

A BAC and BIBAC-based Physical Map of the Soybean Genome

Chengcang Wu^a, Shuku Sun^a, Padmavathi Nimmakayala^a, Felipe A. Santos^a, Rachael Springman^a, Kejiao Ding^a, Khalid Meksem^b, David A. Lightfoot^b and Hong-Bin Zhang^{a,1}

^aDepartment of Soil and Crop Sciences and Institute for Plant Genomics and Biotechnology, 2123 TAMU, Texas A&M University, College Station, Texas 77843-2123, USA

^bDepartment of Plant Soil and General Agriculture, Southern Illinois University, Carbondale, Illinois 62901-4415, USA

Running Title: A Physical Map of the Soybean Genome

¹To whom correspondence should be addressed. E-mail: hbz7049@tamu.edu; fax 979-862-4790.

ABSTRACT

We report a genome-wide, bacterial artificial chromosome (BAC) and plant-transformation-competent binary large-insert plasmid clone (hereafter BIBAC)-based physical map of the soybean genome. The map was constructed from 78,001 clones from five soybean BAC and BIBAC libraries representing 9.6 haploid genomes and three cultivars. The map consisted of 2,905 BAC/BIBAC contigs, and was estimated to span 1,408 Mb in physical length. The physical length is about 293 Mb greater than the expected 1,115-Mb genome size of the species, indicating that most, if not all, of the contigs remain overlapped. We evaluated the reliability of the map contigs using different contig assembly strategies, independent contig building methods, DNA marker screening results of the BACs and BIBACs and different fingerprinting methods, and the results showed that the map was assembled properly. Furthermore, we have integrated 781 of the contigs spanning 663 Mbp (59.5%) of the soybean genome into the existing soybean composite genetic map using 273 SSR and 115 RFLP markers. This map represents the first genome-wide, BAC/BIBAC-based physical map of soybean and will provide a powerful platform for many areas of soybean genome research, including large-scale genome sequencing, target marker development, gene mapping, and gene and quantitative trait locus (QTL) cloning. The inclusion of BIBACs in the map will further streamline the utility of the map for positional cloning of genes and QTLs, and functional analysis of soybean genomic sequences.

[The supplemental material on the clone fingerprint database and contigs of the physical map is available online at <http://hbx.tamu.edu>. The following individuals kindly provided reagents, samples or unpublished information as indicated in the paper: R. Shoemaker, N. D. Young, Z. Xu, and Y.-L. Chang.]

INTRODUCTION

Soybean, *Glycine max* (L.) Merr., is the world's top legume crop and foremost source of edible plant oil and proteins. To develop tools essential for continued genetic improvement of the crop, DNA marker-based genetic linkage maps have been developed (e.g., Lark et al. 1993; Shoemaker and Specht 1995; Iqbal et al., 2001; Keim et al. 1997; Cregan et al. 1999; <http://soybase.agron.iastate.edu>), 93 genes and more than 900 quantitative trait loci (QTLs) of agronomic importance have been mapped with the genetic maps (<http://soybase.agron.iastate.edu>), several large-insert bacterial artificial chromosome (BAC) and plant-transformation-competent binary plasmid clone (hereafter BIBAC) libraries have been constructed (Marek and Shoemaker 1997; Danesh et al. 1998; Salimath and Bhattacharyya 1999; Meksem et al. 2000), and a large collection of expressed sequence tags (ESTs) have been generated (Shoemaker et al. 2002; <http://soybean.ccg.umn.edu>). However, further advances, such as development of DNA markers for a genomic region of interest for fine mapping of genes and QTLs, isolation of clones containing a gene and/or QTL of interest for positional cloning, large-scale mapping of developed ESTs, and large-scale genome sequencing, are limited due to the shortage of essential and powerful infrastructure.

Genome-wide integrated physical and genetic maps have provided powerful tools and infrastructure for advanced genomics research. They are not only crucial for large-scale genome sequencing (Hodgkin et al. 1995; Adams et al. 2000; Arabidopsis Genome Initiative 2000; International Human Genome Sequencing Consortium 2001), but also provide powerful platforms essential for many other aspects of genome research, including targeted DNA marker development, efficient positional cloning, and high-throughput EST mapping (Zhang and Wu 2001). Whole-genome physical maps have been constructed for *Caenorhabditis elegans* (Coulson et al. 1986; Hodgkin et al. 1995), *Arabidopsis thaliana* (Marra et al. 1999; Mozo et al. 1999; Chang et al. 2001), *Drosophila melanogaster* (Hoskins et al. 2000), human (International Human Genome Mapping Consortium 2001), rice (*Oryza sativa*) (Tao et al. 2001; Chen et al. 2002), and mouse (*Mus musculus*) (Gregory et al. 2002). However, no genome-wide physical map has been developed for soybean.

Several approaches have been developed to construct whole-genome physical maps with large-insert BAC and BIBAC clones (Gregory et al. 1997; Marra et al. 1997; Zhang and Wing 1997; Tao and Zhang 1998; Ding et al. 1999; Zhang and Wu 2001). We helped pioneer the strategies and technologies of whole-genome physical mapping from BAC and BIBAC clones by restriction fingerprint analysis on DNA sequencing gels (Zhang and Wing 1997; Tao and Zhang 1998). The DNA sequencing gel-based fingerprinting method (Coulson et al. 1986; Gregory et al. 1997; Zhang and Wing 1997; Tao and Zhang 1998; Ding et al. 1999; Zhang and Wu 2001) not only has a significantly higher resolution (= one nucleotide) than that of the agarose gel-based method (10 – 500 bp; Marra et al. 1997; Zhang and Wu 2001), but is also economical and highly amenable to analysis by automated DNA sequencers (Gregory et al. 1997; Ding et al. 1999; Z. Xu, Y.-L. Chang, K. Ding and H.-B. Zhang, unpublished) and to high throughput technologies (Zhang and Wu 2001; Z. Xu, Y.-L. Chang, K. Ding and H.-B. Zhang,

unpublished; Z. Xu, S. Sun and H.-B. Zhang, unpublished). Using these techniques and strategies, we previously developed a BAC/BIBAC-based integrated physical and genetic map of *Arabidopsis* (Chang et al. 2001) and a whole-genome BAC-based physical map of *O. sativa* ssp. *indica* rice (Tao et al. 2001).

Soybean has a genome size of 1115 Mb/1C (Arumuganathan and Earle 1991), and approximately 40 - 60% of its genome is repetitive sequence and heterochromatic (Goldberg 1978; Gurley et al. 1979; Singh and Hymowitz 1988). Although the genome of soybean is smaller in size than the genomes of human and mouse, for which BAC-based physical maps have been developed (International Human Genome Mapping Consortium 2001; Gregory et al. 2002), development of a genome-wide physical map of the soybean genome is more difficult. Soybean is a recently diploidized tetraploid (last duplication only 8 MYA) and has an average of 2.55 duplicated segments with as many as six copies per gene (Shoemaker et al. 1996).

Efforts were made to develop a regional, BAC-based physical map of the soybean genome using the soybean cv. Williams 82 and cv. Faribault BAC libraries. However, the map only covered approximately 20% of the soybean genome (Marek et al. 2001). Here we report a genome-wide, BAC and BIBAC-based physical map of the soybean genome. This map was constructed from 78,001 (representing 9.6 haploid genomes) clones from five large-insert BAC and BIBAC libraries of three soybean cultivars, Forrest, Williams 82 and Faribault, using the high-resolution and high-throughput DNA sequencing gel-based fingerprinting method (Zhang and Wing 1997; Chang et al. 2001; Tao et al. 2001). We have also integrated 781 physical map contigs spanning 663 Mb (59.5%) of the soybean genome into the existing soybean composite genetic linkage map by screening the Forrest BAC/BIBAC libraries with 22 DNA markers and fingerprinting the positive BAC clones identified from the Williams 82 and Faribault BAC libraries with 366 DNA markers (Marek et al. 2001). The results have further confirmed the nature of the ancient tetraploid origin of soybean and demonstrated that it is feasible to integrate the physical map with the genetic linkage map using genetically mapped DNA markers. The reliability of the physical map was verified using several different approaches. The results indicate that the physical map will provide a readily used framework for advanced genomics research of soybean and related legume species.

