Comparative Mapping of *Arabidopsis thaliana* and *Brassica oleracea*
Chromosomes Reveals Islands of Conserved Organization

Stanley P. Kowalski,* Tien-Hung Lan,* Kenneth A. Feldmann‡ and Andrew H. Paterson*:

*Department of Soil and Crop Sciences, Texas A&M University, College Station, Texas 77843-2474, ‡Department of Plant Sciences, University of Arizona, Tucson, Arizona 85721, and †Department of Plant and Soil Sciences, University of Delaware, Newark, Delaware 19711

Manuscript received March 7, 1994
Accepted for publication July 11, 1994

ABSTRACT

The chromosomes of *Arabidopsis thaliana* and *Brassica oleracea* have been extensively rearranged since the divergence of these species; however, conserved regions are evident. Eleven regions of conserved organization were detected, ranging from 3.7 to 49.6 cM in *A. thaliana*, spanning 158.2 cM (24.6%) of the *A. thaliana* genome, and 245 cM (29.9%) of the *B. oleracea* genome. At least 17 translocations and 9 inversions distinguish the genomes of *A. thaliana* and *B. oleracea*. In one case *B. oleracea* homoeologs show a common marker order, which is distinguished from the *A. thaliana* order by a rearrangement, indicating that the lineages of *A. thaliana* and *B. oleracea* diverged prior to chromosomal duplication in the Brassica lineage (for at least this chromosome). Some chromosomal segments in *B. oleracea* appear to be replicated, indicating the need for reevaluation of a classical model for Brassica chromosome evolution by duplication. The distribution of duplicated loci mapped for about 13% of the DNA probes studied in *A. thaliana* suggests that ancient duplications may also have occurred in Arabidopsis. The degree of chromosomal divergence between *A. thaliana* and *B. oleracea* appears greater than that found in other congeneric species for which comparative maps are available.

The family Cruciferae comprises 360 genera, organized into 13 tribes (SHULTZ 1936; ROLLINS 1942; AL-SHEHRAB 1973). Basic diploid cytode numbers range from *n* = 7 (*Diplolepis acris*) to *n* = 13 (*Diplolepis harra*), a cytode being defined as a group of taxa sharing a common chromosome complement (HARBERD 1976). A number of allotetraploid cytodes also exist, e.g., *Brassica carinata* (*2n* = 34), *Brassica juncea* (*2n* = 36), *Brassica napar* (*2n* = 38) (U 1935). Nuclear DNA content ranges from 145 million base-pairs (Mbp) per haploid complement (c) in *Arabidopsis thaliana* to 1235 Mbp/c in *B. napar* (ARUMUGANATHAN and EARLE 1991).

In the Cruciferae, numerous challenges complicate accurate botanical classification. Traditional classification based on morphology, although voluminous, has been difficult (HEDGE 1976). In some cases tribal and generic classifications may not accurately reflect true evolutionary relationships (WARWICK and BLACK 1991; HEDGE 1976). Only the classification of the tribes Brassicaceae and Lepidiceae can confidently be viewed as natural, i.e., with clearly defined morphological boundaries which are easily recognized. The remaining 11 tribes have varying degrees of artificial classification, e.g., the tribe Euclidideae, wherein the members seem to have little in common phylogenetically, being recognized by only negligible morphological features of the fruit pods (HEDGE 1976).

*A. thaliana*, *n* = 5 (tribe Sisymbrieae), an extensively utilized model system in plant biochemistry, physiology, and classical and molecular genetics (MEYEROWITZ 1989), is often referred to as a close relative of plants within the genus *Brassica* (tribe Brassicaceae). This relationship is further suggested by extensive conservation of coding sequences between *Brassica* and *Arabidopsis* (LYDIATE et al. 1993). However, the degree of chromosomal divergence between these two genera has not previously been determined.

Genetic linkage maps have been constructed for *A. thaliana* (KOORNEEVE et al. 1983; CHANG et al. 1988; NAM et al. 1989; REITER et al. 1992; HAUER et al. 1993; LISTER and DEAN 1993; MCGRAITH et al. 1993), *Brassica oleracea* (SLOCUM et al. 1990; KIANIAN and QUIROS 1992; LANDRY et al. 1992; KENNARD et al. 1994), and *Brassica rapa* (CHI et al. 1992). The high degree of molecular polymorphism found among the *Brassica* has facilitated restriction fragment length polymorphism (RFLP) mapping within this genus (FIGDORE et al. 1988). Comparative mapping of gene order on the chromosomes of both closely and distantly related species within the Cruciferae could help to clarify the degree of similarity between these various genomes, shed light on macroevolutionary events associated with divergence of these species, and facilitate cross-utilization of genetic resources and molecular tools. Although some comparative analyses between species within the genus *Brassica* are available (*B. oleracea* and *Brassica campestris*, now known as *B. rapa*; SLOCUM 1989; MCGRAITH and QUIROS 1991), global comparisons of genomic structure and organization between more distantly related crucifers, i.e.,...
Brassica and Arabidopsis, have not previously been done.

By applying previously mapped *B. oleracea* DNA probes (Slocum et al. 1990) to *A. thaliana* populations, we have analyzed the relative organization of the chromosomes of *B. oleracea* and *A. thaliana*. Extensive rearrangement distinguishes the chromosomes of *B. oleracea* and *A. thaliana*, although numerous regions of locally conserved linkage and/or homoeology were also apparent. To a lesser extent than *B. oleracea* (Slocum et al. 1990), *A. thaliana* shows evidence of sequence duplication. Our results, together with other results previously published (McGrath et al. 1993) suggest that this may have involved ancient duplication of chromosomes or chromosome segments, in an ancestor of *A. thaliana*.

