

# Cross-legume Advances Through Genomics (CATG)

White Paper:

*Legumes as a Model Plant Family:*

*Genomics for Food and Feed*

Organizing committee:	1
Support:	1
Acknowledgments:	1
Executive summary: General Public	2
Executive summary: Scientific Public	3
I. Introduction	6
A. The multiple roles of legumes	6
B. The path to Santa Fe 2004	9
II. Legumes as a Model Plant Family: Genomics for Food and Feed	11
III. Why genomics?	13
IV. A strategy towards unified legume genomics	14
1. The reference legumes, <i>Medicago truncatula</i> , <i>Lotus japonicus</i> , and <i>Glycine max</i>	14
2. Extended genomic tools for targeted species, <i>Phaseolus</i> and <i>Arachis</i>	15
3. Translational genomic tools for other major legume crops and targeted experimental systems	15
4. Develop additional, selected legume experimental systems	16
V. Available genomics resources	16
VI. Tools and timeline to address strategic goals The timeline is divided into short-term (Years 1 to 3), medium-term (Years 4 to 6), and long-term (Years 7 to 10) periods.	16
A. Structural genomics, genome sequencing, and synteny mapping	16
B. Bioinformatics	17
D. Education, outreach, and recruitment of young scientists	21
E. Genetic/Genomic Resource Center	22
F. The international legume community, especially in the developing world	22
VIII. Future of CATG initiative	23
IX. Outcomes	23
X. Conclusions	24
XI. Literature cited	24
Appendix A	29

**Organizing committee:**

William Beavis (Bioinformatics), Charlie Brummer (Alfalfa), Paul Gepts (Beans; chair), Randy Shoemaker (Soybean), Tom Stalker (Peanut), Norm Weeden (Cool-season legumes) and Nevin Young (Model legumes)

**Support:**

National Science Foundation/Plant Genome & United States Department of Agriculture/ Cooperative State Research, Education, and Extension Service/ National Research Initiative/Plant Genome

**Acknowledgments:**

The community of legume researchers thanks the federal agencies (NSF/Plant Genome and USDA/CSREES/NRI), which provided funding and encouragement for the organization of the CATG meeting that gave rise to this white paper. The organizing committee of the CATG meeting thanks:

- their fellow legume researchers for their enthusiasm and positive spirits, which made the CATG meeting a success;
- D. Zamir, J. Doyle, A. Hirsch, M. Grusak, and E. Pichersky for exciting and stimulating plenary lectures
- their colleague W. Beavis and staff at NCGR for assistance with local arrangements
- G. May and E. Wurtele for their contribution to the Proteomics and Metabolomics section

An abbreviated version of this white paper has been published in *Plant Physiology* (Gepts, P., W.D. Beavis, E.C. Brummer, R.C. Shoemaker, H.T. Stalker, N.F. Weeden, and N.D. Young. 2005. Legumes as a model plant family. Genomics for food and feed Report of the Cross-Legume Advances through Genomics Conference. *Plant Physiol.* 137:1228-1235).

## **Executive summary: General Public**

The legume family includes such crops as soybean, alfalfa, peas, dry and green beans, peanuts, chickpea or garbanzo beans, clover, and lentils. These crops play a very important role as a source of proteins for human food and animal feed, as a source of phytochemicals for human health, as forages, as ground covers, and as timber, in the U.S. and worldwide. Genomics is increasingly becoming a powerful tool for the improvement of crops, including legumes because it accelerates the discovery of novel traits such as higher yield, disease and pest resistance, tolerance to plant stresses such as drought, and value-added traits such as increased nutritional value. It also facilitates the introduction of these traits in new, improved crop varieties. In December 2004, at a conference sponsored by NSF and USDA, legume researchers met to discuss opportunities for cross-legume genomics and establish a 10-year plan for joint genomics projects, which is detailed in this white paper. The overall focus of the initiative will be to develop legume varieties that provide improved food and feed. Furthermore, the initiative identifies priorities for the development of genomics research tools in the different legumes on a 3-year, 6-year, and 10-year basis. The practical outcomes expected from this initiative include:

- crop varieties that keep farmers competitive with higher yield and more desirable protein and oil,
- foods with enhanced mineral and vitamin content,
- foods with better health-promoting and fewer anti-nutritional components,
- an improved stewardship of environmental nitrogen,
- plants that resist disease and pests with fewer chemicals, and
- a cadre of young scientists ready to take legume genomics to the next level, and
- a public that understands the benefits of legumes in the human diet and the environment.

An abbreviated version of this white paper has been published in the journal *Plant Physiology* (Gepts, P., W.D. Beavis, E.C. Brummer, R.C. Shoemaker, H.T. Stalker, N.F. Weeden, and N.D. Young. 2005. Legumes as a model plant family. Genomics for food and feed Report of the Cross-Legume Advances through Genomics Conference. *Plant Physiol.* 137:1228-1235).

## Executive summary: Scientific Public

This white paper describes a plan for a coordinated effort for the development and research involving genomics across the legume family. Genomics has established itself as a major research area, which can fuel not only scientific advances in our understanding of living organisms but also practical advances in human health and agriculture. Given the progress in the development of genomics in the model legumes *Medicago truncatula* and *Lotus japonicus* and in soybean (*Glycine max*), the economically most important legume in the world, there is a major scientific opportunity to translate the genomic advances from these species to other food and feed legumes. Conversely, it is also important to permit the sharing of the considerable genetic and biological information that already exists in many of the legume crops. The crops specifically identified in this white paper include cool-season pulses [pea (*Pisum sativum*), lentil (*Lens culinaris*), chickpea (*Cicer arietinum*), field bean (*Vicia faba*)], warm-season food legumes [peanut (*Arachis hypogea*), common bean (*Phaseolus vulgaris*)], and forage legumes [alfalfa (*Medicago sativa*), clover (*Trifolium* spp.)].

With some 20,000 species, the legume family is one of the largest botanical families. It is second to cereals in agricultural importance. Its most distinctive characteristic is the ability to convert free nitrogen from the air into bound nitrogen that can be used by plants via a symbiotic relationship with rhizobia and mycorrhizal fungi. Recent investigations in the model legumes have started identifying the genes that are responsible for this symbiotic relationship in both the plant host and the bacterial symbiont. Furthermore, investigations into the root rhizosphere are very well-developed in legumes. Genomics provides avenues to transfer this information from the model to crop legumes.

Other interactions involving legume plants are of special scientific and agronomic interest, such as those involving pathogens, insect pollinators, and herbivore predators. Additional characteristics of great interest in the legumes are their contribution to the human diet (chiefly as a source of protein) and human health (as a source of vitamins, minerals, fibers, and human hormone analogs), in spite of the presence of anti-nutritional and allergenic factors in some legumes. Legumes also play an important role as forage crops, in the nitrogen cycle of ecosystems, and, in tropical areas, as timber.

The Legume Crops Genomics Initiative brings together several legume commodity organizations and their respective research communities. It has endeavored to coordinate genomics research across the major food and feed legumes. One of the outcomes of LCGI was the organization of the CATG (Cross-legume Advances Through Genomics) conference on December 14-15, 2004, in Sante Fe, NM, and funded by the

National Science Foundation (Plant Genome Research Program) and the United States Department of Agriculture (National Research Initiative). During this conference, legume researchers established a 10-year plan for cross-legume genomics research, detailed in this white paper.

The overall thrust of the plan is to improve legumes as a source of nutritious food and feed. This broad goal represents a statement directed primarily towards the public, who will be the ultimate beneficiary of genomic activities. It will be supported by research in five more specific areas: (1) the biochemical composition of seeds, particularly with regard to proteins; (2) the biochemical pathways leading to health-promoting phytochemicals; (3) symbiotic relationships with rhizobia and mycorrhizal fungi; (4) the characteristic nitrogen economy of legume plants; and (5) resistance to diseases and pests.

Furthermore, four tiers of legume species were identified, each with specific genomic resources to be developed. Based on phylogenetic arguments, and particularly the degree of synteny, two major foci of legumes were identified, the hologaleginoid clade or cool-season legumes and the phaseoloid/millettioid clade or warm-season legumes. In each of these two foci, one or two reference species were identified, *M. truncatula* and *L. japonicus* in the former and soybean in the latter. In these reference species, investments should be continued in order to develop the full range of genomics resources, including sequencing of the entire genome. For a second group, common bean and peanut, a broad range of genomic resources should be developed, including a physical map, BAC-end sequencing and marker development, anchoring of the genetic and physical map, ESTs of the major organs, chip resources, and sequencing of gene-rich regions.

A third group consisted of all other legume crops in the two foci, including pea, lentil, chickpea, field bean, clover, cowpea (*Vigna unguiculata*), and pigeon pea (*Cajanus cajan*). For these legumes, translational genomic tools should be developed, principally the development of cross-legume markers, and species-specific recombinant inbred populations, genetic maps, EST and BAC libraries. A fourth group includes other legumes not included in the two main foci, such as members of the basal legume clades, but which may be important in the study of the evolution of legume-characteristic traits such as symbiosis, flower and pod development, and polyploidy. In this group, *ad hoc* genomic resources should be developed to allow comparison with reference and crop legumes.

The white paper makes specific recommendations for 3-year-, 6-year-, and 10-year objectives. An important concern of legume researchers is the development within the short term of additional cross-legume markers so that synteny relationships can be

established in more detail. A legume information resource will be developed, which provides a one-stop-shop access to all legume bioinformatics resources and would be able to semantically respond to queries for information and services. Furthermore, it envisions the development of transcriptomic, proteomic, and metabolomic resources, principally in the reference legumes, but eventually also in all major legumes.

Finally, the white paper also makes recommendations for education and outreach to young scientists, farmers and the public at large. A concern, not unique to legumes, remains the long-term fate of the genomics and other biological resources that are being developed. An abbreviated version of this white paper has been published in the journal *Plant Physiology* (Gepts, P., W.D. Beavis, E.C. Brummer, R.C. Shoemaker, H.T. Stalker, N.F. Weeden, and N.D. Young. 2005. Legumes as a model plant family. Genomics for food and feed Report of the Cross-Legume Advances through Genomics Conference. *Plant Physiol.* 137:1228-1235).

## I. Introduction

### A. The multiple roles of legumes

With some 20,000 species, the legumes are the third largest family of higher plants. In comparison with other families with model species, the Gramineae only have some 10,000 species and the Brassicaceae some 3,500 species. This situation represents a challenge for comparative genomics, the identification of model species, and the determination of synteny.

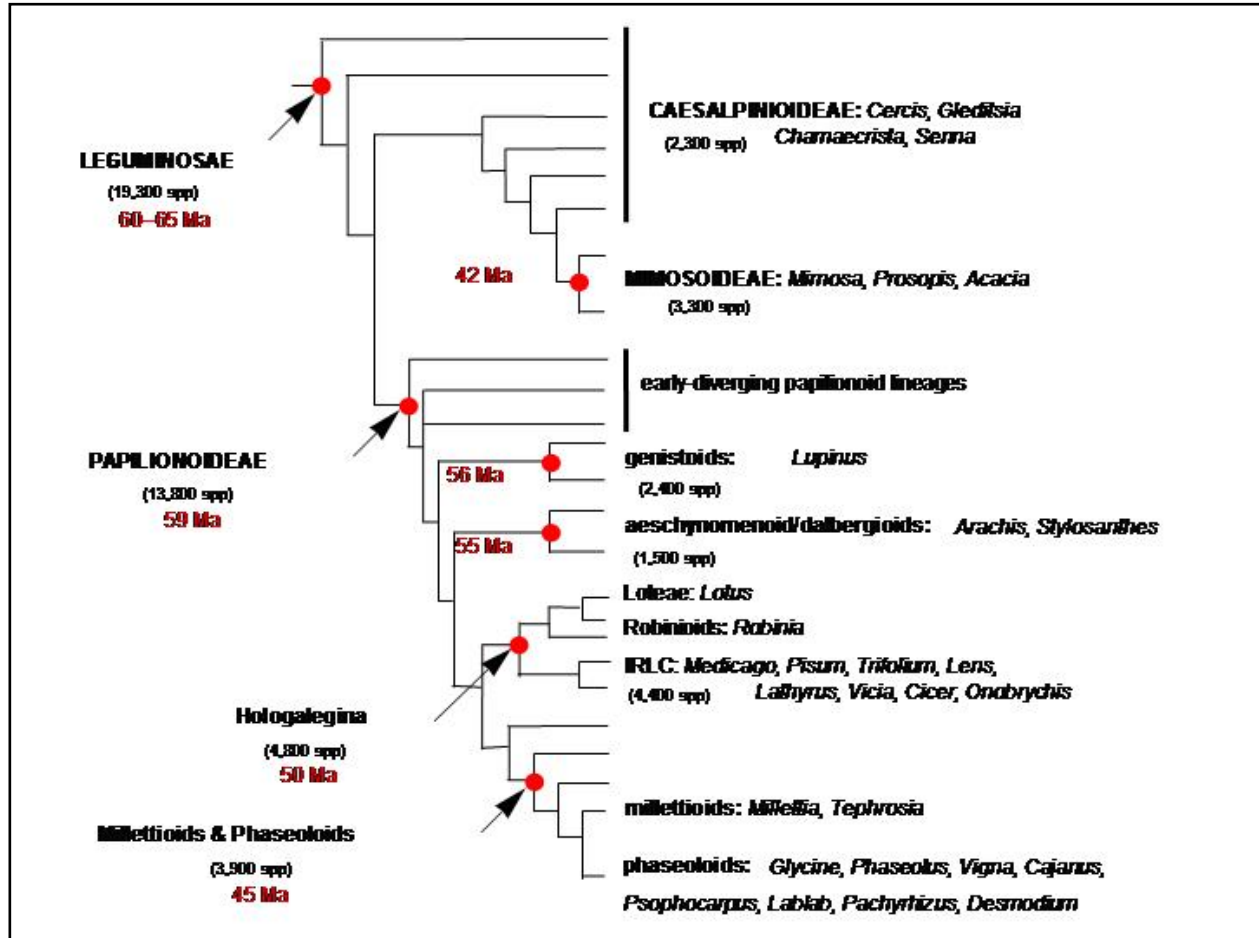
The Leguminosae are second to cereal crops in agricultural importance based on area harvested and total production. In 2004, more than 300 million metric tons of grain legumes were produced on 190 million ha (or about 13 % of total land under cultivation, including arable land and land under permanent crops) ([http://faostat.fao.org/faostat/collections?subset= agriculture](http://faostat.fao.org/faostat/collections?subset=agriculture)). The diverse roles of legume plants are often overlooked. Grain legumes provide about a third of all dietary protein nitrogen and a third of processed vegetable oil for human consumption (Graham and Vance, 2003). Seeds of grain legumes contain at least 20-40% of protein. In many places of the world, legumes complement cereals or root crops, the primary source of carbohydrates, in terms of amino acid composition. Whereas cereal seed proteins are deficient in lysine, legume seed proteins are deficient in sulfur-containing amino acids and tryptophan (Wang et al., 2003). This situation may explain why in most centers of crop domestication, legumes and cereals have been domesticated together (Gepts, 2004). Legumes are also important forages, in temperate [*e.g.*, alfalfa, clover (*Trifolium* spp.)] and tropical regions (*Stylosanthes* sp., *Desmodium* sp.). Of note are tropical legume trees, which play a particularly important role as forage in arid areas (National Academy of Sciences, 1979) and as abundant timber in humid areas (Doyle and Luckow, 2003).