RESULTS

Source DNA Libraries and Fingerprinting

Our previous studies showed that the actual genome coverage of a large-insert BAC library constructed from a single enzyme partial digestion was about 15% lower than the theoretical genome coverage of the library because of the uneven distribution of the enzyme recognition sites in the genome (Tao et al., 2001; Zhang and Wu, 2001). Therefore, it is preferable to use BAC libraries constructed with different enzymes to develop a whole-genome physical map of high genome coverage. Furthermore, given that it is difficult to sub-clone the entire insert of a large-insert BAC into a plant-transformation-competent binary vector for plant transformation (Zhang and Wu 2001),

the incorporation of large-insert plant-transformation-competent BIBACs will further enhance the utility of the physical map for positional cloning and experimental determination of the functions of every segment of the genome. To develop the whole-genome physical map of soybean, we used one BAC (C. Wu and H.-B. Zhang, unpublished; <http://hbz.tamu.edu>) and two BIBAC (Meksem et al., 2000; <http://hbz.tamu.edu>) libraries constructed from the partial digests of soybean cv. Forrest DNA with *Eco* RI, *Bam* HI and *Hind* III, respectively. Each of the *Eco* RI BAC and *Bam* HI and *Hind* III BIBAC libraries consists of 38,400 clones and has an average insert size of 157, 125 and 125 kb, respectively. We also included the other two existing soybean BAC libraries, which were constructed from soybean cv. Williams 82 (Marek and Shoemaker, 1997) and cv. Faribault (Danesh et al., 1998), respectively. The Williams 82 library was constructed with *Hind* III and has an average insert size of 150 kb, and the Faribault library was constructed with *Eco* RI and has an average insert size of 120 kb. The two libraries were screened with 389 SSR markers and 223 RFLP markers, respectively, to construct the regional physical map of the soybean genome (Marek et al., 2001). Inclusion of the positive clones of 267 SSR and 105 RFLP markers in this study will allow not only the integration of the regional physical map with the genome-wide physical map developed in this study, but also the integration of the whole-genome physical map with the existing genetic linkage composite map (Cregan et al., 1999; <http://soybase.agron.iastate.edu/>).

We fingerprinted a total of 84,946 BACs and BIBACs from the five BAC and BIBAC libraries (Table 1) on 1,332 autoradiographs using the DNA sequencing gel-based restriction fingerprinting method (Zhang and Wing, 1997; Chang et al., 2001; Tao et al., 2001). The autoradiographs of the BAC and BIBAC fingerprints were scanned into image files and edited with the Image program of the FPC package (Sulston et al., 1988; Soderlund et al., 1997). Fingerprints for a total of 78,001 clones were successfully obtained, edited and created into FPC database, whereas a total of 6,945 clones (8.18%) were deleted because they either failed in fingerprinting, had no inserts or had inserts that only gave four or fewer bands. Of the 78,001 clones representing 9.580x genome equivalents, the clones equivalent to 3.052x, 2.262x, 4.030x, 0.121x, and 0.115x haploid genomes of soybean were from the Forrest *Hind* III BIBAC library, the Forrest *Bam* HI BIBAC library, the Forrest *Eco* RI BAC library, the Faribault *Hind* III BAC library, and the Williams 82 *Eco* RI BAC library, respectively. To minimize the influence of the lower (> 1 base) resolution fingerprint bands of higher-molecular-weight at the top portion of the autoradiographs on the accuracy of physical map contig assembly, we used only the bands between 58 and 773 bases for the soybean physical map assembly. Consequently, an average band number of 27.63 for the Forrest *Hind* III BIBACs, 23.43 for the Forrest *Bam* HI BIBACs, 39.12 for the Forrest *Eco* RI BACs, 28.46 for the Faribault *Hind* III BACs, and 32.03 for the Williams 82 *Eco* RI BACs (Table 1) was used in the physical map assembly of the soybean genome.

Physical Map Contig Assembly and Manual Editing

The FPC database of 78,001 BAC and BIBAC fingerprints were subjected to overlap analysis using the computer program FPC 4.7 (Soderlund et al., 2000). The FPC program

assembled 4,792 overlapping BAC/BIBAC contigs using the variable cutoffs ranging from $1e-30$ to $1e-10$ and a fixed tolerance of 2, whereas 4,933 clones remained as singletons (Table 2). The physical length of the automated contigs was estimated to be 1481.5 Mb, based on 364,908 unique bands with each being equivalent to 4.06 bp (Table 2). The total physical length of the contigs is obviously larger than the estimated 1115-Mb genome size of soybean (Arumuganathan and Earle, 1991), indicating that most, if not all, of the contigs were expected to overlap adjacent contigs even though the overlaps were not detected under the conditions used, and/or that the genome size of soybean was underestimated.

To verify and extend the contigs, we manually edited each of them using two methods. First, we manually checked every contig and disassembled potential chimeric contigs that were apparently not overlapped according to the clone fingerprint patterns, or that apparently conflicted with either DNA marker data or the existing soybean BAC contig data (Marek et al., 2001). Then all questionable contigs were split or killed. Second, to identify potential junctions between contigs, we searched the entire FPC fingerprint database for matches to the terminal clone fingerprints of every contig using the End Extension function of the FPC program with the cutoffs ranging from $1e-28$ to $1e-10$. We merged the contig pairs if their terminal clones shared 10 or more bands and their overall fingerprint patterns supported the junction. We also coalesced the contig pairs if they hybridized with two or more neighboring DNA markers and could be merged into a single contig using the cutoff values between $1e-15$ and $1e-10$. As a result, the total number of contigs of the physical map was reduced to 2,905, with 4,954 clones (6.35%) remaining as singletons (Table 2). The 2,905 contigs consisted of 346,884 unique bands, collectively spanning 1,408 Mb in physical length. The longest contig (ctg127) contained 319 clones, encompassing 1,345 unique bands and spanning 5.5 Mb in physical length. The fingerprint database of all 78,001 BACs and BIBACs and all contigs of the soybean physical map are posted at <http://hbz.tamu.edu> and made available to the public. Figure 1 shows an example of the contigs of the physical map and the distribution of the BACs and BIBACs from the five soybean libraries within the contig.

Integration of the Soybean Physical Map with the Existing Soybean Genetic Map

To test the feasibility of anchoring the physical map contigs to the existing soybean composite genetic map (Cregan et al., 1999; <http://soybase.agron.iastate.edu/>), to further verify the reliability of the contigs and to increase the utility of the soybean physical map, we anchored a number of the contigs of the physical map to the 20 molecular linkage groups (MLGs) of the soybean genetic map (Cregan et al., 1999; <http://soybase.agron.iastate.edu/>). We screened the Forrest *Eco* RI BAC or *Hind* III BIBAC libraries by colony filter hybridization with 7 RFLP markers and 15 SSR markers. From one to ten positive clones for each probe were identified (Table 3). The results obtained from SSR markers were further confirmed by PCR-based BAC library screening (<http://www.siu.edu/~pbgc/DataBase/datap1.htm>). All of the seven RFLP markers were shown to be multiple-copy in the soybean haploid genome by Southern analysis (also see <http://soybase.agron.iastate.edu/>), and the positive clones identified with each of these DNA markers were located to multiple contigs. In the case of the SSR

markers, the positive BACs identified by each of eight of the 15 SSR markers were observed in a single contig. The positive clones of each of the remaining 7 SSR markers (47%) were observed in two or more contigs, consistent with the agarose gel analysis result of their PCR products indicating that they have multiple loci in the genome.

We also integrated the regional physical map data of soybean (Marek et al., 2001) into the whole-genome physical map constructed in this study. Using the method described above, we fingerprinted the 2,002 positive clones from the Williams 82 and Faribault BAC libraries identified using 267 SSR and 105 RFLP DNA markers (Marek et al., 2001). After editing, fingerprint data were successfully obtained from 1,851 of the 2,002 BACs, which contained 264 SSRs and 102 RFLPs. We then integrated the BAC fingerprint database with our whole-genome BAC fingerprint database and used the combined data for whole-genome physical map contig assembly.

