Achievements and prospects of genomics-assisted breeding in three legume crops of the semi-arid tropics

Rajeev K. Varshney a,b,c,d,⁎, S. Murali Mohana, Pooran M. Gaur a, N.V.P.R. Gangarao e, Manish K. Pandey a,f, Abhishek Bohra g, Shrikant L. Sawargaonkar a, Annapurna Chitikineni a, Paul K. Kimurto h, Pasupuleti Janila a, K.B. Saxena a, Asnake Fikre i, Mamta Sharma a, Abhishek Rathore a, Aditya Pratap g, Shailesh Tripathi j, Subhojit Datta g, S.K. Chaturvedi g, Nalini Mallikarjuna a, G. Anuradha k, Anita Babbar l, Arbind K. Choudhary m, M.B. Mhase n, Ch. Bharadwaj j, D.M. Mannur o, P.N. Harer n, Baozhu Guo f, Xuanqiang Liang d, N. Nadarajan g, C.L.L. Gowda a

a International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad 502324, India
b CGIAR Generation Challenge Programme (GCP), c/o CIMMYT, DF 06600, Mexico
c The University of Western Australia, Crawley 6009, Australia
d Crops Research Institute, Guangdong Academy of Agricultural Sciences (GAAS), Guangzhou 510640, China
e International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Nairobi 39063, Kenya
f Crop Protection and Management Research Unit, US Department of Agriculture - Agricultural Research Service (USDA-ARS), Tifton 31793, USA
g Indian Institute of Pulses Research (IIPR), Kanpur 208024, India
h Egerton University, Egerton 536-20115, Kenya
i Debre Zeit Agricultural Research Centre, Ethiopia
j Debre Zeit Agricultural Research Centre, Ethiopia
k Division of Genetics, Indian Agricultural Research Institute (IARI), New Delhi 110012, India
l Institute of Biotechnology, Acharya NG Ranga Agricultural University (ANGRAU), Hyderabad 500030, India
m Jawaharlal Nehru Krishi Vishwa Vidyalya (JNVV), Jabalpur 482004, India
n Indian Institute of Pulses Research (IIPR), Off-season Research Station, Dharwad 580005, India
o Mahatma Phule Krishi Vidyapeeth (MPKV), Rahuri 413722, India
p Agricultural Research Station, University of Agricultural Sciences (UAS-Raichur), Gulbarga 585101, India

A B S T R A C T

Advances in next-generation sequencing and genotyping technologies have enabled generation of large-scale genomic resources such as molecular markers, transcript reads and BAC-end sequences (BESs) in chickpea, pigeonpea and groundnut, three major legume crops of the semi-arid tropics. Comprehensive transcriptome assemblies and genome sequences have either been developed or underway in these crops. Based on these resources, dense genetic maps, QTL maps as well as physical maps for these legume species have also been developed. As a result, these crops have graduated from ‘orphan’ or ‘less-studied’ crops to ‘genomic resources rich’ crops. This article summarizes the above-mentioned advances in genomics and genomics-assisted breeding applications in the form of marker-assisted selection (MAS) for hybrid purity assessment in pigeonpea; marker-assisted backcrossing (MABC) for introgressing QTL region for drought-tolerance related traits, Fusarium wilt (FW) resistance and Ascochyta blight (AB) resistance in chickpea; late leaf spot (LLS), leaf rust and nematode resistance in groundnut. We critically present the case of use of other modern breeding approaches like marker-assisted recurrent selection (MARS) and genomic selection (GS) to utilize the full potential of genomics-assisted breeding for developing superior cultivars with enhanced tolerance to various environmental stresses. In addition, this article recommends the use of advanced-backcross (AB-backcross) breeding and development of specialized populations such as multi-parent advanced generation intercross (MAGIC) for creating new variations that will help in developing superior lines with broadened genetic base. In summary, we propose the use of integrated genomics and breeding approach in these legume crops to enhance crop productivity in marginal environments ensuring food security in developing countries.

© 2013 Elsevier Inc. Open access under CC BY-NC-ND license.

⁎ Corresponding author at: International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad 502324, India. Tel.: +91 40 30713305; fax: +91 40 30713074. E-mail address: varshney.raj@gmail.com (R.K. Varshney).

0734-9750 © 2013 Elsevier Inc. Open access under CC BY-NC-ND license.
http://dx.doi.org/10.1016/j.biotechadv.2013.01.001
1. Introduction

Legumes form an important constituent of food crops consumed globally and complement cereal crops as a source of dietary protein. In addition to providing important micronutrients to human beings, they also fix atmospheric nitrogen, which consequently increase soil fertility and production of other cereal crops. Legumes are also important source of fodder in many agricultural systems and are grown increasingly on a large-scale in semi-arid tropics (SAT). SAT regions cover many developing countries from Africa, Asia to Latin America, and they are characterized by low and erratic rainfall, prolonged dry seasons, and soils with low fertility. This environment is home to the poor and one-sixth of the world’s human population (http://oar.icrisat.org/5283/1/Impact-Flyer-%20Africa.pdf).

Agriulture in the SAT regions is generally undertaken by smallholder farmers and is the mainstay of their livelihood. Among several food crops, chickpea (Cicer arietinum), pigeonpea ( Cajanus cajan ) and groundnut or peanut ( Arachis hypogaea ) are the leading legume crops to feed underprivileged living in the SAT, which is also called “habitat of the hungry”. As these legume crops are grown in harsh environments and exposed to various biotic and abiotic stresses, their productivity has not increased significantly for the last 50 years (Fig. 1) (FAO, 2012). It is, therefore, important to enhance productivity of these crops to cope up with increased demand by the expanding human population. Although some progress has been made in this direction through conventional breeding methods which may be attributed to insufficient understanding of the underlying genetical or molecular mechanisms conferring resistance/tolerance to biotic/abiotic stresses. Advances in genomics have improved our understanding towards genetic architecture and molecular mechanism for complex traits which led to identification of marker-trait associations for economically important traits in order to enhance selection efficiency in breeding. Tremendous progress made in recent years in genomics research of SAT legume crops namely chickpea, pigeonpea and groundnut has prompted us to review the achievements made so far along with initiatives and future prospects for further genetic enhancement.

2. SAT legume crops and production constraints

2.1. Chickpea (Cicer arietinum L.)

Chickpea, also known as garbanzo bean, is a self-pollinated diploid (2n = 2x = 16) crop with genome size of 740 Mb (Arumuganathan and Earle, 1991). The seeds of chickpea are rich in protein (24.6%), carbohydrate (64.6%) and vitamins (Abu-Salem and Abou, 2011). During 2010, chickpea covered a total of 11.9 Mha worldwide with a global production of 10.9 million tons (Mt) and average yield of 913 kg/ha (FAO, 2012). Several abiotic and biotic stresses pose a big threat to high and stable yields of chickpea in the farmers’ fields. Among abiotic stresses, terminal drought is a major problem for the crop grown under rainfed conditions as it delays flowering and affects seed yield. In addition to the above, this crop is also sensitive to lower temperature (<10 °C) mainly during reproductive period (Bakht et al., 2006) and to salinity (NaCl) during flowering and podding stages (Flowers et al., 2010). Salinity can affect the root nodules by decreasing their number, size and N2-fixation capacity. Important biotic stresses affecting chickpea production are, Fusarium wilt (FW) caused by Fusarium oxysporum f.sp. ciceri, reduces yield up to 90% (Singh and Reddy, 1991) and Ascochyta blight (AB) caused by Ascochyta rabiei (Pass.) Labrousse, may cause total crop loss (Singh and Reddy, 1996). Other biotic stresses of chickpea are Botrytis gray mold (BGM) caused by Botrytis cinerea Pers. ex. Fr., leaf spot by Alternaria spp., black root rot by Fusarium solani, Phytophthora root rot by Phytophthora megasperma and Pythium damping-off by Pythium ultimum, rust by Uromyces and beet western yellow virus (BWYV) causing narrow leaf (Nene and Reddy, 1987). Pod borer or Helicoverpa armigera is the major insect pest of chickpea and it feeds on leaves and developing seeds (Sharma et al., 2005). Because of its complex nature and non-availability of good resistance sources in cultivated gene pool,
breeding for resistance to Helicoverpa in chickpea has remained a serious challenge (Sharma et al., 2008).

