Genome Duplication in Soybean (Glycine subgenus soja)

R. C. Shoemaker,* K. Polzin,† J. Labate,* J. Specht,‡ E. C. Brummer,§ T. Olson,§ N. Young,§ V. Concibido,† J. Wilcox,‡ J.P. Tamulonis,‡§ G. Kochert‡,§§ and H. R. Boerma‡

*Department of Agronomy and USDA-ARS Field Crops Research, Iowa State University, Ames, Iowa 50011, †New Zealand Dairy Research Institute, Private Bag 11029, Palmerston North, New Zealand, ‡Department of Agronomy, University of Nebraska, Lincoln, Nebraska 68583-0915, §Department of Agronomy, Iowa State University, Ames, Iowa 50011-1010, ¶Department of Plant Pathology, University of Minnesota, S. Paul, Minnesota 55108, ¶¶USDA-ARS, Department of Agronomy, Purdue University, West Lafayette, Indiana 47904, §§Department of Crop and Soil Sciences, University of Georgia, Athens, GA 30602 and §§§Department of Botany, University of Georgia, Athens, GA 30602

Manuscript received March 12, 1996
Accepted for publication June 4, 1996

ABSTRACT

Restriction fragment length polymorphism mapping data from nine populations (Glycine max X G. soja and G. max X G. max) of the Glycine subgenus soja genome led to the identification of many duplicated segments of the genome. Linkage groups contained up to 35 markers that were duplicated on other linkage groups. The size of homoeologous regions ranged from 1.5 to 106.4 cM, with an average size of 45.3 cM. We observed segments in the soybean genome that were present in as many as six copies with an average of 2.55 duplications per segment. The presence of nested duplications suggests that at least one of the original genomes may have undergone an additional round of tetraploidization. Tetraploidization, along with large internal duplications, accounts for the highly duplicated nature of the genome of the subgenus. Quantitative trait loci for seed protein and oil showed correspondence across homoeologous regions, suggesting that the genes or gene families contributing to seed composition have retained similar functions throughout the evolution of the chromosomes.

W hen a genetic lineage splits, ancestrally related chromosome segments can usually be detected. Detailed mapping of these chromosome regions has demonstrated conservation of gene order across species and has recently been correlated to conservation of loci conferring complex agronomic traits (Lin et al. 1995; Paterson et al. 1995).

Homoeologous chromosome segments can be generated by polyploidization or by regional duplication. These chromosome segments are syntenic, in that they represent blocks of conserved linkages between the duplicated segments (Lundin 1993). The presence of such regions in a genome tells us much about genome structure and evolution (Bonierbale et al. 1988; Hulbert et al. 1990; Whithkus et al. 1992; Ahn et al. 1992; Ahn and Tanksley 1995; Reinisch et al. 1994) and can provide a means of transferring map information intragenomically from well-mapped to poorly mapped regions (Helentjaris 1993).

Rearrangements have frequently been observed following polyploidization, perhaps because they return the genome to a more stable diploidized state (Leipoldt and Schmidke 1982). Restriction fragment length polymorphism (RFLP) analysis of synthetic polyploids of Brassica showed that genomic changes occurred rapidly, beginning in the F2 generation, and resulted in loss or gain of restriction fragments, as well as the appearance of novel fragments (Song et al. 1995). This process has been proposed to account, in part, for the diversity and success of many ancient polyploid lineages (Song et al. 1995).

Duplications and subsequent rearrangements can result in chromosomes constructed of a patchwork of distinct homoeologous regions (Lundin 1993). Such chromosomes have been observed in Zea mays (Helentjaris et al. 1988) and in B. rapa (Song et al. 1991; Sharpe et al. 1995), with the former exhibiting a relatively simple, and the latter a complex pattern of homoeology. In those species in which the degree of duplication and chromosomal rearrangement has been extensive, ancient linkage relationships may not be conserved, making the identification of homoeologous regions difficult.

