Characterization and acceleration of genome shuffling and ploidy reduction in synthetic allopolyploids by genome sequencing and editing

Xiaohui Zhang [1],*,†, **Shuangshuang Zhang**[1,†], **Zhongping Liu**[1,†], **Wei Zhao**[1],
Xiaoxue Zhang[1], **Jiangping Song**[1], **Huixia Jia**[1], **Wenlong Yang**[1], **Yang Ma**[1], **Yang Wang**[1],
Kabin Xie [2], **Holger Budahn**[3] and **Haiping Wang**[1],*

[1]Key Laboratory of Biology and Genetic Improvement of Horticultural Crops, Ministry of Agriculture and Rural Affairs; Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing, 100081, China, [2]National Key Laboratory of Crop Genetic Improvement and National Center of Plant Gene Research (Wuhan); College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070, China and [3]Institute for Breeding Research on Horticultural Crops, Julius-Kuehn-Institute, Federal Research Centre for Cultivated Plants, D-06484 Quedlinburg, Germany

ABSTRACT

Polyploidy and the subsequent ploidy reduction and genome shuffling are the major driving forces of genome evolution. Here, we revealed short-term allopolyploid genome evolution by sequencing a synthetic intergeneric hybrid (*Raphanobrassica*, RRCC). In this allotetraploid, the genome deletion was quick, while rearrangement was slow. The core and high-frequency genes tended to be retained while the specific and low-frequency genes tended to be deleted in the hybrid. The large-fragment deletions were enriched in the heterochromatin region and probably derived from chromosome breaks. The intergeneric translocations were primarily of short fragments dependent on homoeology, indicating a gene conversion origin. To accelerate genome shuffling, we developed an efficient genome editing platform for *Raphanobrassica*. By editing *Fanconi Anemia Complementation Group M (FANCM)* genes, homoeologous recombination, chromosome deletion and secondary meiosis with additional ploidy reduction were accelerated. *FANCM* was shown to be a checkpoint of meiosis and controller of ploidy stability. By simultaneously editing *FLIP* genes, gene conversion was precisely introduced, and mosaic genes were produced around the target site. This intergeneric hybrid and genome editing platform not only provides models that facilitate experimental evolution research by speeding up genome shuffling and conversion but also accelerates plant breeding by enhancing intergeneric genetic exchange and creating new genes.

INTRODUCTION

Distant hybridization and genome duplication are fundamental driving forces of evolution occurring naturally or artificially in plants and animals (1–3). Half of the world's crops and wild species have undergone at least one recent interspecific hybridization and whole-genome polyploidization event (4). The genome sizes of higher plants vary by 2400-fold from 61 Mb for *Genlisea tuberosa* to 149 Gb for *Paris japonica* (5). The degrees of rearrangement are quite different (6). The underlying genome shuffling, expansion and deletion mechanisms were predicted by comparative genome analysis (7). However, hardly any experimental proof has been provided to verify these theories because the process is too slow to be studied, even exhausting a scientist's whole research career. Therefore, an artificial allopolyploid genome evolution model with accelerated interspecific genome shuffling and purification would be valuable (8,9).

Compared with animals, plants have several key advantages for investigating allopolyploid-dependent genome evolution. First, plants are more tolerant to harmful genome variations. For example, the mutation of chromosome recombination-related genes such as *Fanconi Anemia Complementation Group M* (*FANCM*) causes cancers and other severe syndromes in animals but does not cause obvious vegetative or reproductive growth defects in plants (10,11). Second, plants can produce a large number of

descendants that allow us to screen large populations for rare mutants and infrequent characters. Third, plant tissue culture, transformation, multiplication and growth are simple and inexpensive. Radish and cabbage are among the most important vegetable crops worldwide. Artificial interspecific hybrids between radish and *Brassica* crops have been produced successfully by some research groups, but most of the F$_1$ hybrids are sterile (12). For only very few of them, fertility was restored by genome duplication (13). By subsequent backcrossing steps, the disease resistance genes and other traits were transferred from radish to *Brassica* crops (13). However, due to a low frequency of homoeologous chromosome recombination, the transfer of the target genes as well as the successive reduction of chromosome fragments with unfavorable genes, so-called 'linkage drag', was found to be complicated (14). Therefore, improving interspecific homoeologous chromosome recombination is of great value in plant breeding.

There are two potential ways to improve homoeologous chromosome recombination and shuffling. The first is to induce double strand breaks (DSBs), and the other is to modify the molecular mechanism controlling chromosome pairing and crossover (CO) (15). Compared with physical and chemical methods such as γ-ray and phleomycin, clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated protein (Cas) can be applied very precisely to introduce DSBs at a specific position (16). Studies in yeast and tomato showed that CRISPR/Cas9-mediated DSBs can enhance recombination frequency by several hundred fold in mitosis (17,18). Double DSBs introduced by CRISPR/Cas can also induce chromosomal inversion and translocation in *Arabidopsis* (19–21). On the other hand, numerous genes responsible for chromosome recombination have been identified in yeast, *Arabidopsis* and rice, as well as in insects and mammalin cells (22–24). In interspecific hybrid plants, homoeologous chromosome pairing has been observed, where the D-loop can be produced and double Holliday junction (dHJ) complexes may be generated. Only a very small fraction of the dHJ complexes can finally complete the recombination pathway and generate a CO, and most of the others are repaired, resulting in non-crossover (NCO) (23,25). There are two types of COs, interference sensitive (class I) and interference insensitive (class II) (23). The class II COs can be elevated without a loss of fertility via disruption of the three protein complexes. FANCM and its cofactors MHF1/2 dissolve the D-loop by synthesis-dependent strand annealing (SDSA), resulting in gene conversion instead of CO (26,27). The RECQ4A/B, TOP3α and RMI1 proteins form the BTR complex, which inhibits CO by dissolving the D-loop in a FANCM-independent manner (28). FIGL1 and FLIP form complexes suppressing CO by regulating single-strand DNA invasion (29). Mutation or RNA inference (RNAi) of these genes can improve CO several fold (24,26,30–34).

CRISPR is a native bacterial system for defense against bacteriophages that involves cutting the invasive DNA at the target site, which is homologous to the ∼20 nt short palindromic repeat (35). After moderate modifications, this system was successfully applied in animals and plants for precise genomic mutation, insertion, nucleotide modification and expression regulation (36–38). Several CRISPR/Cas9 vector systems have been well developed for use in plants (37,39–42). Among them, polycistronic tRNA-gRNAs (PTGs) have an advantage in that multiple targets can be edited by assembling up to six single guide RNAs (sgRNAs) in a single polycistronic construct (43,44). Radish is a regeneration-recalcitrant plant, and genome editing technology has not been developed in *Raphanobrassica*, to the best of our knowledge.

Previously, the fertile *Raphanobrassica* line C118 (RRCC) was developed by hybridization of a tetraploid oil radish (*Raphanus sativus* ssp. *oleiferus*) and a tetraploid fodder kale (*Brassica oleracea* convar. *acephala*) (13). In the present study, the genome of the RRCC plant was sequenced, assembled and compared with the parental genomes to gain insight into the early genomic evolution of the intergeneric hybrid. Homologous recombination and genome shuffling were elevated in this RRCC plant by CRISPR/Cas9 editing of recombination-related genes. The present research not only sheds light on the rapid genomic reshaping features of synthetic allopolyploids but also benefits the experimental simulation of allopolyploid genome evolution, as well as germplasm innovation and breeding of new types of crops.

MATERIALS AND METHODS

Plant material

The *Raphanobrassica* line C118 ($2n = 4x = 36$; RRCC) was developed by Professor E. Clauß in 1978 at the Institute for Breeding Research, Quedlinburg (Germany) by an interspecific cross of a tetraploid oil radish (*R. sativus* L. ssp. *oleiferus* DC.; line 2655; $2n = 4x = 36$, RRRR) as the female parent with a pollen mixture of two tetraploid fodder kales (*B. oleracea* L. convar. *acephala*; lines nFMK and PC81; $2n = 4x = 36$, CCCC), as described previously (13). The plants were propagated by selfing in a greenhouse under insect-free conditions. The transgenic plants were grown in 20 cm × 20 cm pots and placed in a greenhouse. The seeds were stored at –20°C.

DNA extraction and sequencing

Young leaf tissues were sampled from a healthy RRCC plant and used for high molecular weight DNA extraction employing a cetyltrimethylammonium bromide (CTAB) method. For PacBio sequencing, a polymerase chain reaction (PCR)-free single-molecule real-time (SMRT) library with a 40 kb insert size was constructed and sequenced on the PacBio Sequel platform using P6 polymerase/C4 chemistry in accordance with the manufacturer's procedure (Pacific Biosciences, CA, USA). For Illumina sequencing, 150 bp paired-end libraries with an insert size of 350 bp were constructed in accordance with the manufacturer's standard protocols (Illumina, CA, USA) and then sequenced on the Illumina NovaSeq 6000 platform. Both sequencings were performed at the Berry Genomics Company (Beijing, China).

High-throughput chromosome conformation capture (Hi-C) sequencing

The young leaves were freshly sampled, dissected into 2–3 mm pieces and subsequently cross-linked with formaldehyde. The cells were lysed to release the nuclei. The cross-linked DNA was digested with MboI and then marked with biotin-14-dCTP during the end-repair process. The adjacent DNA fragments were ligated by T4 polymerase. Subsequently, the proteins at the junctions were digested with proteinase to release the DNA from the DNA−protein cross-linking. Then, the DNA was purified and subsequently randomly sheared into ∼350 bp fragments. The biotin-labeled DNA fragments were isolated using streptavidin MagBeads and then subjected to paired-end library construction according to the manufacturer's instructions (Illumina). The resulting Hi-C libraries were sequenced on an Illumina NovaSeq 6000 by Berry Genomics Company.

RNA sequencing

Total RNA was extracted from bud, flower, silique, stem, leaf, root and callus with TRIzol reagent (Thermo Fisher Scientific, Waltham, MA, USA). RNA samples were used to construct the standard RNA sequencing (RNA-seq) libraries and sequenced on the Illumina NovaSeq 6000 platform by Berry Genomics Company using the standard procedure. The 150 bp paired-end clean reads were mapped to the genome using HISAT2 (45).

Genome assembly

Before genome assembly, the genome size and heterozygosity were estimated via a 23 K-mer frequency analysis with JELLYFISH v2.1.4 (46) using the Illumina data. Using the PacBio data, the genome was *de novo* assembled with Canu (v1.7.1) (47) and polished with Arrow (v2.3.2) (48) with the default parameters. The derived contigs were further polished with Illumina reads using Pilon (v1.22) (49). The high-quality Hi-C paired-end reads were mapped to the assembled contigs using JUICER (50). The chromosome-level scaffolds were constructed, split, sealed and merged with 3D DNA (51). The scaffolds were oriented and placed on the chromosomal groups based on their interaction strengths. The final nomenclature and orientation of the chromosomes were determined based on their synteny to the corresponding cabbage (C1–C9) and radish (R1–R9) chromosomes.

Repeat annotation

Transposable elements (TEs) were annotated with a combination of RepeatMasker (RepBase) (52) and RepeatModeler (53). Full-length long terminal repeats (LTRs) were analyzed using LTR_retriever (54).

Non-coding gene annotations

The tRNAs were annotated with tRNAscan-SE (55). The rRNAs and other non-coding RNAs were annotated by Rfam 14.1 (http://rfam.xfam.org/) (56).

Protein-coding gene annotations

The protein-coding genes were predicted with a combination of homology-based, transcriptome-based and *ab initio* prediction approaches. The protein sequences of *Arabidopsis thaliana*, *A. lyrata*, *Brassica juncea*, *B. napus*, *B. nigra, B. oleracea*, *B. rapa*, *Carica papaya*, *Cucumis sativus*, *Oryza sativa*, *Populus trichocarpa*, *R. sativus*, *Thellungiella salsuginea*, *T. parvula* and *Vitis vinifera* were used for homology-based prediction using GeMoMa 1.4.2 (57). The RNAs from roots, stems, leaves, buds, flowers and siliques were sequenced, mixed and assembled into 95 237 non-redundant transcripts with Cufflinks v2.2.1 (58). These sequences were applied in transcriptome-based gene prediction using Program to Assemble Spliced Alignment (PASA v2.0.1) (59). The *ab initio* prediction was performed using four tools: Augustus (60), SNAP (61), GlimmerHMM (62) and GeneMark-ET (63). All the gene models derived from the three approaches were integrated by EVidenceModeler (EVM) (64). The untranslated regions (UTRs) and alternative splicing variation information were updated using PASA.

The functions of the predicted proteins were annotated by BLAST against the NCBI non-redundant protein (NR), SwissProt, and evolutionary genealogy of genes: Non-supervised Orthologous Groups (eggNOG) databases. The Gene Ontology (GO) terms, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and InterPro signatures were annotated by Blast2GO (65), KOBAS (66) and Inter-ProScan (67), respectively.

Genome evaluation

The completeness and accuracy of the genome assembly were evaluated by BUSCO v3.0.0 (http://busco.ezlab.org), AUGUSTUS (http://augustus.gobics.de) and HMMER (http://hmmer.org/download.html).

Detection of gene deletion, inversion, translocation and tandem duplication

The collinear genes between the R and C genomes were detected with MCScan (68). The protein-coding genes of RRCC were reciprocally BLAST searched against the genes of the Xin-li-mei radish and HDEM cabbage genomes. The genes present in Xin-li-mei or HDEM but with no hits in RRCC were treated as deleted in the RRCC genome. The best-hit gene pairs were termed orthologous genes and were used to detect inversions and translocations. An inversion was identified when a stretch of genes was inversely ordered on the RRCC chromosome in comparison with its radish and cabbage homologous chromosome. When the best-hit genes were shifted to another chromosomal location, they were considered a translocation. The tandemly arranged homologous genes that newly formed in the RRCC genome were identified as tandem duplications.

Pangenome construction and classification of variations

The family-based pangenome of *B. oleracea* was constructed with the same method used for the *Raphanus* pangenome (69). Briefly, the protein sequences of the six

B. oleracea genomes (70–74) were downloaded and then clustered using OrthoMCL (75). The gene families that were shared by all six cabbage genomes were defined as core families. Gene families that were missing in 1–4 genomes were defined as dispensable families. The gene families and singletons that existed in only one genome were defined as accession-specific genes. The gene retention and deletion were assigned to the core, dispensable and specific categories according to the pangenomes.

