Surviving a genome collision: genomic signatures of allopolyploidization in the recent crop species Brassica napus

Birgit Samans, Boulos Chalhoub, Rod J. Snowdon

Published online November 28, 2017

Surviving a Genome Collision: Genomic Signatures of Allopolyploidization in the Recent Crop Species *Brassica napus*

Birgit Samans, Boulos Chalhoub, and Rod J. Snowdon*

Abstract

Polyploidization has played a major role in crop plant evolution, leading to advantageous traits that have been selected by humans. Here, we describe restructuring patterns in the genome of *Brassica napus* L., a recent allopolyploid species. Widespread segmental deletions, duplications, and homeologous chromosome exchanges were identified in diverse genome sequences from 32 natural and 20 synthetic accessions, indicating that homeologous exchanges are a major driver of postpolyploidization genome diversification. Breakpoints of genomic rearrangements are rich in microsatellite sequences that are known to interact with the meiotic recombination machinery. In both synthetic and natural *B. napus*, a subgenome bias was observed toward exchanges replacing larger chromosome segments from the C-subgenome by their smaller, homeologous A-subgenome segments, driving postpolyploidization genome size reduction. Selection in natural *B. napus* favored segmental deletions involving genes associated with immunity, reproduction, and adaptation. Deletions affecting mismatch repair system genes, which are assumed to control homeologous recombination, were also found to be under selection. Structural exchanges between homeologous subgenomes appear to be a major source of novel genetic diversity in de novo allopolyploids. Documenting the consequences of genomic collision by genomic resequencing gives insights into the adaptive processes accompanying allopolyploidization.

Core Ideas

- Homeologous chromosome exchanges drive postpolyploidization genetic diversification in *Brassica napus*.
- Homeologous exchanges cause rapid genome size reduction after allopolyploidization.
- Segmental deletions contain genes associated with adaptive processes and chromosome stability.
- Breakpoints of genomic rearrangements contain motifs associated with meiotic recombination.
- Genes involved in chromosome mismatch repair are subject to selection after polyploidization.

During allopolyploidization, two different genomes collide in the same cell nucleus. Although most plant species originated from at least one round of ancient or recent polyploidizaton (Meyers and Levin, 2006; Otto and Whitton, 2000), the vast majority of allopolyploidization events are arguably evolutionary dead-ends (Mayrose et al., 2015; Soltis et al., 2014). For a de novo allopolyploid to survive, numerous protective systems that normally safeguard species integrity must

B. Samans and R.J. Snowdon, Dep. of Plant Breeding, Interdisciplinary Research Center for Biosystems, Land Use and Nutrition Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig Univ., Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany; B. Chalhoub, Institut National de la Recherche Agronomique – Unité de Recherche en Génomique Végétale, 2 rue gaston Crémieux, Evry 91057, France. Received 28 Feb. 2017. Accepted 1 May 2017. *Corresponding author (Rod.snowdon@agrar.uni-giessen.de).

Abbreviations: Ar, *B. rapa* A-genome; C_o, *B. oleracea* C-genome; CNV, copy number variation; Fst, fixation index; GO, gene ontology; HE, homeologous exchange; MMR, mismatch repair; SNP, single nucleotide polymorphism; TIR-NBS-LRR, Toll interleukin-1 receptor-class nucleotide binding site–leucine-rich repeat

Published in Plant Genome
Volume 10. doi: 10.3835/plantgenome2017.02.0013

break down: Fertilization must first overcome a species barrier, accompanied by a simultaneous failure of either meiotic reduction or mitotic division before or after meiosis. The result is a doubling of the haploid chromosome complements from the two parental gametes. If this leads to a viable seed that is able to germinate, survive, and flower, the absence of self-incompatibility and a rapid restitution of normal chromosome pairing are necessary for the de novo allopolyploid plant to generate viable, fertile offspring. During plant evolution, this improbable sequence of events frequently formed new species, including most major crops, demonstrating the fundamental evolutionary benefits of allopolyploidization (Paterson, 2005; Soltis et al., 2009).

The timing of ancient polyploidizaton events in the Brassicaceae (Cruciferae) family suggests associations with major environmental disruptions (Geiser et al., 2016; Kagale et al., 2014). Polyploidy is thus considered to have contributed to evolutionary adaptation (Otto, 2007); however, the genetic basis underlying polyploid survival and success is still not completely understood. The process of cytological diploidization, which is necessary to form a fertile and competitive allopolyploid embryo, is accompanied by selection of favorable structural and functional genome variants that arise frequently from interchanges between homeologous chromosomes (Soltis and Soltis, 2000; Wendel, 2000). Homeologous exchanges (HE), along with other segmental exchanges during homeologous chromosome pairing, are important drivers of genome restructuring, gene conversion, and adaptive variation. In most flowering plants, these processes took place millions of years ago; however, comparing the genome sequences of recent polyploid species and de novo allopolyploids can potentially provide insights into mechanisms of successful polyploid speciation. *Brassica napus* (genome AACC, $2n = 4x = 38$), an important allopolyploid crop that originated recently from interspecific hybridizations between the diploid parental species *Brassica rapa* L. (AA, $2n = 20$) and *Brassica oleracea* L. (CC, $2n = 18$), has become an important subject for studies of de novo allopolyploidization (Albertin et al., 2006; Nicolas et al., 2009; Sarilar et al., 2013). Synthetic *B. napus*, which can be readily generated with the help of tissue culture, provides a useful model for investigations regarding the genetic and genomic consequences of de novo allopolyploidization.

Cytogenetic (Xiong et al., 2011) and marker-based studies (Gaeta et al., 2007; Song et al., 1995; Szadkowski et al., 2010) revealed extensive HE in synthetic *B. napus*. Chromosome exchanges are also common in natural *B. napus* (Chalhoub et al., 2014; Osborn et al., 2003). Evidence is growing for molecular signatures of selection that have acted on homeologous chromosome segments during adaptation and breeding (Harper et al., 2012; Qian et al., 2014), but their genetic basis has to yet be elucidated. Reference genome sequences are now available for both *B. napus* (Chalhoub et al., 2014) and its parental progenitor species (Liu et al., 2014; Parkin et al., 2014; Wang et al., 2011). High-throughput genomic resequencing methods provide a potential tool to investigate the patterns of chromosome structural variation with unprecedented resolution.

Here, we established a data analysis pipeline to identify genome-wide structural patterns, along with consequent gene loss and duplication, in whole-genome resequencing data from allopolyploid species. We used the pipeline to analyze segmental genome structural variations in a large panel of highly diverse synthetic and natural *B. napus* accessions (Schmutzer et al., 2015). To our knowledge, the 20 synthetic and 32 natural *B. napus* accessions in this panel provide the largest available dataset so far available for detailed comparisons of genome structural features between de novo and natural allopolyploids.

Common areas under selection were identified in the natural *B. napus* accessions, representing successful survivors of allopolyploidization. Genes affected by segmental deletions were enriched for gene ontology (GO) terms related to adaptive processes and chromosome stability. Motifs identified in the breakpoints of genomic rearrangements suggest that structural changes arise primarily from meiotic recombination between chromosomes with extended homeologous sections. Rearrangements were directionally biased, leading to systematic reduction of genome size as a result of preferential loss of longer C-subgenome fragments after nonreciprocal exchanges replacing them with smaller, homeologous segments from the A-subgenome. Genes under strong selection underline the role of the highly conserved mismatch repair system in controlling homeologous recombination in allopolyploid *B. napus*.

Materials and Methods

Genome Resequencing Data from Natural and Synthetic *Brassica napus*

Previously published genomic resequencing data (Schmutzer et al., 2015), based on paired-end 100-bp reads generated on an Illumina HiSeq2000 (Illumina Inc., San Diego, CA), were available from 52 highly diverse *B. napus* genotypes, including 32 natural and 20 synthetic *B. napus* accessions. The 32 natural accessions were selected to broadly represent the genetic diversity present in a species-wide panel of 500 highly diverse *B. napus* accessions (Bus et al., 2011). The panel includes old European and Asian oilseed forms, fodder rapes and kales (all *B. napus* ssp. *napus*) along with rutabaga (or swede) forms (*B. napus* ssp. *napobrassica*). Additionally, 20 synthetic *B. napus* accessions developed by embryo rescue from interspecific hybridizations between highly diverse parental origins were included in the panel (Supplemental Table S1). A detailed description of the development and origin of the synthetic accessions has been described previously (Girke, 2002; Girke et al., 2012; Jesske, 2011; Jesske et al., 2013), as has a detailed description of the sequence data generation (Schmutzer et al., 2015). The

full sequence data for all 52 accessions is archived at the European Nucleotide Archive (http://www.ebi.ac.uk/ena, accessed 3 July 2017) under the project numbers PRJEB5974 and PRJEB6069.

The raw Illumina short read sequencing data were subjected to quality control procedures using the software suite FastQC (Andrews, 2010). Adapters and low-quality bases were removed using *cutadapt* (Martin, 2011), the *fastq_quality_trimmer*, and the *fast*q_*quality_filter* algorithms. Applying the *fastq_quality_trimmer* with the settings −t 28 and −l 90, trimmed bases with a Phred quality score of <28 followed by removing reads with a length of <90 nucleotides. Using the *fastq*_quality_filter algorithm (options:-q 28, -p 90) only reads where more than 90% of the nucleotides had a Phred score of ≥28 were retained. Paired-end reads were synchronized using a custom Perl script provided by the Leibniz Institute for Plant Genetics and Crop Plant Research, Gatersleben, Germany (Schmutzer and Scholz, 2015).

Sequence Alignment to a Concatenated Amphidiploid Reference Genome

Brassica napus (genome AACC, $2n = 4x = 38$) originates from an interspecific hybridization between the diploid parental progenitor species *B. rapa* (AA, $2n = 20$) and *B. oleracea* (CC, $2n = 18$). Because the aim was to monitor structural genome variations in comparison to the diploid progenitors, preprocessed, paired 100-bp reads were mapped to a concatenated amphidiploid genome assembly comprising the conjoined *Brassica* diploid progenitor genomes of the *B. rapa* cultivar Chiifu-401–42 (Wang et al., 2011) (subgenome A) and the *B. oleracea* cultivar TO1000 (Parkin et al., 2014) (subgenome C). A fast-gapped read alignment strategy was implemented with the default options of Bowtie2 (Langmead and Salzberg, 2012). After we applied the *samtools view* algorithm (Li et al., 2009), the alignment files in SAM-format were then converted into binary alignment/map format, excluding unmapped and nonunique reads.