Evolutionary studies and haploid genome analysis have suggested that soybean [Glycine max (L.) Merr. subgenus soja] is an ancient tetraploid whose genome has, over time, become diploidized (Hadley and Hymowitz 1973). The observation that soybean multigene families contain two distinct subgroups of more closely related genes supports this hypothesis (Lee and Verma 1984; Hightower and Meagher 1985; Grandbastien et al. 1986; Nielsen et al. 1989). The construction of a RFLP map has now provided an opportunity to identify homoeologous regions of the genome, although most RFLP markers map as single-locus bimorphisms (i.e., two alleles) in crosses of soybean genotypes (Keim et al.
COMPARATIVE MAPPING OF MUNGBEAN (VIGNA RADIATA) AND COWPEA (V. UNGUIECULATA) HAS SHOWN THAT THESE TWO SPECIES SHARE A GENERAL LINKAGE CONSERVATION, WITH ONLY A FEW REARRANGEMENTS OF MARKER ORDERS (MENANCIO-HAUTEA ET AL. 1993). BOUTIN ET AL. (1995) COMPARED MUNGBEAN, COMMON BEAN (PHASEOLUS VULGARIS), AND SOYBEAN (G. MAX) AND HAVE DEMONSTRATED THAT MUNGBEAN AND COMMON BEAN SHARE RELATIVELY LARGE, CONSERVED LINKAGE BLOCKS. IN CONTRAST, ONLY SHORT, DISPERSED LINKAGE BLOCKS OF THE SOYBEAN GENOME WERE IN COMMON WITH THE TWO SPECIES. FOR EXAMPLE, SEGMENTS FROM AS MANY AS 16 DIFFERENT SOYBEAN LINKAGE GROUPS WERE FOUND ON A SINGLE MUNGBEAN LINKAGE GROUP, WHEREAS SEGMENTS FROM AS MANY AS NINE DIFFERENT SOYBEAN GROUPS WERE FOUND ON A SINGLE COMMON BEAN LINKAGE GROUP. THESE AUTHORS SPECULATED THAT THE GREATER INCIDENCE OF REARRANGEMENT IN THE SOYBEAN MAY BE DUE TO ITS EARLIER DIVERGENCE AND/OR TETRAPLOIDY.

OUR OBJECTIVES IN THIS STUDY WERE TO INTEGRATE INTO ONE COMMON MAP THE LINKAGE MAPS DERIVED FROM NINE DIFFERENT SOYBEAN POPULATIONS INVOLVING BOTH INTRA- AND INTERSPECIFIC CROSSES. MAP INTEGRATION PROVIDES A MEANS FOR DETECTING AND MAPPING LARGE NUMBERS OF DUPLICATED LOCUS. WE USED THE MAP POSITIONS OF DUPLICATED MARKERS TO IDENTIFY HOMELOCOSOME REGIONS WITHIN THE GENOME. IN THIS REPORT, WE ALSO PROVIDE EVIDENCE THAT SEGMENTAL DUPLICATION IN SEVERAL CHROMOSOMAL REGIONS HAS CONTRIBUTED SIGNIFICANTLY TO THE DUPLICATED NATURE OF THE SOYBEAN GENOME AND THAT THE SUBGENUS SOJA MAY HAVE UNDERGONE AN ADDITIONAL ROUND OF GENOME Duplication in its ancient evolutionary past, thus adding credence to its status as an “ancient polyploid” (SOLTIS ET AL. 1993). Finally, we also provide evidence suggesting that genes or gene families contributing to quantitative trait loci (QTL) affecting seed composition have retained similar functions throughout genome duplication and evolution events.

MATERIALS AND METHODS

GENETIC MAPS WERE PREVIOUSLY CONSTRUCTED FROM POPULATIONS DERIVED FROM THE INTRA- AND INTERSPECIFIC CROSSES SHOWN IN TABLE 1. THE INTRA- SPECIFIC CROSSES INVOLVED EVANS AS THE MATERNAL PARENT. THE F$_2$ RECOMBINANT INBRED LINE POPULATION WAS CONSTRUCTED BY ADVANCING 98 LINE FROM EVANS X PI 209:392 F$_2$ PLANTS TO THE F$_3$ GENERATION BY SINGLE-SEED DESCENT. SEVENTY-FIVE SEED FROM EACH LINE WERE PLANTED AND USED AS THE SOURCE OF BULK LEAF TISSUE FOR DNA EXTRACTION AND ANALYSIS. THE CROSS WITH CULTIVAR YOUNG AND PI416937 WAS AN F$_3$ RECOMBINANT INBRED LINE POPULATION. THIS POPULATION CONSISTED OF 120 LINES THAT WERE CREATED BY SINGLE-SEED DESCENT WITH EACH LINE ORIGINATING FROM A DIFFERENT F$_2$ PLANT. THE CROSS WITH PI97100 AND COKER 237 WAS AN F$_2$ POPULATION CONSISTING OF 111 LINES. POPULATIONS DEVELOPED FROM OTHER CROSSES HAVE BEEN DESCRIBED PREVIOUSLY (TABLE 1).


