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, St. 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* × *G. soja* and *G. max* × *G. max*) of the *Glycine* subgenus *soja* genome led to the identification of many duplicated segments of the genome. Linkage groups contained up to 33 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.

WHEN 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; WHITKUS *et al.* 1992; AHN *et al.* 1993; AHN and TANKSLEY 1993; 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 (LEIPOLD and SCHMIDTKE 1982). Restriction fragment length polymorphism (RFLP) analysis of synthetic polyploids of Brassica showed that genomic changes occurred rapidly, beginning in the F_2 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 polypoid 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.*

Corresponding author: Randy C. Shoemaker, USDA-ARS-FCR, Department of Agronomy, Iowa State University, Ames, IA 50011-1010. E-mail: rcsshoe@iastate.edu

1989). A significant number exhibit complex banding patterns on autoradiograms. Independent segregation of cohybridizing RFLP fragments in a segregating population allows the identification of duplicate markers. Those markers constitute duplicate loci in the sense that the DNA sequence used as the probe is to some extent conserved at each chromosomal position.

Comparative mapping of mungbean (*Vigna radiata*) and cowpea (*V. unguiculata*) 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 *inter*specific and *intra*specific crosses. Map integration provides a means for detecting and mapping large numbers of duplicated loci. We used the map positions of duplicated markers to identify homoeologous 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 interspecific and intraspecific crosses shown in Table 1. The intraspecific crosses with the cultivar Evans involved Evans as the maternal parent. The $F_{5:6}$ recombinant inbred line population was constructed by advancing 98 lines from Evans × PI 209.332 F_2 plants to the F_5 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 $F_{4:5}$ 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).

Genomic DNA isolation, restriction enzyme digestion, electrophoresis, blotting, probe preparation and labeling, hybridization, and membrane washing for the *G. max* × *G. soja* populations and the Clark × Harosoy population were conducted using the methods described by KEIM *et al.* (1988). For the other populations, the genomic DNA was prepared by a modified procedure of DELLAPORTA *et al.* (1983) and processed as reported in YOUNG *et al.* (1992). Linkage maps for the nine populations were generated with the computer program MapMaker (LANDER *et al.* 1987). Details of these procedures were reported 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 ascribed 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 loci. This criterion was particularly critical when identifying *cis*-duplicated loci. In those instances in which the same restriction enzyme was not used with a given probe, thereby precluding the restriction fragment comparison, the mapping data themselves were used. If the RFLP in one population 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 individual linkage 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 loci 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 A81-356022 and the *G. soja* plant introduction PI 468.916 were digested using the restriction enzymes *Hin*dIII, *Dra*I, *Eco*RI, *Eco*RV and *Taq*I. Southern hybridizations were carried out using 280 randomly chosen *Pst*I 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 homoeologous if they were located within clearly delineated homoeologous segments as defined by three or more common 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.*,

TABLE 1

Interspecific and intraspecific *Glycine* crosses that led to the mapping populations that were used in this study

Mapping population	Population structure	No. of individuals	No. of markers[a]	Reference
G. max × *G. soja*				
A81-356022 × PI 468.916	F_2	57	547	DIERS *et al.* (1992b)
C1640 × PI 479.750	F_2	59	73	BRUMMER *et al.* (1995)
G. max × *G. max*				
Clark × Harosoy	F_2	60	118	SHOEMAKER and SPECHT (1995)
Evans × PI 90763	F_2	115	10	
Evans × PI 88788	F_2	102	7	
Evans × PI 209.332	$F_{5:6}$ (RIL)	98	53	
Evans × Peking	F_2	110	14	
Young × PI 416937	F_4	120	154	LEE *et al.* (1996); MIAN *et al.* (1996)
PI 97100 × Coker 237	F_2	111	166	LEE *et al.* (1995)

[a] Number of informative markers involved in integrating linkage groups. Informative markers represent (1) anchored loci that are in common among the nine populations and (2) putative duplicate loci. The mapping data of the nine populations were integrated to identify duplicate loci.

oil amount was significantly increased) and not for the other (*e.g.*, protein amount was decreased but not significantly), both traits responded in concert. This suggests that the genetic factors controlling variation for each trait at a QTL may be the same. In this analysis we treated QTL for protein or oil simply as QTL for "seed composition."