Detection of Genomic Rearrangements in Natural and Synthetic *B. napus* Accessions

On the basis of the alignment files, we calculated the genome-wide guanine and cytosine content and median normalized coverage in 1000-bp blocks, assuming five copy number levels by using the commands *getReadCountsFromBAM* and *singlecn.mops* implemented in the R package cn.mops (Klambauer et al., 2012). Five possible copy number levels were assumed because an allotetraploid can carry between zero and four copies of each allele in the five alternative genotypes AAAA, AAAC, AACC, ACCC, or CCCC (Voorrips et al., 2011). Chromosome regions exhibiting copy number variations were identified individually for each line and chromosome by segmenting the normalized 1000-bp blocks, sorted according to their genomic distribution, using the circular binary segmentation algorithm implemented in the R package PSCBS (Olshen et al., 2011). In each instance, the strongest outliers, selected on the basis of their differences from the neighboring values, were smoothed toward the neighboring values using the command *callSegmentationOutliers*, with default options. A mean-level estimate was calculated for each segment. For each chromosome, the overall mean coverage and SD covering all the 1000-bp blocks were calculated. Segments with a minimum length of 10,000 bp and a coverage exceeding the chromosome mean by more than 1.5 SDs were defined as segmental duplications, whereas segments covering at least 10,000 bp, with coverage reduced by more than 1.5 SDs from the chromosome mean, were defined as deleted. All regions defined as either duplicated or deleted were saved in files in bed-format. A list of 32,699 orthologous *B. rapa* Chiifu-401–42 and *B. oleracea* TO1000 gene pairs (Chalhoub et al., 2014) were used to call putative HE. First, we applied the *intersect* algorithm in *bedtools* to assign individual genes of the orthologous gene pairs to the duplicated and deleted regions. When one orthologous gene was in a chromosome segment defined as duplicated and its corresponding ortholog in a region defined as deleted, the respective genomic regions containing this gene pair were defined as an HE block. Chromosome segments in which a putative deletion or duplication was observed without a corresponding exchange in the orthologous region were called as simple duplications or deletions, respectively. The boundaries of the chromosome segments carrying genomic rearrangements were defined by the start of the first gene in the rearranged segment to the end of the last gene. The R script for detection of genomic rearrangements is available as Supplemental File S1.

Gene Annotation

The *B. rapa* Chiifu-401–42 and *B. oleracea* TO1000 transcriptomes were functionally annotated using Blast2GO (Conesa et al., 2005) version 3.0.8 with default settings.

Detection of Enriched GO Terms in Regions Showing Copy Number Variation and HE

The conditional hypergeometric test in the R/Bioconductor (Gentleman et al., 2004) package GOstats version 2.40.0 (Falcon and Gentleman, 2007) was applied to detect functional enriched GO terms among the genes occurring in chromosome regions showing segmental copy number variation and HE.

Characterization of Breakpoints of Genomic Rearrangements

For each accession, putative breakpoints for genomic rearrangements were defined on the basis of the boundaries of the detected chromosomal changes in relation to the *B. rapa* Chiifu-401–42 and *B. oleracea* TO1000 reference genomes. The genomic regions spanning the HE breakpoints were analyzed for common sequence motifs by extracting the sequences 500 bp upstream and downstream of each breakpoint. The motif finder software MEME (Bailey et al., 2009, 2015) was implemented

to search for enriched DNA motifs with oligonucleotide sequences between 8 and 20 bases, applying the options -dna, -nmotifs 10, -minw 8, and -maxw 20. To increase the sensitivity of the analysis and to avoid breakpoint artifacts arising from misassemblies in the reference genomes, we included only positions which appeared in at least five and not more than 46 of the 52 accessions investigated.

Single Nucleotide Polymorphism Calling

Allelic variants were called by mapping the preprocessed 100-bp paired-end reads to the *B. rapa* Chiifu-401–42 and *B. oleracea* TO1000 genome assemblies, using SOAP2 software (http://soap.genomics.org.cn/soapaligner.html, accessed 11 July 2017)with the options -v 2 and -r 1. The alignment file was then converted to binary alignment/map format, followed by filtering of unmapped and nonunique reads by applying the *samtools view* algorithm (Li et al., 2009). High-quality single nucleotide polymorphisms (SNPs) were called by executing the samtools *mpileup* command with the following options: -Q 13 and -q 20. Homozygous SNPs with a minimum depth of five and a minimum quality score of 50 were extracted by applying the Genome Analysis Toolkit variant filtration algorithm (McKenna et al., 2010). By using the *vcftools-merge* command in VCFtools (Danecek, Auton et al., 2011), the SNP positions in all individual lines were merged and used as predefined positions to generate a SNP matrix containing a total of 597,352 SNPs called across the entire panel of natural and synthetic *B. napus* accessions.

Characterization of Selection Signatures

To identify genomic areas under selection, we screened for selective sweeps by using 597,352 high-quality homozygous SNPs called across the entire panel of natural and synthetic *B. napus* accessions with the alignment files generated from mapping to the *B. rapa* Chiifu-401–42 and *B. oleracea* TO1000 reference genome assemblies, respectively. To estimate pairwise differentiation among the natural and synthetic *B. napus* accessions, respectively, the VCFtools suite (Danecek et al., 2011) was used to calculate nucleotide diversity, Tajima's *D* (Tajima, 1989), and the fixation index (Fst) (Holsinger and Weir, 2009) across nonoverlapping 50-kb windows spanning the entire genome. We selected regions with striking selection patterns, filtered by the Fst values of $\geq$90% of the pairwise population distribution and the 10% lowest negative Tajima's *D* values in the natural accessions. Adjacent windows meeting the selection criteria and continuous regions within the 10% tails were merged.

Results

Detection of Genome-wide Structural Changes

A data processing pipeline was established to detect structural changes in short-read resequencing data from *B. napus* (Fig. 1; Supplemental File S1). Whole-genome resequencing data from 32 natural accessions of *B. napus* plus 20 highly diverse synthetic *B. napus* accessions generated via embryo rescue from independent hybridizations of the A-subgenome donor *B. rapa* to the C-subgenome donors *B. oleracea*, *Brassica hilarionis* Post, *Brassica montana* Pourr., or *B. cretica* Lam. were available from a previous study (Schmutzer et al., 2015). These data were reanalyzed for chromosome-scale exchanges after remapping to a synthetic reference, concatenated from *B. rapa* cv. Chiifu (Wang et al., 2011) and *B. oleracea* cv. TO1000 (Parkin et al., 2014), representing the *B. rapa* A-genome (A_r) and the *B. oleracea* C-genome (C_o) donors prior to allopolyploidization (Supplemental Table S1).

Averaged across all accessions, 82% (76–85%) of the reads could be aligned as paired-end reads and 4% (3–6%) as single-end reads, leading to an average genome-wide coverage of 15*x* (9.9–20.4*x*). On the basis of the alignment files, the coverage for consecutive 1000-bp blocks was calculated, accompanied by a normalization procedure to remove systematic biases. The coverage of consecutive 1000-bp blocks was then segmentally classified into unchanged, deleted, or duplicated segments individually for each chromosome. By analyzing all known orthologous gene pairs present in *B. rapa and B. oleracea* (Chalhoub et al., 2014), we classified the homeologous chromosome segments with regard to segmental deletions, duplications, or translocations (HE events from A_r to C_o or from C_o to A_r) in the *B. napus* accessions compared with the expectation from the progenitor genomes. Segments exhibiting structural variants were quantified by the length of the segment from the start of the first affected gene to the end of the last gene.

Frequent Genome Rearrangements in Natural and Synthetic *B. napus*

Natural *B. napus* accessions show areas of genomic rearrangement including deletions, duplications, and HE (Fig. 2; Supplemental Fig. S1–Supplemental Fig. S3; Table S2–Supplemental Table S4). These were unequally distributed, with a significantly higher number of rearrangements on C-subgenome chromosomes. Segmental deletions were considerably more frequent than duplications and translocations. In the A-subgenome, an average of 14 regions were deleted and duplicated per natural *B. napus* accession, with an average length of 106,629 bp for each deleted region and 58,697 bp for duplicated regions. In the C-subgenome, an average of 42 regions with a mean length of 184,215 bp were deleted, whereas 16 regions with a mean length of 106,599 bp were duplicated in the natural accessions (Supplemental Table S2–Supplemental Table S4). Homeologous exchanges were observed ~12 times per line, with an average per accession of seven $A_r >> C_o$ exchanges, comprising duplications of A_r segments and deletions of the corresponding C_o segments, and three $C_o >> A_r$ exchanges, comprising duplications of the Co segment and deletions of the corresponding A_r segment (Supplemental Table S3). Homeologous exchanges were most frequent between homeologous chromosomes A_r1–C_o1, A_r2–C_o2, and A_r9–C_o9. The

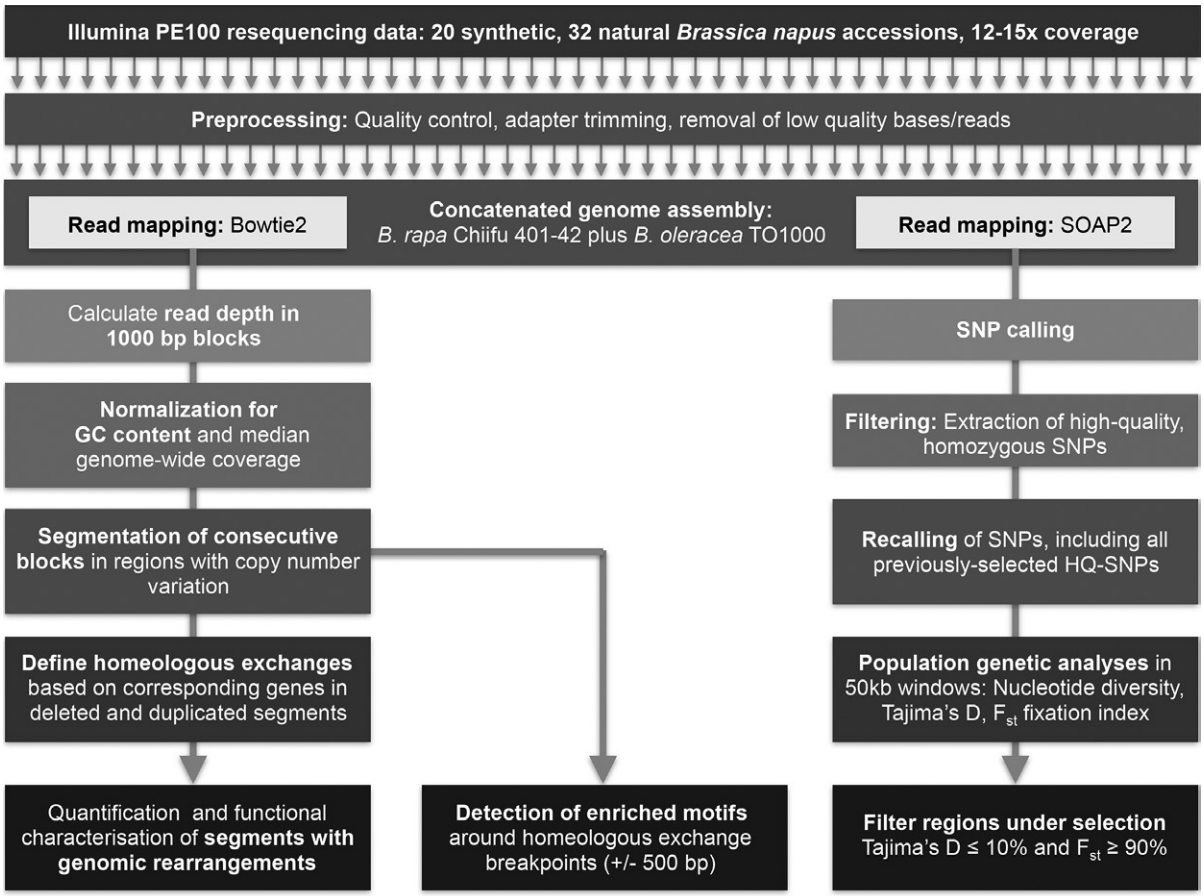

Fig. 1. Data analysis workflow for detecting structural chromosome rearrangements within genomic resequencing data in the allopolyploid genome of *Brassica napus*.

positions of the exchanges were conserved in different accessions, suggesting a nonrandom distribution of HE (Supplemental Fig. S2, Supplemental Fig. S3).