AUTORADIOGRAPHS CORRESPONDING TO EACH PROBE/ENZYME COMBINATION APPLIED TO TWO OR MORE OF THE NINE POPULATIONS WERE EXAMINED TO DETERMINE IF THE RESTRICTION FRAGMENT POLYMORPHISM ASHERED TO A GIVEN MARKER (FOR A SPECIFIC PROBE/ENZYME COMBINATION) WAS IDENTICAL ACROSS ALL POPULATIONS. IF SO, IT WAS TREATED AS AN “ANCHORING” MARKER FOR MAP INTEGRATION PURPOSES AND WAS USED TO DEFINE COMMON MARKERS BETWEEN MAPS. IF THE SEGREGATING POLYMORPHIC FRAGMENTS DIFFERED BETWEEN ANY TWO POPULATIONS, THIS RESULT WAS ATTRIBUTED TO TWO DISTINCT RFLP MARKERS AND WAS ASSUMED TO REPRESENT DUPLICATE LOCUS. THIS CRITERION WAS PARTICULARLY CRITICAL WHEN IDENTIFYING CO-DUPLICATED LOCUS. IN THOSE IN THE SAME RESTRICTION ENZYME WAS USED WITH A GIVEN PROBE, THEREBY PRECLUDING THE RESTRICTION FRAGMENT COMPARISON, THE MARKER DATA THEMSELVES WERE USED. IF THE RFLP IN ONE POPULATION AND THE RFLP IN ANOTHER POPULATION MAPPED TO DIFFERENT LOCATIONS AND THE RFLP IN ANOTHER POPULATION MAPPED TO DIFFERENT LOCATIONS (AS EVIDENCED BY A LOD SCORE OF 3 OR GREATER), AND EACH WAS LINKED TO DIFFERENT ANCHORED MARKERS, THEN THE RFLPS WERE ASSUMED TO BE PUTATIVE DUPLICATE MARKERS. MARKER DATASETS FROM EACH POPULATION WERE VISUALLY EVALUATED AND PRESORTED TO DISTINGUISH SUBSETS OF DATA FOR INTRAPOPULATION GROUPS. THE LINKAGE GROUP DATASETS FROM EACH POPULATION WERE THEN COMBINED, AND AN INTEGRATED MAP WAS CONSTRUCTED USING THE COMPUTER PROGRAM “JOINMAP” (STAM 1993). THE KOSAMBI MAPPING FUNCTION WAS SELECTED AND A MINIMUM LOD SCORE OF 3 WAS REQUIRED FOR A TWO-POINT LINKAGE TO BE INCLUDED IN ANY ANALYSIS. IN ALL CASES THE ORDER OF “ANCHORED MARKERS” DEFINED BY JOINMAP OUTPUT AGREED WITH THE ORDER OF THE LOCUS DEFINED BY THE REFERENCE POPULATION (SHOEMAKER AND OLSON 1993) AND MAPMAKER OUTPUT. THEREFORE, SPECIFICATION OF FIXED SEQUENCES WAS NOT NECESSARY.

TO ESTIMATE THE AVERAGE NUMBER OF RESTRICTION FRAGMENTS DETECTED PER PROBE, GENOMIC DNAs OF THE G. MAX BREEDING LINE AS1-356022 AND THE G. SOJA PLANT INTRODUCTION PI 469:916 WERE DIGESTED USING THE RESTRICTION ENZYMES HINDIII, DOL, EcoRI, EcoRV AND TagI. SOUTHERN HYBRIDIZATIONS WERE CARRIED OUT USING 280 RANDOMLY CHOSEN PIH genomic clones as probes with low stringency wash conditions (0.5× SSC, 0.5% SDS, 60°C). ESTIMATES OF MARKER DUPLICATION BASED ON RESTRICTION FRAGMENT COUNTING WERE OBTAINED BY DETERMINING THE AVERAGE NUMBER OF BANDS OBSERVED ON AUTORADIOGRAMS.

LOCATIONS OF QTL FOR SOYBEAN SEED OIL AND PROTEIN COMPOSITION WERE DETERMINED FROM PREVIOUS STUDIES USING POPULATIONS DESCRIBED IN TABLE 1 AND FOR WHICH ANCHORING RFLP PROBES WERE IN COMMON. CONCORDANCE OF QTL POSITIONS WITH DUPLICATED SEGMENTS WAS DETERMINED. QTL WERE CONSIDERED TO BE HOMELOCOSOME IF THEY WERE LOCATED WITHIN CLEARLY DELINEATED HOMELOCOSOME SEGMENTS AS DEFINED BY THREE OR MORE CONCORDANT MARKERS. NEGATIVE CORRELATION BETWEEN SOYBEAN SEED PROTEIN AND OIL COMPOSITION WAS OBSERVED IN ALL STUDIES (DIERS ET AL. 1992b; LEE ET AL. 1996b; E. C. BRUMMER, UNPUBLISHED RESULTS). THIS WAS CONSISTENT WITH THE STRONG NEGATIVE CORRELATION GENERALLY FOUND BETWEEN THESE TRAITS (BURTON 1985). EVen though variation may have been significant for one trait (e.g.,
of soybean have undergone genome duplication in addition to the pre-
derived an intergrated RFLP map contained ~810 markers
genes, but that large amounts of the genome have undergone
gene duplication in addition to the presumed tetraploidization event, or that soybean has under-
gene duplication in its evolutionary past.

