Patterns of genome duplication within the *Brassica napus* genome

I.A.P. Parkin, A.G. Sharpe, and D.J. Lydiate

Abstract: The progenitor diploid genomes (A and C) of the amphidiploid *Brassica napus* are extensively duplicated with 73% of genomic clones detecting two or more duplicate sequences within each of the diploid genomes. This comprehensive duplication of loci is to be expected in a species that has evolved through a polyploid ancestor. The majority of the duplicate loci within each of the diploid genomes were found in distinct linkage groups as collinear blocks of linked loci, some of which had undergone a variety of rearrangements subsequent to duplication, including inversions and translocations. A number of identical rearrangements were observed in the two diploid genomes, suggesting they had occurred before the divergence of the two species. A number of linkage groups displayed an organization consistent with centric fusion and (or) fission, suggesting this mechanism may have played a role in the evolution of *Brassica* genomes. For almost every genetically mapped locus detected in the A genome a homologous locus was found in the C genome; the collinear arrangement of these homologous markers allowed the primary regions of homoeology between the two genomes to be identified. At least 16 gross chromosomal rearrangements differentiated the two diploid genomes during their divergence from a common ancestor.

Key words: genome evolution, Brassicaeae, polyploidy, homoeologous linkage groups.

Résumé : Les génomes diploïdes ancestraux (A et C) de l'amphidiploïde *Brassica napus* sont fortement dupliqués puisque 73 % des clones génomiques détectent deux séquences ou plus au sein de chacun des génomes diploïdes. Cette duplication exhaustive des locus est attendue chez une espèce qui a évolué via un ancêtre polyploïde. La plupart des locus dupliqués au sein de chacun des génomes diploïdes étaient situés sur des groupes de liaison différents ou sous forme de segments colinéaires de locus liés dont certains avaient subi divers réarrangements suite à la duplication, tels que des inversions et des translocations. Plusieurs réarrangements identiques ont été observés chez les deux génomes, ce qui suggère que ceux-ci étaient survenus avant la divergence des deux espèces. Plusieurs groupes de liaison montraient une structure suggérant des événements de fusion/fission centrique, une observation qui porte à penser que ce mécanisme aurait joué un rôle dans l'évolution des génomes au sein du genre *Brassica*. Pour presque tous les locus situés sur le génome A, un locus homologue a été trouvé sur le génome C et la colinéarité de ces marqueurs homéologues a permis d'identifier les régions primaires d'homéologie entre ces deux génomes. Au moins 16 réarrangements chromosomiques d'importance ont contribué à différencier les deux génomes diploïdes au cours de leur divergence à partir d'un ancêtre commun.

Mots clés : évolution génomique, brassicacées, polyploïdie, groupes de liaison homéologues.

[Traduit par la Rédaction]

Introduction

The extent of genetic, biochemical, and genomic information available for model plant species is growing exponentially, and has culminated in the sequencing of the genome of the model crucifer (dicot) *Arabidopsis thaliana* (*Arabidopsis* Genome Initiative 2000), and is expected to be shortly followed by the genome sequence of the model grass species (monocot) rice (Sasaki and Burr 2000). Owing to this wealth of publicly available information, comparative mapping between the genomes of crop plants and their respective models is becoming a common approach for the identification of markers and candidate genes for mapping studies and to expedite positional gene cloning. However, such comparative mapping studies are reliant upon an accurate picture of the genome organization of the chosen crop species, the majority of which are either acknowledged amphidiploids such as wheat or cotton (Gale and Devos 1998; Jiang et al. 1998) or have evolved through polyploid ancestors, examples of which include soybean and maize (Shoemaker et al. 1996; Gaut et al. 2000). This inherent level of duplication within the genomes of crop species adds an extra level of complexity when attempting to identify regions of homology across species. In defining regions of collinearity between model species and their crop relatives, it is first necessary to define

Received 15 July 2002. Accepted 10 December 2002. Published on the NRC Research Press Web site at http://genome.nrc.ca on 17 March 2003.

Corresponding Editor: G.J. Scoles.

I.A.P. Parkin[1,2], A.G. Sharpe[1], and D.J. Lydiate[1]. John Innes Centre, Norwich Research Park, Colney, Norwich, NR4 7UH.

[1]Present address: Agriculture and Agri-Food Canada, Saskatoon Research Centre, 107 Science Place, Saskatoon, SK S7N 0X2.
[2]Corresponding author (e-mail: parkini@agr.gc.ca).

 doi: 10.1139/G03-006

the extent of the genome duplication found within the genome of the crop itself.

Brassica napus is an amphidiploid formed from the fusion of two diploid genomes, a C-genome progenitor (*Brassica oleracea*) and an A-genome progenitor (*Brassica rapa*) (U 1935). It has been established that since this hybridization event, no major chromosomal rearrangements have differentiated the A and C genomes of *B. napus* from those of their modern-day relatives, *B. rapa* and *B. oleracea*, respectively (Parkin et al. 1995; Bohuon et al. 1996; Parkin and Lydiate 1997). Comparative mapping between the amphidiploid *B. napus* and the model crucifer *A. thaliana* has already established that a number of regions of the model genome are represented by three copies in each of the constituent diploid genomes of *B. napus* (Scheffler et al. 1997; Cavell et al. 1998; Parkin et al. 2002). These studies and others have led to the suggestion that each of the diploid *Brassica* species evolved from a common hexaploid ancestor (Lagercrantz 1998). If indeed each region of the *Arabidopsis* genome is represented three times in each of the diploid *Brassica* genomes, then a comprehensive understanding of the polyploid structure of the *Brassica* genomes must be taken into account in any comparative mapping study.

The Brassicaceae encompass a diverse group of species with a wide range of unique and potentially valuable traits that allow them to inhabit environments ranging from high artic tundra to deserts and tropical forests. Comparative genome analysis allows the identification of collinear regions and the definition of breakpoints in synteny between species. Through such analyses, it is possible to estimate the number and extent of chromosomal rearrangements that have occurred during the evolution of different plant lineages to generate such a range of species. The level of homology between chromosomes of different species should also be an indicator of the opportunity for homoeologous recombination to take place between any two species and hence indicate the likelihood of successful interspecies gene transfer. Such information should assist in exploiting some of the more useful phenotypes from the wild species in their domesticated relatives.

An analysis of the relative map position of duplicate marker loci in the A and C genomes of *B. napus* has uncovered the pattern of genome duplication within the progenitor A and C diploid genomes and identified the regions of primary homology or homoeology between the two genomes. The number of gross chromosomal rearrangements that have differentiated the A and C genomes during their divergence from a common ancestor was determined and contrary to previous data an equivalent number of genome rearrangements has separated each of the three *Brassica* diploid species from one another. A common set of molecular markers had previously been used to map the B genome of the diploid *Brassica nigra* and it was proposed that a larger number of rearrangements separated the B genome from either the A or C genomes compared with that separating the A from the C (Lagercrantz 1998). However, the data presented here suggest that the limited chromosome pairing and recombination reported between the B genome and its related A and C genomes is not necessarily a function of the level of gross chromosomal rearrangements.

