Sanjeev Gupta • Nagasamy Nadarajan
Debjyoti Sen Gupta
Editors

Legumes in the Omic Era

Springer

Editors
Sanjeev Gupta
Indian Institute of Pulses Research
Kanpur, UP, India

Nagasamy Nadarajan
Indian Institute of Pulses Research
Kanpur, UP, India

Debjyoti Sen Gupta
Department of Plant Sciences
North Dakota State University
Fargo, ND, USA

ISBN 978-1-4614-8369-4 ISBN 978-1-4614-8370-0 (eBook)
DOI 10.1007/978-1-4614-8370-0
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2013951148

Chapter 5
Advances in Pigeonpea Genomics

Abhishek Bohra, Rachit K. Saxena, K.B. Saxena, C.V. Sameerkumar, and Rajeev K. Varshney

Abstract Pigeonpea, a member of family *Fabaceae*, is one of the important food legumes cultivated in tropical and subtropical regions. Due to its inherent properties to withstand harsh environments, it plays a critical role in ensuring sustainability in the subsistence agriculture. Furthermore, plasticity in the maturity duration imparts it a greater adaptability in a variety of cropping systems. In the post genomics era, the importance of pigeonpea is further evident from the fact that pigeonpea has emerged as first non-industrial legume crop for which the whole genome sequence has been completed. It revealed 605.78 Mb of assembled and anchored sequence as against the predicted 833 Mb genome consequently representing 72.8 % of the whole genome. In order to perform genetic and genomic analysis various molecular markers like random amplified polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), simple sequence repeat (SSR), diversity array technology (DArT), single feature polymorphism (SFP), and single nucleotide polymorphism (SNP) were employed. So far four transcriptome assemblies have been constructed and different sets of EST-SSRs were developed and validated in a panel of diverse pigeonpea genotypes. Extensive survey of BAC-end sequences (BESs) provided 3,072 BES-SSRs and all these BES-SSRs were further used for linkage analysis and trait mapping. To make the available linkage information more useful, six intra-specific genetic maps were joined together into a single consensus genetic map providing map positions to a total of 339 SSR markers

A. Bohra, Ph.D.
Crop Improvement Division, Indian Institute of Pulses Research, Kanpur, Uttar Pradesh, India

R.K. Saxena, Ph.D. • R.K. Varshney, Ph.D. (✉)
Center of Excellence in Genomics (CEG), International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Hyderabad, Andhra Pradesh 502324, India
e-mail: r.k.varshney@cgiar.org

K.B. Saxena, Ph.D. • C.V. Sameerkumar, Ph.D.
Pigeonpea Breeding, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, Andhra Pradesh, India

S. Gupta et al. (eds.), *Legumes in the Omic Era*, DOI 10.1007/978-1-4614-8370-0_5,

at map coverage of 1,059 cM. However, earlier very few linkage maps were available in the crop because of non-availability of genomic resources. Several quantitative trait loci (QTLs) associated with traits of agronomic interest including QTLs for sterility mosaic disease, fertility restoration, plant type and earliness have been identified and validated. To strengthen the traditional breeding, plenty of genomics tools and technologies are now available for integration in regular pigeonpea breeding schemes. This article presents the progress made in the area of pigeonpea genomics and outlines its applications in crop breeding for pigeonpea improvement.

Keywords Pigeonpea • Genetic map • Quantitative trait loci • Marker assisted selection • Genome sequence

Introduction

Pigeonpea [*Cajanus cajan* (L.) Millsp] is one of the most important legume species belonging to the family *Papilionoideae* and a member of warm season legumes (Millettioid clade). Pigeonpea is grown mainly in Asia, Africa and Central/South America, on ~5 million hectares. India ranks first in pigeonpea production with 2.46 million metric tons (mmt) followed by Myanmar (0.2 mmt) and Malawi (0.18 mmt) (FAOSTAT 2011). In developing world specially in India and Africa, pigeonpea remains one of the potential sources for livelihood generation and providing proteins to the resource poor farmers, whereas, in other countries such as Myanmar and China, it is gaining importance as one of the commodity crop to generate the foreign revenue. The cultivation of pigeonpea mostly in marginal and degraded soils and risk prone environments often causes considerable reduction in crop yield due to several factors. These factors mainly include diseases, insects/pests and abiotic stresses such as drought, salinity and water logging. This has reflected in form of a wide yield gap existing between the potential yield and actual yield realized at farmers' field (see Varshney et al. 2012).

Realizing its importance in subsistence agriculture, sincere efforts have been directed towards genetic improvement of pigeonpea. Significant genetic gains have been achieved in the form of release of several pureline varieties along with cytoplasmic genetic male-sterility (CGMS) based hybrids that has led to the expansion of production area from 2.7 mha (1961) to 5.83 mha (2011) however average yield still remains in the range of 736–755 kg/ha (FAOSTAT 2011). Domestication and breeding methods focusing strictly on self-pollination led to drastic narrowing down of the genetic base therefore further complicated the situation.

In order to experience a quantum jump in the productivity, traditional breeding efforts should be supplemented with the genomics technologies. All the essential prerequisites such as large scale DNA markers e.g. simple sequence repeat markers (SSRs), diversity array technology markers (DArT), single feature polymorphisms (SFPs) and single nucleotide polymorphisms (SNPs), genetic and quantitative trait loci (QTLs) maps, trait specific mapping populations and sequence information (transcriptome and genome assemblies) are now available in pigeonpea for

