

Legume genomes: more than peas in a pod

Nevin Dale Young^{*†}, Joann Mudge^{*} and TH Noel Ellis[‡]

A growing array of sequence-based tools is helping to reveal the organization, evolution and syntenic relationships of legume genomes. The results indicate that legumes form a coherent taxonomic group with frequent and widespread macro- and microsynteny. This is good news for two model legume systems, *Medicago truncatula* and *Lotus japonicus*. Indeed, both models have recently been used to clone and characterize genes for nodulation-related receptors that were originally described in legumes with more complex genomes. Studies of legume genomes have also provided insight into genome size, gene clustering, genome duplications and repetitive elements. To understand legume genomes better, it will be necessary to develop tools for studying under-represented taxa beyond the relatively small group of economically important species that have been examined so far.

Addresses

^{*}Department of Plant Pathology and Department of Plant Biology, 495 Borlaug Hall, University of Minnesota, St. Paul, Minnesota 55108, USA
†e-mail: nevin@umn.edu

[‡]Department of Crop Genetics, John Innes Centre, Colney Lane, Norwich NR4 7UH, UK
e-mail: noel.ellis@bbsrc.ac.uk

Current Opinion in Plant Biology 2003, 6:199–204

1369-5266/03/\$ – see front matter
© 2003 Elsevier Science Ltd. All rights reserved.

DOI 10.1016/S1369-5266(03)00006-2

Abbreviations

BAC	bacterial artificial chromosome
cM	centiMorgan
kbp	kilobase pairs
LG-G	linkage group-G
Mbp	million base pairs
RFLP	restriction fragment length polymorphism
SSR	simple sequence repeat

Introduction

The legumes are a diverse and important family of angiosperms. With more than 650 genera and 18 000 species, legumes are the third largest family of higher plants and are second only to grasses in agricultural importance (reviewed in [1]). Legumes range from tiny herbs to giant trees, dominating many tropical rainforests. They form symbiotic relationships with members of the *Rhizobium* family of bacteria [2], together producing enormous quantities of biological nitrogen. Thus, legumes are a major source of organic fertilizer and an essential component of most agricultural rotations. Legumes provide the largest single source of vegetable

protein in human diets and livestock feed. In addition to their ability to form associations with *Rhizobia*, legumes also form symbiotic relationships with mycorrhizal fungi (as do many other plant families). Studies on legumes increasingly point to common symbiosis pathways for both *Rhizobium* and mycorrhizae [3]. Legumes are also producers of beneficial secondary compounds, many of which have been suggested to have health-promoting properties such as providing protection against human cancers [4]. Not surprisingly, legumes are one of the best-studied plant families in the world.