The integrated RFLP map contained ~810 markers and 25 linkage
groups (Table 3). These linkage groups spanned distances ranging from 10.9 to 193.3 cM and

tained as few as three to as many as 71 markers. Most linkage groups contained numerous markers that
were duplicated on other linkage groups. For example, linkage group g (LG-g) contained 33 markers that were
duplicated on other linkage groups while LG-k contained 26; LG-c2 and LG-a2, 25; LG-f, 24; LG-e, 23;
LG-b1, 22; LG-n, 20; LG-d1, LG-h and LG-l, 19; LG-a1, 16; and LG-b2, 15. All other linkage groups contained 14 or fewer
duplicate markers. For any given linkage group, duplicate markers were present in as few as one
to as many as 17 (mean 8.2) other linkage groups.

Markers were frequently found to be duplicated within the same linkage group. Fifteen incidences of
marker duplication in a cis-configuration were observed. It is possible that some very closely linked cis-
duplicated markers, because they were mapped in separate populations using different restriction enzymes, may
represent a single chromosomal position. Of the nine linkage groups shown to contain cis-configured
duplicate markers, four (LG-c2, LG-f, LG-g, and LG-p)

TABLE 2

<table>
<thead>
<tr>
<th>No. of fragments</th>
<th>No. of probes</th>
<th>Percentage of total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>21</td>
<td>7.5</td>
</tr>
<tr>
<td>2</td>
<td>94</td>
<td>33.6</td>
</tr>
<tr>
<td>3</td>
<td>72</td>
<td>25.7</td>
</tr>
<tr>
<td>4</td>
<td>34</td>
<td>12.1</td>
</tr>
<tr>
<td>&gt;4</td>
<td>58</td>
<td>21.1</td>
</tr>
</tbody>
</table>

Fragments were detected under low stringency conditions by 280 randomly selected soybean 
PstI genomic probes hybridized to restriction enzyme-digested soybean genomic DNA
from the breeding line A81-356022 and plant introduction 
PI 468.916. Fragment numbers represent means over five en-
zymes.
### TABLE 3