Table 5.1 Genomic resources in pigeonpea

Genomic resources	Number	Features	References
Mapping populations	~30	Segregating for *Fusarium* wilt (FW), sterility mosaic disease (SMD), water logging tolerance and plant type	Kumawat et al. (2012), Varshney et al. (2010a), Dhanasekar et al. (2010), Kotresh et al. (2006)
BAC resources			
1) BAC libraries	2	Comprising 34,560 clones each with 11× coverage of pigeonpea genome	Bohra et al. (2011)
2) BAC-end sequences	88,860	A set of >52K non-redundant sequences represented 35 Mb or ~4.3 % of the pigeonpea genome	Bohra et al. (2011)
Second and third generation DNA markers			
1) SSRs			
a) Genomic (gSSRs)	~3,300	BAC library and BES-derived highly informative SSRs	Odeny et al. (2007), (2009), Saxena et al. (2010b), Bohra et al. (2011)
b) Genic or EST-SSRs			
i) Sanger sequencing	84	Average polymorphic information content (PIC) value of 0.40	Raju et al. (2010)
ii) Deep transcriptome sequencing	550	PIC values ranged from 0.46 to 0.72	Dutta et al. (2011)
c) *In silico* mining of draft genome sequence	23,410	Containing tri-, tetra-, penta-, hexa- or compound repeat units	Varshney et al. (2012)
2) SNPs	28,104	Specific to parental combinations derived from 12 genotypes	Varshney et al. (2012)
3) DArTs	29,000	Diversity surveyed for 400 genotypes	Varshney et al. (2010b)
4) SFPs	5,692	Specific for drought tolerance	Saxena et al. (2011)
Genetic maps			
1) DArTs based	1 paternal and 1 maternal	Maps covered 270.0 cM and 451.6 cM of the total genome	Yang et al. (2009)
2) SSRs based	7 (1 inter-specific and 6 intra-specific)	Covering map distances from 466.97 to 930.9 cM	Bohra et al. (2011), (2012), Gnanesh et al. (2011)

(continued)

Table 5.1 (continued)

Genomic resources	Number	Features	References
3) SNPs based	1 intra-specific	Total map length 1520.22 cM	Kumawat et al. (2012)
	1 inter-specific	Total map length 996.21 cM	Saxena et al. (2012)
Transcriptomic resources			
1) ESTs	25,314	Sanger as well as third generation sequencing derived	http://www.ncbi.nlm.nih.gov
2) Transcriptome assemblies	4	Number of transcript assembly contigs (TACs) ranging from 4,557 to 48,726	Kudapa et al. (2012)
Whole genome sequence			
Draft genome sequences of variety 'Asha'	2	~605.78 Mb of genome with ~163.4× coverage	Varshney et al. (2012)
		~511 Mb of the genome with ~10× coverage	Singh et al. (2011)

initiating genomics assisted breeding (GAB) (Bohra et al. 2011; Varshney et al. 2012; Saxena et al. 2012) (Table 5.1). In recent years, several novel molecular breeding methodologies have been proposed for the crop improvement such as marker assisted back-crossing (MABC) and marker assisted recurrent selection (MARS) which offer a precise manner to choose a desired/superior genotype (Varshney et al. 2013). Approaches like multi parent advance generation inter-cross (MAGIC) and introgression libraries (ILs) are offering new avenues to tap natural genetic variation available in wild relatives and landraces into the cultivated gene pool (Varshney et al. 2013).

This chapter provides an overview on availability of genomic resources and the current status of molecular breeding approaches in pigeonpea improvement and explores possibilities to implement emerging molecular genetics and breeding approaches to gain the advancement in pigeonpea research and productivity.

Genome Size

Pigeonpea is a diploid crop with chromosome number $2n = 2x = 22$. The various karyotype studies conducted in pigeonpea (Krishnaswamy and Ayyangar 1935; Naithani 1941; Akinola et al. 1972) have concluded that all the wild relatives of pigeonpea carry the same number of chromosomes. After soybean, pigeonpea became the second member of clade *Phaseoloid* for which the draft genome sequence has become available and based on K-mer statistics the entire genome size was estimated to be 833.07 Mb (Varshney et al. 2012).

Genomic Resources

Mapping Populations

Availability of large segregating populations is an essential requirement for molecular tagging of traits of interest. Several types of bi-parental mapping populations such as F_2, Backcross (BC_1F_1), recombinant inbred lines (RILs), near isogenic lines (NILs) and double haploid (DH) are being employed for genetic map construction and trait mapping. Based on morphological and molecular diversity and targeting the trait segregation a series of mapping populations were generated in pigeonpea under phase I of pigeonpea genomic initiative (PGI). A total of 25 F_2 mapping populations were reported in pigeonpea segregating for several traits such as resistance to sterility mosaic disease (SMD), *Fusarium* wilt (FW), water logging and fertility restoration (Rf). Most of these populations have reached to the RILs and are being deployed for multi-location trials. Details on these mapping populations have been provided by Varshney et al. (2010a). Of these mapping populations, an inter-specific F_2 mapping population (ICP 28 × ICPW 94) was chosen for constructing high density reference genetic map for pigeonpea (Bohra et al. 2011; Saxena et al. 2012). Apart from PGI, few more mapping populations were developed at various national agricultural research centers (Kotresh et al. 2006; Dhanasekar et al. 2010; Ganapathy et al. 2009; Kumawat et al. 2012) (Table 5.2).

Molecular Markers

A wide range of DNA markers have been employed in pigeonpea including RAPD (Ratnaparkhe et al. 1995), RFLP (Sivaramakrishnan et al. 1997, 2002), AFLP (Panguluri et al. 2005), SSR (Saxena et al. 2010a; Bohra et al. 2011), DArT (Yang et al. 2006, 2011), SFP (Saxena et al. 2011) and SNP (Varshney et al. 2012; Saxena et al. 2012) etc. All these marker systems have been used for a variety of applications e.g. estimation of genetic diversity, construction of genetic maps, etc. in pigeonpea. Initially SSRs were preferred over other marker systems due to unavailability of SNPs and several advantages like higher abundance, co-dominant and multi-allelic nature and ease of scoring etc. In pigeonpea, SSRs were generated through (1) enriched library (Burns et al. 2001; Saxena et al. 2010a) (2) *in silico* expressed sequence tags (ESTs) mining (Dutta et al. 2011; Dubey et al. 2011) and (3) surveying BAC-end sequences and whole genome sequence (Bohra et al. 2011; Varshney et al. 2012). The first set of SSRs comprising ten SSRs in pigeonpea was developed by Burns et al. (2001) using CA and CT repeat enriched libraries. However, development of SSRs through enriched libraries remains to be time consuming and of low through put. In this context, sequencing of BAC ends and mining for SSRs had provided potential alternative for large scale SSR discovery.