Plasmid construction

The plasmids pGTR and pRGEB33 were kindly provided by Professor Xie Ka-bin, Huazhong Agricultural University. The polycistronic PTGs were constructed according to the reference method (43,44) and assembled into the pRGEB33 plasmid using the Golden Gate method, generating pFANCM-R, pFANCM-B and pFLIP1. The plasmids were transformed into *Agrobacterium* (strain C58) by electroporation. The oligonucleotides for PTG syntheses are listed in Supplementary Table S1.

Plant transformation

The seeds were surface sterilized with 70% alcohol for 30 s and in 1:1 diluted 84 disinfectant (Aitefu, Jiangsu, China) for 15 min, and then sown in half-strength Murashige and Skoog ($MS_{1/2}$) medium. Approximately 6–7 days after sowing, the cotyledons were dissected from seedlings. Each one was cross-cut from the middle to generate two disc explants, and the hypocotyls were cut into 5 mm length segments. The two types of explants were cultivated in MS_0 medium for 2–3 days. *Agrobacterium* (vector containing) was cultivated in 50 ml of LB liquid medium at 200 rpm and 28°C overnight. The *Agrobacterium* cells were collected by centrifugation at 3000 rpm for 10 min and resuspended in MS_d medium to $OD_{600} = 0.5$. The explants were soaked in the *Agrobacterium* suspension for 10 min and then placed on MS_1 medium for co-cultivation for 2–3 days. Subsequently, the explants were moved to MS_2 medium for 2–3 days to kill the *Agrobacterium* while the explants were allowed to grow without selection. After that, explants were transferred to MS_3 for selection. Regeneration buds were obtained after 2 weeks of cultivation and lasted for 4 weeks or longer. When the morphology was normally developed, the plantlets were dissected from the explants and transplanted onto MS_r medium for rooting. The rooted plants were transplanted into pots filled with peat and vermiculite. The compositions of all of the media are listed in Supplementary Table S2.

DNA extraction and mutant assay

DNA was extracted from leaf tissue using the CTAB method. The primers (UBI10P-F and UBI10P-R) located at the UBIp promoter that drive the PTGs in the pRGEB33 plasmid were used for analysis and detection of the transgenic positive plants. The ∼700–1000 bp fragment covering the target region was PCR amplified from transgenic plants. The PCR products were sequenced using the Sanger method to detect the disruptive peaks at the target region. The primers are listed in Supplementary Table S3.

Cytological studies

Buds ∼3 mm in length were sampled at 9–10 am on sunny days and fixed in Carnoy's fluid for 24 h. Subsequently, the samples were soaked in 95% and 80% alcohol for 30 min and then transferred to 70% alcohol and kept at 4°C. Before tableting, the buds were washed 2–3 times with distilled water to remove the alcohol on the surface. Then, the anthers were dissected from the buds and dissociated with 1 mol/l HCl at 60°C for 8–12 min. The anthers were washed with distilled water 2–3 times to remove excess HCl and then were squashed to release the pollen mother cells (PMCs). An appropriate amount of 4′,6-diamidino-2-phenylindole (DAPI) or Carbol fuchsin (Leagene, Beijing) was added for staining, and the chromosomes were fully spread by pressing. The chromosome morphology was observed with an optical microscope (Leica or Olympus). Three homozygous T_1 lines for each genotype were used in these cytological analyses. The exact numbers of cells analyzed are shown in the respective figures. Statistical significance was calculated with the two-tailed Student's *t*-test.

Detection of targeted gene conversion

To detect gene conversion, the fragment around the cutting site was amplified by high-efficiency thermal asymmetric interlaced PCR (hiTAIL-PCR) according to standard protocols (76). The primers are listed in Supplementary Table S4. The PCR products were cloned into the pEasy-T vector and multiplied in *Escherichia coli* DH5α competent cells (Tsingke, Beijing). The clones were sequenced using M13 primers.

RESULTS

Construction of a synthetic allotetraploid genome

To reveal what happened to the allotetraploid genome, the eighth generation of a fertile synthetic allotetraploid plant, *Raphanobrassica* line C118 (RRCC), was sequenced in this study. The RRCC genome size was estimated to be ∼1.07 Gb by a genomic survey (Supplementary Figure S1). A total of 201 Gb of PacBio long reads (N50 = 37.9 kb) and 100 Gb of Illumina short reads (150 bp, paired-end) were generated, which covered 187× and 93× the depths of the estimated genome, respectively (Supplementary Table S5). Using these data, the genome was assembled into 688 contigs covering a total of 940.7 Mb (pipeline shown in Supplementary Figure S2). The contig N50 = 21.41 Mb and the contig N90 = 1.84 Mb, indicating good continuity of the assembly (Table 1). In total, 885.96 Mb (94.18%) of the entire assembly was anchored onto 18 chromosomes with Hi-C technology (Supplementary Tables S6, S7; Figure 1A, B). The unanchored scaffolds were centromeric repeat-rich sequences, as indicated by the Hi-C interaction strength map (Figure 1B). The chromosomes were numbered and oriented based on their synteny to the corresponding cabbage (C) and radish (R) chromosomes (Supplementary Figure S3). The GC content was determined to be 36.51%, which is comparable with that of the cabbage and radish genomes (69,72). A total of 99.41% of the Illumina reads were properly paired-mapped to the genomes,

Table 1. Statistics of the genome assembly and annotation

Total size of contigs (Mb)	940.7
Number of contigs	688
Longest contig (Mb)	44.99
N50 contig length (Mb)	21.41
N90 contig length (Mb)	1.84
L50 contig count	16
L90 contig count	61
Total size of scaffolds (Mb)	940.87
Number of scaffolds	1119
N50 scaffold length (Mb)	53.89
N90 scaffold length (Mb)	28.79
L50 scaffold count	8
L90 scaffold count	17
Number of gaps	343
Total gap length (Mb)	0.17
Anchored chromosomes (Mb)	885.96
Anchored chromosomes (% of assembly)	94.18%
BUSCO completeness	98.40%
Repeat (% of genome)	56.01%
Number of protein-coding genes	82 356

indicating the high accuracy of the assembly (Supplementary Table S8). The benchmarking universal single-copy ortholog (BUSCO) assessment showed that 98.4% of the 1440 single-copy genes were present in this genome (Supplementary Table S9). Both the contig length and BUSCO completeness were higher than those of the recently released radish and cabbage genomes (Supplementary Table S10) (69,70,73,74). These results indicated the good accuracy and integrity of this assembly.

A total of 527.01 Mb (56.01% of the genome) was annotated as repeats. Among these, retrotransposons and DNA transposons accounted for 37.5% and 8.5% of the total genome assembly, respectively (Supplementary Table S11). This was comparable with the fractions of repeats in the radish and cabbage genomes (69,74). The LTR assembly index (LAI) was 19.51 (Supplementary Table S11), indicating the high quality of the genome assembly (77). We annotated the non-coding RNAs in the RRCC genome and revealed 4517 rRNAs, 2437 tRNAs, 3876 small nuclear RNAs, 383 microRNAs and several to hundreds of other kinds of RNAs (Supplementary Table S12). These numbers of non-coding RNAs are reasonable compared with the corresponding numbers in the radish and cabbage genomes (69,78).

With a comprehensive strategy combining protein homology-based prediction, RNA sequencing (RNA-seq)-based prediction and *ab initio* prediction, a total of 82 356 protein-coding genes were predicted (Supplementary Table S13). The mean lengths of the transcripts and coding sequences (CDSs) were 1456 bp and 1284 bp, respectively, comparable with those of cabbage and radish genomes (69,70,73,74). A total of 80 687 of the genes (97.97%) were annotated by at least one of the NR, GO, eggNOG, KEGG, InterPro and Swiss-Prot databases (Supplementary Table S14).

Deletion is quick while rearrangement is slow in the allotetraploid genome

The radish genome (R genome) and cabbage genome (C genome) were estimated to be 456–574 Mb and 566–

659 Mb, respectively (69,70,73,74). Therefore, the original RRCC genome size should be 1.02–1.23 Gb based on the combination of the RR and CC genomes. The present genome was estimated to be 1.07 Gb, which is much smaller than the median of the expected interval, indicating a high tendency for genomic deletion.

The C genome assembly consisting of the C1–C9 pseudochromosomes showed a length of 533.53 Mb, similar to that of recently released high-quality cabbage genome assemblies (522.88–544.39 Mb) (70,73,74). The R genome, assembled into R1–R9 pseudochromosomes, was 352.43 Mb in length, ~65.9–162.25 Mb less than the recently released radish (*R. sativus*) genome assemblies (418.33–514.68 Mb) (69) (Figure 1C). Because the R and C subgenomes of RRCC plants were derived from the same tissues and were sequenced and assembled by the same pipeline, the different assembly sizes indicated that the R genome shrank significantly faster than the C genome after hybridization. The C1–C9 chromosomes harbored 44 555 protein-coding genes, far fewer than the total gene number (59 064–62 232) of five *B. oleracea* genome assemblies but comparable with the total number of protein-coding genes (43 868 and 45 758) in the cabbage D134 and 02-12.v1 genomes (70,73,74,78). Otherwise, the R1–R9 chromosomes contained 36 261 protein-coding genes, at least 6000 genes fewer than the lowest number of genes in radish genomes (42 319–52 190) (69) (Figure 1D). This indicates that gene deletion occurred in both the C and R subgenomes but was more frequent in the R subgenome in this RRCC hybrid.

We used the RRCC genes as queries for reciprocal Blast against cabbage HDEM and radish Xin-li-mei gene sets. The best hit gene pairs were retained and then manually checked for genes with *in situ* retention, tandem duplication, inversion and translocation. Compared with the HDEM genome, the C subgenome of RRCC plants harbored 31 551 *in situ*-retained genes, 1720 tandemly duplicated genes, 686 inverted genes, 7608 translocated genes (homologs of 4714 HDEM genes) and 2994 genes without homologs in the HDEM genome, which accounted for 70.81, 3.86, 1.54, 17.08 and 6.72% of the C subgenome, respectively. For a total of 28 174 HDEM genes, no *in situ* homologous genes in the C subgenome of RRCC plants were detected. We then extracted the orthologs of these absent genes from the *ex situ* loci, and 1123 best hit genes were translocated or inverted within the C subgenome. The 1363 best hit genes were detected in the R subgenome, indicating that these genes may have been translocated from the C to R genome after interspecific hybridization. The 12 best hit genes were located in unanchored contigs, and the remaining 25 671 genes were treated as genes lost from the C genome, accounting for 41.89% of the total HDEM genes (Supplementary Table S15).

In comparison with the Xin-li-mei genome, the R subgenome of RRCC plants harbored 27 321 (75.35%) *in situ*-retained genes, 1252 (3.45%) tandemly duplicated genes, 744 (2.05%) inverted genes, 4542 (12.53%) translocated genes (homologs of 3362 Xin-li-mei genes) and 2431 (6.70%) genes without homologs in the Xin-li-mei genome. Meanwhile, the 16 238 non-redundant genes of Xin-li-mei were absent among the syntenic loci of the R subgenome of RRCC plants. Among the best-hit orthologs of these absent
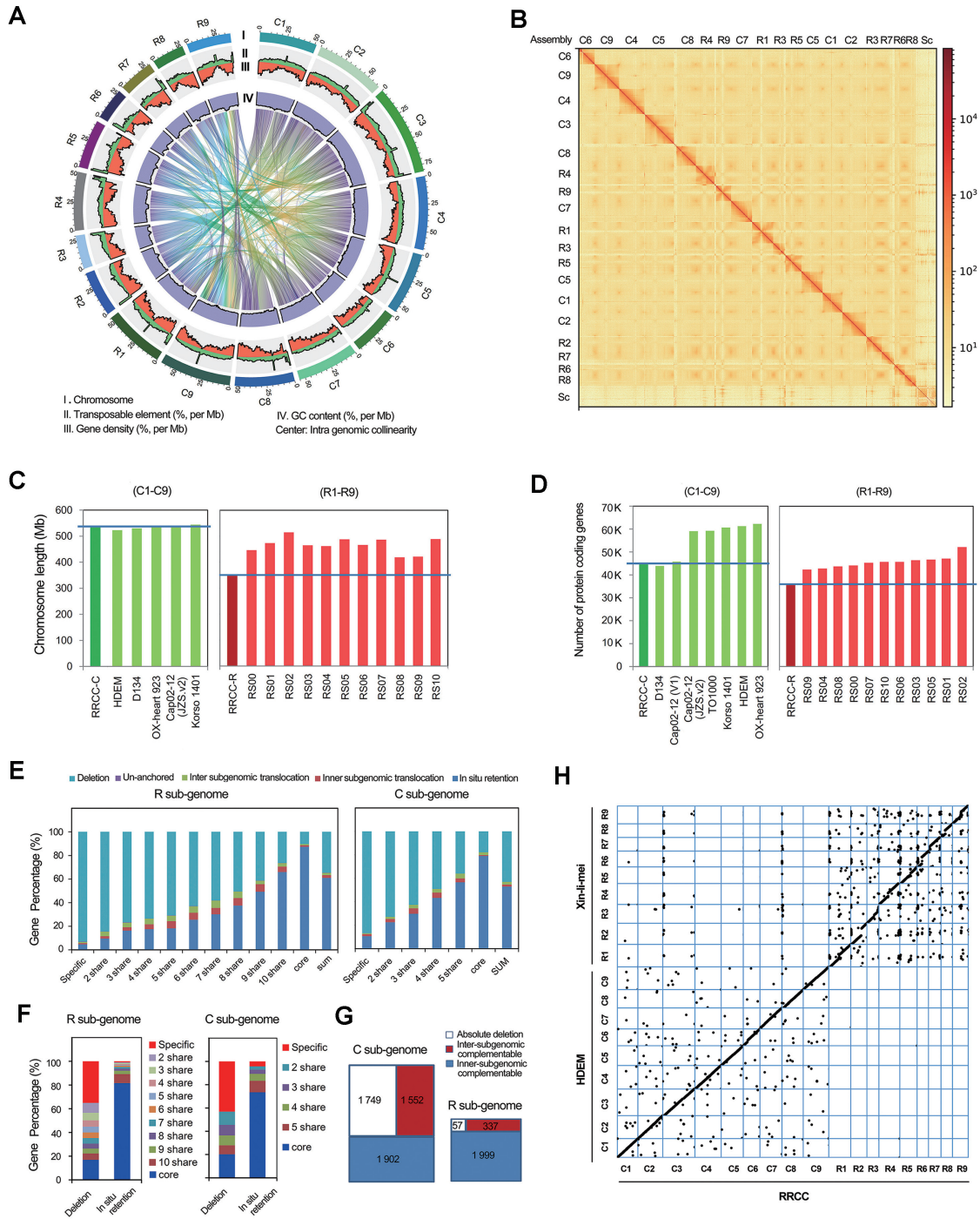
Figure 1. Genome features of RRCC plants. (**A**) Description of the RRCC genome. (**B**) Hi-C strength map. (**C**) Comparison of genome size between the C subgenome of RRCC plants, the five *B. oleracea* genome assemblies (C1–C9), the R subgenome of RRCC plants and the 11 *Raphanus* genome assemblies (R1–R9). (**D**) Comparison of gene numbers between the C subgenome of RRCC plants, the *B. oleracea* genome assemblies (C1–C9), the R subgenome of RRCC plants and the *Raphanus* genome assemblies (R1–R9). (**E–G**) Gene deletion attributed to the pangenome. (**E**) The relationship between the variations and gene categories. (**F**) The percentages of core, dispensable and specific genes that were deleted and retained. (**G**) Categories of deleted genes. (**H**) Genomic synteny of RRCC to the cabbage (HDEM) and radish (Xin-li-mei) chromosomes.