To test for conservation of function between duplicated regions, all possible pairwise homoeologous comparisons were classified into discrete classes: (1) both homoeologues contained QTL for seed protein or oil, (2) one contained a QTL and one did not or (3) neither homoeologue contained a QTL. This method of classification allowed us to test for the significance of differences between two proportions using the 2 × 2 G-test as recommended by SOKAL and ROHLF (1981). The two proportions were defined as (1) number of pairs sharing QTL state (presence *vs.* absence) divided by the total number of pairs, and (2) the number of pairs not sharing QTL state divided by the total number of pairs.

RESULTS

Identification of duplicate markers: Analysis of the average number of fragments generated by each of 280 randomly chosen *PstI* genomic probes tested against five restriction enzyme digests have shown that ~92.5% detect two or more fragments under low stringency conditions (Table 2). These data suggested that more than 90% of the nonrepetitive sequences in soybean may be present in two or more copies. Approximately 58.9% of the RFLP probes detected three or more fragments in soybean (Table 2). This observation suggested that the high level of duplication is not simply due to greater conservation of duplicated loci from ancestral genomes, but that large amounts of the genome have undergone genome duplication in addition to the presumed tetraploidization event, or that soybean has undergone an additional round of genome 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 contained 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

Copy-number of soybean sequences detected with genomic probes

No. of fragments	No. of probes	Percentage of total
1	21	7.5
2	94	33.6
3	72	25.7
4	34	12.1
>4	58	21.1

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 enzymes.

TABLE 3

Homoeologous relationships among soybean linkage groups

Reference linkage group	Total cM of reference linkage group	Total number of markers	No. of markers duplicated elsewhere	No. of associated linkage groups[a]	Homoeologue[b]	Size of homoeologous segment[c]	No. of homoeologous markers[d]
a1	99.6	25	16	8	a2	45.0	5
a2	132.8	55	25	10	a1	70.5	5
					c2	27.7	3
					e	27.3	5
b1	143.4	53	22	10	g	41.9	3
					h	53.3	7
					q	14.6	3
b2	69.6	28	15	8	d1	20.4	3
					e	65.2	3
c1	150.9	25	10	7	—	—	—
c2	142.0	49	18	11	a2	88.4	3
					g	82.7	4
d1	193.3	49	19	13	b2	65.0	3
					n	8.7	3
					p	37.6	5
d2	44.9	11	6	3	f	17.7	3
e	127.7	54	23	17	a2	98.2	5
					b2	52.2	3
					g	47.5	3
					k	29.6	3
f	124.5	55	24	9	e	30.7	4
					h	18.1	3
g	146.9	71	33	13	a2	50.9	3
					b1	87.8	3
					c2	34.3	5
					e	9.0	3
					k	47.6	6
h	131.5	33	19	7	b1	106.4	7
					f	29.1	3
					k	43.2	3
i	82.5	25	13	10	o	26.9	3
j	109.0	40	14	12	l	33.5	4
					k	54.2	3
k	186.6	55	26	14	e	92.0	3
					g	42.7	6
					h	25.2	3
					j	12.7	3
					l	88.4	3
l	121.4	41	19	9	j	46.0	4
					k	31.2	3
					n	42.5	6
m	118.9	26	6	8	—	—	—
n	96.0	40	20	10	d1	1.5	3
					l	45.5	6
o	114.5	23	8	5	i	85.2	3
p	84.5	18	12	3	d1	46.8	6
q	25.2	9	5	4	b1	15.5	3
r	49.7	14	4	6	—	—	—
s	21.4	4	2	2	—	—	—
w	11.4	3	1	1	—	—	—
y	10.9	4	3	4	—	—	—

[a] The number of linkage groups to which markers on the tested linkage group possess duplicate loci.
[b] Linkage group to which homoeologous relationships are observed with the reference linkage group.
[c] Size, in cM, of the homoeologous segment on the reference linkage group.
[d] 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 c1 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 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 homoeolgous 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.