Table 5.2 Trait mapping in pigeonpea

Name of population	Type of population	Size of population	Targeted trait	Marker system used	Markers found associated with trait	Phenotypic variance explained	References
Bulked segregants analysis (BSA) based							
GS1×ICPL 87119	F_2	254	Fusarium wilt	RAPD	OPM03704, OPAC11500	–	Kotresh et al. (2006)
TT 44-4×TDI 2004-1	F_2	84	Plant type	RAPD	$OPF04_{700}$, $OPA09_{1375}$	–	Dhanasekar et al. (2010)
TTB 7×BRG 3	F_2	121	SMD resistance	AFLP	$E\text{-}CAA/M\text{-}GTG_{150}$, $E\text{-}CAA/M\text{-}GTG_{60}$	–	Ganapathy et al. (2009)
Genetic map and QTL based							
ICP 8863×ICPL 20097	$F_{2:3}$	190	SMD resistance	SSR	CcM1982, CcM1447 (qSMD1)	9.2	Gnanesh et al. (2011)
					CcM0588, CcM2781 (qSMD2)	8.3	Gnanesh et al. (2011)
TTB 7×ICP 7035	$F_{2:3}$	130	SMD resistance	SSR	CcM2149, CcM0468 (qSMD3)	12.32	Gnanesh et al. (2011)
					CcM1825, CcM1895 (qSMD4)	24.72	Gnanesh et al. (2011)
					CcM0970, CcM2485 (qSMD5)	15.93	Gnanesh et al. (2011)
					CcM0416, CcM2337 (qSMD6)	10.58	Gnanesh et al. (2011)
ICPA 2039×ICPR 2447	F_2	188	Fertility restoration	SSR	CcM1522, CcM1821 (QTL-RF-1)	14.85	Bohra et al. (2012)
					CcM0047, CcM2332 (QTL-RF-2)	15.84	Bohra et al. (2012)
ICPA 2043×ICPR 2671	F_2	188	Fertility restoration	SSR	CcM2542, CcM1277 (QTL-RF-3)	20.89	Bohra et al. (2012)
ICPA 2043×ICPR 3467	F_2	188	Fertility restoration	SSR	CcM0374, CcM1506 (QTL-RF-4)	24.17	Bohra et al. (2012)
Pusa Dwarf×HDM04-1	$F_{2:3}$	186	Plant type and earliness	SSR and SNPs	ASNP1310-ASNP2099 (*q*PH4.1)	28.0	Kumawat et al. (2012)
					ASNP1310-ASNP2099 (*q*FL4.1)	51.4	Kumawat et al. (2012)
					ASNP1664-ASSR295 (*q*PB4.1)	19.5	Kumawat et al. (2012)
					ASSR100-ASSR206 (*q*SB5.1)	10.4	Kumawat et al. (2012)
					ASSR100-ASSR206 (*q*MT5.1)	25.9	Kumawat et al. (2012)
					ASSR100-ASSR206 (*q*PD5.1)	18.9	Kumawat et al. (2012)

In pigeonpea, extensive survey of BAC-end sequences (BESs) provided 3,072 BES-SSRs and all these BES-SSRs were further used for linkage analysis and trait mapping (Bohra et al. 2011, 2012; Gnanesh et al. 2011). In addition, a detailed microsatellite survey of whole genome sequence of pigeonpea has identified thousands of SSRs (Singh et al. 2011; Varshney et al. 2012).

In addition to SSRs, DArT offers great potential because of its sequence-independent nature and ensures whole genome profiling in a high throughput and cost effective manner. In pigeonpea, development of 5,376 DArT features helped in assessment of genetic diversity in a panel of 96 genotypes from 20 different *Cajanus* species (Yang et al. 2006). However, in the post genomics era, owing to the amenability to high throughput detection and precise genotyping, SNP has emerged as preferred class of DNA markers over SSRs. Thousands of SNPs were identified in pigeonpea to undertake genome wide association studies (GWAS) and genome wide selection (GWS) (Varshney et al. 2012; Saxena et al. 2012). Recently cost effective SNP genotyping assays such as competitive allele-specific polymerase chain reaction (KASPar) assays were developed for a total of 1,616 SNPs and designated as PKAMs (pigeonpea KASPar assay markers). Further utility of all these KASPar based SNPs were successfully demonstrated in genetic mapping and comparative analysis in pigeonpea (Saxena et al. 2012). In a similar instance 752 SNPs were successfully used to genotype a panel of 110 accessions (wild as well as cultivated) using GoldenGate assay and provided valuable evidences about gene flow, phylogeny and domestication bottlenecks occurred in pigeonpea (Kassa et al. 2012).

Furthermore, with an aim to leverage the DNA marker catalog, microarray-based markers such as single feature polymorphism (SFP) were also discovered for various parental combinations in pigeonpea. For example, the number of identified SFPs ranged from 780 to 854 between parents of several mapping populations. In total, a novel set of markers comprising 5,692 SFPs was reported (Saxena et al. 2011).

BAC Libraries

BAC libraries harbor large inserts of DNA ranging from 100 to 350 kb with an average insert size of 150 kb. The large size of DNA inserts ensures better coverage of the genome. These offer several advantages like ease of handling, high stability, non-chimeric nature and better transformation efficiency over other vectors such as yeast artificial chromosomes (YACs) and cosmids (Farrar and Donnison 2007). BAC libraries represent a potential genomic resource extensively used for (1) physical map construction, (2) comparative genome analysis via searching for macrosyntenic blocks across species, (3) map-based or positional cloning to isolate genes/QTLs responsible for economically important traits, (4) large scale DNA marker discovery through BAC-end sequencing, and (5) assembling of raw sequence reads into genome assembly for an organism. In pulses, several BAC libraries have been

reported and are being constructed for chickpea, lentil, pigeonpea, mungbean, cowpea, field pea and common bean etc. (Yu 2012). In pigeonpea, two BAC libraries were constructed by using *Hind*III and *Bam*HI restriction enzymes. Each of the libraries was composed of 34,560 clones. The average insert size of *Hind*III library was 120 kb while the *Bam*HI library had an average insert size of 115 kb. These clones collectively represented ~11× coverage of the pigeonpea genome. The sequences adjacent to the insertion sites are generally known as BESs and potential resources for identifying minimally overlapping clones (Kelley et al. 1999). With this perspective, randomly selected 50,000 BAC clones were targeted for end sequencing which generated a set of 88,860 high quality BESs (Bohra et al. 2011).

