

1

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

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

10  
11 Received: 30 July 2009 / Accepted: 5 August 2009  
12 © The Author(s) 2009. This article is published with open access at Springerlink.com

13 **Abstract** Pigeonpea (*Cajanus cajan*), an important 24  
14 food legume crop in the semi-arid regions of the world 25  
15 and the second most important pulse crop in India, has 26  
16 an average crop productivity of 780 kg/ha. The 27  
17 relatively low crop yields may be attributed to 28  
18 non-availability of improved cultivars, poor crop 29  
19 husbandry and exposure to a number of biotic and 30  
20 abiotic stresses in pigeonpea growing regions. Narrow 31  
21 genetic diversity in cultivated germplasm has further 32  
22 hampered the effective utilization of conventional 33  
23 breeding as well as development and utilization of 34  
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

A1 R. K. Varshney · R. K. Saxena · A. Dubey ·  
A2 K. B. Saxena · R. K. Srivastava · H. D. Upadhyaya ·  
A3 N. Mallikarjuna  
A4 International Crops Research Institute for the Semi-Arid  
A5 Tropics (ICRISAT), Patancheru, Andhra Pradesh 502324,  
India

A6 R. K. Varshney (✉)  
A7 Genomics Towards Gene Discovery Subprogramme,  
A8 Generation Challenge Program (GCP), c/o CIMMYT,  
A9 Int APDO Postal 6-641, 06600 Mexico DF, Mexico  
A10 e-mail: r.k.varshney@cgiar.org

A11 R. V. Penmetsa · B. Rosen · N. Carrasquilla-Garcia ·  
A12 J. Gao · G. E. Bruening · D. R. Cook  
A13 Department of Plant Pathology, University of California,  
A14 354 Hutchison Hall, One Shields Avenue, Davis,  
A15 CA 95616-8680, USA

A16 S. Dutta · T. R. Sharma · B. P. Singh · N. K. Singh  
A17 National Research Centre on Plant Biotechnology,  
A18 IARI Campus, New Delhi 110 012, India

A19 P. L. Kulwal · K. B. Wanjari  
A20 Dr. Panjabrao Deshmukh Agricultural University  
A21 (PDAU), Krishinagar, Akola, Maharashtra 444 104, India

A22 S. Datta  
A23 Indian Institute of Pulses Research (IIPR), Kanpur, Uttar  
A24 Pradesh 208024, India

A25 A. D. Farmer · G. D. May  
A26 National Center for Genome Resources (NCGR), 2935  
A27 Rodeo Park Drive East, Santa Fe, NM 87505, USA

A28 B. Fakrudin  
A29 University of Agricultural Sciences, Dharwad, Karnataka  
A30 580005, India

A31 M. N. Singh  
A32 Banaras Hindu University, Varanasi, Uttar Pradesh  
A33 221005, India

A34 M. Yuan · G. He  
A35 Tuskegee University, Tuskegee, AL 36088, USA

- 35 development of both genetic as well as genomic  
36 resources in this crop through effective collaborations  
37 and coordination of genomics activities across several  
38 institutes and countries. For instance, 25 mapping  
39 populations segregating for a number of biotic and  
40 abiotic stresses have been developed or are under  
41 development. An 11X-genome coverage bacterial  
42 artificial chromosome (BAC) library comprising of  
43 69,120 clones have been developed of which 50,000  
44 clones were end sequenced to generate 87,590 BAC-  
45 end sequences (BESs). About 10,000 expressed  
46 sequence tags (ESTs) from Sanger sequencing and  
47 ca. 2 million short ESTs by 454/FLX sequencing have  
48 been generated. A variety of molecular markers have  
49 been developed from BESs, microsatellite or simple  
50 sequence repeat (SSR)-enriched libraries and mining  
51 of ESTs and genomic amplicon sequencing. Of about  
52 21,000 SSRs identified, 6,698 SSRs are under analysis  
53 along with 670 orthologous genes using a GoldenGate  
54 SNP (single nucleotide polymorphism) genotyping  
55 platform, with large scale SNP discovery using Solexa,  
56 a next generation sequencing technology, is in pro-  
57 gress. Similarly a diversity array technology array  
58 comprising of ca. 15,000 features has been developed.  
59 In addition, >600 unique nucleotide binding site  
60 (NBS) domain containing members of the NBS-  
61 leucine rich repeat disease resistance homologs were  
62 cloned in pigeonpea; 960 BACs containing these  
63 sequences were identified by filter hybridization, BES  
64 physical maps developed using high information  
65 content fingerprinting. To enrich the genomic  
66 resources further, sequenced soybean genome is being  
67 analyzed to establish the anchor points between  
68 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.
- Keywords** Molecular markers · Genetic mapping ·  
Trait mapping · Genomics · Next generation  
sequencing · Gene discovery · Crop improvement
- 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 fam-  
ily *Leguminosae*. *C. cajan* is the only domesticated

- A36 A. Kilian  
A37 Diversity Arrays Technology Pty Ltd, 1 Wilf Crane  
A38 Crescent, Yarralumla, ACT 2600, Australia
- A39 C. D. Town  
A40 J. Craig Venter Institute (JCVI), 9704 Medical Center  
A41 Drive, Rockville, MD 20850, USA
- A42 R. McCombie  
A43 Cold Spring Harbor Laboratory, Watson School  
A44 of Biological Sciences, 1 Bungtown Road, Cold Spring  
A45 Harbor, NY 11724, USA
- A46 S. A. Jackson  
A47 Department of Agronomy, Purdue University, West  
A48 Lafayette, IN 47907-2054, USA

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

121 In 1980s, van der Maesen revised the taxonomy of  
 122 both the genera and merged the genus *Atylosia* in to  
 123 *Cajanus* (van der Maesen 1980). The revised genus  
 124 *Cajanus* currently comprises of 18 species from Asia,  
 125 15 species from Australia, and one species from West  
 126 Africa. Of these, 13 are found only in Australia, 8 in  
 127 the Indian subcontinent, and 1 in West Africa, with  
 128 the remaining 14 species occurring in more than one  
 129 country. Based on growth habit, leaf shape, hairiness,  
 130 structure of corolla, pod size, and presence of  
 131 strophiole, van der Maesen (1980) grouped the genus  
 132 *Cajan* into six sections. The 18 erect species were  
 133 placed under three sections: seven species in *Atylosia*,  
 134 nine species in section *Fruticosa*, and two species in  
 135 section *Cajanus* that consists of the cultivated species  
 136 along with its progenitor, *C. cajanifolius*. Eleven  
 137 climbing and creeping species were arranged in two  
 138 sections, *Cantharospermum* (5) and *Volubilis* (6) and  
 139 the remaining three trailing species were classified  
 140 under *Rhynchosoides*. Three *Cajanus* species have  
 141 been further subdivided into botanical varieties;  
 142 *C. scarabaeoides* into var. *pedunculatus* and var.  
 143 *scarabaeoides*; *C. reticulatus* into var. *grandifolius*,  
 144 var. *reticulatus*, and var. *maritimus*; and *C. volubilis*  
 145 into var. *burmanicus* and var. *volubilis*.

