**Genetics: Published Articles Ahead of Print, published on July 14, 2005 as 10.1534/genetics.105.042093**

*Arabidopsis*/*Brassica napus* comparative map

Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*.

Isobel A. P. Parkin*, Sigrun M. Gulden*, Andrew G. Sharpe*, Lewis Lukens†, Martin Trick§, Thomas C. Osborn‡ and Derek J. Lydiate*

*Saskatoon Research Centre, Agriculture and Agri-Food Canada, 107 Science Place, Saskatoon, SK., S7N 0X2, Canada.

†University of Guelph, Guelph, Ontario, N1G 2W1, Canada.

§John Innes Centre, Colney Lane, Norwich, NR4 7UH, UK.

‡Department of Agronomy, University of Wisconsin, 1575 Linden Drive, Madison, WI 53706, USA.

Sequence data from this article has been deposited with the EMBL/GenBank Data Libraries under accession nos. XXX-XXXX

Running Title: *Arabidopsis*/*Brassica napus* comparative map

Key Words: *Brassica* evolution, *Arabidopsis* model genome, collinearity, polyploidy

Corresponding Author:

Dr. Isobel A. P. Parkin

Saskatoon Research Centre, Agriculture and Agri-Food Canada,

107 Science Place, Saskatoon, SK., S7N 0X2, CANADA.

Tel: 1 306 956 7283

Fax: 1 306 956 7247

Email: ParkinI@agr.gc.ca

ABSTRACT

Over 1000 genetically linked RFLP loci in *Brassica napus* were mapped to homologous positions in the *Arabidopsis* genome based on sequence similarity. Blocks of genetically linked loci in *B. napus* frequently corresponded to physically linked markers in *Arabidopsis*. This comparative analysis allowed the identification of a minimum of 21 conserved genomic units within the *Arabidopsis* genome which can be duplicated and rearranged to generate the present day *B. napus* genome. The conserved regions extended over lengths as great as 50 cM in the *B. napus* genetic map, equivalent to approximately 9 Mb of contiguous sequence in the *Arabidopsis* genome. There was also evidence for conservation of chromosome landmarks, particularly centromeric regions, between the two species. The observed segmental structure of the *Brassica* genome strongly suggests that the extant *Brassica* diploid species evolved from an hexaploid ancestor. The comparative map assists in exploiting the *Arabidopsis* genomic sequence for marker and candidate gene identification within the larger, intractable genomes of the *Brassica* polyploids.

## INTRODUCTION

***ARABIDOPSIS*** *thaliana* (hereafter referred to as *Arabidopsis*) is one of almost 3500 species which make up the monophyletic family of the Brassicaceae (PRICE *et al.*, 1994). *Arabidopsis* thus shares recent common ancestry with a large number of species of significant economic importance, including a diverse range of vegetable and oil producing crops, the majority of which are *Brassica* species. *Arabidopsis* is an excellent model system for the Brassicaceae, with a small and relatively simple genome, efficient transformation system, diverse range of genetic and genomics resources, and a completed genome sequence (ARABIDOPSIS GENOME INITIATIVE, 2000).

Over the past ten years, plant comparative mapping has taken prominence as a powerful tool firstly for uncovering the processes and rate of genome evolution and secondly for allowing the transfer of genetic resources between species. Comparative mapping has been most extensively applied to the grasses where the genetic maps of eleven species, including the model monocot rice, have been aligned. These include 11 diverse species varying dramatically in haploid chromosome number, genome size, and phenotype (reviewed in DEVOS and GALE, 2000). Perhaps the most striking observation from the cereal studies was the extensive genome conservation observed between species that diverged millions of years ago. Using rice as the basal genome, fewer than 30 conserved blocks were identified, which could be rearranged and/or duplicated to form each of the other grass genomes. Comparative mapping studies among members of the Brassicaceae have been more ambiguous in their conclusions, leading to on-going discussions with regards to the level of genome duplication prevalent in modern day *Brassica* cultivars and the extent of the genome rearrangements which have occurred in the

evolution of these cultivars from a common ancestor (LAGERCRANTZ, 1998; LAN *et al.*, 2000; LUKENS *et al.*, 2003).

The present study focuses on the genome of the oilseed crop *Brassica napus*, which is an amphidiploid species formed from multiple independent fusion events between ancestors of the diploids *B. rapa* (A genome donor) and *B. oleracea* (C genome donor) (PARKIN *et al.,* 1995; PALMER *et al*., 1983; U, 1935). Polyploidy is a prevalent evolutionary mechanism within angiosperms since it has been estimated that 30-70% of modern plant species have evolved through a polyploid ancestor (reviewed in WENDEL, 2000). Polyploidy can occur either through the duplication of whole chromosome complements or the fusion of related chromosome complements, and stabilisation of the newly expanded karyotype must then take place to ensure normal diploid inheritance. Diploidisation of the novel polyploid can occur through chromosomal restructuring or genetic control of illegitimate recombination events or a combination of both mechanisms. It is widely accepted that the progenitor diploid genomes of *B. napus* are ancient polyploids and that large scale chromosome rearrangements have occurred since their evolution from a lower chromosome number progenitor (SCHMIDT *et al.*, 2001). What is more contentious is whether the diploids evolved through a hexaploid ancestor or whether they were formed via segmental duplication of one or two ancestral genomes (LUKENS *et al.*, 2004). *B. napus*, a relatively young amphidiploid, is somewhat of an anomaly since it has been established that no major chromosomal rearrangements have occurred since the fusion of the progenitor A and C genomes, but homoeologous recombination events between these two related genomes are common in newly resynthesised *B. napus* lines and have been observed at low levels in established canola cultivars (UDALL *et al*, 2004; PARKIN *et al.*, 1995; SHARPE

*et al.*, 1995). It has yet to be established if *B. napus* has evolved or inherited a locus controlling homologous pairing similar to the *Ph1* locus in hexaploid wheat (JENCZEWSKI *et al.*, 2003).

Comparative mapping between *B. napus* and *Arabidopsis* has thus far targeted small regions of the *Arabidopsis* genome, generally identifying three collinear segments in each of the diploid genomes for every region of *Arabidopsis* studied thereby promoting the idea that the diploid *Brassica* species may have evolved through a hexaploid ancestor (CAVELL *et al.*, 1998; OSBORN *et al.*, 1997; PARKIN *et al.*, 2002). However, at the same time regions suggesting a more complex relationship between the two species were also identified (OSBORN *et al.*, 1997; PARKIN *et al.*, 2002). In the earliest published global comparison between one of the diploid *Brassica*s, *B. nigra* (black mustard), and *Arabidopsis*, an extensive number of rearrangements were invoked to explain how the two extant diploid genomes evolved from a common hexaploid ancestor (LAGERCRANTZ, 1998). There have been four global comparisons of the genomes of *B. oleracea* and *Arabidopsis*. Although all have been limited by a low density of common loci, three identified extensive synteny between the two genomes but were inconclusive in assessing the level of duplication of the collinear segments (LAN *et al.*, 2000; BABULA *et al.*, 2003; LUKENS *et al.*, 2003). A more recent comparison of the *B. oleracea* and *Arabidopsis* genomes refuted the possibility of a hexaploid ancestor, citing evidence of syntenous blocks ranging in copy number from one to seven (LI *et al*., 2003).

The present study describes a comprehensive comparison of a *Brassica* genome with that of *Arabidopsis*. Sequences of 359 probes derived from *Brassica* and *Arabidopsis* that detect 1,232 genetically mapped loci in *B. napus,* were used to query the *Arabidopsis* genome, revealing 550 homologous sequences and their inferred chromosomal positions. The data provides strong evidence to support the hypothesis that the *Brassica* diploid genomes evolved

through a hexaploid ancestor and suggests conservation of some centromeric regions between the two species. The postulated ancestor appears to have been formed from duplication events which occurred subsequent to the putative global duplication events which took place between 65 and 90 million years ago during the evolution of *Arabidopsis* (LYNCH and CONERY, 2000; SIMILLION *et al*., 2002; RAES *et al.*, 2003). The resultant genetic and physical comparative map can be used not only to infer genome rearrangements during the evolution of the *Brassica* species but also to identify regions of the *Arabidopsis* genome which may harbour genes of interest and should potentiate the exploitation of *Arabidopsis* genomics tools in *Brassica* research.

## MATERIALS AND METHODS

**Genetic Linkage Analysis:** Genetic linkage analysis in *B. napus* was carried out as described previously except hybridisations with *Arabidopsis* clones were washed only at low stringency (2 X SSC, 0.1% SDS) (SHARPE *et al.*, 1995). The *B. napus* population consisted of 60 doubled haploid lines derived from crosses between a winter *B. napus* breeding line (CPB87/5) and a newly resynthesised *B. napus* line (SYN1) as described in PARKIN *et al.* (1995). The genetic map also includes loci positioned through previously described map alignments with a second linkage map of *B. napus* and one of *B. oleracea* (PARKIN and LYDIATE, 1997; BOHUON *et al.*, 1996). Briefly, common parental genotypes allowed corresponding loci to be identified between the maps through the inheritance of identical RFLP alleles. Loci mapped in only one population which co-segregated with such common loci were positioned at that locus in the combined map. Loci mapped in only one population positioned between common loci were

placed in the corresponding interval in the combined map based on their relative position in the map of origin. The RFLP probes consisted of 213 *Brassica* genomic clones (pN, pO, pR, pW: SHARPE *et al.*, 1995), 61 *Brassica* cDNA clones (CA, es), 88 *Arabidopsis* cDNA clones (I, N, R, Z: SILLITO *et al.*, 2000) and six cloned *Brassica* or *Arabidopsis* genes (ACYL, CONSTANS, FCA, HS1, oleosin: pC2, Δ9 desaturase: pC3). The genetic linkage map was constructed using Mapmaker v3 with a LOD score of 4.0 (LANDER *et al*, 1987) and the linkage groups were drawn using Mapchart (VOORRIPS 2002). Irregularities in meiotic pairing in the resynthesised *B. napus* parental line of the doubled haploid population used for the initial and the additional mapping, caused a non-disjunction event which prevented the accurate mapping of further loci to linkage group N16 (PARKIN *et al.*, 1995). A limited map of N16 derived from the alignment of N16 from *B. napus*, described in SHARPE *et al.* (1995), and O6 from *B. oleracea*, described in BOHUON *et al.* (1996) has been used in the present analysis. A similar alignment of N16 and O6 was discussed in RYDER et al. (2001).

**Sequence Analysis:** *Brassica* genomic or cDNA clones were sequenced from each end using the BigDye$^{TM}$ v2 Terminator cycle sequencing kit according to the instructions of the manufacturer and subsequently the reactions were run out on an automated ABI377 DNA Sequencer (Applied Biosystems, Foster City, CA, USA). The *Brassica* sequences were analysed using Sequencher (Gene Codes Corp, Ann Arbor, MI, USA) to trim vector sequence, identify overlaps and generate contigs. *Brassica* and *Arabidopsis* sequences were analysed for homology to the TIGR *Arabidopsis* pseudo chromosome genomic sequence version 5.0 (ftp://ftp.tigr.org/pub/data/a_thaliana/) using the BLAST programs of the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/) housed on a Linux server. Low complexity sequences were filtered in the BLAST analysis, and default values for cost

(mismatch cost = -3.0), reward (match reward= 1.0), and wordsize (11 bp) were selected. The default gap opening penalty (5.0), and the gap extension penalty (2.0) were also selected. Perl script was used to extract the base pair position in the *Arabidopsis* genomic sequence of each HSP (high scoring segment pairs), identified with BLASTN, for each clone where the primary HSP had an E value of less than or equal to 1E-07 (Table I, supplemental data).

