Comparative genetic mapping of allotetraploid cotton and its diploid progenitors

C.L. Brubaker, A.H. Paterson, and J.F. Wendel

Abstract: Allotetraploid cotton species (Gossypium) belong to a 1–2 million year old lineage that reunited diploid genomes that diverged from each other 5–10 million years ago. To characterize genome evolution in the diploids and allotetraploids, comparative RFLP mapping was used to construct genetic maps for the allotetraploids (AD genome; \( n = 26 \)) and diploids (A and D genomes; \( n = 13 \)). Comparisons among the 13 suites of homoeologous linkage groups permitted comparisons of synteny and gene order. Two reciprocal translocations were confirmed involving four allotetraploid At genome chromosomes, as was a translocation between the two extant A genome diploids. Nineteen locus order differences were detected among the two diploid and two allotetraploid genomes. Conservation of colinear linkage groups among the four genomes indicates that allopolyploidy in Gossypium was not accompanied by extensive chromosomal rearrangement. Many inversions include duplicated loci, suggesting that the processes that gave rise to inversions are not fully conservative. Allotetraploid At and Dt genomes and the A and D diploid genomes are recombinationally equivalent despite a nearly two-fold difference in physical size. Polyploidization in Gossypium is associated with enhanced recombination, as genetic lengths for allotetraploid genomes are over 50% greater than those of their diploid counterparts.

Key words: restriction fragment length polymorphism (RFLP), Gossypium, evolution, polyploidy.

Introduction

Polyploidy is common in plants (Grant 1981; Lewis 1980; Leitch and Bennett 1997; Masterson 1994; Soltis and Soltis 1993; Stebbins 1950, 1971) and probably has been involved in the evolution of all eukaryotes (Leipoldt and Schmidtke 1982; Sidow 1996; Spring 1997; Wolfe and Shields 1997). In synthetic allopolyploids, illegitimate chromosome pairing often disturbs meioses, and thus it is thought that evolutionary mechanisms which promote exclusively bivalent pairing will be selectively favored in the critical first few generations following natural allopolyploid formation. However, prior to the restoration of diploid-like chromosome behavior, interactions among homoeologous chromosomes may result in the rapid accumulation of structural rearrangements that alter gene order and synteny (Ahn et al. 1993; Feldman et al. 1997; Leipoldt and Schmidtke 1982; Liu et al. 1998; Song et al. 1995).

The cotton genus, Gossypium L., is an ideal system for examining genome evolution in polyploids. Gossypium com-
prises approximately 45 diploid and five allopolyploid species distributed throughout the arid and semi-arid regions of Africa, Australia, Central and South America, the Indian subcontinent, Arabia, the Galápagos, and Hawaii (Fryxell 1979, 1992). The diploid Gossypium species fall into eight cytological groups, or genomes, designated A through G. And K (Beasley 1940; Edwards and Mirza 1979; Endrizzi et al. 1985; Phillips and Strickland 1966; Stewart, 1994). The five allopolyploid species, indigenous to the New World, derive from a single allopolyploidization event that united the Old World A genome with the New World D genome, in an A genome cytoplasm (Wendel 1989; Wendel and Albert 1992).

Judging from differences in meiotic pairing, Gossypium allopolyploids appear to be more fully “differentiated” from one another than are the descendants of their diploid progenitors (Endrizzi 1962; Mursal and Endrizzi 1976). Specifically, allopolyploid-derived haploids form an average of less than one bivalent per cell during meiosis, whereas chromosomes of extant A and D diploid F1’s average 5.8 and 7.8 bivalents at metaphase I (Endrizzi and Phillips 1960; Mursal and Endrizzi 1976; Skovsted 1937). These data suggest that stabilization of the newly evolved allopolyploid involved genic or genomic modifications that inhibit homeologous pairing while promoting exclusively bivalent formation between homologues (Kimber 1961).

To examine whether genome evolution in Gossypium allopolyploids was also accompanied by chromosome rearrangements, we constructed RFLP maps for the A and D diploid genomes, using a common set of nuclear probes, most of which were previously or simultaneously mapped in the allopolyploids (Reinisch et al. 1994). Comparisons between diploid genomes, between diploid genomes and their corresponding allopolyploid genomes, and between allopolyploid genomes allowed us to address whether allopolyploidization in Gossypium was accompanied by rapid chromosomal structural change and compare recombination rates among four homoeologous genomes.

Materials and methods

The A and D genomic linkage maps were based on the multilocus genotypes of 58 A genome (G. herbaceum L. = A1 –97, × G. arboreum L., A2 –47) and 62 D genome (G. trilobum (Moc. & Sessé ex. DC) Skovsted × G. raimondii Ulbr.) F2 individuals. The AD genomic linkage map was based on a G. hirsutum race ‘palmeri’ × G. barbadense K101 F2 population (see Reinisch et al. 1994 for details). DNA extractions followed Paterson et al. (1993) or Brubaker and Wendel (1994). Restriction digestion, electrophoresis, blotting procedures, RNA probe preparation, hybridization protocols, and autoradiography followed Brubaker and Wendel (1994).

