Genetic and Molecular Analysis of Soybean Genome

Harada K.[1], A. Kaga[1], Y. Katayose[1], Y. Tsubokura[1], S. Sato[2], S. Watanabe[1], Z. Xia[1], M. Hayashi[1], H. Kanamori[3], T. Shimizu[3], K. Machita[3], H. Ikawa[3], T. Ito[3], K. Kurita[3], K. Ito[3], J. Wu[1], T, Matsumoto[1], S. Tabata[2], T. Sasaki[1]

[1]*National Institute of Agrobiological Sciences* (*NIAS*) Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan.

[2]*Kazusa DNA Research Institute*, Kazusa Kamatari 2-6-7, Kisarazu, Chiba 292-0818, Japan.

[3]*Institute of Society for Techno-Innovation of Agriculture, Forestry and Fisheries*, Kamiyokoba Ippaizuka 446-1,Tsukuba,Ibaraki 305-0854, Japan.

Corresponding author: Kyuya Harada. Kannondai 2-1-2, Tsukuba 305-8602, Ibaraki, Japan. Tel: 81-298-38-7452, Fax: 81-298-38-7452,

E-mail: haradaq@nifty.com

Abstract

To obtain precise marker information for soybean genetic maps, we are currently constructing a high-density linkage map using a single F_2 mapping population from the cross between Enrei and Peking. Enrei which is an elite cultivar with a high quality for food processing is being sequenced as a representative cultivar in Japan, while Peking is resistant to many diseases and displays an outstanding stress resistance under wet conditions. New 948 SSR markers were designed to large scaffolds (>100kb) of U.S. cultivar Williams 82, Glyma 0, and were incorporated into the map in order to obtain a precise assembly of the genome sequences in the Japanese soybean genetic background. Currently, 1,860 markers have been located on the map. The marker location on the linkage map was compared with that of the chromosome-scale assembly of the soybean genome, Glyma 1.0 (http://www.phytozome.org/soybean.php), updated in December 2008. In total, 1,837 markers (98%) were anchored on the genome assembly. Generally, the marker order on the linkage map revealed a close agreement with that of physical data. Currently, the causes of the large differences are being determined; reverse orientation of marker order on the Gm05, Gm11, Gm13, Gm14, Gm19 and other minor differences between Williams 82 and Enrei × Peking. This map will provide information to identify regions in the sequence assembly where additional information is required to resolve the marker order.

In order to develop new breeding strategies for soybean production, we launched a comprehensive analysis of the soybean genome using the Japanese cultivar Enrei. BAC libraries were constructed and the end-sequences of approximately 100,000 BAC clones were analyzed and used for the construction of a physical map of the genome. To increase the saturation of the map, BLAST analysis between the BAC-end sequences and the genome assembly of another soybean cultivar, Williams 82 (Glyma1.0 assembly) was also carried out. This physical map will be used for characterizing the genome structure of Japanese soybean cultivars in order to accelerate the isolation of agronomically important genes and to develop new breeding strategies for increasing the production of local cultivars. We are also constructing a soybean genome database to facilitate comparative genome analysis.

To elucidate the characteristics of the soybean genome in reference to the genome of *L. japonicus*, synteny between the genomes was investigated. Macrosynteny between soybean and *L. japonicus* was analyzed by mapping the same cDNA clones on the maps of both species by the RFLP method, and by identifying the positions of orthologs on the *L. japonicus* map for cDNA markers located on the soybean map. Relatively large synteny blocks were observed between a few linkage groups of *L. japonicus* and soybean. The major parts of the soybean linkage groups consisted of mosaics of smaller segments syntenic with the *L. japonicus* genome. The presence of many homoeologous regions on different soybean linkage groups was suggested from the distribution of paralogs and orthologs. To investigate the microsynteny between soybean and *L. japonicus*, three soybean BAC clones were selected for the *GmNFR1a*, *GmNFR1b* and *Nts1* genes mapped on the macrosyntenic regions of the linkage groups D1b, B2 and H, respectively. We revealed a significantly high level of collinearity between these BAC clones and corresponding homologous genomic regions of *L. japonicus*.

INTRODUCTION

The Fabaceae family is the third largest family of angiosperm plants including around 20,000 species of legumes. The subfamily Papilionoideae includes agriculturally important species such as soybean, pea, common bean, mungbean, and cowpea, and model legumes, *Lotus japonicus* and *Medicago truncatula*.

Soybean is currently the most important legume in the world, because its seeds contain large amounts of edible protein and oil, and it is widely used for food, livestock feed, industrial materials and biofuel.

The genome size of soybean (*n*=20) is estimated at 1.12Gb (Arumunganathan and Earle 1991), a value about 2.0 times higher than that of *Vigna* species. It was suggested that the soybean genome is the product of a diploid ancestor (*n*=11), which underwent aneuploid loss (*n*=10), and subsequent polyploidization (Lackey 1980). The occurrence of

two rounds of genome duplication or hybridization and rearrangements was estimated by many researchers (Shoemaker *et al*. 1996, Shoemaker *et al*. 2002, Blanc and Wolfe 2004, Schlueter *et al*. 2004).