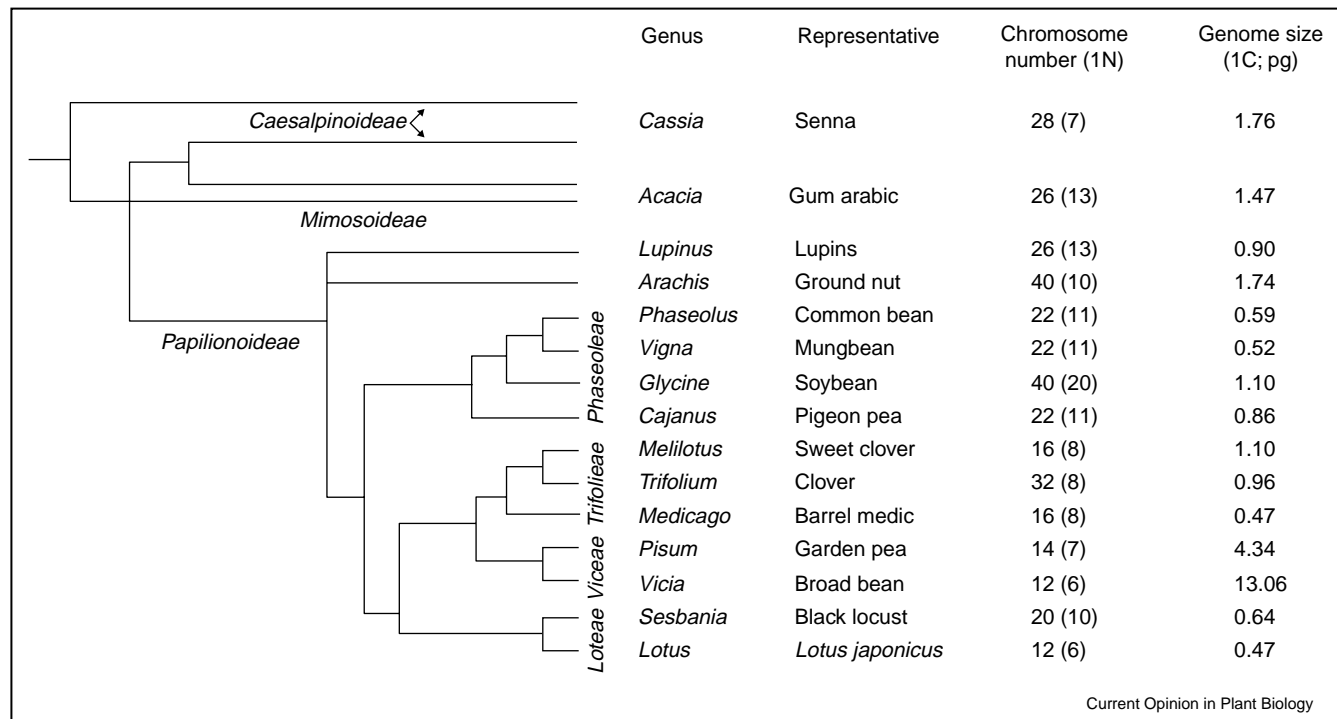
In this commentary, we focus on recent studies of structural, comparative and functional genomics. These studies reveal impressive levels of macro- and microsynteny among legume genomes, and even with the model plant *Arabidopsis thaliana* [5–9]. Genomic research is also adding to our knowledge of the duplicated nature of legume genomes, especially that of soybean, and providing persuasive evidence for non-homogeneous organization of these genomes involving gene-rich islands [7,8,10].

Legume phylogeny and molecular evolution

Classically the legumes are divided into three subfamilies, Mimosoideae, Caesalpinoideae, and Papilionoideae, although recent results indicate that the Caesalpinoideae is actually paraphyletic (Figure 1; reviewed in detail in [1]). Most cultivated legumes are found within the Papilionoideae, the subfamily with largest total number of genera. There are two major groups of cultivated species in the Papilionoideae: the so-called ‘tropical’ or ‘phaseoloid’ legumes (including the genera *Phaseolus*, *Vigna*, *Glycine* and *Cajanus*) and the ‘temperate’ or ‘galegoid’ legumes (including the genera *Melilotus*, *Trifolium*, *Medicago*, *Pisum*, *Vicia*, *Lotus*, *Cicer* and *Lens*). Of course, other legume genera are important to people — forest and lumber trees in Mimosoideae, such as *Acacia* and *Albizia*; other forage crops and shade trees in the Papilionoideae, such as *Robinia* (locust bean); and finally, important ornamental, timber and even food plants in Caesalpinoideae, such as *Poinciana* (flame tree) and *Tamarindus* (the source of tamarind fruit).

Molecular phylogeny, primarily based on sequence data for chloroplast *rbcL* (ribulose biphosphate carboxylase), a gene widely used in studies of plant phylogeny [11,12], has clarified the relationships among many legume lineages but has also raised new questions. The most basal dichotomy involves the tribe Cercideae (which includes the ornamental ‘Judas tree’ *Cercis siliquastrum*). It is clear that the Caesalpinoideae are basal and so are expected to contain much of the genetic and evolutionary

Figure 1



A phylogeny of legumes, featuring the three major subfamilies and details about selected crop species in the Papilionoideae. Estimates of chromosome number and genome size come from the Angiosperm C-value Database (<http://www.rbgekew.org.uk/cval/homepage.html>, Release 1.0).

‘diversity’ present in the entire legume family. By contrast, cultivated legume crops are generally within taxa that have diverged relatively recently. Thus, the cultivated genera, which have been studied most often, are not a good representation of legumes generally or even of the Papilionoideae subfamily.