Three hundred and twenty-one mapped DNA probes were selected from six libraries: anonymous PurI nuclear fragments from G. raimondii ("G"-series; n = 75), G. herbaceum subsp. africanum (Mauer) Watt ("A"-series; n = 143), and G. hirsutum cultivar TM-1 ("M"-series (n = 1) and "P"-series (n = 16)); low copy-number nuclear sequences (Zhao et al. 1996) from G. barbadense L. cv. Pima S6 ("PX"-series; n = 8); and anonymous cDNA clones from drought-stressed G. hirsutum accession T25 ("pAR"-series; n = 78) (Reinisch et al. 1994). RFLPs between G. herbaceum and G. arboresum and between G. raimondii and G. trilobum were revealed using eight (EcoRI, HindIII, PstI, DraI, CfoI, BamHI, Xbal, and EcoRV) and two restriction enzymes (EcoRI and HindIII), respectively, the difference reflecting the relative levels of polymorphism between the A genome parents and the D genome parents. We also assayed allelic segregation among the 58 A genome F2 progeny at seven isozyme loci encoded by six enzyme systems: aconitate hydratase (ACO1), arginyl-specific aminopeptidase (ARG1), leucyl-specific aminopeptidase (LEU1), 6-phosphogluconate dehydrogenase (PGD1 and PGD3), triosephosphate isomerase (TPI1). Sample preparation, starch-gel electrophoresis, and nomenclature follow Wendel et al. (1989).

Genetic interpretation of the RFLP phenotypes and nomenclature followed the system of Reinisch et al. (1994). Loci are designated by probe, and arbitrarily assigned lower-case letters to distinguish multiple segregating loci revealed by the same probe. The letters A and D denote the A and D genome linkage groups, respectively, followed by arbitrarily assigned numbers. MapMaker Macintosh 2.0 (Lander et al. 1987) was used to infer the genomic linkage maps (Figs. 1–11). Initial linkage groups were inferred from two-point analyses using a minimum LOD score of 4.0 and a maximum theta value of 0.40. Linkage groups were ordered by selecting a subset of six or seven loci whose most likely order differed from the next most likely order by a LOD of two or more. Remaining loci were added to the initial map using the “try” function. The “ripple” function confirmed the local order around new loci and identified regions of uncertain order. Linkage-1 (Sutier et al. 1983) was used to identify loci whose segregation ratios differed significantly (P < 0.05) from Mendelian expectations.

Results

A genome map

The A genome map (Figs. 1–11) comprises 161 loci encoded by 152 nuclear probes and 6 isozyme loci mapping to 18 linkage groups (hereafter LGs). The 18 LGs encompass 856 cM. Individual LGs include 2 to 24 loci (average = 8.94).

There was no discernible bias in allelic or genotypic frequencies. Segregation ratios at only 14 loci (8.7%) from seven LGs deviate significantly from Mendelian expectations (P < 0.05). This number is only slightly higher than that expected from chance alone and none of the loci in question occurred in discrete blocks.

Gossypium arboreum and G. herbaceum differ by a reciprocal translocation (Gerstel 1953) and therefore the F1 is a translocation heterozygote. As a result, at least one of the A-genome LGs should contain sets of loci for which three point comparisons create unresolvable contradictions (e.g., locus A maps 5 cM from loci B and C, but locus B and C segregate independently). These contradictory data will be magnified in multipoint comparisons, making map orders ambiguous. Linkage group A5 (Fig. 8) fits this description. Comparisons to homoelogous LGs from the D, A1 (“A” genome of the allopolyploids), and D1 (“D” genome of the allopolyploids) genomes reinforces this inference.

D genome map

The D genome map (Figs. 1–11) comprises 306 loci encoded by 269 nuclear probes mapping to 17 LGs. The 17 LGs encompass 1486 cM. Individual LGs include 2 to 44 loci (average = 8.0). In contrast to the A genome, there is evidence for segregation distortion within three genomic regions. Segregation ratios at 37 loci (12.1%) from nine LGs deviate significantly (P < 0.05) from Mendelian expectations. Twenty-four of these loci map to three blocks on LGs D1, D5, and D9. On LG D1 (Fig 1), three loci (pAR168b, A1559, and pAR173b)
have a lower than expected number of heterozygotes, coupled with a consistent, albeit non-significant, under-representation of \textit{G. raimondii} alleles. On the same chromosome, loci between \textit{G1006} and \textit{pAR168a}, particularly \textit{pAR21a} and \textit{pAR24b}, have fewer \textit{G. raimondii} homozygotes than expected. The loci in the center of LG D5 (Fig. 8) have fewer \textit{G. raimondii} alleles than expected. This is coupled with a slight over-representation of heterozygotes toward the ends of the region. Finally, a 28.8 cM region on the lower end of D9 (Fig. 7) contains five loci with an excess of heterozygotes.