Legumes also provide essential minerals required by humans (Grusak, 2002a) and produce health-promoting secondary compounds that can protect against human cancers (Grusak, 2002b; Bathena and Velasquez 2002; Madar and Stark, 2003) and protect the plant against the onslaught of pathogens and pests (Dixon et al., 2002; Ndakidemi and Dakora, 2003). In addition to their blood cholesterol-reducing effect (*e.g.*, Andersen et al., 1984), grain legumes generally also have a hypoglycemic effect, reducing the increase in blood glucose after a meal and, hence, blood insulin. Legumes are, therefore, included in the diet of insulin-dependent diabetics (Jenkins et al., 2003). Certain legumes, however, produce anti-nutritional factors, such as trypsin inhibitors

and phytohemagglutinins (Berger et al. 2003; Gupta, 1987; Wang et al. 2003), and allergens, the latter being a severe problem in peanut (Spergel and Fiedler, 2001). Genomics approaches, including metabolomics and proteomics, are essential to understanding the metabolic pathways that produce these anti-nutritional compounds and to eliminating these factors from the plant.

The molecular signaling taking place between legumes and rhizobial symbionts, pollinators, and herbivores and carnivores, suggests that legumes are an excellent model for the study of molecular signaling among organisms. Legume crops are of great significance because they produce substantial amounts of organic nitrogen fertilizer resulting from a symbiosis between the plant and bacterial symbionts (*e.g.*, Hirsch, 2004; Jensen and Hauggaard-Nielsen, 2003). Rapid progress is being made in unraveling the molecular basis of this nitrogen-fixing symbiotic relationship, principally in the two model legumes, *M. truncatula* and *L. japonicus* (Cullimore and Dénarié 2003; Mergaert et al., 2003; Penmetsa et al., 2003; Mitra et al., 2004; Cook, 2004; Oldroyd et al., 2005). Translating this success to other legume species can be realized with genomic approaches (Stougaard 2001; Broughton et al., 2003). In contrast with other botanical families, wind-pollinated species are extremely rare in the legumes, which are largely insect-pollinated (or self-fertilized). Although not unique to the legumes, insect pollination is accompanied by adaptations in the plant host such as the development of specific morphological traits and the production of volatile attractants. Morphological traits include specific inflorescence types such as racemes and pseudoracemes and a zygomorphic (bilateral) flower symmetry (Endress 2001; Tucker 2003). Floral volatiles have been studied in several legumes, including and not limited to, alfalfa and clover (Henning et al. 1992; Tava and Pecetti 1997). Interestingly, leaf volatiles also play an important role in communications with insects, particularly as a defense mechanism to attract predators of herbivores (Pichersky and Gershenzon 2002) in lima bean (Arimura et al. 2000) and *L. japonicus* (Ozawa et al. 2002).

Traditionally, the legume family has been divided into three subfamilies, Caesalpinieae, Mimosoideae, and Papilionoideae. The grain legumes are included in the latter subfamily. Within the Papilionoideae, there are four important clades, which group most of the economically important food and feed legumes (Fig. 1; Doyle and Luckow, 2003). The genistoid clade includes the genus *Lupinus* and the aescynomenoid / dalbergioid clade the peanut. The third and fourth clades have a common ancestor. The Hologalegina clade is split into two subclades, one that includes



**Figure 1.** Simplified schematic tree of legume family (modified from Doyle and Luckow, 2003). The three subfamilies (Caesalpinioideae, Mimosoideae, and Papilionoideae, and major subclades identified by recent molecular phylogenetic studies are shown in boldface (Kajita et al. 2001; Wojciechowski et al. 2004) and their positions are indicated by filled circles [estimated number of taxa from Lewis et al. (in press) and ages (in millions of years) from Lavin et al. (2005)].

the Loteae (*L. japonicus*) and the other that includes species with a chloroplast DNA characterized by the loss of one copy of the inverted repeat found in most angiosperm plants (hence called the IRLC or Inverted Repeat Loss Clade). The IRLC includes many species with temperate adaptation (cool-season legumes), such as alfalfa (and *M. truncatula*) chickpea, faba bean, lentil, and pea. The fourth clade, the phaseoloid/millettoid clade, includes several legumes that are better adapted to more tropical climates (warm-season legumes), such as common bean, cowpea [*Vigna unguiculata* (L.) Walp.], pigeon pea [*Cajanus cajan* (L.) Millsp.], and soybean, (Doyle and Luckow, 2003).

Phylogenetic relationships within the legume family are reflected in relatively high similarity or synteny at the genome level among the cool-season legumes, including Medicago and pea (Gualtieri and Bisseling 2002; Gualtieri et al. 2002; Borisov et al. 2003; Kalo et al. 2004) or between the warm-season legumes common bean and soybean (Boutin et al. 1995; Lee et al. 2001), but limited synteny is present among other legumes (for example, between cool-season and warm-season legumes) (Yan et al. 2004; Choi et al. 2004). Comparative genomics to assess synteny can facilitate back-and-forth use of genomics resources between different legume species, making the research cost effective, efficient, and non-duplicative. Despite studies assessing the overall levels of synteny (*e.g.*, Cannon et al. 2003; Yan et al. 2003, 2004; Choi et al. 2004), no systematic determination of (micro- and macro-)syntenic relationships among legume species has been attempted as has been accomplished in cereals. Thus, a detailed determination of these relationships is a critical need of the major legume crops.

Bringing the genomic and biological knowledge in the aforementioned reference legumes to bear on other food and feed legumes of major economic importance, including cool-season pulses (pea, lentil, chickpea), warm-season food legumes (peanut, common bean), and forage legumes (alfalfa, clover) represents a major scientific opportunity. Each legume presents unique features that could be examined with genomic resources developed in the model legumes or soybean together with additional resources to be developed in the crop itself. Examples of these features include the geotropic peg and pod development of peanut (Pattee et al. 1998), drought tolerance and vernalization in chickpea (Abbo et al. 2003), forage nutritive value in alfalfa (Riday et al. 2002), biologically interesting mutants in pea (*e.g.*, DeMason and Villani 2001; Hofer et al. 2001; Novak 2003), evolution of domestication (Koinange et al. 1996), co-evolution of host and pathogen/pests in common bean (*e.g.*, Gepts and Bliss 1985; Geffroy et al. 1999, 2000; Guzmán et al. 1995; Mkandawire et al. 2004; Aguilar et al. 2004), and differential resistance of legume species to closely related groups of pathogens such as *Fusarium* and rust (Gray et al., 1999; O'Donnell and Gray, 1997; Tenuta, 2004).

## B. The path to Santa Fe 2004

The “Cross-Legume Advances Through Genomics” (CATG) meeting in Santa Fe, was one outcome of the Legume Crops Genome Initiative (LCGI). The LCGI is bringing legume genomics researchers together with the major legume commodity industries,

including the alfalfa, bean, cool-season legumes (pea, lentil, and chickpea), peanut, and soybean industries. An initial meeting was convened by R. Boerma (Univ. of Georgia, Athens), J. St. John (USDA/ARS), and J. Yezak-Molen (AgSource) on July 30-31, 2001. This meeting resulted in a document identifying priority research areas in legume genomics ([http://129.186.26.94/Legume\\_Initiative/LegGenomicsPaper10Oct01.html](http://129.186.26.94/Legume_Initiative/LegGenomicsPaper10Oct01.html)). These included:

- sequencing of the genomes or gene-rich regions of strategic legume species,
- physical map development and refinement,
- establishment of high throughput, stable and transient gene transfer systems,
- development of proteomics and metabolomics,
- development of DNA markers for synteny analysis, comparative mapping and breeding,
- characterization and utilization of legume biodiversity, including quantitative trait loci (QTL) discovery
- research on specific topics such as crop-microbe co-evolution, multi-gene family evolution, and domestication of legume crops, and
- the development of a legume information resource to support the other priorities

Following this meeting a steering committee of LCGI was put into place, which included representatives of the legume commodities, their respective research communities, and a representative of the model legume research community.

Research towards achieving these research objectives has progressed, although the focus has been primarily on the development of resources and information, principally in two species considered to be “model legumes” (*M. truncatula* and *L. japonicus*) and in soybean (*Glycine max*), the major economic legume in the U.S. (Parniske, 2004; Riely et al., 2004; Stacey et al., 2004; Sumner et al., 2003; VandenBosch and Stacey, 2003). The development of similar genomics research in other legumes has lagged, which has limited the impact of genomics of the three reference species. Conversely, these three species could benefit from the sharing of biological information that already exists in other legume crops.

In addition to the annual Plant and Animal Genome conferences, legume researchers also met or will meet at the biennial International Conference on Legume Genomics and Genetics (2002: Minneapolis; 2004: Dijon; 2006: Brisbane) as well as at meetings for individual species such as soybean, *M. truncatula*, and beans.

The efforts of the LCGI led to the organization of a conference on legume

genomics held on December 14 and 15, 2004, in Santa Fe, NM (<http://catg.ucdavis.edu>). The main goal of this conference was to forge a common plan with specific objectives for **cross-legume genomics research**. The objectives of the conference were to: (1) identify a unifying goal for an international cross-legume genome project; (2) identify cross-cutting themes to help integrate the different legume crop genomics programs, including a unified legume genomics information system, nutritional and health-related aspects of legumes, and detailed synteny and comparative genomics of legumes; and (3) outline specific components and milestones for this initiative.

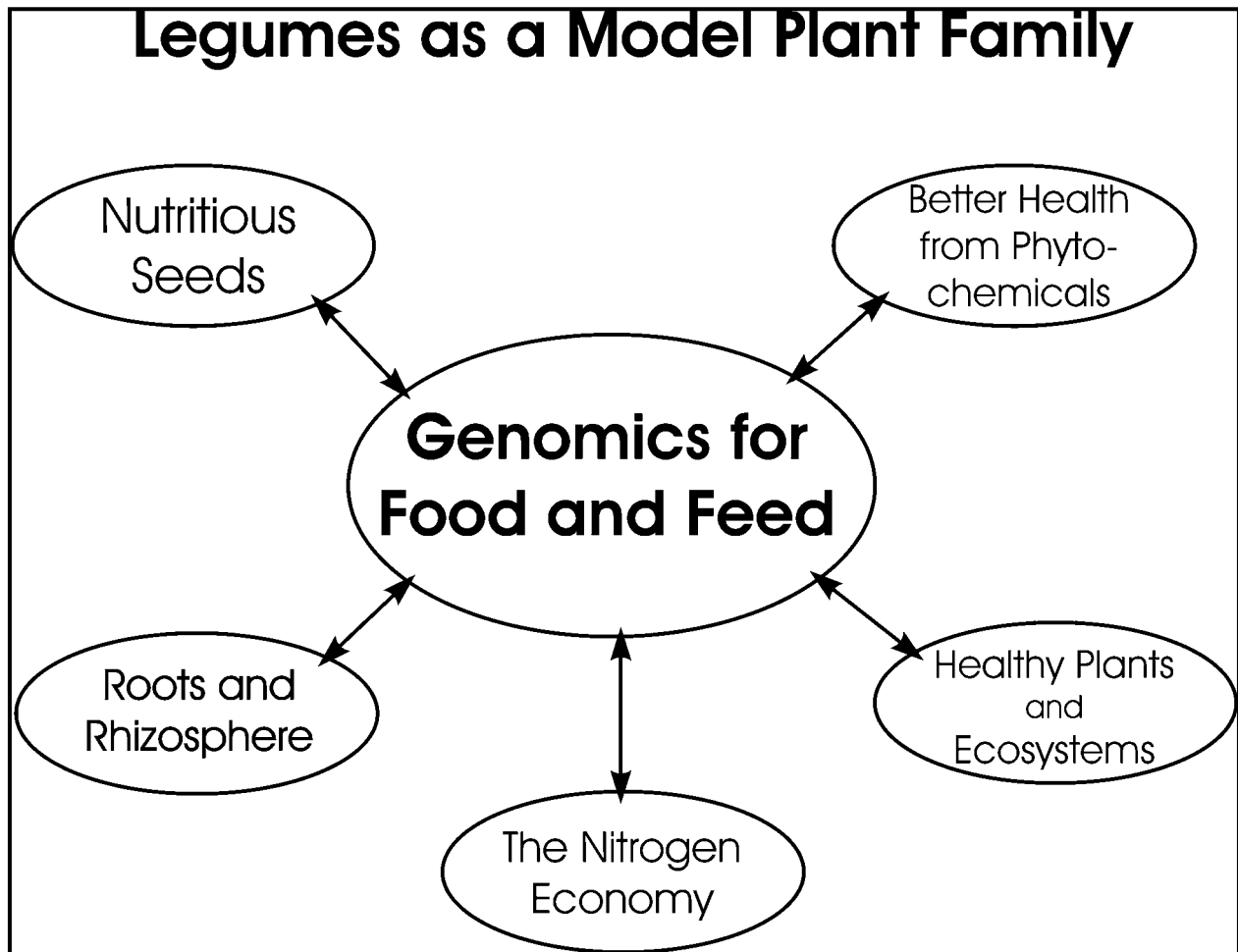
The organizing committee consisted of researcher representatives on the LCGI steering committee: Charlie Brummer (alfalfa), Paul Gepts (beans; chair), Randy Shoemaker (soybean), Tom Stalker (peanut), Norm Weeden (cool-season legumes) and Nevin Young (model legumes). In addition, Bill Beavis participated as a bioinformatics representative and helped with local arrangements. About 50 individuals attended the two-day conference (see Appendix A). They represented the researcher communities of the leading legume communities as well as various funding agencies.