Molecular phylogenies provide a context for posing hypotheses that can be tested using genomic techniques. For example, what are the mechanisms underlying the origin (and loss) of zygomorphic flowers, compound leaves, or nodulation (which is suggested to have arisen three separate times) [11,12] in legume evolution? If losses and gains have occurred independently and several times in legume evolution, were the same proteins involved each time? Knowledge of legume evolution also helps to formulate questions about what has happened to the genes and gene families that were previously associated with nodulation. Basal taxa, such as non-nodulating Caesalpinoideae lineages (which are capable of mycorrhizal symbiosis), can provide a useful ‘out-group’ for molecular and cellular studies that compare rhizobial and mycorrhizal symbioses. Indeed, most aspects of legume biology, from ploidy number to floral diversity, can be examined from the perspective of evolutionary relationships.

Organization of legume genomes

The nuclear genomes of legumes vary greatly in size, from 370 million base pairs (Mbp) in *Lablab niger* to the enormous genome of *Vicia faba* at more than 13 000 Mbp (<http://www.rbgekew.org.uk/cval/homepage.html>; Figure 1). Generally speaking, legume genomes tend to be intermediate in size compared with those of other higher plants. More than 50% of the known legume genome sizes listed on the comprehensive plant C-value website are smaller than 1300 Mbp. Most of the cultivated species that are familiar to researchers are modest in genome size (Figure 1); mung bean, cowpea, common bean, chick pea, and clover all have haploid genomes smaller than 1000 Mbp. The model legumes, *Lotus japonicus* and *Medicago truncatula*, both have compact genomes of approximately 470 Mbp (<http://www.rbgekew.org.uk/cval/homepage.html>).

Despite the modest genome size of most legumes, it is striking that some genera (notably *Lathyrus* and *Vicia*) have genome sizes that vary by a factor of 10 or more. This is associated with extensive differences in the abundance of retroelements, which account for substantial proportions of these genomes [13]. Much of the genome content in these taxa appears to be of recent origin, and this is consistent with the observation that genomes that differ

greatly in size (e.g. pea and alfalfa) still have essentially collinear genomes. From a practical perspective, the recent amplification of retroelements in these genomes is consistent with the utility of retroelements as polymorphic markers [14].

Polymorphism that is associated with retrotransposon insertion sites in pea appears to predate the divergence of the genus, implying that this largely inbreeding species has had extensive recombination between ancient lineages [14]. Similar observations have been noted in maize, in which polymorphism that is associated with nested transposon insertions has been described [15]. This haplotype variation also included the loss of several genes, and the presence of null alleles in pea has likewise been described [16]. Thus, the emerging picture for the organization of the maize genome might also be a good structural model for large genome legumes.

The abundance of retroelements in large legume genomes is also associated with their extensive C-methylation, and differences in C-methylation between lineages can be heritable [17]. The potential for retroelements, as mobile promoters, to alter patterns of gene expression is well established, as is their potential for the generation of new functions [18].

Genome duplications in legumes

There are mechanisms besides retroelements that generate multicopy sequences in plant genomes. Recent studies of *A. thaliana* highlight the crucial role of whole-genome and large-scale segmental duplications in the evolution of higher plants [19,20]. Large-scale genome duplication in *Zea mays* has also been demonstrated and even tested statistically [21]. Indeed, large-scale genome duplications are clearly a common feature in eukaryotic genome evolution; evidence for duplications is growing in organisms from yeast to human [22,23]. From a functional point of view, large-scale genome duplications create the potential for new diversity and the possibility of subfunctionalization [24]. A review outlining some of the exciting progress in the area of genome duplication was published in a recent issue of *Current Opinions in Genetics and Development* [25].

Genome duplication has been examined in detail in soybean, in which some of the first molecular research on this subject was carried out [26]. Soybean has a complex genome with a 'paleopolyploid' evolutionary past. Early studies of genome duplication in soybean utilized multicopy (but still low copy) restriction fragment length polymorphism (RFLP) markers to infer the existence of syntenic and/or collinear runs of markers on two or more linkage groups. The results demonstrated that the soybean genome consists of many homoeologous segments, some of which are present as many as six times. Frequent cases of nested duplications suggested multiple rounds of duplication.

