भाकृअनुप
ICAR

हर कदम, हर डगर
किसानों का हमसफर
भारतीय कृषि अनुसंधान परिषद

Agri search with a human touch

ADVANCES IN PULSES GENOMIC RESEARCH

Indian Institute of Pulses Research
Kanpur – 208 024

Advances in Pulses Genomic Research

K.R. Soren

P.G. Patil

Alok Das

Abhishek Bohra

Subhojit Datta

S.K. Chaturvedi

N. Nadarajan

भाकृअनुप
ICAR

Indian Institute of Pulses Research
Kanpur - 208 024

Publication No. : 3/2012

Printed : September, 2012

Published by : Dr. N. Nadarajan, Director
Indian Institute of Pulses Research, Kanpur

Edited by : Mr. Diwakar Upadhyaya

Correct citation : K.R. Soren, P.G. Patil, Alok Das, Abhishek Bohra, S. Datta, S.K. Chaturvedi and N. Nadarajan (2012): Advances in Pulses Genomic Research. Indian Institute of Pulses Research, Kanpur

---

Printed at : Army Printing Press, 33 Sadar Cantt., Lucknow
Ph.: 0522-2481164

Preface

Genomic research in plant system is the basic foundation to understand the structural and functional genomics of the crop. Large number of genomic resources have been developed in recent year in pulse crops. This bulletin 'Advances in Pulses Genomic Research' contains information from the authentic and highly regarded sources, about the basics of the pulse crops, its importance with respect to nutritional security and soil health, impact of climate change for biotic and abiotic stresses on crop productivity and importance of genetic and genomic resources in crop improvement programme. The bulletin also focues on plant genome analysis, particularly development of molecular marker technology for genomic fingerprinting, genetic diversity analysis, application of high throughput marker system in establishment of genetic / trait maps for biotic and abiotic stresses. The use of bacterial artificial chromosome (BAC) libraries, cloning of agronomically important genes and application of marker systems in marker assisted breeding (MABC) for introgression of QTL / genes have also been discussed for the benefit of the clientele. With the technological explosion, widespread use of automated, robotic means of genome wide sequencing (GBS), now legume genome sequencing has revealed specific genes, genome duplications and single nucleotide polymorphism (SNP) whose functions were largely unknown and important in shaping plant genetic networks..

Authors' are sure that the present bulletin will be useful to the legumes research community and academicians involved in such activities in different corners of the country and will bridge the gap between the science and researcher. The help rendered by our project staff including Research Associates and Senior and Junior Research Fellows in compilation of the present bulletin was tremendous. Authors are thankful to the technical staff Mr. A.P. Singh and Mr. Malkhan Singh for their assistance. Authors also thankful to Mr. Diwakar Upadhyaya for editing and bringing out the bulletin in proper shape.

Authors

Contents

Introduction

Food legumes belong to the third largest family (Fabaceae or Leguminosae) of flowering plants and the second most important plant family in agriculture. Being rich source of vegetable protein for human diet, these crops are known as poor man's meat. These are of especially interest because of their capacity to fix atmospheric nitrogen through mutualistic interactions with soil rhizobial bacteria, a trait that is both ecologically and agriculturally important. Indeed, without the nitrogen fixed each year by legumes, humans would need to consume 288 billion kg of additional fuel in the Haber-Bosch process to generate anhydrous ammonia for agriculture given their importance to people; legumes are now the target of extensive sequence-based genomics research, which is revolutionizing our understanding of legume evolution and its connection to biologically important traits. Among the leguminous crops, chickpea, pigeonpea, mungbean, urdbean, lentil, field pea and common bean are major crops in India. In the legume research, tremendous progress has been achieved by the legume research group in terms of development of genetic as well as genomic resources development in last one decade. With the advancement and high throughput of whole genome sequencing technology, tremendous advancements in the area of next generation sequencing (NGS) have resulted in availability of whole genome sequence not only for model legume species such as *Glycine max* (soybean), *Medicago truncatula*, *Lotus japonicas,* but also for low profile resource poor legume such as *Cajanus cajan* (pigeonpea). Keeping this in view, this technical bulletin mainly focuses on recent genomics advancements witnessed in major pulses.

Historical Aspects

Development of molecular techniques for crop improvement has provided ample scope to understand crop genetics and structure of crop genomes. The application of molecular markers mainly deals with the DNA sequence variations among the crop species and creation of new sources of variation by introducing new and favourauble traits from landraces and related crop species. Molecular markers are useful in selecting target alleles in the population with minimum linkage drag around the target gene, and reduce the number of generations required to recover a very high percentage of the recurrent parent genetic background. Among the various molecular markers developed to date, RFLPs were developed first and were initially used for human genome mapping (Botstein *et al.* 1980). Later,

the improvements in marker detection systems to identify markers linked to useful traits, has enabled great advances in recent years. These marker techniques were very well adopted for mapping plant genomes. Though restriction fragments length polymorphism (RFLP) markers have been the basis for most of the work in crop plants, valuable markers have been generated from random amplification polymorphic DNA (RAPD) and amplified fragments length polymorphism (AFLP). Simple sequence repeats (SSR) or microsatellite markers have been developed recently for major crop plants and this marker system is predicted to lead even more rapid advances in both marker development and their utilization in breeding programmes. Identification of the markers linked to useful traits has been based on complete linkage maps and bulked segregant analysis. However, construction of partial maps in combination of pedigree and marker information has also proved useful in identifying marker / trait associations. Redirecting the current breeding methods by utilizing molecular markers in breeding programmes has become crucial in the present scenario. A genetic marker is a gene or DNA sequence with a known location on a chromosome and associated with a particular gene or trait. It can be described as a variation, which may arise due to mutation or alteration in the genomic loci that can be observed. A genetic marker may be a short DNA sequence, such as a sequence surrounding a single base-pair change (SNPs) or a long one, like minisatellites.

Some commonly used genetic markers are :

- RFLP (Restriction fragment length polymorphism) [Botstein *et al.,* 1980]
- RAPD (Random amplification of polymorphic DNA) [Williams *et al.,* 1990]
- SSR (Simple sequence repeat) or Microsatellite [Akkaya *et al.,* 1992]
- SNP (Single nucleotide polymorphism) [Jordan and Humphries, 1994]
- AFLP (Amplified fragment length polymorphism) [Vos *et al.,* 1995]
- DArT (Diversity Arrays Technology) [Jaccoud *et al.,* 2001].

Genome Sequencing and Its Relevance

Sequenced legume genomes look very much like those of other dicots, though comparisons with *Arabidopsis* can be complicated by its unusually small genome size and complex duplication history. A closer look at the *Gm* genome finds that > 57% of the overall sequence is found in repeat-rich, low-recombination heterochromatin, while most genes (78%) are found in euchromatic chromosome arm. Of course, this also implies that substantial numbers of *Gm* genes (22%) lie within the pericentromeric heterochromatin, a somewhat surprising and potentially important result. As expected, crossovers are profoundly reduced near centromeres, with the ratio of genetic to physical distance dropping by 27-fold between the euchromatic and pericentromeric portions of the genome.

Detailed examination of the genome sequences also provides insights into interesting or unusual gene families. The *Gm* genome is reported to have 283 legume specific gene families, an estimate that increases to 670 with the analysis of the *Mt* genome sequence. Both *Gm* and *Mt* contain higher numbers of nucleotide-binding-siteleucine-rich repeats (NBS-LRRs, also called NB-ARCs-i.e., nucleotide-binding adaptors shared by APAF-1, R proteins, and CED-4) containing disease resistance genes than other plant genomes sequenced to date. In *Mt*, for example, there are 764 NBS-LRR-related genes, with at least 550 expressed based on RNA-Seq (100). Outside of legumes, *O. sativa* is reported to have the largest number so far (519) (98). More than 90% of *Mt* NBS-LRRs reside in clusters that contain on average 7.4 members, including two mega clusters-one on *Mt*06 with 30 NBS-LRRs and another on *Mt*03 with 21. However, the conclusion that NBS-LRRs are overrepresented in legumes (or indeed in any plant family) needs to be tempered by the recent observation that there is considerable variation in NBS-LRR number between different accessions within a single species, including *Gm*. Legumes have higher numbers and increased complexity in other gene families: lipoxygenases, LysM receptor kinases and flavonoid biosynthetic enzymes, such as chalcone synthase. It may be important that LysM receptors and flavonoids are both known to play important roles in nodulation. Finally, all three sequenced legumes contain unusually high numbers of F-box domain genes compared with other plant species, with *Mt* possessing three times the number of F-box domain genes compared with either *Gm* or *Lj*.

In addition, one of the potential uses of whole genome sequence is large scale mining of preferred class of markers such as SSRs and SNPs in a cost effective and time saving manner. For examples in case of pigeonpea, sets of SSRs comprising 3, 09, 502 and 1, 89, 895 SSRs were identified through microsatellite survey of two draft genomes sequences (Varshney *et al*., 2011; Singh *et al*., 2011).