## II. Legumes as a Model Plant Family: Genomics for Food and Feed

The scientists at the CATG conference decided to focus the activities of cross-legume projects on the general theme of the improvement of food and feed. This choice represents a bold vision for the future for legume genomics, as well as a statement directed primarily towards the public, who will be the ultimate beneficiary of genomic activities. This unified theme combines several areas of research (Fig. 2).

First and foremost, it recognizes the importance of **seed legumes (or pulses) as sources of dietary protein** for humans and animals, as well as **health-related phytochemicals** such as **fibers, hormone analogs, and antioxidants**. Genomics provides essential tools to fully understand the molecular and metabolic basis of the synthesis of these compounds, to increase their content in seeds and pods, and to better manipulate interactions between the plant's genetic makeup and its environment. A focus on seeds also underscores the importance of the genomics of reproductive biology in the development of higher-yielding, more nutritious legume cultivars.

One of the signature features of legumes is the **association between plants and rhizobial and mycorrhizal symbionts**. Comparative and functional genomics will provide in-depth understanding of the steps necessary to establish efficient symbioses leading to higher levels of nitrogen fixation and better acquisition of other minerals.



**Figure 2.** Major research areas supporting improved food and feed as a major goal of cross-legume genomics

Studies of the **rhizosphere in legumes** are among the most developed of all botanical families and can lead to significant advances in plant health and growth.

The common currency underlying protein-rich seed/forage and rhizobial symbiosis is nitrogen. The **nitrogen-rich life style** can explain - in part - the success and diversity of the legume family (McKey 1994) and represents a significant contribution to agricultural and natural ecosystems. Comparative and functional genomics will contribute a better understanding of how legume plants acquire, process, and store nitrogen.

Finally, the contributions of legumes could not be fully realized without **low**

**incidence of diseases and pests** that affect the family. Recent advances in our understanding of disease resistance genes will be complemented by the application of genomics to fully understand the mechanisms by which legume resist or tolerate pathogens and pests. Molecular markers developed during genomics projects will assist breeders in developing new, resistant cultivars.

### III. Why genomics?

Genomics is the study of the total set of hereditary material (the genome or complete DNA sequence) of an organism. In contrast with genetics, genomics represents a major paradigm shift in that it represents a comprehensive study of all the genes of an organism, within limits imposed by practical considerations. In addition to studying the DNA sequence of an organism (structural genomics), genomics also studies the transcription of genes into RNA (transcriptomics), the translation of RNA into proteins (proteomics), and the synthesis of metabolic compounds (metabolomics). Furthermore, to fully understand the function of genes, comparative genomics often compares the genomes of two or more organisms for purposes of identifying conserved functional sequences.

In addition to its comprehensive approach, genomics is a preferred approach in genetic improvement of crops for other reasons as well:

- Genomics bridges the scale of analysis between molecular genetics (up to fifty thousand base pairs of DNA) and genetics (several hundred thousand to millions of base pairs, bp). For example, clusters of genes such as disease resistance genes that stretch over several hundred thousand bp can now be analyzed in their entirety. Large DNA fragments can be used to study the chromosomes of species and the location of genes on these chromosomes.
- Interactions among genes and their products in developmental and metabolic pathways can be studied in more detail and better understood, leading to what is now called “systems biology.”
- Information about genes can be transferred between species, including distantly related species such as Arabidopsis and soybean, or humans and plants, because DNA is a “common language” between these disparate living organisms. Thus, advances in one species can be potentially useful to others.
- Because DNA is the common language among organisms, the evolution of genes underlying important traits such as symbiotic nitrogen fixation or reproductive development can be dissected in more depth than ever before. The genes that are

responsible for important nutritional features can be identified and their expression can be increased, decreased, or eliminated, depending on the objectives.

- In turn, the enormous store of genetic diversity in most crop species and their wild relatives can be genetically characterized to, for example, examine the process of domestication, determine the relative importance of selection vs. genetic drift and the extent of gene flow, thus guiding decisions about the management and use of these valuable germplasm resources.
- The better development of molecular markers and their association with agronomically important traits enables plant breeders to increase the efficiency of selection, thereby hastening the development of superior cultivars.

#### IV. A strategy towards unified legume genomics

The path to better food and forage legumes requires a detailed knowledge of the different genes involved in the biochemical pathways leading up to key nutritional compounds, including the expression patterns and levels of these genes and their interactions. The tools of genomics, including bioinformatics and synteny analysis, provide the opportunity not only to obtain this comprehensive type of metabolical information, but also to integrate research efforts across different species. At the CATG conference, participants considered information on legume phylogeny and economic importance - among other factors - to establish a cross-legume genomics strategy and priorities for the development of legume genomic resources over the period of the coming ten years. Specifically, four tiers of investment of genomic tools were recognized in the legumes. These tiers are (in decreasing order of priority and investment):

##### **1. The reference legumes, *Medicago truncatula*, *Lotus japonicus*, and *Glycine max***

Cross-legume genomics research will be organized around two major clades that include most of the economically important legumes. These two clades correspond to the Hologalegina (cool-season legumes) and phaseoloid/millettoid (warm-season legumes) clades (Fig. 1; Doyle and Luckow 2003). Jointly, they represent a large part of the economic legumes, especially for food and feed, and capture approximately 40% of the phenotypic variation among all legumes (J. Doyle, pers. comm.). The relatively broad taxonomic distance separating the two clades (Doyle and Luckow, 2003; Wojciechowsky et al. 2004; Lavin et al., 2005) warrants the development of one or two

reference systems within each one. *M. truncatula* and *L. japonicus* represent models for the cool-season legumes and *G. max* for the warm-season legumes. For each of these reference systems, considerable investments have already been made and should be continued in order to develop the full range of genomics resources, including sequencing of the entire genome (already well under way for *M. truncatula* and *L. japonicus*, Young et al. 2005) or at the very least the gene-rich regions and extensive transcriptomics, proteomics, and metabolomics resources.

## **2. Extended genomic tools for targeted species, *Phaseolus* and *Arachis***

*Phaseolus* and *Arachis* will be targeted for the development of a range of extended genomic resources. These include development of a physical map accompanied by back-end sequencing (BES), molecular markers based on BES, anchoring of physical and genetic maps, ESTs of the major organs especially those involved in reproductive development, microarray and DNA chip resources, and ultimately sequencing of the *Phaseolus* genome and gene-rich regions for *Arachis*. In addition to their economic importance on a worldwide basis, special arguments support more extensive development of genomic tools in these two crops. The small, diploid genome of *Phaseolus* is a key to understanding genome structure and expression in the phaseoloid/millettioid group (Fig. 1), which includes soybean. *Arachis* occupies a phylogenetically more distant position from the two foci proposed here (Fig. 1), providing perspective on the evolution of the foci.

## **3. Translational genomic tools for other major legume crops and targeted experimental systems**

All legumes belonging to the two foci will benefit from the development of translational tools that enable the sharing of genetic and genomics information among the various species. The translation tools consist primarily of cross-legume markers [extending the preliminary work of Brauner et al. (2002), Kalo et al. (2004), Choi et al. (2004), and Zhu et al. (2005), by expanding the number of markers and species across the two major foci and peanut], EST libraries from several critical tissues and BAC libraries with extensive coverage of the genome. In addition, for those species still lacking genetic maps and recombinant inbred populations, development of these resources should be assigned a high priority. The crops recommended for targeting include pea, lentil, chickpea, faba bean, alfalfa, clover, cowpea, pigeon pea, and lupin (*Lupinus* spp.). With progress in sequencing, some genomes in this group may be sequenced at least in part (gene-rich regions). They include the genomes of pea (to permit more productive use of the extensive biochemical and physiological literature

for this species) and chickpea, one of the most drought-tolerant species among legumes of major economic importance.

#### **4. Develop additional, selected legume experimental systems**

Some of the fundamental biological questions, such as the origin of legume-characteristic traits, require an evolutionary approach that encompasses the entire legume family. Such traits include reproductive development (especially floral and pod development), the origin of nodulation, and the evolution and importance of polyploidy. The two foci, while providing coverage for most economic legumes involved in food and feed, do not come close to covering the biodiversity included in the Fabaceae. Hence, to address the issue of legume-characteristic traits, other species may have to be considered, including species in the basal clades of the Papilionoideae, the Caesalpinieae (*e.g.*, *Chamaecrista* sp.), and the Mimosoideae. In these species, *ad hoc* genomic resources targeted toward evolutionary genomics questions of interest will have to be developed to allow comparisons with reference and other legume species.

### **V. Available genomics resources**

Available biological resources and tools for legume genomics are summarized in Table 1. Existing legume data resources are presented in Table 2. They show an uneven distribution in the availability of resources and tools and highlight the need to develop resources so that cross-legume research can become a reality.

### **VI. Tools and timeline to address strategic goals**

The timeline is divided into short-term (Years 1 to 3), medium-term (Years 4 to 6), and long-term (Years 7 to 10) periods.

#### **A. Structural genomics, genome sequencing, and synteny mapping**

Genomics resources need to be developed or expanded within the reference legumes as well as for the major crop legumes to enhance the translation of genomics information across species. The CATG consensus, benchmarked over the next decade, envisions the following:

1. Within three years, a legume synteny project should be completed, in which 500 gene-based markers developed primarily from important biochemical genes will

be mapped across major crop legumes in at least one mapping population per species. These markers should be explicitly tied to whole genome sequences of reference species. Additional segregating populations useful for mapping important traits should be developed. Physical mapping should be initiated by end-sequencing BAC libraries currently available in crop species and fingerprinting BAC libraries that have been constructed from parents of mapping populations. A legume-wide and plant-wide workshop specifically addressing potential solutions to the long-term maintenance and conservation of genomic resources is called for.

2. Within six years, a genome-wide survey of genetic diversity among germplasm accessions of each major legume species should be undertaken, using common markers to facilitate translation of the results. Evaluation of the extent of linkage disequilibrium in exotic and domesticated germplasm should be started.

Phenotypic evaluation of multiple populations per species should be conducted so that the locations of quantitative trait loci for important agronomic traits, such as seed composition and yield, can be identified by genetic and association mapping. The accumulation of mapping information will facilitate the exploration of syntenic region across legumes. Physical maps should be constructed for common bean and diploid *Arachis*. The soybean genome, at least its gene-rich regions, should be sequenced by the fourth or fifth year.

3. Within ten years, the genome sequences of the *Phaseolus* and *Arachis* and selected BAC sequencing in *Pisum* and *Cicer* should be completed. Further genetic mapping in all species will continue.

## B. **Bioinformatics**

1. Current Status. Legume research involves utilization of genetic, genomic, transcript, proteomic, metabolic, and phenotypic information. Legume biologists that work with crop species typically rely on integrative and comparative analyses using information from model (*Arabidopsis*) and reference (*Medicago*, *Lotus* and soybean) species. This information is accessible to legume researchers through dozens of autonomous, web-enabled information resources (*e.g.*, see <http://www.comparative-legumes.org/lisg/lis> and <http://catg.ucdavis.edu/Legume%20information%20resources-wdb.pdf>). Most of these information resources were developed to enable public access to data and information from individual experimental programs, often funded by plant genome initiatives at NSF or USDA. These are typically species-centric and consist of data from a single technology platform (*e.g.*, fingerprinted soybean BACs

assembled into a physical map [<http://soybeangenome.siu.edu>], genomic sequences of the *Medicago* gene space [<http://genome.medicago.org>], and soybean microarray data [<http://soybeangenomics.cropsci.uiuc.edu>]). Some integrate information from a single species (eg. BAC contigs, DNA sequences markers and QTL [<http://bioinformatics.siu.edu>]). The proliferation of distributed, autonomous and often ephemeral information resources and analysis tools is a logical outcome of the distributed nature of the publicly funded plant biological research enterprise. Indeed, because novel bioinformatics ideas will come from multiple institutions, the bioinformatics community should expect and embrace a distributed marketplace.

Although a distributed marketplace provides a fertile environment for development of novel methods and bioinformatic approaches for dissemination of large-scale biological data and the proliferation of distributed, autonomous information resources is a logical outcome to the distributed nature of the publicly supported plant biological research enterprise, it is also a source of logistical frustration to legume biologists. Biologists have to search multiple, often unlinked web sites to find data, information, and analysis tools to support their research. Legume breeders, geneticists, molecular biologists, physiologists and biochemists are quite often interested in identifying candidate genes involved in the expression of a whole plant phenotype. In order to identify candidate genes the information resource should integrate information from linkage maps, QTL, physical maps, annotated genomic and EST sequences, expression arrays, protein arrays, genetic and biochemical networks, and whole plant phenotypes. The Arabidopsis Information Resource (TAIR), is the best known example of an integrated information resource for a single species, but there are others, for example, Soybase (<http://soybase.agron.iastate.edu>).

