Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome

Boulos Chalhoub,1,4 France Denoue,2,3,4 Shengyi Liu,5 Isobel A. P. Parkin,6 Haibao Tang,7,10 Xixin Wang,11 Julien Chiquet,3 Harry Beleram,7 Chaobo Tong,5 Andrew G. Paterson,9 and Tim Whitelaw5

Together with more ancient polyploidizations, this conferred an aggregate 72× genome assembly of 849.7 Mb was obtained with SOAP (29), and include the model plant *Arabidopsis thaliana*. We examined the *B. napus* genome and the consequences of its recent duplication. The constituent *A* and *C* subgenomes are engaged in subtle structural, functional, and epigenetic cross-talk, with abundant homologous exchanges. Incipient gene loss and expression divergence have begun. Selection in *B. napus* oilseed types has accelerated the loss of glucosinolate genes, while preserving expansion of oil biosynthesis genes. These processes provide insights into allopolyploid evolution and its relationship with crop domestication and improvement.

The *B. napus* genome contains 101,040 gene models estimated from 353 Gb of RNA sequencing (RNA-seq) data (tables S15 and S16) in combination with ab initio gene prediction, protein and EST alignments, and transposon masking (7). Of these, 91,167 were confirmed by matches with *B. rapa* and/or *B. oleracea* predicted proteomes. Genes are abundant in distal euchromatin but sparse near centromeric heterochromatin (Fig. 2). RNA-seq data revealed alternative splicing in 48% of genes, with frequent intron retention (82%) and rare exon skipping (3%) (tables S17 and S18 and fig. S11).

The *B. napus* *A* and *C* subgenomes are largely colinear to the corresponding diploid *A* and *C* subgenomes from the best-announced *B. rapa* (29) and *B. oleracea* (30) genomes, based on the ordered scaffolds and their colinearity among the corresponding draft *Arabidopsis* genomes (tables S1 to S5). We further examined the *B. napus* genome and the consequences of its recent duplication. The constituent *A* and *C* subgenomes are engaged in subtle structural, functional, and epigenetic cross-talk, with abundant homologous exchanges. Incipient gene loss and expression divergence have begun. Selection in *B. napus* oilseed types has accelerated the loss of glucosinolate genes, while preserving expansion of oil biosynthesis genes. These processes provide insights into allopolyploid evolution and its relationship with crop domestication and improvement.

The *B. napus* genome contains 101,040 gene models estimated from 353 Gb of RNA sequencing (RNA-seq) data (tables S15 and S16) in combination with ab initio gene prediction, protein and EST alignments, and transposon masking (7). Of these, 91,167 were confirmed by matches with *B. rapa* and/or *B. oleracea* predicted proteomes. Genes are abundant in distal euchromatin but sparse near centromeric heterochromatin (Fig. 2). RNA-seq data revealed alternative splicing in 48% of genes, with frequent intron retention (82%) and rare exon skipping (3%) (tables S17 and S18 and fig. S11).

The *B. napus* *A* and *C* subgenomes are largely colinear to the corresponding diploid *A* and *C* subgenomes from the best-announced *B. rapa* (29) and *B. oleracea* (30) genomes, based on the ordered scaffolds and their colinearity among the corresponding draft *Arabidopsis* genomes (tables S1 to S5). We further examined the *B. napus* genome and the consequences of its recent duplication. The constituent *A* and *C* subgenomes are engaged in subtle structural, functional, and epigenetic cross-talk, with abundant homologous exchanges. Incipient gene loss and expression divergence have begun. Selection in *B. napus* oilseed types has accelerated the loss of glucosinolate genes, while preserving expansion of oil biosynthesis genes. These processes provide insights into allopolyploid evolution and its relationship with crop domestication and improvement.

The *B. napus* genome contains 101,040 gene models estimated from 353 Gb of RNA sequencing (RNA-seq) data (tables S15 and S16) in combination with ab initio gene prediction, protein and EST alignments, and transposon masking (7). Of these, 91,167 were confirmed by matches with *B. rapa* and/or *B. oleracea* predicted proteomes. Genes are abundant in distal euchromatin but sparse near centromeric heterochromatin (Fig. 2). RNA-seq data revealed alternative splicing in 48% of genes, with frequent intron retention (82%) and rare exon skipping (3%) (tables S17 and S18 and fig. S11).