## RESULTS

**Comparative map of *Brassica napus* and *Arabidopsis:*** Genetic linkage mapping of restriction fragment length polymorphisms (RFLPs) identified with 183 *Brassica* and *Arabidopsis* cDNA clones added a further 646 loci to the published aligned map of *B. napus* (BOHUON *et al.*, 1996; PARKIN and LYDIATE, 1997). The complete *B. napus* linkage map is presented in Figure 1 and consists of 1317 genetic loci distributed over nineteen linkage groups with a combined map length of 1,968 cM.

The genetic linkage map was generated from segregating loci detected by 368 DNA clones, 274 of which were derived from anonymous *Brassica* genomic or complementary DNA, five were *Brassica* homologues of known genes and the remaining 89 clones were derived from *Arabidopsis* cDNA. Sequence data was obtained for 267 of the anonymous *Brassica* clones and BLASTN analysis was used to identify homologous loci within the *Arabidopsis* genome for each clone. A fairly low expect (E) value was used as the exclusion cut off (1E-07) (supplemental data, Table I). The low E value was adopted to maximise the number of number of markers positioned, since the majority of the probes were derived from genomic, potentially inter-genic, DNA. Two hundred and fifty-eight of the *Brassica* clones displayed homology to 404 regions

within the *Arabidopsis* genome, with an average sequence identity of 86% over all aligned highest scoring pairs (HSPs). The majority of these hits were to genic regions and the most similar *Arabidopsis* gene was identified for each clone (supplemental data, Table I). A less stringent E value can lead to the identification of a large number of small non-specific regions of homology (LUKENS *et al.*, 2003). Fifty-eight of the 258 clones identified regions of similarity with bit scores lower than 82, a value suggested as a cut-off for identifying orthologous sequences within the *Arabidopsis* genome for *Brassica* markers (LUKENS *et al*., 2003). For 11 clones these lower scoring hits represented their only or primary region of homology within the *Arabidopsis* genome, the data for these clones was included in the comparative analysis described below. The remainder of the low scoring hits represented secondary or tertiary regions of homology which generally fell within predefined duplicated regions within the *Arabidopsis* genome (*ARABIDOPSIS* GENOME INITIATIVE, 2000), and these data did not impact on the comparative analysis. Ten of the *Brassica* genomic clones showed no significant homology to the *Arabidopsis* genomic sequence at an E value of 1E-07, one clone, pR113, mapped to the *Arabidopsis* genome over multiple adjacent HSPs, but with an E value of 1E-06. Subsequent BLASTX analysis of the remaining nine clones identified related sequence for two clones, pR30 and pN87, which showed significant (1E-44 and 1E-94 respectively) homology to an annotated retroelement pol polyprotein sequence (At3g29156). Perhaps not surprisingly neither of these clones mapped to syntenous regions between the two genomes (see below).

To position the *Arabidopsis* clones accurately relative to the *Brassica* sequences all the clones were compared to the *Arabidopsis* pseudo-chromosome sequence using BLASTN analysis. In total, 550 loci were physically positioned within the *Arabidopsis* genome based on sequence identity (an average of one comparative marker every 214 Kb). These same clones

identified 1,232 RFLP loci on the genetic linkage map of *B. napus* (an average of one comparative marker every 1.6 cM). In Figure 1 each of the *B. napus* genetic loci has been colour coded according to the most significant BLASTN hit for the probe which detected that locus. Forty-two percent of the RFLP clones probes showed sequence similarity to more than one region of the *Arabidopsis* genome. Some of the mapped homologous loci in *B. napus* may represent orthologues of these secondary hits within the model genome. *Brassica* loci whose position within a conserved block in *Arabidopsis* was dependant upon such secondary hits are colour coded according to the appropriate duplicate hit and are identified in italics in Figure 1.

All of the *B. napus* linkage groups were composed of loci identified by probes related to sequence from each of the five *Arabidopsis* chromosomes (Table 1 and Figure 1). If the *Brassica* genomes evolved through simple polyploidy from a lower chromosome ancestor similar to *Arabidopsis*, it might be expected that the comparative loci mapped within the *B. napus* genome would be equally represented across the *Arabidopsis* genome. However, the number of loci originating from each *Arabidopsis* chromosome was not evenly distributed with significantly fewer loci than expected detected by probes showing homology to *Arabidopsis* chromosomes 2 and 3 and significantly more loci than expected detected by probes with homology to *Arabidopsis* chromosome 5 ($p<0.001$ for a goodness-of-fit test) (Table 1). This non-random distribution could be a function of a reduction in chromosome number in the *Arabidopsis* lineage and/or a function of gene loss occurring after genome duplication events within *Arabidopsis*.

**Identification of conserved blocks between *Arabidopsis* and *B. napus:*** For each *B. napus* linkage group it was possible to identify blocks of conserved synteny between *B. napus* and

*Arabidopsis* which represent chromosomal segments that have been maintained since the divergence of *Arabidopsis* and *Brassica* from a common ancestor (Figures 1 and 2).

A conserved block is defined as a region that contains several closely linked homologous loci in both the *Arabidopsis* and *Brassica* genomes. Each block has a minimum of four mapped loci with at least one shared locus every 5 cM in *B. napus* and at least one shared locus every 1 Mb in *Arabidopsis*. Using these criteria, each conserved block contained on average 7.8 shared loci and had an average length of 14.8 cM in *B. napus* and 4.8 Mb in *Arabidopsis*. Together the blocks covered almost 90% of the mapped length of the *B. napus* genome. The average physical distance covered in the *Arabidopsis* genome per 1 cM of genetic distance in the *B. napus* genome was calculated for every pair of comparative markers identified within the conserved blocks (Figure 3). The distribution was skewed with 35% of the intervals tested giving a ratio of 1 cM of the *B. napus* genetic map to 100,000 bp or less of *Arabidopsis* sequence, with a median ratio of 1 cM to 160,767 bp.

Based on the conserved blocks, 21 segments were identified within the *Arabidopsis* genome which could be duplicated and rearranged to form the skeleton of the *B. napus* genome (Figure 1, Figure 2). Although coverage of the two genomes is extensive there are areas where marker density is limited, specifically the regions spanning the *Arabidopsis* centromeres (Figure 2). The low copy number sequences utilised in the *Brassica* mapping would be expected to have lower levels of similarity to centromeres, since they tend to be located within gene poor transposable element rich regions (*ARABIDOPSIS* GENOME INITIATIVE, 2000). **Comparative genome organisation:** The organisation of the *B. napus* genome in comparison to the *Arabidopsis* genome as depicted in Figures 1 and 2 has been summarised for each of the linkage groups. Due to the close homology between the A (N1-N10) and C (N11-N19) genomes

of *B. napus*, the primary homoeologues in *B. napus* (described in PARKIN *et al.,* 2003) are indicated in the comparison.

*N1/N11:* These two *B. napus* linkage groups are homologous along their entire length. The top half of each linkage group shows significant homology to the long arm of *Arabidopsis* chromosome 4 (block C4B) with one inversion, previously noted in CAVELL *et al.* (1998), disrupting the collinearity between the two genomes. The inversion appears to be specific to N1/N11, and is not present in the homologous regions of linkage groups N3/N17 and N8/N18 where copies of block C4B were found. The lower half of N1/N11 is homologous to the top arm of *Arabidopsis* chromosome 3 (block C3A). This block is also strongly conserved on N5/N15 and N3/N13. In each case the distal end of the *Arabidopsis* chromosome corresponds with the terminal end of the linkage groups. At the breakpoint between the two large stretches of collinearity there are three markers that span the centromere on *Arabidopsis* C3 and additional markers that do not identify a conserved region. One gross chromosomal rearrangement would be sufficient to generate N1/N11 from the blocks defined in Figure 2.

*N2/N12:* These two linkage groups are homologous along their mapped length. PARKIN *et al.* (2002) previously described the relationship between N2/N12 and *Arabidopsis* C5, where the upper region of N2/N12 is homologous to the top 8 Mb of *Arabidopsis* C5 (block C5A) and an inversion on *Arabidopsis* C5 has moved block C5E to lie adjacent to block C5A. This pattern of C5A-C5E is conserved on linkage groups N3/N13 and N10/N19. The same inversion moved blocks C5B and C5D to the bottom of N2/N12. N2/N12 share a region of homology with *Arabidopsis* C1, block C1E, adjacent to which are five markers that flank the centromere on *Arabidopsis* C4. Two further small conserved regions were identified on N2/N12, C3B and C5F.

One inversion on *Arabidopsis* C5 and three insertion/deletion/translocation events represent the least number of rearrangements, which could generate the present organisation of N2/N12.

*N3(N17)/N13:* The homology of N3/N13 to C5 is described above, below which N3/N13 share homology with *Arabidopsis* C2 (block C2BC). Block C2BC on N3/N13 was defined by a lower density of comparative markers, which were further rearranged by an inversion, compared to the duplicated copies of C2BC found on N4/N14 and N5. The lower end of C2BC on N3/N13, which borders the centromere on C2, lies adjacent to a conserved block originating from the centromeric region of *Arabidopsis* C4 (block C4A). Below C4A, N3/N13 share homology with block C3A as described above. At the junction of C3A, which lies proximal to the centromere on C3, N3 is no longer homologous to N13 but instead shares homology with linkage group N17 and *Arabidopsis* C4 as described above. The remainder of linkage group N13 has no clear region of homoeology in the *B. napus* A genome. However, in relation to *Arabidopsis* this region of N13 shares homology with the blocks flanking the centromere of C3 (C3B-C3C), block C1B and block C4B. In the area which would be homologous to the centromeric region of C3 there are eight markers with homology to different *Arabidopsis* chromosomes, three of which flank the centromere on C2. At least three gross chromosomal rearrangements and two inversions are necessary to generate N3 from the identified conserved blocks, assuming C3ABC has been essentially conserved one additional translocation/insertion would be necessary to generate N13.

*N4/N14/N5*: The majority of N4 and N14 (65% and 75% of the mapped length, respectively) and the upper half of N5 share homology with *Arabidopsis* C2. The organization of N14 suggests that of an isocentric chromosome with the upper and lower arms sharing numerous common markers mapped in inverse orientation with respect to each other. The top of N4 and the homoeologous central section of N14 show small blocks of collinearity with *Arabidopsis* C3,

C4 and C5; N14 has one additional block from C1. Three gross chromosomal rearrangements are sufficient to describe the organisation of N4 and one additional inversion and two translocation/insertions would describe N14.

*N5/N15/N6*: The lower half of N5 and N15 as described above (for N1/N11) are collinear with the long arm of *Arabidopsis* C3. At the centre of N5/N15, the markers originate from *Arabidopsis* C1, with comparative markers flanking the centromere on C1. This central region on N15 is part of a larger block which is collinear with the upper arm of *Arabidopsis* C1 and the homoeologous region of *B. napus* N6. One and two large chromosomal rearrangements would generate the present organisation of N15 and N5 respectively.