On the other hand, many legume researchers, especially evolutionary biologists and computational biologists, are interested in establishing structural and functional gene models. Such comparative biological questions that transcend species need information resources that integrate a type of data or information across species. There are extensive, well known, well supported, international data repositories that store data-types across all species, e.g., genomic sequences GenBank (<http://www.ncbi.nlm.nih.gov/Genbank/>), transcript sequences, dbEST (<http://www.ncbi.nlm.nih.gov/dbEST/>), abundance of mRNA transcripts, GEO [<http://www.ncbi.nlm.nih.gov/geo/>], protein families

(<http://www.sanger.ac.uk/Software/Pfam/>). In addition there have been efforts to develop information resources in which such data are restricted to plant species, e.g., ESTs (PlantGDB [[www.plantgdb.org](http://www.plantgdb.org)], TIGR gene indices [[www.tigr.org/tdb/tgi/plant.shtml](http://www.tigr.org/tdb/tgi/plant.shtml)]), mRNA arrays, (BarleyBASE [[www.barleybase.org](http://www.barleybase.org)]), GeneX-Lite [<http://www.ifom-firc.it/MICROARRAY/genex-lite/>] SGMD [<http://psi081.ba.ars.usda.gov/SGMD/default.htm>]), TM4 [<http://www.tigr.org/software/tm4/>]) protein and metabolic networks and pathways (PathDB [[www.ncgr.org/pathdb](http://www.ncgr.org/pathdb)], MetaCyc [[www.metacyc.org](http://www.metacyc.org)] and ArMet [<http://www.armet.org/>]). All of these are capable of accommodating information from legumes and many are open source.

If the legume biologist is working with a species for which information is incomplete, then s/he will need both integrative and comparative information resources. There have been a few efforts to integrate multiple types of biological information across multiple species; examples include LIS (<http://www.comparative-legumes.org>) and Gramene (<http://www.gramene.org>).

## 2. Recommendations

Scientists at the CATG conference made the following recommendations:

- a. Within three years, Develop a virtual, easy-to-navigate “one-stop” legume information network. By “one-stop” we refer by analogy to Google and how it can be seen as a single, yet non-exclusive, information resource. Such a system should be able to semantically respond to queries for both information and services. Current and future legume information resources can be viewed as components to this network. Attribution to the sources of the data will help assure a willingness to ‘register’ with the network and provide a peer-review mechanism for quality assurance. While it was beyond the scope of this meeting to recommend a particular technological approach to assure these attributes, there was considerable enthusiasm expressed for the utilization of the BioMOBY technology platforms (Schiltz et al., 2004) as a possible solution. Species-level biologist/curator of data and information and cross-species working groups composed of biologist/curators will be put in place.
- b. Within six years, legume information resources will accommodate proteomic and metabolomic data and information as well as analysis and visualization tools for these emerging sources of information.
- c. Within ten years, the legume community should become a leader in the development of new bioinformatics tools. There is an opportunity to do so with development of a one-stop legume information network. There are also

opportunities to develop novel analysis and visualization tools for both integrative and comparative research questions. For example, there are needs for: 1) Visualization of comparative maps at the level of linkage maps, physical maps (BAC-level), and homologous (orthologous) sequences; 2) The development of researcher-centric (breeders, geneticists, biochemists, molecular biologists) interfaces; 3) These tools can be extended to develop common biological and bioinformatic frameworks for all plants and animals to maximize the benefits of comparative biology.

d. Throughout the ten-year period, as new resources become available, workshops, education, and ongoing feedback from user community will be needed for all bioinformatics efforts.

### C. Proteomics and Metabolomics

Breakthroughs in understanding the relationship between genotype and phenotype will come about through proteomics and metabolomics. Legumes are especially appropriate for research in the areas of proteomics and metabolomics. As information on DNA sequences becomes more common in legumes, especially the reference legumes, the next steps will naturally focus increasingly on the expression of DNA sequence. Transcript profiling, proteomics and metabolomics are essential tools to fully understand the synthesis of compounds that are at the basis of this research community's focus on food and feed. In addition to protein, carbohydrates, and lipids, an understanding of the secondary metabolism and mineral nutrition of legumes is an important goal. The secondary metabolism involves interactions between legumes and pathogens, symbiotic organisms, predators, and pollinators. It is also the basis for many of the nutritional benefits ascribed to legumes. Extensive transcriptomics (*e.g.*, Colebatch et al., 2004), proteomics (Watson et al., 2004), and metabolomics (Sumner et al., 2003) programs are already under way in the model legumes, *M. truncatula* and *L. japonicus*.

1. Within three years, these resources need to be further developed in the two model legume species in parallel with the ongoing whole-genome sequencing efforts. The modest EST resources in major legume crops (other than soybean) must be expanded to include a broader range of tissues and environmental conditions.
2. Within six years, transcriptomic, proteomic and metabolomic resources should be extended to soybean. Additional resources to be developed include full length cDNA libraries from seed, mature leaf, nodule, flower and pod for common bean

and peanut from which DNA microarray and oligonucleotide “gene chips” can be developed for these two species.

3. Within ten years, selected resources should be extended to major crop legumes. Protein fragment databases, gel fingerprints for the important tissues of the major legumes, chips, including arrays of legume transcription factors and proteins (or antibodies to the proteins) of the developing flower, seed, and root, should be developed.

#### **D. Education, outreach, and recruitment of young scientists**

From the perspective of non-biologists, the most important goal of ‘omics’ technology platforms is to generate large scale biological information that can be leveraged to improve human health and nutrition. Because of their important role in nutrition and the environment, legumes offer many opportunities in education and outreach. As researchers expand their work into legume genomics, they will extend this knowledge to the general public so informed citizens can better understand the benefits of science and participate in setting the agenda for future research.

Ever since Mendel first discovered the fundamental laws of genetics through his studies of garden peas, legumes have helped people to understand biology. Now classroom and laboratory activities can introduce important concepts like genetic inheritance, symbiosis, and seed and flower development to young people. Already, new legume-based projects for K-12 students are being developed, like the “Bacteria and Plants Team Up!” packet created at Indiana University with NSF Plant Genome support and more educational resources like this need to be developed. For example, the role of legumes on a worldwide basis as a source of food, forage, and timber can be highlighted, as well as their role in ecosystem health and the nitrogen cycle. In the process, the benefits of good nutrition can be introduced to young people by making them familiar with the central role of legumes and health-promoting compounds in their diet (e.g., the place of legumes in the USDA food pyramid). Furthermore, research experience in legume genomics and genetics for K-12 teachers can, for example, also be provided through the RET (Research Experience for Teachers) program of NSF.

Even growers and farmers can benefit from learning more about legumes and the discoveries that result from genomics research in areas such as expanded crop rotations, the use of leguminous cover crops, and opportunities to reduce the use of fertilizer. Finally, workshops describing legume genomics and bioinformatics for applied plant scientists, crop breeders, even younger plant scientists in general need

to be developed and expanded. Bringing in younger generations of scientists to legume research is especially critical, helping to insure the discoveries we make today are exploited for society's benefit and continue far into the future.

#### **E. Genetic/Genomic Resource Center**

Many of the legumes, and especially the legume reference species (*M. truncatula*, *L. japonicum* and *G. max*) possess an enviable suite of genomic and genetic resources including tens to hundreds of thousands of ESTs from a wide range of tissues and organs, extensive national and international germplasm collections, mapping populations, large collections of BES, useful physical maps, mutant collections, TILLING populations, and functional genomic tools such as microarrays and GeneChips. For many resources, centers already exist that can accept, curate and distribute materials upon request. However, many resources are maintained and accessed on an ad hoc basis and need continual maintenance on a long-term basis. Screening of mutant populations often occurs with limited coordination. These issues continue to be a problem for all research communities, not only for legumes.

Many resources such as biological clones and seed stocks of mutants and mapping populations need to be regenerated on a regular and recurring basis. TILLING populations suffer from similar requirements. USDA-sponsored germplasm collections are already overburdened and are unable to accept additional responsibilities without a significant change in their infrastructure and funding. The consensus of this workshop was that there is an urgent need for a resolution to these problems. A legume-wide and plant-wide workshop specifically addressing potential resolutions to these problems is called for.

#### **F. The international legume community, especially in the developing world**

A mechanism will be established to ensure international coordination of legume genomics efforts and involving researchers, funding agencies, farmers and other stakeholders. A large proportion of worldwide legume production occurs in developing countries, where legumes fulfill a vital role in human nutrition, as tropical forages, and as timber species. For several species, such as common bean and chickpea, strong collaborative links among scientists throughout the world, including those of developing countries, will provide additional opportunities for funding and research and will significantly increase the impact of legume genomics. The *Phaseolus* genomics initiative (Phaseomics: Broughton et al., 2003), with representatives of over 20 countries, provide an example of current collaboration.

The scope of some projects, such as genome sequencing is beyond the capacity of single countries (e.g., the International Medicago truncatula Genome Sequencing Initiative). In addition, legume genomics projects and research centers in other countries provide opportunities for collaborations. These include the European Grain Legumes Integrated Project, which is part of the Framework 6 Food Quality and Safety research program of the European Union (<http://www.eugrainlegumes.org>), the Australian Centre of Excellence for Integrative Legume Research at the University of Queensland (<http://www.cilr.uq.edu.au>), and the Generation Challenge Program of the Consultative Group for International Agricultural Research (<http://www.generationcp.org/vw/index.php>). The introduction of Asian soybean rust (*Phakopsora pachyrhizi* Sydow) into the U.S. highlights the potential not only for cross-legume genomics research (e.g., between common bean and soybean) but also for international collaborations given the widespread distribution of the pathogen in Asia, Africa, and the Americas.

#### VIII. Future of CATG initiative

The research members of organizing committee of the CATG conference have represented their respective communities of researchers since 2002 on the LCGI steering committee. At the end of this three-year “term,” and having organized the CATG conference, it is appropriate to raise the issue of rotation in researcher membership of the LCGI steering committee. Any change, and the procedure for this change, will have to be discussed with the respective commodity representatives. Future goals and responsibilities of the CATG committee (i.e. the researcher representatives) include:

- Responsibility for planning future meetings
- Extending the initiative to various international communities
- Acting as bridge to producers and commodity groups
- Review progress of actions described in this white paper
- Keeping the collaborative spirit of the Santa Fe 2004 meeting alive

#### IX. Outcomes

The following are outcomes to expect from cross-legume genome research:

- a. Practical outcomes
  - i. Crop varieties that keep farmers competitive with higher yield and more desirable protein and oil

- ii. Foods with enhanced mineral and vitamin content
- iii. Foods with better health-promoting and fewer anti-nutritional components
- iv. Better stewardship of environmental nitrogen
  - v. Plants that resist disease and pests with fewer chemicals
- vi. Public that understands the importance of legumes in the diet and the environment
- b. Scientific Outcomes
  - i. Develop an extensive toolkit of genomic resources for legumes
  - ii. Dissect pathways to secondary metabolites
  - iii. Understand the biological basis of plant-microbe symbiosis and plant-pathogen/pest interactions
  - iv. Exploit the genetic and evolutionary diversity of legumes
    - v. Understand the biological basis for the commonalities and differences among diverse legumes
  - vi. Guide the development of biologist-friendly informatics resources
  - vii. Attract and develop a cadre of young legume scientists

## X. Conclusions

This white paper marks a major departure from previous legume genomic research, which was focused primarily on the development of genomic tools and biological investigations for individual species. The approach proposed in this paper represents a coordinated effort for the development and research involving genomics across the legume family. To paraphrase Bennetzen and Freeling (1997), the legume genomics community is now pursuing a unified legume genome in the hope of achieving synergy in synteny.

## XI. Literature cited

- Abbo, S., S. Lev-Yadun, and N. Galwey. 2002. Vernalization response of wild chickpea. *New Phytol.* [print] 154:695-701.
- Aguilar, O.M., O. Riva, and E. Peltzer. 2004. Analysis of *Rhizobium etli* and of its symbiosis with wild *Phaseolus vulgaris* supports coevolution in centers of host diversification. *Proc. Natl. Acad. Sci. USA* 101:13548-13553.
- Andersen, J.W., L. Story, B. Sieling, W.-J.L. Chen, M.S. Petro, and J. Story. 1984. Hypocholesterolemic effects of oat-bran or bean intake for hypercholesterolemic men. *Am. J. Clin. Nutr.* 40:1146-1155.
- Arimura, G.-i., R. Ozawa, T. Shimoda, T. Nishioka, W. Boland, and J. Takabayashi. 2000. Herbivory-induced volatiles elicit defence genes in lima bean leaves. *Nature (London)* 406:512-515.

