# Pigeonpea genomics initiative (PGI): an international effort to improve crop productivity of pigeonpea (*Cajanus cajan* L.)

R. K. Varshney · R. V. Penmetsa · S. Dutta · P. L. Kulwal · R. K. Saxena · S. Datta ·
T. R. Sharma · B. Rosen · N. Carrasquilla-Garcia · A. D. Farmer · A. Dubey · K. B. Saxena ·
J. Gao · B. Fakrudin · M. N. Singh · B. P. Singh · K. B. Wanjari · M. Yuan · R. K. Srivastava ·
A. Kilian · H. D. Upadhyaya · N. Mallikarjuna · C. D. Town · G. E. Bruening · G. He ·
G. D. May · R. McCombie · S. A. Jackson · N. K. Singh · D. R. Cook

**Abstract** Pigeonpea (*Cajanus cajan*), an important food legume crop in the semi-arid regions of the world and the second most important pulse crop in India, has an average crop productivity of 780 kg/ha. The relatively low crop yields may be attributed to non-availability of improved cultivars, poor crop husbandry and exposure to a number of biotic and abiotic stresses in pigeonpea growing regions. Narrow genetic diversity in cultivated germplasm has further hampered the effective utilization of conventional breeding as well as development and utilization of genomic tools, resulting in pigeonpea being often referred to as an 'orphan crop legume'. To enable genomics-assisted breeding in this crop, the pigeonpea genomics initiative (PGI) was initiated in late 2006 with funding from Indian Council of Agricultural Research under the umbrella of Indo-US agricultural knowledge initiative, which was further expanded with financial support from the US National Science Foundation's Plant Genome Research Program and the Generation Challenge Program. As a result of the PGI, the last 3 years have witnessed significant progress in

R. K. Varshney · R. K. Saxena · A. Dubey ·
K. B. Saxena · R. K. Srivastava · H. D. Upadhyaya ·
N. Mallikarjuna
International Crops Research Institute for the Semi-Arid
Tropics (ICRISAT), Patancheru, Andhra Pradesh 502324,
India

R. K. Varshney (✉)
Genomics Towards Gene Discovery Subprogramme,
Generation Challenge Program (GCP), c/o CIMMYT,
Int APDO Postal 6-641, 06600 Mexico DF, Mexico
e-mail: r.k.varshney@cgiar.org

R. V. Penmetsa · B. Rosen · N. Carrasquilla-Garcia ·
J. Gao · G. E. Bruening · D. R. Cook
Department of Plant Pathology, University of California,
354 Hutchison Hall, One Shields Avenue, Davis,
CA 95616-8680, USA

S. Dutta · T. R. Sharma · B. P. Singh · N. K. Singh
National Research Centre on Plant Biotechnology,
IARI Campus, New Delhi 110 012, India

P. L. Kulwal · K. B. Wanjari
Dr. Panjabrao Deshmukh Agricultural University
(PDAU), Krishinagar, Akola, Maharasthra 444 104, India

S. Datta
Indian Institute of Pulses Research (IIPR), Kanpur, Uttar
Pradesh 208024, India

A. D. Farmer · G. D. May
National Center for Genome Resources (NCGR), 2935
Rodeo Park Drive East, Santa Fe, NM 87505, USA

B. Fakrudin
University of Agricultural Sciences, Dharwad, Karnataka
580005, India

M. N. Singh
Banaras Hindu University, Varanasi, Uttar Pradesh
221005, India

M. Yuan · G. He
Tuskegee University, Tuskegee, AL 36088, USA

development of both genetic as well as genomic resources in this crop through effective collaborations and coordination of genomics activities across several institutes and countries. For instance, 25 mapping populations segregating for a number of biotic and abiotic stresses have been developed or are under development. An 11X-genome coverage bacterial artificial chromosome (BAC) library comprising of 69,120 clones have been developed of which 50,000 clones were end sequenced to generate 87,590 BAC-end sequences (BESs). About 10,000 expressed sequence tags (ESTs) from Sanger sequencing and ca. 2 million short ESTs by 454/FLX sequencing have been generated. A variety of molecular markers have been developed from BESs, microsatellite or simple sequence repeat (SSR)-enriched libraries and mining of ESTs and genomic amplicon sequencing. Of about 21,000 SSRs identified, 6,698 SSRs are under analysis along with 670 orthologous genes using a GoldenGate SNP (single nucleotide polymorphism) genotyping platform, with large scale SNP discovery using Solexa, a next generation sequencing technology, is in progress. Similarly a diversity array technology array comprising of ca. 15,000 features has been developed. In addition, >600 unique nucleotide binding site (NBS) domain containing members of the NBS-leucine rich repeat disease resistance homologs were cloned in pigeonpea; 960 BACs containing these sequences were identified by filter hybridization, BES physical maps developed using high information content fingerprinting. To enrich the genomic resources further, sequenced soybean genome is being analyzed to establish the anchor points between pigeonpea and soybean genomes. In addition, Solexa

sequencing is being used to explore the feasibility of generating whole genome sequence. In summary, the collaborative efforts of several research groups under the umbrella of PGI are making significant progress in improving molecular tools in pigeonpea and should significantly benefit pigeonpea genetics and breeding. As these efforts come to fruition, and expanded (depending on funding), pigeonpea would move from an 'orphan legume crop' to one where genomics-assisted breeding approaches for a sustainable crop improvement are routine.

A. Kilian
Diversity Arrays Technology Pty Ltd, 1 Wilf Crane Crescent, Yarralumla, ACT 2600, Australia

C. D. Town
J. Craig Venter Institute (JCVI), 9704 Medical Center Drive, Rockville, MD 20850, USA

R. McCombie
Cold Spring Harbor Laboratory, Watson School of Biological Sciences, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA

S. A. Jackson
Department of Agronomy, Purdue University, West Lafayette, IN 47907-2054, USA

# Pigeonpea crop

Pigeonpea (*Cajanus cajan* [L.] Millspaugh) is an important food legume (or pulse) crop that is predominantly cultivated in tropical and subtropical regions of the world. It is a diploid ($2n = 22$) crop with a genome size of 808 Mbp. Pigeonpea is a drought tolerant crop with large variation for days to maturity, ranging from extra short (90 days) duration to long duration (300 days). It is generally cultivated as a sole crop or as a mixed crop with short maturing cereals or legumes as well as with long duration crops like cotton and groundnut. Globally pigeonpea is cultivated on 4.64 M ha, with an annual production of 3.43 million tons and a mean productivity of 780 kg/ha. India is the primary pigeonpea growing country in the world, accounting for 3.53 M ha area and 2.51 million tons of production. Pigeonpea seeds have 20–22% protein and are consumed as green peas, whole grain or split peas. The seed and pod husks make a quality feed, whereas dry branches and stems serve as domestic fuel. Fallen leaves from the plant provide vital nutrients to the soil and the plant also enriches soil through symbiotic nitrogen fixation.

# Pigeonpea taxonomy

Pigeonpea belongs to subtribe *Cajaninae* of tribe *Phaseoleae* under sub-family *Papilionoideae* of family *Leguminosae*. *C. cajan* is the only domesticated

species under sub-tribe *Cajaninae*. Within *Phaseoleae*, *Cajaninae* is well distinguished by the presence of vesicular glands on leaves, calyx, and pods. Currently, 11 genera are grouped under *Cajaninae*. The members of the earlier genus *Atylosia* closely resembled the genus *Cajanus* in major vegetative and reproductive characters but they were relegated to two separate genera, mainly on the basis of the presence or absence of seed strophiole.

In 1980s, van der Maesen revised the taxonomy of both the genera and merged the genus *Atylosia* in to *Cajanus* (van der Maesen 1980). The revised genus *Cajanus* currently comprises of 18 species from Asia, 15 species from Australia, and one species from West Africa. Of these, 13 are found only in Australia, 8 in the Indian subcontinent, and 1 in West Africa, with the remaining 14 species occurring in more than one country. Based on growth habit, leaf shape, hairiness, structure of corolla, pod size, and presence of strophiole, van der Maesen (1980) grouped the genus *Cajan* into six sections. The 18 erect species were placed under three sections: seven species in *Atylosia*, nine species in section *Fruticosa*, and two species in section *Cajanu*s that consists of the cultivated species along with its progenitor, *C. cajanifolius*. Eleven climbing and creeping species were arranged in two sections, *Cantharospermum* (5) and *Volubilis* (6) and the remaining three trailing species were classified under *Rhynchosoides*. Three *Cajanus* species have been further subdivided into botanical varieties; *C. scarabaeoides* into var. *pedunculatus* and var. *scarabaeoides*; *C. reticulatus* into var. *grandifolius*, var. *reticulatus*, and var. *maritimus*; and *C. volubilis* into var. *burmanicus* and var. *volubilis*.