The number and size of genomic rearrangements in synthetic *B. napus* accessions were considerably greater than in the natural accessions (Fig. 2; Supplemental Fig. S1–Supplemental Fig. S3). Like the natural accessions, however, the synthetic *B. napus* accessions exhibited more areas with deletions than duplications and a similar, unequal distribution, with more rearrangements in the C-subgenome than in the A-subgenome. An average of ~26 deleted regions per synthetic accession was observed in the A-subgenome, with a mean length of 255,153 bp, and an average of ~27 duplicated regions with a mean length of 187,391 bp. In contrast, the C-subgenome exhibited an average of 51 deletions per accession with a mean length of 417,924 bp, and 46 duplications with a mean length of 115,454 bp (Supplemental Table S2). The highest frequency of deletions in the synthetic *B. napus* accessions was found on chromosomes A02, C01, C02, C04, and C09, whereas duplications were distributed across all chromosomes, but with a tendency toward the distal ends of the chromosomes (Fig. 2; Supplemental Fig. S1).

Overall, we found an average of 23 HEs per synthetic accession, in which the number and length of HEs varied significantly between accessions. As in the natural *B. napus* accessions, $A_r >> C_o$ exchanges were more frequent than $C_o >> A_r$ exchanges (14 versus 9, respectively, on average) (Supplemental Table S2). The HEs in synthetic *B. napus* accessions were also significantly larger than those in natural accessions, with an average length of 1,459,136 bp in the A-subgenome segments of $A_r >> C_o$ exchanges and 2,513,907 bp in the corresponding C-subgenome segments. In $C_o >> A_r$ exchanges, the A-subgenome segments had average lengths of 928,512 bp for the A-subgenome segments and 1,495,210 bp for the corresponding C-subgenome segments. The frequency of HEs was greatest between the homeologous A_r1–C_o1, A_r2–C_o2, A_r3–C_o3, A_r9–C_o8, and A_r9–C_o9, respectively, which have very strong collinearity along the lengths of the chromosomes (Chalhoub et al., 2014) (Supplemental Fig. S2, Supplemental Fig. S3, Supplemental Table S3).

Segmental Deletions are Conserved in Natural *B. napus*

To investigate common patterns of selection we searched for regions with common rearrangement patterns in a high number of natural accessions. Distinct chromosome segments on chromosomes C01, C02, C04 and C09 were deleted in nearly all investigated natural *B. napus* accessions (Fig. 2; Table S4).

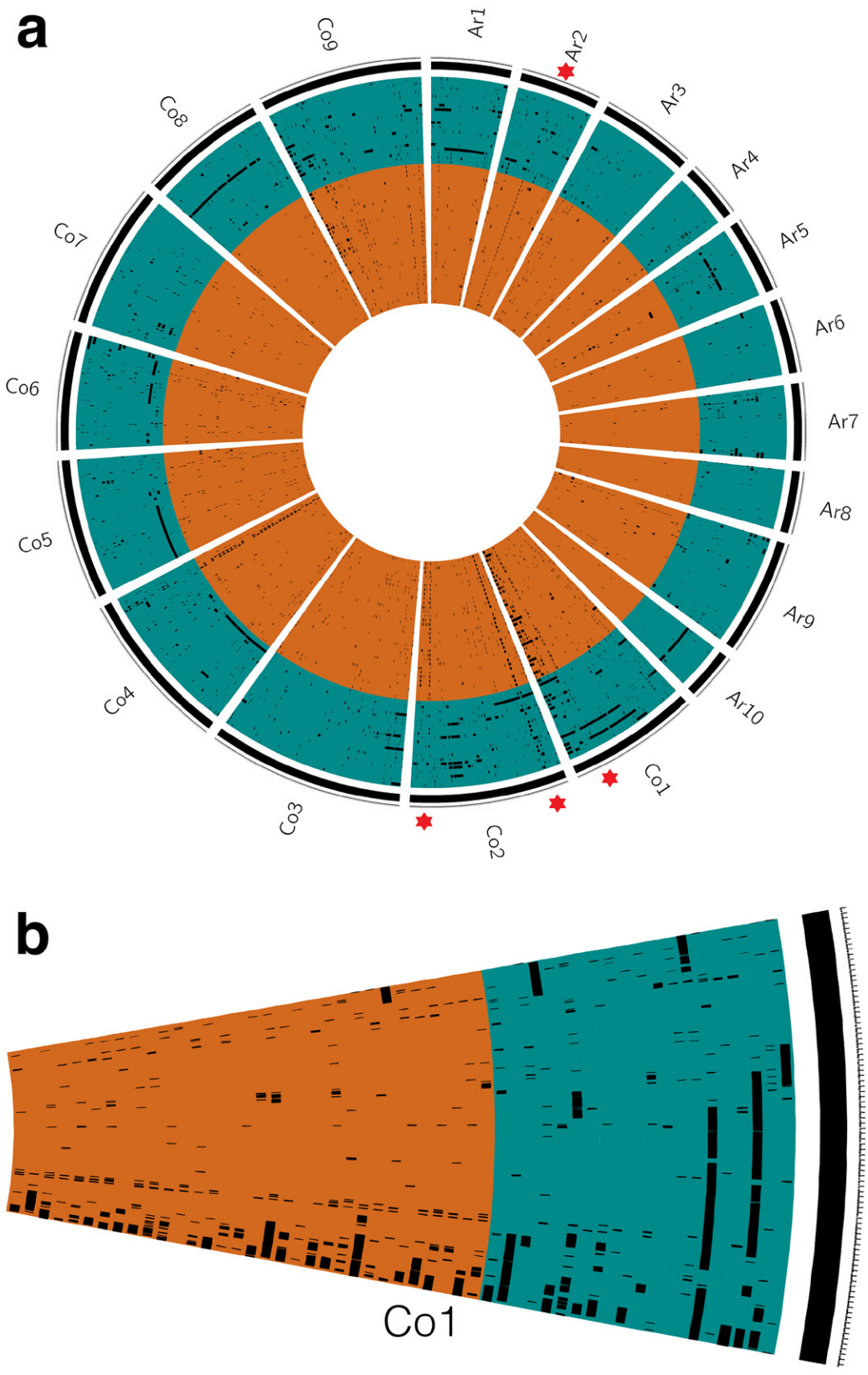

Fig. 2. Circos plots showing the distribution of deletions (indicated by black bars) in 32 natural (orange) and 20 resynthesized (blue) *B. napus* accessions. Each accession is displayed by one concentric circle. (a) Genome-wide overview of all detected deletions in comparison with the 10 chromosomes of the *B. rapa* A-subgenome (chromosomes A_r1–A_r10) and the nine chromosomes of the *B. oleracea* C-subgenome (chromosomes C_o1–C_o10). Red stars indicate regions that are commonly deleted in a high number of natural *B. napus* accessions, indicating postpolyploidisation selection. (b) Detail of chromosome C01 as an example of the widespread segmental deletions identified in all investigated *B. napus* genotypes.

Segmental Deletions are Enriched for Genes Associated with Reproduction, Adaption, and Chromosome Stability

To characterize the functional impact of segmental deletions, we searched for enriched GO terms in each of the 52 natural and synthetic accessions among genes showing deletions in comparison with the *B. rapa* Chiifu-401–42 and *B. oleracea* TO1000 transcriptomes (functionally annotated using Blast2GO version 3.0.8). The conditional hypergeometric test included in the Bioconductor GOstats package (version 2.40.0) was applied to detect functionally enriched GO terms in relation to the complete gene set of the diploid progenitors. Gene ontology terms showing significant enrichment associated with deletions across the 52 accessions are listed in Supplemental Table S5. Twenty-nine of the 32 natural accessions showed enriched deletion of genes involved in cytosol to ER transport, 22 accessions showed enriched deletion of genes for seed development, and 19 for the pectin catabolic process. Overall, 20 of the 32 natural *B. napus* accessions showed significant depletion of genes annotated with the GO term 'telomere capping', and 19 natural accessions showed significant depletion of genes associated with ATP-dependent chromatin remodeling. Both of these processes are involved in the maintenance of chromosome stability and meiotic recombination. Interestingly, analysis of the gene content within two frequently deleted regions on chromosome C09 revealed significant enrichment for Toll interleukin-1 receptor-class nucleotide binding site–leucine-rich repeat (TIR-NBS-LRR) immune receptor genes. In one of these regions, comprising a total of 21 genes, 14 were TIR-NBS-LRR resistance genes and the other region comprised 14 genes, eight of which were TIR-NBS-LRR genes. Between 8 and 14 of the natural *B. napus* accessions carried one or both of these deletions (Table 1).

Segmental Duplications Have Less Selective Influence than Deletions

Genes within segmental duplications were also analyzed individually for enriched GO terms in each of the 52 accessions. In contrast to the clear ontology enrichments observed for deleted genes, no such selective pattern was detected among duplicated genes, which displayed only a few common features. Eight accessions were enriched for the GO term 'positive regulation of seed maturation', which may be associated with artificial selection during breeding (Supplemental Table S6).

