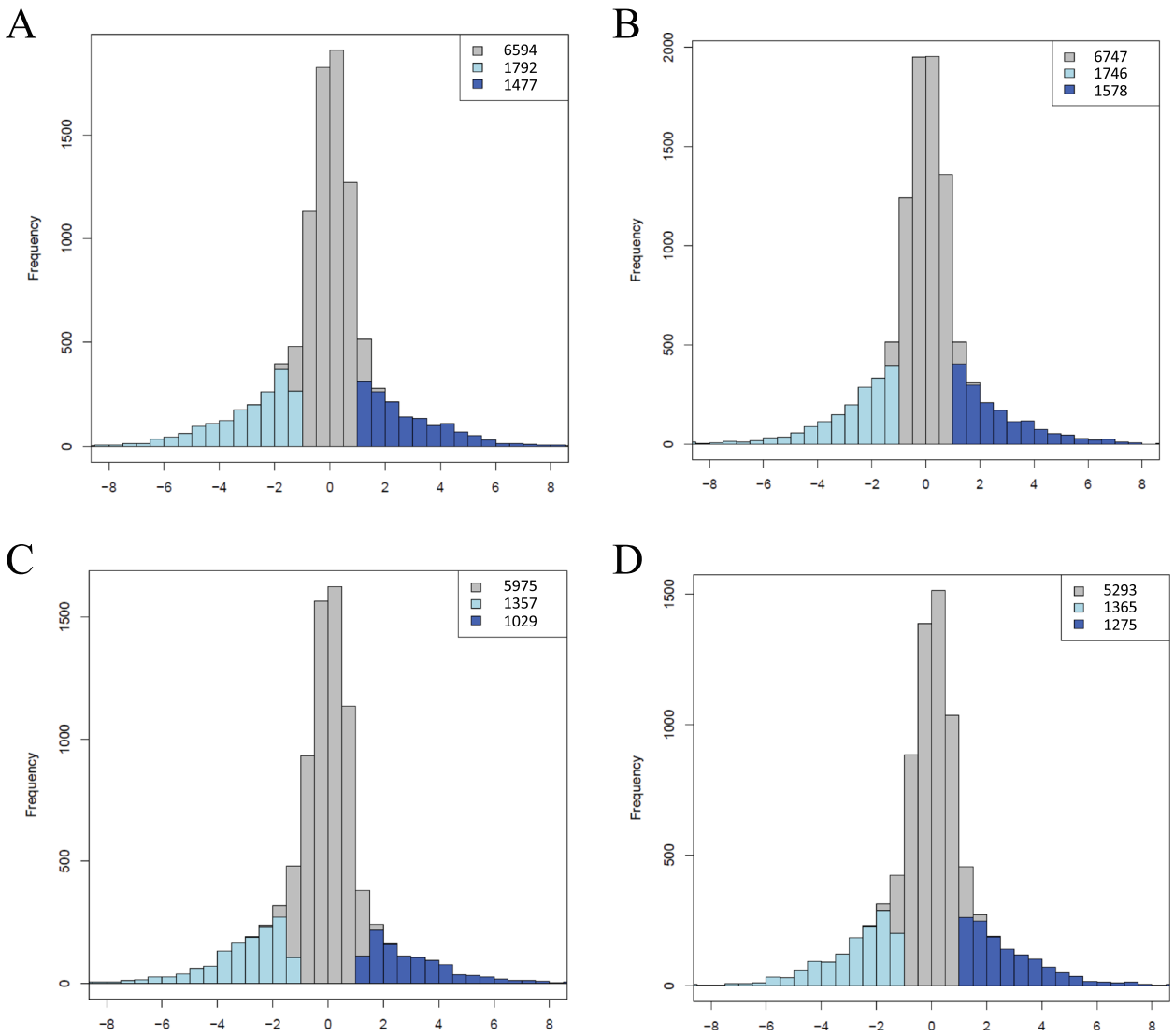

Fig. 4 Histograms showing relative expression levels of homeologous genes in the studied allotetraploid transcriptomes. **a, b** Data from stem apex. **c, d** Data from leaf tissue. **a, c** for *A. alpina*; **b, d** for *A. wilsoniana*. The abscissa is Rh [$\log_2$ (acu-homeolog/asi-homeolog)], and the ordinate is the number of homeolog pairs. Gray columns correspond to homeolog pairs with equal expression level of two parental copies; light blue columns correspond to homeolog pairs with higher expression of the *A. asiatica*-copy, and vise versa, dart blue columns correspond to homeolog pairs with higher expression of the *A. acuminata*-copy (P-value < 0.05, FDR < 0.05). Numbers at the upper right corner indicate the number of homeolog pairs of each category

usually of the 'less dominant' parent [44, 45]. Homeolog expression bias may lead to higher expression of one of the parental gene copies due possibly to a difference between parental subgenomes in number and distribution of transposable elements (usually repressing nearby genes), mismatches between parental copies of *trans*-elements and their target genes, and persistent epigenetic resetting [6, 36, 37, 46, 47]. Maternal effects resulting from one or more of these causes have been reported previously in a number of allopolyploids, e.g. *Gossypium hirsutum* [18, 48], *Spartina anglica* [49], *Triticum aestivum* [50] and *Tragopogon miscellus* [51].