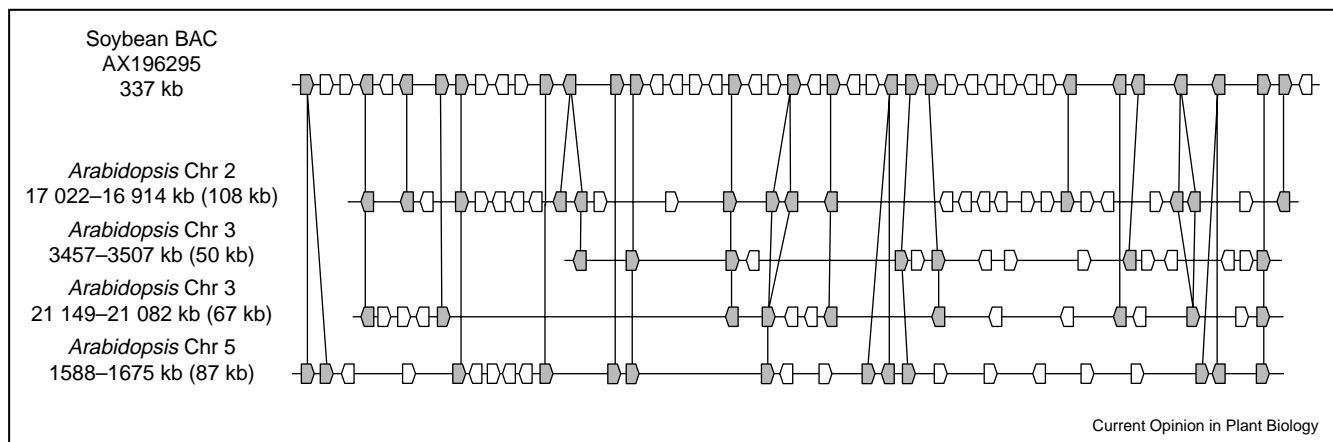
One area of recent interest has been the level of sequence similarity and microsynteny between duplicated segments. In a study of ten bacterial artificial chromosome (BAC) contigs that span a total of 1 Mbp spread throughout a 10-centiMorgan (cM) segment of soybean linkage group-G (LG-G), Foster-Hartnett *et al.* [8] observed that homoeologous BAC contigs, located on unlinked chromosomes, exhibited remarkably similar restriction fragment patterns. Homoeologous BACs also demonstrated 98% sequence identity in several short segments that were compared in detail. In a genome-wide survey of soybean, Yan *et al.* [9] determined that 67% of homoeologous sets of BAC contigs had at least one pair with extensive microsynteny. More than 90% of homoeologous BAC-contig groups in soybean showed at least some level of microsynteny. This high degree of microsynteny suggests that at least one relatively recent large-scale duplication event has occurred in soybean.

Gene clustering in legume genomes

The distribution of repetitive versus low-copy sequences has important implications for genome organization and, by extension, for the feasibility of sequencing gene-rich regions. Using 20 previously mapped gene-rich BACs, Kulikova *et al.* [10] compared genetic maps of *M. truncatula* with the results of fluorescence *in situ* hybridization (FISH) analysis of pachytene chromosomes. The results demonstrated that the *M. truncatula* genome is organized into distinct regions: pericentromeric heterochromatin, which is rich in repeated sequences, is separate from extensive gene-rich euchromatic regions. Subsequent BAC sequencing has suggested that *M. truncatula* gene density is approximately one gene every 6–10 kilobase pairs (kbp) (D Cook, DJ Kim, personal communication), on par with that of *Arabidopsis*. Together, these lines of evidence suggest that just 20–40% of the *M. truncatula* genome contains the majority of its genes. Although these estimates are still too imprecise to allow us to predict the number of genes in *M. truncatula*, they certainly suggest that the total gene number will be comparable to the more than 29 000 previously described in *Arabidopsis* ([19]; C Town, personal communication).