**Identification of orthologous and homoeologous LGs**

Of the 659 potentially orthologous (related by common ancestry and speciation, viz., A vs. D, A vs. At, and D vs. Dt) and homoeologous loci (related by duplication via allotetraploidy, viz., At vs. D), 424 (64\%) fall into 12 suites of loci that unite at least one A, D, At, and Dt LG. A 13th suite of loci unites one D and one A LG with two different LGs previously assigned to the Dt genome. These suites of loci are termed orthologous and homoeologous LGs, the number expected in an allotetraploid derived from two diploids with haploid complements of 13 chromosomess.

For discussion, the putatively orthologous and homoeologous suites of LGs are termed homoeologous assemblages, or HAs, and are denoted by numbers that correspond to their illustrations in Figs. 1–10. The inclusion of multiple LGs for a single genome within one HA implies that they represent segments of the same chromosome (e.g., LG D10 and LG D01 in HA 1; Fig. 1). Figs. 7 and 8 include two and three HAs, respectively, united by A diploid genome translocations and are designated 7A, 7B, 8A, 8B, and 8C (discussed below). Two A genome LGs (A21, A23) and two D genome LGs (D15, D17) could not be included in any of the 13 HAs. Four LGs from the allotetraploid (LG U07, Chr. 17D suppl., LG U02, and LG A04) had loci with putative orthologues in at least one diploid LG but could not be assigned to one of the LGs because of insufficient or contradictory data. These unassigned LGs are shown in Fig. 11, with the map locations of putative orthologues indicated. No loci mapping to allotetraploid LGs U09, A08, D11, or U05 had apparent orthologous loci in either diploid map and are not considered or illustrated. Linkage groups U02 and A04 are unique because all but one of their loci have homoeologues on other LGs, but almost none of these homoeologous loci map to LGs within the same HA. Either these LGs are artifactual or they represent AD genomic regions that are extensively rearranged. This “mosaicism” also occurs to a lesser extent on the upper end of LG U01 (Fig. 5). The lower end of this LG is clearly homoeologous to the A, D, and At LGs in this HA, but potential homoeologues to the loci on the upper end map to a variety of LGs in other HAs, which again suggests extensive rearrangement.

**Deviations from colinearity**

Comparison of locus orders among LGs within HAs revealed 19 locus order differences for which the assumption of colinearity requires accepting an alternate locus order on one or more LGs that is more than 100 times (LOD > 2) less likely than the preferred order (Table 1). Nine of these involve only two loci, and thus only weakly suggest the possibility of a rearrangement. The remaining 10 involve three or more loci and imply one or more inversions. Because these are low-density maps, we clearly have underestimated the actual number of rearrangements.

**Confirmation of At genome translocations**

Homoeologous assemblages 7A & 7B (Fig. 7) comprise two sets of D, A, and D LGs (Chr. 20D/D9/A6 and A14/D12/LG D07) with no evident relationship, except that both contain loci with homoeologs mapping to At LGs Chr. 5A and 4A. The homoeologous relationships shown (Fig. 7) provide evidence that Chr. 5A and 4A have undergone a reciprocal translocation relative to their D, A, and D counterparts (i.e., after polyploidization). Both Chr. 4A and 5A have two non-overlapping sets of loci, with each of the two sets showing homology to one of the two sub-assemblages (HA7A and HA7B) in Fig. 7. This inference is strengthened by classical studies that suggest the At genome will differ from the D, A, and At genomes by two reciprocal translocations, one of which involves At chromosomes 4 and 5 (Brown and Menzel 1950; Gerstel 1953; Gerstel and Sarvella 1956; Menzel and Brown 1954; Brown 1980; Menzel et al. 1982). Furthermore, the two At LGs included in Fig. 7 were assigned by Reinisch et al. (1994) to A, chromosomes 4 and 5.

Inspection of this compound HA revealed several inconsistencies. The A6 and A14 homoeologs to \textit{pAR219b}, \textit{G1033a}, and \textit{A1172} on Chr. 4A are interspersed with loci whose homoeologs map to Chr. 5A. The intercalation of \textit{G1033} and \textit{A1172} with \textit{A1159} and \textit{pAR206} on LG A14 may reflect a post-translocation inversion (the order of these loci on A14 is inverted relative to D12). Loci \textit{pAR219} and \textit{A1543} on Chr. 4A are also intercalated with loci with homoeologs mapping to Chr. 5A. In this case, it is unlikely that an inver-
sion is involved. It may be that the inferred homoeology between pAR219 on Chr. A6 and pAR219b on Chr. 4A is incorrect. Finally the relative placements of pAR138a on Chr. 4A and pAR138b on LG D07 requires the assumption of a post-translocation inversion.