Legume Genomics

It is difficult to believe that massive amount of sequence data is available in plants in such a short time. The pace of change has been so rapid that in less than a decade we have gone from thousands of ESTs in some legume species to having four robust legume reference genomes. At the simplest level, translation of genome data between legume species enables important practical applications: discovery of genetic markers, development of linkage maps, and saturation of genome regions for positional cloning. This is especially true for minor legumes, where many species are important to agriculture but supported by small research communities. At basic level, dissection of genome sequence data reveals the structure, architecture, and evolution of important gene families and enables the identification of orthologous versus paralogous relationships among family (Fig. 1). The basic genomic

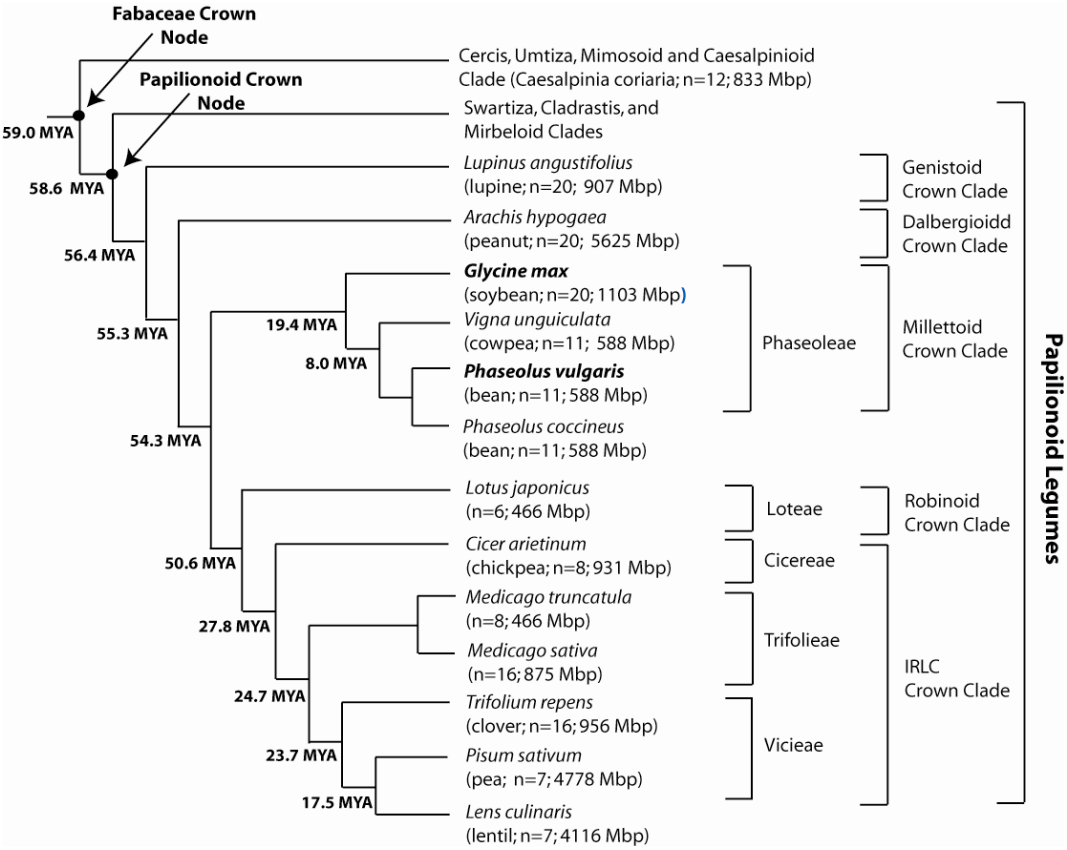

Fig. 1: The phylogeny relationship of major economic legumes relative to the others in the Fabaceae. (Source: Lavin *et al.*, 2005).

Table1: Availability of genomic resources in pulses

Common name	**Chickpea**	**Pigeonpea**	**Mungbean**	**Urdbean**	**Lentil**	**Pea**	**Common bean**
Species	*Cicer arietinum*	*Cajanus cajan*	*Vigna radiata*	*Vigna mungo*	*Lens culinaris*	*Pisum sativum*	*Phaseolus vulgaris*
Ploidy	2n=2x=16	2n=2x=22	2n=2x=22	2n=2x=22	2n=2x=14	2n=2x=14	2n=2x=22
Genome size	740 Mbp	858 Mbp	587Mbp	574Mbp	4063 Mbp	4300Mbp	637 Mbp
SSRs	510 genomic SSRs and 1655 BES-SSRs	~4,000 SSR	310	--	2393	586	500 genomic and EST-SSRs
BAC libraries	3.8X, 10X	11X	3.5x	--	-	--	10–20X
BAC-end sequences	46 270 (33.2 Mbp)	88,860 (56.5 Mbp	--	--	-	-	89 017 (62 Mbp)
ESTs (NCBI db EST release 010712)	44,157	88,860 (56.5 Mbp	829	250	9,513	18,552	1,23,988

information of all the major pulse crops has been shown in Table 1. Complete genome sequences also reveal legume- and species-specific genes whose functions remain largely unknown, although unquestionably important. Gene and genome duplications, so critical in shaping plant genomes, contain intrinsic information that can be exploited to predict function and structure of genetic networks. Additional sequencing and resequencing of legume species will make this possible, but inevitably, it is the research community's capacity to develop imaginative strategies for exploiting massive sequence data that will move legume genomics from the computer to biology. The information on the work done in area of genomic resource development and their utilization has been provided crop wise in this bulletin.

Chickpea (*Cicer arietinum*)

Chickpea is the third most important food legume (pulse) crop grown worldwide after dry bean and field peas in nearly 10 million hectare area. The major producers area across the Asia, Australia, Americas, the Mediterranean basin, East Africa, and the Middle East. (FAOSTAT, 2005). Commercially, the species have been grouped into *desi* and *kabuli* types. The small seeded *desi*-type chickpea accounts for about 85% of world production and mainly grown in India, Pakistan, Iran, Afghanistan and Ethiopia. The less common large seeded *kabuli*-type is grown in Middle East, India, Mexico as well as in North America, Australia and Spain. Chickpea is mostly consumed as a mature pulse (cooked whole, dehulled or as flour), but is also served as a vegetable (immature shoots and seeds). Seeds average about 20% protein, 5% fat and 55% carbohydrate and represent a basic food crop in many developing countries, especially India, where it has a high economic value. Chickpea seeds contain protein, fiber, calcium, potassium, phosphorus, iron, zinc and magnesium along with appreciable quantities of selenium, sodium and copper, which make it one of the nutritionally best composed edible dry legumes for human consumption (Esha, 2010). Chickpea like most other beans is a good source of cholesterol lowering fiber (Pittaway *et al*., 2006). It is a low input crop that often completes its lifecycle in drought and heat stress. While in the developed world it represents a valuable crop for export, in the developing world it provides a protein-rich supplement to cereal-based diets.

Demand for pulses is increasing but economics of production still does not encourage their cultivation on the more productive soils (Saxena *et al.,* 1993). Many of the biotic and abiotic stresses faced by chickpea (Johansen *et al.,* 1994) contribute to the large yield gap between potential yields and realized yields. *Fusarium* wilt, *Ascochyta* blight, root rots, *Botrytis* gray mold and pod borer are among the most important diseases and pests of chickpea.

Major emphasis of chickpea breeding programmes in the country has been on development of varieties with high and stable yield, and resistance to biotic (*Fusarium wilt*, dry root rot, *Helicoverpa* pod borer) and other abiotic stresses. A preliminary genetic map has been developed in this species with respect to *Fusarium* wilt, A*scochyta blight* and other abiotic factor able to identify markers, particularly for resistance to drought stress.

Available genomic and genetic resources

Availability of the chickpea genomic resources is still in its infancy when compared with well researched cereals. Most commonly used molecular markers in chickpea are SSR markers, ESTs and single nucleotide polymorphism (SNP) markers. Till date nearly 800 SSR markers have been developed (Gaur *et al*. 2011),

but the frequency of polymorphic markers is quite low (30%) as compared to the other genera (Choudhary *et al*. 2006; Nayak *et al*. 2010). Although saturated linkage maps are available, but tagging and mapping of biotic and abiotic stress tolerance loci with tightly linked markers are still lacking. As the limited number of genetic resources and high throughput genomic resources are available it is necessary to generate additional genetic resources and genome wide markers (EST-SSR, SNP, CAPS and DArT). To increase the genome wide molecular markers, ICRISAT in collaboration with University of Frankfurt, Germany, has been able to develop 311 SSR markers from SSR-enriched libraries (Nayak *et al*., 2010), and 1344 SSR markers from BAC-end sequence mining approaches in collaboration with University of California, Davis, USA (Nayak *et al*. 2010) in chickpea. Beside this, several EST based sequences from various tissues and developmental stages of chickpea have also been reported (Boominathan *et al*., 2004; Romo *et al*., 2004; Buhariwalla *et al*., 2005; Coram and Pang, 2005; Gao *et al.,* 2008; Varshney *et al*., 2009b, Choudhary *et al*., 2009, Ashraf *et al.,* 2009, Jain and Chattopadhyay, 2010). Although several genes / ESTs involved in various stress responses have been identified based on transcriptomic and proteomic studies (Pandey *et al.,* 2006, 2008; Mantri *et al.,* 2007; Molina *et al.,* 2008, 2011; Varshney *et al.,* 2009), the gene discovery has been very limited in chickpea. So far, only a few candidate genes have been cloned and functionally validated (Kaur *et al.,* 2008; Shukla *et al.,* 2009; Tripathi *et al.,* 2009; Peng *et al.,* 2010).With this EST either from tissue and chickpea data base a few hundred SSR markers have been developed from ESTs (Buhariwalla *et al*., 2005, Varshney *et al*., 2009b, Choudhary *et al*., 2009). As a result of above mentioned efforts, >2000 SSR markers representing the entire chickpea genome are now available at present. Whether this could help in constructing high density genetic linkage maps of chickpea is under question. Therefore, the high throughput SNP genotyping platform using DArT and next generation sequencing (NGS) technology like pyrosequencing (Alderborn *et al.,* 2000; Ching and Rafalski, 2002; Varshney *et al.,* 2008), Mass spectrometry (Rodi *et al.,* 2002), Affymetrix chip (Borevitz *et al.,* 2003), Golden Gate assay (Fan *et al.,* 2003, Rostoks *et al.,* 2006), Roche 454 / FLX, AB Bio system and Illumina / Solexa have been used for whole-genome transcription profiling techniques to identify genomic regions and genes underlying plant stress responses (Varshney *et al*., 2009a; Varshney *et al*., 2010b) to develop large scale SNPs and employing for genotyping to develop highly saturated genetic and transcript maps (Gujaria *et al.,* 2011).

