Genome structural evolution in *Brassica* crops

Zhesi He[1], Ruiqin Ji[1†], Lenka Havlickova[1], Lihong Wang[1], Yi Li[1], Huey Tyng Lee[2], Jiaming Song[3], Chushin Koh[4], Jinghua Yang[5], Mingfang Zhang[5], Isobel A.P. Parkin[6], Xiaowu Wang[7], David Edwards[8], Graham J King[9], Jun Zou[3], Kede Liu[3], Rod J Snowdon[2], Surinder S. Banga[10], Ivana Machackova[11] and Ian Bancroft[1]*

[1] Department of Biology, University of York, Heslington, York, YO10 5DD, UK

[2] Department of Plant Breeding, Justus Liebig University of Giessen, 35392 Giessen, Germany

[3] National Key Laboratory of Crop Genetic Improvement, College of Plant Science & Technology, Huazhong Agricultural University, Wuhan, China

[4] Global Institute for Food Security (GIFS), 110 Gymnasium Place, University of Saskatchewan, Saskatoon, SK S7N 0W9 Canada

[5] Department of Horticulture, College of Agriculture & Biotechnology, Zhejiang University, China, 310058

[6] Agriculture and Agri-Food Canada, 107 Science Place Saskatoon, SK, S7N OX2

[7] Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences (IVF, CAAS), Beijing, China

[8] School of Biological Sciences and the Institute of Agriculture, Faculty of Science, The University of Western Australia, Crawley, WA, Australia

[9] Southern Cross Plant Science, Southern Cross University, Lismore, NSW 2480, Australia

[10] Department of Plant Breeding and Genetics, Punjab Agricultural University, Ludhiana, India

[11] Selgen, a.s., Plant breeding station, Chlumec nad Cidlinou, 503 51, Czech Republic

[†] Current address: Department of Horticulture, Shenyang Agricultural University, Shenyang 110866, China

* Corresponding author Prof Ian Bancroft: ian.bancroft@york.ac.uk

Abstract

The cultivated *Brassica* species include numerous vegetable and oil crops of global importance. Three genomes (designated A, B and C) share mesohexapolyploid ancestry and occur both singly and in each pairwise combination to define the *Brassica* species. With organisational errors (such as misplaced genome segments) corrected, we showed that the fundamental structure of each of the genomes is the same, irrespective of the species in which it occurs. This enabled us to clarify genome evolutionary pathways, including updating the Ancestral Crucifer Karyotype (ACK) block organisation and providing support for the *Brassica* mesohexaploidy occurring via a two-step process. We then constructed genus-wide pan-genomes, drawing from genes present in any species in which the respective genome occurs, which enabled us to provide a global gene nomenclature system for the cultivated *Brassica* species and develop methodology to cost-effectively elucidate the genomic impacts of alien introgressions. Our advances not only underpin knowledge-based approaches to more efficient breeding of *Brassica* crops, it also provides an exemplar for the study of other polyploids.

The cultivated *Brassica* species include vegetable (e.g. cabbage, cauliflower, broccoli, pak choi, mustard greens, turnip and rutabaga), and condiment (e.g. mustard rape) crops, as well as the third largest source of vegetable oil globally (oilseed rape)[1]. Like *Arabidopsis thaliana*, which has been widely used to study fundamental plant science, the *Brassica* species are members of the Brassicaceae (Cruciferae or mustard) family. Ancestral genome reconstruction, based on genome sequences, provides insights into genome evolution[2]. Three genomes and pairwise combinations thereof distinguish the *Brassica* species. The A genome (AA; n=10) occurs in *B. rapa*, the B genome (BB; n=8) in *B. nigra* and the C genome (CC; n=9) in *B. oleracea*. These diploid genomes also occur in each pairwise combination to form the amphidiploid allotetraploid species *B. napus* (AACC; n=19), *B. juncea* (AABB; n=18) and *B. carinata* (BBCC; n=17)[3]. Genome sequences are available for most of the *Brassica* species[4-8] and are consistent with earlier studies that showed the *Brassica* “diploid” genomes to have mesopolyploid ancestry[9-12]. Assessment of the gene content shows a characteristic pattern of one segment showing greater gene retention (least fractionated,

LF) than the other two (which are more fractionated, MF1 and MF2). There are two hypothetical mechanisms for this biased fractionation: (1) a single step hexaploidization followed by expression dominance and loss of silenced genes, and (2) a two-step process, initially involving the formation of a tetraploid by alloploidization and fractionation, before hybridization to introduce a third genome, which shows less fractionation as it has been in a polyploid context for a shorter time.

Genetic studies indicated conserved organisation of the *Brassica* genomes, irrespective of the species in which they occur[13], with extensive collinearity between protein-coding genes[14,15]. However, studies based on whole genome sequencing suggested considerable variability[8]. Adaptation of methods used for high density genic single nucleotide polymorphism (SNP) markers[16,17] enabled the development of a quality assurance process based on scoring SNP markers tagging genes and the identification of inconsistencies in the purported organisation of the genome sequences[18]. Such corrections to the organisation of genome sequences are important for avoiding misleading results in, for example, the assessment of genome collinearity or positional cloning of genes.

Reference genome sequences do not represent the entire gene repertoire of a species. The concept of the pan-genome overcomes this limitation by representing the complete genic makeup of a species[19]. A pan-genome has been constructed for *B. oleracea*, using a reference-guided assembly approach[20]. However, that resource represents only the *Brassica* C genome as it occurs in *B. oleracea* only, i.e. it does not include the genes present in the C genome of *B. napus* and *B. carinata*. The reference-guided *B. napus* pan-genome[21] is similarly limited. Cross-species pan-genomes would better represent the gene pools from which genetic variation can readily be accessed for *Brassica* crop improvement.

Gene flow between species by horizontal transfer provides a source of genetic variation and enables adaptation, particularly in plants[22,23]. Widely used as a traditional plant breeding method for broadening genetic diversity, this “alien introgression” approach has enormous scope for future crop improvement[24]. However, the technical difficulty of assessing the extent of introduced or

exchanged genome segments at the molecular level impairs a deeper understanding of genome changes. An example is the fertility restoration locus for the Ogura cytoplasmic male sterility system (*Rfo*), widely used in oilseed rape, which involves a large segment of the radish genome[25]. The gene responsible, a pentatricopeptide repeat (PPR) gene, is known[26], and genetically linked molecular markers are available[27]. However, a lack of genetic recombination within a very large introgressed chromosomal segment had made the extent of linked radish sequences and the substituted *Brassica* chromosome segment difficult to define[28].

In this study, we corrected errors in the organisation of *Brassica* genome sequences and used the improved resource to elucidate the ancestry and evolution of the *Brassica* A, B and C genomes. Using this knowledge, we defined cross-species pan-genomes to underpin a systematic gene nomenclature system for use by the *Brassica* research community. We also used the new resource to assess the complex genome impacts of alien introgression.

Results

1. Establishing fundamental genome configurations for the *Brassica* genus

Numerous draft genome sequences have been produced for *Brassica* species, with comparative analyses suggesting substantial differences in genome organisation both between and within species. However, these genome assemblies were not well controlled for misassembly that can result either from repetitive sequences contributing to chimeric scaffolds, or from anomalous placement of scaffolds based on sparse genetic linkage maps. Smaller scaffolds can be particularly difficult to position as the linkage maps used to support the original genome sequence assembly will include regions with limited resolution due to lack of polymorphism between the parents or low recombination rates. In polyploids, polymorphisms between homoeologous regions of genomes can be difficult to differentiate from allelic variation and can lead to mis-assignment of scaffolds to incorrect genomes. To address these shortcomings, the availability of a high density of polymorphic markers is important to assess the reliability of mapping by enabling the discrimination of occasional mis-mapping markers from multiple markers correctly mapping a scaffold. Increasing

resolution, by using a larger mapping population to produce more recombination events, adds relatively little if there are additional approaches, such as collinearity with other genomes that can be drawn upon. Advances in high resolution linkage mapping using data from transcriptome re-sequencing[17] enables the quality of assemblies to be revisited. For example, analysis of the first genome assembly for *B. juncea*[8] using genome-ordered graphical genotypes (GOGGs)[18] enabled rapid correction into a much-improved version[29]. This demonstrated the utility of using multiple linkage mapping populations to provide more comprehensive resolution across a genome by complementing gaps in polymorphism and/or recombination, and that these do not necessarily need to be made using the same species as that from which the genome sequence is derived, where the same genome is represented in multiple species. In the present study, we visualized genome assemblies for each diploid species using GOGGs that were generated using SNP markers produced from transcriptome or genome re-sequencing of double haploid (DH) linkage mapping populations. For each diploid genome, two high-density genetic maps were used that represented the two corresponding allotetraploid species in which the respective diploid genome is found (*i.e. B. napus* and *B. juncea* for the *B. rapa* Chiifu v3 A genome, *B. juncea* and *B. carinata* for the *B. nigra* early draft NI100 B genome, *B. napus* and *B. carinata* for the *B. oleracea* TO1000 C genome). Many anomalies were detected, whereby blocks of adjacent markers clearly mapped genetically to positions discordant with their physical placement in the genome assemblies. The markers used for generating GOGGs can all be assigned to gene models, enabling comparative analysis. A distinctive feature of such misplaced genome sequence segments is their lack of collinearity to orthologous genes in both *A. thaliana* and *T. parvula*, whereas they show perfect collinearity to positions indicated by linkage mapping. Thus, such non-collinearity is a good indicator of problematic regions in genome assemblies and precise re-positioning can be achieved using collinearity as a fine-scale guide. We therefore worked manually through the GOGG data (visualized in MS Excel spreadsheets) for all three diploid *Brassica* genome sequences and moved segments manually to achieve congruous linkage mapping and collinearity with *A. thaliana* and *T. parvula* genomes, using the appropriate pairs of allotetraploid linkage mapping datasets. Each potential rearrangement was cross-checked with each of the relevant allotetraploid linkage maps and, remarkably, no mapping conflicts were identified. Our interpretation is that the fundamental