Cytogenetic studies predict that the allotetraploid A genome will differ from the genomes of G. herbaeum and G. arboreum by a second A-specific reciprocal translocation confounded by a third translocation specific to the A diploid genome involving the allotetraploid A genome chromosomes.
1, 2, and 3 (Menzel and Brown 1954; Brown 1980). Thus, it was not surprising that the compound A genome LG A5 has homologies to three A1 LGs, two D genome LGs, and two Dt LGs (Fig. 8). Reinisch et al. (1994) assigned A1 LGs Chr. 1A and Chr. 2A to A1 chromosomes 1 and 2 (Brown 1980), which is fully consistent with this interpretation (see Menzel...
and Brown's (1954) Fig. 1). If one accepts this interpretation of the split affinities of A5, a second Aₜ-specific translocation becomes apparent, this one involving regions of chromosomes marked by LGs Chr. 2A and LG A06. As was the case with Chr. 4A and Chr. 5A in HA7, both of these LGs show split, non-overlapping homologies to two sets of otherwise unrelated assemblages of D, Dt, and A LGs. If these inferences are correct, Aₜ LG A06 corresponds to the allotetraploid Aₜ genome chromosome 3 (Menzel and Brown 1954; Brown 1980).

The only contradictory evidence for the interpretation just given involves loci revealed by probes pAR250 and A1536. The former locus maps to LGs in HA8A (A8 and D5) and HA8B (Chr. 17D), whereas A1536 maps to LGs in HA8B (Chr. 17D) and HA8C (Chr. 1A). Probe A1536 also reveals a locus that maps to A5, but that is not by itself contradictory. In light of the unambiguous homology relationships for a large number of loci, however, this single discrepancy does not weaken the current interpretation.

Homoeologous associations not accounted by HAs

Of the 659 loci that map to at least one other genome, 235 (36%) were not accounted for by the inferred HAs. In Figs. 1–11, locus associations with linkage groups outside of the HA depicted are indicated in brackets next to the pertinent loci. Many of these inter-HA associations occur in discrete blocks of loci, providing clues to additional rearrangements and past chromosome or chromosome-segment duplication events. Reinisch et al. (1994) described five tentative “nested” associations between segments of putative pairs of Aₜ and Dₜ homoeologues. Two of those are supported by the present data: A1759 and P6–57 (Chr. 10A/LG D04 & LG A03/LG D02) and A1751 and P5–61 (Chr. 10A/LG D04 & Chr. 5A). The other three are associated with the translocations and thus “mimic” nested associations. Map comparisons among the four genomes studied here revealed ten “nested” duplications, as follows.

**HA1 versus HA2**

In HA1, A1826 and A1625 map to the upper end of D1; A1826 maps to A15, which is associated with the upper end of A4 to which the other locus, A1625, maps. A1826 and A1625 also, however, map to the lower end of Chr. 22D in HA2.

**HA1 versus HA5**

A number of loci mapped in HA1 have duplicate loci that map to HA5. Specifically, pAR21 and pAR24 map to D1;

<table>
<thead>
<tr>
<th>HA</th>
<th>Linkage groups involved</th>
<th>Loci</th>
</tr>
</thead>
<tbody>
<tr>
<td>HA2</td>
<td>Chr. 22D vs. D8</td>
<td></td>
</tr>
<tr>
<td>HA3</td>
<td>D6 vs. A12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chr. 23D vs. D6</td>
<td></td>
</tr>
<tr>
<td>HA4</td>
<td>A16 vs. D4 &amp; LG D03</td>
<td></td>
</tr>
<tr>
<td>HA5</td>
<td>D7 vs. A13</td>
<td></td>
</tr>
<tr>
<td>HA6</td>
<td>D13 vs. A10</td>
<td></td>
</tr>
<tr>
<td>HA7A</td>
<td>Chr. 20D vs. D9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D9 vs. A6</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HA</th>
<th>Linkage groups involved</th>
<th>Loci</th>
</tr>
</thead>
<tbody>
<tr>
<td>HA8A</td>
<td>LG A06 vs. D5</td>
<td></td>
</tr>
<tr>
<td>HA8B</td>
<td>LG A06 vs. D3 &amp; Chr. 17D</td>
<td></td>
</tr>
<tr>
<td>HA8C</td>
<td>D2 vs. Chr. 15D</td>
<td></td>
</tr>
<tr>
<td>HA10</td>
<td>LG D04 vs. Chr. 10A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chr. 1A vs. D2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D11 vs. A9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D11 vs. Chr. 10A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A9 vs. Chr. 10A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D11 vs. Chr. 10A</td>
<td></td>
</tr>
</tbody>
</table>

**Table 1 (concluded).**
Fig. 3. HA3. The order of pAR288, A1737, A1124, and pAR127 on D6 is inverted relative to A12, as is A1606 and pAR8 on Chr. 23D versus D6.

pAR3 maps to a corresponding region of A4; and G1045b, pAR319a, and pAR21 map to the corresponding region of LG A05. In HA5, these loci variously map to median regions of LG D02, D7, and A13, and the upper region of LG A03.