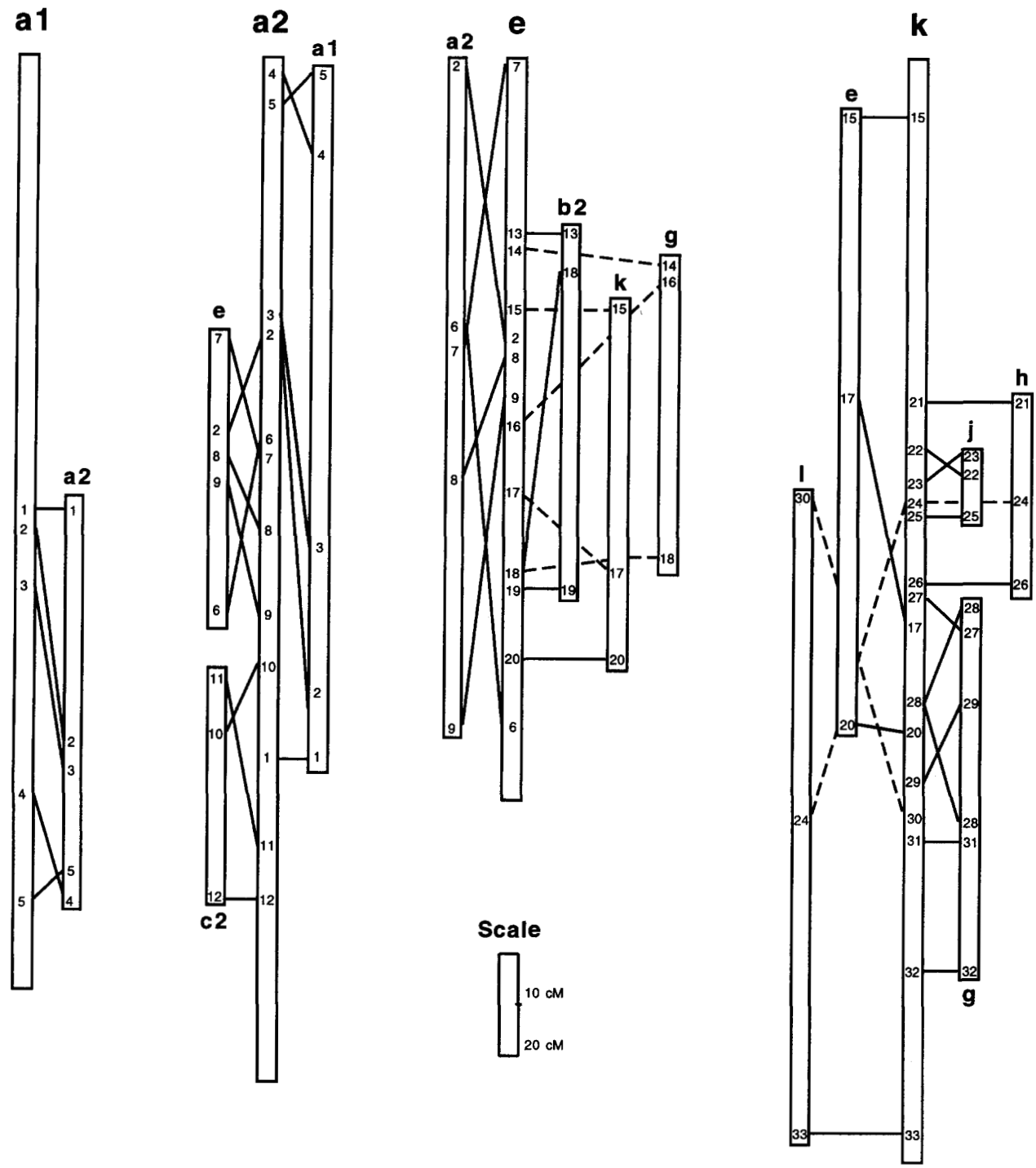

FIGURE 1.—Examples of homoeologous relationships among soybean linkage groups. A homoeologous region is defined by the positions of three (or more) pairs of duplicated RFLP loci, where one of the two markers comprising each pair mapped to a reference linkage group while the other locus of each pair mapped 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.