genes, 968 were translocated within the R subgenome, 832 were translocated to the C subgenome and 55 were not anchored to chromosomes. The remaining 14 383 genes were treated as genes lost from the R genome, accounting for 32.61% of the total Xin-li-mei genes (Supplementary Table S15). Because the RRCC plant was not derived from HDEM and Xin-li-mei, some of the deletions, duplications, inversions and translocations could represent already existing variations in the parent lines before hybridization. If we attribute the 2994 and 2431 RRCC-specific genes without homologs in the HDEM and Xin-li-mei genomes to the pre-hybridization genetic variation between the parents and the references and suppose that HDEM and Xin-li-mei evolved similar numbers of specific genes, we can estimate that ∼88% ($1 - 2\,994/25\,671$) to 83% ($1 - 2\,431/14\,383$) of the deletions occurred in the RRCC genome after intergeneric hybridization.

Recently, pangenomic studies have indicated that genomes can be significantly diverged between different cultivars (69,79). Because HDEM and Xin-li-mei were not the parents of the RRCC plant, we could not distinguish RRCC genes lost since hybridization from differences between the parents and references that existed before hybridization. The conserved (core), variable (dispensable) and specific (specific families and singletons) genes in radishes were identified in a recent pangenome study (69). Therefore, we allocated R subgenome variations to the pangenome classes. Ninety-three percent of the Xin-li-mei-specific genes were absent in the RRCC genome. Eighty-seven percent of the core genes resided *in situ* in the R subgenome of the RRCC plant. The genes with a higher frequency in the pangenome had a higher ratio of *in situ* retention and a lower ratio of loss (Figure 1E). To analyze the C subgenome, we constructed a cabbage pangenome (Supplementary Table S16) with six published genome assemblies (70,73,74,78). Similar to those in the R subgenome, the frequency in the pangenome displayed a positive correlation with *in situ* retention but a negative correlation with gene loss, in the C subgenome (Figure 1E). Seventy-nine percent of the cabbage core genes resided *in situ* in the C subgenome. A total of 86.8% of the HDEM-specific genes were absent in the RRCC genome (Figure 1E). In conclusion, among the *in situ*-retained genes, 81.6% in the R subgenome and 73.5% in the C subgenome were core genes (Figure 1F). This indicates that the core genes were still stable in the interspecific hybrid genome. The interspecific hybridization accelerated the deletion of specific and low-frequency dispensable genes.

There were 5203 *B. oleracea* core genes deleted from the C subgenome. Of these genes, 1902 belong to 1102 gene families and have at least one family member retained in the C genome. The retained family members could functionally complement the lost genes. We termed these deletions innersubgenomic complementary deletions, which accounted for 36.6% of the core genes deleted from the C subgenome. The other 3301 genes, belonging to 271 families and 2654 singletons, showed no family member retention in the C subgenome. For the R subgenome, a total loss of 2433 core genes was recorded. Among these, 1999 genes belonging to 1096 families contain at least one homologous family member retained in the R genome. This indicates that 82.2% of the deleted core genes in the R subgenome were innersubgenomic complementary deletions. The other 434 genes, including 46 families and 266 singletons, had no family member retention in the R subgenome.

We then checked whether these genes strictly deleted from subgenomes have complementary genes in the other subgenome with an OrthoMCL (75) analysis. A total of 1552 out of the 3301 (47%) C subgenomic and 337 out of the 434 (77.6%) R subgenomic genes had homologous genes (same cluster) in the complementary subgenome, indicating that 29.8% of the core genes deleted in the C subgenome and 13.9% of the core genes deleted in the R subgenome could be complemented by inter-subgenomic homologous genes. Only 1749 C genomic and 57 R genomic core genes were absolutely deleted without homolog gene retention, which accounted for 5.89% and 0.22% of the total core genes in the C and R subgenomes, respectively (Figure 1G). This finding indicates that the deletions of core genes were restricted and that the functions were well preserved by the retention of complementary genes. Interspecific hybridization and the subsequent biased gene deletion allow fast shuffling of genomes, mainly of specific and dispensable genes, while even the core genome experiences moderate changes, retaining complete functions by complementation. More core genes in the C subgenome were deleted, consistent with the fact that the cabbage pangenome was constructed with only six genome assemblies, yielding less resolution for core genes than in radish, for which the pangenome was constructed at the genus level from 11 genomes covering all species and subspecies of *Raphanus*.

Both small and large fragmental deletions have occurred within eight generations in the RRCC plant. The chromosome synteny between the RRCC genome and the radish and cabbage genomes indicated that no large fragmental chromosome rearrangement occurred in this allotetraploid plant (Figure 1H; Supplementary Figure S3). The C1–C9 chromosomes showed good synteny with the *B. oleracea* (HDEM) chromosomes (70), while the R1–R9 chromosomes were less syntenic with the *R. sativus* reference (Xinli-mei) genome (69), indicating more small translocations and inversions in R1–R9 than in the C1–C9 chromosomes (Figure 1H; Supplementary Figure S3).

Genomic landscape indicates a chromosome break origin of large-fragment deletions

In both the C and R subgenomes, more than half of the total deletion events included only one gene. With the addition of consecutive genes to the deletion stretches, the case number declines rapidly (Figure 2A). Approximately 95% of the total deletions contained no more than five consecutive genes. This indicates that short-stretch deletions are the major type of genomic deletion in the RRCC genome. Due to the smaller size of the R subgenome in comparison with the C subgenome, the number of deleted genes in the R subgenome was smaller than that in the C subgenome. Nonetheless, long-stretch deletions (containing >20 consecutive genes) were much more frequent in the R subgenome than in the C subgenome (Figure 2A, B). The C subgenome has only one superlong deletion containing >100 genes in a row, while the R subgenome contained
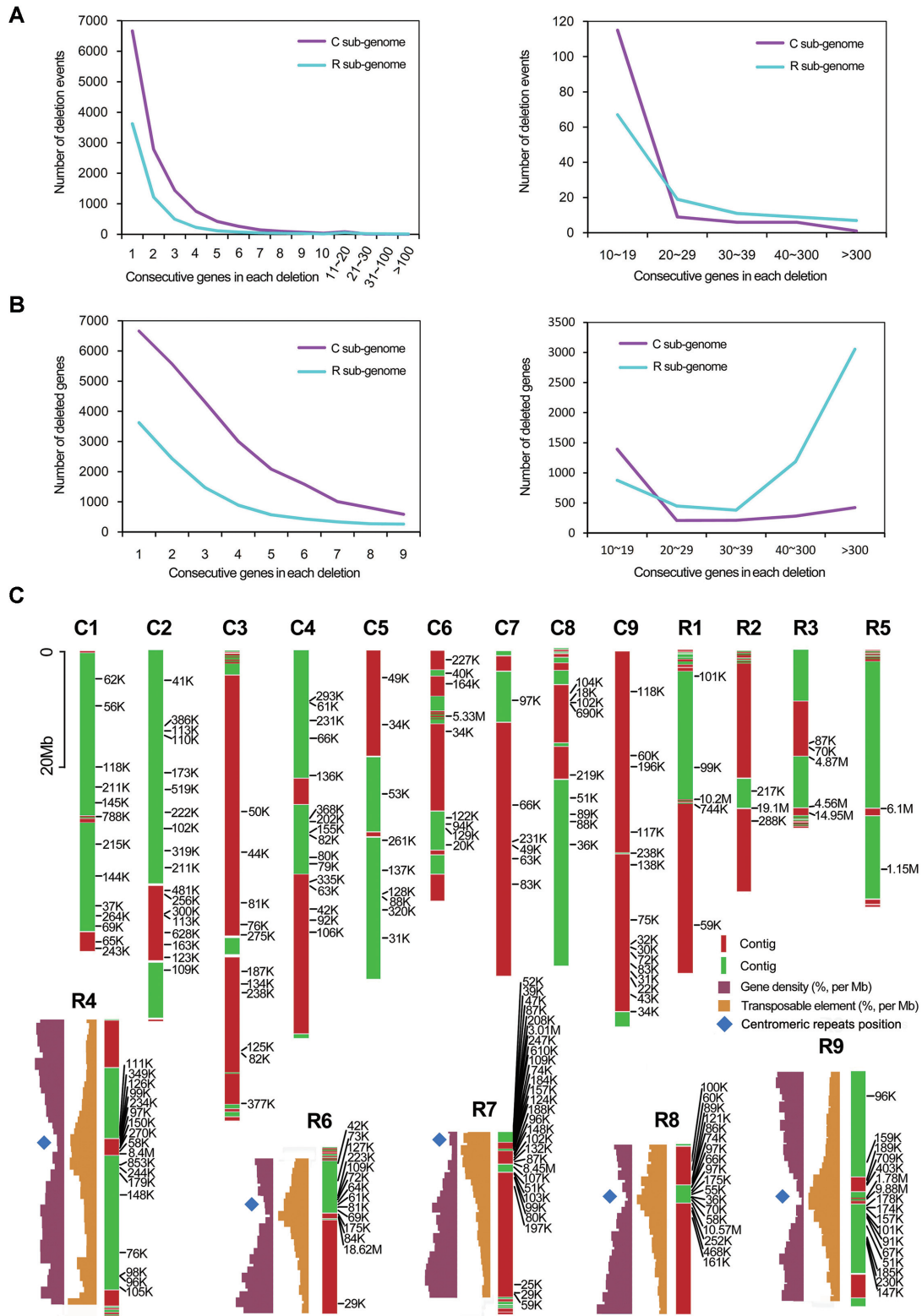
Figure 2. Gene loss in RRCC. (**A**) Number of deletion events according to the loss of consecutive genes. (**B**) Number of genes according to their consecutive deletion. (**C**) Large-fragment deletions and their locations on chromosomes. The green and red colors of the chromosomes indicate the contigs. The purple and ochre bars to the left of the R4 and R6–R9 chromosomes indicate the density of genes and TEs. The blue diamond indicates the centromere position. The density of genes and TEs as well as the centromere positions of the C1–C9, R1–R3 and R5 chromosomes are shown in Supplementary Figure S4.

10 such deletions (Supplementary Table S17). We focused on the long-fragment deletions harboring ≥10 consecutive genes. The C and R subgenomes contained 141 and 111 such deletions, covering a total length of 24.3 and 136.3 Mb, respectively. Each deletion contained 10–545 genes and was 20 kb to 19.14 Mb long. Twelve out of 14 megabase-level superlong deletions in the R subgenome and the only one in the C genome were noticed in the centromere regions and were located near the break points between adjacent contigs (Supplementary Table S17; Figure 2C). This indicates that in regions near the centromere, superlong fragments are lost with higher probability, possibly because the chromosome bridge-derived breaks took place in these regions with higher frequency. This could be attributed to the homoeologous chromosome pairing that occurred in euchromosomal regions causing chromosome bridges, disjunction and subsequent chromosome breaks in the centromeric region. DNA end degradation, repair and joining mechanisms rescued the chromosome but resulted in super-long-fragment deletion. Of course, we cannot exclude the possibility that some of these regions were false deletions derived from misassembly. However, 137 (97% of the total number) C subgenomic and 93 (84% of the total number) R subgenomic large deletions were harbored in the long contigs, indicating that most of the predicted deletions were reliable (Figure 2C). Generally, the deletions were distributed unevenly on the chromosomes. Moderate clusters of deletions were observed in the C subgenome (Figure 2C; Supplementary Figure S4). The R subgenome showed a significantly higher level of clustering of the deletions, especially for chromosomes R4, R6, R7, R8 and R9 (Figure 2C). These deletion clusters resided in repeat-rich, gene-poor centromere regions, further indicating that centromere regions are deletion hotspots (Figure 2C). Centromeres as regions where rearrangements were often enriched have also been revealed in other plants (80,81). These results indicate that the biased subgenomic retention and deletion widely observed in polyploid plants were established at the beginning of hybrid formation.

Intergeneric translocations involve short fragments and are homology dependent

Although chromosomal-level rearrangement has not been identified, the gene-level translocations were truly detected in this allotetraploid. To reveal the features of these gene translocations, the chromosomal synteny of alleles between the R and C subgenomes was first determined. A total of 1153 synteny blocks were revealed, each harboring more than five collinear genes. Because both the R and C genomes contain three ancestral subgenomes, the synteny blocks were seriously redundant (Supplementary Figure S5A). The synteny relationship was manually checked, and the 80 best non-redundant synteny blocks were adopted for the main synteny analysis (Supplementary Figure S5B). Based on these synteny blocks, chromosomal synteny maps displaying the panorama and the main line were drawn (Figure 3A). The non-redundant synteny blocks contained 24 043 pairs of collinear genes (alleles), making up 59.1% of the total genes on the 18 chromosomes. Specifically, 53.8% of the C genome and 65.6% of the R genome genes belonged

to collinear pairs. These collinear pairs served as the backbone for the analysis of inter-subgenome shuffling.