The amount of genomic information available for chickpea has been increasing recently. Efforts have been made to generate genetic linkage maps for chickpea based on microsatellites (Huttel *et al.,* 1999; Radhika *et al.,* 2007; Upadhyaya *et al.,* 2008; Millan *et al.,* 2010; Gaur *et al.,* 2011). Chickpea genome organization and composition have also been analyzed based on the 500 kb sequence from 11 Bacterial

Artificial Chromosome (BAC) clones (Rajesh *et al.,* 2008). Recently, a BAC/BIBAC-based physical map of cultivated chickpea, *Cicer arietinum* (cultivar, Hadas), has also been developed (Zhang *et al.,* 2010). During the advancement of molecular breeding, different mapping populations have been used for development of genetic map, physical map, trait mapping and transcriptomic map which are shown in Table 2 & 3.

Table 2. Molecular genetic maps developed so far in chickpea

Mapping population	Marker loci mapped	References
Inter-specific		
(*C. arietinum* × *C. reticulatum)*	29	Gaur and Slinkard, 1990a, 1990b
ICC 4958 × PI 489777	120	Winter *et al*., 1999
	354	Winter *et al*., 2000
	56	Tekeoglu *et al*., 2002
	296	Pfaff and Kahl, 2003
	521	Nayak *et al*., 2010
	28	Kazan *et al*., 1993
PI 360177 × PI 489777 and PI 360348 × PI 489777		
ICC4959 x PI 489777 PI 360177 × PI 489777 and PI 360348 × PI 489777	91	Simoun and Muehlbuer, 1997
FLIP 84-92C × PI 599072	144	Santra *et al*., 2000
JG 62 × CA-2156	117	Rajesh *et al*., 2002
Hadas × Cr205	93	Abbo *et al*., 2005
ILC 72 × Cr5-10	89	Cobos *et al*., 2006
BDG 112x FLIP 0-166	33	Bharadwaj *et al*., 2010
Inter-specific		
(*C. arietinum* × *C. echinospermum*) Lasseter × PI 527930	83	Collard *et al*., 2003
Intra-specific		
(*C. arietinum* × *C. arietinum)*	103	Cho *et al*., 2002
ICCV 2 × JG 62	55	Udupa and Baum, 2003
ILC 1272 × ILC 3279	69	Flandez-Galvez *et al*., 2003
ICC 12004 × Lasseter CA 2139 × JG62 CA 2156 × JG62	138	Cobos *et al*., 2005
JG 62 × Vijay, Vijay × ICC 4958	273	Radhika *et al*., 2007
ICC 4991 × ICCV 04516	84	Kottapalli *et al*., 2009
WR 315 × C 104 102	102	Sharma *et al*., 2004
Consensus map		
Five narrow crosses (*Desi* × *Kabuli*)	229	Millan *et al*., 2010
Five wide crosses (*C. arietinum* × *C. reticulatum*)	555	Millan *et al*., 2010

Source: Upadhyaya *et al.,* 2011 (updated)

Table 3: Recent update in genomic resources available in chickpea

Type of marker	Number of markers developed	References
Genomic SSR	28	Huttel *et al*., 1999
	174	Winter *et al*., 1999
	95	Sethy *et al.*, 2006
	200	Lichtenzveig *et al.*, 2005
	13	Choudhary *et al*., 2006
	15	Eujayl *et al*., 2004
	85	Sethy *et al.*, 2006a,b
	63	Qadir *et al*., 2007
	311	Nayak *et al*., 2010
	1344	ICRISAT-UC Davis,USA
EST-SSR	60	Choudhary *et al*., 2009
	77	Varshney *et al.*, 2009b
	106	Bhuhariwalla *et al*., 2005
	265	Gaur *et al*., 2011
CAPS	32	Rajesh and Muehlbuer, 2008
	5	Varshney *et al*.,2007
	87	Varshney *et al.*, 2009a
	192	Gujaria *et al.*, 2011
DArT	15,360	DArT Pvt. Ltd, Australia And ICRISAT
SNP	Ca. 9,000 identified and 768 on Golden Gate assay	ICRISAT, UC-Davis, USA and NCGR, USA
	1,893	Gujaria *et al.*, 2011

Source: Upadhyaya *et al.,* 2011

Conventional and molecular breeding efforts

Resistance breeding against the insect pests has not progressed as for other agronomic traits mainly because of unavailability of resistant donors in germplasm pool. Situation is especially complex in case of resistance against polyphagous insect pests. These include structural defence (glandular hairs, trichomes), secondary metabolites (phytoalexines, isoflavonoides, saponine, tanine, alkaloids, oxalic acid and malic acid) and antinutritional compounds (lectin and proteinase inhibitor). In case of resistance breeding for fungal diseases *viz., Fusarium* wilt and *Ascochyta* blight, substantial progress has been made both in conventional breeding and also in terms of genetic as well as genomic resources development. This has paved way for deployment of linked marker in molecular breeding programme with support of marker assisted breeding (MAB) and will add a valuable role in providing resistance against diseases and other yield traits.

Fusarium wilt

Fusarium wilt (*Fusarium oxysporum f.* sp. *ciceri*) is a major disease of chickpea particularly in dry and warm areas. *F. oxysporum* races have been reported from India, Spain and the United States (0, 1A, 1B/C, 2, 3, 4, 5 and 6; Jimennez-Gasco *et al.,* 2004). Among the races, race 0 is least virulent and causes yellowing symptoms, whereas race 5 is most virulent and causes sever leaf chlorosis and plant death. The genetics of wilt resistance has been worked out by different workers. In particular, resistance to race 1a, 2 and 4 is either under control of two and three genes, while resistance to race 3 and 5 is monogenic (Sharma *et al.,* 2005). While Rubino *et al.* (2003) found two genes responsible for resistance to race 0 and genetics of resistance against race 1, race 2 and race 4 have been worked out. Varieties containing resistance to a particular race of *Fusarium* wilt remain resistance to that race even after several years of cultivation in a particular area showing stability of the resistance over time. Over 150 sources for resistance have been identified and some are resistant to more than one race. The genetics of resistance is however complex, since at least for resistance to race 1, a minimum of two out of three detected resistance genes are required. Genetics of the resistance gene were studied in inter- and intra-specific RIL populations to demonstrate the organization of resistance gene for fusarium wilt race 1, 3, 4 and 5 in two adjacent gene cluster in LG 2 flank by STMS marker GA 16 and TA 96 (*foc*1-*foc*4) and TA 96 and TA 27 (*foc*3-*foc*5) respectively. Sequence tagged microsatellite site (STMS) and sequence tagged site (STS) marker linked closely to *Fusarium oxysporum.* sp. *ciceri* race 3 resistance genes in chickpea identified and linkage between three wilt resistance genes was elucidated. Rubino *et al.* (2003) not only showed that two different genes can confer resistance to race 0, but also demonstrated linkage of the *foc* 1 to RAPD marker OPJ0 600 in LG 5 and tightly linked by flank by RAPD and STMS marker (Cobes *et al.,* 2005).

Gowda *et al.* (2009) used a cross (JG 62 x Vijay) to map the *Fusarium* wilt resistance gene. The RIL populations were used for molecular study and field evaluation in sick pot at Mahatma Phule Krishi Vidyapeeth (MPKV), Rahuri and identified the flanking and tightly linked DNA marker for the *Fusarium* wilt resistance. The SSR marker H3A12 and TA110 flanked the *foc* 1 locus at 3.9cM and 2.1cM respectively. While *foc* 2 was mapped 0.2 cM from TA96 and 2.7cM from H3A12. The H1B06y and TA194 marker flank the *foc* 3 locus at 0.2 and 0.7cM. A knowledge of virulence-related difference in pathogenic races at the molecular level is important for breeding wilt resistant chickpea varieties, by utilizing a linkage map as a framework, and the number and genomic positions of the genes conferring quantitative resistance may be determined using QTL analysis. Markers that are tightly linked to resistance QTL may then be utilized in breeding programmes *via* marker-assisted selection.

For tagging resistance genes to *Fusarium* wilt against *race-2* IIPR, Kanpur is already working in this direction and developed three mapping populations [BG256 x WR 315 (F_7), JG 62 x WR 315 (F_3) and K 850 x IPC 2004-52(F_3)]. These tightly linked co-dominant and dominant PCR based markers will be used for deployment in pyramiding of wilt resistance genes in chickpea to develop a cultivar with multiple race resistance to *Fusarium oxysporum f.* sp. *ciceri.*

Ascochyta blight

To study the genetics of resistance to ascochyta blight is as important as mapping is concern. After screening different isolates from different regions of the country, ten prevalent isolates (ab1-ab10) of ascochyta blight (*Ascochyta rabiei*) were reported. The study made in past revealed a confusing understanding that ascochyta blight resistance is controlled either by single gene (Tiwari and Panday, 1985), two genes (Dey and Singh, 1993), two dominant gene and one recessive gene (Tiwari and Pandey, 1986), polygenic (Flandez-Galvez *et al.,* 2003; Cho *et al.,* 2004). This is because it is still to clear whether the resistance genes represent the same or different loci because allelic test were not performed (Winter *et al.,* 2003). In the recent study by Bhardwaj *et al.* (2010) showed digenic recessive, monogenic recessive and digenic dominants and monogenic recessive inheritance of resistance gene in different crosses.