The efforts of Forrest BAC and BIBAC library screening with the addition of the positive BACs of the Williams 82 and Faribault libraries allowed 781 contigs of the physical map to be integrated with the 20 MLGs of the existing soybean genetic map (Cregan et al. 1999; <http://soybase.agron.iastate.edu/>). The 781 contigs encompassed 163,400 unique bands, covering approximately 663 Mb (59.5%) of the 1115-Mb soybean genome. Of the 388 markers (115 RFLPs and 273 SSRs) used in the integration, 26, 28, 10, 9, 5, 28, 34, 4, 8, 22, 29, 36, 24, 5, 25, 24, 9, 6, 11, 24, and 21 were from MLGs A1, A2, B1, B2, C1, C2, D1a+Q, D1b+W, D2, E, F, G, H, I, J, K, L, M, N, O, and unlinked group, respectively (Figure 2). The positive BACs of each of all 115 RFLP markers, except for one (A469), were located to two or more contigs because they have multiple loci in the soybean genome (Marek et al. 2001), whereas the positive clones of each of 82 of the 273 SSR markers (30.0%) were located to a single contig, suggesting a single locus in the soybean genome if a contig is assumed to represent one locus. Therefore, the 83 contigs of the single-locus markers (1 RFLP and 82 SSRs) were unambiguously anchored to the soybean genetic map. In addition, 16 contigs were hybridized with two or more neighboring markers and thus were also unambiguously anchored to the soybean genetic map. The exact positions of the remaining contigs will be refined as additional anchor markers are integrated into the map.

Physical Map Contig Reliability

We evaluated the reliability of the soybean physical map contigs using several approaches. First, we compared contig assembly using two different strategies. We assembled the physical map contigs using individual stepwise cutoff values between $1e-30$ and $1e-10$ and using the cutoff values $1e-25$, $1e-20$, $1e-15$, and $1e-10$, respectively. In the latter contig assembly strategy, the contigs that were obviously chimeric were disassembled and reassembled using higher stringency cutoff values. One thousand contigs were randomly selected from the contigs generated by these two strategies and compared. The result showed that 99.1% of the automated contigs were completely consistent in both clone content and order. In our second approach, we assembled separate contigs from the clones of the individual Forrest *Bam* HI, the Forrest *Eco* RI, the Forrest *Hind* III, and the Williams 82/Faribault libraries. We randomly selected 100

contigs of the clones from each of the three Forrest libraries and all 389 contigs of the Williams 82/Faribault libraries, and compared them with their corresponding contigs from the physical map. Ninety-three percent (93%), 97%, 96% and 96% of the contigs were shown to be in complete agreement in both clone content and order. For our third approach, we compared 141 RFLP-anchored contigs constructed independently by digesting the marker-positive BACs with a restriction enzyme, followed by Southern hybridization with relevant DNA markers (Marek et al., 2001) against the corresponding contigs of the physical map constructed in this study. One hundred twenty-five of the 141 contigs (88.7%) were shown to be completely consistent in both clone content and order. For our fourth approach, we randomly selected ten contigs of the physical map, fingerprinted using a different method with two sets of different enzymes and then re-assembled the contigs. As a result, the fingerprints of the BACs from each contig were very similar, and the same contig was re-assembled (not shown). Finally, we checked the positions of the positive clones of each of the single-locus DNA markers in the physical map. The result showed that the positive clones of every single-locus DNA marker located to the corresponding region of a single contig (e.g., see Table 3), indicating that the contigs are properly assembled.

DISCUSSION

We have successfully fingerprinted 78,0001 clones from five soybean BAC and BIBAC libraries representing a 9.6-fold haploid genome redundancy, created FPC database for the clone fingerprints and constructed a genome-wide physical map of the soybean genome. The map consists of 2,905 contigs, estimated to span 1,408 Mb. The total physical length of the contigs is approximately 293 Mb (26.3%) greater than the 1115-Mb genome size of soybean genome (Arumuganathan and Earle, 1991). This indicates that most, if not all, of the contigs overlap adjacent contigs although the overlaps could not be detected under the conditions used, and/or that the genome size of soybean was underestimated. According to our (Chang et al., 2001; Tao et al., 2001) and other (Marra et al., 1999; International Human Genome Mapping Consortium, 2001; Chen et al., 2002; Gregory et al., 2002) previously reported physical mapping results, the failure to detect overlaps between some contigs is likely to be the main cause of the discrepancy between the total physical length of the contigs and the estimated soybean genome size. Therefore, the physical map contigs could further be merged, and the map could further be refined if additional DNA markers are used to anchor the contigs to the soybean genetic map. The contigs as well as the clone content and order within the contigs have been confirmed by using different contig assembly strategies, independent contig building methods, different fingerprinting methods, and DNA marker screening results of the BACs and BIBACs (see Table 3). These results consistently indicated that the contigs of the soybean physical map were properly assembled and thus, are suitable for different aspects of advanced genome research of soybean and related species.

This study represents the first report of development of a genome-wide physical map of the soybean genome. The physical map not only has provided a platform for large-scale genome sequencing, but also is suitable for effective positional cloning, large-scale EST mapping and comparative genome analysis of soybean and related legume

species. For clone-by-clone shotgun genome sequencing, the physical map has provided an essential, readily usable platform. Minimally overlapping clone tiling paths representing any region of the soybean genome can be directly selected from the constructed contigs. Alternatively, they can also be constructed by electronic chromosome walking using the FPC database generated in this study and the FPC Hitting Tool provided (<http://hbz.tamu.edu>). The minimally overlapping clone paths are essential not only for large-scale clone-by-clone genome sequencing, but also for target marker development, large-scale EST mapping and positional cloning. For whole-genome shotgun sequencing, the physical map has provided a framework for whole-genome sequence map assembly if the ends of BACs and BIBACs of the physical map are sequenced and used as sequence tagged sites (STSs) for sequence contig anchoring. Furthermore, the BIBACs of the physical map have streamlined the positional cloning, genomic sequence functional analysis and gene/QTL engineering by *Agrobacterium*-mediated genetic transformation (Clemente et al., 2000; Donaldson and Simmonds, 2000). Moreover, the U.S. Legume Genomics Initiative had planned to develop a core set of at least 1,000 sequence tagged sites (STSs) that are universal among all legume species (http://129.186.26.94/Legume_Initiative/LegGenomicsPaper10Oct01.html). These markers can be used to integrate the physical and genetic maps between soybean, *M. truncatula* and *Lotus japonicus*, the two model legume species whose genomes are being sequenced. The sequence information obtained from *M. truncatula* and *L. japonicus* thus could be utilized for soybean genomics. As has been done between the mouse and human genomes (Gregory et al., 2002), the soybean physical map can also be used to construct genome-wide comparative physical maps between soybean and the legume model species. High-resolution comparative physical maps will reveal the regions of colinearity and rearrangement and will greatly facilitate map-based cloning of agronomically important genes and QTLs in the legume species.

This study has further confirmed the nature of the ancient tetraploid origin of the soybean genome. In this study, 781 of the 2,905 contigs were integrated with the existing genetic map (Cregan et al., 1999) using 388 DNA markers. Each DNA marker corresponds to an average of 2.0 contigs, with a maximum of 10 contigs per DNA marker. This result is consistent with the average of 2.5 duplicated segments and as many as six copies per gene previously estimated by Shoemaker et al. (1996), thus supporting the hypothesis that the soybean genome is an ancient tetraploid.

The ancient tetraploid nature of the soybean genome has complicated the integration of the physical map contigs with the soybean genetic map (Fig. 2). However, the results obtained here demonstrated that it is feasible to properly integrate the physical map contigs to the existing genetic map and develop a robust integrated physical and genetic map of the soybean genome. First, about 30% of the SSR markers each were shown to anchor only one contig, indicating a single locus for the contig in the soybean genome. Therefore, the contigs containing such SSR markers can be unambiguously anchored to the soybean genetic map. Second, as shown in Figure 2, quite a few of the contigs each contained two or more neighboring DNA markers mapped to the genetic map and thus were also unambiguously anchored to the soybean genetic map even though the DNA markers are multiple-copy in the soybean haploid genome. Therefore,

screening the BACs and BIBACs of the physical map with additional DNA markers, either single-locus or multi-locus, will result in rapidly and unambiguously anchoring the contigs of the physical map to the genetic map, coalescing the neighboring contigs and drastically reducing the total number of contigs representing the physical map. In this regard, the overlaps between the neighboring contigs, despite being not detected under the conditions used for genome-wide map contig assembly, will provide useful information for contig merging. In return, the coalescence of the contigs will further verify the accuracy of the map contigs.