Genetic Maps

Saturated genetic maps have been constructed for several legumes like chickpea (Thudi et al. 2011; Hiremath et al. 2012), cowpea (Muchero et al. 2009; Lucas et al. 2011), common bean (Cordoba et al. 2010), soybean (Hwang et al. 2009) etc. Till 2010, no genetic map was available for pigeonpea due to non-availability of ample amount of genomic resources such as molecular markers and segregating mapping populations and this situation exacerbated by low genetic variation in *Cajanus* primary gene pool. Following the large scale development of BES-SSR and DArT markers, the first generation genetic maps were constructed for an F_2 population derived from an inter-specific cross ICP 28 (*C. cajan*) × ICPW 94 (*C. scarabaeoides*). SSR based genetic map covered a total map length of 930.9 cM with 239 loci with an average inter-marker distance of 3.8 cM (Bohra et al. 2011). In parallel, DArT based genotyping on this parental combination provided a set of 388 polymorphic markers. However, coupling and repulsion phase of polymorphic markers resulted in development of paternal and maternal specific genetic maps with 172 and 122 unique loci, respectively.

The above mentioned genetic maps were derived between *C. cajan* and *C. scarabaeoides*, which does not reflect the level of DNA polymorphism existing in primary or cultivated gene pool of *Cajanus*. Therefore, greater emphasis was directed towards construction of genetic maps for narrow crosses/intra-specific mapping populations. Keeping this view in mind, a total of six SSRs based intra-specific genetic maps with low to moderate marker density were constructed for six F_2 mapping populations (Gnanesh et al. 2011; Bohra et al. 2012). The number of mapped loci were in the range of 59 (ICPB 2049 × ICPL 99050) to 140 (ICPA 2043 × ICPR 3467) while covering map distances of 466.97 cM (TTB 7 × ICP 7035) to 881.57 cM (ICPA 2043 × ICPR 3467). Furthermore, to make the available linkage information more useful, all the six intra-specific genetic maps were joined together into a single consensus genetic map providing map positions to a total of 339 SSR markers at map coverage of 1,059 cM (Bohra et al. 2012). The bin wise polymorphism information content (PIC) values provided for each mapped loci will help geneticists and breeders to select the more informative and precise markers from the region of

interest. Recently one more genetic map based on an intra-specific mapping population (Pusa Dwarf × HDM04-1) was reported for cultivated pigeonpea. This genetic map comprising 296 loci (267 SNPs + 29 SSRs) covered a map length of 1520.22 cM organized into 11 LGs (Kumawat et al. 2012).

Inter-specific mapping population (ICP 28 × ICPW 94) relatively bigger than previously used (167 F_2) mapping populations, used for SNP genotyping through cost effective genotyping platform (KASPar assays) resulted in a much lower genotyping error rate than that obtained with markers like SSRs. A comprehensive genetic map comprising of 875 SNP loci was constructed (Saxena et al. 2012). The total length of this map was 967.03 cM with average marker distance of 1.11 cM. This linkage map was a considerable improvement with the previous pigeonpea genetic linkage maps using SSR and DArT markers. The marker density in this map has almost three times higher than the previous maps. This higher marker density would be useful in determining double recombinants affecting a single marker and in guiding future mapping efforts in pigeonpea.

Trait Mapping

Trait mapping is one of the important pre-requisite for prediction of phenotype from the genotype. As compared to some other legumes like chickpea and common bean not much progress has been witnessed in the area of trait mapping in pigeonpea. Earlier inadequate supply of DNA polymorphisms and lack of saturated genetic maps have posed obstacles in undertaking QTL analysis in pigeonpea. Despite this, some of the traits such as tolerance to SMD and FW and ideal plant type were chosen for mapping using bulked segregants analysis (BSA). BSA was performed using DNA from extremes phenotypes from segregating F_2 populations. The first instance of QTL analysis was reported by Gnanesh et al. (2011) to tag SMD resistance in pigeonpea. This study reported existence of major as well as minor effect QTLs imparting resistance against SMD. The investigation included two F_2 mapping populations which were subjected to linkage and QTL analysis. The results indicated occurrence of six QTLs (designated as qSMD1-6) explaining phenotypic variations in the range of 8.3–24.72 % (Gnanesh et al. 2011) (Table 5.2).

Another successful attempt for mapping a trait using QTL analysis was performed for fertility restoration (Rf). Restoration of fertility in hybrids forms a vital part of CMS based hybrid breeding. Keeping this in mind, QTL analysis was conducted using genotyping and phenotyping data generated from three different F_2 mapping populations showing segregation for fertility restoration. QTL analysis revealed a total of four large effect Rf-QTLs playing important roles in fertility restoration in pigeonpea (Table 5.2). The phenotypic variations governed by the identified QTLs were observed up to 24 % (Bohra et al. 2012). The SSR markers tightly linked with fertility restoration will help not only in search of a potential restorer but also in discriminating between restorer and maintainer. Similarly based on an intra-specific F_2 population and $F_{2:3}$ families (Pusa Dwarf × HDM04-1) several QTLs were

recovered for six different agronomics traits related to plant type and earliness and the phenotypic variation was observed in the range of 3–50 % (Kumawat et al. 2012). These genomic regions can further be chosen as candidates while practicing marker assisted selection (MAS) for plant type and earliness in pigeonpea.