organisation of the *Brassica* A, B and C genomes is conserved across species. The resulting GOGGs are illustrated in Figure 1 while the GOGG data, including details of the markers and gene models, are provided in Supplementary Data 1, 2 and 3. The reorganised genomes, as represented by ordered gene models, are provided in Supplementary Data 4. Due to their polyploid ancestry and strong artificial selection by breeders, some individual accessions of cultivated *Brassica* species are expected to have inherited common structural genome rearrangements, particularly in the allopolyploid species[30]. Genome structural rearrangements are expected to inhibit recombination by blocking pairing in hererozygotes[31], and therefore cannot be elucidated by linkage mapping. Consequently, in chromosome segments where no recombination data useful for linkage mapping was available, the organisation of the corresponding diploid genome sequence in this region was accepted by default.

2. Defining collinearity relationships of the *Brassica* genomes

Shared ancestry with a mesohexaploid progenitor results in extensive collinearity between the *Brassica* A, B and C genomes. Most effort has focused on comparative analysis of the A and C genomes. Linkage mapping with very high densities of markers confirmed large collinearity blocks (extending to the scale of some whole chromosomes, for example A1 and C1) but with disruption of other parts of the genome into small, apparently non-collinear segments[32]. This lack of uniformity in conserved synteny across the genomes may be due partly to noise in the analyses, for example from genome assemblies containing chimeric scaffolds or by errors in linkage mapping with molecular markers, the scoring of which can be complex in polyploids with closely related genomes, such as *B. napus*[16]. In order to develop a clear understanding of the structural genome evolution leading to the organisation of the extant *Brassica* genomes from their most recent common ancestor, we aimed first to identify the most reliable orthology relationships. To do this, we undertook a 3-way BLAST similarity analysis between all gene models in each of the re-assembled *Brassica* A, B and C genomes. We considered as putative orthologous triplets sets of genes that gave reciprocal top BLAST hits with each other. This conservative approach is designed to minimise noise caused by spurious sequence similarities and resulted in the identification of 22,691 triplets. The set of gene models was curated manually, resulting in a set of

21,328 orthologous triplets present in collinear segments across all three genomes, as shown in Supplementary Data 5. Pairwise comparisons of the organisation of the genomes are illustrated in Figure 2. Assessment of molecular markers (as used for the development of GOGGs) could be used to confirm the positioning of small collinear segments by linkage mapping. These results confirm that synteny can be conserved on the chromosome scale, but also that collinearity can break down into relatively small genome segments when rearrangement commences.

3. Identification of conserved ancestral genome blocks

The Brassicaceae provide an excellent model for studying genome structural evolution. Pioneering studies led to the development of a proposed "Ancestral Crucifer Karyotype" (ACK) comprising 24 collinearity blocks (labelled A to X)[33], with greater understanding developing as more genome sequences emerged for the family[34]. Our improved organisation of the *Brassica* genomes was developed without reference to pre-existing knowledge of the ACK collinearity blocks, so a re-evaluation of them provides both the opportunity for new insights and a further quality check on our resource. The genome triplication represented in the *Brassica* species occurred close to the time of separation between the *Brassica* and *T. parvula* lineages, making the *T. parvula* genome a good representative of the pre-triplication genome inherited by *Brassica* species[35]. We therefore used the top *T. parvula* gene model BLAST hits with each of the 21,328 orthologous *Brassica* gene triplets (Supplementary Data 5) to define collinearity blocks that we could subsequently align with the organisation with the previously defined ACK blocks[2]. As shown in Table 1, the results corresponded very well. Our analysis is entirely consistent with the previously-called ancestral blocks, apart from an indication that two small additional blocks can be defined. We label these new blocks as V2 and W2, with the blocks corresponding to the original V and W re-named V1 and W1, respectively, in Table 1. All of our 136 *Brassica* ACK genome blocks could be identified in both *T. parvula* and *A. thaliana* genomes apart from two blocks (which we refer to as Tp blocks 27.5 and 29.5; see Supplementary Data 6) that are not represented in the *T. parvula* genome sequence. This set of 136 ACK genome blocks can be arranged to represent any of the *Brassica* A, B or C genomes or, with 3-fold redundancy, the diploid Brassicaceae (i.e. *A. thaliana* or *T. parvula*) genomes.

4. Inference of the origins of paralogous sub-genome segments

One of the unresolved questions in the evolution of *Brassica* genomes is the mechanism by which one sub-genome exhibits less gene loss (fractionation) than the other two, when compared with an orthologous Brassicaceae genome such as *A. thaliana* or *T. parvula*. We used our improved organisation of the *Brassica* genomes to address this, inferring, at least partially, the chromosomal organisation of their common ancestor prior to speciation. Assuming the rate of structural rearrangement of the subgenomes is approximately constant over time, an approximately equal rearrangement would be expected in each sub-genome if they had come together simultaneously to form a hexaploid, which then preserved one less fractionated sub-genome by a mechanism such as methylation to substantively silence the other two. Alternatively, a two-stage process, in which two genomes fractionated for an extended period of time in a tetraploid before addition of a third genome formed the hexaploid, would result in less rearrangement of the genome that joined most recently, i.e. the less fractionated sub-genome. The approach we used to distinguish between these two scenarios was to manually assess the 134 ACK collinearity blocks with collinearity to *T. parvula* for fusions represented in the *Brassica* A, B and C genomes, which represent their configurations in the ancestors of the *Brassica* genomes. These could be sorted, based on complementarity of break positions between blocks, into larger blocks and a least fractionated block identified based on gene content in *Brassica* as a proportion of all orthologues in the corresponding region of the *T. parvula* genome (Supplementary Data 7). The 136 ACK collinearity blocks (including the two with no *T. parvula* orthologues) were then assembled into their putative configuration in the ancestors of the three *Brassica* genomes, based on adjacencies in the extant *Brassica* genomes (Figure 3, Supplementary Data 8) . The least fractionated blocks almost perfectly represent 7 ancestral chromosomes that exhibit excellent collinearity with the *T. parvula* genome, with only one small block (containing orthologues of Tp6g07740-Tp6g09710, positioned in chromosome 6) having been classified as more fractionated. Ten additional putative ancestral chromosomes were identified that comprise more fractionated blocks, with only one exception: one small block contains orthologues of Tp7g11790-Tp7g11970 (positioned in chromosome 14) that had been classified as least fractionated. These chromosomes exhibited numerous fusions and

rearrangements, indicating extensive rearrangement since formation of the polyploid. Our findings therefore support the two-stage process, with the hexaploid being formed by the fusion of a third genome, very similar in organisation to that of *T. parvula*, with a tetraploid that had already been undergoing fractionation and rearrangement.

5. Construction of cross-species pan-genomes

Having developed a robust understanding of the organisation and evolution of the *Brassica* A, B and C genomes, we next aimed to optimise our knowledge of the gene complement of the genus. Because we have shown that the fundamental organisation of the genomes is the same irrespective of the species, and each genome is shared across three different species of the genus, we aimed to develop the first genus-wide pan-genomes. Doing this in a robust and comprehensive manner required the development of a new approach to pan-genome construction. The selection of the most appropriate underpinning representative of each genome is important for minimising errors and to provide a resilient gene nomenclature for the genus. First, we evaluated genome sequence resources for completeness and fidelity of organisation, using GOGGs. We concluded, on the basis of least errors in organisation, that the most reliable were the genome sequences of *B. rapa* Z1 for the A genome[36], *B. nigra* Ni100 for the B genome[37] and *B. oleracea* HDEM for the C genome[36]. We integrated unanchored scaffolds as listed in Supplementary Data 9 then anchored (based on BLAST similarity) or interpolated (based on collinearity of flanking genes) gene models from 12 additional genome sequence resources, as listed in Table 2. These genome sequences were released previously to our analysis or in parallel with them; we draw upon their gene content. The gene models comprising the pan-genome were given systematic pan-genome names, as approved by the Steering Committee of the Multinational Brassica Genome Project. These are shown in Supplementary Data 10 (for the A genome), Supplementary Data 11 (for the B genome) and Supplementary Data 12 (for the C genome), along with other details of their source genomes. Of the 197,465 gene models in the *Brassica* pan-genomes, 165,698 (83.9%) were already represented in the underpinning genomes. Of these, 104,339 (63.0%) showed significant similarity to orthologues in *T. parvula* and 93,361 (56.3%) showed significant similarity to orthologues in *A. thaliana*. Lower proportions of genes showed significant similarity for the B and C

genomes (58.8% and 57.1%, respectively) than did gene models from the A genome (68.3%). A substantially lower proportion of the 31,767 gene models integrated from additional genome sequences to form the pan-genomes showed significant similarity to orthologues in *T. parvula* (12.0%, 12.5% and 10.1% for A, B and C genomes, respectively).