A linkage map of soybean was constructed mainly based on restriction fragment length polymorphism (RFLP) using random genomic clones as probes (Keim *et al*. 1990, Lark *et al*. 1993, Shoemaker and Specht 1995, Mansur *et al*. 1996). However, the relatively low degree of polymorphism in addition to the large genome size and complex genome structure has slowed the construction of a linkage map covering a large region of the soybean genome. Simple sequence repeat (SSR) markers were integrated into RFLP linkage maps (Akkaya *et al*. 1995). Cregan *et al*. (1999) integrated soybean linkage maps into 20 homologous linkage groups using SSR markers based on their correspondence with some known linkage maps. Five maps were integrated with the map of Cregan *et al*. (1999) using JoinMap, and a composite map with 1015 SSR, 709 RFLP, 73 RAPD, 6 AFLP, 10 isozyme markers, 24 classical markers and 12 other markers has been developed (Song *et al* 2004). Choi *et al*. (2007) located 1,141 genes onto one or more of the three maps based on single nucleotide polymorphism (SNP). A high-density integrated map of 1,810 SSR or sequence-tagged site (STS) markers using three recombinant inbred populations was developed (Hwang *et al*. 2009). A genetic map composed of 509 RFLP, 318 SSR, 318 AFLP, 97 AFLP-derived STS, 29 BAC-end or EST-derived STS, 1RAPD and five morphological markers using a single F2 population was constructed (Xia *et al*. 2008).

Genome-wide integrated physical and genetic maps will provide powerful tools for research into various aspects of the soybean genome, including large-scale genome sequencing, target marker development, high-throughput EST mapping and efficient positional cloning. Integrated genetic and physical mapping of Forrest (http://soybeangenome.siu.edu/) and Williams 82 (http://soybase.agron.iastate.edu/) was conducted using restriction enzyme fragment fingerprints of BAC clones.

A high level of synteny on entire linkage groups was observed between mungbean (*Vigna radiata*) and cowpea (*Vigna unguiculata*) (Menanciohautea *et al*. 1993). Though comparable levels of synteny were identified between mungbean and common bean (*Phaseolus vulgaris*), synteny blocks were more limited between these species and soybean (Boutin *et al*. 1995). Lee *et al*. (2001) observed a high level of conservation between chromosomes of mungbean and common bean, and the chromosome segments of soybean. Choi *et al*. (2004) reported genome-wide macrosynteny among legumes (*M. truncatula*, *M. sativa*, *L. japonicus*, *Pisum sativum*, *Cicer arietinum*, mungbean, common bean and soybean), using a large set of cross-species gene-specific markers. Though the length of the synteny blocks was reduced by chromosome rearrangements in some regions, chromosomes from a variety of Papilionoid species could be aligned, based on the chromosomes of *M. truncatula*. However, only small synteny blocks were observed between *M. truncatula* and soybean.

The order and orientation of at least six genes were found to be conserved in a 70kb region, including apyrase genes from soybean and *M. truncatula* (Cannon *et al*. 2003). Choi *et al*. (2004) examined two BAC clones from the region containing the cyst nematode resistance gene, *rhg1*, of soybean and homologous BAC clones of *M. truncatula*, and revealed that the order and orientation of fourteen genes were conserved between these genomes. They also analyzed ten homologous pairs of BAC clones from *M. truncatula* and TAC clones from *L. japonicus* and found that 72 genes (82%) were syntenic between the genomes. Mudge *et al*. (2005) uncovered two large soybean regions surrounding the cyst nematode resistance genes, *rhg*1 and *Rhg*4, that exhibited a synteny with *M. truncatula*.

Microsynteny analysis between homoeologous regions surrounding the N-hydroxycinnamoyl/benzoyltransferase (HCBT) genes (Schlueter *et al*. 2006), fatty acid desaturase 2 (FAD2) genes (Schlueter *et al*. 2007a), and LysM kinase genes (Zhang *et al*. 2007) has been performed.

Construction of a soybean linkage map based on a single F$_2$ mapping population derived from Enrei ×Peking

The marker order on the integrated maps from several different population is sometimes obscure. To obtain precise marker information for soybean genetic maps, we are currently constructing a high density linkage map using a single F$_2$ mapping population from the cross between Enrei and Peking. Enrei is an elite cultivar with a high quality for food processing. On the other hand, Peking displays many useful characteristics such as cyst nematode resistance, soybean mosaic virus resistance, Phytophthora rot resistance and a high stress resistance in germination under wet conditions.

Initially, genomic SSR markers developed by Cregan *et al*. (http://soybase.org/resources/ssr.php),

including markers that have been used for MAS and mapping by breeders and researchers were anchored on the linkage map to cover the genome widely. Subsequently, newly developed EST-SSR markers by Hisano *et al*. (http://www.kazusa.or.jp/soymarker/) were integrated into the map. The genome sequence of a U.S. cultivar Williams 82, Glyma 0, has been available since January 2008 and chromosome-scale assembly of the soybean genome, Glyma 1.0, updated in December 2008 (Soybean Genome Project, DoE Joint Genome Institute, http://www.phytozome.net/soybean.php). New SSR markers were designed to large scaffolds (>100kb) of Glyma 0 or to Glyma 1.0 and have been incorporated into the map in order to obtain a precise assembly of the genome sequences in the Japanese soybean genetic background. Additionally, SNP markers from 5,551 STS containing SNPs developed by Choi *et al*. (http://bfgl.anri.barc.usda.gov/soybean/) are being integrated. Currently, 1,860 markers (282 USDA SSR markers, 383 EST-SSR markers, 247 SNP markers and 948 newly designed SSR markers) have been located on the map (Fig.1, left white bar). This map spans a total length of 2913 cM with an average marker distance of 1.6 cM and maximum distance of 6.3 cM. Severe segregation distortion (P<0.001) was observed on the Gm06 and Gm11 (Fig.1, blue line). The map length is longer (21%) than that of the soybean consensus map 3.0 developed by Cregan *et al*. (http://soybase.org/LG2Xsome.php).

