# Estimating genome conservation between crop and model legume species

Hong-Kyu Choi[†‡], Jeong-Hwan Mun[†‡], Dong-Jin Kim[†‡], Hongyan Zhu[†], Jong-Min Baek[§], Joanne Mudge[¶], Bruce Roe[‖], Noel Ellis[††], Jeff Doyle[‡‡], Gyorgy B. Kiss[†§§], Nevin D. Young[¶], and Douglas R. Cook[†§¶¶]

[†]Department of Plant Pathology and [§]College of Agricultural and Environmental Sciences Genomics Facility, University of California, One Shields Avenue, Davis, CA 95616; [¶]Department of Plant Pathology, University of Minnesota, St. Paul, MN 55108; [‖]Advanced Center for Genome Technology, University of Oklahoma, Norman, OK 73019; [††]John Innes Centre, Norwich NR4 7UH, United Kingdom; [‡‡]Department of Plant Biology, Cornell University, Ithaca, NY 14853; and [§§]Biological Research Center, Institute of Genetics, H-6701 Szeged, Hungary

Edited by Susan R. Wessler, University of Georgia, Athens, GA, and approved August 13, 2004 (received for review March 30, 2004)

**Legumes are simultaneously one of the largest families of crop plants and a cornerstone in the biological nitrogen cycle. We combined molecular and phylogenetic analyses to evaluate genome conservation both within and between the two major clades of crop legumes. Genetic mapping of orthologous genes identifies broad conservation of genome macrostructure, especially within the galegoid legumes, while also highlighting inferred chromosomal rearrangements that may underlie the variation in chromosome number between these species. As a complement to comparative genetic mapping, we compared sequenced regions of the model legume *Medicago truncatula* with those of the diploid *Lotus japonicus* and the polyploid *Glycine max*. High conservation was observed between the genomes of *M. truncatula* and *L. japonicus*, whereas lower levels of conservation were evident between *M. truncatula* and *G. max*. In all cases, conserved genome microstructure was punctuated by significant structural divergence, including frequent insertion/deletion of individual genes or groups of genes and lineage-specific expansion/contraction of gene families. These results suggest that comparative mapping may have considerable utility for basic and applied research in the legumes, although its predictive value is likely to be tempered by phylogenetic distance and genome duplication.**

The Fabaceae, or legumes, are cultivated on 180 million hectares, involving ≈12% of Earth's arable land and accounting for ≈27% of the world's primary crop production (1). Their unusual capacity for symbiotic nitrogen fixation underlies their importance as a source of protein in the human diet and of nitrogen in both natural and agricultural ecosystems. Legumes are also increasingly recognized as a source of valuable secondary metabolites. These factors have fueled a significant increase in legume research over the past decade.

The ≈20,000 legume species are divided into three subfamilies: Mimosoideae, Caesalpinioideae, and the numerically and economically dominant Papilionoideae (2). With the notable exception of peanut, the important crop legumes occur in two Papilionoid clades, referred to here as the "phaseoloid" and "galegoid" legumes (Table 1). Despite their close phylogenetic affiliations (Fig. 1), the genetic systems represented within this group are diverse, ranging from simple autogamous diploids to complex out-crossing polyploids. Genome size also varies widely among legumes, with pea having a genome size 10 times that of some related diploid genera.

The large number of important legume species precludes their simultaneous in-depth characterization. Moreover, several crop legumes have one or more characters (e.g., medium to large genomes and/or polyploid nature) that limit their utility as experimental systems. Two legumes with favorable genetic attributes, namely *Medicago truncatula* and *Lotus japonicus*, have been selected as model species and are the focus of large multinational genome projects. The early fruits of working with these well characterized genomes are evident in the recent advances in our understanding of symbiotic nitrogen fixation in both *M. truncatula* and *L. japonicus* (3).

A pressing need in legume genomics is to integrate knowledge gained from the study of model legume genomes with the biological and agronomic questions of importance in the crop species. Comparative genetic mapping is well established in several plant families, most notably the Poaceae (4), where initial studies predicted that synteny would greatly facilitate gene discovery among related species (5, 6). However, even closely related grass species (7, 8), in some cases members of the same species (9), can exhibit significant divergence in genome organization. It is important to know whether similar features are prevalent in other plant families, in particular because the extent of such differences may define the limits of comparative structural genomics as a strategy for applied agriculture.

Here we combined genetic and phylogenetic analyses to map putatively orthologous genes across seven legume species. Complementing the genetic linkage analysis, we surveyed the conservation of genome microstructure between *M. truncatula* and *L. japonicus* and *M. truncatula* and *Glycine max* (soybean) by comparing fully sequenced bacterial artificial chromosome (BAC) clones. The combined genetic, phylogenetic, and genomic analyses demonstrate extensive conservation of gene order and orthology between the crop and model legumes and also identify features of structural divergence between these genomes.

## Materials and Methods

**Plant Material and Mapping Populations.** Plant genotypes used for genetic mapping are shown in Table 1. The core maps for each of the genomes under analysis have been described (6–15). In the cases of *Pisum sativum* and *Vigna radiata*, mapping populations were composed of highly selected genotypes chosen so that recombination break points were spaced evenly throughout the genome (16). BAC clones were mapped in *M. truncatula* by means of simple sequence repeat polymorphisms discovered in the course of shotgun sequencing.

