Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Plant Biology

Orphan legume crops enter the genomics era!

Rajeev K Varshney[1,2], Timothy J Close[3], Nagendra K Singh[4], David A Hoisington[1] and Douglas R Cook[5]

Many of the world's most important food legumes are grown in arid and semi-arid regions of Africa and Asia, where crop productivity is hampered by biotic and abiotic stresses. Until recently, these crops have also suffered from a dearth of genomic and molecular-genetic resources and thus were 'orphans' of the genome revolution. However, the community of legume researchers has begun a concerted effort to change this situation. The driving force is a series of international collaborations that benefit from recent advances in genome sequencing and genotyping technologies. The focus of these activities is the development of genome-scale data sets that can be used in high-throughput approaches to facilitate genomics-assisted breeding in these legumes.

Addresses
[1] International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru 502324, A.P., India
[2] Generation Challenge Programme, c/o CIMMYT, Int APDO Postal 6-641, 06600 Mexico, D.F., Mexico
[3] University of California-Riverside (UC-Riverside), Riverside, CA 92521-0124, USA
[4] National Research Centre on Plant Biotechnology (NRCPB), IARI Campus, New Delhi 110012, India
[5] University of California-Davis (UC-Davis), Davis, CA 95616, USA

Corresponding author: Varshney, Rajeev K (r.k.varshney@cgiar.org) and Cook, Douglas R (drcook@ucdavis.edu)

Current Opinion in Plant Biology 2009, **12**:202–210

This review comes from a themed issue on
Genome studies and molecular genetics
Edited by Masahiro Yano and Roberto Tuberosa

Available online 20th January 2009

1369-5266/$ – see front matter

DOI 10.1016/j.pbi.2008.12.004

Introduction — the importance of legumes

The legumes comprise the third largest family of flowering plants and provide humans with important sources of food, fodder, oil, and fiber products. These roughly 18 000 allied species [1] are divided into three subfamilies: the basal and paraphyletic assemblage of Cesalpinoid species, and the monophyletic Mimosoid, and Papilionoid clades. The most conspicuous feature of the legume family is the capacity of most species to fix atmospheric nitrogen to ammonia in collaboration with nitrogen-fixing bacteria known as 'rhizobia'. The resulting ready supply of reduced nitrogen makes legumes pivotal components of both natural and agricultural ecosystems, and underlies their typically high protein content and consequently their value as grain and fodder. On a global scale legumes provide roughly one-third of human kind's dietary protein nitrogen. Legumes are also important sources of mineral micro-nutrients and macro-nutrients [2], as well as health promoting secondary metabolites [3] — interestingly, many of these same metabolites protect plants against an onslaught of pathogens and pests [4].

The Papilionoideae is the numerically dominant subfamily of legumes and includes essentially all major legume crops. With the notable exceptions of peanut (*Arachis hypogaea* in the dalbergioid clade) and lupin (*Lupinus* spp. in the genistoid clade), the major crop legumes are members of two Papilionoid clades (Table 1), known as the hologalegina or 'cool-season/temperate legumes' and millettioid or 'warm-season/tropical legumes' [5••]. Three cool-season legumes (chickpea, *Cicer arietinum*; pea, *Pisum sativum*; and lentil, *Lens culinaris*) were among the earliest domesticated plant species, forming part of the so-called 'grain ensemble' that was brought into cultivation in the Near East during Neolithic times [6]. Members of the sister millettioid clade include the world's most important food legume species, *Phaseolus vulgaris* or common bean, and the most important legume oil seed crop, *Glycine max* or soybean. Beyond sheer production statistics, several of these Papilionoid legumes are vital components of the agricultural systems in resource poor areas of the world, with key examples including cowpea (*Vigna unguiculata*), pigeonpea (*Cajanus cajan*), common bean, lentil, chickpea, and groundnut.

Orphan legumes: needs and opportunities

With the exception of soybean, to various extents legume crops have suffered from poorly developed infrastructure (both knowledge and physical capacity) for genetic and genomic analysis — they have literally been 'orphans' from the genomics revolution. The lack of such infrastructure has limited the application of enabling biotechnologies for crop improvement. In particular, there is a significant need, first, to increase the availability of genomic data and resources in key species; second, to decrease the barriers that limit adoption of complex genomic data sets by crop improvement specialists; and third, to improve the capacity for the uptake of new biotechnologies by training the next generation of scientists to navigate both basic and applied plant science, and thus span the 'gap' (in the sense of Figure 1) between genomics and breeding.

Table 1

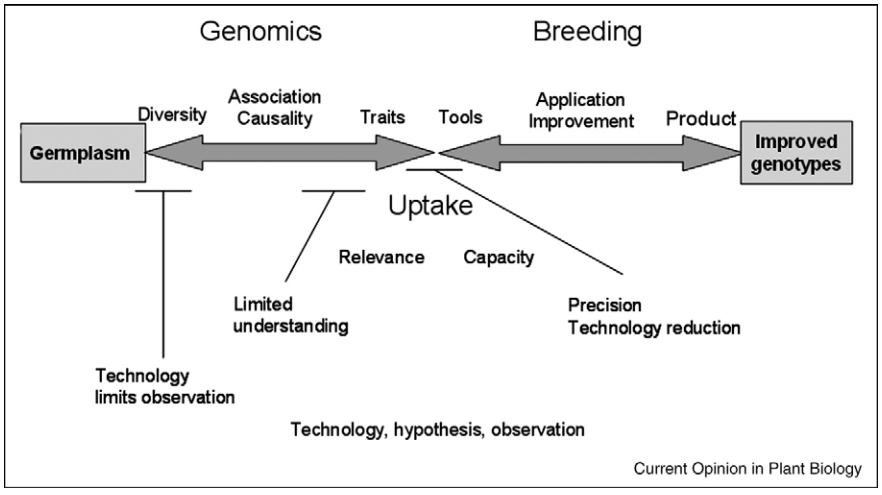
Major Papilionoid legume crop species and 2007 FAO production statistics