For the majority of the translocations occurring within the subgenomes, it is not easy to distinguish which happened before or after hybridization. Therefore, we focused on the translocations between the R and C subgenomes. We used the 1363 C to R translocation candidates and the 832 R to C translocation candidates for reciprocal Blast against the combined data of both cabbage HDEM and radish Xin-li-mei gene sets. After this second round of screening, a total of 775 genes were found to be translocated from the C to R subgenome, and 439 genes were translocated in the opposite direction. These translocations were generally located in the R to C synteny blocks, indicating that they were caused by a homoeologous sequence-dependent pathway (Figure 3). However, we did not observe large-fragment COs between the C and R subgenomes (Figure 1H). Thus, the majority of the intergenomic translocations were probably gene conversions. The conversion segments included no more than two genes. This is in accordance with the results in rapeseed genomes, in which intergenomic conversion ranged from single nucleotides to whole genes (82). The chromosomal end-to-end syntenic regions showed a higher frequency of intergenomic gene translocation. The reason could be that these regions had higher pairing efficiency.

Genomic deletion and rearrangement can be elevated by editing *FANCM* genes

Genome shuffling is dependent on chromosome breaks and homoeologous recombination. We did not find COs between the R and C chromosomes. This result is consistent with previous knowledge that there is strong suppression of homoeologous recombination in intergeneric hybrids (13). Previous studies in *Arabidopsis* have shown that mutation of *FANCM* can significantly improve CO efficiency while maintaining plant fertility (24,26). *FANCM* is an ancient gene shared by microorganisms, animals and plants (83,84). The RRCC genome contains two *FANCM* genes (RB0C5c024012 and RB0R5c065956) on chromosomes C5 and R5 (Figure 4A). Therefore, we designed two and four gRNAs to target each of the two genes. Each three gRNAs were tandemly arranged in PTGs and inserted into the *Cas9*-expressing vector pRGEB33 (44) to form two plasmids, pFANCOM-B and pFANCOM-R (Figure 4B). The construct pFANCOM-B was designed to simultaneously target homologous R and C genomic genes. A total of 174 independent transgenic plants harboring the two constructs were obtained (Figure 4C–F; Supplementary Table S18). The mutation rate of the T_0 generation transgenic plants was estimated by restriction enzyme digestion and sequencing of the PCR products of the targeted DNA fragments. All six gRNAs successfully induced mutations at the targeted positions, with a frequency ranging between 6.3% and 62.2% (Supplementary Table S18). Because all of the mutations were chimeric or heterozygous in the T_0 generation (Figure 4G; Supplementary Figure S6), the plants were self-pollinated to generate homozygous mutations. From the T_1 plants, we successfully selected homologous mutations of both *FANCOM-R* and *FANCOM-C*, as well as the
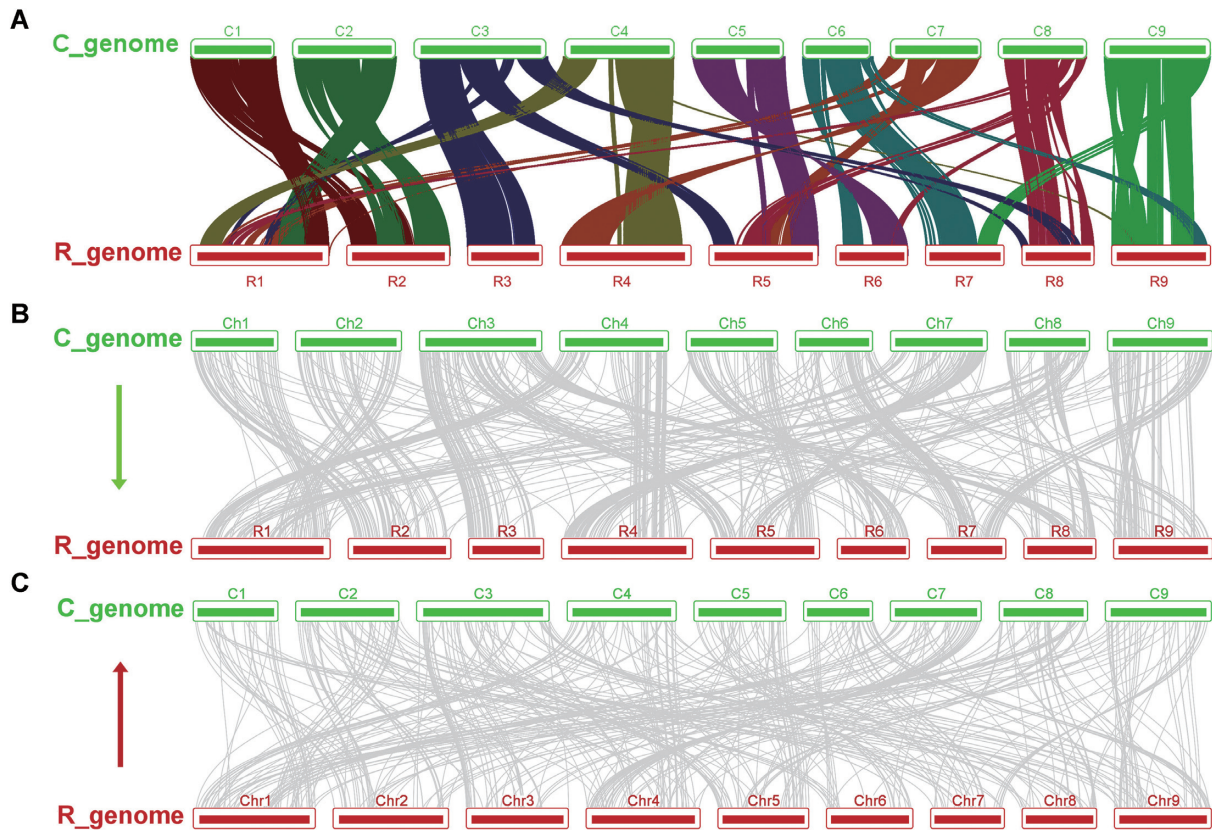
Figure 3. Gene translocation between the C and R subgenomes. (**A**) Collinear gene pairs between the R and C subgenomes; (**B**) translocation from the C to R subgenome; (**C**) translocation from the R to C subgenome.

double mutation. In the eight homologous T_1 lines generated from four independent pFANCOM-R transgenic T_0 lines, the mutations deleted three (L16), four and three (L94s), four (L14s) or eight nucleotides (L101s) that caused one amino acid deletion or frameshifts that produced nearby stop codons, which caused 161 and 158 amino acid truncations of the FANCOM-R protein, respectively (Figure 4H; Supplementary Figure S7). The three homozygous T_1 generation L14s lines representing *fancm-r* were used in further analysis. For the plants harboring the pFANCOM-B construct, the mutations at gRNA1 and gRNA3 were selected. The three T_1 homologous mutants (L7s) harbored deletion of two nucleotides and a nucleotide transversion which caused the 47 amino acids replaced by 31 totally different amino acids at the C-terminus of the FANCM-C protein, which were used representing the *fancm-c* (Figure 4H; Supplementary Figure S8). Derived from another T_0 transgenic line, the homologous T_1 lines (L10s) harboring double mutations of both the *FANCOM-R* and *FANCOM-C* genes were screened out and termed *fancm-r&c* for short. In these plants, the gRNA1 caused an 8 nt deletion in the third exon that resulted in a large fragment truncation (1187 amino acids) of the FANCM-R protein, while gRNA3 caused a 47 amino acid variation at the C-terminus of the FANCM-C protein by a 2 nt deletion in the last exon (Figure 4H; Supplementary Figures S7 and S8). All of these mutant plants displayed normal morphology with no observable difference from the wild type (Supplementary Figure S9).

Mutation of *FANCM* genes affected the meiotic processes. Defects were observed from the diakinesis to the tetrad stages (Figure 5A). In metaphase I of PMCs, the double mutants showed a chiasma frequency of 38.5 ± 0.4 per cell, which was significantly higher than the 35.3 ± 0.4 per cell of *fancm-r*, the 30.8 ± 0.5 per cell of *fancm-c* and the 30.2 ± 0.7 per cell of the wild type (Figure 5B; Supplementary Figure S10). The double mutation showed significantly more chiasmata than the single-mutation plants, indicating that both genes are involved in CO processes. The *fancm-r* plants generated many more chiasmata per cell than the wild-type plants. However, in the *fancm-c* mutants, the meiotic chiasmata number per cell was not significantly different from that in the wild type. These findings indicated that *FANCM-R* conferred stronger functional complementation for the control of chiasmata number. The PMCs of the mutants produced a large amount of chromosomal stickiness that is also called interbivalent chromatin connections, a phenomenon also observed in *Arabidopsis* (32). If a bivalent chromosome stuck to another chromosome, there was at least one interbivalent chromatin connection, and the number of independent chromosome clusters decreased by one. Because many large chromosomal masses were formed by the stickiness of more than two bivalents, the exact numbers of interbivalent chromatin connections could not be determined. Therefore, we adopted the most conservative statistical method, in which the number of interbivalent chromatin connections was 18 (total
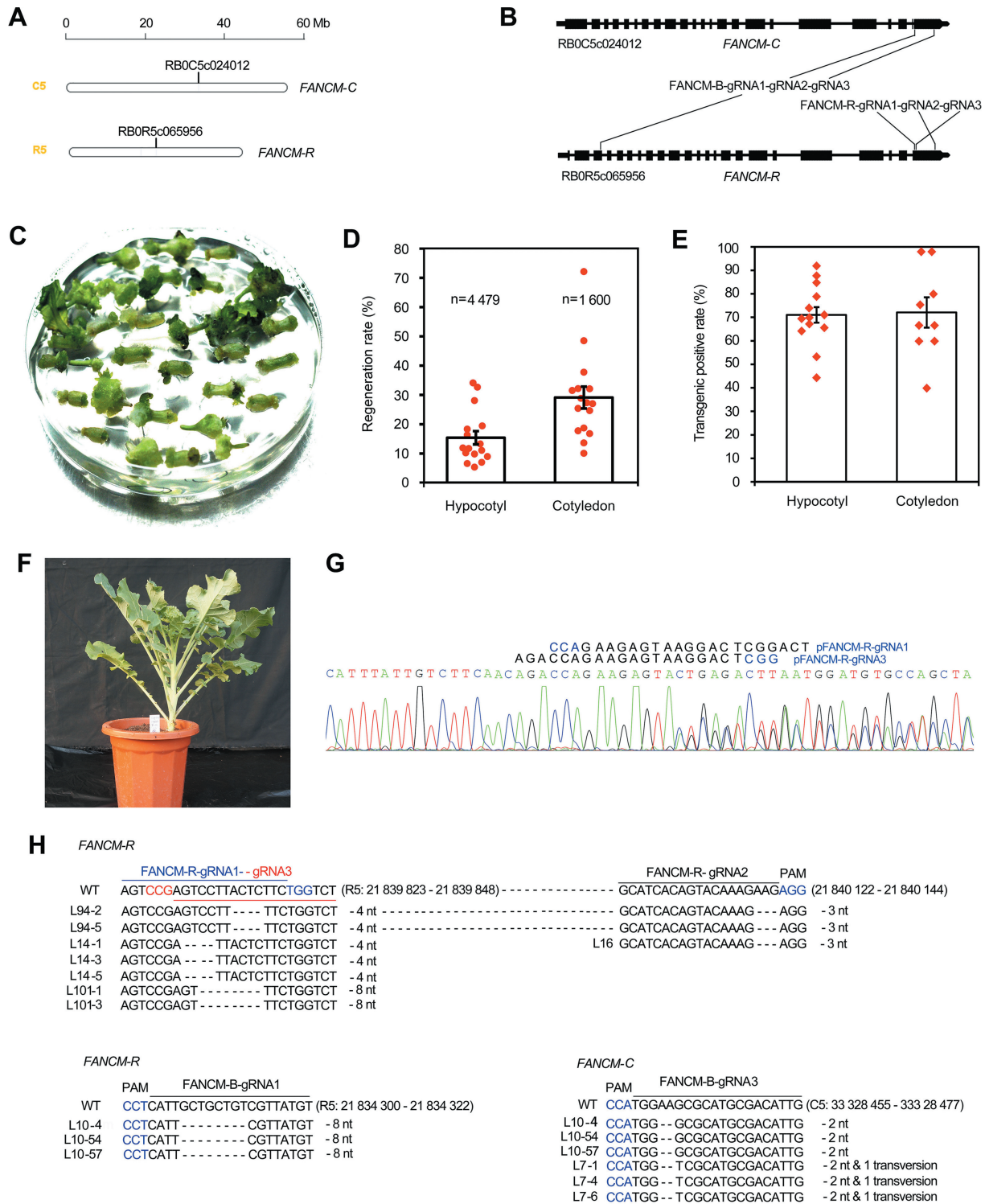
Figure 4. Targeted mutations of *FANCM* genes by CRISPR/Cas9 editing. (**A**) A pair of *FANCM* genes located on chromosomes C5 and R5. (**B**) Two CRISPR/Cas9 vectors targeting the two *FANCM* genes. (**C**) Regeneration from hypocotyl explants. (**D**) Regeneration rate of hypocotyls and cotyledon explants. (**E**) Positive rate of the regenerated plant. (**F**) A transgenic plant. (**G**) A chromatogram of a T_0 plant showing a heterozygous mutation. (**H**) Mutations in T_1 plants.
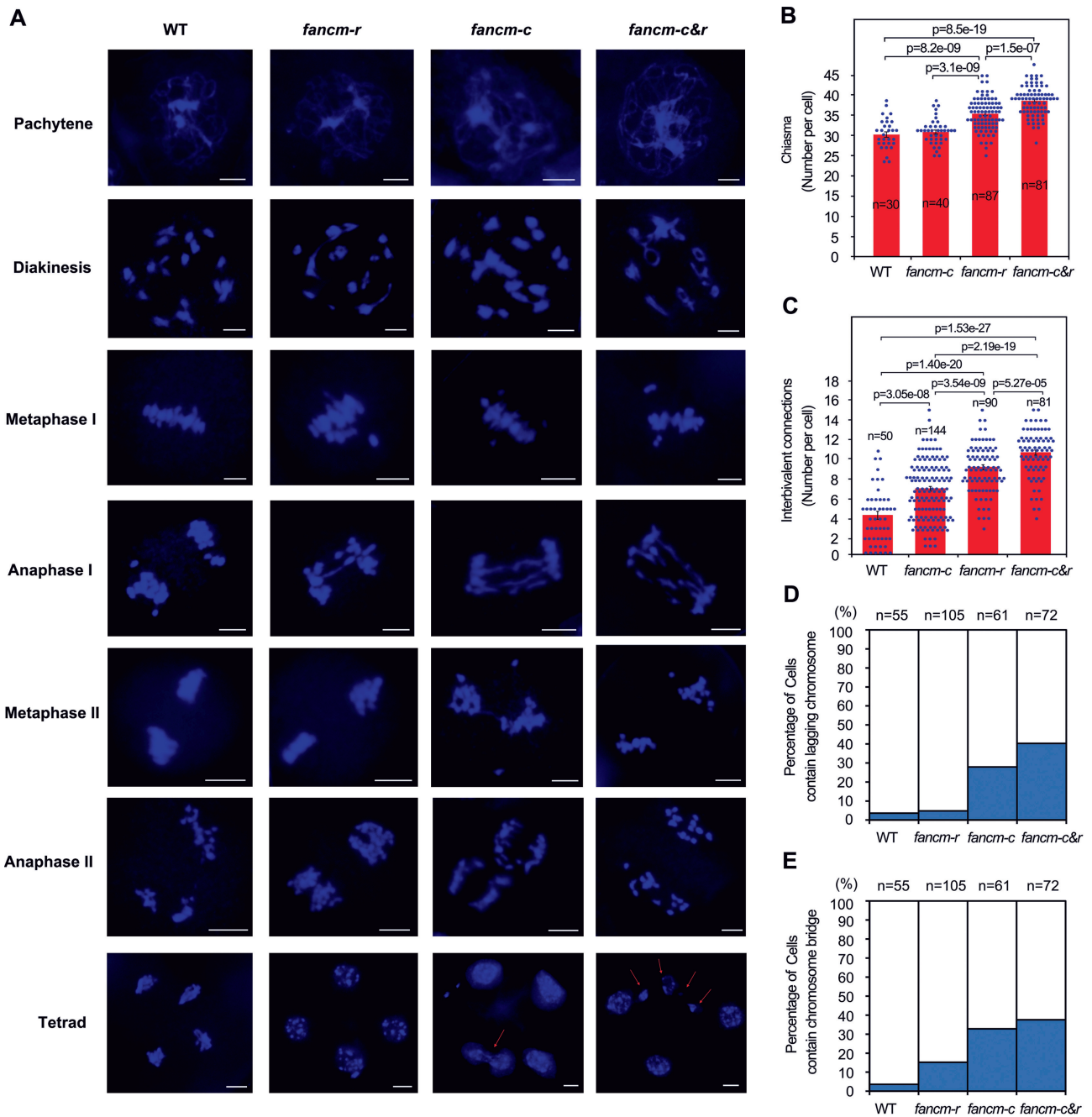
Figure 5. Meiotic chromosomes and chromosome behaviors in *fancm* mutants. (**A**) Chromosome spread of pollen mother cells. (**B**) Statistical analysis of chiasmata at diakinesis in *fancm* mutants. (**C**) Statistics of interbivalent connections at diakinesis in *fancm* mutants. (**D**) Statistics of chromosome lagging at anaphase. (**E**) Statistics of interbivalent collection at anaphase. The scale bar indicates 20 μm.