Materials and methods

Genetic maps

The aligned genetic linkage map of *B. napus*, as described in Parkin and Lydiate (1997), was derived from the maps of the N-fo-61-9 and N-o-72-8 populations and was composed of 455 defined RFLP loci detected by 162 informative genomic probes. Parkin et al. (1995) identified the 10 linkage groups of the *B. rapa* A-genome progenitor (N1–N10) and the 9 linkage groups of the *B. olerecea* C-genome progenitor (N11–N19) of *B. napus*. As described in Bohuon et al. (1996), the nine C genome linkage groups from the N-fo-61-9 map of *B. napus* have been aligned with the nine homologous groups of a *B. oleracea* genetic linkage map, which positioned a further 169 loci on the C genome linkage groups. The probe nomenclature follows that of Sharpe et al. (1995), pC, pN, pO, pR, or pW, indicating the library of origin, followed by an identifying number, and a letter designating an individual locus. For the extra loci mapped in Bohuon et al. (1996), the probe name is suffixed by the capital letter E and a number that designates individual loci. Owing to chromosomal rearrangements in the parental lines that generated the N-fo-61-9 and N-o-72-8 populations (Parkin et al. 1995; Sharpe et al. 1995), N16 was difficult to map accurately and has not been represented in any of the figures.

Duplication of marker loci

One hundred thirty-four probes previously mapped in the N-fo-61-9 population using *Eco*RI to detect polymorphic loci were hybridized to filters with DNA from the parents of N-fo-61-9 cut with four extra restriction enzymes (*Xba*I, *Bam*HI, *Hin*dIII, and *Eco*RV) using the same conditions as in Sharpe et al. (1995). The number of loci both polymorphic and monomorphic detected by these probes was estimated by assessing and collating data from their Southern hybridization banding patterns.

Results

Level of duplication of marker loci

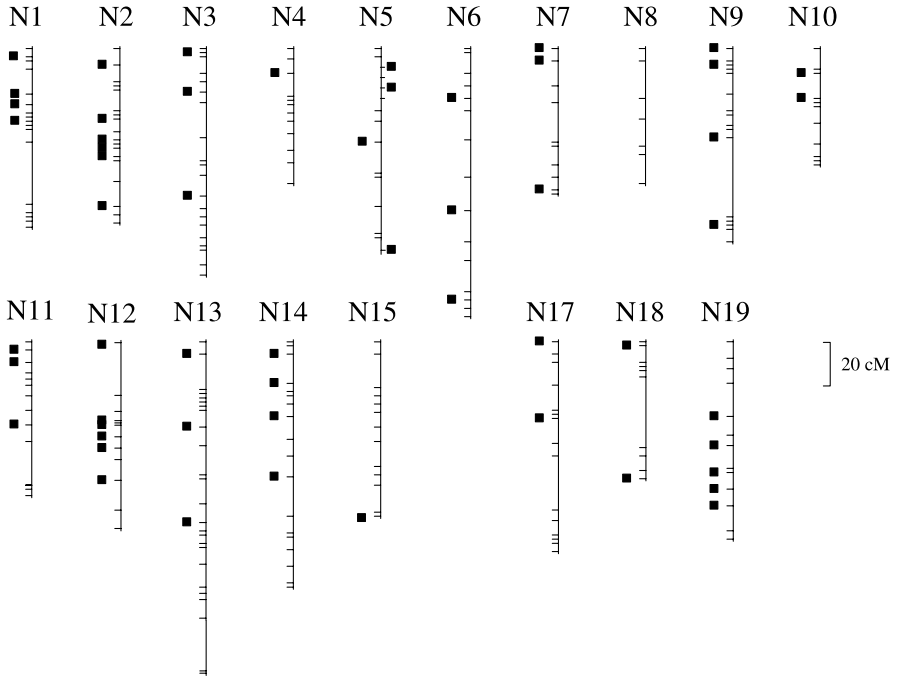
The majority of the probes (84%) detected even numbers of loci in *B. napus* and equal numbers of loci in the A and C genomes (Table 1). This suggests that loci in *B. napus* usually occur in homoeologous pairs (one originating from *B. rapa* and one from *B. oleracea*). A very high proportion of probes, 73%, detected loci with duplicate homologues within the A and (or) C genome. This means that 87% of the loci are within duplicated regions within the A and (or) C genome. Only 1 out of 134 probes detected a single locus in *B. napus*, and only 27% of the probes detected one locus in each diploid genome and could be considered unique. The aligned map of *B. napus* (Parkin and Lydiate 1997) allowed the localization of most of the loci detected by these probes and these loci were apparently distributed at random throughout the *B. napus* genome (Fig. 1).

These data suggest that the single-copy loci represent regions whose duplicate copies have either been lost from the genome or have been mutated to the extent that they are no longer recognized under these experimental conditions. The presence of single-copy loci within duplicated sections of the genome and the random distribution of such loci within

Table 1. The number of clones detecting unique, duplicated, triplicated, and quadrupled loci within the *B. napus* genome.

	Unequal numbers in A and C genome	Equal numbers in A and C genome	Total (%)
(One) or two loci	(1) –	35	36 (27)
(Three) or four loci	(7) 3	48	58 (43)
(Five) or six loci	(5) 1	25	31 (23)
(Seven) or eight loci	(4) 1	4	9 (7)
Total loci (%)	22 (16)	112 (84)	

Fig. 1. The distribution of single-copy loci (■) within the A (N1–N10) and C (N11–N19) genomes of *B. napus*. The vertical lines represent linkage groups; loci placed to the right of the lines were mapped only in the N-o-72-8 population.



the genome is consistent with numerous localized deletions or insertions. However, on some linkage groups, notably N2 and N12, clusters of single-copy loci can be seen (Fig. 1) that are consistent with larger deletions.

Distinguishing the patterns of duplication within the *B. napus* genome

The diploid origin of each of the linkage groups of *B. napus* has been determined (Parkin et al. 1995), which made it possible to distinguish between intra- and intergenomic duplication within the *B. napus* genome.

Intragenomic duplication within the A genome (B. rapa)

Table 2 lists the duplicate loci observed within the A genome of *B. napus* and hence shows the amount and distribution of intragenomic duplication within and between the *B. rapa* linkage groups. The size of the observed duplications ranged from a single locus to a maximum of 10 adjacent loci. Although entire linkage groups were not duplicated, in a number of instances, duplicated loci covered large expanses of linkage groups (Figs. 2*a* and 2*b*). For example, 82% of the mapped length of N1, the region extending from pW225c to pO12e, was collinear with the lower region of N3 (Fig. 2*a*). Sixteen regions had three or more loci duplicated within the A genome (Table 2). Six of these duplications showed conserved locus orders (N1 vs. N9, N2vs. N3, N2 vs. N10, N3 vs. N5, N3 vs. N10, and N5 vs. N6) and of these, four showed limited changes in recombination distances across the linkage groups (N1, 5 cM vs. N9, 6cM; N2, 16 cM vs. N3, 17 cM; N2, 14 cM vs. N10, 12 cM; N3, 21 cM vs. N10, 33 cM; data not shown). Four of the duplicated regions appeared to have undergone inversions (N1/N3, N2/N7, N2/N9, N3/N8) and the remainder had been subject to more complicated rearrangements. In the majority of cases, the duplicated regions appeared to be contracted or expanded in terms of genetic size between linkage groups, which appeared to be a function of the rearrangements that took place subsequent to the duplications.

The largest conserved regions were triplicated (Figs. 2*a* and 2*b*). The homologous regions between the linkage groups N2, N3, and N10 displayed conserved locus order and examples of possible localized deletion events or divergence of duplicated sequences. For example, the probe pN121, under the hybridization conditions used, detected two loci in the *B. rapa* genome and two in the *B. oleracea* genome. The two loci on the A genome were mapped to

Table 2. A summary of the homologous duplicated RFLP-defined loci detected within the A genome.

