High-resolution molecular karyotyping uncovers pairing between ancestrally-related *Brassica* chromosomes

Annaliese S. Mason*¹,², Jacqueline Batley¹,², Philipp Emanuel Bayer¹,³, Alice Hayward¹,², Wallace A. Cowling⁵ and Matthew N. Nelson⁴,⁵

*corresponding author: annaliese.mason@uq.edu.au; Tel: +617 336 59556; Fax: + 617 3365 3556; Level 2 John Hines Building, The University of Queensland, Brisbane 4072, Australia

¹ School of Agriculture and Food Sciences, ² Centre for Integrative Legume Research and ³ Australian Centre for Plant Functional Genomics, The University of Queensland, Brisbane 4072, Australia

⁴ School of Plant Biology and ⁵ The UWA Institute of Agriculture, The University of Western Australia, 35 Stirling Highway, Crawley 6009, Australia.

*Figures and Tables: One Table, Seven Figures (all colour)*

*Total Word Count: 5714*

*Section Word Counts: Summary: 193; Introduction: 582; Materials and Methods: 1514; Results: 1499; Discussion: 1876; Acknowledgements: 50.*

Summary

- How do chromosomal regions with differing degrees of homology and homeology interact at meiosis? We provide a novel analytical method based on simple genetics principles which can help to answer this important question. This method interrogates high-throughput molecular marker data in order to infer chromosome behavior at meiosis in interspecific hybrids.

- We validated this method using high-resolution molecular marker karyotyping in two experimental *Brassica* populations derived from interspecific crosses between *B. juncea, B. napus* and *B. carinata*, using a single nucleotide polymorphism (SNP) chip.

- This analysis method successfully identified meiotic interactions between chromosomes sharing different degrees of similarity: (1) full-length homologues; (2) full-length homeologues; 3) large sections of primary homeologues; and (4) small sections of secondary homeologues.

- This analytical method can be applied to any allopolyploid species or fertile interspecific hybrid in order to detect meiotic associations. This genetic information can then be used to identify which genomic regions share functional homeology (that is, retain enough similarity to allow pairing and segregation at meiosis). When applied to interspecific hybrids for which reference genome sequences are available, the question of how differing degrees of homology and homeology affect meiotic interactions may finally be resolved.

**Keywords**: meiosis, polyploidy, recombination, molecular karyotyping, *Brassica*
**Introduction**

Successful transfer of genetic information from one generation to the next is essential for all organisms. The mechanism for this transfer in most sexually-reproducing eukaryotic species is meiosis. Chromosome behavior during meiosis must be strictly controlled: in order to ensure correct segregation of chromosomes into daughter cells, each chromosome must pair with its homologous partner. However, homologue recognition is one of the least-well-understood meiotic processes (Tiang et al., 2012). In most species, the broader process of meiosis is the same: each homologous chromosome must find its partner, associate with it and undergo the reciprocal “crossing-over” process of genetic exchange, creating one or more physical ties (chiasma) between the chromosomes to ensure correct first division disjunction (Wilson et al., 2005). During homologous chromosome pairing, homologues are roughly aligned before close-range “homology checking” and elimination of associations based on repetitive DNA sequences is then thought to occur (Bozza & Pawlowski, 2008). Although homologous chromosome pairing relies on DNA sequence homology (Bozza & Pawlowski, 2008), the exact relationship between DNA sequence similarity and homologue recognition is unknown (Tiang et al., 2012): how similar do genomic sequences have to be to initiate chromosome pairing at meiosis? In particular, how are homologues recognised, and how are meiotic interactions regulated between genomic regions with different levels of sequence homology?

This question can be addressed in allopolyploid species that are formed when two related diploid species hybridise. Polyploidy is common in many if not most plant and animal species lineages (Otto & Whitton, 2000; Leggatt & Iwama, 2003; Van de Peer et al., 2009; Jiao et al., 2011). Some angiosperm families such as the Brassicaceae have undergone multiple polyploidy events such that the present day species contain homeologous (that is, ancestrally-related) chromosomal regions with a spectrum of sequence similarities (Fig. 1). High-resolution molecular karyotyping is now becoming feasible in non-model species with the increasing availability of high-density genotyping arrays and genotyping-by-sequencing (Elshire et al., 2011; Poland et al., 2012; Cavanagh et al., 2013; Edwards et al., 2013). Therefore, the stage is set for significant advances in our understanding of just how similar DNA sequences have to be for homologue recognition to occur, and of how hybrids and polyploids regulate meiosis when several genomic regions with different levels of sequence
homology exist. However, the analytical tools for interrogating large molecular genotyping datasets to answer these questions remain under-developed.

In this study, we address this issue by developing a novel method that is capable of interrogating large molecular karyotyping data sets to rapidly assess chromosome behavior at meiosis in allopolyploid species or interspecific hybrids. This approach uses a set of simple statistical assumptions (Fig. 2) to infer chromosome pairing behavior. A hybrid between two species/genotypes is crossed to a third species/genotype to make a testcross, and then the resulting progeny are assessed for allele inheritance. Segregation of alleles from the two original parent species is inferred to result from chromosome pairing during meiosis in the hybrid. To validate this method we used the well-characterised Brassica system (Fig. 1), for which a high-density single nucleotide polymorphism (SNP) genotyping array has recently been developed (http://www.illumina.com). We show that this SNP array can be applied to detect homologous and homeologous chromosome pairing in segregating progeny derived from hybrids between the Brassica allotetraploid species (Fig. 3). This analysis is supported by a previous cytogenetic study of the same genotypes (Mason et al., 2010) and confirms the validity of our molecular karyotype analysis pipeline, providing novel, interesting data related to chromosome pairing behaviour in Brassica.

Materials and Methods

Plant material and growth conditions

Two experimental populations were developed by intercrossing homozygous lines of three allotetraploid Brassica species as described previously by Mason et al. (2012) and summarised in Fig. 3. Two to five F₁ hybrids per hybrid type (B. napus “N1” × B. juncea “J1”; B. juncea “J1” × B. napus “N1”; B. juncea “J1” × B. carinata “C1” and B. juncea “J1” × B. carinata “C2”;) were used to generate testcross progeny; generation of the F₁ hybrids is described in Mason et al. (2011). No significant differences in allele inheritance were observed between “N1J1” and “J1N1” reciprocal hybrids (B. juncea × B. napus (F₁: A¹A²B¹C²) or between the “J1C1” and “J1C2” genotypes (B. juncea × B. carinata (F₁: B⁵B¹A¹C⁵)): these were subsequently considered as two single populations for the purposes of the analyses. A total of 41 (B. juncea × B. napus (F₁: A¹A²B¹C²)) × B. carinata experimental progeny (out of 42 seeds planted, germination rate 98%) and 62 (B. juncea × B. carinata (F₁: A¹A²B¹C²)) × B. carinata experimental progeny (out of 63 seeds planted, germination rate 97%) were used for genotyping.
B^2 B^3 A(C^2) × B. napus experimental progeny (out of 73 seeds planted, germination rate 85%),
derived from meiotically reduced gametes of A^2 A^2 B^2 C^0 and B^2 B^3 A^3 C^0 hybrids (Mason et al.,
2012), comprised the final experimental populations in this study (Fig. 3).

