Genome Sequencing and Genome Resources in Model Legumes

Shusei Sato, Yasukazu Nakamura, Erika Asamizu, Sachiko Isobe, and Satoshi Tabata*

Kazusa DNA Research Institute, Kisarazu, Chiba 292–0818 Japan

Since the completion of the sequencing of the Arabidopsis (*Arabidopsis thaliana*) genome in 2000, it has become clear that information about the genome of a particular plant species can have dramatic benefits in promoting plant molecular genetics in general. To create a similar situation in legumes, two species with small genomes, *Lotus japonicus* (Japanese trefoil) and *Medicago truncatula* (barrel medic), with diploid genomes of 470 to 500 Mb in size, were chosen as references, and genome sequencing was launched at the beginning of this decade (Young et al., 2005). Although the genomes of both species have yet to be fully sequenced, a significant amount of information about their gene structures as well as physical and genetic maps has been made public. In addition, a variety of material resources, such as genomic and cDNA libraries, have been generated during the course of the sequencing work and as a result of sequence analysis. These information and material resources have already contributed to our understanding of genetic systems of biological importance, especially with respect to legume-specific phenomena (Stacey et al., 2006), and are expected to continue to augment research in this field as sequencing continues and is eventually completed for both genomes. Here, we briefly describe the current status of the genome sequencing of these two model legumes and summarize the information and material resources currently available to facilitate dissemination and exploitation of the resources within the various fields of legume biology.

STRUCTURAL ANALYSIS OF THE GENOMES OF MODEL LEGUMES

Details of the genome sequencing projects in *L. japonicus* and *M. truncatula* have been described previously (Young et al., 2005). Both projects basically adopted the same strategy, which was designed to preferentially sequence genespaces within the genomes. Multiple seed points on each genome were chosen on the basis of the sequences of known protein-encoding genes, and then corresponding bacterial artificial chromosome (BAC) or P1 phage-derived, transformation-competent artificial chromosome (TAC) clones (Liu et al., 1999) were selected by PCR or hybridization. The selected clones were sequenced with high accuracy before being subjected to gene modeling and annotation using a combination of computer programs. As of January 2007, 176 Mb (89 Mb finished, 9 Mb at phase 2, and 78 Mb at phase 1) and 189 Mb (122 Mb finished, 37 Mb at phase 2, and 30 Mb at phase 1) nonredundant sequences of the *L. japonicus* and the *M. truncatula* genomes, respectively, had been released. These correspond to approximately 40% of the entire genomes of both *L. japonicus* and *M. truncatula* with estimation of more than 60% coverage of the euchromatic regions, and cover 69% and 58% of public ESTs of *L. japonicus* and *M. truncatula*, respectively. In *L. japonicus*, shotgun sequencing is in progress to obtain draft sequences of the genespaces in the whole genome.

Gene assignment and gene modeling in *L. japonicus* were performed as follows. Similarity searches against several DNA and protein libraries were executed to find homologous regions to known gene sequences. Concurrently, ab initio gene-finding predictions were done, including those for splice sites (NetGene2 and splicepredictor), gene structure (GeneMark.hmm and Genescan), and protein-coding exons (Grail; Mural et al., 1992; Hebsgaard et al., 1996; Burge and Karlin, 1997; Brendel and Kleffe, 1998; Lukashin and Borodovsky, 1998). All of the data, including information about positions, scores, phase, and directions, were parsed and improved by manual curation to create the final gene models. The gap/nap (Huang et al., 1997) program was used for mapping ESTs and protein sequences on genomic sequences. The International *Medicago* Genome Annotation Group has selected a canonical genome annotation procedure. Gene modeling involves a pipeline composed of FGENESH (Salamov and Solovyev, 2000), EuGene (Foissac et al., 2003) gene structure predictor, and Combiner software (Allen et al., 2004). The Institute for Genomic Research (TIGR) PASA (Haas et al., 2003) was used for aligning ESTs on the genome sequences.

