Three Sequenced Legume Genomes and Many Crop Species: Rich Opportunities for Translational Genomics

Steven B. Cannon*, Gregory D. May, and Scott A. Jackson

United States Department of Agriculture-Agricultural Research Service, Corn Insects and Crop Genetics Research Unit, Ames, Iowa 50011 (S.B.C.); National Center for Genome Resources, Santa Fe, New Mexico 87505 (G.D.M.); and Department of Agronomy, Purdue University, West Lafayette, Indiana 47906 (S.A.J.)

This year marks the essential completion of the genome sequences of soybean (*Glycine max*), barrel medic (*Medicago truncatula*), and birdsfoot trefoil (*Lotus japonicus*). The impact of these assembled, annotated genomes will be enormous. Birdsfoot trefoil and barrel medic, both forage crops, are the preeminent laboratory plants used in legume research. Monetarily, soybean is the most valuable protein and edible oil crop in the world, and serves as a model for seed and other developmental processes. These genome sequences contain the vast majorities of gene and regulatory sequences for these plants, as well as information about evolutionary histories over the approximately 54 million years (Mya) since their common ancestor. These genome sequences are made more useful by virtue of the ability to compare between the genomes, and to transfer information from these biological models to other crop species and vice versa. This review will describe the basic characteristics of the sequenced legume genomes, and will highlight examples, opportunities, and challenges for translational genomics across the legumes.

TRANSLATIONAL GENOMICS IN THE LEGUMES

Considering the large number of domesticated legume species (for review, see Graham and Vance, 2003), the potential for translational genomics may be greater in the legumes than in any other plant family. By virtue of their symbiotic associations with nitrogen-fixing bacteria, many of legumes produce seeds unusually rich in protein and oil, as well as a range of secondary metabolites. The relative abundance available of nitrogen for legumes lessens a key constraint that plants face, and arguably increases their chances of adapting in more challenging environments. Examples include drought-tolerant *Acacia* or *Prosopis* species (e.g. honey mesquite [*Prosopis glandulosa*]) in desert ecosystems, or tuberous-rooted *Psoralea* species (e.g. breadroot [*Psoralea esculenta*]) in the dry North American short-grass plains, or the cold-hardy alpine *Lupinus* or *Astragalus* species in alpine or high-desert ecosystems, or the stem-nodulating *Sesbania* species as green manures around waterlogged rice (*Oryza sativa*) paddies (Table I).

It is therefore not surprising that humans in every environment have explored and partially domesticated many species of legumes. Hence, there are many crop and fodder species that could benefit from modern breeding efforts. In a time of rapid climate changes, such species may have a role in agriculture of the future due to drought tolerance, water-use efficiency, tolerance of marginal or eroded soils, resistance to diseases and pests, tolerance of waterlogging or flooding, or ability to penetrate compacted soils.

A key rationale for the expensive endeavor of sequencing the genome of a model plant species is that knowledge from that genome can be transferred to a related—and perhaps more genetically intractable—crop species. Three examples illustrate the application of translational genomics in the legumes.

First, Yang et al. (2008) reported the use of barrel medic to map-base clone the *RCT1* gene that confers resistance to multiple races of anthracnose (*Colletotrichum trifolii*) in alfalfa (*Medicago sativa*). This nicely illustrates the use of a model to investigate a trait that would be difficult to dissect genetically in alfalfa, which is tetraploid and predominantly outcrossing.

Second, information about floral regulatory genes identified in Arabidopsis (*Arabidopsis thaliana*) was used to find a gene probably responsible for the determinacy trait in common bean (*Phaseolus vulgaris*; Kwak et al., 2008). The *TERMINAL FLOWER1* (*TFL1*) mutant in Arabidopsis produces a terminal floral meristem. The ortholog of *TFL1* in *Phaseolus* segregates uniformly with the determinate trait in crosses between vining and bush beans (Kwak et al., 2008).

Third is the identification of the gene underlying Mendel's *I* locus, responsible for the traits of yellow or green seed (Armstead et al., 2007). The stay-green trait was observed and mapped in grass meadow fescue; a candidate gene was then identified in rice by synteny analysis; the mutant in the orthologous gene was then tested for phenotype tests in Arabidopsis; and finally, the orthologous gene was fine mapped in pea (*Pisum sativum*), in mapping populations segregating for cotyledon color polymorphisms (Armstead et al., 2007). Thus, traits, genes, tools, and species were marshaled in linking this trait and the underlying gene in several models and crops.

* Corresponding author; e-mail steven.cannon@ars.usda.gov.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Steven B. Cannon (steven.cannon@ars.usda.gov).

www.plantphysiol.org/cgi/doi/10.1104/pp.109.144659

Table I. *Selected food and model legumes*

Crop legume species are grouped in the three classical legume subfamilies: Caesalpinioideae, Mimosoideae, Papilionoideae; and then by clade and tribe. Sequenced legume genomes are underlined. Model or major crop legumes are in bold text. Primary uses: s = seed; t = tuber or root; p = pod or pod wall; l = leaf; f = forage; m = model. D = Drought tolerant, C = cold tolerant, or F = flooding tolerant; P = perennial. *, Varieties may contain toxins (alkaloids or cyanogenic glycosides) removable in preparation.