Homeologous Chromosome Exchanges Show Strong Directional Bias

Brassica rapa A-genome > > C_o events were observed significantly more frequently than C_o > > A_r translocations (Supplemental Fig. S2, Supplemental Fig. S3, Supplemental Table S2, Supplemental Table S3). Previous studies indicate the potential impact of the cytoplasmic background on the direction of HE in synthetic *B. napus* (Szadkowski et al., 2010). To test the putative impact of the cytoplasmatic background, we included synthetic *B. napus* accessions where either *B. rapa* or *B. oleracea* was used as the maternal cytoplasm donor (Supplemental Table S1 and Supplemental Table S7). Using the number of HEs per accession as a measure for the HE rate per chromosome, significant differences (Mann–Whitney test $p < 0.05$) between synthetic accessions with *B. rapa* or *B. oleracea* cytoplasm were found only for HE between chromosomes A_r1 and C_o1. Of the eight accessions with *B. rapa* as the maternal donor, only two showed $A_r1 > > C_o1$ exchanges, with an average of 0.4 HEs per accession. In contrast, for the nine accessions with *B. oleracea* as the maternal donor, eight showed $A_r1 > > C_o1$ exchanges, with an average of 4.1 HEs per accession. On a genome-wide scale, no significant overall effect of the cytoplasmatic background on the direction of HE was observed.

Homeologous Exchanges Invoke Genome Size Reduction

In both synthetic and natural *B. napus* accessions, the average cumulative size of segmental deletions associated with HE events was consistently greater than the average cumulative size of the corresponding homeologous duplications (Supplemental Table S2). This corresponds to the greater frequency of translocations from the smaller A-subgenome to the larger C-subgenome, causing larger C-subgenome segments to be replaced by their smaller homeologous segments from the A-subgenome. Significant differences were also observed between the gene content of exchanged homeologous segments in the A_r and C_o progenitor subgenomes. This has caused both an import to the C-subgenome of previously absent genes specific to the A-subgenome, along with the complete loss of some genes that are specific to the C-subgenome.

Most Genomic Rearrangements Associate With Postpolyploidization Signatures that are Specific for Diploid Progenitor Genomes

To distinguish genuine homeologous rearrangements from potential deletions or duplications already present in the diploid progenitors of specific *B. napus* accessions in comparison with the diploid reference genomes of *B. rapa* (Chiifu-401) and *B. oleracea* TO1000, we compared the rearrangement patterns of synthetic *B. napus* accessions derived from the same *B. oleracea* or *B. rapa* parents. In these accessions, we expected that some or all of the duplications or deletions observed in the genomes of the diploid parents would be conserved in these synthetics. Although a small number of chromosome regions indeed showed inherited variations, their frequency was very low in comparison with the overall frequency of deletions and duplications (Supplemental Fig. S4). We also assumed that individual characteristics of the diploid reference genomes *B. rapa* (Chiifu-401) and *B. oleracea* TO1000 should lead to the observation of common rearrangement events in large numbers of accessions,

Table 1. Descriptions and positions of genes annotated within two frequently deleted regions of B. napus chromosome C09 harboring a high number of Toll interleukin-1 receptor-class nucleotide binding site–leucine-rich repeat (TIR-NBS-LRR) proteins.

Chromosome region (C09)	Gene ID	Start	Stop	Description†	Number of natural *B. napus* accessions with gene deleted
18.39–18.48 Mbp	*Bo9g061110*	18,391,916	18,394,697	TIR-NBS-LRR class disease resistance protein	10
	Bo9g061120	18,395,252	18,396,120	Ferredoxin-NADP+ reductase	10
	Bo9g061130	18,400,755	18,400,991	Hypothetical protein	10
	Bo9g061140	18,406,887	18,408,883	COP1-interacting protein-like protein	12
	Bo9g061150	18,410,770	18,413,011	Retrotransposon protein, putative, Ty3-gypsy subclass	12
	Bo9g061160	18,416,897	18,418,635	TIR-NBS-LRR class disease resistance protein	14
	Bo9g061160	18,416,897	18,418,635	TIR-NBS-LRR class disease resistance protein	14
	Bo9g061170	18,420,697	18,422,247	TIR-NBS-LRR class disease resistance protein	14
	Bo9g061180	18,426,613	18,427,246	TIR-NBS-LRR class disease resistance protein	14
	Bo9g061190	18,427,565	18,428,064	Hypothetical protein	14
	Bo9g061200	18,447,330	18,451,835	TIR-NBS-LRR class disease resistance protein	14
	Bo9g061210	18,452,153	18,453,389	Hypothetical protein ARALYDRAFT_916966	13
	Bo9g061230	18,461,446	18,462,368	NAC domain containing protein	13
	Bo9g061240	18,463,976	18,471,612	TIR-NBS-LRR class disease resistance protein	13
	Bo9g061250	18,477,762	18,477,911	Not available	13
	Bo9g061260	18,478,429	18,482,296	TIR-NBS-LRR class disease resistance protein	13
46.56–46.73 Mbp	*Bo9g156020*	46,563,039	46,566,297	TIR-NBS-LRR class disease resistance protein	7
	Bo9g156030	46,575,675	46,578,226	TIR-NBS-LRR class disease resistance protein	7
	Bo9g156040	46,578,256	46,579,030	TIR-NBS-LRR class disease resistance protein	7
	Bo9g156050	46,583,309	46,587,817	TIR-NBS-LRR class disease resistance protein	8
	Bo9g156060	46,598,211	46,599,571	Nucleic acid-binding, OB-fold-like protein	8
	Bo9g156070	46,622,193	46,625,251	TIR-NBS-LRR class disease resistance protein	21
	Bo9g156080	46,630,261	46,632,111	TIR-NBS-LRR class disease resistance protein	21
	Bo9g156090	46,644,244	46,644,960	U-box domain-containing protein	21
	Bo9g156100	46,648,331	46,650,116	TIR-NBS-LRR class disease resistance protein	21
	Bo9g156110	46,651,071	46,652,733	TIR-NBS-LRR class disease resistance protein	21
	Bo9g156120	46,656,218	46,657,638	TIR-NBS-LRR class disease resistance protein	21
	Bo9g156130	46,657,780	46,658,834	TIR-NBS-LRR class disease resistance protein	21
	Bo9g156140	46,659,219	46,659,650	Retrotransposon protein, putative, Ty1-copia subclass	21
	Bo9g156150	46,666,540	46,667,155	TIR-NBS-LRR class disease resistance protein	21
	Bo9g156160	46,670,510	46,673,478	Glutathione *S*-transferase T3	21
	Bo9g156170	46,673,589	46,675,894	Ribosomal protein-like protein	21
	Bo9g156180	46,676,278	46,680,181	TIR-NBS-LRR class disease resistance protein	21
	Bo9g156190	46,684,837	46,685,226	Hypothetical protein	21
	Bo9g156200	46,706,743	46,707,476	ARM repeat superfamily protein	21
	Bo9g156210	46,718,027	46,719,626	TIR-NBS-LRR class disease resistance protein	21
	Bo9g156220	46,724,719	46,726,066	TIR-NBS-LRR class disease resistance protein	21

† NADP, nicotinamide adenine dinucleotide phosphate; NAC, *N*-terminal amino acid module; OB, oligonucleotide/oligosaccharide binding motif; ARM, *ARMADILLO*

particularly synthetic accessions sharing parental lines with structural variations compared with the diploid references. Because we did not observe clear patterns fulfilling these criteria, it was assumed that most of the observed rearrangements indeed arose during or after allopolyploidization (Supplemental Fig. S4).

Breakpoints of Genome Rearrangements are Enriched for Known Meiotic Motifs

Natural and synthetic *B. napus* accessions were found to share common genome regions enriched for genomic rearrangements, suggesting a nonrandom distribution. For high-resolution characterization of chromosome regions harboring breakpoints for genomic rearrangements, we used the segments called as either deleted or duplicated. Breakpoints were defined as chromosomal positions where a segment started or ended. To avoid false positive calls caused by assembly errors in the reference genomes, we only included breakpoints that appeared in at least five and at most 45 of the 52 accessions. A total of 3653 chromosome positions were thus extracted containing putative rearrangement breakpoints. For each breakpoint, the genomic sequence 500 bp upstream and downstream was extracted. Using MEME software (with settings -nmotifs 10, -minw 8, and -maxw 20), breakpoint-flanking sequences were

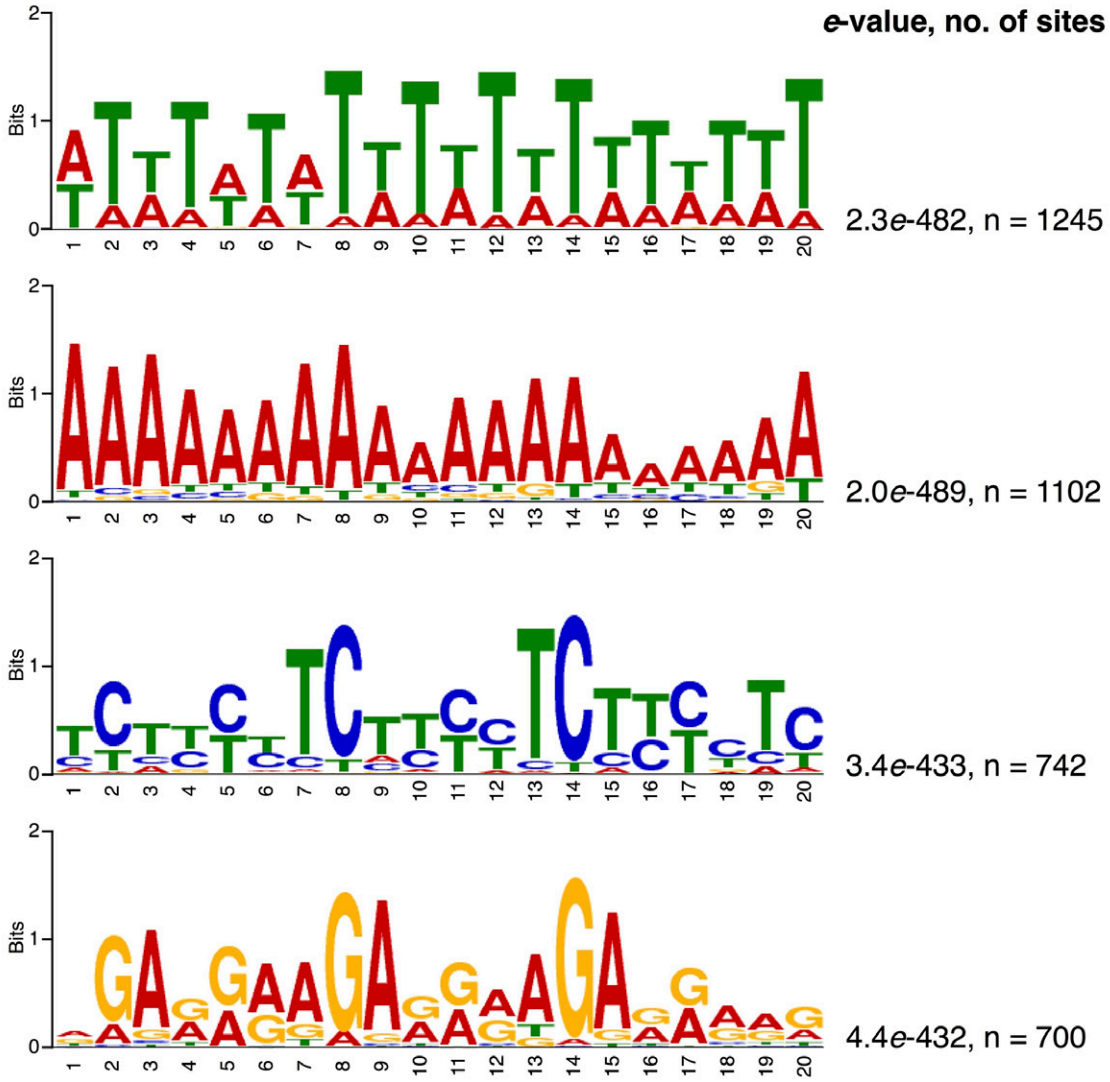

Fig. 3. Motifs enriched around putative breakpoints of genomic rearrangements monitored in 32 natural and 20 synthetic *B. napus* accessions. For each breakpoint, the genomic sequence 500 bp upstream and downstream was extracted and analyzed for enriched DNA motifs using MEME software. The e-value estimates the expected occurrence (with the given log-likelihood ratio or higher) of motifs with the same size and frequency in a similarly sized set of random sequences.