### MATERIALS AND METHODS

#### Population development:
Two *F₂* mapping populations of *A. thaliana* were used in this study: Wassilewskija (WS) × Hannover/Münden (HM), comprised of 118 individuals, and Wassilewskija (WS) × mutant stock M13 (biological ecotype Landsberg, carrying *an*, leaf/silique phenotype, *dis1*, trichome phenotype, and *er*, inflorescence phenotype), comprised of 111 individuals (Kranz and Kirchheim 1987). Individual plants of the indicated ecotypes were hybridized by hand-crossing, hybridity of the *F₂* verified by RFLP analysis, and the *F₂* selfed to generate *F₃* seed. *F₂* and *F₃* generations were grown in a 16-hr photoperiod and 22°C, in a growth chamber.

#### Molecular markers:
DNA markers used were derived from four sources; prefixes are defined as follows: AC: anonymous newly mapped Arabidopsis cDNAs; BC: an anonymous newly mapped *Brassica* Psfl genomic DNA clones; WS: *A. thaliana* genomic clones previously mapped (Chang et al. 1988); see MATERIALS AND METHODS section for additional information.

#### Probe designations are:
- "cDNA," *A. thaliana* cDNA (i.e., "AC" (25 AC clones have been mapped. The map locations of 12 are presented in Figure 1. The remainder will be presented elsewhere.).
- "PstI" *Brassica* genomic clones "EW," "WG," and "WR" (Slocum et al. 1990); "M," *A. thaliana* genomic clones previously mapped (Chang et al. 1988; see MATERIALS AND METHODS section for additional information.

#### Stringency washes for *Brassica* *PstI* genomic DNA clones applied to *A. thaliana* genomic DNA clones were initially at 1 × SSC, 65°C, with 111 of the clones treated in this manner. However, this tended to result in high background, so stringency was subsequently increased to 0.5 × SSC for the remaining 63 *Brassica* *PstI* genomic DNA clones. Clones washed at 1 × SSC hybridized to an average of 4.83 (±0.40) genomic fragments, while clones washed at 0.5 × SSC hybridized to an average of 4.35 (±0.47) genomic fragments, a nonsignificant difference. Consequently, DNA probes treated by these slightly different procedures were pooled in our analyses of the degree of low copy sequence duplication (Table 2).

#### Data analysis:
Determination of recombination fractions utilized MapMaker (Lander et al. 1987, provided by S. Tinge, DuPont), on a Macintosh Quadra 700. Since some of the DNA markers provided informative polymorphisms in only one of the two populations (HM × WS or M13 × WS; Table 1), it was necessary to assemble a composite map of the *Arabidopsis* genome from their respective *F₂* maps, as described elsewhere (Brown and Grant 1991). Specifically, "anchor loci" segregating in both populations are used to infer the relative order of loci segregating in only one of the two populations.

The extent of conservation between the genomes of *A. thaliana* and *B. oleracea* was estimated by previously published methods (Nadeau and Tator 1984). These methods assume that (1) synteny of two or more markers is evidence of linkage, (2) chromosomal rearrangements fixed during evolution are randomly distributed throughout the genome, (3) crossovers are randomly distributed and (4) the distribution of homologous markers is random and independent. In performing the calculation two adjustments are necessary. Markers generally do not occur at the boundaries of chromosomal segments, so the apparent length of conserved segments on the genetic map (see Figure 1) is an underestimate. Therefore, a statistical correction is done; the expected range (i.e., the expected length of a conserved segment) of a random sample taken from a uniform distribution (i.e., the observed markers in a conserved segment) is determined. The second adjustment accounts for the bias toward longer conserved segments, since segments identified by one marker, or unidentified segments, are omitted. Therefore, a correction is made using a truncated Poisson distribution, which determines the probability that a

### TABLE 1

Total number of DNA probes surveyed/mapped in two populations of *A. thaliana*

<table>
<thead>
<tr>
<th>Loci</th>
<th>WS × HM population</th>
<th>Total</th>
<th>WS × M13 population</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cDNA</td>
<td>PstI</td>
<td>&quot;M&quot;</td>
<td></td>
</tr>
<tr>
<td>No data</td>
<td>0</td>
<td>14</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>0 loci</td>
<td>30</td>
<td>154</td>
<td>38</td>
<td>220</td>
</tr>
<tr>
<td>1 locus</td>
<td>13</td>
<td>25</td>
<td>17</td>
<td>57</td>
</tr>
<tr>
<td>2 loci</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3 loci</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>44</td>
<td>174</td>
<td>55</td>
<td>273</td>
</tr>
</tbody>
</table>

 Probe designations are: "cDNA," *A. thaliana* cDNA, i.e., "AC" (25 AC clones have been mapped. The map locations of 12 are presented in Figure 1. The remainder will be presented elsewhere.). "PstI" *Brassica* genomic clones "EW," "WG," and "WR" (Slocum et al. 1990); "M," *A. thaliana* genomic clones previously mapped (Chang et al. 1988; see MATERIALS AND METHODS section for additional information.

* In three cases (AC71, AC155 and AC184), a marker mapped to a single locus in each population, however the respective loci differed. This yielded three additional duplicated loci (see Figure 1, Table 5).
RESULTS.

were examined for hybridization to EcoRI-digested a lane (based on visual assessment). On average, 52% of bridizing" (approximately 1.51 fragments per probe), bridizing fragments representing less than 10% of signal in

DNA sequence conservation between Arabidopsis and for WS genomic DNA from both Brassica genomic fragments were "strongly hybridizing" (approximately 0.95 fragments per probe) (Table 2).