Clade	Tribe	Binomial	Common Name	Uses	Note
Cercidae	Cercideae	*Tylosema esculentum*	Marama bean	s, t	D, P
Detarieae	Detarieae	*Detarium senegalense*	Sweet detar	s	P
Detarieae	Detarieae	*Tamarindus indica*	Tamarind	p	P
Umtiza	Caesalpinieae	*Ceratonia siliqua*	Carob	s, p	P
Caesalpinieae	Caesalpinieae	**Chamaecrista fasciculata**	Partridge pea	m	D, P
Caesalpinieae	Caesalpinieae	*Cordeauxia edulis*	Yeheb nut	s	D, P
Mimosoid	Mimoseae	*Parkia speciosa*	Petai	s, p, f	D, P
Mimosoid	Mimoseae	*Prosopis glandulosa*	Honey mesquite	s, p, f	D, P
Mimosoid	Mimoseae	*Desmanthus illinoensis*	Illinois bundleflower	s, f	D, P
Mimosoid	Mimoseae	*Inga edulis*	Ice-cream bean	p	P
Indigoferoid	Indigofereae	*Cyamopsis tetragonoloba*	Guar/cluster bean	s, p, f	
Genistoid	Genisteae	*Aspalathus linearis*	Rooibos tea	l	D, P
Genistoid	Genisteae	*Lupinus albus*	**White lupin**	s	*
Genistoid	Genisteae	*Lupinus angustifolius*	Narrow-leaved lupin	s	*
Genistoid	Genisteae	*Lupinus luteus*	**Yellow lupin**	s	*
Genistoid	Genisteae	*Lupinus mutabilis*	Andean lupin; tarwi	s	C, *
Genistoid	Genisteae	*Lupinus polyphyllus*	Washington lupin	s, f	C, P, *
Dalbergioid	Aeschynomeneae	*Arachis hypogaea*	**Peanut/groundnut**	s	
Galegoid	Galegeae	*Glycyrrhiza glabra*	Licorice	t	P
Galegoid	Hedysareae	*Caragana arborescens*	Pea shrub	s, p	D, C, P
Galegoid	Cicereae	*Cicer arietinum*	**Chickpea**	s	
Galegoid	Trifolieae	*Trigonella foenum-graecum*	Fenugreek	s	
Galegoid	Trifolieae	**Medicago truncatula**	Barrel medic	f, m	
Galegoid	Vicieae/Fabeae	*Lathyrus sativus*	Grass pea/chickling vetch	s, f	D, *
Galegoid	Vicieae/Fabeae	*Lens culinaris*	**Lentil**	s	
Galegoid	Vicieae/Fabeae	**Pisum sativum**	**Pea**	s, p, f, m	
Galegoid	Vicieae/Fabeae	*Vicia faba*	**Fava bean/broad bean**	s	
Robinioid	Loteae	*Lotus tetragonolobus*	Asparagus pea	p	
Robinioid	Loteae	**Lotus japonicus**	Birdsfoot trefoil	f, m	P
Robinioid	Sesbanieae	*Sesbania* spp.	Agati	f, l, s, p	F, P
Millettioid	Phaseoleae	*Psoralea* spp.	Breadroot, prairie turnip	t	D, P
Millettioid	Phaseoleae	*Apios americana*	Potato bean; groundnut	t	P
Millettioid	Phaseoleae	*Cajanus cajan*	**Pigeonpea**	s, p	D, P
Millettioid	Phaseoleae	*Canavalia ensiformis*	Jack bean/velvet bean	s, p, f	*
Millettioid	Phaseoleae	*Lablab purpureus*	Hyacinth bean	s, p, f	
Millettioid	Phaseoleae	**Glycine max**	**Soybean**	s, m	
Millettioid	Phaseoleae	*Pachyrhizus erosus*	**Jicama/yam bean**	t	
Millettioid	Phaseoleae	*Phaseolus coccineus*	Scarlet runner bean	s, p	
Millettioid	Phaseoleae	*Phaseolus lunatus*	Lima bean	s	*
Millettioid	Phaseoleae	**Phaseolus vulgaris**	**Common bean**	s, p	
Millettioid	Phaseoleae	*Phaseolus acutifolius*	Tepary bean	s, p	D
Millettioid	Phaseoleae	*Macrotyloma geocarpum*	Hausa groundnut	s	D
Millettioid	Phaseoleae	*Psophocarpus* spp.	Winged bean	p, t	
Millettioid	Phaseoleae	*Vigna angularis*	**Adzuki bean**	s	
Millettioid	Phaseoleae	*Vigna aconitifolia*	Moth bean	s	
Millettioid	Phaseoleae	*Vigna mungo* and *radiata*	**Black gram; mung bean**	s	
Millettioid	Phaseoleae	*Vigna subterranea*	Bambara groundnut	s	D
Millettioid	Phaseoleae	*Vigna unguiculata*	**Cowpea/black-eyed pea**	s, p	

MANY MODELS IN THE LEGUMES

Although *Medicago* and *Lotus* are often considered the primary biological models in the legumes, it may be more helpful to think of many models, each making critical contributions to a body of knowledge about legumes as a semiunified genetic system. *Medicago* and *Lotus* will serve as effective models for the cool-season legumes (Young and Udvardi, 2009) and soybean for the many crop species in the Phaseoleae (Gepts et al., 2005). However, these are not one-way streets, since the crops also inform the genomic models.