**Development of Cross-Species Genetic Markers.** The development and genetic mapping of gene-specific markers were as described by Choi *et al.* (10). BLASTN [National Center for Biotechnology Information (NCBI), Bethesda] was used to identify conserved sequences between ESTs of the legume *M. truncatula* and other legume species. Multiple sequence alignments, with the *Arabidopsis* genomic sequence used to infer intron position, facilitated design of PCR primers that anneal to conserved exon sequences and amplify across more diverged introns. Polymorphisms (Table 2, which is published as supporting information on the PNAS web site) were identified by sequencing PCR products from

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: BAC, bacterial artificial chromosome; NCBI, National Center for Biotechnology Information.

[‡]H.-K.C., J.-H.M., and D.-J.K. contributed equally to this work.

[¶¶]To whom correspondence should be addressed. E-mail: drcook@ucdavis.edu.

**Table 1. Attributes of species used for synteny analysis**

| Species | Common name | Genome size, Mbp | *N* | Tribe | Clade | SL | PL | Genotypes |
|---|---|---|---|---|---|---|---|---|
| *M. truncatula* | Barrel medic | 500 | 8 | Trifoleae | Galegoid | 183 | 130 | A17, A20, DZA |
| *M. sativa* | Alfalfa | 1,600 | 16 | Trifoleae | Galegoid | 70 | 68 | Mscw2, Msq93 |
| *Pi. sativum* | Pea | 5,000 | 7 | Viceae | Galegoid | 101 | 68 | JI15, JI281, JI399, JI194 |
| *G. max* | Soybean | 1,100 | 20 | Phaseoleae | Phaseoloid | 56 | 15 | PI209322, Evans |
| *V. radiata* | Mung bean | 520 | 11 | Phaseoleae | Phaseoloid | 62 | 31 | TC1966, VC3890 |
| *Ph. vulgaris* | Common bean | 620 | 11 | Phaseoleae | Phaseoloid | 37 | 22 | BAT93, Jalo |
| *L. japonicus* | Bird's foot trefoil | 500 | 6 | Loteae | | 67 | 44 | *Lotus filicaulis, L. japonicus Gifu* |

SL, sequenced loci; PL, polymorphic loci; *N*, gametic chromosome number.

parental lines (Table 1), followed by manual inspection of alignments and chromatograms. Markers were typically analyzed as cleavable amplified polymorphic sequences (10). Single-nucleotide polymorphisms that did not alter a restriction site were scored by DNA sequencing of PCR products. Resequencing was used to confirm or refute apparently ambiguous data.

**Phylogenetic Analysis.** Neighbor-joining trees were rooted by using the closest *Arabidopsis* sequence as an outgroup or left unrooted where no close homolog was present in *Arabidopsis*. Phylogenetic analyses were conducted by using parsimony options in PAUP* (17). The principal analysis involved 100 searches with random taxon addition and tree bisection–reconnection (TBR) branch swapping, with maxtrees set to increase without limit. Support for branches was assayed by 100 bootstrap replicate searches using simple taxon addition, TBR branch swapping, and maxtrees set to 1,000.

**Microsynteny Analysis.** Accession numbers for sequenced BACs are given in Tables 2–5, which are published as supporting information on the PNAS web site. Homologous transformation-competent BAC (TAC) clones of *L. japonicus* were obtained from NCBI (18). *Ab initio* gene prediction involved the eudicot version of FGENESH (www.softberry.com/berry.phtml?topic = gfind). Gene prediction based on identity to transcribed sequences was obtained by BLASTN against The Institute for Genomic Research *M. truncatula* or *L. japonicus* Gene Index databases (www.tigr.org/tdb/tgi). Additional predicted proteins were identified by means of BLASTX (NCBI) against the NCBI nonredundant protein database. BLASTP (NCBI) was used to compare predicted proteins between *M. truncatula* and *L. japonicus* clone pairs, with a maximum *E* value cutoff of $e^{-10}$ and a median *E* value of $<e^{-100}$ for 533 protein pairs. REPEATMASKER (http://repeatmasker.genome.washington.edu/cgi-bin/RepeatMasker) was used to screen interspersed repeats, transfer RNA, and low-complexity DNA sequences.

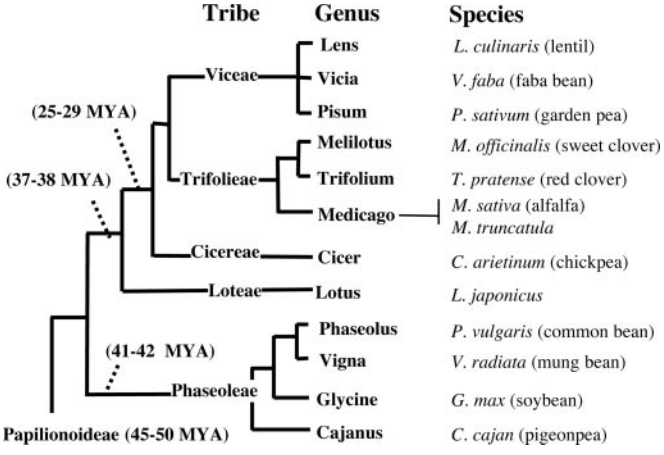


**Fig. 1.** Taxonomic relationships within the two major clades of crop legumes, the prevailing view of phylogeny for the species under analysis, with divergence times estimated based on Penalised Likelihood analysis (2). Most crop legumes occur either within the galegoid clade, including tribes Viceae, Trifolieae, and Cicereae, or within the phaseoloid clade, which is synonymous with the tribe Phaseoleae. MYA, million years ago.