HA2 versus HA3
In HA2, Probes G1155 and G1070 reveal loci mapping to D8 and G1155 reveals a locus on Chr. 22D; in HA3 these probes reveal loci that map to D6 and A12.

HA2 versus HA5
In HA2, pAR101 and PXP2–75 map to the upper regions of Chr. 22D and LG D05 and the former locus maps to D8.

In HA5, one or both of these loci map to the upper region of D7 and the median region of LG U01, which itself represents the most distal region of homology with the other linkage groups in HA5.

HA2 versus HA7A
pAR218, G1045, G1070, P6–25, and pAR179 map to submedian regions of Chr. 22D, D8, and LG D05. In HA7A, homologous loci map to D9 and Chr. 20D, albeit not as a discrete block.

HA4 versus HA5
P10–56 and A1214 map to the lower ends of LG D03 and LG A02, respectively, in HA4. In HA5, duplicates of both
loci map to the lower end of LG D02 whereas only the latter locus is duplicated on LG A03. Both of these chromosome segments show evidence of homology with an additional chromosomal region marked by HA10.

**HA5 versus HA10**

The loci **PXP4–75, G1261, A1759, pAR144, P6–57** are variously distributed on LG U01 + LG D02, D7, and LG A03 in HA5. Duplicates of these loci map to HA10, where one or more of the constituent loci map to LG D04, D11, and Chr. 10A.

**HA6 versus HA8B**

In HA6, **A1267 and A1345 map to A17**; in HA8B duplicates of these loci map to D3. We note that **A1345** also maps to LG A02 on HA4, raising the possibility that A17 is im-
Fig. 5. HA5. The order of pAR21 and pAR319 on D7 is inverted relative to A13.
among the A, D, and Dt linkage groups suggests that these (Figs. 7 and 8). Colinearity in these same genomic regions genomes of the allotetraploid F2 population are recombinational length by 5.8% (Table 2). Similarly, the two and 59%, respectively.

Differentiate the A diploid genome from the A t genome parisons revealed two reciprocal translocations that groups provide genetic evidence for structural rearrange- ments that have altered gene order and synteny. Map com- shorter than their A t and D t counterparts, differing by 52%

Variation in recombination rates among genomes

To evaluate potential differences in recombination rates among the mapped genomes, recombinational distances (cM) between pairs of loci in colinear regions were summed across the A, D, A t, and D t LGs (data available from CLB). To remove any bias due to map density, each summation was calculated directly from recombination fractions (Haldane 1919) for every pair of loci. When summed across all loci, the two diploid F2 populations differ in recombinational length by 5.8% (Table 2). Similarly, the two genomes of the allotetraploid F2 population are recombinationally equivalent (4.8% difference overall). The two diploid F2 populations, however, were genetically shorter than their A and D counterparts, differing by 52% and 59%, respectively.

Discussion

Genome evolution in Gossypium

Comparisons among the four maps demonstrate that gene order and synteny are sufficiently conserved among the A, D, D t, and A t genomes that the expected 13 assemblages of homoeologous linkage groups were easily identified. The resulting maps confirm and clarify some of the inferred homoeologies between linkage groups described previously for allotetraploid cotton (Reinisch et al. 1994).

These comparisons among the A, D, D t, and A t linkage groups provide genetic evidence for structural rearrange- ments that have altered gene order and synteny. Map com- parisons revealed two reciprocal translocations that differentiate the A diploid genome from the A t genome (Figs. 7 and 8). Colinearity in these same genomic regions among the A, D, and D t linkage groups suggests that these translocations occurred in the A t genome subsequent to allotetraploid formation. In addition, RFLP evidence identi- fies the genomic location of a translocation that differenti- ates G. arboreum and G. herbaceum (Fig. 8).

In addition to these translocations, comparisons of gene order among linkage groups within HAs revealed 19 puta- tive inversions. These inversions differentiated two or more of the four genomes in 12 of the 13 HAs (Table 1). In most cases the loci involved in the inversions were not polymor- phic in all four genomes and hence could not be mapped in one or more of the four genomes. Because of this, for 13 of the rearrangements it is unclear whether the event occurred at the diploid or allotetraploid level, or in which particular genome. The remaining six inversions probably arose subsequent to polyploidization. Five of these differentiate D from D t while one differentiates A from A t.

These inferences are remarkably congruent with those based on chromosome pairing behavior in diploid and allotetraploid hybrids. Classical cytogenetic approaches to comparative genome analysis have shown that the two A diploid genome species, G. arboreum and G. herbaceum, differ by a single translocation, and that genomes of these two species differ from the A t genome of allotetraploid cot- ton by two and three translocations, respectively (Brown 1980; Brown and Menzel 1950; Gerstel 1953; Gerstel and Sarvella 1956; Menzel and Brown 1954; Menzel et al. 1982). Beasley (1942) inferred the presence of one or more inversions between the A and A t genomes, claimed that the D and D t chromosomes differ by a minimum of four struc- tural differences (by implication inversions), and suggested that structural differences exist between all of the A and D chromosomes.