For example, *Medicago* and *Lotus* have been vigorously utilized in studies of nodulation, mycorrhiza-

tion, and plant-symbiont signaling (for review, see Oldroyd and Downie, 2008). Barrel medic has been used for study of phenylpropanoid and isoflavonoid pathways and secondary metabolites (Farag et al., 2008), various defense responses (Yang et al., 2008), root architecture (Gonzalez-Rizzo et al., 2006), and aluminum tolerance (Chandran et al., 2008). Soybean has been a model for studies of seed development (for review, see Vodkin et al., 2008), root hair development and early nodulation responses, mineral uptake, and protein and oil biosynthesis. The garden pea is in some senses the founding model for genetics, through Mendel's work in 1866. Pea has been further developed as a model for development traits such as compound leaf form (Tattersall et al., 2005), tendril formation (Hofer et al., 2009), transposon evolution (Jing et al., 2005), defense responses (Gao et al., 2004), starch and sugar synthesis (Barratt et al., 2001), and hormonal control of shoot branching (Beveridge et al., 2009). Common bean has been used to characterize the molecular basis of photoperiod sensitivity and determinacy (Kwak et al., 2008). Scarlet runner bean (*Phaseolus coccineus*), with its very-large seeds, has been developed as a model to study embryo and seed development (Kawashima et al., 2009). The mimosoid legume *Mimosa pudica* (the sensitive plant), whose leaves and petioles respond to touch by closing (thigmonasty), have been studied to determine the basis for their rapid responses (Uehlein and Kaldenhoff, 2008). The leguminous tree honey locust (*Gleditsia triacanthos*) is being used as a model for studying rhizobia in a nonnodulating but nitrogen-fixing species (Lee and Hirsch, 2006). The nodulating prairie legume partridge pea (*Chamaecrista fasciculata*) is being studied both as a model of ecological adaptation to climate change (Shaw et al., 2008) and of floral structure (Tucker, 2003), and to help determine timing of polyploidy early in the legumes (Singer et al., 2009).

TAXONOMIC POSITIONS OF THE SEQUENCED AND AGRONOMICALLY IMPORTANT LEGUMES

The ability to transfer knowledge between species depends on both the evolutionary distances between species, and the rate and nature of changes in the genomes over time. The legumes are, in fact, an old family. For reference, the common ancestor of soybean and pea, estimated at approximately 54 Mya (Lavin et al., 2005), predates by 10 Mya the earliest complete primate fossil, *Darwinius masillae* (Franzen et al., 2009). That early small primate lived in a paratropical rainforest already richly populated by leguminous trees (Engelhardt, 1922).

The legumes are also diverse, with around 20,000 species and 700 genera (Doyle and Luckow, 2003; Lewis et al., 2005). The legumes have traditionally been placed into the three subfamilies: the Papilionoideae, with approximately 70% of species; the Mimosoideae, with approximately 15%; and the remainder in the Caesalpinoideae—though this last subfamily is now known to be comprised of a collection of early diverging legume taxa (Doyle and Luckow, 2003; Lewis et al., 2005).

The papilionoid subfamily includes most crop legumes and the major model legume species, and thus is the taxonomic space across which much of legume comparative genomics and translational genomics will take place (Table I; Fig. 1). Most legume species of agronomic interest fall within four large subdivisions in the Papilionoideae: the galegoid, millettioid, dalbergioid, and genistoid clades (Doyle and Luckow, 2003; Lewis et al., 2005). The papilionoid origin is dated at approximately 59 Mya (Lavin et al., 2005).

The galegoid clade contains the robinioid clade, with birdsfoot trefoil and several allied forage and tree legumes (including *Sesbania*, and *Robinia*, e.g. the black locust tree); and the inverted-repeat-loss clade (IRLC), with barrel medic and the cool-season legumes, including clovers (*Trifolium* spp.), sweetclovers (*Melilotus* spp.), vetches (*Vicia* spp.), pea, chickpea (*Cicer*

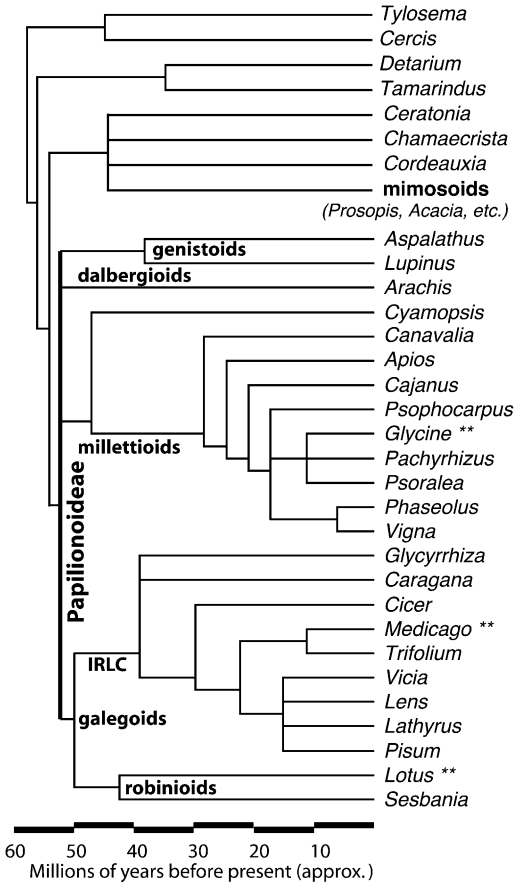

Figure 1. Taxonomic relationships among selected legume genera. Each genus contains one or more food crops, or a genomic model (*Medicago, Lotus, Chamaecrista*). Approximate inferred speciation dates follow the timings in Lavin et al. (2005). The phylogeny is after Lavin et al. (2005) and Lewis et al. (2005). Common names and uses are in Table I.

arietinum), fava (*Vicia* spp.), lentil (*Lens culinaris*), and alfalfa. The galegoid clade is dated at approximately 51 Mya, and the robinioid and IRLC at approximately 48 and 39 Mya, respectively (Lavin et al., 2005).