Clade affiliation	Primary species	Common name	World-wide production[a] (tonnes)
Dalbergioid	*Arachis hypogaea*	Groundnut	34 856 007
Genistoid	*Lupinus* spp.	Lupins	635 337
Hologalegina	*Cicer arietinum*	Chickpea	9 313 043
	Lens culinaris	Lentil	3 873 801
	Pisum sativum	Garden pea (dry + green)	18 393 255
	Melilotus spp.	Sweetclovers	
	Medicago sativa	Alfalfa	
	Trifolium spp.	Clovers	
	Vicia faba	Broad/faba bean	4 868 681
Millettioid	*Cajanus cajan*	Pigeonpea	3 428 610
	Glycine max	Soybean	216 144 262
	Phaseolus vulgaris	Common bean, (dry + green)	28 322 024
	Vigna unguiculata	Cowpea	5 408 431

[a] *Source*: http://faostat.fao.org/.

There is also a pressing need to explore the biological and mechanistic bases of key legume phenotypes. For example, all legume crops are exposed to abiotic and biotic stresses that decrease yield and productivity. Detailed understanding of the molecular mechanisms that underlie these traits could lead to novel and superior mitigating strategies. Tragically, many of these stresses are most severe in developing regions of the world. For example, regional climatic conditions seriously constrain productivity in sub-Saharan Africa and parts of the Indian subcontinent, while economic realities in many of these same areas limit the use of resource-intensive inputs, including irrigation and fertilizer, which could help counteract these constraints. Part of the solution, of course, is that the next generation of improved crop genotypes must be better equipped with endogenous capacities to tolerate such stresses. A range of factors, including marginal soils, suboptimal improved germplasm, and numerous diseases and environmental stresses comprise syndromes of regional and species-related constraints. Thus, while cowpea and pigeonpea are among the most drought tolerant of legume crops, periodic droughts still limit their productivity; moreover, the major constraint to yield in cowpea and pigeonpea is disease, caused by

Figure 1



Spanning the gap between genomics and breeding. In most legume species, crop improvement has occurred through traditional breeding approaches, with limited or no impact from molecular technologies. The genome projects described in this review have the potential to reverse this situation. In particular, these projects will contribute to the characterization of germplasm resources and natural populations, increase the frequency with which genetic variation is correlated with trait variation, and bring increasingly higher throughput and lower cost technologies to bear. In parallel to the development of genomic tools and data sets, it is important to recognize that not all genomic data are relevant to the task of crop improvement — thus, genomics researchers must work with breeders to identify relevant genetic variation. Simultaneously, crop improvement specialists must be better equipped with knowledge and access to infrastructure that are necessary for efficient uptake and application of genomic data. In the end, the result will be a continuum between germplasm resources and improved crop genotypes.

Striga (a parasitic weed) and sterility mosaic virus, respectively. Similarly, while drought can negatively impact productivity in several legume crops, the pod-boring insect (*Helicoverpa armigiera*) is perhaps the most persistent and serious constraint to chickpea and pigeonpea productivity. Finally, the legumes' most notable characteristic and perhaps their chief competitive advantage — symbiotic nitrogen fixation — is strongly constrained by abiotic factors such as drought, salinity, and phosphate availability. The interaction between biotic and abiotic stress is likely to be especially complicating in arid and semi-arid regions of the world, and de-convoluting such interactions is an important long-term challenge for legume improvement.

In the near term, great strides in crop improvement are possible by combining genomic tools with rationale selection of germplasm and precise phenotyping for traits of interest — an approach termed 'genomics-assisted breeding' [7••,8•]. From a longer term perspective, improved knowledge of biological systems will eventually allow the prediction of emergent phenotypes from complex genotypes; such knowledge will enable modeling of complex phenotypic outcomes from species-scale genotyping, and potentially make *in silico* studies a standard prelude to more traditional breeding practices.

Table 2

Status of germplasm collections for the selected legumes

Species	Location of major collections	[a]Total number of accessions	[b]Status of core/minicore collection
Chickpea	ICRISAT	20 140	1956 (211)
	ICARDA	12 776	–
	USDA-ARS	6 195	505
Common bean	CIAT	35 254	1400
	USDA-ARS	8 997	198 and 224
Cowpea	IITA	15 004	2062
	USDA-ARS	6 838	720
	UC Riverside	5 600	–
Groundnut	ICRISAT	15 419	1704 (184)
	USDA-ARS	10 013	831 (112)
Lentil	ICARDA	10 282	972
	USDA-ARS	2 876	280
Pigeonpea	ICRISAT	13 632	1290 (146)
	USDA-ARS	7	–

[a] Information on total germplasm accessions held (as per October 2008) in international centers (e.g. ICRISAT, CIAT, IITA, and ICARDA) and USDA-ARS genebanks taken from SINGER (http://singer.cgiar.org/index.jsp) and USDA-ARS GRIN (http://www.ars-grin.gov/npgs/stats/) databases, respectively.
[b] Information on core collection and minicore collection (in parenthesis) collected directly from respective Genebank curators.