If we accept the congruence between Beasley’s observa- tions and ours as evidence that a minimum of 13 inversions and one translocation occurred during the divergence of the A and D diploid genomes, and set aside the caveats regarding the timing and genomic location of these events, the esti- mated rate at which inversions and translocations arose is 1.4 to 2.8 events per million years (13 inversions and one translocation divided by 5 to 10 million years). Similarly, the rate of fixation for translocations and inversions in the allotetraploids is estimated as 2.6 to 5.2 events per million years (6 inversions and 2 translocations divided by 1 to 2 million years, minus the rate for diplods). This suggests that the rate of fixation of inversions and translocations in Gossypium allotetraploids may be similar to or only moder- ately higher than that which occurred in the A and D diploid genomes. These calculations clearly are approximations and are subject to several sources of error. Nonetheless, the

Table 2. Genetic length differences among the diploid (A, D) and allotetraploid (A t, D t) Gossypium mapping populations.

<table>
<thead>
<tr>
<th>Shorter genome</th>
<th>L1 (cM)a</th>
<th>Longer genome</th>
<th>L2 (cM)a</th>
<th>Percentage differenceb</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>532.73</td>
<td>D</td>
<td>563.77</td>
<td>5.8%</td>
</tr>
<tr>
<td>D</td>
<td>532.40</td>
<td>A t</td>
<td>557.89</td>
<td>4.8%</td>
</tr>
<tr>
<td>A</td>
<td>134.00</td>
<td>A t</td>
<td>203.00</td>
<td>51.5%</td>
</tr>
<tr>
<td>D</td>
<td>769.20</td>
<td>D t</td>
<td>1219.38</td>
<td>58.5%</td>
</tr>
</tbody>
</table>

aL1 and L2 were calculated by summing the genetic distance between each adjacent pair of loci.
bCalculated as L1-L2/L1, where L1 = shorter genome length, and L2 = longer genome length.

Brubaker et al. 1999 NRC Canada
Fig. 6. HA6. The order of A1591, G1016, G1125, A1720, G1130, G1174, pAR49, and pAR8 on D13 is inverted relative to A10.

Fig. 7. This figure includes two HAs (HA7A and HA7B) joined by a reciprocal translocation between A, LGs Chr. 5A and Chr. 4A. A heavily dotted line separates HA7A from HA7B. A lightly dotted line connects two loci whose putative homoeology may contradict the inferred linkage group homologies depicted (see text). (A) HA7A. The orders of pAR169 and pAR65 on Chr. 20D relative to D9 and A1808, A1650, and pAR278 on D9 relative to A6 are inverted. The locations of G1112 and G1066 on Chr. 20D differ from their locations on Chr. 5A and A6, respectively. (B) HA7B. The order of G1033 and A1172 on D12 relative to A14 is inverted.
Fig. 8. This figure includes three HAs (HA8A, HA8B, and HA8C) united by a reciprocal translocation involving A, linkage groups Chr. 2A, and LG A06 and a confounded A linkage group (A5) that arises because the A genome diploid F1 parent of the F2 progeny is heterozygous for a reciprocal translocation. A heavily dotted line separates HA8A from HA8B from HA8C. Lightly dotted lines connect loci whose putative homoeology may contradict the inferred linkage group homoeologies depicted (see text). (A) HA8A. The order of G1164, A1418, and PXP2-60 on LG A06 is inverted relative to D5. (B) HA8B. Some of the HA8B Dt, D, and A loci have putative counterparts mapping to the confounded A genome linkage group, A5. Interspersed among these loci on A5 are loci with homologues mapping to linkage groups in HA8C which otherwise comprise one A1 (Chr. 1A), one D (D2), and one Dt (Chr. 15D) linkage groups. Within HA8B, the order of pAR185, pAR149, and pAR172 on LG A06 is inverted relative to D3 and Chr. 17D. (C) Within HA8C, the order of A1553, A1225, pAR11, P1-3, pAR88, and A1720 on D2 is inverted relative to Chr. 15D, and the locations of A1097 and A1794 on Chr. 1A differ from their positions on D2.

Table 3. Intra-linkage group duplications in the A, D, and AD genetic linkage maps. Probes revealing loci located within or near putative rearrangements are denoted in boldface.

