



Emerging Genomic Tools for Legume Breeding: Current Status and Future Prospects

Manish K. Pandey^{1*†}, Manish Roorkiwal^{1†}, Vikas K. Singh^{1†}, Abirami Ramalingam¹, Himabindu Kudapa¹, Mahendar Thudi¹, Anu Chitikineni¹, Abhishek Rathore¹ and Rajeev K. Varshney^{1,2*}

¹ International Crops Research Institute for the Semi-Arid Tropics, Hyderabad, India, ² The University of Western Australia, Crawley, WA, Australia

OPEN ACCESS

Edited by:

Paul Gepts,
University of California, Davis, USA

Reviewed by:

Ramanjulu Sunkar,
Oklahoma State University, USA
Swarup Kumar Parida,
National Institute of Plant Genome
Research, India

*Correspondence:

Rajeev K. Varshney
r.k.varshney@cgiar.org;
Manish K. Pandey
m.pandey@cgiar.org

[†] These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Plant Genetics and Genomics,
a section of the journal
Frontiers in Plant Science

Received: 24 November 2015

Accepted: 24 March 2016

Published: 02 May 2016

Citation:

Pandey MK, Roorkiwal M, Singh VK,
Ramalingam A, Kudapa H, Thudi M,
Chitikineni A, Rathore A
and Varshney RK (2016) Emerging
Genomic Tools for Legume Breeding:
Current Status and Future Prospects.
Front. Plant Sci. 7:455.
doi: 10.3389/fpls.2016.00455

Legumes play a vital role in ensuring global nutritional food security and improving soil quality through nitrogen fixation. Accelerated higher genetic gains is required to meet the demand of ever increasing global population. In recent years, speedy developments have been witnessed in legume genomics due to advancements in next-generation sequencing (NGS) and high-throughput genotyping technologies. Reference genome sequences for many legume crops have been reported in the last 5 years. The availability of the draft genome sequences and re-sequencing of elite genotypes for several important legume crops have made it possible to identify structural variations at large scale. Availability of large-scale genomic resources and low-cost and high-throughput genotyping technologies are enhancing the efficiency and resolution of genetic mapping and marker-trait association studies. Most importantly, deployment of molecular breeding approaches has resulted in development of improved lines in some legume crops such as chickpea and groundnut. In order to support genomics-driven crop improvement at a fast pace, the deployment of breeder-friendly genomics and decision support tools seems appear to be critical in breeding programs in developing countries. This review provides an overview of emerging genomics and informatics tools/approaches that will be the key driving force for accelerating genomics-assisted breeding and ultimately ensuring nutritional and food security in developing countries.

Keywords: trait mapping, gene discovery, genomics-assisted breeding, high-throughput genotyping, next-generation sequencing

INTRODUCTION

The demand-supply gap of food grain is continuously increasing due to the ever-growing global population which is likely to expand from 7.2 billion to 9.6 billion by 2050 and 10.9 billion by 2100 (Gerland et al., 2014). This burgeoning problem is becoming serious as the current yield increase trends may not be sufficient in dealing with the growing demand (Ray et al., 2013; Fedoroff, 2015). The speedy depletion of natural resources and climate change have badly affected the ongoing efforts to achieve higher productivity. In order to ensure hunger-free society with nutritious food, it is a challenge before the policy makers, farming community, and agriculture scientists to ensure nutritional food security by producing 60% higher food grain by 2050. Among the food grains, the grain legumes are the key sources of protein, minerals, vitamins, iron, zinc, calcium, and magnesium, as well as omega-3 fatty acids. The importance of these legumes is higher where

a large section of the society depends on vegetarian food such as in India. The oilseed legume crops such as groundnut (*Arachis hypogaea*) and soybean (*Glycine max*) play important role in the production of cooking oil and other confectionaries preparations to the consumers. The unique ability to fix atmospheric nitrogen by the legume crops play a crucial role in sustaining the farming system by making available the residual nitrogen to the non-legume crops. Legumes also serve as an excellent source of high quality and nutritious feed to livestock leading to ~20% increase in animal productivity (Tarawali and Ogunbile, 1995). The human civilization has a long association with legume cultivation, i.e., since 6000 years which has contributed significantly toward economical, nutritional, and improving the livelihood of subsistence farmers across the world.

It is well known that proteins are essential macronutrient for growth as well as maintenance of human body and a minimum protein intake of 0.8, 1.5, and 1.0 g protein/kg body weight/day is recommended for adults, children, and adolescents, respectively (Kafatos and Hatzis, 2008). The major protein sources include meat, fish, eggs, poultry, dairy products, legumes, nuts, and grains. Among legumes, the highest proportion of protein content is in soybean (33–45%) followed by common bean (*Phaseolus vulgaris*; 21–39%), winged bean (*Psophocarpus tetragonolobus*; 30–37%), cowpea (*Vigna unguiculata*; 21–35%), groundnut (24–34%), mung bean (*Vigna radiata*), pea (*Pisum sativum*; 21–33%), moth bean (*Vigna aconitifolia*), urd bean (*Vigna mungo*; 21–31%), lentil (*Lens culinaris*; 20–31%), grass pea (*Lathyrus sativus*; 23–30%), chickpea (*Cicer arietinum*; 15–30%), horse gram (*Macrotyloma uniflorum*), pigeonpea (*Cajanus cajan*; 19–29%), and rice bean (*Vigna umbellata*; 18–27%; Salunkhe et al., 1985). The productivity of these legumes could not be enhanced enough to meet the requirement, owing to several biotic and abiotic factors. In addition, the decreasing land and water resources together with climatic fluctuations will make situation worse, leading to protein unavailability to the human population in future. Due to ever increasing global population, especially in Asia where consumption of legumes is more, it is very important to increase the productivity of these legumes in order to meet the protein requirements for the future population.

The grain legumes cover approximately 71.8 million hectares with 50% area in Asia and 25% area in Africa. Although the productivity of legumes is 50% less than cereals, legumes fetch higher returns in the global market¹. Although conventional breeding approaches have been successful to address the issue of low productivity in some legumes, this is not happening at the desired success rate. Therefore, it is very essential to intensify the legume genetic enhancement programs using advanced breeding approaches wherein the potential of genomics needs to be exploited for accelerated development of improved cultivars possessing high yield, genetic resilience against stresses, and enhanced nutritional quality. The next-generation sequencing (NGS) and genotyping technologies need to be used for precise marker-trait association, gene discovery, functional marker development, and their deployment in routine breeding programs. The detailed information on

the availability of genetic and genomic resources in important legumes has been extensively reviewed in many articles (see Varshney et al., 2013c; Bohra et al., 2014; Varshney, 2016). In this review article, we have focused on the availability and deployment of modern and advanced genetic and genomic tools for conducting high resolution trait mapping and molecular breeding in three important legumes largely cultivated in semi-arid tropic (SAT) regions of the world, i.e., chickpea, pigeonpea, and groundnut. Further details on other emerging biological approaches for speedy identification of candidate genes and delivery of improved cultivars have also been provided, though some of these approaches are yet to be exploited in improving productivity, quality, and nutritional richness in these three important legumes.

NEXT – GENERATION GENOTYPING AND SEQUENCING TECHNOLOGIES

Several new genotyping platforms that leverage NGS technologies to discover and simultaneously genotype single nucleotide polymorphisms (SNPs) are currently available. As a result, the sequence-based genotyping (SbG) methods are becoming popular and method of choice for understanding the genetics of complex traits both in plants and animals. So far the SbG technologies have been deployed extensively in genetic mapping, purity testing, establishing marker-trait associations, marker-assisted selection (MAS), and genomic selection (GS) for crop improvement (see Varshney et al., 2014b). These technologies can be broadly classified as (i) amplicon based targeted sequencing, (ii) reduced representation based sequencing, and (iii) hybridization based approaches. Above mentioned technologies are discussed in details at http://www.illumina.com/Documents/products/appspotlights/app_spotlight_ngg_ag.pdf. The choice of platform for genotyping depends on several factors like the scale of genotyping project, genome size, and availability of funds. For instance, MAS, requires flexible, low-cost systems like LGC's automated systems for running KASPTM markers for genotyping smaller numbers of SNPs across large breeding populations. On the other hand, a wide range of options are available for custom genotyping of various number of samples × SNPs from Illumina and Affymetrix Technologies, Fluidigm's Dynamic ArraysTM, Douglas Scientific's Array TapeTM for whole genome scanning, constructing high density genetic maps, and genome-wide association studies (GWAS; see Thompson, 2014). In this review, details have been provided on examples of deployment of SbG technologies for legume research and breeding, with some insights into future technologies that will accelerate legume breeding efforts.

Amplicon based targeted sequencing is amenable for addressing questions in population genetics and systematics that rely on sequence specific genes of known function or diversity levels (Naj et al., 2011). However, lack of inexpensive and fast approaches that allow rapid library preparation using a standard PCR product of specific gene for thousands of samples and sometimes for 100s of loci that are essential for establishing phylogenetic relationships necessitate the development of

¹<http://www.generationcp.org/gcp-research/crops/legumes>

targeted amplicon sequencing approach (Bybee et al., 2011). Several approaches that improve the throughput by barcoding targeted amplicon have been extensively discussed in Mamanova et al. (2011). In fact, amplicon based sequencing targets only limited number of loci and is not useful for fine mapping of complex traits where genome-wide SNPs are essential. Nevertheless, genome-wide SNP calling often hindered by the complexity at genome level in case of some crops such as maize, wheat, groundnut, and soybean. Hence, reduced-representation based sequencing approaches that include reduced-representation libraries (RRLs) or complexity reduction of polymorphic sequences (CROPS), restriction-site-associated DNA sequencing (RAD-seq) and low coverage genotyping were developed and are being deployed. Further, van Orsouw et al. (2007) developed an approach known as CROPSTM which is based on restriction digestion of DNA using methylation sensitive restriction enzyme that reduces the complexity of two or more genetically diverse samples are prepared by AFLP. High conversion rate to genotyping assays makes it more attractive for medium to large scale genotyping particularly for SNPs in low or single copy genome sequences. While Miller et al. (2007) proposed the reduced representation library in which the size reduction was carried out using restriction enzymes, this technique was adapted to include barcoding and multiplexing with Illumina sequencing technologies (Baird et al., 2008). In addition, Double Digest Restriction Associated DNA (ddRAD) sequencing developed by Peterson et al. (2012) exclusively uses size selection to recover an appropriate number of regions, which are distributed randomly throughout the genome and maximizes the ability of multiplexing of several 100s of samples. Nevertheless, the genotyping-by-sequencing (GBS) approach which has been extensively utilized in crop improvement programs (Kim et al., 2016), is more powerful compared to the RAD sequencing approach because the SNP discovery and genotyping can be done at the same time (Poland et al., 2012). Further, compared to the RAD method, GBS is substantially less complicated in terms of generation of restriction fragments with appropriate adapters, fewer DNA purification steps, and no fragments selection steps. Recently, a novel GBS approach called skim-based GBS (skimGBS) that uses low-coverage whole genome sequencing (WGS) for high-resolution genotyping has been developed (Bayer et al., 2015). Using this approach, genome-wide recombination maps can be developed, and the frequency of crossover as well as gene conversion events can be assessed and compared. The SkimGBS, is a two-stage method that requires a reference genome sequence, genomic reads from parental individuals, and individuals of the population. The parental reads are initially mapped to the reference genome and SNPs are called using SGSautoSNP (Lorenc et al., 2012). Subsequently progeny reads are mapped to the same reference and comparison with the parental SNP file enables the calling of SNPs between parental genotypes and progeny.

In recent years, exome sequencing technique is gaining importance as this enables discovery of many low-frequency and rare coding variants that need to be examined systematically for association with complex traits in both plants and animals (Kiezun et al., 2012). Various exome sequencing platforms and

their application for crop improvement as well as health are discussed in detail (see War et al., 2015). Affymetrix and Illumina exome arrays can code the variations like SNPs and InDels and can offer custom content. Nevertheless, Affymetrix exome arrays have capacity for 100,000 additional markers compared to Illumina exome arrays. In addition to the exome arrays, high density SNP arrays have shown great promise as an efficient alternative to GBS approach often constrained by missing data and complex bioinformatics analysis (Unterseer et al., 2014). The higher efficiency in providing genotyping data for the most of the loci across the samples have made this approach most suitable for use in GS breeding, wherein, the consistent data is required not only in training population (TP) but also for all the consequent breeding lines subjected for estimating genomic estimated breeding values (GEBVs). The high density SNP arrays have been successfully developed and deployed in rice (*Oryza sativa*; 50 K SNPs; Singh N. et al., 2015) and maize (*Zea mays*; 616 K SNPs; Unterseer et al., 2014) and will be available in several other important crop plants in coming days. With an objective to exploit the potential of such SNP arrays, the Affymetrix arrays with 60 K SNPs for three legume crops viz., chickpea, pigeonpea, and groundnut were developed by ICRISAT recently (see Varshney, 2016). The availability of these arrays will provide high-throughput, cost-effective, reproducible, and informative SNP genotyping data to facilitate high resolution trait mapping and molecular breeding.

SEQUENCING AND RE-SEQUENCING EFFORTS IN LEGUMES

Decoding of the plant genome sequence provides an opportunity to dissect and understand the mechanism or genetic basis for functional characterization of genes. Considering the importance of genome sequencing for crop improvement, several plant genomes have been decoded (Michael and Jackson, 2013). *Arabidopsis* (*Arabidopsis thaliana*) was the first plant species for which genome sequencing was completed (The Arabidopsis Genome Initiative, 2000) followed by three efforts to sequence *indica* and *japonica* rice (Goff et al., 2002; Yu et al., 2002; International Rice Genome Sequencing Project [IRGSP], 2005) using Sanger sequencing technology. NGS technology based WGS approach could reduce the time and cost of genome sequencing drastically as compared to Sanger sequencing (Schatz et al., 2002). Adoption of NGS technologies made it possible to sequence genome in much less time and cost, which encouraged researchers to even decode the complex genome sequences. By July 2013, >50 plant genomes were sequenced (Michael and Jackson, 2013) and more than 100 have been added since then. It is important to note that majority of these genomes were sequenced using either only NGS technologies or in combination of NGS and Sanger sequencing technologies.