The general features of the deduced protein-encoding genes in the model legumes are summarized as follows. The average length of a coding exon is 304 bp and the average number of introns per gene is

* Corresponding author; e-mail tabata@kazusa.or.jp; fax 81–438–52–3934.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Satoshi Tabata (tabata@kazusa.or.jp).

www.plantphysiol.org/cgi/doi/10.1104/pp.107.097493

3.7 in *L. japonicus*. These features are quite similar to those of *M. truncatula* and Arabidopsis However, the average length of a gene, including introns, is longer in *L. japonicus* (2,883 bp) than in Arabidopsis (1,918 bp) due to the longer average length of an intron (Kato et al., 2003). Average gene density in the manually annotated regions of the *L. japonicus* genome (one gene in every 12.3 kb) is approximately half of the Arabidopsis genome (one gene in every 4.5 kb; TAIR6 Genome Statistics; http://www.arabidopsis.org/portals/genAnnotation/gene_structural_annotation/annotation_data.jsp#data), although this difference is less significant when retrotransposons are taken into account.

COMPARISON OF GENOME STRUCTURES AMONG MODEL PLANTS

Information about genome sequences and their positions in the genomes of both *L. japonicus* and *M. truncatula* have made it possible to explore genome-scale synteny, which provides a basis for comparative genomic studies. Using more than 100 Mb of genome sequences anchored on the genetic linkage maps, a preliminary detailed comparison of sequence similarity and commonalities in genome position between the two genomes has been carried out (Cannon et al., 2006). The results obtained clearly showed strong conservation of genome structures between the two species, despite incomplete genome coverage.

Comparison at the chromosomal level revealed the presence of large-scale synteny blocks (Choi et al., 2004b; Young et al., 2005). Typical examples are *L. japonicus* chr 5 − *M. truncatula* chr 1 and *L. japonicus* chr 1 − *M. truncatula* chr 3 + chr 7, where synteny was observed along the entire chromosomal regions (Cannon et al., 2006). Macrosyntenic relationships between *L. japonicus* and *M. truncatula* are composed of 10 large synteny blocks. Each block spans the whole chromosome arm in most cases, and the sum of the blocks covers over 60% of the genome lengths of both *M. truncatula* and *L. japonicus*. Within the individual synteny blocks, the level of gene colinearity is significant. Gene contents and their order were conserved for approximately 60% of the predicted genes, excluding those related to transposable elements (Cannon et al., 2006). It is likely that more sequence data as well as improvements in sequence accuracy and gene assignment will reveal still higher levels of microsynteny in the future.

During the course of the synteny analysis, segmental duplications within the genomes of *L. japonicus* and *M. truncatula* were often detected as synteny blocks (Cannon et al., 2006). The degree of internal synteny within each genome was approximately one-fifth of what would be expected if a very recent single duplication of the whole genome had occurred. This low degree of conservation may reflect an ancient whole-genome duplication that was later subjected to various types of mutation, including deletions, insertions, base substitutions, and translocations. Microsynteny level within the genomes of *L. japonicus* and *M. truncatula* remains about half of that between the two genomes. According to synonymous substitution and phylogenetic analyses, the putative large-scale duplications would predate the divergence of the two species (Cannon et al., 2006). This speculation is further supported by the presence of paired synteny blocks exhibiting comparable levels of synteny between the two species, each of which would correspond to a descendent of the ancient duplication event.

Significant microsyntenic but not macrosyntenic relationships have been detected between model legumes and other plant species, including Arabidopsis, poplar (*Populus trichocarpa*), and rice (*Oryza sativa*; Stracke et al., 2003; Kevei et al., 2005; Zhu et al., 2006). This limited degree of colinearity between the genomes of different plant species was found to be useful for map-based cloning in the model legumes (Stracke et al., 2003). Furthermore, efforts to identify putative orthologs of *L. japonicus* and *M. truncatula* genes required for nodulation and arbuscular mycorrhizal symbiosis in nonlegume plants on the basis of microsyntenic relations have been reported (Zhu et al., 2006).