Homoeologous relationships among soybean linkage groups

<table>
<thead>
<tr>
<th>Reference linkage group</th>
<th>Total cM of reference linkage group</th>
<th>Total number of markers</th>
<th>No. of markers duplicated elsewhere</th>
<th>No. of associated linkage groups&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Homoeologue&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Size of homoeologous segment&lt;sup&gt;c&lt;/sup&gt;</th>
<th>No. of homoeologous markers&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>a1</td>
<td>99.6</td>
<td>25</td>
<td>16</td>
<td>8</td>
<td>a2</td>
<td>45.0</td>
<td>5</td>
</tr>
<tr>
<td>a2</td>
<td>132.8</td>
<td>55</td>
<td>25</td>
<td>10</td>
<td>a1</td>
<td>70.5</td>
<td>5</td>
</tr>
<tr>
<td>b1</td>
<td>143.4</td>
<td>53</td>
<td>22</td>
<td>10</td>
<td>c2</td>
<td>27.7</td>
<td>3</td>
</tr>
<tr>
<td>b2</td>
<td>69.6</td>
<td>28</td>
<td>15</td>
<td>8</td>
<td>e</td>
<td>27.3</td>
<td>5</td>
</tr>
<tr>
<td>b2&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c1</td>
<td>150.9</td>
<td>25</td>
<td>10</td>
<td>7</td>
<td>d1</td>
<td>20.4</td>
<td>3</td>
</tr>
<tr>
<td>c2</td>
<td>142.0</td>
<td>49</td>
<td>18</td>
<td>11</td>
<td>d2</td>
<td>65.2</td>
<td>3</td>
</tr>
<tr>
<td>d1</td>
<td>51.3</td>
<td>37</td>
<td>29</td>
<td>18</td>
<td>e</td>
<td>37.6</td>
<td>5</td>
</tr>
<tr>
<td>d2&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e</td>
<td>127.7</td>
<td>54</td>
<td>23</td>
<td>17</td>
<td>d2</td>
<td>82.7</td>
<td>4</td>
</tr>
<tr>
<td>f</td>
<td>124.5</td>
<td>55</td>
<td>24</td>
<td>9</td>
<td>e</td>
<td>17.7</td>
<td>3</td>
</tr>
<tr>
<td>g</td>
<td>146.9</td>
<td>71</td>
<td>33</td>
<td>13</td>
<td>f</td>
<td>30.7</td>
<td>4</td>
</tr>
<tr>
<td>h</td>
<td>131.5</td>
<td>33</td>
<td>19</td>
<td>7</td>
<td>g</td>
<td>18.1</td>
<td>3</td>
</tr>
<tr>
<td>i</td>
<td>82.5</td>
<td>25</td>
<td>13</td>
<td>10</td>
<td>g</td>
<td>29.1</td>
<td>3</td>
</tr>
<tr>
<td>j</td>
<td>109.0</td>
<td>40</td>
<td>14</td>
<td>12</td>
<td>h</td>
<td>43.2</td>
<td>3</td>
</tr>
<tr>
<td>k</td>
<td>186.6</td>
<td>55</td>
<td>26</td>
<td>14</td>
<td>k</td>
<td>32.9</td>
<td>3</td>
</tr>
<tr>
<td>l</td>
<td>121.4</td>
<td>41</td>
<td>19</td>
<td>9</td>
<td>l</td>
<td>82.7</td>
<td>4</td>
</tr>
<tr>
<td>m</td>
<td>118.9</td>
<td>26</td>
<td>6</td>
<td>8</td>
<td>m</td>
<td>45.5</td>
<td>6</td>
</tr>
<tr>
<td>n</td>
<td>96.0</td>
<td>40</td>
<td>20</td>
<td>10</td>
<td>n</td>
<td>85.2</td>
<td>3</td>
</tr>
<tr>
<td>o</td>
<td>114.5</td>
<td>23</td>
<td>8</td>
<td>5</td>
<td>o</td>
<td>65.2</td>
<td>3</td>
</tr>
<tr>
<td>p</td>
<td>84.5</td>
<td>18</td>
<td>12</td>
<td>3</td>
<td>p</td>
<td>46.8</td>
<td>6</td>
</tr>
<tr>
<td>q</td>
<td>25.2</td>
<td>9</td>
<td>5</td>
<td>4</td>
<td>q</td>
<td>15.5</td>
<td>3</td>
</tr>
<tr>
<td>r</td>
<td>49.7</td>
<td>14</td>
<td>4</td>
<td>6</td>
<td>r</td>
<td>15.5</td>
<td>3</td>
</tr>
<tr>
<td>s</td>
<td>21.4</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>s</td>
<td>15.5</td>
<td>3</td>
</tr>
<tr>
<td>t</td>
<td>11.4</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>t</td>
<td>15.5</td>
<td>3</td>
</tr>
<tr>
<td>u</td>
<td>10.9</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>u</td>
<td>15.5</td>
<td>3</td>
</tr>
<tr>
<td>v</td>
<td>10.9</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>v</td>
<td>15.5</td>
<td>3</td>
</tr>
<tr>
<td>w</td>
<td>10.9</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>w</td>
<td>15.5</td>
<td>3</td>
</tr>
<tr>
<td>x</td>
<td>10.9</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>x</td>
<td>15.5</td>
<td>3</td>
</tr>
<tr>
<td>y</td>
<td>10.9</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>y</td>
<td>15.5</td>
<td>3</td>
</tr>
</tbody>
</table>

<sup>a</sup>The number of linkage groups to which markers on the tested linkage group possess duplicate loci.

<sup>b</sup>Linkage group to which homoeologous relationships are observed with the reference linkage group.

<sup>c</sup>Size, in cM, of the homoeologous segment on the reference linkage group.

<sup>d</sup>Number of duplicate markers defining the region of homoeology.
contained more than one pair. These markers may have been duplicated congruently or independently. The data do not provide information on this point.

**Identification of homoeologous genomic regions:** The probability that \( n \) randomly distributed duplicate markers are syntenic on linkage groups of equal size in a map of 25 linkage groups is equal to \( (1/25)^{(n-1)} \) (Reinisch et al. 1994). A coincidence of two or more duplicated loci between linkage groups (likelihood 0.04) represents putative evidence of homoeology. However, because this is only a crude estimate and assumptions of equal-sized linkage groups are violated, only a coincidence of three or more duplicated loci between linkage groups (likelihood 0.0016) was considered strong evidence of homoeology between segments rather than synteny by random chance.