The gene density in *L. japonicus* has been estimated as one gene per nine kbp [27] on the basis of the sequencing of 121 gene-containing transformation-competent artificial chromosome (TAC) clones. Similarly, comparisons of hypomethylated RFLPs (as signatures of genic regions) and simple sequence repeats (SSRs) revealed gene-rich regions in soybean [7]. Of more than 2000 BAC-end sequences examined, the RFLP-associated sequences had only half as many repetitive sequences and 50% more genic sequences than the SSR-associated sequences. After studying a segment of soybean sequence of more than 330 kbp in length (Figure 2), Foster-Hartnett *et al.* [8] estimated gene density to be as high as one gene per five kbp (although revised calculations put this value closer

Figure 2



Microsynteny between a 337 kb region of soybean on LG-G and the corresponding genome regions of *Arabidopsis thaliana*. The soybean sequence was downloaded from GenBank (<http://www.ncbi.nlm.nih.gov/GenBank/index.html>) accession AX196295 and annotated using the *Arabidopsis*-trained version of Genscan (<http://genes.mit.edu/GENSCAN.html>). 1-kb sections of the soybean DNA were compared to 1-kb sections of the 17 September, 2002 version of *Arabidopsis* pseudochromosome sequences (<http://mips.gsf.de/>) using TBLASTX and an expectation value of $1E-6$. Only the top four hits for each soybean segment were considered. Shaded arrows represent genes that have a microsyntenic homologue in both species. White arrows represent genes that do not have homologues in *Arabidopsis* or genes whose homologues were in other regions of the *Arabidopsis* genome. Genes and spacing between genes are not drawn to scale, but gene order and orientation are shown correctly.

to one gene per eight kbp). All of these estimates suggest an uneven gene density, with some regions containing higher densities than could be expected if genes were evenly distributed throughout these legume genomes. Nonetheless, gene clustering in legumes may be less extreme than has been described for cereals [28–30].

Early comparative genomic studies in legumes

Vavilov was an early proponent of the idea that related organisms share related gene functions. In his seminal paper [31], this proposition was illustrated by his comparison of *Vicia* and *Lens*. This work presaged the isozyme- and RFLP-based genetic map comparisons of the 1990s, which demonstrated impressive levels of macrosynteny among members of a given taxonomic family. For example, Weeden *et al.* [32] compared the molecular marker maps of lentil (*Lens*) and pea (*Pisum*) and found at least eight conserved genome regions, constituting 40% of the lentil genome. These studies have now been reinforced by the availability of sequence-tagged site (STS) markers [33]. Later work, comparing chickpea (*Cicer*) to lentil and pea, also uncovered substantial genome conservation although at lower levels [34], an observation that is in line with the predicted evolutionary relationships among these genera. Menancio-Hautea *et al.* [35] performed a genome comparison within the phaseoloid group and found multiple linkage groups that were conserved between mung bean (*Vigna radiata*) and cowpea (*V. unguiculata*), as well as several smaller segments of collinearity. Later, in an extensive comparison of *Vigna*,

Phaseolus, and *Glycine*, Boutin *et al.* [36] showed that the *Phaseolus vulgaris* RFLP map could largely be superimposed on the *V. radiata* map, with the average length of collinearity extending 37 cM and in one case more than 100 cM. By contrast, conserved segments between either *V. radiata* and *G. max* or *P. vulgaris* and *G. max* were generally much shorter, averaging just 12–13 cM. Recently, levels of genome conservation among these taxa have been reevaluated, and higher levels of genome conservation with *G. max* have been reported [6].

An ambitious project to extend our understanding of legume macrosynteny by creating a comparative map for the entire legume family is underway (D Cook, DJ Kim, personal communication). By strategically choosing primers in highly conserved gene regions that also flank introns, a large set of low copy, dependable, and reasonably polymorphic primer pairs can be designed and used for linkage mapping in multiple taxa (http://www.intl-pag.org/pag/10/abstracts/PAGX_W100.html). It is hoped that, eventually, a very large set of such markers will be created and used to establish a legume-wide framework for comparative genomic studies. As part of this effort to compare legume genomes, however, more rigorous methods should be adopted to quantify synteny and collinearity [37]. Efforts should also be made to extend comparative studies beyond the narrow subset of Papilionoid genera that have been examined most frequently to date. Initial results suggest that at least some of these markers function in all three legume subfamilies (D Cook, personal communication).

Recent discoveries about legume comparative genomics

From these early studies on macrosynteny, it was clear that legumes resemble other plant families — including Gramineae, Solanaceae, and Brassicaceae — in terms of within-family genome conservation. More recently, the field of comparative genomics has undergone what is essentially a second ‘revolution’ with a new emphasis on microsynteny and microcollinearity. Much of this revolution has been fueled by the complete genome sequence of *Arabidopsis* [19].