INFORMATION AND MATERIAL RESOURCES FOR THE MODEL LEGUMES

Selected information resources for the two model legumes are listed in Table I. Kazusa DNA Research Institute provides a Web database for the *L. japonicus* genome sequence project (http://www.kazusa.or.jp/lotus). Annotated genome and EST sequences as well as a high-density genetic map with marker information are available. The database provides visualized baseline annotations, including ab initio gene predictions and alignment to EST/cDNA and known protein sequences. A central online database for the *M. truncatula* genome project (http://medicago.org/genome/) is maintained by the University of Minnesota and the Center for Computational Genomics and Bioinformatics (CCGB). This Web site offers a Java visualization interface for overlap information and status of sequence assembly of the BAC clones, and furnishes links to other *M. truncatula* sequencing or information centers, which in turn provide access to sequence-related information, such as genetic markers and BAC contigs (Cannon et al., 2005). The Legume Information System (LIS) at the National Center for Genome Resources (NCGR) provides a database that integrates genetic and molecular data from multiple legume species and generates visual presentations of cross-species comparisons performed at genome or linkage map levels (Gonzales et al., 2005).

Genetic mapping of determined sequences has been carried out to enhance the value of such information for molecular genetics and comparative genomics. In *L. japonicus*, an intraspecies genetic linkage map

Table I. *Selected model legume information resources*

Information Resource	Web Site
Lotus information resources	
L. japonicus sequencing project	http://www.kazusa.or.jp/lotus/
L. japonicus Gene Index at Harvard University	http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=l_japonicus
Munich Information Center for Protein Sequences (MIPS) *Lotus* genome database	http://mips.gsf.de/proj/plant/jsf/lotus
Medicago information resources	
M. truncatula sequencing resources	http://www.medicago.org/genome/
M. truncatula Consortium Database Version 2.0 (CCGB)	http://www.medicago.org:8180/MtDB2/
University of Oklahoma BACs and annotations (GBrowse view; chromosomes 1, 4, 6, 8)	http://www.genome.ou.edu/medicago.html
TIGR BACs and annotations (GBrowse view; chromosomes 2, 7)	http://www.tigr.org/tdb/e2k1/mta1/
European *Medicago* consortium (chromosomes 3, 5)	http://medicago.toulouse.inra.fr/
MIPS *Medicago* genome database (UrMeLDB)	http://mips.gsf.de/proj/plant/jsf/medi/
Medicago Gene Index at Harvard University	http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=medicago
Comparative genomics	
LIS at NCGR	http://www.comparative-legumes.org/lis/
LegumeDB	http://ccg.murdoch.edu.au/index.php/LegumeDB

created on the basis of a cross between two accessions, Gifu B-129 and Miyakojima MG-20, has been used as a reference (Hayashi et al., 2001). To anchor the sequenced TAC clones on this map, microsatellite or single nucleotide polymorphism markers were developed using the sequence information obtained (Sato et al., 2001; Nakamura et al., 2002; Kaneko et al., 2003; Asamizu et al., 2003; Kato et al., 2003). As of January 2007, a total of 1,589 clones had been located on the genetic linkage map using 814 microsatellite and 80 dCAPS markers and information on overlapping with the genetically mapped clones. Information about the microsatellite and dCAPS markers is available from the Web site at http://www.kazusa.or.jp/lotus/g_map/B-129xMG-20/table/SSR.html/ and http://www.kazusa.or.jp/lotus/g_map/B-129xMG-20/table/dCAPS.html/, respectively. In the *M. truncatula* genome project, genetic markers and maps were created in collaboration with the University of California at Davis, the University of Minnesota, the French National Institute for Agricultural Research (INRA)-Centre National de la Recherche Scientifique (Toulouse, France), and the Biological Research Center Institute of Genetics in Szeged, Hungary. A core genetic linkage map has been generated using two genotypes, A17 and A20 (Choi et al., 2004a). As of January 2007, 1,039 molecular markers had been located on the core genetic linkage map, and marker-associated BAC clones had been selected for 89% of these markers. Detailed information on these markers can be retrieved from the *M. truncatula* genome project Web site (http://www.medicago.org/genome/map.php). Together with those that overlap with the marker-associated clones, a total of 1,900 sequenced BAC clones (97% of total sequenced clones) have been assigned to their respective genetic loci.