## Results

**Development of Cross-Species Gene-Specific Genetic Markers.** We sought to develop cross-species genetic markers where locus orthology was an explicit aspect of the analysis. One hundred sixty-seven gene-specific PCR primer pairs were tested for amplification and polymorphism by using the parents of available mapping populations in *M. truncatula*, *Medicago sativa*, *Pi. sativum*, *L. japonicus*, *V. radiata*, *Phaseolus vulgaris*, and *G. max* (Table 1). To test the orthology of these genetic markers, we constructed phylogenies for 24 of the markers that produced high-quality sequence data for at least four of the six species under analysis (excluding *M. sativa*) (as shown by example in Fig. 6, which is published as supporting information on the PNAS web site). Combining alignment matrices and analyzing the data by neighbor-joining and maximum parsimony methods yielded a tree depicting genomic relationships. The monophyly of the galeoid and phaseoloid clades was strongly supported, as was the sister relationship of *Phaseolus* and *Vigna*. Analysis of individual gene trees supports the Loteae as sister to the galegoid clade. As a further test of sequence orthology, 11 markers with unambiguous global alignments were analyzed across 95 diploid legume species (Fig. 7, which is published as supporting information on the PNAS web site) spanning the diversity of the Fabaceae. The overall agreement with published phylogenies of the family (2, 19) supports the orthology of these genes and the utility of these exon-derived markers as tools for comparing legume genomes.

**Macrosynteny Analysis Among *M. truncatula*, *M. sativa*, *Pi. sativum*, *G. max*, *V. radiata*, and *Ph. vulgaris*.** For purposes of establishing a comparative genetic map spanning galegoid and phaseoloid species, we analyzed marker segregation in *M. truncatula*, *M. sativa*, *Pi. sativum*, and *V. radiata*. In addition to the markers developed based on phylogenetic criteria, we analyzed 60 primer pairs developed based on homology to genetic markers in *G. max* and 117 additional markers developed for the *M. truncatula* core genetic map (10). In all cases, *M. truncatula* was the central point of comparison. Comparisons between the two *Medicago* species and between *Ph. vulgaris* and *V. radiata* have been presented elsewhere (11, 12) and are included here for the sake of integration.

The pea genome is much larger (≈10 times) than that of *M. truncatula* and has a base chromosome number of 7, compared to 8 in *M. truncatula*. Despite these overt differences, analysis of 57 gene-specific markers reveals broad conservation of genome structure, with the major evident differences being sites of

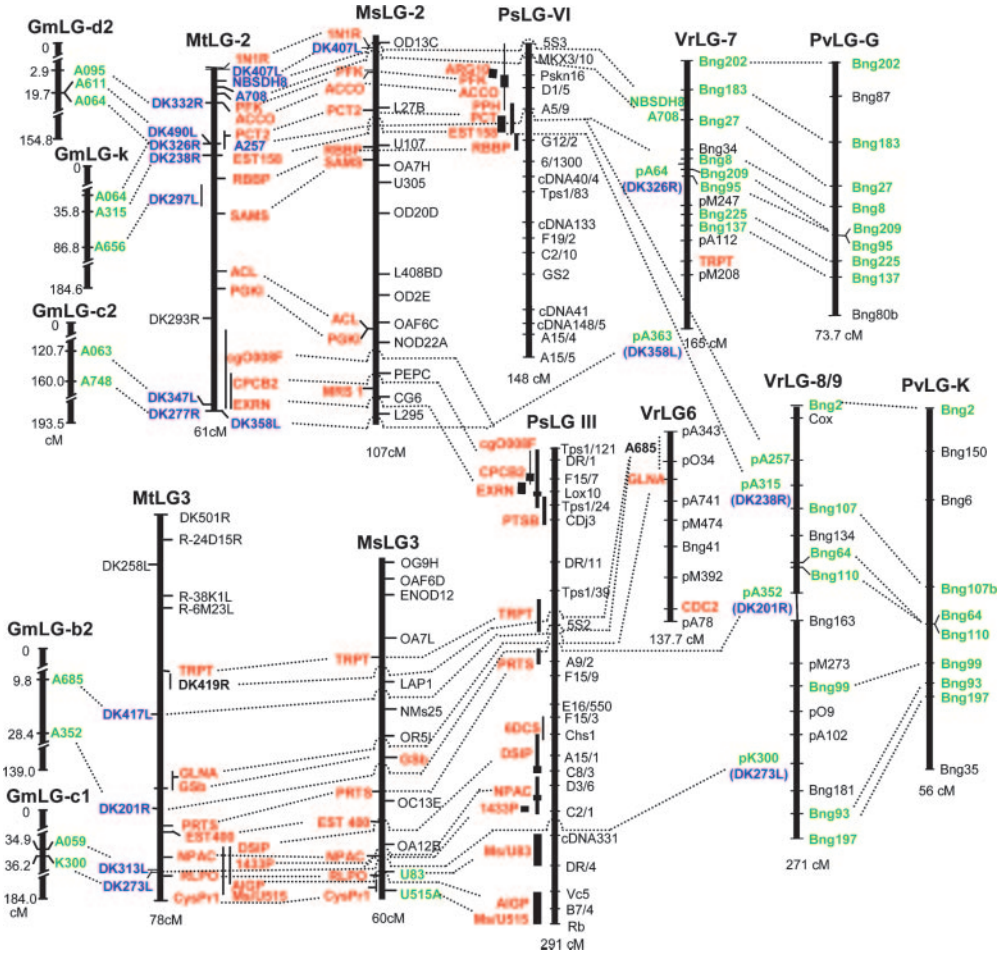


**Fig. 2.** Macrosyntenic relationships among legumes with reference to *M. truncatula* linkage groups 2 and 3. Mt, *M. truncatula*; Ms, *M. sativa*; Ps, *Pi. sativum*; Vr, *V. radiata*; Gm, *G. max*; Pv, *Ph. vulgaris*. Shared comparative markers are denoted by bold lettering. Extrapolated marker positions are denoted by dotted semicircle lines and are drawn only when the colinearity is conserved in the neighboring regions. The comparison between *V. radiata* and *Ph. vulgaris* is based on previous work (11–15).