The *B. napus* *A* and *C* subgenomes are largely colinear to the corresponding diploid *A* and *C* subgenomes from the best-announced *B. rapa* (29) and *B. oleracea* (30) genomes, based on the ordered scaffolds and their colinearity among the corresponding draft *Arabidopsis* genomes (tables S1 to S5). We further examined the *B. napus* genome and the consequences of its recent duplication. The constituent *A* and *C* subgenomes are engaged in subtle structural, functional, and epigenetic cross-talk, with abundant homologous exchanges. Incipient gene loss and expression divergence have begun. Selection in *B. napus* oilseed types has accelerated the loss of glucosinolate genes, while preserving expansion of oil biosynthesis genes. These processes provide insights into allopolyploid evolution and its relationship with crop domestication and improvement.

The *B. napus* genome contains 101,040 gene models estimated from 353 Gb of RNA sequencing (RNA-seq) data (tables S15 and S16) in combination with ab initio gene prediction, protein and EST alignments, and transposon masking (7). Of these, 91,167 were confirmed bymatches with *B. rapa* and/or *B. oleracea* predicted proteomes. Genes are abundant in distal euchromatin but sparse near centromeric heterochromatin (Fig. 2). RNA-seq data revealed alternative splicing in 48% of genes, with frequent intron retention (82%) and rare exon skipping (3%) (tables S17 and S18 and fig. S11).

The *B. napus* *A* and *C* subgenomes are largely colinear to the corresponding diploid *A* and *C* subgenomes from the best-announced *B. rapa* (29) and *B. oleracea* (30) genomes, based on the ordered scaffolds and their colinearity among the corresponding draft *Arabidopsis* genomes (tables S1 to S5). We further examined the *B. napus* genome and the consequences of its recent duplication. The constituent *A* and *C* subgenomes are engaged in subtle structural, functional, and epigenetic cross-talk, with abundant homologous exchanges. Incipient gene loss and expression divergence have begun. Selection in *B. napus* oilseed types has accelerated the loss of glucosinolate genes, while preserving expansion of oil biosynthesis genes. These processes provide insights into allopolyploid evolution and its relationship with crop domestication and improvement.
genomes, with asymmetric gene distribution (42,320 and 48,847, respectively) and 93% of the diploid gene space in orthologous blocks (fig. S12) (7). We identified 34,255 and 38,661 orthologous gene pairs between the $A_n$ and $C_n$ subgenomes and their respective progenitor genomes (fig. S13). Comparison of $A_n-A_n$ and $C_n-C_n$ orthologous gene pairs suggested a divergence 7500 to 12,500 years ago (fig. S14), indicating formation of $B. napus$ after this date. Synteny with Arabidopsis (table S19) confirmed the triplicated mesopolyploid structure (9–11) of the $A_n$ and $C_n$ subgenomes, with the recent allopolyploidy conferring on $B. napus$ an aggregate 72× genome multiplication since the origin of angiosperms (Fig. 1) (7).

**Fig. 1. Recurrent genome duplications in $B. napus$.** Genomic alignments between the basal angiosperm Amborella trichopoda (24), the basal eudicot Vitis vinifera (25), and the model crucifer A. thaliana, as well as $B. rapa$ (9), $B. oleracea$ (10, 11), and $B. napus$, are shown. A typical ancestral region in Amborella is expected to match up to 72 regions in $B. napus$ (69 were detected for this specific region). Gray wedges in the background highlight conserved synteny blocks with more than 10 gene pairs.

**Fig. 2. The genome of the $B. napus$ oilseed cultivar ‘Darmor-bzh’.** The genome comprises 9 chromosomes belonging to the $C_n$ subgenome and 10 to the $A_n$ subgenome, scaled on the basis of their assembled lengths. Tracks displayed are (A) gene density (nonoverlapping, window size = 100 kb for all tracks). Positions showing loss of one or more consecutive genes are displayed (triangles) along with homeologous exchanges, detected as missing genomic segments that have been replaced by duplicates of corresponding homeologous segments (red rectangles). (B and C) Transcription states estimated by RNA-seq in leaves (B) and roots (C) (in nonoverlapping 100-kb windows). (D) DNA transposon density. (E) Retrotransposon density. (F) CpG methylation in leaves (green) and roots (brown); both curves are overlapping. (G) Centromeric repeats (densities exaggerated for visual clarity). Homeologous relationships between $A_n$ and $C_n$ chromosomes are displayed with connecting lines colored according to the $C_n$ chromosomes.
Most orthologous gene pairs in *B. rapa* and *B. oleracea* remain as homeologous pairs in *B. napus* (tables S19 to S25 and figs. S12 to S17). DNA sequence analysis (7) confirmed the loss of 112 *An* and 91 *Cn* genes in *B. napus* ‘Darmor-bzh’ (tables S21 to S26), ~2.6 times higher than the 41 and 37 genes lost in *B. rapa* ‘Chiifu’ and *B. oleracea* ‘TO1000’ respectively (tables S26 and S27; $\chi^2$ test $P = 5.3 \times 10^{-14}$). Further analyses of a *Brassica* diversity set showed that ~47% of *Darmor-bzh* *An* and 31% of *Cn* deleted genes were also deleted in at least one additional progenitor genotype (tables S28 and S29), indicating that their deletion probably predated allopolyploidization of *B. napus* (7). A high proportion (27% to 54%) of the remaining *Darmor-bzh* deleted genes were also deleted from diverse *B. napus* genotypes (tables S28 and S29).