## Breeding and production constraints in pigeonpea

In pigeonpea, plant growth as well as flowering is highly influenced by the environment. Hence, breeding for wider adaptation, a complex phenomenon is a major issue to be tackled. Although related wild species are a rich reservoir of not only resistance genes against various biotic and abiotic stresses but also of genes responsible for yield components such as pods per plant, length of fruiting branches, and number of primary branches per plant, use of inter-specifics in pigeonpea improvement have been limited. This is due to the poor crossability of cultivated *Cajanus cajan* to

species other than the closest species, *Cajanus cajanifolia* and *C. scaraboides*. Biotechnology approaches, such as in vitro rescue and propagation of wide cross hybrids, in conjunction with the use of bridge crosses, may enable the transfer of novel genes from a wider range of germplasm within and outside the genus *Cajanus*. Ongoing efforts using molecular tools to examine taxonomic relationships within subtribe *Cajaninae* should clarify phylogenetic relationships within the subtribe, and may suggest parsimonious routes for trait introgression.

Despite the importance of pigeonpea in semi-arid regions of the world, little concerted research effort has been directed towards pigeonpea crop improvement. A number of factors are responsible for the poor productivity, including lack of improved cultivars, poor crop husbandry, pests, and diseases. Major diseases include *Fusarium* wilt (*Fusarium udum* Butler), sterility mosaic disease (Sterility mosaic virus) and phytophthora blight (*Phytophthora drechsleri*), and pests such as gram pod borer (*Helicoverpa armigera*), Maruca (*Maruca vitrata*), pod fly (*Melanagromyza obtusa*), plume moth (*Exelastis atomosa*) cause substantial reduction to pigeonpea production every year. Furthermore, sensitivity to abiotic stresses like water-logging, common in this rain fed crop during early stages, and stress from low water conditions in the later stages, and salinity also reduce pigeonpea production. Conventional breeding approaches for pigeonpea improvement have been in use for several decades but have had limited success in overcoming these biotic and abiotic challenges to stable crop production (Varshney et al. 2007; Saxena 2008).

Knowledge of genetic inheritance of yield and related traits plays an important role in deciding breeding strategies and methodologies for crop improvement. In comparison to other economically important crops, relatively less effort has been made to understand the genetics of important traits in pigeonpea. Both additive effects and dominant non-additive effects have been reported as being important in determining yield, plant height, and protein content (Saxena and Sharma 1990). Pleiotropic effects of genes, physiological changes, and highly sensitive nature of pigeonpea towards the environmental changes makes it difficult to interpret the inheritance of yield and associated characters (Byth et al. 1981). Like yield, restoration of male fertility in cytoplasmic-genetic male-sterility (CGMS) based

hybrids is also critical and important trait in pigeon-pea as it governs the viability of hybrid system.

## Current status of pigeonpea breeding research

Breeding in pigeonpea has been more challenging compared to other food legumes due to various crop specific traits. Pigeonpea is an often cross pollinated crop, with an insect-aided natural out crossing range from 20 to 70% (Saxena et al. 1990) that has limited the use of efficient selection and mating designs possible in self-pollinating species. Pure line breeding, population breeding, mutation breeding, and wide hybridization have been used for development of new varieties in pigeonpea and have led to incremental improvements in the yield potential of this crop. To overcome this bottleneck, two genetic male-sterility (GMS) systems were discovered in pigeonpea (Reddy et al. 1978; Saxena et al. 1983). Despite a 30% yield advantage over the non-hybrids, the GMS based hybrids could not be commercialized due to high cost of hybrid seed production. The yield-jump observed in the GMS hybrids encouraged the development of the alternative and more efficient cytoplasmic-genetic male-sterility (CGMS) system (Tikka et al. 1997; Saxena and Kumar 2003; Wanjari and Patel 2003). As a result of intensive hybrid development programme at ICRISAT in collaboration with its partners, the first CMS- based hybrid GTH-1 was released in India in 2004. Another CMS-based pigeonpea hybrid, ICPH 2671 was developed using *C. cajanifolius* (*A4* cytoplasm) at ICRISAT in 2005 (Saxena 2008), that has been released as "Pushkal" by Pravardhan Seeds for cultivation in several states of India such as Andhra Pradesh, Karnataka, Madhya Pradesh, and Maharashtra. Continued hybrid-technology based improvements in pigeonpea yield potential, together with on going efforts to breed for resistance to biotic and abiotic stresses (*Fusarium* wilt, sterility mosaic, pod borer, etc.) are likely to lead to increased area under pigeonpea hybrids, contribute to increased crop returns for farmers and sustainable pigeonpea production.

## The pigeonpea genomics initiative

Although pigeonpea improvement through conventional breeding and hybrid technology (Saxena and

Kumar 2003) is ongoing, molecular breeding should accelerate utilization of the substantial variability among the pigeonpea landraces and germplasm lines for various morphological, physiological, and agronomic traits. The genetic basis of most important traits in pigeonpea is not known and to date, no linkage map has been reported. This may be attributed to: (1) low levels of DNA polymorphism within the primary (cultivated) gene pool, and (2) very small number of molecular markers available (Burns et al. 2001; Yang et al. 2006; Odeny et al. 2007, 2009; Saxena et al. 2009a). To address the need for genomic tools in pigeonpea, the pigeonpea genomics initiative (PGI) has focused on the development of a robust set of polymorphic markers including microsatellite or simple sequence repeats (SSRs; Gupta and Varshney 2000), single nucleotide polymorphisms (SNP), and diversity arrays technology (DArT) markers. Use of these molecular markers in diverse mapping populations in pigeonpea will facilitate the construction of a genetic map, mapping, and map based cloning of disease resistance genes, quantitative trait loci (QTL) mapping, and the integration of phenotypic data across the different mapping populations. Simultaneously, there was a need to develop mutant lines and a large DNA-insert library e.g., bacterial artificial chromosome (BAC) library to enable map-based cloning and functional analysis of traits in pigeonpea.

To address these needs, the Indian Council of Agricultural Research (ICAR) and the Government of India, under the umbrella of Indo-US Agricultural Knowledge Initiative (AKI), floated the Pigeonpea Genomics Initiative in November 2006. Initial partners in the initiative were National Research Centre for Plant Biotechnology (NRCPB), New Delhi; Indian Institute of Pulses Research (IIPR), Kanpur; Dr Panjabrao Deshmukh Agricultural University (PDAU), Akola; University of Agricultural Sciences, Dharwad (UAS-D); Banaras Hindu University (BHU), Varanasi; and International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru from India, and the University of California, Davis (UC-Davis) from USA. Subsequently, as a result of funding from the Generation Challenge Program (GCP) of the Consultative Group on International Agricultural Research (CGIAR) or through informal collaborations, additional partners joined the PGI, including National Centre for Genome Resources (NCGR), Santa Fe, New Mexico; Tuskegee

University, Tuskegee, Alabama; Purdue University, West Lafayette, Indiana; The J. Craig Venter Institute (JCVI), Maryland; Cold Spring Harbour Laboratory (CSHL), New York, from the USA, and Diversity Array Technology Pty Ltd., Yaramulla from Australia.