2.2. Pigeonpea (Cajanus cajan (L.) Millspaugh)

Pigeonpea is an often cross-pollinated diploid (2n = 2x = 22) crop with 833.07 Mb genome size (Varshney et al., 2012a). Globally pigeonpea is cultivated on 4.75 Mha yielding 3.68 Mt with an average yield of 774 kg/ha during 2010 (FAO, 2012). Pigeonpea supplement the vegetarian diet in developing countries by ensuring high duration cultivars as compared to long or medium duration genotypes which are otherwise deficient in nutrition. Being a nitrogen fixing crop, the green manure of pigeonpea offers a natural resource for improving soil health by providing organic material rich in nitrogen to the soil (Whiteman and Norton, 1980). The deep root system of pigeonpea also adds adaptive value to the crop making it one of the most tolerant crops of drought prone marginal environments. However, despite of its immense importance in sustainable agriculture and continued breeding efforts directed towards genetic improvement, the global production per hectare of pigeonpea remained static over last three decades. The yield gap between the potential yield and on-farm yield is mainly due to prevalence of various abiotic and biotic stresses in pigeonpea growing areas together with its cultivation in marginal lands with low input supply and lack of efficient management practices (see Varshney et al., 2012b). Among the various diseases, FW caused by Fusarium udum Butler, is the most important disease in Indian Subcontinent and Eastern Africa (Saxena, 2008). Occurrence of wiltting during pod filling stage causes infection to pigeonpea seeds and yield losses up to 50–70% (Marley and Hilllocks, 1996). Another disease severely affecting the pigeonpea yield is sterility mosaic disease (SMD) caused by pigeonpea sterility mosaic virus (PPSMV) causing losses up to 95 to 100% with infection occurring early at ~45 days old plants (Kannaiyan et al., 1984). Apart from wilt and mosaic, Phytophthora blight caused by the fungus Phytophthora drechsleri Tucker f. sp. cajani is another important disease that has got the status of economic concern. However, the disease is limited in distribution and witness more severity in short duration cultivars as compared to long or medium duration genotypes (Ratnaparkhe and Gupta, 2007). Among the variety of insects feeding on pigeonpea, the pod borer, Helicoverpa armigera (Hubner) is the most damaging pest worldwide and its frequent occurrence often results in complete crop failure. Besides Helicoverpa, other pests like Maruca (Maruca vitrata Geyer), pod sucking bugs (Clavigralla horrida Gervar) and podfly (Melanagromyzula chalcosoma Spencer) pose a big threat to pigeonpea production (Shanower et al., 1999). Moreover, infestations from storage pests like bruchids (Callosobruchus chinensis) intensify the situation and result in profound seed damage during storage. Among the abiotic constraints, salinity and water logging severely affect the pigeonpea production (Choudhary et al., 2011; Saxena, 2008). The limited success achieved so far in addressing the problem of production constraints is mainly due to complex mechanism underlying these stresses together with the lack of precise and efficient screening techniques.

2.3. Groundnut (Arachis hypogaea L.)

The cultivated groundnut, self-pollinated crop with tetraploid (2n = 4x = 40) genome, has originated through a single hybridization and polyploidyization event. Successful selection resulted in a highly narrow genetic base of the cultivated species (Young et al., 1998). Even being a tetraploid, cultivated groundnut genetically behaves as diploid due to unusual pairing of AA- and BB- genome chromosomes during meiosis (Stalker, 1991). With the annual production of 37.7 Mt covering 24.1 Mha achieving an average yield of 1564 kg/ha during 2010, this crop stands fourth largest oilseed crop in the world which is cultivated in more than 100 countries (FAO, 2012). The largest producers of groundnut include India and China followed by other countries in Sub-Saharan Africa and Americas. Groundnut seeds are highly nutritious possessing fat (40–50%), protein (20–30%), carbohydrate (10–20%) and several other micronutrients and minerals (Vitamin E, niacin, calcium, phosphorus, magnesium, zinc, iron, riboflavin, thiamine and potassium). It is an excellent cash crop and has multiple uses of each plant part in direct consumption, confectionery preparations, cooking oil and a rich source of protein feed for animals (see Pandey et al., 2012a).

Harnessing genetic yield potential in groundnut is severely challenged by several biotic/abiotic stress factors. Among several biotic stresses, early leaf spot (ELS) (Cercospora arachidicola), late leaf spot (LLS) (Cercosporidium personatum), rust (Puccinia arachidis) and groundnut rosette disease (GRD) (groundnut rosette virus, groundnut rosette assistor virus and SatRNA complex) cause up to 50% yield loss. In addition, aflatoxin contamination deteriorates product quality leading to financial loss to farmers and safety issues to consumers. During pod development and seed filling stages, moisture stress conditions increase susceptibility to produce aflatoxin contamination (by Aspergillus flavus/A. parasiticus). Groundnut bud necrosis and bacterial wilt disease along with nematodes have also been found to be prevalent in some specific regions. Stem and pod rot, caused by Sclerotium rolfsii, is a potential threat to groundnut production in many warm and humid areas, especially where irrigated groundnut cultivation is expanding. Although several chemical treatments are available to control these diseases, host-plant resistance is considered to be the best approach. Terminal drought has been the most important abiotic stress reducing the crop productivity very significantly along with deterioration of quality of the produce in groundnut as it predisposes aflatoxin infection in the field.

3. Genomic resources

Although limited genomic resources were available in these legume crops until 2005, significant progress has been made in the development of large-scale genomic resources (Table 1). This has been possible due to financial support and coordinated efforts of several organizations such as CGIAR’s Generation Challenge Programme, Bill & Melinda Gates Foundation, Indian Council of Agricultural Research (ICAR) and Department of Biotechnology (DBT) of Government of India, US National Science Foundation (NSF), The Peanut Foundation of the American Peanut Council etc. In brief, these efforts have led to the development of large-scale molecular markers, construction of comprehensive genetic maps, establishment of various marker-trait associations and initiation of molecular breeding in these three crops. Not only this, draft genome sequence has become available in pigeonpea (Varshney et al., 2012a) and similar efforts are underway in chickpea, and groundnut. Coordinated efforts and progress on development of genomic resources can be seen on websites of International Initiative on Pigeonpea Genomics (IIPG, http://www.icrisat.org/gt-bt/iipg/Home.html), International Chickpea Genetics and Genomics Consortium (http://www.icrisat.org/gt-bt/ICCGC/home.htm), and International Peanut Genome Initiative (http://www.peanutbioscience.com). An overview on various strategies to develop genomic resources by ICRISAT and its partners has been presented in Fig. 2.

3.1. Molecular markers and genotyping platforms

Although in recent years a range of marker systems including hybridization-based Diversity Array Technology (DArT) and sequence based markers such as single nucleotide polymorphisms (SNPs) have become available, simple sequence repeat (SSR) or microsatellite marker are still preferred marker system especially for genetics and breeding applications. SSRs exhibit polymorphism in terms of variation in the number of repeat units as revealed by amplification of unique sequences flanking these repeat units. They show co-dominant inheritance and
Therefore are suitable for genotyping segregating populations (including F₂). Multi-allelic nature of the markers enables them to detect a large number of allelic variants in the germplasm collection (Gupta and Varshney, 2000).

Until recently, development of SSR markers was largely based on screening of SSR-enriched or size-selected DNA libraries, however mining of ESTs (expressed sequence tags) or BAC-end sequences (BESs) have become popular approaches for development of SSR markers. SSR markers developed from ESTs or DNA sequences are referred to as 'genic SSR' or 'genic markers' (Varshney et al., 2010a). By using a range of different approaches mentioned above, 3000–6000 SSR markers have become available in the target SAT legume crops. For instance, in the case of chickpea, ca. 2000 SSR markers have been developed from genomic DNA libraries (for references see Varshney et al., 2007; Nayak et al., 2010; Gaur et al., 2011). ESTs (Varshney et al., 2009b), 454/FLX transcript reads (Hiremath et al., 2011; Garg et al., 2011a, 2011b) and BESs (Thudi et al., 2011). Similarly, another set of 487 novel functional markers including 125 EST-SSRs, 151 intron targeted primers (ITPs), 109 expressed sequence tag polymorphisms (ESTPs), and 102 SNP markers have been developed at National Institute of Plant Genome Research (NIPGR) (Choudhary et al., 2012). In the case of pigeonpea, a large number of SSR markers have been developed from BESs and 454/FLX sequences. After mining 88,860 BESs, a set of 3072 SSR markers was developed (Bohra et al., 2011). In addition, 3583 SSRs were identified from ESTs (Raju et al., 2010) and 454/FLX sequences (Dube et al., 2011; Dutta et al., 2011). Furthermore, by scanning the draft genome sequence of pigeonpea (see later), 309,052 SSRs have been identified (Varshney et al., 2012a) and they can be used to enrich genetic maps with more number of molecular markers and also to tag QTL/genes for important traits. In the case of groundnut, >6000 SSRs have become available by the international groundnut community (see Feng et al., 2012; Pandey et al., 2012a; Wang et al., 2012). After screening ca. 4500 SSR markers on parental lines of several mapping populations, 199 highly informative SSR markers with polymorphism information content (PIC) value of >0.50 were identified (Pandey et al., 2012b). Similarly, more recently a set of 66 highly informative SSRs (>0.5 PIC) with long TC repeats has been reported (Macedo et al., 2012).

DAfT marker system is another marker resource mainly used for diversity studies, for saturating linkage maps and also for identifying introgressions from other species. ICRISAT in collaboration with DArT Pty Ltd, Australia has developed DAfT arrays with 15,360 features for chickpea, groundnut and pigeonpea crops (see Varshney et al., 2010a). Screening of elite germplasm of the SAT legume crops with these DAfT markers have been found very useful for monitoring diversity studies, for saturating linkage maps and also for identifying introgressions from other species. ICRISAT in collaboration with DArT Pty Ltd, Australia has developed DAfT arrays with 15,360 features for chickpea, groundnut and pigeonpea crops (see Varshney et al., 2010a). Screening of elite germplasm of the SAT legume crops with these DAfT arrays, however, showed very little polymorphism (Thudi et al., 2011). Interestingly, DAfT markers have been found very useful for monitoring the genome introgression in the cultivated species of pigeonpea from the wild species (Malikarjuna et al., 2011). Because of higher abundance and amenability to high-throughput, SNP markers are becoming popular marker system in several crop.

