Theor Appl Genet (2003) 106:1256–1265
DOI 10.1007/s00122-002-1183-y

H. H. Yan · J. Mudge · D.-J. Kim · D. Larsen
R. C. Shoemaker · D. R. Cook · N. D. Young

Estimates of conserved microsynteny among the genomes of *Glycine max*, *Medicago truncatula* and *Arabidopsis thaliana*

Received: 3 January 2002 / Accepted: 28 October 2002 / Published online: 25 January 2003

Abstract A growing body of research indicates that microsynteny is common among dicot genomes. However, most studies focus on just one or a few genomic regions, so the extent of microsynteny across entire genomes remains poorly characterized. To estimate the level of microsynteny between *Medicago truncatula* (*Mt*) and *Glycine max* (soybean), and also among homoeologous segments of soybean, we used a hybridization strategy involving bacterial artificial chromosome (BAC) contigs. A *Mt* BAC library consisting of 30,720 clones was screened with a total of 187 soybean BAC subclones and restriction fragment length polymorphism (RFLP) probes. These probes came from 50 soybean contig groups, defined as one or more related BAC contigs anchored by the same low-copy probe. In addition, 92 whole soybean BAC clones were hybridized to filters of *Hin*dIII-digested *Mt* BAC DNA to identify additional cases of cross-hybridization after removal of those soybean BACs found to be repetitive in *Mt*. Microsynteny was inferred when at least two low-copy probes from a single soybean contig hybridized to the same *Mt* BAC or when a soybean BAC clone hybridized to three or more low-copy fragments from a single *Mt* BAC. Of the 50 soybean contig groups examined, 54% showed microsynteny to *Mt*. The degree of conservation among 37 groups of soybean contigs was also investigated. The results indicated substantial conservation among soybean contigs in the same group, with 86.5% of the groups showing at least some level of microsynteny. One contig group was examined in detail by a combination of physical mapping and comparative sequencing of homoeologous segments. A TBLASTX similarity search was performed between 1,085 soybean sequences on the 50 BAC contig groups and the entire *Arabidopsis* genome. Based on a criterion of sequence homologues <100 kb apart, each with an expected value of ≤1e-07, seven of the 50 soybean contig groups (14%) exhibited microsynteny with *Arabidopsis*.

Keywords *Glycine max* · *Medicago truncatula* · Microsynteny · Genome duplication · Evolution

Communicated by J. Dvorak

H. H. Yan · J. Mudge · D. Larsen · N. D. Young (✉)
Department of Plant Pathology, 495 Borlaug Hall,
1991 Upper Buford Circle, University of Minnesota, St. Paul,
MN 55108, USA
e-mail: neviny@umn.edu
Tel.: 612-625-2225, Fax: 612-625-9728

N. D. Young
Department of Plant Biology, University of Minnesota,
St. Paul, Minnesota, 55108, USA

D.-J. Kim · D. R. Cook
Department of Plant Pathology, University of California,
Davis, California, 95616, USA

R. C. Shoemaker
Department of Agronomy and USDA-ARS-Corn Insect
and Crop Genetics Research Unit, Iowa State University, Ames,
Iowa, 50011, USA

Introduction

Comparative genetic mapping has demonstrated that gene content and order are frequently conserved among related plant species (Gale and Devos 1998; Bennetzen 2000; Keller and Feuillet 2000). For example, colinearity has been extensively documented in the grasses, where fewer than 30 rice linkage blocks can represent the genomes of major cereal crops (Devos et al. 1999; Devos and Gale 2000). Extensive conservation has also been described between *Brassica* species and *Arabidopsis thaliana* (*At*), where segments homoeologous to *Arabidopsis* appear to be duplicated or triplicated in diploid *Brassica* crops (Cavell et al. 1998; Lagercrantz 1998; Jackson et al. 2000). However, map-based comparisons at the centimorgan (cM) level of resolution do not provide a comprehensive picture of similarities in genome microstructure, something that is required for cross-species gene prediction and positional cloning (Tikhonov et al. 1999).

Comparative sequencing of orthologous regions has been carried out among a few plant species (Bennetzen

2000; Ku et al. 2000; Tarchini et al. 2000; Mayer et al. 2001). Chen et al. (1997, 1998) found that the *sh2/a1* region, which spans approximately 140 kilobase pairs (kb) in maize but only 19 kb in rice and sorghum, maintained gene composition and arrangement despite more than 50 million years of divergence. Intergenic sequences in these regions were highly diverged. In the orthologous *adh* regions of maize and sorghum, a similar genomic organization was observed, although both intergenic and genic sequences showed obvious divergence (Tikhonov et al. 1999). By contrast, available data indicate microsynteny between *Arabidopsis* and rice (the two plant models representing dicots and monocots) may be limited to relatively few regions (Devos et al. 1999; Van Dodeweerd et al. 1999; Mayer et al. 2001).

Another important discovery from genomic research has been the existence of extensive duplication in many plant species, ranging in size all the way from small chromosomal segments to entire genomes (Shoemaker et al. 1996; Gaut and Doebley 1997; Ku et al. 2000). For example, the complete sequencing of the *Arabidopsis* genome demonstrated that duplicated segments encompass up to 60% of the 115.4 million base pairs (Mb) of sequenced DNA, with 24 multicopy segments more than 100 kb in size (Blanc et al. 2000; The *Arabidopsis* Genome Initiative 2000). Sequence conservation between duplicated segments was almost completely limited to exons (Blanc et al. 2000; The *Arabidopsis* Genome Initiative 2000). Based on these results, it has been proposed that at least four different large-scale duplication events have occurred in the *Arabidopsis* lineage (Vision et al. 2000). In soybean, more than 90% of the non-repetitive sequences were found to have at least two copies (Shoemaker et al. 1996; Grant et al. 2000). Still, it remains largely unknown what changes may have occurred within homoeologous segments following duplication events.

The Fabaceae (Legume) family comprises some of the world's most important crops, including soybean, alfalfa and pea. Among legumes, *Medicago truncatula* (*Mt*) is widely considered as a model for genomic analysis. *Mt* is a self-fertile, diploid species with eight pairs of chromosomes and a DNA content of approximately 500 Mb (Cook et al. 1997). The main focus of our study was to estimate the nature and extent of conserved microsynteny between soybean and *Mt*, focusing at a physical scale the size of typical bacterial artificial chromosome (BAC) contigs, which in this study was about 180 kb. We also examined microsynteny among soybean homoeologues and between soybean and *Arabidopsis*. The information presented here should be useful in elucidating genomic relationships among legumes and in exploiting microsynteny to complement chromosome walking and positional cloning in crops such as soybean.