One of the first reports of inter-family comparative mapping involved soybean and *Arabidopsis* [5]. Surprisingly, cases of substantial macrosynteny were observed between these two species, although it is likely that macrosynteny between legumes and *Arabidopsis* is probably the exception rather than the rule [9]. Still, the work of Grant *et al.* [5] formed the basis for a more recent comparison of soybean with *V. radiata* and *P. vulgaris* [6]. To uncover higher levels of macrosynteny, it was necessary to consider the duplicated structure of the soybean genome and also to use *Arabidopsis* as a ‘bridging species’. In effect, homoeologous regions of the soybean genome were first compared among themselves, and then to corresponding regions in both *P. vulgaris* and *V. radiata*. In this way, a network of macrosynteny among the legume taxa could be created and then compared to the *Arabidopsis* genome.

Several examples of microsynteny between legumes and *Arabidopsis* have been reported (Figure 2). In one study, Foster-Hartnett *et al.* [8] examined a 10-cM segment of soybean LG-G in which extensive microsynteny with *Arabidopsis* was observed over the entire region. The genome segments in *Arabidopsis* that showed microsynteny with soybean were present as multicopies, with only some of the conserved sequences found in any one of the *Arabidopsis* duplications. These results suggest that regions of microsynteny can be extensive; in this case, extending several cM in soybean and corresponding to segments of up to 2 Mbp in *Arabidopsis*. Other examples of microsynteny between legumes and *Arabidopsis* have also been reported [7,38]. Nevertheless, quantitative comparisons between soybean and *Arabidopsis* indicate that cases of conserved microsynteny may be relatively infrequent. Comparison of more than 1000 soybean BAC-end sequences that had previously been placed into anchored contig groups with the published genome sequence of *Arabidopsis* revealed that only 14% of the sequences showed microsynteny between soybean and *Arabidopsis* [9].

Microsynteny appears to be widespread within legumes. On the basis of the 1000 anchored soybean BAC-end sequences described above, Yan *et al.* [9] found that more than half of soybean BAC contig groups exhibit microsynteny with *M. truncatula*. Of these, more than 80%

showed extensive microsynteny. Microsynteny among legumes also provided the basis for two recent highly significant success stories in gene cloning. Legume microsynteny was essential in the cloning of *NN1* from *Medicago sativa* and *SYM2* from *Lotus japonicus*, along with their respective orthologues *DMI2* in *Medicago truncatula* and *Sym19* in pea [38,39]. Sequence analysis indicates that these related genes are extracellular receptor kinases that have crucial roles in Nod-factor perception. Along with *Nin* [40], these genes represent the first of what are likely to be many nodulation-related genes to be cloned through the use of model legumes. In a similar effort to clone the pea *SYM2* gene (an entirely different gene from the *Lotus japonicus* *SYM2*), Gualtieri *et al.* [41] used a tightly linked pea marker to identify the corresponding region in *M. truncatula*. From this starting point, a fine-structure physical map for *M. truncatula* was constructed, yielding eight DNA markers that could be mapped and oriented by linkage mapping in pea. The results suggest that all eight markers are conserved, tightly linked and collinear.

Conclusions

Many of the tools and resources needed for systematic gene discovery and sequence-based genome analysis are now available in legumes. This has been driven by the economic imperatives of crop legumes coupled with the research community’s interest in the mechanism of symbiotic nitrogen fixation (an interest that initially spawned *M. truncatula* and *L. japonicus* as models). The early indications, reviewed here, suggest that legumes are a coherent taxonomic group with substantial genome conservation. Thus, it should be possible to extend legume genomic resources across taxonomic boundaries to address a wide range of fundamental and applied problems. Moreover, genes seem to be clustered in some (perhaps most) legume taxa. Potentially, this will reduce the genome fraction that needs to be sequenced to discover most genes and their flanking regions.

Acknowledgements

We thank Steve Cannon and Graham Moore for their helpful suggestions on the manuscript. NDY and JM acknowledge the support of National Science Foundation grants DBI-0196179 and DBI-0110206. THNE acknowledges the support of EU projects TEGERM (QLK-CT-2000-01502) and MEDICAGO (QLG2-CT-2000-30676).