Two instances were observed in which homoeologous segments involved seven pairs of duplicate markers, five instances involving six pairs of duplicate markers, six instances of five markers, four instances of four markers, and 28 instances of three markers (Table 3). There were 34 instances in which two pairs of duplicate markers were found to be in common between two different linkage groups. Only linkage groups cl and m, plus the very small two- or three-marker linkage groups \( r, s, w, \) and \( y \) were not identified as containing homoeologous segments.

Sizes of homoeologous segments ranged from 1.5 cM (spanned by three markers) to 106.4 cM (spanned by seven markers) with an average size of 45.3 cM (Table 3). The average density of markers shared between homoeologous segments was one marker per 11.7 cM. Few examples of perfect colinearity between homoeologous regions were observed.

Figure 1 shows examples of reference linkage groups, and the segments of homoeology to other linkage groups, based on the co-occurrence of three or more duplicated loci from the reference group. Based upon comparisons of all reference linkage groups, chromosome segments were duplicated, on average, 2.55 times and a nested pattern of genome duplication often tied together many linkage groups. For example, three markers, spanning ~13 cM of LG-k (markers 22–25) and duplicated on LG-j, are located within a region homoeologous to LG-e (markers 15–20) and within a region homoeologous to LG-h (markers 21–26) (Figure 1). Another example of the apparent multiple duplication of large genomic segments can be seen on reference linkage group LG-e where it seems that copies of the same genomic region can be found in LG-a2, LG-b2, LG-k, and LG-g.

Although duplicate markers were not always directly observed between two particular linkage groups, an inference of homoeology could be made if there was homoeology between each of those two linkage groups and a third linkage group. For example, Figure 1 depicts a region of reference LG-a2 homoeologous with a segment of LG-e. This region of LG-e is also homoeologous to LG-b2, LG-k, and LG-g. Since LG-a1 is also homoeologous with this segment of LG-e (through LG-a2) we can thus infer homoeology of regions of LG-b2, LG-k, and LG-g to LG-a1.

**Concordance of seed composition QTL with homoeologous segments:** Chromosome regions sharing markers also showed a tendency to share QTL for seed composition. Seventeen linkage groups were assayed with multiple markers to identify QTL for seed protein and oil. Major QTL were located on nine linkage groups, with relatively minor QTL on four others. A total of 54 discrete homoeologous comparisons were possible among the 17 linkage groups. From these, 25 pairs contained QTL for seed protein or oil on both homoeologues, 15 pairs contained no QTL on either homoeologue, and 14 pairs exhibited a QTL on only one homoeologue. We found a significant difference between the proportion of 40/54 (shared QTL state) and 14/54 (nonshared QTL state) \( (G = 8.86; P < 0.005, 1 \text{ d.f.}) \) (Sokal and Rohlf 1981).

Figure 2 shows examples of homoeologous regions containing QTL for seed composition. A group of markers identifying QTL for seed protein and oil were located on LG-a1 between markers 19 and 21 (E. C. Brummer, unpublished results). The genomic region encompassing these markers was homoeologous with a region of LG-a2 (Figure 2) also containing QTL for seed protein and oil (E. C. Brummer, unpublished results). In turn, markers defining a 24-cM region of LG-a2 containing QTL for these traits were dispersed along a region of LG-e (markers 7–6) that also contained numerous markers identifying QTL for protein and oil (Figure 2; E. C. Brummer, unpublished results; Diers et al. 1992b). A cluster of markers identifying QTL for protein and oil was also observed on LG-g. This region (markers 35–39) was shown to be homoeologous with a region on LG-c2 that also contained QTL for seed composition. LG-c2, in turn, contained a region homoeologous to LG-a2 (markers 10–12). Other markers were often shared between nonadjacent linkage groups, thus strengthening evidence for these homoeologous relationships even further. However, because one or two markers in common did not meet our criteria for claiming synteny, these data are not shown.