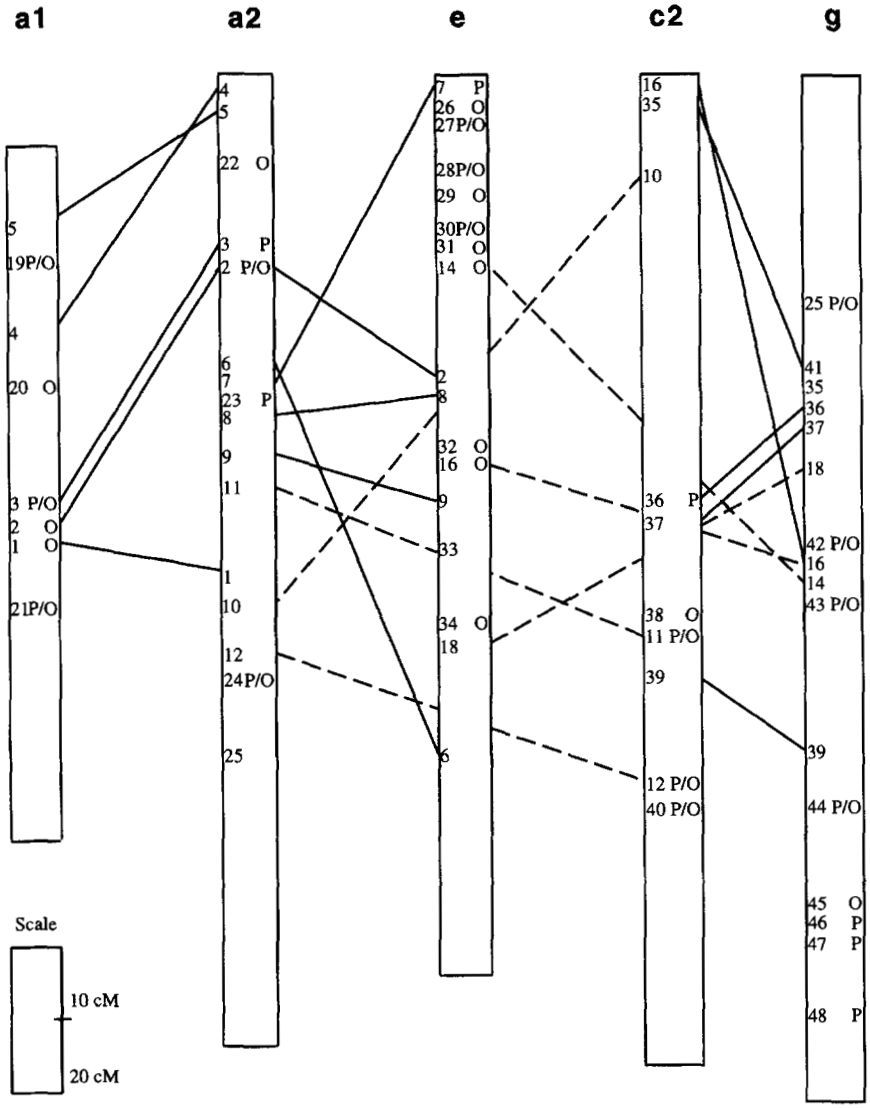

FIGURE 2.—Examples of linkage groups containing QTL for seed protein and oil and for which homoeology is demonstrated. Markers are drawn and distanced to reflect mapping data. The numbers within each vertical bar refer to markers and similar numbers indicate duplicate loci. Lines drawn between linkage 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.