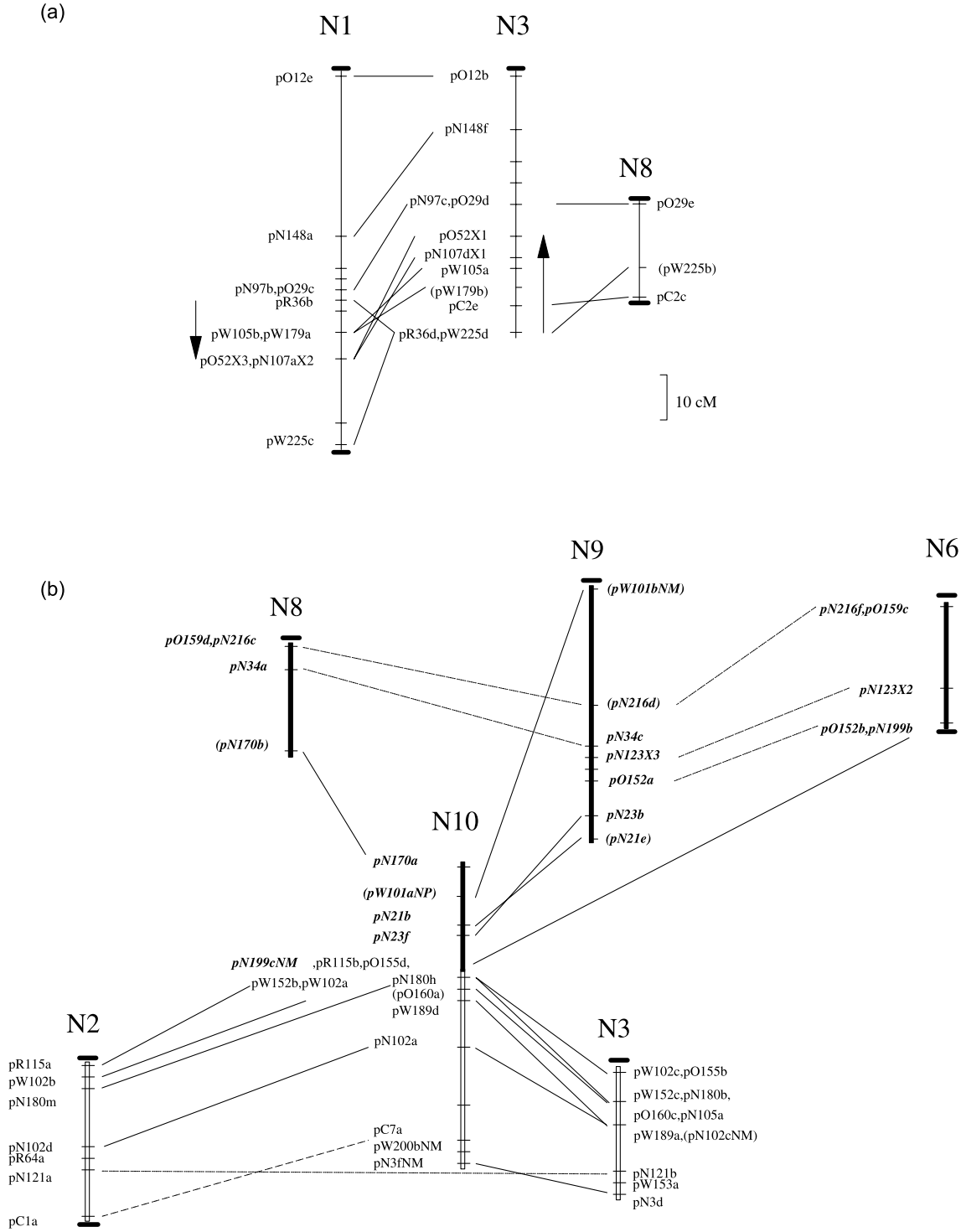
	N3	N4	N5	N6	N7	N8	N9	N10
N1	pW225c-d pO52X3-X1 pN107a-d pW105b-a pW179a-(b) pR36b-d pN97b-c pO29c-d pN148a-f pO12e-b	pN97b-g	pN148a-c pN13b-e	pW145e-c	pW108a-d	pW225c-(b) pO29c-e	pR36b-h pN173a-c pW203c-a	
N2	pN121a-b pN102d-(c) pN180m-b pW102b-c			pW102b-d pO120a-b pO119a-d	(pO86a)-(d) pO3b-d pW191e-f		(pR30a)-d pW180b-a pN181a-(b) pW191e-(d)	pN102d-a pN180m-h pW102b-e pR115a-b
N3	pW144b-c	(pN99b)-dpN97c-g	(pN99b)-(c) pN148f-c pN215b-a	pW102c-d	pO79c-b	pO29d-e pC2e-c pW225d-(b)	pN105a-f pR36d-h pO59a-e	pN3d-f (pN102c)-a pW189a-d pN180b-h pO160c pW152c-b pW102c-a pO155b-d
N4			pN99d-(c)		pW130e-d pN44c-a pN151c-a pN59f/k-g		pW205c-b pW130e-c	
N5				pW137b-(c) pO46b-a pW217a-b pW115e-b				pN53d-(b)
N6				pO9a-d pN199a-b pN101a-e		pN216f-c pO159c-d	pN52b-(c) pO152b-a pN123X2-X3 pN216f-(d)	pN199a/b-c pW102d-e
N7							pW191f-(d) pW130d-c	
N8							pN216c-(d) pN34a-c	(pN170b)-a
N9								pW101b)-a (pN21e)-b pN23b-f

Note: The loci in brackets were mapped only in the N-o-72–8 population. The columns representing linkage groups N1 and N2 and the row representing linkage group N10 have been omitted from the table, since no data was present.

linkage groups N2 and N3 within the region of conserved homology shown in Fig. 2. It is probable that the sequence detected by pN121 was originally found in three copies in the A genome and the third locus on N10 has been deleted or has diverged subsequent to duplication of the whole region. This argument also holds true for pN3, pW153, pW189, pO160, pO155, and pW152. The loci pW200b, pR64a, and pN105a were detected by clones for which all the hybridizing sequences could not be mapped; the unmapped sequences would be expected to lie within the homologous regions. pC1 and pC7 are cDNA clones from two different functional pollen oleosin genes that generate different but related protein products (J. Ross, unpublished data). The presence of loci detected by these clones within the duplicated region suggests they are paralogues that have diverged sufficiently to prevent cross hybridization.