<table>
<thead>
<tr>
<th>Probes revealing two loci within a linkage group</th>
<th>A genome Map location</th>
<th>D genome Map location</th>
<th>AD genome Map location</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1130</td>
<td>A10</td>
<td>D13</td>
<td>(Fig. 6)</td>
</tr>
<tr>
<td>A1124</td>
<td>D6</td>
<td>A12</td>
<td>(Fig. 3)</td>
</tr>
<tr>
<td>A1159</td>
<td>D12</td>
<td>A14, Chr. 5A</td>
<td>(Fig. 7B)</td>
</tr>
<tr>
<td>A1590</td>
<td>D4</td>
<td>A16, LG D03, LG A02</td>
<td>(Fig. 4)</td>
</tr>
<tr>
<td>G1070</td>
<td>D9</td>
<td></td>
<td>(Fig. 7A)</td>
</tr>
<tr>
<td>pAR145</td>
<td>D11</td>
<td></td>
<td>(Fig. 10)</td>
</tr>
<tr>
<td>pAR163</td>
<td>D7</td>
<td>LG D02</td>
<td>(Fig. 5)</td>
</tr>
<tr>
<td>pAR168</td>
<td>D1</td>
<td>A4, LG A05, LG D01</td>
<td>(Fig. 1)</td>
</tr>
<tr>
<td>pAR173</td>
<td>D1</td>
<td>A4, LG A05, LG D01</td>
<td>(Fig. 1)</td>
</tr>
<tr>
<td>pAR202</td>
<td>D9</td>
<td></td>
<td>(Fig. 1)</td>
</tr>
<tr>
<td>pAR173</td>
<td></td>
<td>LG D01</td>
<td>(Fig. 1)</td>
</tr>
<tr>
<td>pAR183</td>
<td></td>
<td></td>
<td>(Fig. 2)</td>
</tr>
<tr>
<td>A1183</td>
<td></td>
<td>Chr. 10A</td>
<td>(Fig. 10)</td>
</tr>
</tbody>
</table>

RFLP evidence is unambiguous with respect to the magnitude of structural changes: allotetraploidy in Gossypium was not accompanied by extensive restructuring. The question naturally arises as to whether this structural conservation is true for other allopolyploids. At present, this is a difficult question to address, as few comparative mapping data exist for diploids and their allotetraploid descendants. Perhaps the best example is Brassica. The three diploid genomes in B. rapa (A), B. nigra (B), and B. oleracea (C) are extensively rearranged relative to each other (Lagercrantz and Lydiate 1996; Quiros et al. 1994), but the A and C genomes of the allotetraploid B. napus are highly conserved relative to B. rapa and B. oleracea (Bohuon et al. 1996; Lydiate et al. 1993; Parkin et al. 1995). Cheung et al. (1997) inferred that major genome rearrangements differentiate B. napus and B. oleracea, but they did not distinguish between the A and C genomes in B. napus. Because the A and C diploid genomes are structurally different, many of the rearrangements they observed probably reflect differences between the C genome of B. oleracea and the A genome of B. napus. Nonreciprocal translocations may arise from homoeologous recombination in doubled haploids of Brassica napus (Sharpe et al. 1995), but the relevance of this observation to rearrangement in natural allopolyploids is not clear. Similarly, it may be that the unexpected gains and losses of restriction fragments observed in synthetic allopolyploids (Song et al. 1995) were partly due to genome rearrangement on the scale detectable by comparative RFLP mapping, but other mechanisms are perhaps more likely (Feldman et al. 1997; Liu et al. 1998; Song et al. 1995). Thus in Brassica, as in Gossypium, there is little direct evidence for structural rearrangement directly associated with polyploidy itself (Parkin et al. 1995).

Genome duplication and paleopolyploidy in Gossypium

One of the more notable features of genetic maps of diploid plants is that many genomic regions appear to be duplicated. For example, approximately 50% of loci in the Brassica A and C genomes are duplicated (Quiros et al. 1994), and in Sorghum 41% of probes detected more than one restriction fragment (Pereira et al. 1994). While there are a number of mechanisms by which loci may become duplicated, the occurrence of conserved linkage blocks shared by two chromosomes within a diploid genome or two chromosomes within a polyploid genome has been taken as evidence of paleopolyploidy (e.g., Chittenden et al. 1994; Helentjaris et al. 1988; Kianian and Quiros 1992; Kowalski et al. 1994; Whitkus et al. 1992). Given an evolutionary history of repeated cycles of polyploidization in angiosperms, ancient duplicated linkage blocks should be detectable in
map comparisons, and may even be detectable across kingdoms (Paterson et al. 1996). Within the grasses and Brassica, chromosomes appear to be mosaics of conserved linkage blocks. In Brassica, the A, B, and C genomes may be paleohexaploids that combine three sets of eight linkage blocks that have been differentially rearranged in each lineage as they diverged from a common ancestor (Lagercrantz and Lydiate 1996). Similarly, most grass genomes can be reconstructed using 19 linkage blocks (Moore et al. 1995; Moore et al. 1997). This suggests that diploid genome evolution may be accompanied by extensive shuffling of conserved linkage blocks that are flanked by centromeric and telomeric sites (Moore et al. 1997).
In *Gossypium*, previous cytogenetic, biochemical, and genetic mapping evidence suggested that the diploid species are paleopolyploids (reviewed by Reinisch et al. 1994). Thus, in the present study we expected to see blocks of loci shared by two or more linkage groups in one HA to be reiterated in linkage groups located in another HA. Summed across HAs, there were 235 loci (36% of those mapped) in the A, D, D_{0}, and A_{0} genomes that had putative orthologs...
and homoeologs on linkage groups in other HAs. For the most part, however, these duplicated loci did not fit the simple mapping pattern expected for ancient duplicated chromosome segments. The 10 nested duplications that fit the predictions of paleopolyploidy (see “Results”) accounted for 76, or about one third, of the loci under consideration. Moreover, there was no simple pattern of duplicated loci among the various HAs. It is probable that some duplications were generated by processes other than paleopolyploidy. Map densities also may be insufficient to identify ancient, duplicated linkage groups. In addition, it may be that the amount of time that has elapsed since paleopolyploidization has been long enough that subsequent mutational change has disrupted or obscured most of the ancient linkage groups. In this respect, we note that the entire tribe to which *Gossypium* belongs (the Gossypieae) is based on a chromosome number of \( n = 13 \), implying that paleopolyploidization antedates the origin of the tribe, which may be 20 to 40 million years old (Seelanan et al. 1997, and unpublished data).