*N6/N17:* The lower half of N6 shows homology to sections of *Arabidopsis* C5 and C3. The region from block C5B to the bottom of N6 is homoeologous but inverted with respect to N17. There are two markers on N6/N17 (CA129 and es1732) that identify sequences on the short arm of *Arabidopsis* C2; there was insufficient marker data from this region to identify a conserved block, however fine mapping of a dwarf gene in *B. rapa* has subsequently aligned this region of N6 with the short arm of *Arabidopsis* C2 (MUANGPROM and OSBORN, 2004**)**. It is to be expected that for regions such as these, flanking the *Arabidopsis* centromeres where there is a dearth of comparative markers, further conserved blocks will be identified. The comparison of N6/N17 to *Arabidopsis* is complex relative to other *B. napus* linkage groups and at least five and six chromosomal rearrangements need to be invoked to generate N6 and N17, respectively.

*N7(N16)/N17:* The top of N7/N17 are homologous to the short arm of *Arabidopsis* C2 including comparative markers which flank the centromere on C2. Homoeology between N7/N17 breaks down after block C1B, where the lower half of N7 is homologous with N16 and *Arabidopsis* C1. Due to the constraints of the mapping population (refer to Materials and Methods) there are

limited markers mapped to N16, making the number of rearrangements difficult to interpret. The data suggests that at least three translocations/deletions/insertions of conserved blocks have taken place to give N7 and at least one chromosomal rearrangement gave rise to N16.

*N8/N18/N9:* The whole of N8 appears to be homoeologous with N18, and is syntenous with *Arabidopsis* C1C, C4B and C1AB; however block C1AB is inverted on N18 with respect to N8. The remainder of N18 is homoeologous to the lower portion of N9 and is syntenous with *Arabidopsis* C3D, C2B, a fraction of C1B and C1A. The latter block forms part of an internal duplication on N18. One insertion of block C4B into the centromeric region lying between C1AB and C1C and two inversions (in C4B) could describe N8. The same insertion of C4B found on N8, duplication of C1A and translocation/insertion of C3D would generate N18.

*N9/N19/N10*: N10 and N19 share a region which is syntenous with *Arabidopsis* C5 as described above (for N2/N12). The end of C5E, which coincides with the break in homology between N10/N19, separates a region of apparent conservation between the two species from one which is fragmented. The tops of N9 and N19 share loci from comparative markers which are assigned to a number of blocks, running from the top of N9/N19 in the order C4A-C5B-C5F-C1D-C5D-C4A. There is no clear region of homology in the *B. napus* C genome for the top of N10, which is syntenous with *Arabidopsis* C1. N9 has the most complex segmental pattern of all the linkage groups necessitating at least nine chromosomal rearrangements to generate the mapped group. One inversion on C5 (as described for N2/N12) and one translocation would explain N10, one inversion and six further rearrangements would explain N19.

At least seventy-four translocations, fusions, deletions or inversions of the 21 conserved segments found within the *Arabidopsis* genome are necessary to generate the present day *B. napus* genome. However, 28 of these rearrangements are common to both the A and C genomes

of *B. napus*, suggesting they occurred prior to their divergence from a common ancestor. As described above, a number of the breakpoints between conserved segments correspond to previously defined translocation end points which separate the A and C genomes of *B. napus* (PARKIN *et al.*, 2003). In a number of instances the junctions of conserved blocks coincide with telomeric or centromeric regions of *Arabidopsis* suggesting centric fission and fusion have played a role in the chromosomal restructuring.

**Duplication within the *Brassica* Genome:** Counting the number of times a single *Arabidopsis* region is found within the *B. napus* genome provides an estimate of the level of genome duplication within *Brassica* compared to the model genome. Each conserved chromosomal segment was represented between four and seven times within the *B. napus* genome (Table 2). However, the organisation of the different duplicated copies of each block varied with respect to each other, either by the presence of additional rearrangements (see description for N1/N11 above) or by the number of comparative markers (see description for N3(N17)/N13 above). In *Arabidopsis*, 81% of the comparative loci positioned on the genome mapped to conserved regions present in at least six copies within the *B. napus* genome (Table 2). Eighty-six percent of the mapped length of the *B. napus* genome, which was arranged in conserved blocks, was found in at least six copies (Table 2). These results corroborate previous suggestions based on more limited data that the *Brassica* diploid genomes have evolved through a hexaploid ancestor. However, the presence of seven copies of some *Arabidopsis* regions within the *B. napus* genome suggests that further segmental duplication events may have occurred subsequent to any polyploidy event(s).

**Consequences of duplication within the *Arabidopsis* genome:** The majority of the conserved *Arabidopsis* blocks, including those known to be part of duplicated regions within *Arabidopsis*,

are each found between five and seven times within the *B. napus* genome. Effectively this means that the duplicated regions of the *Arabidopsis* genome are found between ten and fourteen times within the *B. napus* genome, similarly recent physical mapping carried out in *B. napus* identified twelve regions within the *B. napus* genome homologous to a small duplicated region of the *Arabidopsis* genome (RANA *et al.*, 2004). These data suggest the large segmental genomic duplications found within *Arabidopsis* occurred in the common ancestor of the two lineages prior to the formation of a *Brassica* hexaploid ancestor. These data are also consistent with the fact that the last round of genome duplication is believed to have occurred in *Arabidopsis* between 65 and 90 million years ago (LYNCH and CONERY, 2000; SIMILLION *et al*., 2002; RAES *et al.*, 2003) whereas the separation of the *Arabidopsis* and *Brassica* lineages is dated somewhere between 12 and 24 million years ago (KOCH *et al*, 2000).

Since the divergence of these two species one would expect the independent loss of redundant duplicate genes from both species. Several such losses were observed from the *Arabidopsis* genome. For example on N1 and N11, the upper parts of the linkage groups are collinear with the long arm of *Arabidopsis* chromosome 4 (Figure 1). Nonetheless, a number of *Brassica* loci were identified by probes (IC06, CA87, pN52, pN67) originating from *Arabidopsis* chromosome 2. Although these probes were found in regions identified as being duplicated between chromosomes 2 and 4 of *Arabidopsis* (http://www.tigr.org/tdb/e2k1/ath1/ Arabidopsis_genome_duplication.shtml) they showed no homology to *Arabidopsis* chromosome 4 sequence. Thus, *Brassica* has maintained duplicate copies of these sequences within the region equivalent to chromosome 4, whereas *Arabidopsis* has lost them.

In some instances the duplications evident within the *Arabidopsis* genome have made it difficult to identify the most similar region shared between the two species. For example, loci on

*B. napus* linkage group N19, show strong homology to both chromosome 5 block C and to the duplicated region on *Arabidopsis* chromosome 1 block D (Figure 4).

**Conservation of chromosome landmarks between the two species:** The position of each *Brassica* centromere has yet to be accurately determined relative to the genetic linkage maps. However, RFLP mapping of artefactual telocentric chromosomes in *Brassica* aneuploids placed the centromere of linkage group N12 between markers pW177E3 and pO5b, the centromere of group N13 between pW181a and pN96b and the centromere of group N14 between markers pN151b and pW130a (KELLY, 1996). Additionally, integration of the cytogenetic and genetic linkage maps of *B. oleracea* positioned the centromere of linkage group O1 (equivalent to N11) between markers pN152E1 and pO168E1 (HOWELL *et al.*, 2002).

In the proposed centromeric region of N12 four coincident markers were mapped with homology to *Arabidopsis* sequences that span the centromere on chromosome 4, suggesting conservation of chromosome position between the species. It is possible that with sufficient marker data the *Arabidopsis* centromeric positions could be used to predict functional and ancestral centromeric regions in *Brassica* chromosomes. The latter would arise, since a hexaploid derived from a lower chromosome progenitor, which likely had between 5 and 8 chromosomes, would have originally had between 15 and 24 functional centromeres, which were then reduced to 10 and 9 in the *Brassica* A and C genomes respectively. As in the case of N12, there were a number of instances where the density of markers across the *Arabidopsis* centromere was insufficient to identify a conserved block in *B. napus*. However, the loci identified by these same markers were tightly linked in *B. napus* and in the case of N11, N12 and N13 there was further cytological evidence suggesting the centromere location. These putative centromeric regions have each been indicated in Figure 1. As evidenced by numerous small

segments of collinearity flanking these provisional centromeric regions on N11, N12 and N14, it appears the neighbouring regions are prone to rearrangements and evolve rapidly compared to more distal regions.

The karyotype of *B. oleracea* indicates that linkage group O7 (equivalent to N17) is an acrocentric chromosome and has a strongly hybridising 45S locus at the terminus of the short arm (HOWELL *et al.*, 2002). This region of N17 shows homology to the short arm of *Arabidopsis* chromosome 2 and coincidently one of the two NOR regions of *Arabidopsis* also maps to the terminus of the short arm of chromosome 2 (FRANZ *et al.*, 1998).

## DISCUSSION

In the present study, by allowing minor disruptions in conserved regions it was possible to identify 21 conserved blocks within *Arabidopsis* which could be replicated and rearranged to cover almost 90% of the mapped length of *B. napus*. A minimum number of 74 gross rearrangements, with 38 in the A genome and 36 in the C genome, can be estimated to have separated the two lineages since their divergence 14-24 million years ago (mya) (KOCH et al, 2000). This lies between two previously published figures derived from *Brassica Arabidopsis* comparative mapping, which were 19 chromosomal rearrangements separating *B. oleracea* from *Arabidopsis* (LAN *et al.*, 2000) and 90 separating *B. nigra* from *Arabidopsis* (LAGERCRANTZ, 1998). Detecting rearrangements is influenced by a number of variables including the number and type of available comparative markers, the level of polymorphism within a mapping population and the method of determining synteny between species. For LAN *et al.* (2000) the lower figure was probably due to a low density of comparative markers and for LAGERCRANTZ *et al.* (1998) the much higher figure was due in part to the approach used to

identify syntenous regions, with no allowance made for minor disruptions of collinearity, and was exacerbated by the inclusion of markers thought to be single copy in *Arabidopsis* but now known to be multi-copy. Comparing estimates of the level of rearrangements in lineages is problematic because of the inherent difficulties in comparing between data sets and due to variation in the estimated divergence times. With that proviso, considering the data presented here, the level of rearrangement observed in the Brassiceae tribe, as represented by the A and C genomes of *B. napus*, is relatively high when compared with related species from the Brassicaceae family. Recently the genetic maps of *Capsella rubella* (Lepideae tribe) and *Arabidopsis lyrata* (Sisymbrieae tribe) have been compared to the sequence map of *A. thaliana* (BOIVIN *et al.*, 2004; KUITTINEN *et al*., 2004). Based on the comparison to the *A. thaliana* genome, analysis of the two maps indicates equivalent linkage group organisation, with the eight chromosomes of *C. rubella*, A-H, aligning with the *A. lyrata* chromosomes, AL1-AL8, respectively. This demonstrates that both species evolved from a common ancestor. *A. lyrata* and *C. rubella* are estimated to have diverged from *Arabidopsis* 5 mya and 10 mya, respectively (BOIVIN *et al.*, 2004; KUITTINEN *et al*., 2004). A limited number of major chromosomal rearrangements, approximately 6-13, separate these two species from *A. thaliana.* In addition, no major rearrangements have separated *A. lyrata* from *C. rubella*. Although it is not possible to align all the conserved blocks identified in this study with the *C. rubella* and *A. lyrata* genomes, the junctions of a number of the rearrangements identified between these two species and *A. thaliana* correspond to the ends of conserved blocks identified in this study. However, none of the chromosomal rearrangements which separate *A. lyrata* and *C. rubella* from *A. thaliana* appear to be common to the Brassiceae lineage.