An additional measure of the extent of DNA sequence duplication was provided by estimating the number of undoubtedly "single-copy" DNA clones in each genome. Clones were designated single copy when each of the five restriction enzyme digests tested showed only one genomic fragment which hybridized to the clone. In B. oleracea, only two (2.5%) of the 80 Brassica PstI genomic DNA clones tested were deemed single copy, by this criterion. One of these (EW7G06) detected an RFLP, and was mapped to chromosome 2 of A. thaliana (Figure 1).

In A. thaliana, 24 (14%) of 174 Brassica PstI genomic DNA clones tested were deemed single copy. Six of these detected RFLPs, four mapping to chromosome 1 (EW7E08, EW7D03, EW8E09, EW9A05), one to chromosome 3 (EW6G12), and one to chromosome 5 (ES5H06) of A. thaliana (Figure 1). Six (14%) of 44 cDNAs tested were deemed "single copy". One of these (AC97) detected an RFLP which mapped to chromosome 2 of A. thaliana (Figure 1). We acknowledge that this underestimates the frequency of single-copy clones in each genome, as the occurrence of restriction sites within the genomic region spanned by the probe will, in some cases, generate two restriction fragments from a single locus.

TABLE 2

<table>
<thead>
<tr>
<th>Hybridization of homologous and heterologous DNA probes to EcoRI-digested genomic DNA from A. thaliana and B. oleracea</th>
<th>Fragments strongly hybridizing</th>
<th>Fragments weakly hybridizing</th>
<th>Total no. of probes tested</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brassica PstI genomic DNA clones</td>
<td>139</td>
<td>120</td>
<td>86</td>
</tr>
<tr>
<td>B. oleracea</td>
<td>165</td>
<td>202</td>
<td>174</td>
</tr>
<tr>
<td>Arabidopsis cDNA clones</td>
<td>58</td>
<td>49</td>
<td>44</td>
</tr>
<tr>
<td>A. thaliana</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*a* "Strongly" and "weakly" hybridizing fragments are defined in RESULTS.

DNA sequence organization in B. oleracea and A. thaliana: Low copy DNA sequence repertoire was largely conserved between A. thaliana and B. oleracea. A subset of 80 PstI-digested Brassica genomic DNA clones (which had been previously mapped; Slocum et al. 1990) were examined for hybridization to EcoRI-digested genomic DNA from both A. thaliana (WS ecotype) and B. oleracea (rapid cycling, self-compatible). Of these, 71 (89%) hybridized to A. thaliana, of which 22 (28%) were placed on the Arabidopsis composite map. Many additional clones were screened only on Arabidopsis, providing 26 additional mapped markers (Figure 1).

Brassica PstI genomic DNA clones and Arabidopsis cDNAs hybridized to similar numbers of Arabidopsis genomic restriction fragments. Forty-four random cDNAs (isolated from an A. thaliana cDNA library) hybridized to an average of 2.43 EcoRI genomic fragments in Arabidopsis, not significantly different from the 2.11 fragments which hybridized to the Brassica PstI genomic DNA clones (Table 2). This reemphasizes the extent of DNA sequence conservation between Arabidopsis and Brassica, at least in hypomethylated regions of the genome, i.e., the regions in which cDNAs and PstI genomic clones occur.

The degree of low copy DNA sequence duplication in the genomes of A. thaliana and B. oleracea was assessed by counting the number of EcoRI genomic fragments hybridizing to each Brassica PstI genomic DNA clone, for WS (A. thaliana) and for a plant derived from several generations of selfing of a self-compatible population of B. oleracea. Brassica PstI genomic DNA clones hybridized to an average of 2.91 genomic DNA fragments in B. oleracea, and 2.11 fragments in A. thaliana (Table 2). EcoRI genomic fragments were designated as "weakly" or "strongly" hybridizing, with "weakly hybridizing" fragments representing less than 10% of signal in a lane (based on visual assessment). On average, 52% of the 250 Brassica genomic fragments were "strongly hybridizing" (approximately 1.51 fragments per probe), and 45% of the 367 Arabidopsis genomic fragments were strongly hybridizing (approximately 0.95 fragments per probe) (Table 2).

An additional measure of the extent of DNA sequence duplication was provided by estimating the number of undoubtedly "single-copy" DNA clones in each genome. Clones were designated single copy when each of the five restriction enzyme digests tested showed only one genomic fragment which hybridized to the clone. In B. oleracea, only two (2.5%) of the 80 Brassica PstI genomic DNA clones tested were deemed single copy, by this criterion. One of these (EW7G06) detected an RFLP, and was mapped to chromosome 2 of A. thaliana (Figure 1).

In A. thaliana, 24 (14%) of 174 Brassica PstI genomic DNA clones tested were deemed single copy. Six of these detected RFLPs, four mapping to chromosome 1 (EW7E08, EW7D03, EW8E09, EW9A05), one to chromosome 3 (EW6G12), and one to chromosome 5 (ES5H06) of A. thaliana (Figure 1). Six (14%) of 44 cDNAs tested were deemed "single copy". One of these (AC97) detected an RFLP which mapped to chromosome 2 of A. thaliana (Figure 1). We acknowledge that this underestimates the frequency of single-copy clones in each genome, as the occurrence of restriction sites within the genomic region spanned by the probe will, in some cases, generate two restriction fragments from a single locus.

Genetic maps of A. thaliana and B. oleracea: The composite RFLP linkage map of A. thaliana is presented in Figure 1. The map spans 187.5 cM on chromosome 1, 92.2 cM on chromosome 2, 91.6 cM on chromosome 3, 118.8 cM on chromosome 4 and 151.4 cM on chromosome 5, for a total recombinational length of 641.5 cM, with an average spacing of 6.1 cM between loci. The map is slightly longer than some previously reported Arabidopsis maps [501 cM, Chang et al. (1988); 493 cM, Nam et al. (1989)], although not significantly different from the longest [630.4 cM, Reiter et al. (1992)]. We note that WS was a common parent in the present map and that of Reiter et al. (1992), suggesting that WS may be more recombinogenic than the more commonly used Landsberg and Columbia ecotypes. The aggregate length of intervals between anchor loci in the composite map was 331.1 cM, approximately intermediate between the two component maps, 402.1 cM (WS × M13) and 296.5 cM (WS × HM).