### Table 1

<table>
<thead>
<tr>
<th>Common name</th>
<th>Chickpea</th>
<th>Pigeonpea</th>
<th>Groundnut</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species</td>
<td>Cicer arietinum</td>
<td>Cajanus cajan</td>
<td>Arachis hypogaea</td>
</tr>
<tr>
<td>Ploidy</td>
<td>2n = 2x = 16</td>
<td>2n = 2x = 22</td>
<td>2n = 4x = 40</td>
</tr>
<tr>
<td>Estimated genome size</td>
<td>740 Mbp</td>
<td>833.07 Mbp</td>
<td>2890 Mbp</td>
</tr>
<tr>
<td>BAC libraries</td>
<td>10× Thudi et al. (2011)</td>
<td>11× Bohra et al. (2011)</td>
<td>ca. 5.3×-BIBAC (BB); ca. 7.4×-BIBAC (AA)</td>
</tr>
<tr>
<td>EST SSRs</td>
<td>44,707 (see Choudhary et al. (2012))</td>
<td>24,176 Kudapa et al. (2012)</td>
<td>6000 (see Pandey et al., 2012a)</td>
</tr>
<tr>
<td>TILLING population</td>
<td>5000 mutant Mₚ lines (Unpublished data)</td>
<td>ca.5000 mutant lines (Varshney et al. (2010c))</td>
<td>3400 mutant Mₚ lines Knoll et al. (2011)</td>
</tr>
<tr>
<td>DAfT clones</td>
<td>5397 Thudi et al. (2011)</td>
<td>15,360 Varshney et al. (2011b); Yang et al. (2006, 2011)</td>
<td>ca. 15,000 Kilian (2008), Varshney et al. (2010a)</td>
</tr>
<tr>
<td>SNPs</td>
<td>9000</td>
<td>10,000</td>
<td>&gt;2000 SNPs, 768-SNP (see Pandey et al., 2012a)</td>
</tr>
<tr>
<td>Genetic maps</td>
<td>24 (15 inter-specific &amp; 9 intra-specific)</td>
<td>Reference genetic map, six intra-specific maps, one consensus map and DAFT based maternal and paternal maps Argout et al. (2011b, 2011, 2012), Yang et al. (2011)</td>
<td>Diploid (AA)-3, Diploid (BB)-2, Tetraploid-13 maps and one reference consensus map (see Pandey et al., 2012a)</td>
</tr>
<tr>
<td>Physical maps</td>
<td>Not available</td>
<td>Available Varshney et al. (2012a)</td>
<td>–</td>
</tr>
<tr>
<td>Complete genome sequence</td>
<td>In progress</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

26,082 potential SNPs have been identified in the population with a transcriptome assembly of chickpea (see later), ICC 4958 and ICC 1882 parental genotypes of intraspecific cross mapping. These have been used recently to identify large scale SNPs. For instance, based on sequence similarity from ESTs/genes of soybean, Medicago truncatula, and Lotus japonicus species were used for allele-specific sequencing of parental genotypes of mapping populations of chickpea (Hiremath et al., 2012) and pigeonpea (Kassa et al., 2012; Saxena et al., 2012). Next generation sequencing (NGS) technologies including 454/FLX and Illumina/Solexa have been used recently to identify large scale SNPs. For instance, based on alignment of ~37 million Illumina/Solexa tags generated from 10 parental genotypes against a transcriptome assembly (CcTA version 1.0) has identified a total of 12,141 SNPs in pigeonpea (Dubey et al., 2011). Furthermore, comparison of transcript reads from 12 different pigeonpea genotypes against the genome assembly has resulted in identification of 28,104 novel SNPs (Varshney et al., 2012a). In the case of groundnut, comparison of >350 Mb 454/FLX-sequencing based transcriptome data from 17 tetraploid genotypes against a reference transcriptome of ‘Tifrunner’ resulted in identification of a total of 8486 SNPs with moderately stringent filtering (http://nespal.org/oziasakinslab/projects/plant-biotechnology-peanut-grasses/peanut-snp-discovery/).

Once SNPs are identified, development of an appropriate SNP genotyping platform is very critical to make the SNP genotyping cost-effective. In the SAT legume crops, a range of SNP genotyping platforms have become available. For instance, University of California-Davis, USA in collaboration with some partner institutes has developed Illumina GoldenGate assays for genotyping 768 SNPs in chickpea, pigeonpea and diploid Arachis species. Similarly, The University of Georgia, USA has also developed an Illumina GoldenGate SNP array comprising of 1536-SNPs with high confidence for Arachis species. These assays are most suitable when relatively large number of SNPs (>500) need to be genotyped with a large number of samples. However, in the case of certain molecular breeding applications which generally require less number of markers (<400), GoldenGate based SNP arrays are not very cost-effective (see Hiremath et al., 2012). Therefore, VeraCode assays have been developed for 96-plex and 48-plex SNP sets for chickpea and pigeonpea, respectively, which can be used on Illumina’s BeadXpress system (RK Varshney et al., unpublished results).
KASPar assay is another important SNP genotyping technology from Kbiosciences (www.kbioscience.co.uk) that is very flexible and can be used to genotype any number of samples with any number of SNPs. Very recently, KASPar assays have been developed for 2005 SNPs in chickpea (Hiremath et al., 2012), 1616 SNPs in pigeonpea (Saxena et al., 2012) and 96 SNPs in groundnut.

3.2. Transcriptome and genome assemblies

Transcriptome sequencing is the first step to access the gene contents of a species. In absence of low-cost NGS technologies for sequencing the genome, transcriptome sequencing using Sanger sequencing technology has been a popular approach to access the gene contents in a range of crop species. In the SAT legumes also, Sanger sequencing technology has been a popular approach to access the gene contents of a species. With absence of low-cost NGS technologies for sequencing, as a result, 435,018 transcript reads were generated (Hiremath et al., 2011). Collected from >20 different developmental stage tissues of the plant. As a result, 435,018 transcript reads were generated (Hiremath et al., 2011). Analysis of these transcript reads with Sanger ESTs generated by Varshney et al. (2009b) as well as those available in public domain provided comprehensive transcript assembly of chickpea (Ca TA) with 103,215 tentative unique sequences (TUSs) (Hiremath et al., 2011). In another study, Garg et al. (2011b) developed hybrid assembly with 34,760 tentative consensus (TCs). More recently, transcriptome of a wild chickpea, C. reticulatum (genotype PI488777) has also been sequenced using CS-FLX Roche 454 NGS technology (Jhanwar et al., 2012). In this study, both de novo assembly and reference-based assembly approaches were explored to develop an optimized assembly and 37,265 C. reticulatum tentative consensus transcripts were (CrTC) was reported (http://www.nipgr.res.in/cdbt.html).

In the case of pigeonpea, 494,353 transcript reads were generated from normalized and pooled RNA samples collected from >20 different developmental stage tissues of the plant. As a result, 435,018 transcript reads were generated (Hiremath et al., 2011). Analysis of these transcript reads with Sanger ESTs generated by Varshney et al. (2009b) as well as those available in public domain (http://www.ncbi.nlm.nih.gov/sites/ Entrez) (Hiremath et al., 2011). Similarly, 9888 ESTs were generated from drought- and salinity-challenged cDNA libraries in the case of chickpea (Varshney et al., 2009b). Analysis of these ESTs together with the then available ESTs in public domain, 9569 and 5085 unigenes were defined for chickpea and pigeonpea, respectively.

To advance development of transcriptomic resources, 454/FLX sequencing was undertaken on normalized and pooled RNA samples collected from >20 different developmental stage tissues of the plant. As a result, 435,018 transcript reads were generated (Hiremath et al., 2011). Analysis of these transcript reads with Sanger ESTs generated by Varshney et al. (2009b) as well as those available in public domain (http://www.ncbi.nlm.nih.gov/sites/ Entrez) (Hiremath et al., 2011). In another study, Garg et al. (2011b) developed hybrid assembly with 34,760 tentative consensus (TCs). More recently, transcriptome of a wild chickpea, C. reticulatum (genotype PI488777) has also been sequenced using CS-FLX Roche 454 NGS technology (Jhanwar et al., 2012). In this study, both de novo assembly and reference-based assembly approaches were explored to develop an optimized assembly and 37,265 C. reticulatum tentative consensus transcripts were (CrTC) was reported (http://www.nipgr.res.in/cdbt.html).

In the case of pigeonpea, 494,353 transcript reads were generated from normalized and pooled RNA samples collected from >20 different developmental stage tissues of the plant. As a result, 435,018 transcript reads were generated (Hiremath et al., 2011). Analysis of these transcript reads with Sanger ESTs generated by Varshney et al. (2009b) as well as those available in public domain (http://www.ncbi.nlm.nih.gov/sites/ Entrez) (Hiremath et al., 2011). Similarly, 9888 ESTs were generated from drought- and salinity-challenged cDNA libraries in the case of chickpea (Varshney et al., 2009b). Analysis of these ESTs together with the then available ESTs in public domain, 9569 and 5085 unigenes were defined for chickpea and pigeonpea, respectively.