Materials and methods

Sequence analysis and probe isolation

Two BAC libraries were constructed in soybean, one from *Hin*dIII partially digested genomic DNA of variety "Williams 82" (Marek and Shoemaker 1997), the other from *Eco*RI partially digested genomic DNA of variety "Faribault" (Danesh et al. 1998). Ordered sets of overlapping BAC clones (BAC "contigs") were built from these two BAC libraries using mapped RFLP clones plus a few BAC-ends as hybridization probes (Marek et al. 2001). Due to frequent duplication in the soybean genome (Shoemaker et al. 1996; Marek et al. 2001), multiple BAC contigs were generally identified with each soybean probe (see Results). In this study, all the related contigs uncovered by a given probe were assigned to a single "contig group". Contigs from each contig group were then termed "contig 1", "2", "3", and so on in arbitrary order.

BAC ends and several BAC subclones were sequenced by the University of Minnesota Advanced Genetic Analysis Center on ABI PRISM 377 sequencers (Perkin Elmer) using BigDye terminator cycle sequencing kits (Perkin Elmer) (Marek et al. 2001). These sequences, plus the underlying RFLP sequences retrieved from GenBank or sequenced as part of this project, were analyzed against: (1) the NR databases at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/BLAST/) and the National Center for Genome Resources (http://Seqsim.ncgr.org/newBlast.html) using BLASTX; (2) mapped *Arabidopsis* clones in GenBank comprising a minimum tiling path identified by The Institute for Genomic Research (http://www.tigr.org/tdb/e2k1/ath1/) using TBLASTX; and (3) soybean ESTs in a database stored at the Center for Computational Genomics and Bioinformatics at the University of Minnesota (http://web.ahc.umn.edu/biodata/nsfsoy/) using BLASTN. In this report, sequences with database hits were defined as those that showed significant alignments in any of these searches with an expected value of ≤1e-08 in BLASTX or TBLASTX, or ≥95% identity over at least 100 bp in BLASTN, excluding any sequences annotated or discovered to be repetitive DNA in our own database of soybean BAC-end sequences (J. Mudge, personal communication). Remaining sequences were considered to have no database hits.

Primers were designed for 181 BAC-derived (BAC-ends and subclones) sequences. The PCR products, 150–500 base pairs (bp) in length, were amplified from DNA of the corresponding BAC clones. Products were gel-purified before use in hybridization.

Library screening, Southern hybridization and fingerprinting

Low-copy soybean BAC-derived sequences and soybean RFLP probes that had been used to anchor the soybean contigs were hybridized to the *Mt* BAC library at 65 °C. This *Mt* library was constructed with *Hin*dIII partially digested genomic DNA from the variety "Jemalong", consisting of 30,720 clones and representing approximately 5× the haploid genome coverage (Nam et al. 1999). Filters were washed twice in 2 × SSC/0.1% SDS, 10 min each, at 65 °C, and then exposed at –70 ° C for 4–7 days. Such low-stringency washings and long exposures allowed for easy localization of positive BACs, which were subsequently verified by preparation of positive *Mt* BACs using the alkaline-lysis method (Marra et al. 1997) and Southern hybridization. In parallel experiments, soybean BAC-derived probes were also hybridized to filters containing *Eco*RI- and *Hin*dIII-digested genomic DNA from soybean cv Faribault and *M. truncatula* cv Jemalong.

BAC contig construction

In most cases, *Mt* BACs identified by a soybean probe were fingerprinted in order to build BAC contigs (Marra et al. 1997). Briefly, 4.4-µl of BAC DNA was digested with *Hin*dIII at 37 °C for 3–4 h, run in a 1.0% gel (2 × GGB buffer, 80 mM of Tris base/40 mM of sodium acetate/4 mM of EDTA/52 mM of glacial acetic acid, Wong et al. 1997) for 15 h at 12 °C, and stained for

about 1 h with Sybr Gold (Molecular Probes) diluted to 1:10,000 in 2 × GGB buffer. Images were acquired by scanning gels on a FluorImager STORM 840 (Molecular Dynamics). In some cases, images were also captured and analyzed on Polaroid films in order to resolve small (<2 kb) fragments. BACs identified by the same probe were assigned to a single contig if: (1) they shared five or more common restriction fragments, and (2) the original anchoring probe hybridized to a restriction fragment of the same size.

Cross-hybridization with whole soybean BAC clones

In addition to BAC-ends, some whole soybean BAC clones were used in cross-hybridizations both to soybean BACs from duplicated contigs and to positive *Mt* BACs. First, the level of repetitive DNA in these clones was evaluated by hybridizing about 100 ng of *Hae*III-digested DNA from each BAC clone to filters containing *Eco*RI- and *Hin*dIII-digested genomic DNA of soybean and *Mt*. If the results of these experiments indicated essentially no repetitive DNA in the *Mt* genome, these soybean BAC clones were then hybridized to filters containing *Hin*dIII-digested *Mt* BACs that had been previously identified by a BAC-end or RFLP from the same soybean contig group.

To analyze cross-hybridization among soybean duplicated contigs, filters containing soybean BACs that had been digested with *Eco*RI, *Eco*RV, *Hin*dIII, *Dra*I or *Taq*I (depending on the enzyme originally used to map the anchoring probe) were probed with a BAC clone from a different contig within the same contig group. While it was difficult to identify soybean BACs entirely free of soybean repetitive DNA for use as probes, we did try to chose soybean BAC clones containing low levels of repetitive DNA based on the hybridization pattern of soybean genomic DNA when probed with each of the BACs. To further monitor for repetitive sequences among cross-hybridizing fragments, the filters used to identify cross-hybridization between soybean BACs were subsequently hybridized with radiolabeled soybean genomic DNA. About 100 ng of genomic DNA was separately digested with three enzymes, *Eco*RI, *Hin*dIII and *Hae*III, then pooled and labeled with ^{32}P-dCTP. X-ray films were developed after 2-day exposure at –70 °C. Cross-hybridized fragments containing highly and middle repetitive DNA sequences could thereby be identified based on signal intensity and excluded when microsynteny was estimated. Except for the original BAC library screenings, all filters were washed in 2 × SSC/0.1% SDS, 1 × SSC/0.1% SDS and 0.5 × SSC/0.1% SDS for 15 min each at 65 °C.

Criteria for microsynteny

Two types of microsynteny were inferred from results with short low-copy soybean probes (RFLP, BAC-end, BAC subclone) or whole BAC probes. Conserved microsynteny between soybean and *Mt* was inferred when: (1) at least two distinct probes (BAC-end, subclone, or RFLP) from the same soybean contig hybridized to the same *Mt* BAC (type I), or (2) at least three non-repetitive cross-hybridizing restriction fragments were observed when a whole soybean BAC clone was hybridized to a *Mt* BAC that had previously been detected by a low-copy probe from the same soybean contig (type II). Microsynteny among duplicated contigs within a soybean contig group was inferred when: (1) at least two distinct low-copy probes (BAC-end, subclone, or RFLP) from one contig hybridized to BAC(s) in a duplicated contig (type I), or (2) at least three cross-hybridizing fragments were observed in BACs of duplicated contigs when a whole soybean BAC clone was used as probe (type II).