A total of 110 loci (including 11 duplicated loci and one triplicated locus) corresponding to 97 DNA probes were mapped (Figure 1): 22 loci are "anchors" (mapped in both populations); 28 loci were mapped in the HM × WS F₂ population only; 60 loci were mapped in the M13 × WS F₂ population only. The map shows the linkage arrangement of 49 loci detected by 44 previously mapped Brassica PstI genomic DNA clones (Slocum et al. 1990). Also, four Brassica PstI genomic DNA
Clones not previously mapped are presented ("unknown" designation, Figure 1).

Three criteria were employed in assessing conservation of chromosome organization between *A. thaliana* and *B. oleracea*, and in developing a model (Figure 1, Tables 3–5) which attempts to account for rearrangements which distinguish the two species:

**Criterion 1. Linkage:** Linkage is defined as two or more markers which are linked in *A. thaliana* and also linked in *B. oleracea* and are uninterrupted by markers mapping to other *B. oleracea* chromosomes.

**Criterion 2. Homoeology:** Homoeology is defined as two or more markers linked in *A. thaliana*, which fall on homoeologous regions of *B. oleracea*, and are uninterrupted by markers mapping to non-homoeologous regions of *B. oleracea* chromosomes. Within a region showing evidence of homoeology (e.g., upper end of *A. thaliana* chromosome 3, with *B. oleracea* C3 and C8), some probes mapped to only one of the homoeologous regions. We infer that such cases are explained by lack of an RFLP at the corresponding homoeologous locus (rather than lack of a corresponding locus).

**Criterion 3. Synteny:** Synteny is defined as two or more markers from a particular *B. oleracea* chromosome mapping to a common *A. thaliana* chromosome. Syntenic markers in Arabidopsis were often separated by intervening markers from other *B. oleracea* chromosomes or linkage groups. It was inferred that a single inversion, rather than two translocations, was more likely to be the means by which syntenic markers became separated. This assumption is based on the general observation that closely related taxa more frequently differ by inversions than by translocations [Lycopersicon esculentum and Solanum tuberosum (Bonierbale et al.].
1988; Tanksley et al. 1992); B. oleracea subspecies (Kianian and Quiros 1992); homoeologous chromosomes of Gossypium (Reinisch et al. 1994) and our observation that Brassica probes which are syntenic in A. thaliana tend to be closely linked in B. oleracea (Table 5), also see below. In cases where A. thaliana chromosomes were associated with two or more putatively non-homoeologous regions of the B. oleracea chromosomes, stronger conservation was assumed to be with the B. oleracea chromosome (or homoeologous group) showing a greater number of syntenic loci. The model for proposed rearrangements reflects this conservation, in that lines indicating progressively weaker levels of conservation are drawn progressively further from the chromosome (Figure 1). In cases involving duplicated loci at distal sites on a linkage group (all relevant duplications were on Brassica chromosomes), we inferred conservation (or synteny) to be over the shortest possible distance.

Using these three criteria, we have proposed a model which minimizes the number of rearrangements neces-
When two alternate marker orders were equally likely for A. thaliana and B. oleracea, the model, we have assumed that inversion occurs more readily than translocation (as described above). When two alternate marker orders were equally likely for A. thaliana, the marker order most parsimonious with that in B. oleracea was used.

Figures 1 and 2 illustrate, and Tables 3 and 4 summarize, the degree of conservation between the linkage maps of A. thaliana and B. oleracea. Note that C1-C9 designate B. oleracea linkage groups (chromosomes) as per SLOCUM et al. (1990). The organization of individual chromosomes of A. thaliana relative to B. oleracea is described below. Chromosome numbers separated by a slash are deemed homoeologous in the relevant regions (based on mapping of duplicated loci), e.g., C3/8 refers to a homoeologous region of C3 and C8.

Chromosome 1 of A. thaliana shows association with parts of C1/6, C3/5, C4 and C9 of B. oleracea, implying that a minimum of three translocations differentiate the A. thaliana and B. oleracea chromosomes.

A region spanning the five markers EW7E08, EW8F11, EW2B01, EW7D03 and EW9A06, corresponds to a region of C3 and C5 of B. oleracea. Conservation in this region is inferred based on both linkage and homoeology (criteria 1 and 2).

Linkage (criterion 1) is observed between markers EW2C08 and WG2C07a (C6 in B. oleracea), between markers EW3F01a and EW6C09a (C1/6 in B. oleracea), and tentatively between markers WG3F04 and EW8E09 (C9 in B. oleracea), although this region may be interrupted by a marker (WG4E07) which maps to C3 of B. oleracea. An alternate order placing WG3F04 adjacent to EW8E09 was less likely, but could not be ruled out at a statistically significant level. Since the order shown is preferred by LOD 1.08 we have tentatively accounted for this event with an inversion.

Three inversions are proposed, involving regions corresponding to C3/5, C6 and C9 of B. oleracea (Figure 1).

Chromosome 2 of A. thaliana shows association with parts of C2, C3, C6 and C7 of B. oleracea, implying that a minimum of three translocations differentiate the A. thaliana and B. oleracea chromosomes.

Linkage was detected (criterion 1) between markers WG2C07 and WG2G02, corresponding to C6 of B. oleracea. Although WG2C07 also maps to C7 of B. oleracea, conservation was assumed to be with the more numerous B. oleracea C6 markers (criterion 3).