- Bennetzen, J., and M. Freeling. 1997. The unified grass genome: synergy in synteny. *Genome Res.* 7:301-306.
- Berger, J.D., L.D. Robertson, and P.S. Cocks. 2003. Agricultural potential of Mediterranean grain and forage legumes: 2) Anti-nutritional factor concentrations in the genus *Vicia*. *Genetic Resources & Crop Evolution* [print] 50:201-212.
- Bhathena, S.J., and M.T. Velasquez. 2002. Beneficial role of dietary phytoestrogens in obesity and diabetes. *American Journal of Clinical Nutrition* [print] 76:1191-1201.
- Borisov, A.Y., L.H. Madsen, V.E. Tsyganov, Y. Umehara, V.A. Voroshilova, A.O. Batagov, N. Sandal, A. Mortensen, L. Schauser, N. Ellis, I.A. Tikhonovich, and J. Stougaard. 2003. The Sym35 gene required for root nodule development in pea is an ortholog of nin from *Lotus japonicus*. *Plant Physiology* (Rockville) [print] 131:1009-1017.
- Boutin, S., N. Young, T. Olson, Z. Yu, R. Shoemaker, and C. Vallejos. 1995. Genome conservation among three legume genera detected with DNA markers. *Genetics* 38:928-937.
- Brauner, S., R.L. Murphy, J.G. Walling, J. Przyborowski, and N.F. Weeden. 2002. STS markers for comparative mapping in legumes. *J. Amer. Soc. Hort. Sci.* 127:616-622.
- Broughton, W.J., G. Hernandez, M. Blair, S. Beebe, P. Gepts, and J. Vanderleyden. 2003. Beans (*Phaseolus* spp.) - model food legumes. *Plant Soil* 252:55-128.
- Cannon, S.B., W.R. McCombie, S. Sato, S. Tabata, R. Denny, L. Palmer, M. Katari, N.D. Young, and G. Stacey. 2003. Evolution and microsynteny of the apyrase gene family in three legume genomes. *Molecular Genetics and Genomics* 270:347-361.
- Choi, H.-K., J.-H. Mun, D.-J. Kim, H. Zhu, J.-M. Baek, J. Mudge, B. Roe, N. Ellis, J. Doyle, G.B. Kiss, N.D. Young, and D.R. Cook. 2004. Estimating genome conservation between crop and model legume species. *Proc. Natl. Acad. Sci. USA* 101:15289-15294.
- Colebatch, G., G. Desbrosses, T. Ott, L. Krusell, O. Montanari, S. Kloska, J. Kopka, and M.K. Udvardi. 2004. Global changes in transcription orchestrate metabolic differentiation during symbiotic nitrogen fixation in *Lotus japonicus*. *Plant Journal* 39:487-512.
- Cook, D.R. 2004. Unraveling the mystery of Nod factor signaling by a genomic approach in *Medicago truncatula*. *Proc. Natl. Acad. Sci. USA* 101:4339-4340.
- Cullimore, J., and J. Denarié. 2003. How legumes select their sweet talking symbionts. *Science* (Washington D C) [print] 302:575-578.
- DeMason, D.A., and P.J. Villani. 2001. Genetic control of leaf development in pea (*Pisum sativum*). *Int. J. Plant Sci.* 162:493-511.
- Dixon, R.A., L. Achnine, P. Kota, C.-J. Liu, M.S.S. Reddy, and L. Wang. 2002. The phenylpropanoid pathway and plant defence: A genomics perspective. *Molecular Plant Pathology* [print] 3:371-390.
- Doyle, J.J., and M.A. Luckow. 2003. The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. *Plant Physiology* (Rockville) [print] 131:900-910.
- Endress, P.K. 2001. Evolution of floral symmetry. *Curr. Opin. Plant Biol.* 4:86-91.
- Geffroy, V., D. Sicard, J. de Oliveira, M. Sévignac, S. Cohen, P. Gepts, C. Neema, and M. Dron. 1999. Identification of an ancestral resistance gene cluster involved in the coevolution process between *Phaseolus vulgaris* and its fungal pathogen *Colletotrichum lindemuthianum*. *Mol. Plant-Micr. Inter.* 12:774-784.
- Geffroy, V., M. Sévignac, J. De Oliveira, G. Fouilloux, P. Skroch, P. Thoquet, P. Gepts, T. Langin, and M. Dron. 2000. Inheritance of partial resistance against *Colletotrichum lindemuthianum* in *Phaseolus vulgaris* and co-localization of QTL with genes involved in specific resistance. *Mol. Plant-Micr. Inter.* 13:287-296.

- Gepts, P. 2004. Domestication as a long-term selection experiment. *Plant Breed. Rev.* 24 (Part 2):1-44.
- Gepts, P., and F.A. Bliss. 1985. F<sub>1</sub> hybrid weakness in the common bean: differential geographic origin suggests two gene pools in cultivated bean germplasm. *Journal of Heredity* 76:447-450.
- Graham, P.H., and C.P. Vance. 2003. Legumes: Importance and constraints to greater use. *Plant Physiol.* 131:872-877.
- Gray, L.E., L.A. Achenbach, R.J. Duff, and D.A. Lightfoot. 1999. Pathogenicity of *Fusarium solani* f. sp. *glycines* isolates on soybean and green bean plants. *J. Phytopathology* 147:281-284.
- Grusak, M.A. 2002. Enhancing mineral content in plant food products. *Journal of the American College of Nutrition* 21:178S-183S.
- Grusak, M.A. 2002. Phytochemicals in plants: genomics-assisted plant improvement for nutritional and health benefits. *Curr Opin Biotechnol* 13:508-511.
- Gualtieri, G., and T. Bisseling. 2002. Microsynteny between the *Medicago truncatula* SYM2-orthologous genomic region and another region located on the same chromosome arm. *Theoretical & Applied Genetics* [print] 105:771-779.
- Gualtieri, G., O. Kulikova, E. Limpens, D.-J. Kim, R. Cook Douglas, T. Bisseling, and R. Geurts. 2002. Microsynteny between pea and *Medicago truncatula* in the SYM2 region. *Plant Molec. Biol.* 50:225-235.
- Gupta, Y.P. 1987. Anti-nutritional and toxic factors in food legumes: a review. *Qualitas Plantarum Plant Foods for Human Nutrition* 37:201-228.
- Guzmán, P., R.L. Gilbertson, R. Nodari, W.C. Johnson, S.R. Temple, D. Mandala, A.B.C. Mkandawire, and P. Gepts. 1995. Characterization of variability in the fungus *Phaeoisariopsis griseola* suggests coevolution with the common bean (*Phaseolus vulgaris*). *Phytopathology* 85:600-607.
- Henning, J., Y. Peng, M. Montague, and L. Teuber. 1992. Honey bee (Hymenoptera: Apidae) behavioral response to primary alfalfa (Rosales: Fabaceae) floral volatiles. *J. Econ. Entomol.* 85:233-239.
- Hirsch, A.M. 2004. Plant-microbe symbioses: A continuum from commensalism to parasitism. *Symbiosis* 37:345-363.
- Hirsch, A.M., M.R. Lum, and J.A. Downie. 2001. What makes the rhizobia-legume symbiosis so special? *Plant Physiol.* 127:1484-1492.
- Hofer, J.M.I., C.W. Gourlay, and T.H.N. Ellis. 2001. Genetic control of leaf morphology: A partial view. *Ann. Bot.* 88:1129-1139.
- Jenkins, D.J.A., C.W.C. Kendall, A. Marchie, A.L. Jenkins, L.S.A. Augustin, D.S. Ludwig, N.D. Barnard, and J.W. Anderson. 2003. Type 2 diabetes and the vegetarian diet. *American Journal of Clinical Nutrition* [print] 78:610S-616S.
- Jensen, E.S., and H. Hauggaard-Nielsen. 2003. How can increased use of biological N<sub>2</sub> fixation in agriculture benefit the environment? *Plant & Soil* 252:177-186.
- Kajita, T., H. Ohashi, Y. Tateishi, C.D. Bailey, and J.J. Doyle. 2001. rbcL and legume phylogeny, with particular reference to Phaseoleae, Millettieae, and allies. *Syst. Bot.* 26:515-536.
- Kalo, P., A. Seres, S.A. Taylor, J. Jakab, Z. Kevei, A. Kereszt, G. Endre, T.H.N. Ellis, and G.B. Kiss. 2004. Comparative mapping between *Medicago sativa* and *Pisum sativum*. *Molecular Genetics and Genomics* 272:235-246.
- Koinange, E.M.K., S.P. Singh, and P. Gepts. 1996. Genetic control of the domestication syndrome in common-bean. *Crop Science* 36:1037-1045.
- Kris-Etherton, P.M., K.D. Hecker, A. Bonanome, S.M. Coval, A.E. Binkoski, K.F. Hilpert, A.E. Griel, and T.D. Etherton. 2002. Bioactive compounds in foods: Their role in the prevention of cardiovascular disease and cancer. *American Journal of Medicine* 113:71S-88S.

- Lavin, M., P.S. Herendeen, and M.F. Wojciechowski. 2005. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of the major family lineages immediately following an Early Tertiary emergence. *Syst. Biol.* in press.
- Lee, J.M., D. Grant, C.E. Vallejos, and R.C. Shoemaker. 2001. Genome organization in dicots. II. Arabidopsis as a 'bridging species' to resolve genome evolution events among legumes. *Theoretical & Applied Genetics* [print] 103:765-773.
- Lewis, G.P., B.D. Schrire, B.A. Mackinder, and M. Lock., (eds.). Legumes of the worldpp. Royal Botanic Garden, Kew, UK in press.
- Madar, Z., and A.H. Stark. 2002. New legume sources as therapeutic agents. *British Journal of Nutrition* [print] 88:S287-S292.
- McKey, D. 1994. Legumes and nitrogen: the evolutionary ecology of a nitrogen-demanding lifestyle, p. 211-228 *Advances in Legume Systematics*, Vol. 5.
- Mergaert, P., K. Nikovics, Z. Kelemen, N. Maunoury, D. Vaubert, A. Kondorosi, and E. Kondorosi. 2003. A novel family in *Medicago truncatula* consisting of more than 300 nodule-specific genes coding for small, secreted polypeptides with conserved cysteine motifs. *Plant Physiology (Rockville)* [print] 132:161-173.
- Mitra, R.M., C.A. Gleason, A. Edwards, J. Hadfield, J.A. Downie, G.E.D. Oldroyd, and S.R. Long. 2004. A Ca<sup>2+</sup>/calmodulin-dependent protein kinase required for symbiotic nodule development: Gene identification by transcript-based cloning. *Proc. Natl. Acad. Sci. USA* 101:4701-4705.
- Mkandawire, A.B.C., R.B. Mabagala, P. Guzman, P. Gepts, and R.L. Gilbertson. 2004. Genetic diversity and pathogenic variation of common blight bacteria (*Xanthomonas campestris* pv. *phaseoli* and *X. campestris* pv. *phaseoli* var. *fuscans*) suggests pathogen coevolution with the common bean. *Phytopathology* 94:593-603.
- National Academy of Sciences. 1979. Tropical legumes: resources for the future National Academy of Sciences, Washington D.C.
- Ndakidemi, P.A., and F.D. Dakora. 2003. Review: Legume seed flavonoids and nitrogenous metabolites as signals and protectants in early seedling development. *Functional Plant Biology* 30:729-745.
- Novak, K. 2003. Allelic relationships of pea nodulation mutants. *Journal of Heredity* [print] 94:191-193.
- O'Donnell, K., and L.E. Gray. 1995. Phylogenetic relationships of the soybean sudden death syndrome pathogen *Fusarium solani* f. sp. *phaseoli* inferred from rDNA sequence data and PCR primers for its identification. *Mol Plant Microbe Interact.* 8:709-716.
- Oldroyd, G.E.D., M.J. Harrison, and M. Udvardi. 2005. Peace talks and trade deals. Keys to long-term harmony in legume-microbe symbioses. *Plant Physiol.* 137:1205-1210.
- Ozawa, R., T. Shimoda, M. Kawaguchi, G.-i. Arimura, J.-i. Horiuchi, T. Nishioka, and J. Takabayashi. 2000. *Lotus japonicus* infested with herbivorous mites emits volatile compounds that attract predatory mites. *Journal of Plant Research* 113:427-433.
- Parniske, M. 2004. Molecular genetics of the arbuscular mycorrhizal symbiosis. *Curr. Opin. Plant Biol.* 7:414-421.
- Pattee, H.E., H.T. Stalker, and F.G. Giesbrecht. 1998. Reproductive efficiency in reciprocal crosses of *Arachis monticola* with *A. hypogaea* subspecies. *Peanut Science* 25:7-12.
- Penmetsa, R.V., A. Frugoli Julia, S. Smith Lucinda, R. Long Sharon, and R. Cook Douglas. 2003. Dual genetic pathways controlling nodule number in *Medicago truncatula*. *Plant Physiology (Rockville)* [print] 131:998-1008.
- Pichersky, E., and J. Gershenzon. 2002. The formation and function of plant volatiles: Perfumes for pollinator attraction and defense. *Curr. Opin. Plant Biol.* 5:237-243.