The millettioid clade contains the Phaseoleae tribe, with common beans, soybean, and cowpea (*Vigna unguiculata*), pigeonpea (*Cajanus cajan*), mungbean (*Vigna radiata*), adzuki bean (*Vigna angularis*), tepary bean (*Phaseolus acutifolius*), lima bean (*Phaseolus lunatus*), and hyacinth bean (*Lablab purpureus*). This clade also has several less widely known genera that have been used for food and probably at least partly domesticated: the African Bambara groundnut (*Vigna subterranea*) and Hausa groundnut (*Macrotyloma geocarpum*); jicama (*Pachyrizus erosus*); American groundnut (*Apios americana*); and American breadroot or prairie turnip (*Pediomelum*). The Bambara and Hausa groundnuts are interesting evolutionarily: they have evolved a pod-burying mechanism similar to that seen in the independently evolved trait in peanut (*Arachis hypogaea*). Similarly, jicama and breadroot have apparently independently evolved root storage organs. All are of agronomic interest because of their high drought tolerance. The millettioid origin is dated at approximately 52.8 Mya (Lavin et al., 2005), with *Phaseolus* and *Glycine* separating at approximately 19 Mya (Lavin et al., 2005).

The dalbergioid clade contains numerous tropical trees (e.g. rosewood [*Dalbergia* spp.]), as well as peanut. The genistoid clade includes many tropical and temperate genera, including lupins (*Lupinus* spp.), several species of which have been independently domesticated in both the old and new worlds.

Outside of the papilionoid legumes are numerous early diverging clades. The largest of these, the mimosoid and allied clades (including some traditionally placed in the caesalpinoid subfamily), includes the genera *Acacia*, *Prosopis*, and *Parkinsonia*—all of which are dominated by drought-tolerant trees and shrubs. Additionally, all include species that have been used for human food and animal fodder. *Acacia* are key species in African savannas, and *Prosopis* (e.g. honey mesquite) and *Parkinsonia* (e.g. paloverde) fill a similar niche in southwestern North American deserts and scrublands. The Illinois bundleflower (*Desmanthus illinoensis* Michx.) is a North American prairie legume under evaluation as a perennial seed crop (Vail et al., 1992). The North American prairie legume partridge pea is being used as a model species for investigating evolutionary and developmental patterns in an early diverging legume. *Chamaecrista* (Singer et al., 2009) has traditionally been classified in the caesalpinoid legume subfamily, but now is seen as either part of, or sister to, the mimosoid group, along with several tropical or subtropical legumes that have been used for food or medicine: carob (*Ceratonia siliqua*) and senna leaf (*Senna* spp.).

Other remaining early diverging clades with species of agronomic interest are the Detarieae clade, including the Indian tamarind (*Tamarindus indica*) and sweet detar (*Detarium senegalense*); and two highly drought-tolerant southern African perennials: the nut-producing yeheb nut shrub (*Cordeauxia edulis*), and the nut- and tuber-producing vine marama bean (*Tylosema esculentum*). Both the yeheb nut and marama bean have considerable potential as food crops for arid regions (Graham and Vance, 2003). The marama bean was the second-most-important plant foots of the !Kung bushmen in the Kalahari, with nutritional qualities similar to peanut and soybean, but thriving in harsh desert conditions (Vietmeyer, 1978). Similarly, the yeheb nut has been a staple food of nomadic groups in Somalia and Ethiopia.

BASIC CHARACTERISTICS OF THE SEQUENCED LEGUME GENOMES

The estimated size of the soybean genome is 1,115 Mb. The current assembly (Glyma1.01, available at http://www.phytozome.net) consists of 950 Mb in 20 chromosome pseudomolecule sequences, and 23 Mb in additional smaller, unanchored scaffold sequence assemblies (Soybean Genome Sequencing Consortium, http://www.phytozome.net/soybean.php).

An important feature of the soybean genome—one that was known before the sequencing project began, on the basis of marker and cytogenetic information and targeted sequencing—is that the genome underwent polyploidy approximately 13 Mya (for review, see Shoemaker et al., 2006). Further, soybean and other papilionoid legumes show evidence of an older, shared duplication. The older duplication, dated using a combination of rates of synonymous-site changes (Ks values) and fossil-calibrated species phylogenies in the legumes, is estimated to have occurred at approximately 59 Mya (Pfeil et al., 2005; Schlueter et al., 2007). These duplications are widely evident, both in the number of highly similar duplicated genes, and in large areas of correspondence (synteny) between chromosomal regions.

The estimated size of the *Medicago* genome is between 471 and 583 Mb in size (Medicago Genome Sequence Consortium, 2007). The Mt2.0 genome assembly (from 2007) is 240 Mb, excluding between-scaffold gaps. This assembly differs from the soybean genome sequence in that the Medicago Genome Sequence Consortium used a bacterial artificial chromosome (BAC)-by-BAC sequencing approach, and explicitly focused on sequencing the euchromatic chromosome arms. Remaining repetitive sequence has been sampled by lower-coverage high-throughput sequencing. The 3.0 assembly, anticipated for release in late 2009, is 277 Mb and estimated to span 80% of the euchromatin (Nevin Young, personal communication). Smaller assemblies generated by the lower-coverage sample sequencing will also be available though will not be included in the BAC-based pseudomolecules.