Until recently, legumes namely chickpea, pigeonpea, and groundnut were considered orphan crops as not much genomic resources were available (Varshney et al., 2012b). With the advent of NGS technologies, efforts were made to develop

genomic resources for these crops (Varshney et al., 2013c). Among these legumes, genome sequence of the pigeonpea was first to be completed in 2012 (Varshney et al., 2012a). Illumina NGS technology along with Sanger based bacterial artificial chromosome (BAC) end sequences was used for assembling the pigeonpea genome. ICPL 87119 (also known as 'Asha,' an inbred line and a widely cultivated medium duration pigeonpea genotype) was sequenced using Illumina sequencing technology to generate 237.2 Gb of paired-end (PE) reads. Filtered 130.7 Gb high quality Illumina sequencing data and Sanger sequencing data for >88 K BACs were used to assemble 605.78 Mb representing 72.7% of the pigeonpea genome. Gene annotation using combination of *de novo* gene prediction and homology-based methods led to identification of ~48 K genes, though it seems to be on a higher side, in pigeonpea genome (Varshney et al., 2012a). All these genes were functionally annotated using combination of a range of approaches, which resulted in assigning tentative gene function to more than 95% of genes and less than 4% genes could not be functionally annotated. In another effort to sequence the pigeonpea genome, long sequence read of 454 GS-FLX pyrosequencing technology were used for assembling ~548 Mb of pigeonpea genome (Singh N.K. et al., 2012).

Next-generation sequencing approach was also used for sequencing the chickpea genome of CDC Frontier (a *kabuli* chickpea variety) by International Chickpea Genome Sequencing Consortium (ICGSC). Around 153 Gb sequence data were generated using Illumina sequencing technology of which 87.65 Gb of high-quality sequence data were assembled into 544.73 Mb of genomic sequence scaffolds representing 74% of chickpea genome (Varshney et al., 2013d). Using a combination of approaches, more than 28 K non-redundant gene models were predicted across the chickpea genome of which around 25 K (89.73%) could be functionally annotated. Along with the draft chickpea genome, sequence data were generated for 90 chickpea accessions using whole genome re-sequencing (WGRS) and RAD-sequencing approaches (Varshney et al., 2013d). In another effort to sequence the chickpea genome, ICC 4958 *desi* chickpea genotype was targeted for generating a draft genome assembly using NGS platforms along with BAC end sequences and a genetic map. A total of 13.35 Gb high quality sequencing data from 454/Roche GS FLX Titanium platform and 43.7 Gb of quality-filtered PE Illumina sequence data on ICC 4958 along with BAC end sequences were used to assemble ~520 Mb of chickpea genome (Jain et al., 2013). Very recently, an improved version of *desi* chickpea cultivar ICC 4958 with 2.7-fold increase in the length of pseudomolecules was reported (Parween et al., 2015). This improved assembly could reduce the gaps in the existing genome assembly and predicted the presence of more than 30 K protein-coding genes. In a similar effort, Ruperao et al. (2014) used chromosomal sequencing approach to validate the available *desi* and *kabuli* chickpea genome assemblies. Isolation and NGS based sequencing of individual chromosomes helped in validating the genome assemblies at chromosome level, which identified small misassembled regions in *kabuli* and large misassembled region in *desi* draft genomes (Ruperao et al., 2014).

Very recently, the International Peanut Genome Initiative (IPGI) successfully sequenced the genomes of diploid groundnut progenitors. The progenitors representing A-genome (*Arachis duranensis*, accession V14167) and B-genome (*Arachis ipaensis*, accession K30076) together represent the tetraploid genome of cultivated groundnut (*Arachis hypogaea*). In this context, a total of 216 Gb WGS data at 154.4X coverage for accession V14167 while 168.8 Gb data at 120.6X coverage for accession K30076 were generated (Bertioli et al., 2016). In addition, 155.5 Mb transcriptomic sequence data for accession V14167 and 175.6 Mb data for accession K30076 were generated from different plant tissues and growth stages. The above mentioned sequences were assembled into 10 pseudomolecules for each genome. The B-genome (1.37 Gb) had larger genome size than A-genome (1.2 Gb). The availability of the genome sequence will provide access to 97% of groundnut genes in their genomic context to the global groundnut research community leading to development of better understanding about the complex groundnut genome and accelerated development of more productive, climate, and stress resilient groundnut varieties.

The draft genome sequence serves as the foundation for deploying genomics in crop improvement for accelerating the rate of genetic gains by identifying the genes responsible for economically important traits. The draft genome sequence also helps to understand the genome architecture and unravel the basic mechanism involve in the stress responses. The availability of draft genome sequence has enabled the undertaking of large scale genome re-sequencing projects for identification of genetic variations, single base mutations, insertions, and deletions. Following the completion of draft reference genome, cataloging the sequence variations by comparing the genome sequence at individual and population level helps in understanding the mechanism of plant's response to different stresses. Large scale germplasm resources available in genebanks world-wide provide opportunity to the global plant science community to mine superior alleles (McCouch et al., 2013). ICRISAT's genebank has about 50,000 accessions of cultivated species and wild relatives of chickpea, pigeonpea, and groundnut from 133 countries (Gowda et al., 2013). In order to identify new sources of genetic variation and allelic variants of candidate gene(s) associated with beneficial traits by exploring the huge genetic diversity available in the genebank, ICRISAT has initiated the efforts to re-sequence the germplasm for identification of novel alleles. In the case of pigeonpea, 292 *Cajanus* accessions from reference set were re-sequenced using WGRS approach and were used to identify 13.8 million sequence variations (SNPs and InDels) with average of 41.1 variations per Kb (Saxena et al., 2015). Detailed analysis of re-sequencing data provided genetic variation patterns across the pigeonpea genome along with identification of genomic regions that are expected to play an important role during domestication and selection. Besides improved understanding of the genome, such genetic variations together with phenotypic data are also being used for undertaking marker-trait association analysis for identification of markers associated with trait of interest. In addition, 104 pigeonpea hybrid parental lines were also re-sequenced using

WGRS approach. The hybrid parental lines includes, CMS lines, maintainer and restorer lines. The generated genotyping datasets along with phenotyping of parental lines and there of derived hybrids will be utilized for construction of heterotic pool in pigeonpea.

In the case of chickpea, WGRS approach was used for resequencing >400 chickpea genotypes and were analyzed to identify ~4.7 million SNPs, >500,000 Indels and CNVs. Resequencing data on >100 elite chickpea varieties were used for developing first generation Hapmap of chickpea. Re-sequencing data on 300 lines form chickpea reference set along with available phenotyping data is being used for identification of markers associated with trait of interest. Considering the utility/application of WGRS in the crop improvement program, very recently ICRISAT had launched “The 3000 Chickpea Genome Sequencing Initiative” where 3000 lines from the global composite collection of chickpea from genebank of ICRISAT and ICARDA are being re-sequenced for identification of novel alleles (see Varshney, 2016).

ADVANCED/MULTI-PARENT MAPPING POPULATIONS FOR HIGH RESOLUTION MAPPING

For achieving accelerated genetic gains, a number and combination of genomics-assisted breeding (GAB) approaches need to be used. Marker-assisted backcrossing (MABC) for improvement of single or multiple traits (gene pyramiding) is the most successful GAB approach which has helped in the development of several improved varieties and lines in many crops. However in order to deploy GAB, the identification of genomic regions either through linkage mapping or association mapping is mandatory for the traits of interest. There are three trait mapping approaches using forward genetics namely (1) linkage mapping/quantitative trait locus (QTL) mapping, (2) association mapping/linkage disequilibrium (LD) mapping, and (3) joint linkage-association mapping (JLAM). Majority of the linkage mapping studies used either F_2 , recombinant inbred lines (RILs), backcross inbred lines (BILs), near isogenic lines (NILs), advanced intercrossed lines (AILs), or double haploid populations in crop plants. On the other hand, germplasm sets were used for association mapping. All the above mentioned mapping populations had a major disadvantage that only few traits can be mapped. Realizing the long time spent in developing these populations, the recent trend has shifted in development and deployment of multi-parent mapping populations. These populations include nested association mapping (NAM) and multi-parent advanced generation inter-cross (MAGIC) populations (Varshney and Dubey, 2009) (**Figure 1**). Gene discovery activities in the past were limited by the extent of availability of genomic data in a crop but with NGS technology there is a need to redesign the way to understand the genetics of traits of interest. These populations have advantages of both bi-parental (high power of QTL detection) and association mapping (high

resolution; Gupta et al., 2014). Therefore, these two populations achieve higher level of polymorphism, high resolution genetic mapping and QTL identification, and handling several traits in one go.

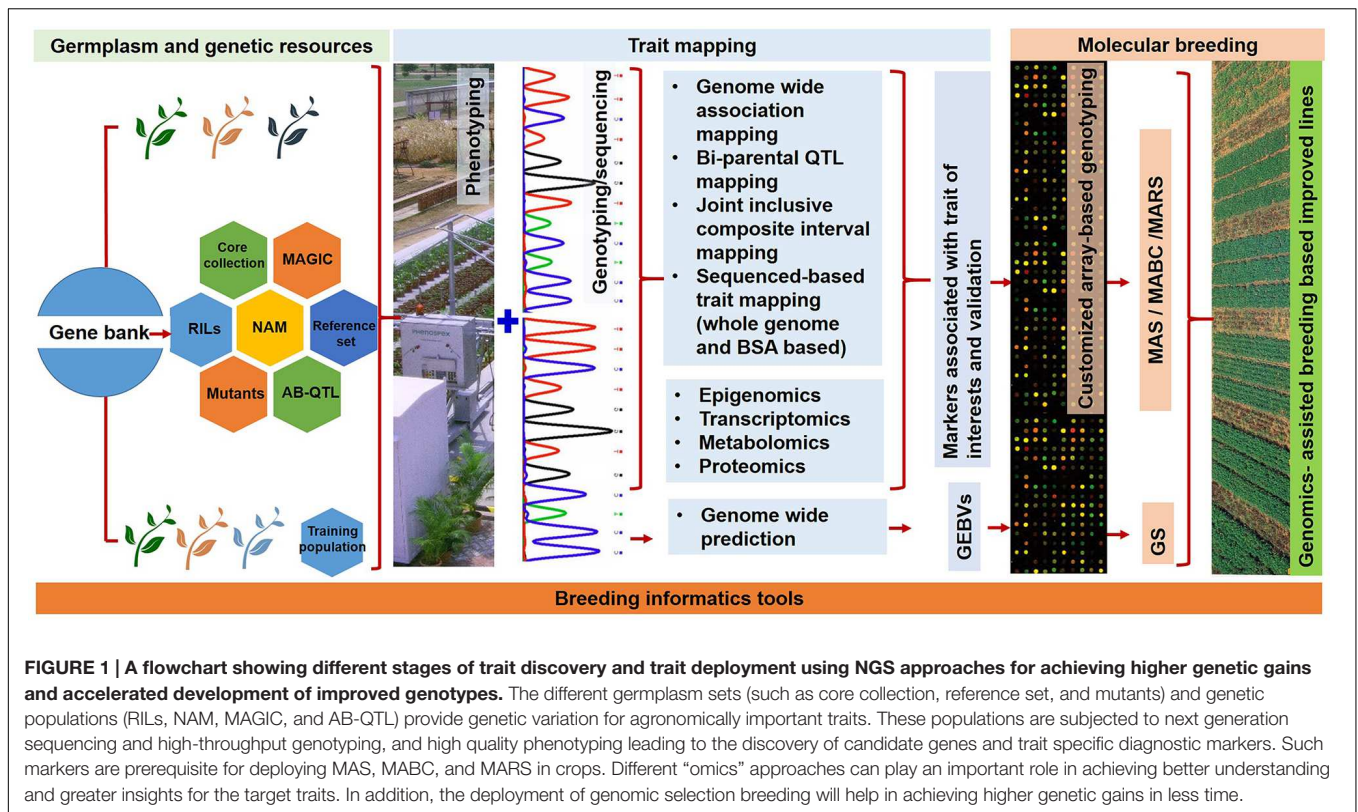
Advanced Intercrossed Line (AIL)

Development of large bi-parental population to achieve higher recombination events is not a cost-effective approach due to substantial cost involved in developing, phenotyping, and genotyping. In order to achieve higher resolution within a small population, it is important to increase the recombination events for the target loci. In this context, AIL population is very promising approach where in the two variant lines selected in F_2 generation are randomly and sequentially intercrossed to several generations (Darvasi and Soller, 1995). Such crossing efforts break linkage even between very closely linked loci due to several recombination events. Development and deployment of such populations may provide more precise location and estimates of genomic/quantitative trait loci, and therefore, a good genetic resource for fine mapping. Such mapping populations were largely deployed for fine mapping in several plant species (Huang X. et al., 2010). However such populations have not been deployed frequently in these legume crops.

Nested Association Mapping (NAM)

This approach takes benefits of both mapping approaches, i.e., linkage and association mapping and has shown ability to identify and resolve functional markers for complex traits. It is important to note that the success depends upon frequency of functional marker alleles, magnitudes of their genetic effects, disequilibrium among functional and non-functional markers, statistical analysis methods, and mating design (Guo et al., 2010). Technically, the NAM population makes use of both primitive and recent recombination events to take advantage of low marker density requirements, allele richness, high mapping resolution, and high statistical power. Since this approach involves several parental genotypes which led to generation of lines from nested mating designs, and thereby more chance of resolving the functional markers as compared to the bi-parental populations.