Molecular karyotyping using the Illumina 60K Brassica SNP chip
Recently, an Illumina Infinium 60,000 SNP array was developed and released for Brassica
napus (http://www.illumina.com/). This chip comprised 52,157 SNPs distributed across the A
and C genome chromosomes. B genome inheritance could not be characterised in our
populations due to lack of available high-throughput molecular tools for this genome.
Hybridisation protocols were run according to manufacturer’s instructions for all samples in
the population plus controls (parent species and F_1 hybrids), and chips were scanned using an
Illumina HiScanSQ. Genotyping data were visualised using Genome Studio V2011.1
(Illumina, Inc. (San Diego)).

SNPs were filtered in Genome Studio to retain genome-specific SNPs that were: a)
polymorphic between the B. juncea and B. napus A genomes (A^n and A^j) and did not amplify
in B. carinata (2n = B^2 B^3 C^2 C^0), or were b) polymorphic between the B. napus and B. carinata
C genomes (C^n and C^c) and did not amplify in B. juncea (2n = A^j A^j B^j B^j), for the genotypes
used in the experimental populations. Filtering for polymorphism between the segregating
parent alleles yielded 4,883 SNPs for the A genome, and 7,215 SNPs for the C genome.
These 12,098 SNPs were then manually screened in Genome Studio: SNPs which did not
show three clear genotype clusters (AA, AB, BB) in the population were removed, leaving 3
705 A-genome and 6,489 C-genome SNPs (10,194 SNPs in total). After filtering and quality
control measures were implemented, the total number of SNP markers retained for
segregation analysis was 3,046 in the B. napus A^n genome, 4,643 in the B. napus C^n genome,
3,055 in the B. juncea A^j genome and 4,687 in the B. carinata C^c genome (Supplementary
Table 1). On average, 406 SNP markers were retained per chromosome, with a minimum of
124 (B. napus C^n5) and a maximum of 1,279 (B. carinata C^c4) (Supplementary Information).

Marker genotype calls were imported into Microsoft Excel 2010 (Microsoft Corporation)
where SNPs were sorted by chromosome, with reference to the published and available draft
reference Brassica diploid genome sequences: B. rapa (Wang et al., 2011) and B. oleracea
(http://brassicadb.org/brad/; (Cheng et al., 2011)). Further data cleaning was undertaken to
remove SNPs that: a) showed allelic segregation patterns inconsistent with the determined
genomic locations; b) appeared to detect two or more loci; c) had an excess of no-calls (NCs)
or d) were null for one of the parental alleles.

In the $A_j^A A_n^B C_n^C$ population, several blocks of markers ($A01_015 - A01_37; A05_300 - A05_305; A10_001 - A10_005; C1_041 - C1_053; C4_1278 - C4_1279; C8_0001 - C8_0004$) amplified both an $A$- and a $C$-genome allele in $B. napus$, but only an $A$-genome or a $C$-genome allele in $B. juncea$ and $B. carinata$ respectively. These markers were eliminated from the $A_j^A A_n^B C_n^C$ ($B. juncea \times B. napus$) population analyses, but retained in the $B_c^B A_j^A C_c^C$ ($B. juncea \times B. carinata$) population analyses. For the known $B. napus$ $A7 - C6$ interstitial homeologous reciprocal translocation in the parental genotype “Surpass 400” (Osborn et al., 2003), markers were named according to their physical location, such that “A7” in $B. napus$ comprised what would be part of $C6$ in $B. oleracea$ and $B. carinata$, and “C6” in $B. napus$ comprised part of what would be $A7$ in $B. juncea$ and $B. rapa$ (see Supplementary Table 1). A putative duplication in $B. napus$ covered 17% of $A_n^{10}$, from markers $A10_{284}$ to $A10_{341}$, but was not obviously linked to any other genomic region.

The SNP data set was highly saturated: many markers were redundant in terms of detecting chromosome segment inheritance and recombination events across each population. Hence, in order to facilitate the segregation analyses and crossing-over analyses, SNP marker data was reduced to one representative marker per haplotype block representing unique information per chromosome across each population (i.e. marker bins). SNPs providing redundant information across the population were removed using an R script that treated missing values (NCs) as wild cards (either 1 or 0) and retained markers with the fewest missing values (Supplementary Information). In the $B_c^B A_j^A C_c^C$ testcross population, an average of 19 marker bins per $A_j^A$-genome chromosome (range 3 to 33) and 18 marker bins per $C_c^C$-genome chromosome (range 3 to 29) were retained after elimination of redundant SNPs. In the $A_j^A A_n^B C_n^C$ testcross population, an average of 58 marker bins (range 36 to 93) were retained in the $B. juncea$ $A_j^A$ and $B. napus$ $A_n^B$ genomes and an average of six marker bins (range 2 to 18) were retained in the $B. napus$ $C_n^C$ genome after elimination of redundant SNPs.