The estimated size of the *Lotus* genome is 472 Mb (Sato et al., 2008). The *Lotus* sequencing project is

proceeding with a similar strategy as that of *Medicago*: clone-by-clone, with additional whole-genome shotgun sequence for additional coverage. Total coverage reported for the 1.0 assembly, including both clone-by-clone and selected whole-genome shotgun assemblies, was 314.1 Mb, or 67% of the genome size (Sato et al., 2008). Considering the clone-by-clone coverage used in the chromosome pseudomolecules and not counting gaps, the coverage is 175 Mb. Coverage in the next assembly is expected to rise to approximately 200 Mb within the pseudomolecules (Shusei Sato, personal communication).

EXTENT OF SYNTENY IN THE LEGUMES

Most genes in a papilionoid legume species are likely to be found within a relatively large (several hundred kb to several Mb) syntenic region with respect to any other given papilionoid species. This is good news for positional cloning: If a gene and phenotype are experimentally associated in one species, then an orthologous gene is likely to be found in a similar neighborhood in another legume species. Between barrel medic and birdsfoot trefoil, for example, approximately 10 large-scale blocks contain the majority of the euchromatic space of each genome (Cannon et al., 2006). Many of the same blocks are also conserved in peanut (Bertioli et al., 2009). However, comparisons of either *Medicago* or *Lotus* to themselves show limited synteny (Cannon et al., 2006), suggesting substantial rearrangement shortly following the early legume polyploidy—an event that probably predated at least the separation of the genistoid (e.g. *Lupinus*) and millettioid (e.g. bean and pea) clades (Bertioli et al., 2009).

SOME SIMILARITIES AND DIFFERENCES AMONG THE GENOMES

The legume genome sequences have uncovered both striking similarities and differences. Since articles are in preparation for each of the genome projects, it is inappropriate to describe the genome features in detail, but several features have been evident from early versions of the assemblies (Cannon et al., 2006; Medicago Genome Sequence Consortium, 2007; Innes et al., 2008; Sato et al., 2008).

Among the similarities are that gene densities in euchromatic regions are generally similar in the three genomes, despite the relatively large differences in the genome sizes. The genome size differences are due primarily to two factors. First, much of the genomic DNA generated in the 13 Mya *Glycine* duplication remains. This means that any given *Medicago* or *Lotus* region is likely to correspond well with two *Glycine* regions. In many syntenic regions, the gene densities are similar (within approximately 30%; Mudge et al., 2005; Cannon et al., 2006; Innes et al., 2008). The second factor producing different genome sizes is the size of the repeat-rich, gene-poor, recombination-suppressed pericentromeres in the three genomes. In soybean, these are remarkably large, comprising nearly two-thirds of the total genome sequence (Schmutz et al., 2009).

An intriguing difference between the sequenced legume genomes is that the centromeres do not, in general, appear to correspond. This highlights that the pericentromeric regions are relatively more labile than the euchromatic regions. The pericentromeres are evidently capable of shifting locations, and expanding or contracting. It is not yet clear to what extent shifts of location may be due to breakages or fusions, as opposed to the adoption of new centromere locations within intact chromosomes.

TRANSFORMATIVE TECHNOLOGIES MAKING USE OF GENOME SEQUENCES

The genome sequences of *Glycine*, *Medicago*, and *Lotus* will be important tools for basic research in these species, particularly when combined with other new genomics technologies. Primary among the new genomics technologies are high-throughput sequencing, which enable essentially complete, high-resolution transcriptome characterization and genome resequencing. Other transformative technologies include highly multiplexed mapping techniques, which can produce dense maps in essentially single reactions, and can genotype thousands of cultivars or ecotypes (Rostoks et al., 2006). The genome sequences and new genomic technologies are enabling more rapid exploration of the broad range of molecular and organismal biology in these species, including diversity characterization, gene and regulatory-site identification, genome structure and change, and plant development and physiology.

High-throughput sequencing technologies facilitate the resequencing of legume species with reference genomes. Both *Medicago* and soybean have very large resequencing and haplotype map (HapMap) projects under way (N. Young and S. Jackson, respectively, personal communication). Taking the *Medicago* HapMap project as an illustration, 384 diverse *Medicago* genetic lines will be resequenced using the Illumina paired-end read sequencing platform. Association-mapping techniques will be used to create a HapMap. An extensive collection of single nucleotide polymorphisms (SNPs) between *Medicago* lines will enable the prediction of genome segments with shared ancestry. These haplotypes can then be associated statistically with variation in traits of interest.

Next-generation sequencing technologies may have their greatest impact on minor crop or emerging model legume species (e.g. *Chamaecrista*). Researchers working in legume crops such as chickpea, pigeonpea, common bean, and alfalfa have been early adapters of these technologies with a goal toward understanding

population-level genetic diversity and the development of molecular (SNP) markers. In some instances, little or no genomic data exist for these species. A typical experiment involves the establishment of a reference sequence using Roche 454 RNA sequencing of a selected reference plant using a pooled or normalized RNA sample. Populations or cultivars are then characterized through Illumina RNA sequencing. These Illumina sequence reads are aligned to either the 454 reference or to the existing legume genome sequences to determine both genetic diversity and transcript abundance differences between individuals or populations. As a result, informatics databases, such as the Legume Information System (www.comparativelegumes.org), already accommodate short-read DNA sequences and facilitate intra- and interlegume species comparisons.