The marker location on the linkage map (Fig.1, left white bar) was compared with that of the chromosome-scale assembly of the soybean genome, Glyma 1.0 (Fig.1, black right bar). In total, 1,837 markers (98%) were anchored on the genome assembly. The other 23 markers were localized either on the unexpected chromosome due to duplicated loci (8), or unassigned scaffolds (9) or no hit (6). Generally, the marker order on the linkage map revealed a close agreement with that of physical data. Although the relationship between the genetic distance and physical distance appears to be ~360 kb/cM, assuming a genome size of approximately 1.1 Gb, positional biases (50kb-7M/cM) were found. In each chromosome, a genomic region with a highly suppressed recombination between markers, possibly the pericentromeric region, was observed and the extent differed among chromosomes (Fig.1). The ratio of genetic and physical distance
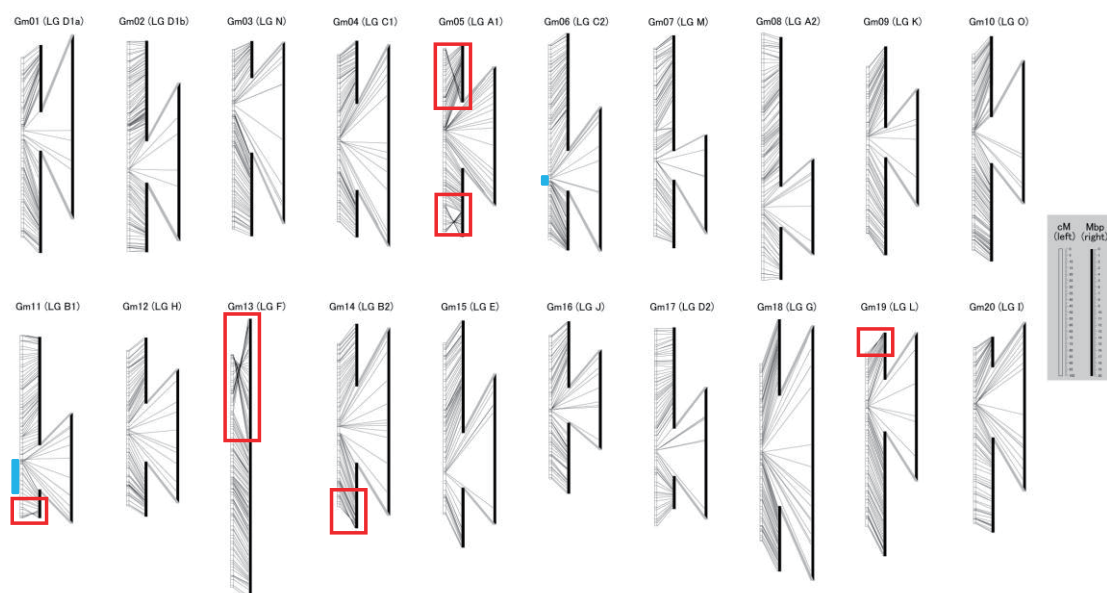

Fig.1. A genetic linkage map constructed based on an F_2 population derived from the cross between Enrei and Peking (left). Physical map of Williams 82, Glyma 1.0, was aligned (right) and marker locations are connected by black lines. In each chromosome, the physical map revealed a region of highly suppressed recombination between markers on the linkage map. These regions which are shown on the same scale at the right side may correspond to the pericentromeric regions. Chromosomal regions in which the order of DNA markers is different from sequence in the physical map are highlighted in red boxes.
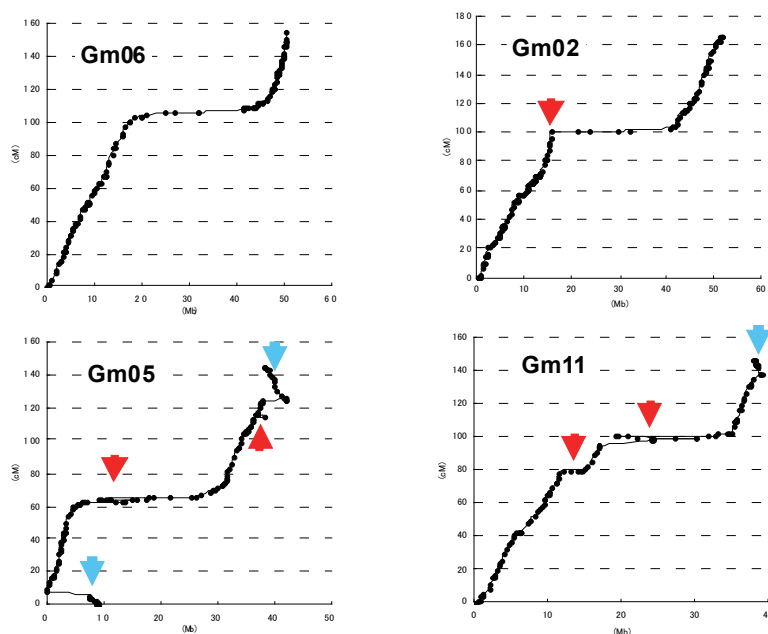
Fig.2. Variation in genetic/physical distance ratio. X and Y axes show the physical distance and genetic distance, respectively. Small dots indicate marker locations. Blue arrows designate reverse orientation of marker order on the Gm05 and Gm11 and red arrows indicate the regions where the orders of marker positions are being determined.

varied considerably depending on the chromosomal region (Fig.2). Currently the causes of the large differences are being determined; reverse orientation of marker order on the Gm05, Gm11, Gm13, Gm14, Gm19 (Fig.1, red) and the other minor differences between Williams 82 and Enrei × Peking. This map will provide information to identify regions in the sequence assembly where additional information is required to resolve the marker order.