The NAM populations are currently being developed at ICRISAT (see Varshney, 2016). In the case of chickpea, with an objective to generate new breeding material with enhanced diversity for high resolution mapping of target traits, ICRISAT along with its partner is developing NAM population having ICC 4958 as common female parent. In total 14 different crosses were initiated and generated F_1 s. These F_1 s are being advanced to generate minimum of 200 lines from each cross. In pigeonpea, for mapping different traits (FW, SMD, yield and yield related traits and seed protein content), NAM population is being developed. In this regard 10 different F_1 combinations were generated by crossing Asha variety (ICPL 87119) with 10 different elite pigeonpea lines. As a result a total of 10 NAM- $F_{2:3}$ populations were developed. These developed populations are currently being genotyped through GBS based approach and phenotyped for the target traits for identification of marker-trait associations. Similarly, in the case of groundnut, two NAM populations, i.e., one each for Spanish type (ICGV 91114 and 22 testers) and



Virginia type (ICGS 76 and 21 testers) are being developed at ICRISAT.

Multi-parent Advanced Generation Inter-Cross (MAGIC)

The MAGIC population is generated from multiple parents (4–8) of diverse origins including alien backgrounds possessing desired features to bring multiple favorable alleles in multiple combinations in the population. This population allows both coarse and fine mapping and the complex architecture of many traits which are associated with crop yield and quality can be deduced using epistatic interactions (Cavanagh et al., 2008). MAGIC populations, in future, will serve as important resources for the discovery, isolation, and transfer of essential genes to facilitate crop improvement.

The MAGIC populations are currently being developed at ICRISAT (see Varshney, 2016). In the case of chickpea a MAGIC population with around 1200 lines were developed using eight parents including cultivars and elite breeding lines (ICC 4958, ICCV 10, JAKI 9218, JG 11, JG 130, JG 16, ICCV 97105, and ICCV 00108) from India and Africa. These eight diverse parents were crossed in 28 two-way, 14 four-way, and 7 eight-way crosses for accumulation of recombination events to allow genome reshuffle to bring greater genetic diversity (see Varshney, 2016). In order to access the allele diversity in the MAGIC population, the population have been re-sequenced using WGRS approach and generated a total of 4.67 Tb clean sequence data.

Alignment of re-sequence data to the reference genome led to identification of 1 million SNP variants. These SNPs are being used for further understanding genome diversity and haplotype analysis.

In pigeonpea, MAGIC population is also being developed to enhance the genetic base and to identify the marker trait associations. A total of eight diverse founder parents (ICP 5529, HPL 24, ICP 7035, ICP 8863, ICP 4486, ICP 11605, ICP 7426, and ICP 14209) were selected for the development of MAGIC population. Half di-allele crossing approach (28, two-way F₁s) followed by funnel based mating design (14 four-way and 7 eight-way F₁s) was utilized for the development of MAGIC lines. Currently, 7 eight-way F₁s are being selfed in the controlled conditions for the development of high resolution MAGIC lines. In the case of groundnut, three MAGIC populations are under development targeting different trait combinations. The first MAGIC population (ICGV 88145, ICGV 00308, ICGV 91114, ICGV 06040, ICGV 00440, ICGV 05155, GPBD 4, and 55-437) targets different traits such as fresh seed dormancy, oil content, seed mass, kernel Fe and Zn content, aflatoxin tolerance, stem rot tolerance, and PBNB tolerance. The second MAGIC population targets the different components of aspergillus resistance and aflatoxin contamination while the third one targets different component traits of drought tolerance. The genotypes for aflatoxin resistance include 55-437, ICG 51, ICGV 12014, U4-7-5, VRR245, ICGV 88145, ICGV 89104, and ICGV 97278 while for drought tolerance include ICGV 02022, ICG 7190, ICGV 97183, ICG 3053, ICG 14482, ICG 11515, TAG 24, and ICGV 02266.

SEQUENCING-BASED TRAIT MAPPING IN LEGUMES

Advances in NGS technologies and a steeper drop down in the cost of sequencing provides a significant opportunity for trait mapping at sequence level and selection of plants at nucleotide levels. Traditional approaches of trait mapping were time-consuming and costly in comparison to NGS-based approaches. Recently, a large number of sequencing-based approaches have been proposed and used for trait mapping, which can be mainly classified in two groups (i) trait mapping through sequencing of complete populations and (ii) trait mapping through sequencing of pooled samples (Figure 1). The brief description of the specific approaches and current status in different legumes are given below.

Trait Mapping through Sequencing of Complete Populations

The GBS and WGRS of mapping populations provide large scale genome-wide SNPs for conducting high resolution trait mapping.

The GBS (Elshire et al., 2011) has been found promising approach for rapid identification of large number of genome-wide SNPs for diversity assessment, trait mapping, GWAS, and GS in several crops (see He et al., 2014). GBS has gained popularity due to low genotyping cost for providing high-density genotyping data. In addition, it also provides the edge over other sequencing-based methods as this does not require the prior genome information. GBS has been extensively used for genetic diversity analysis (Fu et al., 2014), developing dense genetic maps (Kujur et al., 2015b; Li et al., 2015), refine the target genomic regions (Jaganathan et al., 2015), marker-trait associations (Romay et al., 2013) as well as deployment in GS breeding (Poland et al., 2012 in wheat, Crossa et al., 2013 in maize, and Huang et al., 2014 in oat). Applications of GBS in crop improvement have been extensively reviewed in many research papers (see Heschamps et al., 2012; Poland and Rife, 2012; He et al., 2014; Kim et al., 2016). For instance, GBS approach was used successfully for enriching the existing genetic map of chickpea developed with 241 SSR loci (Varshney et al., 2014c) to 1,007 loci including 828 SNPs identified using GBS along with earlier mapped SSRs (Jaganathan et al., 2015). Interestingly GBS could saturate the “QTL-hotspot” region identified by Varshney et al. (2014c) that harbors QTLs for drought tolerance component traits, by integrating 45 additional GBS-SNPs narrowing down the genomic region from 29 cM to 14 cM (Jaganathan et al., 2015). Additionally, GBS based approach was used for development of high-density linkage map in chickpea comprising of 1,336 SNPs (Deokar et al., 2014). In another study, one genetic map each for *desi* and *kabuli* type chickpea were reported with 3,625 SNPs and 2,177 SNPs, respectively. These high-density linkage maps were then used for identification of QTLs controlling seed weight in chickpea (Kujur et al., 2015b). GBS and Skim sequencing approaches also demonstrated their utility in improving the genome assemblies of both *desi* and *kabuli* chickpeas (Ruperao et al., 2014), which serve as a better reference for legume biology and comparative genomics. Complex admixed domestications

patterns of chickpea were reported using high-throughput GBS approach (Kujur et al., 2015a). GBS approach is also being utilized for generating genotyping data on ~20 mapping populations segregating for diverse targeted traits including fusarium wilt (FW) and sterility mosaic disease (SMD) resistance, seed protein content, yield and its associated traits in pigeonpea. Similarly, in groundnut the mapping population (TAG 24 × GPBD 4) is being genotyped using GBS approach for identification of candidate genomic regions for leaf rust, late leaf spot, and other agronomic important traits (see Varshney, 2016).

The WGRS of entire mapping population or diverse germplasm set is one of the promising approaches for many diverse studies; including identification of candidate genomic region/gene and GWAS (see Huang et al., 2013). The classical example of GWAS through WGRS was reported by Huang Y.-F. et al. (2010) in rice which identified marker trait associations for 14 agronomic traits after generating ~3.6 million SNPs by sequencing 517 rice landraces map. Similarly, to dissect the genetic architecture of oil biosynthesis in maize kernels, a total of 368 maize inbred lines were sequenced to perform GWAS (Li et al., 2013). This study identified 26 loci associated with oil concentration explaining up to 83% phenotypic variation. In the context of legumes, WGRS for performing GWAS is currently being utilized in 292 lines of reference set and 104 hybrid parental lines to define heterotic pool in pigeonpea (see Varshney et al., 2015b). Similarly, in the case of chickpea ~300 lines of reference set and 100 elite varieties have been sequenced for performing GWAS for economically important traits. In addition to this “The 3,000 Chickpea Genome Sequencing Initiative” has also been started to capture superior alleles for the targeted traits (see Varshney, 2016). In addition to GWAS, the first successful example of WGRS in segregation mapping population for identification of candidate gene(s) was reported in rice by Huang et al. (2009). Sequencing of RIL population with 150 individuals at lower coverage (~0.02× coverage) resulted in identification of 1,493,461 SNPs with average density of 25 SNPs/Mb or 1 SNP every 40 kb. The detailed analysis lead identification of 100-kb region containing the rice “green revolution” gene (Huang et al., 2009). In another study for pinpointing genes for root-knot nematode (RKN) resistance in soybean, a total of 246 RILs were sequenced at an average of 0.19X depth and identified two candidate genes (*Glyma10g02150* and *Glyma10g02160*) associated with RKN resistance in soybean (Xu et al., 2013). In this context, examples of WGRS of mapping population in the ICRISAT-mandate legume crops are still not reported, however, in near future with low cost of sequencing, it will be feasible to generate WGRS data for entire mapping population in these legumes as well.

It is important to note that the sequencing of mapping populations/genotypes through GBS approach suffers from the limitation of missing regions of the genome whereas the WGRS overcomes the missing data challenge. On the other hand, the WGRS offers better data quality, however, it is also a costly process. To save the cost of sequencing, WGRS can be done at lower depth and in that scenario the approach is referred as skim sequencing (Golicz et al., 2015). Using this approach,

Bayer et al. (2015) characterized the distribution of crossover and non-crossover recombination in rape mustard (*Brassica napus*) and chickpea using SkimGBS. GWAS based gene enrichment analysis of skim sequenced data of RIL population in chickpea was useful to split the earlier identified “QTL-hotspot” in two sub-regions viz. “QTL-hotspot_a” and “QTL-hotspot_b” of 139.22 and 153.36 Kb sizes, respectively (Kale et al., 2015). For targeting direct candidate genes from the segregating mapping populations and or diverse germplasm set, exome sequencing approach has been proposed. This is the most promising approach for sequencing all the protein-coding genes in a genome (known as the exome). This approach is also useful for those crops which are having higher genome size like groundnut, maize, barley (*Hordeum vulgare*) and wheat (*Triticum aestivum*; see Singh D. et al., 2012; War et al., 2015). Exome sequencing approach is currently being utilized to sequence 250 lines of groundnut for understanding the role of candidate genes for the targeted traits. Identification of non-synonymous SNPs substitution in the identified candidate genomic regions or through identification of putative nsSNPs through principal component analysis between a different set of genotypes is one of the promising approach. Recently, this approach has been used for identification of nsSNPs of the target candidate genes for sheath blight resistance in rice (Silva et al., 2012) and more complex trait like the drought tolerance in maize (Xu et al., 2014). Based on the advantages of this approach, it has also been utilized for the identification of candidate genes for FW and SMD resistance in pigeonpea (Singh V.K. et al., 2015).

Trait Mapping through Pooled Sequencing

There are currently five trait mapping approaches wherein the sequencing of complete population is not required and the analysis is done on the sequences generated on pooled samples. These include QTL-Seq, MutMap, Seq-BSA, Indel-Seq, and BSR-Seq.

The “QTL-Seq” is the first, and most promising approach which has been successfully applied in crop plants with higher genome size. It has been utilized in the localization of the genomic regions for blast resistance and seedling vigor in rice (Takagi et al., 2013a); flowering QTL in cucumber (Lu et al., 2014) and fruit weight and locule number loci in tomato (Illa-Berenguer et al., 2015). In the case of legumes, this approach has also been found successful in the localization of QTLs/candidate genes for 100 seed weight in chickpea (Das et al., 2015). This study reported the identification of coding SNP in potential seed weight-governing candidate gene *CSN8*. Similarly, based on precise phenotyping of the cross ICC 4958 × ICC 1882, and sequencing of extreme bulks along with resistant parent were used to define candidate genes for 100-seed weight and root trait ratio (see Varshney, 2016). This approach has also found promising for identification of candidate genomic regions for late leaf spot and rust resistance in groundnut (see Varshney, 2016). In addition, this approach was tested in pigeonpea for localization of genomic regions for days to flowering and obcordate leaf shape (see Varshney, 2016).

The second approach called “MutMap” is a simple and robust NGS-based approach which was proposed in rice for the identification of candidate genes from promising EMS-induced mutants (Abe et al., 2012). The main advantage of this approach is that the mapping population developed for MutMap experiment requires crossing of selected mutant plant with the wild type, which minimizes the background noise. Thereafter, sequencing of extreme pool samples from segregating mapping populations along with a wild type parent are utilized for calculation of the genome-wide SNP index. Identification of SNP-index through specialized pipelines is useful for identification of candidate genes. Few more variants of MutMap approach are MutMap+ (Fekih et al., 2013; mapping without development of mapping population) and MutMap-Gap (Takagi et al., 2013b; useful approach for identification of candidate genes in the gap region, which was not sequenced through genome sequencing). Recently, MutMap approach has been found useful in identifying candidate gene for salinity tolerance leading to development of salt tolerance line through MABC (Takagi et al., 2015). In the context of legumes, this approach is being utilized for identification of candidate genes for leaf and plant type mutants in chickpea. Identified promising mutants in the genetic background of ICC 4958 are selected and crossed with wild-type parent. The developed M₂ populations are being phenotyped for the targeted traits. Based on the phenotypic datasets extreme pools will be constructed for sequencing and candidate gene identification. In addition to chickpea, EMS induced mapping populations are being developed in the case of pigeonpea and groundnut.

