Admixture of divergent genomes facilitates hybridization across species in the family Brassicaceae

Hosub Shin[1,2]*, Jeong Eun Park[1]*, Hye Rang Park[1]*, Woo Lee Choi[1], Seung Hwa Yu[3], Wonjun Koh[1], Seungill Kim[3,4], Hye Yeon Soh[3], Nomar Espinosa Waminal[1,5], Hadassah Roa Belandres[5], Joo Young Lim[1], Gibum Yi[1,2], Jong Hwa Ahn[1], June-Sik Kim[1,6], Yong-Min Kim[7], Namjin Koo[7], Kyunghee Kim[1], Sampath Perumal[1], Taegu Kang[1], Junghyo Kim[3], Hosung Jang[1,2], Dong Hyun Kang[1], Ye Seul Kim[1], Hyeon-Min Jeong[3], Junwoo Yang[1], Somin Song[1], Suhyoung Park[8], Jin A. Kim[9], Yong Pyo Lim[10], Beom-Seok Park[11], Tzung-Fu Hsieh[12], Tae-Jin Yang[1,2,3,6], Doil Choi[1,2,3,6], Hyun Hee Kim[5], Soo-Seong Lee[13] and Jin Hoe Huh[1,2,3,6]

[1]Department of Agriculture, Forestry and Bioresources, College of Agriculture and Life Science, Seoul National University, Seoul 08826, South Korea; [2]Plant Genomics and Breeding Institute, Seoul National University, Seoul 08826, South Korea; [3]Interdisciplinary Program in Agricultural Genomics, Seoul National University, Seoul 08826, South Korea; [4]Department of Environmental Horticulture, University of Seoul, Seoul 02504, South Korea; [5]Department of Life Science, Chromosome Research Institute, Sahmyook University, Seoul 01795, South Korea; [6]Research Institute of Agriculture and Life Science, Seoul National University, Seoul 08826, South Korea; [7]Korea Bioinformation Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon 34141, South Korea; [8]Department of Horticultural Crop Research, National Institute of Horticultural and Herbal Science, Rural Development Administration, Wanju, Jeollabuk-do 55365, South Korea; [9]Department of Agricultural Biotechnology, National Academy of Agricultural Science, Rural Development Administration, Jeonju, Jeollabuk-do 54874, South Korea; [10]Department of Horticulture, Chungnam National University, Daejeon 34134, South Korea; [11]Hongik Bio, Pyeongtaek, Gyeonggi-do 17977, South Korea; [12]Plants for Human Health Institute, North Carolina State University, North Carolina Research Campus, Kannapolis, NC 27695, USA; [13]BioBreeding Institute, Anseong, Gyeonggi-do 17544, South Korea

Author for correspondence:
Jin Hoe Huh
Email: huhjh@snu.ac.kr

Summary

- Hybridization and polyploidization are pivotal to plant evolution. Genetic crosses between distantly related species are rare in nature due to reproductive barriers but how such hurdles can be overcome is largely unknown. Here we report the hybrid genome structure of xBrassicoraphanus, a synthetic allotetraploid of *Brassica rapa* and *Raphanus sativus*.
- We performed cytogenetic analysis and *de novo* genome assembly to examine chromosome behaviors and genome integrity in the hybrid. Transcriptome analysis was conducted to investigate expression of duplicated genes in conjunction with epigenome analysis to address whether genome admixture entails epigenetic reconfiguration.
- Allotetraploid xBrassicoraphanus retains both parental chromosomes without genome rearrangement. Meiotic synapsis formation and chromosome exchange are avoided between nonhomologous progenitor chromosomes. Reconfiguration of transcription network occurs, and less divergent *cis*-elements of duplicated genes are associated with convergent expression. Genome-wide DNA methylation asymmetry between progenitors is largely maintained but, notably, *B. rapa*-originated transposable elements are transcriptionally silenced in xBrassicoraphanus through gain of DNA methylation.
- Our results demonstrate that hybrid genome stabilization and transcription compatibility necessitate epigenome landscape adjustment and rewiring of *cis–trans* interactions. Overall, this study suggests that a certain extent of genome divergence facilitates hybridization across species, which may explain the great diversification and expansion of angiosperms during evolution.

Introduction

Genome hybridization and polyploidization have served as major driving forces in plant evolution (Wendel, 2000; Soltis & Soltis, 2009, 2016; Van de Peer *et al.*, 2017; Cheng *et al.*, 2018). However, strong hybridization barriers exist in nature to prevent gene flow between different species in plants and animals (Abbott *et al.*, 2013). Several mechanisms have been proposed to explain the postzygotic barriers resulting from genome incompatibility between distantly related species (Lafon-Placette & Köhler, 2015;

*These authors contributed equally to this work.

Dion-Cote & Barbash, 2017). Among them, a 'genome shock' is proposed as one of the critical causes of genome destabilization upon hybridization, restructuring the hybrid genome through changes of chromosomal organization or mobilization of transposable elements (TEs) (McClintock, 1984). Another is a 'transcriptome shock' that incurs extensive changes of parental gene expression patterns in the hybrid (Hegarty *et al.*, 2006; Buggs *et al.*, 2011).

Despite such negative consequences of hybridization between distantly related species, novel species can be naturally or artificially produced on rare occasions while overcoming the hybridization barrier, the mechanism of which is largely unknown. The family Brassicaceae contains a variety of agronomically important crop species such as broccoli, cabbage, cauliflower, oilseed rape, radish and turnip, in addition to the model plant *Arabidopsis*. Within the genus, *Brassica* is well known for hybridization between different species (interspecific hybridization). For instance, three diploid species, *Brassica rapa* (*Br*; AA), *B. nigra* (BB) and *B. oleracea* (*Bo*; CC), can hybridize with each other, generating allotetraploid species *B. napus* (AACC), *B. juncea* (AABB) and *B. carinata* (BBCC), as epitomized by the model of the 'Triangle of U' (Nagaharu, 1935).

Hybridization between species in the family Brassicaceae is not restricted to interspecific hybridization. Since first noted by Sageret in 1826 (Oost, 1984), intergeneric hybrids between *Brassica* and *Raphanus* have been sporadically reported (Karpechenko, 1928; Mcnaughton, 1973; Dolstra, 1982) but failed to survive. *Brassica* and *Raphanus* are estimated to have diverged 7–22.4 million years ago (Ma), although whether they belong to different genera remains controversial (Mitsui *et al.*, 2015; Jeong *et al.*, 2016; Kim *et al.*, 2018). Recently developed x*Brassicoraphanus* (x*B*) (AARR; $2n = 4x = 38$) is an intergeneric allotetraploid between *Br* (AA; $2n = 2x = 20$) and *Raphanus sativus* (*Rs*) (RR; $2n = 2x = 18$) (Lee *et al.*, 2011). Unlike most newly synthesized interspecific/intergeneric hybrids, x*B* is self-fertile and genetically stable, displaying phenotypic uniformity in successive generations (Supporting Information Fig. S1). The genetic and phenotypic stability of x*B* is exceptional given that many allopolyploids often display a high degree of genome instability and sterility issues, indicating that the hybridization barrier was overcome immediately after the two genomes merged.

We hypothesized that allopolyploidization events have somewhat ameliorated deleterious shock phenomena such as genome and transcriptome shocks, and thereby overcome an intrinsic hybridization barrier between distantly related species. Here we report the genome structure, chromosome behaviors and transcriptome/epigenome profiles of x*B*. We observed inhibition of meiotic nonhomologous interactions, adjustment of homoeologous gene expressions and gain of DNA methylation. All these probably contribute to genome stability and transcription network compatibility in x*B*. This study further proposes the possible mechanisms by which two divergent genomes can successfully merge into a novel species during the evolution of angiosperms.

Materials and Methods

Plant materials

x*Brassicoraphanus* cv BB1 (x*B*), *Brassica rapa* L. cv Chiifu-401-42 (*Br*) and *Raphanus sativus* L. cv WK10039 (*Rs*) were chosen as primary plant materials for genome and cytogenetic analyses. x*B* was originally derived from the microspore culture of a synthetic hybrid between commercial cultivars *B. rapa* cv Jeonseung and *R. sativus* cv Taebaek, and maintained for > 10 generations by self-pollination (Lee *et al.*, 2011; Park *et al.*, 2020). *Br* and *Rs* are inbred lines whose genome sequences are available (Wang *et al.*, 2011; Jeong *et al.*, 2016), propagated by self-pollination. Sterilized seeds were germinated and grown on 1× Murashige & Skoog (MS) medium (Duchefa, Haarlem, the Netherlands) in a growth chamber under 16 h of fluorescent light at 20 ± 10 µmol m^{-2} s^{-1} and 22°C for 14 d. The seedlings including shoots and roots were harvested together for whole genome-sequencing (seq), RNA-seq, bisulfite (BS)-seq, chromatin immunoprecipitation (ChIP)-seq and small RNA-seq. For tissue-specific transcriptome analysis, RNA was extracted from leaf, hypocotyl and root tissue of seedlings and from opened flowers of *Br*, *Rs* and x*B*. For cold treatment, 14 d after sowing, seedlings of *Br*, *Rs* and x*B* were grown at 4°C for 5 wk. RNA was extracted and stored at −20°C until use.

Genome sequencing, assembly and genome size estimation

Paired-end and mate-pair sequencing libraries with insert sizes of 200 bp, 400 bp, 3 kb, 8 kb, 5 kb, 10 kb and 15 kb were constructed using KAPA Library Prep Kit (Roche, Bazel, Switzerland) and Illumina Mate Pair Library Kit (Illumina, San Diego, CA, USA) following the manufacturers' instructions (Table S1). The libraries were sequenced on an Illumina HiSeq 2000 platform. Prokaryotic sequences, duplicated reads, low-quality reads and low-frequency reads were filtered out (Table S1). The preprocessed sequences were assembled using SOAPDENOVO2 (Luo *et al.*, 2012) with the best *k*-mer values for each library. To increase the length of scaffolds, serial scaffolding processes were carried out using SOAPDENOVO2 (Luo *et al.*, 2012) and SSPACE (Boetzer *et al.*, 2011). Gaps in the scaffolds were reduced further using SOAPDENOVO GAPCLOSER (Luo *et al.*, 2012) and PLATANUS (Kajitani *et al.*, 2014) (Table S2). Completeness of the final assembly was validated using BUSCO v.5.2.2 (Simao *et al.*, 2015) with the embryophyta ortholog database. In the *k*-mer analysis, counting *k*-mer occurrence of 19-mers was performed using JELLYFISH (Marcais & Kingsford, 2011). The genome size of x*B* was estimated by flow cytometry analysis (FACSCalibur; BD Biosciences, Franklin Lakes, NJ, USA) as previously described (Huang *et al.*, 2013). Genome data were visualized with CIRCOS (Krzywinski *et al.*, 2009).

Chloroplast genome assembly

The chloroplast genome was *de novo* assembled from the 1× coverage of whole-genome sequencing reads. The chloroplast genome was annotated with GESEQ (Tillich *et al.*, 2017) and

manually curated. The chloroplast genome was visualized using OrganellarGenomeDRAW (Lohse et al., 2013).

Assignment of scaffolds to A_{xB} and R_{xB} subgenomes

Whole-genome sequencing reads of Rs and Br from the Brassica Database (BRAD) were mapped to the xB scaffolds using Bowtie (Langmead et al., 2009). The number of mapped reads was counted and the scaffolds were assigned to A_{xB} and R_{xB} subgenomes, based on a comparison of the number of parental reads (A_{xB} subgenome: >99% ratio of mapped reads from Br; R_{xB} subgenome: >99% ratio of mapped reads from Rs). Next, assigned xB scaffolds were anchored to the reference chromosomes of Br and Rs to build xB pseudochromosomes.