Two inversions are proposed, involving regions corresponding to C2 and C5 of B. oleracea, as illustrated (Figure 1).

Chromosome 3 of A. thaliana shows association with parts of C2, C3/8, C5 and C9 of B. oleracea, implying that a minimum of three translocations differentiate the A. thaliana and B. oleracea chromosomes.

Linkage (criterion 1) was detected between markers EW2E07, EW5H02 and EW8E11, but differing from their order in B. oleracea by an inversion between EW2E07 and EW5H02 (Figure 2). For each of the three markers involved, an additional locus was mapped in B. oleracea but to three different regions,
possibly representing independent duplications (criterion 3).

Both linkage and homoeology (criteria 1 and 2) were detected for markers EW1D09, EW6G12, EW2C06, EW5F05, EW8F03, EW2D02 and EW4H05, which map to homoeologous regions of B. oleracea C3 and C8. The order of these markers along both B. oleracea homoeologs is conserved, but differs in A. thaliana in that EW8F03 (B. oleracea loci 200 a,c) and EW6G12 (B. oleracea loci 66a,b) co-segregate in B. oleracea (on both homoeologs), but are separated by an interval of 5.7 cM, which includes markers EW2C06 and EW5F05, in A. thaliana. Thus, marker order in these regions is conserved between B. oleracea homoeologs, but different between B. oleracea and A. thaliana (Figure 2). All relevant markers in this region were mapped in a common population (WS × HM), and alternate orders placing EW8F03 (B. oleracea loci 200 a,c) adjacent to EW6G12 (B. oleracea loci 66a,b) were rejected by a LOD of 7.44 or greater. For four of the markers in this region (EW1D09, EW8F03, EW2D03, EW4H05), additional loci were mapped in B. oleracea, but to four different regions, possibly representing independent duplications (criterion 3).

The two cases described above were the only inversions detected.

Chromosome 4 of A. thaliana appears to differ between HM and WS by a reciprocal translocation (KOWALSKI et al. 1994), therefore the map of chromosome 4 is not a composite but from the M13 × WS F2 population (Figure 1). Chromosome 4 shows association with parts of C1, C6 and C7 of B. oleracea, implying that a minimum of three translocations differentiate the A. thaliana and B. oleracea chromosomes.

Linkage (criterion 1) was detected between markers EW7C10 and EW2B10, corresponding to C6 of B. oleracea,
TABLE 6
Distribution of duplicated loci among the chromosomes of *A. thaliana*

<table>
<thead>
<tr>
<th>Chr.</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AC120a-AC120a</td>
<td>AC184a-AC184b</td>
<td>EW4A05b-EW4A05a</td>
<td>M291b-M291a</td>
</tr>
<tr>
<td></td>
<td>W02G07a-W02G07b</td>
<td></td>
<td>AC85b-AC85a</td>
<td>AC85b-AC85c</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>EW6C09b-EW6C09b</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>AC155b-AC155a</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>AC79b-AC79a</td>
<td></td>
</tr>
</tbody>
</table>

and between markers EW7E12 and EW4A05a, corresponding to C1 of *B. oleracea*. Although EW7E12 also maps to C7 of *B. oleracea*, conservation was assumed to be with the more prevalent *B. oleracea* C1 markers (criterion 3). In addition, conservation was assumed to be with locus 21b of *B. oleracea*, based on the proximity of loci (21b-93a spans 32 cM, 21b-93b spans 3 cM; criterion 3).

One inversion is proposed, corresponding to C1 of *B. oleracea* (Figure 1).

Chromosome 5 of *A. thaliana* shows association with parts of C1, C2, C4, C5, C6 and C7 of *B. oleracea*, implying that a minimum of five translocations differentiate the *A. thaliana* and *B. oleracea* chromosomes.

Linkage (criterion 1) was detected between markers EW5D12 and EW6C09b, corresponding to C6 of *B. oleracea*. Although EW5D12 and EW6C09b also map to C5 and C1 (respectively) of *B. oleracea*, conservation was assumed to be with the more prevalent *B. oleracea* C6 markers (criterion 3). A second region of linkage was also detected, between EW4D04 and EW5H06, corresponding to C4 of *B. oleracea*. Although EW4D04 also maps to C1 of *B. oleracea*, conservation was assumed to be with the more prevalent *B. oleracea* C4 markers (criterion 3).

One inversion is proposed, corresponding to C4 of *B. oleracea* (Figure 1).

The total recombinational length of conserved regions between the genomes of *A. thaliana* and *B. oleracea* are 158.2 and 245 cM, respectively (Table 4). This represents 24.6% of the genome of *A. thaliana* (based on our map length of 641.5 cM) and 29.9% of the genome of *B. oleracea* (based on the published length of 820 cM; SLOCUM et al. [1990]).

Chromosomal inversions appear to account for synteny of unlinked markers: We have inferred (see above) that synteny, where two or more markers mapping to different regions of a particular *A. thaliana* chromosome also map to a common *B. oleracea* chromosome, does not occur simply by chance. Rather, we have proposed that inversion is the most likely means by which such markers have become separated (or joined). Previous studies have also suggested this (BONIERBALE et al. 1988; TANKSLEY et al. 1992; KIANIAN and QUIROS 1992; REINISCH et al. 1994). If markers syntenic in *A. thaliana* were conserved with regions of *B. oleracea* chromosomes, one would expect such markers to be close together in *B. oleracea*. Syntenic markers on distal regions of *A. thaliana* chromosomes were much closer together on *B. oleracea*, as measured by recombination (Figure 1, Table 5). This further supports the inference that such markers reflect localized regions of conservation between *A. thaliana* and *B. oleracea*, and that these regions are distinguished by inversions.