From Table 1 it can be seen that 7% of the clones appeared to detect sequences that had been replicated four

Fig. 2. Intragenomic duplication with the A genome of *B. napus*. The vertical lines or rectangles represent sections of linkage groups, the thicker horizontal lines indicate that the linkage group continues beyond that point. Loci in parentheses were mapped only in the N-o-72-8 population. (*a*) Triplicated segment between linkage groups N1, N3, and N8. The arrows indicate an inversion of markers in one homologous region compared with the second. (*b*) All regions of A-genome intragenomic homology for N10. Markers with mapped loci in two of the duplicated copies but not in the third or fourth region are indicated by dashed lines. The vertical rectangles representing the linkage groups have been shaded to indicate two duplicated regions, one (in black) duplicated four times and the second (no shading) three times.



times; within the limits of this analysis, the maximum number of linkage groups to share parallel duplications is four (Fig. 2*b*). Not all of the duplicated loci are found in each of the four duplicated regions. This is probably the result of unmapped monomorphic loci and the deletion and (or) divergence of duplicate loci as discussed previously. Figure 2*b*

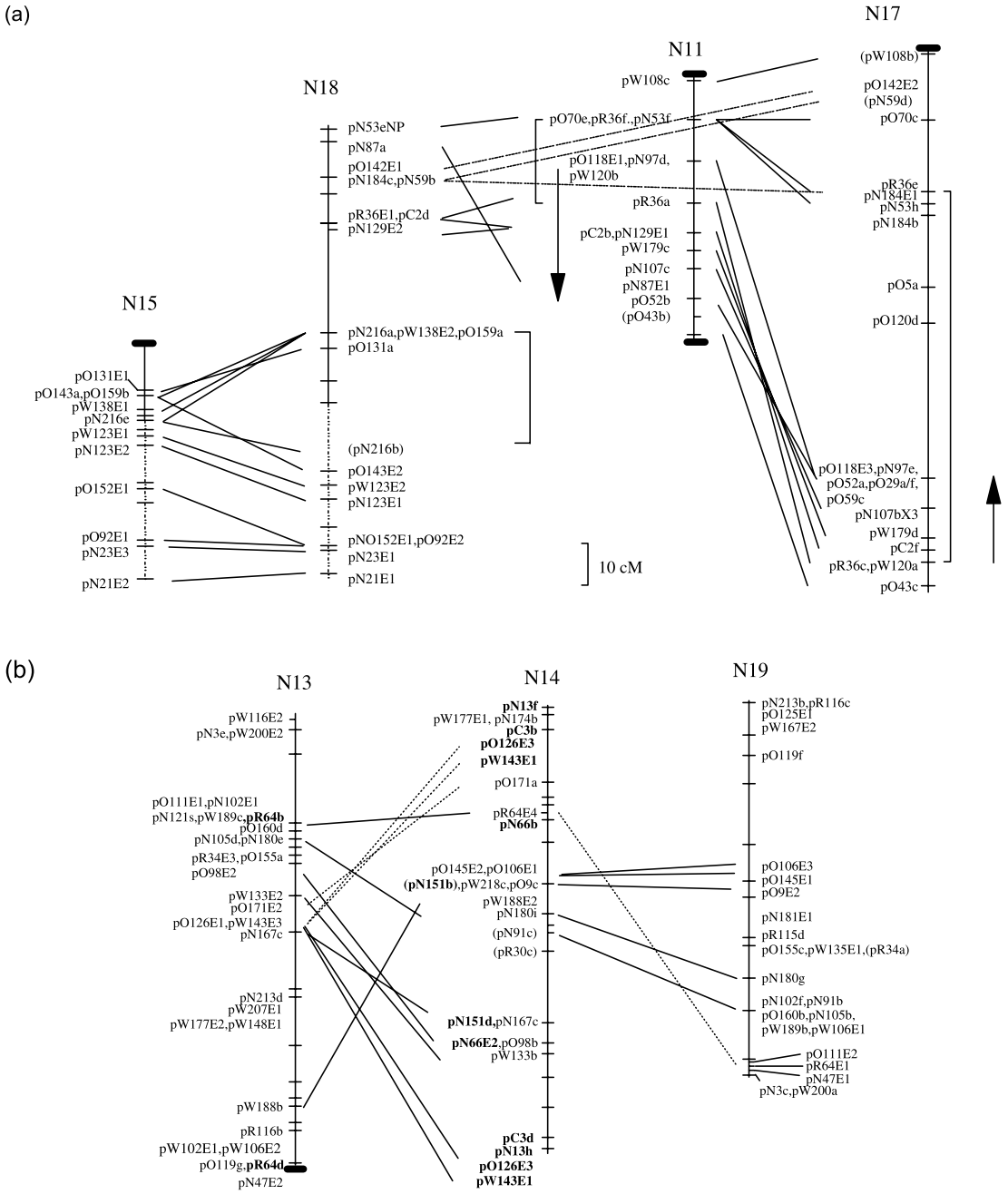
Table 3. A summary of the homologous duplicated RFLP-defined loci detected within the C genome.

	N11	N12	N13	N14	N15	N16	N17	N18	N19
N11	pR36a-f	pO17a-E1 pN121E2-q pN13c-E2	pN186c-E2 pO43b-E1 pN107c-X5 pN47E3-E2 pN53f-g pN121E2-s pO12c-a pW172b-E2	pN97d-(f) (pN173d)-b pW136c-E2 pN13b-f/h	pW145a-d pN52c-a pN47E3-a pO12c-d pW172b-E1		(pO43b)-c pO52a-b pN107c-b pW179c-d pC2b-f pR36a-c pO118E1-E3 pN97d-e pW120b-a pO70e-c pN53f-h pR36f-e pN121E2-E3 pW108c-(b)	pN87E1-a pC2b-d (pN129a)-E2 pR36a-E1 pN53f-e	pW157b-e pW239b-E1 pN52c-f pN87E1-E3 pN47E3-E4 pO70e-d pO168E1-E2
N12			pW116E1-E2 pN121q-s pN102e-E1 pN180d-e pW102b-E1 pW106E3-E2 pR34E2-E3 pN105e-d pW207b-E1 (pW148b)-E1 pW177E3-E2 pO119e-g	pN13E2-f/h pN180d-i pN174E2-b (pR30b)-(c) pW177E3-E1 pW218a-c	pO136E1-E2 pO153E2-E1	pN180d-E4	pN121q-E3 pO85a-b pO59f-c pO120c-d pW167E1-b	pR86E1-E2	pN102e-f pN180d-g pR115c-d pW106E3-E1 pR34E2-(a) pN105e-b pW135E2-E1 pN181c-E1 pO125E2-E1 pW167E1-E2 pO119e-f
N13			pR64b-d	pN180e-i pO98E2-b pW133E2-b pO171E2-a pO126E1-E2/E3 pW143E3-E1/E2 pN167c-(a) pW188b-E2 pR64d-E4	pN120E3-a pW172E2-E1 pO12a-d pN148e-d pO123b-a pN47E2-a	pN180e-E4 pO10E3-E1/E2 pR64d-E2 pW115f-(a)	pN121s-E3 pN53g-h pN20E1-E2 pO142E3-E2 pR93E2-E1 pO87E1-E2 pN107X5-b pW225E1-a pO43E1-c	pN120E3-E1 pN53g-e pW207E1-E3 pW188b-c pO142E2-E1 pO52E1-E2 pR54a-E4	pW200E2-a pN3e-c pO111E1-E2 pN102E1-f pN105d-b pO160d-b pN180e-g pW189c-b pR34E3-(a) pO155a-c pN213d-b pW106E2-E1 pO119g-f pR116b-c pR64d-E1 pN47E2-E4

N14	pN13f-h pC3b-d pO126E2-E3 pW143E2-E1 pN66b-E2 (pN151b)-d	pN91E1-E2	pR94E2-b pR64E4-E2 pO9c-b pN180i-E4 (pW130a)-(f)	(pN97f)-e pN64b-E2 pN59c-(d)	pN120a-E1 pW188E2-c (pW130a)-b pN59c-b	pR64E4-E1 pO106E1-E3 pO145E2-E1 pO9c-E2 pN180i-g (pN91c)-b
N15		pN113a-(b)	pW197d-E2 pO128E1-E2	pN101g-c pN216e-E2 pO131E1-E2 pW197d-f	pN21E2-E1 pN23E3-c pO92E1-E2 pO152E1-E2 pN123E2-E1 pW123E1-E2 pN216e-a pW138E1-E2 pO143a-E2 pO159b-a pO131E1-a pN59E2-b	pN52a-f pN101g-h pN47a-E4 pW114a-E2
N16			pO10E1-E2 pO104E2-E3	pW228E1-E2 pO104E2/E3-E1 pW197E2-f	pR54E1-E4 (pW130f)-b	pN180E4-g pR64E2-E1
N17				pR36e-c	(pN59d)-b pR36e/c-E1 pO142E2-E1 pO131E2-a pN53h-e pN184b-c pN216E2-a pW104E1-E2 pO29a-b pC2f-d	pO70c-d pN101c-h pW167b-E2
N18					pN216a-b pN23c-E1	pN87a-E3 pN173E1-E2