Gene and TE annotation

Gene annotations of xB and Rs were performed following the previous annotation pipeline with minor modifications (Kim et al., 2014). Briefly, the annotation pipeline consisted of repeat masking, mapping of different protein sequence sets and mapping of RNA-seq reads. Independent ab initio predictions were performed with Augustus (Stanke et al., 2008). The EvidenceModeler (Haas et al., 2008) software combines ab initio gene predictions with protein and transcript alignments into weighted consensus gene structures. Gene annotation of Br was downloaded from Ensembl plant (ftp://ftp.ensemblgenomes.org/pub/plants/release-31/gff3/brassica_rapa/) and an additional 1700 genes were annotated using Exonerate (Slater & Birney, 2005). Functional annotation was performed through Blastp against the SwissProt and Plant RefSeq databases. TE-related repeat sequences were predicted by RepeatModeler (Smit & Hubley, 2008) and RepeatMasker (Smit et al., 2015).

Fluorescence in situ hybridization (FISH) analysis

The sequences of 5S rDNA, 45S rDNA, RsCent1, RsCent2, BrCent1, BrCent2, RsSTRa, RsSTRb, BrSTRa, BrSTRb and telomere were used as probes (Table S3). The probes were labeled by nick translation with different fluorochromes. Root mitotic chromosome spreads and FISH procedures were performed according to a previous method (Waminal & Kim, 2012). For directly labeled probes, slides were immediately used for FISH after fixation with 4% paraformaldehyde, without subsequent pepsin and RNase pretreatment. Images were captured with an Olympus BX53 fluorescence microscope equipped with a Leica DFC365 FS CCD camera and processed using Cytovision v.7.2 (Leica Microsystems, Wetzlar, Germany).

Resynthesized allodiploid and allotetraploid xBrassicoraphanus plants

Resynthesized allodiploid xBrassicoraphanus plants were produced from a cross between Br as a seed parent and Rs as a pollen donor. Thirty-day-old immature hybrid ovules were cultured on 1× MS medium supplemented with 2% (w/v) sucrose and 0.8%

(w/v) plant agar. The plates were placed at 24°C in a growth chamber for 2 wk and then seedlings were vernalized at 4°C in a cold chamber for 4 wk with 16 h : 8 h, light : dark. The plants were transferred to pots in the glasshouse with the same light conditions. A 0.3% colchicine solution was applied to the shoot apical meristem for 2 d to induce chromosome doubling.

Resynthesized allodiploid and allotetraploid B. napus plants

Resynthesized allodiploid B. napus plants were produced from a cross between Br as a seed parent and B. oleracea var. Capitata as a pollen donor. Ovary culture was performed as described by Inomata (1977) with modifications. Ovaries at 4 d after pollination were explanted on 1× MS medium supplemented with 5% (w/v) sucrose, 300 mg l^{-1} casein hydrolysate and 0.8% (w/v) plant agar at 24°C in a growth chamber. Four weeks after explantation, hybrid ovules were transferred to B5 medium with vitamin supplemented with 2% (w/v) sucrose, 300 mg l^{-1} casein hydrolysate and 0.8% (w/v) plant agar. Ovules were incubated in the dark at 24°C for 3 d under 16 h : 8 h, light : dark conditions. Seedlings were vernalized at 4°C in a cold chamber for 4 wk under 16 h : 8 h, light : dark. The plants were transferred to pots in the glasshouse with the same light conditions. A 0.3% colchicine solution was applied to the shoot apical meristems for 2 d to obtain allotetraploids.

Production of antibody and immunolocalization of meiotic proteins

The coding regions of BrASY1 and BrZYP1 genes were PCR-amplified from cDNA of young flowering buds from Br (Table S4). The fragments of BrASY1 (708 bp) and BrZYP1 (1332 bp) were inserted into the pET-28a expression vector (Novagen, Darmstadt, Germany) and transformed into Escherichia coli Rosetta2 (DE3) strains (Novagen). The transformed E. coli cells were grown at 30°C in 1 l of LB medium in the presence of 50 μg ml^{-1} of kanamycin and 50 μg ml^{-1} of chloramphenicol until an OD$_{600}$ of 0.4 was reached. Recombinant protein expression was induced with 1 mM isopropyl β-D-thiogalactopyranoside at 16°C for 16 h. Cells were centrifuged (4°C) at 6700 $\textbf{\textit{g}}$ for 15 min and the pellet was resuspended in 100 ml of ice-cold column buffer (50 mM Tris–HCl, pH 7.4, 100 mM NaCl, 10% glycerol, 0.1 mM dithiothreitol, 0.1 mM phenylmethylsulfonyl fluoride (PMSF)). Cells were lysed by sonication for 5 min on ice (output power, 4; duty cycle, 50%; Branson Sonifer 250; Branson). Inclusion bodies were collected by centrifugation (4°C) at 9300 $\textbf{\textit{g}}$ for 25 min and dissolved in 4 M urea. The soluble lysate was purified with a 5 ml HisTrap FF column (GE Healthcare, Piscataway, NJ, USA) with a linear gradient of ice-cold column buffer (50 mM Tris–HCl, pH 7.4, 100 mM NaCl, 10% glycerol, 0.1 mM dithiothreitol, 250 mM imidazole). The purified BrASY1 and BrZYP1 proteins were used to produce polyclonal antibodies from rabbit and rat, respectively, by Youngin Frontier (Korea), and the quality of the antibody was validated by western blotting. Immunolocalization was performed as described by Chelysheva et al. (2013). In brief,

primary antibodies anti-BrASY1 and anti-BrZYP1 were used at a dilution of 1 : 250 in PBST (0.1% Triton-X 100 in 1× PBS) containing 1% BSA and the secondary antibodies (goat antirabbit IgG H&L, Alexa Fluor 488 and donkey antirat IgG H&L, Alexa Fluor 594) were used at a dilution of 1 : 500. Images were captured with an Axioskop2 microscope equipped with an Axiocam 506 color CCD camera (Carl Zeiss, Jena, Germany) and processed using Adobe Photoshop CS6 (Adobe Systems Inc., Mountain View, CA, USA).

Identification of homoeologous gene pairs

The reciprocal best Blast hit method was used to determine orthologous gene pairs with over 80% identity and coverage between A and R genomes. The regions that share a common order of five or more orthologs in A and R scaffolds with a distance cutoff of 50 genes were defined as syntenic regions. Then, the reciprocal best Blast hit method was performed again on the genes within the syntenic regions of A and R subgenomes, and the homoeologous gene pairs were established. Pairwise sequence alignment using ClustalW (Larkin et al., 2007) was performed to remove gene pairs with low sequence identity and coverage. Finally, only gene pairs with > 80% identity and > 80% coverage were defined as syntenic homoeologs between A and R subgenomes.

RNA-seq analysis

Total RNA was extracted with RNeasy Plant Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol. The DNase-treated RNA samples, including two replicates for each of seedling, leaf, hypocotyl and flower, and one replicate for root of xB and its progenitors, were used for constructing RNA-seq libraries (Zhong et al., 2011). RNA-seq was performed on an Illumina HiSeq 2000 platform. The obtained raw reads were filtered using Fastx-Toolkit and low-quality reads ($Q < 20$) were removed. The filtered reads were mapped to the Br, Rs and xB genomes using Tophat (Trapnell et al., 2009) with default parameters (Table S5). The mapped read counts were calculated using HTSeq (Anders et al., 2015). Genes not expressed in any tissues were excluded from the orthologous or homoeologous gene pair sets for differentially expressed gene (DEG) analysis. Statistical tests of DEGs were performed using EdgeR (Robinson et al., 2010) with false discovery rate (FDR) < 0.05 and fragments per kilobase of transcript per million mapped reads (FPKM) $\log_2$ fold change > 1. Orthologous/homoeologous gene expression patterns were categorized as relative expression levels in the context of A_{Br}–A_{xBr}–R_{xB}–R_{Rs}. Differential expressions for every gene pair of A_{Br} vs R_{Rs}, A_{Br} vs A_{xBr}, A_{xB} vs R_{xB} and R_{xB} vs R_{Rs} were assessed and classified into 27 groups of relative expression changes except ambiguous patterns. The Gene Ontology (GO) terms of the xB genome were annotated by Blast2GO using the nonredundant sequence database from NCBI with e-value parameter $< 1e^{-15}$. Statistical comparison of GO term accumulation was conducted using the R package topGO (Alexa & Rahnenfuhrer, 2010) with P-values of Fisher's exact test ($P < 0.001$). Motifs of ABRE (BACGTGK, B = C, G or T; K = G or T) and

DRE/CRT (RCCGAC, R = A or G) were searched in 500 bp upstream regions of genes using Fimo (Grant et al., 2011) with parameters '--verbosity 1 --thresh 0.01'.

BS-seq analysis

Genomic DNA (5 μg) extracted from 14-d-old seedlings was used to construct the BS-seq library with the KAPA Library Kit (Roche) and EpiTect Bisulfite Kit (Qiagen) according to the manufacturers' instructions. The libraries were sequenced using the Illumina HiSeq 2000. Raw reads were filtered using Fastx-Toolkit and low-quality reads ($Q < 20$) were removed. Reads were mapped onto the xB genome using Bison (Ryan & Ehninger, 2014), with the parameters '--very-sensitive --score-min "L,-0.6,-0.6"'. Only cytosine sites with 4× coverage read depth were accepted for subsequent analysis. Differentially methylated cytosines (DMCs) and regions (DMRs) were identified as described previously (Kim et al., 2019). In brief, DMCs were identified using Fisher's exact test ($P < 0.05$) between the levels of methylation in xB and the progenitors Br and Rs. DMRs were identified based on the regions with a length ≥ 200 bp, five or more DMCs, and mean methylation difference ≥ 0.3 for CG, ≥ 0.15 for CHG and ≥ 0.1 for CHH. Regions 1 kb up- and downstream of the gene body and repeat were divided into 50 bp windows, and weighted DNA methylation levels were represented with a metaplot (Schultz et al., 2012). Methylation data were visualized with the Integrated Genome Browser (Freese et al., 2016).

ChIP-seq analysis

ChIP was performed following Lee et al. (2007). Chromatin was immunoprecipitated with antibody against histone H3K9me2 (Abcam, Cambridge, MA, USA). ChIP-seq libraries were constructed as described in the Illumina ChIP Sequencing Kit (Illumina). DNA fragments of c. 600 bp were excised from an agarose gel and amplified for cluster generation and sequencing. All DNA libraries were sequenced on a HiSeq2500 platform (Illumina) with single-end reads. The sequencing reads were quality-controlled with Fastx-Toolkit and aligned to the xB genome using Bowtie (Langmead et al., 2009) with parameters '-best -m1'. H3K9me2-enriched regions were defined using Sicer (Zang et al., 2009) (window size = 500, gap size = 600, FDR = 0.01) and overlapping regions between two biological replicates were identified using the MergePeaks module of the Homer software (Heinz et al., 2010).

Small RNA-seq analysis

Small RNA libraries were constructed using the Illumina TruSeq Small RNA sample Prep Kit (Illumina). The libraries were sequenced on the HiSeq 2000 platform (Illumina). Adaptor sequences were trimmed using Cutadapt (Martin, 2011) with parameters '-g TACAGTCCGACGATC -a TGGAATTCTC GGGTGCCAAGG -m 18 -M 30'. Low-quality sequences were removed using Fastx-Toolkit with parameters '-q 20 -p 100'.

Quality-trimmed read sequences ranged from 18 to 30 nt were mapped to the x*B* genome using Bowtie (Langmead *et al.*, 2009) with parameters '-best -v 0'. Mapped reads were classified into rRNA, small nucleolar RNA, small nuclear RNA, signal recognition particle RNA, and tRNA using the Rfam database v.12.1 (Nawrocki *et al.*, 2014). Prediction of microRNA (miRNA) was performed with miRDeep-P (Yang & Li, 2011) and ShortStack (Axtell, 2013), and the secondary structure was predicted using RNAFold. Candidate miRNAs were annotated by alignment to the miRBase database v.21 (Kozomara & Griffiths-Jones, 2013).

Northern blot analysis

Total RNA (10 μg) was electrophoresed on a 1% formaldehyde denaturing gel and transferred onto a Hybond N+ membrane (GE Healthcare). The probes *BrGypsy*, *BrCopia* and *Actin* were amplified by PCR and randomly labelled with [α-^{32}P]dCTP (Perkin Elmer, Waltham, MA, USA) using a Klenow fragment (3′ → 5′ exo−) (New England Biolabs, Ipswich, MA, USA). Hybridization was performed at 65°C overnight in the prehybridization solution containing 6× saline sodium citrate buffer, 5× Denhardt's reagent and 1% sodium dodecyl sulfate. After hybridization, the membrane was washed and exposed to an X-ray film (Fujifilm, Tokyo, Japan). Primer sequences are detailed in Table S6.