inferred chromosomal rearrangements. An average of eight colinear genetic markers were present for each pea chromosome, and in only one case (i.e., marker PTSB) did we identify the possible translocation of an orthologous gene. Instead, we identified a limited number of chromosomal translocation events such as that shown for the top terminal region of PsLGIII and the bottom portion of MtLG2 (Fig. 2). MtLG6 could not be effectively integrated into the *Pi. sativum* genetic map, due to a lack of comparative markers in this linkage group. This result corresponds with the previous observation that MtLG6 is rich in heterochromatic DNA (20) and relatively poor in transcribed genes (10). The absence of a corresponding single linkage group in pea suggests that chromosomal rearrangements involving *M. truncatula* chromosome 6 might be responsible for the difference in chromosome number between these two species.

*V. radiata* (mung bean) and *Ph. vulgaris* (common bean) represent closely related members of the *Phaseoleae*, both with a chromosome number of each other. The genetic analysis of 22 gene-specific markers and 16 PCR markers converted from restriction fragment length polymorphisms (11, 12) revealed a combination of marker colinearity, inferred translocations or duplications, and nonsyntenic loci (Figs. 2 and 8, which is published as supporting information on the PNAS web site). Twenty-nine of the 38 markers tested revealed evidence of conserved gene order between *M. truncatula* and *V. radiata*. The remaining nonsyntenic markers either mapped distal to regions of colinearity or their positions interrupted syntenic markers. In many cases, syntenic genes bracketed inferred rearrangements, such as the possible duplication/translocation of markers pA257 and pA315 on *Vigna* LG8/9 (Fig. 2) and the putative single gene translocation of markers ARG10 and PPH on *Pi. sativum* LGVI to *Vigna* LG11 and LG1, respectively (Fig. 2). In the most extreme case, *Vigna* LG1 (Fig. 8*F*) contains seven markers that map to four different *M. truncatula* linkage groups, with two possible cases of conserved synteny (i.e., the syntenic markers PNDKN1, SUSY, and CPOX2 map to *M. truncatula* LG8, whereas the syntenic markers MMK1 and pA487 map to *M. truncatula* LG4). In contrast, broad synteny was observed among MtLG1-PsLGII-VrLG4, indicating that this linkage group might be ancestral to the galegoid and phaseoloid legumes (Fig. 8*A*).

Soybean is also a member of the phaseoloid clade but has a polyploid genome that is predicted to have undergone duplication since its divergence from other *Phaseoleae*. Sixty loci were mapped in common between *M. truncatula* and soybean, with the majority of markers derived from homologs of soybean restriction fragment length polymorphism probes identified in *M. truncatula* (10). A complicating feature of cultivated soybean is a low level of polymorphism, as evidenced in Table 1, which when taken together with the high level of soybean gene duplication significantly reduced our ability to interpret the comparative map between these two species. Nevertheless, 38% of the markers revealed putative synteny between *M. truncatula* and

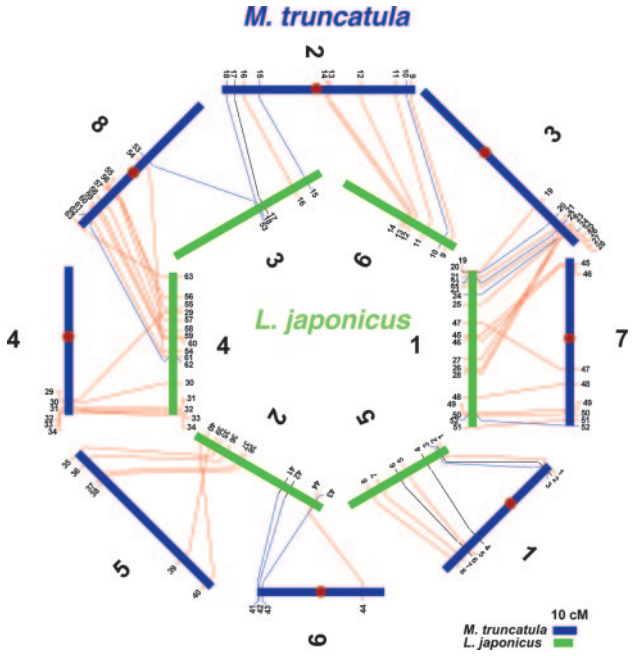


**Fig. 3.** Macrosyntenic relationship of *M. truncatula* and *L. japonicus*. Sixty-three pairs of sequenced BAC clones (Table 3), representing putatively orthologous loci with known genetic map positions, were used to construct a comparative map between *M. truncatula* and *L. japonicus*. Based on the independent selection of clones in each species, many clone pairs possess only partial overlap. Line color indicates the number of conserved genes between two clones: black, two; blue, three to four; red, five or more.