To advance development of transcriptomic resources, 454/FLX sequencing was undertaken on normalized and pooled RNA samples collected from >20 different developmental stage tissues of the plant. As a result, 435,018 transcript reads were generated (Hiremath et al., 2011). Analysis of these transcript reads with Sanger ESTs generated by Varshney et al. (2009b) as well as those available in public domain (http://www.ncbi.nlm.nih.gov/sites/ Entrez) (Hiremath et al., 2011). Similarly, 9888 ESTs were generated from drought- and salinity-challenged cDNA libraries in the case of chickpea (Varshney et al., 2009b). Analysis of these ESTs together with the then available ESTs in public domain, 9569 and 5085 unigenes were defined for chickpea and pigeonpea, respectively.

To advance development of transcriptomic resources, 454/FLX sequencing was undertaken on normalized and pooled RNA samples collected from >20 different developmental stage tissues of the plant. As a result, 435,018 transcript reads were generated (Hiremath et al., 2011). Analysis of these transcript reads with Sanger ESTs generated by Varshney et al. (2009b) as well as those available in public domain (http://www.ncbi.nlm.nih.gov/sites/ Entrez) (Hiremath et al., 2011). Similarly, 9888 ESTs were generated from drought- and salinity-challenged cDNA libraries in the case of chickpea (Varshney et al., 2009b). Analysis of these ESTs together with the then available ESTs in public domain, 9569 and 5085 unigenes were defined for chickpea and pigeonpea, respectively.

To advance development of transcriptomic resources, 454/FLX sequencing was undertaken on normalized and pooled RNA samples collected from >20 different developmental stage tissues of the plant. As a result, 435,018 transcript reads were generated (Hiremath et al., 2011). Analysis of these transcript reads with Sanger ESTs generated by Varshney et al. (2009b) as well as those available in public domain (http://www.ncbi.nlm.nih.gov/sites/ Entrez) (Hiremath et al., 2011). Similarly, 9888 ESTs were generated from drought- and salinity-challenged cDNA libraries in the case of chickpea (Varshney et al., 2009b). Analysis of these ESTs together with the then available ESTs in public domain, 9569 and 5085 unigenes were defined for chickpea and pigeonpea, respectively.
3.3 Genetic maps and trait mapping

Availability of large-scale genomic resources, as mentioned above, has led to development of either the first generation or comprehensive genetic maps in the SAT legume crops. Analysis of these genetic maps together with phenotyping of the respective segregating populations for the traits of interest to the breeders has facilitated identification of molecular markers associated with several agronomically important traits.

An inter-specific mapping population derived from a cross ICC 4958 x PI 489777 has been used as a reference mapping population in chickpea and majority of molecular markers have been used to integrate in the genetic map of this population (see Upadhaya et al., 2011). By using the SSR markers from SSR-enriched libraries and BEs, together with DArT markers and genic molecular markers (GMMs) (Gujaria et al., 2011), an integrated genetic map with 1291 marker loci has been developed (Thudi et al., 2011). In parallel, an advanced gene-rich map of chickpea comprising of 406 loci (including 177 gene-based markers) spanning 1497.7 cM genetic distance has been developed by Choudhary et al. (2011) using the same reference population. By developing large-scale KASPar assays for SNP genotyping, Hiremath et al. (2012) has developed a second-generation genetic map comprising 1328 marker loci including 625 novel CKAMs (Chickpea KASPar Assay Markers), 314 TOG-SNPs and 389 published marker loci comprising 1328 marker loci including 625 novel CKAMs (Chickpea KASPar Assay Markers), 314 TOG-SNPs and 389 published marker loci and 314 published marker loci.

Table 2

List of markers associated with major QTL/genes for different traits in chickpea.

<table>
<thead>
<tr>
<th>Traits studied</th>
<th>QTL/genes</th>
<th>Markers linked</th>
<th>PVE (%)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agronomic &amp; yield</td>
<td>Prostrate</td>
<td>TA34–TA142</td>
<td>95.2</td>
<td>Aryamanesh et al. (2010)</td>
</tr>
<tr>
<td>Plant growth habit</td>
<td>Hg/hg</td>
<td>OPB172apo–OPA102101</td>
<td>–</td>
<td>Cobos et al. (2009)</td>
</tr>
<tr>
<td>Days to flowering</td>
<td>Q3–1</td>
<td>TA6–NCPG12</td>
<td>22.0</td>
<td>Rehman et al. (2011)</td>
</tr>
<tr>
<td>QTL</td>
<td>TA142–TA64</td>
<td>45.0</td>
<td>Aryamanesh et al. (2010)</td>
<td></td>
</tr>
<tr>
<td>NCPG176</td>
<td>TA6–OPBP172apo</td>
<td>45.2</td>
<td>Aryamanesh et al. (2010)</td>
<td></td>
</tr>
<tr>
<td>Days to maturity</td>
<td>Q3–1</td>
<td>TA6–NCPG12</td>
<td>33.0</td>
<td>Rehman et al. (2011)</td>
</tr>
<tr>
<td>Seed coat thickness</td>
<td>QTL</td>
<td>B/b–TA61</td>
<td>20.0</td>
<td>Cobos et al. (2009)</td>
</tr>
<tr>
<td>Seed size</td>
<td>QTL</td>
<td>GAA47–STMS11</td>
<td>32.0</td>
<td>Cobos et al. (2009)</td>
</tr>
<tr>
<td>NCPG128</td>
<td>UBC465–TA2x</td>
<td>–</td>
<td>Radhika et al. (2007)</td>
<td></td>
</tr>
<tr>
<td>Double podding</td>
<td>Sfp</td>
<td>NCPG133–UBC249r</td>
<td>–</td>
<td>Radhika et al. (2007)</td>
</tr>
<tr>
<td>Harvest index</td>
<td>Q3–1</td>
<td>HSA08–TA8</td>
<td>13.0</td>
<td>Rehman et al. (2011)</td>
</tr>
<tr>
<td>Q3–1</td>
<td>TA6–NCPG12</td>
<td>25.0</td>
<td>Rehman et al. (2011)</td>
<td></td>
</tr>
</tbody>
</table>

Abiotic stress

| Root traits | QTL | ICCM0249, TAA170, GA24, STMS11 | 30.0 | Varshney et al. (Unpublished) |
| Drought tolerance score | Q3–1 | TA6–NCPG12 | 27.0 | Rehman et al. (2011) |
| Canopy temperature differential | Q1–1 | HSA08–TA8 | 15.0 | Rehman et al. (2011) |

Biotic stress

| Resistance to Ascochyta blight | QTL | OPAR192apo–OPAC042120 | 23.7 | Millan et al. (2003), Cobos et al. (2005) |
| Ar19 | UBC733B–UBC181A | 42.5 | Rakshit et al. (2003) |
| QTL | TR58–TA52 | 22.6 | Iruela et al. (2006) |
| QTL | GAA47 | 34.0 | Iruela et al. (2006) |
| QTL | TA146–TA72 | 21.0 | Iruela et al. (2006) |
| QTL | TA2–TA146 | 29.0 | Anbessa et al. (2009) |
| QTL | STMS11–TA170 | 26.0 | Aryamanesh et al. (2010) |
| Resistance to Fusarium wilt | Foc9 | OP2090–TR59 | 73.0 | Cobos et al. (2005) |
| Foc1 | TA110–GHA12 | – | Gowda et al. (2009) |
| Foc2 | H3A12–TA96 | – | Gowda et al. (2009) |
| Foc3 | TA96–TA194; TA194–H1B06 | – | Sharma et al. (2004), Gowda et al. (2009) |
| Foc5 | TA59–TA96 | 46.5 | Cobos et al. (2009) |
| Resistance to Botrytis gray mold | QTL | TA118–TA159 | 48.0 | Anuradha et al. (2011) |
| Resistance to rust | Uca1 uca1 | TA18–TA180 | 73.7 | Madrid et al. (2008) |
using second and third generation sequencing platforms. Availability of more than 3000 SSR markers facilitated development of inter- as well as intra-specific genetic maps using several F2 mapping populations. The first reference genetic map was developed using an inter-specific population i.e. ICP C (C. cajan) × ICWP 94 (C. scarabaeoides) comprising 79 F2 individuals. This F2 genetic map consists of a total of 239 SSR markers spanning a map distance of 930.9 cM over 11 linkage groups (LGs) (Bohra et al., 2011). Furthermore, DArT genotyping of the same mapping population also resulted in development of DAfT based paternal- (122 unique markers mapped at a distance of 270 cM) and maternal-specific (172 unique loci mapped at 451.6 cM) genetic maps (Yang et al., 2011). Furthermore, after developing KASPar assays for pigeonpea, a dense genetic map comprising 875 SNP loci with an average inter-marker distance of 1.11cM has been developed on an extended F2 population (Saxena et al., 2012). Apart from these inter-specific genetic maps, some more SSR-based specific genetic maps with low to moderate marker density were made available for cultivated pigeonpea. These intra-specific genetic maps have been developed based on six F2 populations viz. TTB × ICP 7035 (84 loci, 466.97 cM), ICP 8863 × ICP 20097 (120 loci, 534.89 cM), ICP 2043 × ICP 99050 (59 loci, 860.02 cM), ICP 2043 × ICP 3467 (140 loci, 881.57 cM), ICP 2039 × ICP 2447 (78 loci, 570.53 cM), ICP 2043 × ICP 2671 (111 loci, 677.97 cM). Based on these six populations, a consensus map comprising of 339 SSR loci with 1059 cM genetic distance has been developed (Bohra et al., 2012). This represents the first instance of merging multiple genetic maps in pigeonpea.