#### 146 **Breeding and production constraints in pigeonpea**

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

158 species other than the closest species, *Cajanus* 158  
*cajanifolia* and *C. scaraboides*. Biotechnology 159  
 approaches, such as in vitro rescue and propagation 160  
 of wide cross hybrids, in conjunction with the use of 161  
 bridge crosses, may enable the transfer of novel genes 162  
 from a wider range of germplasm within and outside 163  
 the genus *Cajanus*. Ongoing efforts using molecular 164  
 tools to examine taxonomic relationships within 165  
 subtribe *Cajaninae* should clarify phylogenetic rela- 166  
 tionships within the subtribe, and may suggest pars- 167  
 monious routes for trait introgression. 168

169 Despite the importance of pigeonpea in semi-arid 169  
 regions of the world, little concerted research effort has 170  
 been directed towards pigeonpea crop improvement. A 171  
 number of factors are responsible for the poor produc- 172  
 tivity, including lack of improved cultivars, poor crop 173  
 husbandry, pests, and diseases. Major diseases include 174  
*Fusarium* wilt (*Fusarium udum* Butler), sterility 175  
 mosaic disease (Sterility mosaic virus) and phytoph- 176  
 thora blight (*Phytophthora drechsleri*), and pests such 177  
 as gram pod borer (*Helicoverpa armigera*), Maruca 178  
 (*Maruca vitrata*), pod fly (*Melanagromyza obtusa*), 179  
 plume moth (*Exelastis atomosa*) cause substantial 180  
 reduction to pigeonpea production every year. Further- 181  
 more, sensitivity to abiotic stresses like water-logging, 182  
 common in this rain fed crop during early stages, and 183  
 stress from low water conditions in the later stages, and 184  
 salinity also reduce pigeonpea production. Convent- 185  
 ional breeding approaches for pigeonpea improvement 186  
 have been in use for several decades but have had 187  
 limited success in overcoming these biotic and abiotic 188  
 challenges to stable crop production (Varshney et al. 189  
 2007; Saxena 2008). 190

191 Knowledge of genetic inheritance of yield and 191  
 related traits plays an important role in deciding 192  
 breeding strategies and methodologies for crop 193  
 improvement. In comparison to other economically 194  
 important crops, relatively less effort has been made 195  
 to understand the genetics of important traits in 196  
 pigeonpea. Both additive effects and dominant non- 197  
 additive effects have been reported as being impor- 198  
 tant in determining yield, plant height, and protein 199  
 content (Saxena and Sharma 1990). Pleiotropic 200  
 effects of genes, physiological changes, and highly 201  
 sensitive nature of pigeonpea towards the environ- 202  
 mental changes makes it difficult to interpret the 203  
 inheritance of yield and associated characters (Byth 204  
 et al. 1981). Like yield, restoration of male fertility in 205  
 cytoplasmic-genetic male-sterility (CGMS) based 206

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

### 209 **Current status of pigeonpea breeding research**

210 Breeding in pigeonpea has been more challenging  
211 compared to other food legumes due to various crop  
212 specific traits. Pigeonpea is an often cross pollinated  
213 crop, with an insect-aided natural out crossing range  
214 from 20 to 70% (Saxena et al. 1990) that has limited the  
215 use of efficient selection and mating designs possible  
216 in self-pollinating species. Pure line breeding, popu-  
217 lation breeding, mutation breeding, and wide hybrid-  
218 ization have been used for development of new  
219 varieties in pigeonpea and have led to incremental  
220 improvements in the yield potential of this crop. To  
221 overcome this bottleneck, two genetic male-sterility  
222 (GMS) systems were discovered in pigeonpea (Reddy  
223 et al. 1978; Saxena et al. 1983). Despite a 30% yield  
224 advantage over the non-hybrids, the GMS based  
225 hybrids could not be commercialized due to high cost  
226 of hybrid seed production. The yield-jump observed in  
227 the GMS hybrids encouraged the development of the  
228 alternative and more efficient cytoplasmic-genetic  
229 male-sterility (CGMS) system (Tikka et al. 1997;  
230 Saxena and Kumar 2003; Wanjari and Patel 2003). As  
231 a result of intensive hybrid development programme at  
232 ICRISAT in collaboration with its partners, the first  
233 CMS- based hybrid GTH-1 was released in India in  
234 2004. Another CMS-based pigeonpea hybrid, ICPH  
235 2671 was developed using *C. cajanifolius* (A4 cyto-  
236 plasm) at ICRISAT in 2005 (Saxena 2008), that has  
237 been released as “Pushkal” by Pravardhan Seeds for  
238 cultivation in several states of India such as Andhra  
239 Pradesh, Karnataka, Madhya Pradesh, and Maharash-  
240 tra. Continued hybrid-technology based improvements  
241 in pigeonpea yield potential, together with on going  
242 efforts to breed for resistance to biotic and abiotic  
243 stresses (*Fusarium* wilt, sterility mosaic, pod borer,  
244 etc.) are likely to lead to increased area under  
245 pigeonpea hybrids, contribute to increased crop returns  
246 for farmers and sustainable pigeonpea production.

### 247 **The pigeonpea genomics initiative**

248 Although pigeonpea improvement through conven-  
249 tional 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 agro-  
nomic 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 sim-  
ple 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 popula-  
tions 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 chromo-  
some (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 part-  
ners 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 infor-  
mal collaborations, additional partners joined the  
PGI, including National Centre for Genome Resources  
(NCGR), Santa Fe, New Mexico; Tuskegee

299 University, Tuskegee, Alabama; Purdue University,  
 300 West Lafayette, Indiana; The J. Craig Venter  
 301 Institute (JCVI), Maryland; Cold Spring Harbour  
 302 Laboratory (CSHL), New York, from the USA, and  
 303 Diversity Array Technology Pty Ltd., Yaramulla  
 304 from Australia.  
 305 The overall aim of AKI-PGI consortium is to  
 306 convert this so called orphan tropical legume crop  
 307 into one with genetic and genomic resources and  
 308 allow for knowledge-based rapid pigeonpea variety  
 309 improvement. The work plan of the consortium was  
 310 grouped in drafted into four phases (Fig. 1): Phase I,  
 311 which deals with genetic and genomic resource  
 312 development; Phase 2, dealing with developing  
 313 genetic maps, trait mapping and initiation of gene  
 314 discovery and transcriptome/genome sequencing;  
 315 Phase 3, plans for large scale genome sequencing,  
 316 and Phase 4, utilizing the genomic and genetic  
 317 resources for crop improvement through molecular  
 318 breeding or transgenic approaches. Importantly, the  
 319 consortium selected Asha (ICPL 87119), a widely  
 320 cultivated medium duration Indian variety that is  
 321 resistant to *Fusarium* wilt and sterility mosaic  
 322 disease, as the reference strain for the development  
 323 of genetic as well as genomic resources. Although the  
 324 PGI was started in late 2006, significant progress has  
 325 already been made under the Phases 1 and 2  
 326 (described below) and further progress in activities  
 327 under Phase 2, 3, and 4 will depend on the extent of  
 328 funding from ICAR and/or other funding agencies.