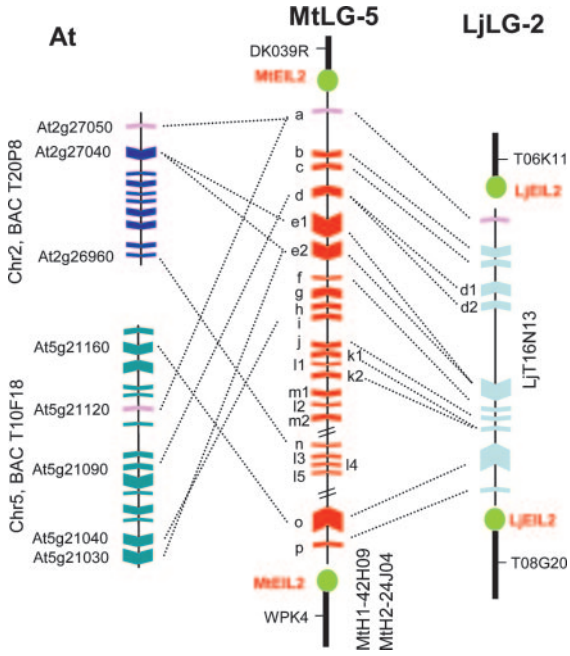


**Fig. 4.** Microsynteny between *M. truncatula* and *L. japonicus*. Microsynteny at the MtLG5 locus MtEIL2–1. Gene annotations are given in Table 4. Genes shown in pink correspond to genetic markers. Letters denote unique gene annotations, with numbers denoting tandem duplication.

soybean, identifying 11 colinear blocks between the two genomes (Figs. 2 and 8). Yan *et al.* (21, 22) report genome-wide conserved microsynteny between the genomes of *M. truncatula* and soybean, with 54% of 50 soybean contig groups showing conserved microsynteny to *M. truncatula*. Five of the extensively microsyntenic contigs (21) were mapped in *M. truncatula* in this study, three of which, namely *A095*, *A064*, and *A315*, were mapped in regions showing putative synteny between *M. truncatula* and soybean.

**Microsynteny Among Papilionoid Legume Genomes.** To determine the extent to which macrosytenic relationships identified by genetic mapping are indicative of conserved genome microstructure, we compared putatively orthologous large insert clone pairs [i.e., BAC or transformation-competent BAC (TAC) clones] between *M. truncatula* and *L. japonicus* and between *M. truncatula* and soybean. The Loteae are a sister group to the galegoid legumes (Fig. 1), and, thus, *L. japonicus* has a more recent ancestry to *M. truncatula* than to soybean. Sixty-three sequenced BAC and TAC clone pairs containing an average of nine microsyntenic gene pairs were mapped between the *M. truncatula* and *L. japonicus* genomes. As shown in Fig. 3, the genomes are highly syntenic, with macrosynteny punctuated by rearrangements that frequently involve translocation of chromosome arms (Fig. 3), reflecting the difference of six chromosomes in *L. japonicus* vs. eight chromosomes in *M. truncatula*.

Ten clone pairs with broadly spaced genetic positions in the two genomes were selected for comparison of microsynteny. Gene content was predicted by a combination of BLASTN against legume EST databases and *ab initio* prediction by using the dicot version of FGENESH. BLASTP was used for comparison among species. Counting tandem duplications as single homologs and excluding mobile DNAs, 91 and 84 genes were identified in *M. truncatula* and *L. japonicus*, respectively, with 72 (82%) conserved homologs (see Table 4 for a complete list of predicted genes). With four exceptions, all homologs were present in conserved order and transcriptional orientation. Tandem duplication increased the number of predicted genes in *L. japonicus* and *M. truncatula* by 12% and 17%, respectively, with only one example of the same homolog duplicated in both species. Moreover, of 18 transposon sequences identified in *L. japonicus* and 8 identified in *M. truncatula*, only a single example of a syntenic transposon was observed.

The example of a 141-kb region of *M. truncatula* at genetic marker *MtEIL* is shown in Fig. 4. All 16 predicted *M. truncatula* genes possess strong similarity to annotated genes in the *Arabidopsis* genome. A remarkable feature of this region is the frequent occurrence of local gene duplication, including two argonaut-like genes (*MtEIL-e1–2*), two blue-copper-binding proteins (*MtEIL-k1–2*), five kinase-like genes (*MtEIL-l1–4*), and two I-box-binding factors (*MtEIL-m1–2*). Analysis of the corresponding 97-kb segment from *L. japonicus* (*LjEIL*) revealed region-wide colinearity with the *MtEIL* contig. Ten distinct genes were identified in *LjEIL*, with only a single case of tandem duplication (*LjEIL-d1–2*). All 10 *L. japonicus* genes and a transfer $RNA^{Leu}$ had homologs in the *MtEIL* region. Six of the 16 distinct homologs from the *MtEIL* and *LjEIL* regions exhibit a network of microsynteny with two homeologous regions of *Arabidopsis* chromosomes 2 and 5, respectively (Fig. 4).