The third approach namely “Seq-BSA” is NGS-based simple and robust approach for identification of candidate SNPs in the targeted genomic regions. This approach works on the calculation of genome-wide SNP-index of both the extreme bulks using high trait parent as reference parent assembly using QTL-seq pipeline (Takagi et al., 2013a). Identified SNPs that were monomorphic for high trait parent and high trait bulk will show SNP index of ‘0’ due to the presence of a similar genomic region of a particular locus. However, identification of the SNP index value of ‘1’ in low trait bulk, with the same genomic positions might be the putative SNPs linked to the target traits. This approach has been successfully utilized for the identification of putative SNPs associated with FW and SMD in pigeonpea (Singh V.K. et al., 2015).

The fourth approach namely “Indel-Seq” also has emerged as the promising trait mapping approach which is largely based on insertions and deletions. To date, the proposed approaches for identification of genomic regions are based on identification of SNPs and thereafter utilization of different statistical approaches for a declaration of candidate genomic regions/genes. However, the presence of insertions and deletions in the candidate genomic region has been largely ignored and not targeted for trait mapping in any of the approaches. This approach has larger practical utility with the fact that most of the cloned genes in rice and other crop possess Indels in the reported candidate genes.

In case of fifth approach namely “Bulked segregant RNA-Seq (BSR-Seq),” the strength of RNA-seq approach was combined with BSA and new genetic mapping approach to identify

candidate genes for the target trait. This strategy has been successfully applied for the identification of *glossy3* genes of maize (Liu et al., 2012). Similar to this, RNA-seq of extreme pooled samples at high coverage were used to localize the candidate gene for grain protein content (GPC) gene *GPC-B1* in wheat (Trick et al., 2012). This approach has been found useful for the crops, which is having higher genome size (e.g., wheat 17 Gb and maize 2.3 Gb). This approach has more advantages in terms of cost saving as WGRS at higher coverage will be more costly than RNA-seq based experiments. Based on the merits of RNA-Seq, we are optimistic that this approach will be useful on the legumes with higher genome size (e.g., groundnut 2.5 Gb).

Trait Mapping for Epigenetic Factors

Epigenetic markers associated with heritable epi-alleles for a particular trait can be deployed in crop breeding i.e., epigenomics-assisted breeding (EAB) (Figure 1). Whole genome bisulfite sequencing (WGBS) and chromatin immunoprecipitation-sequencing (ChIP-Seq) are the two important approaches for mapping epigenetic factors. In case of WGBS, the NGS based technologies enable us to find genome-wide 5-methylcytosine, in rapid and precise manner. Methylation of DNA cytosine plays a significant role in many cellular processes, including expression of the gene. DNA methylations have been playing an important role in the understanding of the molecular mechanism of heterosis. NGS-based WGBS approach has been found promising for understanding the molecular mechanism of heterosis in *Arabidopsis*, rice, and maize (He et al., 2010, 2013; Shen et al., 2012). Additionally WGBS approach has been successfully used to understand the segregation pattern of methylation in F₂ generation (Schmitz et al., 2013). This approach is being utilized for understanding the molecular mechanism of heterosis in pigeonpea along with profiling of siRNA and transcriptome of parental lines and thereof derived hybrids.

In addition to WGBS, remodeling of chromatin through histone modification plays a significant role in the expression of many important genes. Specific histone modification has been found associated with expression of genes, plays a significant role in heterosis. In this context NGS-based, ChIP-seq approach was developed to identify the binding sites of DNA-associated proteins. This approach has been widely utilized in *Arabidopsis*, rice, and maize to find an association with heterosis (He et al., 2010, 2013; Shen et al., 2012). Along with WGBS, ChIP-seq approach is also being used in the case of pigeonpea hybrids and parental lines for understanding the molecular mechanism of heterosis.

TRANSCRIPTOMICS APPROACHES FOR GENE DISCOVERY AND MARKER DEVELOPMENT

Identification of genes and pathways that are responsible for tolerance to various abiotic and biotic stresses is crucial to enhance productivity of legumes (Figure 1). Transcriptome

sequencing is a better alternative approach to genome sequencing for targeted expressed gene sequencing. Furthermore, global gene expression analysis provides insights into gene function and molecular basis of various cellular components and transcriptional programs. Therefore, in legumes, to begin with, efforts have been focused on the development of cDNA libraries, generation of expressed sequence tags (EST), gene expression analysis, and the *in silico* mining of functional information from EST data sets even before genome sequences became available. The transcriptome sequencing has been applied for various other functional genomics approaches such as gene expression profiling, genome annotation and discovery of non-coding RNA, etc. (Morozova and Marra, 2008).

Extensive efforts have been made initially in legume transcriptomics and an abundance of ESTs from a range of tissues, including from plants challenged by different stresses have been generated in soybean (1.5 million ESTs, Vodkin et al., 2004), Medicago (*Medicago truncatula*; 280,000, Cheung et al., 2006) and Lotus (*Lotus japonicas*; 242,000, Asamizu et al., 2004). Initially, Sanger sequencing based ESTs were generated in majority of cases. For example, in the case of chickpea, cDNA libraries resulting in 20,162 ESTs have been generated from plants under drought and salinity stress conditions (Varshney et al., 2009). In the case of pigeonpea, responsive ESTs (9,888) for FW and SMD were generated (Raju et al., 2010). In addition, EST libraries have also been constructed using the suppression subtractive hybridization (SSH) technique, and utilizing this approach in chickpea, 477 drought-responsive ESTs were generated from root tissues (Buhariwalla et al., 2005). Further, Deokar et al. (2011) also generated 3,062 unigenes from SSH libraries of root and shoot tissues from drought tolerant- and sensitive- genotypes in chickpea. Similarly, in pigeonpea, 182 unique ESTs were generated from drought-stressed and unstressed pigeonpea seedlings using SSH (Qiao et al., 2012). However, major disadvantage of this method is that it is technically demanding and labor intensive.

In the recent years, several sequencing platforms available at low cost have already been demonstrated for use in generation of a huge set of transcript reads from a range of developing and stress-responsive tissues in different crop legumes. For instance, in chickpea, an improved transcriptome assembly has been generated based on FLX/454 sequencing together with Sanger ESTs comprised 103,215 Transcript Assembly Contigs (TACs) with an average contig length of 459 bp (Hiremath et al., 2011). In a different study, using the Illumina sequencing platform, another 53,409 contigs representing ~28 Mb of unique transcriptome sequence were assembled (Garg et al., 2011b). The same group, by using both FLX/454 and Illumina sequencing technologies, defined another set of 34,760 contigs representing ~4.8% (35.5 Mb) of the chickpea genome (Garg et al., 2011a). Furthermore, by analyzing sequencing data from three different platforms (Sanger, FLX/454, and Illumina), hybrid comprehensive assemblies have been generated in the case of pigeonpea (Kudapa et al., 2012) and chickpea (Kudapa et al., 2014). Several versions of transcriptome assemblies have been developed,

in fact, for many different legume crops (see Varshney et al., 2015a). Furthermore, the National Center for Genome Resources (NCGR) in cooperation with the U.S. Department of Agriculture (USDA)-supported Legume Information System² offers a comprehensive collection of transcriptome assemblies for several legumes.

Transcriptome assemblies generated using different sequencing technologies or in combination of two or more sequencing technologies provide valuable transcriptomic resources such as functional markers [EST-SSRs, SNPs, Intron Spanning regions (ISRs), etc.] for use in crop breeding programs. For example, a total of 1,682 and 4,099 SNPs were identified in soybean (Deschamps and Campbell, 2010) and common bean (Wu et al., 2014), respectively. In some studies, ISR markers (flanking intron junctions) were identified based on alignment to related genome sequences (Hiremath et al., 2011; Kudapa et al., 2012) and have found wide application in generating highly informative dense genetic maps with well distributed markers (Bordat et al., 2011). Furthermore, gene expression profiling data from the developed transcriptome assemblies would enable identification of candidate genes associated with different traits of interest including stress responsive genes. In common bean, a total of 441 salt responsive transcription factors (TF) were identified from a set 2,678 TFs classified under 59 TF families (Hiz et al., 2014) responsive. Information on the transcriptomic resources and candidate genes should provide insights into the molecular mechanisms of stress tolerance and ultimately help to develop improved stress tolerant legume varieties.

PROTEOMICS AND METABOLOMICS

Advances in ‘omics’ technologies provide opportunities through which new datasets can be produced for crop plants. This is particularly important in crop research to identify candidate genes and pathways involved in important agronomic traits, especially through the integration of genomics and functional ‘omics’ data together with genetic and phenotypic information (Langridge and Fleury, 2011) (**Figure 1**). The above developments will ensure higher integration of ‘omics’ information with crop breeding leading to evolution from GAB to omics-assisted breeding (OAB) in coming years.

Proteomics Approaches in Legumes

Proteomics involves research on cellular proteomes in which sets of protein species found in a biological unit such as cells/tissues/organs/organelle, at a particular developmental stage or external condition are analyzed (Jorin et al., 2006). The increased range of proteome coverage and improvements in quantitative measurements have been vital for studying proteome composition, modulation, and modifications for development stages and stress–response mechanisms in plant system. In crops research, proteomics pipelines are

increasingly being used, especially to study traits and stress response mechanisms, specific to crop systems. The proteomic information obtained, combined with the accurate identification of genetic determinants underlying a trait of interest together with improvements in the quality of genomic information, can be integrated into advanced breeding programs (Vanderschuren et al., 2013).

The important areas of plant proteomics include proteome mapping, comparison of protein profiles for different genotypes/biological units/stress factors (comparative proteomics), identification of posttranslational modifications (PTMs) and interaction networks through protein–protein interactions (PPI; Hu et al., 2015; Katam et al., 2015). Translational proteomics has great potential to be implemented in crop research for increasing agricultural production through the use of methodology and knowledge (Jorin-Novo et al., 2015). Information obtained through orthoproteomics (different species) and comparative proteomics (different genotypes) approaches will have importance in advanced breeding programs through coordination and standardization of proteomic approaches and high quality crop proteomics databases (Vanderschuren et al., 2013).

Advances in mass spectrometry (MS) applications in terms of speed, accuracy, sensitivity, and software tools have been instrumental for high-throughput protein quantification. Quantitative proteomics platforms have emerged due to advances in MS technologies for high-throughput protein quantifications and involve gel-based or gel-free, ‘shot-gun’ and ‘label-based’ (isotopic/isobaric) or ‘label-free’ approaches (Abdallah et al., 2012; Hu et al., 2015). Gel-based proteomics involving two dimensional gel-electrophoresis (2D-GE) and difference-GE (2D-DIGE) for protein separation followed by MS analyses, have underpinned the understanding of proteomic changes during growth and development as well as stress response mechanisms in plants. Although 2D-GE has been the workhorse for protein expression analysis due to its high resolving power, 2D-DIGE, which involves pre-electrophoretic labeling of samples with fluorescent dyes, and allows separation of proteins on the same gel, is quantitatively more accurate. (Schulze and Usadel, 2010; Vadivel and Kumaran, 2015). Shotgun proteomics, is a ‘bottom-up, peptide-centric’ strategy, mostly involve coupled liquid chromatography tandem MS (LC-MS/MS) platforms, which is efficient for high-throughput analyses of cell or organelle proteome by providing an overview of the major proteins (Jorin-Novo et al., 2015). Selected reaction monitoring (SRM) of peptides is a sensitive and specific targeted proteomic approach, to quantify the abundance of selected target proteins. It can also be an important method for biomarker validation of candidate proteins in dissecting molecular mechanisms underlying a particular trait for crops (Jacoby et al., 2013). PPI network analysis is an important aspect of proteomics by providing important information on the molecular mechanisms of signal transduction, protein complexes, stress responses as well as insight into developmental and physiological processes. The strategies involved include yeast-two hybrid (Y2H), affinity purification MS (AP-MS) and

²<http://www.comparative-legumes.org>

biomolecular fluorescence complementation (BiFC; Zhang et al., 2010).

Proteomics studies in legumes have been carried out mainly in the model systems, *Medicago* and *Lotus* (Watson et al., 2003; Larrainzar et al., 2007; Colditz, 2013; Lee et al., 2013; Dam et al., 2014; Ino et al., 2014) or soybean (Hossain et al., 2013; Hossain and Komatsu, 2014). Lesser but important proteomic information is available from the crop legumes such as protein differential expression to understand abiotic and biotic stress responses and proteome reference maps. Although there is limited information available currently, proteomics based datasets for the crop legumes will be enhanced with the increasing availability of legume genome sequences and transcriptome datasets (Dubey et al., 2011; Hiremath et al., 2011; Kalavacharla et al., 2011; Varshney et al., 2013c). Examples of proteome reference maps available in crop legumes include different subcellular membrane proteins from chickpea (Jaiswal et al., 2012), mature seed proteins and vegetative tissue proteins from pea (Schiltz et al., 2004; Bourgeois et al., 2009) and leaf proteins from groundnut (Katam et al., 2010).

In crop legumes, differential expression analyses and comparative proteomics approaches have provided insight to stress responses. Some examples include dehydration in chickpea (Bhushan et al., 2007; Pandey et al., 2008; Subba et al., 2013), early phases of cold stress in chickpea (Heidarvand and Maali-Amiri, 2013), drought stress in common bean (Zadraznik et al., 2013), fungal infection in pea (Barilli et al., 2012), drought stress in groundnut (Basha et al., 2007; Kottapalli et al., 2008), and salt stress in groundnut calli (Jain et al., 2006). For studying PTMs, in chickpea, a nucleus-specific phosphoproteome map was generated in developing seedlings (Kumar et al., 2014) and phosphoproteins were also analyzed in root tips of common bean subjected to osmotic stress (Yang et al., 2013).

Metabolomics Approaches

Metabolomics approaches involve the identification and quantification of low molecular weight metabolites in an organism, at a particular developmental stage in a specific organ/tissue/cell and have been successfully implemented to investigate the molecular phenotypes of plants in response to abiotic stress (Arbona et al., 2013). High-throughput screening of metabolites are advantageous compared to targeted reverse genetic approaches for plant metabolic engineering because the former method provides a more comprehensive understanding of metabolic networks in connection with developmental stages of phenotypes and capable of screening out unwanted traits (Ferne and Schauer, 2009).