The fact that the majority of the identified conserved segments are found in at least six copies in *B. napus,* and 81% of the comparative loci, which define the conserved blocks in *Arabidopsis,* are mapped to these triplicated regions, is consistent with a proposed hexaploid ancestor for the diploid *Brassica* progenitor. However, it could still be argued that the observed pattern of duplicated segments is the result of several smaller independent segmental duplications following a single whole genome duplication event, a mode of evolution which would require a significant number of independent duplication events. Polyploidy has been a prevalent mechanism of evolution within the angiosperms and it has been estimated that 30-70% of species having undergone at least one round of chromosome doubling during their evolutionary development (reviewed in WENDEL, 2000). There is also well documented evidence for extensive chromosomal rearrangements in newly resynthesised *Brassica* polyploids (SONG *et al*, 1995; PARKIN *et al*, 1995). Thus genome triplication followed by a small number of insertions/deletions/translocations would provide the simplest explanation for the present structure of the *Brassica* diploid genome.

In this study, the overall picture is one of conservation of gene content and gene order between the genomes of *Arabidopsis* and *B. napus*. The average length of the conserved blocks identified between the two species was 14.8 cM in *B. napus* and 4.8 Mb in *Arabidopsis*. However, for at least seven *B. napus* linkage groups half their mapped length was equivalent to one conserved region of the *Arabidopsis* genome. Undoubtedly the *Brassica* genomes have undergone restructuring during their evolution from a common ancestor of *Arabidopsis*, but this has not prevented the maintenance of large stretches of similarity, in some cases equivalent to 9 Mb of contiguous *Arabidopsis* genomic sequence. In a number of instances the comparative mapping provisionally suggests correspondence of centromere positions between the two

species. The large conserved regions found across the different genomes, punctuated by numerous smaller blocks of similarity suggest there are preferential regions for chromosome breakage and subsequent rearrangements.

The publication of the genome sequence of *Arabidopsis* has opened up many avenues of research with the expectation that these endeavours would have applications in the study of the more complex genomes of crop plants (*ARABIDOPSIS* GENOME INITIATIVE, 2000). The complete sequence allowed the resolution of the exact physical positions for some 30,000 genes, 50% of which have no known function, and any of which could hold the key to understanding a number of important agronomic traits. The comparative map suggests that the model genome of *Arabidopsis* can be widely exploited to infer the genetic basis of traits in its economically valuable *Brassica* crop relatives. In the identified conserved regions, the *Arabidopsis* genomic sequence should be an excellent resource for identifying useful markers, targeting the genic regions, since they show on average 86% sequence identity. Accurately mapping the genes controlling target phenotypes in large segregating *Brassica* populations should allow candidate genes to be inferred from the *Arabidopsis* sequence. However, due to the duplicated nature of the *Brassica* genomes it will be difficult to predict whether any particular *Arabidopsis* gene will have been maintained in all the duplicate copies. Comparative genomic sequencing in other plant species suggest that there will have almost certainly been numerous rearrangements at the level of microsynteny (BENNETZEN and RAMAKRISHNA, 2002). Limited physical mapping in *B. oleracea* only identified one potential inversion and one gene in a non-syntenic position; however, there was obvious interspersed gene loss from the different triplicated regions (O'NEILL and BANCROFT, 2000). In addition, recent physical mapping in the *B. napus* genome uncovered a similarly small number of disruptions in the microsynteny but evidence of

changes in gene content between the homologous *Brassica* segments compared to the homologous *Arabidopsis* regions (RANA *et al.*, 2004). Genomic sequence data of such regions from *Brassica* species will allow the extent to which the duplicate copies have been conserved to be determined, provide insights into the mechanism underlying the rearrangements differentiating the different copies and allow an estimate of the relative age of the different duplication events.

This research was funded by the Saskatchewan Agri-Food Innovation Fund. We would like to thank Cambridge Plant Breeders, Twyfords and Advanta who supported the development of the *Brassica* RFLP probes (pN, pO, pR) and the construction of the genetic linkage maps of *B. napus* described in Parkin *et al.* (1995) and Sharpe *et al* (1995). We would like to thank Christopher Lewis from AAFC Saskatoon Research Centre, who assisted with the Blast analysis and provided Perl scripts for handling the data. We would also like to thank Drs Stephen Robinson and Steve Barnes for critical reading of the manuscript.

## LITERATURE CITED

*ARABIDOPSIS* GENOME INITIATIVE**,** 2000 Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408:** 796-815.

AXELSSON, T., C. M. BOWMAN, A. G. SHARPE, D. J. LYDIATE, and U. LAGERCRANTZ, 2000 Amphidiploid *Brassica juncea* contains conserved progenitor genomes. Genome **43:** 679-688.

BABULA, D., M. KACZMAREK, A. BARAKAT, M. DELSENY, C.F. QUIROS and J. SADOWSKI, 2003 Chromosomal mapping of *Brassica oleracea* based on ESTs from *Arabidopsis thaliana*: complexity of the comparative map. Mol. Genet. Genom. **268**: 656-665.

BENNETZEN, J. L. and W. RAMAKRISHNA, 2002 Numerous small rearrangements of gene content, order and orientation differentiate grass genomes. Plant Mol. Biol. **48:** 821-827.

BOHUON, E. J. R., D. J. KEITH, I. A. P. PARKIN, A. G. SHARPE, and D. J. LYDIATE, 1996 Alignment of the conserved C genomes of *Brassica oleracea* and *Brassica napus*. Theor. Appl. Genet. **93:** 833-839.

BOIVIN, K., A. ACARKAN, R. S. MBULU, O. CLARENZ, and R. SCHMIDT, 2004 The *Arabidopsis* genome sequence as a tool for genome analysis in Brassicaceae. A comparison of the *Arabidopsis* and *Capsella rubella* genomes. Plant Physiol. **135:** 735-744.

CAVELL, A. C, D. J. LYDIATE, I. A. P. PARKIN, C. DEAN, and M. TRICK, 1998 Collinearity between a 30-centimorgan segment of *Arabidopsis thaliana* chromosome 4 and duplicated regions within the *Brassica napus* genome. Genome **41:** 62-69.

DEVOS, K. M. and M. D. GALE, 2000 Genome relationships: The grass model in current research. Plant Cell **12:** 637-646.

FRANZ, P., S. ARMSTRONG, C. ALONSO-BLANCO, T. C. FISCHER, R. A. TORRES-RULZ, and G. JONES, 1998 Cytogenetics of the model system *Arabidopsis thaliana*. Plant J. **13:** 867-876.

HOWELL, E. C., G. C. BARKER, G. H. JONES, M. J. KEARSEY, G. J. KING, E. P. KOP, C. D. RYDER, G. R. TEAKLE, J. G. VICENTE, and S. J. ARMSTRONG, 2002 Integration of the cytogenetic and genetic linkage maps of *Brassica oleracea*. Genetics **161:** 1225-1234.

JENCZEWSKI, E., F. EBER, A. GRIMAUD, S. HUET, M. O. LUCAS, H. MONOD, and A. M. CHEVRE, 2003 *PrBn*, a major gene controlling homoeologous pairing in oilseed rape (*Brassica napus*) haploids. Genetics **164:** 645-653.

KELLY, A. 1996 The genetic basis of petal number and pod orientation in oilseed rape (*B. napus*). PhD Thesis, University of Newcastle, UK.

KOCH, M. A., B. HAUBOLD, and T. MITCHELL-OLDS, 2000 Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). Mol. Biol. Evol. **17:** 1483-1498.

KUITTINEN, H., A. A. DE HAAN, C. VOGL, S. OIKARINEN, J. LEPPALA, M. KOCH, T. MITCHELL-OLDS, C. H. LANGLEY, and O. SAVOLAINEN, 2004 Comparing the linkage maps of the close relatives *Arabidopsis lyrata* and *A. thaliana*. Genetics **168:** 1575-1584.

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.

LAN, T. H., T. A. DELMONTE, K. P. REISCHMANN, J. HYMAN, S. P. KOWALSKI, J. M$^{c}$FERSON, S. KRESOVICH, and A. H. PATERSON, 2000 An EST-enriched comparative map of *Brassica oleracea* and *Arabidopsis thaliana*. Genome Res. **10:** 776-788.

LANDER, E. S., J. ABRAHAMSON, A. BARLOW, M. DALEY, S. LINCOLN, and L. NEWBERG, 1987 MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics **1:** 174-181.

LI, G., M. GAO, B. YANG, and C. F. QUIROS, 2003 Gene for gene alignment between the *Brassica* and *Arabidopsis* genomes by direct transcriptome mapping. Theor Appl Genet. **107:** 168-180.

LUKENS, L., F. ZOU, D. LYDIATE, I. PARKIN, and T. OSBORN, 2003 Comparison of a *Brassica oleracea* genetic map with the genome of *Arabidopsis thaliana*. Genetics **164:** 359-372.

LUKENS, L. N., P. A. QUIJADA, J. UDALL, J. C. PIRES, M. E. SCHRANZ, AND T. C. OSBORN, 2004 Genome redundancy and plasticity within ancient and recent *Brassica* crop species. Biol. J. Linnean Soc. **82:** 665-674.

LYNCH, M. and J. S. CONERY, 2000 The evolutionary fate and consequences of duplicate genes. Science **290:** 1151-1155.

MUANGPROM, A. and T. C. OSBORN, 2004 Characterization of a dwarf gene in *Brassica rapa*, including the identification of a candidate gene. Theor Appl Genet. **108:** 1378-1384.

NADEAU, J. H. and B. A. TAYLOR, 1984 Lengths of chromosomal segments conserved since divergence of man and mouse. Proc. Natl. Acad. Sci. USA **81:** 814-818.

O'NEILL, C., and I. BANCROFT, 2000 Comparative physical mapping of segments of the genome of *Brassica oleracea* var. *alboglabra* that are homoeologous to sequenced regions of chromosomes 4 and 5 of *Arabidopsis thaliana*. Plant J. **23:** 233-243.

OSBORN, T. C., C. KALE, I. A. P. PARKIN, A. G. SHARPE, M. KUIPER, D. J. LYDIATE, and M. TRICK, 1997 Comparison of flowering time genes in *Brassica rapa*, *B. napus* and *Arabidopsis thaliana*. Genetics **146:** 1123-1129.

PALMER, J. D., C. R. SHIELDS, D. B. COHEN, T. J. ORTON, 1983 Chloroplast DNA evolution and the origin of amphidiploid *Brassica* species. Theor. Appl. Genet. **65**: 181-189.

PARKIN, I.A.P., A. G. SHARPE, D. J. KEITH, and D. J. LYDIATE, 1995 Identification of the A and C genomes of the amphidiploid *Brassica napus* (oilseed rape). Genome **38:** 1122-1131.