Data analysis, data visualisation and statistics
Fisher’s exact test for count data with a significance level of $p < 0.05$ was used to test if pairs of marker alleles were segregating with one another, as would be expected from e.g. the presence of the two alleles at a single homologous locus (Fig. 2). To test whether any given two marker alleles present anywhere in the genome(s) were segregating with each other, every possible combination of two alleles was concatenated across individuals, such that each individual was scored for presence of both alleles (1/1), presence of only one of the pair of alleles (1/0 or 0/1) or absence of both alleles (0/0). The number of observations in each category was then summed and compared to expected values derived from sums for $2 \times 2$ contingency tables (Fig. 2). Expected values were derived from normal Mendelian expectations for segregation of individuals with both alleles present (1/1), one allele present (1/0 or 0/1) or neither allele present (0/0) for each pair of alleles (1:1:1:1), using contingency tables to adjust for alleles with an excess of 0 or 1 scores. Every marker allele with a known genome location was tested against every other marker allele with a known genome location. Deviation from this ratio as a result of linkage was expected to result in an excess of individuals with scores of 0/0 and 1/1 due to co-segregation of pairs of alleles (Fig. 2). Deviation from this ratio as a result of segregation of the two alleles was expected to result in an excess of individuals with scores of 0/1 or 1/0 (only one allele present of the pair) (Fig. 2). A significant deviation which resulted in an excess of individuals with a score of 0/1 or 1/0 (one allele present for each pair) was deduced to be the result of segregation of the two alleles (Fig. 2). Significant ($p < 0.05$) results of the pairwise Fisher’s exact tests for count data are presented in figures only where the sum of 0/1 and 1/0 scores was greater than the sum of 0/0 and 1/1 scores, as expected from segregation of alleles. P-values are presented uncorrected for multiple testing, and instead with a range of p-values encompassing rigorous correction cut-offs, as the assumption of independence is invalid for tests of linked alleles. However, the stringent Bonferroni corrections for multiple testing values assuming independent tests were $p < 0.00014$ for the $B^cB^iA^jC^e$ hybrid population and $p < 0.000041$ for the $A^jA^iB^cC^e$ hybrid population for the final set of alleles used for analysis of each population. In the absence of meiotic crossing over in $A^jA^iB^cC^e$ and $B^iB^jA^cC^e$ hybrids, marker alleles on each haploid chromosome in the experimental progeny were expected to be either all present or all absent. Recombination was manifested as changes in the presence or absence of marker
alleles along haploid chromosomes. Recombination events were recorded for all A- and C-genome chromosomes in each population.

R version 3.0.0 (The R Project for Statistical Computing) was used to carry out the statistical analyses, to reduce duplicate markers to single representatives per block and to generate the associated figures (Supplementary Text S1). Heatmap figures showing allele associations were generated using R software package “gplots”, and linkage disequilibrium was calculated using the “LD” function in software package “genetics”.

Results

Molecular karyotyping

A total of 3 046 SNPs polymorphic between the B. napus and B. juncea A genomes (A^a and A^j), and a total of 4 643 SNPs polymorphic between the B. napus and B. carinata C genomes (C^a and C^c), were selected for final analysis. Linkage disequilibrium values (r, p-value and D') were calculated for every pair of alleles in each of the two populations to provide additional information, and linkage disequilibrium was identified between both adjacent alleles and segregating alleles (Supplementary Tables S4, S5). Using the molecular karyotyping pipeline in R (Supplementary Text S1), every SNP allele in each of the two testcross populations (Fig. 3) was tested for significant segregation (Fig. 2) with every other SNP allele in the same population. This approach allowed discrimination between linkage disequilibrium due to segregation of alleles, and linkage disequilibrium due to co-segregation of alleles (linkage). Increasing significance corresponded to increasing numbers of individuals in the population with either one allele or the other (but not both or neither) for any two alleles.

A-genome allele segregation in AABC hybrids

In testcross population A (Fig. 3), either an A^j or an A^a allele was present at most homologous loci, as evidenced by the strong diagonal of significant segregation between A^j and A^a alleles in Fig. 4. In the A genome, which was present as homologous chromosomes from B. juncea (A^j) and B. napus (A^a), most segregating allele pairs corresponded to homologous loci in the A genome (Fig. 4). Most B. juncea A^j-genome chromosomes segregated with high fidelity with their B. napus A^a-genome homologues. An exception was part of chromosome A7,
where strong associations between $A^j$ and $A^n$ alleles were absent and this region instead showed an $A^j7 - C^n6$ association.

**Segregation in the BBAC hybrids between alleles sharing primary homeology**

In testcross population B (Fig. 3), *B. juncea* A-genome alleles ($A^j$) segregated with *B. carinata* C-genome alleles ($C^c$) with which they shared primary homeology (Fig. 5; see Fig. 1 for detailed primary homeologous relationships). These associations between homeologous alleles were of similar significance to those between two alleles at a single homologous locus ($p < 0.00000001$, Fig. 5).

Four $A^j$-$C^c$ chromosome pairs showed significant segregation of homeologous alleles along their entire length: $A^j1/C^c1$, $A^j2/C^c2$, $A^j3/C^c3$ and $A^j7/C^c6$ (Fig. 5), as indicated by the diagonal red lines across the length of the chromosome blocks. Chromosomes $A^j4$ and $C^c4$ showed segregation of alleles along the whole length of $A^j4$, but with decreasing significance ($0.00000001 < p < 0.001$) towards the end of $C^c4$ (from $\sim40 - 55$ Mbp). $A^j9$ and $C^c8$ alleles segregated along the entire length of $C^c8$, with the top 8 Mbp of $A^j9$ segregating instead with the top of $C^c9$. Clear evidence of trivalent formation was apparent between chromosomes $A^j9$, $C^c8$ and $C^c9$, with $A^j9$ chromosomes showing matching allele segregation patterns and crossovers with $C^c8$ at one end and $C^c9$ at the other end of the chromosome. Likewise, strongly significant segregation was observed between alleles on part of $A^j10$ and $C^c9$, with a switch in $C^c9$ preference to $A^j9$.

Other chromosomes showed more fragmentary allele segregation across regions of A-C homeology (Fig. 5, Supplementary Table S4). For example, $A^j5$ alleles preferentially segregated with $C^c5$ alleles, but segregation with $C^c4$ and $C^c6$ alleles was also observed. $A^j8$ alleles showed little tendency to segregate with $C^c$ alleles, but significant segregation associations were observed with the top of $C^c8$ and the bottom of $C^c9$. Similarly, alleles on chromosome $C^c7$ showed only weak segregation with the alleles in the region of primary homeology at the bottom of chromosome $A^j3$.

**Segregation between alleles sharing secondary homeology**

Evidence was also obtained for weakly significant segregation between alleles in regions of secondary homeology (derived from ancient polyploidy events) within each haploid genome
in both testcross populations, i.e. autosyndesis (Aj-Aj: within the B. juncea A genome; Cc-Cc within the B. carinata C genome and Cn-Cn: within the B. napus C genome) (Fig. 6, Supplementary Figures S4, S5).