The occurrence of ascochyta blight has been reported in more than 40 countries of the world and has become one of the major constraints in chickpea cultivation. It has been reported that under epidemic condition the disease may lead to the 100% crop loss (Nene and Reddy, 1987). The genetic diversity of a range isolates of *Ascochyta rabiei* collected throughout the chickpea growing regions of Australia between 1996 and 2000 was quite narrow, compared to the large amount of variation found among *A. rabiei* isolates assessed from Canada, USA and Syria. Under the resistance breeding programme several resistance breeding line *viz.,* GG1267, GL90168, GL96010 and GL98010 have been reported and being used in breeding against ascochyta blight and to study the genetics of ascochyta resistance (Bhardwaj *et al.,* 2010).

Molecular linkage maps were constructed of chickpea genome to identify quantitative trait loci (QTLs) controlling ascochyta blight resistance (ABR) (Millan *et al.,* 2003, Rakshit *et al.,* 2003., Santra *et al.,* 2000). An intraspecific population of chickpea (*C. arietinum*) established from a cross between a highly susceptible cultivar 'Lasseter' and a resistant inbred line 'ICC 12004' was used to construct the first anchored linkage map of chickpea genome. On this map 7 QTLs were located that conditioned ABR. Resistance Gene Analogue (RGA) and chickpea Sequence Tagged Microsatellite Site (STMS) markers closely flanked major resistance QTLs with two

markers identified 0.1 cM from the largest QTL peak (QTL-3) for both glass house and field resistance. An interspecific population was also established from a cross between 'Lasseter' and a resistant *C. echinospermum* accession (PI 527930). A genome map was constructed consisting of 8 linkage groups with 4 QTLs identified that conditioned ABR. At the seedling stage, using interval mapping, two QTLs were identified for resistance and two QTLs were associated with adult-plant resistance. The two QTLs for seedling and adult resistance on linkage group 2, that were flanked by a STMS marker (TR 20,) appeared to be common to the QTL-3 on linkage group III in chickpea map.

Abiotic stress

Towards mitigation of drought in chickpea IIPR, Kanpur is working in collaboration with ICRISAT, Hyderabad to introgress genes / QTLs responsible for drought tolerance in chickpea, as drought is the number one production constraint in chickpea production. Introgression of drought QTL from ICC 4958 into elite chickpea cultivar DCP 92-3 and KWR 108 using marker assisted back crossing (MABC) is the main target. One chromosomal region containing several QTLs for drought tolerance traits bracketed by three SSR markers has already been identified under the Phase I of the Tropical Legumes 1(TL 1) project and is being introgressed into elite lines of Kenya and Ethiopia.

Pigeonpea (*Cajanus cajan*)

Pigeonpea (2n=2 x = 22) is an important grain legume crop of tropical and sub-tropical regions of the world. It is the main source of dietary protein to a large fraction of vegetarian population in developing countries. Globally, India is one of the major pigeonpea growing countries covering about 3.4 million hectares, with average production and productivity of 2.89 m tonnes and 741 kg / ha, respectively (DAC 2011). Despite concerted breeding efforts for several decades, expected progress has not been witnessed in pigeonpea genetic improvement. The underlying reasons are narrow genetic diversity utilization, lack of precise information on available plant genetic resources, poor crop husbandry and exposure to a number of biotic and abiotic stresses.

Available genomic and genetic resources

In order to supplement breeding efforts, molecular breeding has gained more importance for last several years in addressing these challenges. However, molecular breeding in pigeonpea still remains in infancy stage due to lack of appropriate genetic resource and dearth of DNA polymorphism in cultivated pigeonpea pool. In order to leverage the pigeonpea genomic resource, significant progress has been made during the last 3 years in area of large number genomics and genetic resources development through national and international collaborative efforts which facilitated genetic mapping and reverse genetic analysis (Varshney *et al.,* 2010). For instance, 25 mapping populations segregating for various biotic and abiotic stresses have been generated. A total of 88, 860 BAC- end sequences (BESs) were generated through end-sequencing of 50,000 BAC clones and 3,072 SSR markers were synthesised form these BESs to facilitate genetic mapping and hybrid purity testing (Bohra *et al*., 2011). Utilizing the potential of second and third generation sequencing platforms, functional or transcriptomics resources like 10,000 ESTs and 2 million short ESTs were generated using Sanger sequencing and 454 / FLX sequencing, respectively (Dubey *et al*., 2011; Dutta *et al*., 2011). As a result, a total of four transcriptome assemblies have been developed in pigeonpea (Kudapa *et al.,* 2012). To facilitate high throughput genotyping a diversity array technology (DArT) comprising of 15,000 features has been developed (Yang *et al*., 2006; 2011).

Availability of mapping populations and genetic maps is the basic requirement for identifying marker-trait association. The markers thus identified are subsequently used for practicing marker assisted selection (MAS) for crop improvement. However, in case of pigeonpea extreme scarcity of polymorphic markers led to development of genetic maps with low to moderate marker densities. The first genetic map in pigeonpea was reported for an inter-specific F_2 population [ICP 28 (*C. cajan*) × ICPW 94 (*C. scarabaeoides*)] (Yang *et al*., 2011). The first SSR based genetic map was also based on the same inter-specific population (Table 4) and comprised of 239 SSR loci covering a map length of 930.90 cM (Bohra *et al.,* 2011). For cultivated pigeonpea, intra-specific genetic maps were developed from two different F_2 populations *viz*., TTB 7 × ICP 7035 and ICP 8863 × ICPL 20097 (Gnanesh *et al*., 2011. Furthermore, to merge the molecular information from several cultivated genetic map, first SSR based consensus genetic map was synthesized from six different mapping populations. This consensus genetic map provided map locations to a total of 339 SSR loci and spanned a total map distance of 1054 cM as shown in Fig. 2 (Bohra *et al.,* 2012).

Apart from genetic mapping and QTL analysis, emphasis has also been given to whole genome sequencing of pigeonpea. Undertaking of NGS platforms has successfully led to the availability of *de novo* whole genome sequencing which does

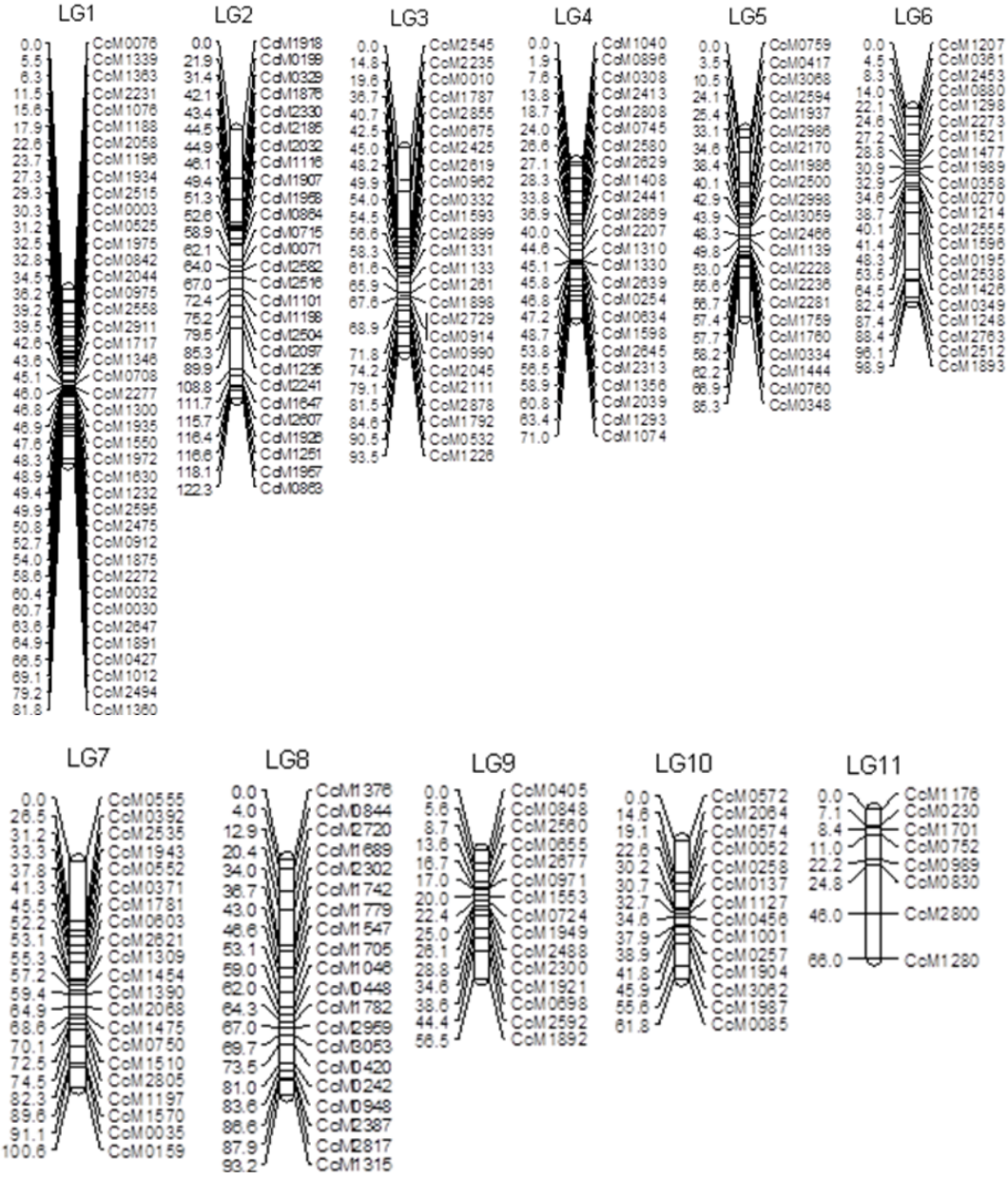

Figure 2: The first SSR based genetic map of pigeonpea based on inter-specific F_2 population (ICP 28 × ICPW 94) (Bohra *et al*., 2011; BMC Plant Biology, 11:56)