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-

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 survivability. 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

AHN, S., and S. D. TANKSLEY, 1993 Comparative linkage maps of the rice and maize genomes. Proc. Natl. Acad. Sci. **90:** 7980–7984.

AHN, S., J. A. ANDERSON, M. E. SORRELLS and S. D. TANKSLEY, 1993 Homoeologous relationships of rice, wheat and maize chromosomes. Mol. Gen. Genet. **241:** 483–490.

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

BONIERBALE, M. W., R. L. PLAISTED and S. D. TANKSLEY, 1988 RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics **120:** 1095–1103.

BOUTIN, S., N. YOUNG, T. OLSON, Z.-H. YU, R. SHOEMAKER *et al.,* 1995 Genome conservation among three legume genera detected with DNA markers. Genome **38:** 928–937.

BRUMMER, E. C., A. D. NICKEL, J. R. WILCOX and R. C. SHOEMAKER, 1995 Mapping the *Fan* locus controlling linolenic acid content in soybean oil. J. Hered. **86:** 245–247.

BURTON, J. W., 1985 Breeding soybeans for improved protein quantitiy and quality, pp. 361–367 in *Proceedings of the 3rd World Soybean Research Conference,* edited by R. SHIBLES. Westview Press, Boulder, CO.

CRANE, C. F., W. B. BEVERSDORF and E. T. BINGHAM, 1982 Chromosome pairing and associations at meiosis in haploid soybean (*Glycine max*). Can. J. Genet. Cytol. **24:** 293–300.

DELLAPORTA, S. L., J. WOOD and J. B. HICKS, 1983 A plant DNA minipreparation; version II. Plant Mol. Biol. Rept. **1:** 19–21.

DIERS, B. W., P. KEIM and R. C. SHOEMAKER, 1992a Mapping of the *Phytophthora* resistance loci in soybean with restriction fragment length polymorphism markers. Crop Sci. **32:** 377–383.

DIERS, B. W., P. KEIM, W. R. FEHR and R. C. SHOEMAKER, 1992b RFLP analysis of soybean seed protein and oil content. Theoret. Appl. Genet. **83:** 608–612.

GRANDBASTIEN, M. A., S. BERRY-LOWE, B. W. SHIRLEY and R. B. MEAGHER, 1986 Two soybean ribulose-1,5-bisphosphate carboxylase small subunit genes share extensive homology even in distant flanking sequences. Plant Mol. Biol. **7:** 451–465.

HADLEY, H. H., and T. HYMOWITZ, 1973 Speciation and cytogenetics, pp. 97–116 in *Soybeans: Improvement, Production, and Uses,* Ed. 1, edited by B. E. CALDWELL. American Society of Agronomy, Madison, WI.

HELENTJARIS, T., 1993 Implications for conserved genomic structure among plant species. Proc. Natl. Acad. Sci. USA **90:** 8308–8309.

HELENTJARIS, T., D. WEBER and S. WRIGHT, 1988 Identification of the genomic locations of duplicate nucleotide sequences in maize by analysis of restriction fragment length polymorphisms. Genetics **118:** 353–363.

HIGHTOWER, R. C., and R. B. MEAGHER, 1985 Divergence and differential expression of soybean actin genes. EMBO J. **4:** 1–8.

HULBERT, S., T. RICHTER, J. AXTEL and J. BENNETZEN, 1990 Genetic mapping and characterization of sorghum and related crops by means of maize DNA probes. Proc. Natl. Acad. Sci. USA **87:** 4251–4255.

KEIM, P., T. C. OLSON and R. C. SHOEMAKER, 1988 A rapid protocol for isolating soybean DNA. Soybean Genet. Newsl. **15:** 147–148.

KEIM, P., R. C. SHOEMAKER and R. G. PALMER, 1989 Restriction fragment length polymorphism diversity in soybean. Theor. Appl. Genet. **77:** 786–792.