The largest clusters of segregation between regions of secondary homeology were observed for chromosomes Aj2 and Aj10 and for chromosomes Cc2 and Cc9 (Fig. 6a, 6b). The Aj2 – Aj10 association spanned approximately 5 – 10 Mbp based on putative SNP positions, whereas the Cc2 – Cc9 association was significant over 35 – 45 Mbp. Another association was also present between Cc3 and Cc9 over a range of approximately 15 – 35 Mbp. Only two weak regions of segregation between Cn alleles were detected within the haploid B. napus C genome in the AjAaBcCn hybrid type. One of these associations was between Cn2 and Cn9 (Fig. 6c), as was observed for Cc2 and Cc9.

**Recombination events: distribution and frequency**

Chromosome segregation as predicted by the molecular karyotyping analysis using R (Supplementary Text S1) was independently validated by manual inspection of chromosome segregation and recombination events: Brassica Aj and Cc allele segregation patterns between known primary homeologues were assessed (Table 1). In general, predicted associations from molecular karyotyping could be validated by manual assessment of allele segregation and evidence of crossover formation for each chromosome pair (Fig. 5, Table 1). The location of recombination breakpoints along individual chromosomes was assessed for chromosomes that were neither completely absent nor completely present (i.e. recombinant chromosomes). The distribution of breakpoints along each chromosome is shown in Supplementary Figures S1 to S5.

**Aj-An recombination frequencies**

Homologous Aj and An chromosomes were highly recombined in the AjAaBcCn testcross population, averaging 2.4 crossover breakpoints per chromatid per individual (Fig. 7a). As many as eight breakpoints per chromatid were observed for A3 and A9. A-genome chromatids were transmitted intact without recombination only 8% of the time on average (range 0-15%), except for An7 (unrecombined 22% of the time) which appeared to have the structural An7-Cn6 translocation variant common to many B. napus genotypes. The region on An7 corresponding to the translocation had no recombination breakpoints within it, but many
recombination breakpoints flanking this region (Supplementary Figure S2). Otherwise, recombinant breakpoints were distributed along the lengths of the chromosomes, with a general tendency towards increased frequency approaching distal regions (Supplementary Figure S1, Supplementary Figure S2).

\[ C^n \] recombination frequencies

Recombination involving secondary homeologues in the haploid \( B. napus \) C genome in the \( A^1A^nB^1C^n \) testcross population occurred at much lower frequencies, averaging 0.26 \( C^n \) chromosome breakpoints per plant (Fig. 7b). \( C^n \) chromosomes were either wholly present or wholly absent (unrecombined, either lost or transmitted intact as univalents) 59 – 95% of the time, except for 46% of \( C^n6 \) chromatids which had undergone recombination with chromosome \( A^17 \) (again as a result of the \( A^n7-C^n6 \) translocation variant). \( C^n6 \) had a proximal cluster of recombination breakpoints. However, few recombination breakpoints were observed overall in \( C^n \) chromosomes, and these tended to be distally located (Supplementary Figure S3).

\[ A^1-C^c \] recombination frequencies

Recombination was frequent between primary homeologues in the \( A^1 \) and \( C^c \) genomes in the \( B^1B^1A^1C^c \) hybrid progeny (Fig. 7c) with an average of 0.46 chromosome breakpoints observed per chromosome per plant. On average, 58% of \( A^1 \) and \( C^c \) chromosomes were transmitted without detectable recombination events. This was not significantly different from the 50% expected to result from a single crossover event per homologous chromosome pair per meiosis, whereby two out of four chromatids would be expected to be transmitted without evidence of recombination (\( p=0.09 \), Student’s paired t-test against the mean). However, some chromosomes were far less likely to recombine than others (Fig. 7c): chromosomes \( C^c7 \) and \( A^18 \) each had only two detectable recombination events across the whole population. Recombination breakpoints were distributed distally on every chromosome, except in chromosomes \( A^17 \) and \( C^c6 \) which had proximal clusters of breakpoints (Supplementary Figures S4, S5).

Detecting non-random co-inheritance of alleles across genomes

In addition to above segregation and recombination analyses, we also sought evidence of positive associations between alleles that would indicate significant co-inheritance of
chromosomes. In addition to the expected co-inheritance of alleles physically located on the same chromosomes, we found some parts of the genome were significantly associated in this way (Supplementary Table 2, Supplementary Table 3). This generally occurred when two chromosomes both associated with a third chromosome (Supplementary Table 2, Supplementary Table 3). Examples included Cc2 and Cc3 (both paired with Cc9), Cc8 and Cc9 (both paired with Aj9), Cc7 and Cc3 (both paired with Aj3) and Aj2 and Cc9 (both paired with Cc2).

Transmission frequencies of marker alleles from AjAnBjCn and BcBjAjCc hybrids to progeny

In the B. juncea × B. napus (AjAnBjCn) hybrids, most Aj, An and Cn genome alleles were transmitted to the testcross population according to normal Mendelian expectations (50% frequency; p > 0.05, Pearson’s χ² test), (Supplementary Fig. S6). Some bias towards retention of Aj or An alleles was observed over short regions (Supplementary Fig. S6). Alleles from the B. napus chromosome Cn8 and half of chromosome Cn4 were retained more often than expected by chance, and no chromosome was lost more often than the expected 50% frequency (Supplementary Fig. S7). In the B. juncea × B. carinata (BcBjAjCc) testcross population, B. juncea Aj-genome chromosomes were retained more often than B. carinataCc genome chromosomes for every chromosome except Aj7 and Cc6 (Supplementary Fig. S8).

Discussion

In this study, we developed a novel open source analytical pipeline suitable for inferring chromosome segregation in large data sets (see Supplementary Text S1). This molecular karyotyping approach uses a set of simple genetic principles (Fig. 2) to infer chromosome pairing behavior. A hybrid between two species or two genotypes is crossed to another species/genotype to make a testcross (Fig. 3), and then the resulting progeny are assessed for allele inheritance using high-throughput SNP molecular markers. Segregation of alleles from the original two species can then be inferred to result from chromosome pairing during meiosis in the hybrid without prior knowledge of chromosome relationships. This method can provide new insight into homeologous and homologous relationships in species complexes by determining homologous and homeologous relationships between chromosomes and
estimating stability of chromosome inheritance. By assessing functional chromosome interactions it would also allow empirically-based predictions of transfer of alleles between genomes in interspecific hybrids, a matter of importance for estimating the risk of transgene escape from transgenic crop species to their wild relatives (Chèvre et al., 1997). In future, and with the increasingly availability of genotyping-by-sequencing approaches, this analytical method may provide us with the answer to the question of how similar DNA sequence has to be to pass homology checks and initiate chromosome pairing during meiosis.