The overall aim of AKI-PGI consortium is to convert this so called orphan tropical legume crop into one with genetic and genomic resources and allow for knowledge-based rapid pigeonpea variety improvement. The work plan of the consortium was grouped in drafted into four phases (Fig. 1): Phase I, which deals with genetic and genomic resource development; Phase 2, dealing with developing genetic maps, trait mapping and initiation of gene discovery and transcriptome/genome sequencing; Phase 3, plans for large scale genome sequencing, and Phase 4, utilizing the genomic and genetic resources for crop improvement through molecular breeding or transgenic approaches. Importantly, the consortium selected Asha (ICPL 87119), a widely cultivated medium duration Indian variety that is resistant to *Fusarium* wilt and sterility mosaic disease, as the reference strain for the development of genetic as well as genomic resources. Although the PGI was started in late 2006, significant progress has already been made under the Phases 1 and 2 (described below) and further progress in activities under Phase 2, 3, and 4 will depend on the extent of funding from ICAR and/or other funding agencies.

## Achievements in pigeonpea genetics and genomics

### Genetic resources

The availability of appropriate genetic resources is a pre-requisite for the effective use of genomics derived tools in any crop species (Varshney et al. 2005a). Therefore the PGI consortium planned from the beginning to develop a suitable set of genetic resources. Significant progress has been made during the last <3 years in developing a large number of populations and for genetic mapping and reverse genetic analysis.

### Mapping populations

Although some mapping populations were available at the onset of the PGI, several partner institutes initiated efforts to develop a rational set of mapping populations suited for the molecular tagging of various biotic and abiotic stresses in pigeonpea (Table 1). Regionally adapted elite cultivars of interest to PGI partners were evaluated with SSR markers to select a diverse set of parents (Saxena et al. 2009b), and where phenotypes reflecting the main production constraints in pigeonpea such as *Fusarium* wilt, sterility mosaic disease (SMD), drought, and water logging, were segregating (Table 2). Furthermore, with an objective of developing high density reference genetic maps, one inter-specific [ICP 28 (*C. cajan*) × ICPW 94



Fig. 1 Scheme of PGI for generating genetic and genomic resources under four phases
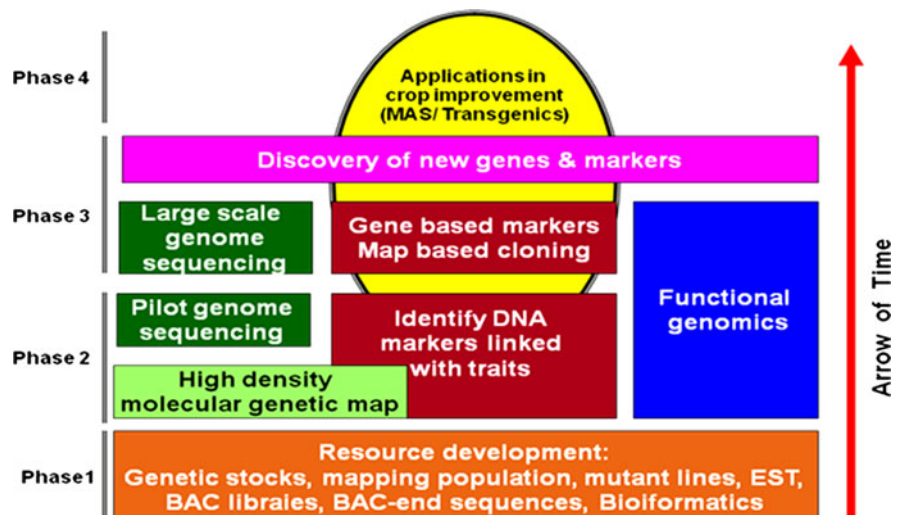
**Table 1** Current status on development of pigeonpea mapping populations at different collaborating centers

| Sr. no. | Crossing parents | Generation | Size of population | Important segregating traits |
|---|---|---|---|---|
| *International Crops Research Institute for the Semi Arid Tropics (ICRISAT), Hyderabad* | | | | |
| 1 | ICPB 2049 × ICPL 99050 | $F_3$ | 360 | *Fusarium* wilt |
| 2 | ICPL 20096 × ICP 332 | $F_3$ | 360 | *Fusarium* wilt and sterility mosaic |
| 3 | ICPL 20097 × ICP 8863 | $F_3$ | 360 | Sterility mosaic |
| 4 | ICPL 87119 × ICPL 87091 | $F_2$ | 600 | *Fusarium* wilt and sterility mosaic |
| 5 | ICP 7035 × ICPL 332 | $F_2$ | 400 | Sterility mosaic |
| 6 | ICPL 88034 × ICPL 84023 | $F_1$ | 10 | Water-logging |
| 7 | ICP 28 × ICPW 94 | $F_2$ | 80 | Pod borer |
| 8 | ICPA 2043 × ICPR 3467 | $F_2/BC_1F_1$ | 252/140 | Fertility restoration |
| 9 | ICPA 2043 × ICPR 2671 | $F_2/BC_1F_1$ | 238/138 | Fertility restoration |
| 10 | ICPA 2039 × ICPR 2447 | $F_2/BC_1F_1$ | 221/196 | Fertility restoration |
| 11 | ICPA 2039 × ICPR 2438 | $F_2/BC_1F_1$ | 225/183 | Fertility restoration |
| *Dr. Panjabrao Deshmukh Agricultural University (PDAU), Akola* | | | | |
| 12 | TAT10 × BSMR736 | $F_3$ | 225 | *Fusarium* wilt and sterility mosaic, morphological traits |
| 13 | Asha × TV1 | $F_3$ | 304 | *Fusarium* wilt and sterility mosaic |
| 14 | AKT 8811 × BSMR 736 | $F_3$ | 246 | *Fusarium* wilt and sterility mosaic |
| 15 | GT 288 × C 11 | $F_3$ | 329 | *Fusarium* wilt and morphological traits |
| *University of Agricultural Sciences (UAS), Dharwad* | | | | |
| 16 | Gullyal white × Maruti | $F_3$ | 394 | *Fusarium* wilt, morphological traits, seed colour |
| 17 | Gullyal white × BSMR 736 | $F_3$ | 329 | Sterility mosaic, seed colour |
| 18 | Asha × Andola black | $F_2$ | – | Drought tolerance |
| 19 | Asha × Gulyal red | $F_2$ | – | *Fusarium* wilt |
| *Indian Institute of Pulses Research (IIPR), Kanpur* | | | | |
| 20 | Asha × UPAS 120 | $F_3$ | 216 | For Reference map and *Fusarium* wilt |
| 21 | Bahar × 67B | $F_3$ | 200 | Morphological traits |
| 22 | IPA6-1 × UPAS 120 | $F_3$ | 200 | *Fusarium* wilt |
| *Banaras Hindu University (BHU), Varanasi* | | | | |
| 23 | MAL 13 × MA Deo 74 | $F_3$ | ∼1,000 | Sterility mosaic |
| 24 | NDA 1 × MA 6 | $F_3$ | ∼1,000 | Sterility mosaic |
| 25 | MAL 13 × ICPL 9150 | $F_3$ | ∼1,000 | Morphological traits |

(*C. scarabaeoides*)] and one intra-specific (Asha × UPAS120) mapping populations have also been developed.

As mentioned earlier, ICRISAT in collaboration with various partners has been successful in developing hybrids in pigeonpea; ICRISAT is developing populations for mapping of the fertility restorer (*Rf*) gene for A4 cytoplasm. Identification of fertility restorer lines for a particular cytoplasm is an important requirement for sustainable pigeonpea hybrid production. In this context, additional eight mapping populations ($BC_1F_1$ and $F_2$) have been

developed at ICRISAT for the mapping of *Rf* gene (Table 1). Molecular markers tightly linked with *Rf* gene will help breeders for marker assisted introgression (MSI) of fertility restorer loci into other elite cultivars using marker assisted selection.