Comparative global gene expression patterns of allopolyploids independently derived from the same parent species

Previous research on *Dactylorhiza* showed that three sibling allotetraploid species derived from the same two parental species were divergent epigenetically and in gene expression, and it was suggested that these

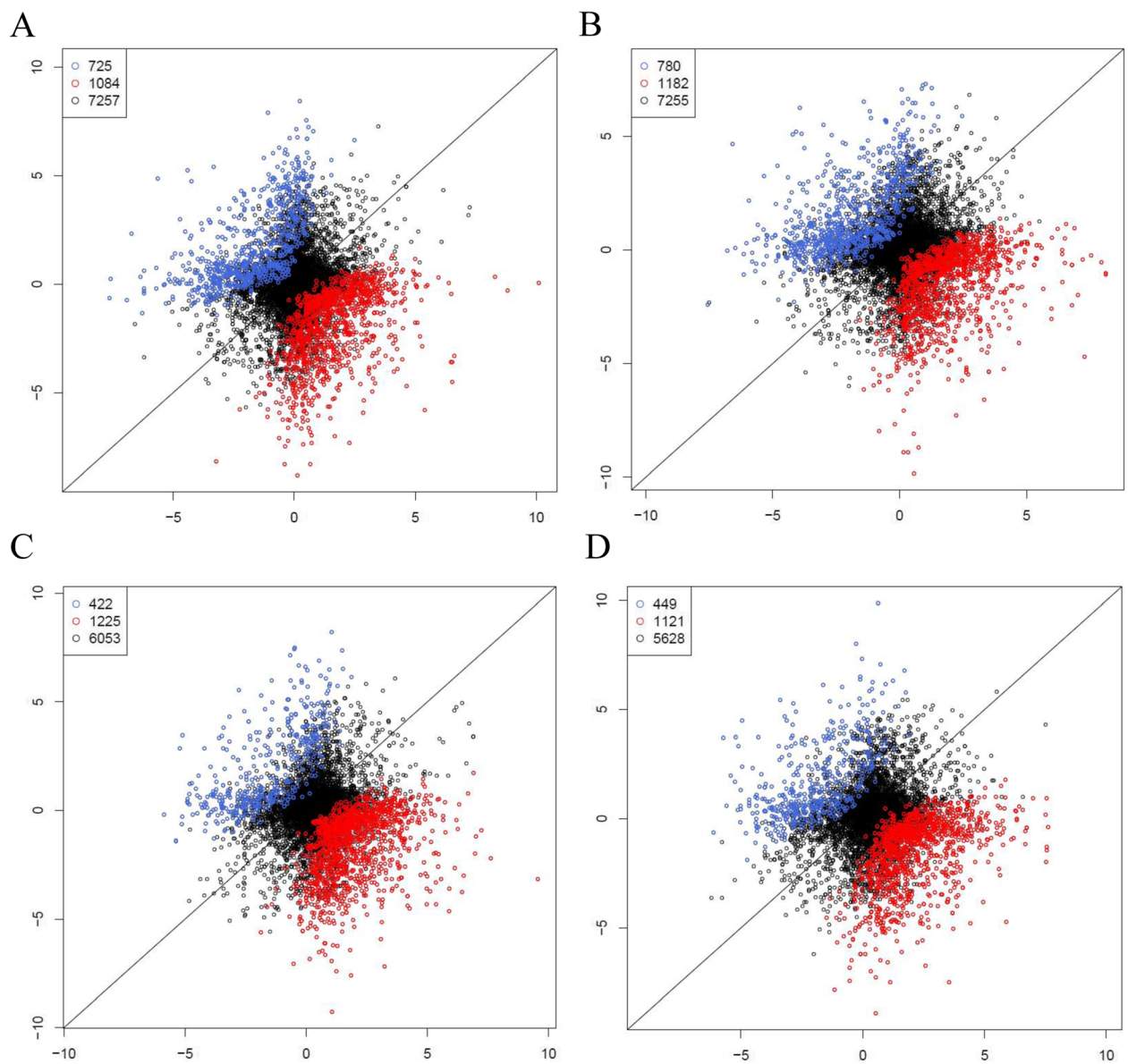

Fig. 5 Scatter plots showing expression bias of homeologs in the studied allotetraploid transcriptomes. **a, b** Data from stem apex. **c, d** Data from leaf tissue. **a, c** for *A. alpina*. **b, d** for *A. wilsoniana*. The abscissa is Rp [$\log_2$ (*A. acuminata*/*A. asiatica*)], and the ordinate is Rh [$\log_2$ (acu-homeolog/asi-homeolog)]. Black circles correspond to homeolog pairs without expression bias; blue circles correspond to the homeolog pairs with expression bias toward *A. acuminata* (namely C-bias); and red circles correspond to homeolog pairs with biased expression of the *A. asiatica*-copy (namely S-bias) (P-value < 0.05, FDR < 0.01, Fisher's exact text). Numbers at the upper left corner indicate the number of homeolog pairs of each category

differences may have been important in their adaptation to different habitats [28, 29]. Similarly, the two *Achillea* tetraploids studied here originated independently due to multiple contacts between the same two parental species in different geographical regions, with population genetic analysis showing them to be genetically well-differentiated [30]. Comparing expressed genes in the two tetraploids, we found a high proportion of species-specific expression (29.4% in *A. alpina* and 33.9% in *A. wilsoniana*) (Table 2). These species-specific expressed genes were enriched for GO terms pertaining to "defense response" (Fig. 2; Additional file 7). Polyploidy may confer adaptive novelties, as indicated in the aforementioned orchids and in *Achillea* [9, 13, 28, 29, 39, 52–54]. Species-specific and differential expression of genes of *A. alpina* and *A. wilsoniana* might have partly originated as a result of the independent allopolyploidization events that gave rise to these two species, but also to independent post-speciation events due to natural selection and/or genetic drift.

Despite the species-specific and differentially expressed genes between the two allotetraploids, they display similar transcriptome changes in comparison to their diploid progenitors, e.g., maternal effects of *A. asiatica* have influenced both tetraploid transcriptomes as suggested by inheritance patterns of gene expression, parental contributions to tetraploid transcriptomes, and homeolog expression bias.

Conclusion