The two main metabolomics profiling strategies using MS and nuclear magnetic resonance (NMR) have been described in the literature. A wider coverage of the great number of metabolites in plants has been obtained through the combination of several analytical techniques that usually consist of a separation techniques coupled to MS (for detection; Arbona et al., 2013). The regularly applied methods are gas-chromatography-MS (GC-MS), GC-time of flight-MS (GC-TOF-MS) and LC-MS are regular methods. GC-MS allows the identification and

quantification of large number of primary metabolites, GC-TOF-MS is noted for its fast scan times for better resolved peaks and sample throughput, while LC-MS allows a broader range of primary and secondary metabolites to be measured (Ferne and Schauer, 2009). Other techniques include flow injection analysis coupled to MS (FIA/MS) and Fourier Transform Infrared spectroscopy (FTIR; Arbona et al., 2013).

In legumes, most of the metabolic profiling analyses have been carried out in the model systems, *Medicago* and *Lotus*. A non-targeted, comparative metabolomic was used to study drought tolerance in *Lotus* genotypes in which conserved and unique metabolic responses were identified, with GC coupled to electron impact ionization (EI)-TOF-MS for metabolic profiling (Sanchez et al., 2012). A couple of other examples of metabolic profiling in *Lotus* include analysis of flavonoids (Suzuki et al., 2008) and long term salt stress response mechanisms (Sanchez et al., 2011). To study legume-rhizobia symbiosis in *Medicago*, which involves rapid metabolic changes in both partners, untargeted quantitative, LC-electrospray ionization (ESI)-TOF-MS was used for metabolic profiling followed by targeted LC-ESI-QTrap MS of selected candidates metabolites (Zhang et al., 2012). The study was useful in identifying that oxylipin pathway being involved in nodulation factor signaling in the early stages of symbiosis. In *Medicago*, some of the other recent examples include the study of new pathways and alternative mechanism for phenylpropanoid and isoflavonoid biosynthesis (Farg et al., 2008), development and symbiosis-dependent primary and secondary metabolism in roots (Schliemann et al., 2008), metabolites involved in symbiosis (Ye et al., 2013). Among the crop legumes, fewer metabolic profiling studies available and include chickpea-*Fusarium* interaction (Kumar et al., 2015), developing pea seed (Vigeolas et al., 2008), phosphorus stressed common bean (Hernandez et al., 2009).

The integration of metabolomics with transcriptomics datasets, high-throughput phenotyping and bioinformatics platforms, for profiling of large and genetically diverse populations, will enable the identification of novel metabolic QTLs and enhance the identification of candidate genes for trait of interest. Also metabolomics used as an additional tool with genomics-assisted selection strategies for crop-improvement, reduces the times spent for the discovery of new traits and allelic variations (Ferne and Schauer, 2009).

BREEDING INFORMATICS TOOLS

Breeding informatics tools are pre-requisites for effective and efficient application in each and every step in any molecular breeding program, starting from planning for field experiments to taking breeding decision for selection of plants in making the crosses. The GAB experiments involve identification of suitable germplasm from large scale diversity experiments to develop mapping population and identification of quantitative trait loci through genotyping and phenotyping using either family based mapping approach or germplasm based approach (association mapping). Identification of marker-trait associations through detailed analysis of genotypic and phenotypic data and finally

applying markers in molecular breeding programs depends on a sequential use of a number of decision support tools (Xu, 2010). In this context, Breeding Management System (BMS) of Integrated Breeding Platform (IBP)³ has been developed to help breeders to manage their day-to-day activities through all phases of their breeding programs. The BMS Workbench is useful from straightforward phenotyping to complex genotyping. It provides all the tools needed to conduct modern breeding in one comprehensive package including, project planning, germplasm management, germplasm evaluation, molecular analysis, data analysis, and breeding decision support (Varshney et al., 2015b).

MOLECULAR BREEDING FOR ACCELERATED LEGUME IMPROVEMENT

The conventional breeding approaches led to development of large number of improved cultivars over the time for legumes. However, these improved cultivars lag far behind in attempting to match the ever increasing human population and thus putting more pressure on the ongoing breeding programs across the world. Further, the limiting natural resources in terms of land and water in addition to climate change and uncertain rains have further made the ongoing breeding programs on the high pressure. Nevertheless, last decade has witnessed a speedy development in the area of NGS and high-throughput genotyping technologies. These developments led to development of trait-linked markers for several traits using first generation of markers (RFLP, RAPD, SSR) and mapping populations (F₂, RIL, NIL, BIL, DH and germplasm sets). There are three GAB approaches namely MABC, marker-assisted recurrent selection (MARS), and GS which are currently been deployed for developing improved lines (Figure 1).

The MABC has been the most preferred and result oriented molecular breeding approach for improving existing popular genotypes for one or two traits and pyramiding of few genes/QTLs. This approach has become very popular and widely accepted among breeders for these reasons; optimum time utilization and resources, opportunity to perform selection at early growth stages for multiple genes/loci, ability to break linkage between desirable and non-desirable traits, and most importantly no phenotyping requirement except for final trait confirmation. In the case of chickpea, MABC has been successfully used for introgressing a “QTL-hotspot” harboring several QTLs controlling several drought tolerance related root traits in the elite chickpea variety JG 11 (drought tolerant variety; Varshney et al., 2013a). Introgression lines have shown improved performance under rainfed as well irrigated environment as compared to recipient parent. In another attempt in chickpea, this approach was used for introgression of resistance to FW and ascochyta blight (AB) into the elite chickpea cultivar C 214 (Varshney et al., 2013b). Further, now the efforts are underway to pyramid the resistance to FW and AB by crossing introgression lines. Similarly in the case of groundnut, diagnostic markers were used to develop an improved cultivar called ‘NemaTAM,’ which

is the first RKN resistant groundnut variety and was released for cultivation in the USA (Simpson et al., 2003). In another effort, high oleate trait was improved in the nematode resistant cultivar “Tifguard” leading to development of second cultivar “Tifguard High O/L” possessing resistance to nematode and high oleate trait (Chu et al., 2011). Most recently, three elite and popular cultivars (TAG 24, JL 24, and ICGV 91114) were improved for rust resistance while three elite cultivars were improved for oil quality (Varshney et al., 2014a; Janila et al., 2016). The above achievements have demonstrated significance of molecular breeding and many research organizations across the world are deploying molecular breeding in their crop improvement programs. Further availability of linked markers in future from trait mapping pipeline will provide more options to the breeders in accumulating favorable alleles for multiple traits in a single genetic background using MABC approach.

Another molecular breeding approach termed MARS promises improve to accumulate superior alleles for quantitative traits such as drought resistance, yield, etc., which are controlled by several QTLs, each with a small effect on the phenotype and the ideal genotype cannot be attained through MABC (Varshney et al., 2013c). In such situations, MARS is quite useful to target more number of minor as well as major QTLs (Ribaut and Ragot, 2007). MARS approach requires estimation of marker effects for traits of interest and then favorable alleles linked to these traits are traced in two or three recombination cycles to combine favorable alleles (Bernardo and Charcosset, 2006). Multiallelic MARS which involves creating a series of bi-allelic lines and finally assembling them in the final inbred line has also been suggested for traits where the number of target loci is small (Ribaut and Ragot, 2007). In the case of chickpea, in order to pyramid favorable alleles for drought tolerance, two MARS crosses were initiated (JG 11 × ICCV 04112 and JG 130 × ICCV 05107). Based on QTL analysis on genotyping data on F₃ lines and phenotyping data on F₅ lines, superior lines were selected using OptiMAS ver. 1.0 (Valente et al., 2013). Although, MARS have been proven successful in private breeding programs for enhancing genetic gain, it could not achieve similar level of success in chickpea.

Another emerging molecular breeding approach called GS holds great promise in achieving higher genetic gains in lesser time for complex traits. Unlike MABC and MARS, the GS approach does not require population development and identification of linked markers prior to its deployment and has potential to enhance genetic gains for even complex traits such as yield under drought stress. This approach deploys evenly distributed genetic markers across the genome to predict GEBV using multiple methods with varying degrees of complexity, computational efficiency, and predictive accuracy (Meuwissen et al., 2001). For achieving higher genetic gains, GS showed its potential in accumulating 1000s of favorable alleles leading to development of resilient crop varieties with high yield potential under unfavorable conditions. Most importantly, GS can reduce breeding cost by 22.4% and breeding duration which is a remarkable achievement (König et al., 2009). ICRISAT has taken initiatives by constituting TPs in chickpea and groundnut with 320 and 314 elite breeding lines, respectively. These populations

³<https://www.integratedbreeding.net/>

have been genotyped with DArT arrays (15,360 features) and are also being used currently to generate multi-season and multi-location phenotyping data on important traits. In parallel, the effort is also underway to genotype these populations with 60 K crop specific SNP array. The high-throughput genotyping and phenotyping data on these sets will be used for development and train appropriate GS models to initiate GS breeding in chickpea and groundnut. Similarly in case of pigeonpea wherein the emphasis is more on development of hybrids, a set of cytoplasmic male sterility (CMS) lines and restorer lines have been included in the TP. In this context, a total of 550 test cross F₁s have been developed through hybridization using 55 restorer lines and 10 CMS lines for phenotyping at multiple locations. Based on the genotypic datasets of parental lines and phenotypic datasets of test cross F₁s, the GS models will be defined for use in GS breeding to develop improved parental lines and new hybrids in pigeonpea.

CONCLUSION

The demand-supply gap for the legumes is perpetually increasing widening day-by-day which will lead to a huge shortfall in the supply to the ever increasing global population in coming years. The only option is to maximize the efforts toward developing improved high yielding cultivars possessing resistance/tolerance to the major stresses especially in context of climate change. The cost-effective sequencing technologies have introduced a new era in genomics and breeding by pinpointing the genes responsible for distinct phenotypes leading to selection of plants based on genotyping information. The knowledge generated through all the “omics” studies need to be integrated in breeding so that breeders can move toward “knowledge-based breeding” from “chance breeding.” In recent years, several emerging genomic technologies have been developed to foster trait mapping, gene discovery, and trait improvement programs in several crop plants. Some of these technologies have been already deployed in the three legume crops as

discussed in this review. In addition, important information generated through transcriptomics, proteomics, metabolomics, and epigenomics have greatly benefitted the scientific community in developing better understanding of the traits and crops leading to development of effective strategy for achieving higher genetic gains in less time. It is obvious that the current and the upcoming technologies will further assist legume improvement programs in a more cost-effective, user-friendly, and less time consuming manner.

AUTHOR CONTRIBUTIONS

MP planned the MS content, coordinated with other coauthors, contributed in drafting special sections, brought flow between different sections, and finalizing the manuscript. MR, VS, AbiR, HK, MT, AC, and AbhR contributed specific sections of this manuscript. MP, MR, and VS worked on this MS in several rounds to bring in proper shape. RV was involved in planning the content and finalizing the manuscript.

ACKNOWLEDGMENTS

The authors are thankful to CGIAR's Generation Challenge Programme, Bill & Melinda Gates Foundation (Tropical Legumes I, II, and III), The Peanut Foundation, United States Agency for International Development (USAID), Biotechnology Industry Partnership Programme (BIPP), MARS Chocolate Inc., Australia-India Strategic Research Fund (AISRF), Indo-German Science & Technology Centre (IGSTC), Indian Council of Agricultural Research (ICAR), Department of Biotechnology (DBT), of Government of India, and World Bank Assisted Watershed Development Project II (KWDP-II) by Government of Karnataka, India. The work reported in this article was undertaken as a part of the CGIAR Research Program on Grain Legumes. ICRISAT is a member of the CGIAR.

REFERENCES

- Abdallah, C., Dumas-Gaudot, E., Renault, J., and Sergeant, K. (2012). Gel-based and gel-free quantitative proteomics approaches at a glance. *Int. J. Plant Genomics* 2012:17. doi: 10.1155/2012/494572
- Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., et al. (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat. Biotechnol.* 30, 174–178. doi: 10.1038/nbt.2095
- Arbona, V., Manzi, M., de Ollas, C., and Gómez-Cadenas, A. (2013). Metabolomics as a tool to investigate abiotic stress tolerance in plants. *Int. J. Mol. Sci.* 14, 4885–4911. doi: 10.3390/ijms14034885
- Asamizu, E., Nakamura, Y., Sato, S., and Tabata, S. (2004). Characteristics of the *Lotus japonicus* gene repertoire deduced from large-scale expressed sequence tag (EST) analysis. *Plant Mol. Biol.* 54, 405–414. doi: 10.1023/B:PLAN.0000036372.46942.b8
- Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., et al. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3:e3376. doi: 10.1371/journal.pone.0003376
- Barilli, E., Rubiales, D., and Castillejo, M. Á. (2012). Comparative proteomic analysis of BTH and BABA-induced resistance in pea (*Pisum sativum*) toward infection with pea rust (*Uromyces pisi*). *J. Proteomics* 75, 5189–5205. doi: 10.1016/j.jprot.2012.06.033
- Basha, S. M., Katam, R., and Naik, K. S. S. (2007). Differential response of peanut genotypes to water stress. *Peanut Sci.* 34, 96–104. doi: 10.1111/j.1365-3040.2009.01933.x
- Bayer, P. E., Ruperao, P., Mason, A. S., Stiller, J., Chan, C. K., Hayashi, S., et al. (2015). High-resolution skim genotyping by sequencing reveals the distribution of crossovers and gene conversions in *Cicer arietinum* and *Brassica napus*. *Theor. Appl. Genet.* 128, 1039–1048. doi: 10.1007/s00122-015-2488-y
- Bernardo, R., and Charcosset, A. (2006). Usefulness of gene information in marker-assisted recurrent selection: a simulation appraisal. *Crop Sci.* 46, 614–621. doi: 10.2135/cropsci2005.05-0088
- Bertioli, D. J., Cannon, S. B., Froenicke, L., Huang, G., Farmer, A. D., Cannon, E. K. S., et al. (2016). The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nat. Genet.* 48, 438–446. doi: 10.1038/ng.3517