subsequently analyzed for enrichment with DNA motifs 8 to 20 bp long. Genomic sequences around putative breakpoints were found to be significantly enriched for poly-A or poly-T stretches and the palindromic GAA–CTT microsatellite sequence (Fig. 3). Both of these motifs are associated with recombination hotspots in *Arabidopsis thaliana* (L.) Heynh. (Choi et al., 2013; Wijnker et al., 2013). By sequencing a bacterial artificial chromosome clone spanning a putative breakpoint for a known HE, we confirmed the breakpoint junction at the transcription start site of an ortholog of the *A. thaliana* gene *SIMILAR TO RCO1* (*SRO5*; *Bra029254/Bo2g164210*). Within the promoter sequence of this gene, we identified the GAA–CTT recombination hotspot predicted by the sequence enrichment analysis (Supplemental Fig. S4).

Selective Sweeps Reveal Signatures of Selection for Agronomic Traits

Changes in nucleotide diversity and skewed allele frequency spectra provide additional information on selection processes beyond structural chromosome rearrangements. Supplemental Fig. S6 summarizes the results of the sliding window analysis of the calculated nucleotide diversity, Tajima's *D* and the Fst were calculated in 50-kb intervals across the entire genome using 597,352 SNPs from the entire panel of natural and synthetic *B. napus* accessions. Tajima's *D* describes the number of rare variants at a genomic locus, indicating the regions under selection, whereas genome-wide Fst values can reveal chromosome regions contributing to subpopulation differentiation. Details on 244 regions under putative selection, containing a total of 2006 genes, are provided in detail in Supplemental Table S8. These selective sweeps show unequal subgenomic distribution,

Table 2. *Brassica napus* genes found in regions under selection that are involved in the DNA repair system. Genes were identified on the basis of their functional annotation with gene ontology terms using Blast2GO.

Genomic position of selective sweep	Chr.†	Start	Stop	Gene ID	Description
A01:16,550,001–166,000,00	A01	16,555,277	16,556,421	*Bra031527*	Serpin family protein
A02:800,001–850,000	A02	809,190	821,032	*Bra028755*	Anaphase-promoting complex subunit 1-like
A04:12,850,001–12,900,000	A04	12,892,203	12,896,592	*Bra035691*	DNA repair protein rad51b
A05:21,400,001–21,450,000	A05	21,436,900	21,438,621	*Bra027399*	Nucleosome assembly protein 1
A08:20,650,001–20,700,000	A08	20,693,198	20,698,863	*Bra030524*	DNA repair helicase UVH6
A08:20,650,001–20,700,000	A08	20,656,215	20,656,733	*Bra030537*	Thylakoid lumenal protein chloroplastic-like
A08:20,700,001–20,750,000	A08	20,722,339	20,727,390	*Bra030517*	WEE1-like protein kinase-like
A09:27,950,001–28,000,000	A09	27,995,877	27,998,036	*Bra006956*	Growth-regulating factor 4
BolScaffold00950:1–50,000	BolScaffold00950	7,934	12,918	*Bo00950s020*	ATP-dependent DNA helicase pif1
C02:14,550,001–14,600,000	C02	14,561,051	14,564,572	*Bo2g051060*	DNA mismatch repair protein 2
C02:15,300,001–15,350,000	C02	15,341,352	15,342,734	*Bo2g053890*	Histone chaperone ASF1b-like
C02:25,200,001–25,250,000	C02	25,204,403	25,208,840	*Bo2g092740*	Nucleoside-triphosphatase nucleotide binding protein
C05:9,600,001–9,650,000	C05	9,624,775	9,626,321	*Bo5g027490*	DNA glycosylase superfamily protein
C05:9,600,001–9,650,000	C05	9,627,877	9,632,180	*Bo5g027500*	Transducin WD-40 repeat-containing protein
C07:16,000,001–16,050,000	C07	16,026,117	16,028,388	*Bo7g043530*	DNA-damage-repair toleration protein DRT111
C07:33,750,001–33,800,000	C07	33,749,965	33,757,578	*Bo7g087110*	Structural maintenance of chromosomes protein 4-like
C09:6,800,001–6,850,000	C09	6,807,683	6,809,809	*Bo9g022220*	Wound-responsive family protein

† Chr., chromosome

with 172 areas under selection in the C-subgenome and 72 in the A-subgenome. Analysis for enriched GO terms revealed a range of terms strongly associated with traits of relevance for domestication and breeding, including leaf senescence, carbohydrate biosynthesis, lipid metabolism, response to osmotic stress, vernalization response, and root development (Supplemental Table S9).

Meiotic Mismatch Repair Genes are Located in Regions Associated with Selective Sweeps

Allopolyploidization was found to strongly impact the genomic structure of *B. napus*, via the rearrangement processes that result from homeologous recombination during meiosis. On the other hand, although natural *B. napus* accessions retain a number of HE, they generally exhibit diploid-like meiotic behavior. Strict control of homeologous pairing during meiosis ensures correct chromosome partitioning, guaranteeing fertility and genome stability. The high rate of HEs in synthetic accessions, of which only a selected few are conserved in natural *B. napus*, suggests strong selection to reinstate normal chromosome pairing behavior. Despite the evolutionary importance of this process and its implications in allopolyploid plant evolution, the mechanisms preventing homeologous recombination in allopolyploid species is still unclear. We manually inspected the list of genes under selection for putative candidates involved in the control of homeologous pairing during meiosis. Among the genes under selection (Tajima's $D \leq 10\%$; Fst $\geq 90\%$), we found 17 genes involved in the meiotic mismatch repair (MMR) system, including orthologs of *MSH2*, *RAD51b*, and *NAP1* as putative candidates for control of homeologous recombination (Table 2).

Subgenomic Bias for Nucleotide Diversity

As expected because of their diverse origin and absence of stringent agronomic selection, overall nucleotide diversity was higher in the synthetic *B. napus* accessions than in the natural accessions (Fig. 4; Supplemental Fig. S6). However, both groups showed an overall lower nucleotide diversity in the C-subgenome than in the A-subgenome, and differences between chromosomes within the subgenomes. Nucleotide diversity in synthetic accessions showed consistent levels across the A-subgenome, which were slightly reduced in chromosome A09. Chromosomes C01, C02, and C09, which are frequently affected by homeologous Ar > > Co exchanges in synthetic *B. napus*, show a clearly reduced diversity (Fig. 4). Overall, the nucleotide diversity levels of the natural accessions are lower in the C-subgenome, whereas chromosome-wise nucleotide diversity is particularly reduced on chromosomes A01, A02, A09, C05, C07 and C09 (Fig. 4). This corresponds to the results from large-scale genetic diversity studies using genome-wide SNP markers (Qian et al., 2014; Schiessl et al., 2015). However, reduced C-subgenome diversity in synthetic *B. napus* derived from highly diverse C-subgenome donors appears to contradict the hypothesis that A-subgenome diversity is primarily higher because of continued admixture with *B. rapa* subsequent to allopolyploidization (Qian et al., 2014). Instead, imbalanced subgenomic diversity already appears to have arisen from strong selection acting on particular C-subgenome chromosomes in very early generations after de novo allopolyploidization.

Discussion

Analysis of resequencing data in a large panel of synthetic and natural *B. napus* provided unprecedented insights into the patterns and consequences of genome

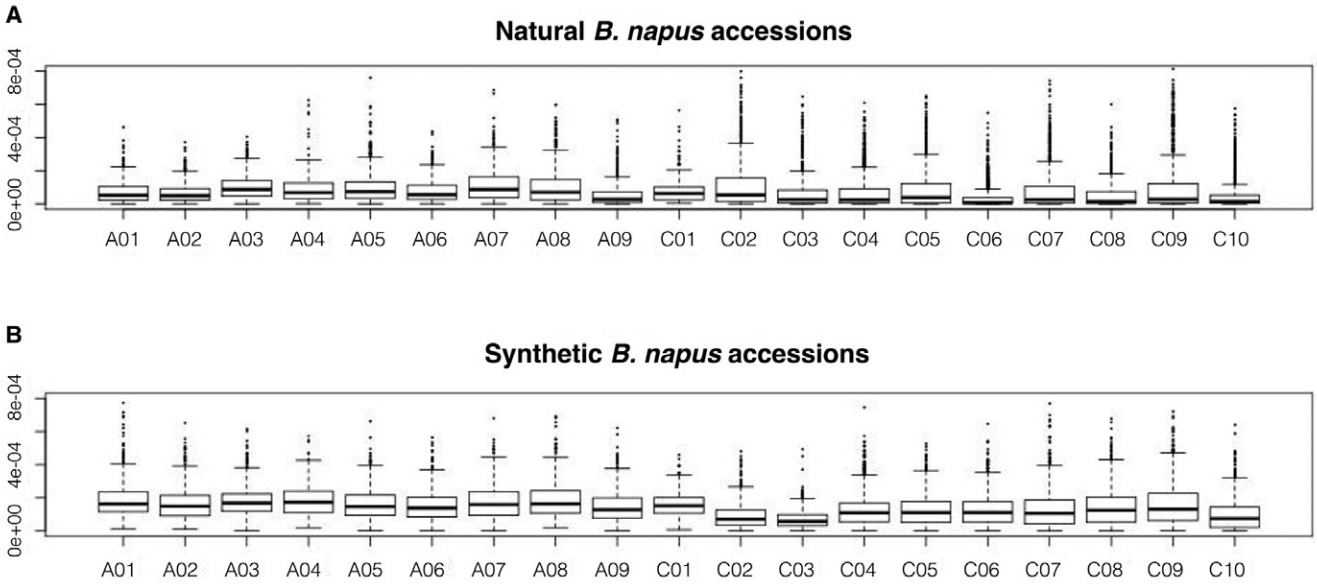

Fig. 4. Box plots displaying distribution of nucleotide diversity per chromosome, measured genome-wide across 50-kb windows of the 32 natural and 20 synthetic *B. napus* accessions.

restructuring as a consequence of polyploidization. We developed an analysis procedure based on short-read genomic resequencing datasets to detect structural genome variation in homeologous chromosome segments of allopolyploids compared with their diploid parental genomes. The pipeline delivers a robust strategy to detect structural genome variation in *B. napus*, in spite of the non-uniform chromosome coverage associated with short-read resequencing data. Nonuniform coverage, caused by, for example, differences between repeat-rich and repeat-poor regions, can bias segmentation processes when a highly sensitive segmentation algorithm is used. To overcome this potential problem we used circular binary segmentation algorithm that tests for change-points by using a maximal *t*-statistic with a permutation reference distribution. The result was a robust handling of the noisy data from short-read sequencing. By setting a conservative 10-kb threshold as the minimum length for calling rearrangements, we provided additional protection against calling false positive duplicated or deleted regions.