6. Using the pan-genomes to elucidate the genomic impacts of alien trait introgression

Alien introgression is an important approach to broadening the genetic diversity of crops by exchanging genetic material between related species. The genomic impacts have been difficult to assess, restricting the application of the system to targets of high commercial value, such as the introduction of components of F_1 hybrid production systems. However, there is growing demand for novel genetic resistances to both biotic and abiotic stresses, and strong allopolyploidization and breeding bottlenecks necessitate the introduction of novel genetic diversity for such traits from secondary or tertiary gene pools. Having generated a comprehensive platform for describing genomic variation in the cultivated *Brassica* species, we tested whether it could be used to support genome analysis of alien introgression lines. As a first example, we assessed the widely-used Ogura fertility restorer system. The restorer gene carried by the *Rfo* locus (the orthologue of which in our pan-genome is C09p019070.1_BolHDE) was introgressed into *B. napus* (oilseed rape) from the closely-related species *Raphanus sativus* (radish), however the extent of the *Brassica* genome which was replaced has not been described at the gene level. As a second example, we screened a population of lines with putative introgressions into *B. juncea* from *Brassica fruticulosa*. Because well-assembled genome sequences are seldom available for wild relatives, we implemented a method termed “curing” to edit a *Brassica* pan-genome to more closely represent that of a donor species[38]. For the introgression into *B. napus* (which contains the *Brassica* A and C genomes) we cured the *Brassica* B pan-genome with *R. sativus* genome sequence reads and re-named this reference genome R. For the introgression into *B. juncea* (which contains the *Brassica* A and B genomes) we cured the *Brassica* C pan-genome with *B. fruticulosa* genome sequence reads and re-named this reference genome F. We then mapped genome sequence reads from recipient and donor parents, and introgression lines, to the appropriate triple reference sequence (ACR or ABF) and processed as described previously for the visualization of homoeologous exchanges in

polyploid species[30]. The resulting Genome Display Tile Plots (GDTPs) are shown in Figure 4, while a quantification of reads mapping to each gene in the reference is provided in Supplementary Data 13 for the radish intogression and Supplementary Data 14 for the *B. fruticulosa* intogression. This approach clearly delineated the extent of the *R. sativus* introgression into *B. napus* and a *B. fruticulosa* introgression into *B. juncea*. Other putative *B. fruticulosa* introgression lines showed no evidence of capturing genomic DNA from the donor species, but did show segmental losses from the recipient genome.

As there is a genome sequence available for *R. sativus*, we repeated the analysis for the *Rfo* introgression using the radish genome sequence as the R genome reference. This yielded the same result as the cured reference sequence (Extended Data Figure 1), but with the advantage that visualization was also possible based on the genome order of the radish genes. Testing for the radish introgression, similar results could also be obtained using mRNAseq reads instead of gDNA reads (Extended Data Figure 2), making the approach cost-effective even for species with very large genomes where genome re-sequencing would be prohibitively expensive.

Discussion

Rigorous assessment of genome sequences available for *Brassica* species enabled us to show conclusively that that the fundamental organisation of the *Brassica* A, B and C genomes is conserved across the multiple species in which they occur. By identifying triplets of orthologous genes across all three genomes, we were able to define collinearity relationships between the genomes, demonstrating a wide range of sizes for blocks of collinearity, ranging from a few genes to whole chromosomes. Using the resource, in particular the collinearity observed with the genome of *T. parvula* (as the best extant representative of the genome structure of the ancestor of *Brassica* species prior to genome triplication), we were able to re-visit analyses of ancestral genome evolution based on collinear genome blocks traceable to the ACK. We identified two additional blocks (V2 and W2) and showed the existence of blocks in the *Brassica* genomes for which no orthologous segments exist in the *T. parvula* genome sequence, providing 136 genome blocks that

can be rearranged to represent the structure of any of the *Brassica* genomes. Using these structures, we were able to infer the structure of the least fractionated *Brassica* sub-genome, showing that prior to rearrangement it comprised 7 chromosomes of similar structure to those present in *T. parvula*. In contrast, the more fractionated sub-genomes appear to be derived from two genomes that had undergone previous rearrangement, supporting the hypothesised 2-step derivation of the mesohexaploid structure, initially involving the formation of a tetraploid and a period of rearrangement and fractionation, before further hybridization to introduce a third genome, which shows less fractionation as it has been in a polyploid context for a shorter time.

In addition to understanding more of the evolutionary process by which the *Brassica* genomes evolved, we developed resources to underpin future research and breeding in the many commercially important *Brassica* crops. First, we established the first genus-wide pan-genomes for any species. Of the 197,465 gene models in the combined *Brassica* A, B, and C pan-genomes, the majority (165,698) were already represented in the underpinning genomes and most of these (>55%) showed significant similarity to orthologues in *T. parvula*. A notably lower proportion of the 31,767 gene models integrated from additional genome sequences to form the pan-genomes showed such similarity to orthologues in *T. parvula* (<13%), which may indicate that these contain a greater proportion of spurious gene models. This genus-wide pan genome resource provides a framework for describing *Brassica* gene content, via a systematic new nomenclature approved by the Multinational *Brassica* Genome Project. We demonstrated the practical utility of this resource to support crop breeding to introduce traits from related species by developing a novel approach for the first cost-effective analysis of alien introgressions at the level of gene sequences, providing examples of a radish introgression into *B. napus* and a *B. fruticulosa* introgression into *B. juncea*.

References

1 USDA-FAS. *USDA Oilseeds: World markets and trade,* 2020).

2 Murat, F. *et al.* Understanding Brassicaceae evolution through ancestral genome reconstruction. *Genome biology* **16***,* 262 (2015).

3 Nagaharu, U. Genome analysis in Brassica with special reference to the experimental formation of B. napus and peculiar mode of fertilization. *Jpn J Bot* **7**, 389-452 (1935).

4 Wang, X. *et al.* The genome of the mesopolyploid crop species Brassica rapa. *Nature genetics* **43**, 1035 (2011).

5 Liu, S. *et al.* The Brassica oleracea genome reveals the asymmetrical evolution of polyploid genomes. *Nature communications* **5**, 3930, doi:10.1038/ncomms4930 (2014).

6 Parkin, I. *et al.* Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid Brassica oleracea. *Genome Biology* **15**, R77 (2014).

7 Chalhoub, B. *et al.* Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. *Science* **345**, 950-953, doi:10.1126/science.1253435 (2014).

8 Yang, J. *et al.* The genome sequence of allopolyploid Brassica juncea and analysis of differential homoeolog gene expression influencing selection. *Nature genetics* **48**, 1225-1232, doi:10.1038/ng.3657 (2016).

9 Lagercrantz, U. & Lydiate, D. J. Comparative genome mapping in Brassica. *Genetics* **144**, 1903-1910 (1996).

10 O'neill, C. M. & Bancroft, I. Comparative physical mapping of segments of the genome of Brassica oleracea var. alboglabra that are homoeologous to sequenced regions of chromosomes 4 and 5 of Arabidopsis thaliana. *The plant journal* **23**, 233-243 (2000).

11 Yang, T.-J. *et al.* Sequence-level analysis of the diploidization process in the triplicated FLOWERING LOCUS C region of Brassica rapa. *The Plant cell* **18**, 1339-1347 (2006).

12 Town, C. D. *et al.* Comparative genomics of Brassica oleracea and Arabidopsis thaliana reveal gene loss, fragmentation, and dispersal after polyploidy. *The Plant cell* **18**, 1348-1359 (2006).

13 Parkin, I. A., Sharpe, A. G., Keith, D. J. & Lydiate, D. J. Identification of the A and C genomes of amphidiploid Brassica napus (oilseed rape). *Genome / National Research Council Canada = Genome / Conseil national de recherches Canada* **38**, 1122-1131, doi:10.1139/g95-149 (1995).

14 Rana, D. *et al.* Conservation of the microstructure of genome segments in Brassica napus and its diploid relatives. *Plant Journal* **40**, 725-733, doi:10.1111/j.1365-313X.2004.02244.x (2004).

15 Cheung, F. *et al.* Comparative analysis between homoeologous genome segments of Brassica napus and its progenitor species reveals extensive sequence-level divergence. *The Plant cell* **21**, 1912 - 1928 (2009).

16 Trick, M., Long, Y., Meng, J. & Bancroft, I. Single nucleotide polymorphism (SNP) discovery in the polyploid Brassica napus using Solexa transcriptome sequencing. *Plant biotechnology journal* **7**, 334-346, doi:10.1111/j.1467-7652.2008.00396.x (2009).

17 Bancroft, I. *et al.* Dissecting the genome of the polyploid crop oilseed rape by transcriptome sequencing. *Nature biotechnology* **29**, 762-766, doi:10.1038/nbt.1926 (2011).

18 He, Z. & Bancroft, I. Organization of the genome sequence of the polyploid crop species Brassica juncea. *Nature genetics* **50**, 1496-1497 (2018).

19 Vernikos, G., Medini, D., Riley, D. R. & Tettelin, H. Ten years of pan-genome analyses. *Current opinion in microbiology* **23**, 148-154 (2015).