For trait mapping in pigeonpea, some preliminary mapping efforts have been initiated with F2 mapping populations. For instance, bulked segregant analysis (BSA) approach was used for mapping of FW resistance with RAPD markers (Kotresh et al., 2006), mapping of SMD resistance with AFLP marker system (Ganapathy et al., 2009), and ideal plant type with RAPD markers (Dhanasekar et al., 2011). Availability of SSR based genetic maps for F2 populations coupled with extensive phenotyping data facilitated identification of QTLs/marker(s) for various traits of economic importance. For SMD resistance, a total of four major QTLs (qSMD3-qSMD6) were recovered from population TTB × ICP 7035 and one of the underlying QTL qSMD4 explained a phenotypic variance (PV) up to 24%. Similarly some minor QTLs, qSMD1 and qSMD2 (governing PV <10%) were also discovered from another F2 population viz. ICP 8863 × ICP 20097 for SMD resistance (Gnanesh et al., 2010). Availability of SSR based genetic maps for F2 populations coupled with extensive phenotyping data facilitated identification of QTLs/marker(s) for various traits of economic importance. For SMD resistance, a total of four major QTLs (qSMD3-qSMD6) were recovered from population TTB × ICP 7035 and one of the underlying QTL qSMD4 explained a phenotypic variance (PV) up to 24%. Similarly some minor QTLs, qSMD1 and qSMD2 (governing PV <10%) were also discovered from another F2 population viz. ICP 8863 × ICP 20097 for SMD resistance (Gnanesh et al., 2010) (Table 3). SSR markers associated with SMD resistance offer rapid recovery of SMD resistant genotypes from large segregating populations and would open tremendous opportunities for practicing markers-assisted introgression of resistant allele/QTL(s) into susceptible pigeonpea cultivars. Similarly, to map fertility restoration (RF) genes, genotyping and phenotyping data from three F2 populations namely ICPA 2039 × ICWP 2447, ICPA 2043 × ICWP 2671 and, ICWA 2043 × ICWP 3467 were also subjected to QTL analyses (Bohra et al., 2012). All four QTLs (QTL-RF-1 to QTL-RF-4) identified from these three populations can be called ‘major QTLs’ that contributed 14.85% to 24.17% PV. Identification of such linked SSR markers with fertility restoration will supplement hybrid breeding through quick and precise discrimination between B- and R- lines which is otherwise time consuming and labor intensive. In addition, marker-assisted introgression of gene(s)/QTL(s) imparting fertility into the elite genetic background would allow easy conversion of a pure line to a potential restorer in a time saving manner. For facilitating hybrid breeding and adoption, diagnostic SSR markers have also been developed for purity assessment of two hybrids (Bohra et al., 2011; Saxena et al., 2010).

In the case of groundnut, the first SSR-based genetic linkage map for cultivated groundnut was developed on TAG 24 × ICGV 96031 RIL population (RIL-1) (Varshney et al., 2005c). It is now considered as a reference map for cultivated groundnut and has been saturated up to 191 SSR loci (Ravi et al., 2011). Four more genetic maps based on RIL populations segregating for drought tolerance related traits (RIL-2: ICGS 76 × CMSG 84-1 with 119 SSR loci and RIL-3: ICGS 44 × ICGS 76 with 82 SSR loci) and foliar diseases (RIL-4: TAG 24 × GPBD 4 with 188 SSR loci and RIL-5: TG 26 × GPBD 4 with 181 SSR loci) were developed (Gautami et al., 2012a; Sujay et al., 2012). In order to place maximum markers on a single map, one consensus map each for drought tolerance related traits (RILs 1–3; 2840.8 cM) with 293 SSR loci (Gautami et al., 2012a) and foliar disease resistance (RILs 4–5; 1452.09 cM) with 225 SSR loci (Sujay et al., 2012) was developed. Among other genetic maps developed by the groundnut community, an integrated genetic map with 175 marker loci based on three RIL populations (Hong et al., 2010) and another map with 325 marker loci based on two RIL populations (Qin et al., 2012) are noteworthy. More recently, a genetic linkage map consisting of 318 loci onto 21 LGs and covering a total of 1674.4 cM has been developed using an F2 population (Tifrunner × GT200) (Wang et al., 2012). Furthermore, collaborative efforts of several international partners have resulted in the construction of consensus map with 897 SSR marker loci using genotyping data of 11 mapping populations. This map possesses 20 LGs (a01–a10 and b01–b10) spanning a map distance of 3863.6 cM with an average map density of 4.4 cM. This map was divided into 20 cM long 203 BINs and these BINs carry one (a10_02, a10_08 and a10_09) to 20 (a10_04) loci with an average of 4 marker loci per BIN (Gautami et al., 2012b). In addition to this, a dense genetic map using F2 mapping population (Nakatuyenitaka × Y-0311) with 1,114 loci distributed on 21 LGs covering a total 2,166.4 cM map distance has been developed (Shirasawa et al., 2012). These dense genetic/consensus maps are very useful resource while selecting highly informative and uniformly distributed markers for background selection, construction of new genetic maps and diversity analysis.

In terms of trait mapping in groundnut also, significant efforts have been made. Comprehensive QTL analysis of genotyping data and phenotyping data for drought tolerance traits (e.g. transpiration, transpiration efficiency, biomass, specific leaf area, pod weight, total dry matter, SPAD chlorophyll meter reading, total dry weight, shoot dry weight and harvest index) on three mapping populations detected 153 main effect and 25 epistatic QTLs for drought tolerance related traits (Gautami et al., 2012a; Ravi et al., 2011; Varshney et al.,

### Table 3

List of markers associated with major QTLs for different traits in pigeonpea.

<table>
<thead>
<tr>
<th>Traits studied</th>
<th>QTL/genes</th>
<th>Markers linked</th>
<th>PVE (%)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agronomic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fertility restoration</td>
<td>QTL-RF-1</td>
<td>CcM1522-CcM1821</td>
<td>14.85</td>
<td>Saxena et al. (2011), Bohra et al. (2012)</td>
</tr>
<tr>
<td></td>
<td>QTL-RF-2</td>
<td>CcM0047-CcM2332</td>
<td>16.27</td>
<td>Saxena et al. (2011), Bohra et al. (2012)</td>
</tr>
<tr>
<td></td>
<td>QTL-RF-3</td>
<td>CcM2542-CcM1277</td>
<td>20.89</td>
<td>Saxena et al. (2011), Bohra et al. (2012)</td>
</tr>
<tr>
<td></td>
<td>QTL-RF-4</td>
<td>CcM0374-CcM1506</td>
<td>24.17</td>
<td>Saxena et al. (2011), Bohra et al. (2012)</td>
</tr>
<tr>
<td>Biotic stress</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resistance to sterility mosaic disease</td>
<td>qSMD3</td>
<td>CcM2149-CcM0468</td>
<td>12.32</td>
<td>Gnanesh et al. (2011)</td>
</tr>
<tr>
<td></td>
<td>qSMD4</td>
<td>CcM1825-CcM1895</td>
<td>24.72</td>
<td>Gnanesh et al. (2011)</td>
</tr>
<tr>
<td></td>
<td>qSMD5</td>
<td>CcM0970-CcM2485</td>
<td>15.93</td>
<td>Gnanesh et al. (2011)</td>
</tr>
<tr>
<td></td>
<td>qSMD6</td>
<td>CcM0416-CcM2337</td>
<td>10.58</td>
<td>Gnanesh et al. (2011)</td>
</tr>
</tbody>
</table>

---
2009c) (Table 4). Similarly, QTL analysis based on extensive phenotyping data generated on two RIL populations (TAG 24 × GPBD 4 and TG 26 × GPBD 4) for rust and LLS resistance respectively for 7–8 seasons (2004–2010) at University of Agricultural Sciences-Dharwad (India) and genotyping data (207 marker loci each) resulted in identification of a total of 28 QTLs for leaf spot (LLS; 10.1 to 67.8% PV) and 13 QTLs for rust (2.5 to 82.9% PV) (Khedekar et al., 2010; Sujay et al., 2012). More significantly, a major QTL each for LLS (62.34% PV) and rust (82.96% PV) resistance were identified. The associated markers for rust and LLS were validated in alternate mapping populations and germplasm set. In addition, QTL analysis using phenotyping data on important nutritional and oil quality generated on TG 26 × GPBD 4 resulted in detection of a total of seven QTLs for protein content (2.5–9.8% PV), eight QTLs for oil content (1.5–10.2% PV) and six common QTLs for oleic and linoleic acid (3.3–9.7%) (Sarvamangala et al., 2011). Similarly, two QTLs were mapped for tomato spotted wilt virus (TSWV) using two RIL populations (Qin et al., 2012).