This study demonstrated that it is feasible to construct a genome-wide, BAC and BIBAC-based physical map of a large, ancient or allopolyploid genome using the high-resolution and high-throughput DNA sequencing gel-based fingerprinting method (Zhang and Wing, 1997; Tao and Zhang, 1998; Chang et al., 2001; Tao et al., 2001; Zhang and Wu, 2001). We found from this study that the bottleneck to the construction of a genome-wide, robust integrated physical and genetic integrated map of a large, complex genome is not technical, but of a shortage in funding. The soybean contig map was developed at a total cost of about \$800,000 (\$440,000 direct cost) in six scientist-years. If sufficient funding was available, an additional 2,000 DNA markers from the soybean composite genetic map would be used as probes to screen the BACs and BIBACs of the physical map, most, if not all, of the contigs would be anchored to the soybean genetic map, and the map would be significantly enhanced.

METHODS

Source BAC and BIBAC Libraries and DNA Probes

The two BIBAC (Meksem et al., 2000) and one BAC (C. Wu and H.-B. Zhang, unpublished) libraries of soybean cv. Forrest and the BAC libraries of soybean cv. Williams 82 (Marek and Shoemaker, 1997) and cv. Faribault (Danesh et al., 1998) were used to develop the BAC and BIBAC-based physical map of the soybean genome. The two Forrest BIBAC libraries were constructed in the *Bam* HI and *Hind* III sites of pCLD04541, respectively, a binary vector that was designed for *Agrobacterium*-mediated transformation in plants (Jones et al., 1992; Tao and Zhang, 1998). The Forrest BAC library was constructed in the *Eco* RI site of the vector pECBAC1 (Frijters et al., 1997), a derivative of pBeloBAC 11 (Kim et al., 1996). The Williams 82 BAC library was constructed in the *Hind* III site of the vector pBeloBAC11, having an average insert size of 150 kb (Marek and Shoemaker, 1997), and the Faribault BAC library was constructed in the *Eco* RI site of the vector pECBAC4 (Frijters et al., 1997), having an average insert size of 120 kb (Danesh et al., 1998). These libraries, constructed from the partial digests of soybean nuclear DNA with three restriction enzymes (*Hind* III, *Bam* HI and *Eco* RI) in two different vector systems (bacterial F-factor based – pBeloBAC 11, pECBAC1 and pECBAC4, and bacterial P1 plasmid-based – pCLD04541), are complementary in genome coverage and thus would greatly facilitate development of the physical map. The small sizes (about 7.5 kb) of the BAC library vectors, pBeloBAC 11 (Kim et al., 1996), pECBAC1 and pECBAC4 (Frijters et al., 1997), will facilitate the utility of the physical map for clone-by-clone-based shotgun genome sequencing. The

transformability of the pCLD04541 BIBACs in plants will facilitate the utility of the physical map in the positional cloning of genes and QTLs important to agriculture, functional analysis of soybean genomic sequences, and gene/QTL engineering by genetic transformation. These libraries are permanently maintained in 384-well microplates and are publicly available at the GENEfinder Genomic Resources (<http://hbz.tamu.edu>).

The DNA marker probes were selected from the soybean composite genetic map (Cregan et al., 1999; <http://soybase.agron.iastate.edu/>). RFLP probes were purchased from Biogenetic Services, Inc. (Brookings, South Dakota). The inserts of the probe clones were isolated, purified on agarose gels and used as probes. SSR probes were generated by PCR amplification of the Forrest genomic DNA with the SSR primer pairs obtained from the Soybase (<http://soybase.agron.iastate.edu/>). The PCR products of the SSR primers were analyzed and purified on agarose gels. Seven hundred thirty-two RFLP-positive and 1270 SSR-positive BAC clones were selected from the Williams 82 (860 BACs) and Faribault (1142 BACs) BAC libraries (Marek et al., 2001).

BAC and BIBAC Fingerprinting and Contig Assembly

BAC and BIBAC DNA were isolated and fingerprinted according to Chang et al. (2001) and Tao et al. (2001). Briefly, BAC or BIBAC clones maintained in a 384-well microplate were inoculated in four 96-deep well plates containing 1 ml LB medium plus 12.5 µg/ml (w/v) chloramphenicol for BACs or 15 µg/ml (w/v) tetracycline for BIBACs and grown at 37°C with shaking at 250 rpm overnight. BAC or BIBAC DNA was isolated and purified in the 96-deep well plates and then in 12-microtube strips using a modified alkaline lysis method (Tao et al., 2001). The DNA was double-digested with *Hind* III and *Hae* III, end-labeled with [³³P] dATP using reverse transcriptase at 37°C for 2 h, and then subjected to 3.5% (w/v) polyacrylamide DNA sequencing gel electrophoresis at 85 W for 100 min. The gel was dried and autoradiographed.

The fingerprints on the autoradiographs were scanned into image files using a UMAX Mirage D-16L scanner and edited using Image 4.0 of the FPC package (Soderlund et al., 1997). To reduce the influence of the relatively lower resolution bands (>1 base) of larger fragments (>773 bases) at the top portion of a gel on the accurate contig assembly, only the fragments ranging from 58 to 773 bases were used for contig assembly. There was no vector-derived band of the BIBAC clones present in the fingerprint fragment range, and the vector bands derived from the BAC clones were manually removed from the data files. The clones that failed in fingerprinting, had no inserts, or produced four or fewer bands were excluded during fingerprint editing because the number of bands per fingerprint was insufficient to be included for contig assembly. Consequently, 78,001 clones equivalent to 9.580 x soybean haploid genomes were used to assemble the physical map contigs.

To select the tolerance and cutoff values that were suitable for physical map contig assembly of the soybean genome, we first assembled BAC/BIBACs contigs using different tolerance and cutoff values using the FPC program version 4.7 (Soderlund et al., 2000) and then analyzed the contigs containing the RFLP and SSR positive clones,

especially the positive clones of single-locus SSR markers (a single band on a 3.0% Metaphor agarose gel). We assumed that if the contigs were properly assembled, the positive clones of a single-locus DNA marker should be assembled to the same region of a contig. Using this criterion, we conducted a series of tests and finally, tolerance = 2 and variable cutoff values between $1e-30$ and $1e-10$ were selected for the physical map contig assembly.

The BAC/BIBAC contigs of the soybean genome physical map were assembled as follows. We first assembled automated contigs under the above criteria and manually edited every automated contig to ensure that they were accurate. Then we joined automated contigs into larger contigs using lower stringency cutoff values ranging between $1e-28$ - $1e-10$. The contig pairs were merged if their terminal clones shared 10 or more bands and their overall fingerprint patterns supported joining. We also coalesced the contig pairs if they were anchored to the genetic map (Cregan et al., 1999) by two or more neighboring DNA markers and formed a single contig under the cutoff values between $1e-15$ and $1e-10$.

BAC and BIBAC Library Screening and Integration of DNA Marker-containing BACs

The soybean BAC and BIBAC libraries or the clones of the physical map contigs were double-spotted on Hybond N + membrane (Amersham, Piscataway, NJ) in a 3 x 3 format using the Biomek 2000 robotic workstation (Beckman, Fullerton, CA). The high-density colony filters were prepared according to Zhang et al. (1996). The probes were labeled with [32 P]dCTP and the colony hybridization was performed according to Zhang et al. (1996). The filters were washed three times in 0.1% SDS, 0.5x SSC at 65°C, 30 min each wash.

The positive BAC clones of the Williams 82 and Faribault BAC libraries identified with 264 SSR and 102 RFLP markers (Marek et al., 2001) were successfully fingerprinted, edited, and assembled into contigs along with the Forrest BACs and BIBACs as above. The purposes of this experiment were to integrate the contigs of the Williams 82 and Faribault BAC clones into the genome-wide BAC/BIBAC physical map of the soybean genome under construction and to anchor the BAC/BIBAC contigs that contained the Williams 82 and/or Faribault BACs to the soybean composite genetic map (Cregan et al., 1999; <http://soybase.agron.iastate.edu/>).

ACKNOWLEDGMENTS

The TAMU authors thank Randy Shoemaker, USDA/ARS and Iowa State University and Nevin D. Young, University of Minnesota for kindly providing the marker-anchored positive clone lists and the soybean Williams 82 and Faribault BAC libraries; and Chantel Scheuring for critically reading the manuscript. This study was supported in part by the National Science Foundation – Plant Genome Program (Award #: 9872635), the Illinois Soybean Operating Board (Award #: ISPOB-24-198-3) and Texas Agricultural Experiment Station (Acc. #: 8536-203104).