Microsynteny between the soybean and *Arabidopsis* genomes was based on TBLASTX analysis described (http://mips.gsf.de/). We established a relatively non-stringent criterion for microsynteny, considering that soybean and *Arabidopsis* genomes diverged about 90 million years ago (Grant et al. 2000). Conserved microsynteny was inferred when homologues of two or more soybean sequences from the same soybean contig group were separated by less than 100 kb in *Arabidopsis*, each with an expected value of ≤1e-07.

If any soybean contig within a contig group showed microsynteny to *Mt*, this soybean contig group was considered to show microsynteny to *Mt*. This same criterion was applied to comparisons among duplicated contigs within a soybean contig group and between soybean and *At* (see Tables 1 and 2).

Isolation of DNA paralogues in soybean

In order to isolate paralogues from contig groups within soybean, primers that were originally designed according to BAC-end sequences in one soybean contig were used to amplify DNA from BACs located in a duplicated contig. If the PCR product was similar in size to the original BAC-end sequence, this pair of primers were then used directly in a sequencing reaction, with the DNA from BACs in a duplicated contig as a template. Each clone (including the original BAC-ends) was sequenced at least twice, and sequence alignments were performed with Sequencher 3.1.1 (Gene Codes Co., Ann Arbor, Michigan).

Results

Microsynteny between soybean and *M. truncatula*

An average of 20 BAC-derived sequences were obtained from each soybean contig group and analyzed using BLASTX, TBLASTX and BLASTN. Physically linked

Table 1 Summary of conserved microsynteny estimates between soybean (Sb) contig groups, *M. truncatula* (*Mt*) and *A. thaliana* (*At*), as well as between soybean duplicated contigs

Item	Sb × Sb[a]			Sb × *Mt*[a]			Sb × *At*[b]
	Type I	Type II	Type I or Type II	Type I	Type II	Type I or Type II	
No. of contig groups with conserved microsynteny	23 (62.2%)	29 (78.4%)	32 (86.5%)	15 (30.0%)	23 (46.0%)	27 (54.0%)	7 (14.0%)
No. of contig groups without conserved microsynteny	14 (37.8%)	8 (21.6%)	5 (13.5%)	35 (70.0%)	27 (54.0%)	23 (46.0%)	43 (86.0%)
Total contig groups	37	37	37	50	50	50	50

[a] Type I: microsynteny detected with two or more linked soybean BAC-end or RFLP probes. Type II: microsynteny detected with whole soybean BAC probes

[b] Based on TBLASTX analysis, microsynteny inferred when homologues of sequences from one soybean contig group were less than 100 kb apart in *At*, each with an expected value of ≤1e-07

Table 2 Microsynteny between soybean (Sb) contig groups, *M. truncatula* (*Mt*) and *A. thaliana* (*At*), as well as within the soybean genome

No.	Anchoring probe	No. Sb contigs	Sb × Sb synteny[a]	Sb × *Mt* No. contigs tested	Sb × *Mt* synteny[a]	Sb × *At* synteny[a,b]
1	pA588	2	* I	2	* II	no
2	pA036	2	* I	1	no	no
3	pA450	2	* I	1	no	no
4	pA110	2	* I II	2	* I	no
5	pA064	4	* I II	3	* I II	no
6	pA078	2	* I II	2	* I II	no
7	pA086	3	* I II	3	* I II	no
8	pA109	3	* I II	2	* I II	no
9	pA112	4	* I II	3	* I II	no
10	pA401	4	* I II	2	* I II	*
11	pA810	3	* I II	3	* I II	*
12	pA885	2	* I II	1	* I II	no
13	pA257	2	* I II	2	* II	*
14	pA486	3	* I II	2	* II	*
15	pA702	2	* I II	2	* II	no
16	pB030	4	* I II	3	* II	no
17	26_J06R	3	* I II	2	no	no
18	Bng171	2	* I II	2	no	no
19	pA124	4	* I II	3	no	no
20	pA802	2	* I II	2	no	no
21	pA878	4	* I II	4	no	no
22	pB031	2	* I II	2	no	no
23	pB208	2	* I II	1	no	no
24	pA059	3	* II	2	* I	no
25	pA708	4	* II	3	* I II	no
26	pA095	3	* II	2	* II	no
27	pA963	5	* II	4	* II	no
28	pB039	2	* II	1	* II	no
29	pA077	4	* II	3	no	no
30	pA264	2	* II	1	no	no
31	pA398	2	* II	1	no	no
32	pA489	2	* II	2	no	no
33	52_G04R	3	no	1	no	no
34	pA481	3	no	2	no	no
35	pB132	4	no	3	no	no
36	pB212	3	no	2	no	no
37	pK493	2	no	2	no	no
38	pA315	1	nd	1	* I II	no
39	NBSD_H1	3	nd	1	* II	no
40	pA069	3	nd	1	* II	no
41	pA703	1	nd	1	* II	*
42	Satt309	9	nd	1	* II	no
43	27_A14F	1	nd	1	no	no
44	pA104	1	nd	1	no	no
45	pA567	1	nd	1	no	no
46	pB053	1	nd	1	no	no
47	pC006	4	nd	1	no	no
48	NBSD_H8	6	nd	2	* I	*
49	pA085	1	nd	1	* I	no
50	AFLP_170	7	nd	1	* I II	*

[a] Microsynteny indicated by *. I and II: type I and type II microsynteny respectively (see Table 1 for details); no: without microsynteny; nd: not determined

[b] Microsynteny inferred when homologues of sequences from one Sb contig group were less than 100 kb apart in *At*, each with an expected value of ≤1e-07

sequences within a soybean contig (each approximately 180 kb in length) allowed us to search for instances of microsynteny with *Mt*. A total of 181 BAC-derived probes were developed from 50 soybean contig groups. To avoid possible artifacts of microsynteny caused by soybean repetitive DNA, all BAC-derived probes were first surveyed by hybridization to restriction-digested soybean and *Mt* genomic DNA. Signals in the lanes of *Mt* genomic DNA were mostly weak or absent. In soybean, 37 probes hybridized to multiple fragments (often with smear signals) and presumably belonged to middle or highly repetitive DNA sequences, while the remaining 144 appeared to consist of low-copy sequences only. All 144 low-copy BAC-derived probes along with 43 RFLP probes from 50 soybean contig groups – 114 with database hits and 83 without – were then used to screen the *Mt* BAC library. On average 2.7 *Mt* BACs were detected with each probe, although 62 soybean probes (33.2%) did not detect any *Mt* BAC. Microsynteny was inferred when two or more low-copy probes from a single soybean contig hybridized to the same *Mt* BAC (type I). By this definition, a total of 15 soybean contig groups (30%) showed conserved microsynteny to *Mt* (Tables 1 and 2). Within this group, there were 22.8% (26/114) soybean

sequences with database hits and 12.1% (10/83) with no database hits.