- Riday, H., E.C. Brummer, and K.J. Moore. 2002. Heterosis of forage quality in alfalfa. *Crop Science* 42:1088-1093.
- Riely, B.K., J.M. Ane, R.V. Penmetsa, and D.R. Cook. 2004. Genetic and genomic analysis in model legumes bring Nod-factor signaling to center stage. *Curr. Opin. Plant Biol.* 7:408-413.
- Schiltz, G., D. Gessler, and L. Stein. 2004. Semantic MOBY. Position paper for the W3C workshop on Semantic Web for Life Sciences [Online] <http://lists.w3.org/Archives/Public/public-swls-ws/2004Sep/att-0036/smoby-w3c-sw-ls.pdf> (verified Febr. 4, 2005, 2005).
- Spergel, J.M., and J.M. Fiedler. 2001. Natural history of peanut allergy. *Current Opinion in Pediatrics* [print] 13:517-522.
- Stacey, G., L. Vodkin, W.A. Parrott, and R.C. Shoemaker. 2004. National Science Foundation-Sponsored Workshop Report. Draft Plan for Soybean Genomics. *Plant Physiol.* 135:59-70.
- Stougaard, J. 2001. Genetics and genomics of root symbiosis. *Curr. Opin. Plant Biol.* [print] 4:328-335.
- Sumner, L.W., P. Mendes, and R.A. Dixon. 2003. Plant metabolomics: Large-scale phytochemistry in the functional genomics era. *Phytochemistry* [print] 62:817-836.
- Tava, A., and L. Pecetti. 1997. Volatiles from *Medicago sativa* complex flowers. *Phytochemistry* 45:1145-1148.
- Tenuta, A. 2004. Soybean rust infosheet [Online]. Available from Ontario Ministry of Agriculture and Food [http://www.gov.on.ca/OMAFRA/english/crops/facts/soybean\\_rust.htm](http://www.gov.on.ca/OMAFRA/english/crops/facts/soybean_rust.htm) (verified Febr. 4, 2005).
- Tucker, S.C. 2003. Floral development in legumes. *Plant Physiol.* 131:911-926.
- VandenBosch, K.A., and G. Stacey. 2003. Summaries of legume genomics projects from around the globe. Community resources for crops and models. *Plant Physiology (Rockville)* [print] 131:840-865.
- Wang, T.L., C. Domoney, C.L. Hedley, R. Casey, and M.A. Grusak. 2003. Can we improve the nutritional quality of legume seeds? *Plant Physiol.* 131:886-891.
- Watson, B.S., Z.T. Lei, R.A. Dixon, and L.W. Sumner. 2004. Proteomics of *Medicago sativa* cell walls. *Phytochemistry* 65:1709-1720.
- Wojciechowski, M.F., M. Lavin, and M.J. Sanderson. 2004. A phylogeny of legumes (Leguminosae) based on analyses of the plastid *matK* gene resolves many well-supported subclades within the family. *Amer. J. Bot.* 91:1846-1862.
- Yan, H.H., J. Mudge, D.J. Kim, R.C. Shoemaker, D.R. Cook, and N.D. Young. 2004. Comparative physical mapping reveals features of microsynteny between *Glycine max*, *Medicago truncatula*, and *Arabidopsis thaliana*. *Genome* 47:141-155.
- Yan, H.H., J. Mudge, D.J. Kim, D. Larsen, R.C. Shoemaker, D.R. Cook, and N.D. Young. 2003. Estimates of conserved microsynteny among the genomes of *Glycine max*, *Medicago truncatula* and *Arabidopsis thaliana*. *Theoretical & Applied Genetics* [print] 106:1256-1265.
- Young, N.D., S.B. Cannon, S. Sato, D. Kim, D.R. Cook, C.D. Town, B.A. Roe, and S. Tabata. 2005. Sequencing the genespaces of *Medicago truncatula* and *Lotus japonicus*. *Plant Physiol.* 137:1174-1181.
- Zhu, H., H.-K. Choi, D.R. Cook, and R.C. Shoemaker. 2005. Bridging model and crop legumes through comparative genomics. *Plant Physiol.* 137:1189-1196.

Table 1. Major biological features and currently available genomic resources of selected legume species (prepared by C. Brummer, N. Young, N. Weeden, R. Shoemaker, P. Gepts, and T. Stalker)

Hologaleginoid clade			Phaseoloid/Millettoid clade		Aeschynomeneae tribe
Alfalfa ( <i>Medicago sativa</i> )	Barrel medic ( <i>Medicago truncatula</i> )	Cool-season legumes: <i>Ps, Lc, Ca</i> <sup>a</sup>	Soybean ( <i>Glycine max</i> )	Beans ( <i>Phaseolus</i> spp.)	Peanut ( <i>Arachis hypogea</i> )
<b>Special biological features</b>					
<ul style="list-style-type: none"> <li>• <i>M. sativa</i> is one of a complex of species (or sub-species) that exist as both diploids and tetrasomic tetraploids, and a full range of hybrids are extant at both ploidies.</li> <li>• Close relatives: <i>M. truncatula</i>, and various other annual and perennial species, from diploid to hexaploid, and in base chromosome number (<math>x=7</math>, <math>x=8</math>)</li> <li>• Range of secondary metabolites, with different profile than <i>M. truncatula</i></li> <li>• Severe inbreeding depression precludes development of inbred lines</li> <li>• Clonal propagation</li> </ul>	<ul style="list-style-type: none"> <li>• Several genes in the Rhizobium and Mycorrhizae symbiosis pathway cloned and characterized</li> <li>• Symbiotic partner, Sinorhizobium genome, fully sequenced</li> <li>• Highly syntenic with cool season legumes; moderately syntenic with soybean and other Phaseoleae</li> <li>• Small number of translocations and inversions with other legume model, <i>Lotus japonicus</i></li> <li>• Systems biology program, integrating transcriptome, proteome, and metabolome</li> </ul>	<ul style="list-style-type: none"> <li>• Large collection of morphological, developmental and symbiotic mutations</li> <li>• Broad background knowledge in physiology. Coding sequences identified for plant height, seed shape and size, nodulation control, leaf morphology and flowering time. QTLs identified for many traits.</li> <li>• Extensive synteny with <i>Medicago truncatula</i></li> </ul>	<ul style="list-style-type: none"> <li>• Ancient paleopolyploid (14 MY and 45-55 MY)</li> <li>• Accumulates high levels of seed protein</li> <li>• Large seed size</li> <li>• Advanced embryonic systems methodologies</li> <li>• Adapted to wide range of latitudes</li> <li>• Close relatives are one annual (<i>G. soja</i>) and many perennials (<i>Glycine</i> sp.)</li> <li>• Allergen-free germplasm</li> <li>• Produces many bioactive seed compounds such as isoflavones, saponins, sterols, sphingolipids and</li> </ul>	<ul style="list-style-type: none"> <li>• Knowledge of phylogenetic and genealogic (domestication and ecological pattern); closely related domesticated species with contrasting growth habit, reproductive system, and ecological adaptation</li> <li>• Inheritance of the domestication syndrome is known (QTLs, major genes): growth habit, pod dehiscence, seed dormancy, size, and pigment</li> <li>• Co-evolution</li> </ul>	<ul style="list-style-type: none"> <li>• Reproductive development with temporal and special separation of fertilization and fruit development</li> <li>• Knowledge of species diversity and genomic groups; varying ecological adaptations of cultivated and wild species</li> <li>• Many disease and insect pathogens associated with cultivation</li> <li>• Allergens are major storage proteins and as yet unidentified proteins</li> </ul>

- via stem cuttings possible
- Self-incompatibility mechanisms are present but poorly defined
  - Pollinated by a diversity of bees

- research
- Produces many of the same bioactive compounds found in food legumes, including isoflavonoids and triterpenes

lectin

between host vs. pathogens, Rhizobium, bruchid insects

<i>Genome Features</i>							
Genome size (Mbp/haploid genome)	800-900 <sup>b</sup>	470 - 520	<i>Ps</i> : 4,400 <i>Lc</i> : 4,200 <i>Ca</i> : 950		1,115	588 (450-650 range)	2800
No. of chromosomes	<ul style="list-style-type: none"> <li>• <math>2n = 4x = 32</math></li> <li>• Small chromosomes of similar size</li> </ul>	<ul style="list-style-type: none"> <li>• <math>2n = 2x = 16</math></li> <li>• Distinctly separate euchromatic gene-space and pericentromeric heterochromatin</li> </ul>	<ul style="list-style-type: none"> <li>Ps: <math>2n = 2x = 14</math></li> <li>Lc: <math>2n = 2x = 16</math></li> <li>Ca: <math>2n = 2x = 14</math></li> </ul>		<ul style="list-style-type: none"> <li>• <math>2n = 2x = 40</math></li> <li>• wide range in chromosome size</li> <li>• 5 chromosomes with at least one heterochromatic arm</li> </ul>	<ul style="list-style-type: none"> <li>• <math>2n = 2x = 22</math></li> <li>• polytene chromosomes</li> </ul>	$2n = 4x = 40$ ; many diploid relatives
Cytogenetics	C-banding to distinguish among germplasms	<ul style="list-style-type: none"> <li>• High resolution pachytene analysis</li> <li>• Detailed FISH analysis with gene-rich BAC clones</li> <li>• Gene density estimated at 1 gene per 6.4 kbp throughout gene-space</li> <li>• Total gene number estimated at between 37,000 and 46,000</li> </ul>			<ul style="list-style-type: none"> <li>• Jackson: FIBER FISH with genetically-anchored BACs</li> <li>• Knapp: FIBER FISH with cDNA clones</li> <li>• Hymowitz &amp; Singh: Karyotypes, trisomics, monosomics</li> </ul>	<ul style="list-style-type: none"> <li>• Pedrosa-Harrand: FISH on pachytene chromosomes with RFLP clone pools and BACs</li> </ul>	

<i>Genetic Resources</i>						
Germplasm banks	<ul style="list-style-type: none"> <li>• USDA Germplasm collections (USDA-ARS, Pullman, WA): ca. 4000 accessions of <i>M. sativa</i>; well characterized for many traits, except ploidy</li> <li>• Large amounts of descriptive data on germplasm in GRIN</li> </ul>	<ul style="list-style-type: none"> <li>• INRA Germplasm resource (Montpellier, FR)</li> <li>• USDA Germplasm collection (USDA-ARS, Pullman, WA)</li> <li>• Australian Germplasm collection (SARDI, Adelaide, AUS)</li> <li>• Numerous phenotypically defined genes, especially mutants in nodulation/nitrogen fixation pathway</li> </ul>	<p>Ps:</p> <ul style="list-style-type: none"> <li>• USDA, Pullman, WA (n = 3,600)</li> <li>• John Innes Institute, Norwich, UK (n = 3,030)</li> </ul> <p>Lc:</p> <ul style="list-style-type: none"> <li>• USDA, Pullman, WA (n = 2,790)</li> <li>• ICARDA, Aleppo, Syria (n = 7,400)</li> </ul> <p>Ca:</p> <ul style="list-style-type: none"> <li>• USDA, Pullman, WA (n = 4670)</li> <li>• ICRISAT, Andhra Pradesh, India (n = 17,260)</li> </ul>	<p>USDA Germplasm collections (USDA-ARS, Urbana, IL):</p> <ul style="list-style-type: none"> <li>• Domestic Cultivar Collection: - "old", released prior to 1945 (208 accessions including some Asian landraces) - "modern", released by public institutions after 1945 (474 cultivars) - "proprietary" (60 cultivars)</li> <li>• Asian Vegetable Research and Development Center ~ 14,000 <i>G. max</i></li> <li>• China ~ 26,000 <i>G. max</i></li> <li>• Global Collections - <i>Glycine max</i> ~ 170,000 accessions in ~ 160 institutions in ~ 70 countries - <i>Glycine soja</i> ~ 10,000 - <i>Glycine</i> spp (perennials) ~3,500</li> </ul>	<ul style="list-style-type: none"> <li>• CIAT, Cali Colombia (n ~ 30,000): five domesticated species, wild species</li> <li>• USDA, Pullman, WA (n ~ 13,000): primarily <i>P. vulgaris</i>, wild and domesticated</li> <li>• Gatersleben, Germany</li> <li>• National Botanical Garden, Belgium: Wild Phaseoleae - Phaseolinae collection</li> </ul>	<ul style="list-style-type: none"> <li>• ICRISAT, Hyderabad, India: n ~ 14,000 <i>A. hypogaea</i>; n ~ 200 <i>Arachis</i> spp. accessions</li> <li>• CENARGEN, Brasilia, Brazil: n ~ 1000 <i>Arachis</i> spp. accessions</li> <li>• USDA, Griffin, GA: n ~ 10,500 <i>A. hypogaea</i>; n ~ 750 <i>Arachis</i> spp. accessions consisting of about 70 species</li> </ul>
Core collections					Yes	Yes

Genetic stocks		<ul style="list-style-type: none"> <li>• Primary experimental genotype "Jemalong" accepted worldwide</li> <li>• Multiple fast neutron mutant collections</li> <li>• Agrobacterium rhizogenes / RNAi-based root gene knockouts underway</li> <li>• Tnt1 transposon-based gene knockouts underway</li> <li>• Large-scale gene-tilling project</li> </ul>	<p>Ps:</p> <ul style="list-style-type: none"> <li>• USDA, Pullman, WA (Marx collection of 100 genetically defined mutations)</li> <li>• John Innes Institute, Norwich, UK (Type lines for 250 genetically defined mutations)</li> </ul>	<ul style="list-style-type: none"> <li>• Genetic Stock Collection: <ul style="list-style-type: none"> <li>- Type Collection (194 lines)</li> <li>- Isoline Collection (593 lines)</li> <li>- Germplasm Releases (178 lines maintained)</li> </ul> </li> <li>• USDA Collection: <ul style="list-style-type: none"> <li>- <i>Glycine max</i> ~ 16,750</li> <li>- <i>Glycine soja</i> ~ 1,100</li> <li>- 16 perennial species ~900</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• USDA, Pullman, WA: ~ 100</li> <li>• Genotype BAT93 accepted as standard genomics genotype</li> <li>• Broughton: Tilling project underway</li> </ul>	
Mapping populations	Few mapping populations (primarily F1, F2 and BC1) have been characterized; several are in existence and could be clonally propagated	Multiple mapping populations, primarily RILs and F2/F3	<ul style="list-style-type: none"> <li>• Ps: 20-30 RI populations, including two reference mapping populations with over 2000 molecular markers (RFLPs, RAPDs, ISSRs, microsatellites, and sequence-tagged genes)</li> <li>• Lc: at least 20 RI populations, including at least 3 derived from different</li> </ul>	Many mapping populations (BC populations, RILs, etc.) well populated with markers and publicly available of homoeologous BACs	<p>I. Core RI map: BAT93 x Jalo EEP558, some 500 RFLP markers</p> <p>II. 15-20 additional RI maps</p>	<ul style="list-style-type: none"> <li>• one <i>A. hypogaea</i> × amphidiploid species cross (370 RFLP markers mapped);</li> <li>• one diploid A-genome × A-genome cross of which few plants remain (117 RFLP and 167 RAPD markers mapped)</li> </ul>