The dearth of genomic resources that has characterized most legume crops, especially those of primary importance in the developing world, is beginning to change as these species are adopted into the genomics era. In the past three years several national and international initiatives have emerged to tackle this challenge. Although the explicit objectives vary from project to project, in aggregate these activities will provide genomic data sets, derivative knowledge, and new technologies that have the potential to transform molecular strategies for legume crop improvement.

Germplasm as a starting point

Managed germplasm collections are available for many orphan legume species (Table 2), and characterization of genetic diversity within these collections is a necessary prelude to their efficient use. The results of such genetic analyses permit legume researchers to distil large collections of individual lineages to smaller subsets, including representative core collections (for review see [9]). Such subsets may be constructed to encompass the majority of genetic and phenotypic diversity in a given species, or they may be selected to represent desired genetic structures, such as recombinant inbred lines, purpose-driven association panels, or even natural populations. Recent technological advances in the areas of DNA sequencing [10] and genotyping [11••] are serving to redefine the scope of germplasm characterization. Importantly, the combination of high-throughput genotyping with precise and focused phenotyping will facilitate efforts to associate molecular markers with agronomic traits. When distilled to assays of acceptable cost and scale, the stage will be set to more effectively incorporate germplasm collections of the world's orphan legumes into genomics-assisted breeding programs (Figure 1).

Comparative genomics as a strategy to leverage data from the reference legume genomes

In the 1990s, *Medicago truncatula* [12] and *Lotus japonicus* [13] emerged as model species to accelerate the study of legume biology. Their respective small diploid genomes, autogamous nature, short generation times, and prolific seed production made *Medicago* and *Lotus* excellent choices for undertaking genome analyses, and a range of powerful molecular, genetic, and genomic tools have been developed in each species [14•]. Examples of such tools include genetic and physical maps [15–17]; comprehensive EST collections, and detailed expression atlases [18,19,20••]; proteome [21] and small RNA catalogs [22]; resources for forward and reversed genetics [23–26]; bioinformatics tools and databases [27,28]; metabolomic profiling [29]; and genome-wide sequence data [30,31•,32••].

Owing to phylogenetic relationships within the legume family [5••], the investments made in *Medicago* and *Lotus* genomics have fuelled research to transfer knowledge of genome structure and function from the well-characterized reference legumes (including soybean) to related

Table 3

Overview on genomic resources in selected legume crop species

Common name	Cowpea	Chickpea	Pigeonpea	Groundnut	Lentil	Common bean
Species	*Vigna unguiculata*	*Cicer arietinum*	*Cajanus cajan*	*Arachis* spp.	*Lens culinaris*	*Phaseolus vulgaris*
Ploidy	$2n = 2x = 22$	$2n = 2x = 16$	$2n = 2x = 22$	$2n = 2x = 20$, $2n = 4x = 40$	$2n = 2x = 14$	$2n = 2x = 22$
Genome size	620 Mbp	740 Mbp	858 Mbp	$2n = 1260$ Mbp (*A. duranensis* — AA genome; *A. ipanensis* — BB genome), $4n = 2890$ Mbp	4063 Mbp	637 Mbp
SSRs (in use)	768 BES-SSRs[a,b]	510 genomic SSRs[e] (see [9]), 1 655 BES-SSRs[a,b]	130 genomic SSRs[b,e] (see [9])	~700[a] (see [9]), >2000 EST-SSRs[g]	~100[h] (e.g. [48])	~500 genomic and EST-SSRs (see [49])
BAC libraries	6X[d], 10X[a], 17X[c]	3.8X (see [9]), 7X (see [9]), 10X[a]	11X[a]	4X *A. hypogaea*[a] 6.5X *A. hypogaea* (see [9]) 7.4X *A. duranensis* [47] 5.3X *A. ipaensis*	–	10–20X [50]
BAC-end sequences	50 120 (36.7 Mbp)[a,e], 30 000[c]	46 270 (33.2 Mbp)[a,e]	85 785 (56.5 Mbp)[a,b,e]	41 856 (28.6 Mbp)[a]	–	89 017 (62 Mbp) [40•]
ESTs	183 658[e]	7355[b,e], 20 159[b], 435 184 454/FLX[b]	933[e], 9888[b], ~15 000[f], 496 705 454/FLX[b]	59 288[e]	1[e]	83 448[e]
Genetic maps						
Broad crosses	++	++	No	AA (2X) genome: ++	+	++
Narrow crosses	+	+	No	BB (2X) genome: ++, AABB (4X): +	+	+
Physical map	Yes[c]	No	No	In progress	No	Yes [40•]

[a] *Source of information*: UC Davis (DR Cook).
[b] *Source of information*: ICRISAT (RK Varshney).
[c] *Source of information*: UC Riverside (TJ Close, MC Luo).
[d] *Source of information*: University of Virginia (M Timko).
[e] *Source of information*: In public domain (e.g. NCBI).
[f] *Source of information*: NRCPB (NK Singh).
[g] *Source of information*: University of Georgia (SJ Knapp).
[h] *Source of information*: USDA-ARS/Washington State University (PN Rajesh).