Functional and Comparative Genomics

Functional genomics has shown remarkable impacts on plant genetics and breeding. In the context, collection of ESTs represents an excellent genomic resource to carry out functional genomics studies. In pigeonpea, a total of 25,576 ESTs have been deposited in NCBI database (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html). In parallel, recent advancements made in the area of next generation sequencing technologies have helped generation of massive transcriptome sequence data. For instance, in the case of pigeonpea, a total of four transcriptome assemblies have been constructed using Illumina GA IIx, FLX/454 and Sanger sequencing (Raju et al. 2010; Dutta et al. 2011; Dubey et al. 2011; Kudapa et al. 2012). Among these, the two most comprehensive assemblies were designated as *Cajanus cajan* transcriptome assembly version 1 and 2 (CcTA v1: Dubey et al. 2011 and CcTA v2: Kudapa et al. 2012) comprising 48,726 and 21,434 transcript assembly contigs (TACs), respectively. The robust transcriptome assembly offers tremendous scope for predicting gene content, function and large scale mining of genic or functional molecular markers (GMM or FMM). For instance, different sets of EST-SSRs were developed from these transcriptome assemblies and validated in a panel of diverse pigeonpea genotypes (Raju et al. 2010; Dutta et al. 2011). Since the functional markers remain highly conserved across genera during the course of evolution, these form the basis for comparative genome analysis.

Comparative genomics remains a powerful approach to harness genomic information form related genera. In pigeonpea, successful cloning of approximately 600 unique nucleotide-binding site (NBS) domain and leucine-rich repeat (LRR) domain sequences was performed using degenerate primers targeting the NBS-LRR sequences from model legume *Medicago truncatula* (Varshney et al. 2010a). NBS-LRR represents the most abundant class of resistance genes in plants (Varshney et al. 2009). Therefore, availability of cloned NBS-LRR fragments would shed light into the fate of NBS-LRR resistance genes during divergence of Millettioid and Galegoid clades within the subfamily *Papilionoideae*. Similarly, comparative analysis of the CcTA v2 with genome sequence of soybean (*Glycine max*) provided a set of 128 intron spanning region (ISR) markers. Mapped SNPs were also used to discover the synteny blocks in each of the 11 pigeonpea linkage groups to their counterparts of four legumes chromosomes (soybean, cowpea, *Medicago* and *Lotus*), implying certain co-linearity for the syntenic chromosome/linkage pairs. Conserved sequences were identified among five legume species (pigeonpea, soybean, cowpea, *Medicago* and *Lotus*) (Saxena et al. 2012). The data from comparative genome analysis should facilitate studies on genome evolution and analysis of structural genome, but more significantly would be helpful in understanding and validation

of functional inference of genes in pigeonpea. The identification of gene functions is difficult in non-model species including pigeonpea, thus functional genome analysis will have to rely heavily on the establishment of orthologies from model species by using comparative genomics analysis.

Genome Sequencing

With the availability of draft genome sequence, pigeonpea has shown a quantum jump in its status and joined the league of model/genomic resource rich crops. Pigeonpea has become the first orphan and non-industrial grain legume to have a draft genome sequence (Varshney et al. 2012). Next generation sequencing platforms such as Illumina GA and HiSeq 2000 were used to sequence elite pigeonpea cultivar Asha (ICPL 87119). Using a *de novo* genome assembly and with the help of bacterial artificial chromosome (BAC)-end sequences and available genetic maps, 605.78 Mb was assembled into scaffold with N50 = 516.06 kb. Based on estimated genome size of 833 Mb using a K-mer analysis, 72.8 % of the genome was assembled. Gene prediction analysis suggested presence of 48,680 genes with an average transcript length of 2,348 bp and 3.59 exons per gene. A total of 46,750 genes (96.4 %) could be assigned based on functional ontology and 1,930 genes (3.96 %) are of unknown function. In terms of non-coding RNAs (ncRNAs), 763 tRNA, 862 miRNA, 329 rRNA and 363 snRNA were encountered in <0.5 % assembled genome. In another sequencing effort, 454 GS-FLX sequencing technology was used to assemble ~511 Mb sequence data for Asha variety (Singh et al. 2011). In this study, 47,004 protein coding genes including 1,213 disease resistance/defence related genes and 152 abiotic stress tolerance genes were predicted.

Analysis of genome assembly (Varshney et al. 2012) for repetitive DNA showed presence of transposable elements (TEs) in 49.61 % of assembled genome. Comparison of pigeonpea genome with soybean, grape, *Medicago truncatula* and *Lotus japonicus* genomes revealed 4,311 clusters of genes that were common to all five eudicot genomes whereas 3,068 gene families were specific to the pigeonpea genome. Pigeonpea genome contains higher number (111) of drought responsive genes than soybean, *Medicago truncatula* and *Lotus japonicus*. These genes are suitable candidates for allele mining in global germplasm collection of pigeonpea so that superior alleles and haplotypes for drought tolerance can be implemented in pigeonpea crop improvement (Varshney et al. 2012).

Genome sequence will be useful in utilizing gene sequences for molecular breeding as well as genetic engineering approaches for crop improvement to minimize yield gap in farmers' field. It will not only facilitate comparative analysis with other members of warm-season Millettioids and cool-season Galegoids but also in understanding the phylogeny and evolution within the legume family as a whole. Furthermore, identified candidate drought responsive genes can be utilized for improving other legume crops such as soybean and common bean, which are adversely affected by drought stress.