Results

Genomic features of x*Brassicoraphanus*

x*B* is a fertile and genetically stable intergeneric allotetraploid synthesized from a cross between *Br* and *Rs*. The x*B* genome was *de novo* assembled using 195.0 Gb of Illumina shotgun reads (Fig. 1a; Tables 1, S1). Flow cytometry analysis estimated the size of the x*B* genome as 998.3 Mb, close to the sum of the *Br* (485 Mb) and *Rs* (510 Mb) genomes (Wang *et al.*, 2011; Jeong *et al.*, 2016) (Fig. S2). We assembled 692.8 Mb of sequence covering *c.* 70% of the x*B* genome, on which 97.56% of sequence reads (67 343 270/69 029 900) were successfully mapped back. The final assembly contained 87 861 annotated genes with a benchmarking universal single-copy orthologs (Busco) completeness score of 99.7% (Tables 1, S7). Repeat sequences accounted for 39.19% (255.8 Mb) of the assembled x*B* genome with long terminal repeats (LTRs) being predominant (Table S8). The assembled chloroplast genome of x*B* (153 482 bp) was 99.9% identical to that of *Br*, indicating its maternal origin (Fig. S3; Table S9). In the x*B* genome (692.8 Mb), 335.5 and 343.5 Mb of scaffolds were assigned to the *Br* and *Rs* reference genomes (referred to as A_{Br} and R_{Rs} hereafter), respectively (Wang *et al.*, 2011; Jeong *et al.*, 2016), comprising two subgenomes of x*B* (referred to as A_{xB} and R_{xB} hereafter) (Table 1; Fig. S4). Of 44 432 and 41 733 annotated genes initially assigned to the A_{xB} and R_{xB} subgenomes, a total of 17 393 syntenic homoeologous pairs were identified in 291 synteny blocks with > 80% identity to each other and > 80% coverage. DEGs whose expressions are up- or downregulated relative to the progenitors emerged evenly throughout the x*B* genome (Fig. 1a). DNA methylation is predominant in repeat-enriched regions at all CG, CHG and CHH (H = A, T or C) contexts (Figs 1a, S5). DMRs are regions where DNA methylation levels in x*B* are significantly different (absolute difference > 0.3 for CG, > 0.15 for CHG and > 0.1 for CHH) from those of *Br* and *Rs*, and *c.* 60.2% of hyper-DMRs are confined to repeat regions (Fig. S5; Table S10). Approximately 75.8% of H3K9me2 repressive histone marks are also enriched in repeat regions (Figs 1a, S6; Table S11). Small RNAs (18–30 nt) are distributed throughout the entire x*B* genome and significantly associated with DNA methylation (Fig. 1a). A06 chromosome of x*B* is exemplified in Fig. 1(b), showing distributions of genes, repeats, DNA methylation, H3K9me2 and small RNAs, along with DEGs and DMRs compared to the progenitor (Fig. 1b). Cytological observation revealed a total of 19 chromosome pairs present in x*B* without aneuploidy and/or chromosome rearrangements (Fig. 1c). Previous studies showed that many synthetic allopolyploid plants such as rapeseed, tobacco and wheat went through massive chromosome reconstruction leading to transgressive gain or loss of chromosomes and/or aneuploidy over generations (Xiong *et al.*, 2011; Zhang *et al.*, 2013; Chen *et al.*, 2018; Sosnowska *et al.*, 2020). However, our findings indicate that x*B* retains both A_{Br} and R_{Rs} genomes in the single nucleus without structural aberrations, but at the same time experiences substantial changes in transcriptome and epigenome profiles after hybridization.

Avoidance of homoeologous interactions between A and R chromosomes

Interspecific hybridization often involves extensive homoeologous exchanges during meiosis, eventually causing nonhomologous recombination in immediate offspring (Szadkowski *et al.*, 2010, 2011; Xiong *et al.*, 2011; Zhang *et al.*, 2013; Grandont *et al.*, 2014; Chen *et al.*, 2018; Sosnowska *et al.*, 2020). We previously showed that allodiploid x*B* meiocytes rarely undergo chromosome exchanges during meiosis, as demonstrated by immunolocalization of HUMAN ENHANCER OF INVASION 10 (HEI10), a component of the ZMM complex that mediates meiotic crossover (Park *et al.*, 2020). To verify whether absence of meiotic crossover accompanied by reduction of HEI10 foci is due to the absence of nonhomologous interactions between *Br*- and *Rs*-originated chromosomes, we examined synapsis formation of meiotic chromosomes by immunolocalization of ASYNAPTIC1 (ASY1) and ZIPPER1 (ZYP1). ASY1 is the axial/lateral element of meiotic chromosomes loaded onto chromatids before synapsis (Armstrong *et al.*, 2002), and ZYP1 is the central element of synaptonemal complex present in synapsed chromosomes (Higgins *et al.*, 2005). We found that ASY1 was correctly loaded onto the entire axis of all euploid and allodiploid pachytene chromosomes at meiotic prophase I (Fig. 2). ZYP1 also co-localized with ASY1 in all euploid pachytene chromosomes (Fig. 2). Allodiploid *B. napus* (AC) produced discontinuous stretches of ZYP1 signals, indicating partial synapsis between A and C chromosomes (Fig. S7). Notably, however, ZYP1 was barely associated with allodiploid x*B* (AR) pachytene chromosomes (Fig. 2), suggesting that

(a)

(b)

(c)
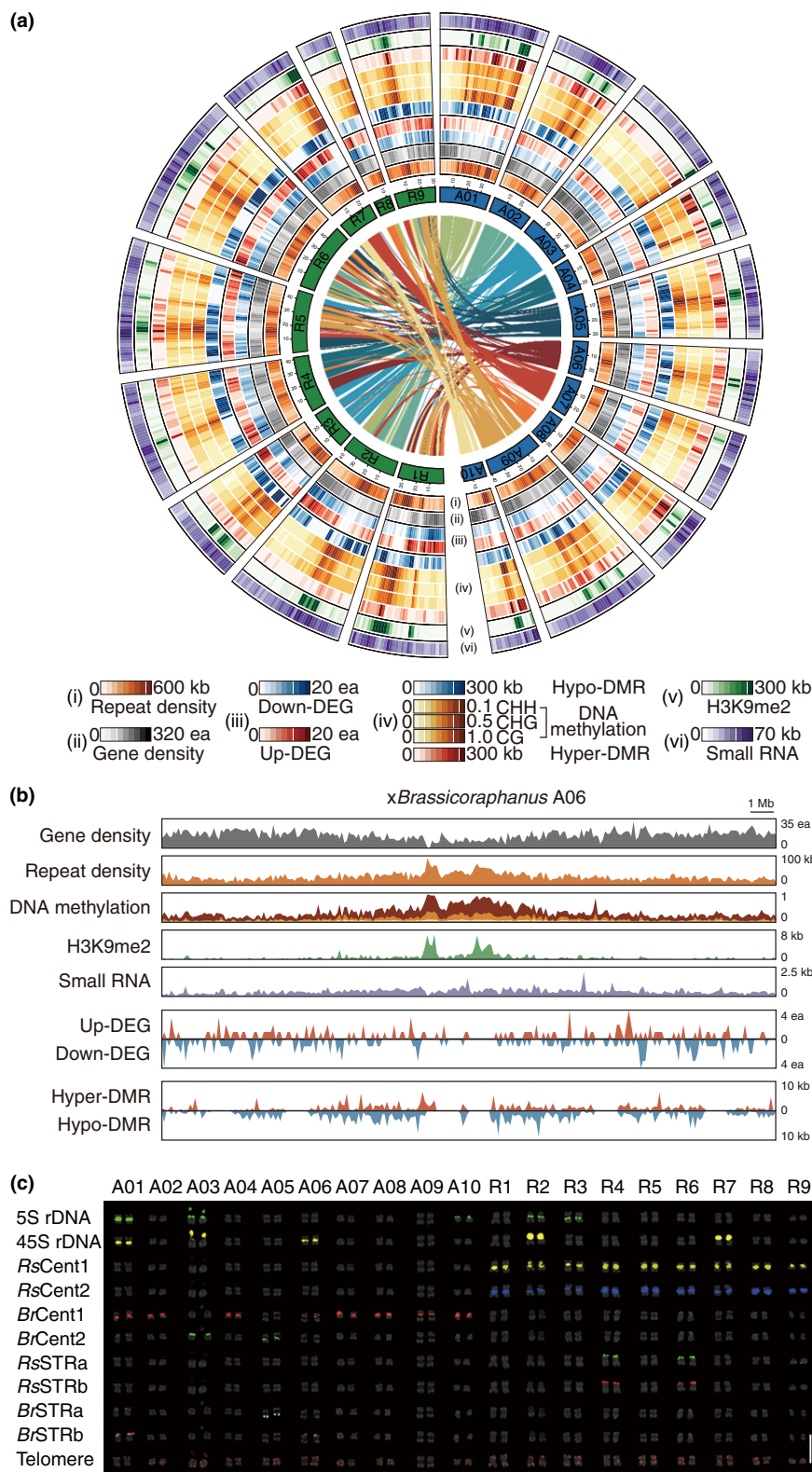
Fig. 1 Genome structure of x*Brassicoraphanus* (x*B*). (a) The x*B* genome comprises 10 A_{xB} and nine R_{xB} chromosomes. The data tracks represent (i) repeat density; (ii) gene density; (iii) differentially expressed genes (DEGs) between x*B* and its progenitor seedlings; (iv) CG, CHG and CHH methylation levels and differentially methylated regions (DMRs); (v) H3K9me2 repressive histone mark; and (vi) small RNAs. Lines in the inner circle represent syntenic relationships between A_{xB} and R_{xB}. (b) Distributions of genes, repeats, DNA methylation, H3K9me2 and small RNAs on chromosome A06 of x*B*, and DEGs and DMRs relative to A_{Br} in 100 kb bins. (c) Multicolor fluorescence *in situ* hybridization (FISH) karyograms of x*B* with specific probes for 5S rDNA, 45S rDNA, centromeric tandem repeats (Cent), short tandem repeats (STR) and telomere repeats. Bar, 10 μm.

crossover between nonhomologous chromosomes was strongly avoided in x*B* probably due to the absence of synapsis formation (Park *et al.*, 2020). These findings demonstrate that *Br* and *Rs* chromosomes share little structrual similarity, and thus orthology-dependent homoeologous interactions are prevented during meiosis while minimizing nonhomologous exchanges, which would otherwise lead to aneuploidy and/or chromosome reshuffling. This also supports our observation that both *Br* and *Rs* genomes exist in entirety without losses in allotetraploid x*B* after hybridization (Fig. 1b).

Table 1 Summary of the x*Brassicoraphanus* genome assembly.

Assembly information	Contig			Scaffold
Total length/number	652.44 Mb/68 454 ea			692.83 Mb/20 299 ea
Average/median	9.53 kb/2.40 kb			34.13 kb/901 bp
Max./Min. length	190.62 kb/200 bp			16.46 Mb/213 bp
N50	28 581 bp (6854th)			4479 746 bp (49th)
N90	5982 bp (24 969th)			166 698 bp (284th)
GC content	35.75%			33.68%

Scaffold assignment	Total number	Assigned to A_{xB} subgenome	Assigned to R_{xB} subgenome	Unassigned
No. of scaffolds	20 299	7790	7364	5145
Cumulative size (bp) (% of total assembly)	692 831 961 (100%)	335 554 805 (48.43%)	343 544 771 (49.59%)	13 732 385 (1.98%)
No. of scaffolds assigned to reference chromosomes	213	129	84	
Size of scaffolds assigned to reference chromosomes (bp) (% of total assembly)	581 691 615 (83.96%)	279 795 674 (83.38%)	301 895 941 (87.87%)	

Species	Protein-coding loci	Total CDS length (bp)	Average CDS length (bp)	Average exon length (bp)	Average intron length (bp)
x*Brassicoraphanus*	87 861	106 896 611	1216	244	196
Brassica rapa[a]	42 601	49 456 892	1172	233	209
Raphanus sativus[b]	52 326	67 790 376	1295	252	170

[a]Ensembl Plants Database (40 901 genes) with additional annotations from this study.
[b]Reannotation on the reference genome of *R. sativus* (Jeong *et al*., 2016) in this study.