food and feed legumes. It is noteworthy that, with the exception of polyploidy in select species (e.g. soybean, cultivated groundnut, and alfalfa), the most recent whole genome duplication in the Papilionoideae is predicted to be at least as ancient as the divergence of the Hologalegina and Millettioid clades [31•]. One might predict, therefore, that genome structure and content have been relatively stable since divergence of orphan legume species from the related reference species. Indeed, several studies have shown conserved synteny among the cool-season legumes, including between *M. truncatula* and alfalfa [16] and pea [33], as well as between the major Papilionoid clades listed in Table 1 [34•,35,36•]. Conservation of genome structure and function between legume species should facilitate the use and reuse of genomics resources between different legume species, as in the case of cross-species molecular markers [37], and in the case of cross-species use of oligonucleotide arrays [38•]. Moreover, the benefit of comparative biology to the study of agronomic traits has been recently demonstrated in the case of cross-species transfer of disease resistance between *M. truncatula* and alfalfa [39].

The existing legume comparative maps are based on small numbers of orthologous markers and thus are imprecise tools for translation between species. One goal of current research is to develop considerably more detailed comparative genetic maps, based on hundreds to thousands of conserved genes. For example, a project led by a coalition of UC-Davis, Tuskegee University, and the National Center for Genome Resources (NCGR) focuses on allele resequencing of 1369 orthologous genes across the orphan legume species, creating a syntenic network of 10 orphan and reference legume genomes. The current data set of polymorphic genes represents ~35 000 validated SNP.

Genome-specific genetic resources

As a complement to the efforts on cross-species genome resources, even larger efforts are being directed toward the development of species-specific genomic tools and data sets. These efforts are being driven in part by reduced sequencing costs, advances in automation, and the advent of high-throughput genotyping platforms (Figure 1), leading to the situations described below where considerable progress is on the horizon.

Bacterial artificial chromosome libraries as primary species-specific resources

Recent efforts have produced BAC libraries that represent several fold genome coverage, often in multiple genotypes, for most of the target legume species (Table 3). Production of BAC libraries has been combined with medium-scale BAC end sequencing and bioinformatics analyses, yielding between 35 Mbp and 65 Mbp of genome sequence data per species. In the cases of cowpea, common bean, and diploid peanut (*Arachis duranensis*) BAC library production has been combined with genome-wide physical mapping efforts. In cowpea, for example, 60 000 BAC clones were subjected to high information content fingerprinting (HICF) and assembled into a 10X physical map. Efforts are underway to anchor the cowpea physical map to the emerging SNP-based genetic linkage map (see below). Related activities in common bean have yielded a 9X draft physical map [40•], including the sequencing of ~89 000 BAC ends. The resulting sequence data represent 62 Mbp of genome sequence, or an estimated 9.5% of the common bean genome. BAC-based resources will have great utility for subsequent genome analyses, because they provide the basis for a physical interpretation of other genetic and genomics resources within each species, and they will facilitate more detailed analysis of high value regions of the genomes of orphan legumes.

Simple sequence repeats or microsatellites

Owing to their multi-allelic and codominant nature, simple sequence repeats (SSRs) have often been the markers of choice in plant genetics and breeding [41]. SSR markers have been developed from both genomic and transcript data sets. Transcript-associated SSRs have the advantage of mapping annotated genes, but the disadvantage of comparatively low polymorphism rates [42]. More recently BAC end-sequence data sets have been mined for SSRs (e.g. [17]), facilitating the integration of genetic, physical, and genome sequence resources. Fortuitously, SSRs are over-represented on BACs whose end sequences are low-copy and/or gene-containing, presumably biasing genetic maps toward the gene-containing euchromatin — the genome fraction most likely to control agronomic phenotypes. Integration of these newly isolated SSR markers into genetic maps is ongoing in chickpea, cowpea, pigeonpea, common bean, and peanut; the result should be increased linkage between physical and genetic map resources, and will provide a useful complement to gene-based SNP markers.

Genome-scale analyses of NBS–LRR disease resistance gene homologs

Although plant genomes contain numerous genes that confer disease resistance, the most abundant class of disease resistance genes contains a centrally located nucleotide-binding site (NBS) domain and a carboxy-terminal leucine-rich repeat (LRR) domain. Researchers at the University of California-Davis and Tuskegee University have used the known diversity of NBS domain proteins identified in *M. truncatula* [43,44] to develop PCR primers for deep sampling of NBS domains across the Fabaceae, including cowpea, pigeonpea, common bean, chickpea, peanut, lentil, lupin, and redbud (*Cercis occidentalis*, a basal Cesalpinoid legume). To date >3000 unique NBS domains have been identified. In parallel to gene cloning, BAC-based physical maps are

being produced around singleton and clustered disease resistance gene homologs, while BAC end sequencing is providing the basis for the development of SSR and SNP genetic markers. The goal is a comprehensive molecular-genetic resource of candidate disease resistance genes in each of the target species, providing tools for molecular breeding as well as more fundamental studies of resistance gene evolution.