20 Golicz, A. A. *et al.* The pangenome of an agronomically important crop plant Brassica oleracea. *Nature communications* **7**, 13390 (2016).

21 Dolatabadian, A. *et al.* Characterization of disease resistance genes in the Brassica napus pangenome reveals significant structural variation. *Plant biotechnology journal* (2019).

22 Mallet, J. Hybridization as an invasion of the genome. *Trends in ecology & evolution* **20**, 229-237 (2005).

23 Arnold, M. L. Transfer and origin of adaptations through natural hybridization: were Anderson and Stebbins right? *The Plant cell* **16**, 562-570 (2004).

24 Zamir, D. Improving plant breeding with exotic genetic libraries. *Nature reviews. Genetics* **2**, 983-989, doi:10.1038/35103589 (2001).

25 Delourme, R., Horvais, R., Vallée, P. & Renard, M. Double low restored F1 hybrids can be produced with the Ogu-INRA CMS in rapeseed. in *Proc. 10th Int. Rapeseed Congress.*

26 Brown, G. G. *et al.* The radish Rfo restorer gene of Ogura cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats. *The Plant journal : for cell and molecular biology* **35**, 262-272, doi:10.1046/j.1365-313X.2003.01799.x (2003).
27 Hu, X. *et al.* Mapping of the Ogura fertility restorer gene Rfo and development of Rfo allele-specific markers in canola (Brassica napus L.). *Molecular breeding* **22**, 663-674 (2008).
28 Feng, J. *et al.* Physical localization and genetic mapping of the fertility restoration gene Rfo in canola (Brassica napus L.). *Genome / National Research Council Canada = Genome / Conseil national de recherches Canada* **52**, 401-407 (2009).
29 Yang, J., Ji, C., Liu, D., Wang, X. & Zhang, M. Reply to: 'Organization of the genome sequence of the polyploid crop species Brassica juncea'. *Nature genetics* **50**, 1497-1498, doi:10.1038/s41588-018-0240-7 (2018).
30 He, Z. *et al.* Extensive homoeologous genome exchanges in allopolyploid crops revealed by mRNAseq-based visualization. *Plant biotechnology journal* **15**, 594-604, doi:doi:10.1111/pbi.12657 (2017).
31 Crown, K. N., Miller, D. E., Sekelsky, J. & Hawley, R. S. Local inversion heterozygosity alters recombination throughout the genome. *Current Biology* **28**, 2984-2990. e2983 (2018).
32 Bancroft, I., Fraser, F., Morgan, C. & Trick, M. Collinearity analysis of Brassica A and C genomes based on an updated inferred unigene order. *Data in brief* **3**, 51-55 (2015).
33 Schranz, M., Lysak, M. & Mitchell-Olds, T. The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. *Trends Plant Sci* **11**, 535 - 542 (2006).
34 Lysak, M. A., Mandáková, T. & Schranz, M. E. Comparative paleogenomics of crucifers: ancestral genomic blocks revisited. *Current Opinion in Plant Biology* **30**, 108-115 (2016).
35 Cheng, F. *et al.* Deciphering the diploid ancestral genome of the mesohexaploid Brassica rapa. *The Plant cell* **25**, 1541-1554 (2013).
36 Belser, C. *et al.* Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nature plants* **4**, 879 (2018).
37 Perumal, S. *et al.* A high-contiguity Brassica nigra genome localizes active centromeres and defines the ancestral Brassica genome. *Nature Plants* **6**, 929-941 (2020).
38 Higgins, J., Magusin, A., Trick, M., Fraser, F. & Bancroft, I. in *BMC genomics* Vol. 13 247 (2012).
39 Camacho, C. *et al.* BLAST+: architecture and applications. *BMC bioinformatics* **10**, 421 (2009).
40 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics* **25**, 1754-1760 (2009).
41 Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
42 Zhang, L. *et al.* Improved Brassica rapa reference genome by single-molecule sequencing and chromosome conformation capture technologies. *Horticulture research* **5**, 1-11 (2018).
43 Zou, J. *et al.* Genome-wide selection footprints and deleterious variations in young Asian allotetraploid rapeseed. *Plant biotechnology journal* (2019).
44 Song, J.-M. *et al.* Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of Brassica napus. *Nature Plants*, 1-12 (2020).
45 Lee, H. *et al.* Chromosome-scale assembly of winter oilseed rape Brassica napus. *Frontiers in Plant Science* **11**, 496 (2020).

Acknowledgments

This work was supported by: UK Biotechnology and Biological Sciences Research Council BB/L002124/1 and BB/R019819/1 to IB, the China Agriculture Research System CARS-25-A-03 and the Natural Science Foundation of Liaoning Province, China 2013020071 to RJ, grants

031B0890A from BMBF and SN14/22-1 from DFG to RJS and HTL, Australia Research Council Project LP160100030 to DE, National Natural Science Foundation of China 31970564 to JZ, and Indian Council of Agricultural Research F.No.27(5)/2007-HRD and Department of Biotechnology and Government of India BT/01/CEIB/12/I/03 to SSB.

Author contributions

IB conceived the work. ZH, RJ, LH, IM and IB designed the experiments. ZH, RJ, LH, LW, YL, HYL, JS, CK, JY, MZ, IAPP, XW, DE, GJK, JZ, KL, RJS, SSB acquired, analysed and/or interpreted the data. ZH and IB drafted the manuscript.

Competing interests statement

The authors declare no competing interests.

Figure legends

Figure 1. Genome-ordered graphical genotypes for the *Brassica* A, B and C genomes as represented in allotetraploid species. Graphical genotypes are shown for transcriptome or genome SNP markers scored across three doubled haploid (DH) linkage mapping populations: (1) 119 lines of the Varuna x Heera (VHDH) mapping population for A genome *B. juncea* and B genome *B. juncea* (Heera alleles in coral, Varuna alleles in blue and missing scores in white). (2) 45 lines of the Tapidor x Ningyou 7 (TNDH) mapping population for A genome *B. napus* and C genome *B. napus* (Ningyou 7 alleles in coral, Tapidor alleles in blue and missing scores in white). (3) 93 lines of the Yellowcross x Whiteban (YWDH) mapping population for B genome *B. carinata* and C genome *B. carinata* (Whiteban alleles in coral, Yellowcross alleles in blue and missing scores in white). The multi-coloured bars are colour-coded by the top BLAST sequence similarity match to the chromosomes in *Arabidopsis thaliana* (left bar) and *Thellungiella parvula* (right bar) of the

Brassica gene model in which each respective SNP is scored (light blue = chromosome 1, orange = chromosome 2, dark blue = chromosome 3, green = chromosome 4, red = chromosome 5, salmon = chromosome 6, yellow = chromosome 7, light grey = no BLAST hit with E-value < $1e^{-30}$).

Figure 2. Collinearity of *Brassica* A, B and C genomes. 21,328 triplets of orthologous genes in the *Brassica* A, B and C genomes are plotted by their order in the respective genomes: (a) *Brassica* A and C genomes, (b) *Brassica* A and B genomes and (c) *Brassica* C and B genomes.

Figure 3. Inferred structure of chromosomes in the nascent triplicated ancestral genome. Partial hypothesised genome structure is depicted in relation to genome blocks collinear between *Brassica* and *T. parvula*, with the range of *T. parvula* orthologues indicated for each block and colour-coded by chromosome (Tp1 = light blue; Tp 2 = orange; Tp 3 = dark blue, Tp 4 = green, Tp 5 = red, Tp 6 = yellow, Tp 7 = beige). The genome segments identified in *Brassica* for which there is no identified *T. parvula* orthologue are shown as grey blocks. Hypothesised ancestral chromosomes 1 to 7 comprise *Brassica* blocks identified as least fractionated and 8 to 17 comprise *Brassica* blocks identified as more fractionated, except those illustrated with *T. parvula* orthologue names in red font. Arrows indicate the relative orientations of the inferred contiguous blocks. Chromosome numbers are assigned arbitrarily, except 1 to 7, which are assigned numbers to match the numbering of the orthologous *T. parvula* chromosomes.

Figure 4. Visualization of genomic impacts of alien introgression into allotetraploid Brassica species. Genome Display Tile Plots were generated based on the relative abundance of genome sequence reads mapping to three reference genomes, two of which are genomes of the introgression recipient species and the third a cured genome representing the introgression donor species. Quantification is represented in CMYK colour space for orthologous gene triplets. **(a)**

Rapeseed Ogura hybrid system with a radish (*Raphanus sativus*) introgression in *Brassica napus*. The cyan component represents abundance of the *Brassica* A genome orthologue, the yellow component that of the *Brassica* C genome orthologue and the magenta component that of the radish (R) genome orthologue, produced by curing. The triplets are plotted in *Brassica* C genome order, along with controls comprising parental species and *in silico* combinations to render a diagnostic colour key. Three plants representing the male sterile (CMS) plants of the hybrid system (no introgression) and three plants containing the radish introgression harbouring the restorer (Rfo) gene are illustrated. **(b)** Mustard rape (*Brassica juncea*) lines with *Brassica fruticulosa* introgression. The cyan component represents abundance of the *Brassica* A genome orthologue, the yellow component that of the *Brassica* B genome orthologue and the magenta component that of the *B. fruticulosa* (F) genome orthologue, produced by curing. The triplets are plotted in *Brassica* A genome order, with only A1 shown, along with controls comprising parental species and *in silico* combinations to render a diagnostic colour key. Four putative introgression lines are illustrated, including one with no introgression in this chromosome (AD-19-003), one that has inherited a large heterozygous deletion of the orthologous part of the *Brassica* B genome (AD-19-014), one that has inherited a small homozygous deletion of the *Brassica* A genome (AD-19-027) and one confirmed introgression line which has inherited a substitution of part of the *Brassica* A genome by *B. fruticulosa* sequences (AD-19-037).