EST information is crucial for accurate gene annotation of the genome sequence; it gives information about gene structure, alternative splicing, expression patterns, and transcript abundance. The *L. japonicus* EST information and transcriptome data are provided by the *Lotus* EST database at http://www.kazusa.or.jp/en/plant/lotus/EST/. The EST sequences were clustered into groups to reduce redundancy and were annotated with the BLASTX search results against a protein database UniRef100 (Apweiler et al., 2004). Transcriptome analyses of the *L. japonicus* nodulation process have been performed by two methods, cDNA macroarrays and serial analysis of gene expression (Kouchi et al., 2004; Asamizu et al., 2005). Expression data obtained from the array experiments are linked to the EST information and are provided from the same site. Integration of the EST and transcriptome data into the *Lotus* genome annotation database is in progress. *M. truncatula* EST databases have been created, such as the *Medicago* EST Navigation System (Journet et al., 2002), the Dana-Farber Cancer Institute *Medicago* Gene Index, which is the successor of the TIGR *Medicago* Gene Index (Lee et al., 2005), and the MtDB3 Nimbus Database available at http://medicago.org:8180/MtDB3/nimbus/project.do?project=MtDB3. This database provides links to overall information about the EST sequences, e.g. the UniRef100, hidden Markov model, and gene ontology terms.

As the genome projects of the two model legumes have advanced, material resources, such as a large number of cDNA clones and genomic libraries, have been developed. These resources constitute extremely valuable tools for genetic and physiological studies on individual biological phenomena when they are made available to the research community. Miyazaki University has established a resource center for *L. japonicus* and soybean (*Glycine max*) that is financially supported by the National Bioresource Project of Japan. They distribute TAC/BAC genomic and cDNA

clones generated during the course of *Lotus* genome analysis, as well as seeds of major experimental accessions and wild accessions collected from all over Japan. LegumeBase (http://www.legumebase.agr.miyazaki-u.ac.jp) provides a list of resources and detailed information. Recombinant inbred lines between accessions Miyakojima MG-20 and Gifu B-129 are also available, and phenotypic and genotyping data on 205 established lines can be accessed online through LegumeBase. John Innes Center provides a screening service of a TILLING population of approximately 5,000 M2 lines and a database (http://www.lotusjaponicus.org) of phenotypic descriptions and photographs of various mutants (Perry et al., 2003).

Most material resources for *M. truncatula* are distributed from the institutes where the resources were developed. cDNA and BAC libraries generated by European groups are collected and distributed to the research community from the Centre National de Ressources Génomiques Végétales in France (http://toulouse.inra.fr/cnrgv/). In the United States, mth1 and mth2 BAC libraries and derived individual clones are provided by the Clemson University Genomics Institute (http://www.genome.clemson.edu/), and EST clones from 14 distinct cDNA libraries are available from the Samuel Roberts Noble Foundation (http://www.noble.org/PlantBio/). Germlines and other genetic resources, such as recombinant inbred lines and mutant lines, are maintained at four resource centers, the South Australian Research and Development Institute, the U.S. Department of Agriculture National Plant Germplasm System, INRA, and the Samuel Roberts Noble Foundation (United States). Detailed information about the above resources has been summarized in an online-accessible article (Nair et al., 2006).

FROM MODELS TO CROPS

An essential role for model plants is to serve as a source of the accumulated knowledge to transfer to crop plants by means of common gene sequences and DNA markers. Though reliable and efficient procedures for knowledge transfer have yet to be established, intensive trials to utilize the information of model legumes in crops have been undertaken for a variety of purposes. One straightforward and promising approach—direct gene transfer from a model plant to a crop—is exemplified by a report using the isoflavone synthase gene of *M. truncatula* to engineer genistein glucoside production in alfalfa (*Medicago sativa*; Deavours and Dixon, 2005). An alternative approach is based on the use of DNA markers primarily for the identification of synteny among legumes, followed by the development of selection markers.

DNA markers that allow cross-species mapping, referred to as ''anchor markers,'' are essential for comparative genome analysis. The anchor markers are usually developed from the protein-encoding regions of the genome to define unique loci in the genetic linkage maps of multiple species. An intensive study in which anchor markers successfully detected orthologous loci in multiple legume genomes has been reported (Choi et al., 2006). Efforts to integrate the informative anchor markers into linkage maps of a variety of crop legumes are also in progress (Sato et al., 2005; Nelson et al., 2006; Phan et al., 2007). Furthermore, an automated bioinformatic pipeline for the development of anchor markers has been developed (Fredslund et al., 2006).