**Achievements in pigeonpea genetics and genomics**

329

Genetic resources

330

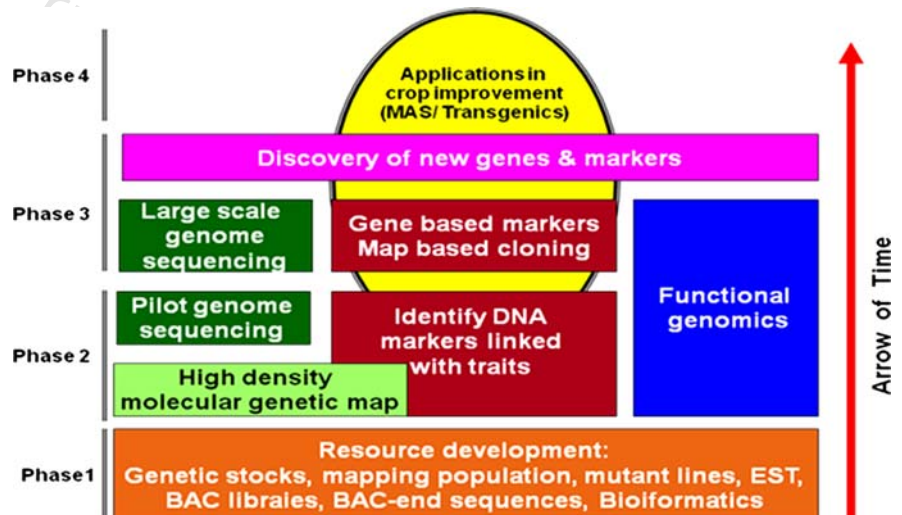
The availability of appropriate genetic resources is a  
 331 pre-requisite for the effective use of genomics  
 332 derived tools in any crop species (Varshney et al.  
 333 2005a). Therefore the PGI consortium planned from  
 334 the beginning to develop a suitable set of genetic  
 335 resources. Significant progress has been made during  
 336 the last <3 years in developing a large number of  
 337 populations and for genetic mapping and reverse  
 338 genetic analysis.  
 339

Mapping populations

340

Although some mapping populations were available at  
 341 the onset of the PGI, several partner institutes initiated  
 342 efforts to develop a rational set of mapping popula-  
 343 tions suited for the molecular tagging of various biotic  
 344 and abiotic stresses in pigeonpea (Table 1). Region-  
 345 ally adapted elite cultivars of interest to PGI partners  
 346 were evaluated with SSR markers to select a diverse  
 347 set of parents (Saxena et al. 2009b), and where  
 348 phenotypes reflecting the main production constraints  
 349 in pigeonpea such as *Fusarium* wilt, sterility mosaic  
 350 disease (SMD), drought, and water logging, were  
 351 segregating (Table 2). Furthermore, with an objective  
 352 of developing high density reference genetic maps,  
 353 one inter-specific [ICP 28 (*C. cajan*) × ICPW 94  
 354

**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
<i>International Crops Research Institute for the Semi Arid Tropics (ICRISAT), Hyderabad</i>				
1	ICPB 2049 × ICPL 99050	F <sub>3</sub>	360	<i>Fusarium</i> wilt
2	ICPL 20096 × ICP 332	F <sub>3</sub>	360	<i>Fusarium</i> wilt and sterility mosaic
3	ICPL 20097 × ICP 8863	F <sub>3</sub>	360	Sterility mosaic
4	ICPL 87119 × ICPL 87091	F <sub>2</sub>	600	<i>Fusarium</i> wilt and sterility mosaic
5	ICP 7035 × ICPL 332	F <sub>2</sub>	400	Sterility mosaic
6	ICPL 88034 × ICPL 84023	F <sub>1</sub>	10	Water-logging
7	ICP 28 × ICPW 94	F <sub>2</sub>	80	Pod borer
8	ICPA 2043 × ICPR 3467	F <sub>2</sub> /BC <sub>1</sub> F <sub>1</sub>	252/140	Fertility restoration
9	ICPA 2043 × ICPR 2671	F <sub>2</sub> /BC <sub>1</sub> F <sub>1</sub>	238/138	Fertility restoration
10	ICPA 2039 × ICPR 2447	F <sub>2</sub> /BC <sub>1</sub> F <sub>1</sub>	221/196	Fertility restoration
11	ICPA 2039 × ICPR 2438	F <sub>2</sub> /BC <sub>1</sub> F <sub>1</sub>	225/183	Fertility restoration
<i>Dr. Panjabrao Deshmukh Agricultural University (PDAU), Akola</i>				
12	TAT10 × BSMR736	F <sub>3</sub>	225	<i>Fusarium</i> wilt and sterility mosaic, morphological traits
13	Asha × TV1	F <sub>3</sub>	304	<i>Fusarium</i> wilt and sterility mosaic
14	AKT 8811 × BSMR 736	F <sub>3</sub>	246	<i>Fusarium</i> wilt and sterility mosaic
15	GT 288 × C 11	F <sub>3</sub>	329	<i>Fusarium</i> wilt and morphological traits
<i>University of Agricultural Sciences (UAS), Dharwad</i>				
16	Gullyal white × Maruti	F <sub>3</sub>	394	<i>Fusarium</i> wilt, morphological traits, seed colour
17	Gullyal white × BSMR 736	F <sub>3</sub>	329	Sterility mosaic, seed colour
18	Asha × Andola black	F <sub>2</sub>	–	Drought tolerance
19	Asha × Gulyal red	F <sub>2</sub>	–	<i>Fusarium</i> wilt
<i>Indian Institute of Pulses Research (IIPR), Kanpur</i>				
20	Asha × UPAS 120	F <sub>3</sub>	216	For Reference map and <i>Fusarium</i> wilt
21	Bahar × 67B	F <sub>3</sub>	200	Morphological traits
22	IPA6-1 × UPAS 120	F <sub>3</sub>	200	<i>Fusarium</i> wilt
<i>Banaras Hindu University (BHU), Varanasi</i>				
23	MAL 13 × MA Deo 74	F <sub>3</sub>	~1,000	Sterility mosaic
24	NDA 1 × MA 6	F <sub>3</sub>	~1,000	Sterility mosaic
25	MAL 13 × ICPL 9150	F <sub>3</sub>	~1,000	Morphological traits