- Bhushan, D., Pandey, A., Choudhary, M. K., Datta, A., Chakraborty, S., and Chakraborty, N. (2007). Comparative proteomics analysis of differentially expressed proteins in chickpea extracellular matrix during dehydration stress. *Mol. Cell. Proteomics* 6, 1868–1884. doi: 10.1074/mcp.M700015-MCP200
- Bohra, A., Pandey, M. K., Jha, U. C., Singh, B., Singh, I. P., Datta, D., et al. (2014). Genomics-assisted breeding in the major pulse crops of developing countries: present status and prospects. *Theor. Appl. Genet.* 127, 1263–1291. doi: 10.1007/s00122-014-2301-3
- Bordat, A., Savoies, V., Nicolas, M., Salse, J., Chauveau, A., Bourgeois, M., et al. (2011). Translational genomics in legumes allowed placing in silico 5460 unigenes on the pea functional map and identified candidate genes in *Pisum sativum* L. *G3* 2, 93–103. doi: 10.1534/g3.111.000349
- Bourgeois, M., Jacquin, F., Savoies, V., Sommerer, N., Labas, V., Henry, C., et al. (2009). Dissecting the proteome of pea mature seeds reveals the phenotypic plasticity of seed protein composition. *Proteomics* 9, 254–271. doi: 10.1002/pmic.200700903
- Buhariwalla, H. K., Jayashree, B., Eshwar, K., and Crouch, J. H. (2005). Development of ESTs from chickpea roots and their use in diversity analysis of the *Cicer* genus. *BMC Plant Biol.* 5:16. doi: 10.1186/1471-2229-5-16
- Bybee, S. M., Bracken-Grissom, H., Haynes, B. D., Hermansen, R. A., Byers, R. L., Clement, M. J., et al. (2011). Targeted amplicon sequencing (TAS): a scalable next-gen approach to multi-locus, multitaxa phylogenetics. *Genome Biol. Evol.* 3, 1312–1323. doi: 10.1093/gbe/evr106
- Cavanagh, C., Morell, M., Mackay, L., and Powell, W. (2008). From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Curr. Opin. Plant Biol.* 11, 215–221. doi: 10.1016/j.pbi.2008.01.002
- Cheung, F., Haas, B. J., Goldberg, S. M. D., May, G. D., Xiao, Y., and Town, C. D. (2006). Sequencing *Medicago truncatula* expressed sequenced tags using 454 life sciences technology. *BMC Genomics* 7:272. doi: 10.1186/1471-2164-7-272
- Chu, Y., Wu, C. L., Holbrook, C. C., Tillman, B. L., Person, G., and Ozias-Akins, P. (2011). Marker-assisted selection to pyramid nematode resistance and the high oleic trait in peanut. *Plant Genome* 4, 110–117. doi: 10.3835/plantgenome2011.01.0001
- Colditz, F. (2013). “*Medicago truncatula* root proteomics,” in *Molecular Microbial Ecology of the Rhizosphere*, Vol. 1 and 2, eds F. J. de Bruijn and F. Colditz (Toulouse: Wiley-Blackwell Publisher), 271–278.
- Crossa, J., Beyene, Y., Kassa, S., Pérez, P., Hickey, J. M., Chen, C., et al. (2013). Genomic prediction in maize breeding populations with genotyping-by-sequencing. *G3 (Bethesda)* 6, 1903–1926. doi: 10.1534/g3.113.008227
- Dam, S., Dyrland, T. F., Ussatjuk, A., Jochimsen, B., Nielsen, K., Goffard, N., et al. (2014). Proteome reference maps of the *Lotus japonicus* nodule and root. *Proteomics* 14, 230–240. doi: 10.1002/pmic.201300353
- Darvasi, A., and Soller, M. (1995). Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics* 141, 1199–1207.
- Das, S., Upadhyaya, H. D., Bajaj, D., Kujur, A., Badoni, S., Kumar, L. V., et al. (2015). Deploying QTL-seq for rapid delineation of a potential candidate gene underlying major trait-associated QTL in chickpea. *DNA Res.* 22, 193–203. doi: 10.1093/dnares/dsv004
- Deokar, A. A., Kondawar, V., Jain, P. K., Karuppaiyil, S. M., Raju, N. L., Vadez, V., et al. (2011). Comparative analysis of expressed sequence tags (ESTs) between drought-tolerant and-susceptible genotypes of chickpea under terminal drought stress. *BMC Plant Biol.* 11:70. doi: 10.1186/1471-2229-11-70
- Deokar, A. A., Ramsay, L., Sharpe, A. G., Diapari, M., Sindhu, A., Bett, K., et al. (2014). Genome wide SNP identification in chickpea for use in development of a high density genetic map and improvement of chickpea reference genome assembly. *BMC Genomics* 15:708. doi: 10.1186/1471-2164-15-708
- Deschamps, S., and Campbell, M. A. (2010). Utilization of next-generation sequencing platforms in plant genomics and genetic variant discovery. *Mol. Breed.* 25, 553–570. doi: 10.1007/s11032-009-9357-9
- Dubey, A., Farmer, A., Schlueter, J., Cannon, S. B., Abernathy, B., Tuteja, R., et al. (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. doi: 10.1093/dnares/dsr007
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6:e19379. doi: 10.1371/journal.pone.0019379
- Farag, M. A., Huhman, D. V., Dixon, R. A., and Sumner, L. W. (2008). Metabolomics reveals novel pathways and differential mechanistic and elicitor-specific responses in phenylpropanoid and isoflavonoid biosynthesis in *Medicago truncatula* cell cultures. *Plant Physiol.* 146, 387–402. doi: 10.1104/pp.107.108431
- Fedoroff, N. V. (2015). Food in a future of 10 billion. *Agric. Food Secur.* 4:11. doi: 10.1186/s40066-015-0031-7
- Fekih, R., Takagi, H., Tamiru, M., Abe, A., Natsume, S., Yaegashi, H., et al. (2013). MutMap+: genetic mapping and mutant identification without crossing in rice. *PLoS ONE* 8:e68529. doi: 10.1371/journal.pone.0068529
- Fernie, A. R., and Schauer, N. (2009). Metabolomics-assisted breeding: a viable option for crop improvement? *Trends Genet.* 25, 39–48. doi: 10.1016/j.tig.2008.10.010
- Fu, Y. B., Cheng, B., and Peterson, G. W. (2014). Genetic diversity analysis of yellow mustard (*Sinapis alba* L.) germplasm based on genotyping by sequencing. *Genet. Resour. Crop Evol.* 61, 579–594. doi: 10.1007/s10722-013-0058-1
- Garg, R., Patel, R. K., Jhanwar, S., Priya, P., Bhattacharjee, A., Yadav, G., et al. (2011a). Gene discovery and tissue-specific transcriptome analysis in chickpea with massively parallel pyrosequencing and web resource development. *Plant Physiol.* 156, 1661–1678. doi: 10.1104/pp.111.178616
- Garg, R., Patel, R. K., Tyagi, A. K., and Jain, M. (2011b). De novo assembly of chickpea transcriptome using short reads for gene discovery and marker identification. *DNA Res.* 18, 53–63. doi: 10.1093/dnares/dsq028
- Gerland, P., Raftery, A. E., Ševčíková, H., Li, N., Gu, D., Spoorenberg, T., et al. (2014). World population stabilization unlikely this century. *Science* 346, 234–237. doi: 10.1126/science.1257469
- Goff, S. A., Ricke, D., Lan, T. H., Presting, G., Wang, R., Dunn, M., et al. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). *Science* 296, 92–100. doi: 10.1126/science.1068275
- Golicz, A. A., Bayer, P. E., and Edwards, D. (2015). Skim-based genotyping by sequencing. *Methods Mol. Biol.* 1245, 257–270. doi: 10.1007/978-1-4939-1966-6_19
- Gowda, C. L. L., Upadhyaya, H. D., Sharma, S., Varshney, R. K., and Dwivedi, S. L. (2013). Exploiting genomic resources for efficient conservation and use of chickpea, groundnut, and pigeonpea collections for crop improvement. *Plant Genome* 6, 1–28. doi: 10.3835/plantgenome2013.05.0016
- Guo, B., Sleper, D. A., and Beavis, W. D. (2010). Nested association mapping for identification of functional markers. *Genetics* 186, 373–383. doi: 10.1534/genetics.110.115782
- Gupta, P. K., Kulwal, P. L., and Jaiswal, V. (2014). Association mapping in crop plants: opportunities and challenges. *Adv. Genet.* 85, 109–147. doi: 10.1016/B978-0-12-800271-1.00002-0
- He, G., Chen, B., Wang, X., Li, X., Li, J., He, H., et al. (2013). Conservation and divergence of transcriptomic and epigenomic variation in maize hybrids. *Genome Biol.* 14:R57. doi: 10.1186/gb-2013-14-6-r57
- He, G., Zhu, X., Elling, A. A., Chen, L., Wang, X., Guo, L., et al. (2010). Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids. *Plant Cell* 22, 17–33. doi: 10.1105/tpc.109.072041
- He, J., Zhao, X., Laroche, A., Lu, Z.-X., Liu, H., and Li, Z. (2014). Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Front. Plant Sci.* 5:484. doi: 10.3389/fpls.2014.00484
- Heidarvand, L., and Maali-Amiri, R. (2013). Physio-biochemical and proteome analysis of chickpea in early phases of cold stress. *J. Plant Physiol.* 170, 459–469. doi: 10.1016/j.jplph.2012.11.021
- Hernandez, G., Valdés-López, O., Ramírez, M., Goffard, N., Weiller, G., Aparicio-Fabre, R., et al. (2009). Global changes in the transcript and metabolic profiles during symbiotic nitrogen fixation in phosphorus-stressed common bean plants. *Plant Physiol.* 151, 1221–1238. doi: 10.1104/pp.109.143842
- Heschamps, S., Llaca, V., and May, G. D. (2012). Genotyping-by-sequencing in plants. *Biology* 1, 460–483. doi: 10.3390/biology1030460
- Hiremath, P. J., Farmer, A., Cannon, S. B., Woodward, J., Kudapa, H., Tuteja, R., et al. (2011). Large-scale transcriptome analysis in chickpea (*Cicer arietinum* L.), an orphan legume crop of the semi-arid tropics of Asia and Africa. *Plant Biotechnol. J.* 9, 922–931. doi: 10.1111/j.1467-7652.2011.00625.x
- Hiz, M. C., Canher, B., Niron, H., and Turet, M. (2014). Transcriptome analysis of salt tolerant common bean (*Phaseolus vulgaris* L.) under saline conditions. *PLoS ONE* 9:e92598. doi: 10.1371/journal.pone.0092598