Successful detection of homologous, homeologous and autosyndetic chromosome segregation

High-resolution molecular karyotyping of two novel interspecific Brassica populations was carried out to validate this approach. A spectrum of genome interactions was revealed, ranging from almost completely regular pairing and recombination of homologues (A-genome chromosomes from B. napus and B. juncea in the A1A8B7C8 hybrids; Fig. 4), to mostly regular pairing and reduced recombination frequencies between primary homeologues (A- and C-genome chromosomes in the B1B7A1C8 hybrids; Fig. 5), down to low-level autosyndetic chromosomal interactions between secondary homeologues (within the haploid B. napus Cn genome, Fig. 6c). In other words, this molecular karyotyping analysis revealed exactly which regions of both ancient (secondary) and recent (primary) homeology in the Brassica A- and C-genomes have retained enough sequence similarity to form pairing associations at meiosis (Fig. 4, 5 and 6), and this was validated by manual inspection of allele segregation patterns (Table 1, Fig. 7).

Molecular karyotyping detects segregation between alleles at homologous loci

Segregation of homologous alleles belonging to the B. juncea and B. napus A1 and An genomes (Fig. 4) supports the status of these genomes as predominantly un-rearranged relative to each other (Parkin et al., 1995; Axelsson et al., 2000). However, one major discrepancy was observed: an A8n7-C8n6 translocation in the B. napus line used in this experiment (Osborn et al., 2003) resulted in associations between A87 and part of C86 (Fig. 4, Supplementary Table S3, Supplementary Figures S1 and S3) rather than between A87 – A8n over the region of the translocation. Several genomic regions showed amplification of two B. napus alleles at each locus rather than one B. napus allele (A8n1, C8n1, A8n5, C8n4, A8n10 and C8n8), but amplified only one allele per locus in B. juncea and B. carinata. This phenomenon may have been due to non-specificity of SNP primer binding and/or divergence between the
genomic sequences of *B. napus*, *B. juncea* and *B. carinata*. Therefore, these regions were removed from the analysis except the largest, which comprised 30% of chromosome A<sup>n</sup>10 (5 Mb). However, it is more probable that this observation is due to the presence of duplication/deletion events between the A<sup>n</sup> and C<sup>n</sup> genomes in the *B. napus* parent, as similar chromosome rearrangements have been observed in other cultivars of this young allopolyploid species (Udall *et al.*, 2005).

*Molecular karyotyping predicts segregation between regions of primary homeology*

In the B<sup>c</sup>B<sup>j</sup>A<sup>j</sup>C<sup>c</sup> hybrids, A<sup>j</sup>-C<sup>c</sup> allosyndetic (between-genome) recombination between primary homeologous regions approached the high levels observed between homologues in natural *Brassica* species. Nicolas *et al.* (2007) observed that only half of the chromosome rearrangements transmitted from *B. napus* haploids resulted from recombination between regions of primary homeology, with the rest resulting from recombination between other regions of the genome. By contrast, in our B<sup>c</sup>B<sup>j</sup>A<sup>j</sup>C<sup>c</sup> population the vast majority of crossover events appeared to derive from regions of primary homeology: for chromosome pairs A<sup>j</sup>1 – C<sup>c</sup>1, A<sup>j</sup>2 – C<sup>c</sup>2 and A<sup>j</sup>3 – C<sup>c</sup>3 we predicted 94 – 100% pairing between primary homeologues (Table 1), and in general only a few convincing associations between regions of secondary homeology were detected (Fig. 6). The contrasting results of these two studies are surprising, but may be due to differences in genome structure or genetic factors between the two population types (Leflon *et al.*, 2010; Suay *et al.*, 2013), or to different selection pressures in the generation of the experimental populations.

Our study suggests that the most frequently formed A<sup>j</sup>-C<sup>c</sup> bivalents are likely to be collinear pairs A<sup>j</sup>1-C<sup>c</sup>1, A<sup>j</sup>2-C<sup>c</sup>2, A<sup>j</sup>3-C<sup>c</sup>3 (Fig. 5), supporting previous analyses of primary A-C homeology (Parkin *et al.*, 1995). However, the chromosome pair A<sup>j</sup>7-C<sup>c</sup>6, which lacks whole-chromosome collinearity (Parkin *et al.*, 2003), was also very strongly associated across their entire length (Fig. 5). This is surprising, as the *B. juncea* A genome and the *B. carinata* C genome lack the interstitial A<sup>n</sup>7 – C<sup>n</sup>6 reciprocal translocation found in many genotypes of *B. napus* (Osborn *et al.*, 2003), including the genotype used in this study to generate the A<sup>j</sup>A<sup>n</sup>B<sup>c</sup>/C<sup>n</sup> hybrid population. In fact, several A<sup>j</sup>-C<sup>c</sup> chromosome regions that would be predicted to show segregation based on primary homeology (Fig. 1) showed little to no association (Fig. 5, Table 1). For instance, A<sup>j</sup>7 and C<sup>c</sup>6 also share primary homeology with A<sup>j</sup>6 and C<sup>c</sup>7 (Fig. 1), but no A<sup>j</sup>6 - C<sup>c</sup>6 or A<sup>j</sup>7 - C<sup>c</sup>7 segregation was observed (Fig. 5). Hence,
factors other than primary sequence homeology must be influencing the \( A^7 - C^6 \) association, and the lack of association between other known homeologues. Concentration of recombination breakpoints around the 8-10 Mbp region of \( A^7 \) and the 15-25 Mbp region of \( C^6 \) in this population (Supplementary Figure S4, S5) suggests an unusual mode of bivalent formation between these two chromosomes: other \( A^1-C^c \) chromosome pairs appeared to form mostly distal associations (Supplementary Figure S4, S5).

Genetic factors present on single chromosomes that affect genome-wide recombination frequency have recently been found in *Brassica* (Suay et al., 2013). However, genetic factors influencing recombination between particular chromosomes have yet to be identified in the *Brassica* genus, although chromosome-specific genetic factors have been identified in other species (Jackson et al., 2002). Hence, certain chromosomes or chromosome pairs (e.g. \( A^7 - C^6 \)) may harbour, or be the target of, specific genetic factors encouraging crossover formation involving that chromosome. The “choice” of pairing partner may also be influenced by a combination of availability of other suitable partners (Nicolas et al., 2008), genetic factors and the location of the homeologous sequence in relation to the centromeres as well as degree of sequence similarity (Nicolas et al., 2012).