Evaluation of whole BAC probes for the presence of repetitive sequences

To compare genome organization across entire contigs (approximately 180 kb), the use of short probes like RFLPs or BAC subclones was insufficient. Therefore, we tested 106 whole soybean BAC clones as potential probes in hybridization against Southern filters of digested *Mt* BAC clones. First, we eliminated soybean BAC clones containing repeated sequences in *Mt*. The hybridization patterns of whole BAC probes against filters with digested soybean genomic DNA indicated that 72 of the BAC probes contained regions of highly or middle repetitive DNA in soybean. However, 92 soybean BAC clones hybridized to fewer than five visible fragments in digested *Mt* genomic DNA, and most of them were very weak in intensity. Presumably, most soybean repeat sequences were absent or sufficiently diverged in *Mt* such that they produced little or no signal in cross-hybridization experiments. The other 14 soybean BAC clones hybridized to five to ten *Mt* fragments and were not analyzed further.

Identifying microsynteny with *Mt* using whole BAC clones

Thirty six BAC clones from 23 soybean contig groups (out of a total of 50) hybridized to three or more fragments from digested *Mt* BACs, indicating the existence of type II conserved microsynteny between these soybean and *Mt* BACs (Tables 1 and 2). For example, soybean BAC 025_G19 hybridized to eight bands in *Mt* BAC 021_E16, while soybean BAC 139_N10 hybridized to a total of 12 bands in *Mt* BAC 007_D24 (Fig. 1B, D). As noted earlier, these soybean and *Mt* BACs contained little or no cross-hybridizing repetitive DNA based on Southern hybridization of genomic DNA filters by whole BAC probes or that of BAC DNA filters probed with genomic DNA. As only one BAC clone from each soybean contig was used as probe in these cross-hybridizations, there were a few cases where we failed to detect type II microsynteny even when type I microsynteny had been observed. Combining the results of whole BAC probes, BAC-derived probes and RFLPs, a total of 27 (54%) out of the 50 soybean contig groups showed at least some conserved microsynteny with the *Mt* genome (Tables 1 and 2).

There were eight instances of duplicated contigs from soybean showing conserved microsynteny to the same *Mt* BAC. These contigs were anchored by soybean probes pA064, pA078, pA086, pA095, pA257, pA486, pA702 and pA810. In addition, we found five soybean contigs that each showed conserved microsynteny to *Mt* BACs from two different contigs. These were soybean contigs anchored by probes pA078, pA257, pA703, pA963 and AFLP_170.

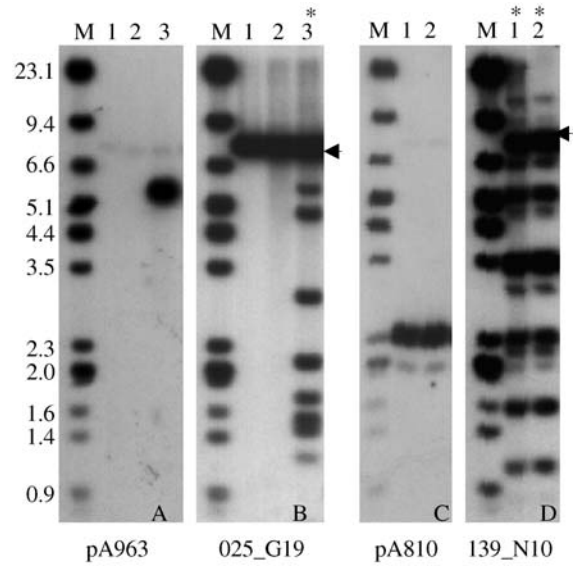

Fig. 1A–D Two examples of cross-hybridization experiments to study microsynteny between soybean and *Mt*. DNA from *Mt* BAC clones was digested with *Hin*dIII, transferred to nylon membranes and hybridized with soybean RFLP probes (**panels A and C**) or whole BAC probes (**panels B and D**). *Arrows* indicate vector; *M*, molecular-weight standards (in kb). *Mt* BACs showing multiple cross-hybridizing bands are marked by *. (**A**) *Lanes 1 to 3*: hybridization with soybean RFLP pA963. Probe pA963 hybridized to a 5.6-kb band in *Mt* BAC 021_E16 (*lane 3*). The two *Mt* BACs in *lanes 1 and 2* showed no hybridization and were not analyzed further. (**B**) *Lanes 1 to 3*: hybridization with soybean BAC 025_G19 from contig 4 of probe pA963. BAC 025_G19 hybridized to eight bands in *Mt* BAC 021_E16 (*lane 3*). (**C**) Hybridization with soybean RFLP pA810. Probe pA810 hybridized with two fragments in *Mt* BAC 007_D24 (2.3 kb and 1.9 kb, *lanes 1 and 2*). (**D**) Hybridization with soybean BAC 139_N10 from contig 2 of probe pA810. BAC 139_N10 hybridized to a total of twelve bands in *Mt* BAC 007_D24 (*lanes 1 and 2*)

Microsynteny within soybean contig groups

We also used hybridization to evaluate the extent of conserved microsynteny among contigs within each of 37 soybean contig groups. These contig groups were chosen because they had between two and five separate BAC contigs previously identified by a low-copy RFLP or BAC-end probe (Fig. 2). A total of 108 low-copy BAC-end sequences (54 with database hits and 54 without) were hybridized to filters containing digested BACs from the corresponding contig groups. Twenty seven of the probes with database hits (50.0%) and 21 probes without database hits (39.1%) revealed type I microsynteny in 23 contig groups (62.1%) (Tables 1 and 2).

In addition, 90 whole soybean BAC clones were used as hybridization probes. These hybridizations revealed six contig groups with three or four cross-hybridizing fragments and 25 contig groups with five or more cross-hybridizing fragments. For example, soybean BAC