KEIM, P., B. DIERS, T. C. OLSON and R. C. SHOEMAKER, 1990 RFLP mapping in soybean: association between marker loci and variation in quantitative traits. Genetics **126:** 735–742.

KILEN, T. C., E. E. HARTWIG and B. L. KEELING, 1974 Inheritance of a second major gene for resistance to phytophthora rot in soybeans. Crop Sci. **14:** 260–262.

LANDER, E., P. GREEN, J. ABRAHAMSON, A. BARLOW, M. DALY *et al.,* 1987 MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics **1:** 174–181.

LEE, J. S., and D. P. S. VERMA, 1984 Structure and chromosomal arrangement of leghemoglobin genes in kidney bean suggest divergence in soybean leghemoglobin gene loci following tetraploidization. EMBO J. **3:** 2745–2752.

LEE, S. H., M. A. BAILEY, M. A. R. MIAN, E. R. SHIPE, D. A. ASHLEY *et al.,* 1996 Identification of quantitative trait loci for plant height, lodging and maturity in a soybean population segregating for growth habit. Theor. Appl. Genet. **92:** 516–523.

LEE, S. H., M. A. BAILEY, M. A. R. MIAN, T. E. CARTER JR., D. A. SHELY *et al.,* 1996a Molecular markers associated with soybean plant height, lodging, and maturity across locations. Crop Sci. **36:** 728–734.

LEE, S. H., M. A. BAILEY, M. A. R. MIAN, T. E. CARTER, E. R. SHIPE *et al.,* 1996b RFLP loci associated with soybean seed protein and oil content across populations and locations. Theor. Appl. Genet. (in press).

LEIPOLD, M., and J. SCHMIDTKE, 1982 Gene expression in phylogenetically polyploid organisms, pp. 219–236 in *Genome Evolution,* edited by G. DOVER and R. FLAVELL. Academic Press, New York.

LIN, Y.-R., K. SCHERTZ and A. PATERSON, 1995 Comparative analysis of QTLs affecting plant height and maturity across the Poacae, in reference to an interspecific sorghum population. Genetics **141:** 391–411.

LUNDIN, L. G., 1993 Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse. Genomics **16:** 1–9.

MCGRATH, J., and C. QUIROS, 1991 Ineritance of isozyme and RFLP loci in *Brassica campestris* and comparison with *B. oleracea.* Theor. Appl. Genet. **82:** 668–673.

MENANCIO-HAUTEA, D., C. FATOKUN, L. KUMAR, D. DANESH and N. D. YOUNG, 1993 Comparative genome analysis of mungbean (*Vigna radiata* l. Wilczek) and cowpea (*V. unguiculata* L. Walpers) using RFLP mapping data. Theor. Appl. Genet. **86:** 797–801.

MCCOUCH, S. R., G. KOCHERT, Z. H. YU, Z. Y. WANG, G. S. KHUSH, *et al.,* 1988 Molecular mapping of rice chromosomes. Theoret. Appl. Genet. **76:** 815–829.

MIAN, M. A. R., M. A. BAILEY, D. A. ASHLEY, R. WELLS, T. E. CARTER JR. *et al.,* 1996 Molecular markers associated with water use efficiency and leaf ash in soybean. Crop Sci. (in press).

MUELLER, E. H., K. L. ATHOW and F. A. LAVIOLETTE, 1978 Inheritance of resistance to four physiologic races of *Phytophthora megasperma* var. *sojae.* Phytopathology **68:** 1318–1322.

NIELSEN, N. C., C. D. DICKINSON, T.-J. CHO, V. H. THANH, B. J. SCALLON, *et al.,* 1989 Characterization of the glycinin gene family in soybean. Plant Cell **1:** 313–328.

O'BRIEN, S. J. (Editor), 1993 *Genetic Maps.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

PALMER, R. G., and T. C. KILEN, 1987 Qualitative genetics and cytogenetics, pp. 135–209 in *Soybeans: Improvement, Production, and Uses,* Ed. 2, edited by J. R. WILCOX. American Society of Agronomy, Madison, WI.