References

1. Doyle JJ: **Leguminosae**. In *Encyclopedia of Genetics*. Edited by Brenner S, Miller JH. San Diego: Academic Press; 2001:1081-1085.
2. Schultze M, Kondorosi A: **Regulation of symbiotic root nodule development**. *Annu Rev Genet* 1998, **32**:33-57.
3. Vanrhijn P, Fang Y, Galili S, Shaul O, Atzmon N, Wininger S, Eshed Y, Lum M, Li Y, To V *et al.*: **Expression of early nodulin genes in alfalfa mycorrhizae indicates that signal transduction pathways used in forming arbuscular mycorrhizae and rhizobium-induced nodules may be conserved**. *Proc Natl Acad Sci USA* 1997, **94**:5467-5472.
4. Stark A, Madar Z: **Phytoestrogens: a review of recent findings**. *J Ped Endocrin Metab* 2002, **15**:561-572.

5. Grant D, Cregan P, Shoemaker RC: **Genome organization in dicots: genome duplication in *Arabidopsis* and synteny between soybean and *Arabidopsis*.** *Proc Natl Acad Sci USA* 2000, **97**:4168-4173.
6. Lee JM, Grant D, Vallejos CE, Shoemaker RC: **Genome organization in dicots. II. *Arabidopsis* as a 'bridging species' to resolve genome evolution events among legumes.** *Theor Appl Genet* 2001, **103**:765-773.
7. Marek LF, Mudge J, Darnielle L, Grant D, Hanson N, Paz M, Huihuang Y, Denny R, Larson K, Foster-Hartnett D *et al.*: **Soybean genomic survey: BAC-end sequences near RFLP and SSR markers.** *Genome* 2001, **44**:472-581.
8. Foster-Hartnett D, Mudge J, Danesh D, Yan H, Larsen D, Denny R, Young ND: **Comparative genomic analysis of sequences sampled from a small region on soybean molecular linkage group 'G'.** *Genome* 2002, **45**:634-645.
9. Yan H, Mudge J, Kim DJ, Shoemaker RC, Cook DR, Young ND: **Estimates of conserved microsynteny among the genomes of *Glycine max*, *Medicago truncatula* and *Arabidopsis thaliana*.** *Theor Appl Genet*, in press.
10. Kulikova O, Gualtieri G, Geurts R, Kim DJ, Cook D, Huguet T, de Jong JH, Fransz PF, Bisseling T: **Integration of the FISH pachytene and genetic maps of *Medicago truncatula*.** *Plant J* 2001, **27**:49-58.
11. Doyle JJ, Doyle JL, Ballenger JA, Dickson EE, Kajita T, Ohashi H: **A phylogeny of the chloroplast gene *rbcl* in the Leguminosae – taxonomic correlations and insights into the evolution of modulation.** *Am J Bot* 1997, **84**:541-554.
12. Kajita T, Ohashi H, Tateishi Y, Bailey CD, Doyle JJ: ***rbcl* and legume phylogeny, with particular reference to Phaseoleae, Millettieae, and allies.** *Syst Bot* 2001, **26**:515-536.
13. Pearce SR, Harrison G, Li D, Heslop-Harrison JS, Kumar A, Flavell AJ: **The *Ty1-copia* group retrotransposons in *Vicia* species: copy number, sequence heterogeneity and chromosomal location.** *Mol Gen Genet* 1996, **250**:305-315.
14. Pearce SR, Knox MR, Ellis THN, Flavell AJ, Kumar A: **Pea *Ty1-copia* group retrotransposons: transpositional activity and use as markers to study genetic diversity in *Pisum*.** *Mol Gen Genet* 2000, **263**:898-907.
15. Fu H, Dooner HK: **Intraspecific violation of genetic colinearity and its implications in maize.** *Proc Natl Acad Sci USA* 2002, **99**:9573-9578.
16. Ellis THN, Turner L, Hellens RP, Lee D, Harker CL, Enard C, Domoney C, Davies DR: **Linkage maps in pea.** *Genetics* 1992, **130**:649-663.
17. Knox MR, Ellis THN: **Stability and inheritance of methylation states at *PstI* sites in *Pisum*.** *Mol Gen Genet* 2001, **265**:497-507.
18. Nekrutenko A, Li W-H: **Transposable elements are found in a large number of human protein-coding genes.** *Trends Genet* 2001, **17**:619-621.
19. *Arabidopsis* Genome Initiative: **Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*.** *Nature* 2000, **408**:796-815.