Gene discovery, functional genomics, and genotyping

Functional genomics has revolutionized biological research in several crop species and is predicted to have a similar impact on plant breeding [7••] — especially as a means to identify genes underlying agronomic traits. In many species, collections of expressed sequence tag (EST) data have provided an important starting point for functional genomics strategies. Although hundreds of thousands ESTs are available in *Medicago*, *Lotus*, and soybean, until recently transcript sequence data were scarce in the orphan legumes. Current efforts are changing this situation.

In cowpea, Sanger EST sequencing projects have yielded a large number of ESTs. The respective cowpea ESTs, now publicly available from NCBI, were from cDNA libraries of 9 diverse genotypes produced by researchers at the UC Riverside (141 538 ESTs; sequenced mainly at the Department of Energy Joint Genome Institute, USA), and 2 normalized cDNA libraries produced in a project headed by researchers at IITA, derived from 4 African breeding genotypes (41 505 ESTs; sequenced at the JCVI). In parallel, the Kirkhouse Trust has funded low pass genome sequencing of hypomethylated cowpea DNA, representing ~160 Mbp of sequence information that contains partial structures for thousands of genes [45].

In the case of chickpea, 80 238 26-bp tags representing 17 493 unique transcripts (UniTags) from drought-stressed and nonstressed control roots have been generated using SuperSAGE technology for the analysis of gene expression in chickpea roots in response to drought [46]. Sanger sequencing has been used to a limited extent to access the chickpea and pigeonpea transcriptomes (~27 000 and 13 000 ESTs, respectively) (Table 3). More recently, 454/FLX sequencing was used at ICRISAT in collaboration with JCVI and NCGR to obtain 435 184 and 496 705 sequence reads for chickpea and pigeonpea, respectively, providing 44 852 and 48 519 contigs. These sequence data provide access to a significant fraction of the total transcriptomes of chickpea and pigeonpea, and are expected to aid in the analysis of drought tolerance, including candidate gene discovery and the development of molecular markers for breeding applications [42].

Slightly more extensive Sanger sequencing has been conducted in the peanut genomes, with ~54 000 ESTs available for cultivated peanut (*A. hypogaea*) and ~6000 ESTs in the diploid *A. stenosperma*. Recently, approximately 1 000 000 454/FLX sequence reads have been generated for two *A. duranensis* genotypes at the University of Georgia, in collaboration with JCVI. These sequences are expected to represent at least 40 000 unigenes.

Next generation genotyping and sequencing technologies

New sequencing and genotyping technologies will play an increasingly significant role in the genomics of orphan legumes. Of particular importance is the ability of these technologies to sequence at great depths, which will reduce the barrier to SNP discovery that has plagued the narrow germplasm base of many orphan legumes. As described below, large-scale SNP discovery efforts are being combined with massively parallel genotyping platforms, which will accelerate linkage mapping and whole genome association (WGA) studies.

Researchers at the University of California-Riverside constructed multiple sequence alignments from ESTs derived from multiple cowpea genotypes, and identified approximately 8500 SNPs. One thousand five hundred and thirty-six SNPs were chosen and the first Illumina GoldenGate assay has been prepared for cowpea. The respective cowpea ESTs were mainly from 11 diverse genotypes compiled by researchers at the University of California Riverside (141 538 ESTs), and 2 normalized cDNA libraries produced at IITA in Nairobi, Kenya, derived from 4 African breeding genotypes (41 505 ESTs).

For SNP discovery in chickpea, Solexa 1 Gbp technology was used to sequence root cDNAs from parents of a mapping population segregating for drought tolerance. This work was conducted as a collaborative effort involving the NCGR, University of California-Davis, and ICRISAT. One-half run of Solexa sequencing yielded 5.2×10^6 and 3.6×10^6 sequence reads for each genotype, respectively. Owing to the absence of extensive chickpea genome sequence, genomic data from *M. truncatula* and transcriptome sequence data from other legume species were used to align and analyze the Solexa data sets for SNP discovery.

In the case of pigeonpea, researchers at NCGR and ICRISAT are using Solexa 1 Gbp sequencing to analyze cDNA from 10 lines that are the parents of key mapping populations. In parallel, ICRISAT and JCVI are sequencing cDNA libraries of a single pigeonpea genotype using 454/FLX technology. The combination of 454/FLX cDNA reads, 55 Mbp of pigeonpea BAC end data, and sequence data from the closely related soybean genome, should facilitate assembly and SNP discovery among the more numerous but shorter Solexa reads.

The large quantity of 454/FLX reads currently available for multiple genotypes of diploid groundnut (*A. duranensis*) is

expected to yield ~20 000 SNPs. These same *A. duranensis* genotypes, as well as those of other AA genome species provided by researchers at the Catholic University in Brazil, are being analyzed using the orthologous marker resources developed at UC Davis. These SNP data sets will be combined to develop an Illumina SNP genotyping platform containing 2X 1536 SNPs that will enable detailed molecular-genetic analysis of the *Arachis* genome.