Tables

Table 1. Ancestral Crucifer Karyotype genome blocks identified across the *Brassica* genomes.

ACK block	At start	Tp start	At end	Tp end
A	AT1G01030	Tp1g00050	AT1G19530	Tp1g17400
B	AT1G19570	Tp1g17410	AT1G42990	Tp1g32200
C	AT1G43130	Tp1g32290	AT1G56180	Tp1g41870
D	AT1G64670	Tp2g00040	AT1G56230	Tp2g06750
E	AT1G64960	Tp5g19220	AT1G80950	Tp5g36020
F	AT3G01040	Tp3g00010	AT3G25520	Tp3g23020

G	AT2G04050	Tp3g23100	AT2G06200	Tp3g24630
H	AT2G11520	Tp3g26330	AT2G20900	Tp3g34170
I	AT2G20920	Tp4g00010	AT2G29710	Tp4g12510
J	AT2G29980	Tp4g12540	AT2G48140	Tp4g30080
K	AT2G01070	Tp2g12540	AT2G04038	Tp2g14790
L	AT3G25545	Tp2g14810	AT3G30975	Tp2g18310
M	AT3G42640	Tp5g18280	AT3G49730	Tp5g12350
N	AT3G49790	Tp5g12330	AT3G63530	Tp5g00010
O	AT4G00080	Tp6g00060	AT4G05420	Tp6g04790
P	AT4G12600	Tp6g04910	AT4G08320	Tp6g09620
Q	AT5G29560	Tp2g19970	AT5G23090	Tp2g24020
R	AT5G23000	Tp6g22390	AT5G01010	Tp6g41230
S	AT5G42100	Tp7g00010	AT5G35220	Tp7g07090
T	AT4G12650	Tp7g09040	AT4G16630	Tp7g15250
U	AT4G16765	Tp7g15470	AT4G40080	Tp7g37560
V1	AT5G42220	Tp7g08390	AT5G42420	Tp7g07940
V2	AT5G42490	Tp2g06900	AT5G47780	Tp2g12520
W1	AT5G47820	Tp2g18320	AT5G49620	Tp2g19930
W2	AT5G49660	Tp6g10770	AT5G60800	Tp6g22380
X	AT5G60820	Tp2g23880	AT5G67630	Tp2g30710

At start = BLAST hits of *Brassica* CDS gene models to *Arabidopsis thaliana* CDS gene models: lowest gene model number

At end = BLAST hits of *Brassica* CDS gene models to *Arabidopsis thaliana* CDS gene models: highest gene model number

Tp start = BLAST hits of *Brassica* CDS gene models to *Thellungiella parvula* CDS gene models: lowest gene model number

Tp end = BLAST hits of *Brassica* CDS gene models to *Thellungiella parvula* CDS gene models: highest gene model number

529 Table 2. Sources of 197,465 gene models in cross-species *Brassica* pan-genomes.

Source	**Reference**	**Code**	**Order added**	**A genome gene models**		**B genome gene models**		**C genome gene models**	
				Total	Tp hits	Total	Tp hits	Total	Tp hits
B. rapa Z1	Belser et al[36] 2018	BraZAA	1	45819	31322				
B. nigra Ni100	Perumal et al[37] 2020	BniNIA	1			59422	34940		
B. oleracea HDEM	Belser et al[36] 2018	BolHDE	1					60457	34539
B. rapa Chiifu v3	Zhang et al[42] 2018	BraCHC	2	2863	592				
B. nigra YZ12151	Yang et al[8] 2016	BniYZA	2			598	69		
B. oleracea Pangenome	Golicz et al[20] 2016	BolBOP	2					1888	581
B. rapa R-o-18	King et al unpublished	BraROA	3	3387	172				
B. oleracea 02-12	Liu et al[5] 2014	BolBOA	3					594	146
B. napus Pangenome	Dolatabadian et al[21] 2019	BnaBNP	4	2409	201			7724	314
B. napus Ningyou 7	Zou et al[43] 2019	BnaNYA	5	1241	111			2950	161
B. napus ZS11	Song et al[44] 2020	BnaZSA	6	500	108			775	125
B. napus Express 617	Snowdon et al[45]	BnaEXA	7	111	33			184	41

B. juncea T84v2	Yang et al[8] 2016	BjuTUA	8	858	121	1630	222		
B. juncea AU213	Yang et al unpublished	BjuAUA	9	370	73	661	80		
B. carinata 080798EM	Parkin et al unpublished	BcaBCA	10			1319	157	1705	231
Underpinning genome				45819	31322	59422	34940	60457	34539
Added				11739	1411	4208	528	15820	1599
Total				57558	32733	63630	35468	76277	36138
Tp hits = BLAST hits of *Brassica* CDS gene models to *Thellungiella parvula* CDS gene models									

Methods

Reorganising genome resources

GOGGs analysis was performed on each of the A, B, C genome resources (*B. rapa* Chiifu v3 for the A genome, *B. nigra* Ni100 for the B genome and *B. oleracea* TO1000 for the C genome), essentially as described previously[18]. The MS Excel (version 2016) spreadsheets representing the GOGGs were displayed with gene models ordered by their coordinate in the genome sequence resource, at 1-pixel row height and screen shots compiled into MS PowerPoint slides for visualization. These were scanned manually for discontinuities in the orderly transition within each line between alleles originating from each of the two parents. When such anomalies (which resemble what would be observed in linkage mapping as apparent double recombinants at the same point in many individual mapping lines) were detected, we manually generated a standardized formatted table for correcting misplaced segments of genomes, with correct placement based on where in the genome the pattern of alleles present in each line of the mapping population best matches. Fine placement (*i.e.*, between precisely which two gene models the segment should be inserted) was determined to preserve collinearity with the *A. thaliana* genome,

based on top BLAST similarity to *A. thaliana* CDS gene models. These tables and the original genome resource files were then taken into an automated R script, Genome_Sequence_Reorganise (deposited on GitHub https://github.com/hezhesi/Genome_Sequence_Reorganise), to generate the revised genome sequence and annotation files. This improvement was undertaken as an iterative process to refine genome organisation. As an example, the final iteration of construction is illustrated in Extended Data Figure 3 to show mis-placed chromosome segments and unanchored sequence scaffolds, with the editing file for construction of the final pseudomolecules presented as Supplementary Data 15.

Gene anchoring and interpolation

R (version 3.6) scripts were used to anchor or interpolate gene models from additional genome sequence resources with the following steps. Based on the quality and completeness of the genome resources, the scripts were run for genome resources, in the order shown in Table 2. The process was:

i. Align gene models onto the pseudomolecules, align CDS gene models (separate by chromosomes) onto the previous round of globalABC CDS models on each chromosome using BLASTN[39] (Version: 2.6.0+). Only top hits and Evalue < 1e-30 were included.
ii. Sort them by their locations. Also using AT and TP models as a guidance to determine if they are anchored correctly.
iii. Add all extra CDS models and integrate a new list
iv. Delete overlapped models on the existing gene space.
v. Generate final ordered CDS model list for next round of processing.

Reference genome curing with donor species

"Curing" was performed on gene models from the *Brassica* pan-genome with DNA re-sequencing data from donor species. This method was first developed and described by Higgins et al[38]. Instead of running it on RNA-seq data for a transcriptome reference, DNA-seq data were used and

a cured reference set of gene models was created. The 150 base reads were split into three files, each containing a set of 50 base reads using the Perl script illumina_split_read.pl. Using other Perl scripts from Higgins et al[38], with default parameters, iterative mapping and comparing with consensus was performed over six cycles after which there was no significant gain in alignment efficiency. This process resulted in a cured *Raphanus* sativus L. (R) derived from the *Brassica* B genome and *B. fruticulosa* (F) derived from the *Brassica* C genome. Then the combined ACR or ABF genomes were used as a reference for read mapping.

Genome quantification with cured reference

Genome sequence reads were 150-base paired-end Illumina HiSeqX reads obtained from DNA purified from leaves. BWA sequence-alignment program[40] (Version: 0.7.17-r1188) was used for mapping genomic reads, using the appropriate 3-genome DNA gene model reference combination. SAMtools[41] (Version: 1.10) was used to index mapping results and score counts of each gene model for each sample, then R was used to calculate the normalised as reads per kb per million aligned reads (RPKM) values. In silico reads were simulated from gene models sequence file using simulator program wgsim version 1.6 (https: //github.com/lh3/wgsim) with number of read pairs being 1000000 and read length being 150.