The present comparative transcriptome analysis revealed that two independently originated *Achillea* allotetraploid species exhibited difference in gene expression, some of which was inevitably produced by randomly combined effects of hybridization and polyploidization, but some others must have occurred and maintained under natural selection and/or genetic drift during their tens of thousand years of evolution [30]. Particularly, the species-specific expressed genes enriched for GO terms pertaining to "defense response" suggested differential adaptation during their post-speciation evolution. On the other hand, they showed similar transcriptome changes in comparison to their diploid progenitors. This similarity may be expected when the combinations of genomes merged by different allopolyploidization events were similar [37]. More detailed studies are now required to determine the adaptive significance of differences in gene expression between these two allotetraploid species, which have been revealed by our analysis.

Methods

Plant materials

Plants used for this study were grown in laboratory incubators (16 h: 8 h light-dark cycles at 23 °C) for 3–4 months from achenes collected from natural populations of the four *Achillea* species in China. Achenes of the allotetraploid *A. alpina,* were sampled from Arxan Mountain in the northeast (N 47°17′, E 120°27′; 860 m), where both diploid species also occur in sympatry. Achenes of the other tetraploid, *A. wilsoniana* in the southwest, were collected from Taibai Mountain (N 33°59′, E 107°17′; alt. 2094 m) in Qinling mountain range, approximately where this allotetraploid was originated. Because populations of diploid species *A. acuminata* in NE China and in Qinling mountains are genetically differentiated [30], achenes of this species were collected from both Arxan Mt. and Taibai Mt.. In contrast, we collected achenes of diploid *A. asiatica* only from Arxan Mt. as populations of this species across E Asia are genetically similar [30].

Tissues analyzed were stem apex and the first fully-spread leaf beneath the stem apex. Samples of different tissues were separately snap frozen in liquid nitrogen and stored at − 80 °C. Three replicates of each tissue were obtained, with each replicate containing samples pooled from several plant individuals so that there was sufficient RNA for analysis.

RNA extraction, cDNA library construction and RNA sequencing

Total RNA was extracted using a RNeasy Plus Mini Kit (Qiagen, Hilden, Germany). RNA concentration and quantification were determined using the NanoDrop 2000 spectrophotometer (Thermo Scientific, USA). RNA-sequencing libraries of each sample were constructed and sequenced on an Illumina Hiseq 2000 platform with 100 bp paired-end reads by the Biodynamic Optical Imaging Center (BIOPIC) of Peking University (Beijing, China). The sequencing data have been deposited with links to BioProject accession number PRJNA669168 in the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA669168).