By utilizing the anchor markers as well as orthologous gene sequences, a growing number of studies have demonstrated extensive synteny among model and crop legumes, as summarized previously (Zhu et al., 2005). The syntenic relations identified in eight legume species have been compiled into a simplified consensus map. As an example of the extension of synteny analysis for the development of applications, a fine genetic map around a region of a virus-resistance gene (*Rsv4*) in soybean has been constructed using DNA markers generated by comparison of the genome sequences of soybean and *L. japonicus* (Hwang et al., 2006). Aubert et al. have reported the construction of a functional map in pea (*Pisum sativum*) that is composed of markers derived from the genes of both pea and *M. truncatula* (Aubert et al., 2006). It is hoped that this map will be used to identify genes responsible for phenotypic trait variability by the candidate gene approach.

To support comparative genome analysis among legumes, databases that integrate genetic and genomic data from multiple legume species have been developed. LIS (http://www.comparative-legumes.org/lis/) and LegumeDB (http://ccg.murdoch.edu.au/index.php/LegumeDB; Gonzales et al., 2005; Moolhuijzen et al., 2006) are examples. These databases provide curated genetic maps, and genomic and transcriptomic data from various sources, allowing cross-species comparison with well-organized viewers. On the other hand, for a comparison of the genomes of a wider range of plant species, the Phytome project (www.phytome.org) provides an informative and user-friendly platform (Hartmann et al., 2006). This database collects protein-encoding gene sequences from a variety of angiosperms, including model and crop legumes, and emphasizes phylogenomics, thus facilitating the identification of orthologous and paralogous sequences.

PERSPECTIVES

The increasing rate of accumulation of genomic sequence information allied to the increasing availability of associated material resources is drastically accelerating the pace of investigation of the genetic backgrounds of individual biological phenomena in model legumes. Furthermore, the development of these resources has facilitated the introduction of ''omics'' approaches that are crucial for our comprehensive

understanding of whole genetic systems in legume plants. Comparison of the knowledge obtained by performing comparative genomics between legumes and other model plants, such as Arabidopsis and rice, would not only allow the identification of legume-specific systems but also would provide insights into plant genome evolution.

Model plants hold the promise that their genomic and genetic information will be used for crop breeding. A typical example are the *Poaceae*, where the sequencing of the rice genome is encouraging the development of molecular genetics approaches to other *Poaceae* crops, such as maize (*Zea mays*), sorghum (*Sorghum bicolor*), wheat (*Triticum aestivum*), and barley (*Hordeum vulgare*; Jaiswal et al., 2006). Accumulating sequence information and comparative genomics among *Poaceae* crops has identified and permitted the screening of many kinds of quantitative trait loci related to important agronomic traits in marker-assisted selection breeding programs (Koebner and Summers, 2003; Tamasaki et al., 2005; Ashikari and Matsuoka, 2006). On the other hand, the situation is somewhat different in the case of the legume family mainly for two reasons. First, the two model plants *L. japonicus* and *M. truncatula* are wild and noncultivated species, and, second, legume crops are more diverse in terms of phylogeny and potential uses (Graham and Vance, 2003). For this reason, the sources and types of information acquired from these models should be carefully sifted according to the targeted crops and breeding objectives. For example, it is generally accepted that knowledge about certain shared characteristics of legumes, such as the pathways involved in symbiosis with rhizobia and the synthesis of flavonoids and glycosides, is considerably transferable from models to crops. However, such model information may not be very valuable for improving certain specific traits, such as the seed yield and oil contents of bean crops, e.g. soybean and pea, because it is likely that such traits have already been acquired by these crops during the course of their domestication but not by wild species. In contrast, information from these two models may be useful for forage legumes, e.g. alfalfa and clovers, even for the study of holistic agronomic traits, such as yield and growth habits, because *L. japonicus* and *M. truncatula* are related to the forage legume species *Lotus corniculatus* (birdsfoot trefoil) and alfalfa, respectively.

The transfer of knowledge acquired from model plants, mainly through orthologous gene sequences and DNA markers, results in identification and isolation of the corresponding genes using genomic and/or cDNA libraries and common DNA markers on genetic linkage maps in crop plants. Such resources have been vigorously developed for model legumes but not intensively for crop legumes so far. A remarkable exception is sequencing of the soybean genome (http://www.energy.gov/news/2979.htm). The creation of a larger number of ''common words'' will be crucial for the more efficient transfer of knowledge and for facilitating the exchange of knowledge between researchers working on legumes. In addition, organization of a public system to share the information and material resources within the legume community, e.g. genome databases, bioinformatics tools, and resource centers, is urgently required.