20. Vision TJ, Brown DG, Tanksley SD: **The origins of genomic duplications in *Arabidopsis*.** *Science* 2000, **290**:2114-2117.
21. Gaut BS: **Patterns of chromosomal duplication in maize and their implications for comparative maps of the grasses.** *Genome Res* 2001, **11**:55-66.
22. Lynch M, Conery JS: **The evolutionary fate and consequences of duplicate genes.** *Science* 2000, **290**:1151-1155.
23. Li WH, Gu Z, Wang H, Nekrutenko A: **Evolutionary analyses of the human genome.** *Nature* 2001, **409**:847-849.
24. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J: **Preservation of duplicate genes by complementary, degenerative mutations.** *Genetics* 1999, **151**:1531-1545.
25. Sankov D: **Gene and genome duplication.** *Curr Opin Genet Dev* 2001, **11**:681-684.
26. Shoemaker RC, Polzin K, Labate J, Specht J, Brummer EC, Olson T, Young ND, Concibido V, Wilcox J, Tamulonis JP *et al.*: **Genome duplication in soybean (*Glycine* subgenus *soja*).** *Genetics* 1996, **144**:329-338.
27. Nakamura Y, Kaneki T, Asamizu E, Kato T, Sato S, Tabata S: **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* 2002, **9**:63-70.
28. Sandhu D, Champoux JA, Bondareva SN, Gill KS: **Identification and physical localization of useful genes and markers to a major gene-rich region on wheat group 1S chromosomes.** *Genetics* 2001, **157**:1735-1747.
29. Wicker T, Stein N, Albar L, Feuillet C, Schlagenhauf E, Keller B: **Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution.** *Plant J* 2001, **26**:307-316.
30. Sandhu D, Gill KS: **Gene-containing regions of wheat and the other grass genomes.** *Plant Physiol* 2002, **128**:803-811.
31. Vavilov NI: **The law of homologous series in variation.** *J Genet* 1922, **12**:47-88.
32. Weeden NF, Muehlbauer FJ, Ladizinsky G: **Extensive conservation of linkage relationships between pea and lentil genetic maps.** *J Hered* 1992, **83**:123-129.
33. Brauner S, Murphy RL, Walling JG, Przyborowski J, Weeden NF: **STS markers for comparative mapping in legumes.** *J Am Soc Hort Sci* 2002, **127**:616-622.
34. Simon CJ, Muehlbauer FJ: **Construction of a chickpea linkage map and its comparison with maps of pea and lentil.** *J Hered* 1997, **88**:115-119.
35. Menancio-Hautea D, Fatokun CA, Kumar L, Danesh D, Young ND: **Comparative genome analysis of mungbean (*Vigna radiata* (L.) Wilczek) and cowpea (*V. unguiculata* (L.) Walpers) using RFLP mapping data.** *Theor Appl Genet* 1993, **86**:797-810.
36. Boutin SR, Young ND, Olsen TC, Yu ZH, Shoemaker RC, Vallejos CE: **Genome conservation among three legume genera detected with DNA markers.** *Genome* 1995, **38**:928-937.
37. Gaut BS: **Evolutionary dynamics of grass genomes.** *New Phytol* 2002, **154**:15-28.
38. Endre G, Kereszt A, Kevei Z, Mihacea S, Kaló P, Kiss GB: **A receptor kinase gene regulating symbiotic nodule development.** *Nature* 2002, **417**:962-966.
39. Stracke S, Kistner C, Yoshida S, Mulder L, Sato S, Kaneko T, Tabata S, Sandal N, Stougaard K, Szczyglowski K, Parniske M: **A plant receptor-like kinase required for both bacterial and fungal symbiosis.** *Nature* 2002, **417**:959-962.
40. Schauser L, Roussis A, Stiller J, Stougaard J: **A plant regulator controlling development of symbiotic root nodules.** *Nature* 1999, **402**:191-195.
41. Gualtieri G, Kulikova O, Limpens E, Kim DJ, Cook DR, Bisseling T, Geurts R: **Microsynteny between pea and *Medicago truncatula* in the *SYM2* region.** *Plant Mol Biol* 2002, **50**:225-235.