### Mutant population

Rapid acquisition of genomic sequence data has elevated a new discipline, functional genomics, which focuses on determination of gene function. To facilitate functional studies in pigeonpea that

**Table 2** Features of the parental genotypes used for developing mapping populations

| Genotype | Salient features |
|---|---|
| ICPB 2049 | Susceptible to *Fusarium* wilt |
| ICPL 99050 | Resistant to *Fusarium* wilt |
| ICPL 20096 | Resistant to *Fusarium* wilt and sterility mosaic |
| ICP 332 | Susceptible to *Fusarium* wilt and sterility mosaic |
| ICPL 20097 | Resistant to sterility mosaic |
| ICP 8863 (Maruti) | Erect, mid late, highly resistant to *Fusarium* wilt and susceptible to SMD, an extensively grown variety in Northern Karnataka and Maharashtra region of India, red seeded genotype |
| ICPL 87119 (Asha) | A high yielding popular variety, matures late, red seeded, susceptible to terminal drought stress in the field; resistant to *Fusarium* wilt and sterility mosaic |
| ICPL 87091 | Susceptible to *Fusarium* wilt and sterility mosaic |
| ICP 7035 | Resistant to sterility mosaic |
| ICPL 88034 | Susceptible to water logging |
| ICPL 84023 | Tolerant to water logging |
| TAT 10 | Erect, extra early, susceptible to *Fusarium* wilt and sterility mosaic |
| BSMR 736 | Spreading, mid-late, green stem, red seeded with yellow flowers; resistant to *Fusarium* wilt and highly resistant to SMD |
| TV 1 | Semi spreading, early, susceptible to *Fusarium* wilt and sterility mosaic |
| AKT 8811 | Semi spreading, early, tolerant to *Fusarium* wilt and susceptible to sterility mosaic |
| GT 288 | Erect, early, susceptible to *Fusarium* wilt and sterility mosaic, white seeded |
| C 11 | Spreading, mid late, resistant to *Fusarium* wilt and susceptible to sterility mosaic, red seeded |
| Gullyal white | A local genotype highly susceptible for *Fusarium* wilt and SMD, flowers early, medium duration, light brown stem with light red flowers, white seeded; good dhal quality |
| Gullyal red | A local genotype highly susceptible for *Fusarium* wilt and SMD, flowers early, medium duration, light brown stem with light red flowers, red seeded, good milling quality, known for field drought tolerance |
| Andola black | A local genotype, flowers early, field tolerance to drought stress |
| UPAS 120 | Indeterminate; Early, susceptible to wilt |
| Bahar | Compact; Late, yellow flower; flat and deep purple pods; brown seeds; susceptible to wilt |
| 67B | Determinate; dwarf, early ($\sim 100$ cm), susceptible to wilt |
| IPA6-1 | Indeterminate; late, tall (>250 cm), resistant to wilt |
| MAL 13 | Spreading; green stem; light yellow flowers; green pods; brown seeds; resistant to sterility mosaic |
| MA Deo 74 | Compact; yellow flower with purplish streaks; green pods with brown seeds; susceptible to sterility mosaic |
| NDA 1 | Compact; yellow flower with purplish streaks; green pods with brown oval seeds; susceptible to sterility mosaic |
| MA6 | Spreading; yellow flower; dark purple pods; brown-slightly rectangular seeds, resistant to sterility mosaic |
| ICPL 9150 | Compact, purple stem; yellow flower; green pods; creamy white seeds; moderately resistant to sterility mosaic |

would follow from genome sequencing, PGI has initiated the development of a TILLING (Targeted Induced Local Lesions in Genomes; McCallum et al. 2000)-based reverse genetic resource for pigeonpea. TILLING is a reverse genetic approach where a library of saturation mutagenized individuals, each with several hundred-low 1,000 s of point mutations, are screened by high-throughput genotyping to identify individuals harboring a range of single nucleotide induced variants in genes of interest. The reference genotype Asha (ICPL 87119) was mutagenized with

ethyl methane sulfonate (EMS) mutagen to develop TILLING population at BHU (Banaras Hindu University). In the pilot experiments, BHU treated 3,000 seeds in each of four different concentrations of EMS (0.01, 0.02, 0.03, and 0.04 M) during the year 2007–2008. As expected, the germination (70%) and pollen fertility (87.9%) were higher with 0.01 M treatment of EMS and reduced down with the subsequent doses. To date, a total of 505 single plant $M_1$ lines yielded fertile $M_2$ seed. In the $M_2$ generation, a number of chlorophyll and plant form (very dwarf, dwarf,

fasciated and tall) mutants have been identified. Efforts to significantly expand such mutant populations to several 1,000–10,000 plant lines are underway.

Genomic resources

A significant amount of genomic resources have become available for pigeonpea during last 3 years (see Varshney et al. 2009a), some of them are discussed below.

*Large-insert genomic DNA library*

A BAC library was produced at UC-Davis for the PGI reference genotype-"Asha". The library was constructed based on partial digestion of high molecular weight DNA separately with *Hin*dIII and *Bam*H1 restriction enzymes. Size-purified fragments were ligated into vector pCC1BAC and transformed into competent cells of *E. coli* strain EPI300. Insert sizes were estimated based on pulsed field gel electrophoresis of *Not*I-digested BAC DNA, as shown in Fig. 2. The *Hin*dIII library is composed of 34,560 clones with an estimated average insert size of 120,000 bp, while the *Bam*H1 library is composed of 34,560 clones with an estimated average insert size of 115 kb. Taken together, the combined clones represent ~11X coverage of the pigeonpea genome. This BAC-library is an important resource for marker development (described below) as well as for physical mapping and to seed future genome sequencing. A set of randomly selected 50,000 BAC clones was subjected to end sequencing by UC-Davis and ICRISAT in collaboration with JCVI, USA. In total, 87,590 BAC end sequences (BESs) were generated, with most of the 50,000 BAC clones containing high quality sequence from both ends. The combined data represent ~56 Mbp of DNA sequences, which were submitted to the National Center for Biotechnology Information (NCBI) Genome Survey Sequence (GSS) database for public access.

*cDNA libraries and expressed sequence tags*

Transcriptome sequencing has been a popular approach to efficiently identify the transcribed portion of the genome (Sreenivasulu et al. 2002). With an objective to identify genes associated with
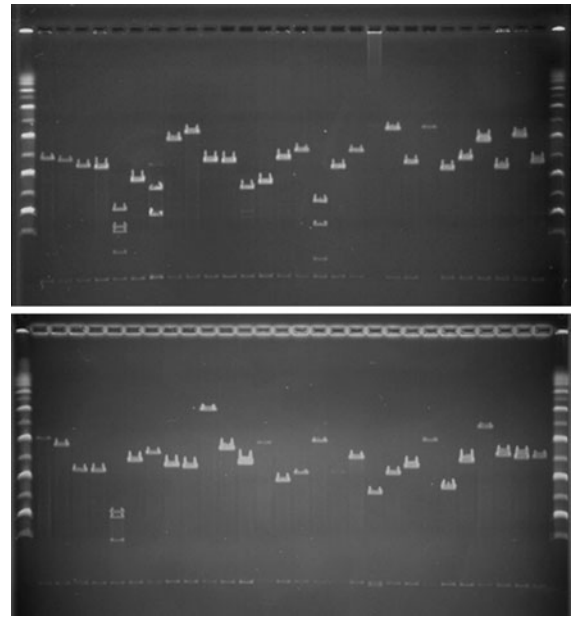


**Fig. 2** Insert sizes in pigeonpea BAC libraries. 28 randomly selected clones from the *Bam*HI (*top panel*) and *Hin*dIII (*bottom panel*) libraries were digested with NotI that cuts at either ends of the insert cloning site. Digested BAC DNA was analyzed by pulse field gel electrophoresis. The first and last *lanes* on each image are molecular weight ladders, starting at 25 kbp. The common fragment at the bottom of each *lane* corresponds to vector (~7 kbp)

*Fusarium* wilt and sterility mosaic diseases, a total of 16 cDNA libraries were generated from *Fusarium udum* and Sterility mosaic virus challenged root tissues of four genotypes (ICPL 20102 and ICP 2376 for FW; ICP 7035 and TTB 7 for SMD). Several thousand cDNA clones from these cDNA libraries were sequenced at ICRISAT (Raju et al., unpublished) to obtain a total of 5,680 expressed sequence tags (ESTs) from FW challenged and 3,788 ESTs from SMD challenged tissues and submitted to NCBI.