ACKNOWLEDGMENT

We thank Drs. Giles Oldroyd and Nevin Young for updated information on the *M. truncatula* genome sequencing project.

Received February 3, 2007; accepted March 10, 2007; published June 6, 2007.

LITERATURE CITED

Allen JE, Pertea M, Salzberg SL (2004) Computational gene prediction using multiple sources of evidence. Genome Res **14:** 142–148

Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, et al (2004) UniProt: the Universal Protein knowledgebase. Nucleic Acids Res **32:** D115–D119

Asamizu E, Kato T, Sato S, Nakamura Y, Kaneko T, Tabata S (2003) Structural analysis of a *Lotus japonicus* genome. IV. Sequence features and mapping of seventy-three TAC clones which cover the 7.5 mb regions of the genome. DNA Res **10:** 115–122

Asamizu E, Nakamura Y, Sato S, Tabata S (2005) Comparison of the transcript profiles from the root and the nodulating root of the model legume *Lotus japonicus* by serial analysis of gene expression. Mol Plant Microbe Interact **18:** 487–498

Ashikari M, Matsuoka M (2006) Identification, isolation and pyramiding of quantitative trait loci for rice breeding. Trends Plant Sci **11:** 344–350

Aubert G, Morin J, Jacquin F, Loridon K, Quillet MC, Petit A, Rameau C, Lejeune-Hénaut I, Huguet T, Burstin J (2006) Functional mapping in pea, as an aid to the candidate gene selection and for investigating synteny with the model legume *Medicago truncatula*. Theor Appl Genet **112:** 1024–1041

Brendel V, Kleffe J (1998) Prediction of locally optimal splice sites in plant pre-mRNA with applications to gene identification in *Arabidopsis thaliana* genomic DNA. Nucleic Acids Res **26:** 4748–4757

Burge C, Karlin S (1997) Prediction of complete gene structures in human genomic DNA. J Mol Biol **268:** 78–94

Cannon SB, Crow JA, Heuer ML, Wang X, Cannon EK, Dwan C, Lamblin AF, Vasdewani J, Mudge J, Cook A, et al (2005) Databases and information integration for the *Medicago truncatula* genome and transcriptome. Plant Physiol **138:** 1–3

Cannon SB, Sterck L, Rombauts S, Sato S, Cheung F, Gouzy J, Wang X, Mudge J, Vasdewani J, Schiex T, et al (2006) Legume genome evolution viewed through the *Medicago truncatula* and *Lotus japonicus* genomes. Proc Natl Acad Sci USA **103:** 14959–14964

Choi HK, Kim D, Uhm T, Limpens E, Lim H, Mun JH, Kalo P, Penmetsa RV, Seres A, Kulikova O, et al (2004a) A sequence-based genetic map of *Medicago truncatula* and comparison of marker colinearity with *M. sativa*. Genetics **166:** 1463–1502

Choi HK, Luckow MA, Doyle J, Cook DR (2006) Development of nuclear gene-derived molecular markers linked to legume genetic maps. Mol Genet Genomics **276:** 56–70

Choi HK, Mun JH, Kim DJ, Zhu H, Baek JM, Mudge J, Roe B, Ellis N, Doyle J, Kiss GB, et al (2004b) Estimating genome conservation between crop and model legume species. Proc Natl Acad Sci USA **101:** 15289–15294

Deavours BE, Dixon RA (2005) Metabolic engineering of isoflavonoid biosynthesis in alfalfa. Plant Physiol **138:** 2245–2259

Foissac S, Bardou P, Moisan A, Cros MJ, Schiex T (2003) EuGene'Hom: a generic similarity-based gene finder using multiple homologous sequences. Nucleic Acids Res **31:** 3742–3745

Fredslund J, Madsen LH, Hougaard BK, Sandal N, Stougaard J, Bertioli D, Schauser L (2006) GeMprospector—online design of cross-species genetic marker candidates in legumes and grasses. Nucleic Acids Res **34:** W670–W675

Graham PH, Vance CP (2003) Legumes: importance and constraints to greater use. Plant Physiol **131:** 872–877