Chromosome associations detected for \( A^8, C^8, A^9, C^9 \) and \( A^{10} \) were strongly indicative of multivalent formation or partner swapping (Fig. 5), consistent with the status of these linkage groups as the most rearranged between the A and C genomes (Parkin et al., 2003). At least one multivalent association per cell involving A- and C-genome chromosomes only was observed by cytogenetic means by Mason et al. (2010) for this \( B^aB^bA^jC^c \) hybrid type: 0.5 \( A^j-A^j-C^c \) (0-2 per cell), 0.3 \( C^c-C^c-A^j \) (0-2 per cell) and 0.1 \( A^1-A^1-C^c-C^c \) (0-1 per cell). These multivalents are most likely attributable to \( C^8, A^9, C^9 \) and \( A^{10} \) based on these current findings.

**Molecular karyotyping predicts segregation between regions of secondary homeology**

Significant autosyndesis (within-genome pairing) in both the A and C genomes was predicted on the basis of segregation between alleles (Fig. 6). These associations are likely due to shared ancestral karyotype blocks resulting from the ancient genomic triplication before the divergence of the *Brassica* A and C genomes (Parkin et al., 2005; Schranz et al., 2006; Cheng et al., 2013). The three strongest associations observed were between \( A^j2 \) and \( A^{10} \), between...
C2 and C9 and between C3 and C9, all in the B\textsuperscript{c}A\textsuperscript{j}C\textsuperscript{c} hybrid population. These chromosomes all share two large conserved ancestral karyotype blocks in the same orientation and terminal to the end of the chromosome (Schranz et al., 2006; Cheng et al., 2013). An association between C\textsuperscript{a}2 and C\textsuperscript{a}9 was also observed in the A\textsuperscript{j}A\textsuperscript{n}B\textsuperscript{j}C\textsuperscript{n} hybrid population. These autosyndetic associations suggest that associations may form between regions of ancient homeology even in the presence of recent homeologues.

Recombination frequency and allelic selection pressure

Recombination between the B. juncea and B. napus A genomes was greatly enhanced relative to that observed in natural and resynthesised B. napus (Nicolas et al., 2008). This is probably due to the presence of additional univalent chromosomes during meiosis in this hybrid type (Mason et al., 2010) as observed previously in other Brassica hybrids (Leflon et al., 2010). Hence, this hybrid type may offer a useful bridge in Brassica for breeders aiming to break up regions of undesirable linkage disequilibrium, following a trend of interspecific hybridisation for crop improvement in this genus (Rygulla et al., 2007; Navabi et al., 2010; Chen et al., 2011; Zou et al., 2011). In the B\textsuperscript{c}B\textsuperscript{j}A\textsuperscript{j}C\textsuperscript{c} hybrid-derived progeny, A-genome loci were retained in preference to C-genome loci (Supplementary Fig. S8). As B. juncea was the maternal parent in the initial crosses, this may indicate a cytoplasmic effect on preferential genome inheritance, as suggested previously in crosses between natural and synthetic B. napus (Szadkowski et al., 2010). Several genomic regions also appeared to be under retentive selection pressure in both hybrid populations (Supplementary Figures S6, S7, S8), and may harbor alleles related to success in interspecific hybridisation or other processes conferring a viability advantage.

Selection for particular chromosome complements through viability advantage

Viability advantage for particular karyotypes could affect the success of the molecular karyotyping approach. For instance, Xiong et al (2011) found strong “dosage compensation” between homeologues, such that presence of four copies of primary homeologue pair A1/C1 (e.g. A1/A1/C1/C1, A1/A1/A1/A1, A1/C1/C1/C1 etc.) was selected for in advanced generations of synthetic B. napus. Similar dosage compensation of homeologous chromosome sets has been observed in neo-allopolyploid Tragopogon miscellus (Chester et al., 2012; Chester et al., 2013). Hence, a similar effect could occur in our B\textsuperscript{c}B\textsuperscript{j}A\textsuperscript{j}C\textsuperscript{c} hybrids to eliminate gametes which have not inherited a copy of either chromosome A1 or C1 for
example, to give the false appearance of chromosome segregation and hence pairing at meiosis. However, recombination events between primary homeologues were manually validated (Table 1), and putative crossover frequency was not significantly different to predictions for one crossover event per chromosome per gamete. Therefore, we predict that the majority of segregation associations identified through the molecular karyotyping analysis are the result of actual chromosome pairing and segregation. In systems with unknown chromosome homeology, even dosage compensation effects will provide information about chromosome similarity, as only primary homeologues are predicted to provide “dosage compensation” (Xiong et al., 2011).

Conclusions

In future, as reference genome sequences for polyploid species (including Brassica crop species) become available, we will have the opportunity to identify at base-pair resolution the chromosomal regions initiating homeologous pairing at meiosis. We provide a simple analytical pipeline that can, once whole genome sequences become available, be interrogated by whole-chromosome comparative sequence analysis to deduce exactly how similar genomic regions need to be to pair and segregate at meiosis. Similar approaches using our analysis pipeline may also be taken in highly complex polyploid species, helping us understand how chromosome pairing is controlled in complex polyploids with homeology from ancient and common ancestors.

Acknowledgements

This work was made possible by prior support by the Australian Research Council Linkage Project LP0667805, with industry partners Council of Grain Grower Organisations Ltd and Norddeutsche Pflanzenzucht Hans-Georg Lembke KG. ASM was supported and additional work funded by an Australian Research Council Discovery Early Career Researcher Award (DE120100668).

References


Figure Legends

Figure 1: Rearrangements between the *Brassica* A and C genomes (relative to the A genome). Regions of primary homeology between the A and C genomes are represented by different colours. Secondary homeology within the A genome is indicated by letters in light, medium and dark grey font representing the order of ancestral karyotype blocks resulting from the ancestral Brassiceae genome triplication (most fractionated subgenome 1 = light grey, most fractionated subgenome 2 = medium grey and least fractionated subgenome = dark grey, see Cheng et al. (2013)). Approximate centromere positions in the C genome are based on cytogenetic information (Xiong & Pires, 2011). Data for this figure were synthesised from Parkin et al. (2003), Parkin et al. (2005), Schranz et al. (2006), Xiong and Pires (2011) and Cheng et al. (2013).