In addition to traditional Sanger sequencing to obtain these FW and SMD associated ESTs, next generation sequencing (NGS) technology (Varshney et al. 2009b), was employed to identify whole plant ESTs. cDNA prepared at ICRISAT from 15 tissues representing different developmental stages of the Pusa Ageti genotype were pooled, and normalized to minimize redundancy and enhance efficiency of capture of rare transcripts. ICRISAT in collaboration with JCVI sequenced the normalized cDNA pool using 454/FLX sequencing, a next generation sequencing technology that offers higher throughput relative to

Sanger sequencing. Of 4,96,705 sequence reads 2, 87,766(>50%) were longer than 200 bp. Cluster analysis of these sequences, done at NCGR in collaboration with ICRISAT, provided 48,519 contigs. Similarly, NRCPB has employed 454/FLX sequencing on the cDNA pools from two cultivars, Asha and UPAS120. As a result of this a total of 1,696,724 sequence reads (566 Mb) with an average read length of 333 bp were generated from these two genotypes at NRCPB. Sequence analysis of these genotypes provided 42,000 unique sequences of which 25,000 were common between these two genotypes.

Together these transcript sequences represent a significant fraction of the pigeonpea transcriptome, and should be a useful resource for marker development as well as gene discovery and functional studies.

### Microsatellite/simple sequence repeat markers

Simple sequence repeat markers are the markers of choice for plant genetics and breeding applications (Gupta and Varshney 2000). In case of pigeonpea, however, only 140 SSR markers were available in public domain before the establishment of PGI (Burns et al. 2001; Odeny et al. 2007, 2009). As <10% SSR markers show polymorphism in the intra-specific germplasm, development of an intra-specific map with moderate marker density (with about 300 markers in each intra-specific population) would require availability of at least 3,000 SSR markers. To develop the larger number of SSR markers, three approaches are being used at ICRISAT in collaboration with UC-Davis and other collaborating centers.

*SSR-enriched library* Simple sequence repeat-enriched library have been a popular method to isolate SSRs in several plant species (Gupta and Varshney 2000). Therefore, several genomic DNA libraries enriched for five SSR repeat motifs (CT, TG, AG, AAG, and TCG) were generated from Asha variety using bead capture enrichment protocol (Glenn and Schable 2005). Initially, 1,728 clones were picked from two libraries and 82 clones were sequenced. This pilot experiment provided 36 SSRs from which 23 primer pairs were synthesized of which 16 provided scorable amplification products. Screening of 40 elite genotypes with these 16 markers indicated moderate polymorphism information

content (PIC) values (Saxena et al. 2009a; Table 3). This approach of SSR marker development, however, was subsequently abandoned as SSRs developed in parallel from BAC-end sequences (see below) were found to be more effective.

*BAC-end sequences derived SSRs* In species where BAC-end sequences are available, development of SSR markers from the BAC-end sequences (BESs) is very cost effective (Shultz et al. 2008; Temnykh et al. 2001). SSR development from BAC ends obviates the need for a priori assumptions regarding the nature of the repeat motifs within a species, and offers genome-wide coverage as all repeat types are systematically sampled in the randomly selected BACs. Therefore, all 87,590 pigeonpea BESs were screened with *MISA* (*MI*cro*SA*tellite) search module for identification of SSRs. In total, 18,149 SSRs were identified in 14,001 BESs representing 6,590 BAC clones. 3,124 BESs contained more than one SSR and 2,111 SSRs were present in compound form. Mono- and di-nucleotide were the most abundant classes of SSRs, followed by tri- and tetra-nucleotides SSRs; penta- and hexa-nucleotide SSRs occurred at lower proportions. From a total of 6,590 primer pairs designed 3,072 primer pairs were synthesized and tested. Amplified products were obtained for 2,565 primer pairs (Table 3) and are currently being used at ICRISAT to identify polymorphism in a set of 24 pigeonpea genotypes that are parents of different mapping populations.

*Expressed sequence tags derived SSRs* With the availability of pigeonpea ESTs from transcriptome sequencing described above, it has been possible to identify SSRs from EST sequences. Although expressed sequence tags derived SSR (EST-SSR) markers have been useful for assaying functional genetic diversity in the germplasm, these markers display lower levels of polymorphism relative to SSRs derived from genomic sequences. In case of pigeonpea, the unigene set of 5,085 genes assembled from cluster analysis of the available Sanger ESTs, was searched for the presence of SSRs at ICRISAT to identify 3,583 EST-SSRs that occurred at a frequency of 1/800 bp. 698 ESTs contained more than one SSR and 1,729 SSRs were found as compound SSRs. The majority (3,498, 97.6%) of EST-SSRs were mono-nucleotide repeats, with only a limited number of SSRs of other repeat classes [di- (40), tri- (33), tetra-

**Table 3** Advances in development of SSR markers in pigeonpea under PGI

| Features | SSR enriched library | BAC-end sequences | ESTs | | | Total |
| --- | --- | --- | --- | --- | --- | --- |
| | | | Sanger sequencing | 454 Sequencing | | |
| | | | | ICRISAT | NRCPB | |
| Number of clones | 3,072 | 50,000 | 10,376 | – | – | – |
| Sequence surveyed | 82 | 87,590 | 5,085 | 188,741 | 1,696,724 | – |
| Amount of sequence data (kb) | 60.8 | 56,506.90 | 2,878.30 | 30,876.16 | 56,600.00 | – |
| SSRs identified | 36 | 18,149 | 3,583 | 87,314 | – | 130,850 |
| SSRs frequency (in kb) | 1/1.68 | 1/3.11 | 1/0.80 | 1/0.35 | – | – |
| Primer pairs designed | 23 | 6,590 | 383 | *In progress* | – | 6,996 |
| Primer pairs synthesized | 23 | 3,072 | 84 | – | – | 3,179 |
| Primers amplified | 16 | 2,565 | 52 | – | 400 | 3,033 |

(9), penta- (2), and hexa (1)-nucleotide SSRs]. Primer pairs were designed for 383 SSRs including some mononucleotide SSRs as Saxena et al. (2009a) reported a moderate level of polymorphism for mononucleotide SSRs. Of 84 randomly selected EST-SSR targeting primer pairs, 52 (61.9%) primer pairs showed scorable amplification of which 15 (28.8%) markers showed polymorphism in a set of 40 genotypes. These 15 EST-SSRs identified 2–7 alleles, with the PIC value ranging from 0.20 to 0.70.

As large amount of transcript data have been generated from three genotypes by using 454 sequencing approach, SSR mining has been undertaken in these datasets. For instance, 87,314 SSRs have been identified in 188,741 sequences at ICRISAT. While 12,168 454 sequences contained more than one SSR, 12,679 SSRs were found as compound SSRs (Table 3). In addition to this, by using 454 sequences of two genotypes (Asha and UPAS120), a set of 400 potential polymorphic SSR markers has been identified at NRCPB. Thus in principle, a larger number of primer pairs can be synthesized for SSRs identified in short read sequences generated by 454/FLX machine sequences for enhancing the repertoire of SSR markers for pigeonpea. Although a large number of SSRs could be developed from transcript data, ongoing efforts are focused on the use of the >3,000 set (Table 3) that are predominantly BES-SSRs. Such genomic sequence derived SSRs are more polymorphic relative to EST-SSRs (Varshney et al. 2005b), and offer the additional advantage of allowing for the anchoring of the source BACs to physical and genetic maps.

In summary, the pigeonpea community has an access to >3,000 SSR markers (Table 3). Availability of these markers together with other classes of markers should be a good resource for developing genetic maps and assessment of genetic diversity (Varshney et al. 2009a).

### Single nucleotide polymorphism markers

In recent years, the development and use of SNP markers in plant genetics and breeding has gained popularity compared to SSRs, as SNPs are more abundant and amenable for high-throughput genotyping (Varshney 2009).