Homoeologous expression adjustments in x*B*

It is assumed that speciation between *Br* and *Rs* has occurred earlier than between *Br* and *Bo*, although exact speciation timing is controversial (Mitsui *et al.*, 2015; Jeong *et al.*, 2016; Kim *et al.*, 2018). Pairwise comparison of coding sequences (CDS) of all orthologs revealed 95.7% sequence identity between *Br* and *Bo* within the same genus but 91.9% (*Br* vs *Rs*) and 92.0% (*Bo* vs *Rs*) across the genera (Fig. 3a,b). The same analysis in the tribe Camelineae also showed similar sequence divergence for interspecific (93.5% for *Arabidopsis thaliana* vs *A. lyrata*) and intergeneric (89.7% for *A. lyrata* vs *Capsella rubella*, and 90.3% for *A. thaliana* vs *C. rubella*) relationships (Fig. 3a,b). Since *c*. 8.1% sequence divergence exists between *Br* and *Rs* orthologous gene pairs, such divergence allowed us to clearly distinguish *Br*- and *Rs*-originated transcripts in x*B* (Fig. 3a,b). In the x*B* seedling transcriptome, about half of the reads (51.4%) were assigned to A_{xB} and the other half to R_{xB} (48.6%), indicating that both subgenomes contribute equally to the x*B* transcriptome (Fig. S8a). Similar proportions of A_{xB} and R_{xB} transcripts were also present in four different tissues (leaf, hypocotyl, root and flower; Fig. S8a).

Both *Br* and *Rs* genomes are retained, and thus orthologous pairs become homoeologous to each other in x*B* (Fig. 3c). Among 28 751 genes commonly expressed in *Br* and x*B*, the majority were expressed at similar levels but 2703 (9.40%) genes were differentially expressed (> 2-fold) between *Br* and x*B* seedlings (1251 upregulated DEGs (up-DEGs) and 1452 downregulated DEGs (down-DEGs) in x*B*; Fig. 3d,e). Differential

expression between *Rs* and x*B* was more prominent, with 4767 (20.96%) from 22 741 *Rs*-derived genes being dissimilarly expressed between *Rs* and x*B* seedlings (2395 up-DEGs and 2372 down-DEGs in x*B*; Fig. 3d,e). In addition, expression levels of *Br*-originated genes expressed in *Br* and x*B* seedlings were more positively correlated ($R = 0.9367$) than those of *Rs*-originated genes expressed in *Rs* and x*B* seedlings ($R = 0.8403$). These findings indicate that the majority of genes retain parental gene expression levels in x*B*, although *Br*-originated genes have a greater tendency to maintain their parental expression levels than *Rs*-originated genes. In other words, *Br* genome retains 'maintenance expression' over *Rs*, where *Br*-originated expression levels are preferentially inherited to the x*B* hybrid genome.

Of 15 376 syntenic gene pairs expressed in x*B* and its progenitors, 5701 orthologous pairs (37.07%) were differentially expressed (> 2-fold) between *Br* and *Rs* seedlings (2440 up- and 3261 down-DEGs in *Br* relative to *Rs*; Fig. 3f). This indicates that *Br* and *Rs* have distinct expression profiles for phenotypic divergence. In x*B* seedlings, however, only 3655 (23.77%) homoeologous pairs were differentially expressed (1553 up- and 2102 down-DEGs in A_{xB} relative to R_{xB}; Fig. 3f). Moreover, expression levels of A_{xB} and R_{xB} homoeologous pairs in x*B* seedlings were more highly correlated ($R = 0.8667$) than those of A_{Br} and R_{Rs} orthologous pairs between *Br* and *Rs* seedlings ($R = 0.7628$) (Fig. 3g). This suggests that distinct expressions of many orthologous genes are adjusted to similar levels in the context of homoeologous relationship in x*B*. Such expression adjustment was also observed in tissue-specific expression profiles (Fig. S8b).
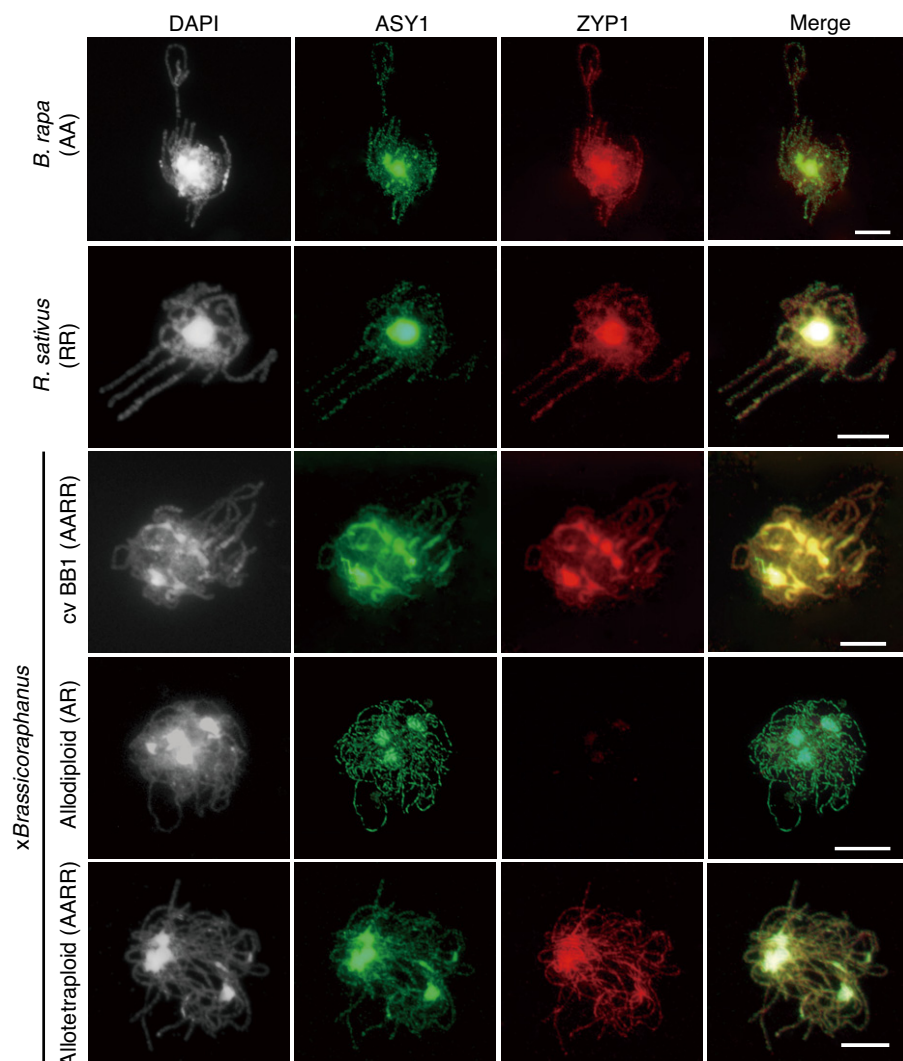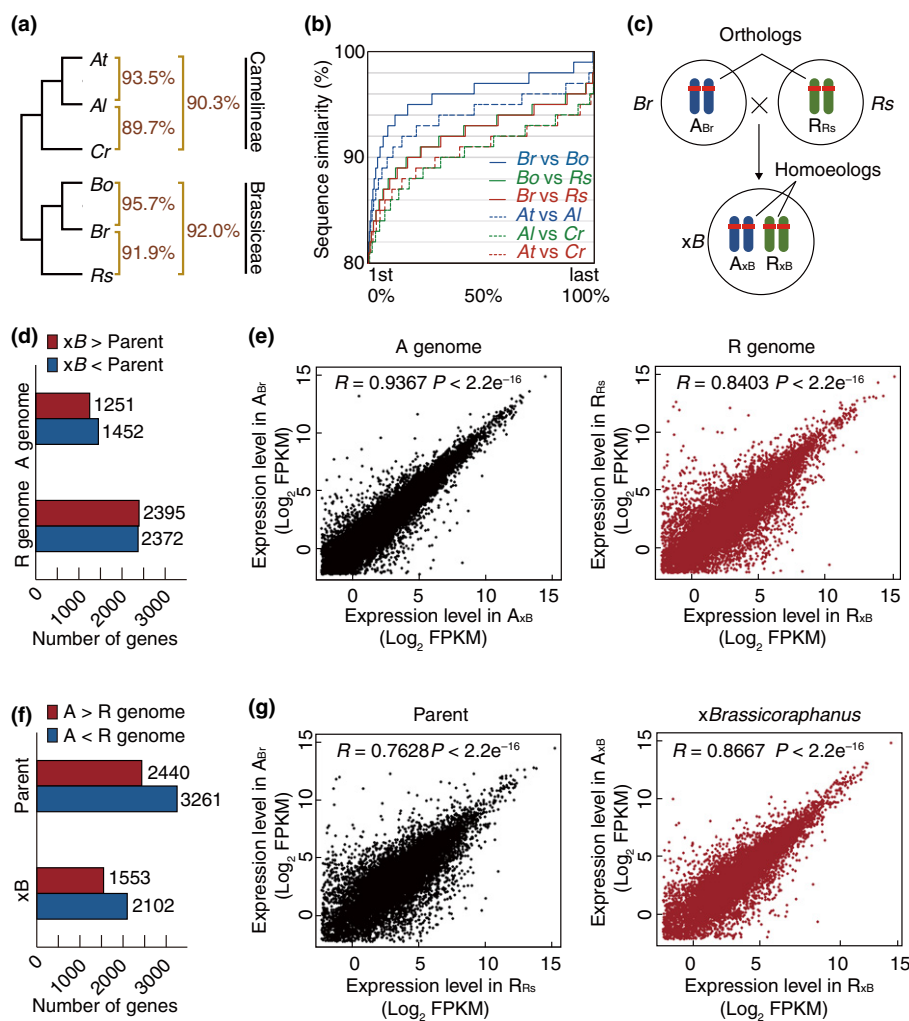
Fig. 2 Chromosome behaviors of x*Brassicoraphanus* (x*B*). Coimmunolocalization of ASYNAPTIC1 (ASY1; green) and ZIPPER1 (ZYP1; red) at pachytene in *Brassica rapa* (*Br*) (AA), *Raphanus sativus* (*Rs*) (RR), x*B* cv BB1 (AARR), and resynthesized allodiploid (AR) and allotetraploid (AARR) x*B*. Chromosomes were stained with 4′,6-diamidino-2-phenylindole (DAPI; white) and the overlay of three signals is shown (merge). Bar, 10 μm.