Gonzales MD, Archuleta E, Farmer A, Gajendran K, Grant D, Shoemaker R, Beavis WD, Waugh ME (2005) The Legume Information System (LIS): an integrated information resource for comparative legume biology. Nucleic Acids Res **33:** D660–D665

Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK Jr, Hannick LI, Maiti R, Ronning CM, Rusch DB, Town CD, et al (2003) Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res **31:** 5654–5666

Hartmann S, Lu D, Phillips J, Vision TJ (2006) Phytome: a platform for plant comparative genomics. Nucleic Acids Res **34:** D724–D730

Hayashi M, Miyahara A, Sato S, Kato T, Yoshikawa M, Taketa M, Hayashi M, Pedrosa A, Onda R, Imaizumi-Anraku H, et al (2001) Construction of a genetic linkage map of the model legume *Lotus japonicus* using an intraspecific F2 population. DNA Res **8:** 301–310

Hebsgaard SM, Korning PG, Tolstrup N, Engelbrecht J, Rouze P, Brunak S (1996) Splice site prediction in *Arabidopsis thaliana* DNA by combining local and global sequence information. Nucleic Acids Res **24:** 3439–3452

Huang X, Adams MD, Zhou H, Kerlavage AR (1997) A tool for analyzing and annotating genomic sequences. Genomics **46:** 37–45

Hwang TY, Moon JK, Yu S, Yang K, Mohankumar S, Yu YH, Lee YH, Kim HS, Kim HM, Maroof MAS, et al (2006) Application of comparative genomics in developing molecular markers tightly linked to the virus resistance gene *Rsv*4 in soybean. Genome **49:** 380–388

Jaiswal P, Ni J, Yap I, Ware D, Spooner W, Youens-Clark K, Ren L, Liang C, Zhao W, Ratnapu K, et al (2006) Gramene: a bird's eye view of cereal genomes. Nucleic Acids Res **34:** D717–D723

Journet EP, van Tuinen D, Gouzy J, Crespeau H, Carreau V, Farmer MJ, Niebel A, Schiex T, Jaillon O, Chatagnier O, et al (2002) Exploring root symbiotic programs in the model legume *Medicago truncatula* using EST analysis. Nucleic Acids Res **30:** 5579–5592

Kaneko T, Asamizu E, Kato T, Sato S, Nakamura Y, Tabata S (2003) Structural analysis of a *Lotus japonicus* genome. III. Sequence features and mapping of sixty-two TAC clones which cover the 6.7 Mb regions of the genome. DNA Res **10:** 27–33

Kato T, Sato S, Nakamura Y, Kaneko T, Asamizu E, Tabata S (2003) Structural analysis of a *Lotus japonicus* genome. V. Sequence features and mapping of sixty-four TAC clones which cover the 6.4 mb regions of the genome. DNA Res **10:** 277–285

Kevei Z, Seres A, Kereszt A, Kalo P, Kiss P, Toth G, Endre G, Kiss GB (2005) Significant microsynteny with new evolutionary highlights is detected between *Arabidopsis* and legume model plants despite the lack of macrosynteny. Mol Genet Genomics **274:** 644–657

Koebner RM, Summers RW (2003) 21st century wheat breeding: plot selection or plant detection? Trends Biotechnol **21:** 59–63

Kouchi H, Shimomura K, Hata S, Hirota A, Wu GJ, Kumagai H, Tajima S, Suganuma N, Suzuki A, Aoki T, et al (2004) Large-scale analysis of gene expression profile during early stages of root nodule formation in a model legume, *Lotus japonicus*. DNA Res **11:** 263–274

Lee Y, Tsai J, Sunkara S, Karamycheva S, Pertea G, Sultana R, Antonescu V, Chan A, Cheung F, Quackenbush J (2005) The TIGR Gene Indices: clustering and assembling EST and known genes and integration with eukaryotic genomes. Nucleic Acids Res **33:** D71–D74

Liu YG, Shirano Y, Fukaki H, Yanai Y, Tasaka M, Tabata S, Shibata D (1999) Complementation of plant mutants with large genomic DNA fragments by a transformation-competent artificial chromosome vector accelerates positional cloning. Proc Natl Acad Sci USA **96:** 6535–6540