Genomics-Assisted Breeding

To enhance the crop productivity of pigeonpea, it is important to implement recently developed biotechnological tools such as molecular markers and genetic maps in the breeding programs. These are pre-requisites for genomics-assisted breeding applications such as marker-assisted selection (MAS) (Varshney et al. 2009). With the development and availability of molecular markers and dense molecular genetic maps, MAS is in routine in breeding programs in several major crop species. However, in pigeonpea full potential of molecular breeding still needs to be realized to reap the benefits of colossal amount of molecular information generated through whole genome sequencing. Though, traditional pigeonpea breeding has provided enough genetic gains in the form of release of several elite cultivars, the pace of improvement is not adequate. Wild relatives of pigeonpea representing the untapped reservoir of tremendous genetic variation offer greater scope for broadening of genetic base in pigeonpea. However the undesirable alleles associated with the wild germplasm i.e. linkage drag hampers the speedy recovery of superior performance. Some novel molecular breeding methods such as advanced backcross QTL (AB-QTL) analysis permitting identification as well as transfer of wild type superior alleles into elite cultivars help greatly by generating broad based breeding materials including introgression lines (ILs), near isogenic lines (NILs), chromosome segment substitution lines (CSSLs) etc. Some efforts have also been initiated at ICRISAT using *C. scarabaeoides* as donor to discover superior alleles of various economically important QTLs through AB-QTL approach (Varshney et al. 2013).

Apart from this, whole genome opens new avenues for re-sequencing and genome wide SNP genotyping of landraces/core/reference sets/composite collection (Upadhyaya et al. 2011) (Fig. 5.1). This will greatly assist in discovery of alleles and unlocking the alleles/loci undergoing selection pressure during the process of domestication. In addition, reference genome would facilitate precise identification of recombination blocks using high throughput genotyping platforms and methods such as genotyping by sequencing (GBS). GBS can be employed to tap the potential of nested association mapping (NAM) ensuring benefits of both association mapping (historical recombination) as well as linkage analysis (bi-parental recombination). NAM would provide insights into the molecular basis underlying various QTLs governing several complex traits. In crops like pigeonpea, some of the other schemes relying on multi-parent crossing would be very effective in providing opportunities for extensive recombination. For instance, creation of multi-parent advanced generation intercross (MAGIC) lines in pigeonpea will help not only in accumulation of superior alleles from various genetic backgrounds but also in fine mapping of the region of interest (Kover et al. 2009). Access to the genome wide SNP/SSR markers together with availability of a training population with a robust historical phenotyping data would allow identification of a genotype with higher breeding value through genomic selection (GS) bypassing extensive field testing/repeated phenotyping (Varshney et al. 2013).

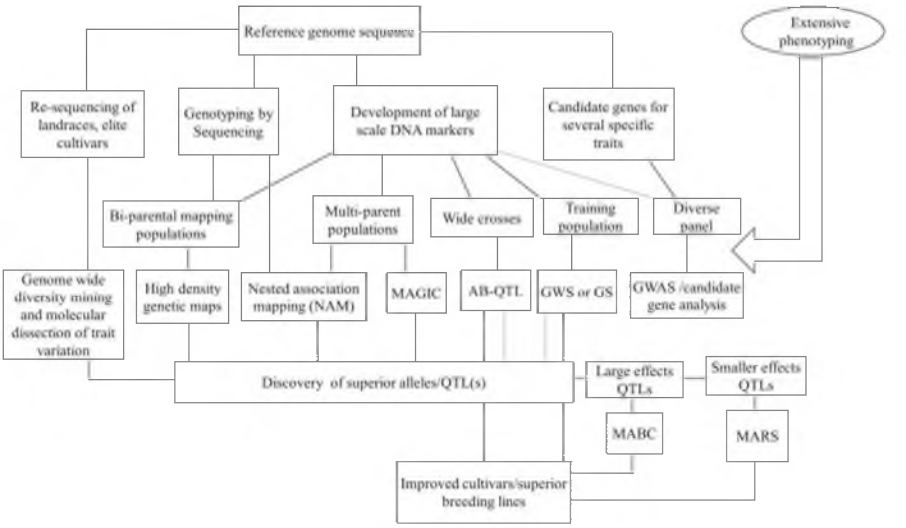

Fig. 5.1 An integrated approach to harness reference genome sequence for pigeonpea genetic information

Conclusion and Perspectives

With the availability of draft genome sequence, pigeonpea has marked its presence among sequenced legumes such as *Lotus*, *Medicago* and soybean enabling more focus on basic research and translational genomics for crop improvement. In the context, high density genetic maps along with precise phenotyping platforms would facilitate identification of genomic regions/QTLs associated with traits such as tolerance to abiotic and biotic stress and fertility restoration. Since exploitation of hybrid vigor seems to a potential alternative to overcome the existing yield barriers, elaborated understanding about the molecular basis of heterosis would allow easy access to the genes imparting hybrid vigor. Furthermore, re-sequencing of several genotypes including landraces, wild relatives and cultivars would ensure recovery of novel haplotypes associated with domestication and other important phenomenon. The deployment of these genomic tools into regular breeding programmes in the form of MABC, MARS and GS would help greatly in bridging the yield gap in pigeonpea through enhancement of productivity in the resource poor and risk prone environment.

Acknowledgements Authors from ICRISAT are thankful to United States Agency for International Development (USAID) for financial support to utilize genome sequence information for pigeonpea improvement. This work has been undertaken as part of the CGIAR Research Program on Grain Legumes.

References

Akinola JO, Pritchard AJ, Whiteman PC (1972) Chromosome number in pigeonpea (*Cajanus cajan* (L.) Millsp.). J Aust Inst Agric Sci 38:305

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

Bohra A, Saxena RK, Gnanesh BN, Saxena KB, Byregowda M, Rathore A, KaviKishor PB, Cook DR, Varshney RK (2012) An intra-specific consensus genetic map of pigeonpea [*Cajanus cajan* (L.) Millspaugh] derived from six mapping populations. Theor Appl Genet 125:1325–1338

Burns MJ, Edwards KJ, Newbury HJ, Ford-Lloyd BV, Baggott 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

Cordoba JM, Chavarro C, Schlueter JA, Jackson SA, Blair MW (2010) Integration of physical and genetic maps of common bean through BAC-derived microsatellite markers. BMC Genomics 11:436

Dhanasekar P, Dhumal KH, Reddy KS (2010) Identification of RAPD marker linked to plant type gene in pigeonpea. Indian J Biotechnol 9:58–63