*Conserved orthologous sequence based SNPs* With the goal of linking the pigeonpea genome to genomes of other crop and model legume species, UC-Davis in collaboration with its partners have developed ortholog-targeting gene based genetic markers. Orthologous genes were identified from analysis of EST collections of three reference genome legumes, *Medicago truncatula*, *Lotus japonicus,* and *Glycine max*. Low copy transcript clusters with high nucleotide similarity were used to develop 1,536 intron-spanning degenerate primer pairs, with genome sequence of *M. truncatula* and poplar guiding intron-exon junction predictions. Sanger sequencing in *C. cajan* (ICP 28) and *C. scarabaeoides* (ICPW 94), the parents of an inter-specific mapping population developed at ICRISAT, provided high quality sequence for 1,206 loci, and 6,639 single nucleotide polymorphisms (SNPs) in 679 unique conserved orthologous
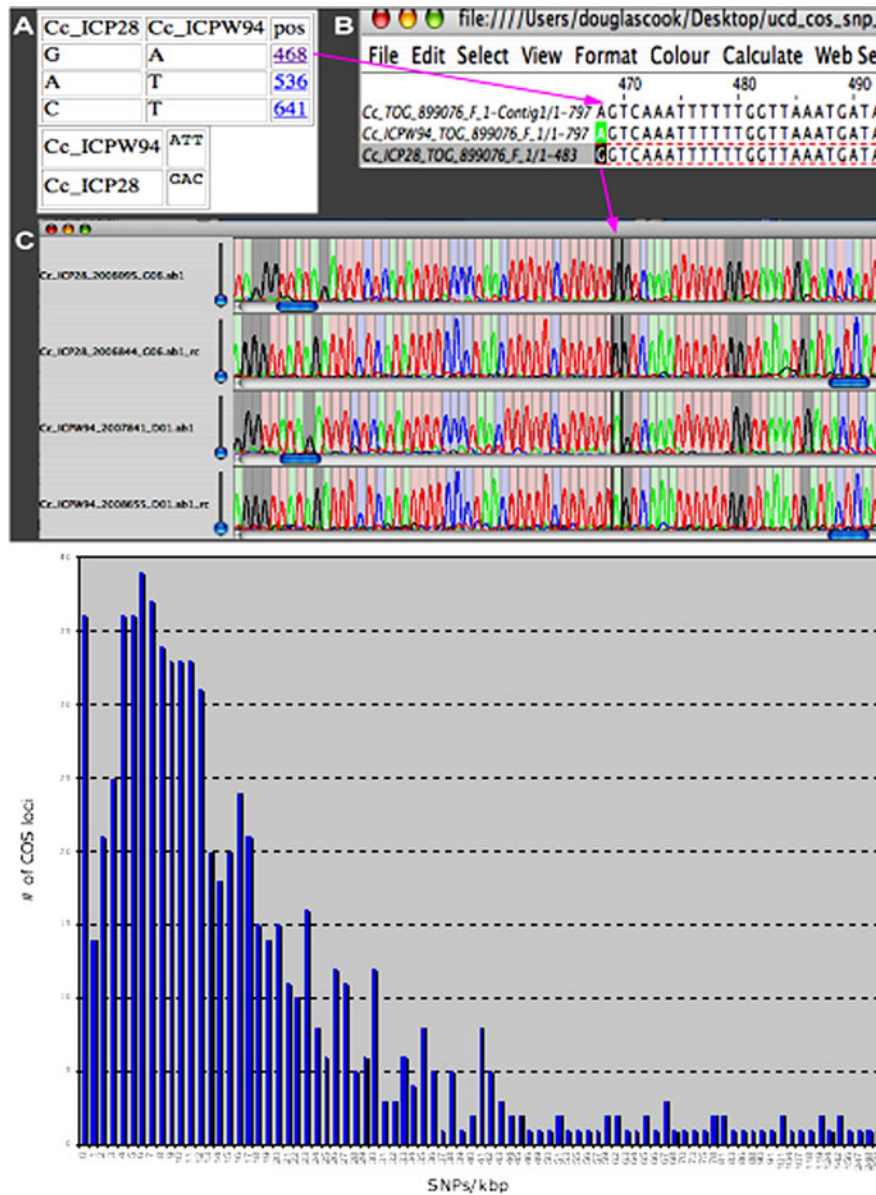
**Fig. 3** Validation and distribution of SNPs in conserved orthologous sequence (COS) markers in pigeonpea. *Top panel*: User interface for manual verification of computationally predicted single nucleotide polymorphisms. Example of SNP position 468 in ortholog 899076 between pigeonpea genotypes ICP28 and ICPW94. **a** SNP table of computationally predicted SNPs; "pos" = nucleotide position. **b** Multiple sequence alignment (MSA) of the region surrounding SNP pos 468 (highlighted *green* [A], *black* [G]). *Uppercase* and *lowercase letters* distinguish high and low quality scores, respectively. **c** Chromatogram window of region flanking the predicted SNP selected in *panel A*. MSA and chromatograms are automatically adjusted and the corresponding SNP highlighted when a new SNP is selected in *panel A*. *Bottom panel*: Frequency distribution of SNP rate in COS sequences between pigeonpea genotypes ICP28 and ICPW94. Counts of SNPs were normalized against amplification product length to obtain SNP/kbp polymorphism rate

sequence (COS) loci with the SNP frequency of 9.8 SNP per gene (Fig. 3). All validated SNPs were assessed for suitability for analysis via GoldenGate genotyping, which yielded 670 unique loci amenable to the high-throughput and parallel "oligo pool all" (OPA) genotyping assay. A single SNP per 670 unique loci, with an additional SNP for 98 loci were selected for production of a 768 SNP OPA assay. Genotyping

with this pigeonpea 768 GoldenGate assay in inter-specific mapping population and a broad set of additional germplasm samples is ongoing, and expected to provide a high density SNP map, and indicate the extent to which the current SNP assay could function in additional populations and genotypes.

*Next generation sequencing based SNPs* Recent developments in NGS technologies are catalyzing the development of SNP markers even in those crops with little or no sequence information (Varshney et al. 2009b). ICRISAT and NCGR have been working to use the Solexa NGS technology to sequence transcriptomes of ten pigeonpea genotypes that are parents of six mapping populations. The availability of reference genome sequence data vastly improves the effectiveness of NGS approaches. In pigeonpea, transcript assembly ($\sim 48,000$ transcript contigs) developed from 454/FLX NGS and Sanger ESTs will facilitate alignment of Solexa sequence data. Multiple sequence alignment (MSA) of Solexa datasets from the ten genotypes, together with reference sequence, should provide a large number of high confidence SNPs for high frequency alleles. SNPs identified from such NGS approaches, together with re-sequencing of COS loci in additional genotypes should allow for the development of additional SNP sets (for example, a 1,536 GoldenGate SNP assay or even larger) for mapping of several hundred SNPs in different intra-specific mapping populations.

*Diversity array technology markers*

Diversity array technology provides a sequence independent and high throughput approach to develop dominant-type markers, and provides a cost-effective whole-genome profiling (Jaccoud et al. 2001). Further, as the same platform is used for discovery and scoring of polymorphic markers it is a quite cost effective and user friendly approach for genotyping of polymorphic markers. DArT has been developed as a technology for whole-genome profiling in over 40 crop species. In pigeonpea, a pilot DArT array, comprising of 5,376 features was developed by Yang et al. (2006). When this array was used to analyze 96 genotypes representing 20 species of *Cajanus*, cultivated genotypes did not show much polymorphism. Under the framework of a recent project sponsored by

Generation Challenge Programme, DArT Pty Ltd in collaboration with ICRISAT, has upgraded the DArT array for pigeonpea that has >15,000 features. These DArT markers are being used to genotype several mapping populations to develop integrated and high-density genetic maps of pigeonpea.

*Isolation and characterization of nucleotide binding site domains*

With an objective of understanding evolution of disease resistance during legume speciation, UC-Davis designed degenerate primer pairs based on NBS-leucine rich repeat disease resistance (NBS-LRR) homologs. Using NBS-LRR sequences from the reference legume *Medicago truncatula*, a set of 544 primers were developed and used in the pigeonpea reference genotype Asha. In excess of 600 unique nucleotide binding site (NBS) sequences were cloned and sequenced, resulting in a large collection of disease resistance genes from both major clades, toll interleukin receptor-like (TIR) and non-TIR NBS domains (Rosen et al., unpublished data). In addition to phylogenetic analysis of NBS sequences, probes from representative NBS clades were used to hybridize to high density nylon filters containing ordered arrays of the pigeonpea BAC library. From 69 probes analyzed to date ($\sim 70\%$ of total), 960 BAC clones were identified, which have been BAC end sequenced and used for physical map construction. A total of 1,805 BAC end sequences were obtained and were submitted to the NCBI GSS database; mining of this BAC end sequence data yielded 151 SSRs. 756 of the 960 BAC clones yielded data when subjected to high information content fingerprinting (HICF) that was analyzed by means of fingerprint contig (FPC) software. 91% of the 756 BAC clones assembled into 90 BAC contigs, while 67 clones remained as singleton BAC clones. An integrated view of BAC contig data, with BAC end SSR annotation has been presented in Fig. 4. Together, these data form the basis for an SSR molecular marker resource that is linked to NBS-LRR contigs, acting as strong candidates for molecular breeding and disease resistance trait dissection.