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

358 As mentioned earlier, ICRISAT in collaboration  
 359 with various partners has been successful in devel-  
 360 oping hybrids in pigeonpea; ICRISAT is developing  
 361 populations for mapping of the fertility restorer (*Rf*)  
 362 gene for *A4* cytoplasm. Identification of fertility  
 363 restorer lines for a particular cytoplasm is an  
 364 important requirement for sustainable pigeonpea  
 365 hybrid production. In this context, additional eight  
 366 mapping populations (BC<sub>1</sub>F<sub>1</sub> and F<sub>2</sub>) have been

developed at ICRISAT for the mapping of *Rf* gene 367  
 (Table 1). Molecular markers tightly linked with *Rf* 368  
 gene will help breeders for marker assisted intro- 369  
 gression (MSI) of fertility restorer loci into other elite 370  
 cultivars using marker assisted selection. 371

#### *Mutant population* 372

Rapid acquisition of genomic sequence data has 373  
 elevated a new discipline, functional genomics, 374  
 which focuses on determination of gene function. 375  
 To facilitate functional studies in pigeonpea that 376

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

Genotype	Salient features
ICPB 2049	Susceptible to <i>Fusarium</i> wilt
ICPL 99050	Resistant to <i>Fusarium</i> wilt
ICPL 20096	Resistant to <i>Fusarium</i> wilt and sterility mosaic
ICP 332	Susceptible to <i>Fusarium</i> wilt and sterility mosaic
ICPL 20097	Resistant to sterility mosaic
ICP 8863 (Maruti)	Erect, mid late, highly resistant to <i>Fusarium</i> 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 <i>Fusarium</i> wilt and sterility mosaic
ICPL 87091	Susceptible to <i>Fusarium</i> 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 <i>Fusarium</i> wilt and sterility mosaic
BSMR 736	Spreading, mid-late, green stem, red seeded with yellow flowers; resistant to <i>Fusarium</i> wilt and highly resistant to SMD
TV 1	Semi spreading, early, susceptible to <i>Fusarium</i> wilt and sterility mosaic
AKT 8811	Semi spreading, early, tolerant to <i>Fusarium</i> wilt and susceptible to sterility mosaic
GT 288	Erect, early, susceptible to <i>Fusarium</i> wilt and sterility mosaic, white seeded
C 11	Spreading, mid late, resistant to <i>Fusarium</i> wilt and susceptible to sterility mosaic, red seeded
Gullyal white	A local genotype highly susceptible for <i>Fusarium</i> 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 <i>Fusarium</i> 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 (~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

377 would follow from genome sequencing, PGI has  
 378 initiated the development of a TILLING (Targeted  
 379 Induced Local Lesions in Genomes; McCallum et al.  
 380 2000)-based reverse genetic resource for pigeonpea.  
 381 TILLING is a reverse genetic approach where a  
 382 library of saturation mutagenized individuals, each  
 383 with several hundred-low 1,000 s of point mutations,  
 384 are screened by high-throughput genotyping to iden-  
 385 tify individuals harboring a range of single nucleotide  
 386 induced variants in genes of interest. The reference  
 387 genotype Asha (ICPL 87119) was mutagenized with

ethyl methane sulfonate (EMS) mutagen to develop 388  
 TILLING population at BHU (Banaras Hindu Uni- 389  
 versity). In the pilot experiments, BHU treated 3,000 390  
 seeds in each of four different concentrations of EMS 391  
 (0.01, 0.02, 0.03, and 0.04 M) during the year 2007– 392  
 2008. As expected, the germination (70%) and pollen 393  
 fertility (87.9%) were higher with 0.01 M treatment 394  
 of EMS and reduced down with the subsequent doses. 395  
 To date, a total of 505 single plant M<sub>1</sub> lines yielded 396  
 fertile M<sub>2</sub> seed. In the M<sub>2</sub> generation, a number of 397  
 chlorophyll and plant form (very dwarf, dwarf, 398

399 fasciated and tall) mutants have been identified.  
 400 Efforts to significantly expand such mutant popula-  
 401 tions to several 1,000–10,000 plant lines are  
 402 underway.

#### 403 Genomic resources

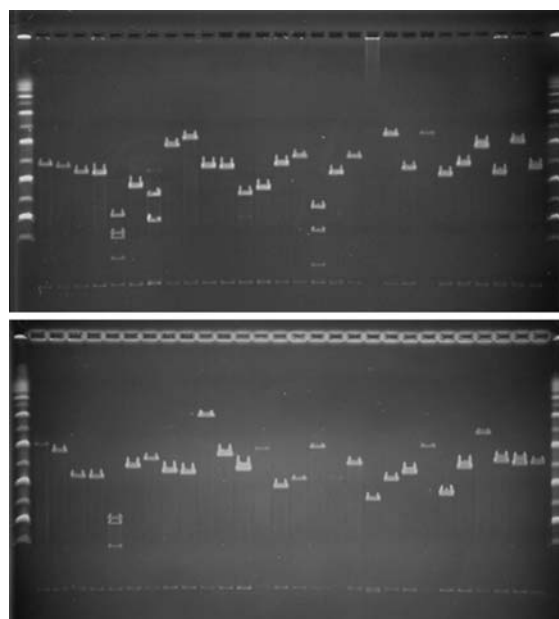
404 A significant amount of genomic resources have  
 405 become available for pigeonpea during last 3 years  
 406 (see Varshney et al. 2009a), some of them are  
 407 discussed below.

#### 408 Large-insert genomic DNA library

409 A BAC library was produced at UC-Davis for the  
 410 PGI reference genotype-“Asha”. The library was  
 411 constructed based on partial digestion of high  
 412 molecular weight DNA separately with *Hind*III and  
 413 *Bam*HI restriction enzymes. Size-purified fragments  
 414 were ligated into vector pCC1BAC and transformed  
 415 into competent cells of *E. coli* strain EPI300. Insert  
 416 sizes were estimated based on pulsed field gel  
 417 electrophoresis of *Not*I-digested BAC DNA, as  
 418 shown in Fig. 2. The *Hind*III library is composed of  
 419 34,560 clones with an estimated average insert size of  
 420 120,000 bp, while the *Bam*HI library is composed of  
 421 34,560 clones with an estimated average insert size of  
 422 115 kb. Taken together, the combined clones repre-  
 423 sent ~11X coverage of the pigeonpea genome. This  
 424 BAC-library is an important resource for marker  
 425 development (described below) as well as for phys-  
 426 ical mapping and to seed future genome sequencing.  
 427 A set of randomly selected 50,000 BAC clones was  
 428 subjected to end sequencing by UC-Davis and  
 429 ICRISAT in collaboration with JCVI, USA. In total,  
 430 87,590 BAC end sequences (BESs) were generated,  
 431 with most of the 50,000 BAC clones containing high  
 432 quality sequence from both ends. The combined data  
 433 represent ~56 Mbp of DNA sequences, which were  
 434 submitted to the National Center for Biotechnology  
 435 Information (NCBI) Genome Survey Sequence (GSS)  
 436 database for public access.