- Hossain, Z., and Komatsu, S. (2014). "Soybean proteomics," in *Plant Proteomics Methods Mol. Biol.*, 1072, eds J. V. Jorrin-Novo, S. Komatsu, W. Weckwerth, and S. Wienkoop (New York, NY: Springer Science+Business Media), 315–331.
- Hossain, Z., Khatoon, A., and Komatsu, S. (2013). Soybean proteomics for unravelling abiotic stress response mechanism. *J. Proteome Res.* 12, 4670–4684. doi: 10.1021/pr400604b
- Hu, J., Rampitsch, C., and Bykova, N. V. (2015). Advances in plant proteomics toward improvement of crop productivity and stress resistance. *Front. Plant Sci.* 6:209. doi: 10.3389/fpls.2015.00209
- Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A., et al. (2009). High-throughput genotyping by whole-genome resequencing. *Genome Res.* 19, 1068–1076. doi: 10.1101/gr.089516.108
- Huang, X., Lu, T., and Han, B. (2013). Resequencing rice genomes: an emerging new era of rice genomics. *Trends Genet.* 29, 225–232. doi: 10.1016/j.tig.2012.12.001
- Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., et al. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* 42, 961–967. doi: 10.1038/ng.695
- Huang, Y.-F., Madur, D., Combes, V., Ky, C. L., Coubriche, D., Jamin, P., et al. (2010). The genetic architecture of grain yield and related traits in *Zea mays* L. revealed by comparing intermated and conventional populations. *Genetics* 186, 395–404. doi: 10.1534/genetics.110.113878
- Huang, Y.-F., Poland, J. A., Wight, C. P., Jackson, E. W., and Tinker, N. A. (2014). Using genotyping-by-sequencing (GBS) for genomic discovery in cultivated oat. *PLoS ONE* 9:e102448. doi: 10.1371/journal.pone.0102448
- Illa-Berenguer, E., Van Houten, J., Huang, Z., and van der Knaap, E. (2015). Rapid and reliable identification of tomato fruit weight and locule number loci by QTL-seq. *Theor. Appl. Genet.* 128, 1329–1342. doi: 10.1007/s00122-015-2509-x
- Ino, Y., Ishikawa, A., Nomura, A., Kajiwara, H., Harada, K., and Hirano, H. (2014). Phosphoproteome analysis of *Lotus japonicus* seeds. *Proteomics* 14, 116–120. doi: 10.1002/pmic.201300237
- International Rice Genome Sequencing Project [IRGSP] (2005). The map-based sequence of the rice genome. *Nature* 436, 793–800.
- Jacoby, R. P., Millar, A. H., and Taylor, N. L. (2013). Application of selected reaction monitoring mass spectrometry to field-grown crop plants to allow dissection of the molecular mechanisms of abiotic stress tolerance. *Front. Plant Sci.* 4:20. doi: 10.3389/fpls.2013.00020
- Jaganathan, D., Thudi, M., Kale, S., Azam, S., Roorkiwal, M., Gaur, P. M., et al. (2015). Genotyping-by-sequencing based intra-specific genetic map refines a "QTL-hotspot" region for drought tolerance in chickpea. *Mol. Genet. Genomics* 290, 559–571. doi: 10.1007/s00438-014-0932-3
- Jain, M., Misra, G., Patel, R. K., Priya, P., Jhanwar, S., Khan, A. W., et al. (2013). A draft genome sequence of the pulse crop chickpea (*Cicer arietinum* L.). *Plant J.* 74, 715–729. doi: 10.1111/tpj.12173
- Jain, S., Srivastava, S., Sarin, N. B., and Kav, N. N. (2006). Proteomics reveals elevated levels of PR 10 proteins in saline-tolerant peanut (*Arachis hypogaea*) calli. *Plant Physiol. Biochem.* 44, 253–259. doi: 10.1016/j.plaphy.2006.04.006
- Jaiswal, D. K., Ray, D., Subba, P., Mishra, P., Gayali, S., Datta, A., et al. (2012). Proteomic analysis reveals the diversity and complexity of membrane proteins in chickpea (*Cicer arietinum* L.). *Proteome Sci.* 10:59. doi: 10.1186/1477-5956-10-59
- Janila, P., Pandey, M. K., Shasidhar, Y., Variath, M. T., Sriswathi, M., Khera, P., et al. (2016). Molecular breeding for introgression of fatty acid desaturase mutant alleles (ahFAD2A and ahFAD2B) enhances oil quality in high and low oil containing peanut genotypes. *Plant Sci.* 242, 203–213. doi: 10.1016/j.plantsci.2015.08.013
- Jorrin, J. V., Rubiales, D., Dumas-Gaudot, E., Recorbet, G., Maldonado, A., Castillejo, M. A., et al. (2006). Proteomics: a promising approach to study biotic interaction in legumes, A review. *Euphytica* 147, 37–47. doi: 10.1007/s10681-006-3061-1
- Jorrin-Novo, J. V., Pascual, J., Sánchez-Lucas, R., Romero-Rodríguez, M. C., Rodríguez-Ortega, M. J., Lenz, C., et al. (2015). Fourteen years of plant proteomics reflected in Proteomics: moving from model species and 2DE-based approaches to orphan species and gel-free platforms. *Proteomics* 15, 1089–1112. doi: 10.1002/pmic.201400349
- Kafatos, A., and Hatzis, C. (2008). *Clinical Nutrition for Medical Students*. Crete: University of Crete.
- Kalavacharla, V., Liu, Z., Meyers, B. C., Thimmapuram, J., and Melmaiee, K. (2011). Identification and analysis of common bean (*Phaseolus vulgaris* L.) transcriptomes by massively parallel pyrosequencing. *BMC Plant Biol.* 11:135. doi: 10.1186/1471-2229-11-135
- Kale, S. M., Jaganathan, J., Ruperao, P., Chen, C., Punna, R., Kudapa, H., et al. (2015). Prioritization of candidate genes in QTL-hotspot region for drought tolerance in chickpea (*Cicer arietinum* L.). *Sci. Rep.* 5:15296. doi: 10.1038/srep15296
- Katam, K., Jones, K. A., and Sakata, K. (2015). Advances in proteomics and bioinformatics in agriculture research and crop improvement. *J. Proteomics Bioinform.* 8, 39–48. doi: 10.1016/j.jprot.2013.05.036
- Katam, R., Basha, S. M., Suravajhala, P., and Pechan, T. (2010). Analysis of peanut leaf proteome. *J. Proteome Res.* 9, 2236–2254. doi: 10.1021/pr901009n
- Kiezun, A., Garimella, K., Do, R., Stitzel, N. O., Neale, B. M., McLaren, P. J., et al. (2012). Exome sequencing and the genetic basis of complex traits. *Nat. Genet.* 44, 623–630. doi: 10.1038/ng.2303
- Kim, C., Guo, H., Kong, W., Chandrani, R., Shuang, L.-S., and Paterson, A. H. (2016). Application of genotyping by sequencing technology to a variety of crop breeding programs. *Plant Sci.* 242, 14–22. doi: 10.1016/j.plantsci.2015.04.016
- König, S., Simianer, H., and William, A. (2009). Economic evaluation of genomic breeding programs. *J. Dairy Sci.* 92, 382–391. doi: 10.3168/jds.2008-1310
- Kottapalli, K. R., Payton, P., Rakwal, R., Agrawal, G. K., Shibato, J., Burow, M., et al. (2008). Proteomics analysis of mature seed of four peanut cultivars using two-dimensional gel electrophoresis reveals distinct differential expression of storage, anti-nutritional, and allergenic proteins. *Plant Sci.* 175, 321–329. doi: 10.1016/j.plantsci.2008.05.005
- Kudapa, H., Bharti, A. K., Cannon, S. B., Farmer, A. D., Mulaosmanovic, B., Kramer, R., et al. (2012). A comprehensive transcriptome assembly of pigeonpea (*Cajanus cajan* L.) using Sanger and second-generation sequencing platforms. *Mol. Plant* 5, 1020–1028. doi: 10.1093/mp/ssr111
- Kudapa, H., Azam, S., Sharpe, A. G., Taran, B., Li, R., Deonovic, B., et al. (2014). Comprehensive transcriptome assembly of chickpea (*Cicer arietinum*) using sanger and next generation sequencing platforms: development and applications. *PLoS ONE* 9:e86039. doi: 10.1371/journal.pone.0086039
- Kujur, A., Bajaj, D., Upadhyaya, H. D., Das, S., Ranjan, R., Shree, T., et al. (2015a). A genome-wide SNP scan accelerates trait-regulatory genomic loci identification in chickpea. *Nat. Sci. Rep.* 5:11166. doi: 10.1038/srep11166
- Kujur, A., Upadhyaya, H. D., Shree, T., Bajaj, D., Das, S., Saxena, M., et al. (2015b). Ultra-high density intra-specific genetic linkage maps accelerate identification of functionally relevant molecular tags governing important agronomic traits in chickpea. *Nat. Sci. Rep.* 5:9468. doi: 10.1038/srep09468
- Kumar, R., Kumar, A., Subba, P., Gayali, S., Barua, P., Chakraborty, S., et al. (2014). Nuclear phosphoproteome of developing chickpea seedlings (*Cicer arietinum* L.) and protein-kinase interaction network. *J. Proteomics* 105, 58–73. doi: 10.1016/j.jprot.2014.04.002
- Kumar, Y., Dholakia, B. B., Panigrahi, P., Kadoo, N. Y., Giri, A. P., and Gupta, V. S. (2015). Metabolic profiling of chickpea-*Fusarium* interaction identifies differential modulation of disease resistance pathways. *Phytochemistry* 116, 120–129. doi: 10.1016/j.phytochem.2015.04.001
- Langridge, P., and Fleury, D. (2011). Making the most of 'omics' for crop breeding. *Trends Biotechnol.* 29, 33–40. doi: 10.1016/j.tibtech.2010.09.006
- Larrainzar, E., Wienkoop, S., Weckwerth, W., Ladrera, R., Arrese-Igor, C., and González, E. M. (2007). *Medicago truncatula* root nodule proteome analysis reveals differential plant and bacteroid responses to drought stress. *Plant Physiol.* 144, 1495–1507. doi: 10.1104/pp.107.101618
- Lee, J., Lei, Z., Watson, B. S., and Sumner, L. W. (2013). Sub-cellular proteomics of *Medicago truncatula*. *Front. Plant Sci.* 4:112. doi: 10.3389/fpls.2013.00112
- Li, H., Peng, Z., Yang, X., Wang, W., Fu, J., Wang, J., et al. (2013). Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nat. Genet.* 45, 43–50. doi: 10.1038/ng.2484
- Li, H., Vikram, P., Singh, R. P., Kilian, A., Carling, J., Song, J., et al. (2015). A high density GBS map of bread wheat and its application for dissecting complex disease resistance traits. *BMC Genomics* 16:216. doi: 10.1186/s12864-015-1424-5
- Liu, S., Yeh, C. T., Tang, H. M., Nettleton, D., and Schnable, P. S. (2012). Gene mapping via bulked segregant RNA-Seq (BSR-Seq). *PLoS ONE* 7:e36406. doi: 10.1371/journal.pone.0036406

- Lorenc, M. T., Hayashi, S., Stiller, J., Lee, H., Manoli, S., Ruperao, P., et al. (2012). Discovery of single nucleotide polymorphisms in complex genomes using SGAutoSNP. *Biology* 1, 370–382. doi: 10.3390/biology1020370
- Lu, H., Lin, T., Klein, J., Wang, S., Qi, J., Zhou, Q., et al. (2014). QTL-seq identifies an early flowering QTL located near flowering locus T in cucumber. *Theor. Appl. Genet.* 127, 1491–1499. doi: 10.1007/s00122-014-2313-z
- Mamanova, L., Coffey, A. J., Scott, C. E., Kozarewa, I., Turner, E. H., Kumar, A., et al. (2011). Target-enrichment strategies for next generation sequencing. *Nat. Methods* 7, 111–118. doi: 10.1038/nmeth.1419
- McCouch, S., Baute, G. J., Bradeen, J., Bramel, P., Bretting, P. K., Buckler, E., et al. (2013). Agriculture: feeding the future. *Nature* 499, 23–24. doi: 10.1038/499023a
- Meuwissen, T. H., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.
- Michael, T. P., and Jackson, S. (2013). The first 50 plant genomes. *Plant Genome* 6. doi: 10.3835/plantgenome2013.03.0001in
- Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A., and Johnson, E. A. (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res.* 17, 240–248. doi: 10.1101/gr.5681207
- Morozova, O., and Marra, M. A. (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics* 92, 255–264. doi: 10.1016/j.ygeno.2008.07.001
- Naj, A. C., Jun, G., Beecham, G. W., Wang, L.-S., Vardarajan, B. N., Buros, J., et al. (2011). Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat. Genet.* 43, 436–441. doi: 10.1038/ng.801
- Pandey, A., Chakraborty, S., Datta, A., and Chakraborty, N. (2008). Proteomics approach to identify dehydration responsive nuclear proteins from chickpea (*Cicer arietinum* L.). *Mol. Cell. Proteomics* 7, 88–107. doi: 10.1074/mcp.M700314-MCP200
- Parween, S., Nawaz, K., Roy, R., Pole, A. K., Venkata Suresh, B., Misra, G., et al. (2015). An advanced draft genome assembly of a desi type chickpea (*Cicer arietinum* L.). *Sci. Rep.* 5:12806. doi: 10.1038/srep12806
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., and Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE* 7:e37135. doi: 10.1371/journal.pone.0037135
- Poland, J. A., Brown, P. J., Sorrells, M. E., and Jannink, J.-L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* 7:e32253. doi: 10.1371/journal.pone.0032253
- Poland, J. A., and Rife, T. W. (2012). Genotyping-by-sequencing for plant breeding and genetics. *Plant Genome* 5, 92–102. doi: 10.3835/plantgenome2012.05.0005
- Qiao, G., Wen, X., Yu, L., and Ji, X. (2012). Identification of differentially expressed genes preferably related to drought response in pigeonpea (*Cajanus cajan*) inoculated by arbuscular mycorrhizae fungi (AMF). *Acta Physiol. Plant.* 34, 1711–1721. doi: 10.1007/s11738-012-0966-2
- Raju, N. L., Gnanesh, B. N., Lekha, P., Jayashree, B., Pande, S., Hiremath, P. J., et al. (2010). The first set of EST resource for gene discovery and marker development in pigeonpea (*Cajanus cajan* L.). *BMC Plant Biol.* 10:45. doi: 10.1186/1471-2229-10-45
- Ray, D. K., Mueller, N. D., West, P. C., and Foley, J. A. (2013). Yield trends are insufficient to double global crop production by 2050. *PLoS ONE* 8:e66428. doi: 10.1371/journal.pone.0066428
- Ribaut, J. M., and Ragot, M. (2007). Marker-assisted selection to improve drought adaptation in maize: the backcross approach, perspectives, limitations, and alternatives. *J. Exp. Bot.* 58, 351–360. doi: 10.1093/jxb/erl214
- Romay, M. C., Millard, M. J., Glaubitz, J. C., Peiffer, J. A., Swarts, K. L., Casstevens, T. M., et al. (2013). Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biol.* 14, R55. doi: 10.1186/gb-2013-14-6-r55
- Ruperao, P., Chan, C. K., Azam, S., Karafátová, M., Hayashi, S., Cížková, J., et al. (2014). A chromosomal genomics approach to assess and validate the desi and kabuli draft chickpea genome assemblies. *Plant Biotechnol. J.* 12, 778–786. doi: 10.1111/pbi.12182
- Salunkhe, D. K., Kadam, S. S., and Chavan, J. K. (1985). *Post-Harvest Biotechnology of Food Legumes*. Boca Raton, FL: CRC press, 160.
- Sanchez, D. H., Pieckenstein, F. L., Escaray, F., Erban, A., Kraemer, U. T. E., Udvardi, M. K., et al. (2011). Comparative ionomics and metabolomics in extremophile and glycophytic *Lotus* species under salt stress challenge the metabolic pre-adaptation hypothesis. *Plant Cell Environ.* 34, 605–617. doi: 10.1111/j.1365-3040.2010.02266.x
- Sanchez, D. H., Schwabe, F., Erban, A., Udvardi, M. K., and Kopka, J. (2012). Comparative metabolomics of drought acclimation in model and forage legumes. *Plant Cell Environ.* 35, 136–149. doi: 10.1111/j.1365-3040.2011.02423.x
- Saxena, R. K., Upadhyaya, H. D., Yu, O., Kim, C., Rathore, A., Seon, D., et al. (2015). “Comprehensive genome sequencing analysis of 292 *Cajanus* accessions,” in *Proceedings of the Plant and Animal Genome XXI Conference Jan, 10-14, 2015, San Diego, CA*.
- Schatz, M. C., Witkowski, J., and McCombie, W. R. (2002). Current challenges in de novo plant genome sequencing and assembly. *Genome Biol.* 13:243.
- Schiltz, S., Gallardo, K., Huart, M., Negroni, L., Sommerer, N., and Burstin, J. (2004). Proteome reference maps of vegetative tissues in pea. An investigation of nitrogen mobilization from leaves during seed filling. *Plant Physiol.* 135, 2241–2260. doi: 10.1104/pp.104.041947
- Schliemann, W., Ammer, C., and Strack, D. (2008). Metabolite profiling of mycorrhizal roots of *Medicago truncatula*. *Phytochemistry* 69, 112–146. doi: 10.1016/j.phytochem.2008.02.004
- Schmitz, R. J., He, Y., Valdés-López, O., Khan, S. M., Joshi, T., Urlich, M. A., et al. (2013). Epigenome-wide inheritance of cytosine methylation variants in a recombinant inbred population. *Genome Res.* 23, 1663–1674. doi: 10.1101/gr.152538.112
- Schulze, W. X., and Usadel, B. (2010). Quantitation in mass-spectrometry-based proteomics. *Ann. Rev. Plant Biol.* 61, 491–516. doi: 10.1146/annurev-arplant-042809-112132
- Shen, H., He, H., Li, J., Chen, W., Wang, X., Guo, L., et al. (2012). Genome-wide analysis of DNA methylation and gene expression changes in two *Arabidopsis* ecotypes and their reciprocal hybrids. *Plant Cell* 24, 875–892. doi: 10.1105/tpc.111.094870
- Silva, J., Scheffler, B., Sanabria, Y., De Guzman, C., Galam, D., Farmer, A., et al. (2012). Identification of candidate genes in rice for resistance to sheath blight disease by whole genome sequencing. *Theor. Appl. Genet.* 124, 63–74. doi: 10.1007/s00122-011-1687-4
- Simpson, C. E., Starr, J. L., Church, G. T., Burrow, M. D., and Paterson, A. H. (2003). Registration of ‘NemaTAM’ peanut. *Crop Sci.* 43:1561. doi: 10.2135/cropsci2003.1561
- Singh, D., Singh, P. K., Chaudhary, S., Mehla, K., and Kumar, S. (2012). Exome sequencing and advances in crop improvement. *Adv. Genet.* 79, 87–121. doi: 10.1016/B978-0-12-394395-8.00003-7
- Singh, N., Jayaswal, P. K., Panda, K., Mandal, P., Kumar, V., Singh, B., et al. (2015). Single-copy gene based 50K SNP chip for genetic studies and molecular breeding in rice. *Sci. Rep.* 5:11600. doi: 10.1038/srep11600
- Singh, N. K., Gupta, D. K., Jayaswal, P. K., Mahato, A. K., Dutta, S., Singh, S., et al. (2012). The first draft of the pigeonpea genome sequence. *J. Plant Biochem. Biotechnol.* 21, 98–112. doi: 10.1007/s13562-011-0088-8
- Singh, V. K., Khan, A. W., Saxena, R. K., Kumar, V., Kale, S. M., Chitikineni, A., et al. (2015). Next-generation sequencing for identification of candidate genes for *Fusarium* wilt and sterility mosaic disease in pigeonpea (*Cajanus cajan*). *Plant Biotechnol. J.* doi: 10.1111/pbi.12470 [Epub ahead of print].
- Subba, P., Kumar, R., Gayali, S., Shekhar, S., Parveen, S., Pandey, A., et al. (2013). Characterisation of the nuclear proteome of a dehydration-sensitive cultivar of chickpea and comparative proteomic analysis with a tolerant cultivar. *Proteomics* 13, 1973–1992. doi: 10.1002/pmic.201200380
- Suzuki, H., Sasaki, R., Ogata, Y., Nakamura, Y., Sakurai, N., Kitajima, M., et al. (2008). Metabolic profiling of flavonoids in *Lotus japonicus* using liquid chromatography Fourier transform ion cyclotron resonance mass spectrometry. *Phytochemistry* 69, 99–111. doi: 10.1016/j.phytochem.2007.06.017
- Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., et al. (2013a). QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J.* 74, 174–183. doi: 10.1111/tpj.12105