RNA-Seq de novo assembly and annotation

The number and quality of raw reads from each library were evaluated with FastQC v. 0.11.2 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Adapter sequences, low quality bases (Q < 20) and unknown nucleotides (Ns) were trimmed using Trimmomatic v. 0.32 [55]. After trimming, both end of reads with length above 25 bases were kept for assembly. To minimize technical bias, all filtered clean reads from three biological replicates were used to conduct de novo transcriptome assembly by Trinity (v. r2013-02-25) [56]. Redundant transcripts were removed using CD-HIT (v. 4.6) [57] with 90% identity, and the longest transcript in each group was retained. Positively expressed genes were defined using an empirical cutoff value (FPKM > 1), and those with more than 200 bp were chosen as reliable unigenes [58]. The recognition of ORFs (open reading frames) and lncRNA were conducted on unigenes by TransDecoder (v. 2.0.1) (http://transdecoder.github.io/) and CNCI, separately, and unigenes with a complete or partial ORF were prepared for subsequent gene expression analysis [59].

Functions of unigenes were identified by searching against NCBI NR databases using locally installed BLASTX with an E-value cutoff of 1e-5, and the best alignment results were assigned as annotations of unigenes. The same strategy was applied to searches in Uniprot and KEGG databases.

Ortholog and homeolog identification

All unigene sequences were aligned using BLASTN with cutoff E-value of 1e-10, with orthologous gene families identified by OrthoFinder (v. 0.4.0) [60, 61]. BLAST similarity searches were performed for pairwise

comparisons between libraries. Orthologous genes were standardized by setting E-value ≤5e-100, alignment length ≥ 200 bp, and identity ≥90%.

Previous data showed that the two progenitors of the studied allotetraploids are genetically distinct, and the allotetraploid species have maintained their parental subgenomes relatively intact [30, 32, 33]. This made it easy to identify homeologous gene copies in the allotetraploid species using the single nucleotide polymorphisms (SNPs) between the two diploid species. Clean reads of diploid and allotetraploid species were mapped to the assembled unigenes of the two allotetraploids, and SNPs were identified by SAMtools (v0.1.17) [62]. Only SNPs that could tell the genomes of the parental species *A. acuminata* and *A. asiatica* apart were chosen, and clean reads in the allotetraploids exhibiting parental SNPs were parsed into homeolog-specific bins using custom perl scripts so that reads in the tetraploids were designated as of *A. acuminata*- or *A. asiatica*-type.

Differential expression among species

Species-specific expressed and species-coexpressed genes were identified using orthologous gene families as units. Only genes coexpressed in the allopolyploid and both of parental species were further analyzed for gene expression levels. The number of clean reads mapped onto each gene was counted by RSEM (v. 1.1.13) [63] and the expression level of an unigene was determined as the average of three biological replicates. The analysis for differential expression between an allopolyploid and each of its diploid progenitors was performed using edgeR (v. 2.2.5) in R software (v. 2.13) with the trimmed mean of M-values (TMM) to normalize read counts within and across libraries. Benjamini and Hochberg (BH) methods were used to adjust *p*-values to account for significance of differentially expressed genes (DEGs) [64, 65]. DEGs were identified by absolute value of $\log_2$ (fold change) > 1 and FDR < 0.05 using a negative binomial test. DEGs among the allotetraploid and its diploid progenitors were assigned to 12 categories modified from Rapp et al. (2009) [48] containing intermediate expression of the polyploids between that of the parents, expression-level dominance, transgressive expression, and conserved (equal in all species).

Analyses of homeolog expression bias

To calculate the expression levels of homeologs in the allotetraploid, read number mapped onto putative parental interspecific SNPs was counted and the average of those read number was calculated when more than one such SNP occurred in one fragment. To understand the homeolog-specific contributions to the allotetraploid gene expression, the analysis of differential expression was assessed between the two parental homeologs via a negative binomial test in edgeR package with the criterion of absolute value of $\log_2$ (fold change) > 1, FDR < 0.05 and *P*-value < 0.05. To further quantify expression level differences, we defined the relative expression of homeologs (Rh) as: Rh = $\log_2$ (acu-homeolog/asi-homeolog), where acu-homeolog or asi-homeolog is the expression level of the corresponding homeolog. This measurement can be computed for any homeolog pair with non-zero read counts (testable homeolog pairs); when Rh > 0, it indicates a higher expression level of the *A. acuminata*-homeolog than the *A. asiatica*-homeolog, and vice versa, when Rh < 0, the *A. asiatica*-homeolog may be expressed higher.

To examine the homeolog expression bias, we further defined the relative expression level of orthologous pairs of genes in parental species (Rp) as: Rp = $\log_2$ (*A. acuminata*/*A. asiatica*) and compared Rh with Rp using Fisher's exact tests with the criterion of absolute value of $\log_2$ (fold change) > 1, FDR < 0.01 and P-value < 0.05. When Rh > Rp, it indicates expression bias toward diploid *A. acuminata*, and vice versa, when Rh < Rp, expression bias toward diploid *A. asiatica*.