Beside nonreciprocal homeologous translocations, reciprocal homeologous translocations might also be present. As homeologous translocations are balanced exchanges that do not go along with changes in read depth, they cannot be quantified via the established pipeline. For the detection of homeologous translocations, alternative methods are necessary, for example, long-read sequencing, optical mapping, or genetic mapping when blocks of markers with a known physical position map to the linkage groups of their respective homeolog (Stein et al., 2017).

By applying this pipeline, we confirmed the widespread presence of HE events in the allopolyploid *B. napus* genome. The resolution provided by the next-generation DNA sequencing technology, in association with the genome assemblies of *B. napus* and their progenitor species, gave unprecedented insights into the extent, physical distribution, and genetic consequences of chromosome exchanges during allopolyploidization. The hypothesis that HE events cause intersubgenomic restructuring during de novo polyploidization was first construed from observations of lost and gained restriction fragments in newly resynthesized *B. napus* accessions (Sharpe et al., 1995; Song et al., 1995), then later supported by molecular cytogenetic studies (Xiong et al., 2011), which implied homeologous chromosome losses or gains and chromosome restructuring. Similar phenomena were also observed in a very recent polyploid *Tragopodon* species (Chester et al., 2012), and have been associated with karyotypic and phenotypic plasticity in synthetic allohexaploids of *A. thaliana* (Matsushita et al., 2012). In synthetic wheat (*Triticum aestivum* L.) allotetraploids, extensive chromosome restructuring was followed by immediate and persistent karyotype stabilization (Zhang et al., 2013). Detailed knowledge of how such restructuring is driven and how it shapes selection in a new polyploid has nevertheless remained elusive.

We confirmed that vast changes in genome structure are common in synthetic *B. napus* accessions; frequently, entire chromosomes are affected. These massive rearrangements are likely to explain the severe disruption of fertility and performance commonly seen in synthetic *B. napus*, which can hinder trait introgression by breeders from the diploid progenitors. In contrast, we found that comparatively few structural variants were retained in natural, cultivated *B. napus* accessions but some of these were strikingly conserved. This implies that natural *B. napus* derived from strong early selection for traits influencing fitness and fertility, which are essential for the success of a new species.

In wheat, there is evidence that selection for successful allopolyploid combinations prefers specific

subgenome donors that are less prone to aberrant homeologous recombination (Zhang et al., 2013). A similar scenario may have restricted allopolyploid donors in *Brassica*, where vast diversity exists among the species representing the A and C subgenomes. An alternative hypothesis for the relatively low number of rearrangements retained by selection is that natural *B. napus* originates from specific allopolyploidization events carrying only a few advantageous rearrangements. The relatively small size of the conserved rearrangements, compared with the massive rearrangements in synthetic *B. napus*, implies rare double-recombinants with advantageous properties. Because the exact diploid progenitors of modern *B. napus* have not yet been discovered; however, testing of these hypotheses is difficult. Interestingly, segmental deletions of specific chromosome regions appear to be particularly highly conserved, suggesting early functional gene loss within homeologous loci for which duplications confer a selective disadvantage. The A-subgenome bias of these deleted regions is associated with strong selection against C-subgenome gene homeologs. The positions of these regions on chromosome arms with very high levels of HE in synthetic *B. napus* (Supplemental Fig. S2, Supplemental Fig. S3, Supplemental Table S4) suggest that these are hotspots for HE and that the deletions have already been strongly selected immediately after allopolyploidization. Postpolyploidization gene silencing mechanisms, such as epigenetic changes, might also play a similar role; however, the immediate and irreversible influence of gene loss induced by nonreciprocal HE appears to be a key feature that drove evolutionary selection in *B. napus*.

The observation of strong selection for deletions involving specific homeologous nucleotide binding site–leucine-rich repeat resistance gene loci concurs with studies demonstrating that drastic fitness costs can be associated with incompatible resistance allele combinations (Alcazar et al., 2014; Stirnweis et al., 2014). Nevertheless, care must be taken when interpreting this observation, as nucleotide binding site–leucine-rich repeat gene clusters are often highly duplicated and can collapse into single loci in reference assemblies, complicating the interpretation of copy-number variation.

Polyploidization is often followed by a diploidization process resulting in single-copy genes by deletion of previously duplicated genes (Woodhouse et al., 2010). Gene loss is generally non-random. Which genes are retained as single copies depends on their genomic position or the specific functional category (Albalat and Canestro, 2016). We observed both a subgenome-bias for deletions, with a significantly higher number of deletions in the C subgenome, and also a bias toward distinct functional categories. Different studies found gene losses biased by gene function to remain preferentially single-copy genes that are over-represented in the functional categories DNA repair, recombination, enzyme activity, kinase activity, transport, tRNA ligation, defense, and categories associated with domestication processes (Duarte et al., 2010). In concordance with these observations, we also found deleted genes enriched for GO terms associated with transport and domestication processes (Supplemental Table S5).

Genes involved in the MMR system were found to be under selection in the natural accessions. The MMR system, which is extremely conserved across eukaryote kingdoms, is known to be involved in preventing homeologous recombination (Hunter et al., 1996; Rayssiguier et al., 1989), as they can discriminate homology from homeology. In interspecific yeast hybrids between *Saccharomyces cerevisiae* and its close relative *Saccharomyces paradoxus*, inactivation of the DNA mismatch repair gene *MSH2* [a homolog of the bacterial *MutS* gene, which controls the fidelity of genetic exchanges (Stambuk and Radman, 1998)] induced tolerable levels of homeologous recombination between mismatched heteroduplexes without overly adverse effects on spore viability, aneuploidy, or fitness (Hunter et al., 1996). A similar mechanism may be necessary to overcome the interspecific barrier between *B. rapa* and *B. oleracea*. In combination with *MSH3*, *MSH6*, or *MSH7*, *MSH2* has variable binding specificity to different chromosome mismatches (Chambers et al., 1996; Hunter et al., 1996). Correspondingly, *A. thaliana* germ-cell lineages showed a threefold increase in homeologous recombination after disruption of *AtMSH2*, underlining the role of this gene in preventing exchanges between diverged chromosomes (Lafleuriel et al., 2007). In hexaploid wheat, the pairing control loci *Ph1* and *Ph2* suppress crossovers between homeologous chromosomes (Mello-Sampayo, 1971). The *Ph2* locus contains a *T. aestivum* ortholog of *MSH7*, a homolog of the *MSH2* interaction partner *MSH6*. Changes in the distribution of (GAA)n sequences have been associated with chromosomal rearrangements during the evolution of wild and domesticated allopolyploid wheat (Adonina et al., 2015). This corresponds to our observed enrichment of poly-A or poly-T stretches and GAA–CTT motifs adjacent to the breakpoints of homeologous rearrangements. These are coincident with regions of low nucleosome density, which facilitate the access of recombination proteins (Pan et al., 2011), but also with the presence of H2A.Z nucleosomes. The latter promote DNA double-strand breaks during meiosis by forming overlapping chromosomal foci with the recombinases DMC1 and RAD51 (Choi et al., 2013). Temporary inactivation of the MMR system in a de novo allopolyploid may encourage postpolyploidization HE, helping to overcome the interspecies reproductive barrier. Subsequent selection of exchanges that reinstate normal homologous pairing may be an essential requirement for establishing a stable, fertile allopolyploid. Although complex functional studies are necessary to confirm the individual roles of specific genes, the prevalence of MMR genes in areas under postpolyploidization selection suggest a key role in regulating survival of de novo allopolyploids.

Interestingly, we observed no clear patterns of selection acting on genes influenced by segmental homeologous duplications. This suggests that gene loss played a

more important role than duplication or neofunctionalization in the success of allopolyploid *B. napus*.

Our data confirmed predictions based on genetic mapping studies that homeologous genome restructuring in *B. napus* increases with physical distance from the centromeres and the degree of collinearity between chromosomes (Nicolas et al., 2012). In synthetic *B. napus* accessions, we observed the highest frequency of rearrangements on the homeologous chromosome pair A01–C01. These two long chromosomes display highly conserved synteny along their entire lengths. High rearrangement frequencies were also observed between other highly syntenic chromosome arms, particularly between chromosomes A02–C02, A03–C03, A09–C08, and A09–C09. Similar gene conversions and exchanges at the single nucleotide level as a consequence of allopolyploidization were described for *B. napus* (Chalhoub et al., 2014), with the same directional bias observed in the present data. Comparisons of allopolyploid cotton (*Gossypium hirsutum* L.) with its progenitor species also revealed that homeologous gene conversion was biased in favor of one of the progenitor subgenomes (Paterson et al., 2012), with the agronomically inferior D-subgenome demonstrating a greater tendency to overwrite genes from the A-subgenome.

The preference of HE events to replace large C-subgenome chromosome segments by their smaller homeologous A-subgenome regions is consistent with observations associating genome size reductions with ancient neopolyploidization events (Leitch and Bennett, 2004). The corresponding patterns we found in natural and synthetic *B. napus* indicate that selection for genome size reduction has an immediate impact during de novo allopolyploidization. Nonequivalent subgenome evolution and biased genome fractionation are common observations in polyploids (Guo et al., 2014; Woodhouse et al., 2014), and preferential reduction of the larger parental subgenome was also observed in Triticeae polyploids. Because directional bias for HE impacts the abundance, distribution, and frequency of subgenome-specific DNA repeats, it is tempting to associate genome size reduction with selection acting on DNA repeat sequences (Hu et al., 2011). Differences in homeologous gene content between exchanged sequences promote immediate subgenome-specific gene loss, with significant functional implications. Copy number variations induced by HE, which can influence specific *B. napus* copies of *FLOWERING LOCUS C*, were implicated in subspeciation (Chalhoub et al., 2014).