Construction of soybean BAC-based physical map and database

To develop the basic resources for analyses of the whole genome, we have constructed two bacterial artificial chromosome (BAC) libraries from the Japanese soybean cultivar Enrei using the restriction enzymes *Hin*dIII and *Mbo*I. We have sequenced Enrei BAC end sequences since 2007 to develop a BAC-based physical map and database by using the Williams 82 genome assembly. We sequenced about 80,000 clones from a *Hin*dIII-library and about 20,000 clones from a *Mbo*I-library. To construct a BAC-based physical map of the Enrei genome, BLASTn analysis between the BAC-end sequences of Enrei and genome assembly of Williams 82 (Glyma1 assembly, http://www.phytozome.org/soybean.php) was carried out. This physical map will be used for characterizing the genome structure of Japanese

soybean cultivars in order to accelerate the isolation of agronomically important genes and to develop new breeding strategies for increasing the yield of Japanese cultivars. We are also constructing a soybean genome database, "DaizuBase", to facilitate comparative genome analysis. Unified map of DaizuBase indicates the relationship between the linkage map and physical maps (Fig. 3). GBrowse shows aligned Enrei BAC clones on Glyma1 assembly (Fig. 4). Although presently, only participants in the DD-project (Genomics for Agricultural Innovation of MAFF) can utilize DaizuBase, consultation will be opened by the end of 2009. We introduced a next generation sequencer GS-FLX (Roche/454) to decode the Enrei genome for the characterization of the genome structure of Japanese (domestic) soybean. The high sequencing ability will enable us to conduct a comparative analysis of the soybean genome structure.

Characterization of soybean genome by synteny analysis with *Lotus japonicus*

The rapid accumulation of genome sequence information for *L. japonicus* and *M. truncatula* provides a unique opportunity for comparative studies between the model legume and leguminous crops.

Macrosynteny between soybean and *L. japonicus* was analyzed by mapping the same cDNA clones on the maps of both species by the RFLP method, and by iden-
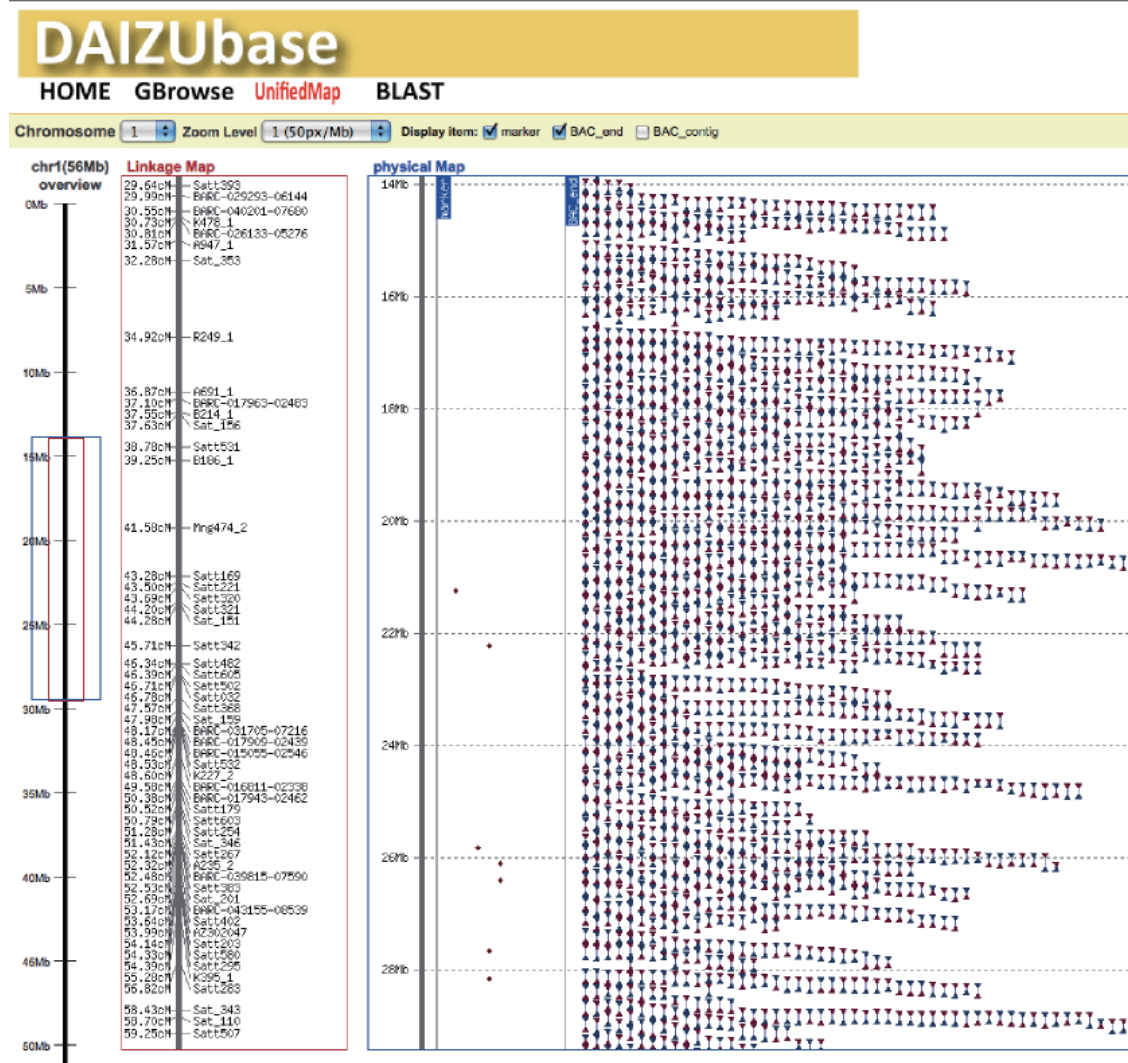
Fig.3. Screen shot of UnifiedMap on DaizuBase. This shot offers information about DNA markers, BAC clones and the relationship between the linkage map and physical maps.

tifying the positions of orthologs on the *L. japonicus* map for the cDNA markers located on the soybean map.