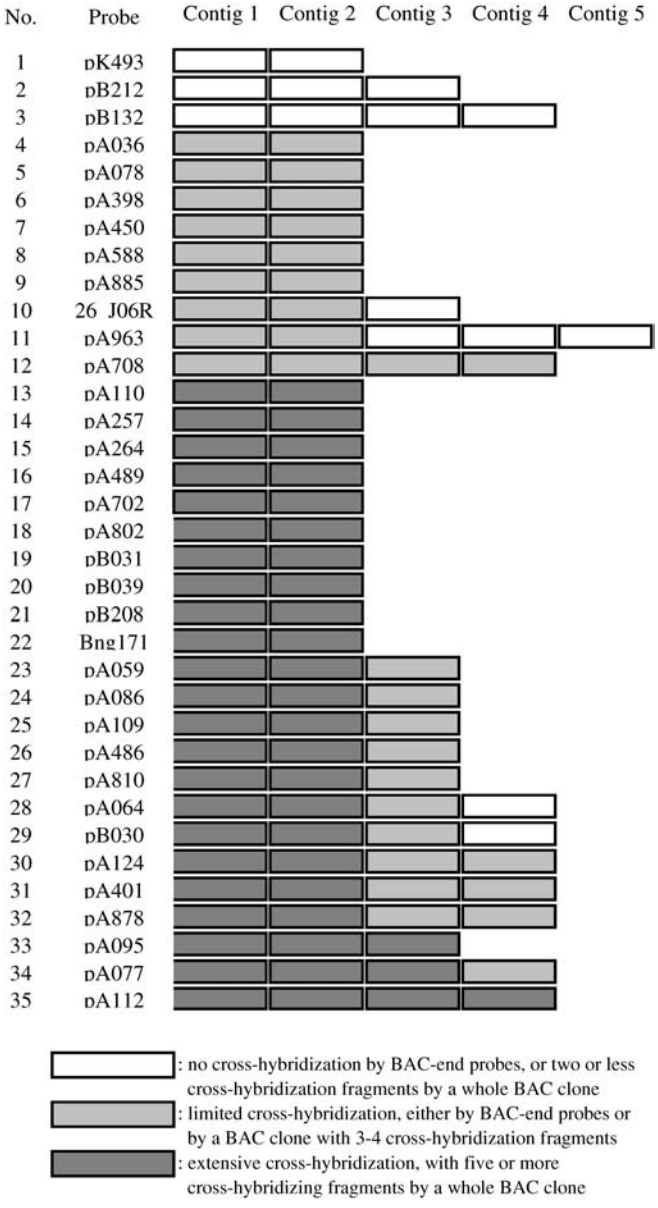

Fig. 2 Graphical representation of homoeology, as inferred from cross-hybridization experiments, among soybean BAC contigs. Both BAC-ends and whole BAC clones were used as probes in cross-hybridization experiments (see Materials and methods). For contigs in rows 1 to 3, no cross-hybridization observed; in rows 4–12, at least two contigs exhibited limited cross-hybridization; in rows 13 to 35, at least two contigs exhibited extensive cross-hybridization. In several cases (contig groups in rows 23–32) additional contigs with no or limited cross-hybridization were also observed. Contig groups in rows 33 to 35 exhibited other patterns of cross-hybridization

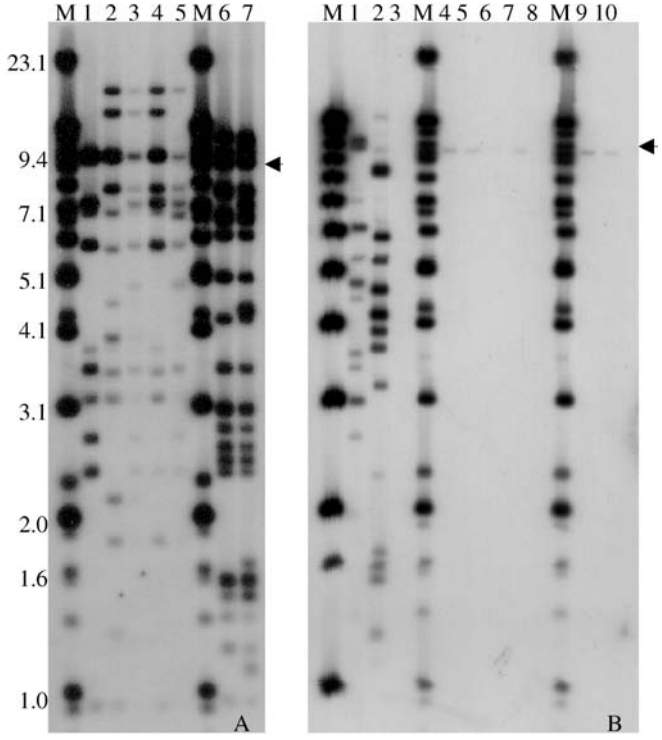

Fig. 3A, B Detailed analysis of microsynteny between two soybean contigs anchored by RFLP probe pB039. The BAC DNA filter was hybridized with a whole soybean BAC probe (panel **A**) or soybean genomic DNA as the probe (panel **B**) (See Materials and methods). BAC DNA was digested with *Eco*RI; the *arrow* indicates the vector (9.4 kb); *M*, 1-kb marker or Marra marker (in kb). Contig 1 contained five BACs (*lanes 1–5* in panel **A**, *4–8* in panel **B**) and contig 2 contained two BACs (*lanes 6–7* in panel **A**, *9–10* in panel **B**). Two soybean BACs in *lanes 1–2* of panel **B** were included as positive controls known to contain middle to highly repeated sequences, while *lane 3* in panel **B** is blank. **(A)** BAC 159_M12 (*lane 7*) in contig 2 as a probe hybridized to 6–11 fragments from homoeologous BACs in contig 1. **(B)** Soybean genomic DNA as a probe, with no hybridizing fragments visible, indicated that all fragments showing cross-hybridization by BAC 159_M12 contain only single- or low-copy DNA sequences

159_M12 in contig 2 of the pB039 contig group hybridized to at least six bands in each of the BACs from pB039 contig 1 (Fig. 3A). In all of these contigs groups, the level of soybean repetitive DNA in cross-hybridizing fragments was evaluated by hybridizing soybean genomic DNA to the restriction digested BAC DNA filters. In the case of 26 contig groups, the original cross-hybridizing bands (those indicating the existence of type II microsynteny) were barely visible after 2 days exposure (Fig. 3B; data not shown). Based on this result, we inferred that highly and middle repetitive DNA is probably not present in the cross-hybridizing fragments of these contig groups. In three contig groups (anchored by pA077, pA264 and pB031), two of the previously identified cross-hybridizing bands were found to consist of repetitive DNA. However, in these three contig groups, there were five or six additional cross-hybridizing fragments that were not visible when hybridized with soybean genome DNA. In two contig groups (anchored by probes pA481 and 52_G24R), three to eight bands were found to contain repetitive DNA, so they were not included as being microsyntenic in soybean. Altogether, a total of 29 contig groups (26 + 3) exhibited type II microsynteny in soybean.