Validation of DEGs by reverse transcription real-time quantitative PCR (RT-qPCR)

To confirm the differential gene expression presented by the RNA-Seq data, we performed reverse transcription Real-Time quantitative PCR (RT-qPCR) analysis on several randomly selected genes. Gene-specific primer pairs were designed by the Primer Premier 5.0. Tissue/organ samples were the same as for the RNA-Seq analysis. Three independent batches of RNA were isolated as biological replicates. The Fast Quant RT kit (with gDNase) (Tiangen Biotech, Beijing, China) was used for cDNA synthesis following the manufacturer's instructions. Then SYBR Premix Ex TaqTM (Tli RNaseH Plus) (Takara) was used for qPCR reactions. PCR reactions were performed on a 7500/7500 Fast Real-Time PCR System (Applied Biosystems) with the following program: 95 °C for 5 min, and then 40 PCR cycles at 95 °C for 5 s; 60 °C for 34 s. The glucose 6-phosphate dehydrogenase (G6PDH) and protein phosphatase 2A (PP2A) genes, which were confirmed having similar expression level among all the studied species by RNA-seq and qRT-PCR, were used as the internal reference genes. A relative quantitative method (delta-delta Ct) was used to evaluate the expression level of candidate genes. Primers used in this study are listed in Table S3.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-021-07566-6.

Additional file 1: Supplementary Fig. S1. Correlation analysis of transcriptome data from different samples.

Additional file 2: Supplementary Fig. S2. Patterns of gene expression detected by RT-qPCR to verify the RNA-Seq data.

Additional file 3: Supplementary Table S1. Information of reads in the transcriptome data.

Additional file 4: Supplementary Table S2. Relative expression level and expression bias of homeologs in the studied *Achillea* allotetraploid species.

Additional file 5: Supplementary Table S3. Primers used in RT-qPCR assays.

Additional file 6. Full information of enriched GO terms of genes specifically coexpressed by each allotetraploid species with each of its diploid progenitors as shown in Fig. 1. The top ten most enriched GO terms related to biological process (BP) of genes specifically coexpressed of each allotetraploid species with its sympatric diploid *A. acuminata* are shown in Fig. 2.

Additional file 7. Full information of enriched GO terms of (1) species-specifically expressed genes when the two allotetraploid species are compared, and (2) population-specifically expressed genes when two populaitons of the diploid parental *A. acuminata* are compared. The top ten most enriched GO terms related to biological process (BP) are shown in Fig. 2.

Acknowledgements
We are grateful to Richard J. Abbott (University of St Andrews, UK) for constructive advices and comments for this manuscript.

Authors' contributions
DC designed and performed the experiments, interpreted the RNA-Seq data and wrote the draft of the manuscript. PCY analyzed the RNA-Seq data. YPG conceived the research, led the manuscript writing. All authors read and approved the final manuscript.

Funding
This work was supported financially by the National Natural Science Foundation of China (Grant Nos. 32070234 and 31570215), and the 111 project (B13008) of China.

Availability of data and materials
The data generated and analyzed by this study have been deposited with links to BioProject accession number PRJNA669168 in the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA669168).

Declarations

Ethics approval and consent to participate
The plants under this study are not rare or endangered. The samples were collected in their wild populations in non-protected areas; no any legal authorization/license is required.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Key Laboratory of Biodiversity Science and Ecological Engineering of the Ministry of Education, and College of Life Sciences, Beijing Normal University, Beijing, China. [2]Beijing Tangtang Tianxia Biotechnology Co., Ltd, Beijing, China.

Received: 9 November 2020 Accepted: 30 March 2021
Published online: 13 April 2021