Relatively large synteny blocks were observed between *L. japonicus* LGs (*Lj*LGs) 1, 3 and 4 and soybean LGs (*Gm*LGs) L, H and A1, respectively (Fig. 5). However, the major parts of *Gm*LGs consisted of mosaics of smaller segments syntenic with *L. japonicus*. On the other hand, the L-G synteny map indicated that each of the *Lj*LGs harbored syntenic regions with two or more *Gm*LGs (Fig. 5). In particular, each region of *Lj*LGs often overlapped with syntenic segments of different *Gm*LGs, suggesting that such overlapped segments of *Gm*LGs reflected the homoeologous duplicated regions. Such possible duplications were observed between A1 and A2, A1 and D2, A1 and K, A2 and M, B1 and B2, B1 and C2, B1 and D1b, B2 and G, C1 and C2, C2 and B1, D1a and D1b, E and K, F and M, G and L, G and

M, I and L, I and O, L and O, and N and O (Fig. 5). In fact, the same DNA markers or paralogs were located at two or three positions mostly on different chromosomes. They are indicated in Fig. 5 and include the putative Nod-factor receptor gene and its paralog, *GmNFR1a* (D1b) and *GmNFR1b* (B2), and the autoregulation gene *Nts1* (H) and its paralog *GmCLV1A* (B1).

To investigate the microsynteny between soybean and *L. japonicus*, three soybean BAC clones were selected for the *GmNFR1a*, *GmNFR1b* and *Nts1* genes, mapped on the macrosyntenic regions of *Gm*LGs D1a, B2 and H, respectively (Fig. 5). Nucleotide sequences of the BAC clones WBb76L15 (*GmNFR1a*), WBb98N11 (*GmNFR1b*) and WBb225L1 (*Nts1*) were determined by the shotgun method with a 5-time redundancy, and assembled sequences of 124,649 bp, 132,374 bp and 129,251 bp were obtained, respectively. The nucleotide
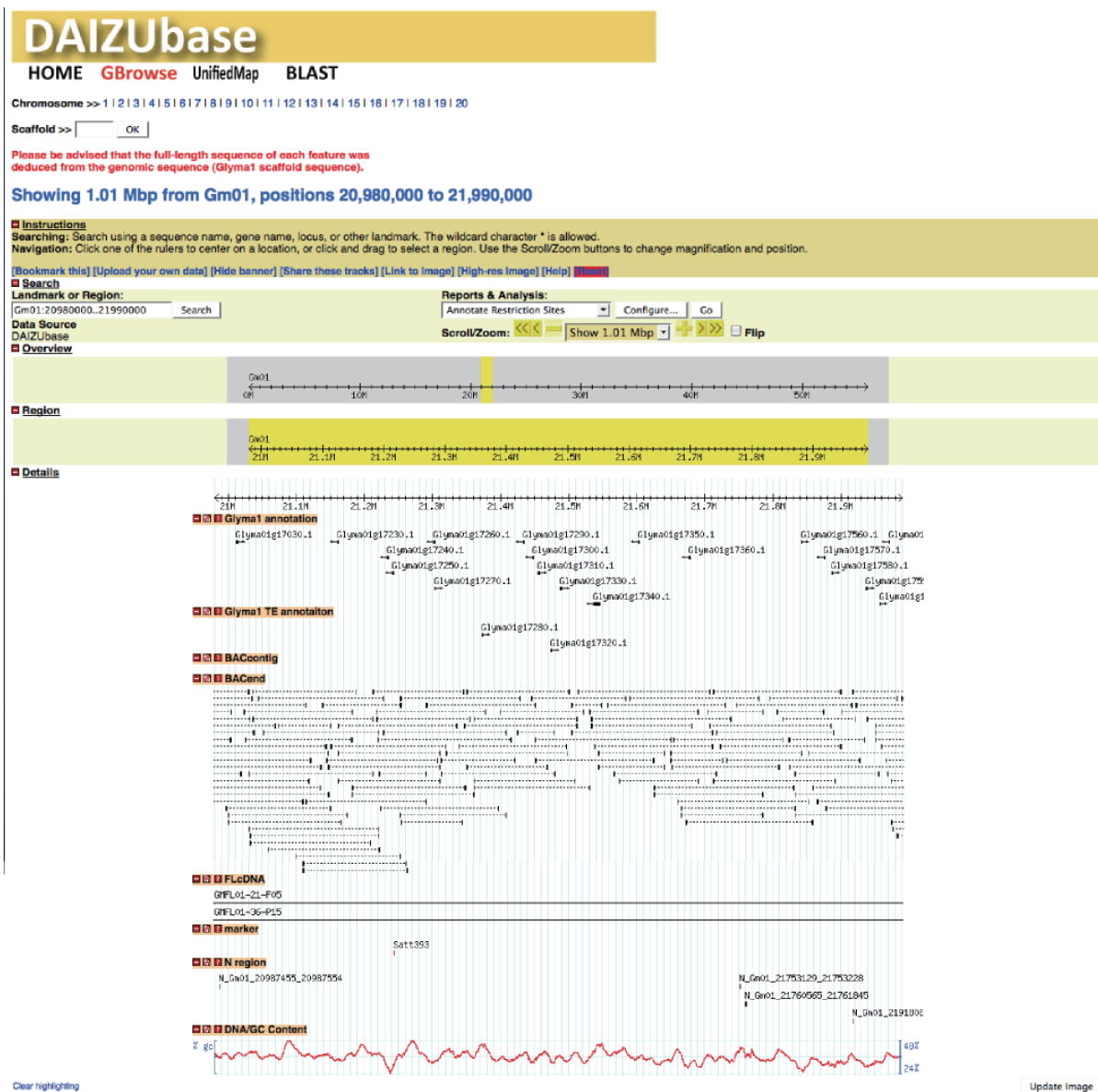
Fig.4. Screen shot of GBrowse of DaizuBase. BAC contigs are shown on each chromosome of the Williams 82 genome assembly.

sequences obtained were subjected to manual assignment of protein-encoding genes, and as a result, a total of 44 complete gene structures and 9 partial and pseudo-gene structures could be identified. The average gene density of each clone ranged from one gene every 8.3 kb to 9.6kb. This density was lower than that in previous estimates (Young *et al*. 2003, Triwitayakorn *et al*. 2005, Mudge *et al*. 2005) but slightly higher than that in a recent report surveyed by 17 BAC sequences (Schlueter *et al*. 2007b).