#### 437 cDNA libraries and expressed sequence tags

438 Transcriptome sequencing has been a popular  
 439 approach to efficiently identify the transcribed  
 440 portion of the genome (Sreenivasulu et al. 2002).  
 441 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 *Hind*III (bottom panel) libraries were digested with *Not*I 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 442  
 16 cDNA libraries were generated from *Fusarium* 443  
*udum* and Sterility mosaic virus challenged root 444  
 tissues of four genotypes (ICPL 20102 and ICP 2376 445  
 for FW; ICP 7035 and TTB 7 for SMD). Several 446  
 thousand cDNA clones from these cDNA libraries 447  
 were sequenced at ICRISAT (Raju et al., unpub- 448  
 lished) to obtain a total of 5,680 expressed sequence 449  
 tags (ESTs) from FW challenged and 3,788 ESTs 450  
 from SMD challenged tissues and submitted to NCBI. 451

In addition to traditional Sanger sequencing to 452  
 obtain these FW and SMD associated ESTs, next 453  
 generation sequencing (NGS) technology (Varshney 454  
 et al. 2009b), was employed to identify whole plant 455  
 ESTs. cDNA prepared at ICRISAT from 15 tissues 456  
 representing different developmental stages of the 457  
 Pusa Ageti genotype were pooled, and normalized to 458  
 minimize redundancy and enhance efficiency of 459  
 capture of rare transcripts. ICRISAT in collaboration 460  
 with JCVI sequenced the normalized cDNA pool using 461  
 454/FLX sequencing, a next generation sequencing 462  
 technology that offers higher throughput relative to 463

- 464 Sanger sequencing. Of 4,96,705 sequence reads 2,  
465 87,766(>50%) were longer than 200 bp. Cluster  
466 analysis of these sequences, done at NCGR in  
467 collaboration with ICRISAT, provided 48,519 contigs.  
468 Similarly, NRCPB has employed 454/FLX sequencing  
469 on the cDNA pools from two cultivars, Asha and  
470 UPAS120. As a result of this a total of 1,696,724  
471 sequence reads (566 Mb) with an average read length  
472 of 333 bp were generated from these two genotypes at  
473 NRCPB. Sequence analysis of these genotypes pro-  
474 vided 42,000 unique sequences of which 25,000 were  
475 common between these two genotypes.
- 476 Together these transcript sequences represent a  
477 significant fraction of the pigeonpea transcriptome,  
478 and should be a useful resource for marker develop-  
479 ment as well as gene discovery and functional  
480 studies.
- 481 *Microsatellite/simple sequence repeat markers*
- 482 Simple sequence repeat markers are the markers of  
483 choice for plant genetics and breeding applications  
484 (Gupta and Varshney 2000). In case of pigeonpea,  
485 however, only 140 SSR markers were available in  
486 public domain before the establishment of PGI  
487 (Burns et al. 2001; Odeny et al. 2007, 2009). As  
488 <10% SSR markers show polymorphism in the intra-  
489 specific germplasm, development of an intra-specific  
490 map with moderate marker density (with about 300  
491 markers in each intra-specific population) would  
492 require availability of at least 3,000 SSR markers.  
493 To develop the larger number of SSR markers, three  
494 approaches are being used at ICRISAT in collabora-  
495 tion with UC-Davis and other collaborating centers.
- 496 *SSR-enriched library* Simple sequence repeat-  
497 enriched library have been a popular method to  
498 isolate SSRs in several plant species (Gupta and  
499 Varshney 2000). Therefore, several genomic DNA  
500 libraries enriched for five SSR repeat motifs (CT, TG,  
501 AG, AAG, and TCG) were generated from Asha  
502 variety using bead capture enrichment protocol  
503 (Glenn and Schable 2005). Initially, 1,728 clones  
504 were picked from two libraries and 82 clones were  
505 sequenced. This pilot experiment provided 36 SSRs  
506 from which 23 primer pairs were synthesized of  
507 which 16 provided scorable amplification products.  
508 Screening of 40 elite genotypes with these 16 markers  
509 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*  
(*MicroSATellite*) 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	454 Sequencing		Total
				Sanger sequencing		
					ICRISAT	
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	<i>In progress</i>	–	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.

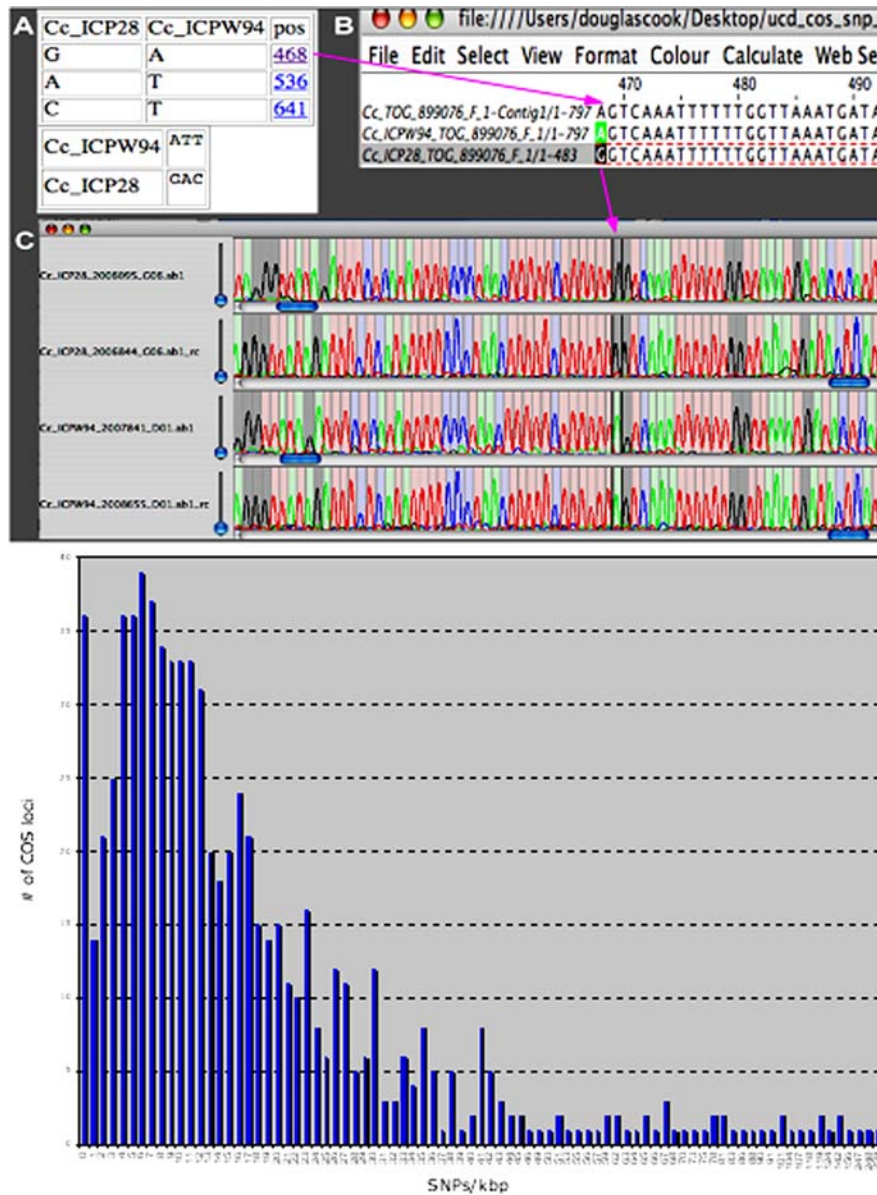
As large amount of transcript data have been generated from three genotypes by using 454 sequencing approach, SSR mining has been undertaken in these datasets. For instance, 87,314 SSRs have been identified in 188,741 sequences at ICRISAT. While 12,168 454 sequences contained more than one SSR, 12,679 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