As shown in Fig. 2, microsynteny inferrred from cross-hybridization experiments was frequently ob-

Table 3 Sequence identity among nine sets of paralogues isolated from two pB039-anchoring soybean contigs

Original contig	BAC-end in the original contig[a]	Database hit	Duplicated contig	BAC in duplicated contig[a]	Length[b] (bp)	Deletion/ insertion[c] (%)	Mutation[d] (%)	Identity (%)
Contig 1	Gm_UMb001_106_M06F	No	Contig 2	Gm_UMb001_159_M12	233	22 (9.4)	22 (9.4)	81.1
Contig 1	Gm_ISb001_027_O11F	Yes	Contig 2	Gm_UMb001_159_M12	290	4 (1.4)	12 (4.1)	94.5
Contig 1	Gm_UMb001_031_P04F	Yes	Contig 2	Gm_UMb001_159_M12	380	5 (1.3)	12 (3.2)	95.5
Contig 1	Gm_UMb001_031_P04R	No	Contig 2	Gm_UMb001_159_M12	303	2 (0.7)	19 (6.3)	93.1
Contig 1	Gm_UMb001_015_B21F	Yes	Contig 2	Gm_UMb001_159_M12	294	3 (1.0)	12 (4.1)	94.9
Contig 1	Gm_UMb001_057_P20F	Yes	Contig 2	Gm_UMb001_159_M12	645	3 (0.5)	21 (3.3)	96.3
Contig 2	Gm_ISb001_069_K02R	Yes	Contig 1	Gm_UMb001_088_I02	323	6 (1.9)	11 (3.4)	94.7
Contig 2	Gm_UMb001_008_M14R	No	Contig 1	Gm_UMb001_088_I02	455	4 (0.9)	23 (5.1)	94.1
Contig 2	Gm_UMb001_014_O09R	Yes	Contig 1	Gm_UMb001_088_I02	495	1 (0.2)	17 (3.4)	96.4
					3,418	50 (1.5)	149 (4.4)	94.2

[a] Both forward and reverse primers from each BAC-end in the original contig were used in sequencing reaction on BAC from the duplicated contig
[b] Aligned length between BAC-end sequence in the original contig and isolated paralogue from the duplicated contig
[c] Removal or addition of base pairs in either of the two compared sequences
[d] Base pair change

served within soybean contig groups. However, usually just two of the contigs within any contig group showed extensive cross-hybridization. The remaining contigs within any contig group generally had only limited or no cross-hybridization (apart from the original anchoring probe). There were three exceptions to this general observation. In the contig groups identified by probes pA095 and pA077, extensive cross-hybridization was observed among three contigs for each group. In the case of the four contigs uncovered by probe pA112, two different sub-groups were formed, with contigs 1 and 2 showing extensive cross-hybridization and contigs 3 and 4 separately showing extensive cross-hybridization (Fig. 2).

Detailed analysis of the pB039 BAC contig group

In order to characterize relationships among microsyntenic BAC contigs and at the same time provide evidence for the authenticity of results detected with whole soybean BAC clones, we examined the pB039 contig group in further detail. This group consisted of three contigs: soybean contigs 1 and 2 plus one BAC contig in *Mt*, all anchored by RFLP pB039. In our original cross-hybridization experiments, these three contigs exhibited type II microsynteny (detected by BAC clone 159_M12 in soybean contig 2). We first obtained additional low-copy BAC-end probes by re-screening the soybean and *Mt* BAC libraries with all available unique soybean and *Mt* BAC-end probes in this contig group. We then used all 43 unique BAC-end probes in cross-hybridization experiments: 15 from soybean contig 1, nine from soybean contig 2 and 19 from *Mt* contig 1. Eight probes showed cross-hybridization between soybean contig 1 and *Mt* contig 1, 11 between soybean contig 2 and *Mt* contig 1. Between the two soybean contigs, 14 probes showed cross-hybridization. We went on to isolate and sequence paralogous sequences between the two soybean contigs. We tested 12 pairs of BAC-end primers in sequencing reactions in which BACs from the duplicated contig were used as templates, with nine pairs of primers giving clear, single products. When the original BAC-end sequences were aligned with the isolated paralogous sequences, these nine sets of paralogues showed an average of 94.2% identity over 3,418 bp, ranging from 81.1 to 96.4% (Table 3). Sequence divergence was limited, resulting from insertion/deletion events (1.5%) and point mutations (4.4%). These data suggest that the three contigs associated with pB039 shared extensive microsynteny. In other experiments (data not shown) three additional contig groups (by pA685, Bng154 and pC063) are also being examined in detail through the use of comparative restriction mapping, cross hybridization and comparative sequencing of homoeologous segments. For each of these contig groups, type II microsynteny has been confirmed, with 4–20 probes showing cross-hybridization between soybean duplicated contigs and 2–5 probes showing cross-hybridization between soybean and *Mt* (data not shown).

Conserved microsynteny between soybean and *Arabidopsis*

A total of 1,085 sequences from the 50 soybean contig groups were searched against *Arabidopsis* sequences using TBLASTX. Based on the criterion of the expected value ≤1e-07 and physical separation between sequences less than 100 kb in *Arabidopsis*, seven contig groups were found to show conserved microsynteny to *At* (Tables 1 and 2). In the case of one contig anchored by NBSD_H8, three different microsyntenic regions were identified in *Arabidopsis*. In detail, BAC-ends 102_G14F and 166_A24R each have homologs on A*rabidopsis* chromosomes 1 (23.9 kb apart) and 2 (6.3 kb apart), while BAC-ends 033_G16F and 166_A24R also have homologs on *Arabidopsis* chromosome 1 (14.5 kb apart).

Discussion

Genomic conservation between soybean and *M. truncatula*

We conducted a preliminary investigation into the extent of microsynteny between 50 randomly chosen soybean BAC contig groups involving 93 contigs in total and the *Mt* genome, using a hybridization-based approach. Microsynteny with *Mt* was found for 15 contig groups using physically linked low-copy sequences as probes (type I), and for an additional 12 contig groups using whole soybean BAC as probes (type II) (Tables 1 and 2). We minimized complications due to repetitive DNA by excluding those soybean BAC clones that hybridized to five or more fragments (or a light smear of fragments) in *Mt* genomic DNA. Still, some cases of cross-hybridizing fragments revealed by a whole soybean BAC probe might still correspond to different fragments of the same gene, if that gene had one or more internal *Hin*dIII sites (*Mt* BACs were digested with *Hin*dIII). Some cross-hybridizing fragments could also be duplicates of a gene on a single *Mt* BAC clone. For these reasons, we could have overestimated the extent of conserved microsynteny. By contrast, the size of our "window" for testing microsyteny was confined to just a single BAC clone – or fewer than five short sequences distributed along a single BAC contig. For this reason, the extent of the genomic region assayed for conserved microsynteny was constrained, potentially leading to an underestimate of conserved microsynteny.

In the case of 18 soybean contig groups, one soybean BAC hybridized to at least five or more fragments from a single *Mt* BAC. It seems likely that at least some of these BAC clones are cases of extensive microsynteny with the *Mt* genome. By completely sequencing such pairs of soybean and *Mt* BACs, the degree of local conservation in gene repertoire and organization, the nature and structure of intergenic sequences, and the extent of duplications and rearrangements could be examined. In fact, some of the pairs of BACs described in this study are now in the process of complete sequencing (B. Roe, University of Oklahoma, personal communication).

Only about 16.6 Mb (assuming an average BAC contig size of 180 kb) or just approximately 1.4% of the soybean genome was examined in this study, so we still do not know the extent to which these results are representative of the soybean or *Mt* genomes as a whole. Indeed, most of the soybean contigs examined in this study were anchored by RFLP clones from a *Pst*I genomic library (Keim et al. 1990), and therefore were presumably associated with hypomethylated and low-copy DNA (Burr et al. 1988). Nonetheless, our observation that between 30 and 54% of soybean contig groups show microsynteny to *Mt* (Table 1) may be a good indicator for the existence of widespread microsynteny between these two species. The use of probes from additional soybean contig groups would help to verify this conclusion.