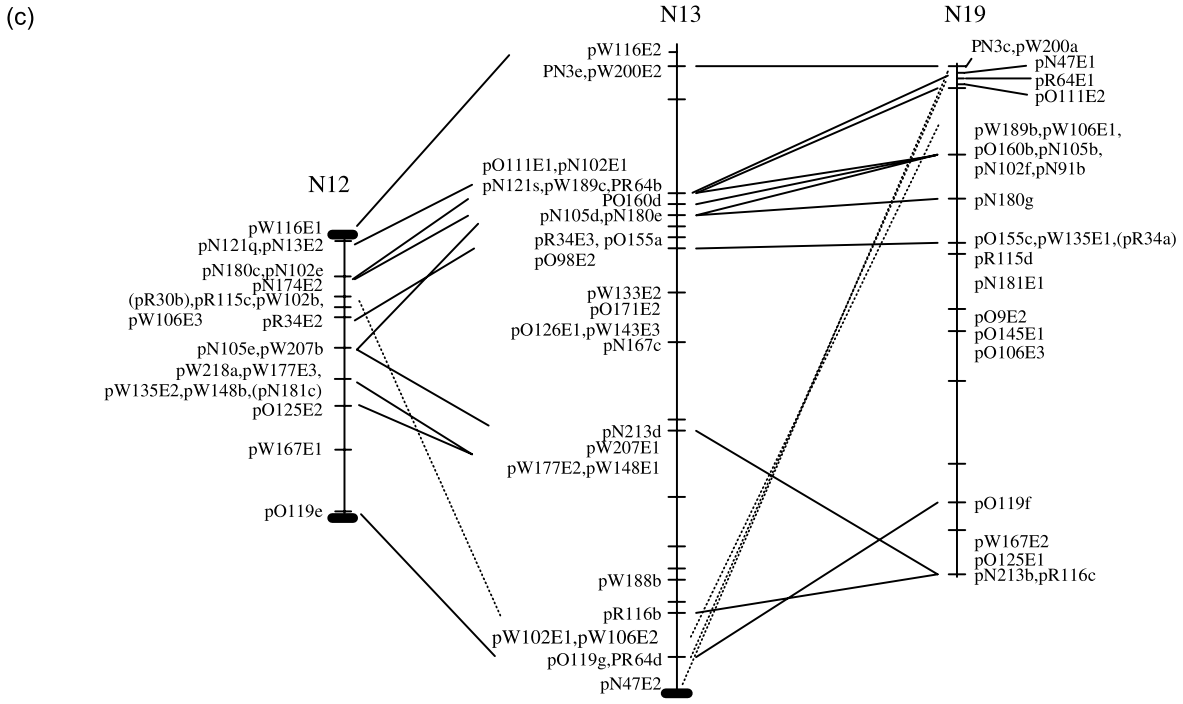
Note: The loci in brackets were mapped only in the N-o-72–8 population.

Fig. 3. Intra-genomic duplication within the C genome of *B. napus*. The vertical lines represent sections of linkage groups, the thicker horizontal lines indicate that the linkage group continues beyond that point. Loci in parentheses were mapped only in the N-o-72-8 population. Markers with mapped loci in two of the duplicated copies but not in the third or fourth region are indicated by dashed lines. (*a*) Triplicated segment between linkage groups N11, N17, and N18. The arrows indicate an inversion of markers in one homologous region compared with the second. Intra chromosomal duplications of markers, pR36 on N11 and N17, and pN216 on N18 are indicated. An extra section (dotted line) was added to N18 through alignment with the *B. oleracea* linkage map (Bohuon et al. 1996). (*b*) Homology between linkage groups N13, N14, and N19. N14 has six internally duplicated markers, which are shown in bold. Owing to the presence of this intra chromosomal duplication the collinear marker order is apparently disrupted and this is indicated by the dotted lines. (*c*) Triplicated segment between linkage groups N12, N13, and N19. pR64 is internally duplicated on N13 (shown in bold) and maybe part of a larger duplication that disrupts the collinearity, indicated by the dotted lines.



shows all the homologous regions of N10 found within the A genome. N10 appears to be composed of two conserved homologous regions, the top portion apparently replicated four times and the lower three times. In both linkage maps of *B. napus*, the junction between the two conserved regions is notable for lying at a concentration of coincident loci, which could be due to the suppression of recombination associated with centromeric regions. Because the putative

Fig. 3. (*concluded*).



centromere lies at the interchange between the two duplicated regions, it is possible that the observed organization of N10 is a result of centric fusion.

Duplicated loci were predominantly found on distinct linkage groups. Only four pairs of duplicate loci were mapped within the same A-genome linkage groups, one pair on N3 and three pairs on N6.

Intra-genomic duplication within the C genome (B. oleracea)

Table 3 lists the duplicate loci found within the C genome of *B. napus*, linkage groups N11 to N19, and shows the degree and distribution of intragenomic duplication within the C genome. The duplicated regions contained between 1 and 16 loci. As with the A genome, the vast majority of duplicate loci were on different linkage groups; however, a greater number of linkage groups, seven in total, had internal duplications. A previous study of *B. oleracea* detected such intra-chromosomal duplications on three of their nine linkage groups and it was suggested that such duplications may be common in *Brassica* species (Lan et al. 2000). The intrachromosomal duplication of loci detected by pR36 on both N11 and N17 probably indicates that the duplication of pR36 occurred before the duplication of a larger region (Fig. 3*a*). All of the intrachromosomal duplications involved either one or two loci apart from the duplication on N14. The distribution of the six pairs of duplicated loci on N14 suggest the ancient remnants of an isochromosome that has undergone further rearrangement, like the insertion of the middle segment, which showed conserved marker order with N19 (Fig. 3*b*). The only disruption to the conserved order between N14 and N19 is the position of pR64 on N14; however, if pR64 was also internally duplicated on N14, the order would be identical.