Reconfiguration of transcription network

Previous studies analyzed the changes of expression levels based on the sum of homoeologous pairs in allopolyploids relative to the parents, and determined additive or nonadditive expressions of duplicated genes (Rapp *et al.*, 2009; Grover *et al.*, 2012; Yoo *et al.*, 2014; Li *et al.*, 2020; Shan *et al.*, 2020; Wei *et al.*, 2021). In this study, we further investigated how orthologous pairs were adapted to a new nuclear environment by monitoring changes of expression patterns of homoeologous genes in x*B* relative to the progenitors (Fig. 4a). After filtering ambiguous patterns, 12 150 out of 15 376 syntenic homoeologous pairs were selected in *Br*, *Rs* and x*B* seedlings and classified into 27 categories of relative expressions (Fig. 4a). Among them, 7631 (62.80%) gene pairs were expressed at similar levels in every genome context, and their expressions are regarded as being 'constant' (gray in Fig. 4a). By contrast, dynamic changes were observed in 1113 (9.16%) homoeologous pairs, which were unequally expressed in x*B* and also expressed differently from at least one of the progenitor counterparts (yellow in Fig. 4a). It is notable that 1435 (11.81%) pairs showed 'biased' expressions with significant differences

between *Br* and *Rs*, while maintaining distinct progenitor expression levels in subgenomes A_{xB} and R_{xB} (blue in Fig. 4a). Interestingly, expressions of 1971 (16.22%) homoeologous pairs were adjusted to similar levels in x*B*, although their expressions were different between A_{Br} and R_{Rs} progenitors (red in Fig. 4a). Such 'convergent' expressions were more prominent for R_{Rs}-originated genes (1483/1971). We assumed that 'convergent' expressions might result from similar *cis*-regulatory sequences between homoeologous pairs under the same transcriptional control in x*B*. We analyzed sequence similarities between homoeologous gene pairs of the categories of 'convergent' vs 'biased' expression (Fig. 4b). CDSs of both 'convergent' and 'biased' homoeologous pairs showed a high level of sequence identity (92.54% vs 92.00%; Fig. 4b). By contrast, the upstream *cis*-elements were notably divergent between homoeologous pairs. Interestingly, 'convergent' homoeologous pairs shared less diverged *cis*-element sequences than 'biased' ones (68.42% vs 63.29%; Fig. 4b). These findings support our hypothesis that the upstream regulatory sequences of the orthologs have diverged after speciation but retain essential *cis*-elements that are probably under control of the same *trans*-acting regulators in x*B*. This also suggests that

Fig. 3 Transcriptome changes in xBrassicoraphanus (xB). (a) Phylogenetic relationship and sequence divergence in tribes Camelineae and Brassiceae. Percentages between species represent their coding sequence similarity of orthologous gene pairs. At, Arabidopsis thaliana; Al, A. lyrata; Cr, Capsella rubella; Bo, Brassica oleracea; Br, B. rapa; Rs, Raphanus sativus. (b) Distribution of sequence similarities of interspecific/intergeneric orthologs. Horizontal axis indicates orthologous gene pairs sorted in ascending order of sequence similarity. (c) Relationship between orthologous and homoeologous genes in progenitors and xB. (d) Number of differentially expressed genes (DEGs) in xB relative to the progenitors (A_{Br} vs A_{xB} and R_{Rs} vs R_{xB}). (e) Scatter plots comparing gene expression levels between A_{Br} and A_{xB} (black), and R_{Rs} and R_{xB} (red). (f) Number of DEGs of orthologous pairs between A_{Br} and R_{Rs}, and homoeologous pairs between A_{xB} and R_{xB}. (g) Scatter plots comparing gene expression levels between A_{Br} and R_{Rs} (black), and A_{xB} and R_{xB} (red).

both A and R genomes still maintain compatibility in the transcription system to prevent a 'transcriptome shock' (Hegarty *et al.*, 2006; Buggs *et al.*, 2011), but divergence in regulatory elements should entail reconfiguration of the overall expression network in the hybrid genome of xB.

Coordinated expression of homoeologous genes in response to external stimuli

Gene Ontology enrichment analysis was performed for three categories of homoeologous expression – 'constant', 'biased' and 'convergent'. 'Constant' homoeologous pairs have enrichment for GO terms such as 'cell differentiation', 'developmental cell growth' and 'cell cycle' (Fig. 5a; Table S12), suggesting that cell function-related genes maintain consistent expression patterns after hybridization. However, the 'biased' homoeologous pairs did not display GO enrichment for specific functions ($P > 0.001$). Notably, the 'convergent' homoeologous pairs had GO enrichment for diverse responses such as 'response to hormone', 'response to stress', 'response to biotic stimulus' and 'response to abiotic stimulus' (Fig. 5a; Table S12). This suggests

that the homoeologous pairs coordinately expressed in response to various stimuli tend to have similar *cis*-elements, although they are distinctly expressed in the progenitors. Moreover, the motifs of stress-responsive *cis*-elements such as abscisic acid-responsive element (ABRE; BACGTGK, B = C, G or T; K = G or T) (Lieberman-Lazarovich *et al.*, 2019) and dehydration-responsive element/C-repeat element (DRE/CRT; RCCGAC, R = A or G) (Suzuki *et al.*, 2005) were found abundantly in the upstream sequence of 'convergent' homoeologous pairs (Fig. 5b). This indicates that the genes involved in cellular signaling may require essential *cis*-elements to properly respond to external stimuli.

We treated 14-d-old seedlings of Br, Rs and xB to cold for 4 wk and monitored expression changes of orthologous/homoeologous genes. Of 15 376 orthologs, 1579 genes were differentially regulated by cold in Br seedlings, with 956 up-DEGs and 623 down-DEGs (Fig. 5c). In cold-treated Rs seedlings, 2378 genes were differentially expressed, with 1093 up-DEGs and 1285 down-DEGs (Fig. 5c). Among them, only small fractions of orthologous genes (182 up- and 91 down-DEGs; 9.75% and 5.01%) were similarly regulated in both Br and Rs (Fig. 5c). In xB seedlings, 2657 genes were differentially regulated by cold
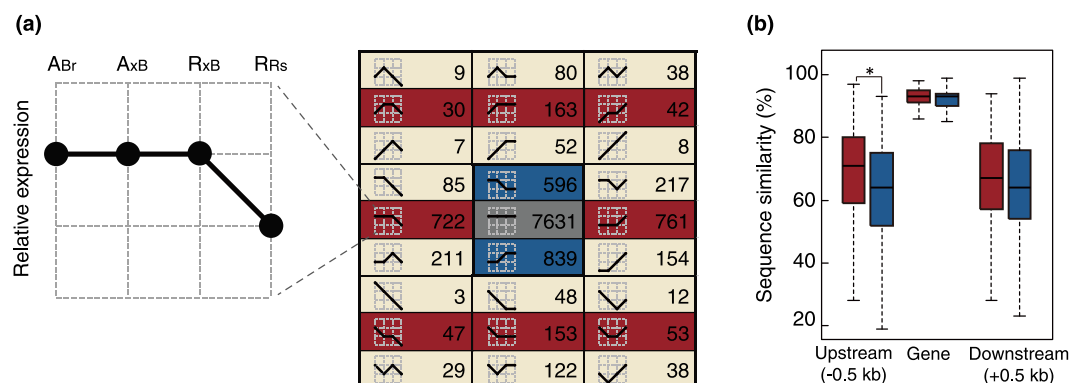
Fig. 4 Expression patterns of homoeologous pairs in x*Brassicoraphanus* (x*B*). (a) Classification of expression patterns of homoeologs in the x*B* relative to progenitor orthologs. The gray, blue and red blocks represent gene pairs showing 'constant', 'biased' and 'convergent' expression, respectively. (b) Sequence similarities of genic and adjacent upstream/downstream regions of orthologous genes showing convergent (red) and biased (blue) expression in x*B* subgenomes. The median value of sequence similarity is represented as a horizontal line inside the box, and the boxplot whiskers denote the range of values (Wilcoxon's rank-sum test: *, $P < 2.2e^{-10}$).

treatment. Specifically, 1431 *Br*-derived orthologs were differentially expressed (661 up-DEGs and 770 down-DEGs in A_{xB}) and 1226 *Rs*-derived orthologs were differently regulated (562 up-DEGs and 664 down-DEGs in R_{xB}) after cold treatment (Fig. 5c). Notably, a larger fraction (261 up- and 378 down-DEGs; 27.13% and 35.80%) of A_{xB} and R_{xB} homoeologous pairs were identified as common DEGs in x*B* (Fig. 5c). These observations indicate that many orthologous/homoeologous pairs are distinctly regulated in *Br* and *Rs* progenitors but their expressions are systematically coordinated in the x*B* hybrid genome in response to cold exposure. We also found that expressions of A_{Br} and R_{Rs} orthologous genes had a weak correlation regardless of expression categories (Fig. 5d). Interestingly, A_{xB} and R_{xB}

'convergent' homoeologous pairs showed a strong correlation ($R = 0.620$), whereas 'biased' ones did not ($R = 0.195$) (Fig. 5e). These data suggest that evolutionarily divergent homoeologous pairs still share essential motifs in *cis*-elements that can be subjected to the same *trans*-acting regulation, conceivably responsible for coordinated expression in response to environmental cues in hybrids.

Silencing of TEs stabilizes the x*B* hybrid genome

Resynthesized hybrids often experience epigenetic alterations (Greaves *et al.*, 2015). We investigated methylation profiles in coding genes and repeat regions. In coding regions, DNA
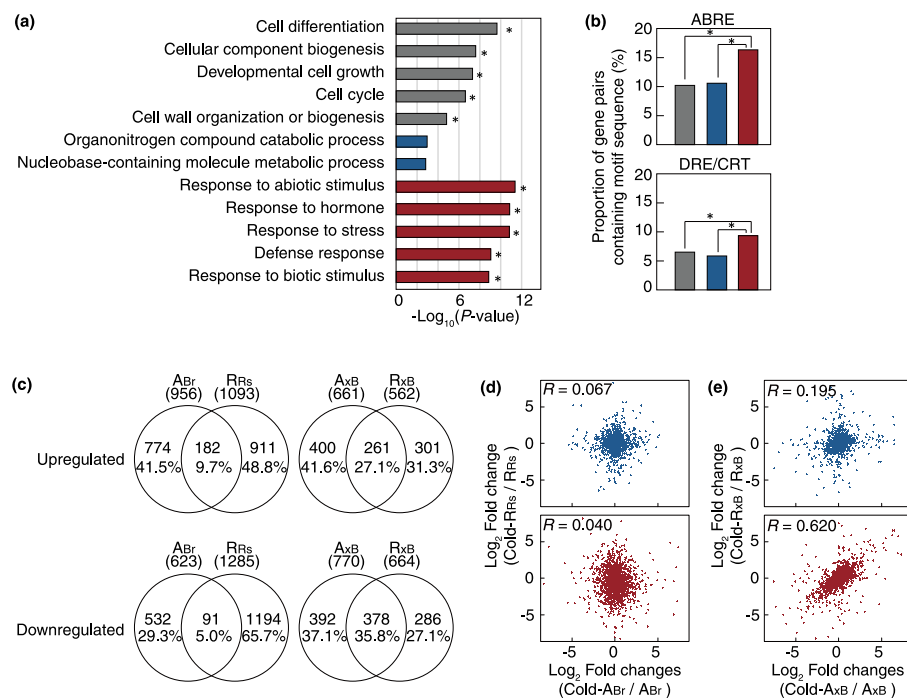
Fig. 5 Expression of homoeologous genes in response to external stimuli. (a) Gene ontology enrichments of 'constant' (gray), 'biased' (blue) and 'convergent' (red) homoeologous pairs (Fisher's exact test: *, $P < 0.001$). (b) Proportion of 'constant' (gray), 'biased' (blue) and 'convergent' (red) homoeologous pairs containing conserved sequences of abscisic acid-responsive element (ABRE) and dehydration-responsive element/C-repeat element (DRE/CRT) (Fisher's exact test: *, $P < 0.001$). (c) Venn diagram of cold-induced differentially expressed genes between A_{Br} and R_{Rs} orthologs (left) and between A_{xB} and R_{xB} homoeologs (right). (d) Scatter plots of cold-induced expression changes of A_{Br} and R_{Rs} orthologous genes showing 'biased' (blue) and 'convergent (red)' expressions. (e) Scatter plots of cold-induced expression changes of A_{xB} and R_{xB} homoeologous genes showing 'biased' (blue) and 'convergent' (red) expressions.

methylation levels are high in the gene body, decrease towards the 5′ and 3′ ends, and increase again beyond translation start and termination sites in all Br, Rs and xB seedlings (Fig. 6a). Notably, A_{Br} and R_{Rs} progenitor genomes have distinct CG methylation patterns in coding genes, with A_{Br} being more densely methylated than R_{Rs}. This methylation asymmetry is inherited to A_{xB} and R_{xB} subgenomes (Fig. 6a). TEs are heavily methylated in general, especially near-complete CG methylation in all species (Fig. 6b). TEs also have higher CHG and CHH methylation levels than coding genes. Interestingly, Br and Rs TEs have distinct CHG methylation profiles, with more CHG methylation at Rs TEs (Fig. 6b). However, such asymmetry is abolished in xB, where Br-derived TEs have an increased CHG methylation level comparable to Rs-derived TEs (Fig. 6b). This suggests that TEs from Br acquired more CHG methylation after hybridization possibly via $trans$-acting mechanisms. We analyzed small RNAs in Br, Rs and xB seedlings, and found that c. 30–50% of small RNAs were 24 nt RNAs as potential short-interfering RNAs (siRNAs) (Fig. S9a). siRNAs were highly associated with hyper-DMRs in xB but loosely with hypo-DMRs, indicating a strong correlation between 24 nt RNA and DNA methylation (Fig. 6c). About 12% of 24- nt RNAs from Br and