**Are intrachromosomal duplications in *Gossypium* inversion footprints?**

Twelve probes revealed 13 pairs of loci that mapped to a single A, D, or AD linkage group (Table 3). Although intrachromosomal duplication of a locus may not be especially noteworthy, we draw attention to the correlation between these duplications and the rearrangements detected. Ten of the 13 locus pairs are located within or near putative rearrangements. Of the ten putative rearrangements involving three or more loci, six include duplicated loci on at least one linkage group. Only two intra-linkage group duplications are not associated with putative inversions. This suggests that the duplications may be “footprints” of inversions and that the process or processes that give rise to inversions...
may not be fully conservative, giving rise to duplications or deficiencies. Although duplications are evident, identifying deletions requires finding probes that differentially hybridize to the genomes being mapped. This possibility was not evaluated because probes were selected to maximize intergenomic comparisons.

Allopolyploidy in *Gossypium* is associated with increased recombination

A growing body of evidence indicates that recombination rates are not correlated with chromosome size. Maize has six times as much DNA per nucleus as rice and three times as much as sorghum, yet all three genomes are recombinationally similar across conserved linkages (Ahn and Tanksley 1993; Binelli et al. 1992; Whitkus et al. 1992). Similarly, *Capsicum* genomes are four times larger than those of tomato, yet the two species are recombinationally similar (Tanksley et al. 1988). Conversely, conserved linkages between potato and tomato differ in recombinational length by a factor of 1.4 to 3.6 but are of similar physical sizes (Bonierbale et al. 1988; Tanksley et al. 1992).

In *Gossypium*, chromosomes in the A genome diploids are nearly twice as large as those in the D genome diploids, and the A1 and D1 genome chromosomes retain the sizes of their diploid antecedents in allotetraploid cells (Endrizzi et al. 1985; Skovsted 1934). These size differences parallel incongruities in nuclear DNA content. The A genome has almost twice as much DNA per nucleus as does the D genome (2C = 4 vs. 2 pg; Bennett et al. 1982; Edwards et al. 1974; Edwards and Endrizzi 1975; Kadir 1976; H.J. Price, personal communication) with near additivity in the allotetraploids (2C = 5.6–5.8 pg; Gomez et al. 1993; Michaelson et al. 1991; H.J. Price, personal communication). Despite this nearly two-fold difference in size, recombination in linkage blocks conserved between the A and D diploid genomes and between the A1 and D1 allotetraploid genomes are essentially equivalent (5.8% and 4.8%, respectively; Table 2). This result verifies previous reports of a lack of correlation between genome size and total recombination in a particularly satisfying way, in that the two allotetraploid genomes are in the same nucleus, thereby controlling for all of the various life-history, population genetic, and ecological covariables that might be suspected of effecting recombination rates.

An unexpected result was that although there is no significant difference in recombination between genomes that vary in size by a factor of two, at either the diploid or allotetraploid level, there was an increase in recombination in the allotetraploid genomes (Table 2). The D1 genome was 58.5% larger, recombinationally, than its diploid counterpart, with a similar increase in recombination in the A1 genome (51.5% greater than A). These differences are remarkably similar to each other, suggesting that the responsible mechanism operates genome-wide in the allotetraploid. Although these results are based on only a single allotetraploid mapping population, Antonio et al. (1996) demonstrated that genetic distance for 300 markers in five different populations of rice did not vary significantly in any of 12 chromosomes. Our results should be considered suggestive and in need of verification with additional crosses, but at present, they suggest that allotetraploidy in *Gossypium* has been accompanied by an increased frequency of recombination. Whether this is true for polyploids in general, relative to their diploid ancestors, is worthy of investigation, as is the question of why this might be the case. At present, a satisfactory explanation for enhanced recombination in allotetraploids is wanting.

Acknowledgements

This work was supported by the National Science Foundation (JFW), U.S. Department of Agriculture (AHP), the Texas Agricultural Experiment Station (AHP), and Texas Higher Education Coordination Board (AHP).

References


© 1999 NRC Canada