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

PARKIN, I. A. P., D. J. LYDIATE, and M. TRICK, 2002 Assessing the level of collinearity between *Arabidopsis thaliana* and *Brassica napus* for *A. thaliana* chromosome 5. Genome **45:** 356-366.

PARKIN, I. A., A. G. SHARPE, and D. J. LYDIATE, 2003 Patterns of genome duplication within the *Brassica napus* genome. Genome **46:** 291-303.

PRICE, R. A, J. D. PALMER, and I. A. AL-SHEHBAZ, 1994 Sytematic Relationships of *Arabidopsis*: A molecular and Morphological Perspective. In '*Arabidopsis*' (Meyerowitz, E.M. and Somerville, C.R., eds.) Cold Spring Harbor Laboratory Press, pp 7-19.

RAES, J., K. VANDEPOELE, C. SIMILLION, Y. SAEYS and Y.VAN DE PEER, 2003 Investigating ancient duplication events in the *Arabidopsis* genome. J. Struct. Funct. Genomics, **3**:117-129.

RANA, D., T. VAN DEN BOOGAART, C.M. O'NEILL, L. HYNES, E. BENT, L. MACPHERSON, J.Y. PARK, Y.P. LIM and I. BANCROFT, 2004 Conservation of the microstructure of genome segments in *Brassica napus* and its diploid relatives. Plant J. **40**: 725-733.

RYDER, C.D., L.B. SMITH, G.R. TEAKLE and G.J. KING, 2001 Contrasting genome organisation: two regions of the *Brassica oleracea* genome compared with collinear regions of the *Arabidopsis thaliana* genome. Genome. **44**: 808-817.

SCHMIDT, R., A. ACARKAN, and K. BOIVIN, 2001 Comparative structural genomics in the Brassicaceae family. Plant Physiol. Biochem. **39:** 253-262.

SHARPE, A. G., I. A. P. PARKIN, D. J. KEITH, and D. J. LYDIATE, 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*). PhD Thesis, University of East Anglia, UK.

SILLITO, D., I. A. PARKIN, R. MAYERHOFER, D. J. LYDIATE, and A. G. GOOD, 2000 *Arabidopsis thaliana*: a source of candidate disease-resistance genes for *Brassica napus*. Genome **43:** 452-460.

SIMILLION, C., K. VANDEPOELE, M.C. VAN MONTAGU, M. ZABEAU and Y.VAN DE PEER, 2002 The hidden duplication past of *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. U.S.A. **99**:13627-13632.

SONG, K., P. LU, K. TANG, and T. C. OSBORN, 1995 Rapid changes in synthetic polyploids of *Brassica* and its implications for polyploidy evolution. Proc. Natl. Acad. Sci. USA **92:** 7719-7723.

U, N., 1935 Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilisation. Jpn. J. Bot. **7:** 389-452.

UDALL, J., P. QUIJADA, and T. C. OSBORN, 2004 Detection of chromosomal rearrangements derived from homoeologous recombination in four mapping populations of *Brassica napus* L. Genetics **169**: 967-979.

VOORRIPS, R. E., 2002 MapChart: Software for the graphical presentation of linkage maps and QTLs. J. Hered. **93:** 77-78.

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

TABLE 1: Number of loci originating from each *Arabidopsis* chromosome, based on sequence homology, for each *B. napus* linkage group. ND: Not determined.

| | C1 | C2 | C3 | C4 | C5 | ND | Total |
|---|---|---|---|---|---|---|---|
| N1 | 4 | 5 | 14 | 29 | 2 | 6 | 60 |
| N11 | 3 | 6 | 13 | 27 | 7 | 7 | 63 |
| N2 | 14 | 3 | 3 | 4 | 34 | 9 | 67 |
| N12 | 12 | 4 | 4 | 7 | 41 | 2 | 70 |
| N3 | 9 | 19 | 17 | 24 | 36 | 2 | 107 |
| N13 | 19 | 16 | 25 | 17 | 42 | 11 | 130 |
| N4 | 6 | 14 | 6 | 7 | 7 | 2 | 42 |
| N14 | 13 | 33 | 10 | 6 | 13 | 6 | 81 |
| N5 | 13 | 13 | 16 | 1 | 5 | 5 | 53 |
| N15 | 36 | 3 | 16 | 2 | 7 | 4 | 68 |
| N6 | 29 | 4 | 3 | 5 | 21 | 1 | 63 |
| N17 | 17 | 11 | 9 | 23 | 18 | 4 | 82 |
| N7 | 32 | 9 | 7 | 3 | 3 | 1 | 55 |
| N16 | 17 | 0 | 3 | 0 | 3 | 1 | 24 |
| N8 | 20 | 1 | 1 | 17 | 4 | 3 | 46 |
| N18 | 31 | 7 | 12 | 7 | 8 | 7 | 72 |
| N9 | 27 | 10 | 9 | 12 | 17 | 1 | 76 |
| N19 | 10 | 4 | 5 | 8 | 57 | 7 | 91 |
| N10 | 14 | 2 | 3 | 2 | 40 | 6 | 67 |
| Total (%) | 326 (26) | 164 (13) | 176 (13) | 201 (16) | 365 (30) | 85 | 1317 |
| Expected[a] | 313 | 209 | 240 | 187 | 281 | | |

a. Expected number of loci to originate from each Arabidopsis chromosome based on random distribution of loci across five chromosomes with approximate sizes: C1 – 30 Mb; C2 – 20 Mb (not including the NOR region); C3 – 23 Mb; C4 – 18 Mb (not including the NOR region) and C5 – 27 Mb.

TABLE 2: Description of each conserved block found in *Arabidopsis* (refer to Figures 1 and 2).

| Conserved Block[a] | No. of times replicated | No. of comparative markers in Arabidopsis (non-syntenous)[b] | No. of comparative loci in *B. napus*[c] | Total cM coverage in *B. napus* | Length of conserved block (Mb) in *Arabidopsis*[d] |
|---|---|---|---|---|---|
| **C1A** | **7** | **31 (1)** | **88** | **192.1** | **6.68** |
| **C1B** | **7** | **22 (2)** | **54** | **64.7** | **5.32** |
| **C1C** | **7** | **14 (1)** | **28** | **38.1** | **4.12** |
| C1D | 4 | 10 (1) | 18 | 24.1 | 2.49 |
| C1E | 4 | 23 (2) | 42 | 106.6 | 6.07 |
| C2A | 5 | 12 (2) | 19 | 42.1 | 8.71* |
| **C2B** | **6** | **11 (1)** | **24** | **34.15** | **3.42** |
| **C2C** | **6** | **31 (3)** | **65** | **186** | **6.35** |
| **C3A** | **6** | **31 (1)** | **88** | **244.6** | **9.27** |
| C3B[e] | 5-8 | 7 | 13 | 24.75 | 1.66 |
| C3C[e] | 1-4 | 6 | 8 | 4.35 | 3.07 |
| **C3D** | **6** | **18 (1)** | **34** | **36.85** | **4.7** |
| **C4A** | **6** | **11** | **31** | **25.65** | **7.09*** |
| **C4B'** | **6** | **8** | **18** | **5.85** | **1.45** |
| **C4B** | **6** | **35 (2)** | **106** | **244.85** | **8.96** |
| **C5A** | **6** | **44 (3)** | **142** | **245.7** | **7.55** |
| **C5B** | **6** | **7** | **25** | **55** | **1.98** |
| C5C | 4 | 7 (2) | 11 | 27.55 | 3.54 |
| **C5D** | **6** | **16 (1)** | **26** | **45.05** | **2.42** |

| | | | | | |
|---|---|---|---|---|---|
| **C5E** | **6** | **18 (3)** | **52** | **45.35** | **4.32** |
| C5F | 5 | 6 (1) | 14 | 13.15 | 1.99 |
| Totals | 115 (121) | 368 (27) | 906 | 1706.55 | 101.16 |
| (%) | | | | (86.7%) | (85.3%) |

a. Conserved blocks are indicated in Figures 1 and 2, those blocks that are present in at least three copies in each of the diploid *Brassica* genomes are indicated in bold text.

b. Number of comparative loci originating from the conserved block in *Arabidopsis*, that are mapped in a conserved region within the A and/or C genomes of *B. napus*. Loci within the conserved block that have not been mapped to a syntenous position in *B. napus* are indicated in parentheses.

c. Total number of mapped loci within *B. napus* which originate from the conserved block and are found in a syntenous position.

d. Physical length of the designated conserved block in *Arabidopsis* as shown in Figure 2. The complete block may not be represented in each of the duplicate copies. * indicates blocks which include centromeric regions in *Arabidopsis*.

e. Limited marker data in the region flanking the centromere on *Arabidopsis* chromosome three makes it difficult to accurately identify these regions within the *B. napus* genome

FIGURE LEGENDS

FIGURE 1: Genetic linkage map of *Brassica napus*. Linkage groups are arranged according to the regions of primary homology between the A (N1-N10) and C (N11-N19) genomes (Parkin *et al.*, 2003), with cM distances indicated to the left of each group. Each genetic locus is coloured according to the presumed *Arabidopsis* homologue: light blue - chromosome 1; orange - chromosome 2; dark blue - chromosome 3; green - chromosome 4 and red - chromosome 5. Loci in italics were found within conserved blocks based on secondary or tertiary hits within the *Arabidopsis* genome. Loci duplicated within a *B. napus* linkage group are indicated by vertical lines to the right of the group. Identified genome blocks showing conservation of marker content and marker order between the *Arabidopsis* and *B. napus* genomes are shown to the left of each linkage group. Each block is coloured and labelled according to the identified homologous region in *Arabidopsis* (see Figure 2). Inversions identified in *Brassica* relative to Arabidopsis are indicated by arrows. Regions of the *B. napus* genome which have been tentatively aligned with *Arabidopsis* centromeric regions are indicated by hashed blocks.

FIGURE 2: A representation of the *Arabidopsis* genome based on the primary location of each sequenced *B. napus* RFLP marker on the *Arabidopsis* pseudochromosomes (Mb distances are indicated to the right of the chromosomes). Duplicate marker locations are indicated in parentheses. Blocks of markers found to be genetically linked in *B. napus* are indicated by shading and capital letters (A-F). In the majority of cases C4B is conserved as a complete block, but in two instances, on N4 and N14, a small section of the block was observed and is represented by C4B'.

FIGURE 3: Distribution of the physical distance in *Arabidopsis* compared to the genetic distance in *B. napus* for each pair of linked comparative markers found with the conserved blocks.

FIGURE 4: Alignment of segment of *B. napus* linkage group N19 with both *Arabidopsis* chromosome 1 and chromosome 5, highlighting the difficulty in identifying the most related *Arabidopsis* region where there are ancient duplications in the model genome.