In essence, what we are witnessing is the democratization of legume genomics. That is, in a matter of weeks and for relatively little cost, individual laboratories alone can provide plant breeders and the scientific community with data resources and research tools that were once reserved for model or major crop species.

WHERE TO GO FROM HERE?

Looking forward at what can and should be done to leverage the investment in the sequencing of the three legume genomes, we reflect on these key areas: basic plant biology; legume evolution and domestication; and plant improvement through better use of translational information, more sophisticated selection methods, and development of hybrids.

The plant biological process that can uniquely be addressed in legumes is nodulation. Many nodulation genes have been cloned in the legumes, but having the entire genome and all the genes for three legumes will allow access to the entire repertoire of nodulation-related genes. Tools to elucidate function, including insertion lines, RNAi, TILLING, and others, must continue to be developed (see related reports, this issue). In addition to nodulation, soybean seed development and oil biosynthesis can now be dissected at a much higher genetic resolution. It will be interesting to see how these fields unfold in the next few years.

Genome sequences will enable researchers to better understand legume evolution and domestication, and to examine how genes change within a genomic context. Polyploidy is a recurring event in the legumes. Within the *Glycine* genus, there have been polyploidy events at 59 and 13 Mya; and in some *Glycine* species, an event at approximately 50,000 years ago (Doyle et al., 2003). Thus, there is an opportunity to study the effects of polyploidy on genome structure, gene function, and the process of recurrent diploidization. Parallel to evolution is domestication. Many legumes have undergone human domestication, some more than once (Kwak and Gepts, 2009). Thus, with three nearly complete genome sequences and others yet to come, we can begin to understand the genomic architecture of domestication and what genes/traits are commonly selected for during domestication (e.g. shattering) versus those that may be particular to certain lineages.

Genome sequences can be leveraged to improve orphan crops. Many legumes are important regional food sources, but investment in genetic and genomic tools is limited—for example, common bean, pigeonpea, cowpea, and lentil (Varshney et al., 2009). Beyond these species, there are other even less well characterized but worth investigating, such as the drought-tolerant perennial African yeheb nut, or the cold-tolerant Andean "tarwi" lupin (*Lupinus mutabilis*). However, molecular markers, even gene-based markers, can be developed in reference genomes and then deployed in the related species. For instance, gene-based markers can be developed for pigeonpea using the *Medicago* genome as a proxy by aligning pigeonpea ESTs to the soybean genome and then develop intron-spanning primers to screen for polymorphisms in pigeonpea. This approach has been used to develop legume anchor markers that have been used in common bean, birdfoot trefoil, barrel medic, and *Arachis* (Hougaard et al., 2008), and in six crop or model legume species from the Phaseoleae and galegoid clades (Choi et al., 2006). Another approach for utilizing genome sequences for orphan crops is to sequence the orphan genome using next-generation sequencing approaches that will not result in reference-level sequences, but will be able to use the reference genomes (soybean, *Lotus*, or *Medicago*) to help assemble or organize the sequence scaffolds.

Crop improvement can also be facilitated with better molecular tools for selection in breeding. Resequencing is a cost-effective approach to find thousands of SNPs that can then be used to develop or integrate genomic selection tools that hold the promise of accelerating breeding programs. In addition, the resolution offered by the vast number of markers from such marker programs allows far more precise backcrossing than before and should temporally accelerate such programs.

Lastly, there is room for substantial legume crop improvement through enhanced understanding of the molecular basis of hybrid vigor. Hybrid vigor is the basis of much of the yield increases in maize (*Zea mays*) over the last 100 years, and more recently in rice (Garcia et al., 2008). In legumes, there have been few examples of hybrid vigor in theory or in practice. Pigeonpea is a recent example (Kalaimagal et al., 2008). Having access to all the genes should allow scientists to begin to understand the basis of hybrid vigor and to begin to exploit it in other species to help meet the global challenge of feeding the world. Other barriers to harnessing hybrid vigor include the challenges of developing male-sterile lines, and managing pollination biology in crops that are insect pollinated. Both have been challenges in soybean (Palmer et al., 2003).

CONCLUSION

The legumes are remarkably well positioned in the genomic era. There are three essentially completed genome sequences in species related to large numbers of crops and forages. Extensive genetic and genomic tools have been developed for many crops and models. A great deal of work remains: to characterize more genes and traits, to better determine correspondences across the genomes, and to extend new genomic tools to orphan species. Some of the most critical work does not rely on new high-throughput sequencing or genomic technologies. This includes characterizing and managing germplasm collections and breeding lines in many species; developing mapping populations for various traits of interest in less-studied species; working with indigenous farmers to ensure that the products of centuries of plant domestication are not lost; investigating protocols for hybrid seed production in various legumes; and working to maintain and develop understudied legumes for use in diverse, challenging growing environments around the globe. With energy supplies diminishing and higher nitrogen prices inevitably following, and the human population rising while the amount of arable land declines or degrades, legume researchers now have both great opportunity and responsibility to help develop crops for a changing world.

ACKNOWLEDGMENTS

We thank Nevin Young, Shusei Sato, Randy Shoemaker, and Andrew Severin for helpful discussions about the manuscript.