Sequence comparison of WBb225L1 (*Nts1*) and corresponding *HAR1* genomic region of *L. japonicus* revealed a highly conserved microsynteny (Fig. 6A). Among 14 predicted genes, including partially predicted and pseudo-genes in the *HAR1*

region of the *L. japonicus* genome, 11 (78%) genes had homologs on WBb225L1 (Fig. 6A). On the other hand, 69% (11/16) of the genes of WBb225L1 predicted in the overlapped region shared homologs in the *HAR1* region of the *L. japonicus* genome. Similarly, a highly conserved microsynteny was identified based on a comparison between the *NFR1* genomic region of *L. japonicus* and homoeologous regions of soybean *NFR1*, WBb76L15 (*GmNFR1a*) and WBb98N11 (*GmNFR1b*) (Fig. 6B). Among the 19 predicted genes, including partially predicted and pseudo-genes in the *NFR1* region of the *L. japonicus* genome, 14 (74%) genes showed homologs in the *GmNFR1* genomic regions (Fig. 6B). On the other hand, 73% (11/15) and 79% (15/19) of the predicted

genes on WBb76L15 and WBb98N11, respectively, displayed homologs in the *NFR1* region of the *L. japonicus* genome. The level of gene conservation between the homoeologous regions of *GmNFR1a/GmNFR1b/NFR1* was even higher compared with that in the *Nts1/HAR1* region. In 89% (25 out of 28) of the predicted genes in the overlapped regions, the order and orientation were conserved. In most cases, genes with complete structures were identified in both regions (ie, homoeologs). In such cases, most of the homoeologs showed a 90% or higher nucleotide sequence identity in exons and even introns.

We are currently constructing a synteny map based on the sequences of soybean and *Lotus japoni-*
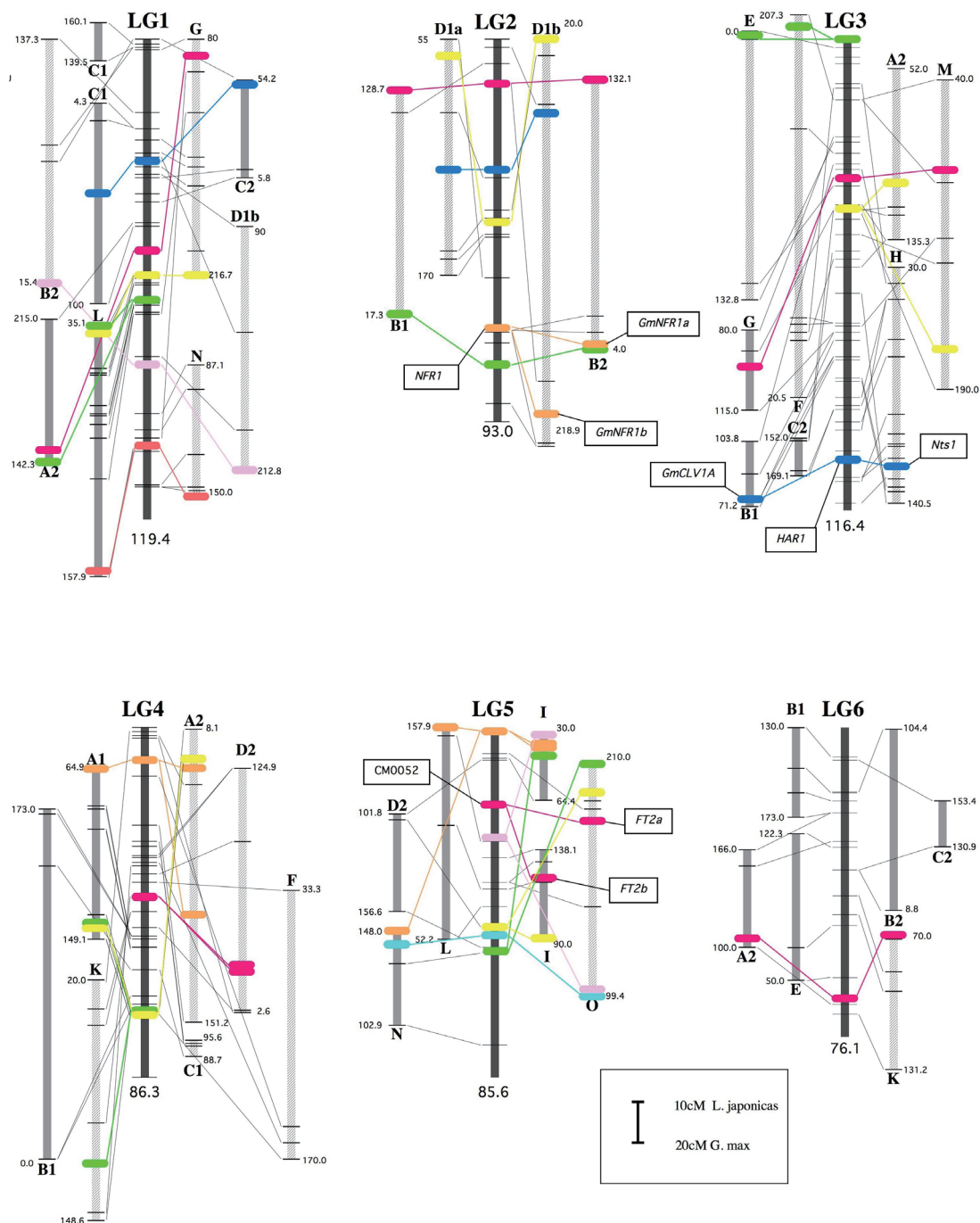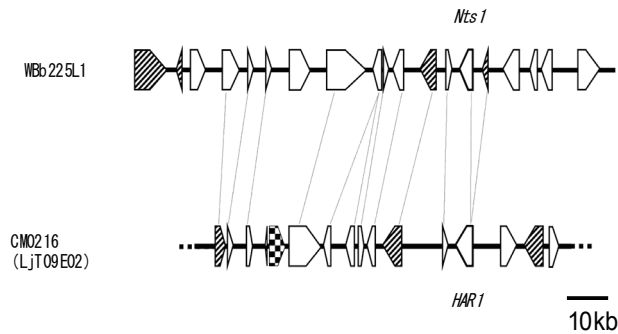