Similar analyses were conducted for two BAC clones at the *rgh1* locus of soybean and putatively orthologous BAC clones from *M. truncatula*. In total, 22 genes were identified in *M. truncatula* and 21 genes in soybean (Table 5). Fourteen homologs (63%) were conserved between the two genomes, all in the same order and transcriptional orientation. Two cases of tandem duplication were observed, one in each species, and no transposon sequences were identified within the syntenic region of either species. With one exception, all of the predicted genes possessed homology to predicted or known proteins in *Arabidopsis*, suggesting that the genes absent from the syntenic regions of *M. truncatula* or soybean are likely to be present elsewhere in

their respective genomes. This single colinear region of the legume genomes shares abbreviated stretches of microsynteny with three separate regions of the *Arabidopsis* genome, as shown in Fig. 9, which is published as supporting information on the PNAS web site.

## Discussion

The idea that conserved genome synteny can facilitate transfer of knowledge among related species of plants is best articulated in the case of the grasses (5, 6). It is increasingly clear, however, that there are many exceptions and complexities to the "rule" of conserved grass genome synteny, because even in regions of genetically defined synteny the insertion/deletion and duplication of genes can contribute to significant divergence (e.g., refs. 7 and 8). Recent studies using the genome of *Arabidopsis* as a reference document a history of genome duplication in angiosperms (23–25), followed by significant erosion of local gene content. In retrospect, it appears that genome synteny is unlikely to be well conserved beyond the taxonomic level of plant families (4). As a consequence, testing the "grass model" of genome synteny for other agronomically important families is an important objective for plant sciences.

In the case of the legume family of plants, there are numerous species with a long history of traditional breeding but with limited molecular characterization, and there are several species that are characterized at both the genetic and genomic levels. Determining the extent of synteny (and the frequency and nature of its exceptions) among these legume genomes was the focus of this study. We report that synteny is high among closely related species, and that the degree of synteny declines with increasing phylogenetic distance. Although this study is unusual in its use of explicit phylogenetic measures to assess gene orthology and its incorporation of a large genome sequence data set, the features of genome conservation and divergence that we describe are typical of those observed in comparative analysis of both plant and animal genomes.

The high level of conservation between the genomes of *M. truncatula* and *Pi. sativum* is particularly striking given the 10 times larger genome (26) and one less chromosome in *Pi. sativum*. Much of the expansion in *Pi. sativum* genome size may be due to retroelements (27), but, whatever the mechanism, it has done little to disrupt macrosyntenic relationships. *M. truncatula* LG6 could not be effectively integrated into the *Pi. sativum* genetic map. MtLG6 is interesting for several reasons: (*i*) it is the shortest and most highly heterochromatic of the chromosomes (20), (*ii*) it is underrepresented for randomly selected EST markers (5), and (*iii*) it is remarkably rich in resistance gene analogs (RGAs) (28). The inability to establish synteny in this study between MtLG6 and *Pi. sativum* is undoubtedly due to the low frequency of non-RGA EST markers in MtLG6 (10) and the fact that the genetic maps of pea (13) are not well populated by RGA markers. However, parallel analyses conducted by Kaló *et al.* (29) suggest that *M. sativa* LG6 is syntenic with several regions in the pea genome, in particular PsLGVI and PsLGVII (Fig. 7*d*). The absence of a corresponding single linkage group in pea suggests that chromosomal fission/fusion events involving *Medicago* chromosome 6 might be responsible for the reduction of chromosome number in *Pi. sativum*.

*L. japonicus* (tribe Loteae) and *M. truncatula* represent the two best-characterized legume genomes. Although there are no important crop legumes within the Loteae, the relatively recent divergence and sister-clade relationship to the galegoid legumes (Fig. 1) offers a potentially useful point of comparison to *M. truncatula*. We determined that *M. truncatula* and *L. japonicus* share a remarkably high level of conserved macrosynteny, dominated by a few large chromosome arm-size rearrangements. The availability of fully sequenced large insert clones [i.e., BACs and transformation-competent BAC (TACs)] at each of these genetically syntenic loci provided an opportunity to evaluate the correlation between genetic macrosynteny and sequence microsynteny. Conserved microsynteny was characterized by ≈80% of close homologs in the same order and transcriptional orientation, similar to values obtained between human and mouse (30) and within the range identified for the grasses (7). The current analysis also reveals significant divergence between these two legume genomes, with the insertion or deletion of individual or groups of genes accounting for ≈20% divergence of gene content in microsyntenic intervals. Species-specific tandem duplication of genes accounted for an additional 12–17% divergence of gene content, and each species possessed a unique distribution of mobile DNAs. Of 21 tandemly duplicated genes, only one duplication was reciprocal. Similarly, of 26 cases of mobile DNAs, only one mobile DNA was potentially syntenic. The lack of ancestral tandem duplication is suggestive of either efficient removal of tandem duplicates that predate speciation or a recent increase in the rate of tandem duplication. Moreover, the observation that tandemly duplicated genes are occasionally interspersed with single copy genes (Table 4) suggests that purification of duplicates by homologous recombination would simultaneously eliminate the intervening single copy gene(s). Such a mechanism could explain, at least in part, the loss of gene homologs from microsyntenic regions.