Homeologous exchanges (HEs), including crossovers and noncrossovers, are frequent between *B. napus* subgenomes and range in size from large segments to single SNPs (7) (Fig. 3, figs. S17 to S24, and tables S30 to S39). At the chromosome segment level, HEs are characterized by replacement of a chromosomal region with a duplicated copy from the corresponding homeologous subgenome (7). We identified 17 HEs, 14 *Cn* to *An* and 3 *An* to *Cn* (Fig. 3, fig. S19, and tables S30 and S31). Sequences from seven diverse *B. napus* genotypes revealed both shared and specific segmental HEs. These are of varying sizes and are most frequent between chromosomes *A*$_1$-*C*$_1$, *A*$_2$-*C*$_2$, and *A*$_9$-*C*$_9$ (table S32, Fig. 3, and fig. S19). Larger HEs found in the synthetic *B. napus* H165 affect, for example, most of chromosomes *A*$_1$-*C*$_1$ and *A*$_2$-*C*$_2$ (Fig. 3 and fig. S19).

Fig. 3. HEs between *B. napus* chromosomes *A*$_2$ and *C*$_2$. (A) Coverage depth obtained along the *A*$_2$ chromosome after mapping Illumina sequence reads from seven natural and one resynthesized *B. napus* genotypes (named on the right) to the reference genome of *B. napus* ‘Darmor-bzh’. (B and C) Coverage depth obtained for *A*$_2$ and *C*$_2$ chromosomes, respectively, after mapping >21 genome-equivalents of Illumina sequence reads from *B. napus* Darmor-bzh on the *B. rapa* and *B. oleracea* genome assemblies concatenated together. (D) Similar to (A), where the *C*$_2$ chromosome of Darmor-bzh is displayed. Segmental HEs are revealed based on sequence read coverage analysis, where a duplication (red) is revealed by significantly greater coverage for a given segment than the rest of the genome (black) and a deletion (blue) by little or no coverage for the corresponding homeologous segment. Sizes of chromosomes are indicated in Mb. *A*$_{1}$-to-*C*$_{1}$ converted genes (at 60% or more conversion sites) are plotted as blue dots on *A*$_2$ (B) and red dots on *C*$_2$ (C). *C*$_{1}$-to-*A*$_{1}$ converted genes are plotted as blue dots on *C*$_2$ (C) and red dots on *A*$_2$ (B). Open circles denote entirely converted genes using the same color code. Light gray lines connecting (A), (B), (C), and (D) indicate orthology relationships, and dark gray lines highlight segmental HEs in Darmor-bzh (names and descriptions detailed in table S31). Further HEs occurring between other homeologous chromosomes are shown in fig. S19. Black arrows in (A) indicate HEs involving GSL and FLC genes.
diversification of winter, spring, and Asian types of oilseed rape, rutabaga, and kale vegetables (Fig. 3B, fig. S19, and table S33).

We also identified 37 Cn to An and 56 An to Cn whole-gene conversions (22) (table S34).

At the single-nucleotide scale, exchanges between homeologous subgenomes account for ~86% of allelic differences between B. napus and its progenitors, with nearly ~1.3 times more between homeologous subgenomes account foring behavior, and improve pathogen resistance.

Genetic variation for reduced seed GSLs also appears to be under breeding-directed selection. GSLs are sulfur-rich secondary metabolites important for plant defense and human health (19). However, high levels in seeds form toxic breakdown products in animal feedstuffs. All 22 GSL catabolism genes identified in B. rapa and B. oleracea (10) are conserved in B. napus (7), and orthologs of only three Cn and one An GSL biosynthesis genes are missing (table S49). One deleted homeologous pair, corresponding to orthologs of B. oleracea Bt2g161590 and B. rapa Br020531, colocalize with a deletion (QTIs) for total aliphatic GSL content (20) and corresponds to a HE in which a segment of A2, with a missing C2 homeolog, has replaced the C2 homeolog (Fig. 3). Two additional QTIs for alliphatic GSL content (21) colocalize with a deletion of the B. rapa Br0353092 ortholog on A9 and its nondeleted homeolog on C9 (BnaC09g03500D, fig. S17).