Mapping of multiple genetic loci in *A. thaliana*: Of the 97 DNA probes, 12 (12.5%) mapped to more than one locus in *A. thaliana*: 11 to two loci and one to three loci (Table 6). Three DNA probes detect genetically linked duplicated sites on regions spanning 69.3 cM of chromosome 1 and 68.5 cM of chromosome 5 (respectively), and differing in order by an inversion (Figure 3).

DISCUSSION

Chromosomal organization of *A. thaliana* and *B. oleracea*: Although extensive chromosomal rearrangements have occurred since the divergence of *B. oleracea* and *A. thaliana*, islands of conserved organization are discernible. At least one conserved region was detected on each of the five chromosomes (Figure 1, Table 3). In total, we have identified 11 regions spanning 24.6% of the *A. thaliana* genetic map which are closely conserved with 29.9% of the *B. oleracea* genetic map.

Using previously published methods (NADEAU and TAYLOR 1984), we estimate that chromosomal segments with an average length of 21.3 cM in *A. thaliana* are uninterrupted by rearrangements distinguishing them from their order in *B. oleracea*. This calculation predicts that approximately 25 chromosomal rearrangements have occurred since divergence of these two species, at a rate of 2.5 rearrangements per million years. This estimate is in close agreement with our proposed model (26 rearrangements: 17 translocations and nine inversions; Figure 1, Table 4). We conservatively assume that the divergence of *A. thaliana* and *B. oleracea* occurred 10 million years ago [paleopalynological evidence indicates that the plant order Capparales, including the families Capparaceae, Resedaceae and Cruciferae, first appeared during the upper Miocene, approximately 10 million years ago (MULLER 1981, 1984)]. By the same
method, although with a much lower density of genetic markers, NADEAU and TAYLOR (1984) estimated that chromosomal segments with an average length of 8 cM had been conserved between the mouse and human genomes, and that 180 rearrangements had occurred since divergence of the human and mouse genomes (approximately 70 million years ago).

Despite concordance between the number of rearrangements observed herein and the number predicted by the method of NADEAU and TAYLOR (1984), it is likely that mapping of additional DNA probes will define a few more rearrangements. The relationship between the number of homologous markers mapped and the number of conserved segments identified approximately follows an asymptotic regression \[ W = A - B(e^{-c}) \] (NADEAU 1989; SNEDECOR and COCHRAN 1989). With continued mapping of homologous markers between A. thaliana and B. oleracea, it should be possible to clearly determine the actual number of rearrangements which distinguish the chromosomes of these organisms.

Relative to other plant species for which equivalent comparisons can be made, the chromosomes of B. oleracea and A. thaliana appear to have diverged relatively rapidly. The genomes of rice and maize, which diverged approximately 50 million years ago (BENNETZEN and FREELING 1993) have 32 conserved linkage segments along their 12 and 10 chromosomes (respectively), comprising 70% of the rice genome (AHN and TANKSLEY 1993), based on genetic mapping at moderately higher density of common markers than that reported here (markers averaging 8.5 cM apart, vs. 14.6 cM in our study). The A and D genomes of cotton, which diverged approximately 6–11 million years ago (WENDEL 1989) are distinguished by one translocation and six inversions along 11 (of an expected 13) homologous chromosome pairs for which sufficient data is available (REINISCH et al. 1994). The genomes of maize and sorghum, which diverged approximately 20 million years ago (BENNETZEN and FREELING 1993) are distinguished by at least nine inversions along the 10 chromosomes, although analysis of several chromosomes is incomplete (WHITkus et al. 1999). The minimum number of 26 rearrangements which we estimate to distinguish the five chromosomes of A. thaliana from the nine chromosomes of B. oleracea suggests a level of rearrangement paralleled only by the rice-maize comparison, although rice and maize diverged at least 40 million years (5X) earlier.

**Ancient duplication of Arabidopsis chromosome segments:** Based on the degree of gene and sequence duplication, as well as evidence for duplication of chromosomal segments, A. thaliana appears to have undergone ancient duplication of chromosomes or chromosome segments. Considerable duplication of individual sequences in A. thaliana is revealed by both Brassica PstI genomic DNA clones, and Arabidopsis cDNAs, with each showing an average of two EcoRI fragments (Table 2). A minimum estimate of the frequency of DNA sequence duplication in A. thaliana can be obtained from the frequency of DNA probes detecting RFLPs at two (or more) unlinked loci, which account for about 12.4% of the DNA probes we mapped. McGrath et al. (1993) previously reported "that more than 15% of the genes in the A. thaliana genome may be encoded by multiple loci." However, such estimates are confounded with levels of DNA polymorphism in particular mapping populations. If the likelihood of detecting an RFLP at one locus is \( x \), the likelihood of detecting RFLPs at each of two unlinked loci is \( x^2 \). At low levels of \( x \), \( x^2 \) will approach zero, and it will rarely be possible to map duplicated loci (CHITTENDEN et al. 1994) (Table 1). A maximum estimate for the frequency of duplicated loci can be obtained from the frequency of DNA probes which detect only one genomic fragment in digests with several different restriction enzymes—we estimate this frequency at 14% (see RESULTS). In our study, mapped duplicate loci are found on each chromosome, with no cases of proximal duplication found (Figure 1, Table 6).

Although modest levels of DNA polymorphism among Arabidopsis ecotypes (KING et al. 1993) make it difficult to study possible chromosomal duplications, we did find some evidence supporting at least one such event. Despite the suggestion that many DNA probes may be duplicated in A. thaliana, only a small fraction could be mapped to two loci (Table 1). Nevertheless, we identified one region of chromosome 1 which may be homoeologous with a region of chromosome 5 (Figure 3).