Rs have a pairwise sequence identity and may share the same targets across the genomes (Fig. S9b). Indeed, 10.4% of 24 nt RNAs from xB also have indistinguishable origins (Fig. S9c). This suggests that, in the xB hybrid genome, R_{xB}-originated siRNAs induce gain of CHG methylation at TEs on A_{xB} possibly via RNA-directed DNA methylation (RdDM) (Law & Jacobsen, 2010). DNA transposons are widespread throughout the xB genome with little association with DMRs (Fig. 6d). LTRs that account for c. 30% of repeats (Table S8) were also heavily methylated. Notably, it was clear that LTRs on A_{xB} had higher methylation levels at the CHG context than A_{Br} (Figs 6d, S10, S11). This suggests that DNA methylation profiles have changed in a subgenome-specific manner, for which R_{xB}-originated siRNAs might induce gain of CHG methylation in $trans$ at LTRs of the same kind on A_{xB}. As exemplified in Fig. 6(e), the $Gypsy$ element on A_{xB} was found to have higher CHG methylation levels than A_{Br} at the scaffold level, although CG and CHH methylation levels are nearly identical. Northern blot analysis verified that $Copia$ and $Gypsy$ elements were moderately expressed in Br but silenced in xB seedlings (Fig. 6f). These findings suggest that RdDM-mediated DNA methylation induces TE silencing across subgenomes, which in turn stabilizes the xB hybrid genome.
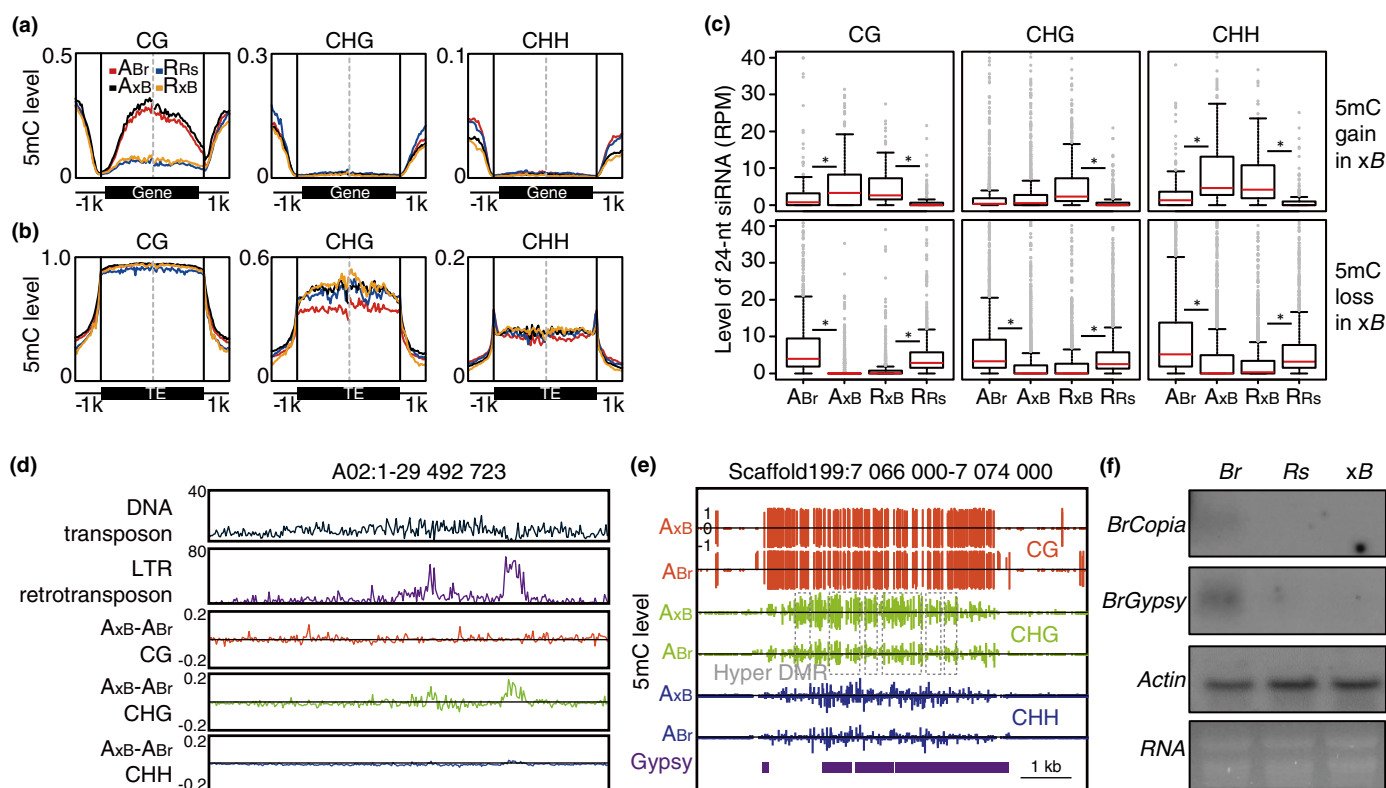

Fig. 6 Relationships between DNA methylation, small RNA and transposable element (TE) expression in x$Brassicoraphanus$ (xB). Distribution of DNA methylation at gene body (a) and TE regions (b) in xB subgenomes (A_{xB} and R_{xB}) and its progenitor genomes (A_{Br} and R_{Rs}). (c) Expression levels of 24 nt RNAs at CG, CHG and CHH differentially methylated regions in xB subgenomes (A_{xB} and R_{xB}) and the progenitor genomes (A_{Br} and R_{Rs}). The expression level of 24 nt RNAs was calculated as reads per million (RPM). The median value of sequence similarity is represented as a horizontal line inside the box, and the boxplot whiskers denote the range of values (two-tailed Student's t-test: *, $P < 5.0e^{-5}$). (d) Distributions of DNA transposons, long terminal repeats (LTRs) and DNA methylation difference between A_{Br} and A_{xB} across chromosome A02 in 100 kb bins. (e) An example of methylation distributions at hyper-methylated $Gypsy$ class LTRs in A_{xB} and A_{Br}. (f) Northern blot for $BrCopia$ and $BrGypsy$. Actin was used as a loading control.

Discussion

Hybridization barriers serve as a mechanism to prevent gene flow between species (Abbott *et al.*, 2013). In particular, the postzygotic hybridization barrier after fertilization is often manifested as hybrid inviability or sterility (Dion-Cote & Barbash, 2017). Hybrid sterility is generally associated with a failure in meiosis. Normal meiosis requires the formation of synapsis between homologous chromosome pairs, but when they are abolished or formed between multiple and/or nonhomologous chromosomes, the chromosomes segregate abnormally, resulting in sterile gametes and aneuploidy (Martinez-Perez & Colaiacovo, 2009). Aneuploidy and/or chromosome rearrangements are frequently observed in resynthesized allopolyploids between closely related species (Xiong *et al.*, 2011; Zhang *et al.*, 2013; Chen *et al.*, 2018). This is mainly caused by the collinearity/homology between less divergent parental chromosomes. For instance, A1/C1, A2/C2 and parts of A5/C4 (A from *Br* and C from *Bo*) chromosomes are homologous to each other (Parkin *et al.*, 2005), and most phenotypic variations and aneuploidy in resynthesized *B. napus* lines are caused by homoeologous interactions, mostly between nonhomologous chromosomes (Gaeta *et al.*, 2007; Xiong *et al.*, 2011; Grandont *et al.*, 2014). However, the presence of full compliments of both *Br* and *Rs* chromosomes in x*B* demonstrates that a merger of divergent genomes may avoid such harmful interactions, while producing fertile gametes after polyploidization.

The Dobzhansky–Muller model proposes that hybrid incompatibility arises from negative epistatic interactions between the incompatible alleles in nascent hybrids (Moyle & Nakazato, 2010). This may operate as a primary postzygotic barrier, which also appears true for early generations of x*B*. Most synthetic F₁ individuals of x*B*, including resynthesized x*B* plants in this study, display reproductive defects such as pollen inviability and seed abortion (Lee *et al.*, 2011; Shin *et al.*, 2021). Underdeveloped embryos must be *in vitro* rescued, and thus only a handful of progenies can be finally fixed as stable lines. During this stabilization process over multiple generations, the x*B* hybrid genome is assumed to have undergone constant transgenerational changes in genome structure and epigenetic landscape until the genome environment settles down to a stable condition. Consequently, deleterious or incompatible alleles have possibly been removed by homoeologous exchanges or silenced by epigenetic mechanisms.

Although *Br* and *Rs* genomes are estimated to have diverged 7–22.4 Ma (Mitsui *et al.*, 2015; Jeong *et al.*, 2016; Kim *et al.*, 2018), we cannot rule out the possibility that the two genomes still retain syntenic regions similar enough to trigger homoeologous exchanges. We previously reported that some genomic regions of *Br* and *Rs* share structural similarity, particularly between A8 and R8, although most other regions are fragmented and diversified (Park *et al.*, 2020). In addition, we recently showed that several x*B* lines derived from the same intergeneric cross endured chromosome instability and pollen inviability, while occasionally producing tetrads with an extra micronucleus during microsporogenesis (Shin *et al.*, 2021). This is probably a consequence of homoeologous chromosome exchanges at meiotic prophase I leading to the formation of univalent or multivalent

chromosomes (Shin *et al.*, 2021). This suggests that chromosome instability persists in successive generations while generating non-functional haploid gametes with aberrant karyotypes. Such chromosome rearrangement may occur in a stochastic manner involving diverse combinations of nonspecific homoeologous interactions, and we presume that once the reorganized genome acquires a stable state by losing a sticky region that is prone to constantly interact with a wrong counterpart, the descendants are able to secure genome stability and finally survive.

Hybridization between species inevitably entails changes in *cis–trans* interactions bringing about alterations in the transcription network (Hu & Wendel, 2019). Therefore, extensive changes in parental expression profiles are expected, and when such changes are intolerable, the hybrid will undergo a 'transcriptome shock', manifested as hybrid dysgenesis (Martienssen, 2010) or outcrossing depression (Frankham *et al.*, 2011). x*B* experienced moderate expression changes of progenitor genes after hybridization but still maintains a transcription network between subgenomes compatible enough to generate novel or intermediate phenotypes. Our four-point expression analysis revealed that 'convergent' homoeologs share similar *cis*-elements, and expression levels of a larger fraction of *Rs*-derived homoeologs were adjusted to *Br*-derived ones. This suggests that *Br*-originated *trans*-acting factors probably play dominant roles for coregulation of homoeologous pairs in x*B* (Hu & Wendel, 2019). Notably, stress response-related motifs are enriched in the *cis*-elements of 'convergent' homoeologs, suggesting that transcriptional regulation is primarily mediated by *trans*-acting factors sharing common homoeologous targets that are involved in diverse responses. Such reconfiguration of the transcription network is conceivably crucial to the adaptation of newly synthesized hybrids.

Br has higher gene-body CG methylation levels than *Rs*, which is inherited to each subgenome in x*B*. This indicates that differential gene-body methylation is maintained after hybridization and this methylation asymmetry may contribute to 'maintenance expression' of A$_{xB}$ through unknown mechanisms. TEs are heavily methylated in general, but also showed asymmetric CHG methylation between *Br* and *Rs*. Intriguingly, *Br*-originated LTRs gained CHG methylation comparable to *Rs* LTRs in x*B*, suggesting that repeat-originated siRNAs trigger hypermethylation via RdDM *in trans* and TE silencing (Wendel *et al.*, 2016). This may prevent hyperactivation of TEs and subsequent genome destabilization, which would otherwise culminate in a 'genomic shock' as initially proposed by McClintock (1984).