Dubey A, Farmer A, Schlueter J, Cannon SB, Abernathy B, Tuteja R, Woodward J, Shah T, Mulasmanovic B, Kudapa H, Raju NL, Gothalwal R, Pande S, Xiao Y, Town CD, Singh NK, May GD, Jackson S, Varshney RK (2011) Defining the transcriptome assembly and its use for genome dynamics and transcriptome profiling studies in pigeonpea (*Cajanus cajan* L.). DNA Res 18:153–164

Dutta S, Kumawat G, Singh BP, Gupta DK, Singh S, Dogra V, Gaikwad K, Sharma TR, Raje RS, Bandhopadhya TK, Datta S, Singh MN, Bashasab F, Kulwal P, Wanjari KB, Varshney RK, Cook DR, Singh NK (2011) Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [*Cajanus cajan* (L.) Millspaugh]. BMC Plant Biol 11:17

FAOSTAT. database (2011) http://faostat.fao.org/site/567/DextopDEfault.aspx?PageID=567#ancor. Accessed 14 Mar 2013

Farrar K, Donnison IS (2007) Construction and screening of BAC libraries made from *Brachypodium* genomic DNA. Nat Protoc 2:1661–1674

Ganapathy KN, Byregowda M, Venkatesha SC, Rama Chandra R, Gnanesh BN, Girish G (2009) Identification of AFLP markers linked to sterility mosaic disease in pigeonpea *Cajanus cajan*(L.) Millsp. Int J Integr Biol 7:145–149

Gnanesh BN, Bohra A, Sharma M, Byregowda M, Pande S, Wesley V, Saxena RK, Saxena KB, KaviKishor PB, Varshney RK (2011) Genetic mapping and quantitative trait locus analysis of resistance to sterility mosaic disease in pigeonpea [*Cajanus cajan* (L.) Millsp.]. Field Crops Res 123:53–61

Hiremath PJ, Kumar A, Penmetsa RV, Farmer A, Schlueter JA, Chamarthi SK, Whaley AM, Carrasquilla-Garcia N, Gaur PM, Upadhyaya HD, KaviKishor PB, Shah TM, Cook DR, Varshney RK (2012) Large-scale development of cost-effective SNP marker assays for diversity assessment and genetic mapping in chickpea and comparative mapping in legumes. Plant Biotechnol J 10:716–732

Hwang TY, Sayama T, Takahashi M, Takada Y, Nakamoto Y, Funatsuki H, Hisano H, Sasamoto S, Sato S, Tabata S, Kono I, Hoshi M, Hanawa M, Yano C, Xia Z, Harada K, Kitamura K, Ishimoto M (2009) High-density integrated linkage map based on SSR markers in soybean. DNA Res 16:213–225

Kassa MT, Penmetsa RV, Carrasquilla-Garcia N, Sarma BK, Datta S, Upadhyaya HD, Varshney RK, Von Wettberg EJB, Cook DR (2012) Genetic patterns of domestication in pigeonpea (*Cajanus cajan* (L.) Millsp.) and wild *Cajanus* relatives. PLoS One 7:e39563

Kelley JM, Field CE, Craven MB, Bocskai D, Kim UJ, Rounsley SD, Adams MD (1999) High throughput direct end sequencing of BAC clones. Nucleic Acids Res 27:1539–1546

Kotresh H, Fakrudin B, Punnuri S, Rajkumar B, Thudi M, Paramesh H, Lohithswa H, Kuruvinashetti M (2006) Identification of two RAPD markers genetically linked to a recessive allele of a *Fusarium* wilt resistance gene in pigeonpea (*Cajanus cajan* (L.) Millsp.). Euphytica 149:113–120

Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM et al (2009) A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. PLoS Genet 5:e1000551

Krishnaswamy N, Ayyangar GNR (1935) Chromosome number in *Cajanus indicus* Spreng. Curr Sci 3:614

Kudapa H, Bharti AK, Cannon SB, Farmer AD, Mulaosmanovic B, Kramer R, Bohra A, Weeks NT, Crow JA, Tuteja R, Shah T, Dutta S, Gupta DK, Singh A, Gaikwad K, Sharma TR, May GD, Singh NK, Varshney RK (2012) A comprehensive transcriptome assembly of pigeonpea (*Cajanaus cajan* L.) using Sanger and Second-generation sequencing platforms. Mol Plant 5:1020–1028

Kumawat G, Raje RS, Bhutani S, Pal JK, Mithra SVCR, Kishor Gaikwad K, Sharma TR, Singh NK (2012) Molecular mapping of QTLs for plant type and earliness traits in pigeonpea (*Cajanus cajan* L. Millsp.). BMC Genet 13:84

Lucas MR, Diop NN, Wanamaker S, Ehlers JD, Roberts PA, Close TJ (2011) Cowpea soybean synteny clarified through an improved genetic map. Plant Genome 4:218–225

Muchero W, Diop NN, Bhat PR, Fenton RD, Wanamaker S, Pottorff M, Hearne S, Cisse N, Fatokun C, Ehlers JD, Roberts PA, Close TJ (2009) A consensus genetic map of cowpea [*Vigna unguiculata* (L) Walp.] and synteny based on EST-derived SNPs. Proc Natl Acad Sci U S A 106:18159–18164

Naithani SP (1941) Cytological studies on Indian pulses, Part 1, the somatic chromosomes and the pro-chromosomes of *Cajanus*. Proc Natl Acad Sci India 11:67

Odeny DA, Jayashree B, Ferguson M, Hoisington D, Crouch J, 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

Panguluri SK, Janaiah J, Govil JN, Kumar PA, Sharma PC (2005) AFLP fingerprinting in pigeonpea (*Cajanus cajan* L. Millsp.) and its wild relatives. Genet Res Crop Evol 53:523–531

Raju NL, Gnanesh BN, Lekha P, Jayashree B, Pande S, Hiremath PJ, Byregowda M, Singh NK, Varshney RK (2010) The first set of EST resource for gene discovery and marker development in pigeonpea (*Cajanus cajan*L.). BMC Plant Biol 10:45