Analysis of introgressions by genome re-sequencing

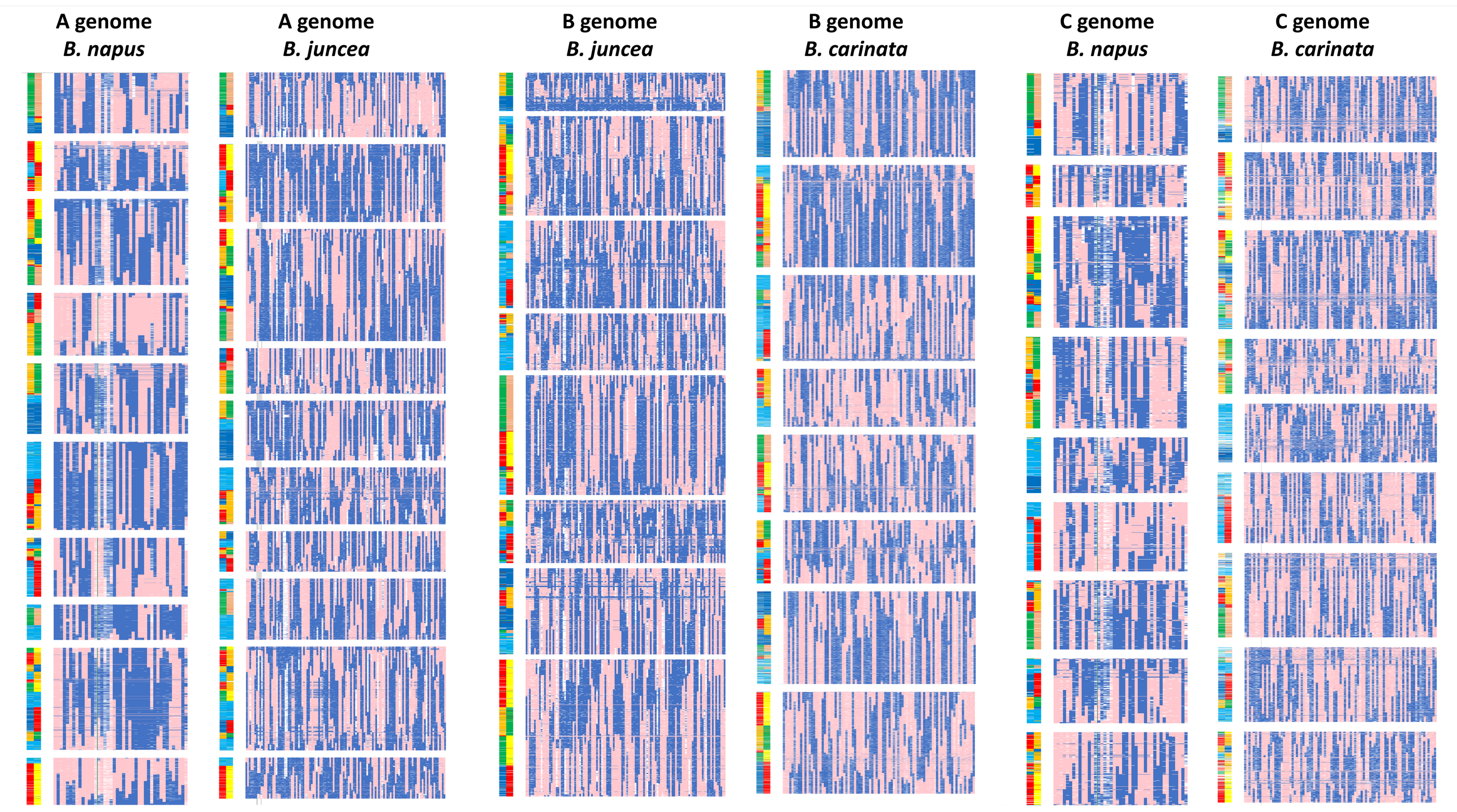
Genome representation was analysed using DNA purified from leaves, as described above. The visualization approach based on Genome Display Tile Plots (GDTPs) is essentially the same as that used for TDTPs described in He et al 2017[30], except that DNA gene models (i.e. including introns and UTRs) are used as the reference sequences and genome sequence reads are mapped. Only genes with significant signals (mean RPKM across the set of plants analysed in the experiment > 0.01) were used for further analysis. Tile plots were used to visualize genome redundancy data using quantitative representation of DNA gene models.

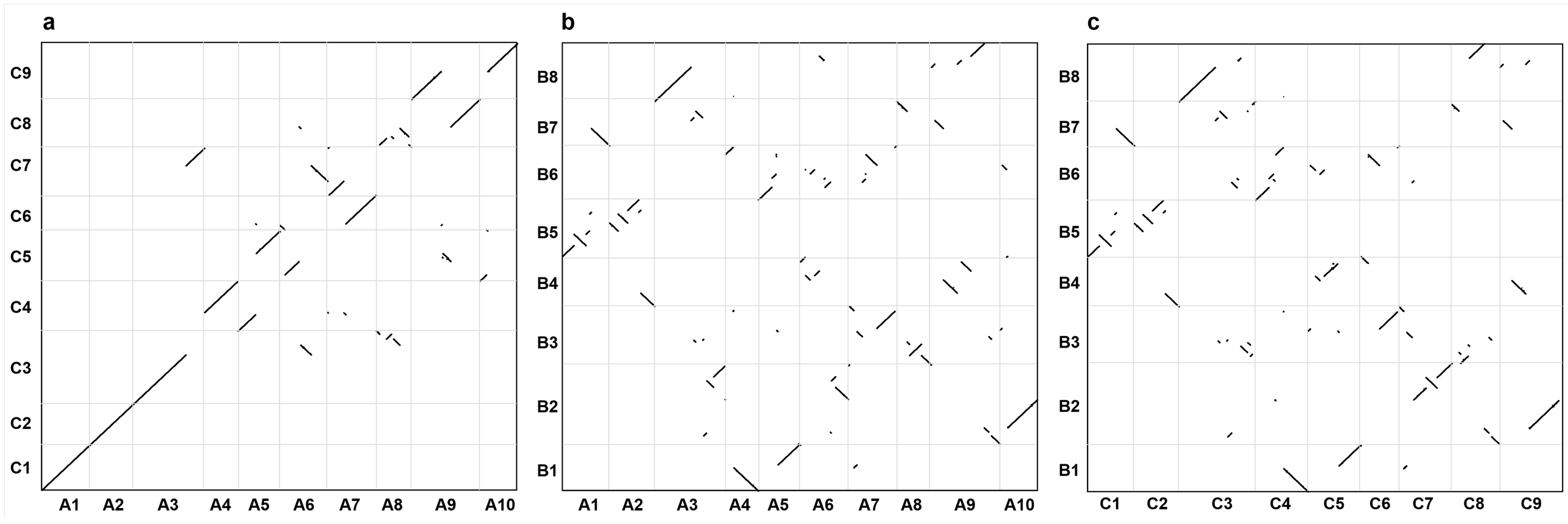
Data availability

Raw sequence reads of *R. sativus* introgression samples can be found under NCBI BioProject accession ID PRJNA507350. Raw sequence reads of *B. fruticulosa* introgression samples can be found under NCBI BioProject accession ID PRJNA673122. Raw genome re-sequencing reads for the *B. carinata* mapping population YWDH can be found under NCBI BioProject accession ID PRJNA722822. R-o-18 genome assembly information can be found under NCBI BioProject ID PRJNA649364.

Code availability

The R script Genome_Sequence_Reorganise has been deposited on GitHub (https://github.com/hezhesi/Genome_Sequence_Reorganise).


A genome
B. napus
A genome
B. juncea
B genome
B. juncea
B genome
B. carinata
C genome
B. napus
C genome
B. carinata


a
C9
C8
C7
C6
C5
C4
C3
C2
C1
A1
A2
A3
A4
A5
A6
A7
A8
A9
A10
b
B8
B7
B6
B5
B4
B3
B2
B1
A1
A2
A3
A4
A5
A6
A7
A8
A9
A10
c
B8
B7
B6
B5
B4
B3
B2
B1
C1
C2
C3
C4
C5
C6
C7
C8
C9

1

Tp1g00060	Tp1g02080
Tp1g02100	Tp1g06340
Tp1g06370	Tp1g06720
Tp1g06760	Tp1g10420
Tp1g10520	Tp1g10850
Tp1g10880	Tp1g11870
Tp1g11930	Tp1g15230
Tp1g15260	Tp1g19420
Tp1g19450	Tp1g27540
Tp1g27600	Tp1g28010
Tp1g28650	Tp1g30260
Tp1g34890	Tp1g41430
Tp1g32510	Tp1g34820

2

Tp2g00040	Tp2g06750
Tp2g06890	Tp2g07820
Tp2g07870	Tp2g20060
Tp2g20080	Tp2g23870
Tp2g23880	Tp2g24950
Tp2g24970	Tp2g29200
Tp2g29210	Tp2g30530

3

Tp3g00420	Tp3g02230
Tp3g02250	Tp3g20770
Tp3g20810	Tp3g24570
Tp3g26140	Tp3g34060

4

Tp4g00120	Tp4g08120
Tp4g08210	Tp4g09040
Tp4g09740	Tp4g11380
Tp4g11490	Tp4g11900
Tp4g11930	Tp4g16510
Tp4g16530	Tp4g22740
Tp4g22770	Tp4g28140
Tp4g28200	Tp4g30030