Syntenic relationships were significantly more convoluted between the galegoid and phaseoloid clades. Twenty-five percent of genetically mapped orthologous genes were potentially nonsyntenic, resulting in smaller regions of colinearity than those observed between *M. truncatula* and *Pi. sativum* or between *M. truncatula* and *L. japonicus*. This fragmentation of synteny might be expected based solely on the differences in chromosome number between these two clades. Despite the relatively large number of genetic markers used for comparison, synteny between *M. truncatula* and soybean was difficult to characterize. We attribute this situation to a combination of recent duplication and low rates of polymorphism in the soybean genome. Nevertheless, comparison of putatively orthologous BAC clone pairs

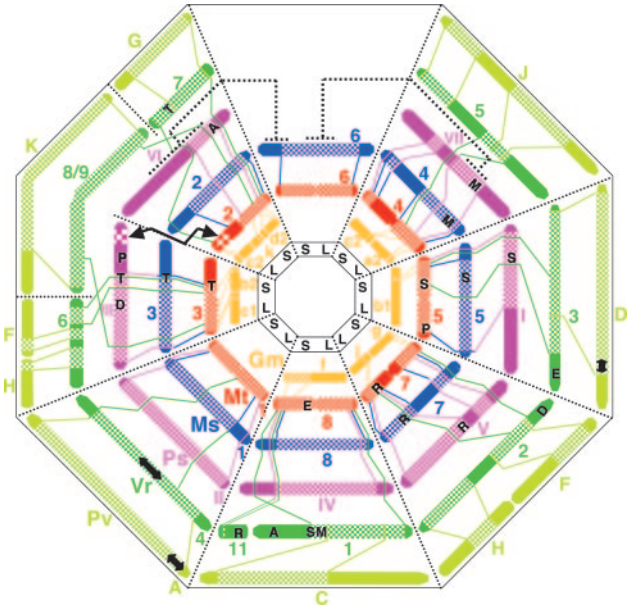


**Fig. 5.** Consensus comparative map data for six legume species. Species abbreviations are as in Fig. 2. S and L denote short and long arms of each chromosome in *M. truncatula* (20). Synteny blocks are drawn to scale based on genetic distance. Solid lines, postulated rearrangements; double-headed arrows, postulated inversions. Gene-specific markers that disrupt synteny are S-*SHMT*, R-*RNAH*, T-*TRPT*, M-*MMK1*, P-*PTSB*, A-*ARG10*, D-*6DCS*, and E-*REP*.

revealed significant conservation of microsynteny between *M. truncatula* and soybean, consistent with previous comparison of other genome regions between these two species (22, 31). In all cases, conservation of microsynteny was significantly greater between legume genomes than between legume genomes and the corresponding regions of the *Arabidopsis* genome (Figs. 4 and 9).

The genetic mapping of orthologous genes across multiple species provides an opportunity to propose an integrated view of legume synteny. As shown in Fig. 5, the results suggest broad macrosynteny, particularly within the galegoid or phaseoloid legumes, punctuated by chromosomal rearrangements that increase in frequency with phylogenetic distance. The inclusion of phylogenetic measures in marker analysis also aided the inference of genomic rearrangements. For example, in *M. truncatula*, orthologous marker *PTSB* maps to LG5, which is highly conserved with PsLGI (Fig. 2). In *Pi. sativum*, *PTSB* maps to a nonsyntenic region of PsLGIII, near the point of an inferred chromosomal rearrangement. The combination of *PTSB* orthology to a marker on MtLG5 and disrupted synteny of the flanking markers are consistent with a complex genome rearrangement involving both chromosomal fragment translocation and single gene translocation. Such events are likely to be significantly more frequent between the galegoid and phaseoloid legumes, because several markers that map to conserved regions in the galegoid legumes occur in nonsyntenic regions of *V. radiata* (Figs. 2, 5, and 8). This result is consistent with significant genomic changes that must underlie differences in chromosome number between galegoid (typically eight chromosomes) and phaseoloid (typically 11 chromosomes) lineages.

## Conclusion

Several studies suggest that the practical utility of comparative mapping in plants may be limited to within-family comparisons (4). Factors contributing to this situation are chromosomal rearrangements that result in the progressive fractionation of the genome into increasingly smaller conserved segments and the high frequency of gene loss from within-genome segmental duplications (e.g., refs. 23, 24, 32, 33). The present analysis suggests that the same factors that contribute to divergence between other plant families are also operative within the legumes.

Although the current study documents substantial conservation between the genomes of crop and model legumes, it also reveals features of genome divergence. The degree to which genome synteny can facilitate cross-species analysis of gene function will depend both on the conservation of gene order and content, as well as on the frequency with which similar traits have a common genetic basis in different species. An indication that this latter criterion might not always be met is suggested by a recent study of branching in foxtail millet (34). By contrast, similar studies of symbiotic nitrogen fixation in legumes (reviewed in ref. 3) demonstrate that functionally conserved genes occupy syntenic positions across the diversity of legume species analyzed in this study. Moreover, even large and rapidly evolving gene families, such as the nucleotide-binding site-leucine-rich repeat resistance gene homologs, can occupy ancestral genome locations among legumes (28). Testing the extent to which inferences made from comparative genomics can be translated to practical applications in crop improvement represents one of the major current challenges facing plant biology.

This work was supported by National Science Foundation Grant 0110206 (to D.R.C., N.D.Y., and D.-J.K.); the Samuel Roberts Noble Foundation (to B.R.); the European Union MEDICAGO project (QLG2-CT-2000-30676) and the Biotechnology and Biological Sciences Research Council (to N.E.); and Nemzeti Kutatás es Feilesztesi Program and the (HNDRP) [Országos Tudományos Kutatási Alap Programok (OTKA) and T038211, to G.B.K.].