We identified 425 nucleotide binding site leucine-rich repeat (NBS-LRR) sequences encoding resistance genes homologs (245 on Cn and 180 on An). Of these, 75% (153 An and 224 Cn) are syntenic to An and Cn progenitors (7) (table S50 and figs. S33 and S34). We confirmed the absence of five NBS-LRR genes from the An subgenome, three from the Cn subgenome, and three from B. rapa (An), with none absent from B. oleracea Cn. This variation may reflect differential selection for resistance to diseases.

B. napus morphotypes show broad adaptation to different climatic zones and latitudes. A key adaptive gene controlling vernalization and photoperiod responses, FLOWERING LOCUS C (FLC) is expanded from one copy in A. thaliana to four in B. rapa and B. oleracea and nine or more in B. napus (7) (table S51). Different FLC homologs lie within HEs, from C2 to A2 in the Asian semimature oilseed forms Yadul and Abruamasi (Fig. 3) and C9 to A10 in late-flowering swedes (fig. S19 and table S33). These loci correspond to important QTIs for vernalization requirement and flowering time (22).

Human culture and breeding of B. napus morphotypes may have selected favorable HEs, causing subgenome restructuring of regions containing genes controlling valuable agronomic traits such as those shown here for oil biosynthesis, seed GSL content, disease resistance, and flowering. Because B. napus is a young allotypic beginning gene loss and genome reorganization, further partitioning of expression may become a key determinant for the long-term preservation of its duplicated genes (23). The integrative genomic resources that we report provide unique perspectives on the early evolution of a domesticated allotypic and will facilitate the manipulation of useful variation, contributing to sustainable increases in oilseed crop production to meet growing demands for both edible and biofuel oils.

REFERENCES AND NOTES
6. H. L. Liu, Rpsedeed Genetics and Breeding (China Agricultural University, Beijing, 2000), pp. 1–81.
7. Supplementary materials for this article are available on Science Online.
22. X. Zou et al., PLOS ONE 7, e45751 (2012).

ACKNOWLEDGMENTS
Sequence Read Archive accession numbers of B. napus sequencing data are ERP005275 and PRJEB6609, and those of B. rapa and B. oleracea data are PRJNA243988 and PRJNA50627. The B. napus assembly is available at ENA (European Nucleotide Archive), in the WGS section for contigs (accession numbers CCCW010000001 to CCCW010044187) and the CON section for scaffolds, chromosomes, and annotation (accession numbers LK031197 to LT032985). The B. napus genome is also available at CoGe (http://genomevolution.org/CoGe/) and at (www.genoscope.cns.fr/brassicanapus) the Genoscope Genome Database, with additional tools for comparative genomic analysis. The B. napus segregating populations Darmor-bzh × Yadul and Darmor × Brisot are available at INRA-IGEPP, Rennes, France, under a material transfer agreement. This project was funded by the French ANR (Agence Nationale de la Recherche, www.agence-nationale-recherche.fr) 2009 (ANR-09-GENM-021) to B.C., P.W., D.B., and R.D., with additional funding from Soft-Protot for bioinformatic personnel (J.J.); the National Basic Research Program of China (2011CB101300) for S.L., Y.Z., C.G., and W.H.; and the Canadian Canola Genome Sequencing Initiative (http://saac-ucsd.ca/scanseq/, Genome Alberta, and industry partners) for I.A.P., A.G.S., C.K., and C.S. Research members are B.C., S.L., I.A.P., X.W., I.B., R.D., J.B., D.E., Y.Z., W.H., A.G.S., A.H.P., C.G., and P.W. Additional acknowledgments and author contributions are included in the supplementary materials.

SUPPLEMENTARY MATERIALS
www.sciencemag.org/content/345/6199/950/suppl/DC1

RESEARCH | REPORTS

Corrected 4 September 2014; see full text.
Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome


Science 345 (6199), 950-953.
DOI: 10.1126/science.1253435

The genomic origins of rape oilseed

Many domesticated plants arose through the meeting of multiple genomes through hybridization and genome doubling, known as polyploidy. Chalhoub et al. sequenced the polyploid genome of *Brassica napus*, which originated from a recent combination of two distinct genomes approximately 7500 years ago and gave rise to the crops of rape oilseed (canola), kale, and rutabaga. *B. napus* has undergone multiple events affecting differently sized genetic regions where a gene from one progenitor species has been converted to the copy from a second progenitor species. Some of these gene conversion events appear to have been selected by humans as part of the process of domestication and crop improvement.

*Science*, this issue p. 950