Documenting the short- and long-term consequences of genomic collision, by genome-scale comparisons between large, diverse panels of de novo (synthetic) and natural accessions of the model allopolyploid plant species *B. napus* gave unprecedented insights into the adaptive processes accompanying allopolyploidization. Substantial structural exchanges between homeologous subgenomes have been described before, but their high frequency and extent was previously unknown. Structural genome variants appear to be a major source of novel genetic diversity in de novo allopolyploids, presumably helping to overcome the strict genetic bottleneck imposed by allopolyploidization. This may explain the paradoxical evolutionary success of polyploid plants, which are ubiquitous despite the origin of many allopolyploid species from just a few rare founder events. Interestingly, subgenome bias appears to select for exchanges that replace larger segments by smaller homeologous segments, potentially driving genome size reduction immediately after de novo allopolyploidisation. Conserved motifs identified at homeologous breakpoints support previous hypotheses that normal meiotic exchanges between homeologous chromosomes are the key driver of postallopolyploidization chromosome restructuring. Although relaxation of strict homologous pairing control is essential to facilitate the creation of diversity via HE, a subsequent reinstatement of stable chromosome pairing is necessary to ensure fertility and species viability. Loss of homeologous duplicates of MMR system genes in early generations after allopolyploidization may provide a means to relax and reinstate homeologous recombination.

Supplemental Information

Supplemental Table S1: Overview of re-sequencing results from 32 natural and 20 resynthesized *B. napus* accessions.

Supplemental Table S2: Number, average, and cumulative length of genomic rearrangements in 32 natural and 20 synthetic *B. napus* accessions.

Supplemental Table S3: Number of homoeologous exchanges between distinct homeologous chromosome regions.

Supplemental Table S4: Overview of regions affected by genomic rearrangements in 32 natural and 20 resynthesized *B. napus* accessions.

Supplemental Table S5: Gene ontology (GO) terms over-represented in deleted genes of 52 *B. napus* accessions.

Supplemental Table S6: Gene ontology (GO) terms over-represented in duplicated genes of 52 *B. napus* accessions.

Supplemental Table S7: Number of homoeologous exchanges (HE) in synthetic *B. napus* accessions with *B. rapa* or *B. oleracea* as the maternal or paternal genome donor, respectively.

Supplemental Table S8: Genes in regions under selection.

Supplemental Table S9: Over-represented gene ontology (GO) terms in genes under selection.

Supplemental Figure S1: Circos plot showing duplicated genomic areas in 32 natural and 20 synthetic *B. napus* accessions.

Supplemental Figure S2: Circos plot showing homeologous exchanges (HE) associated with duplications of chromosome segments from the A subgenome and deletions of the corresponding homeologous chromosome segment from the C-subgenome in 32 natural and 20 synthetic *B. napus* accessions.

Supplemental Figure S3: Circos plot showing homeologous exchanges (HE) associated with duplications of

chromosome segments from the C-subgenome and deletions of the corresponding homeologous chromosome segment from the A-subgenome in 32 natural and 20 synthetic *B. napus* accessions.

Supplemental Figure S4: Sequence characterization of a homeologous exchange breakpoint.

Supplemental Figure S5: Circos plot summarizing population genetics parameters in 32 natural and 20 synthetic *B. napus* accessions, respectively.

Supplemental File S1: R script for detection of genomic rearrangements.

Conflict of Interest Disclosure

The authors declare that they have no competing financial interests.

Acknowledgments

BS and RJS acknowledge funding from the German Federal Ministry of Education and Research (BMBF) Plant Biotechnology Program, grant number 0315964: "Precision Breeding for Yield Gain in Oilseed Rape (Pre-Breed Yield)". Public access to the raw sequencing data for the study was provided by the PreBreed-Yield consortium partners German Seed Alliance GmbH, Norddeutsche Pflanzenzucht Hans-Georg Lembke KG, Deutsche Saatveredelung AG, KWS SAAT SE, Limagrain GmbH, Syngenta Seeds GmbH, and Bayer Crop Science Raps GbR and coordinated by Thomas Schmutzer and Uwe Scholz (Leibniz Institute for Plant Genetics and Crop Plant Research Gatersleben, Germany). BS and RJS conceived the study. BS developed analysis tools and analyzed the data. BC provided validation data. BS and RJS wrote the manuscript.

References

Adonina, I.G., N.P. Goncharov, E.D. Badaeva, E.M. Sergeeva, N.V. Petrash, and E.A. Salina. 2015. (GAA)n microsatellite as an indicator of the A genome reorganization during wheat evolution and domestication. Comp. Cytogenet. 9:533–547. doi:10.3897/CompCytogen.v9i4.5120

Albalat, R., and C. Canestro. 2016. Evolution by gene loss. Nat. Rev. Genet. 17:379–391. doi:10.1038/nrg.2016.39

Albertin, W., T. Balliau, P. Brabant, A.M. Chevre, F. Eber, C. Malosse, et al. 2006. Numerous and rapid nonstochastic modifications of gene products in newly synthesized *Brassica napus* allotetraploids. Genetics 173:1101–1113. doi:10.1534/genetics.106.057554

Alcazar, R., M. von Reth, J. Bautor, E. Chae, D. Weigel, M. Koornneef, et al. 2014. Analysis of a plant complex resistance gene locus underlying immune-related hybrid incompatibility and its occurrence in nature. PLoS Genet. 10(12):e1004848. doi:10.1371/journal.pgen.1004848.

Andrews, S. 2010. FastQC: A quality control tool for high throughput sequence data. Babraham Bioinformatics. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Bailey, T.L., M. Boden, F.A. Buske, M. Frith, C.E. Grant, L. Clementi, et al. 2009. MEME SUITE: Tools for motif discovery and searching. Nucleic Acids Res. 37:W202–W208. doi:10.1093/nar/gkp335

Bailey, T.L., J. Johnson, C.E. Grant, and W.S. Noble. 2015. The MEME suite. Nucleic Acids Res. 43:W39–W49. doi:10.1093/nar/gkv416

Bus, A., N. Korber, R.J. Snowdon, and B. Stich. 2011. Patterns of molecular variation in a species-wide germplasm set of *Brassica napus*. Theor. Appl. Genet. 123:1413–1423. doi:10.1007/s00122-011-1676-7

Chalhoub, B., F. Denoeud, S.Y. Liu, I.A.P. Parkin, H.B. Tang, X.Y. Wang, et al. 2014. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. Science 345:950–953. doi:10.1126/science.1253435

Chambers, S.R., N. Hunter, E.J. Louis, and R.H. Borts. 1996. The mismatch repair system reduces meiotic homeologous recombination and stimulates recombination-dependent chromosome loss. Mol. Cell. Biol. 16:6110–6120. doi:10.1128/MCB.16.11.6110

Chester, M., J.P. Gallagher, V.V. Symonds, A.V. Cruz da Silva, E.V. Mavrodiev, A.R. Leitch, et al. 2012. Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon miscellus* (Asteraceae). Proc. Natl. Acad. Sci. USA 109:1176–1181. doi:10.1073/pnas.1112041109

Choi, K., X. Zhao, K.A. Kelly, O. Venn, J.D. Higgins, N.E. Yelina, et al. 2013. *Arabidopsis* meiotic crossover hot spots overlap with H2A.Z nucleosomes at gene promoters. Nat. Genet. 45:1327–1336. doi:10.1038/ng.2766

Conesa, A., S. Gotz, J.M. Garcia-Gomez, J. Terol, M. Talon, and M. Robles. 2005. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676. doi:10.1093/bioinformatics/bti610

Danecek, P., A. Auton, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, et al. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158. doi:10.1093/bioinformatics/btr330

Duarte, J.M., P.K. Wall, P.P. Edger, L.L. Landherr, H. Ma, J.C. Pires, et al. 2010. Identification of shared single copy nuclear genes in *Arabidopsis, Populus, Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. BMC Evol. Biol. 10:61. doi:10.1186/1471-2148-10-61

Falcon, S., and R. Gentleman. 2007. Using GOstats to test gene lists for GO term association. Bioinformatics 23:257–258. doi:10.1093/bioinformatics/btl567

Gaeta, R.T., J.C. Pires, F. Iniguez-Luy, E. Leon, and T.C. Osborn. 2007. Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. Plant Cell 19:3403–3417. doi:10.1105/tpc.107.054346

Geiser, C., T. Mandakova, N. Arrigo, M.A. Lysak, and C. Parisod. 2016. Repeated whole-genome duplication, karyotype reshuffling, and biased retention of stress-responding genes in Buckler mustard. Plant Cell 28:17–27. doi:10.1105/tpc.15.00791

Gentleman, R.C., V.J. Carey, D.M. Bates, B. Bolstad, M. Dettling, S. Dudoit, et al. 2004. Bioconductor: Open software development for computational biology and bioinformatics. Genome Biol. 5:R80. doi:10.1186/gb-2004-5-10-r80

Girke, A. 2002. Neue Genpools aus resynthetisiertem Raps (*Brassica napus* L.) für die Hybridzüchtung. PhD diss., Georg-August-Universität Göttingen, Germany.

Girke, A., A. Schierholt, and H.C. Becker. 2012. Extending the rapeseed gene pool with resynthesized *Brassica napus* II: Heterosis. Theor. Appl. Genet. 124:1017–1026. doi:10.1007/s00122-011-1765-7

Guo, H., X. Wang, H. Gundlach, K.F. Mayer, D.G. Peterson, B.E. Scheffler, et al. 2014. Extensive and biased intergenomic nonreciprocal DNA exchanges shaped a nascent polyploid genome, *Gossypium* (cotton). Genetics 197:1153–1163. doi:10.1534/genetics.114.166124

Harper, A.L., M. Trick, J. Higgins, F. Fraser, L. Clissold, R. Wells, et al. 2012. Associative transcriptomics of traits in the polyploid crop species *Brassica napus*. Nat. Biotechnol. 30:798–802. doi:10.1038/nbt.2302

Holsinger, K.E., and B.S. Weir. 2009. Genetics in geographically structured populations: Defining, estimating and interpreting F-ST. Nat. Rev. Genet. 10:639–650. doi:10.1038/nrg2611

Hu, T.T., P. Pattyn, E.G. Bakker, J. Cao, J.F. Cheng, R.M. Clark, et al. 2011. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. Nat. Genet. 43:476–481. doi:10.1038/ng.807

Hunter, N., S.R. Chambers, E.J. Louis, and R.H. Borts. 1996. The mismatch repair system contributes to meiotic sterility in an interspecific yeast hybrid. EMBO J. 15:1726–1733.

Jesske, T. 2011. *Brassica*-Wildarten als neue genetische Ressource für die Rapszüchtung. PhD diss.,Georg-August-Universität Göttingen, Germany.