Close inspection of previously published results provides independent corroboration of an ancestral chromosome (or segment) duplication in Arabidopsis. McGrath et al. (1993) report that three DNA probes detect RFLP loci duplicated on Arabidopsis chromosome 1 ([515A, 559B, 711], spanning 9.7 cM), and chromosome 5 ([559A, 515B, 711B, spanning 108.9 cM). Each of the duplicated regions are close to the respective homoeologous regions on our map, based on anchor loci common to chromosome 1 ([M235, M213],

---

**Figure 3:** Segmental duplication of chromosomes 1 and 5 in A. thaliana. Distances are in centiMorgans. Loci connected by a line are detected by the same DNA marker.
and tightly linked reference loci (separated by 1.5 cM on the CHANG et al. (1988) map, on chromosome 5 (M331 on the MCGRATH et al. map; M268 on our map)). In the MCGRATH et al. (1993) map, putative homoeologous regions of the genome may also be present between regions on chromosomes 2 (579B, 415E, 574B, 173A, 415A; spanning 31.3 cM) and 3 (574A, 173B, 415C, 579C; spanning 53.1 cM), and between regions on chromosomes 3 (415C, 579C, 713C; spanning 58.4 cM) and 4 (713B, 579A, 415F, 713D, 415B; spanning 52.8 cM). It must be noted that duplicate loci which contradict both our evidence and that of MCGRATH et al. (1993) have also been reported, with duplications between chromosomes 1 (m281a, g2488b; spanning 18.7 cM) and 3 [g2488a, m281a; spanning 48.5 cM; HAUGE et al. (1993)]. However, the bulk of the evidence suggests that chromosome 1 contains a segment having undergone duplication, possibly including sequences which have been rearranged, or duplicated by other mechanisms such as replicative transposition (VOYTAS and AUSUBEL 1988). The proposal that A. thaliana is a “paleopolyploid” is consistent with like proposals for several other species which show strict bivalent pairing at meiosis (based on cytological evidence) that the earliest number of x and diploid cotton (REMSCH et al. 1994) suggests (based on cytological evidence) that the earliest chromosomal evolution proceeded in both ascending and descending progressions of basic chromosome numbers, e.g., from 6 to 5 to 4, etc., and from 7 to 6 or 7 to a lower chromosome number, followed by (at least) segmental duplication.

Consequences of ancestral duplication of Arabidopsis chromatin for physical mapping and chromosome walking: Beyond its evolutionary consequences, duplication in the A. thaliana genome has ramifications for molecular manipulations of large DNA. Duplication of large genomic regions could dramatically complicate long range restriction mapping and physical mapping in these regions, as is the case in polyploids (REINSCH et al. 1994). Some cases of apparently chimeric YACs, i.e., a YAC with distal ends mapping to unlinked regions of the genome, may really be a result of genomic duplication. Based on estimates that 15% (MCGRATH et al. 1993) to 12.5% (herein) of Arabidopsis low copy DNA may occur at two or more sites with discernible homology, a like fraction of corresponding YAC ends would be expected to detect RFLPs at different (putatively homoeologous) sites. While it is well established that megabase DNA cloning is subject to chimeras, estimates of chimeric frequency based upon RFLP mapping may be inflated as a result of ancestral duplication of Arabidopsis chromatin. As (putative) regions of ancestral duplication in Arabidopsis are better delineated by additional mapping, it will be easier to distinguish “truly chimeric” YACs from artifacts resulting from ancient duplications.

Chromosomal divergence among the Cruciferae: Arabidopsis and Brassica diverged from a common ancestor with less chromosomal duplication than B. oleracea. The relative orders of DNA markers along homoeologous chromosomal regions permit us to infer whether specific chromosomal rearrangements predate, or postdate, duplication of Brassica chromosomes. The segment of chromosome 3 of A. thaliana spanning markers EW1D09, EW6G12, EW2C06, EW5F05, EW8F03, EW2D03 and EW4H05 displays nearly complete linkage conservation with homoeologous regions on C8 and C3 of B. oleracea, with the exception of EW6G12 (66a,b) and EW8F03 (200a,c). Although these markers co-segregate on both C8 and C3 of B. oleracea (C8: 66b/200a; C3 66a/200c), they are separated by a distance of 5.7 cM, and two other markers, in A. thaliana. The simplest explanation for this would be that the prototypical B. oleracea and A. thaliana chromosomes differed by a rearrangement in this region, and that chromosomal duplications then propagated this region in B. oleracea, i.e., the rearrangement predates duplication of the Brassica chromosomes. SADOWSKI et al. (1994) reported a complex of three tightly linked genes in A. thaliana, mapping to a single locus. Each of these probes maps to duplicated loci in B. oleraceae, co-segregating on one homoeolog, but with duplicated sites dispersed over three chromosomes. The simplest explanation for this would be that the prototypical B. oleracea and A. thaliana chromosomes showed close linkage of these markers, and that one B. oleracea homoeolog has been rearranged subsequent to duplication.

Based on cytological evidence, the fundamental number of chromosomes in the genus Brassica has been suggested to be n = 6 (PRAKASH and HINATA 1980). Cyto- logical (ROBBELEN 1960), and more recently molecular evidence (SONG et al. 1990), suggests that the evolution of Brassica and closely related genera has proceeded in ascending order of chromosome number, from the n = 6 ancestor to n = 7 (Diplotaxis erucoides), to n = 8 (hypothetical “bridge” species), to n = 9 (B. oleracea). This model proposes that B. oleracea is a secondary polyploid, with a basic chromosomal complement of AABBCCDEF or ABBCCDEEF (HAAG 1988; ROBBELEN 1960). Linkage mapping of B. oleracea has supported evidence for the existence of duplicated chromosomal
segments, with nearly half of the DNA probes mapping to two or more loci, many of the duplicated loci being clustered on specific pairs of linkage groups. However, homoeology spanning entire pairs of linkage blocks was not found, indicating the occurrence of translocations during the evolution of *B. oleracea* (Slocum et al. 1990). Comparative mapping of the genomes of *B. oleracea* (*n* = 9) and *B. campestris* (*n* = 10) revealed predominant conservation of gene order along the chromosomes (Slocum 1989), suggesting that most rearrangements differentiating Arabidopsis from Brassica occurred before these Brassicas diverged.