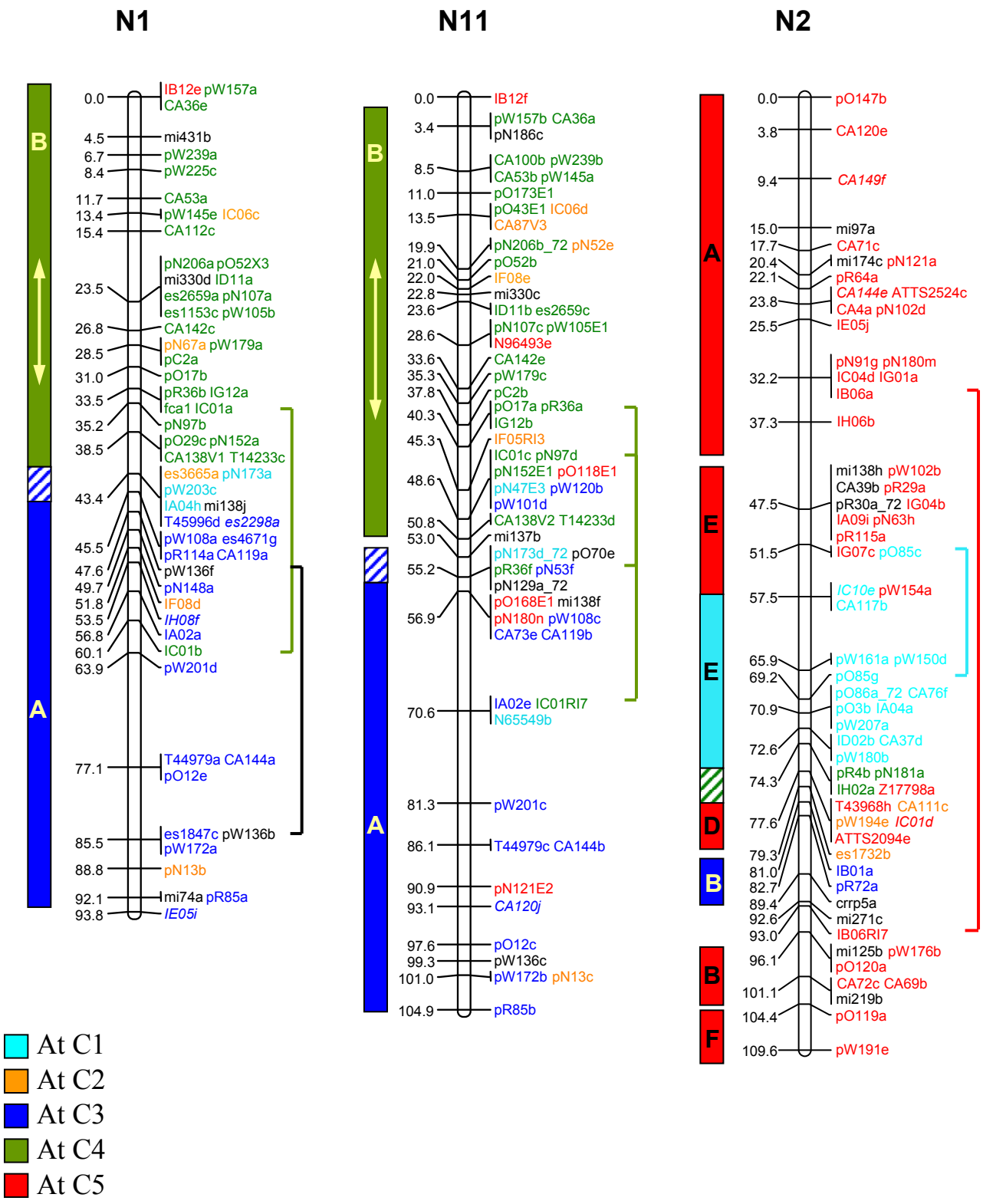



Figure 1_Parkin et al.

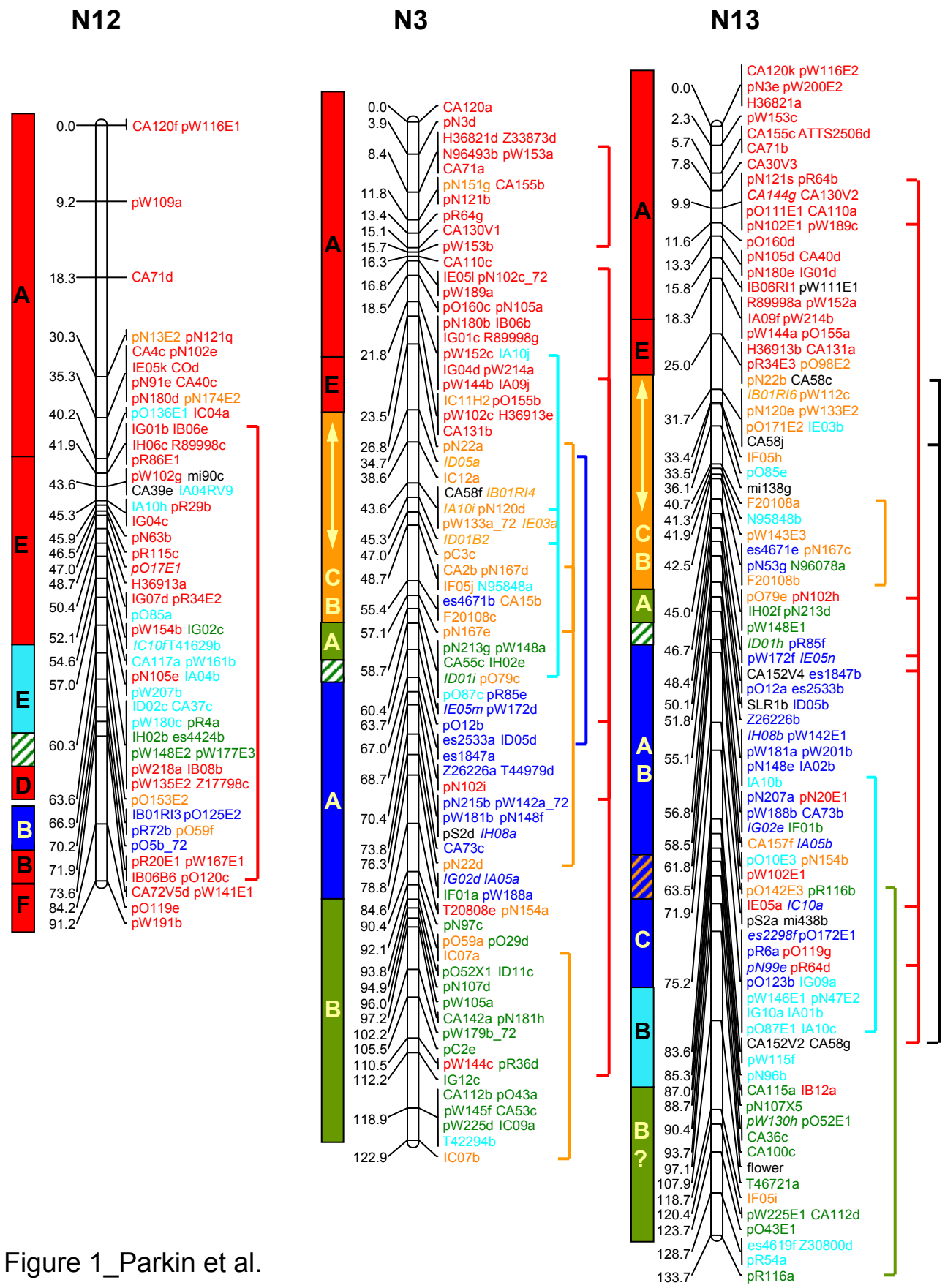



Figure 1_Parkin et al.

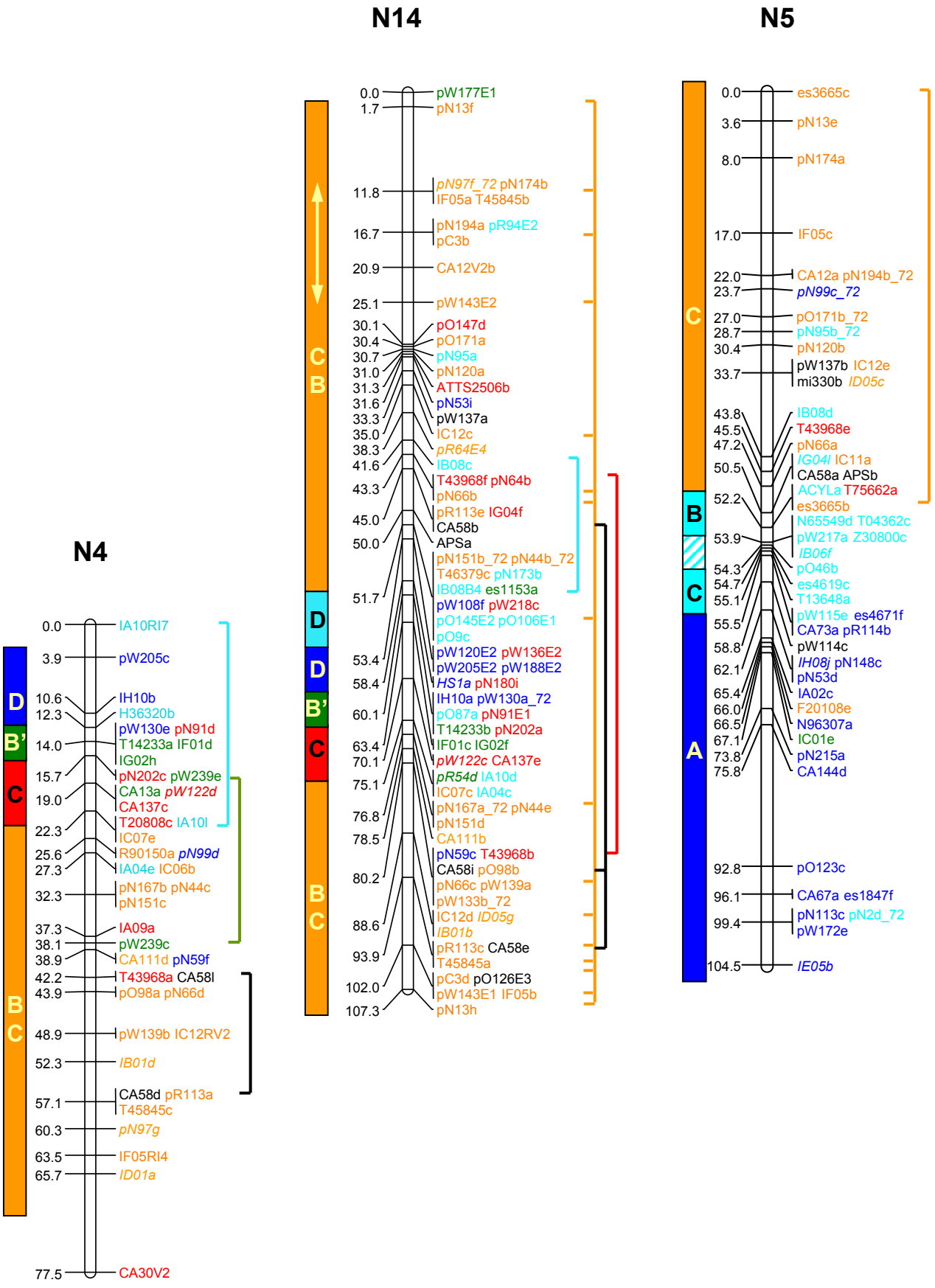



Figure 1_Parkin et al.