**DISCUSSION**

We have used nine different soybean populations to identify homoeologous relationships among chromosomes and have compiled data from 11 populations to identify QTL for seed composition. Our results indicate that (1) large portions of the soybean genome seem to have undergone duplication, (2) more than one round of duplication may have occurred (triplicate and quadruplicate markers are evident), and (3) genes conferring quantitative differences in seed composition show concordance with homoeologous regions.
FIGURE 1.—Examples of homoeologous relationships among soybean linkage groups. A homoeologous region is defined by the positions of three (or more) pairs of duplicated RFLP loci, where one of the two markers comprising each pair mapped to a reference linkage group while the other locus of each pair mapped to another linkage group. The reference linkage groups were those of the A81-356022 × PI 468,916 interspecific cross and in the figure the linkage group designations (a1, a2, e and k) are shown above the group in large bold type. The numbers within the open vertical bar correspond to markers and similar numbers indicate duplicate loci. Markers are graphically ordered and distanced in the reference linkage group to reflect the mapping data. The vertical open bars to the left and right of the reference linkage group correspond to segments of other linkage groups that possess putative homoeology with the reference linkage group. Homoeologous linkage group designations are shown in smaller bold type above or below the bars. Lines were drawn to connect the two markers of each duplicate pair primarily to show, where applicable, incidences of rearrangement between or among the homoeologous segments. Dashed lines connect nonadjacent groups. The scale marker applies only to the reference linkage group. All linkage groups and homoeologous relationships can be viewed through the SoyBase homepage on the WWW at http://129.186.26.94.
One byproduct of RFLP map construction has been the ability to estimate the degree of duplication within plant genomes, generally by counting the number of bands detected by cDNA or random genomic fragments under medium or low stringency conditions. Estimates of 26–80% sequence duplication (probes hybridizing to two or more bands) in rice (McCouch et al. 1988) and B. rapa genomes (McGrath and Quiros 1991; Song et al. 1991) indicate that plant genomes can differ significantly in their degree of duplication. Several studies indicate that the soybean genome is highly duplicated. Observations of significant chromosome pairing during meiosis in haploid soybean is indicative of extensive homoeology (Crane et al. 1982). In addition, many soybean phenotypes are known with digenic inheritance (Palmer and Kilen 1987). In this study, counts of restriction fragment bands on autoradiograms suggested that more than 90% of nonrepetitive soybean sequences are probably present in two or more copies. This high level of duplication is expected from an ancient tetraploid, such as the soybean, but it is much higher than the frequency of duplicated sequences observed for other supposed ancient tetraploids such as maize (28.6%) (Helentjaris et al. 1988). However, Ahn and Tanksley (1993) and Lee et al. (1992), who have used (as probes) sequences conserved between rice and sorghum and rice and maize, respectively, have demonstrated that the extent of duplication may be higher in maize than previously reported. The observation that ~60% of 280 RFLP probes detect three or more bands in soybean (based on counting bands in five different restriction enzyme digests under low stringency conditions) suggests that this higher level of duplication is not simply due to greater conservation of the duplicated loci present in the ancient tetraploid, but that much of the genome has undergone duplications in addition to the tetraploidization event. The high incidence of duplicated segments in common between three or more linkage groups, as we observed for soybean, could be paleohomoeologic evidence (Reinisch et al. 1994) of an additional round of genome duplication. It could also be explained by a high rate of segmental duplica-

**FIGURE 2.**—Examples of linkage groups containing QTL for seed protein and oil and for which homoeology is demonstrated. Markers are drawn and distanced to reflect mapping data. The numbers within each vertical bar refer to markers and similar numbers indicate duplicate loci. Lines drawn between linkage groups indicate duplicated markers. Dashed lines connect duplicate markers of nonadjacent linkage groups. The inverse relationship between quantities of protein and oil suggest that similar genetic factors control QTL for either trait (see text). Markers adjacent to the letters P and/or O are markers associated with QTL for protein and/or oil. For purposes of this study these QTL were considered collectively as seed composition QTL. The independent studies identifying the QTL are cited in the text.
tion during the diploidization of the soybean genome following a single tetraploidization event. We know of no easy way to unambiguously resolve these possibilities.

The distribution of duplicate markers previously suggested that the soybean genome was significantly rearranged during its evolution (Keim et al. 1990). Although it is likely that many of the rearrangements do represent actual chromosomal shuffling, some observed rearrangements are possibly artifacts arising from the inability to simultaneously detect and map all duplicated, triplicated, or quadruplicated markers present in the genome. Indeed, incomplete mapping of such markers could give the appearance of rearrangement simply because a random marker from each of several duplicate markers (that constitute homoeologous linkage groups) may not have yet been found to be polymorphic in the crosses tested to date. If, for example, a molecular map had been developed for a diploidized ancient octaploid, but not all of the quadruplicate loci had been identified and mapped, one might conclude that “rearrangement” had occurred, even if it had not (Lundin 1993).

It is unlikely that perceived rearrangements were due to structural differences in chromosomes among the various populations. Cytogenetic variation in soybean is uncommon, and none of the populations exhibited symptoms of sterility associated with cytogenetic aberrations.

In this study we showed that even though many homoeologous relationships likely remain unresolved, as evidenced by the many duplicate markers seemingly not associated with a homoeologous “segment,” much of the genome structure is consistent with a polyploid origin.