REFERENCES

- Arabidopsis Genome Initiative** (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796-815.
- Arumuganathan, K., and Earle, E.D.** (1991). Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.* **9**, 208-219.
- Chang, Y.L., Tao, Q., Scheuring, C., Meksem, K., and Zhang, H.-B.** (2001). An integrated map of *Arabidopsis thaliana* for functional analysis of its genome sequence. *Genetics* **159**, 1231-1242.
- Chen, M., Presting, G., Barbazuk, W.G., Goicoechea, J.L., Blackmon, B., Fang, G., Kim, H., Frisch, D., Yu, Y., Sun, S., et al.** (2002). An integrated physical and genetic map of the rice genome. *Plant Cell* **14**, 537-545.
- Clemente, T.E., LaVallee, B.J., Howe, A.R., Conner-Ward, D., Rozman, R.J., Hunter, P.E., Broyles, D.L., Kasten, D.S., and Hinchee, M.A.** (2000). Progeny analysis of glyphosate-selected transgenic soybeans derived from *Agrobacterium*-mediated transformation. *Crop Sci.* **40**, 797-803.
- Coulson, A., Sulston, J., Brenner, S., and Karn, J.** (1986). Toward a physical map of the genome of the nematode *C. elegans*. *Proc. Natl. Acad. Sci. USA* **83**, 7821-7825.
- Cregan, P.B., Jarvik, T., Bush, A.L., Shoemaker, R.C., Lark, K.G., Kahler, A.L., Kaya, N., VanToai, T.T., Lohnes, D.J., Chung, J., and Specht, J.E.** (1999). The integrated map of the soybean genome. *Crop Sci.* **39**, 1464-1490.
- Danesh, D., Penuela, S., Mudge, J., Denny, R.L., Nordstrom, H., Martinez, J.P., and Young, N.D.** (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.
- Ding, Y., Johnson, M.D., Colayco, R., Chen, Y.J., Melnyk, J., Schmitt, H., and Shizuya, H.** (1999). Contig assembly of bacterial artificial chromosome clones through multiplexed fluorescence-labeled fingerprinting. *Genomics* **56**, 237-246.
- Donaldson, P.H., and Simmonds, D.H.** (2000). Susceptibility to *Agrobacterium tumefaciens* and cotyledonary node transformation in short-season soybean. *Plant Cell. Rep.* **19**, 478-484.
- Frijters, A.C., Zhang, Z., van Damme, M., Wang, G. L., Ronald, P.C., and Michelmore, R.W.** (1997). Construction of a bacterial artificial chromosome library containing large EcoRI and Hind III genomic fragments of lettuce. *Theor. Appl. Genet.* **94**, 390-399.
- Goldberg, R.B.** (1978). DNA sequence organization in the soybean plant. *Biochem. Genet.* **16**, 45-68.
- Gregory, S. G., Howell, G.R., and Bentley, D.R.** (1997). Genome mapping by fluorescent fingerprinting. *Genome Res.* **7**, 1162-1168.
- Gregory, S.G., Sekhon, M., Schein, J., Zhao, S., Osoegawa, K., Scott, C.E., Evans, R.S., BurrIDGE, P.W., Cox, T.V., Fox, C.A., et al.** (2002). A physical map of the mouse genome. *Nature* **418**, 743-750.
- Gurley, W.B., Hepburn, A.G., and Key, J.L.** (1979). Sequence organization of the soybean genome. *Biochem. Biophys. Acta* **561**, 167-183.

- Hodgkin, J., Plasterk, R.H.A., and Waterston, R.H.** (1995). The nematode *Caenorhabditis elegans* and its genome. *Science* **270**, 410-414.
- Hoskins, R.A., Nelson, C.R., Berman, B.P., Laverly, T.R., George, R.A., Ciesiolka, L., Naeemuddin, M., Arenson, A.D., Durbin, J., David, R.G., et al.** (2000). A BAC-based physical map of the major autosomes of *Drosophila melanogaster*. *Science* **287**, 2271-2274.
- Iqbal, M.J., Meksem, K., Njiti, V.N., Kassem, A., and Lightfoot, D.A.** 2001. Microsatellite markers identify three additional quantitative trait loci for resistance to soybean sudden-death syndrome (SDS) in Essex × Forrest RILs. *Theor. Appl. Genet.* **102**, 187-192.
- International Human Genome Mapping Consortium** (2001). A physical map of the human genome. *Nature* **409**, 934-941.
- Jones, J.G., Shumukov, L., Carland, F., English, J., Scofield, S.R., Bishop, G.J., and Harrison, K.** (1992). Effective vectors for transformation, expression of heterologous genes, and assaying transposon excision in transgenic plants. *Transgenic Res.* **1**, 285-297.
- Keim, P., Schupp, J.M., Travis, S.E., Clayton, K., Zhu, T., Shi, L., Ferreira, A., and Webb, D.M.** (1997). A high density genetic map of soybean based upon AFLP markers. *Crop Sci.* **37**, 537-543.
- Kim, U.-J., Birren, B.W., Slepak, T., Mancino, V., Boysen, C., Kang, H.L., Simon, M.I., and Shizuya, H.** (1996). Construction and characterization of a human bacterial artificial chromosome library. *Genomics* **34**, 213-218.
- Lark, K.G., Weisemann, J.M., Matthews, B.F., Palmer, R., Chase, K., and Macalma, T.A.** (1993). Genetic map of soybean (*Glycine max* L.) using an intraspecific cross of two cultivars: 'Minsoy' and 'Noir 1'. *Theor. Appl. Genet.* **86**, 901-906.
- Marek, L.F., and Shoemaker, R.C.** (1997). BAC contig development by fingerprint analysis in soybean. *Genome* **40**, 420-427.
- Marek, L.F., Mudge, J., Darnielle, L., Grant, D., Hanson, N., Paz, M., Yan, H., Denny, R., Larson, K., Foster-Hartnett, D., et al.** (2001). Soybean genomic survey: BAC-end sequences near RFLP and SSR markers. *Genome* **44**, 572-581.
- Marra, M.A., Kucaba, T.A., Dietrich, N.L., Green, E.D., Brownstein, B., Wilson, R.K., McDonald, K.M., Hillier, L.W., McPherson, J.D., Waterston, R.H.** (1997). High-throughput fingerprint analysis of large-insert clones. *Genome Res.* **7**, 1072-1084.
- Marra, M., Kucaba, T., Sekhon, M., Hillier, L., Martienssen, R., Chinwalla, A., Chinwalla, A., Crockett, J., Fedele, J., Grover, H., et al.** (1999). A map for sequence analysis of the *Arabidopsis thaliana* genome. *Nature Genet.* **22**, 265-270.
- Meksem, K., Ruben, E., Zobrist, K., Zhang, H.-B., and Lightfoot, D.A.** (2000). Two large insert libraries for soybean: Applications in cyst nematode resistance and genome wide physical mapping. *Theor. Appl. Genet.* **101**, 747-755.
- Mozo, T., Dewar, K., Dunn, P., Ecker, J.R., Fischer, S., Kloska, S., Lehrach, H., Marra, M., Martienssen, R., Meier-Ewert, S., and Altmann, T.** (1999). A complete BAC-based physical map of the *Arabidopsis thaliana* genome, *Nature Genet.* **22**, 271-275.