5

Tp5g00110	Tp5g12320
Tp5g12350	Tp5g13990
Tp5g14010	Tp5g16130
Tp5g16200	Tp5g17110
Tp5g17130	Tp5g18280
Tp5g19220	Tp5g33970
Tp5g34000	Tp5g35900

6

Tp6g00060	Tp6g03600
Tp6g03660	Tp6g07730
Tp6g07740	Tp6g09710
Tp6g10770	Tp6g35180
Tp6g35190	Tp6g35900
Tp6g35920	Tp6g40960

7

Tp7g00050	Tp7g00360
Tp7g01440	Tp7g06850
Tp7g07940	Tp7g08390
Tp7g12590	Tp7g15360
Tp7g16710	Tp7g17380
Tp7g17390	Tp7g26340
Tp7g15380	Tp7g16670
Tp7g26370	Tp7g37520

8

Tp1g00050	Tp1g13310
Tp1g13360	Tp1g17330
Tp1g17640	Tp1g27210
Tp1g27440	Tp1g31540

9

Tp2g06900	Tp2g30710
Tp6g00880	Tp6g07500

10

Tp3g00010	Tp3g22270
Tp2g02650	Tp2g06750
Tp5g11300	Tp5g16450
Tp5g10450	Tp5g11270

11

Tp4g00080	Tp4g30070
Tp7g00910	Tp7g01970
Tp7g02330	Tp7g04930

12

Tp6g26950	Tp6g41140
Tp6g14530	Tp6g26910
Tp6g11200	Tp6g14300
Tp5g19300	Tp5g35860

13

Tp5g00050	Tp5g10430
Tp5g12430	Tp5g13210
Tp5g13240	Tp5g17840
Tp5g22160	Tp5g29240
Tp5g29380	Tp5g36020
Tp5g00010	Tp5g09560
Tp7g00010	Tp7g04200
Tp7g04300	Tp7g07090
Tp7g09040	Tp7g11450
Tp7g11490	Tp7g12550
Tp7g12990	Tp7g13430
Tp5g09800	Tp5g12390
Tp7g13430	Tp7g15270
Tp7g15490	Tp7g17370
Tp7g17430	Tp7g20000
Tp7g20100	Tp7g36000
Tp7g36130	Tp7g37550

14

Tp1g01280	Tp1g06880
Tp1g07060	Tp1g18390
Tp1g19310	Tp1g22020
Tp1g22110	Tp1g26320
Tp7g12030	Tp7g12420
Tp7g34020	Tp7g37420
Tp7g21540	Tp7g22030
Tp7g33130	Tp7g33820
Tp7g22200	Tp7g26990
Tp7g27110	Tp7g31380
Tp7g31510	Tp7g33020
Tp7g18040	Tp7g21470
Tp7g17410	Tp7g17860
Tp7g15470	Tp7g15750
Tp7g15870	Tp7g17350
Tp7g12490	Tp7g15250
Tp1g26420	Tp1g27430
Tp7g11790	Tp7g11970
Tp7g05910	Tp7g07670
Tp1g27620	Tp1g32200
Tp7g09420	Tp7g11750
Tp1g32290	Tp1g32670
Tp1g32260	Tp1g36750
Tp1g36760	Tp1g41870

15

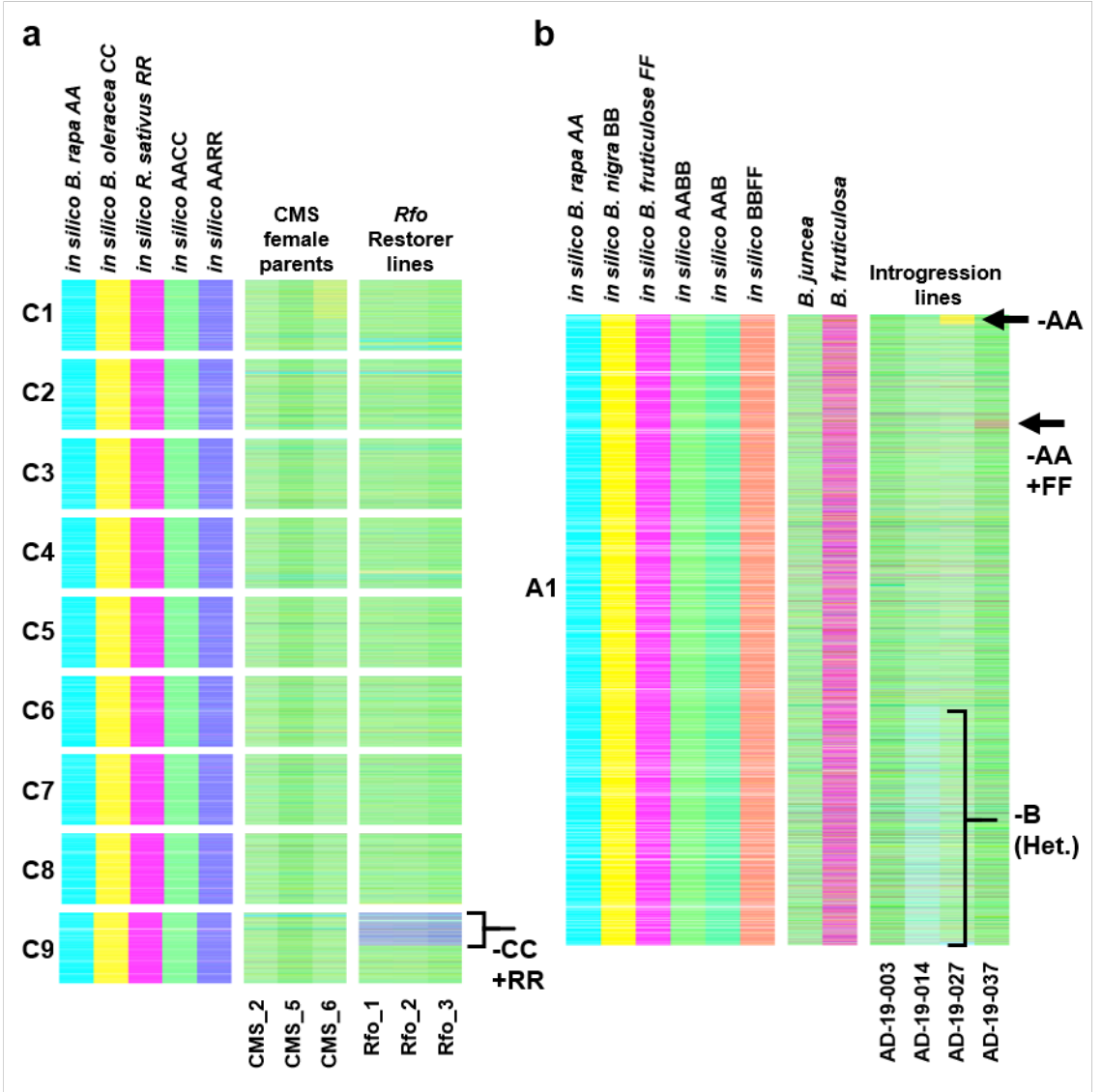
Tp3g28090	Tp3g33960
Tp2g00090	Tp2g01220
Tp2g02230	Tp2g03320
Tp2g04510	Tp2g05220
Tp2g06950	Tp2g15340
Tp2g15660	Tp2g19860
Tp2g19980	Tp2g20180
Tp2g20240	Tp2g30770