Conclusions

Recent progress in the development of genome-scale data sets for several legume species offers important new possibilities for crop improvement. This progress will enable biotechnologists to more rapidly and precisely target genes that underlie key agronomic traits, and with such knowledge to develop molecular assays that are both relevant and of appropriate scale for breeding applications. Among the most important agronomic targets are a series of abiotic and biotic stresses that limit crop productivity, especially in the marginal physical and economic environments that define much of Africa and parts of Asia. In this context, an important but underutilized asset of several legume species is their extensive germplasm collections. These collections reflect global genetic diversity in each species and as such they are storehouses of potential genetic solutions to a range of agronomic constraints. Molecular analysis of germplasm collections with new-generation genomic tools will accelerate trait discovery through methods such as linkage and association mapping. Moreover, organized genome resources, including physical maps and functional genomics tools, will facilitate the isolation of genes for resistance/tolerance to biotic/abiotic stresses. Ultimately the availability of high-throughput and cost-effective genotyping platforms, combined with automation in phenotyping methodologies, will increase the uptake of genomic tools into breeding programs, and thus usher in an era of genomics-enabled molecular breeding in these legumes.

Acknowledgements

Authors are grateful to Generation Challenge Program (RKV, TJC, DAH, and DRC), National Science Foundation (DRC), Pigeonpea Genomics Initiative of Indian Council of Agricultural Research (ICAR), Government of India under the umbrella of Indo-US Agricultural Knowledge Initiative (AKI) (RKV, NKS, and DRC), National Fund of ICAR (RKV and DAH), and Department of Biotechnology of Government of India (RKV and DAH). Thanks are also due to several individuals (Greg May, MingCheng Luo, Yong Gu, Frank You, Sarah Hearne, Morag Ferguson, Richard Bishop, Jean Hanson, Christopher Town, Jun Zhuang, Jeffrey Ehlers, Philip Roberts, Michael Timko, Steve Knapp, and David Bertioli) for providing access to information before publication.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Lewis G, Schrire B, Mackinder B, Lock M (Eds): *Legumes of the World*. Kew: Royal Botanic Gardens; 2005, p. 592.

2. Grusak MA: **Photochemicals in plants: genomics-assisted plant improvement for nutritional and health benefits**. *Curr Opin Biotechnol* 2002, **13**:508-511.

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

4. He XZ, Dixon RA: **Genetic manipulation of isoflavone 7-*O*-methyltransferase enhances biosynthesis of 4′-*O*-methylated isoflavonoid phytoalexins and disease resistance in alfalfa**. *Plant Cell* 2000, **12**:1689-1702.

5. •• Wojciechowski MF: **Reconstructing the phylogeny of legumes (Leguminosae): an early 21st century perspective**. In *Advances in Legume Systematics*, vol. 10. Edited by Klitgaard BB, Bruneau A. Kew: Royal Botanic Gardens; 2003:5-25.

Wojciechowski has been pioneer to provide the molecular phylogeny of papilionoid legumes for understanding the evolutionary history of events that underlie the origin and diversification of this family of ecologically and economically important flowering plants. In this article, Wojciechowski, on the basis of extended molecular data set available till 2002, reconstructed the phylogeny of the Leguminosae and its close relatives that advanced our knowledge of legume biology and facilitated comparative studies of plant structure and development, plant–animal interactions, plant–microbial symbiosis, and genome structure and dynamics.

6. Zohary D, Hopf M: *Domestication of Plants in the Old World. The Origin and Spread of Cultivated Plants in West Asia, Europe and the Nile Valley*. edn 3. New York: Oxford University Press Inc.; 2000.

7. •• Varshney RK, Graner A, Sorrells ME: **Genomics-assisted breeding for crop improvement**. *Trends Plant Sci* 2005, **10**:621-630.

An outstanding article that presents a holistic approach of use of various genomic tools and approaches to predict the phenotype from the genotype and the approach was termed as 'genomics-assisted breeding'. By improving precision of prediction of phenotype from the genotype with a higher accuracy and efficiency, selection efficiency in breeding program can be enhanced to develop the improved genotypes from the germplasm.

8. • Varshney RK, Hoisington D, Tyagi AK: **Advances in cereal genomics and applications in crop breeding**. *Trends Biotechnol* 2006, **24**:490-499.

This article provides the highlights on advances recently made in cereal genomics, for example genetic and physical mapping, genome/EST sequencing, and genomics applications in cloning/isolation of genes for trait of interest and marker-assisted selection leading to development or release of improved genotypes/varieties.

9. Varshney RK, Hoisington DA, Upadhyaya HD, Gaur PM, Nigam SN, Saxena K, Vadaz V, Sethy NK, Bhatia S, Aruna R *et al.*: **Molecular genetics and breeding of grain legume crop for the semi-arid tropics**. In *Genomics Assisted Crop Improvement Vol. II 2007: Genomics Applications in Crops.* Edited by Varshney RK, Tuberosa R. The Netherlands: Springer; 2007:207-242.

10. Hudson M: **Sequencing breakthroughs for genomic ecology and evolutionary biology**. *Mol Ecol Resour* 2008, **8**:3-17.

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

This is the first study on utilizing Illumina Golden Gate assay in any crop. By assaying 1524 genome-wide SNPs, the study demonstrated that, after accounting for population substructure, the level of linkage disequilibrium exhibited in elite northwest European barley, a typical inbred cereal crop, can be effectively exploited to map traits by using whole-genome association scans with several hundred to thousands of biallelic SNPs.

12. Cook DR: ***Medicago truncatula* — a model in the making!** *Curr Opin Plant Biol* 1999, **2**:301-304.

13. Handberg K, Stougaard J: ***Lotus japonicus*, an autogamous, diploid legume species for classical and molecular genetics**. *Plant J* 1992, **2**:487-496.