- Salimath, S.S., and Bhattacharyya, M.K.** (1999). Generation of a soybean BAC library, and identification of DNA sequences tightly linked to the Rps-1k disease resistance gene. *Theor. Appl. Genet.* **98**, 712-720.
- Shoemaker, R., Keim, P., Vodkin, L., Retzel, E., Clifton, S.W., Waterston, R., Smoller, D., Coryell, V., Khanna, A., Erpelding, J., et al.** (2002). A compilation of soybean ESTs: generation and analysis. *Genome* **45**, 329-338.
- Shoemaker, R.C., and Specht, J.E.** (1995). Integration of the soybean molecular and classical genetic linkage groups. *Crop Sci.* **35**, 436-446.
- Shoemaker, R.C., Polzin, K.C., Labate, J., Specht, J.E., Brummer, E.C., Olson, T., Young, N., Concibido, V., Wilcox, J., Tanmulonis, J.P., Kochert, G., and Boerma, H.R.** (1996). Genome duplication in soybean (*Glycine* subgenus soja). *Genetics* **144**, 329-338.
- Singh, R.J., and Hymowitz, T.** (1998) The genomic relationship between *Glycine max* (L.) Merr. and *G. soja* Sieb. and Zucc. as revealed by pachytene chromosomal analysis. *Theor. Appl. Genet.* **76**, 705-711.
- Soderlund, C., Longden, I., and Mott, R.** (1997). FPC: a system for building contigs from restriction fingerprinted clones. *CABIOS* **13**, 523-535.
- Soderlund, C., Humphray, S., Dunham, A., and French, L.** (2000). Contigs built with fingerprints, markers, and FPC V4.7. *Genome Res.* **10**, 1772-1787.
- Sulston, J., Mallett, F., Staden, R., Durbin, R., Horsnell, T., and Coulson, A.** (1988). Software for genome mapping by fingerprinting techniques. *CABIOS* **4**, 125-132.
- Tao, Q., and Zhang, H.-B.** (1998). Cloning and stable maintenance of DNA fragments over 300 kb in *Escherichia coli* with conventional plasmid-based vectors. *Nucleic Acids Res.* **26**, 4901-4909.
- Tao, Q., Chang, Y.L., Wang, J., Chen, H., Islam-Faridi, M.N., Scheuring, C., Wang, B., Stelly, D.M., and Zhang, H.-B.** (2001). BAC-based physical map of the rice genome constructed by restriction fingerprint analysis. *Genetics* **158**, 1711-1724.
- Zhang H.-B., and Wu, C.** (2001). BAC as tools for genome sequencing. *Plant Physiol. Biochem.* **39**, 195-209.
- Zhang, H.-B., and Wing, R.A.** (1997). Physical mapping of the rice genome with BACs. *Plant Mol. Biol.* **35**, 115-127.
- Zhang, H.-B., Choi, S., Woo, S.-S., Li, Z., and Wing, R.A.** (1996). Construction and characterization of two rice bacterial artificial chromosome libraries from the parents of a permanent recombinant inbred mapping population. *Mol. Breed* **2**, 11-24.

WEB SITE REFERENCES

- <http://soybean.ccg.umn.edu/>, The Center for Computational Genomics and Bioinformatics, University of Minnesota.
- <http://hbz.tamu.edu>, the GENEfinder Genomic Resources, Texas A&M University.
- <http://soybase.agron.iastate.edu/>, Soybase.
- <http://www.siu.edu/~pbgc/DataBase/datap1.htm>, Plant Biotechnology and Genome Core Facility, Southern Illinois University at Carbondale.
- http://129.186.26.94/Legume_Initiative/LegGenomicsPaper10Oct01.html, Soybase.

Table 1 BACs and BIBACs fingerprinted and used for the soybean physical map.

Libraries	Cloning site	No. of clones fingerprinted	Average insert size (kbp)	No. of clones after edited	Average no. of bands	Redundancy Genome equivalent
Forrest						
BIBAC-H	<i>Hin</i> d III	30,720	125	27,221	27.63	3.052 x
BIBAC-B	<i>Bam</i> HI	21,504	125	20,181	23.43	2.262 x
BAC-E	<i>Eco</i> RI	30,720	157	28,620	39.12	4.030 x
Faribault						
BAC	<i>Eco</i> RI	1,142*	120	1,125	28.46	0.121 x
Williams 82						
BAC	<i>Hin</i> d III	860*	150	854	32.03	0.115 x
Combined libraries		84,946		78,001		9.580 x

*The BACs were identified with 267 SSR markers and 105 RFLP markers by the laboratories of R. Shoemaker, Iowa State University and N. Young, University of Minnesota (Marek et al. 2001) from the soybean cv. Williams 82 (Marek and Shoemaker 1997) and cv. Faribault (Danesh et al. 1998) BAC libraries. The distribution of the markers in the soybean genetic map (Cregan et al. 1999) were from Marek et al. (2001).

Table 2 Status of the soybean physical map before and after manual editing.

	Automatic	Manually edited
Date	Jan. 2002	Aug. 2002
Number of clones in FPC database	78,001	78,001
Number of singletons	4,933	4,954
Number of contigs	4,792	2,905
Contigs containing:		
> 200 clones	0	6
101 – 200 clones	0	88
51 – 100 clones	21	321
26 – 50 clones	1,176	504
10 – 25 clones	1,606	920
3 – 9 clones	1,777	850
2 clones	212	216
Unique bands of the contigs	364,570	346,884
Physical length of the contigs in Mb	1,480*	1,408*
DNA markers in the contigs	0	388

*Each fingerprint band was estimated to represent an average of 4.06 kb. It was estimated by the average insert size of all BAC and BIBAC clones divided by the average number of bands per clone used for the map contig assembly.

Table 3 BACs and BIBACs identified with RFLP and SSR markers and their locations in the soybean physical map.

Probe	MLG	Clone/contig	contig
pA186 ^a	F	E59J08/438, E71M02/542, E24A12/1500, E12C14/S	438, 542, 1500, S
pA504 ^a	F	E48A22/599, E63H03/599, E49N17/687, E60E23/687, E64H12/779, E75I15/779, E42F17/1766	599, 687, 779, 1766
pA708 ^a	F	E77E03/238, E22N05/368, E54G20/437, E56B13/1203, E63G21/1203, E76A16/1203, E41N07/1226, E44J07/1226, E78B05/1285, E64A17/S	S, 238, 368, 437, 1203, 1226, 1285
pA757 ^a	F	E18K20/373, E11E16/S, E19G06/S, E65K13/S, E80C11/S	373, S
pB212 ^a	F	E66G10/137, E66G10/845, E71L20/845, E57B10/S	845, 137, S
pK644 ^a	F	E41M09/559, E01H14/723, E34L17/723, E58O17/723, E58A20/845, E71L 20/845	723, 845, 559
pBLT065 ^a	A2	H32G20/79, H37B08/846, H60O19/1944, H61H01/1944	1994, 79, 846
Satt170 ^b	C2	H14P07/749, H67C07/749	749
Satt174 ^b	A1	H38B14/810, H50K08/810, H73B14/810	810
Satt183 ^b	J	H13K18/1753	1753
Satt217 ^b	G	H16M16/768, H19K01/768, H26P18/768, H33E24/768, H72H09/768	768
Satt221 ^b	D1a	H12F21/74, H36F17/74, H43L01/74, H66C02/74	74
Satt237	N	H20O17/333, H34D23/333, H77L12/333, H55P10/533	333, 533
Satt288	G	H41K10/97, H19K03/790, H41K09/790, H76K19/790	790, 97
Satt350 ^b	D1b	H45K18/31, H50J09/31	31
Satt408	D1a	H04L17/268, H35C16/268, H52N01/268, H05C13/1787	268, 1787
Satt439 ^b	D1a	H33D18/1009, H75D06/1009, H75M24/1009	1009
Satt471	A1	H45A20/346, H72C07/346, H72D06/1494	346, 1494
Satt511	A1	H53D05/113, H75A01/113, H26F13/310	113, 310
Satt570	G	H52N14/540, H52N13/2262	540, 2262
Satt596	J	H49A16/636, H53K07/636, H69A02/636, H74K24/753, H53K08/1412, H56D12/1858	636, 753, 1412, 1858
Satt610 ^b	G	H42E01/799, H67I11/799	799

S: Singleton; MLG: molecular linkage group; a: the DNA markers are multiple-copy in the soybean haploid genome as indicated by Southern analysis; and b: the DNA markers are single-copy in the soybean haploid genome as indicated by agarose gel analysis. The copy numbers of the remaining markers were not estimated.



Figure 1 Example of the BAC/BIBAC contigs of the soybean physical map. This contig (ctg16) was anchored to the molecular linkage group MLG F of the soybean genetic map by the SSR marker Satt343f (Creagan et al., 1999). The highlighted clones were the positive clones of the SSR marker. The contig contains 120 BAC and BIBAC clones and 651 unique bands, estimated to span about 2,643 kb in physical length. The clones prefixed with “IS” or “UM” were from the Williams 82 and Faribault BAC libraries, respectively; and the clones prefixed with “E”, “B” or “H” were from the Forrest *Eco* RI BAC, and *Bam* HI and *Hind* III BIBAC libraries, respectively.