620 sequence (COS) loci with the SNP frequency of 9.8  
 621 SNP per gene (Fig. 3). All validated SNPs were  
 622 assessed for suitability for analysis via GoldenGate  
 623 genotyping, which yielded 670 unique loci amenable

to the high-throughput and parallel “oligo pool all” 624  
 (OPA) genotyping assay. A single SNP per 670 unique 625  
 loci, with an additional SNP for 98 loci were selected 626  
 for production of a 768 SNP OPA assay. Genotyping 627

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

635 *Next generation sequencing based SNPs* Recent  
636 developments in NGS technologies are catalyzing the  
637 development of SNP markers even in those crops with  
638 little or no sequence information (Varshney et al.  
639 2009b). ICRISAT and NCGR have been working to  
640 use the Solexa NGS technology to sequence  
641 transcriptomes of ten pigeonpea genotypes that are  
642 parents of six mapping populations. The availability of  
643 reference genome sequence data vastly improves the  
644 effectiveness of NGS approaches. In pigeonpea,  
645 transcript assembly (~48,000 transcript contigs)  
646 developed from 454/FLX NGS and Sanger ESTs will  
647 facilitate alignment of Solexa sequence data. Multiple  
648 sequence alignment (MSA) of Solexa datasets from  
649 the ten genotypes, together with reference sequence,  
650 should provide a large number of high confidence  
651 SNPs for high frequency alleles. SNPs identified from  
652 such NGS approaches, together with re-sequencing of  
653 COS loci in additional genotypes should allow for the  
654 development of additional SNP sets (for example, a  
655 1,536 GoldenGate SNP assay or even larger) for  
656 mapping of several hundred SNPs in different intra-  
657 specific mapping populations.

#### 658 *Diversity array technology markers*

659 Diversity array technology provides a sequence  
660 independent and high throughput approach to develop  
661 dominant-type markers, and provides a cost-effective  
662 whole-genome profiling (Jaccoud et al. 2001). Fur-  
663 ther, as the same platform is used for discovery and  
664 scoring of polymorphic markers it is a quite cost  
665 effective and user friendly approach for genotyping  
666 of polymorphic markers. DARt has been developed  
667 as a technology for whole-genome profiling in over  
668 40 crop species. In pigeonpea, a pilot DARt array,  
669 comprising of 5,376 features was developed by Yang  
670 et al. (2006). When this array was used to analyze 96  
671 genotypes representing 20 species of *Cajanus*, culti-  
672 vated genotypes did not show much polymorphism.  
673 Under the framework of a recent project sponsored by

Generation Challenge Programme, DARt Pty Ltd in  
674 collaboration with ICRISAT, has upgraded the DARt  
675 array for pigeonpea that has >15,000 features. These  
676 DARt markers are being used to genotype several  
677 mapping populations to develop integrated and high-  
678 density genetic maps of pigeonpea. 679

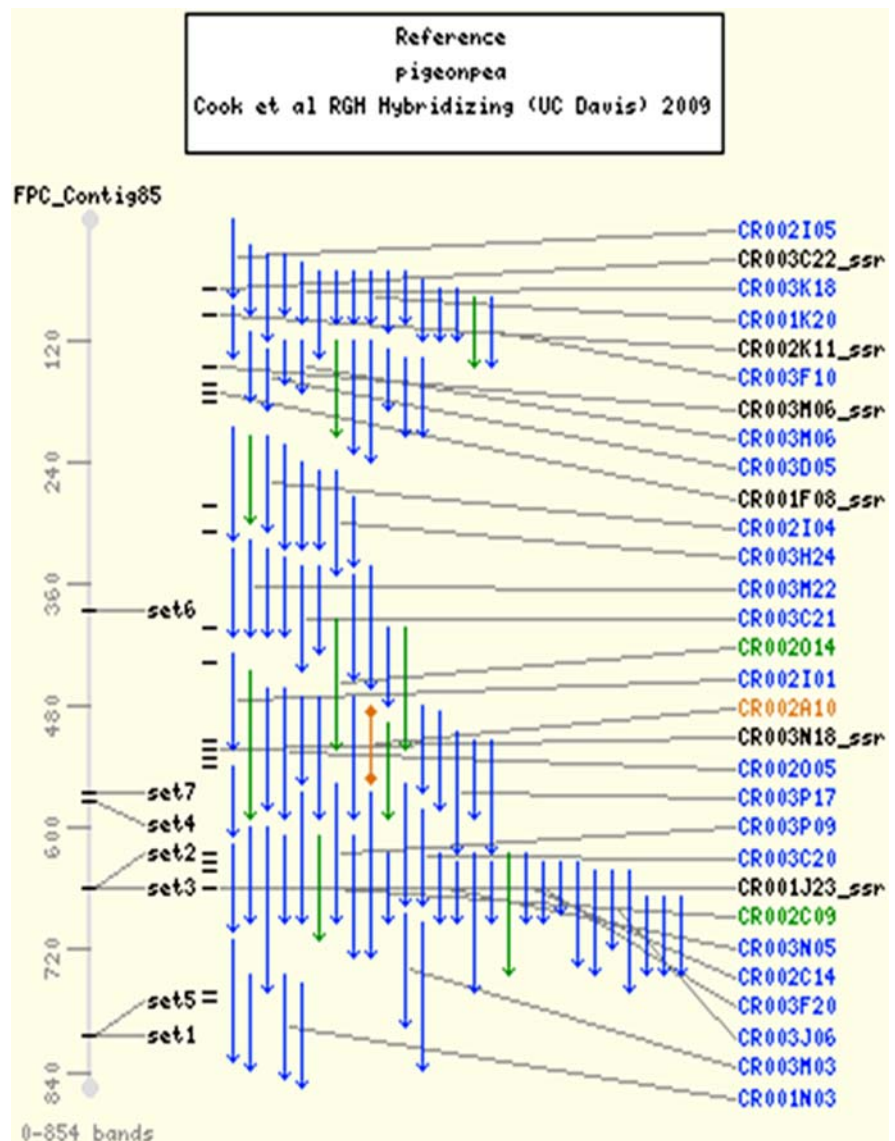
#### 680 *Isolation and characterization of nucleotide binding* 681 *site domains*