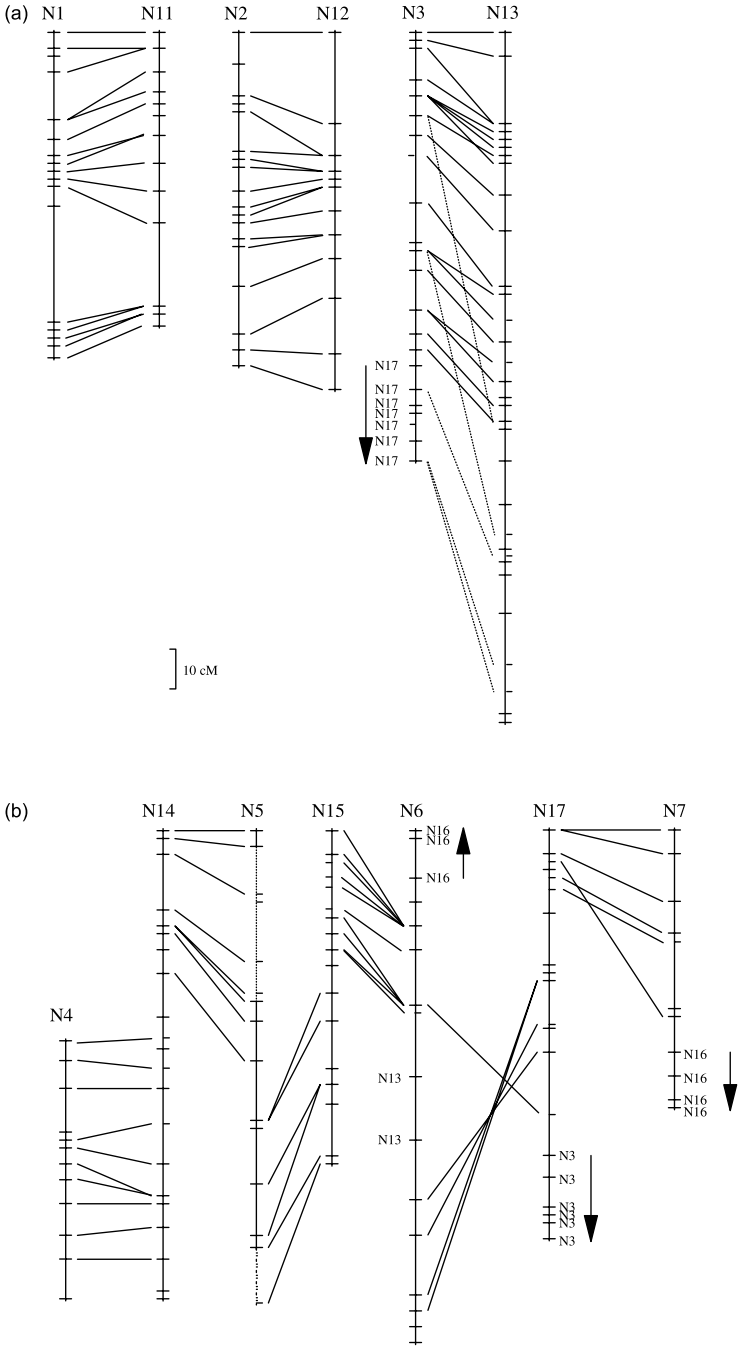
Twenty-eight regions had three or more loci duplicated within the C genome (Table 3). Five of these displayed conserved locus order between the linkage groups. No whole linkage groups of the C genome were found to be duplicated. The loci duplicated in linkage groups N13 and N19 appeared to extend over the entire length of linkage group N19 (Fig. 3*c*). The top of N19 from marker locus pN3 to locus pO155 was collinear with N13; however, a number of loci from this region on N19 mapped to the bottom of N13. Interestingly, one of these loci, pR64, was internally duplicated on N13. It is possible that a larger undetected region is duplicated on N13. This would explain the position of pW102, which disrupts the homology between N12 and N13 (Fig. 3*c*). N12, N13, and N19 probably arose from a common ancestral chromosome, which is substantiated by the strong collinearity exhibited between these three linkage groups and chromosome V of the model crucifer *Arabidopsis thaliana* (Parkin et al. 2002).

A large conserved region is triplicated between linkage groups N11, N17, and the top of N18 (Fig. 3*a*). The bottom segment of N18 shows strong homology to a large portion of linkage group N15, but did not appear to be triplicated.

Comparison of intragenomic duplication within the A and C genomes

The organization of a number of regions affected by intragenomic duplications appeared to have similar structures in both the A and C genomes. The probes that detected loci on N1 and N3 also detected loci on N11 and N17. The replicated region was collinear in both cases and an internal inversion between the markers detected by pO52 and pR36 was perfectly conserved in both diploid genomes (Figs. 2*a* and 3*a*). Similarly, duplicated sets of homologous loci have been mapped on the A-genome linkage groups N2, N3, and

Fig. 4. The primary regions of homoeology between the A and C genomes of *B. napus*. Where possible each of the linkage groups from the A and C genomes (vertical lines) has been aligned with its primary homoeologue. Where the groups are not aligned with their homoeologues, arrows are drawn to indicate whether the marker loci are in the same orientation (pointing down) or are inverted (pointing up) with respect to the homologous region. The linkage map is based on that from Parkin and Lydiate (1997) extra loci (half horizontal bars) and extra sections (dotted lines) which have been added through alignment with Bohuon et al. (1996) are indicated. Additional loci were added to linkage group N9 through alignment with the *B. napus* genetic linkage map described in Sharpe (1997) again indicated by a dotted line. A potential inversion event separating N8 and N18 is indicated by a dotted arrow.

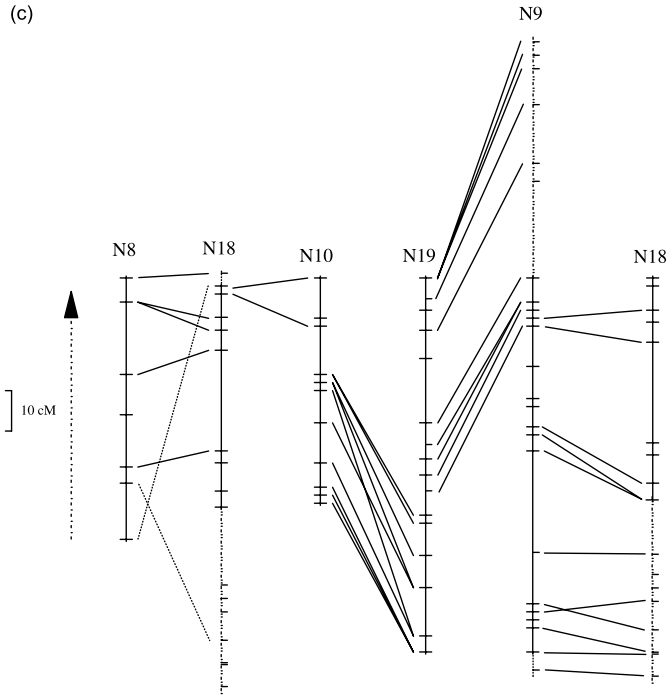


N10, and the C-genome linkage groups N12, N13, and N19 (Figs. 2*b* and 3*c*). These examples of conserved intra-genomic rearrangements of the A and C genome identify regions that have remained unchanged during the divergence of *B. oleracea* and *B. rapa* from a common ancestor and demonstrate that the rearrangements predate their evolution into distinct species.

Inter-genomic duplication and identifying regions of primary homoeology

Regions of homology were found between the A and C genomes ranging from single pairs of loci to whole linkage groups. Each linkage group from each diploid genome shared homologous loci with a number of linkage groups from the other diploid genome as anticipated from the extent

Fig. 4. (*concluded*).



of intragenomic duplication within each genome. However, the degree of conserved synteny varied considerably. It was possible to align the A- and C-genome linkage groups based on the longest stretches of collinearity (Fig. 4). These regions were termed the primary regions of homoeology because they probably have the most direct evolutionary relationship and evolved from the same chromosomal segments of the common ancestor of *B. oleracea* and *B. rapa*. This designation of syntenous groups has been confirmed in most instances by the localization of homoeologous recombination events (Parkin et al. 1995; Sharpe et al. 1995; R. Ruiter, U. Lagercrantz, and D.J. Lydiate, unpublished). Linkage groups N1 with N11 and N2 with N12 of the A and C genomes, respectively, appeared completely syntenous (Fig. 4). Groups N3 and N13 showed extensive collinearity apart from the bottom sections of each group, which were completely different. The remaining homoeologous regions tended to be between half linkage groups, suggesting that translocations have taken place during the divergence of *B. oleracea* and *B. rapa*. Two of the translocation end points, on N14 and N17, are coincident with the centromere position, which provides further circumstantial evidence for centric fission and (or) fusion playing a role in the evolution of *Brassica* chromosomes (Kelly 1996). If all the breakpoints between regions of homology are considered to be as the result of chromosomal rearrangements, then the A and C genomes have been differentiated by at least 16 observable events. There are at least 16 possible translocation endpoints differentiating the genomes and at least 2 of the translocated regions have been inverted, i.e., between linkage groups N6 and N17 and between groups N6 and N16. There is also some evidence to suggest that there has been an inversion between linkage groups N8 and N18.