Many previous studies report that global DNA methylation changes occur in allopolyploid plants such as *Arabidopsis suecica* (Madlung *et al.*, 2002), *B. napus* (Xu *et al.*, 2009), wheat (Shaked *et al.*, 2001) and rice (Li *et al.*, 2019). Importantly, even autotetraploid rice plants show an increase in CHG and CHH methylation levels at TEs associated with 24 nt siRNA clusters (Zhang *et al.*, 2015), suggesting the role of RdDM for gain of DNA methylation and silencing of TEs and nearby genes. The current model of RdDM delineates the establishment of CHH methylation triggered by repeat-originated small RNAs (Law & Jacobsen, 2010). However, the regulation of CHG methylation via RdDM is largely elusive. One possible mechanism of CHG

hypermethylation at TEs is that DNA methyltransferase such as chromomethyl transferase 3 is also under the control of RdDM, which may utilize the RdDM components to access either directly or indirectly to targets for CHG methylation. Another is that CHH-methylated regions may provide a unique chromatin signature for subsequent CHG methylation and maintenance of TE silencing. A recent study proposed that silencing of distal LTRs in *Arabidopsis* and tomato requires an additional phase of CMT-mediated CHG methylation following RdDM-mediated CHH methylation (Wang & Baulcombe, 2020). A similar mechanism might have worked in multiple steps for silencing of *Br*-originated LTRs in x*B*, for which siRNAs derived from *Rs* LTRs conceivably play crucial roles for gain of CHG methylation *in trans*.

It is believed that the more distantly related the species, the stronger the hybridization barrier. Contrary to this assumption, our findings strongly suggest that, as long as the physiology and transcriptional regulatory networks are compatible, a certain extent of genome divergence promotes hybridization between distant species. Therefore, a trade-off between genome divergence and transcriptome compatibility is meaningful to facilitate hybridization between species without causing genome destabilization and/or a conflict in the transcription network. This concept also proposes that interspecific/intergeneric hybridization may occur more frequently in nature than previously thought, and the 'triangle of U' model (Nagaharu, 1935) can be further expanded to the intergeneric level.

After whole genome duplication or hybridization between the different species followed by chromosome doubling (allopolyploidization), polyploid plants generally undergo gradual but substantial genome reconstruction, including massive chromosome rearrangement, differential deletion or retention of duplicated genes and biased genome fragmentation (Cheng *et al.*, 2018). This eventually leads to a decrease in both chromosome number and genome size, with most of the polyploid properties being lost. Extensive changes in genome structure and gene repertoire accompanied by subfunctionalization/neofunctionalization of duplicated genes also contribute to the formation of new species with novel phenotype and function, which sometimes outperform the diploid progenitors with greater ecological fitness. Thus, the evolution of land plants, especially the angiosperms, is not a one-way process. Rather, it is likely to comprise the recurrent cycles of hybridization, diversification, diploidization and reunification among species in the same lineage (Wendel, 2015). Furthermore, understanding the highly dynamic and flexible process of hybridization and polyploidization should provide a clue to Darwin's 'abominable mystery' (Darwin, 1903) questioning the great diversification and expansion of angiosperms within a short geological timeframe.

Acknowledgements

Author contributions

JHH conceived the project. HS, JEP, HRP and JHH designed the study. HS, JEP, WLC, HYS and GY performed molecular biology experiments and analyzed the data. JEP performed FACS analysis. HRP and WK performed immunolocalization experiments. HS, SHY, SK, JHA and J-SK performed bioinformatics analysis. NEW, HRB and S Perumal performed FISH analysis. Y-MK and NK performed annotation analysis. KK and T-JY analyzed chloroplast genomes. S Park, JAK, YPL and S-SL provided plant materials. HS, JEP, HRP, WLC, SHY, WK, HYS, JYL, GY, TK, JK, HJ, DHK, YSK, H-MJ, JY and SS prepared plant materials. WLC, SHY, JYL, B-SP, T-FH, T-JY, DC, HHK and S-SL commented on the manuscript. HS, JEP, HRP and JHH wrote the manuscript with help from all coauthors.

ORCID

Doil Choi https://orcid.org/0000-0002-4366-3627
Woo Lee Choi https://orcid.org/0000-0002-5875-8434
Tzung-Fu Hsieh https://orcid.org/0000-0001-7584-3721
Jin Hoe Huh https://orcid.org/0000-0002-9367-9280
Hyun Hee Kim https://orcid.org/0000-0002-2422-643X
June-Sik Kim https://orcid.org/0000-0002-4703-609X
Seungill Kim https://orcid.org/0000-0001-6287-8693
Yong-Min Kim https://orcid.org/0000-0001-9210-5843
Soo-Seong Lee https://orcid.org/0000-0002-6564-0944
Joo Young Lim https://orcid.org/0000-0001-7575-2125
Yong Pyo Lim https://orcid.org/0000-0003-2978-9378
Jeong Eun Park https://orcid.org/0000-0003-4579-1179
Sampath Perumal https://orcid.org/0000-0003-3403-1625
Hosub Shin https://orcid.org/0000-0001-6488-8430
Nomar Espinosa Waminal https://orcid.org/0000-0001-5119-152X
Tae-Jin Yang https://orcid.org/0000-0002-9676-8801
Gibum Yi https://orcid.org/0000-0002-7630-1892

Data availability

The sequencing data for genomic, transcriptomic and epigenomic analyses are available as BioProject ID PRJNA353741, PRJNA353738, PRJNA394950 and PRJNA353316. The assembled x*Brassicoraphanus* genome is available as BioProject ID PRJNA353741. The chloroplast genome of x*B* is deposited in GenBank under accession no. MN928713.

References

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

Alexa A, Rahnenfuhrer J. 2010. TOPGO: enrichment analysis for gene ontology. R package v.2.22. doi: 10.8129/B9.bioc.topGO.

Anders S, Pyl PT, Huber W. 2015. HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**: 166–169.

Armstrong SJ, Caryl AP, Jones GH, Franklin FC. 2002. Asy1, a protein required for meiotic chromosome synapsis, localizes to axis-associated chromatin in *Arabidopsis* and *Brassica*. *Journal of Cell Science* **115**: 3645–3655.

Axtell MJ. 2013. SHORTSTACK: comprehensive annotation and quantification of small RNA genes. *RNA* **19**: 740–751.

Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**: 578–579.

Buggs RJA, Zhang LJ, Miles N, Tate JA, Gao L, Wei W, Schnable PS, Barbazuk WB, Soltis PS, Soltis DE. 2011. Transcriptomic shock generates evolutionary novelty in a newly formed, natural allopolyploid plant. *Current Biology* **21**: 551–556.

Chelysheva LA, Grandont L, Grelon M. 2013. Immunolocalization of meiotic proteins in Brassicaceae: method 1. *Methods in Molecular Biology* **990**: 93–101.

Chen S, Ren F, Zhang L, Liu Y, Chen X, Li Y, Zhang L, Zhu B, Zeng P, Li Z et al. 2018. Unstable allotetraploid tobacco genome due to frequent homeologous recombination, segmental deletion, and chromosome loss. *Molecular Plant* **11**: 914–927.

Cheng F, Wu J, Cai X, Liang JL, Freeling M, Wang XW. 2018. Gene retention, fractionation and subgenome differences in polyploid plants. *Nature Plants* **4**: 258–268.

Darwin C. 1903. In: Darwin F, Seward AC, eds. *More letters of Charles Darwin: a record of his work in a series of hitherto unpublished letters, vol. 2, Chapter VII.* New York, NY, USA: Appleton & Co., 515.

Dion-Cote AM, Barbash DA. 2017. Beyond speciation genes: an overview of genome stability in evolution and speciation. *Current Opinion in Genetics & Development* **47**: 17–23.

Dolstra O. 1982. *Synthesis and fertility of Brassicaraphanus and ways of transferring Raphanus characters to Brassica.* PhD thesis, Wageningen University & Research, Wageningen, the Netherlands.

Frankham R, Ballou JD, Eldridge MD, Lacy RC, Ralls K, Dudash MR, Fenster CB. 2011. Predicting the probability of outbreeding depression. *Conservation Biology* **25**: 465–475.

Freese NH, Norris DC, Loraine AE. 2016. Integrated genome browser: visual analytics platform for genomics. *Bioinformatics* **32**: 2089–2095.

Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC. 2007. Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. *Plant Cell* **19**: 3403–3417.

Grandont L, Cunado N, Coriton O, Huteau V, Eber F, Chevre AM, Grelon M, Chelysheva L, Jenczewski E. 2014. Homoeologous chromosome sorting and progression of meiotic recombination in *Brassica napus*: ploidy does matter! *Plant Cell* **26**: 1448–1463.

Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**: 1017–1018.

Greaves IK, Gonzalez-Bayon R, Wang L, Zhu A, Liu PC, Groszmann M, Peacock WJ, Dennis ES. 2015. Epigenetic changes in hybrids. *Plant Physiology* **168**: 1197–1205.

Grover CE, Gallagher JP, Szadkowski EP, Yoo MJ, Flagel LE, Wendel JF. 2012. Homoeolog expression bias and expression level dominance in allopolyploids. *New Phytologist* **196**: 966–971.

Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EVIDENCEMODELER and the program to assemble spliced alignments. *Genome Biology* **9**: R7.

Hegarty MJ, Barker GL, Wilson ID, Abbott RJ, Edwards KJ, Hiscock SJ. 2006. Transcriptome shock after interspecific hybridization in *Senecio* is ameliorated by genome duplication. *Current Biology* **16**: 1652–1659.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Molecular Cell* **38**: 576–589.

Higgins JD, Sanchez-Moran E, Armstrong SJ, Jones GH, Franklin FC. 2005. The *Arabidopsis* synaptonemal complex protein ZYP1 is required for chromosome synapsis and normal fidelity of crossing over. *Genes & Development* **19**: 2488–2500.

Hu G, Wendel JF. 2019. *Cis-trans* controls and regulatory novelty accompanying allopolyploidization. *New Phytologist* **221**: 1691–1700.

Huang H, Tong Y, Zhang QJ, Gao LZ. 2013. Genome size variation among and within *Camellia* species by using flow cytometric analysis. *PLoS ONE* **8**: e64981.

Inomata N. 1977. Production of interspecific hybrids between *Brassica campestris* and *Brassica oleracea* by culture in vitro of excised ovaries: I. Effects of yeast extract and casein hydrolysate on the development of excised ovaries. *Japanese Journal of Breeding* **27**: 295–304.

Jeong Y-M, Kim N, Ahn BO, Oh M, Chung W-H, Chung H, Jeong S, Lim K-B, Hwang Y-J, Kim G-B et al. 2016. Elucidating the triplicated ancestral genome structure of radish based on chromosome-level comparison with the *Brassica* genomes. *Theoretical and Applied Genetics* **129**: 1357–1372.

Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, Yabana M, Harada M, Nagayasu E, Maruyama H et al. 2014. Efficient *de novo* assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Research* **24**: 1384–1395.

Karpechenko GD. 1928. Polyploid hybrids of *Raphanus sativus* L. x*Brassica oleracea* L. *Zeitschrift für Induktive Abstammungs- und Vererbungslehre* **48**: 1–85.

Kim C-K, Seol Y-J, Perumal S, Lee J, Waminal NE, Jayakodi M, Lee S-C, Jin S, Choi B-S, Yu Y et al. 2018. Re-exploration of U's triangle *Brassica* species based on chloroplast genomes and 45S nrDNA sequences. *Scientific Reports* **8**: 7353.

Kim JS, Lim JY, Shin H, Kim BG, Yoo SD, Kim WT, Huh JH. 2019. ROS1-dependent DNA demethylation is required for ABA-inducible *NIC3* expression. *Plant Physiology* **179**: 1810–1821.

Kim S, Park M, Yeom S-I, Kim Y-M, Lee JM, Lee H-A, Seo E, Choi J, Cheong K, Kim K-T et al. 2014. Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. *Nature Genetics* **46**: 270.

Kozomara A, Griffiths-Jones S. 2013. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research* **42**: D68–D73.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. CIRCOS: an information aesthetic for comparative genomics. *Genome Research* **19**: 1639–1645.