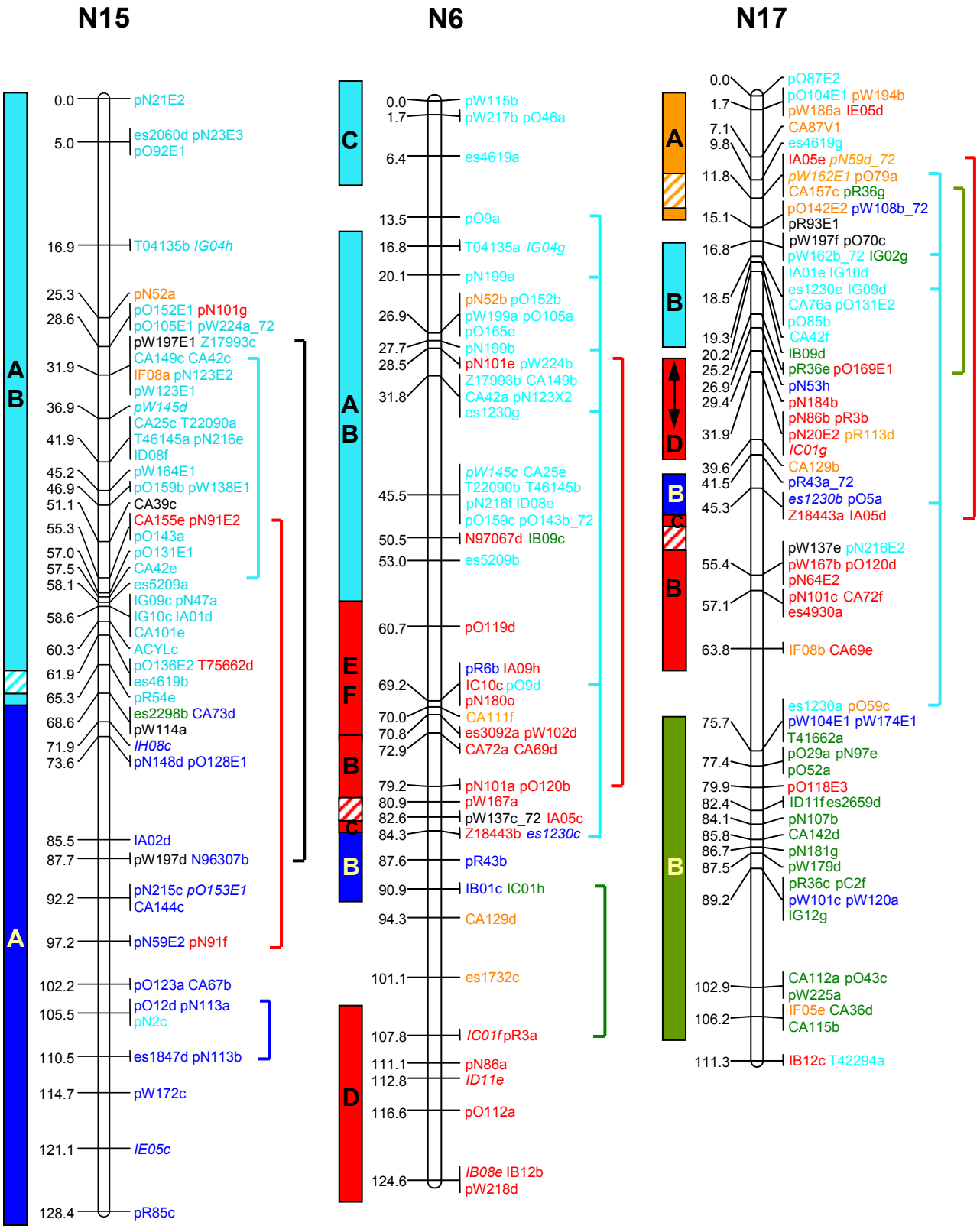



Figure 1_Parkin et al.

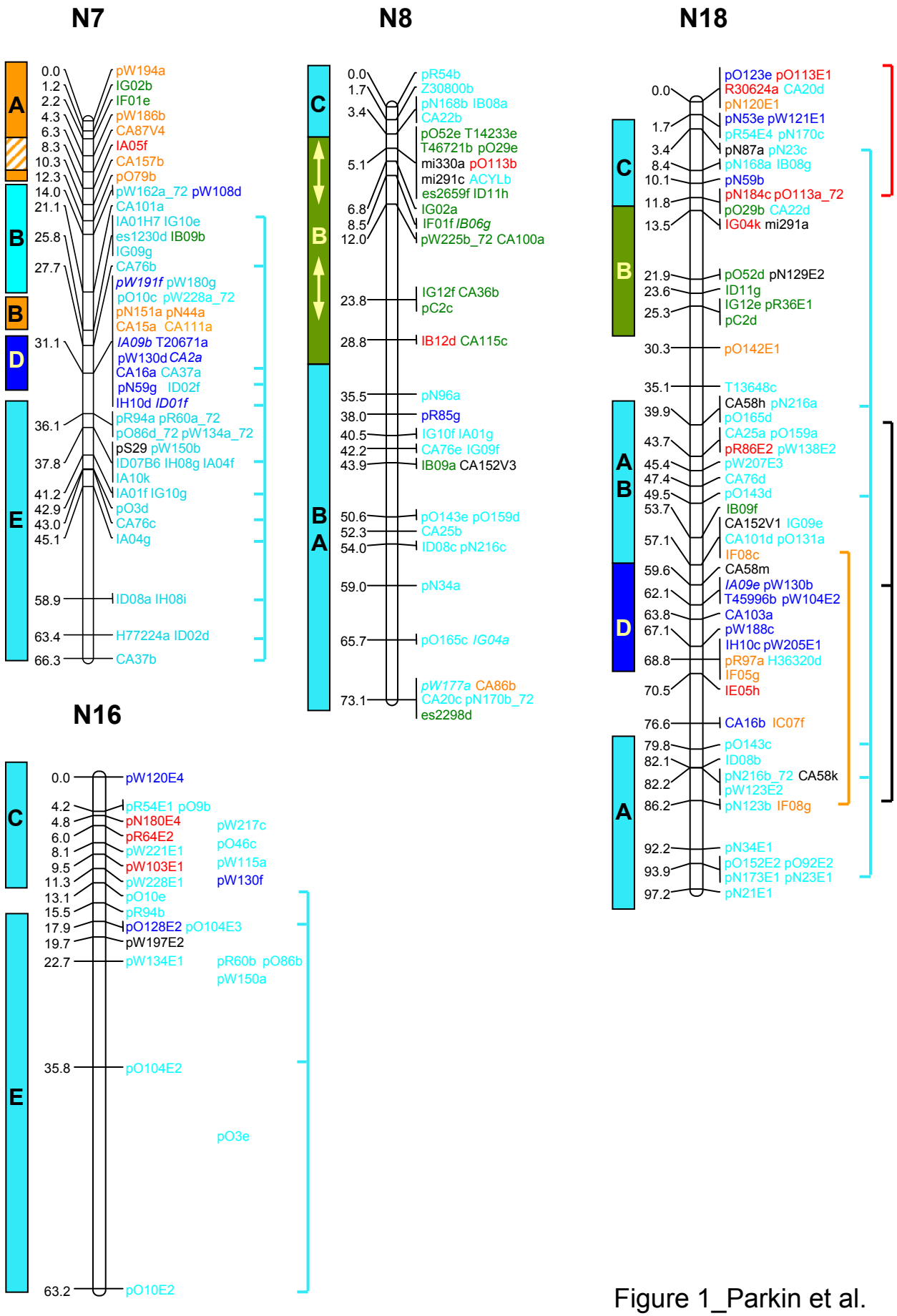



Figure 1_Parkin et al.

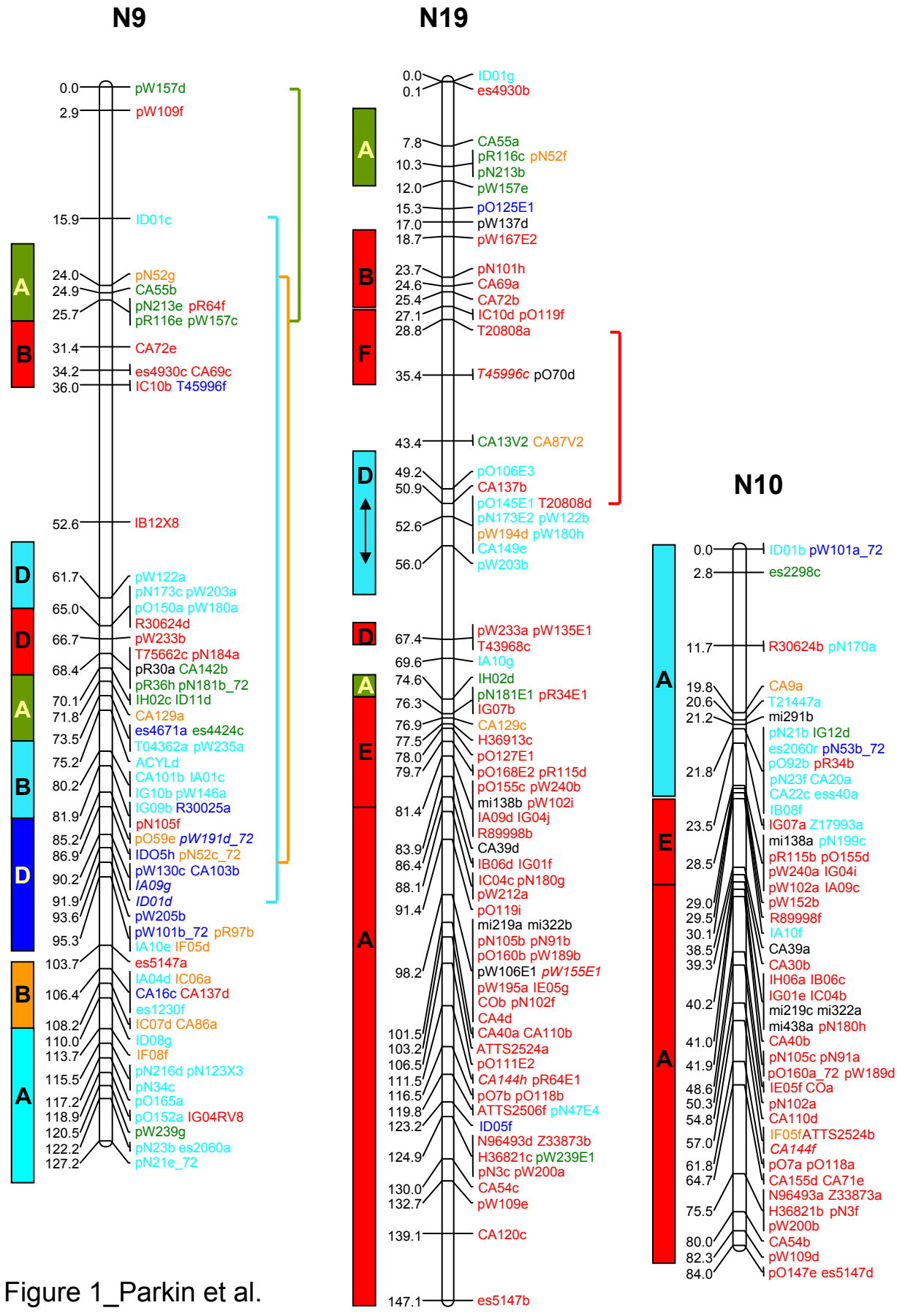



Figure 1_Parkin et al.