Fig. 5. L-G synteny map of six *Lj*LGs accompanied by macrosyntenic segments *Gm*LGs. The anchoring marker pairs were connected by strait lines. *Nts1, HAR1*: autoregulation genes for nodule number. *NFR1*: Nod-factor receptor gene. *GmNFR1a, GmNFR1b*: putative Nod-factor genes.
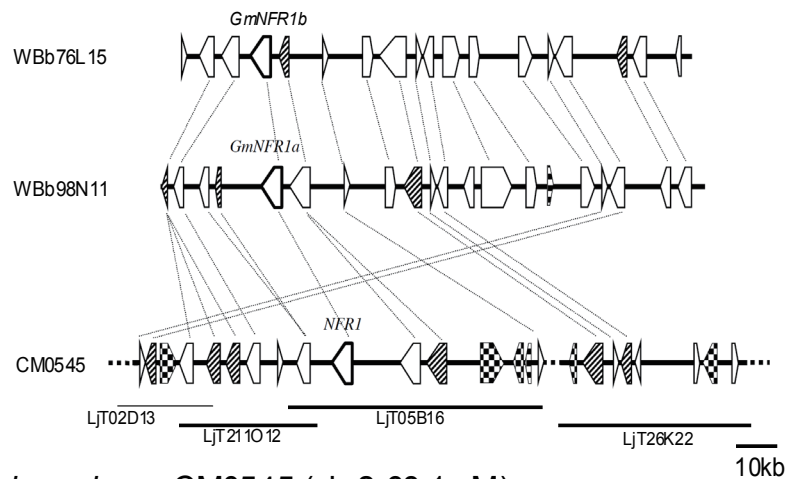
(A)

G. max : WBb225L1 (LG H)

L. japonicus : CM0216 (chr3 75.6 cM)

(B)

G. max : WBb76L15 (LG. B2), WBb98N11 (LG. D1b)

L. japonicus : CM0545 (chr2 68.1 cM)

Fig. 6. Microsynteny of (A) *Nts1/HAR1* orthologous and (B) *NFR1* orthologous regions in soybean and *L. japonicus*. Each block arrow on the horizontal lines represents a gene and each dotted line connects homologs. Shaded and dotted arrows indicate the truncated or mutated genes and retrotransposons, respectively.

References

Akkaya, M.S., R.C. Shoemaker, J.E. Specht, A.A. Bhagwat and P.B. Cregan (1995) Integration of simple sequence repeat DNA markers into a soy bean linkage map. Crop Sci. 35: 1439-1445.

Arumuganathan, K. and E.D. Earle (1991) Estimation of nuclear DNA content of plants by flow cytometry. Plant Mol. Biol. Rep. 9: 229-241.

Blanc, G. and K.H. Wolfe (2004) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell 16: 1667-1678.

Boutin, S.R., N.D. Young, T.C. Olson, Z.H. Yu, R.C. Shoemaker and C.E. Vallejos (1995) Genomeconservation among three legume genera detected with DNA markers. Genome 38: 928-937.

Cannon, S.B., W.R. McCombie, S. Sato, S. Tabata, R.Denny, L. Palmer, M. Katari, N. D. Young and G. Stacey (2003) Evolution and microsynteny of the

apyrase gene family in three legume genomes. Mol. Gen. Genomics 270: 347-361.

Choi, H.K., J.H. Mun, D.J. Kim, H. Zhu, J.M. Baek, J. Mudge, B. Roe, N. Ellis, J. Doyle, G.B. Kiss, N.D. Young and D.R. Cook (2004) Estimating genome conservation between crop and model legume species. Proc. Natl. Acad. Sci. U S A 101:15289-15294.

Choi, I.Y., D.L. Hyten, L.K. Matukumalli, Q. Song, J.M. Chaky, C.V. Quigley, K. Chase, K.G. Lark, R.S.Reiter, M.S. Yoon, E.Y. Hwang, S.I. Yi, N.D. Young, R.C. Shoemaker, C.P. van Tassel, J.E. Specht and P.B. Cregan (2007) A soybean transcript map: gene distribution, haplotype and single-nucleotide polymorphism analysis. Genetics 176: 685-696.

Cregan, P.B., T. Jarvik, A.L. Bush, R.C. Shoemarker, K.G. Lark, A.L. Kahler, N. Kaya, T.T. Van Todai, D.G.Lohnes, J. Chung and J.E. Specht (1999) An integrated genetic map of the soybean genome. Crop Sci. 39:1464-1490.