The incidence of multi-locus polymorphism between any given pair of soybean genotypes is relatively low (Keim et al. 1989). Thus duplicate markers are not frequently discovered when a single mapping population is examined (Keim et al. 1990). However, mapping populations constructed from different crosses can identify duplicate RFLP markers if the autoradiogram for a given probe exhibits multiple restriction fragments. Integration of independently derived genetic maps from different populations provides a way to detect many more duplicate markers than would be possible using any given single population.

Lack of variation for some QTL alleles prevents the identification of all QTL for any given trait in a single population. By comparing the locations of QTL identified in three independent studies and 11 independent populations representing Northern and Southern soybean germ plasm (Diers et al. 1992b; Lee et al. 1996b; E. C. Brummer unpublished results), we were able to identify a much larger array of QTL than would be possible with a single population.

We found a strong relationship between the presence of a QTL for seed composition on one member of a homoeologous pair and a similar QTL on the other homoeologue. These relationships often were maintained within nested duplications. The fact that these QTL reside within interrelated homoeologous regions suggests that they may share a common ancestral gene.

Certainly many genes, following tetraploidization and subsequent duplication events, undergo mutation to eliminate or alter their function (Pickett and Meeks-Wagner 1995). The finding in soybean that QTL for major agronomic characters were seemingly retained across homoeologous regions is noteworthy. A similar association was seen in maize where 50% of the QTL intervals for plant height resided within seven pairs of homoeologous regions (Lin et al. 1995). This suggests that the genes for these traits, or their functions, have been conserved, either by stabilizing selection or parallel evolution, in both soybean and maize.

Rearrangement following polyploidization can result in genome organizations resembling a scrambled patchwork (Leipold and Schmidtke 1982; Lundin 1993). Similarly, mutation, gene silencing, and change in function of redundant genes is also common following polyploidization (Pickett and Meeks-Wagner 1995). Still, in two widely diverse taxa we see evidence that not only are ancient homoeologous segments conserved, but that QTL for major agronomic traits are also apparently conserved within these regions. This seems highly unlikely to occur by chance. Persistence of genetic redundancy should only occur through negative or purifying selection (loss of function of one gene copy results in a selective disadvantage), or positive selection for the roles of all copies of the gene (Pickett and Meeks-Wagner 1995).

It is possible that natural selection factors have helped to maintain function of redundant QTL for plant height and quantity of seed protein and oil. It is not difficult to envision selective advantages of plant height in competition for sunlight, or quantity of seed protein providing better germination and seedling survival. It is equally plausible that the domesticators of these crops have provided positive selection pressure on the traits. If this is true, we could predict that other homoeologous complexes within plant genomes will contain homoeologous QTL for additional traits that provide domestication value or competitive advantage.

It has been speculated that genetic redundancy caused by tetraploidization or gene duplication events provides the genetic diversity needed to permit the physiological and morphological complexity observed in higher eukaryotes (Pickett and Meeks-Wagner 1995). The concordance of homoeologous segments with QTL for complex agronomic traits in both dicots (this study) and monocots (Lin et al. 1995) supports a universal role of homoeologous loci and/or genetic redundancy in quantitative inheritance.

The development of a detailed molecular genetic map of the Glycine subgenus soja genome has shown
that much of the organization of the genome is consistent with a polyploid origin and has suggested a possible additional round of genome duplication in the genome's past. In this study we have only cursorily compared locations of QTL from one phenotypic trait with regions of genome homoeology and have found a high degree of concordance. The identification of homoeologous regions of soybean chromosomes and the association of homoeologous regions with other phenotypic traits should increase our knowledge of the evolution of this legume genome and increase our understanding of the inheritance and genetic mechanisms controlling complex traits in plants.

The authors are grateful to E. POLLAK and S. MACKENZIE for valuable discussion. This work is a joint contribution of Midwest Area, U.S. Department of Agriculture (USDA)-ARS and Journal Paper No. 16591 of the Iowa Agricultural and Home Economics Experiment Station, Ames, IA. 50011-1010, Project 3236. Names are necessary to report factually on available data; however, the USDA neither guarantees nor warrants the standard of the product and the use of the name by the USDA implies no approval of the product to the exclusion of others that may also be suitable. This work was supported in part by grant no. 58-3625-2-401 from the United Soybean Board.

LITERATURE CITED


BAILEY, N., 1995 Statistical methods in biology. Press Syndicate of the University of Cambridge. The Pitt Building, Trumpleston Street, Cambridge CB2 1RP.


PATerson, A. Y.-R., L. LI, K. SCHERTZ, J. DOEPRY et al., 1995...


Shoemaker, R. C., and J. E. Specht, 1995 Integration of the soybean molecular and classical genetic linkage groups. Crop Sci. 35: 436–446.


Communicating editor: D. Charlesworth