682 With an objective of understanding evolution of  
683 disease resistance during legume speciation, UC-  
684 Davis designed degenerate primer pairs based on  
685 NBS-leucine rich repeat disease resistance (NBS-  
686 LRR) homologs. Using NBS-LRR sequences from the  
687 reference legume *Medicago truncatula*, a set of 544  
688 primers were developed and used in the pigeonpea  
689 reference genotype Asha. In excess of 600 unique  
690 nucleotide binding site (NBS) sequences were cloned  
691 and sequenced, resulting in a large collection of  
692 disease resistance genes from both major clades, toll  
693 interleukin receptor-like (TIR) and non-TIR NBS  
694 domains (Rosen et al., unpublished data). In addition  
695 to phylogenetic analysis of NBS sequences, probes  
696 from representative NBS clades were used to hybridize  
697 to high density nylon filters containing ordered arrays  
698 of the pigeonpea BAC library. From 69 probes  
699 analyzed to date (~70% of total), 960 BAC clones  
700 were identified, which have been BAC end sequenced  
701 and used for physical map construction. A total of  
702 1,805 BAC end sequences were obtained and were  
703 submitted to the NCBI GSS database; mining of this  
704 BAC end sequence data yielded 151 SSRs. 756 of the  
705 960 BAC clones yielded data when subjected to high  
706 information content fingerprinting (HICF) that was  
707 analyzed by means of fingerprint contig (FPC) soft-  
708 ware. 91% of the 756 BAC clones assembled into 90  
709 BAC contigs, while 67 clones remained as singleton  
710 BAC clones. An integrated view of BAC contig data,  
711 with BAC end SSR annotation has been presented in  
712 Fig. 4. Together, these data form the basis for an SSR  
713 molecular marker resource that is linked to NBS-LRR  
714 contigs, acting as strong candidates for molecular  
715 breeding and disease resistance trait dissection.

#### 716 *Harnessing the potential of comparative genomics*

717 Comparative genomics offers the promise of lever-  
718 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



719 more rapidly advance genetics in species with fewer  
 720 genetic/genomic resources (O'Brien et al. 1999).  
 721 Development of COS markers, as mentioned above,  
 722 is one of those examples of application of compar-  
 723 ative genomics approaches. The nearest reference  
 724 genome (i.e., one with extensive genome sequence)  
 725 for pigeonpea is soybean (*Glycine max*), another  
 726 *phaseoloid* legume. Despite this phylogenetic rela-  
 727 tionship, leveraging soybean to advance pigeonpea  
 728 genomics may be hampered by polyploidy (Shoe-  
 729 maker et al. 1996; Doyle et al. 2004; Walling et al.  
 730 2006) and extensive whole or partial genome dupli-  
 731 cation 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 (McClellan et al., unpub-  
 lished 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 com-  
 pletion (<http://www.phytozome.org/soybean>), may  
 more readily benefit genetics and crop improvement in  
 pigeonpea.

732  
 733  
 734  
 735  
 736  
 737  
 738  
 739  
 740  
 741  
 742  
 743  
 744

- 745 Towards genome sequencing of pigeonpea 790
- 746 At the inception of the PGI, a clone-by-clone approach 791
- 747 was proposed to sequence the pigeonpea genome in 792
- 748 Phase 3. However, recent developments in sequencing 793
- 749 technologies (Mardis 2008) suggest possible revisions 794
- 750 to this approach. NGS technologies can “democra- 795
- 751 tize” genome sequencing for crops with little political 796
- 752 and/or research support (Varshney et al. 2009b) such 797
- 753 as pigeonpea. Although complete genome sequence 798
- 754 would require more extensive resources, we anticipate 799
- 755 that existing genomics resources (BACs, BES, tran- 800
- 756 script sequences, high density molecular maps) 801
- 757 together with NGS technology could allow the 802
- 758 assembly of a significant fraction of the low copy 803
- 759 genic portions of the genome (euchromatin) in the near 804
- 760 term, which in itself should revolutionize molecular 805
- 761 breeding in pigeonpea. Ongoing rapid advances in 806
- 762 sequencing technologies are likely to make complete 807
- 763 genome sequencing in pigeonpea achievable in the not 808
- 764 too distant future. 809
- 810
- 765 **Summary and outlook** 811
- 766 In many crop species, genomic tools and approaches 812
- 767 have enhanced the precision and efficiency of predic- 813
- 768 tion of phenotype from genotype (Varshney et al. 814
- 769 2005a) and have been instrumental in developing 815
- 770 superior genotypes and varieties (Varshney et al. 2006, 816
- 771 2009c). However, until recently appropriate genomic 817
- 772 tools were not available in pigeonpea. The PGI 818
- 773 consortium, during the last 3 years, has been very 819
- 774 successful in developing a significant amount of 820
- 775 genetic and genomic resources in pigeonpea, with the 821
- 776 majority of the data already in the public domain, or 822
- 777 nearly so. Generation of a variety of genetic and 823
- 778 genomic tools such as mapping populations, mutant 824
- 779 population, different kinds of molecular markers (e.g., 825
- 780 SSRs, DArTs, SNPs, COSs) at moderately large scale, 826
- 781 BAC library, Sanger, and 454/FLX ESTs, unigenes, 827
- 782 NBS-LRR genes, etc. in pigeonpea, from the PGI 828
- 783 should have tremendous impact on pigeonpea breed- 829
- 784 ing. For instance, molecular markers together with 830
- 785 mapping populations would provide the markers 831
- 786 associated with the trait through linkage mapping 832
- 787 approach. High-throughput marker genotyping plat- 833
- 788 forms such as DArT markers, GoldenGate assays for 834
- 789 SNPs will enable the community to undertake 835
- association mapping to identify the markers/gene(s) 790
- associated with complex traits by harnessing the 791
- genetic variation present in the natural populations 792
- and germplasm collections. BAC library and BESs will 793
- help develop physical maps to anchor genome 794
- sequence data, and to clone genes in concert with 795
- transcriptomics resources. Molecular markers identi- 796
- fied from these approaches that are associated with 797
- traits of importance to breeders should accelerate 798
- pigeonpea improvement via marker assisted selection 799
- (MAS) or transgenic approaches. Largely through the 800
- efforts of the PGI, pigeonpea should move from its 801
- current status of an ‘orphan’ legume to a well- 802
- resourced crop species. 803
- In summary, modern molecular breeding methods 804
- together with the power of genomics and genetic 805
- resources developed through the PGI should revolu- 806
- tionize pigeonpea crop improvement, and conse- 807
- quently benefit farmers and consumers of this 808
- important pulse crop of India and the semi-arid 809
- regions of the world. 810
- Acknowledgments** Authors are thankful to several research 811
- scholars and post-doctoral fellows working in the laboratories 812
- of collaborating research institutes of PGI as well as a number 813
- of collaborators engaged directly or indirectly in research 814
- activities of PGI. Authors would like to extend their sincere 815
- thanks to Dr. Mangla Rai, Director General, Indian Council of 816
- Agricultural Research (ICAR) and Secretary, Department of 817
- Agricultural Research and Extension (DARE) for his support in 818
- initiating the PGI under the Indo-US Agricultural Knowledge 819
- Initiative (AKI). Thanks are also due to Dr. Satya Prakash 820
- Tiwari, Deputy Director General (Education), ICAR; Dr. Jean- 821
- Marcel Ribaut, Director, Generation Challenge Programme 822
- (GCP, [www.generationcp.org](http://www.generationcp.org)); Dr. William Dar, Director 823
- General, ICRISAT and Dr. Dave Hoisington, Deputy Director 824
- General-Research, ICRISAT for their strong support for pi- 825
- geonpea genomics activities. Research in the laboratories of 826
- authors presented in this article is supported by ICAR under the 827
- umbrella of Indo-US AKI, Generation Challenge Program 828
- (GCP), and the National Science Foundation (USA). 829
- Open Access** This article is distributed under the terms of the 830
- Creative Commons Attribution Noncommercial License which 831
- permits any noncommercial use, distribution, and reproduction 832
- in any medium, provided the original author(s) and source are 833
- credited. 834
- 835
- References** 836
- Burns MJ, Edwards KJ, Newbury HJ, Ford LBR, Baggot CD 837
- (2001) Development of simple sequence repeat (SSR) 838
- markers for the assessment of gene flow and genetic 839