Harnessing the potential of comparative genomics

Comparative genomics offers the promise of lever-aging genomic information from related species to

**Fig. 4** Physical mapping around NBS-LRR genes using BACs. 960 pigeonpea BACs, identified from Southern hybridization of pigeonpea RGH derived probes to high-density BAC library filters, were fingerprinted via HIC fingerprinting and assembled by FPC. One such contig (# 85) is shown. Note that BAC clones in *blue* have sequence data at both ends, those in *green* have sequence data at one end only, and those in *orange* lack sequence data. All BAC contigs have sequence data associated with them, often at many distinct points separated by 5–10 kbp
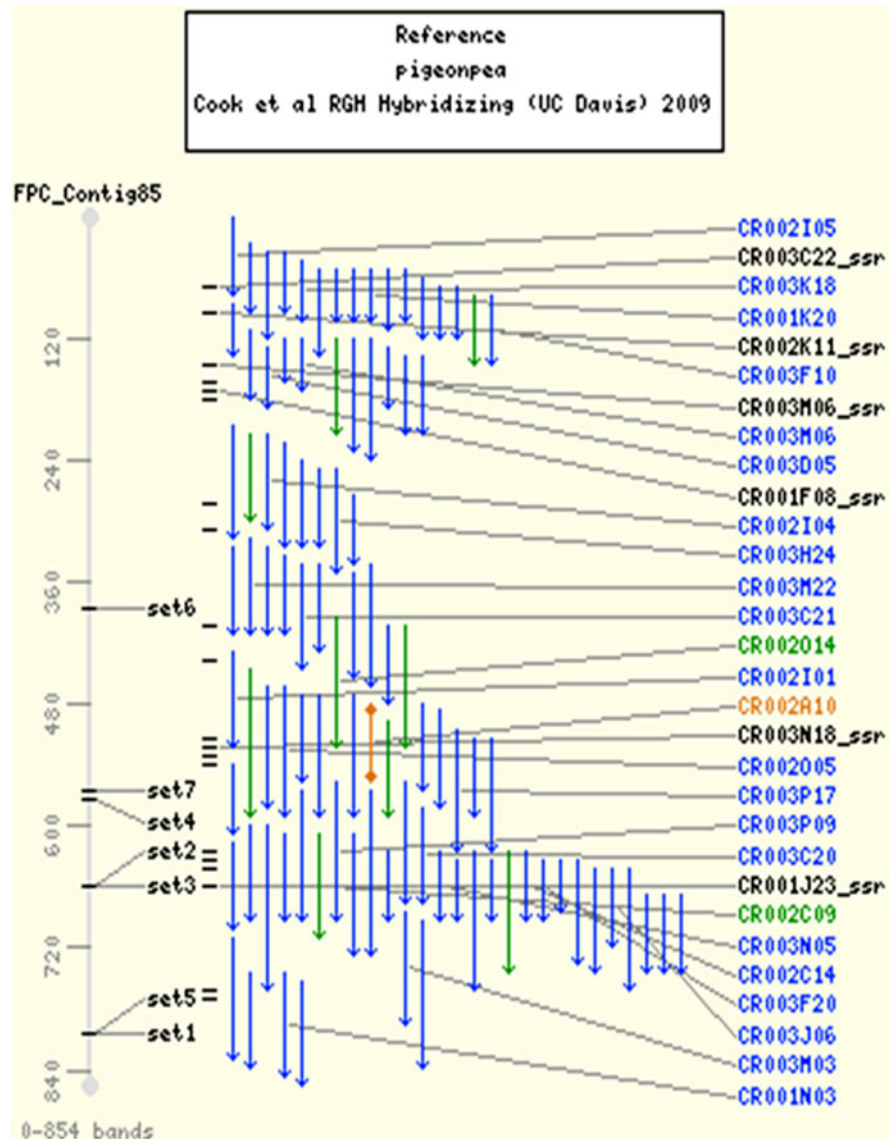


more rapidly advance genetics in species with fewer genetic/genomic resources (O'Brien et al. 1999). Development of COS markers, as mentioned above, is one of those examples of application of comparative genomics approaches. The nearest reference genome (i.e., one with extensive genome sequence) for pigeonpea is soybean (*Glycine max*), another *phaseoloid* legume. Despite this phylogenetic relationship, leveraging soybean to advance pigeonpea genomics may be hampered by polyploidy (Shoemaker et al. 1996; Doyle et al. 2004; Walling et al. 2006) and extensive whole or partial genome duplication and gene fractionation (Schlueter et al. 2006,

2007; Innes et al. 2008). Thus comparisons to any one segment may not be as informative as comparisons to both duplicated segments (McClean et al., unpublished data). Two other reference legume species *Medicago truncatula* and *Lotus japonicus* have extensive genome sequence data, but are estimated to have diverged from the *phaseoloid* clade ~45–55 *my*, compared to 15–20 *my* for the pigeonpea-soybean split (Lavin et al. 2005), which suggests that soybean genome sequencing which is nearing completion (http://www.phytozome.org/soybean), may more readily benefit genetics and crop improvement in pigeonpea.

Towards genome sequencing of pigeonpea

At the inception of the PGI, a clone-by-clone approach was proposed to sequence the pigeonpea genome in Phase 3. However, recent developments in sequencing technologies (Mardis 2008) suggest possible revisions to this approach. NGS technologies can "democratize" genome sequencing for crops with little political and/or research support (Varshney et al. 2009b) such as pigeonpea. Although complete genome sequence would require more extensive resources, we anticipate that existing genomics resources (BACs, BES, transcript sequences, high density molecular maps) together with NGS technology could allow the assembly of a significant fraction of the low copy genic portions of the genome (euchromatin) in the near term, which in itself should revolutionize molecular breeding in pigeonpea. Ongoing rapid advances in sequencing technologies are likely to make complete genome sequencing in pigeonpea achievable in the not too distant future.

## Summary and outlook

In many crop species, genomic tools and approaches have enhanced the precision and efficiency of prediction of phenotype from genotype (Varshney et al. 2005a) and have been instrumental in developing superior genotypes and varieties (Varshney et al. 2006, 2009c). However, until recently appropriate genomic tools were not available in pigeonpea. The PGI consortium, during the last 3 years, has been very successful in developing a significant amount of genetic and genomic resources in pigeonpea, with the majority of the data already in the public domain, or nearly so. Generation of a variety of genetic and genomic tools such as mapping populations, mutant population, different kinds of molecular markers (e.g., SSRs, DArTs, SNPs, COSs) at moderately large scale, BAC library, Sanger, and 454/FLX ESTs, unigenes, NBS-LRR genes, etc. in pigeonpea, from the PGI should have tremendous impact on pigeonpea breeding. For instance, molecular markers together with mapping populations would provide the markers associated with the trait through linkage mapping approach. High-throughput marker genotyping platforms such as DArT markers, GoldenGate assays for SNPs will enable the community to undertake association mapping to identify the markers/gene(s) associated with complex traits by harnessing the genetic variation present in the natural populations and germplasm collections. BAC library and BESs will help develop physical maps to anchor genome sequence data, and to clone genes in concert with transcriptomics resources. Molecular markers identified from these approaches that are associated with traits of importance to breeders should accelerate pigeonpea improvement via marker assisted selection (MAS) or transgenic approaches. Largely through the efforts of the PGI, pigeonpea should move from its current status of an 'orphan' legume to a well-resourced crop species.

In summary, modern molecular breeding methods together with the power of genomics and genetic resources developed through the PGI should revolutionize pigeonpea crop improvement, and consequently benefit farmers and consumers of this important pulse crop of India and the semi-arid regions of the world.