Received July 11, 2009; accepted September 14, 2009; published September 16, 2009.

LITERATURE CITED

Armstead I, Donnison I, Aubry S, Harper J, Hortensteiner S, James C, Mani J, Moffet M, Ougham H, Roberts L, et al (2007) Cross-species identification of Mendel's *I* locus. Science **315:** 73

Barratt DH, Barber L, Kruger NJ, Smith AM, Wang TL, Martin C (2001) Multiple, distinct isoforms of sucrose synthase in pea. Plant Physiol **127:** 655–664

Bertioli DJ, Moretzsohn MC, Madsen LH, Sandal N, Leal-Bertioli SC, Guimaraes PM, Hougaard BK, Fredslund J, Schauser L, Nielsen AM, et al (2009) An analysis of synteny of *Arachis* with *Lotus* and *Medicago* sheds new light on the structure, stability and evolution of legume genomes. BMC Genomics **10:** 45

Beveridge CA, Dun EA, Rameau C (2009) Pea has its tendrils in branching discoveries spanning a century from auxin to strigolactones. Plant Physiol **151:** 985–990

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

Chandran D, Sharopova N, VandenBosch KA, Garvin DF, Samac DA (2008) Physiological and molecular characterization of aluminum resistance in *Medicago truncatula.* BMC Plant Biol **8:** 89

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

Doyle JJ, Doyle JL, Rauscher JT, Brown AHD (2003) Diploid and polyploid reticulate evolution throughout the history of the perennial soybeans (Glycine subgenus Glycine). New Phytol **161:** 121–132

Doyle JJ, Luckow MA (2003) The rest of the iceberg: legume diversity and evolution in a phylogenetic context. Plant Physiol **131:** 900–910

Engelhardt H (1922) Die alttertiare flora von messel bei darmstadt. *In* Abhandlungen der Hessischen Geologischen Landesanstalt zu Darmstadt, Band VII, Heft 4. Hessischer Staatsverlag, Darmstadt, Germany, pp 21–23

Farag MA, Huhman DV, Dixon RA, Sumner LW (2008) Metabolomics reveals novel pathways and differential mechanistic and elicitor-specific responses in phenylpropanoid and isoflavonoid biosynthesis in *Medicago* truncatula cell cultures. Plant Physiol **146:** 387–402

Franzen JL, Gingerich PD, Habersetzer J, Hurum JH, von Koenigswald W, Smith BH (2009) Complete primate skeleton from the Middle Eocene of Messel in Germany: morphology and paleobiology. PLoS One **4:** e5723

Gao Z, Eyers S, Thomas C, Ellis N, Maule A (2004) Identification of markers tightly linked to *sbm* recessive genes for resistance to Pea seedborne mosaic virus. Theor Appl Genet **109:** 488–494

Garcia AA, Wang S, Melchinger AE, Zeng ZB (2008) Quantitative trait loci mapping and the genetic basis of heterosis in maize and rice. Genetics **180:** 1707–1724

Gepts P, Beavis WD, Brummer EC, Shoemaker RC, Stalker HT, Weeden NF, Young ND (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

Gonzalez-Rizzo S, Crespi M, Frugier F (2006) The *Medicago truncatula* CRE1 cytokinin receptor regulates lateral root development and early symbiotic interaction with *Sinorhizobium meliloti*. Plant Cell **18:** 2680–2693

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

Hofer J, Turner L, Moreau C, Ambrose M, Isaac P, Butcher S, Weller J, Dupin A, Dalmais M, Le Signor C, et al (2009) Tendril-less regulates tendril formation in pea leaves. Plant Cell **21:** 420–428

Hougaard BK, Madsen LH, Sandal N, de Carvalho Moretzsohn M, Fredslund J, Schauser L, Nielsen AM, Rohde T, Sato S, Tabata S, et al (2008) Legume anchor markers link syntenic regions between *Phaseolus vulgaris*, *Lotus japonicus*, *Medicago truncatula* and *Arachis*. Genetics **179:** 2299–2312

Innes RW, Ameline-Torregrosa C, Ashfield T, Cannon E, Cannon SB, Chacko B, Chen NW, Couloux A, Dalwani A, Denny R, et al (2008) Differential accumulation of retroelements and diversification of NB-LRR disease resistance genes in duplicated regions following polyploidy in the ancestor of soybean. Plant Physiol **148:** 1740–1759

Jing R, Knox MR, Lee JM, Vershinin AV, Ambrose M, Ellis TH, Flavell AJ (2005) Insertional polymorphism and antiquity of PDR1 retrotransposon insertions in *pisum* species. Genetics **171:** 741–752

Kalaimagal T, Muthaiah A, Rajarathinam S, Malini S, Nadarajan N, Pechiammal I (2008) Development of new cytoplasmic-genetic male-sterile lines in pigeonpea from crosses between *Cajanus cajan* (L.) Millsp. and *C. scarabaeoides* (L.) Thouars. J Appl Genet **49:** 221–227

Kawashima T, Wang X, Henry KF, Bi Y, Weterings K, Goldberg RB (2009) Identification of cis-regulatory sequences that activate transcription in the suspensor of plant embryos. Proc Natl Acad Sci USA **106:** 3627–3632

Kwak M, Gepts P (2009) Structure of genetic diversity in the two major gene pools of common bean (*Phaseolus vulgaris* L., Fabaceae). Theor Appl Genet **118:** 979–992