Lukashin AV, Borodovsky M (1998) GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res **26:** 1107–1115

Moolhuijzen P, Cakir M, Hunter A, Schibeci D, Macgregor A, Smith C, Francki M, Jones MG, Appels R, Bellgard M (2006) LegumeDB bioinformatics resource: comparative genomic analysis and novel cross-genera marker identification in lupin and pasture legume species. Genome **49:** 689–699

Mural RJ, Einstein JR, Guan X, Mann RC, Uberbacher EC (1992) An artificial intelligence approach to DNA sequence feature recognition. Trends Biotechnol **10:** 66–69

Nair R, Hughes SJ, Ellwood S, Oliver R, Green SL, Delalande M, Wen J, Oldroyd G (2006) *M. truncatula* stock centre. *In* The *Medicago truncatula* Handbook. Samuel Roberts Noble Foundation. http://www.noble.org/MedicagoHandbook/ (December 1, 2006)

Nakamura Y, Kaneko T, Asamizu E, Kato T, Sato S, Tabata S (2002) Structural analysis of a *Lotus japonicus* genome. II. Sequence features and mapping of sixty-five TAC clones which cover the 6.5-mb regions of the genome. DNA Res **9:** 63–70

Nelson MN, Phan HT, Ellwood SR, Moolhuijzen PM, Hane J, Williams A, O'Lone CE, Fosu-Nyarko J, Scobie M, Cakir M, et al (2006) The first gene-based map of *Lupinus angustifolius* L.-location of domestication genes and conserved synteny with *Medicago truncatula*. Theor Appl Genet **113:** 225–238

Perry JA, Wang TL, Welham TJ, Gardner S, Pike JM, Yoshida S, Parniske M (2003) A TILLING reverse genetics tool and a web accessible collection of mutants of the legume *Lotus japonicus*. Plant Physiol **131:** 866–871

Phan HT, Ellwood SR, Hane JK, Ford R, Materne M, Oliver RP (2007) Extensive macrosynteny between *Medicago truncatula* and *Lens culinaris* ssp. culinaris. Theor Appl Genet **114:** 549–558

Salamov A, Solovyev V (2000) *Ab initio* gene finding in *Drosophila* genomic DNA. Genome Res **10:** 516–522

Sato S, Isobe S, Asamizu E, Ohmido N, Kataoka R, Nakamura Y, Kaneko T, Sakurai N, Okumura K, Klimenko I, et al (2005) Comprehensive structural analysis of the genome of red clover (*Trifolium pratense* L.). DNA Res **12:** 301–364

Sato S, Kaneko T, Nakamura Y, Asamizu E, Kato T, Tabata S (2001) Structural analysis of a *Lotus japonicus* genome. I. Sequence features and mapping of fifty-six TAC clones which cover the 5.4 Mb regions of the genome. DNA Res **8:** 311–318

Stacey G, Libault M, Brechenmacher L, Wan J, May GD (2006) Genetics and functional genomics of legume nodulation. Curr Opin Plant Biol **9:** 110–121

Stracke S, Sato S, Sandal N, Koyama M, Kaneko T, Tabata S, Parniske M (2003) Exploitation of colinear relationships between the genomes of Lotus japonicus, Pisum sativum and Arabidopsis thaliana, for positional cloning of a legume symbiosis gene. Theor Appl Genet **108:** 442–444

Tamasaki M, Tenaillon MI, Bi IV, Schroeder SG, Sanchez-Villeda H, Doebley JF, Gaut BS, McMullenn MD (2005) A large-scale screen for artificial selection in maize identified candidate agronomic loci for domestication and crop improvement. Plant Cell **7:** 2859–2872

Young ND, Cannon SB, Sato S, Kim D, Cook DR, Town CD, Roe BA, Tabata S (2005) Sequencing the genespaces of *Medicago truncatula* and *Lotus japonicus*. Plant Physiol **137:** 1174–1181

Zhu H, Choi HK, Cook DR, Shoemaker RC (2005) Bridging model and crop legumes through comparative genomics. Plant Physiol **137:** 1189–1196

Zhu H, Riely BK, Burns NJ, Ane JM (2006) Tracing nonlegume orthologs of legume genes required for nodulation and arbuscular mycorrhizal symbioses. Genetics **172:** 2491–2499