Discussion

Because unmapped monomorphic loci are not considered, the level of genomic duplication reported in genetic mapping studies tends to be an underestimate and is a direct reflection of the choice of marker, the number of restriction enzymes employed (in the case of RFLPs), and the choice of parents for the mapping population. In *Brassica*, levels of genome duplication estimated from mapped loci have been between 34 and 54% (Slocum et al. 1990; Bohuon et al. 1996; Truco et al. 1996). In the present study, 73% of the genomic clones were found to detect two or more duplicate sequences within each of the diploid genomes of *B. napus*. The high proportion of duplicated loci in the diploid *Brassica* species is in all probability the result of evolution from a single progenitor of lower chromosome number. The single-copy probes invariably detected a single locus in both the A and C genome. The detection of these intergenomic duplicate copies suggests that the organization of the unique loci predates the speciation of *B. oleracea* and *B. rapa*. The lack of loci unique to one or other genome could indicate the vital nature of the current loci within the *B. napus* background or could be the result of the recent evolution of the diploids from a common polyploid ancestor (Lagercrantz and Lydiate 1996; Parkin et al. 2002).

It would be difficult to conclude from the presented data that each portion of the genome is represented in six copies within the *B. napus* genome, which would be anticipated if the diploid *Brassica* genomes had evolved from a hexaploid ancestor. However, the largest regions showing intragenomic homology within each of the diploid genomes were triplicated (Figs. 2*a*, 2*b*, 3*a*, and 3*c*). These regions are also known to be collinear with *A. thaliana*, which suggests that they have undergone limited rearrangements and have retained much of the structure of the ancestral crucifer genome. Why there would be selection for the conservation of particular regions is an interesting question. It is possible that these regions enable (or contribute) to the concerted expression of sets of genes, that there are physical impediments preventing further large-scale rearrangements, similar to the suppression of recombination associated with inversions, or that there is a higher level of allelic variation associated with these regions, which makes their discovery through genetic mapping more likely.

It should be noted that in Scheffler et al. (1997), comparative mapping between *A. thaliana* chromosome III and *B. napus* identified three pairs of homologous regions in each diploid genome (N1/N11, N3/N13, and N5/N15). However, in the present study, for the same region we could only detect limited homology (one or two markers) between the expected regions of the A genome. Similarly for the C genome, although four markers from the region (Table 3) uncovered the expected homology between N13 and N15, only two markers from this region were found on the third region of linkage group N11 (Table 3). This illustrates one of the problems that can arise from using limited data points when trying to infer evolutionary relationships from comparative mapping studies. It is probable that further regions of duplication will be uncovered when markers are added to the linkage maps.

The RFLP probes generally hybridized to an even number of loci in *B. napus*, usually two, four, or six (Table 1), re-

flecting the presence of equivalent loci in the two diploid genomes. The conserved pairs of homoeoloci found in the *B. napus* genome probably indicate that relatively little duplication and (or) deletion of loci has occurred since the divergence of *B. rapa* and *B. oleracea* from a common ancestor. These data reinforce the conclusions of previous phylogenies based on polymorphisms in organellar DNA (Warwick and Black 1991) and nuclear DNA (Song and Osborn 1992), which suggested that the divergence of the diploid species *B. rapa* and *B. oleracea* was a recent event in the evolution of the *Brassica* genus. The arrangement of homoeloci is also reflected in the map position of loci controlling different traits in the *B. napus* genome, where the traits are determined by at least two loci, one originating from the A genome and the second from the C genome. Examples include loci controlling glucosinolate biosynthesis (Parkin et al. 1994) and stamenoid petal production (Fray et al. 1997). The fate of duplicate genes in polyploid genomes is an interesting question. Duplicate copies could either be deleted or silenced, redesigned to fill a new functional role, or maintain the same function (reviewed in Wendel 2000). In relatively young polyploids like hexaploid bread wheat, which is thought to have evolved about 8000 years ago, it has already been shown that in most instances homoeologous isozyme loci remain functional with no obvious tendency towards gene silencing (Hart and Langston 1987). The maintenance of homeoloci in amphidiploid genomes may provide a selectional advantage for the colonization of new environments, especially to predominantly in-breeding populations where the gene diversity normally provided through out crossing is not available. It is also possible that a proportion of the duplicate copies must be preserved to maintain functional protein complexes that are formed from the interactions of more than one gene product (Wendel 2000). Although the probes used in this study were generated from genomic *Pst*I clones, and although there is a tendency for such probes to derive from genic regions (Michalek et al. 1999; I.A.P. Parkin, unpublished data), only the correlation of expression data with mapped loci will confirm if *B. napus* has retained functional copies of the majority of homoeloci inherited from the A and C genomes.

When considering the relationships between duplicated regions within polyploid genomes, the many potential cycles of genome duplication and the consequent rearrangements that take place to bring about rediploidization of the genome at each round can confuse the identification of truly homoeologous regions. To align the homoeologous regions of the A and C genome of *B. napus* we were assisted by previously observed recombination events that had taken place between the A and C genomes (Parkin et al. 1995; Sharpe et al. 1995). The present study comparing genetic maps of the A and C genomes has shown that at least 16 gross chromosomal rearrangements separate the two. This number is significantly higher than the previously published figure of five rearrangements and is similar to the number believed to separate both the A and C genomes from the *Brassica* B genome of *B. nigra*, 10 and 12 chromosomal rearrangements, respectively (Lagercrantz 1998). Because the number of rearrangements separating the three *Brassica* genomes are very similar, it is interesting to note that homoeologous recombination is relatively common between the A and C genomes (Parkin et al. 1995; Sharpe et al. 1995; R. Ruiter, U. Lagercrantz, and D.J. Lydiate, unpublished), yet was not in evidence in a newly resynthesized *B. juncea* line (Axelsson et al. 2000) and limited chromosome association was observed in cytogenetic studies of meiosis in AB amphihaploids (Attia and Röbbelen 1986). Because it is unlikely that structural rearrangements are limiting such recombination, the inhibition of homoeologous recombination between the A and B genomes could be explained by variation at the level of the DNA nucleotide sequence. It has already been shown that the rate of nucleotide substitution between *B. nigra* and either *B. rapa* or *B. oleracea*, for different mitochondrial DNA sequences, was almost double that found between *B. rapa* and *B. oleracea* (Yang et al. 1999). Further evidence for elevated sequence divergence between the B genome and its relative diploid genomes can be inferred from a study identifying genome-specific repetitive sequences to use as probes in *Brassica* species (Somers and Demmon 2002). This study succeeded in isolating B genome specific probes, but did not isolate any probes that could distinguish between the A and C genomes.