Ratnaparkhe MB, Gupta VS, Ven Murthy MR, Ranjekar PK (1995) Genetic fingerprinting of pigeonpea (*Cajanus cajan* (L.) Millsp) and its wild relatives using RAPD markers. Theor Appl Genet 91:893–898

Saxena RK, Prathima C, Saxena KB, Hoisington DA, Singh NK, Varshney RK (2010a) Novel SSR markers for polymorphism detection in pigeonpea (*Cajanus* spp.). Plant Breed 129:142–148

Saxena RK, Saxena KB, Varshney RK (2010b) Application of SSR markers for molecular characterization of hybrid parents and purity assessment of ICPH 2438 hybrid of pigeonpea [*Cajanus cajan* (L.) Millspaugh]. Mol Breed 26:371–380

Saxena RK, Cui X, Thakur V, Walter B, Close TJ, Varshney RK (2011) Single feature polymorphisms (SFPs) for drought tolerance in pigeonpea (*Cajanus* spp.). Funct Integr Genomics 11:651–657

Saxena RK, Penmetsa RV, Upadhyaya HD, Kumar A, Carrasquilla-Garcia N, Schlueter JA, Farmer A, Whaley AM, Sarma BK, May GD, Cook DR, Varshney RK (2012) Large-scale development of cost-effective single-nucleotide polymorphism marker assays for genetic mapping in pigeonpea and comparative mapping in legumes. DNA Res 19:449–461

Singh NK, Gupta DK, Jayaswal PK, Mahato AK, Dutta S, Singh S, Bhutani S, Dogra V, Singh BP, Kumawat G, Pal JK, Pandit A, Singh A, Rawal H, Kumar A, Prashat RG, Khare A, Yadav R,

Raje RS, Singh MN, Datta S, Fakrudin B, Wanjari KB, Kansal R, Dash PK, Jain PK, Bhattacharya R, Gaikwad K, Mohapatra T, Srinivasan R, Sharma TR (2011) The first draft of the pigeonpea genome sequence. J Plant Biochem Biotechnol 21:98–112

Sivaramakrishnan S, Seetha K, Rao AN, Singh L (1997) RFLP analysis of cytoplasmic male sterile lines in pigeonpea (*Cajanus cajan* L. Millsp.). Euphytica 126:293–299

Sivaramakrishnan S, Seetha K, Reddy LJ (2002) Diversity in selected wild and cultivated species of pigeonpea using RFLP of mtDNA. Euphytica 125:21–28

Thudi M, Bohra A, Nayak SN, Varghese N, Shah TM, Penmesta RV, Thirunavukkarasu N, Gudipati S, Gaur PM, Kulwal PL, Upadhyaya HD, Kavikishor PB, Winter P, Kahl G, Town CD, Kilian A, Cook DR, Varshney RK (2011) Novel SSR markers from BAC-end sequences, DArT arrays and a comprehensive genetic map with 1,291 marker loci for chickpea (*Cicer arietinum* L.). PLoS One 6:e27275

Upadhyaya HD, Reddy KN, Sharma S, Varshney RK, Bhattacharjee R, Singh S, Gowda CLL (2011) Pigeonpea composite collection and identification of germplasm for use in crop improvement programmes. Plant Genet Resour 9:97–108

Varshney RK, Close TJ, Singh NK, Hoisington DA, Cook DR (2009) Orphan legume crops enter the genomics era. Curr Opin Plant Biol 12:1–9

Varshney RK, Penmetsa RV, Dutta S, Kulwal PL, Saxena RK, Datta S, Sharma TR, Rosen B, Carrasquilla-Garcia N, Farmer AD, Dubey A, Saxena KB, Gao J, Fakrudin B, Singh MN, Singh BP, Wanjari KB, Yuan M, Srivastava RK, Kilian A, Upadhyaya HD, Mallikarjuna N, Town CD, Bruening GE, He G, May GD, McCombie R, Jackson SA, Singh NK, Cook DR (2010a) Pigeonpea genomics initiative (PGI): an international effort to improve crop productivity of pigeonpea (*Cajanus cajan* L.). Mol Breed 26:393–408

Varshney RK, Glaszmann JC, Leung H, Ribaut JM (2010b) More genomic resources for less-studied crops. Trends Biotechnol 28:452–460

Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, Schlueter JA, Donoghue MTA, Azam S, Fan G, Whaley AM, Farmer AD, Sheridan J, Iwata A, Tuteja R, Penmetsa RV, Wu W, Upadhyaya HD, Yang SP, Shah T, Saxena KB, Michael T, McCombie WR, Yang B, Zhang G, Yang H, Wang J, Spillane C, Cook DR, May GD, Xu X, Jackson SA (2012) Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. Nat Biotechnol 30:83–89

Varshney RK, Murali Mohan S, Gaur PM, Gangarao NVPR, Pandey MK, Bohra A, Sawargaonkar SL, Kimurto PK, Janila P, Saxena KB, Fikre A, Sharma M, Pratap A, Tripathi S, Datta S, Chaturvedi SK, Mallikarjuna N, Anuradha G, Babbar A, Choudhary AK, Mhase MB, Bharadwaj CH, Mannur DM, Harer PN, Guo B, Liang X, Nadarajan N and Gowda CLL (2013) Achievements and prospects of genomics-assisted breeding in three legume crops of the semi-arid tropics. Biotechnol Adv S0734-9750(13)00003-7

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

Yang SS, Xu WW, Tesfaye M, Lamb JAFS, Jung HJG, Samac DA, Vance CP, Gronwald JW (2009) Single-feature polymorphism discovery in the transcriptome of tetraploid alfalfa. Plant Genome 2:224–232

Yang SY, Saxena RK, Kulwal PL, Ash GJ, Dubey A, Harpe JDI, Upadhyaya HD, Gothalwal R, Kilian A, Varshney RK (2011) The first genetic map of pigeonpea based on diversity arrays technology (DArT) markers. J Genet 90:103–109

Yu K (2012) Bacterial chromosome libraries of pulse crops—characteristics and applications. J Biomed Biotechnol 2012:493186