References
1. Hegarty MJ, Barker GL, Wilson ID, Abbott RJ, Edwards KJ, Hiscock SJ. Transcriptome shock after interspecific hybridization in *Senecio* is ameliorated by genome duplication. Curr Biol. 2006;16(16):1652–9. https://doi.org/10.1016/j.cub.2006.06.071.
2. Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, et al. Evolutionary genetics of genome merger and doubling in plants. Ann Rev Genet. 2008;42(1):443–61. https://doi.org/10.1146/annurev.genet.42.110807.091524.
3. Buggs RJ, Zhang L, Miles N, Tate JA, Gao L, Wei W, et al. Transcriptomic shock generates evolutionary novelty in a newly formed natural allopolyploid plant. Curr Biol. 2011;21(7):551–6. https://doi.org/10.1016/j.cub.2011.02.016.
4. Springer NM, Li Q, Lisch DJPC. Creating order from chaos: Epigenome dynamics in plants with complex genomes. Plant Cell. 2016;2016:314–25.
5. Van de Peer Y, Mizrachi E, Marchal K. The evolutionary significance of polyploidy. Nat Rev Genet. 2017;18(7):411–24. https://doi.org/10.1038/nrg.2017.26.
6. Wendel JF, Lisch D, Hu G, Mason AS. The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation. Curr Opin Genet Dev. 2018;49:1–7. https://doi.org/10.1016/j.gde.2018.01.004.
7. Anderson E, Stebbins GL. Hybridization as an evolutionary stimulus. Evolution. 1954;8(4):378–88. https://doi.org/10.1111/j.1558-5646.1954.tb01504.x.
8. Grant V. Plant Speciation. New York: Columbia University Press; 1981. https://doi.org/10.7312/gran92318.
9. Otto SP. The evolutionary consequences of polyploidy. Cell. 2007;131(3):452–62. https://doi.org/10.1016/j.cell.2007.10.022.
10. Chen ZJ. Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. Ann Rev Plant Biol. 2007;58(1):377–406. https://doi.org/10.1146/annurev.arplant.58.032806.103835.
11. Jiao YN, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, et al. Ancestral polyploidy in seed plants and angiosperms. Nature. 2011;473(7345):97–100. https://doi.org/10.1038/nature09916.
12. Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, et al. Hybridization and speciation. J Evol Biol. 2013;26(2):229–46. https://doi.org/10.1111/j.1420-9101.2012.02599.x.
13. Soltis DE, Visger CJ, Soltis PS. The polyploidy revolution then … and now: Stebbins revisited. Amer J Bot. 2014;101(7):1057–78. https://doi.org/10.3732/ajb.1400178.
14. Kashkush K, Feldman M, Levy AA. Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. Genetics. 2002;160(4):1651–9.
15. Wang J, Tian L, Lee H-S, Wei NE, Jiang H, Watson B, et al. Genomewide nonadditive gene regulation in Arabidopsis allotetraploids. Genetics. 2006;172(1):507–17. https://doi.org/10.1534/genetics.105.047894.
16. Zhang D, Pan Q, Tan C, Zhu B, Ge X, Shao Y, et al. Genome-wide gene expressions respond differently to A-subgenome origins in *Brassica napus* synthetic hybrids and natural allotetraploid. Front Plant Sci. 2016;7:1508.
17. Wu J, Lin L, Xu M, Chen P, Liu D, Sun Q, et al. Homoeolog expression bias and expression level dominance in resynthesized allopolyploid *Brassica napus*. BMC Genomics. 2018;19:1–13.
18. Yoo M, Szadkowski E, Wendel J. Homoeolog expression bias and expression level dominance in allopolyploid cotton. Heredity. 2013;110(2):171–80. https://doi.org/10.1038/hdy.2012.94.
19. Mandáková T, Lysak MA. Post-polyploid diploidization and diversification through dysploid changes. Curr Opin Plant Biol. 2018;42:55–65. https://doi.org/10.1016/j.pbi.2018.03.001.
20. Soltis DE, Soltis PS. Polyploidy: recurrent formation and genome evolution. Trends Ecol Evol. 1999;14(9):348–52. https://doi.org/10.1016/S0169-5347(99)01638-9.
21. Soltis D, Buggs R, Barbazuk W, Schnable P, Soltis P. On the origins of species: does evolution repeat itself in polyploid populations of independent origin? Cold Spring Harb Symp Quant Biol. 2009;74:215–23. https://doi.org/10.1101/sqb.2009.74.007.
22. Tate JA, Ni Z, Scheen AC, Koh J, Gilbert CA, Lefkowitz D, et al. Evolution and expression of homoeologous loci in *Tragopogon miscellus* (Asteraceae), a recent and reciprocally formed allopolyploid. Genetics. 2006;173(3):1599–611. https://doi.org/10.1534/genetics.106.057646.
23. Buggs RJA, Doust AN, Tate JA, Koh J, Soltis K, Feltus FA, et al. Gene loss and silencing in *Tragopogon miscellus* (Asteraceae): comparison of natural and synthetic allotetraploids. Heredity. 2009;103(1):73–81. https://doi.org/10.1038/hdy.2009.24.
24. Tate JA, Joshi P, Soltis KA, Soltis PS, Soltis DE. On the road to diploidization? Homoeolog loss in independently formed populations of the allopolyploid *Tragopogon miscellus* (Asteraceae). BMC Plant Biol. 2009;9(1):80–9. https://doi.org/10.1186/1471-2229-9-80.