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

This interesting review focuses on the most recent discoveries relating to how the symbiosis is established. Two general developments have contributed to the recent explosion of research progress in this area: first, the adoption of two genetic model legumes, *Medicago truncatula*

and *Lotus japonicus*, and second, the application of modern methods in functional genomics (e.g. transcriptomic, proteomic, and metabolomic analyses).

15. Pedrosa A, Sandal N, Stougaard J, Schweizer D, Bachmair A: **Chromosomal map of the model legume *Lotus japonicus***. *Genetics* 2002, **161**:1661-1672.

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

17. Mun JH, Kim DJ, Choi HK, Gish J, Debellé F, Mudge J, Denny R, Endré G, Dénarié J, Kiss GB *et al.*: **Distribution of microsatellites in the genome of *Medicago truncatula*: a resource of genetic markers that integrate genetic and physical maps**. *Genetics* 2006, **172**:2541-2555.

18. Fedorova M, van de Mortel J, Matsumoto PA, Cho J, Town CD, VandenBosch KA, Gantt JS, Vance CP: **Genome-wide identification of nodule-specific transcripts in the model legume *Medicago truncatula***. *Plant Physiol* 2002, **130**:519-537.

19. Asamizu E, Nakamura Y, Sato S, Tabata S: **Characteristics of the *Lotus japonicus* gene repertoire deduced from large-scale expressed sequence tag (EST) analysis**. *Plant Mol Biol* 2004, **54**:405-414.

20. •• Benedito VA, Torres-Jerez I, Murray JD, Andriankaja A, Allen S, Kakar K, Wandrey M, Verdier J, Zuber H, Ott T *et al.*: **A gene expression atlas of the model legume *Medicago truncatula***. *Plant J* 2008, **55**:504-513.

An impressive paper that provides a global view of gene expression in all major organ systems of *M. truncatula*, with special emphasis on nodule and seed development. Preferential expression of many legume-specific genes in nitrogen-fixing nodules indicates that evolution endowed them with special roles in this unique and important organ. Comparative transcriptome analysis of *Medicago* versus *Arabidopsis* revealed significant divergence in developmental expression profiles of orthologous genes, which indicates that phylogenetic analysis alone is insufficient to predict the function of orthologs in different species.

21. Imin N, De Jong F, Mathesius U, van Noorden G, Saeed NA, Wang XD, Rose RJ, Rolfe BG: **Proteome reference maps of *Medicago truncatula* embryogenic cell cultures generated from single protoplasts**. *Proteomics* 2004, **4**:1883-1896.

22. Wen J, Parker BJ, Weiller GF: ***In silico* identification and characterization of mRNA-like noncoding transcripts in *Medicago truncatula***. *In Silico Biol* 2007, **7**:485-505.

23. Penmetsa RV, Cook DR: **Production and characterization of diverse developmental mutants of *Medicago truncatula***. *Plant Physiol* 2000, **123**:1387-1397.

24. Tadege M, Wen J, He J, Tu H, Kwak Y, Eschstruth A, Cayrel A, Endre G, Zhao PX, Chabaud M *et al.*: **Large-scale insertional mutagenesis using the Tnt1 retrotransposon in the model legume *Medicago truncatula***. *Plant J* 2008, **54**:335-347.

25. Ivashuta S, Liu J, Liu J, Lohar DP, Haridas S, Bucciarelli B, VandenBosch KA, Vance CP, Harrison MJ, Gantt JS: **RNA interference identifies a calcium-dependent protein kinase involved in *Medicago truncatula* root development**. *Plant Cell* 2005, **17**:2911-2921.

26. Horst I, Welham T, Kelly S, Kaneko T, Sato S, Tabata S, Parniske M, Wang TL: **TILLING mutants of *Lotus japonicus* reveal that nitrogen assimilation and fixation can occur in the absence of nodule-enhanced sucrose synthase**. *Plant Physiol* 2007, **144**:806-820.

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

28. Town CD: **Annotating the genome of *Medicago truncatula***. *Curr Opin Plant Biol* 2006, **9**:122-127.

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

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

31. • Cannon SB, Sterck L, Rombauts S, Sato S, Cheung F, Gouzy J, Wang X, Muge J, Vasdewani J, Schiex T *et al.*: **Legume genome evolution viewed through the *Medicago truncatula* and *Lotus japonicus* genomes**. *Proc Natl Acad Sci U S A* 2006, **103**:14959-14964.

A comprehensive study on synteny comparisons between two model legumes, that is *M. truncatula* (Mt) and *L. japonicus* (Lj) including details about chromosome relationships, large-scale synteny blocks, microsynteny within blocks, and genome regions lacking clear correspondence. The study also indicates similar and largely homogeneous gene densities, although gene-containing regions in Mt occupy 20–30% more space than Lj counterparts and also a duplication (within the Rosid I clade) predating speciation.

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

This recent article presents the first opportunity to look into the complex and unique genetic system of legumes. The article provides the structural features of the *L. japonicus* genome based on 315.1 Mbp sequences, that is 67% of the genome and likely to cover 91.3% of the gene space. The study reports assigning of a total of 10 951 complete and 19 848 partial structures of protein-encoding genes to the genome. Synteny analysis showed traces of whole-genome duplication and the presence of synteny blocks with other plant genomes to various degrees.