			<p>interspecific crosses. RI populations are in USA, Australia, Israel, and Syria</p> <ul style="list-style-type: none"> <li>• Ca: numerous RI populations, including at least two derived from different interspecific crosses. RI populations being studied at ICARDA, ICRISAT, Melbourne, Horsham and Pullman</li> </ul>			
<b>Bioinformatics (for a more detailed description, see Table 2)</b>						
<b>BAC Libraries</b>						
Libraries	None	Jemalong - <i>Eco</i> RI, <i>Hind</i> III libraries, for physical mapping and genome sequencing	<ul style="list-style-type: none"> <li>• Ps: Two libraries (Pullman and Versailles)</li> <li>• Lc: none reported, although at least one is in production</li> <li>• Ca: Two libraries reported and a second as yet</li> </ul>	5: <ul style="list-style-type: none"> <li>• Williams 82 BACs (<i>Hind</i>III and <i>Bst</i>I) - USDA-ARS at Iowa and Missouri</li> <li>• Fairibault - Minnesota</li> <li>• Forrest - Southern Illinois</li> <li>• PI 437654 - Clemson</li> </ul>	5: <ul style="list-style-type: none"> <li>• Vanhouten &amp; Mackenzie: male sterility: <i>P. vulgaris</i> Sprite (Andean)</li> <li>• Kami and Gepts: phylogenetically arrayed: 3 in <i>P. vulgaris</i> (DGD1962, wild ancestral; G02771, wild Mesoamerican; BAT93, domesticated)</li> </ul>	1: Paterson, from Florunner component lines (n = 180,000)

			unpublished in Germany		• PI 229358 - Georgia	Mesoamerican), 1 in <i>P. lunatus</i> (Henderson)	
Physical map	No	FPC-based physical map consisting of ~1300 contigs and covering >95% of genome			• In development	No	No
Sequenced BACs , genome sequences		<ul style="list-style-type: none"> <li>• n ~ 155,000 BAC-end sequences</li> <li>• 102 Mbp or more than 40% of gene space sequenced (Sept. '04)</li> <li>• Chloroplast genome sequence complete</li> <li>• Nearly all repeat classes identified; repeat database underway</li> </ul>			Genomic sequence: several Mbp mostly from disease resistance loci, but one pair of homologous BACs	Individual BACs: <ul style="list-style-type: none"> <li>• Melotto et al.: around R gene cluster (Sprite)</li> <li>• Kami and Gepts: APA locus (G02771)</li> </ul>	

***Molecular Markers***

Types	<ul style="list-style-type: none"> <li>• RFLPs : 200- 300; SSRs: ~500+ (most from <i>M. truncatula</i> EST sequencing); SNPs/STSs: few; AFLPs: abundant; few isoenzymes and classical genes</li> <li>• QTL studies on aluminum tolerance, cold tolerance, biomass</li> </ul>	<ul style="list-style-type: none"> <li>• Approximately 500 FPC contigs anchored to genetic map through BAC-based SSRs and CAPS markers</li> <li>• SSRs and CAPS: ~500; AFLPs: ~300</li> </ul>	<ul style="list-style-type: none"> <li>• Ps: Broad range of biochemical and molecular markers, allozymes, RFLPs, RAPDs, AFLPs, ISSRs, STSs, RBIPs and microsa-tellites</li> <li>• Lc: Similar to <i>P. sativum</i> except fewer RFLPs,</li> </ul>		<ul style="list-style-type: none"> <li>• RFLPs : ~550</li> <li>• SSRs: ~1015</li> <li>• SNPs/STSs: ~1,000 SNP from 1,000 unigenes;</li> <li>• ~ 8,000 SNPs identified in 2,000 STSs</li> <li>• AFLPs: 650</li> <li>• many additional isoenzymes and classical genes</li> </ul>	RFLPs, RAPDs, AFLPs, ISSRs, STSs, SSRs, SNPs. Additional SSRs needed	RFLPs, RAPDs, AFLPs; 70 mapped SSRs
-------	--	---	---	--	--	--	-------------------------------------

	production, cell wall composition, and other agronomic traits have been or are being conducted		RAPDs, RBIPs and almost no microsatellites <ul style="list-style-type: none"> <li>• Ca: Similar to <i>P. sativum</i> except fewer polymorphic allozymes, RFLPs, AFLPs, RBIPs and STSs. Excellent set of microsatellites</li> </ul>			
<i>ESTs, CDS</i>						
Tissue and no.	<ul style="list-style-type: none"> <li>• ESTs: ~10,000 (S. Laberge, St. Foy, Quebec)</li> <li>• Macroarrays (S. Laberge)</li> <li>• Arrays from <i>M. truncatula</i> work well with alfalfa</li> </ul>	<ul style="list-style-type: none"> <li>• ESTs: ~190,000, constituting ~37,000 unigenes</li> <li>• 16K cDNA microarrays as well as long-oligo arrays available</li> <li>• AFFYMETRIX GeneChip in progress</li> </ul>		<ul style="list-style-type: none"> <li>• ESTs: &gt; 340,000, from more than 80 different cDNA libraries</li> <li>• AFFYMETRIX Soybean GeneChip (28,000 genes)</li> <li>• Microarrays (36,000 unigenes)</li> <li>• Long-oligo arrays in process</li> </ul>	<ul style="list-style-type: none"> <li>• <i>P. coccineus</i>: 21,000 dev. embryo</li> <li>• <i>P. vulgaris</i>: 25,000; various tissues</li> </ul>	1400 from leaves, pods
CDS		Genes in the Rhizobium and Mycorrhizae symbiosis pathway cloned and characterized			~ 1,000; disease response, seed proteins	

<i>Organizations, Consortia or Initiatives</i>							
	National Alfalfa Alliance: generally inactive relative to genomics	<ul style="list-style-type: none"> <li>• International <i>Medicago truncatula</i> Steering Committee</li> <li>• International <i>Medicago truncatula</i> Genome Sequencing Initiative</li> </ul>			United Soybean Board, Better Bean Initiative	<ul style="list-style-type: none"> <li>• Bean Improvement Cooperative: meets every other year, publishes annual report</li> <li>• Phaseomics: meets every other year; focused on genomics; has developed strategic plan</li> </ul>	<ul style="list-style-type: none"> <li>• Allergen Initiative: to eliminate peanut allergy problem. USDA, State Experiment Station, and private institution initiative.</li> <li>• Peanut Research Genomics Initiative: Cooperative research initiative among 33 U.S. scientists to develop priorities of peanut genomics and seek funding</li> </ul>

<sup>a</sup> Ps: *Pisum sativum*, pea; Lc: *Lens culinaris*, lentil; Ca: *Cicer arietinum*, chickpea

<sup>b</sup> There is some confusion in the literature on this number, which is often given as ca. 1500 Mbp. Domesticated alfalfa is a tetraploid, so its gamete (1C) is actually diploid. Because all alfalfa chromosome sets are the same (i.e., alfalfa is an autotetraploid), the true size of its genome should be one-half the size of the tetraploid genome (or the size of a diploid alfalfa genome). Hence, the true genome size is the lower value.

For further information, see Legume Crop Genomics, edited by Richard F. Wilson, H. Thomas Stalker, E. Charles Brummer, H. T. Stalker, E. C. Brummer, published by AOCS Press

**Table 2. Legume Information Resources** (Prepared by W. Beavis)

---

**Multi-Species Information Resources<sup>a</sup>**

---

**Legumes.org** (<http://www.legumes.org/>)

- portal to other legume resources

**Legume Information System** (<http://comparative-legumes.org/>)

- Medicago, Lotus, Soybean, Peanut, Lentil, Alfalfa, Pea, Phaseolus beans, Chickpea, Cowpea, [Arabidopsis]
- Linkage maps, QTL, Physical maps, ESTs, EST clusters, Consensus EST contigs, Assembled Genomic Contigs, Annotations (GO, Interpro, Blocks+), Pathways, Whole Plant Phenotypes
- BLAST, motif analyses, ESTs by library, Placement of EST Consensus contigs on Assembled Genomic Contigs, Gene Predictions, Comparative views of functional annotations, Comparative Sequence Viewer, Comparative linkage and physical maps, portal to most legume resources

**NCBI** (<http://www.ncbi.nih.gov/>)

- All species with sequence data
- Primarily data. EST, GSS and Trace data (where available); genomic clone and WGS sequence assemblies; STS and other sequence-based marker data; protein sequences (predicted and experimental), protein structure and domains; Unigene clusters for highly-represented species; limited SAGE expression data (only Medicago); limited genetic map data.(only Soybean)
- Wide variety of BLAST-based services; ePCR; similarity-based annotation of Unigene sequences, per-library analysis and Digital Differential Display of Unigene representation; mRNA-to-genomic alignment (Spidey); integrated map viewer; precomputed sequence similarities; third-party annotation

**PlantGDB** (<http://www.plantgdb.org/>)

- Medicago, Soybean, Alfalfa, Pea, Peanut, Phaseolus beans, Lupine, [Arabidopsis and many other plant species]
- Assembled EST contigs, Genomic Contigs
- BLAST, Placement of EST Consensus contigs on Assembled Genomic Contigs and GeneSeqer gene prediction

**The TIGR Plant Repeat Databases** (<http://www.tigr.org/tdb/e2k1/plant.repeats/index.shtml>)

- Medicago, Lotus, Soybean, [Arabidopsis and other plant species]
- Compilation and identification of repeat sequences in Medicago

**TIGR Gene Indices** (<http://www.tigr.org/tdb/tgi/>)

- Medicago, Lotus, Soybean, [Arabidopsis and many more plant species]
- ESTs, EST clusters, Consensus EST contigs (TC), Assembled Genomic Contigs, Annotation of TC (TOGs-tentative ortholog assignment, GO annotations, tentative annotation) Pathways, Whole Plant Phenotypes
- BLAST, ESTs by library, Placement of EST Consensus contigs on Assembled Genomic Contigs and in KEGG Pathway maps via putative EC functions, Identification of ORFs, Identification of unique

oligos, identification of alternative splice sites.

**openSputnik comparative genomics platform** (<http://sputnik.btk.fi/ests>)

- Medicago, Lotus, Soybean, *Phaseolus coccineus* (other plant species)
- ESTs, clusters and consensus sequences
- EST distribution reports by library property and functional assignment; consensus and predicted peptide functional annotation; SNP mining; BLAST services

---

### Species-specific Information Resources

---

#### Alfalfa

**AlfaGenes** (<http://ukcrop.net/perl/ace/search/AlfaGenes>)

- genetic maps, molecular markers, ESTs, pathways, molecular markers, enzymes, metabolites, phenotypes, germplasm

#### Chickpea and Lentil

**CoolGenes** (<http://ukcrop.net/perl/ace/search/CoolGenes>)

- genetic maps data for chickpea and lentil

#### Lotus

**Kazusa *Lotus japonicus*** (<http://www.kazusa.or.jp/lotus/>)

- Linkage maps, Annotated clustered ESTs, Annotated genomic clones, seed stocks, RI lines
- BLAST against TAC/BAC-end sequences

**The *Lotus japonicus* Website** (<http://www.lotusjaponicus.org/>)

- Mutant stocks
- Tilling service

**EU Project LOTUS** (<http://www.mpimp-golm.mpg.de/lotus/>)

- EU Research Project overview

The Lotus Newsletter Home Page (<http://dmz.inia.org.uy/sitios/lnl/>)

The Lotus Newsletter (volumes 23-32 1991-2001) (<http://www.psu.missouri.edu/lnl/>)

#### Medicago

**MENS - Medicago EST Navigation System** (<http://medicago.toulouse.inra.fr/Mt/EST/>)

- Annotated and Assembled ESTs, cluster and library expression patterns, maps to KEGG pathways
- BLAST and pattern scan services

**MtDB** (<http://www.medicago.org/>)

- EST clusters, linkage maps, BAC overlaps, BAC registry, genomic clone marker associations, annotated repetitive elements
- ORF prediction, cluster comparison to TIGR TCs, BLAST services, alignment against unigenes from various species, genome-to-genome comparisons against Arabidopsis

**Sequencing *Medicago truncatula* at the University of Oklahoma**

(<http://www.genome.ou.edu/medicago.html>)

- Annotated BACs from WGS assembly, BAC Clone Descriptions
- BLAST

**Medicago Genome Project at UC Davis** (<http://mtgenome.ucdavis.edu/>)

- Physical map and linkage map, BAC-end sequences, SSR, molecular markers
- Contig viewer, BLAST

**TIGR Medicago truncatula Database** (<http://www.tigr.org/tdb/e2k1/mta1/>)

- Annotated and assembled genomic contigs,
- BLAST, Gene prediction, Assigned gene ontologies

**Medicago genome database UrMeLDB** (<http://mips.gsf.de/proj/plant/jsf/medi/index.jsp>)

- Annotated and Assembled genomic contigs,

**Medicago truncatula functional genomics and bioinformatics** (<http://medicago.vbi.vt.edu/>)

- Genes, Proteins, Metabolites

**The Samuel Roberts Noble Foundation 2D-PAGE Protein Database**

(<http://www.noble.org/2DPAGE/Search.asp>)

- 2D-PAGE gel archives
- Browse and search based on multiple criteria including species, genotype/treatment, tissue source, subcellular fraction and date of analysis

**Australian National University - 2D-PAGE Database** (<http://semele.anu.edu.au/>)

- 2D-PAGE gel images
- Browse and search for 2D images from Medicago

**Mt PROTEOMICS - Medicago truncatula Proteomics site** (<http://www.mtproteomics.fr.st/>)

- Proteome analysis of *Medicago truncatula* roots in interaction either with the mycorrhizal fungus *Glomus mosseae* or the bacterium *Sinorhizobium meliloti*;
- online images of gel results

**Phaseolus Beans**

**Phaseomics: Phaseolus genome initiative** (<http://www.phaseolus.net/>)

- Overview of initiative
- Portal to bean related resources

**BeanGenes** (<http://beangenes.cws.ndsu.nodak.edu/>)

- genetic maps, molecular markers, gene classification, phenotypes, pathogens, germplasm

**BeanRef** (<http://www.neruo.it/Beanref/>)

- genetic resources, cytogenetics, physiology, phytopathology, production-consumption, databases

**Soybean**

**Soybase** (<http://soybase.org/>)

- genetic and QTL maps, molecular markers, ESTs, pathways, germplasm, phenotypes
- soybean bulletin boards

**SIU Soybean Genome** (<http://soybeangenome.siu.edu/>)

- Soybean Forrest Physical Map; FPC contigs anchored to genetic map along with QTL, EST and marker data

**Soybean Genomics Initiative** (<http://soybean.ccg.umn.edu/>)

- cDNA and BAC library info, SAGE tag data, EST clusters,
- BLAST

**A Functional Genomics Program for Soybean**

(<http://www.cbc.umn.edu/ResearchProjects/Soybean/Functional.html>)

- ESTs

**NSF Soybean Functional Genomics Project**

(<http://soybeangenomics.crops.umn.edu/>)

- Microarray project overview and results

**Soybean Genomics and Microarray Database** (<http://psi081.ba.ars.usda.gov/SGMD/Default.htm>)

- EST and analyzed microarray data

**The Public Soybean EST Project at WUSTL** (<http://genome.wustl.edu/est/index.php?soybean=1>)

- EST and Trace Data

**TAMU Soybean Genome** ([http://hbz7.tamu.edu/homelinks/phymap/soybean/soy\\_home.htm](http://hbz7.tamu.edu/homelinks/phymap/soybean/soy_home.htm))

- BAC/BIBAC-based physical map of the soybean genome from five complementary BAC and BIBAC libraries of 3 *Glycine max* cvs.