not require a physical map. Utilizing the same approach recently two draft genome sequences of pigeonpea cultivar 'Asha (ICPL 87119)' were made available through 454 GS-FLX (Singh *et al*., 2011) and Illumina GA Hi Seq 2000 (Varshney *et al*., 2011). The former attempt sequenced a total of 511 Mb of pigeonpea genome with 10 x

coverage, while the latter attempt captured 605 Mb of genome with 160 x coverage. The whole genome sequence provided insights into evolutional forces like genome duplication and domestication, gene content and proportion of repetitive elements. Moreover, the whole genome sequence provided access to a large set of DNA markers such as SSRs and SNP.

Table 4: SSR based genetic maps in pigeonpea

Name of F_2 mapping population	Number of total mapped markers	Total map length (cM)	Average inter-marker distance (cM)	References
ICP 28 × ICPW 94	239	930.9	3.8	Bohra *et al.*, 2011
ICP 8863 × ICPL 20097	120	534.89	4.45	Bohra *et al.*, 2012
ICPA 2043 × ICPR 3467	140	881.57	6.29	Bohra *et al.*, 2012
ICPA 2043 × ICPR 2671	111	677.97	6.1	Bohra *et al.*, 2012
ICPA 2039 × ICPR 2447	78	570.53	7.31	Bohra *et al.*, 2012
TTB 7 × ICP 7035	78	466.97	5.98	Bohra *et al.*, 2012
ICPB 2049 × ICPL 99050	59	586.02	9.93	Bohra *et al.*, 2012
Consensus genetic map	339	1059	3.12	Bohra *et al.*, 2012

Major biotic constraints

The major constraints affecting pigeonpea production are mainly biotic stresses like *Fusarium* wilt (*Fusarium udum* Butler), sterility mosaic disease (sterility mosaic virus) and phytophthora blight (*Phytophthora drechsleri*) and pests such as pod borer complex (*Helicoverpa armigera*, *Maruca vitrata*), pod fly (*Melanagromyza obtusa*), plume moth (*Exelastis atomosa*) and abiotic stresses like drought and sensitivity to salinity. Apart from this, lower harvest index of all the available pigeonpea cultivars also results in less productivity. Hence, achieving an ideal plant type will help greatly in overcoming the existing yield barriers. With this background the main aims are: 1) Efficient utilization of genetic resources like mapping populations segregating for various biotic/abiotic stresses and plant type, developed during AKI-PGI project; 2) To exploit the genomics resources like SSR, EST-SSRs, SNPs and the information generated from two whole sequencing projects in searching and designing of new primers like genomic SSRs, genic SSRs, SNPs and related genes for mapping genes / QTLs for different traits of interests; 3) Genomics enabled molecular breeding through marker assisted selection for improving the traits of interests, and 4) Gene pyramiding and releasing of high yielding pigeonpea varieties through MAS/ MABC.

Fusarium wilt

Fusarium wilt (*Fusarium udum* Butler) is the most important soil borne disease of pigeonpea. The disease first appears in patches in a field, it can even extend to entire field if pigeonpea is repeatedly cultivated in the same field. The yield loss due to wilt depends on the stage at which the plants wilt; it can approach 100% when wilt occurs at the pre-pod stage, about 67% when wilt occurs at maturity, and 30% when it occurs at the pre-harvest stages (Kannaiyan and Nene, 1984).

Conventional breeding for resistance to *Fusarium* wilt was problematic because of lack of accurate inheritance of resistance, inaccurate phenotyping and selection. Till date there is only one RAPD marker linked to *Fusarium* wilt resistant genes (Kotresh *et al.,* 2006). Under the Indo-US-AKI project number of mapping population have been generated for the different traits where IIPR, Kanpur has a population in F_6 generation for *Fusarium* wilt (Asha x UPAS 120). There is an urgent need to speed up the progress of deployment of genetic and genomics resources for mapping genes / QTLs for this trait.

Sterility mosaic disease (SMD)

Sterility mosaic disease of pigeonpea is caused by pigeonpea sterility mosaic virus (PPSMV) and transmitted by eriophyid mite (*Acœria cajani*) (Kumar *et al.,* 2003). It is a most important disease (Kannaiyan *et al*., 1984) causing yield losses upto 95%. Control of disease through chemical method was found economically not feasible and non-ecofriendly (Nene *et al.,* 1989). Development of resistant cultivars has been receiving top priority of pigeonpea breeders. Few instances of molecular tagging of SMD resistance genes were reported using F_2 mapping population and two AFLP markers E-CAA / M-GTG_{150}, CAA / M-GTG_{60} linked in coupling phase to susceptible dominant allele were discovered (Ganpathy *et al*., 2009). Similarly one major QTL *viz*., qSMD4 explaining 24.72% of phenotypic variance is identified from F_2 mapping population TTB 7 × ICP 735. Fine mapping of identified genes / QTLs using high throughput marker systems can be performed and the tightly linked markers can be deployed in MAB aiming at crop improvement.

Phytophthora blight

Phytophthora blight is another important fungal diseases caused by *Phytophthora drechsleri* resulting in substantial yield reduction. Despite of its economic importance, precise information on genetics and mapping of the trait has not been reported so far. Therefore, there is tremendous scope for deployment of molecular markers for mapping genes / QTLs in resistance breeding against this disease.

Mungbean (*Vigna radiata*)

Mungbean (2n=2x=22) is an important legume crop widely cultivated in Asia. The crop is utilized in several ways, where seeds, sprouts and young pods are consumed as sources of protein, amino acids, vitamins and minerals, and plant parts are used as fodder and green manure. Mungbean protein is easily digested without flatulence. It is an important protein source for people in the cereal-based society. It adapts well to various cropping systems owing to its ability to fix atmospheric nitrogen (N_2) in symbiosis with soil bacteria of *Rhizobium* spp., rapid growth, and early maturity. The area under mungbean has increased to the 3.55 mha in the country and mainly in Rajasthan, Maharashtra, Andhra Pradesh, Karnataka, Orissa and Bihar. The annual production of mungbean is about 1.8 million tonnes.

Genetic linkage maps of mungbean have been constructed using RFLP, RAPD, and SSR markers were used to map genes for bruchids resistance (Dan *et al.,* 2010) and seed colour to identify QTLs for seed weight, hard-seededness, powdery mildew resistance and cercospora leaf spot resistance. Recently, Isemura *et al.,* (2012) have constructed a genetic linkage map using transferable SSR markers from azuki bean, mungbean, cowpea, and common bean and EST-SSR markers from soybean (Glycine max), using a BC_1F_1 population derived from a cross between cultivated and wild mungbean to identify 105 QTLs and genes for 38 domestication traits in mungbean.

Major biotic constraints

Major constraints affecting mungbean production are mainly biotic stresses like mungbean yellow mosaic virus (MYMV), leaf crinkle and powdery mildew. Beside these crops also possess serious problem of heavy flower drops, *etc.*

Mungbean yellow mosaic disease

Mungbean yellow mosaic disease of mungbean is caused by mungbean yellow mosaic virus (MYMV) and transmitted by whitefly. It is a most important disease causing enormous yield loss. Towards the tagging of genes Gupta *et al.,* (2006) mapped the one marker ($ISSR811_{1357}$) which is tightly linked to the MYMV resistance gene at 6.8 cM.. The $ISSR811_{1357}$ marker was sequenced and sequence characterized amplified region (SCAR) primers were designed (YMV1-F and YMV1-R) to amplify the marker. Screening for the SCAR marker in the RIL population distinguished the MYMV resistant and susceptible plants, agreeing well with the phenotypic data. The $ISSR811_{1357}$ marker was validated using diverse blackgram genotypes differing in their MYMV reaction. The marker will be useful for the development of MYMV-resistant genotypes.

Powdery mildew disease

Powdery mildew disease caused by the fungus *Erysiphe polygoni* is a common foliar disease of mungbean. The disease may cause yield loss up to 40%. Using VC3890A as a resistance source, Young *et al.* (1985) found three QTLs on three different LGs associated with the resistance. These QTLs together accounted for 58% of the trait variation. Chaitieng *et al.* (2006) used VC1210A as a resistance source to map the resistance gene. Initial mapping with 98 framework RFLP probes failed to identify any association with the resistance. However, subsequent identification using amplified fragment length polymorphism (AFLP) markers and bulked segregant analysis (BSA) resulted in 4 bands linking to the resistance. These bands were then cloned and used as probes for RFLP analysis of which finally 5 RFLPs were found associated with the resistance. The five RFLPs constituted a new LG. A major QTL, *PMR1*, associated with the resistance on this LG accounted for 68% of the trait variation.