4. Genomics-assisted breeding (GAB)

Genomics-assisted breeding refers to integration and use of genomic tools in breeding practices for developing superior lines with enhanced biotic or abiotic stress tolerance and improved yield. The objective of GAB is to establish and utilize relationship between genotype and phenotype for crop improvement. GAB includes a range of approaches including genomics, transcriptomics and proteomics to identify the molecular markers associated with traits of interest to the breeders that help prediction of phenotype from the genotype to assist breeding (Fig. 3). With the advent of next-generation sequencing (NGS) technologies (Varshney et al., 2009c) and high-throughput genotyping technologies (Varshney, 2011), it has been possible to use the genome-wide marker profile/allele data for prediction of phenotype of progenies for selection to the new cycle in breeding programs. To breed for the traits controlled by major QTL genes (e.g. disease resistance), marker-assisted backcrossing (MABC) approach has been considered a good approach (Ribaut and Hoisington, 1998). However, majority of traits targeted by breeders e.g. drought tolerance or durable resistance to multiple races of pathogens are controlled by several QTLs or genes. For instance, in the case of groundnut, 153 main effect and 25 epistatic interaction QTLs with small phenotypic variation were identified that confer drought tolerance (Gautami et al., 2012a; Ravi et al., 2011). In such cases, retaining desirable gene combinations or pyramiding of several QTLs through MABC approach is a challenging task (Peleman and Voort, 2003). Hence, marker-assisted recurrent selection (MARS) has been proposed as better approach (Ribaut and Ragot, 2007). Furthermore, genome-wide selection or genomic selection (GS) approach, due to possibility of generating genome-wide marker data through use of high-throughput genotyping or NGS approaches, is emerging as a powerful approach for identifying desirable progenies for making the crosses (Bernardo and Yu, 2007; Jannink et al., 2010). In some cases, superior alleles for a given trait e.g. disease resistance are identified and transferred from the wild species to a leading variety/cultivar. In such cases, advanced back-cross QTL (AB-QTL) approach has been proposed by Tanksley and Nelson (1996) for simultaneous discovery and transfer of superior alleles from wild species to develop improved lines. These approaches are being used in the SAT legume crops for improving a range of traits. Some of these examples have been listed below.

4.1. Marker-assisted backcrossing (MABC)

MABC involves introgression of specific trait(s) from a donor parent into the genetic background of a recurrent parent (generally leading variety) using molecular markers (Hospital, 2005). The product of MABC is a line/cultivar containing only the major gene/QTL from the donor parent, while retaining the whole genome of the recurrent parent (see Gupta et al., 2010). MABC approach generally involves transfer of a limited number of trait loci including transgenes from one genetic background (donor genotype) to the other genetic background (elite variety). This approach can also be used to generate near-isogenic lines (NILs) or chromosome segment substitution lines (CSSLs) for genomics research, which are populations that are often used for genetic analysis of genes/QTLs and alien gene introgressions (Lorieux, 2005; Varshney et al., 2010b). Gene pyramiding is an important application of MABC in which a few different genes for the same trait (e.g. resistance to different races of a pathogen) or for different traits are brought together in one genetic background using molecular markers.

As mentioned earlier, availability of molecular markers associated with traits of interest has provided an opportunity to initiate MABC for some traits in the SAT legume crops. Groundnut is probably the first among three legumes crops discussed in this article in which, MABC has been used to develop and release an improved variety. For instance, markers linked with root-knot nematode (Meloidogyne arenaria) resistance were used for introgression through the amphidiploid pathway into cultivated groundnut (Simpson et al., 2001). It was found relatively easy to identify linked markers due to sequence divergence between diploid and tetraploid genomes (Chu et al., 2007; Nagy et al., 2010) in groundnut. DNA fragment carrying nematode resistance was introgressed simultaneously selecting a recessive AhFAD28 allele (controls high ratio of oleic: linoleic acid (O/L)) using these linked markers as foreground selection markers (Chu et al., 2011). These efforts led to release of improved Tiftguard variety “Tiftguard High O/L” (Chu et al., 2011). As SSR markers linked with resistance to leaf rust have also been identified in groundnut recently, MABC approach has been initiated to introgress a major QTL contributing 82.96% PV for leaf rust into the genetic background of three elite cultivars namely ICCV 91114, JL 24 and TAG 24. By using 2–3 rounds of backcrossing and selfing, BC2F2 and BC3F2 homozygous lines have been developed at ICRISAT.

In the case of chickpea, two major MABC projects are underway. Under Accelerated Crop Improvement Programme (ACIP) project sponsored by Department of Biotechnology, Government of India, MABC approach is being used for introgressing resistance to two races (foc2 and foc4) independently and pyramiding of resistance to two races (foc1 and foc3) for FW and two QTLs conferring resistance to AB, Jawaharlal Nehru Krishi Vishvavidyalaya (JNKV), Mahatma Phule Krishi Vidyapeeth (MPKV) and Agricultural Research Station (ARS)-Gulbarga (all in India) are transferring resistance to foc4 from WR 315 genotype in leading varieties namely JG 74, Phule G12 and Annigeri-1, respectively. Indian Institute of Pulses Research (IIPR), India is engaged in introgressing resistance to foc2 from the resistant genotype Vijay in to an elite variety, Pusa 256. ICRISAT (India) on the other hand is pyramiding resistances for foc1 and foc3 from WR 315 and 2 QTLs for AB resistance from ILC 3279 line into C 214. At present, homozygous BC2F2.4 lines resistant for both FW and AB diseases in the preliminary evaluations are available. Different partner institutes have generated a range of backcross progenies followed by both foreground selection and background selection. In another initiative called as Tropical Legume-I (TL-I) of CGIAR Generation Challenge Programme in collaboration with Bill & Melinda Gates Foundation, significant efforts have been made to develop drought tolerant progenies (BC3F3:4) in the genetic background of JG11, a leading variety in India by transferring a genomic region containing several QTLS for drought tolerance traits from ICC 4958 genotype. Phenotypic evaluation of these lines is underway in India, Kenya and Ethiopia. Inspired by MABC work in JG11 genetic background, IIPR, Indian Agricultural Research Institute (IARI), Egerton University and Ethiopian Institute of Agricultural Research (EIAR) have also initiated MABC program for introgressing the drought tolerance genomic region from ICC 4958 in the leading varieties from their respective regions. While the work at IIPR and IARI is funded through DBT, Government of
India, TLI-Phase 2 of CGIAR GCP is funding molecular breeding work at Egerton University and EIAR.

4.2. Marker-assisted recurrent selection (MARS)

MARS involves estimation of marker effects from genotyping F$_2$ or F$_3$ population and phenotyping F$_2$ derived F$_3$ or F$_3$ progenies, followed by two or three recombination cycles based on presence of marker alleles for small effect QTLs (Eathington et al., 2007). In the first step of MARS, de novo QTL identification is carried out initially, i.e. QTLs are identified in the breeding population itself, generally derived from good×good crosses. Subsequently, the lines carrying superior alleles for maximum QTLs are crossed to pyramid superior alleles in one genetic background. Recombined lines are then subjected to a final phenotypic screening to select the best lines for multi-location field testing to release them as varieties. MARS is particularly useful for capturing the several genomic regions especially to target more number of minor as well as major QTLs. Therefore, genetic gain achieved is higher by MARS as compared to the MABC program (Bernardo and Charcosset, 2006).

The recurrent-selection method is routinely used mainly in cross-pollinated crops like maize and, this process can be improved with the help of molecular markers; therefore, the process is called marker-assisted recurrent selection (MARS). While several multinational companies have been using MARS in crops like maize and soybean, only a few public sector institutes have started to use MARS in crops like wheat (Charvet et al., 2001), sorghum (Abdallah et al., 2009) and rice (Grenier et al., 2012). Some efforts have been initiated to use MARS in the case of chickpea also, for assembling favorable alleles for drought tolerance using ICCV 04112×ICCV 93954 and ICCV 05107×ICCV 94954 crosses. IARI and IIPR also have initiated MARS in...
chickpea by using Pusa 372 × JG130 and DCP92-3 × ICCV 10 crosses. These efforts are expected to result in superior lines with enhanced drought tolerance.