Genome duplication and microsynteny

In this study, we found eight cases of homoeologous contigs in soybean with conserved microsynteny to the same *Mt* BAC. This result is not unexpected since all of these soybean contigs showed at least some microsynteny with one another (Table 2; Fig. 2). This two-to-one relationship is most likely the result of homoeologous sequences retained in both soybean contigs after duplication. It is also conceivable that the two soybean contigs show microsynteny to different subregions of the same *Mt* BAC, which would imply the existence of two sets of orthologous genes. We also uncovered five cases of a single soybean contig being microsyntenic to two *Mt* BACs in different contigs. Such one-to-two relationships potentially reflect segmental duplications in the *Mt* genome. Southern hybridization of *Mt* and soybean BAC DNA filters supports this inference. If we estimate the copy number for each soybean probe in both the soybean and *Mt* genomes by treating bands of different sizes as distinct copies, we observe a minimal estimate of 2.6 copies per probe in *Mt* compared with 4.3 copies in soybean. Some duplication events may be ancient and predate the divergence of soybean and *Mt*. Direct evidence for this comes from soybean probes pA078 and pA257. Both of these RFLPs anchored two contigs microsyntenic within the soybean genome, each with contig I having two counterparts (represented by two different BACs) in the *Mt* genome, where they are presumably homoeologous.

Exceptions to conserved microsynteny

Not surprisingly, there were 23 soybean contig groups that lacked any indication of microsynteny with *Mt*, as well as 62 probes from 34 contig groups that did not detect *Mt* BACs at all. These observations raise several possibilities: (1) the counterparts do not exist in the common ancestor or were deleted from the *Mt* lineage after divergence from soybean; (2) the counterparts in *Mt* or the probe sequences in soybean have evolved to the point that similarity is unrecognizable by the hybridization method used; (3) the counterparts were not included in this *Mt* BAC library; or (4) rearrangements have disrupted their physical linkage, since in many cases multiple sequences from one soybean contig had homologs in *Mt*, but they were not physically close to one another. Potentially, further study of genomic regions that have diverged between soybean and *Mt* will reveal the basis of disruptive modes of genome evolution that may have occurred.

Implications of microsynteny

Microsynteny between the *Mt* and soybean genomes has great practical implications. Gene cloning in soybean has been difficult, partly because of its large genome size, abundant repetitive DNA and extensive duplications (Shoemaker et al. 1996; Marek et al. 2001). Compared

to soybean and most other agriculturally important legumes, *Mt* has a relatively compact and simple genome (Kulikova et al. 2001). It is also the target of extensive genome sequencing and functional genomic studies (Cook 1999). Our results suggest it may frequently be possible to use the *Mt* genome as a reference for isolating genes in soybean and other legumes, assuming the counterparts reside in conserved microsyntenic regions. In fact, this approach has recently been used in pea (*Pisum sativum*), where conserved macro- and micro-synteny with the *Mt* genome have been used to construct a fine genetic map around the *Sym2* locus (Gualtieri et al. 2002). In alfalfa (*Medicago sativa*), high levels of synteny between *M. sativa* and *Medicago truncatula* in the Nod region have been successfully used to clone a gene essential for Nod-factor perception in alfalfa (Endre et al. 2002).

Genomic conservation between soybean and *Arabidopsis*

Because the complete *Arabidopsis* genome has been sequenced and deposited in public databases, detailed analysis of genome conservation is now possible (The *Arabidopsis* Genome Initiative 2000). Sequence comparisons between the *Arabidopsis* genome and the soybean genomic regions examined here contribute to our understanding of the approximate 90 million years of evolution separating these two taxa (Grant et al. 2000). Macrosynteny between three soybean linkage groups (A2, L and J) and *Arabidopsis* has been reported previously (Grant et al. 2000). We observed microsynteny to *Arabidopsis* for seven of the 50 soybean contig groups (14%) (Tables 1 and 2), implying a much higher level of genome divergence between soybean and *Arabidopsis* than between soybean and *Mt*. As expected, these seven contig groups also showed conserved microsynteny with the *Mt* genome. Presumably, they are all descended from common genomic segments. Nevertheless, a lack of contiguous sequence data in soybean and *Mt* prevents us from making detailed comparisons. Moreover, the existence of extensive duplications and frequent gene loss in *Arabidopsis* (Ku et al. 2000) make it difficult to identify the real orthologs among these three species.

Genomic conservation between duplicated contigs in soybean

An earlier study that compared the relative positions of RFLP probes across nine different mapping populations revealed that the soybean genome is highly duplicated (Shoemaker et al. 1996). Here we examined microsynteny among several homoeologous regions, i.e., groups of two to five homoeologous contigs anchored by the same low-copy sequence. We observed limited or extensive cross-hybridization among such contigs in 32 of 37 contig groups. These results imply the existence of several large-scale or whole genome duplications that may have occurred in the lineage leading to present day soybean. Multicopy contigs that share extensive cross-hybridization could represent the products of recent duplication events, or of an ancient duplication evolving at a slow rate. Of the 29 cases of cross-hybridization in soybean detected with whole BAC probes, 20 were also detected by BAC-end probes. Therefore, we believe that a majority of the cross-hybridizations detected by whole BAC probes do represent real cases of conserved microsynteny, although we can not completely exclude the presence of repetitive DNA in cross-hybridizing fragments. A preliminary sequence comparison between nine sets of paralogues from two homoeologous contigs (anchored by pB039) indicated an average similarity of 94.2%, with both insertion/deletion events and point mutations being observed.

Results from the cross-hybridization experiments provide evidence for the existence of microsynteny between 30 and 54% of the soybean contig groups and the *Mt* genome. As legume genomic research moves forward, additional soybean BAC contigs and DNA sequence data will become available, making it possible to examine additional cases of microsynteny between the two genomes (and with *Arabidopsis*). Obviously, a high level of fine-scale genome conservation provides a good reason to capitalize on *Mt* to help understand the organization and evolution of legume genomes. Our data suggest that most contig-size homoeologous segments have maintained high levels of conservation within the soybean genome. Further efforts will be needed to understand just how far microsyntenic regions extend, or for the principal mechanism(s) responsible for the divergence observed among homoeologous segments.

Acknowledgements We thank S. Cannon, R. Denny, D. Danesh, D. Foster-Hartnett, K. Larson and L. Marek for helpful discussions and for providing valuable materials. We also thank R. Staggs and T. Schmidt for assistance with some computational analysis. This research was supported by National Science Foundation grants DBI 98-72664 and DBI 98-72565. This paper is published as part of the series of the Minnesota Agricultural Experiment Station.