PATERSON, A., Y.-R. LIN, Z. LI, K. SCHERTZ, J. DOEBLEY *et al.,* 1995

Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. Science **269:** 1714–1718.

PELLEGRINI, M., and R. B. GOLDBERG, 1979 DNA sequence organization in soybean investigated by electron microscopy. Chromosoma **75:** 309–326.

PICKETT, B. G., and D. R. MEEKS-WAGNER, 1995 Seeing double: appreciating genetic redundancy. Plant Cell **7:** 1347–1356.

REINISCH, A., J. DONG, C. BRUBAKER, D. STELLY, J. WENDEL *et al.*, 1994 A detailed RFLP map of cotton, *Gossypium hirsutum* × *Gossypium barbadense:* chromosome organization and evolution in a disomic polyploid genome. Genetics **138:** 829–847.

SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989 *Molecular Cloning.* Cold Spring Harbor Press, Cold Spring Harbor, NY.

SHARPE, A. G., I. A. P. PARKIN, D. J. KEITH and D. J. LYDIATE, 1995 Frequent nonreciprocal translocations in the amphidiploid genome of oilseed rape (*Brassica napus*). Genome **38:** 1112–1121.

SHOEMAKER, R. C., and T. C. OLSON, 1993 Molecular linkage map of soybean (*Glycine max* L. Merr.), pp. 6.131–6.138 in *Genetic Maps,* Ed. 6, edited by S. J. O'BRIEN. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

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

SOKAL, R., and F. J. ROHLF, 1981 *Biometry: The Principles and Practice of Statistics in Biological Research.* W. H. Freeman and Co., New York.

SOLTIS, P., J. DOYLE and D. SOLTIS, 1993 Molecular data and polyploid evolution in plants, pp. 177–201 in *Molecular Systematics of Plants,* edited by P. SOLTIS, D. SOLTIS and J. DOYLE. Routledge, Chapmann and Hall, New York.

SONG, K. M., J. Y. SUZUKI, M. K. SLOCUM, P. H. WILLIAMS and T. C. OSBORN, 1991 A linkage map of *Brassica rapa* (syn. *campestris*) based on restriction fragment length polymorphism loci. Theoret. Appl. Genet. **82:** 296–304.

SONG, K., P. LU, K. TANG and T. C. OSBORN, 1995 Rapid genome change in synthetic polypoids of *Brassica* and its implications for polyploid evolution. Proc. Natl. Acad. Sci. USA **92:** 7719–7723.

SPRINGER, P. S., K. J. EDWARD and J. L. BENNETZEN, 1994 DNA class organization on maize ADH yeast artificial chromosomes. Proc. Natl. Acad. Sci. USA **91:** 863–867.

STAM, P., 1993 Construction of integrated genetic linkage maps by means of a new computer package. Join Map. Plant J. **3:** 739–744.

TANKSLEY, S. D., R. BERNATZKY, N. L. LAPITAN and J. P. PRINCE, 1988 Conservation of gene repertoire but not gene order in pepper and tomato. Proc. Natl. Acad. Sci. USA **85:** 6419–6423.

WADSWORTH, G. J., M. G. REDINBAUGH and J. G. SCANDALIOS, 1988 A procedure for the small-scale isolation of plant RNA suitable for RNA blot analysis. Anal. Biochem. **172:** 279–283.

WHITKUS, R., J. DOEBLEY and M. LEE, 1992 Comparative genome mapping of sorghum and maize. Genetics **132:** 1119–1130.

YOUNG, N., L. KUMAR, D. MENANCIO-HAUTEA, D. DANESH, N. TALEKAR *et al.*, 1992 RFLP mapping of a major bruchid resistance gene in mungbean (*Vigna radiata* L., Wilczek). Theor. Appl. Genet. **84:** 839–844.

Communicating editor: D. CHARLESWORTH