Even with the relatively large number of rearrangements observed to have separated the A and C genomes, the level of synteny was striking, with little or no minor rearrangements of loci detected at this level of resolution. For almost every region of each diploid genome it was possible to identify a region of primary homoeology, which suggests that the genomic content of the A and C genomes are equivalent and that the difference in chromosome number is a result of the observed rearrangements. The ability to align the A and C genomes of the *Brassica* genus with confidence gives researchers primarily interested in *B. napus* the ability to use information and tools developed in each of the diploid species, *B. rapa* or *B. oleracea*, to facilitate work in the more complex amphidiploid genome. It also provides the necessary background information for extending these comparative mapping studies to the model genome of *A. thaliana* and the observed homoeology can act as an internal control in such mapping studies.

Ackowledgements

This research was supported by Advanta Seeds and Cambridge Plant Breeders – Twyfords.

References

Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature (London), **408**: 796–815.

Attia, T., and Röbbellen, G. 1986. Cytogenetic relationship within cultivated *Brassica* analysed in amphihaploids from the three diploid ancestors. Can. J. Genet. Cytol. **28**: 323–329.

Axelsson, T., Bowman, C.M., Sharpe, A.G., Lydiate, D.J., and Lagercrantz, U. 2000. Amphidiploid *Brassica juncea* contains conserved progenitor genomes. Genome, **43**: 679–688.

Bohuon, E. J. R., Keith, D. J., Parkin, I. A. P., Sharpe, A. G., and Lydiate, D. J. 1996 Alignment of the conserved C genomes of *Brassica oleracea* and *Brassica napus*. Theor. Appl. Genet. **93**: 833–839.

Cavell, A.C, Lydiate, D.J., Parkin, I.A.P., Dean, C., and Trick, M. 1998. Collinearity between a 30-centimorgan segment of

Arabidopsis thaliana chromosome 4 and duplicated regions within the *Brassica napus* genome. Genome, **41**: 62–69.

Fray, M.J., Puangsomlee, P., Goodrich, J., Coupland, G., Evans, E. J., Arthur, A. E., and Lydiate, D. J. 1997.The genetics of stamenoid petal production in oilseed rape (*Brassica napus*) and equivalent variation in *Arabidopsis thaliana*. Theor. Appl. Genet. **94**: 731–736.

Gale, M.D., and Devos, K.M. 1998. Plant comparative genetics after 10 years. Science (Washington, D.C.), **282**: 656–659.

Gaut, B.S., Le Thierry, d'E.M., Peek, A.S., and Sawkins, M.C. 2000. Maize as a model for the evolution of plant nuclear genomes, Proc. Natl. Acad. Sci. U.S.A. **97**: 7008–7015.

Hart, G.E., and Langston, P.J. 1987. Chromosomal location and evolution of isozyme structural genes in hexaploid wheat. Heredity, **39**: 263–277.

Jiang, C., Wright, R.J., El-Zik, K.M., and Paterson, A.H. 1998. Polyploid formation created unique avenues for response to selection in gossypium, Proc. Natl. Acad. Sci. U.S.A. **95**: 4419–4424.

Kelly, A. 1996. The genetic basis of petal number and pod orientation in oilseed rape (*B. napus*). Ph.D. thesis, University of Newcastle, Newcastle upon Tyne, U.K.

Lagercrantz, U. 1998. Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that *Brassica* genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. Genetics, **150**: 1217–1228.

Lagercrantz, U., and Lydiate, D.J. 1996. Comparative genome mapping in *Brassica*. Genetics, **144**: 1903–1910.

Lan, T-H., Delmonte, T.A., Reischmann, K.P., Hyman, J., Kowalski, S.P., McFerson, J., Kresovich, S., and Paterson, A.H. 2000. An EST-enriched comparative map of *Brassica oleracea* and *Arabidopsis thaliana*. Genome Res. **10**: 776–788.

Michalek, W., Kunzel, G., and Graner, A. 1999. Sequence analysis and gene identification in a set of mapped RFLP markers in barley (*Hordeum vulgare*). Genome, **42**: 849–853.

Parkin, I.A.P., and Lydiate, D. J. 1997. Conserved patterns of chromosome pairing and recombination in *Brassica napus* crosses. Genome, **40**: 496–504.

Parkin, I., Magrath, R., Keith, D., Sharpe, A., Mithen, R., and Lydiate, D. 1994. Genetics of aliphatic glucosinolates. II. Hydroxylation of alkenyl glucosinolates in *Brassica napus*. Heredity, **72**: 594–598.

Parkin, I.A.P., Sharpe, A.G., Keith, D.J., and Lydiate D.J. 1995. Identification of the A and C genomes of the amphidiploid *Brassica napus* (oilseed rape). Genome, **38**: 1122–1131.

Parkin, I.A.P., Lydiate, D. J., and Trick, M. 2002. Assessing the level of collinearity between *Arabidopsis thaliana* and *Brassica napus* for *A. thaliana* chromosome 5. Genome, **45**: 356–366.

Sasaki, T., and Burr, B. 2000. International rice genome sequencing project: the effort to completely sequence the rice genome. Curr. Opin. Plant Biol. **3**: 138–141.

Scheffler, J.A., Sharpe, A.G., Schmidt, H., Sperling, P., Parkin, I.A.P., Lydiate, D.J., and Heinz, E. 1997. Desaturase multigene families of *Brassica napus* arose through genome duplication. Theor. Appl. Genet. **94**: 583–591.

Sharpe, A.G., Parkin, I.A.P., Keith, D.J., and Lydiate D.J. 1995. Frequent non-reciprocal translocations in the amphidiploid genome of oilseed rape. Genome, **38**: 1112–1121.

Sharpe, A.G. 1997. Marker assisted breeding in oilseed rape (*Brassica napus*). Ph.D. thesis, University of East Anglia, Norwich, Norfolk, U.K.

Shoemaker, R.C., Polzin, K., Labate, J., Specht, J., Brummer, E.C., Olson, T., Young, N., Concibido, V., Wilcox, J., Tamulonis, J.P., Kochert, G., and Boerma, H.R. 1996. Genome duplication in soybean (*Glycine* subgenus *soja*), Genetics, **144**: 329–338.

Slocum, M.K., Figdore, S.S., Kennard, W.C., Suzuki, J.Y., and Osborn, T.C. 1990. Linkage arrangement of restriction fragment length polymorphism loci in *Brassica oleracea*. Theor. Appl. Genet. **80**: 57–64.

Somers, D.J., and Demmon, G. 2002. Identification of repetitive, genome-specific probes in crucifer oilseed species. Genome, **45**: 485–492.

Song, K., and Osborn, T.C. 1992. Polyphyletic origins of *Brassica napus*: new evidence based on organelle and nuclear RFLP analyses. Genome, **35**: 992–1001.

Truco, M.J., Hu, J., Sadowski, J., and Quiros, C.F. 1996. Inter- and intra-genomic homology of the Brassica genomes: implications for their origin and evolution. Theor. Appl. Genet. **93**: 1225–1233.

U, N. 1935 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.

Warwick, S. L., and Black, L. D. 1991. Molecular systematics of *Brassica* and allied genera (Subtribe Brassicinae, Brassiceae) — chloroplast genome and cytodeme congruence. Theor. Appl. Genet. **82**: 81–92.

Wendel, J.F. 2000. Genome evolution in polyploids. Plant Mol. Biol. **42**: 225–249.

Yang, Y.W., Lai, K.N., Tai, P.Y., and Li, W.H. 1999. Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between *Brassica* and other angiosperm lineages. J. Mol. Evol. **48**: 597–604.