Urdbean (*Vigna mungo*)

Urdbean (2n=22) also known as blackgram, is the third most important pulse crop in India after chickpea and pigeonpea, which is grown in various agro-ecological conditions and season under diverse cropping system. It is grown over an area of 3.26 million ha with annual production of 1.76 million tonnes. It is widely cultivated in Indian sub-continent and to lesser extent in Thailand, Australia and other Asian and South pacific countries (Poehlman, 1991). In India, Andhra Pradesh, Karnataka, Maharashtra, Rajasthan, Tamil Nadu and Uttar Pradesh are the major blackgram growing states. Blackgram is an important self-pollinating diploid grain legume crop belonging to family *leguminosae* and nutritionally rich in protein (26%), which is almost three times than that of cereals. It is also rich in essential amino acids such as arginine, leucine, lysine, isoleucine, valine and phenylalanine, which complements to cereal based diet. In addition to being an important source of human food and animal feed, it also plays an important role in sustaining soil fertility by improving soil physical properties and fixing atmospheric nitrogen.

Towards the genomic resources development much work has not been done in this crop, but by using genomic resources from azuki bean, common bean and mungbean, genetic map were constructed (Chaiteing *et al.,* 2006).

Lentil (*Lens culinaris*)

Lentil is a self-pollinating diploid (2n=14) annual cool season grain legume produced as a high protein food source throughout the world. In India lentil occupies about 1.48 m ha with annual production of 1.03m tonnes. Several lentil genome maps are available and recent progress towards a consensus map has been made by employing robust locus markers that are derived from the model legume *Medicago truncatula* and other legume genomes. Such markers are co-dominant and will likely be useful across a broad lentil genetic background for marker-assisted trait selection. Candidate trait-associated genes are under investigation, particularly for disease resistance, and these are soon likely to become available for validation against pathogen populations and in differing environments using transgenic approaches. However, further effort is required to develop a robust and high-throughput full regeneration system for transformant lentil plants. In last few decades many genetic resources have been developed as shwon in Table 5.

Table 5: List of some of the published genetic maps in lentil

Type of genetic map	Cross name	Type of popu-lation	Type of markers mapped	Total number of loci mapped	Total map length (cM)	Reference
Inter-specific	*L. ervoides* × *L. culinaris*	F_2	Morphological, Isozymes, RFLPs	64	560	Havey and Muehlbauer, 1989
Inter-subspecific	*L. culinaris ssp. orientalis* × *L. culinaris*	F_2	RAPD, RFLP, morphological	33	206	Eujayl *et al*., 1997
Inter-subspecific	*L. culinaris ssp. orientalis* × *L. culinaris*	RIL	RAPDs, AFLPs, RFLPs, morphological	177	1073	Eujayl *et al*., 1998
Intra-specific	ILL5588 × ILL7537 (**first cultivated map**)	F_2	RAPDs, ISSRs RGAs	114	784	Rubeena *et al*., 2003
Inter-subspecific	*L. culinaris* ssp. *culinaris* × *L. c.* ssp. *orientalis*.	F_2	RAPDs, ISSRs, AFLPs, SSRs, morphological	200	2172	Duran *et al*., 2004
Intra-specific	ILL 5588 × L 692-16-1(s)	RIL	SSRs, AFLPs	283	751	Hamwieh *et al*., 2005
Intra-specific	ILL5722 × ILL5588 (**first gene-based genetic map**)	RIL	ITAPs, SSRs	97	928	Phan *et al*., 2007
Intra-specific	Eston × PI320937	RIL	AFLP, RAPD, SSRs	207	1868	Tullu *et al*., 2008
Intra-specific	Precoz × WA8949041	RIL	AFLP, ISSR, RAPD, morphologic markers	166	1396	Tanyolac *et al*., 2010

The near future of lentil genomics will include further candidate gene characterization through transcriptome and reverse genetic techniques. These studies will be conducted to uncover genes responsive to biotic and abiotic stimuli as well as those governing desirable seed quality traits, such as size, shape and colour. Towards development of genomic resources in lentil, research group from Australia has generated a total of 15,354 contigs and 68,715 singletons by *de novo* assembly. The complete sequence was analysed against genome drafts of the model legume species *Medicago truncatula* and *Arabidopsis thaliana* to identify 12,639, and 7,476 unique matches, respectively. When compared with the genome of Glycine max, a total of 20,419 unique hits were observed corresponding to 1.38×10^6 ESTs, 31% of the known gene space. A total of 25,592 lentil unigenes were subsequently annotated from GenBank. Simple sequence repeat (SSR) containing ESTs were identified from consensus sequences and a total of 2,393 primer pairs were designed (Kaur *et al.,* 2011; BMC genomics).

Biotic constraints

Foliar diseases are the most serious biotic stresses affecting lentil crop. *Ascochyta* blight caused by *A. lentis* is problematic to various degrees in all lentil growing regions of the world, but especially damaging in Canada (Ahmed and Morrall, 1996; Ahmed *et al.,* 1996), Australia and Middle Eastern countries (Johansen *et al.,* 1994). Symptoms of *Ascochyta* blight include lesions on all above ground parts of the plant, stem girdling, pod and seed lesions, and resistance has been found in the germplasm (Andrahennadi *et al.,* 1996). Other major biotic stresses of lentil include Anthracnose caused by *C. truncatum*, Botrytis grey mold caused by *Botrytis fabae* and *B. cinerea*, Stemphylium blight caused by *Stemphylium botryosum*, lentil rust caused by *U. fabae* and Sclerotinia white mold caused by *S. sclerotiorum.* Stemphylium blight is a major threat to lentil in South Asia and North America (ICARDA, 2004; Vandenberg: Personal communication). The pathogen causes leaf blight, plant defoliation and death.

Lentil rust caused by *U. fabae* is widespread in South Asia, Morocco and Ethiopia and is characterized by lesions on the stems and leaves, leaf drop and premature plant death (Ahmed and Morral, 1996; Ahmed *et al.,* 1996). Losses from the disease, estimated up to 70%, have been reported (Erskine and Sarker, 1997; Negussie *et al.,* 1998). Resistance to the disease has been identified in germplasm line ILL5588 and is currently being used as a source of resistance in breeding programmes and in studies to determine the inheritance of resistance and to map the important genes.

Sclerotinia white mold caused by *S. sclerotiorum* is responsible for extensive damage to lentil crop in areas that are relatively moist and humid. Dense crop canopies also contribute to the severity of the disease. Root rots and wilt usually

attack lentil plants in the seedling stage and cause seed rot, damping off, wilt, destruction of the root system and rotting of lower stems. These disease problems have been extensively reviewed (Khare, 1981; Kraft *et al.,* 1988, 1994). *Fusarium* wilt caused by *F. oxysporum* is considered to be the most damaging soil borne disease of lentil worldwide (Khare, 1981) and several races of the pathogen can cause the disease. Hamwieh *et al.,* (2005) developed a genetic map of lentil that comprises over 300 molecular markers and used the map to determine the location of the gene for resistance to *Fusarium* wilt.

Abiotic constraints

Stresses that affect lentil are cold, drought, heat, salinity, nutrient deficiency and nutrient toxicity. Of these stresses, drought and terminal heat are considered as most important worldwide (Turner *et al.,* 2001). Cold stress was considered important in the West Asia-North Africa (WANA) region. Salinity is an important stress factor in the Indian sub-continent and to some extent in WANA. Nutrient deficiency and nutrient toxicity is of lesser importance world wide but important in localized region.

Field pea (*Pisum sativum)*

Pea (2n=2x=24) is a member of the galegoid (cool-season) legume clade. In India, pea is an important *rabi* pulse crop grown in about 0.76 m ha with annual production of 0.70 m tonnes. Uttar Pradesh and Madhya Pradesh are the leading field pea growing states that contribute 49% to whole country production. Bihar, Assam, Maharashtra and Orissa are also major field pea growing states. It is widely cultivated in northern temperate regions. Besides their use as a fresh or canned green vegetable, peas are also grown for animal fodder and for consumption as a split dry pulse or *dal* (used much like lentil). Peas are genetically highly diverse, with genetic contributions coming from three inter fertile species or sub-species native to the Mediterranean basin and the Near-East.

Available genomic and genetic resources

Recent investigations using next generation sequencing (NGS) data confirmed the occurrence of highly diverse families of repeats and revealed that about 50–60% of pea nuclear DNA is made up of highly to moderately repeated sequences (Novak *et al*., 2010). In these studies, similarity based clustering of sequence reads obtained from low-pass whole genome sequencing provided a global overview about repeat composition, identifying *Ty3/gypsy* LTR-retrotransposons as the main

component of the pea repeats. Ogre elements alone were estimated to represent 20–33% of the pea genome. Other lineages of *Ty3/gypsy* and *Ty1/copia* elements as well as other types of repeats were found at low frequency in the genome. Ogre retrotransposons represent a phylogenetically distinct clad of the Tat lineage of *Ty3/gypsy* elements and were first identified in pea. Due to their size and amplification to high copy numbers, Ogre elements are important genome constituents in other legume species. Pea repeats have been the subject of a number of studies focusing on individual elements, including LTR-retrotransposons Cyclops, PDR, PIGY, Angela, MITE elements Zaba, Stowaway, and a group of centromeric retrotransposons. Pea has a high number of diverse satellite repeats; some of the satellites provide useful cytogenetic markers allowing discrimination of individual chromosomes within the karyotype.