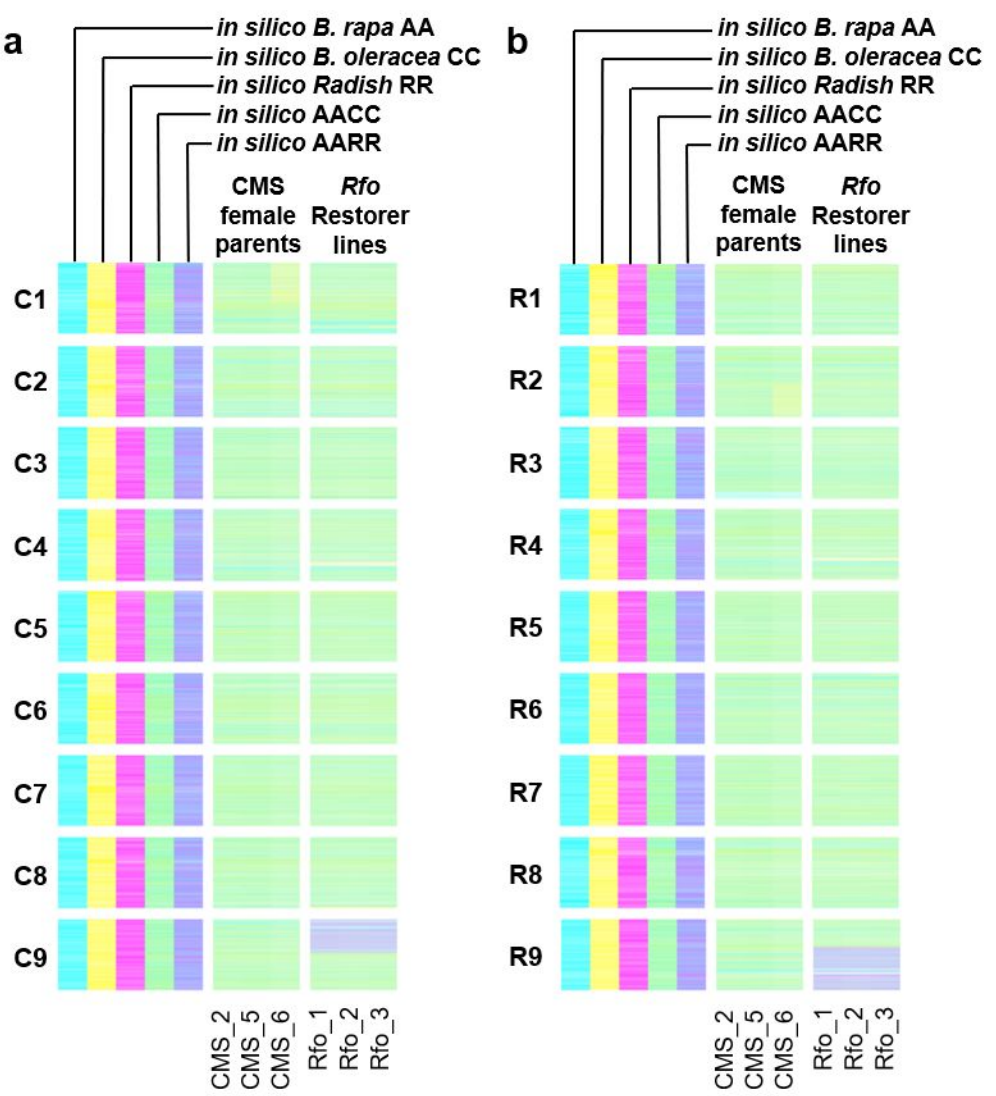
16

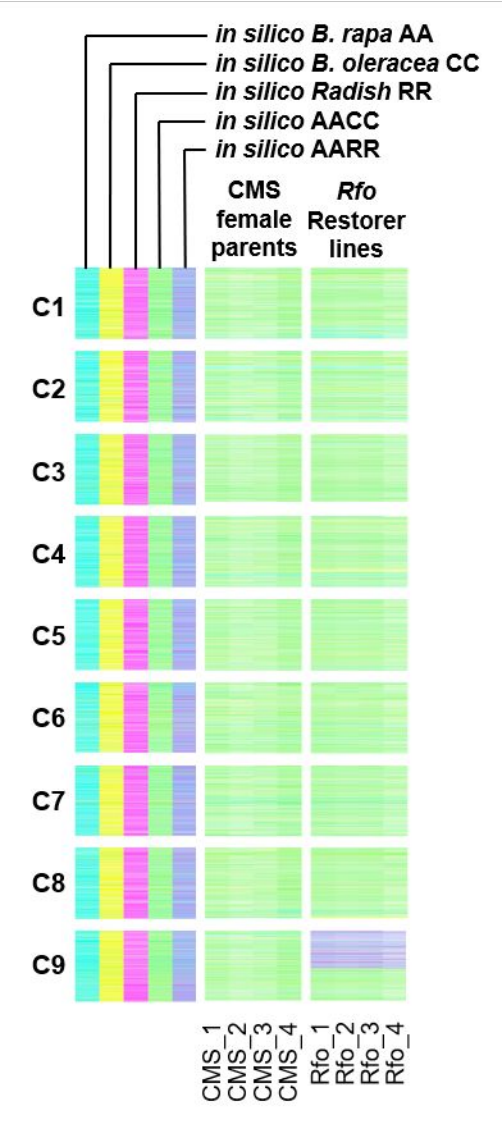
Tp3g00110	Tp3g05080
Tp3g05110	Tp3g06990
Tp3g07100	Tp3g07720
Tp3g07780	Tp3g08460
Tp3g08600	Tp3g20410
Tp3g20540	Tp3g22540
Tp3g22570	Tp3g30400
Tp3g30410	Tp3g32510
Tp3g32920	Tp3g34170

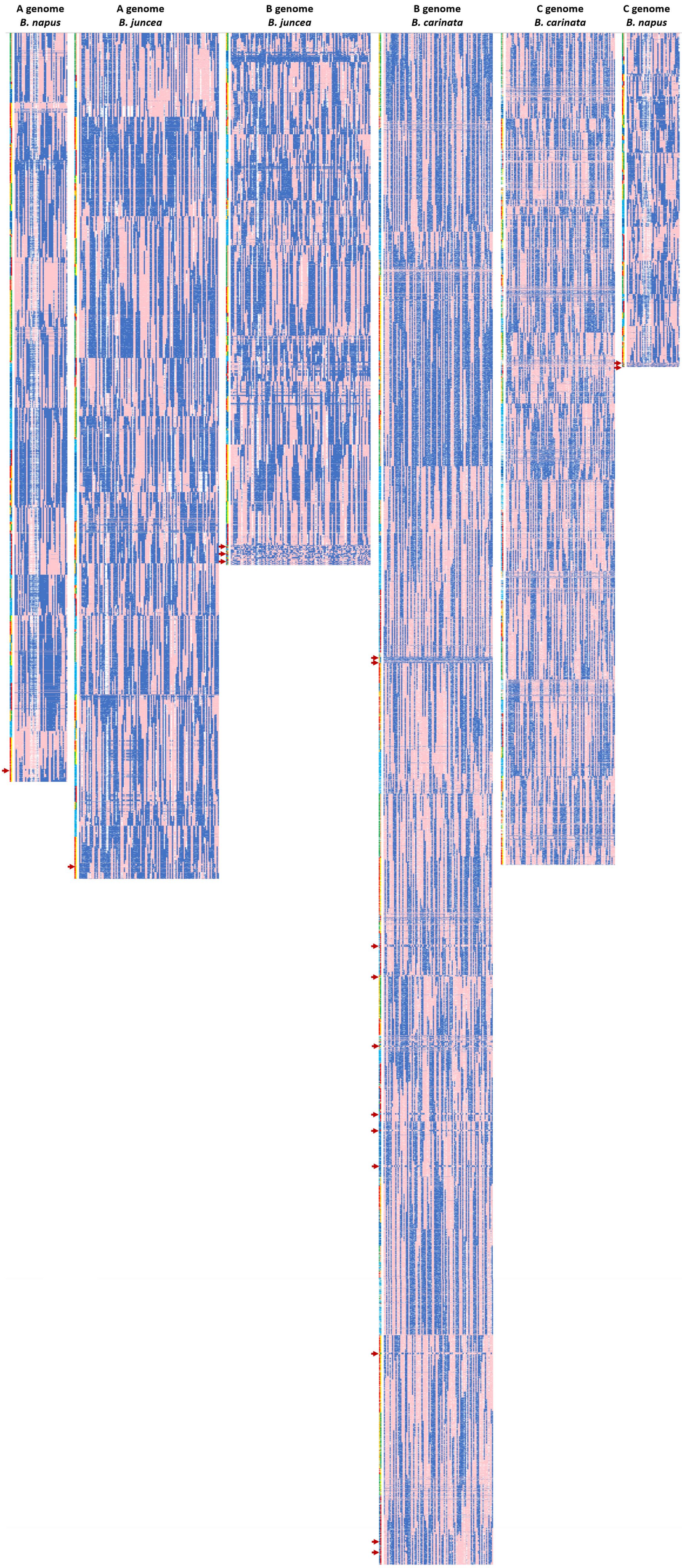
17

Tp6g40460	Tp6g41230
Tp6g12750	Tp6g40050
Tp4g12540	Tp4g29770
Tp4g00010	Tp4g12510
Tp6g10970	Tp6g12640
Tp6g00190	Tp6g10610


a
in silico B. rapa AA
in silico B. oleracea CC
in silico R. sativus RR
in silico AACC
in silico AARR
CMS female parents
Rfo Restorer lines
C1
C2
C3
C4
C5
C6
C7
C8
C9
-CC
+RR
CMS_2
CMS_5
CMS_6
Rfo_1
Rfo_2
Rfo_3
b
in silico B. rapa AA
in silico B. nigra BB
in silico B. fruticulose FF
in silico AABB
in silico AAB
in silico BBFF
B. juncea
B. fruticulosa
Introgression lines
A1
-AA
-AA
+FF
-B
(Het.)
AD-19-003
AD-19-014
AD-19-027
AD-19-037


a
in silico B. rapa AA
in silico B. oleracea CC
in silico Radish RR
in silico AACC
in silico AARR
CMS female parents
Rfo Restorer lines
C1
C2
C3
C4
C5
C6
C7
C8
C9
CMS_2
CMS_5
CMS_6
Rfo_1
Rfo_2
Rfo_3
b
in silico B. rapa AA
in silico B. oleracea CC
in silico Radish RR
in silico AACC
in silico AARR
CMS female parents
Rfo Restorer lines
R1
R2
R3
R4
R5
R6
R7
R8
R9
CMS_2
CMS_5
CMS_6
Rfo_1
Rfo_2
Rfo_3


in silico B. rapa AA
in silico B. oleracea CC
in silico Radish RR
in silico AACC
in silico AARR
CMS female parents
Rfo Restorer lines
C1
C2
C3
C4
C5
C6
C7
C8
C9
CMS_1
CMS_2
CMS_3
CMS_4
Rfo_1
Rfo_2
Rfo_3
Rfo_4


A genome
B. napus
A genome
B. juncea
B genome
B. juncea
B genome
B. carinata
C genome
B. carinata
C genome
B. napus