References

Bennetzen JL (2000) Comparative sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions. Plant Cell 12:1021–1029

Blanc G, Barakat A, Guyot R, Cooke R, Delseny M (2000) Extensive duplication and reshuffling in the *Arabidopsis* genome. Plant Cell 12:1093–1101

Burr B, Burr R, Thompson K, Albertson M, Stuber C (1988) Gene mapping with recombinant inbreds in maize. Genetics 118:519–526

Cavell AC, Lydiate DJ, Parkin IAP, Dean C, Trick M (1998) Colinearity between a 30–centimorgan segment of *Arabidopsis thaliana* chromosome 4 and duplicated regions within the *Brassica napus* genome. Genome 41:62–69

Chen M, SanMiguel P, de Oliveira AC, Woo S-S, Zhang H, Wing RA, Bennetzen JL (1997) Microcolinearity in *sh*2-homologous regions of the maize, rice and sorghum genomes. Proc Natl Acad Sci USA 94:3431–3435

Chen M, SanMiguel P, Bennetzen JL (1998) Sequence organization and conservation in *sh2/ a1*-homologous regions of sorghum and rice. Genetics 148:435–443

Cook DR (1999) *Medicago truncatula* – a model in the making. Curr Opin Plant Biol 2:301–304

Cook DR, VandenBosch K, de Bruijn FJ, Huguet T (1997) Model legumes get the nod. Plant Cell 9:275–280

Danesh D, Penuela S, Mudge J, Denny RL, Nordstrom H, Martinez JP, Young ND (1998) A bacterial artificial chromosome library for soybean and identification of clones near a major cyst nematode resistance gene. Theor Appl Genet 96:196–202

Devos KM, Gale MD (2000) Genome relationships: the grass model in current research. Plant Cell 12:637–646

Devos KM, Beales J, Nagamura Y, Sasaki T (1999) *Arabidopsis*–rice: will colinearity allow gene prediction across the eudicot–monocot divide? Genome Res 9:825–829

Endre G, Kereszt A, Kevei Z, Mihacea S, Kaló P, Kiss GB (2002) A receptor kinase gene regulating symbiotic nodule development. Nature 417:962–966

Gale MD, Devos KM (1998) Plant comparative genetics after 10 years. Science 282:656–659

Gualtieri G, Kulikova O, Limpens E, Kim DJ, Cook DR, Bisseling T, Geurts R (2002) Microsynteny between pea and *Medicago truncatula* in the *SYM2* region. Plant Mol Biol 50:225–235

Gaut BS, Doebley JF (1997) DNA sequence evidence for the segmental allotetraploid origin of maize. Proc Natl Acad Sci USA 94:6809–6814

Grant D, Cregan P, Shoemaker RC (2000) Genome organization in dicots: genome duplication in *Arabidopsis* and synteny between soybean and *Arabidopsis*. Proc Natl Acad Sci USA 97:4168–4173

Jackson SA, Cheng Z, Wang ML, Goodman HM, Jiang J (2000) Comparative fluorescence in situ hybridization mapping of a 431-kb *Arabidopsis thaliana* bacterial artificial chromosome contig reveals the role of chromosomal duplications in the expansion of the *Brassica rapa* genome. Genetics 156:833–838

Keim P, Diers B, Olson T, Shoemaker RC (1990) RFLP mapping in soybean: association between marker loci and variation in quantitative traits. Genetics 126:735–742

Keller B, Feuillet C (2000) Colinearity and gene density in grass genomes. Trends Plant Sci 5:246–251

Ku H-M, Vision T, Liu J, Tanksley SD (2000) Comparing sequenced segments of the tomato and *Arabidopsis* genomes: large-scale duplication followed by selective gene loss creates a network of systeny. Proc Natl Acad Sci USA 97:9121–9126

Kulikova O, Gualtieri G, Geurts R, Kim DJ, Cook DR, Huguet T, de Jong JH, Fransz PF, Bisseling T (2001) Integration of the FISH-pachytene and genetic maps of *Medicago truncatula*. Plant J 27:49–58

Lagercrantz U (1998) Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that *Brassica* genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. Genetics 150:1217–1228

Marek LF, Shoemaker RC (1997) BAC contig development by fingerprint analysis in soybean. Genome 40:420–427

Marek LF, Mudge J, Darnielle L, Grant D, Hanson N, Paz M, Yan H, Denny R, Larson K, Foster-Hartnett D, Cooper A, Danesh D, Larsen D, Schmidt T, Staggs R, Crow JA, Retzel E, Young ND, Shoemaker RC (2001) Soybean genomic survey: BAC-end sequences near RFLP and SSR markers. Genome 44:572–581

Marra MA, Kucaba TA, Dietrich NL, Green ED, Brownstein B, Wilson RK, McDonald KM, Hillier LW, McPherson JD, Waterston RH (1997) High throughput fingerprint analysis of large-insert clones. Genome Res 7:1072–1084

Mayer K, Murphy G, Tarchini R, Wambutt R, Volckaert G, Pohl T, Dusterhoft A, Stiekema W, Entian KD, Terryn N, Lemcke K, Haase D, Hall CR, van Dodeweerd AM, Tingey SV, Mewes HW, Bevan MW, Bancroft I (2001) Conservation of microstructure between a sequenced region of the genome of rice and multiple segments of the genome of *Arabidopsis thaliana*. Genome Res 11:1167–1174

Nam Y-W, Penmetsa RV, Endre G, Uribe P, Kim D, Cook DR (1999) Construction of a bacterial artificial chromosome library of *Medicago truncatula* and identification of clones containing ethylene-response genes. Theor Appl Genet 98:638–646

Shoemaker RC, Polzin K, Labate J, Specht J, Brummer EC, Olson T, Young N, Concibido V, Wilcox J, Tamulonis JP, Kochert G, Boerma HR (1996) Genome duplication in soybean (*Glycine* subgenus *soja*). Genetics 144:329–338

Tarchini R, Biddle P, Wineland R, Tingey S, Rafalski A (2000) The complete sequence of 340 kb of DNA around the rice *adh1-adh2* region reveals interrupted colinearity with maize chromosome 4. Plant Cell 12:381–391

The *Arabidopsis* Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

Tikhonov AP, SanMiguel PJ, Nakajima Y, Gorenstein NM, Bennetzen JL, Avramova Z (1999) Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum. Proc Natl Acad Sci USA 96:7409–7414

Van Dodeweerd A-M, Hall CR, Bent EG, Johnson SJ, Bevan MW, Bancroft I (1999) Identification and analysis of homoeologous segments of the genomes of rice and *Arabidopsis thaliana*. Genome 42:887–892

Vision TJ, Brown DG, Tanksley SD (2000) The origins of genomic duplications in *Arabidopsis*. Science 290:2114–2117

Wong GKS, Yu J, Thayer EC, Olson MV (1997) Multiple-complete-digest restriction fragment mapping: generating sequence-ready maps for large-scale DNA sequencing. Proc Natl Acad Sci USA 94:5225–5230