### 4.3. Genomic selection (GS)

Genomic selection (GS) or genome-wide selection (GWS) unlike MABC or MARS approaches targets identification of superior lines with higher breeding value in a breeding program based on genome-wide marker profile data. As breeding values are estimated using the genome-wide marker data, these are generally referred as genomic-estimated breeding values (GEBVs). In brief, GS employs two populations: (i) ‘training population’, that is generally comprised of breeding lines that were/ are in use in a breeding program and phenotyping data, not for some traits, but for overall performance (e.g. yield and yield components) are available across the environments, and (ii) ‘candidate population’, which is generally being used currently by breeders. This population may be derived from the parental lines that are present in the training population. In the first step of GS, all the individuals of the training population are genotyped with large number of markers by considering linkage disequilibrium (LD) in the breeding germplasm collection. Based on historical phenotyping data and genotyping data, statistical models are developed for estimating GEBVs of the lines. Subsequently, marker genotyping data generated on the candidate population are used with the models or depending on the skills of the crop breeders, can be selfed for suitable progenies that can either be used for next cycle of crossing or depending on the skills of the crop breeders, can be selfed for field

<table>
<thead>
<tr>
<th>Traits studied</th>
<th>QTL Genes</th>
<th>Markers linked</th>
<th>PVE (%)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agronomic &amp; yield</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flowering date</td>
<td>qPD02.1</td>
<td>AHC52736-AHC51251</td>
<td>19.5</td>
<td>Shirasawa et al. (2012)</td>
</tr>
<tr>
<td>Angle of branch</td>
<td>qA005.1</td>
<td>AHC5324-AHC52622</td>
<td>11.9</td>
<td>Shirasawa et al. (2012)</td>
</tr>
<tr>
<td>Length of main stem</td>
<td>qLMS04.2</td>
<td>AHC5215-AHC53725</td>
<td>19.2</td>
<td>Shirasawa et al. (2012)</td>
</tr>
<tr>
<td>qLMS05.2</td>
<td>AHC5220-AHC52450</td>
<td>15.7</td>
<td>Shirasawa et al. (2012)</td>
<td></td>
</tr>
<tr>
<td>Length of the longest branch</td>
<td>qL2B06.2</td>
<td>AHT0697-AHTC1H7</td>
<td>21.1</td>
<td>Shirasawa et al. (2012)</td>
</tr>
<tr>
<td>qL2B01.2</td>
<td>AHT0745-AHT0816</td>
<td>14.2</td>
<td>Shirasawa et al. (2012)</td>
<td></td>
</tr>
<tr>
<td>Number of branches</td>
<td>qNB06.2</td>
<td>AHT0697-AHT0074</td>
<td>15.6</td>
<td>Shirasawa et al. (2012)</td>
</tr>
<tr>
<td>Weight of plant</td>
<td>qWP06.2</td>
<td>AHT0697-AHTC1H7</td>
<td>11.8</td>
<td>Shirasawa et al. (2012)</td>
</tr>
<tr>
<td>Mature pod wt/plant</td>
<td>qWMP09.2</td>
<td>AHC5642-AHC52635</td>
<td>28.1</td>
<td>Shirasawa et al. (2012)</td>
</tr>
<tr>
<td>Length of pod</td>
<td>qPL05.1</td>
<td>AHT0601-AHC5143</td>
<td>28.2</td>
<td>Shirasawa et al. (2012)</td>
</tr>
<tr>
<td>qPL06.2</td>
<td>AHT0745-AHT0816</td>
<td>20.5</td>
<td>Shirasawa et al. (2012)</td>
<td></td>
</tr>
<tr>
<td>Pod thickness</td>
<td>qPT07.1</td>
<td>AHC51803-AHT0025</td>
<td>21.7</td>
<td>Shirasawa et al. (2012)</td>
</tr>
<tr>
<td>Pod width</td>
<td>qPW07.1</td>
<td>AHT0602-pPGPseq26b</td>
<td>15.2</td>
<td>Shirasawa et al. (2012)</td>
</tr>
<tr>
<td>qPW08.2</td>
<td>AHC51286-AHC52249</td>
<td>25.5</td>
<td>Shirasawa et al. (2012)</td>
<td></td>
</tr>
<tr>
<td>Pod constriction</td>
<td>qCP09.2</td>
<td>AHC50362-AHT0726</td>
<td>18.1</td>
<td>Shirasawa et al. (2012)</td>
</tr>
<tr>
<td>Seed weight</td>
<td>qVS08.2</td>
<td>AHT0846-AHT0974</td>
<td>19.1</td>
<td>Shirasawa et al. (2012)</td>
</tr>
<tr>
<td>Stem diameter</td>
<td>SDD2</td>
<td>pPGPseq2G3-TCTA02</td>
<td>24.1</td>
<td>Liang et al. (2009)</td>
</tr>
<tr>
<td>Total dry weight (TDW)</td>
<td>Total DWVWW09_AhIX</td>
<td>TCC504-GM1949</td>
<td>22.39</td>
<td>Gautami et al. 2012a</td>
</tr>
<tr>
<td>Harvest index (HI)</td>
<td>Hi Control 08_AhIX</td>
<td>GM1922-GM2050</td>
<td>40.1</td>
<td>Gautami et al. 2012a</td>
</tr>
<tr>
<td>Shoot dry weight (SDW)</td>
<td>ShDWWS08_AhVIII</td>
<td>GM1979-GM1919</td>
<td>22.09</td>
<td>Gautami et al. 2012a</td>
</tr>
<tr>
<td>Haulm weight</td>
<td>HaulmWVWU08_IV</td>
<td>TC1D02-TC1E05</td>
<td>33.36</td>
<td>Ravi et al. (2011)</td>
</tr>
<tr>
<td>Canopy conductance</td>
<td>ISO04_IVa</td>
<td>19H03-PM418</td>
<td>22.24</td>
<td>Ravi et al. (2011)</td>
</tr>
</tbody>
</table>

**Biotic stress**

<table>
<thead>
<tr>
<th>Trait</th>
<th>QTL Genes</th>
<th>Markers linked</th>
<th>PVE (%)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leaf rust</td>
<td>qLRT01</td>
<td>IPAHM103</td>
<td>55.2</td>
<td>Khedikar et al. (2010)</td>
</tr>
<tr>
<td>Late leaf spot (LLS)</td>
<td>qLLS03</td>
<td>GM2009-GM1536</td>
<td>82.27</td>
<td>Sujay et al. (2012)</td>
</tr>
<tr>
<td>Aspergillus flavus invasion</td>
<td>aJ01</td>
<td>GM1953-GM2057</td>
<td>62.35</td>
<td>Sujay et al. (2012)</td>
</tr>
<tr>
<td>Tomato spotted wilt virus (TSV)</td>
<td>qTSWV1</td>
<td>IPAHM287</td>
<td>12.9</td>
<td>Qin et al. (2012)</td>
</tr>
<tr>
<td>Aphid vector of groundnut</td>
<td>qMn1</td>
<td>Sep3D0F</td>
<td>35.8</td>
<td>Qin et al. (2012)</td>
</tr>
<tr>
<td>Nematode resistance</td>
<td>Rna</td>
<td>S197, GM565</td>
<td>-</td>
<td>Chu et al. (2007), Nagy et al. (2010)</td>
</tr>
</tbody>
</table>

**Oil and protein**

<table>
<thead>
<tr>
<th>Trait</th>
<th>QTL Genes</th>
<th>Markers linked</th>
<th>PVE (%)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein content</td>
<td>qPT01</td>
<td>TCE02-TCE010</td>
<td>10.2</td>
<td>Sarvarangala et al. (2011)</td>
</tr>
<tr>
<td>Oily content</td>
<td>qPT01</td>
<td>TCOH03-TCA1A04</td>
<td>10.7</td>
<td>Sarvarangala et al. (2011)</td>
</tr>
<tr>
<td>High oleate trait</td>
<td>FAD2A, FAD2B</td>
<td>af19:10568, bf19:1FAD</td>
<td>89.7</td>
<td>Chu et al. (2007, 2009), Shirasawa et al. (2012)</td>
</tr>
</tbody>
</table>
evaluation in targeted environments and can be advanced for multi-
location field trials (Fig. 3).

In summary, GS minimizes time-duration and cost by reducing the
frequency of extensive phenotyping and bypasses the need for QTL
mapping. GS can also reduce the selection cycle length of a breeding
program that could take several seasons to develop reliable pheno-
types. However, use of appropriate statistical model is very critical
for estimating the GEBVs with higher precision. Among different
models of GS, GEBVs predicted using either best linear unbiased pre-
diction (BLUP) or Bayesian methods are more effective according to
simulation studies (Bernardo and Yu, 2007). In addition to a Bayesian
method, Bayes B, another method called wBSR (weighed Bayesian
Shrinkage Regression) which reduces computational burden on MCMC-based Bayesian methods is considered to be a method of choice
for genomic selection (Takeshi and Hiroyoshi, 2010). Several studies
considered prediction using GS models, for instance Crossa et al.
(2010) used two data sets including a historical wheat phenotypic
data from trials evaluated in ten environments and another data set
pertaining to maize for two diseases (Eisserohilum turcicum and
Cercospora zae-maydis) from five environments. In both the cases
models used marker data and a gain in the predictive ability was
observed. Other groups including University of Oulu (Karkkinen and
Silampaa, 2012), Cornell University (Jammink et al., 2010), University of
Minnesota (Bernardo and Yu (2007), Hohenheim University (Piepho,
2009) also developed statistical models and/or pursued applications
of GS in breeding of some major crops like maize, wheat, etc. Though
GS has not been used in any legume species at present, due to
availability of: (i) historical phenotyping data on several breeding
lines (that can be used for training population), (ii) big linkage disequi-
librium (LD) blocks in breeding populations, and (iii) genome wide
marker genotyping system like DAfT and SNP markers, GS seems to
be a deployable approach in coming future at least in chickpea and
groundnut.