## References

Burns MJ, Edwards KJ, Newbury HJ, Ford LBR, Baggot CD (2001) Development of simple sequence repeat (SSR) markers for the assessment of gene flow and genetic

diversity in pigeonpea (*Cajanus cajan*). Mol Ecol Notes 1:283–285

Byth DE, Wallis ES, Saxena KB (1981) Adaptation and breeding strategies for pigeonpea. Proceedings of ICRISAT/ICAR international workshop on pigeonpea, 15–19 December, vol. 1, pp 450–465

Doyle JJ, Doyle JL, Rauscher JT, Brown AHD (2004) Diploid and polyploid reticulate evolution throughout the history of the perennial soybeans (*Glycine* subgenus *Glycine*). New Phytol 161:121–132

Glenn TC, Schable NA (2005) Isolating microsatellite DNA loci. Methods Enzymol 395:202–222

Gupta PK, Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113: 163–185

Innes RW, Ameline-Torregrosa C, Ashfield T, Cannon E, Cannon SB, Chacko B, Chen W, Couloux A, Dalwani A, Denny R, Deshpande S, Egan AN, Glover N, Hans CS, Howell S, Ilut D, Jackson S, Lai H, Mammadov J, Del Campo SM, Metcalf M, Nguyen A, O'Bleness M, Pfeil BE, Podicheti R, Ratnaparkhe MB, Samain S, Sanders I, Segurens B, Sevignac M, Sherman-Broyles S, Thareau V, Tucker DM, Walling J, Wawrzynski A, Yi J, Doyle JJ, Geffroy V, Roe BA, Maroof MA, Young ND (2008) Differential accumulation of retroelements and diversification of NB-LRR disease resistance genes in duplicated regions following polyploidy in the ancestor of soybean. Plant Physiol 148:1740–1759

Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Res 29:e25

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

Mardis ER (2008) Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet 9:387–402

McCallum CM, Comai L, Greene EA, Henikoff S (2000) Targeting induced local lesions in genomes (TILLING) for plant functional genomics. Plant Physiol 123:439–442

O'Brien SJ, Menotti RM, Murphy WJ, Nash WG, Wienberg J, Stanyon R, Copeland NG, Jenkins NA, Womack JE, Marshall GJA (1999) The promise of comparative genomics in mammals. Science 286:458–481

Odeny DA, Jayashree B, Ferguson M, Hoisington D, Cry LJ, Gebhardt C (2007) Development, characterization and utilization of microsatellite markers in pigeonpea. Plant Breed 126:130–136

Odeny DA, Jayashree B, Gebhardt C, Crouch J (2009) New microsatellite markers for pigeonpea (*cajanus cajan* (L.) millsp.). BMC Res Notes 2:35

Reddy BVS, Green JM, Bise SS (1978) Genetic male sterility in pigeonpea. Crop Sci 18:362–364

Saxena KB (2008) Genetic improvement of pigeonpea—a review. Trop Plant Biol 1:159–178

Saxena KB, Kumar RV (2003) Development of a cytoplasmic nuclear male sterility system in pigeonpea using *C. scarabaeoides* (L.) thours. Ind J Genet Plant Breed 63: 225–229

Saxena KB, Sharma D (1990) Pigeonpea: genetics. In: Nene YL, Hall SD, Sheila VK (eds) The pigeonpea. CAB International, Wallingford, pp 137–158

Saxena KB, Wallis ES, Byth DE (1983) A new gene for male-sterility in pigeonpea (*Cajanus cajan* (L.) Millsp.). Heredity 51:419–421

Saxena KB, Singh L, Gupta MD (1990) Variation for natural out-crossing in pigeonpea. Euphytica 46:143–148

Saxena RK, Prathima C, Saxena KB, Hoisington DA, Singh NK, Varshney RK (2009a) Novel SSR markers for polymorphism detection in pigeonpea (*Cajanus* spp.). Plant Breed. doi: 10.1111/j.1439-0523.2009.01680.x

Saxena RK, Saxena KB, Kumar RV, Hoisington DA, Varshney RK (2009b) SSR-based diversity in elite pigeonpea genotypes for developing mapping populations to map resistance to *Fusarium* wilt and sterility mosaic disease. Plant Breed (in press)

Schlueter JA, Scheffler BE, Schlueter SD, Shoemaker RC (2006) Sequence conservation of homeologous bacterial artificial chromosomes and transcription of homeologous genes in soybean (*Glycine max* L. Merr.). Genetics 174:1017–1028

Schlueter JA, Lin JY, Schlueter SD, Vasylenko SIF, Deshpande S, Yi J, O'Bleness M, Roe BA, Nelson RT, Scheffler BE, Jackson SA, Shoemaker RC (2007) Gene duplication and paleopolyploidy in soybean and the implications for whole genome sequencing. BMC Genomics 8:330

Shoemaker RC, Polzin K, Labate J, Specht J, Brummer EC, Olson T, Young N, Concibido V, Wilcox J, Tamulonis JP, Kochert G, Boerma HR (1996) Genome duplication in soybean (*Glycine* subgenus *soja*). Genetics 144:329–338

Shultz JL, Kazi S, Bashir R, Afzal JA, Lightfoot DA (2008) The development of BAC-end sequence-based microsatellite markers and placement in the physical and genetic maps of soybean. Theor Appl Genet 114:1081–1090

Sreenivasulu N, Kishor PBK, Varshney RK, Altschmied L (2002) Mining functional information from cereal genomes—the utility of expressed sequence tags. Curr Sci 83:965–973

Temnykh S, De Clerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res 11:1441–1452

Tikka SBS, Panwar LD, Chauhan RM (1997) First report of cytoplasmic genic male sterility in pigeonpea (*Cajanus cajan* (L) Millsp.) through wide hybridization. GAU Res J 22:160–162

van der Maesen LJG (1980) India is the native home of the pigeonpea. In: Arends JC, Boelema G, de Groot CT, Leeuwenberg AJM (eds) Libergratulatorius in honerem HCD deWit. Wageningen Agricultural University, The Netherlands, pp 257–262 miscellaneous paper

Varshney RK (2009) Gene-based marker systems in plants: high throughput approaches for discovery and genotyping. In: Jain SM, Brar DS (eds) Molecular techniques in crop improvement, vol 2. Springer, The Netherlands. doi: 10.1007/978-90-481-2967-6_5

Varshney RK, Graner A, Sorrells ME (2005a) Genomics-assisted breeding for crop improvement. Trends Plant Sci 10:621–630

Varshney RK, Graner A, Sorrells ME (2005b) Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23:48–55

Varshney RK, Hoisington D, Tyagi AK (2006) Advances in cereals genomics and applications in crop breeding. Trends Biotechnol 24:490–499

Varshney RK, Hoisington DA, Upadhyaya HD, Gaur PM, Nigam SN, Saxena K, Vadez V, Sethy NK, Bhatia S, Aruna R, Gowda MVC, Singh NK (2007) Molecular genetics and breeding of grain legume crops for the semi-arid tropics. In: Varshney RK, Tuberosa R (eds) Genomics assisted crop improvement, vol II: genomics applications in crops. Springer, Dordrecht, pp 207–242

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

Varshney RK, Nayak SN, May GD, Jackson SA (2009b) Next generation sequencing technologies and their implications for crop genetics and breeding. Trends Biotechnol. doi:10.1016/j.tibtech.2009.05.006

Varshney RK, Thudi M, May GD, Jackson SA (2009c) Advances in legume genomics applied to breeding. Plant Breed Rev (in press)

Walling JG, Shoemaker RC, Young ND, Mudge J, Jackson SA (2006) Chromosome level homeology in paleopolyploid soybean (Glycine max) revealed through integration of genetic and chromosome maps. Genetics 172:1893–1900

Wanjari KB, Patel MC (2003) Fertility restorers isolated from germplasm for cytoplasmic male sterility in pigeonpea. PKV Res J 27:111–113

Yang S, Ash G, Harper J, Varling J, Wenzl P, Huttner E, Kilian A (2006) Low level of genetic diversity in cultivated pigeonpea compared to its wild relatives is revealed by diversity arrays technology. Theor Appl Genet 113:585–595