bivalents) minus the number of independent chromosomal masses in each PMC. The number of interbivalent chromatin connections per cell was determined for the *fancm-r&c* double mutant, with 10.51 ± 0.26 being significantly higher than the 9.01 ± 0.25 observed for the *fancm-r* mutant (*P*-value $= 5.27 \times 10^{-5}$, *t*-test), the 6.80 ± 0.24 observed for the *fancm-c* mutant (*P*-value $= 2.19 \times 10^{-19}$, *t*-test) and the 4.08 ± 0.41 observed for the wild type (*P*-value $= 1.53 \times 10^{-27}$, *t*-test) (Figure 5C). This indicates that homoeologous recombination was increased ~2.6-fold

in the *fancm-c&r* double mutants. The interbivalent chromatin connection number in *fancm-c* was moderate (1.67-fold) but statistically significantly (*P*-value $= 3.05 \times 10^{-8}$, *t*-test) higher than that for the wild type. However, the *fancm-c* mutant did not show a significant change in the total number of chiasmata but displayed more univalents; 28.8% ($n = 59$) of the cells in metaphase I and 27.8% ($n = 61$) of the cells in anaphase I contained one or more univalent chromosomes (free or lagging chromosomes). Lagging chromosomes were observed in 5.7% and 1.9% of anaphase

I cells in *fancm-r* and wild-type plants, respectively (Figure 5D). These results indicate the divergent functions of *FANCM-C* and *FANCM-R*. Mutation of *FANCM-C* affected the CO distribution, which was also observed in other plants, such as lettuce (*Lactuca sativa*) (85). The total number of COs did not change much, but their distribution shifted from some of the homologous chromosomes to the inter-subgenomic chromosomes, which resulted in the loss of some obligate COs (the one or more COs per homologous chromosome pair required for synapsis) while producing severe non-homologous COs that cause chromosome stickiness. A total of 32.8% ($n = 61$) of the *fancm-c* anaphase cells contained chromosome bridges, significantly higher than the 15.2% ($n = 105$) observed for *fancm-r* and the 3.6% ($n = 55$) observed for the wild type (Figure 5E). The *fancm-r* formed more chiasmata but accumulated fewer chromosome bridges, indicating that the two homologous genes were functionally biased. In all cases, double mutants displayed more serious defects, indicating that the two genes were partially complementary. The chromosome bridges indicate unresolved interbivalent chromatin connections. Further phases of meiosis will break the connections and release chromosome fragments, which form the free chromosomes in anaphase II and micronuclei in tetrads (Figures 5 and 6). In *fancm-r* mutants, only 2.1% and 1.0% of the tetrads ($n = 97$) showed free chromosomes and micronuclei, respectively. In the others, 96.9% of the total tetrads were normal, which is nearly the same as the frequency in the wild type (Figure 6A). The proportion of normal tetrads declined to 50.5% ($n = 105$) in *fancm-c* and 30.2% ($n = 149$) in the *fancm-c&r* double mutant (Figure 6A). In the *fancm-c* mutant and *fancm-c&r* double mutant, 9.5% and 10.0% of the tetrads contained free chromosomes or fragments, 11.4% and 12.1% of the tetrads contained one micronucleus, 12.4% and 8.1% of the tetrads contained two micronuclei and 5.7% and 5.4% of the tetrads contained three and more micronuclei, respectively (Figure 6A, B). This indicates that the C rather than the R genomic *FANCM* played the major role in the resolution of the interchromosome connections. The micronuclei will be eliminated from the pollen. Therefore, mutation of *fancm-c* will cause highly efficient genomic deletion and chromosome rearrangement (Figure 6C, pathway I).

FANCM is a checkpoint of meiosis and controls ploidy stability

The above-mentioned features of *fancm* mutants were within our expectations based on knowledge of *FANCM* functions in other organisms. Surprisingly, 10.5% of the *fancm-c* and 34.2% of the *fancm-c&r* tetrads entered an additional round of meiosis without genome duplication (we call this secondary meiosis hereafter) (Figure 6A, B). The tetrad cells showed additional rounds of typical metaphase, anaphase and tetrad stages, which we call metaphase III, anaphase III and tetrad III, respectively, producing smaller nuclei with fewer chromosomes (Figure 6C, pathway II). After this process, the tetrad genome (RC, allodiploid) content was further divided into two or four parts, which resulted in 1/2 RC (allohaploid) and 1/4 RC (semiploid) genomes. Compared with chromosome lag and fragment deletion, this is a much more violent pathway for ploidy reduction. At the same time, this process avoided aneuploid formation, which benefits fertility recovery and enables interspecific gene flow into the original parental gene pool without changing ploidy (here, parental gene pools are the diploid cabbage and radish gene pools; the secondary meiosis of RRCC allotetraploids generates haploid gametes, which matches the ploidy of the gametes of cabbage and radish). The natural occurrence and artificial induction of unreduced gametes that occasionally result in genome duplications have been well recognized for decades. However, the additional ploidy reduction of the gametes in meiosis has not been reported elsewhere; thus, the present study is the first to record this phenomenon, to the best of our knowledge. *FANCM* is an ancient, important gene that has been extensively investigated in a wide range of hosts, including humans, plants and non-human animals (Supplementary Figure S11). Its role in maintaining ploidy stability in meiosis has not been recognized before. One possible reason could be that most of the studies were carried out on diploid genomes; thus, the tetrads were haploid and lacked the homoeologous chromosome for synapsis. Hence, secondary metaphase I cannot occur, and secondary meiosis cannot be initiated even though restriction was lost. However, the *fancm* mutants in *B. napus* (AACC genome) did not show secondary meiosis (86). When the *FANCM-R* gene was functional, the *fancm-c* mutant still displayed the secondary meiosis phenotype, indicating that secondary meiosis is not solely the byproduct of the loss of function of *fancm* genes. The *fancm-c* mutants caused only the 47 amino acids at the C-terminus to be replaced by 31 amino acids, and the remaining 1327 amino acids of the total 1374 amino acids (96.6%) remained unchanged. Therefore, secondary meiosis may also be associated with gain of function of the *fancm-c* gene. However, it is necessary to generate a complete mutant of the *fancm-c* gene (e.g. targeting the first exons) and to complement with full-length and C-terminal replaced *FANCM-C* to prove this hypothesis. Secondary meiosis in polyploids can produce haploid gametes that can quickly generate diploid individuals, facilitating genome restabilization. *FANCM* serves as a checkpoint of meiosis; therefore, the manipulation of this gene has great application value. Taken together, these results show that the chromosome elimination pathway and the secondary meiosis pathway together led to multiple kinds of genome size variation in *fancm-c* and *fancm-c&r* mutants (Figure 6C). Microscopic examination indicated that pollen fertility declined significantly in the *fancm-c* and *fancm-c&r* mutants. However, these plants produced seeds successfully by both self- and cross-pollination, although they yielded fewer seeds than the wild-type control, indicating that these mutants are valuable for genetic research and plant breeding (Supplementary Figure S12). In particular, this could be a powerful way to generate new breeds, with a significant selection screen for the ideal or required traits from a large number of their offspring. In application, it may be necessary to cross the selected germplasm with the wild type to stabilize the genome.
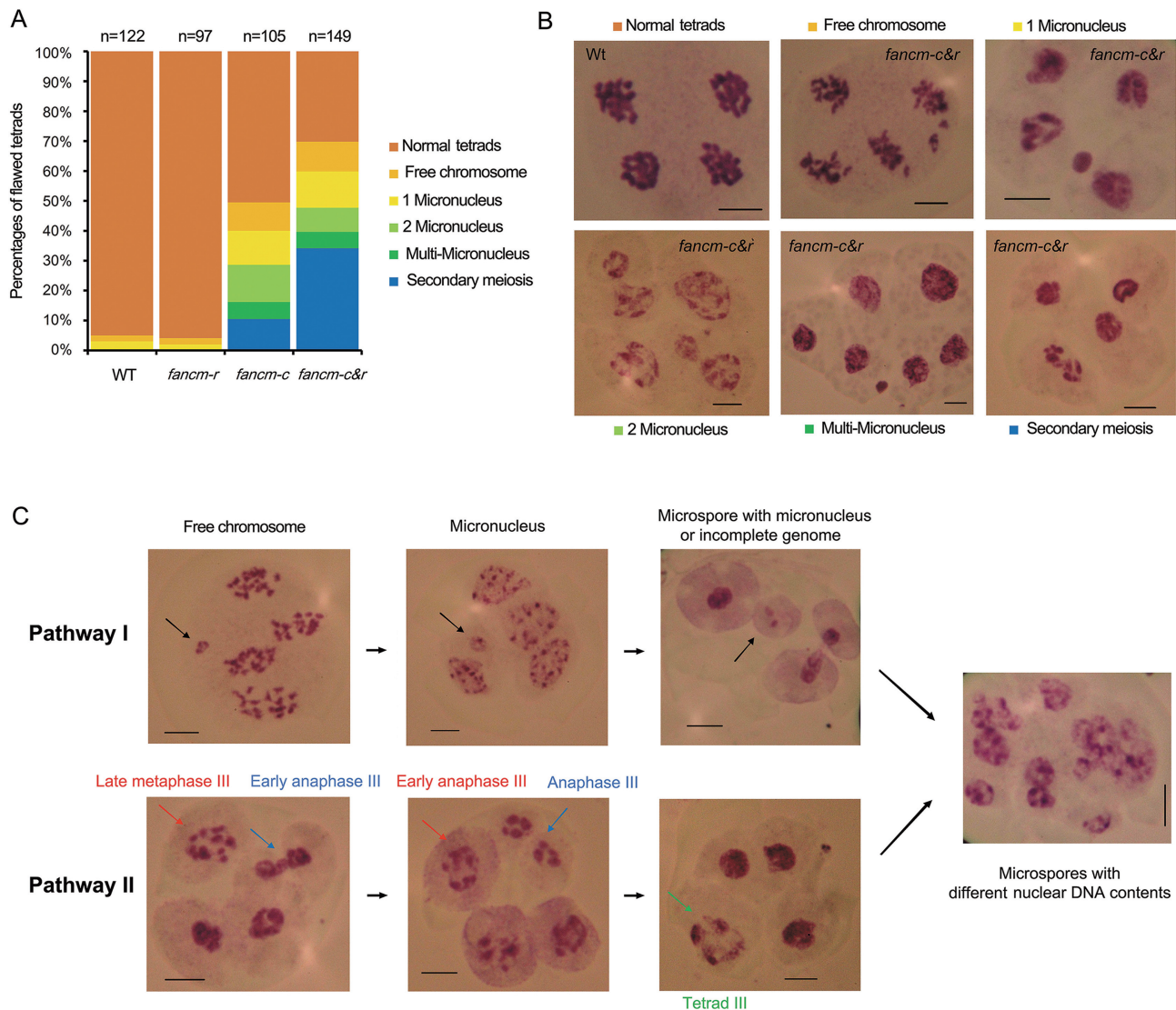
Figure 6. Chromosome behaviors at anaphase II show two pathways leading to genome size variation in *fancm* mutants. (**A**) Statistics of tetrads containing free chromosomes, micronuclei and secondary meiosis in anaphase II. (**B**) Chromosome spread showing normal tetrad cells, tetrads with free chromosomes, tetrads with one, two and multiple micronuclei, and tetrads with secondary meiosis. The scale bar indicates 10 μm. (**C**) Two pathways leading to genome size variation in *fancm* mutants. Pathway I, the chromosome lagging pathway results in the micronucleus and aneuploidy variation; pathway II, the secondary meiosis pathway of tetrads (RC) results in monoploidy (1/2 RC) and 1/2 ploidy (1/4 RC). Combining the two pathways results in diverse genome contents of the microspores in the *fancm* mutants. The scale bar indicates 10 μm.