33. Kalo P, Seres A, Taylor SA, Jakab J, Kevei Z, Kereszt A, Endre G, Ellis THN, Kiss GB: **Comparative mapping between *Medicago sativa* and *Pisum sativum***. *Mol Genet Genom* 2004, **272**:235-246.

34. • Choi HK, Mun JH, Kim DJ, Zhu H, Baek JM, Mudge J, Roe B, Ellis N, Doyle J, Kiss GB *et al.*: **Estimating genome conservation between crop and model legume species**. *Proc Natl Acad Sci U S A* 2004, **101**:15289-15294.

Probably the first comprehensive comparative mapping study in legumes that evaluated, after combining the molecular and phylogenetic analyses, genome conservation both within and between the two major clades of crop legumes. The study suggested considerable utility of comparative mapping for basic and applied research in the legumes, although its predictive value is likely to be tempered by phylogenetic distance and genome duplication.

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

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

A recent paper that reports on development and mapping of legume anchor markers in common bean and AA genome of allotetraploid peanut. New legume markers enabled the alignment of genetic linkage maps through corresponding genes and provided an estimate of the extent of synteny and collinearity among common bean, AA genome of *Arachis*, *L. japonicus*, and *M. truncatula*.

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

38. • Das S, Bhat PR, Sudhakar C, Ehlers JD, Wanamaker S, Roberts PA, Cui X, Close T: **Detection and validation of single feature polymorphisms in cowpea (*Vigna unguiculata* L. Walp) using a soybean genome array**. *BMC Genom* 2008, **9**:107.

A recent study that shows that the Affymetrix soybean genome array is a satisfactory platform for the identification of some 1000s of SFPs for cowpea and in related legumes. This study provides an example of extension of genomic resources from a well-supported species to an orphan crop.

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

40. • Schlueter JA, Goicoechea JL, Collura K, Gill N, Lin JY, Yu Y, Kudrna D, Zuccolo A, Vallejos CE, Muñoz-Torres M *et al.*: **BAC-end sequence analysis and a draft physical map of the common bean (*Phaseolus vulgaris* L.) genome**. *Trop Plant Biol* 2008, **1**:40-48.
On the basis of analysis of 89 017 BAC-end sequences (62.58 Mbp) covering approximately 9.54% of the genome and 1404 shotgun sequences, this study revealed that approximately 49.2% of the common bean genome contains repetitive sequence and 29.3% is genic. Compared to other legume BAC-end sequencing projects, it appears that *P. vulgaris* has higher predicted levels of repetitive sequence. Furthermore, assembling of fingerprinting data for 41 717 BACs provided a draft physical map consisting of 1183 clone contigs and 6385 singletons with ~9X coverage of the genome.

41. Gupta PK, Varshney RK: **The development and use of microsatellite markers for genetics and plant breeding with emphasis on bread wheat**. *Euphytica* 2000, **113**:163-185.

42. Varshney RK, Graner A, Sorrells ME: **Genic microsatellite markers in plants: features and applications**. *Trends Biotechnol* 2005, **23**:48-55.

43. Zhu HY, Cannon S, Young ND, Cook DR: **Phylogeny and genomic organization of the TIR and non-TIR NBS–LRR resistance gene family in *Medicago truncatula***. *Mol Plant Microbe Interact* 2002, **15**:529-539.

44. Cannon SB, Zhu HY, Baumgarten A, Spangler R, May G, Cook DR, Young ND: **Diversity, distribution and ancient taxonomic relationships within the TIR and non-TIR NBS–LRR resistance gene subfamilies**. *J Mol Evol* 2002, **54**:548-562.

45. Timko MP, Rushton PJ, Laudeman TW, Bokowiec MT, Chipumuro E, Cheung F, Town CD, Chen X: **Sequencing and analysis of the gene-rich space of cowpea**. *BMC Genom* 2008, **9**:103.

46. Molina C, Rotter B, Horres R, Udupa S, Besser B, Bellarmino L, Baum M, Matsumura H, Terauchi R, Kahl G, Winter P: **SuperSAGE: the drought stress-responsive transcriptome of chickpea roots**. *BMC Genom* 2008, **9**:553.

47. Guimaraes PM, Garsmeur O, Proite K, Leal-Bertioli SCM, Seijo G, Chaine C, Bertioli DJ, D'Hont A: **BAC libraries construction from the ancestral diploid genomes of the allotetraploid cultivated peanut**. *BMC Plant Biol* 2008, **8**:14.

48. Hamwieh A, Udupa SM, Choumane W, Sarker A, Dreyer F, Jung C, Baum M: **A genetic linkage map of lentil based on microsatellite and AFLP markers and localization of fusarium vascular wilt resistance**. *Theor Appl Genet* 2005, **110**:669-677.

49. McClean P, Gepts P, Kamir J: **Genomics and genetic diversity in common bean**. In *Legume Crop Genomics.* Edited by Wilson RF, Stalker HT, Brummer EC. USA: AOCS Press; 2004:60-82.

50. Kami J, Poncet V, Geffroy V, Gepts P: **Development of four phylogenetically-arrayed BAC libraries and sequence of the APA locus in *Phaseolus vulgaris***. *Theor Appl Genet* 2006, **112**:987-998.