Our results, in conjunction with previously published results (Slocum et al. 1990), suggest that some regions of the *B. oleracea* genome have been triplicated. Chromosome 3 of *A. thaliana* (in the region spanning markers *EW1D09, EW6G12, EW2C06, EW5F05, EW5F03, EW2D03, EW4H05*) is clearly homoeologous to both C8 and C3 of *B. oleracea*; however, C3 and C1 of *B. oleracea* also appear to be homoeologous in the same region [Figure 2, also Slocum et al. (1990)]. If chromosomal segments of C1, C3 and C8 are homoeologous (e.g., triplicated), the model which postulates *B. oleracea* as a secondary polyploid, with a basic chromosomal complement of AABBCCDEF or ABBCCDEEF (Haga 1938; Robbeelen 1960) becomes inadequate to explain the evolution of *B. oleracea* chromosomes. An alternate hypothesis to triplication would invoke paleopolyploidy, suggesting that the close association between Brassica C3 and C8 is the result of a recent duplication, and the association between Brassica C1 and C3 is ancient. Neither of these hypotheses satisfactorily explain why no loci were found with duplicated RFLPs on C1 and C8, however this could occur just by chance failure to detect the appropriate RFLPs.

Kianian and Quiros (1992) proposed that the high level of duplications and chromosomal rearrangements in the genome of Brassica might impart enhanced "flexibility to change and evolve." Our suggestion, that Arabidopsis and Brassica appear to have diverged at the chromosomal level relatively rapidly, support this proposal. We note that the Cruciferae has a center of diversity in the temperate zone of the northern hemisphere, with the main radiation center proposed to have been from the eastern portion of the Irano-Turanian phytogeographical region (Hedge 1976). The appearance of the Cruciferae is coincident with the advent of glaciation and major cooling, approximately 10 million years ago (Traverse 1988). Therefore, this family evolved in an environment characterized by rapid climatic changes, with alternating glacial-interglacial cycles. It is tantalizing to speculate that the apparent plasticity of the Cruciferae genome (numerous duplications and rearrangements) may have conferred a selective advantage in taxa subjected to dramatic (at least, on an evolutionary time scale) changes in climate (McClintock 1984). Alternately, physical subdivision of populations as a result of glaciation may simply have facilitated fixation of rare chromosomal rearrangements in small local populations.

**Utility of a comparative map of the Cruciferae:** Comparative maps of *A. thaliana* and *B. oleracea* permit cross-utilization of tools and resources which have been developed for each, a particular boon to the Brassica research community. Map-based cloning of orthologous genes may be much easier in *A. thaliana* than in *Brassica* spp., due to the small genome size and low level of repetitive DNA present in its genome (Meyerowitz 1989). One case in which a Brassica cDNA was used to complement an Arabidopsis mutation is already published (Arondel et al. 1992). More detailed fine-scale comparative mapping of these related crucifers may facilitate use of *A. thaliana* YAC islands (Schmidt et al. 1992) for positional cloning in *Brassica*. Finally, additional mapping of heterologous markers between *A. thaliana* and *B. oleracea* will provide much new basic information, helping us to understand the process of evolution in greater detail.

Contemporary molecular tools now permit detailed studies of "chromosomal archaeology." Stebbins (1966) had foreseen these advances nearly three decades ago, when he suggested that "The opportunities for profitable investigations of this sort are by no means at an end, and new techniques may extend them to degrees of clarity and certainty which at present can hardly be imagined."

We thank R.L. Scholl for the HM and WS Arabidopsis seed, E. Meyerowitz, J. Braam and Pioneer Hi-Bred Production Ltd. for DNA clones, Clontech Inc. for an Arabidopsis cDNA library, and L. Curtenden, J. Dong, N. Forthoezel, C. Lan and R. Wing for technical assistance. This research was funded by the Texas Agricultural Experiment Station (REP Award 111225-8108 to A.H.P.), Texas Higher Education Coordinating Board Science Foundation grant DME9108442 to A.F., and the Delaware Agricultural Experiment Station (to A.H.P.). Probes designated M are available through the Arabidopsis Genetic Stocks Center at Ohio State University. Probes designated AC will be placed in the Arabidopsis Genetic Stocks Center, and requests can be directed there. Requests for Brassica probes should be directed to Pioneer Hi-Bred Production Ltd.

**LITERATURE CITED**


world cruciferae, pp. 83-103. 1983

Assessment of the degree of restriction fragment length polymorphism in Brassica oleracea, pp. 1028-1032.


Stearn, J. L., 1966 Chromosomal variation and evolution; polyploidy and chromosome size and number shed light on evolutionary processes in higher plants. Science 152: 1463-1469.


Communicating editor: M. R. HANSON

510 S. P. Kowalski et al.


Landry, B. S., N. Hubert, R. GEGE, M. S. CHANG, S. E. LINCHELON et al., 1992 A genetic map for Brassica oleracea based on RFLP markers detected with expressed DNA sequences and mapping of re- sistance genes to race 2 of Plasmopora brassicae (Borrisoni). Genome 35: 409-420.