---

**Symbionts**

---

**Molecular Basis of Mycorrhizal Symbioses** (<http://www.genetik.uni-bielefeld.de/MolMyk/>)

- Project status, macro/microarrays, Annotated EST clusters

**RhizoBase- the Genome database for Rhizobia** (<http://www.kazusa.or.jp/rhizobase/>)

- Genes, maps, Annotated genome, Annotated KEGG pathways,
- Gene finding (FrameD), BLAST, PatScan

**pEXO megaplasmid** (<http://life.biology.mcmaster.ca/brian/Rhizobium/pEXO.html>)

- Physical map, Sequence and annotation of *Rhizobium meliloti*

**R. etli DataBase** (<http://www.cifn.unam.mx/retlidb/>)

- regulatory signals of transcription initiation, promoters, regulatory binding sites and terminators, as well as information on genes clustered in operons

## Other Web-enabled Legume Information Resources

---

Noble - Center for Medicago Genomics  
Medicago Genomics Group - Wageningen University Plant Sciences  
Integrated structural, functional and comparative genomics of the model legume *Medicago truncatula*  
*Medicago truncatula* Gene Nomenclature  
Medicago Genetics Group at the Institute of Genetics Biological Research Center,  
SoyCAP: Soybean Coordinated Agricultural Project  
Soybean Genomics and Improvement Laboratory (SGIL),  
National Center for Soybean Biotechnology  
Soybean Genetics Newsletter: Home Page  
Agricultural Network Information Center (AgNic)Soybean Information  
Soybean Genetics - Keim genetics lab  
Hymowitz's Soybean Genetics Lab University of Illinois at Urbana-Champaign  
North Central Soybean Research Program  
UW Soybean Research and Extension  
Soybean Sudden Death Syndrome and Cyst Nematode Research  
Agriculture and Agri-Food Canada Research Program - Soybean Genomics  
Agricultural Network Information Center (AgNic)Soybean Information  
United Soybean Board  
Soybean/Maize Germplasm, Pathology, and Genetics Research Unit  
NSRL : National Soybean Research Laboratory  
The Shoemaker Lab  
International Consortium to Sequence the Soybean Genome  
Comparative Analysis of Legume Genome Evolution  
The Gepts lab  
USDA ARS Rhizobium Germplasm Resource Center

---

<sup>a</sup> The summary should not be interpreted as a complete matrix. For example, the Legume Information System has data and information from a dozen species, but for most species the information consists of only linkage maps, QTL and mapped genetic markers. Only a few species, e.g., Medicago and Soybean have all data and information types in the LIS.

## Appendix A

List of registered participants  
CATG Meeting, December 14-15, 2004 at Sante Fe, NM

Name		Institution			Telephone	E-mail	Disc. Session	
Eric Archuleta	NCGR		2935 Rodeo Rd.	Santa Fe, NM 87505	USA	995-4419	eea@ncgr.org	b c
Bill Beavis	NCGR		2935 E Rodeo Park Dr	Santa Fe, NM 87505	USA	995-4412	wdb@ncgr.org	a c
William John Broughton	University of Geneva, LBMPs,		30 quai Ernest-Ansermet,	1211 Genève 4	Switzerland	379 3108	william.broughton@bioveg.unige.ch	b d
Edward Charles Brummer	Iowa State University	1204 Agronomy Hall	100 Osborn Dr.	Ames, IA 50011	USA	(515) 294-1415	brummer@iastate.edu	a c
Douglas Randal Cook	University of California, Davis		Department of Plant Pathology	Davis, CA 95616	USA	(530) 754-6561	drcook@ucdavis.edu	a d
Perry B. Cregan	USDA, ARS	Soybean Genomics and Improvement Laboratory	Building 006, Room 100	Beltsville, MD 20705	USA	(301) 504-5070	creganp@ba.ars.usda.gov	a d
Darleen Audrey DeMason	University of California		Botany and Plant Sciences	Riverside, CA 92521	USA	827-3580	demason@ucr.edu	b d
Jeff Joseph Doyle	Cornell University	Department of Plant Biology	228 Plant Science Building	Ithaca, NY 14853	USA	(607) 255-7972	jjd5@cornell.edu	a d
Andrew D Farmer	National Center for Genome Resources		2935 Rodeo Park Drive East	Santa Fe, NM 87505	USA	(505) 995-4464	adf@ncgr.org	b d
Michael Ernest Fromm	University of Nebraska-Lincoln	N300 Beadle Center	1901 Vine Street	Lincoln, NE 68588-0660	USA	(402) 472-2635	mfromm2@unl.edu	b c
Kamal Gajendran	National Center for Genome Resources		2935 Rodeo Park Drive East	santa fe, NM 87505	USA	(505) 995 4409	kg@ncgr.org	b c
Paul Gepts	University of California, Davis		Agronomy and Range Science	Davis, CA 95616	USA	(530) 752-7743	plgepts@ucdavis.edu	a d
Michael D Gonzales	NCGR		2935 Rodeo Park Drive East	Santa Fe, NM 87505	USA	995-4436	mg@ncgr.org	b c
David Grant	Iowa State University		G304 Agronomy Hall	Ames, IA 50011	USA	(515) 294 1205	dgrant@iastate.edu	a c
Tom D Grebb	US Dry Bean Council		PO Box 215	Quincy, WA 98848	USA	(509) 398-1790	tom@centralbean.com	a c
Michael A Grusak	USDA-ARS Children's Nutrition Research Center	Dept. of Pediatrics, Baylor College of Medicine	1100 Bates Street	Houston, TX 77030	USA	(713) 798-7044	mgrusak@bcm.tmc.edu	a c
Baozhu Guo	USDA-ARS/University of Georgia		Coastal Plain Exp. Station	Tifton, GA 31793	USA	387-2334	bguo@tifton.usda.gov	b d
Guohao He	Tuskegee University		106 Milbank Hall	Tuskegee, AL 36088	USA	727-8459	hguohao@tuskegee.edu	b d
Gina Hernández	UNAM (National University of Mexico)		Ap. Postal 565-A	Cuernavaca, MOR 62210	Mexico	(52 777) 3139877	gina@cifn.unam.mx	b d
Ann M. Hirsch	UCLA	Dept. of MCD Biology	405 Hilgard Avenue	Los Angeles, CA 90095-1606	USA	(310) 206-8673	ahirsch@ucla.edu	b d
Scott A. Jackson	Purdue University		915 W. State St.	West Lafayette, IN 47907	USA	(765) 496-3621	sjackson@purdue.edu	b d
Ed K Kaleikau	USDA-CSREES	Plant Genome Program, STOP 2241	1400 Independence Avenue SW	Washington, DC 20250-2241	USA	(202) 401-1931	ekaleikau@csrees.usda.gov	a c
David A Lightfoot	SIUC	Genomics Core Facility, Bldg 66	MC4415	Carbondale, IL 62901	USA	(618) 453-1797	ga4082@siu.edu	b c
Sally A Mackenzie	University of Nebraska	N305 Beadle Center		Lincoln, Nebraska 68588-0660	USA	(402) 472 6997	smackenzie2@unl.edu	a c
Kanyand Matand	Langston University		P.O. box 383	Langston, OK 73050	USA	(405) 466 3836	kmataand@luresext.edu	b c
Gregory D. May	The Samuel Roberts Noble Foundation		2510 Sam Noble Parkway	Ardmore, Oklahoma 73401	USA	(580) 224-6650	gdmay@noble.org	b d
Phillip E. McClean	North Dakota State University	Department of Plant Sciences	Loftsgard Hall	Fargo, North Dakota 58105	USA	231-8443	phillip.mcclean@ndsu.nodak.edu	a c
Mary H Montoya	NCGR		2935 Rodeo Park Drive East	Santa Fe, NM 87505	USA	(505) 995		b c
Fred J Muehlbauer	USDA-ARS	Washington State University	303 Johnson Hall	Pullman, WA 99164-6434	USA	(509) 335-9521	muehlbauer@wsu.edu	a d
Henry T. Nguyen	University of Missouri	National Center for Soybean Biotechnology	1-87 Agriculture Building	Columbia, MO 65211	USA	(573) 882-5494	nguyenhenry@missouri.edu	b c
Peggy Ozias-Akins	University of Georgia	Tifton Campus	PO Box 748	Tifton, GA 31793-0748	USA	(229) 386-3902	ozias@tifton.uga.edu	a c
Eran Pichersky	University of Michigan	Molecular, Cellular, and Developmental Biology	830 N. University Street	Ann Arbor, MI 48109	USA	(734) 9363522	leix@umich.edu	b d

Tim Porch	USDA/ARS/TARS		2200 P.A. Campos Ave, Suite 201	Mayaguez, PR 00680	USA	831-3435x254	maytp@ars-grin.gov	a	c
Ian McCandliss Ray	New Mexico State University	Dept. of Agronomy & Horticulture	MSC 3Q, PO Box 30003	Las Cruces, New Mexico 88003	USA	646-3819	iaray@nmsu.edu	a	c
Ernest F. Retzel	University of Minnesota / Center for Computational Genomics	664 Children's Rehabilitation Center	426 Church St. S.E.	Minneapolis, MN 55455	USA	(612) 626-0495	ernest@ccgb.umn.edu	b	d
Deborah A Samac	USDA-ARS and University of Minnesota	495 Borlaug Hall	1991 Upper Buford Circle	St. Paul, MN 55108	USA	625-1243	dasamac@umn.edu	b	d
Randy C. Shoemaker	USDA-ARS	Iowa State University	Department of Agronomy	Ames, Iowa 50011	USA	294-6233	rcsshoe@iastate.edu	b	c
Susan Rundell Singer	Carleton College	Department of Biology	One North College St.	Northfield, MN 55057	USA	(507) 646-4391	ssinger@carleton.edu	b	d
James Eugene Specht	University of Nebraska	Dep. of Agronomy & Horticulture	322 Keim Hall	Lincoln, NE 68583-0915	USA	(402) 472-1536	jspecht1@unl.edu	a	c
Gary Stacey	University of Missouri		271E Life Science Center	Columbia, MO 65211	USA	(573) 884-4752	staceyg@missouri.edu	b	d
H. Thomas Stalker	N. C. State University	Department Crop Science	Box 7629	Raleigh, NC 27695	USA	(919) 515-2647	Tom_Stalker@ncsu.edu	a	c
Fedora Sutton	NSF		4201 Wilson Blvd Suite 615	Arlington, VA 22230	USA	292-7166	fsutton@nsf.gov	b	c
Ann Marie Thro	USDA3443	Waterfront Building	800 9th Street, SW	Washington, DC 20024	USA	401-6702	athro@csrees.usda.gov	b	c
Chris Town	The Institute for Genomic Research		9712 Medical Center Drive	Rockville, MD 20850	USA	(301) 795-7523	cdtown@tigr.org	a	d
Norman F. Weeden	Montana State University	Dept. Plant Sciences and Plant Pathology	119 ABS Facility	Bozeman, Montana 59717	USA	994-7622	nweeden@montana.edu	a	d
Richard Ferrol Wilson	USDA-Agricultural Research Service	Room 4-2214 George Washington Carver Center	5601 Sunnyside Avenue	Beltsville, Maryland 20705-5139	USA	504-4670	rww@ars.usda.gov	b	c
Martin (Marty) F. Wojciechowski	Arizona State University	School of Life Sciences		Tempe, AZ 85287-4501	USA	727-7767	mwojciechowski@asu.edu	a	d
Eve Syrkin Wurtele	Iowa State	GDCB	441 Bessey Hall	Ames, IA 50014	USA	(515) 294-8989	mash@iastate.edu	b	c
Nevin D. Young	University of Minnesota	495 Borlaug Hall	1991 Upper Buford Circle	St. Paul, MN 55108	USA	(612) 625-9728	nevin@umn.edu	b	d
Dani Zamir	The Hebrew University		P.O. Box 12	Rehovot, 76100	Israel	972-8-9489092	zamir@agri.huji.ac.il	a	d