1. Graham, P. H. & Vance, C. P. (2003) *Plant Physiol.* **131,** 872–877.
2. Doyle, J. J. & Luckow, M. A. (2003) *Plant Physiol.* **131,** 900–910.
3. Riely, B., Ane, J.-M., Penmetsa, R. V. & Cook, D. R. (2004) *Curr. Opin. Plant Biol.* (2004) **7,** 408–413.
4. Delseny, M. (2004) *Curr. Opin. Plant Biol.* **7,** 126–131.
5. Devos, K. M. & Gale, M. D. (2000) *Plant Cell* **12,** 637–646.
6. Bennetzen, J. L. & Freeling, M. (1997) *Genome Res.* **7,** 301–306.
7. Bennetzen, J. L. & Ramakrishna, W. (2002) *Plant Mol. Biol.* **48,** 821–817.
8. Song, R., Llaca, V. & Messing, J. (2002) *Genome Res.* **12,** 1549–1555.
9. Fu, H. & Dooner, H. K. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 9573–9578.
10. Choi, H. K., Kim, D., Uhm, T., Limpens, E., Lim, H., Mun, J. H., Kalo, P., Penmetsa, R. V., Seres, A., Kulikova, O., *et al.* (2004) *Genetics* **166,** 1463–1502.
11. Vallejos, C. E., Sakiyama, N. S. & Chase, C. D. (1992) *Genetics* **131,** 733–740.
12. Boutin, S. R., Young, N. D., Olson, T. C., Yu, Z. H., Shoemaker, R. C. & Vallejos, C. E. (1995) *Genome* **38,** 928–937.
13. Ellis, T. H. N. & Poyser, S. J. (2002) *New Phytol.* **153,** 17–25.
14. Humphry, E., Konduri, V., Lambrides, J., Magner, T., McIntyre, L., Aitken, B. & Liu, J. (2002) *Theor. Appl. Genet.* **105,** 160–166.
15. Lambrides, C. J., Lawn, R. J., Godwin, I. D., Manners, J. & Imrie, B. C. (2000) *Aust. J. Agric. Res.* **51,** 415–425.
16. Vision, T. J., Brown, D. G., Shmoys, D. B., Durrett, R. T. & Tanksley, S. D. (2000) *Genetics* **155,** 407–420.
17. Swofford, D. L. (2003) PAUP*, Phylogenetic Analyses Using Parsimony (*and Other Methods) (Sinauer, Sunderland, MA), Version 4.0b10.
18. Sato, S., Kaneko, T., Nakamura, Y., Asamizu, E., Kato, T. & Tabata, S. (2001) *DNA Res.* **8,** 311–318.
19. Wojciechowski, M. F. (2004) *Advances in Legume Systematics, Part 10: Higher Level Systematics*, eds. Klitgaard, B. B. & Bruneau, A. (Royal Botanic Gardens, Kew, U.K.), pp. 5–35.
20. Kulikova, O., Gualtieri, G., Geurts, R., Kim, D., Cook, D. R., Huguet, T., de Jong, J. H., Fransz, P. F. & Bisseling, T. (2001) *Plant J.* **27,** 49–58.
21. Yan, H., Mudge, J., Kim, D., Shoemaker, R. C., Cook, D. R. & Young, N. D. (2003) *Theor. Appl. Genet.* **106,** 1256–1265.
22. Yan, H. H., Mudge, J., Kim, D., Larsen, D., Denny, R., Shoemaker, R. C., Cook, D. R. & Young, N. D. (2004) *Genome* **47,** 141–155.
23. Bowers, J. E., Chapman, B. A., Rong, J. & Paterson, A. H. (2003) *Nature* **422,** 233–438.
24. Ku, H. M., Vision, T., Liu, J. & Tanksley, S. D. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 9121–9126.
25. Vision, T. J., Brown, D. G. & Tanksley, S. D. (2000) *Science* **290,** 2114–2117.
26. Greilhuber, J. & Ebert, I. (1994) *Genome* **37,** 646–655.
27. Vershinin, A. V., Allnutt, T. R., Knox, M. R., Ambrose, M. J. & Ellis, T. H. N. (2003) *Mol. Biol. Evol.* **20,** 2067–2075.
28. Zhu, H. Y., Cannon, S., Young, N. D. & Cook, D. R. (2002) *Mol. Plant–Microbe Interact.* **15,** 529–539.
29. Kaló, P., Seres, A., Taylor, S. A., Jakab, J., Kevei, Z., Kereszt, A., Endre, G., Ellis, T. H. N. & Kiss, G. B. (2004) *Mol. Genet. Genomics*, in press.
30. Mural, R. J., Adams, M. D., Myers, E. W., Smith, H. O., Miklos, G. L. G., Wides, R., Halpern, A., Li, P. W., Sutton, G. G., Nadeau, J., *et al.* (2002) *Science* **296,** 1661–1671.
31. Cannon, S. B., McCombie, W. R., Sato, S., Tabata, S., Denny, R., Palmer, L., Katari, M. & Young, N. D. (2003) *Mol. Genet. Genomics* **270,** 347–361.
32. Zhu, H. Y., Kim, D., Baek, J. M., Choi, H. K., Ellis, L., Kuester, H., McCombie, W. R., Peng, H. M. & Cook, D. R. (2003) *Plant Physiol.* **131,** 1018–1026.
33. Vandepoele, K., Simillion, C. & Van de Peer, Y. (2002) *Trends Genet.* **18,** 606–608.
34. Doust, A. N., Devos, K. M., Gadberry, M. D., Gale, M. D. & Kellog, E. A. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 9045–9050.