Lafon-Placette C, Kohler C. 2015. Epigenetic mechanisms of postzygotic reproductive isolation in plants. *Current Opinion in Plant Biology* **23**: 39–44.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**: R25.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R et al. 2007. CLUSTAL W and CLUSTAL X v.2.0. *Bioinformatics* **23**: 2947–2948.

Law JA, Jacobsen SE. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics* **11**: 204–220.

Lee J, He K, Stolc V, Lee H, Figueroa P, Gao Y, Tongprasit W, Zhao H, Lee I, Deng XW. 2007. Analysis of transcription factor HY5 genomic binding sites revealed its hierarchical role in light regulation of development. *Plant Cell* **19**: 731–749.

Lee S-S, Lee S-A, Yang J, Kim J. 2011. Developing stable progenies of ×*Brassicoraphanus*, an intergeneric allopolyploid between *Brassica rapa* and *Raphanus sativus*, through induced mutation using microspore culture. *Theoretical and Applied Genetics* **122**: 885–891.

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.

Lieberman-Lazarovich M, Yahav C, Israeli A, Efroni I. 2019. Deep conservation of *cis*-element variants regulating plant hormonal responses. *Plant Cell* **31**: 2559–2572.

Lohse M, Drechsel O, Kahlau S, Bock R. 2013. OrganellarGenomeDRAW-a suite of tools for generating physical maps of plastid and mitochondrial

genomes and visualizing expression data sets. *Nucleic Acids Research* 41: W575–W581.

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y. 2012. SOAPDENOVO2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaScience* 1: 18.

Madlung A, Masuelli RW, Watson B, Reynolds SH, Davison J, Comai L. 2002. Remodeling of DNA methylation and phenotypic and transcriptional changes in synthetic *Arabidopsis* allotetraploids. *Plant Physiology* 129: 733–746.

Marcais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27: 764–770.

Martienssen RA. 2010. Heterochromatin, small RNA and post-fertilization dysgenesis in allopolyploid and interploid hybrids of *Arabidopsis*. *New Phytologist* 186: 46–53.

Martin M. 2011. CUTADAPT removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17: 10–12.

Martinez-Perez E, Colaiacovo MP. 2009. Distribution of meiotic recombination events: talking to your neighbors. *Current Opinion in Genetics & Development* 19: 105–112.

McClintock B. 1984. The significance of responses of the genome to challenge. *Science* 226: 792–801.

Mcnaughton IH. 1973. Synthesis and sterility of *Raphanobrassica*. *Euphytica* 22: 70–88.

Mitsui Y, Shimomura M, Komatsu K, Namiki N, Shibata-Hatta M, Imai M, Katayose Y, Mukai Y, Kanamori H, Kurita K et al. 2015. The radish genome and comprehensive gene expression profile of tuberous root formation and development. *Scientific Reports* 5: 10835.

Moyle LC, Nakazato T. 2010. Hybrid incompatibility "snowballs" between *Solanum* species. *Science* 329: 1521–1523.

Nagaharu U. 1935. Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Japanese Journal of Botany* 7: 389–452.

Nawrocki EP, Burge SW, Bateman A, Daub J, Eberhardt RY, Eddy SR, Floden EW, Gardner PP, Jones TA, Tate J et al. 2014. RFAM 12.0: updates to the RNA families database. *Nucleic Acids Research* 43: D130–D137.

Oost E. 1984. x*Brassicoraphanus* Sageret or x*Raphanobrassica* Karpechenko? *Cruciferae Newsletter* 9: 11–12.

Park HR, Park JE, Kim JH, Shin H, Yu SH, Son S, Yi G, Lee S-S, Kim HH, Huh JH. 2020. Meiotic chromosome stability and suppression of crossover between non-homologous chromosomes in x*Brassicoraphanus*, an intergeneric allotetraploid derived from a cross between *Brassica rapa* and *Raphanus sativus*. *Frontiers in Plant Science* 11: 851.

Parkin IA, Gulden SM, Sharpe AG, Lukens L, Trick M, Osborn TC, Lydiate DJ. 2005. Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. *Genetics* 171: 765–781.

Rapp RA, Udall JA, Wendel JF. 2009. Genomic expression dominance in allopolyploids. *BMC Biology* 7: 18.

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

Ryan DP, Ehninger D. 2014. Bison: bisulfite alignment on nodes of a cluster. *BMC Bioinformatics* 15: 337.

Schultz MD, Schmitz RJ, Ecker JR. 2012. 'Leveling' the playing field for analyses of single-base resolution DNA methylomes. *Trends in Genetics* 28: 583–585.

Shaked H, Kashkush K, Ozkan H, Feldman M, Levy AA. 2001. Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. *Plant Cell* 13: 1749–1759.

Shan S, Boatwright JL, Liu X, Chanderbali AS, Fu C, Soltis PS, Soltis DE. 2020. Transcriptome dynamics of the inflorescence in reciprocally formed allopolyploid *Tragopogon miscellus* (Asteraceae). *Frontiers in Genetics* 11: 888.

Shin H, Park HR, Park JE, Yu SH, Yi G, Kim JH, Koh W, Kim HH, Lee SS, Huh JH. 2021. Reduced fertility caused by meiotic defects and micronuclei formation during microsporogenesis in x*Brassicoraphanus*. *Genes & Genomics* 43: 251–258.

Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31: 3210–3212.

Slater GS, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6: 31.

Smit AF, Hubley R. 2008. REPEATMODELER Open-1.0. [WWW document] URL http://www.repeatmasker.org [accessed 21 April 2022].

Smit A, Hubley R, Green P. 2015. REPEATMASKER Open-4.0. [WWW document] URL http://www.repeatmasker.org [accessed 21 April 2022].

Soltis PS, Soltis DE. 2009. The role of hybridization in plant speciation. *Annual Review of Plant Biology* 60: 561–588.

Soltis PS, Soltis DE. 2016. Ancient WGD events as drivers of key innovations in angiosperms. *Current Opinion in Plant Biology* 30: 159–165.

Sosnowska K, Majka M, Majka J, Bocianowski J, Kasprowicz M, Ksiazczyk T, Szala L, Cegielska-Taras T. 2020. Chromosome instabilities in resynthesized *Brassica napus* revealed by FISH. *Journal of Applied Genetics* 61: 323–335.

Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics* 24: 637–644.

Suzuki M, Ketterling MG, McCarty DR. 2005. Quantitative statistical analysis of cis-regulatory sequences in ABA/VP1- and CBF/DREB1-regulated genes of *Arabidopsis*. *Plant Physiology* 139: 437–447.

Szadkowski E, Eber F, Huteau V, Lode M, Huneau C, Belcram H, Coriton O, Manzanares-Dauleux MJ, Delourme R, King GJ et al. 2010. The first meiosis of resynthesized *Brassica napus*, a genome blender. *New Phytologist* 186: 102–112.

Szadkowski E, Eber F, Huteau V, Lode M, Coriton O, Jenczewski E, Chevre AM. 2011. Polyploid formation pathways have an impact on genetic rearrangements in resynthesized *Brassica napus*. *New Phytologist* 191: 884–894.

Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. 2017. GESEQ - versatile and accurate annotation of organelle genomes. *Nucleic Acids Research* 45: W6–W11.

Trapnell C, Pachter L, Salzberg SL. 2009. TOPHAT: discovering splice junctions with RNA-Seq. *Bioinformatics* 25: 1105–1111.

Van de Peer Y, Mizrachi E, Marchal K. 2017. The evolutionary significance of polyploidy. *Nature Reviews Genetics* 18: 411.

Waminal NE, Kim HH. 2012. Dual-color FISH karyotype and rDNA distribution analyses on four Cucurbitaceae species. *Horticulture Environment and Biotechnology* 53: 49–56.

Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun JH, Bancroft I, Cheng F et al. 2011. The genome of the mesopolyploid crop species *Brassica rapa*. *Nature Genetics* 43: 1035–1039.

Wang Z, Baulcombe DC. 2020. Transposon age and non-CG methylation. *Nature Communications* 11: 1221.

Wei Y, Li G, Zhang S, Zhang S, Zhang H, Sun R, Zhang R, Li F. 2021. Analysis of transcriptional changes in different *Brassica napus* synthetic allopolyploids. *Genes* 12: 82.

Wendel JF. 2000. Genome evolution in polyploids. *Plant Molecular Biology* 42: 225–249.

Wendel JF. 2015. The wondrous cycles of polyploidy in plants. *American Journal of Botany* 102: 1753–1756.

Wendel JF, Jackson SA, Meyers BC, Wing RA. 2016. Evolution of plant genome architecture. *Genome Biology* 17: 37.

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 Y, Zhong L, Wu X, Fang X, Wang J. 2009. Rapid alterations of gene expression and cytosine methylation in newly synthesized *Brassica napus* allopolyploids. *Planta* 229: 471–483.

Yang XZ, Li L. 2011. MIRDeep-P: a computational tool for analyzing the microRNA transcriptome in plants. *Bioinformatics* 27: 2614–2615.

Yoo MJ, Liu X, Pires JC, Soltis PS, Soltis DE. 2014. Nonadditive gene expression in polyploids. *Annual Review of Genetics* 48: 485–517.

Zang CZ, Schones DE, Zeng C, Cui KR, Zhao KJ, Peng WQ. 2009. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* 25: 1952–1958.

Zhang H, Bian Y, Gou X, Zhu B, Xu C, Qi B, Li N, Rustgi S, Zhou H, Han F et al. 2013. Persistent whole-chromosome aneuploidy is generally associated with nascent allohexaploid wheat. *Proceedings of the National Academy of Sciences, USA* 110: 3447–3452.

Zhang J, Liu Y, Xia EH, Yao QY, Liu XD, Gao LZ. 2015. Autotetraploid rice methylome analysis reveals methylation variation of transposable elements and their effects on gene expression. *Proceedings of the National Academy of Sciences, USA* **112**: E7022–E7029.

Zhong S, Joung JG, Zheng Y, Chen YR, Liu B, Shao Y, Xiang JZ, Fei Z, Giovannoni JJ. 2011. High-throughput illumina strand-specific RNA sequencing library preparation. *Cold Spring Harbor Protocols* **2011**: 940–949.

Supporting Information

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

Fig. S1 Phenotypes of x*Brassicoraphanus* intermediate between *B. rapa* and *R. sativus*.

Fig. S2 Flow cytometry analysis and genome size estimation of x*Brassicoraphanus*.

Fig. S3 Chloroplast genome of x*Brassicoraphanus*.

Fig. S4 Comparison of the x*Brassicoraphanus* genome with its parental genomes.

Fig. S5 Genome-wide DNA methylation in x*Brassicoraphanus*.

Fig. S6 H3K9me2 modification of x*Brassicoraphanus*.

Fig. S7 Chromosome interactions in resynthesized x*Brassicoraphanus* and *Brassica napus*.

Fig. S8 Transcriptome analysis of x*Brassicoraphanus*.

Fig. S9 Small RNA analysis of x*Brassicoraphanus*.

Fig. S10 DNA methylation metaplots in transposable elements.

Fig. S11 Distribution of DNA methylation changes in A and R subgenomes of x*Brassicoraphanus*.

Table S1 Summary of genomic reads from x*Brassicoraphanus*.

Table S2 Statistics of x*Brassicoraphanus* genome assembly.

Table S3 Primers and oligonucleotide for fluorescence *in situ* hybridization (FISH) probes.

Table S4 Primers for production of antibodies.

Table S5 Number of read pairs mapped on *B. rapa*, *R. sativus* and x*Brassicoraphanus* genomes.

Table S6 Primers for northern blot probes.

Table S7 Benchmarking universal single-copy orthologs (BUSCO) analysis to estimate the completeness of the genome assembly.

Table S8 Annotation of repeat sequences in the x*Brassicoraphanus* genome.

Table S9 Chloroplast genome annotations of x*Brassicoraphanus* and progenitors.

Table S10 Detected differentially methylated regions in x*Brassicoraphanus*.

Table S11 Histone H3K9me2 peak regions in *B. rapa*, *R. sativus* and x*Brassicoraphanus*.

Table S12 Gene ontology analysis of homoeologous gene pairs showing biased, convergent and constant expression.

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.