MLG A1			MLG A2		
cM	Marker	Contigs	cM	Marker	Contigs
0	A487	2493 216 1409 437 502 1030 2205 2511	0	K636_1	2075 1435 778 1163 2075 2263 2377
5.7	Satt572	575 735 925	8.6	Satt390	223 865
18	Satt593	222	11.4	Sct_067	400 967 1278
19.1	Satt248	59 143 622 2031	19.3	Satt207	455 1791 276
19.1	Satt364	58 126 1733 1809	20	Satt480	455 302
19.1	Satt382	227 356 875 2167 2749	23.3	Satt493	480 1848
19.1	Satt454	157	23.3	Satt589	43
19.1	Satt471	346 292 1494	25.6	Satt177	1174 33 1792
19.1	Satt526	211	33.2	Satt315	316
27.1	Satt300	953 1586 2415	49.3	BLT065	1994 79 846
27.1	Satt591	105 68 65	50	Satt187	1938
28.5	Satt155	980	58.5	Satt424	265 1290 1297
48.2	Satt050	s	73.5	Satt341	1555
49.8	B030_2	577 642 953 247 899 674	78.3	Sat_115	1659
52	A096_1	502 1865 1952 259 483	89.9	Satt377	176
69.9	Satt385	770 44 2199	93.7	Satt525	488 200
73.1	A975_1	754 1222 2902	96.2	Satt233	209 1941 145 1791
76.7	A256_2	1346 1668 29 1390 1848 2011 2075 1163 1435	108.7	Satt327	631
84.3	Satt174	810 2002	108.7	Satt329	331
86.4	Satt211	357	108.7	Satt508	750
86.9	Satt200	370 171	119.6	Satt421	371 445 628 425 2561
87.3	Satt225	588 793	119.6	Satt470	151 112
87.3	Satt236	113 1318 1999 2156 2373	123	Sat_040	479 430 1211
87.3	Satt511	113 183 310	124.9	Satt333	425
90.1	A104_1	298 462 783	132.4	Satt133	74
?	Satt512	385	138.2	Satt455	2391 57 387 505 541 560 617
			173.5	Satt538	1332
			184	Satt429	479 331

Figure 2 – Wu et al. (1/7)

Figure 2 BAC/BIBAC contigs of the soybean physical map anchored by DNA markers to the existing soybean composite genetic map (<http://soybase.agron.iastate.edu/>; Cregan et al., 1999). The soybean genetic map consists of 20 molecular linkage groups (MLGs). The DNA markers used are listed in the middle column, positions (cM) of the markers in the linkage group in the left column and anchored contig(s) in the right column of each molecular linkage group. The letter “s” indicates singleton. Note that the contigs shown in rectangles were hybridized with two or more neighboring DNA markers, and one (A469) RFLP and 82 SSR markers each were hybridized to only one contig. These contigs were unambiguously anchored to the genetic map, whereas the exact positions of the remaining contigs remain to be further refined as additional markers are used to anchor the map contigs. The names of the DNA markers were after Cregan et al. (1999) and <http://soybase.agron.iastate.edu/>, the markers prefixed with “Satt”, “Sat” or “Sct” representing SSR markers and the remaining markers being RFLP markers. The markers suffixed with “_1, _2, or _3” indicate that they reside at two or more loci in the soybean genome.

MLG B1			MLG C2		
cM	Marker	Contigs	cM	Marker	Contigs
11.4	A588_1	1398 633 739 810 1820	25.4	Satt227	159 278 616 2561
20.7	A702_1	2522 756 996	29.2	Sat_062	468
24.7	A109_1	319 744 437 637 1587 2617	29.2	Satt432	316
32.3	A129_1	1163 289 800 772	45.9	Satt520	952 441
32.3	A632_1	1925 2328	48.2	Satt291	80 328 1200 1857
34.1	B031_1	1896 18 139 778	55.3	Satt457	1096
51.8	A089_2	2376 1524	55.5	A338_1	1096 847 457 743 813 1058 1581 1683
59.3	A006_1	1058 395 1760	77.9	Satt170	749 261 854 1591
113.8	A567_1	1645 160 259 1952	83.7	Satt305	749 74 261 485 854
?	A036_2	888	84.6	Satt322	108
MLG B2			112.8	Satt450	227 768
cM	Marker	Contigs	121.6	B160_1	2705 336 549
7.3	A043_1	1784 311 782 652 744	122.4	A635_1	414 2007
22.3	A242_1	343 27 1233	127.6	Satt376	49 2014 405 513 885
66.1	A329_1	216 550 2404 42 779 1844 2493	128.5	Sat_076	8 248 773
73.5	A509_1	802 838 78 1595	138.8	Satt277	723
104.9	B124_1	93 538 59 1347	139.7	Satt365	325 612 1895
110.9	A741_1	888 1397 1992	140.9	Satt557	498 110 424 1216
116.8	A516_1	289 2531	141.7	Satt134	73 1659
131.6	A230_1	547 677 2859	141.7	Satt289	313 110 267
139	A404_2	826	143.6	Satt100	868 328
MLG C1			145.8	Satt319	8 209 636
cM	Marker	Contigs	165.8	Satt460	843
33.1	A078_1	1836 674 772 953	168.3	Sct_028	7
34.9	A059_1	67 844 914 247 848	170.8	Satt433	565
36.2	K300_1	838 1124 2903 1837 29	171.9	Satt316	534 1827
105.2	A538_2	59 1868 669 8271973	193.5	Satt357	309
119	A538_3	262 1868 669 8271973	?	B208	521 5

Figure 2 – continued (2/7)

MLG D1a			MLG D1b		
cM	Marker	Contigs	cM	Marker	Contigs
0	A398_1	1637			
8.3	Satt184	s	14.9	A481_1	965 1865 1935 1918 2033
44.6	Satt482	2025	66.9	A586_3	709 656
49.1	Satt532	859	112.8	Satt506	318 10 496
49.1	Satt605	74 136	113.1	Satt350	31
49.8	Satt320	260			
49.8	Satt342	73 322 1859			
51.5	Satt221	74 496 197 272			
52.1	Satt502	74 119 268 614 1074 1247 2517			
53.5	Satt169	330 339			
54.7	Satt321	330 517			
54.7	Satt548	267 153 230			
63	A077_1	2522 463 886 2496			
67	Satt402	585 632 1313			
67.5	Satt603	166 1446			
67.8	Satt254	467 87 2290			
68.6	Satt267	909 1096			
69.8	Satt295	1019 442 1488			
69.8	Satt515	646 759			
73.1	Satt203	652 965			
74	Satt580	112 10 94 308			
75.9	Satt370	832			
80.3	Satt507	575			
83.5	Sat_106	362 2118			
84.2	Satt198	1740			
87.9	Satt077	1009 157 411			
89.3	Satt439	1009			
89.3	Satt436	43 118 2185			
94.1	Sat_036	508 1332			
173	Satt071	1135 2069			
174.2	Satt407	658 2134			
176.6	Satt408	268 226 1787			
178.2	Satt129	1358 48 581 714			
178.5	Satt147	s			

Unlinked	
Marker	Contigs
A469	258
Sat_144	1531 184 197
Sat_145	77 23 73
Sat_157	344
Sat_162	846 294 406 434
Sat_163	2236
Sat_167	201
Sat_185	462 255 768 1405
Sat_199	1205
Sat_200	520 2302
Sat_201	114 190 2006 2332
Sat_205	521 5
Sat_206	521 945 2309
Satt621	669 238
Satt632	538 846
Satt647	205 557
Satt648	79 58 223 488
Satt653	977 147 849
Satt654	861 259 808
Satt675	632 480 1313 1332 2680
Satt679	78 112 365 2859

Figure 2 – continued (3/7)