- 840 diversity in pigeonpea (*Cajanus cajan*). Mol Ecol Notes  
841 1:283–285
- 842 Byth DE, Wallis ES, Saxena KB (1981) Adaptation and  
843 breeding strategies for pigeonpea. Proceedings of ICRI-  
844 SAT/ICAR international workshop on pigeonpea, 15–19  
845 December, vol. 1, pp 450–465
- 846 Doyle JJ, Doyle JL, Rauscher JT, Brown AHD (2004) Diploid  
847 and polyploid reticulate evolution throughout the history  
848 of the perennial soybeans (*Glycine* subgenus *Glycine*).  
849 New Phytol 161:121–132
- 850 Glenn TC, Schable NA (2005) Isolating microsatellite DNA  
851 loci. Methods Enzymol 395:202–222
- 852 Gupta PK, Varshney RK (2000) The development and use of  
853 microsatellite markers for genetic analysis and plant  
854 breeding with emphasis on bread wheat. Euphytica 113:  
855 163–185
- 856 Innes RW, Ameline-Torregrosa C, Ashfield T, Cannon E,  
857 Cannon SB, Chacko B, Chen W, Couloux A, Dalwani A,  
858 Denny R, Deshpande S, Egan AN, Glover N, Hans CS,  
859 Howell S, Ilut D, Jackson S, Lai H, Mammadov J, Del  
860 Campo SM, Metcalf M, Nguyen A, O'Bleness M, Pfeil  
861 BE, Podicheti R, Ratnaparkhe MB, Samain S, Sanders I,  
862 Segurens B, Seviganc M, Sherman-Broyles S, Thareau V,  
863 Tucker DM, Walling J, Wawrzynski A, Yi J, Doyle JJ,  
864 Geffroy V, Roe BA, Maroof MA, Young ND (2008)  
865 Differential accumulation of retroelements and diversifi-  
866 cation of NB-LRR disease resistance genes in duplicated  
867 regions following polyploidy in the ancestor of soybean.  
868 Plant Physiol 148:1740–1759
- 869 Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity  
870 arrays: a solid state technology for sequence information  
871 independent genotyping. Nucleic Acids Res 29:e25
- 872 Lavin M, Herendeen PS, Wojciechowski MF (2005) Evolu-  
873 tionary rates analysis of leguminosae implicates a rapid  
874 diversification of lineages during the tertiary. Syst Biol  
875 54:575–594
- 876 Mardis ER (2008) Next-generation DNA sequencing methods.  
877 Annu Rev Genomics Hum Genet 9:387–402
- 878 McCallum CM, Comai L, Greene EA, Henikoff S (2000)  
879 Targeting induced local lesions in genomes (TILLING)  
880 for plant functional genomics. Plant Physiol 123:439–442
- 881 O'Brien SJ, Menotti RM, Murphy WJ, Nash WG, Wienberg J,  
882 Stanyon R, Copeland NG, Jenkins NA, Womack JE,  
883 Marshall GJA (1999) The promise of comparative  
884 genomics in mammals. Science 286:458–481
- 885 Odeny DA, Jayashree B, Ferguson M, Hoisington D, Cry LJ,  
886 Gebhardt C (2007) Development, characterization and  
887 utilization of microsatellite markers in pigeonpea. Plant  
888 Breed 126:130–136
- 889 Odeny DA, Jayashree B, Gebhardt C, Crouch J (2009) New  
890 microsatellite markers for pigeonpea (*cajanus cajan* (L.)  
891 millsp.). BMC Res Notes 2:35
- 892 Reddy BVS, Green JM, Bise SS (1978) Genetic male sterility  
893 in pigeonpea. Crop Sci 18:362–364
- 894 Saxena KB (2008) Genetic improvement of pigeonpea—a  
895 review. Trop Plant Biol 1:159–178
- 896 Saxena KB, Kumar RV (2003) Development of a cytoplasmic  
897 nuclear male sterility system in pigeonpea using *C. sca-*  
898 *rabaeoides* (L.) thours. Ind J Genet Plant Breed 63:  
899 225–229
- 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 poly-  
morphism 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 microsat-  
ellite 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 gen-  
omes—the utility of expressed sequence tags. Curr Sci  
83:965–973
- Temnykh S, De Clerck G, Lukashova A, Lipovich L, Cartin-  
hour S, McCouch S (2001) Computational and experi-  
mental 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 honorem  
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

- 960 Varshney RK, Graner A, Sorrells ME (2005a) Genomics-  
961 assisted breeding for crop improvement. *Trends Plant Sci*  
962 10:621–630
- 963 Varshney RK, Graner A, Sorrells ME (2005b) Genic micro-  
964 satellite markers in plants: features and applications.  
965 *Trends Biotechnol* 23:48–55
- 966 Varshney RK, Hoisington D, Tyagi AK (2006) Advances in  
967 cereals genomics and applications in crop breeding.  
968 *Trends Biotechnol* 24:490–499
- 969 Varshney RK, Hoisington DA, Upadhyaya HD, Gaur PM,  
970 Nigam SN, Saxena K, Vadez V, Sethy NK, Bhatia S,  
971 Aruna R, Gowda MVC, Singh NK (2007) Molecular  
972 genetics and breeding of grain legume crops for the semi-  
973 arid tropics. In: Varshney RK, Tuberosa R (eds)  
974 Genomics assisted crop improvement, vol II: genomics  
975 applications in crops. Springer, Dordrecht, pp 207–242
- 976 Varshney RK, Close TJ, Singh NK, Hoisington DA, Cook DR  
977 (2009a) Orphan legume crops enter the genomics era.  
978 *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](https://doi.org/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

979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997  
998

REVISED PROOF