Gene translocation and neofunctionalization were achieved by conversion and can be precisely introduced

The short fragment and homology-dependent nature of gene translocation suggested that translocation was generated by a gene conversion pathway. To confirm this hypothesis, we designed a CRISPR/Cas9 system for cutting multiple sites of *FLIP* genes. The *FLIP* gene is a key inhibitor of chromosome CO that regulates single-strand DNA invasion (29). Dysfunction of *FLIP* can improve single-stranded DNA invasion into the homoeologous DNA double strand, which is an initiation step of gene conversion.

The RRCC genome contains three full-length *FLIP* genes and a pseudogene on chromosomes R6, R8, C8 and C5 (Supplementary Figure S13A). The pseudogene in the C5 chromosome was truncated with 1–4 exons and contained

an ∼2 kb insertion in exon 5 and a large deletion covering the majority of exon 6 that extends to the 5′ end of exon 7 (Figure 7A). We designed one gRNA to simultaneously target the *FLIP* genes on chromosomes R6 and R8 and the pseudogene on chromosome C5 (Supplementary Figure S13B). The CRISPR-gRNA was transformed into the RRCC plant. The allele-specific PCR primers (P1F and P1R) were designed for amplification of fragments harboring each of the targets (Figure 7A). In several transgenic plants, the PCR produced nothing for the *FLIP* gene located on R6. By testing with additional primers located upstream (P2F and P2R) and downstream (P3F and P3R) of the CRISPR cutting site (Figure 7A), we enriched the upstream fragment but not the downstream sequences. To determine what happened in this region, we amplified the
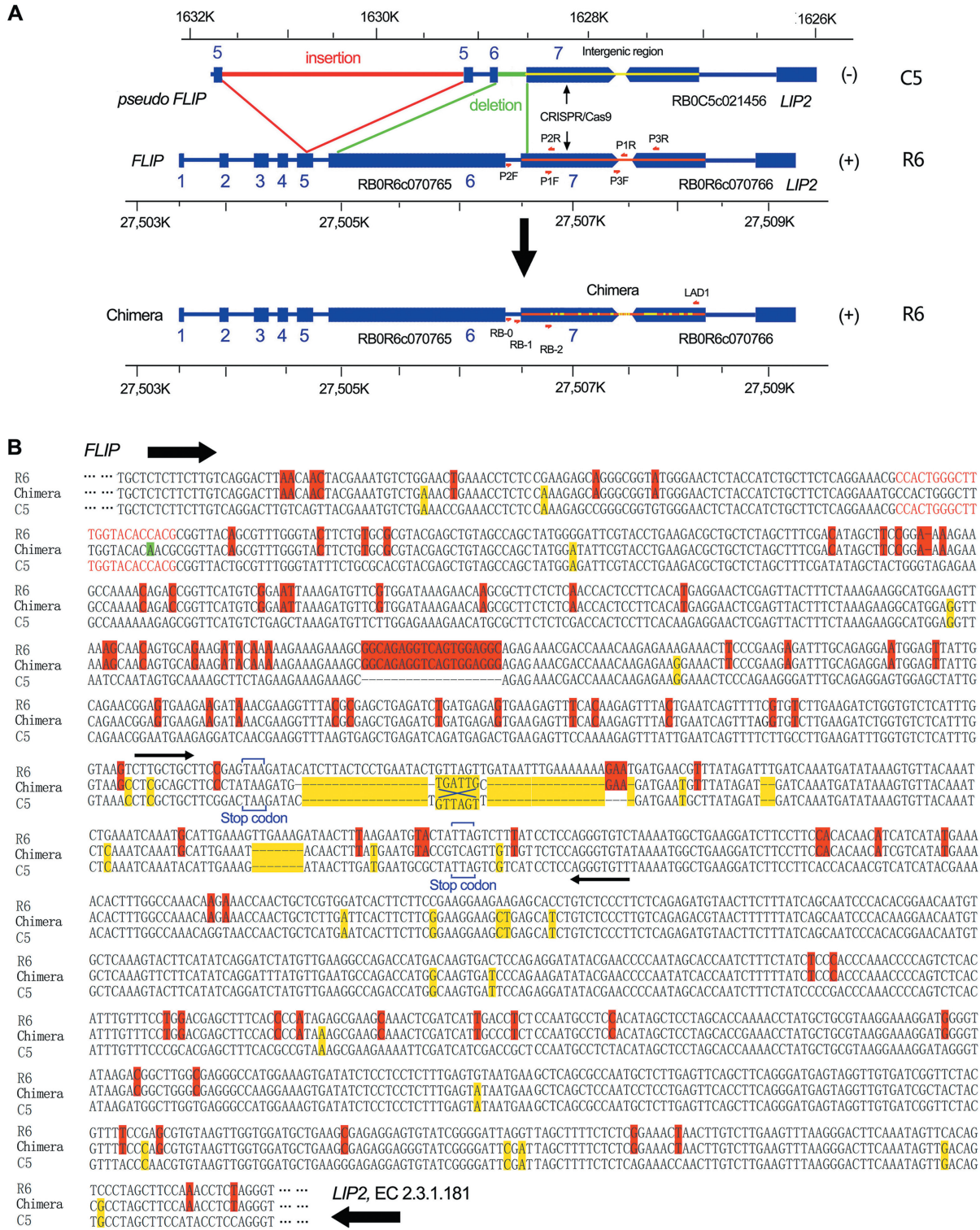
Figure 7. Targeted conversion of *FLIP* genes between C5 and R6. (**A**) Schematic diagram of recombination between the C5 and R6 chromosomal regions. The pseudo-*FLIP* on C5 has a 1–4 exon truncation and contains a large insertion in exon 5 and a large deletion covering the majority of exon 6 that extends to the 5′ end of exon 7. CRISPR/Cas9 targeted the middle of exon 7 of both C5 and R6 *FLIP* genes. *LIP2* and *FLIP* are tail-to-tail adjacent genes with only 100 bp and 146 bp intergenic regions in C5 and R6, respectively. P1F&P1R, P2F&P2R and P3F&P3R are the primer pairs (sequences shown in Supplementary Table S4). Chimeras are indicated by the red and yellow mosaic in the chromosomal region. RB-0, RB-1, RB-2 and LAD1 indicate the TAIL-PCR primers (sequences shown in Supplementary Table S4). (**B**) Detailed sequences of wild C5, R6 and the chimeric genes formed by intergeneric conversion. R6, wild-type sequence from the R6 chromosome; Chimera, chimeric gene; C5, wild-type sequence from the C5 chromosome; red shading indicates SNPs and indels the same as in R6; yellow shading indicates SNPs and indels the same as in C5; recombination indicated by the chimera of R6 and C5 type markers. Red letters indicate the gRNA target, and green shading indicates the single-nucleotide mutation. The 'X' indicates the mirror symmetry of 'TGATTG'. The chimeric fragment covered at least the last exons of *FLIP* and *LIP2* as well as their intergenic region.

unknown flanking fragment via hiTAIL-PCR (76). Sequencing of the PCR product revealed that the *FLIP* genes on chromosomes R6 and C5 have undergone gene conversion, resulting in a chimeric gene on chromosome R6 (Figure 7B). The reason for the failure of the first PCR for the R6 *FLIP* segment was that the reverse primer region was replaced by the allelic C5 sequence. The C5 origin chimeric fragment ranged from several nucleotides to ~200 nt (Figure 7B). The chimeric regions extended at least 1200 bp from the CRISPR cutting site, covering exon 2 of the tail-to-tail adjacent gene encoding octanoyltransferase (LIP2; EC 2.3.1.181) (Figure 7B). The sgRNA target site was mutated by C to A transversion. The chimeric *FLIP* obtained one and three amino acids from the C5 and R6 alleles, respectively, and produced six novel amino acids (Supplementary Figure S14). Chimeric *LIP2* gained four and six amino acids from the C5 and R6 alleles, respectively, and produced 11 novel amino acids (Supplementary Figure S15). The plant showed no significant difference in either morphology or fertility compared with the wild type (Supplementary Figure S16). Compared with CRISPR/Cas9-mediated conversion in tomato (17), the present conversion was not only much more efficient but also qualitatively advanced. In tomato, the homologous sequences are highly similar [2.61 single nucleotide polymorphisms (SNPs) and 0.39 insertions/deletions (indels) per kb], and their recombination is not inhibited naturally. In the present study, R6 and C5 never recombined naturally, and the sequences contained 28.5 and 10.5 times higher densities of SNPs and indels (74.4 SNPs and 4.1 indels per kb). The high efficiency of gene conversion indicates that mutation of the *FLIP* gene played an important role in promoting gene conversion along with precise cutting by CRISPR/Cas. We did not detect gene conversion between either R6 and R8 or R8 and C5. One reason could be that these two pairs of homologous genes did not reside in the major homologous region. R6 and C5 are homoeologous and span half of each of the chromosomes (Figure 3), which facilitates interchromosome pairing. The *FLIP* genes resided at the end of both chromosomes, which also facilitates gene conversion, as mentioned above (Figure 3). These results indicated that the interhomologous translocations were truly potentially derived from gene conversion. This process was elevated by editing *FLIP* and precisely introducing it at the target site by double-strand cutting with CRISPR/Cas9. The gene conversion-generated novel alleles could gain new functions. Thus, this method has broad application value in plant breeding.

DISCUSSION

Interspecific and intergeneric hybridization as well as the generated allopolyploids are valuable for both genetic research and breeding applications. Many synthetic allopolyploid plants, including *Raphanobrassica* (*R. sativus* × *B. oleracea*), have been produced. Except for an interspecific *Cucumis* (*C. sativus* × *C. hystrix*) and an allotetraploid *Glycine* (*G. dolichocarpa*; *G. syndetika* × *G. tomentella* D3), few synthetic allopolyploids have been whole-genome sequenced (87,88). To the best of our knowledge, the present study reports the first genome assembly of an intergeneric hybrid allopolyploid plant. Taking advantage of this high-quality genome assembly, we attempted to answer long-standing questions about allopolyploid genomic changes after distant hybridization.

The number of small deletions (harboring <20 genes) and large-fragment deletions (harboring >20 genes) showed different patterns in the R and C subgenomes, indicating that small and large deletions were generated via different mechanisms. The small deletions formed by small structure variations (SVs) cause gene truncation or the entire gene deletion without affecting the general genome arrangement (69). Therefore, the C subgenome contained more small deletions but maintained better integrity and collinearity.

Among different types of *B. oleracea*, the total gene number diverged significantly, ranging from 43 868 to 62 232 (Supplementary Table S16), indicating that some of the so-called deletion genes were caused by the background variations between the reference (HDEM) and the paternal parent (nFMK or PC81). If we instead use the D134 genome, which contains 43 868 genes, close to the number of genes (44 555) in the C subgenome of RRCC plants, as a reference for comparison, significantly fewer gene deletions could be identified. The genomes of the parents are ideal comparison references that could generate true deletion maps. Unfortunately, the genomes of the parent lines are not currently available. However, the concept of the pangenome was developed in recent years, the pangenomes of *Raphanus* and *B. oleracea* were constructed and multiple *B. oleracea* genomes were released, which inspired us to compare our RRCC genome against the pangenomes of *Raphanus* and *B. oleracea* (69,89). The core and dispensable genes were biased toward retention and deletion, respectively, indicating that selection pressure acted heavily on gene retention and deletion. The same trends were also observed in soybean (88).

The large deletions were enriched in the TE-rich regions. One reason is that these regions contain fewer genes, especially core genes; thus, the deletions conferred less pressure on plant viability. Another reason could be that the TE-rich regions were condensed to form heterochromatin regions, where a DSB, once formed, will hardly be precisely repaired by the SDSA pathway. The degradation of the DSB ends will generate large deletions. The unrepaired double DSB can cause large fragment lags, resulting in superlong deletions, a phenomenon frequently appearing in *fancm-c* and *fancm-c&r* mutants. However, this kind of heterochromatic region rarely forms DSBs in normal cells. In interspecific hybrids, homoeologous recombination or just interbivalent connection in homoeologous euchromosome regions, especially chromosome ends, can cause chromosome bridges in heterochromatic regions located in the middle of chromosomes. These will break when the connected chromosomes segregate to the opposite poles at anaphase. The hotspots of the large deletions were located particularly in several heterochromatin regions of R genome chromosomes. This is consistent with the fact that several chromosomes in the RRCC plant are always collected together as interbivalent connections. In the *fancm-c&r* mutants, interbivalent connections increased at least 2.5-fold and showed high-frequency chromosomal lagging and micronuclei popout. Therefore, manipulating *FANCM* genes in allopoly-

ploids is a powerful method for accelerating genome shuffling.