Major biotic constraint

The productivity of pea is limited by large number of biotic stresses including fungal, viral, bacterial pathogens causing diseases and various insect-pests and nematodes. Among diseases, fungal diseases are most common and the destructive disease are powdery mildew, rust, root rots, wilt, stem pod rot, ascochyta blight, *etc*.

Powdery mildew

Powdery mildew is a common disease of field pea. It occurs all over the world and can cause severe damage in areas where pea is cultivated (Kraft and Pfleger, 2001). Powdery mildew is caused by *Erysiphe pisi*, which is a biotrophic ascomycete's fungus. The infection starts with small diffuse spots on the upper surface of leaflets or stipules and may quickly cover the plant surfaces completely. The disease affects plants in several ways: it has a negative influence on plant weight and height and may cause plant death, thus reducing yield by as much as 25-30%.

Many studies have been conducted on *Pisum* germplasm collections to investigate genetic and trait diversity. Several major world pea germplasm collections have been analyzed by molecular methods and core collections were formed. For example, genetic diversity has been assessed in over 2000 accessions from the Chinese collection using 21 SSR loci, 310 USDA pea accessions have been assessed using 37 RAPD and 15 SSR markers, and INRA France used an extensive set of 121 isozyme, RAPD, EST and SSR markers to genotype 148 accessions. Additional examples include, analysis of 60 pea cultivars grown in Canada with RAPD, ISSR and SSR markers, analysis of the entire JIC pea germplasm collection (3029 accessions) consisting of a broad balance of cultivars (33%), landraces (19%) wild accessions (13%) and genetic stocks (26%), using 45 retrotransposon-based insertion polymorphism (RBIP) markers; and genotyping of 1283 pea accessions

representing much of the cultivated pea diversity held at the Czech National Pea Germplasm collection (CzNPC) using a combination of 25 RBIPs and 10 SSRs.

Rust

Pea rust is a serious disease of worldwide distribution and has become an important pathogen of dry pea from the mid-1980s. The disease is caused either by *Uromyces viciae-fabae(pers.)* j.Schort (syn. *Uromyces fabae* pers. de bary) (Arthur.1934) or *U. pisi* (pers. Wint).

Recently, Rai *et al*. (2011) reported that pea rust caused by *Uromyces fabae* (Pers.) de-Bary is a major problem in warm humid regions causing huge economic loss. They developed a mapping population of 136 F (6:7) recombinant inbred lines (RILs) derived from the cross between HUVP 1 (susceptible) and FC 1 (resistant) pea genotypes and evaluated population in polyhouse as well as also in field conditions during two consecutive years. Assessment of rust reaction of RILs were carried out with the help of infection frequency (IF) and area under disease progress curve (AUDPC). A linkage map was constructed with 57 polymorphic loci selected from 148 simple sequence repeats (SSRs), 3 sequence tagged sites (STS), and 2 random amplified polymorphic (RAPD) markers covering 634 cM of genetic distance on the seven linkage groups of pea with an average interval length of 11.3 cM. Rai *et al*. (2011) revealed one major (Qruf) and one minor (Qruf1) QTL for rust resistance on LGVII by using composite interval mapping (CIM). They also found two flanking SSR markers AA505 and AA446 (10.8cM) for major QTL. The minor QTL was environment specific and it was detected only in the polyhouse. It was flanked by SSR markers, AD146 and AA416 (7. 3 cM). The major QTL Qruf was consistently identified across all the four environments. Therefore, the SSR markers flanking Qruf would be useful for marker assisted selection for pea rust (*U. fabae*) resistance.

Common bean (*Phaseolus vulgar*is)

Common bean (2n = 2 x = 22) is a major legume crop with significant nutritional importance. It is a major source of calories and protein source in many developing countries throughout the world (FAO: http: / / faostat.fao.org /). For countries such as Burundi and Rwanda, with some of the lowest total calory intakes per day, common bean provides about 15% of the total daily calories. It also provides greater than 30% of the daily protein intake per day in these countries. In these poorer countries, malnutrition is one aggravating factor for AIDS patients. In particular, micronutrient deficiencies and HIV-1 disease progression are associated (Baum

et al., 1995). Common bean is a rich source of zinc and iron, two micronutrients depleted from individuals with AIDS (Savarino *et al.,* 1999; Buys *et al.,* 2002). Diets containing foods rich in these micronutrients are suggested to benefit the health status of HIV infected patients (ADA 2004; Kruzich *et al.,* 2004) which in turn delays the onset of AIDS. Common bean also contains a protein that inhibits the HIV-1 reverse transcriptase (Wong *et al.,* 2006). Collectively, these features support the importance of common bean as one of the many factors that can address the AIDS problem through improved nutrition. The immediate benefit of improved nutrition will be to improve food security in these countries (Gillespie and Kadiyala, 2005). When placed in this perspective, the value of common bean is best seen through its role as a crop worthy of aggressive improvement using the tools that can be generated by the availability of a whole-genome sequence.

Available genomic resources

ESTs

A total of 83,000 common bean ESTs are currently available (Ramirez *et al.,* 2005; Melotto *et al.,* 2005; Thibivilliers *et al.,* unpublished). This EST set was analyzed and 11,000 contigs and 9,000 singletons were discovered. Multiple *P. vulgaris* genotypes were used to obtain the EST data which in turn facilitates SNP discovery. EST data was collected from seedling shoots [with or without *Colletotrichum lindemuthianum* (anthracnose) infection], seedling leaves, nodules elicited by *Rhizobium tropici*, roots, leaves (three genotypes) and pods. In *P. coccineus*, ESTs were isolated from the suspensor regions in globular-stage embryos six days after pollination and because of the close relationship between the two species, sequences in *P. vulgaris* can be identified through similarity with *P. coccineus* (Nanni *et al*., 2005).

BAC libraries

Eleven BAC libraries are available in the genus *Phaseolus*, ten in *P. vulgaris* and one in *P. lunatus* (Gepts *et al.,* 2007). Based on their average insert size, library coverage varies (5-12X). The BAT93 library has coverage of 20X, in part because it has been designated as the standard genotype for *Phaseolus* genomics (Broughton *et al.,* 2003). The *Phaseolus* BAC libraries are a phylogenetically ordered set useful for evolutionary studies (Gepts *et al.,* 2007). DGD1962 is a wild bean from northern Peru, representing the presumed ancestral gene pool of the species (Debouck *et al.,* 1993; Kami *et al.,* 1995). The remainder of the libraries is representative of the two evolutionary gene pools. G02771 and G12946 are wild Mexican beans of Mesoamerican origin that contain the three sub-families of the APA seed proteins, which confer resistance to seed weevils. G2333 is a Mexican landrace highly resistant

to anthracnose. BAT93, OAC-HR45 and OAC-HR67 are breeding lines and OAC-Rex is a cultivar from the Mesoamerican gene pool. G19833 is an Andean landrace from Peru, whereas Sprite is a bred Andean variety. Using this array of BAC libraries it is possible to study *Phaseolus* genome evolution both prior to and after domestication and analyze phenotypic changes resulting from specific structural modification at the genome level. Single BAC clones have been fully sequenced, one around the *Co-4* locus for resistance to anthracnose (Melotto *et al.,* 2004), and the other around the *APA* locus (Kami *et al.,* 2006).

Genetic maps

Over 25 linkage maps, mostly low density (markers on average every 10 cM), have been developed for common bean (Kelly *et al.,* 2003; Miklas *et al.,* 2006) from the mapping population developed over the periods (Table 6). To maximize molecular polymorphism, the majority of mapping populations were derived from crosses between domesticated parents belonging to the Andean vs. Middle American gene pools. A highly polymorphic core map utilizing a recombinant inbred population from the cross BAT93 x Jalo EEP558 (Nodari *et al.,* 1992) was developed to coalesce the mapping data (Freyre *et al.,* 1998). BAT93 is a breeding line from the Mesoamerican gene pool and Jalo EEP558 is an Andean bean cultivar resulting from selection in a Brazilian landrace. The two parents show contrasting resistances to pathogens. Some 600 markers have been mapped directly in this population, including 71 RFLPs, 161 AFLPs, 158 RAPDs, 50 ISSRs and 200 microsatellites (Freyre *et al.,* 1998; Papa and Gepts, 2003; González *et al.,* 2005; Blair *et al.,* 2003; Grisi *et al.,* 2007).

Table 6: List of mapping population in common bean

Population	Background markers	Total map distance (cM)	References
BC from XR-235-1-1 x Calima	224 (RFLP)	960	Vellejos *et al*., 1992
$75F_2$ from BAT93 x Jalo EEP558	143 (RFLP, RAPD)	827	Nodari *et al*., 1993
RILs from BAT93 x Jalo EEP558	563 (RFLP, RAPD)	1226	Freyre *et al*., 1998
87 RILs from DOR364 x G19833	100 (SSR)	1720	Blair *et al*., 2003

Major constraints

Majority of the bean production occurs under low input agriculture on small-scale farms in developing countries. Beans produced by these resource-poor farmers are more vulnerable to attack by disease and insect pests and to abiotic stresses including drought and low soil fertility. Under the biotic stress, this crop is mainly encountered by diseases like angular leaf spot, anthracnoze, bean common mosaic virus (BCMV), bean common mosaic necrosis virus (BCMNV), beat curly top virus, bean golden yellow mosaic virus, common bacterial blast (CBB), root rot, white mold, *etc.* Insects like bruchids and leaf hopper are major pests of this crop. Abiotic stress by its nature is more complex physiologically, and this crop has been less well studied than biotic stress resistance in common bean (Rao, 2001).