Figure 2: The relationship between allele presence and absence for two different alleles at unknown genomic locations transmitted on individual chromatids in an example population (n = 16) expected as a result of linkage, segregation or no relationship. The parent of the population is assumed to be a perfect F1 heterozygote, such that its parents have different alleles for every locus in the population, but no information about the relationship of any two alleles is assumed. A) Linkage between two alleles, as would be observed if allele A and B were present at loci close together on the same chromosome. The blue box indicates a recombination event that has occurred between the locus of allele A and allele B to separate parent alleles that were on a single chromatid in the F1 heterozygote parent. B) Segregation between two alleles, as would be observed for two homologous alleles at a single locus, or for alleles at two homeologous loci that were undergoing meiotic pairing. The orange and purple boxes indicate the inheritance of neither and both parental alleles for that allele pair, respectively. If both alleles were present at a single homologous locus, this would represent homologous pairing failure, such that a multivalent or univalent chromosome association resulted in transmission of both or neither allele to a resulting daughter cell. C) No relationship between two alleles, as expected for the majority of alleles randomly tested in a segregating population.

Figure 3: Schematic diagram of interspecific crossing undertaken to produce two interspecific *Brassica* experimental populations, following Mason et al. (2012); (A) (B.
juncea × B. napus) × B. carinata population and (B) (B. juncea × B. carinata) × B. napus population.

**Figure 4:** Segregation of A-genome alleles in a B. juncea × B. napus (A\textsuperscript{j}A\textsuperscript{n}B\textsuperscript{j}C\textsuperscript{n})-derived testcross population. Statistically significant interactions between homologous regions are shown (Fisher’s exact test for count data). The Bonferroni correction for multiple testing at the p < 0.05 significance level in this population is p < 4.12E-05. Only non-redundant SNP alleles are presented, arranged sequentially according to their genetic location (not drawn to scale).

**Figure 5:** Segregation of A- and C-genome alleles in a B. juncea × B. carinata (B\textsuperscript{j}B\textsuperscript{c}A\textsuperscript{j}C\textsuperscript{c})-derived testcross population. Statistically significant interactions between regions of primary homeology are shown (that is, allosyndesis; Fisher’s exact test for count data). The Bonferroni correction for multiple testing at the p < 0.05 significance level in this population is p < 0.00014. Only non-redundant SNP alleles are presented, arranged sequentially according to their genetic location (not drawn to scale).

**Figure 6:** Segregation of alleles within A) the B. juncea A genome and B) the B. carinata C genome in a B. juncea × B. carinata (B\textsuperscript{j}B\textsuperscript{c}A\textsuperscript{j}C\textsuperscript{c})-derived testcross population, and C) the B. napus genome in a B. juncea × B. napus (A\textsuperscript{j}A\textsuperscript{n}B\textsuperscript{j}C\textsuperscript{n})-derived testcross population, showing statistically significant interactions between regions of secondary homeology (that is, autosyndesis; Fisher’s exact test for count data). The Bonferroni correction for multiple testing at the p < 0.05 significance level is p < 0.00014 for A) and B) and p < 4.12E-05 for C). Only non-redundant SNP alleles are presented, arranged sequentially according to their genetic location (not drawn to scale). Circles indicate strong associations putatively based on shared ancestral karyotype blocks R and W.

**Figure 7:** Average number of chromosome breakpoints observed per A- and C-genome chromosome in testcross individuals derived from A) and B) Brassica juncea × B. napus (2n = A\textsuperscript{j}A\textsuperscript{n}B\textsuperscript{j}C\textsuperscript{n}) hybrids and C) Brassica juncea × B. carinata (2n = B\textsuperscript{j}B\textsuperscript{c}A\textsuperscript{j}C\textsuperscript{c}) hybrids.
Supporting Information

Supplementary Table S1: SNP markers used to characterise the (B. juncea × B. napus) × B. carinata and the (B. juncea × B. carinata) × B. napus experimental populations, with information from the Brassica SNP consortium and the two sequenced diploid Brassica genomes.

Supplementary Table S2: Segregation of all alleles within B. juncea × B. carinata (B'B'C) hybrids, showing statistical significance of segregation (red) and co-segregation (blue) interactions (Fisher’s exact test for count data).

Supplementary Table S3: Segregation of all alleles within B. juncea × B. napus (A'B'Cn) hybrids, showing statistical significance of segregation (red) and co-segregation (blue) interactions (Fisher’s exact test for count data).

Supplementary Table S4: Linkage disequilibrium between A- and C-genome markers in a B. juncea × B. carinata (B'B'C) hybrid testcross population: r correlations, p-values and D'.

Supplementary Table S5: Linkage disequilibrium between A- and C-genome markers in a B. juncea × B. napus (A'B'Cn) hybrid testcross population: r correlations, p-values and D'.

Supplementary Figure S1: Frequency of recombination breakpoint locations distributed along the length of each A^n chromosome in a population derived from a B. juncea × B. napus (A^nA'B'Cn) hybrid.

Supplementary Figure S2: Frequency of recombination breakpoint locations distributed along the length of each A^n chromosome in a population derived from a B. juncea × B. napus (A^nA'B'Cn) hybrid.

Supplementary Figure S3: Frequency of recombination breakpoint locations distributed along the length of each C^n chromosome in a population derived from a B. juncea × B. napus (A^nA'B'Cn) hybrid.
**Supplementary Figure S4:** Frequency of recombination breakpoint locations distributed along the length of each A<sup>1</sup> chromosome in a population derived from a *B. juncea × B. carinata (B<sup>j</sup>B<sup>c</sup>A<sup>j</sup>C<sup>c</sup>)* hybrid.

**Supplementary Figure S5:** Frequency of recombination breakpoint locations distributed along the length of each C<sup>c</sup> chromosome in a population derived from a *B. juncea × B. carinata (B<sup>j</sup>B<sup>c</sup>A<sup>j</sup>C<sup>c</sup>)* hybrid.

**Supplementary Figure S6:** Allelic inheritance of SNP markers in the *B. juncea* and *B. napus* A genomes in testcross progeny derived from a *B. juncea × B. napus (A<sup>j</sup>A<sup>n</sup>B<sup>j</sup>C<sup>n</sup>)* hybrid. Only non-redundant SNP alleles are presented, arranged sequentially according to their genetic location (not drawn to scale). Dotted lines indicate approximate cut-offs for significant deviations (p < 0.05) from 50% inheritance (Pearson’s χ<sup>2</sup> test).

**Supplementary Figure S7:** Allelic inheritance of SNP markers in the *B. napus* C genome in testcross progeny derived from a *B. juncea × B. napus (A<sup>j</sup>A<sup>n</sup>B<sup>j</sup>C<sup>n</sup>)* hybrid. Only non-redundant SNP alleles are presented, arranged sequentially according to their genetic location (not drawn to scale). Dotted lines indicate approximate cut-offs for significant deviations (p < 0.05) from 50% inheritance (Pearson’s χ<sup>2</sup> test).