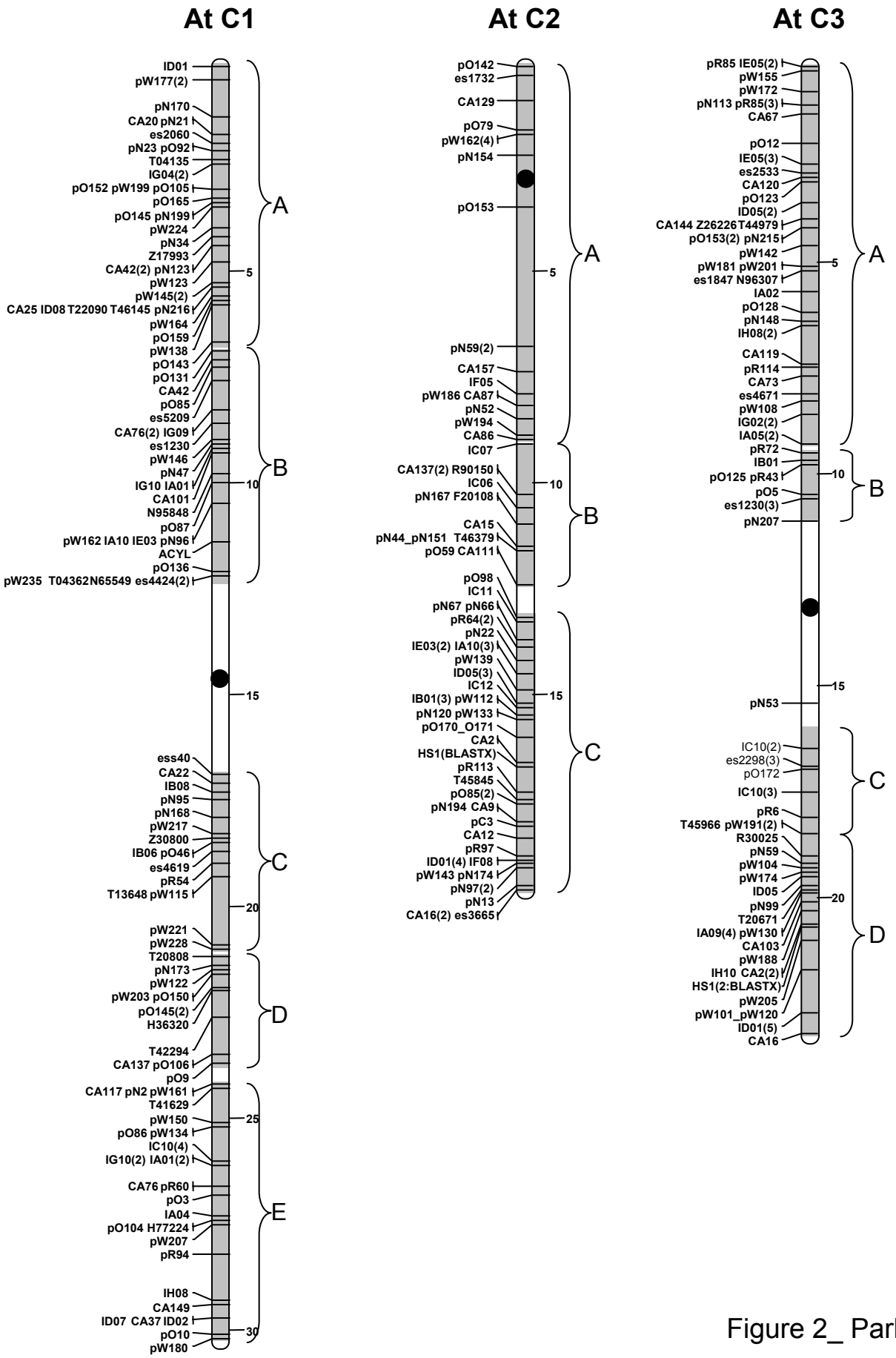



Figure 2_ Parkin et al

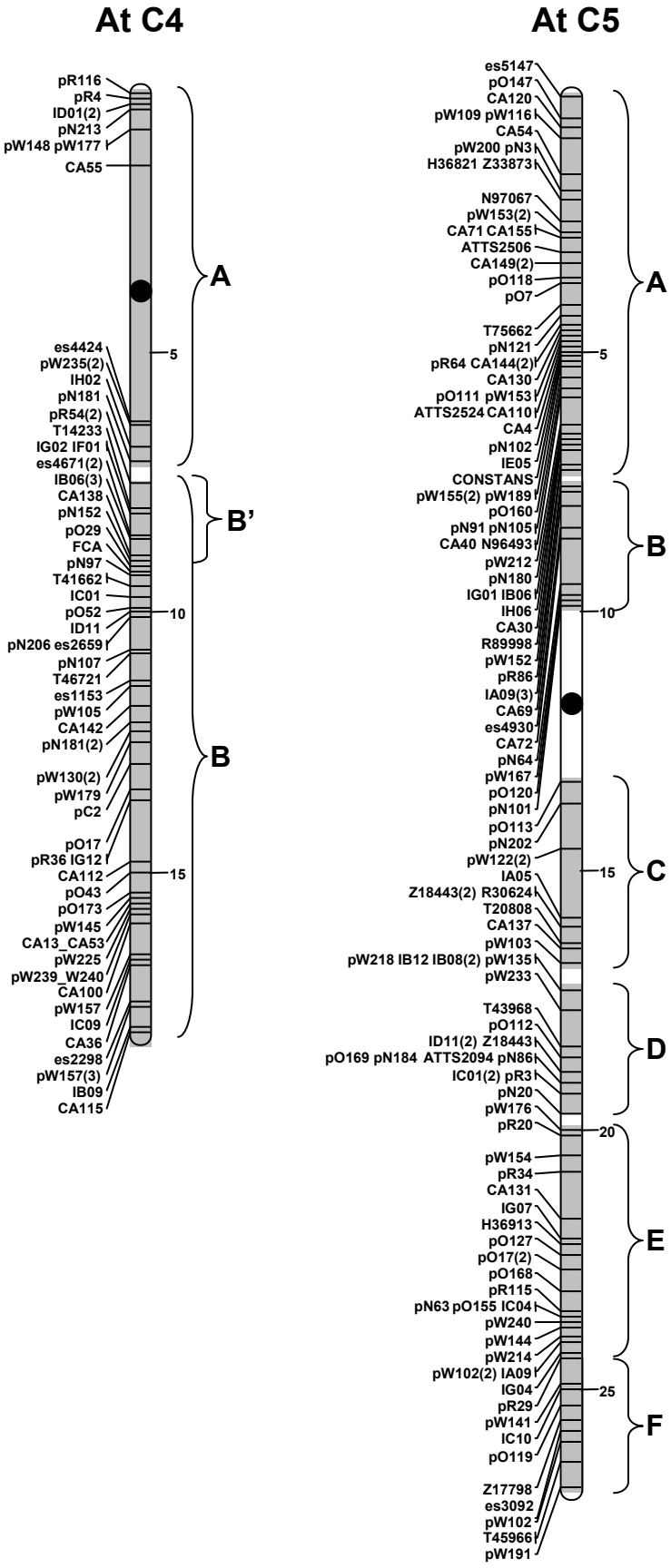
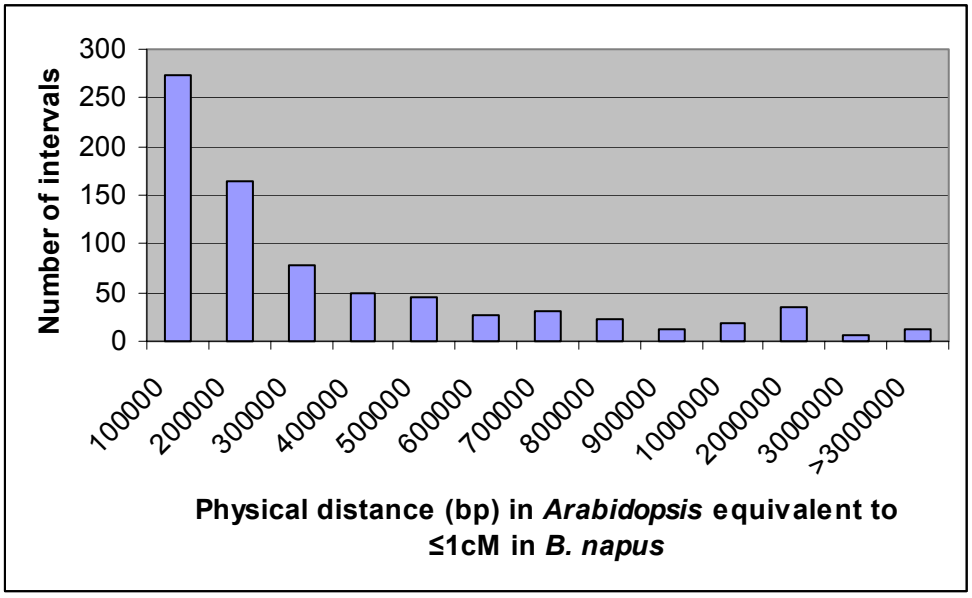



**Figure 2_ Parkin et al**

FIGURE 3.


Number of intervals
300
250
200
150
100
50
0
100000
200000
300000
400000
500000
600000
700000
800000
900000
1000000
2000000
3000000
>3000000
Physical distance (bp) in *Arabidopsis* equivalent to ≤1cM in *B. napus*

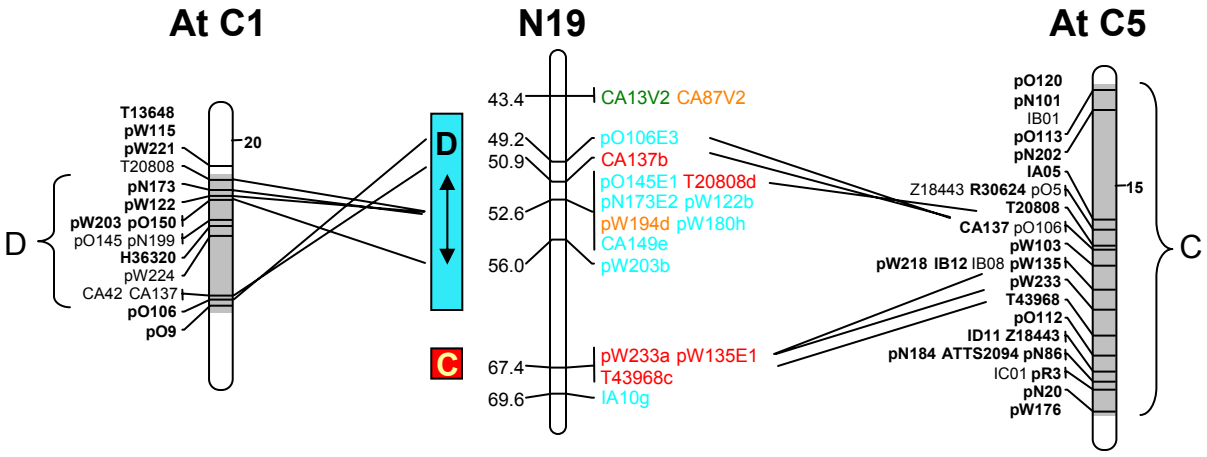



Figure 4_Parkin et al