25. Koh J, Soltis PS, Soltis DE. Homoeolog loss and expression changes in natural populations of the recently and repeatedly formed allotetraploid *Tragopogon mirus* (Asteraceae). BMC Genomics. 2010;11(1):97–121. https://doi.org/10.1186/1471-2164-11-97.
26. Buggs RJA, Chamala S, Wu W, Gao L, May GD, Schnable PS, et al. Characterization of duplicate gene evolution in the recent natural allopolyploid *Tragopogon miscellus* by next-generation sequencing and Sequenom iPLEX MassARRAY genotyping. Mole Ecol. 2010;19:132–46. https://doi.org/10.1111/j.1365-294X.2009.04469.x.
27. Chester M, Gallagher JP, Symonds VV, Cruz da Silva AV, Mavrodiev EV, Leitch AR, et al. Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon miscellus* (Asteraceae). Proc Nat Acad Sci USA. 2012;109:1176–81.
28. Paun O, Bateman RM, Fay MF, Hedrén M, Civeyrel L, Chase MW. Stable epigenetic effects impact adaptation in allopolyploid orchids (*Dactylorhiza*: Orchidaceae). Mole Biol Evol. 2010;27(11):2465–73. https://doi.org/10.1093/molbev/msq150.
29. Paun O, Bateman RM, Fay MF, Luna JA, Moat J, Hedrén M, et al. Altered gene expression and ecological divergence in sibling allopolyploids of *Dactylorhiza* (Orchidaceae). BMC Evol Biol. 2011;11(1):113. https://doi.org/10.1186/1471-2148-11-113.
30. Wan JN, Guo YP, Rao GY. Unraveling independent origins of two tetraploid *Achillea* species by amplicon sequencing. J Syst Evol. 2020;58(6):913–24. https://doi.org/10.1111/jse.12544.
31. Shih C, Fu GX. *Achillea L*. In: Flora Reipublicae Popularis Sinica, vol. 76. Beijing: Science Press; 1983. p. 9–19.
32. Guo YP, Vogl C, van Loo M, Ehrendorfer F. Hybrid origin and differentiation of two tetraploid *Achillea* species in East Asia: molecular, morphological and ecogeographical evidence. Mole Ecol. 2006;15:133–44.
33. Guo YP, Tong XY, Wang LW, Vogl C. A population genetic model to infer allotetraploid speciation and long-term evolution applied to two yarrow species. New Phytol. 2013;199(2):609–21. https://doi.org/10.1111/nph.12262.
34. Grover C, Gallagher J, Szadkowski E, Yoo M, Flagel L, Wendel J. Homoeolog expression bias and expression level dominance in allopolyploids. New Phytol. 2012;196(4):966–71. https://doi.org/10.1111/j.1469-8137.2012.04365.x.
35. Buggs RJ, Wendel JF, Doyle JJ, Soltis DE, Soltis PS, Coate JE. The legacy of diploid progenitors in allopolyploid gene expression patterns. Philos Trans R Soc B Biol Sci. 2014;369(1648):20130354. https://doi.org/10.1098/rstb.2013.0354.
36. Yoo MJ, Liu X, Pires JC, Soltis PS, Soltis DE. Nonadditive gene expression in polyploids. Ann Rev Genet. 2014;48(1):485–517. https://doi.org/10.1146/annurev-genet-120213-092159.
37. Bottani S, Zabet NR, Wendel JF, Veitia RA. Gene expression dominance in allopolyploids: hypotheses and models. Trends Plant Sci. 2018;23(5):393–402. https://doi.org/10.1016/j.tplants.2018.01.002.
38. Guo YP, Ehrendorfer F, Samuel R. Phylogeny and systematics of *Achillea* (Asteraceae-anthemideae) inferred from the nrITS and plastid *trn*L-F DNA sequences. Taxon. 2004;53(3):657–72. https://doi.org/10.2307/4135441.
39. Ramsey J. Polyploidy and ecological adaptation in wild yarrow. Proc Natl Acad Sci U S A. 2011;108(17):7096–101. https://doi.org/10.1073/pnas.1016631108.
40. Uller T. Developmental plasticity and the evolution of parental effects. Trends Ecol Evol. 2008;23(8):432–8. https://doi.org/10.1016/j.tree.2008.04.005.
41. Donohue K. Completing the cycle: maternal effects as the missing link in plant life histories. Philos Trans R Soc Lond Ser B Biol Sci. 2009;364(1520):1059–74. https://doi.org/10.1098/rstb.2008.0291.
42. Videvall E, Sletvold N, Hagenblad J, Ågren J, Hansson B. Strong maternal effects on gene expression in *Arabidopsis lyrata* hybrids. Mol Biol Evol. 2016;33(4):984–94. https://doi.org/10.1093/molbev/msv342.
43. Gong L, Salmon A, Yoo M-J, Grupp KK, Wang Z, Paterson AH, et al. The cytonuclear dimension of allopolyploid evolution: an example from cotton using rubisco. Mol Biol Evol. 2012;29(10):3023–36. https://doi.org/10.1093/molbev/mss110.