Due to the biotic and abiotic pressure on the crop, development of cultivars with improved resistance to biotic and abiotic stresses is a primary goal of bean breeding programmes throughout the world. The effort is being made towards the tagging of agronomical important traits by the molecular breeders to implement and develop agronomical superior varieties against biotic and abiotic stress through marker assisted breeding (MAS) (Miklas *et al.,* 2006).

Future Prospects of Genomic Research

Significant advancements have been made in the area of pulse genomics during last ten years. This progress has been seen in the form of generation of large scale DNA markers such as SSRs, SNPs and DArTs, highly saturated genetic linkage maps and establishment of association between several traits of economic importance and the underlying QTL(s)/Gene(s). In this way, genomics has enhanced our understanding the complex traits through unraveling molecular basis the underlying mechanisms which would otherwise be difficult through traditional breeding methods. The QTLs expressing large phenotypic effects on important traits have been deployed in routine pulse breeding programmes using MAS and MABC, whereas to tap the smaller effects QTLs some novel genomics based breeding methods like MARS is being practiced in some pulse crops.

In addition, dramatically decreasing cost per data point coupled with the tremendously increased throughput of NGS technologies facilitated *de novo* whole genome assembly of orphan crop like pigeonpea. Like pigeonpea, several collaborative efforts are underway aiming at whole genome sequencing of some

other important pulse crops including chickpea and lentil. Availability of whole genome sequence has provided access to a variety of economically important genes which can be targeted as 'candidates' for future research. Further re-sequencing of landraces and wild cultivars would help tremendously in unraveling genes playing a crucial role in the complex process of domestication. This would probably provide answer related to continuous narrowing down of genetic diversity in cultivated germplasm pool witnessed in most of the pulses. The superior alleles from wild progenitors can also be harnessed using advanced backcross (AB)-QTL method facilitating identification and introgression of superior alleles at the same time.

A reference genome offers ways to practice several high throughput genotyping methods such as restriction site associated DNA (RAD) sequencing and genotyping by sequencing (GBS) allowing simultaneous discovery and mapping of several thousand of genetic markers in a rapid and cost effective way. Apart from this, a quick access to genome wide genetic markers would open new avenues for genome wide association studies (GWAS) and genomic selection (GS) or whole genome selection (WGS). GS allows identification of a superior breeding line without repeated field testing at several locations. Once a robust phenotyping data has been recorded for a training population, then calculated genomics estimated breeding values (GEBVs) can be used for recovery of superior genotypes from breeding populations. However, the availability of robust phenomics platforms still poses a big obstacle in large scale implementation of these high throughput molecular breeding methods.

The developed genomic tools and technologies can be a great supplement to the conventional breeding to experience a quantum leap in the genetic gains of pulse crops. Strong interfaces are needed to be developed among the pulse breeders, biotechnologists and bioinformaticians so that the enormous amount of the genomics information which has been accumulated in last years could be translated into superior high yielding cultivars.

References

1. Bohra A, Dubey A, Saxena RK, Penmetsa RV, Poornima KN, Kumar N, Farmer AD, Srivani G, Upadhyaya HD, Gothalwal R, Ramesh R, Singh D, Saxena KB, Kavi Kishor PB, Singh NK, Town CD, May GD, Cook DR, Varshney RK (2011): Analysis of BAC-end sequences (BESs) and development of BES-SSR markers for genetic mapping and hybrid purity assessment in pigeonpea (*Cajanus* spp.). *BMC Plant Biol.* **11:** 56.

2. Bohra A, Saxena RK, Gnanesh BN, Saxena KB, Byregowda M, Rathore A, KaviKishor PB, Cook DR, Varshney RK (2012): An intra-specific consensus genetic map of pigeonpea [*Cajanus cajan* (L.) Millspaugh] derived from six mapping populations. *Theor Appl Genet.* doi 10.1007 / s00122-012-1916-5.

3. Millan T, Winter P, Ju¨ngling R, Gil J, Rubio J, Cho S, Cobos MJ, Iruela M, Rajesh PN, Tekeoglu M, Kahl G, Muehlbauer FJ (2010): A consensus genetic map of chickpea (*Cicer arietinum* L.) based on 10 mapping populations. *Euphytica* **175:** 175–189.

4. Nayak SN, Zhu H, Varghese N, Datta S, Choi H, Horres R, Jungling R, Singh J, Kavi Kishor PB, Sivaramakrishnan S, Hoisington DA, Kahl G, Winter P, Cook DR, Varshney RK (2010): Integration of novel SSR and gene-based SNP marker loci in the chickpea genetic map and establishment of new anchor points with *Medicago truncatula* genome. *Theor Appl Genet.* **120:** 1415–1441.

5. Phan HTT, Ellwood SR, Hane JK, Ford R, Materne M, Oliver RP (2007): Extensive macrosynteny between *Medicago truncatula* and *Lens culinaris* ssp. *culinaris. Theor Appl Genet.* **114:** 549–558

6. Phillip, N., Miklas, James, D Kelly., Steve E., Beebe and Matthew, W Blair (2006): Common bean breeding for resistance against biotic and abiotic stresses: From classical to MAS breeding. *Euphytica* (2006) **147:** 105–131

7. Rubeena, Taylor PWJ, Ford R (2003) : Molecular mapping the lentil (*Lens culinaris* ssp. *culinaris*) genome. *Theor Appl Genet.* **107:** 910–916.

8. Singh NK, Gupta DK, Jaiswal PK, Mahato AK, Dutta S, Singh S, Bhutani S, Dogra V, Singh BP, Kumawat G, Pal JK, Pandit A, Singh A, Rawal H, Kumar A, Prashat RG, Khare A, Yadav R, Raje RS, Singh MN, Datta S, Fakrudin B, Wanjari KB, Kansal R, Dash PK, Jain PK, Bhattacharya R, Gaikwad K, Mohapatra T, Srinivasan R,Sharma TR (2011) : The first draft of the pigeonpea genome sequence. *J Plant Biochem Biotech.* **21:** 98–112.

9. Takehisa Isemura., Akito Kaga., Satoshi Tabata, Prakit Somta, Peerasak Srinives,Takehiko Shimizu, Uken Jo, Duncan A. Vaughan1, Norihiko Tomooka1(2012) : Construction of a Genetic Linkage Map and Genetic Analysis

of Domestication Related Traits in Mungbean (*Vigna radiata*). *PLoS one* **(7):** e41304.

10. Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, Schlueter JA, Donoghue MTA, Azam S, Fan G, Whaley AM, Farmer AD, Sheridan J, Iwata A, Tuteja R, Penmetsa RV, Wu W, Upadhyaya HD, Yang SP, Shah T, Saxena KB, Michael T, McCombie WR, Yang B, Zhang G, Yang H, Wang J, Spillane C, Cook DR, May GD, Xu X, Jackson SA (2011): Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nat Biotechnology* **30:** 83–89.
11. Varshney RK, Penmetsa RV, Dutta S, Kulwal PL, Saxena RK, Datta S, Sharma TR, Rosen B, Carrasquilla-Garcia N, Farmer AD, Dubey A, Saxena KB, Gao J, Fakrudin B, Singh, MN, Singh BP, Wanjari KB, Yuan M, Srivastava RK, Kilian A, Upadhyaya HD, Mallikarjuna N, Town CD, Bruening GE, He G, May GD, McCombie R, Jackson SA, Singh NK, Cook DR (2010) : Pigeonpea genomics initiative (PGI): An international effort to improve crop productivity of pigeonpea (*Cajanus cajan* L.). *Mol Breed.* **26:** 393–408.
12. Winter P, Pfaff T, Udupa S, Huttel B, Sharma P, Sahi S, Arreguin-Espinoza R, Weigand F, Muehlbauer FJ and Kahl G (1999) : Characterization and mapping of sequence-tagged microsatellite sites in the chickpea (*Cicer arietinum* L.) genome. *Mol Gen Genet.* **262:** 90–101.

Annexure

Information resource	Website
***Lotus* information resources**	
L. japonicus sequencing project	http://www.kazusa.or.jp/lotus/
L. japonicus Gene Index at Harvard University	http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=l_japonicus
Munich Information Center for Protein Sequences (MIPS) *Lotus* genome database	http://mips.gsf.de/proj/plant/jsf/lotus
***Medicago* information resources**	
M. truncatula sequencing resources	http://www.medicago.org/genome/
M. truncatula Consortium Database Version 2.0 (CCGB)	http://www.medicago.org:8180/MtDB2/
University of Oklahoma BACs and annotations (G Browse view; chromosomes 1, 4, 6, 8)	http://www.genome.ou.edu/medicago.html
TIGR BACs and annotations (G Browse view; chromosomes 2, 7)	http://www.tigr.org/tdb/e2k1/mta1/
European *Medicago* consortium (chromosomes 3, 5)	http://medicago.toulouse.inra.fr/
MIPS *Medicago* genome database (UrMeLDB)	http://mips.gsf.de/proj/plant/jsf/medi/
Medicago Gene Index at Harvard University	http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=medicago
Comparative genomics	
LIS at NCGR	http://www.comparative-legumes.org/lis/
Legume DB	http://ccg.murdoch.edu.au/index.php/LegumeDB