Why genome evolution does not show the same speed on different branches of the plant kingdom phylogeny is a long-standing question. Angiosperm genomes have experienced several rounds of polyploidization and shuffling, and continue to undergo fragmental deletion and ploidy reduction, while bryophyte, pteridophyte and gymnosperm genomes have remained relatively stable even over longer evolutionary time scales (4,7). The mechanism controlling the switch from stasis to polyploidization and shuffling is a mystery. DSB misrepair was assumed to be the main source of genome shrinkage, expansion and equilibrium in allopolyploids (90). *FANCM* is a key DSB DNA repair component. As indicated herein and in previous studies, the mutation of *FANCM* can significantly enhance homoeologous exchange (HE), chromosomal deletion and shuffling, and ploidy reduction, which are fundamental for the evolution of allopolyploid genomes. We analyzed FANCM on all branches of the plant kingdom phylogeny. FANCM exists in all kinds of plants, and the protein sequences are highly conserved, with the exception of the C-terminal domain, where large deletions accumulate in angiosperms (Supplementary Figure S11). The mutations in angiosperms could be the reason for the enhanced genome polyploidization and shuffling, which lead to reproductive isolation and subsequent speciation. In the present study, the frameshift of 47 amino acids at the C-terminus in *fancm-c* resulted in severe phenotypes, including not only those reported in the literature, such as enhanced HE, but also novel traits of secondary meiosis that have not been reported. Therefore, FANCM could be one of the governors of genome evolution speed. In gymnosperms, FANCM maintained stronger functions to restrict homoeologous chromosome pairing and genome shuffling. In angiosperms, FANCM is mutated step by step, which increasingly relaxes the restriction of HE and genome shuffling, resulting in corresponding genome polyploidization and ploidy reduction. Thus, *FANCM* is a key gene controlling genome stability and an ideal target of editing for improving allopolyploid genome shuffling for genetic research and homoeologous recombination for plant breeding.

Most of the intergenomic translocations in the RRCC genome occurred between the R and C syntenic regions, indicating that they evolved in a homoeologous sequence-dependent gene conversion pathway rather than in a transposon-like manner. In the *FLIP* cut lines, conversions between R6 and C5 were achieved in the two genes around the cutting site. From the perspective of sequence polymorphism, this kind of gene conversion was not only substantially equivalent to, but also much better than, translocations for genetic diversity. Therefore, homoeologous gene conversion is, more than expected, a mechanism for new gene evolution or neofunctionalization of duplicate gene sets in allopolyploids.

FLIP associates with FIGL1 to inhibit single-strand invasion into homologous double strands (29). By CRISPR cutting multiple sites of the *FLIP* genes, we successfully obtained a new form of chimeric *FLIP* and *LIP2* genes. Double cutting has been used to generate reciprocal chromosome end translocation in the *Arabidopsis* genome (19).

We did not obtain this type of translocation in this study, possibly because we did not screen as many individuals as has been done in *Arabidopsis* research. NCO recombination (conversion) at the designed position was also achieved in tomato tissue cells using the CRISPR cutting method (17). The tomatoes used in that study had no reproductive isolation, and basal homologous recombination was normal during meiosis. Therefore, the improvement of recombination at the target site is inferential and predictable. In the present study, we show for the first time that gene conversion can be precisely introduced and genetic exchanges can be successfully achieved between two reproductively isolated genera. This is a long-awaited technology for genetic research and breeding applications. Although conducted in a much more difficult plant system, our targeted conversion was far more efficient than that in the tomato study. The reason could be that the cutting of *FLIP* relaxed the constraint on single-strand invasion; thus, the *FLIP* break end of R6 had a higher chance of invading the double strand of C5 homologs, which served as a template to form a HJ. The HJs slid from the invading site up- and downstream, which copied the C5 sequence and generated heteroduplex regions. The subsequent mismatch repair in this region may have generated chimeric genes with discontinuous conversion tracts (17). Chimeric *FLIP* and *LIP2* obtained allelic and *de novo* amino acids but avoided frameshift indels, indicating selection pressure to keep the gene functional. Whether these new types of genes improve or reduce functionality or even generate novel functions should be clarified by further experiments. This kind of gene shuffling provides the possibility of screening for variations toward desired directions and provides a broad imagination space for utilization in breeding. In the *FLIP* editing lines, we also successfully generated conversions in the intergenic region and improved sequence novelty. The intergenic regions contain important *cis*-acting regulatory elements, and many important agronomic traits are regulated by variations in intergenic regions. The shuffling of intergenic regions also has great application value. It is worth investigating if any pair of homologous genes cut by CRISPR can produce high-efficiency conversion in the *flip* genotype. If this is true, we can induce neofunctionalization of most of the important genes by design.

The present study indicates that the RRCC intergeneric hybrid can be used as a model for investigating allopolyploid genome evolution. By editing *FANCM*, allopolyploid genome evolution can be accelerated by improving HE, genome shuffling and deletion, and secondary meiosis-derived ploidy reduction. In terms of application, the efficiency of HE was low, and the exchange between the R and C genomes was shown to be difficult in the wild-type RRCC plant. The polyploidy of the interspecific hybrids restricted the backcrossing with their parental species. Therefore, mutants with high-efficiency HE and producing ploidy-reduced gametes are valuable materials for germplasm innovation and plant breeding.

Targeted conversion of homologous genes from intergeneric chromosomes was stimulated by CRISPR cutting of *FLIP*. The chimeric genes and intergenic regions can potentially generate novel functions. Given that recombination between homoeologous chromosomes in the intergeneric

hybrid is rare, this technology is valuable for artificial gene evolution and other applications.

In conclusion, the present study revealed allopolyploid genome evolution using third-generation genome sequencing and comparative genomic analysis. The genome shuffling processes were monitored and improved by genome editing with CRISPR/Cas9. Overall, rapid genome ploidy reduction, highly efficient genome-wide chromosomal shuffling and precise gene conversion technology were established in the RRCC system, which will serve as a model for experimental studies of allopolyploid genome evolution and benefit germplasm innovation and plant breeding.

DATA AVAILABILITY

The assembled genome and annotations of *Raphanobrassica* (RRCC) have been deposited in the GenomeWarehouse (GWH) database of the National Genomics Data Center (BIG Data Center) (https://bigd.big.ac.cn) under accession number GWHBJWS00000000. All original sequence data have been deposited in the Genome Sequence Archive (GSA) database of the National Genomics Data Center under accession number CRA007704.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Mallet,J. (2007) Hybrid speciation. *Nature*, **446**, 279–283.
2. Cerca,J., Petersen,B., Lazaro-Guevara,J.M., Rivera-Colón,A., Birkeland,S., Vizueta,J., Li,S., Li,Q., Loureiro,J., Kosawang,C. *et al.* (2022) The genomic basis of the plant island syndrome in Darwin's giant daisies. *Nat. Commun.*, **13**, 3729.
3. Chen,J., Luo,M., Li,S., Tao,M., Ye,X., Duan,W., Zhang,C., Qin,Q., Xiao,J. and Liu,S. (2018) A comparative study of distant hybridization in plants and animals. *Sci. China Life Sci.*, **61**, 285–309.
4. Alix,K., Gerard,P.R., Schwarzacher,T. and Heslop-Harrison,J.S. (2017) Polyploidy and interspecific hybridization: partners for adaptation, speciation and evolution in plants. *Ann. Bot.*, **120**, 183–194.
5. Wang,X., Morton,J., Pellicer,J., Leitch,I.J. and Leitch,A.R. (2021) Genome downsizing after polyploidy: mechanisms, rates and selection pressures. *Plant J.*, **107**, 1003–1015.
6. Bayat,S., Lysak,M.A. and Mandáková,T. (2021) Genome structure and evolution in the cruciferous tribe Thlaspideae (Brassicaceae). *Plant J.*, **108**, 1768–1785.
7. Cheng,F., Wu,J., Cai,X., Liang,J., Freeling,M. and Wang,X. (2018) Gene retention, fractionation and subgenome differences in polyploid plants. *Nat. Plants*, **4**, 258–268.
8. Gaeta,R.T., Pires,J.C., Iniguez-Luy,F., Leon,E. and Osborn,T.C. (2007) Genomic changes in resynthesized *Brassicanapus* and their effect on gene expression and phenotype. *Plant Cell*, **19**, 3403–3417.
9. Ferreira de Carvalho,J., Stoeckel,S., Eber,F., Lodé-Taburel,M., Gilet,M.M., Trotoux,G., Morice,J., Falentin,C., Chèvre,A.M. and Rousseau-Gueutin,M. (2021) Untangling structural factors driving genome stabilization in nascent *Brassicanapus* allopolyploids. *New Phytol.*, **230**, 2072–2084.
10. Desjardins,S.D., Simmonds,J., Guterman,I., Kanyuka,K., Burridge,A.J., Tock,A.J., Sanchez-Moran,E., Franklin,F.C.H., Henderson,I.R., Edwards,K.J. *et al.* (2022) FANCM promotes class I interfering crossovers and suppresses class II non-interfering crossovers in wheat meiosis. *Nat. Commun.*, **13**, 3644.
11. Panday,A., Willis,N.A., Elango,R., Menghi,F., Duffey,E.E., Liu,E.T. and Scully,R. (2021) FANCM regulates repair pathway choice at stalled replication forks. *Mol. Cell*, **81**, 2428–2444.
12. Harberd,D.J. and McArthur,E.D. (1980) Meiotic analysis of some species and genus hybrids in the Brassiceae. In: Tsunoda,S., Hinata,K. and Gomez-Campo,C. (eds). *Brassica Crops and Wild Allies: Biology and Breeding*. Japan Scientific Societies Press, Tokyo.
13. Peterka,H., Budahn,H., Schrader,O., Ahne,R. and Schutze,W. (2004) Transfer of resistance against the beet cyst nematode from radish (*Raphanus sativus*) to rape (*Brassica napus*) by monosomic chromosome addition. *Theor. Appl. Genet.*, **109**, 30–41.
14. Giancola,S., Marhadour,S., Desloire,S., Clouet,V., Falentin-Guyomarc'h,H., Laloui,W., Falentin,C., Pelletier,G., Renard,M., Bendahmane,A. *et al.* (2003) Characterization of a radish introgression carrying the ogura fertility restorer gene *rfo* in rapeseed, using the *Arabidopsis* genome sequence and radish genetic mapping. *Theor. Appl. Genet.*, **107**, 1442–1451.
15. Rönspies,M., Dorn,A., Schindele,P. and Puchta,H. (2021) CRISPR–Cas-mediated chromosome engineering for crop improvement and synthetic biology. *Nat. Plants*, **7**, 566–573.
16. Sauer,N.J., Narvaez-Vasquez,J., Mozoruk,J., Miller,R.B., Warburg,Z.J., Woodward,M.J., Mihiret,Y.A., Lincoln,T.A., Segami,R.E., Sanders,S.L. *et al.* (2016) Oligonucleotide-mediated genome editing provides precision and function to engineered nucleases and antibiotics in plants. *Plant Physiol.*, **170**, 1917–1928.
17. Filler Hayut,S., Melamed Bessudo,C. and Levy,A.A. (2017) Targeted recombination between homologous chromosomes for precise breeding in tomato. *Nat. Commun.*, **8**, 15605.
18. Sadhu,M.J., Bloom,J.S., Day,L. and Kruglyak,L. (2016) CRISPR-directed mitotic recombination enables genetic mapping without crosses. *Science*, **352**, 1113–1116.
19. Beying,N., Schmidt,C., Pacher,M., Houben,A. and Puchta,H. (2020) CRISPR-Cas9-mediated induction of heritable chromosomal translocations in *Arabidopsis. Nat. Plants*, **6**, 638–645.
20. Schmidt,C., Pacher,M. and Puchta,H. (2019) Efficient induction of heritable inversions in plant genomes using the CRISPR/Cas system. *Plant J.*, **98**, 577–589.
21. Schwartz,C., Lenderts,B., Feigenbutz,L., Barone,P., Llaca,V., Fengler,K. and Svitashev,S. (2020) CRISPR-Cas9-mediated 75.5-Mb inversion in maize. *Nat. Plants*, **6**, 1427–1431.
22. Fayos,I., Mieulet,D., Petit,J., Meunier,A.C., Perin,C., Nicolas,A. and Guiderdoni,E. (2019) Engineering meiotic recombination pathways in rice. *Plant Biotechnol. J.*, **17**, 2062–2077.
23. Lambing,C., Franklin,F.C.H. and Wang,C.J.R. (2017) Understanding and manipulating meiotic recombination in plants. *Plant Physiol.*, **173**, 1530–1542.

24. Mieulet,D., Aubert,G., Bres,C., Klein,A., Droc,G., Vieille,E., Rond-Coissieux,C., Sanchez,M., Dalmais,M., Mauxion,J.P. *et al.* (2018) Unleashing meiotic crossovers in crops. *Nat. Plants*, **4**, 1010–1016.

25. Marsolier-Kergoat,M.C., Khan,M.M., Schott,J., Zhu,X. and Llorente,B. (2018) Mechanistic view and genetic control of DNA recombination during meiosis. *Mol. Cell*, **70**, 9–20.

26. Crismani,W., Girard,C., Froger,N., Pradillo,M., Santos,J.L., Chelysheva,L., Copenhaver,G.P., Horlow,C. and Mercier,R. (2012) FANCM limits meiotic crossovers. *Science*, **336**, 1588–1590.

27. Girard,C., Crismani,W., Froger,N., Mazel,J., Lemhemdi,A., Horlow,C. and Mercier,R. (2014) FANCM-associated proteins MHF1 and MHF2, but not the other Fanconi anemia factors, limit meiotic crossovers. *Nucleic Acids Res.*, **42**, 9087–9095.

28. Seguela-Arnaud,M., Crismani,W., Larcheveque,C., Mazel,J., Froger,N., Choinard,S., Lemhemdi,A., Macaisne,N., Van Leene,J., Gevaert,K. *et al.* (2015) Multiple mechanisms limit meiotic crossovers: TOP3alpha and two BLM homologs antagonize crossovers in parallel to FANCM. *Proc. Natl Acad. Sci. USA*, **112**, 4713–4718.

29. Fernandes,J.B., Duhamel,M., Seguela-Arnaud,M., Froger,N., Girard,C., Choinard,S., Solier,V., De Winne,N., De Jaeger,G., Gevaert,K. *et al.* (2018) FIGL1 and its novel partner FLIP form a conserved complex that regulates homologous recombination. *PLoS Genet.*, **14**, e1007317.