44. Shi XL, Ng DWK, Zhang CQ, Comai L, Ye WX, Chen ZJ. *Cis*-and *trans*-regulatory divergence between progenitor species determines gene-expression novelty in *Arabidopsis* allopolyploids. Nat Commun. 2012;3(1):950. https://doi.org/10.1038/ncomms1954.
45. Combes MC, Hueber Y, Dereeper A, Rialle S, Herrera JC, Lashermes P. Regulatory divergence between parental alleles determines gene expression patterns in hybrids. Genome Biol Evol. 2015;7(4):1110–21. https://doi.org/10.1093/gbe/evv057.
46. Parisod C, Alix K, Just J, Petit M, Sarilar V, Mhiri C, et al. Impact of transposable elements on the organization and function of allopolyploid genomes. New Phytol. 2010;186(1):37–45. https://doi.org/10.1111/j.1469-8137.2009.03096.x.
47. Vicient CM, Casacuberta JM. Impact of transposable elements on polyploid plant genomes. Ann Bot. 2017;120(2):195–207. https://doi.org/10.1093/aob/mcx078.
48. Rapp RA, Udall JA, Wendel JF. Genomic expression dominance in allopolyploids. BMC Biol. 2009;7(1):18. https://doi.org/10.1186/1741-7007-7-18.
49. Chelaifa H, Monnier A, Ainouche M. Transcriptomic changes following recent natural hybridization and allopolyploidy in the salt marsh species *Spartina* × *townsendii* and *Spartina anglica* (Poaceae). New Phytol. 2010;186(1):161–74. https://doi.org/10.1111/j.1469-8137.2010.03179.x.
50. Qi B, Huang W, Zhu B, Zhong XF, Guo JH, Zhao N, et al. Global transgenerational gene expression dynamics in two newly synthesized allohexaploid wheat (*Triticum aestivum*) lines. BMC Biol. 2012;10(1):3. https://doi.org/10.1186/1741-7007-10-3.
51. Soltis DE, Buggs RJ, Barbazuk WB, Chamala S, Chester M, Gallagher JP, et al. The early stages of polyploidy: rapid and repeated evolution in *Tragopogon*. In: Soltis PS, Soltis DE, editors. Polyploidy and genome evolution. Berlin: Springer; 2012. p. 271–92. https://doi.org/10.1007/978-3-642-31442-1_14.
52. Clausen J, Keck DD, Hiesey WM. Experimental studies on the nature of species. III. Environmental Responses of Climatic Races of *Achillea*. Washington: Carnegie Institution of Washington Publication; 1948. No. 581
53. Ramsey J, Robertson A, Husband B, Conti E. Rapid adaptive divergence in New World *Achillea*, an autopolyploid complex of ecological races. Evolution. 2008;62(3):639–53. https://doi.org/10.1111/j.1558-5646.2007.00264.x.
54. Chao DY, Dilkes B, Luo HB, Douglas A, Yakubova E, Lahner B, et al. Polyploids exhibit higher potassium uptake and salinity tolerance in *Arabidopsis*. Science. 2013;341(6146):658–9. https://doi.org/10.1126/science.1240561.
55. Cox MP, Peterson DA, Biggs PJ. SolexaQA: at-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinformatics. 2010;11(1):485. https://doi.org/10.1186/1471-2105-11-485.
56. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29(7):644–52. https://doi.org/10.1038/nbt.1883.
57. Li WZ, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006;22(13):1658–9. https://doi.org/10.1093/bioinformatics/btl158.
58. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. Nat Biotechnol. 2010;28(5):511–5. https://doi.org/10.1038/nbt.1621.
59. Liang S, Luo HT, Bu DC, Zhao GG, Yu KT, Zhang CH, et al. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. Nucleic Acids Res. 2013;41:e166.
60. Shiryev SA, Papadopoulos JS, Schaeffer AA, Agarwala R. Improved BLAST searches using longer words for protein seeding. Bioinformatics. 2007;23(21):2949–51. https://doi.org/10.1093/bioinformatics/btm479.
61. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 2015;16(1):157. https://doi.org/10.1186/s13059-015-0721-2.
62. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. 1000 genome project data processing subgroup. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078–9. https://doi.org/10.1093/bioinformatics/btp352.
63. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12(1):323. https://doi.org/10.1186/1471-2105-12-323.
64. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc: Series B. 1995;57:289–300.
65. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-Seq data. Genome Biol. 2010;11(3):R25. https://doi.org/10.1186/gb-2010-11-3-r25.

Publisher's Note