Jesske, T., B. Olberg, A. Schierholt, and H.C. Becker. 2013. Resynthesized lines from domesticated and wild *Brassica* taxa and their hybrids with *B. napus* L.: Genetic diversity and hybrid yield. Theor. Appl. Genet. 126:1053–1065. doi:10.1007/s00122-012-2036-y

Kagale, S., S.J. Robinson, J. Nixon, R. Xiao, T. Huebert, J. Condie, et al. 2014. Polyploid evolution of the Brassicaceae during the Cenozoic era. Plant Cell 26:2777–2791. doi:10.1105/tpc.114.126391

Klambauer, G., K. Schwarzbauer, A. Mayr, D.A. Clevert, A. Mitterecker, U. Bodenhofer, et al. 2012. cn.MOPS: Mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate. Nucleic Acids Res. 40:E69. doi:10.1093/nar/gks003

Lafleuriel, J., F. Degroote, A. Depeiges, and G. Picard. 2007. Impact of the loss of AtMSH2 on double-strand break-induced recombination between highly diverged homeologous sequences in *Arabidopsis thaliana* germinal tissues. Plant Mol. Biol. 63:833–846. doi:10.1007/s11103-006-9128-5

Langmead, B., and S.L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9:357–359. doi:10.1038/nmeth.1923

Leitch, I.J., and M.D. Bennett. 2004. Genome downsizing in polyploid plants. Biol. J. Linn. Soc. Lond. 82:651–663. doi:10.1111/j.1095-8312.2004.00349.x

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, et al. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. doi:10.1093/bioinformatics/btp352

Liu, S., Y. Liu, X. Yang, C. Tong, D. Edwards, I.A. Parkin, et al. 2014. The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. Nat. Commun. 5:3930. doi:10.1038/ncomms4930

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17:10–12. doi:10.14806/ej.17.1.200.

Matsushita, S.C., A.P. Tyagi, G.M. Thornton, J.C. Pires, and A. Madlung. 2012. Allopolyploidization lays the foundation for evolution of distinct populations: Evidence from analysis of synthetic *Arabidopsis* allohexaploids. Genetics 191:535–547. doi:10.1534/genetics.112.139295

Mayrose, I., S.H. Zhan, C.J. Rothfels, N. Arrigo, M.S. Barker, L.H. Rieseberg, et al. 2015. Methods for studying polyploid diversification and the dead end hypothesis: A reply to Soltis et al. (2014). New Phytol. 206:27–35. doi:10.1111/nph.13192

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, et al. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297–1303. doi:10.1101/gr.107524.110

Mello-Sampayo, T. 1971. Genetic regulation of meiotic chromosome pairing by chromosome 3D of *Triticum aestivum*. Nat. New Biol. 230:22–23. doi:10.1038/newbio230022a0

Meyers, L.A., and D.A. Levin. 2006. On the abundance of polyploids in flowering plants. Evolution 60:1198–1206. doi:10.1111/j.0014-3820.2006.tb01198.x

Nicolas, S.D., M. Leflon, H. Monod, F. Eber, O. Coriton, V. Huteau, et al. 2009. Genetic regulation of meiotic cross-overs between related genomes in *Brassica napus* haploids and hybrids. Plant Cell 21:373–385. doi:10.1105/tpc.108.062273

Nicolas, S.D., H. Monod, F. Eber, A.M. Chevre, and E. Jenczewski. 2012. Non-random distribution of extensive chromosome rearrangements in *Brassica napus* depends on genome organization. Plant J. 70:691–703. doi:10.1111/j.1365-313X.2012.04914.x

Olshen, A.B., H. Bengtsson, P. Neuvial, P.T. Spellman, R.A. Olshen, and V.E. Seshan. 2011. Parent-specific copy number in paired tumor-normal studies using circular binary segmentation. Bioinformatics 27:2038–2046. doi:10.1093/bioinformatics/btr329

Osborn, T.C., D.V. Butrulle, A.G. Sharpe, K.J. Pickering, I.A. Parkin, J.S. Parker, et al. 2003. Detection and effects of a homeologous reciprocal transposition in *Brassica napus*. Genetics 165:1569–1577.

Otto, S.P. 2007. The evolutionary consequences of polyploidy. Cell 131:452–462. doi:10.1016/j.cell.2007.10.022

Otto, S.P., and J. Whitton. 2000. Polyploid incidence and evolution. Annu. Rev. Genet. 34:401–437. doi:10.1146/annurev.genet.34.1.401

Pan, J., M. Sasaki, R. Kniewel, H. Murakami, H.G. Blitzblau, S.E. Tischfield, et al. 2011. A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. Cell 144:719–731. doi:10.1016/j.cell.2011.02.009

Parkin, I.A., C. Koh, H. Tang, S.J. Robinson, S. Kagale, W.E. Clarke, et al. 2014. Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid *Brassica oleracea*. Genome Biol. 15:R77. doi:10.1186/gb-2014-15-6-r77

Paterson, A.H. 2005. Polyploidy, evolutionary opportunity, and crop adaptation. Genetica (The Hague) 123:191–196.

Paterson, A.H., J.F. Wendel, H. Gundlach, H. Guo, J. Jenkins, D. Jin, et al. 2012. Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. Nature 492:423–427. doi:10.1038/nature11798

Qian, L., W. Qian, and R.J. Snowdon. 2014. Sub-genomic selection patterns as a signature of breeding in the allopolyploid *Brassica napus* genome. BMC Genomics 15:1170. doi:10.1186/1471-2164-15-1170

Rayssiguier, C., D.S. Thaler, and M. Radman. 1989. The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch-repair mutants. Nature 342:396–401. doi:10.1038/342396a0

Sarilar, V., P.M. Palacios, A. Rousselet, C. Ridel, M. Falque, F. Eber, et al. 2013. Allopolyploidy has a moderate impact on restructuring at three contrasting transposable element insertion sites in resynthesized *Brassica napus* allotetraploids. New Phytol. 198:593–604. doi:10.1111/nph.12156

Schiessl, S., F. Iniguez-Luy, W. Qian, and R.J. Snowdon. 2015. Diverse regulatory factors associate with flowering time and yield responses in winter-type *Brassica napus*. BMC Genomics 16:737. doi:10.1186/s12864-015-1950-1

Schmutzer, T., B. Samans, E. Dyrszka, C. Ulpinnis, S. Weise, D. Stengel, et al. 2015. Species-wide genome sequence and nucleotide polymorphisms from the model allopolyploid plant *Brassica napus*. Sci. Data 2:150072. doi:10.1038/sdata.2015.72

Schmutzer, T., and U. Scholz. 2015. PreBreed Yield data processing scripts. e!DAL Electronic Data Archive Library. https://doi.ipk-gatersleben.de/DOI/d0e5e591-1c4f-458c-89ec-ce30ee944572/b3848aed-5b32-4852-b74a-2d2534e955ff/2.

Sharpe, A.G., I.A. Parkin, D.J. Keith, and D.J. Lydiate. 1995. Frequent nonreciprocal translocations in the amphidiploid genome of oilseed rape (*Brassica napus*). Genome 38:1112–1121. doi:10.1139/g95-148

Soltis, D.E., V.A. Albert, J. Leebens-Mack, C.D. Bell, A.H. Paterson, C. Zheng, et al. 2009. Polyploidy and angiosperm diversification. Am. J. Bot. 96:336–348. doi:10.3732/ajb.0800079

Soltis, D.E., M.C. Segovia-Salcedo, I. Jordon-Thaden, L. Majure, N.M. Miles, E.V. Mavrodiev, et al. 2014. Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose et al. New Phytol. 202:1105–1117. doi:10.1111/nph.12756

Soltis, P.S., and D.E. Soltis. 2000. The role of genetic and genomic attributes in the success of polyploids. Proc. Natl. Acad. Sci. USA 97:7051–7057. doi:10.1073/pnas.97.13.7051

Song, K., P. Lu, K. Tang, and T.C. Osborn. 1995. Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. Proc. Natl. Acad. Sci. USA 92:7719–7723. doi:10.1073/pnas.92.17.7719

Stambuk, S., and M. Radman. 1998. Mechanism and control of interspecies recombination in *Escherichia coli*. I. Mismatch repair, methylation, recombination and replication functions. Genetics 150:533–542.

Stein, A., O. Coriton, M. Rousseau-Gueutin, B. Samans, S.V. Schiessl, C. Obermeier, et al. 2017. Mapping of homoeologous chromosome exchanges influencing quantitative trait variation in *Brassica napus*. Plant Biotechnol. J. doi:10.1111/pbi.12732

Stirnweis, D., S.D. Milani, S. Brunner, G. Herren, G. Buchmann, D. Peditto, et al. 2014. Suppression among alleles encoding nucleotide-binding-leucine-rich repeat resistance proteins interferes with resistance in F_1 hybrid and allele-pyramided wheat plants. Plant J. 79:893–903. doi:10.1111/tpj.12592

Szadkowski, E., F. Eber, V. Huteau, M. Lode, C. Huneau, H. Belcram, et al. 2010. The first meiosis of resynthesized *Brassica napus*, a genome blender. New Phytol. 186:102–112. doi:10.1111/j.1469-8137.2010.03182.x

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595.

Voorrips, R.E., G. Gort, and B. Vosman. 2011. Genotype calling in tetraploid species from bi-allelic marker data using mixture models. BMC Bioinformatics 12:172. doi:10.1186/1471-2105-12-172

Wang, X., H. Wang, J. Wang, R. Sun, J. Wu, S. Liu, et al. 2011. The genome of the mesopolyploid crop species *Brassica rapa*. Nat. Genet. 43:1035–1039. doi:10.1038/ng.919

Wendel, J.F. 2000. Genome evolution in polyploids. Plant Mol. Biol. 42:225–249. doi:10.1023/A:1006392424384

Wijnker, E., G. Velikkakam James, J. Ding, F. Becker, J.R. Klasen, V. Rawat, et al. 2013. The genomic landscape of meiotic crossovers and gene conversions in *Arabidopsis thaliana*. eLife 2:e01426. doi:10.7554/eLife.01426

Woodhouse, M.R., F. Cheng, J.C. Pires, D. Lisch, M. Freeling, and X. Wang. 2014. Origin, inheritance, and gene regulatory consequences of genome dominance in polyploids. Proc. Natl. Acad. Sci. USA 111:5283–5288. doi:10.1073/pnas.1402475111

Woodhouse, M.R., J.C. Schnable, B.S. Pedersen, E. Lyons, D. Lisch, S. Subramaniam, et al. 2010. Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homologs. PLoS Biol. 8:e1000409. doi:10.1371/journal.pbio.1000409

Xiong, Z., R.T. Gaeta, and J.C. Pires. 2011. Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. Proc. Natl. Acad. Sci. USA 108:7908–7913. doi:10.1073/pnas.1014138108

Zhang, H., Y. Bian, X. Gou, Y. Dong, S. Rustgi, B. Zhang, et al. 2013. Intrinsic karyotype stability and gene copy number variations may have laid the foundation for tetraploid wheat formation. Proc. Natl. Acad. Sci. USA 110:19466–19471. doi:10.1073/pnas.1319598110