Hwang, T.Y., T. Sayama, M. Takahashi, Y. Takada, Y.Nakamoto, H. Funatsuki, H. Hisano, S. Sasamoto, S. Sato, S. Tabata, I. Kono, M. Hoshi, M.Hanawa, C. Yano, Z. Xia, K. Harada, K. Kitamura and M. Ishimoto (2009) High-density integrated linkage map based on SSR markers in soybean. DNA Res. 16:213-225.

Keim, P., B. W. Diers, T.C. Olson and R.C. Shoemaker (1990) RFLP mapping in soybean: Association between marker loci and variation in quantitative traits. Genetics 126: 735-742.

Lackey, J.A. (1980) Chromosome numbers in the Phaseoleae (Fabaceae: Faboideae) and their relation to taxonomy. Amer. J. Bot. 67: 595-602.

Lark, K.G., J.M. Weisemann, B.F. Matthews, R. Palmer, K. Chase and T. Macalma (1993) A genetic map of soybean (*Glycine max* L.) using an intraspecific cross of two cultivars: ' Minsoy' and ' Noir 1'. Theor. Appl. Genet. 86: 901-906.

Lee, J.M., D. Grant, C.E.Vallejos and R.C.Shoemaker (2001) Genome organization in dicots. II. *Arabidopsis* as a 'bridging species' to resolve genome evolution events among legumes. Theor. Appl. Genet. 103: 765-773.

Mansur, L.M., J.H. Orf, K. Chase, T. Jarvik, P.B. Cregan and K.G. Lark (1996) Genetic mapping of agronomic traits using recombinant inbred of soybean. Crop Sci. 36: 1327-1336.

Menanciohautea D, C.A. Fatokun, L. Kumar, D. Danesh and N.D. Young (1993) Comparative genome analysis of mungbean (*V.radiata* L. Wilczek) and cowpea (*V. unguiculata* L.Walpers) using RFLP mapping data. Theor. Appl. Genet. 86:797-810.

Mudge, J., S.B. Cannon, P. Kalo, G.E.D. Oldroyd, B.A. Roe, C.D. Town and N.D. Young (2005) Highly syntenic regions in the genomes of soybean, *Medicago truncatula*, and *Arabidopsis thaliana*. BMC Plant Biology 5: 15-30.

Schlueter, J.A., P. Dixon, C. Granger, D. Grant, L. Clark, J.J. Doyle and R.C. Shoemaker (2004) Mining EST databases to resolve evolutionary events in major crop species. Genome 47: 868-876.

Schlueter, J.A., B.E. Scheffler, S.D. Schlueter and R.C. Shoemaker (2006) Sequence conservation of homeologous bacterial artificial chromosomes and transcription of homeologous genes in soybean (*Glycine max* L. Merr.). Genetics 174: 1017-1028.

Schlueter, J.A., B.E. Scheffler, B. Roe, S.D. Schlueter and R.C. Shoemaker (2007a) The FAD2 family of soybean: insights into the structural and functional divergence of a paleopolyploid genome. The Plant Genome, a Suppl. to Crop Sci. 47(S1).

Schlueter, J.A., J.Y. Lin, S.D. Schlueter, I.F. Vasylenko-Sanders, S. Deshpande, J. Yi, M. O'Bleness, B.A. Roe, R.T. Nelson, B.E. Scheffler, S.A. Jacson and R.C. Shoemaker (2007b) Gene duplication and paleopolyploidy and the implications for whole genome sequencing. BMC Genomics 8:330-345.

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

Shoemaker, R.C., K. Polzin, J. Labate, J. Specht, E.C. Brummer, T. Olson, N. Young, V. Concibido, J. Wilcox, J.P. Tamulonis *et al*. (1996) Genome duplication in soybean (*Glycine* subgenus *soja)*. Genetics 144: 329-338.

Shoemaker, R.C., P. Keim, L. Vodkin, E. Retzel, S.W. Clifton, R. Waterston, D. Smoller, V. Coryell, A. Khanna, J. Erpelding *et al*. (2002) A compilation of soybean ESTs: generation and analysis. Genome 45:329-338.

Song, Q.J., L.F. Marek, R.C. Shoemarker, K.G. Lark, V.C. Concibido, X. Delanny, J.E. Specht and P.B. Cregan (2004) A new integrated linkage map of the soybean. Theor. Appl. Genet. 109: 122-128.

Triwitayakorn, K., V.N. Njiti, M.J. Iqbal, S. Yaegashi, C. Town and D.A. Lightfoot (2005) Genomic analysis of a region encompassing QRfs1 and QRfs2: genes that underlie soybean resistance to sudden death syndrome. Genome 48: 125-138.

Xia, Z., Y. Tsubokura, M. Hoshi, M. Hanawa, C.Yano, K. Okamura, T.A. Ahmed, T. Anai, S.Watanabe, M. Hayashi, T. Kawai, K.G. Hossain, H. Masaki,

K. Asai, N. Yamanaka, N. Kubo, K. Kadowaki, Y. Nagamura, M. Yano, T. Sasaki and K. Harada (2007) An integrated high-density linkage map of soybean with RFLP, SSR, STS, and AFLP markers using a single F_2 population. DNA Res. 14: 257-269.

Young, N.D., J. Mudge and T.H. Ellis (2003) Legume genomes: more than peas in a pod. Curr. Opin. Plant Biol. 6: 199-204.

Zhang, X.C., X. Wu, S. Findley, J. Wan, M. Libault, H.T. Nguyen, S.B. Cannon and G. Stacey (2007) Molecular evolution of LysM motif-type receptor like kinase in plants. Plant Physiol. 144: 623-636.