30. de Maagd,R.A., Loonen,A., Chouaref,J., Pele,A., Meijer-Dekens,F., Fransz,P. and Bai,Y. (2020) CRISPR/Cas inactivation of RECQ4 increases homeologous crossovers in an interspecific tomato hybrid. *Plant Biotechnol. J.*, **18**, 805–813.

31. Fernandes,J.B., Seguela-Arnaud,M., Larcheveque,C., Lloyd,A.H. and Mercier,R. (2018) Unleashing meiotic crossovers in hybrid plants. *Proc. Natl Acad. Sci. USA*, **115**, 2431–2436.

32. Knoll,A., D. Higgins,J.D., Seeliger,K., Reha,S.J., Dangel,N.J., Bauknecht,M., Schröpfer,S., Franklin,F.C. and Puchtaa,H. (2012) The Fanconi anemia ortholog FANCM ensures ordered homologous recombination in both somatic and meiotic cells in *Arabidopsis*. *Plant Cell*, **24**, 1448–1464.

33. Sidhu,G.K., Fang,C., Olson,M.A., Falque,M., Martin,O.C. and Pawlowski,W.P. (2015) Recombination patterns in maize reveal limits to crossover homeostasis. *Proc. Natl Acad. Sci. USA*, **112**, 15982–15987.

34. Wang,K., Wang,C., Liu,Q., Liu,W. and Fu,Y. (2015) Increasing the genetic recombination frequency by partial loss of function of the synaptonemal complex in rice. *Mol. Plant*, **8**, 1295–1298.

35. Marraffini,L.A. and Sontheimer,E.J. (2010) Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature*, **463**, 568–571.

36. Jinek,M., Chylinski,K., Fonfara,I., Hauer,M., Doudna,J.A. and Charpentier,E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816–821.

37. Tang,X., Lowder,L.G., Zhang,T., Malzahn,A.A., Zheng,X., Voytas,D.F., Zhong,Z., Chen,Y., Ren,Q., Li,Q. *et al.* (2017) A CRISPR-Cpf1 system for efficient genome editing and transcriptional repression in plants. *Nat. Plants*, **3**, 17018.

38. Cong,L., Ran,F.A., Cox,D., Lin,S., Barretto,R., Habib,N., Hsu,P.D., Wu,X., Jiang,W., Marraffini,L.A. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819–823.

39. Lawrenson,T., Shorinola,O., Stacey,N., Li,C., Ostergaard,L., Patron,N., Uauy,C. and Harwood,W. (2015) Induction of targeted, heritable mutations in barley and *Brassicaoleracea* using RNA-guided cas9 nuclease. *Genome Biol.*, **16**, 258.

40. Puchta,H. (2016) Applying CRISPR/Cas for genome engineering in plants: the best is yet to come. *Curr. Opin. Plant Biol.*, **36**, 1–8.

41. Liang,Z., Chen,K., Li,T., Zhang,Y., Wang,Y., Zhao,Q., Liu,J., Zhang,H., Liu,C., Ran,Y. *et al.* (2017) Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. *Nat. Commun.*, **8**, 14261.

42. Gil-Humanes,J., Wang,Y., Liang,Z., Shan,Q., Ozuna,C.V., Sanchez-Leon,S., Baltes,N.J., Starker,C., Barro,F., Gao,C. *et al.* (2016) High-efficiency gene targeting in hexaploid wheat using DNA replicons and CRISPR/Cas9. *Plant J.*, **89**, 1251–1262.

43. Xie,K., Minkenberg,B. and Yang,Y. (2015) Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc. Natl Acad. Sci. USA*, **112**, 3570–3575.

44. Ding,D., Chen,K., Chen,Y., Li,H. and Xie,K. (2018) Engineering introns to express RNA guides for Cas9- and Cpf1-mediated multiplex genome editing. *Mol. Plant*, **11**, 542–552.

45. Kim,D., Langmead,B. and Salzberg,S.L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods*, **12**, 357–360.

46. Marçais,G. and Kingsford,C. (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, **27**, 764–770.

47. Koren,S., Walenz,B.P., Berlin,K., Miller,J.R., Bergman,N.H. and Phillippy,A.M. (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.*, **27**, 722–736.

48. Chin,C.S., Alexander,D.H., Marks,P., Klammer,A.A., Drake,J., Heiner,C., Clum,A., Copeland,A., Huddleston,J. and Eichler,E.E. (2013) Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods*, **10**, 563.

49. Walker,B.J., Abeel,T., Shea,T., Priest,M., Abouelliel,A., Sakthikumar,S., Cuomo,C.A., Zeng,Q., Wortman,J., Young,S.K. *et al.* (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*, **9**, e112963.

50. Durand,N.C., Shamim,M.S., Machol,I., Rao,S.S., Huntley,M.H., Lander,E.S. and Aiden,E.L. (2016) Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.*, **3**, 95–98.

51. Dudchenko,O., Batra,S.S., Omer,A.D., Nyquist,S.K., Hoeger,M., Durand,N.C., Shamim,M.S., Machol,I., Lander,E.S., Aiden,A.P. *et al.* (2017) De novo assembly of the *Aedesaegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*, **356**, 92–95.

52. Jurka,J. (2000) Repbase update: a database and an electronic journal of repetitive elements. *Trends Genet.*, **16**, 418–420.

53. Flynn,J.M., Hubley,R., Goubert,C., Rosen,J., Clark,A.G., Feschotte,C. and Smit,A.F. (2020) RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl Acad. Sci. USA*, **117**, 9451–9457.

54. Ou,S.J. and Jiang,N. (2018) LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.*, **176**, 1410–1422.

55. Lowe,T.M. and Eddy,S.R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.*, **25**, 955–964.

56. Kalvari,I., Argasinska,J., Quinones-Olvera,N., Nawrocki,E.P., Rivas,E., Eddy,S.R., Bateman,A., Finn,R.D. and Petrov,A.I. (2018) Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res.*, **46**, D335–D342.

57. Keilwagen,J., Hartung,F., Paulini,M., Twardziok,S.O. and Grau,J. (2018) Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *BMC Bioinformatics*, **19**, 189.

58. Trapnell,C., Roberts,A., Goff,L., Pertea,G., Kim,D., Kelley,D.R., Pimentel,H., Salzberg,S.L., Rinn,J.L. and Pachter,L. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with tophat and cufflinks. *Nat. Protoc.*, **7**, 562–578.

59. Haas,B.J., Delcher,A.L., Mount,S.M., Wortman,J.R., Smith,R.K. Jr, Hannick,L.I., Maiti,R., Ronning,C.M., Rusch,D.B., Town,C.D *et al.* (2003) Improving the arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.*, **31**, 5654–5666.

60. Stanke,M., Steinkamp,R., Waack,S. and Morgenstern,B. (2004) AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res.*, **32**, W309–W312.

61. Majoros,W.H., Pertea,M. and Salzberg,S.L. (2004) TigrScan and glimmerhmm: two open source ab initio eukaryotic gene-finders. *Bioinformatics*, **20**, 2878–2879.

62. Korf,I. (2004) Gene finding in novel genomes. *BMC Bioinformatics*, **5**, 59.

63. Borodovsky,M. and Lomsadze,A. (2011) Eukaryotic gene prediction using GeneMark.hmm-E and GeneMark-ES. *Curr. Protoc. Bioinformatics*, **Chapter 4**, 4.6.1–4.6.10.

64. Haas,B.J., Salzberg,S.L., Zhu,W., Pertea,M., Allen,J.E., Orvis,J., White,O., Buell,C.R. and Wortman,J.R. (2008) Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol.*, **9**, R7.

65. Conesa,A., Götz,S., García-Gómez,J.M., Terol,J., Talón,M. and Robles,M. (2005) Blast2GO: a universal tool for annotation,

visualization and analysis in functional genomics research. *Bioinformatics*, **21**, 3674–3676.

66. Xie,C., Mao,X., Huang,J., Ding,Y., Wu,J., Dong,S., Kong,L., Gao,G., Li,C.Y. and Wei,L. (2011) KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.*, **39**, W316–W322.

67. Jones,P., Binns,D., Chang,H.Y., Fraser,M., Li,W., McAnulla,C., McWilliam,H., Maslen,J., Mitchell,A. and Nuka,G. (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics*, **30**, 1236–1240.

68. Wang,Y., Tang,H., Debarry,J., Tan,X., Li,J., Wang,X., Lee,T., Jin,H., Marler,B., Guo,H. *et al.* (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.*, **40**, e49.

69. Zhang,X.H., Liu,T.J., Wang,J.L., Wang,P., Qiu,Y., Zhao,W., Pang,S., Li,X.M., Wang,H.P., Song,J.P. *et al.* (2021) Pan-genome of *Raphanus* highlights genetic variation and introgression among domesticated, wild, and weedy radishes. *Mol. Plant*, **14**, 2032–2055.

70. Belser,C., Istace,B., Denis,E., Dubarry,M., Baurens,F.C., Falentin,C., Genete,M., Berrabah,W., Chevre,A.M., Delourme,R. *et al.* (2018) Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nat. Plants*, **4**, 879–887.

71. Parkin,I.A., Koh,C., Tang,H., Robinson,S.J., Kagale,S., Clarke,W.E., Town,C.D., Nixon,J., Krishnakumar,V., Bidwell,S.L. *et al.* (2014) Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid *Brassicaoleracea*. *Genome Biol.*, **15**, R77.

72. Cai,X., Wu,J., Liang,J., Lin,R., Zhang,K., Cheng,F. and Wang,X. (2020) Improved *Brassicaoleracea* JZS assembly reveals significant changing of LTRRT dynamics in different morphotypes. *Theor. Appl. Genet.*, **133**, 3187–3199.

73. Guo,N., Wang,S., Gao,L., Liu,Y., Wang,X., Lai,E., Duan,M., Wang,G., Li,J., Yang,M. *et al.* (2021) Genome sequencing sheds light on the contribution of structural variants to *Brassicaoleracea* diversification. *BMC Biol.*, **19**, 93.

74. Lv,H., Wang,Y., Han,F., Ji,J., Fang,Z., Zhuang,M., Li,Z., Zhang,Y. and Yang,L. (2020) A high quality reference genome for cabbage obtained with SMRT reveals novel genomic features and evolutionary characteristics. *Sci. Rep.*, **10**, 12394.

75. Li,L., Stoeckert,C.J. Jr and Roos,D.S. (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.*, **13**, 2178–2189.

76. Liu,Y. and Chen,Y. (2007) High-efficiency thermal asymmetric interlaced PCR for amplification of unknown flanking sequences. *BioTechniques*, **43**, 649–656.

77. Ou,S.J., Chen,J.F. and Jiang,N. (2018) Assessing genome assembly quality using the LTR assembly index (LAI). *Nucleic Acids Res.*, **46**, e126.

78. Liu,S., Liu,Y., Yang,X., Tong,C., Edwards,D., Parkin,I.A., Zhao,M., Ma,J., Yu,J., Huang,S. *et al.* (2014) The *Brassicaoleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nat. Commun.*, **5**, 3930.

79. Bayer,P.E., Golicz,A.A., Tirnaz,S., Chan,C.K.K., Edwards,D. and Batle,J. (2019) Variation in abundance of predicted resistance genes in the *Brassicaoleracea* pangenome. *Plant Biotechnol. J.*, **17**, 789–800.

80. Ma,J., SanMiguel,P., Lai,J., Messing,J. and Bennetzen,J.L. (2005) DNA rearrangement in orthologous *orp* regions of the maize, rice and sorghum genomes. *Genetics*, **170**, 1209–1220.

81. Jiao,W.B. and Schneeberger,K. (2020) Chromosome-level assemblies of multiple *Arabidopsis* genomes reveal hotspots of rearrangements with altered evolutionary dynamics. *Nat. Commun.*, **11**, 989.

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

83. Lorenz,A., Osman,F., Sun,W., Nandi,S., Steinacher,R. and Whitby,M.C. (2012) The fission yeast FANCM ortholog directs non-crossover recombination during meiosis. *Science*, **336**, 1585–1588.

84. Ling,C., Huang,J., Yan,Z., Li,Y., Ohzeki,M., Ishiai,M., Xu,D., Takata,M., Seidman,M. and Wang,W. (2016) Bloom syndrome complex promotes FANCM recruitment to stalled replication forks and facilitates both repair and traverse of DNA interstrand crosslinks. *Cell Discov.*, **2**, 16047.

85. Li,X., Yu,M., Bolanos-Villegas,P., Zhang,J., Ni,D., Ma,H. and Wang,Y. (2021) Fanconi anemia ortholog FANCM regulates meiotic crossover distribution in plants. *Plant Physiol.*, **186**, 344–360.

86. Blary,A., Gonzalo,A., Eber,F., Berard,A., Berges,H., Bessoltane,N., Charif,D., Charpentier,C., Cromer,L., Fourment,J. *et al.* (2018) FANCM limits meiotic crossovers in *Brassica* crops. *Front. Plant Sci.*, **9**, 368.

87. Yu,X., Wang,P., Li,J., Zhao,Q., Ji,C., Zhu,Z., Zhai,Y., Qin,X., Zhou,J., Yu,H. *et al.* (2021) Whole-genome sequence of synthesized allopolyploids in *Cucumis* reveals insights into the genome evolution of allopolyploidization. *Adv. Sci.*, **8**, 2004222.

88. Zhuang,Y., Wang,X., Li,X., Hu,J., Fan,L., Landis,J.B., Cannon,S.B., Grimwood,J., Schmutz,J., Jackson,S.A. *et al.* (2022) Phylogenomics of the genus *Glycine* sheds light on polyploid evolution and life-strategy transition. *Nat. Plants*, **8**, 233–244.

89. Golicz,A.A., Bayer,P.E., Barker,G.C., Edger,P.P., Kim,H., Martinez,P.A., Chan,C.K.K., Severn-Ellis,A., McCombie,W.R., Parkin,I.A.P. *et al.* (2016) The pangenome of an agronomically important crop plant *Brassicaoleracea*. *Nat. Commun.*, **7**, 13390.

90. Schubert,I. and Vu,G.T.H. (2016) Genome stability and evolution: attempting a holistic view. *Trends Plant Sci.*, **21**, 749–757.