Kwak M, Velasco D, Gepts P (2008) Mapping homologous sequences for determinacy and photoperiod sensitivity in common bean (*Phaseolus vulgaris*). J Hered **99:** 283–291

Lavin M, Herendeen PS, Wojciechowski MF (2005) Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. Syst Biol **54:** 575–594

Lee A, Hirsch AM (2006) Signals and responses: choreographing the complex interaction between legumes and alpha- and beta-rhizobia. Plant Signal Behav **1:** 161–168

Lewis G, Schrire B, Mackind B, Lock M (2005) Legumes of the World. Royal Botanic Gardens, Kew, UK

Medicago Genome Sequence Consortium (2007) *Medicago truncatula* genome "Mt2.0" release whitepaper. README: prerelease of Medicago truncatula genome sequences. http://medicago.org/genome/downloads/Mt2/ (August 10, 2007)

Mudge J, Cannon SB, Kalo P, Oldroyd GE, Roe BA, Town CD, Young ND

(2005) Highly syntenic regions in the genomes of soybean, *Medicago truncatula*, and *Arabidopsis thaliana*. BMC Plant Biol **5:** 15

Oldroyd GE, Downie JA (2008) Coordinating nodule morphogenesis with rhizobial infection in legumes. Annu Rev Plant Biol **59:** 519–546

Palmer RG, Ortiz-Perez E, Cervantes-Martinez I, Wiley H, Hanlin SJ, Healy RA, Horner HT, Davis WH (2003) Hybrid Soybean—Current Status and Future Outlook. American Seed Trade Association, Washington, DC

Pfeil BE, Schlueter JA, Shoemaker RC, Doyle JJ (2005) Placing paleopolyploidy in relation to taxon divergence: a phylogenetic analysis in legumes using 39 gene families. Syst Biol **54:** 441–454

Rostoks N, Ramsay L, MacKenzie K, Cardle L, Bhat PR, Roose ML, Svensson JT, Stein N, Varshney RK, Marshall DF, et al (2006) Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. Proc Natl Acad Sci USA **103:** 18656–18661

Sato S, Nakamura Y, Kaneko T, Asamizu E, Kato T, Nakao M, Sasamoto S, Watanabe A, Ono A, Kawashima K, et al (2008) Genome structure of the legume, *Lotus japonicus*. DNA Res **15:** 227–239

Schlueter JA, Lin JY, Schlueter SD, Vasylenko-Sanders IF, Deshpande S, Yi J, O'Bleness M, Roe BA, Nelson RT, Scheffler BE, et al (2007) Gene duplication and paleopolyploidy in soybean and the implications for whole genome sequencing. BMC Genomics **8:** 330

Schmutz J, Cannon SB, Schlueter J, Ma J, Hyten DL, Song Q, Mitros T, Nelson W, May GD, Gill N, et al (2009) Genome sequence of the paleopolyploid soybean (Glycine max (L.) Merr.). Nature (in press)

Shaw RG, Geyer CJ, Wagenius S, Hangelbroek HH, Etterson JR (2008) Unifying life-history analyses for inference of fitness and population growth. Am Nat **172:** E35–47

Shoemaker RC, Schlueter J, Doyle JJ (2006) Paleopolyploidy and gene duplication in soybean and other legumes. Curr Opin Plant Biol **9:** 104–109

Singer SR, Maki SL, Farmer AD, Ilut D, May GD, Cannon SB, Doyle JJ (2009) Venturing beyond beans and peas: What can we learn from *Chamaecrista*? Plant Physiol **151:** 1041–1047

Tattersall AD, Turner L, Knox MR, Ambrose MJ, Ellis TH, Hofer JM (2005) The mutant crispa reveals multiple roles for PHANTASTICA in pea compound leaf development. Plant Cell **17:** 1046–1060

Tucker SC (2003) Floral development in legumes. Plant Physiol **131:** 911–926

Uehlein N, Kaldenhoff R (2008) Aquaporins and plant leaf movements. Ann Bot (Lond) **101:** 1–4

Vail J, Kulakow P, Benson L (1992) Illinois bundleflower: prospects for a perennial seed crop. *In* DD Smith, CA Jacobs, eds, Recapturing a Vanishing Heritage. Proceedings of the Twelfth North American Prairie Conference. University of Northern Iowa, Cedar Falls, IA, pp 31–32

Varshney RK, Close TJ, Singh NK, Hoisington DA, Cook DR (2009) Orphan legume crops enter the genomics era! Curr Opin Plant Biol **12:** 202–210

Vietmeyer ND (1978) The plight of humble crops (some neglected crops of high nutritive value). *In* FAO Review on Agriculture and Development, Vol 62. FAO, Rome, pp 23–27

Vodkin L, Jones S, Gonzales OD, Thibaud-Nissen F, Tutega ZG (2008) Genomics of soybean seed development. *In* G Stacey, ed, Genetics and Genomics of Soybean. Springer, New York, pp 163–184

Yang S, Gao M, Xu C, Gao J, Deshpande S, Lin S, Roe BA, Zhu H (2008) Alfalfa benefits from *Medicago truncatula*: the RCT1 gene from *M. truncatula* confers broad-spectrum resistance to anthracnose in alfalfa. Proc Natl Acad Sci USA **105:** 12164–12169

Young ND, Udvardi M (2009) Translating *Medicago truncatula* genomics to crop legumes. Curr Opin Plant Biol **12:** 193–201