MLG D2			MLG F		
cM	Marker	Contigs	cM	Marker	Contigs
2.9	A095_1	960 1242 457	0	Satt146	337 440 1035 2384
15.1	A257_1	1619 67 128 252 798 340 762 999	0	Satt649	337 402 696 1630
21.2	A401_2	409	1.7	Satt343	16 2680
32.5	A124_1	67 914 1781 2079 348 770 914 1168 2651	1.8	Satt193	15 188
40.7	Satt498	630	12.4	Satt269	2410
106.6	Satt311	25 462 2236	19.3	Satt149	45 1931
154.8	Sct_137	1121	44.5	A401_1	2803 872 2496
?	A708_2	1797	50.7	Satt516	127
			56.4	Satt425	820 1337
			57.5	Satt374	820
			57.6	K644_5	559
			58.6	Sat_133	97 2645
			74.2	K644_3	723
			77	A504_2	599 687 779 1766
			77.5	A186_1	438 542 732 1500
			77.5	A757_1	373
			80.6	Satt114	337 405 1513
			117.1	B212_1	845 1380 137 20 219 400 484 633 844 1820
			117.1	K644_1	845
			123.8	A708_1	437 1203 1226 238 319 368 1285
			126.2	Satt335	1689 420
			128.3	Satt362	1642 664 1408
			133.1	Sct_188	510
			155.8	B148_1	650 311
			160	Satt657	1296
			166	A566_1	170
			178.2	Satt395	537 2505
			181.8	Sat_074	2279
			?	A059_3	643
MLG E					
cM	Marker	Contigs			
6	Satt575	841 1790			
16.3	Satt384	317 959			
31.6	A069_2	66			
33.7	A517_1	1147 1464			
68.4	A086_1	699 114 231 1313 2559			
72.5	A517_2	1147 1464			
73.8	A458_1	869 1408 521			
84.9	Satt573	987 271			
84.9	Satt598	987			
91.8	Satt491	560 1028			
93.3	K477_1	1299 353 1705 2757			
99.9	Satt452	155 1089 1564			
99.9	Satt483	155			
100.4	Satt204	257 388 1639			
100.4	Satt263	114 323 904 172 1870 1932 2384			
100.4	Satt268	651			
134.7	Satt231	471			
134.7	Satt553	470			
136.5	A111_2	1219			
138	A711_1	75 319 1337 2114 2517			
?	A538_1	79			
?	A963	1398 327 457 463 798 1337 1752 2776			

Figure 2 – continued (4/7)

MLG G			MLG H		
cM	Marker	Contigs	cM	Marker	Contigs
-2.5	Satt163	69 177 315 594	10	Satt666	520
1.9	Satt309	1972	15	Satt635	6 503 1290 1374
5.1	Satt610	799	27.6	Satt568	s
7.7	Satt570	462 2262 25 81 331 540 2561	31.9	A089_1	85 1569 2173
13.2	Satt217	768 408 545 841 1281 2331	32.6	A036_1	s
13.6	Satt130	841	41.1	Satt192	s
14.2	Satt235	150 612 2328	45.2	Scft009	1310 318
17	K069_1	1887 1285 2283	56.3	A703_1	1517 246 549 770
25.9	Satt324	95	61.4	A404_1	137 20 104 552 56 400
36	A112_1	1231 1299	68.5	Satt469	56 495 842
61.3	Satt594	300 119 215 735	68.5	Satt541	56 842
63.3	Satt427	272	70.2	Sat_122	107 411 564 1427
66.1	Satt533	219 299 419 612 641	71.1	Sat_118	1103 1492
66.1	Satt564	1041 230 590	77.3	Satt279	213 211 1007
67	Sat_094	318 645	77.3	Satt314	98
67	Satt504	10 101	78.1	Satt222	94 2194
71.6	Satt303	1169 1408 2727	78.1	Satt253	42 621 1839
72.4	Satt352	74 654	105.8	A810_1	1661 885 1421 154 2377 263
72.4	Satt566	74 654	110.4	Satt302	149 154 936 1661 2377
74.2	Satt131	48 91 1679	116.9	Satt142	494 1877
77.4	Satt340	200 714	116.9	Satt293	494 571
78.5	Satt501	959 84 257	125.3	Satt181	586 72 691 1554 2198
94.6	A073_1	462 783 2190	128.9	Satt317	1461 1972
96.4	Satt138	1601	145.3	A570_1	1208 464 951 1313 1398 1531 2560
100.1	Satt505	1502			
100.8	Satt400	621 803	MLG I		
103.2	Satt503	284	cM	Marker	Contigs
103.2	Satt517	402 620 2906	43.9	A515_1	298 637 1587 646 2788
112.9	Satt288	790 97	54	A955_1	710 2077 550 1935 52 150 1344 2582
120.7	A885_1	129 1313 464 538 1231	93.3	A007_1	2033 170 1286
121.7	K493_1	320 263 885	98.4	B039_1	564 652 744 1587
135	Satt472	217	108.2	A644_1	911 799 2458
138.4	Sat_117	965			
160	A378_1	2114 798 391 721 796 1424			
160.7	A586_2	646			
166	A681_1	252 1332			

Figure 2 – continued (5/7)

MLG J			MLG K		
cM	Marker	Contigs	cM	Marker	Contigs
9.8	Satt405	410 79	4.3	Satt539	21 330 1945
10.5	Satt249	s	17.9	Satt242	1178
13.1	Satt287	1408 714 720	20.3	Sat_119	880 1178
19.5	Sct_046	390 248 1332	35.8	A315_1	463 1705
25.4	A060_1	307 343 1677	64.6	Satt349	461 54 145 1813 2561
28.2	B046_2	67 1168 2011	69.7	Satt381	950
36.3	A204_1	960 2205 2511 1183	70.9	Satt417	411 341 396
37.9	B074_1	104 56 927 2283	70.9	Satt552	426 502 2453
50.5	K384_1	1205	71.9	Satt046	428 690
61.3	Sct_065	294 472 18 48 130 389 2017	72.8	Satt375	411
63.6	Satt414	238	72.8	Satt544	82 324 1297 1408 1940
63.6	Satt596	636 88 145 405 753 1412 1858	73.6	Satt518	10 115 239 632 1567
63.6	Scaa003	122 265	82.2	Satt326	1030 563
67.2	Satt280	453 254 2164	85.8	Satt240	1260 397 1018 2561
67.2	Satt456	453 292 2563	86.9	Sat_044	1604 367
68	Satt406	1295 5	120	Satt273	62 384 660 1172 1673
70.8	Satt380	107 258	122.2	Sat_111	237 461 1642 2289
72.6	Satt183	1753 136 231 501 844	126.1	A199_1	697
74.4	Satt529	179 1734	144.3	Satt475	690 411
75	Satt132	736	145.1	Satt260	10 163 690 1753 2399
75.8	Sct_001	714 227 604 705 1610	164.8	Satt196	1247
105.5	Satt244	1614 238 596	169.4	K266_1	547 677 847 2442
108	Satt547	164 194 2400	175.6	Sat_126	1018
118	Satt431	1721 1841	184.6	Satt588	1055 501
124.8	A199_2	709 1424 1569			

Figure 2 – continued (6/7)

MLG L			MLG O		
cM	Marker	Contigs	cM	Marker	Contigs
33.3	A264_1	77 610 349 1569 2269	0	Sat_132	807 519
34	A106_1	66 914	2.4	Satt358	s
34	A450_2	2053	12.6	Satt492	974
45.1	A023_1	1677 709 1347 2360 960 1242	60.5	Satt420	1273
56.7	B046_1	1210	65.1	Satt576	308 162 218 479 587 925
57.6	A071_1	28 246 1237 786 1517	65.8	Satt633	308 77 119 123 202 496 748 945 1569
110.5	A489_1	491 1914 1285	66.7	Satt094	896
128.5	K385_1	778 1993 2134	67.5	Satt466	506 291 1606
138.3	A802_2	1163 201	67.5	Satt550	730 1300 2327 224
			68	Satt608	614
MLG M			68.4	Satt188	72 2430
cM	Marker	Contigs	68.4	Satt479	103 479 506 1606
100.5	A226_1	462 899 128 521 999	68.4	Satt585	672 135 359
100.5	K024_1	247 899	69.2	Satt262	146 404 2338
101.3	A715_1	400 914 1859	69.2	Satt473	81 146 280 1492
151.8	K227_1	526 770 555	78	Satt173	363 555 1247 1887
154.5	A064_1	1030 321 999 437	78	Satt345	1 119 274 363 670 883 1597
164.9	A504_1	779 1619 28 216 959 2000	81.7	Satt478	1088 148 1149
			103.8	Satt477	361 454 598 615 1901
MLG N			104.6	A878_1	770 170 324 526 1466 1594
cM	Marker	Contigs	125	Satt581	237
17.2	Satt009	621 1292 114 521	132	B157_1	450 672
35.4	Satt584	360 1749 2436	153.3	Satt153	1314
36.3	Satt393	1278 1749 2436	155.1	Satt243	1691
36.3	Satt485	166			
43.7	Satt125	1164			
54.8	Sat_033	245			
90.1	Satt237	333 533			
95.5	Sat_091	934			
118.5	Satt410	122 517 769 829			
128.2	K494_1	1359 1300 1797 181 353 556 709			
137	A802_1	336 2045			

Figure 2 – continued (7/7)