4.4. Advanced backcross QTL analysis based breeding (AB-breeding)

All above-mentioned molecular breeding approaches (MABC, MARS and GS) are useful only when the superior alleles for the trait
of interest are available in the breeding germplasm collection i.e., in
primary gene pool. However, presence of genetic variability for a
particular trait may not be available in primary gene pool such as
resistance to pod borer in the case of chickpea and pigeonpea. In
such cases, breeders need to utilize the potential of wild relatives
that are considered reservoirs of superior alleles for traits that
might have been lost during domestication and breeding. However,
this is not a straight forward approach as most breeders are reluctant
to use the wild relatives for transferring traits of interest from wild
relatives to the cultivars because of not having efficient tracking for
desired and non-desired alleles in breeding lines. To solve above
problem, advanced-backcross QTL based breeding (AB-breeding)
approach is the most suitable for introducing novel alleles from wild
relatives to the cultivated species in a controlled manner. In AB-
breeding approach, a selected wild species is backcrossed to a cultivar
or a variety and then, selection is imposed in segregating BC$_2$F$_2$ or in
BC$_2$F$_3$ population to identify and preserve individuals with desirable
traits in the population. Both genotyping and phenotypic data are
generated with this segregating BC$_2$F$_2$ or BC$_2$F$_3$ and, these data sets
will be subjected to QTL analysis to identify QTL, QTL associated
markers and, also to check whether any of these QTL are involved in
trait improvement in the progenies that are preserved. Therefore,
AB-QTL strategy involves the parallel discovery and transfer of
desired QTL from an unadapted germplasm into selected breeding
lines (Tanksley and Nelson, 1996). In addition, AB-QTL strategy
postpones the QTL mapping up to BC$_2$ or BC$_3$ generations to avoid
problems associated with incompatibility and pollen fertility in the
initial backcross populations as well as to ensure maximum genome
recovery from the recurrent parent. AB-breeding has been initially
practiced in a vegetable crop like tomato, where crosses between
wild tomato species and elite tomato lines were generated and QTLs
for various fruit characters were identified and introgressed success-
fully (Fulton et al., 2000). The precision of QTL identification increases
with a backcross population like BC$_2$ or BC$_3$ and, it offers adequate
statistical power and ensures sufficient similarity to the recurrent
parent. In addition, it also provides an opportunity to select for
QTL-near isogenic lines (NILs) in a short time span. Using QTL-NILs,
the QTL effects can be established and NILs may serve either as im-
proved varieties or as parents for use in hybridization programs and
for studies related to heterosis. In the case of chickpea, attempts
were made for making wide crosses using cultivated and wild
chickpea (C. arietinum × C. reticulatum) which resulted in selection of
progenies with increased seed yield (Jaiswal et al., 1986). Similarly,
Singh and Ocampo (1997) identified transgressive segregation for agro-
nomic traits in an F$_2$ population of a cross, C. arietinum × C. reticulatum.
These reports offer some opportunities in exploiting genetic variation
present in wild species of chickpea and hence, fresh initiatives has been
initiated at ICRISAT to introgress stress resistance through AB-breeding
approach (N Mallikarjuna, ICRISAT, personal communication).

Realizing the scope for AB-breeding in improvement of pigeonpea,
initiatives have been taken at ICRISAT to develop two backcross
populations (ICPL 87119 × ICPW 29 and ICPW 87119 × ICPW 12) for
AB-QTL analysis and their subsequent use in AB-breeding. ICPW 29 is
an accession of C. cajaniformis species and ICPW 12 is an accession of
C. acutifolius species. At present, for both the crosses, BC$_2$F$_3$ seeds have
been generated for multifacation phenotyping and selection.

In groundnut, tetraploidization event restricted the sharing of
genomic regions between wild and cultivated groups due to differ-
ence in ploidy levels, which has created a serious genetic bottleneck
i.e., narrow genetic base. Though conventional approaches were
used for attempting wide crosses through different ways such as
use of autotetraploids and allotetraploids, these efforts seriously posed
problems of fertility barrier, linkage drag and, a great difficulty in track-
ing introgressed alien genomic regions (see Bertoloi et al., 2011). Of
the above three important barriers, at least the later two (linkage drag
and tracking of alien genomic regions) can be efficiently handled by
integrating genomics into routine breeding programs to diversify the
narrow primary gene pool of groundnut. AB-breeding can help in track-
ing alien genomic regions and hence, the linkage drag can easily
be taken care of. Two major studies by Simpson et al. (1993) and
Fa‘vero et al. (2006) reported development of three amphiploids
using a rage of wild AA and BB genome species like A. cardenasiis,
A. diogoi and A. batizocioi. A. ipaensis, A. duranensis, A. Gregoryi and
A. linearifolium. More recently, in order to diversify the primary gene
pool and conduct AB-QTL analysis, ICRISAT has developed a set of 17
amphloid and autotetraploid groundnut (Mallikarjuna et al., 2011).

Furthermore, two AB-QTL mapping populations namely ICG 91114
(cultivated) × ISATGR 1212 (A. duranensis ICG 8123 × A. ipaensis ICG
8206, synthetic amphidiplod) and ICG 87846 (cultivated) × ISATGR
265-5A (A. kempff-mercadoi ICG 8164 × A. hoehnii ICG 8190, synthetic
ampidiplod) have been developed (Mallikarjuna et al., 2011). These
populations are segregating for several biotic, abiotic and agronomic
traits. A subset of 183 and 184 BC$_2$F$_3$ individuals, respectively have
also been genotyped with DAfT markers to construct genetic maps.
These two populations are planned to be phenotyped for several eco-
logical traits in multiple locations and seasons. A successful effort for
gene-wise segment introgressions from a synthetic amphidiplod
(A. duranensis × A. ipaensis) to a cultivated variety (Fluer 11) using molecu-
lar markers has already been reported (Foncêka et al., 2009). The
backcross BC$_2$F$_1$ and BC$_2$F$_3$ lines carrying the wild genomic segments
with maximum recurrent parent genomic regions provided optimal dis-
tribution of the synthetic genome introgressions (Foncêka et al., 2009).

Keeping in view, the low genetic diversity in all the three legume
crops, this approach is required urgently for diversifying the primary
gene pool with favorable alleles to enhance the chances of further crop improvement. Availability of large number of markers in recent years has ensured limiting linkage drag through stringent background selection and tracking the presence of non-desirable genomic region from the wild relatives.

5. Summary and outlook

Until 2005, there was a shortage of genomics resources in these three SAT legume crops and therefore these crops were often referred to as ‘orphan legume crops’. Nevertheless, the collaborative and coordinated efforts of the legume community made during the last 5 years, supplemented with generous financial support from several agencies, however, contributed to development of large-scale genomic resources in these crops. As a result, these crops are no longer ‘orphan legumes’ and have become ‘genomic resource rich’ crops (Varshney et al., 2009a, 2010b). These resources have also been used to understand the genetics of traits of several traits and as a result, approaches like MABC, MARS and AB-breeding are being used in these crops. GS seems to be a potential approach to be used very soon in chickpea and groundnut. While genome sequence has become available in pigeonpea, molecular breeding approaches have not yet been initiated as it may take more efforts to identify major QTLs for traits like FW, SMD and fertility restoration.

As genome sequence is expected to become available soon for chickpea and groundnut also, all the three legume crops mentioned in this article will have many more opportunities for practicing GAB. Genome sequence will facilitate re-sequencing of breeding populations, germplasm collections in faster and probably cheaper manner. In addition, availability of genome sequence will enable easy detection of variation at nucleotide level even among closely related parental lines and enhance exploitation of this variation for trait improvement in the breeding program. Analysis of genome-wide allelic data with the phenotyping data on germplasm collections will provide the alleles and haplotypes associated with the traits. These advances will also encourage the legume breeders to develop specialized populations like nested association mapping (NAM) populations or multi-parents advanced generation intercross ( MAGIC) populations. These populations may not be just useful for fine mapping of traits but also in the development of lines with enhanced genetic diversity (in the case of MAGIC populations) which is very much required in these low-diversity species. In this case, AB-population will also be very helpful to enhance genetic base of the legumes.

Once genome-wide allelic and haplotype data available on at least leading germplasm including breeding lines, varieties or segregating populations and haplotype-trait association are established, it may be possible for legume breeders to undertake breeding-by-design. Rich databases of genome sequences/haplotypes, phenotypes, marker-trait associations may facilitated legume breeders to select the parental lines and consider different crossing schemes, so that superior lines with enhanced resistance to diseases and tolerance to abiotic stresses with other market quality traits can be generated. It can be anticipated that coming years will be more exciting for integrating GAB tools and approaches in conventional breeding programs. While generating sequence or genotyping data is expected to be trivial, the legume scientists need to work on precise and cost-effective phenotyping and developing the decision support tools and breeders-friendly databases to ensure undertaking integrated breeding approaches for crop improvement.

Acknowledgements

The work presented in this article is a contribution from several research projects sponsored by CGIAR Generation Challenge Programme (Theme Leader Discretionary Grant), The Bill & Melinda Gates Foundation (Tropical Legumes I & II), the Department of Biotechnology (Centre of Excellence and Accelerated Crop Improvement projects), the Ministry of Science & Technology (Indo-German Science & Technology Cooperation and Australia–India Science & Research Foundation projects) and the Ministry of Agriculture (National Fund, Indian Council of Agricultural Research) of Government of India. Thanks are also due to several colleagues at ICRISAT and partners in collaborating centers engaged in genomics and molecular breeding research in SAT legumes.

References