**Supplementary Figure S8:** Allelic inheritance of SNP markers in testcross progeny derived from a *B. juncea × B. carinata (B<sup>c</sup>B<sup>j</sup>A<sup>j</sup>C<sup>c</sup>)* hybrid. Only non-redundant SNP alleles are presented, arranged sequentially according to their genetic location (not drawn to scale). Dotted lines indicate approximate cut-offs for significant (p < 0.05) deviations from 50% inheritance (Pearson’s χ<sup>2</sup> test).

**Supplementary Text S1:** R scripts used to analyse segregation of alleles from *B. juncea × B. carinata (B<sup>c</sup>B<sup>j</sup>A<sup>j</sup>C<sup>c</sup>)* and *B. juncea × B. napus (A<sup>j</sup>A<sup>n</sup>B<sup>j</sup>C<sup>n</sup>)* hybrids.
Table 1: Proportion of chromosomes segregating with primary homeologue(s) based on matching patterns of allelic inheritance (presence of one chromosome and absence of the other, with or without recombination between the two) in a population of n = 62 plants derived from *Brassica juncea* × *B. carinata* (*BjBcAjCc*) hybrids segregating for *Brassica juncea* A genome (*Aj*) and *B. carinata* C-genome (*Cc*) alleles.

<table>
<thead>
<tr>
<th></th>
<th>C^c1</th>
<th>C^c2</th>
<th>C^c3</th>
<th>C^c4</th>
<th>C^c5</th>
<th>C^c6</th>
<th>C^c7</th>
<th>C^c8</th>
<th>C^c9</th>
<th>Unrecombined chromosomes^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A^1</td>
<td>94%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>39%</td>
</tr>
<tr>
<td>A^2</td>
<td>100%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>37%</td>
</tr>
<tr>
<td>A^3</td>
<td></td>
<td>94%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>47%</td>
</tr>
<tr>
<td>A^4</td>
<td></td>
<td></td>
<td>65%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>81%</td>
</tr>
<tr>
<td>A^5</td>
<td></td>
<td></td>
<td>29%</td>
<td>53%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>63%</td>
</tr>
<tr>
<td>A^6</td>
<td></td>
<td></td>
<td></td>
<td>47%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>84%</td>
</tr>
<tr>
<td>A^7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>97%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50%</td>
</tr>
<tr>
<td>A^8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>97%</td>
</tr>
<tr>
<td>A^9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>52%</td>
</tr>
<tr>
<td>A^10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50%</td>
</tr>
</tbody>
</table>

Unrecombined chromosomes^2 37% 37% 45% 50% 50% 56% 97% 63% 68%

^1 Percentages that add up to more than 100% are strongly suggestive of multivalent formation involving that chromosome.

^2 Proportion of non-recombinant chromosomes (transmitted without recombination breakpoints indicating crossover formation), regardless of segregation with primary homeologue. Indicates maximum possible percentage univalent inheritance for each chromosome.
a): Linkage in coupling/co-segregation

Allele A 0 1 1 0 0 1 0 0 0 0 1 1 0 1 1
Allele B 0 1 1 0 0 1 0 1 0 0 0 1 1 0 1 1

b): Segregation

Allele B 0 1 1 0 0 1 0 1 0 0 0 1 1 0 1 1
Allele C 1 0 0 1 1 0 1 0 0 1 1 1 0 1 0 0

c): Unlinked (no relationship)

Allele A 0 1 1 0 0 1 0 0 0 0 1 1 0 1 1
Allele D 0 0 1 1 1 0 0 1 0 1 1 0 1 0 1 0

Fisher’s exact test for count data (2 × 2 contingency tables)

- Allele A
  - | Allele B | Allele A |
  - | 1 0 |
  - | 1 0 |
  - | 0 7 |
  - | 0 1 |
  - | p = 0.0014 |

- Allele C
  - | Allele B | Allele C |
  - | 1 0 |
  - | 1 7 |
  - | 7 1 |
  - | p = 0.0101 |

- Allele D
  - | Allele A | Allele D |
  - | 1 0 |
  - | 4 4 |
  - | 5 3 |
  - | p = 1 |
**B. juncea** × **B. napus**

- **F₁ hybrid** × **B. carinata**

  - **Homologous chromosomes**

  - **A<sup>j</sup>A<sup>n</sup>B<sup>j</sup>C<sup>n</sup> testcross population**
    - Alleles assessed: A<sup>n</sup>, A<sup>j</sup>, C<sup>n</sup>
    - A = 10 chromosomes
    - B = 8 chromosomes
    - C = 9 chromosomes

**B. juncea** × **B. carinata**

- **F₁ hybrid** × **B. napus**

  - **Homeologous chromosomes**

  - **B<sup>c</sup>B<sup>j</sup>A<sup>j</sup>C<sup>c</sup> testcross population**
    - Alleles assessed: A<sup>j</sup>, C<sup>c</sup>
    - A = 10 chromosomes
    - B = 8 chromosomes
    - C = 9 chromosomes
A genome

B. napus A genome

B. juncea A genome

p < 0.00000001
1E - 8 ≤ p < 1E - 5
1E - 5 ≤ p < 1E - 3
0.001 ≤ p < 0.05
p ≥ 0.05
Average number of chromosome breakpoints detected per plant

(a) B. napus

(b) B. juncea

(c) B. carinata
Figure S1: Frequency of recombination breakpoint locations distributed along the length of each $A^n$ chromosome in a population derived from a $B. \text{juncea} \times B. \text{napus} (A^nA\text{j}B\text{j}C^n)$ hybrid.
Figure S2: Frequency of recombination breakpoint locations distributed along the length of each \textit{A}^i chromosome in a population derived from a \textit{B. juncea} x \textit{B. napus} (\textit{A}^n\textit{A}^i\textit{B}^n\textit{C}^n) hybrid.
Figure S3: Frequency of recombination breakpoint locations distributed along the length of each $C^n$ chromosome in a population derived from a $B. \text{juncea} \times B. \text{napus}$ ($A^nA^nB^nC^n$) hybrid.
Figure S4: Frequency of recombination breakpoint locations distributed along the length of each A\textsuperscript{i} chromosome in a population derived from a B. juncea x B. carinata (B\textsuperscript{J}B\textsuperscript{C}A\textsuperscript{J}C\textsuperscript{C}) hybrid.
Figure S5: Frequency of recombination breakpoint locations distributed along the length of each Cc chromosome in a population derived from a B. juncea x B. carinata (B^BcA^Cc) hybrid.