- Takagi, H., Tamiru, M., Abe, A., Yoshida, K., Uemura, A., Yaegashi, H., et al. (2015). MutMap accelerates breeding of a salt-tolerant rice cultivar. *Nat. Biotechnol.* 33, 445–449.
- Takagi, H., Uemura, A., Yaegashi, H., Tamiru, M., Abe, A., Mitsuoka, C., et al. (2013b). MutMap-Gap: whole-genome resequencing of mutant F2 progeny bulk combined with de novo assembly of gap regions identifies the rice blast resistance gene Pii. *New Phytol.* 200, 276–283. doi: 10.1111/nph.12369
- Tarawali, G., and Ogunbile, O. A. (1995). Legumes for sustainable food production in semi-arid savannahs. *ILEIA Newslett.* 11, 18–23.
- The Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815. doi: 10.1038/35048692
- Thompson, M. J. (2014). High-throughput SNP genotyping to accelerate crop improvement. *Plant Breed. Biotechnol.* 2, 195–212. doi: 10.9787/PBB.2014.2.3.195
- Trick, M., Adamski, N. M., Mugford, S. G., Jiang, C. C., Febrer, M., and Uauy, C. (2012). Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat. *BMC Plant Biol.* 12:14. doi: 10.1186/1471-2229-12-14
- Unterseer, S., Bauer, E., Haberer, G., Seidel, M., Knaak, C., Ouzunova, M., et al. (2014). A powerful tool for genome analysis in maize: development and evaluation of the high density 600 k SNP genotyping array. *BMC Genomics* 15:823. doi: 10.1186/1471-2164-15-823
- Vadivel, A., and Kumaran, A. (2015). Gel based proteomics in plants: time to move on from the tradition. *Front. Plant Sci.* 6:369.
- Valente, F., Gauthier, F., Bardol, N., Joets, J., Charcosset, A., and Moreau, L. (2013). OptiMAS: a decision support tool for marker-assisted assembly of diverse alleles. *J. Hered.* 104, 586–590. doi: 10.1093/jhered/est020
- Vanderschuren, H., Lentz, E., Zainuddin, I., and Gruissem, W. (2013). Proteomics of model and crop plant species: status, current limitations and strategic advances for crop improvement. *J. Proteomics* 93, 5–19. doi: 10.1016/j.jprot.2013.05.036
- van Orsouw, N. J., Hogers, R. C. J., Janssen, A., Yalcin, F., Snoeijs, S., Verstege, E., et al. (2007). Complexity reduction of polymorphic sequences (CRoPS™): a novel approach for large-scale polymorphism discovery in complex genomes. *PLoS ONE* 2:e1172. doi: 10.1371/journal.pone.0001172
- Varshney, R. K. (2016). Exciting journey of 10 years from genomes to fields and markets: some success stories of genomics-assisted breeding in chickpea, pigeonpea and groundnut. *Plant Sci.* 242, 98–107. doi: 10.1016/j.plantsci.2015.09.009
- Varshney, R. K., Chen, W., Li, Y., Bharti, A. K., Saxena, R. K., Schlueter, J. A., et al. (2012a). Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nat. Biotechnol.* 30, 83–89. doi: 10.1038/nbt.2022
- Varshney, R. K., and Dubey, A. (2009). Novel genomic tools and modern genetic and breeding approaches for crop improvement. *J. Plant Biochem. Biotechnol.* 18, 127–138. doi: 10.1007/BF03263311
- Varshney, R. K., Gaur, P. M., Chamarthi, S. K., Krishnamurthy, L., Tripathi, S., Kashiwagi, J., et al. (2013a). Fast-track introgression of QTL-hotspot for root traits and other drought tolerance traits in JG 11, an elite and leading variety of chickpea (*Cicer arietinum* L.). *Plant Genome* 6. doi: 10.3835/plantgenome2013.07.0022
- Varshney, R. K., Hiremath, P. J., Lekha, P., Kashiwagi, J., Balaji, J., Deokar, A. A., et al. (2009). A comprehensive resource of drought- and salinity-responsive ESTs for gene discovery and marker development in chickpea (*Cicer arietinum* L.). *BMC Genomics* 10:523. doi: 10.1186/1471-2164-10-523
- Varshney, R. K., Kudapa, H., Pazhamala, L., Chitikineni, A., Thudi, M., Bohra, A., et al. (2015a). Translational genomics in agriculture: some examples in grain legumes. *Crit. Rev. Plant Sci.* 34, 169–194. doi: 10.1080/07352689.2014.897909
- Varshney, R. K., Mohan, S. M., Gaur, P. M., Chamarthi, S. K., Singh, V. K., Srinivasan, S., et al. (2013b). Marker-assisted backcrossing to introgress resistance to *Fusarium* wilt (FW) race 1 and *Ascochyta* blight (AB) in C 214, an elite cultivar of chickpea. *Plant Gen.* 7. doi: 10.3835/plantgenome2013.10.0035
- Varshney, R. K., Mohan, S. M., Gaur, P. M., Gangarao, N. V. P. R., Pandey, M. K., Bohra, A., et al. (2013c). Achievements and prospects of genomics-assisted breeding in three legume crops of the semi-arid tropics. *Biotechnol. Adv.* 31, 1120–1134. doi: 10.1016/j.biotechadv.2013.01.001
- Varshney, R. K., Pandey, M. K., Janila, P., Nigam, S. N., Sudini, H., Gowda, M. V. C., et al. (2014a). Marker-assisted introgression of a QTL region to improve rust resistance in three elite and popular varieties of peanut (*Arachis hypogaea* L.). *Theor. Appl. Genet.* 127, 1771–1781. doi: 10.1007/s00122-014-2338-3
- Varshney, R. K., Ribaut, J. M., Buckler, E. S., Tuberosa, R., Rafalski, J. A., and Langridge, P. (2012b). Can genomics boost productivity of orphan crops? *Nat. Biotechnol.* 30, 1172–1176. doi: 10.1038/nbt.2440
- Varshney, R. K., Singh, V. K., Hickey, J., Xun, X., Marshall, D. F., Wang, J., et al. (2015b). Analytical and decision support tools for genomics-assisted breeding. *Trends Plant Sci.* doi: 10.1016/j.tplants.2015.10.018 [Epub ahead of print].
- Varshney, R. K., Song, C., Saxena, R. K., Azam, S., Yu, S., Sharpe, A. G., et al. (2013d). Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat. Biotechnol.* 31, 240–246. doi: 10.1038/nbt.2491
- Varshney, R. K., Terauchi, R., and McCouch, S. R. (2014b). Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. *PLoS Biol.* 12:e1001883. doi: 10.1371/journal.pbio.1001883
- Varshney, R. K., Thudi, M., Nayak, S. N., Gaur, P. M., Kashiwagi, J., Krishnamurthy, L., et al. (2014c). Genetic dissection of drought tolerance in chickpea (*Cicer arietinum* L.). *Theor. Appl. Genet.* 127, 445–462. doi: 10.1007/s00122-013-2230-6
- Vigeolas, H., Chinoy, C., Zuther, E., Blessington, B., Geigenberger, P., and Domoney, C. (2008). Combined metabolomic and genetic approaches reveal a link between the polyamine pathway and albumin 2 in developing pea seeds. *Plant Physiol.* 146, 74–82. doi: 10.1104/pp.107.111369
- Vodkin, L. O., Khanna, A., Robin, S. R., Steven, J., Clough, S. J., Gonzalez, D. O., et al. (2004). Microarrays for global expression constructed with a low redundancy set of 27,500 sequenced cDNAs representing an array of developmental stages and physiological conditions of the soybean plant. *BMC Genomics* 5:73. doi: 10.1186/1471-2164-5-73
- War, A., Robert, C., Hume, D., Archibald, A., Deeb, N., and Watson, M. (2015). Exome sequencing: current and future perspectives. *G3* 5, 1543–1550. doi: 10.1534/g3.115.018564
- Watson, B. S., Asirvatham, V. S., Wang, L., and Sumner, L. W. (2003). Mapping the proteome of barrel medic (*Medicago truncatula*). *Plant Physiol.* 131, 1104–1123. doi: 10.1104/pp.102.019034
- Wu, J., Wang, L., Li, L., and Wang, S. (2014). De novo assembly of the common bean transcriptome using short reads for the discovery of drought-responsive genes. *PLoS ONE* 9:e109262. doi: 10.1371/journal.pone.0109262
- Xu, J., Yuan, Y., Xu, Y., Zhang, G., Guo, X., Wu, F., et al. (2014). Identification of candidate genes for drought tolerance by whole-genome resequencing in maize. *BMC Plant Biol.* 14:83. doi: 10.1186/1471-2229-14-83
- Xu, X., Zeng, L., Tao, Y., Vuong, T., Wan, J., Boerma, R., et al. (2013). Pinpointing genes underlying the quantitative trait loci for root-knot nematode resistance in paleopolyploid soybean by whole genome resequencing. *Proc. Natl. Acad. Sci. U.S.A.* 110, 13469–13474. doi: 10.1073/pnas.1222368110
- Xu, Y. (2010). Molecular plant breeding. *Crop Sci.* 50:2196. doi: 10.2135/cropsci2010.12.0004br
- Yang, Z. B., Eticha, D., Führs, H., Heintz, D., Ayoub, D., Van Dorselaer, A., et al. (2013). Proteomic and phosphoproteomic analysis of polyethylene glycol-induced osmotic stress in root tips of common bean (*Phaseolus vulgaris* L.). *J. Exp. Bot.* 64, 5569–5586. doi: 10.1093/jxb/ert328
- Ye, H., Gemperline, E., Venkateshwaran, M., Chen, R., Delaux, P. M., Howes-Podoll, M., et al. (2013). MALDI mass spectrometry-assisted molecular imaging of metabolites during nitrogen fixation in the *Medicago truncatula*-*Sinorhizobium meliloti* symbiosis. *Plant J.* 75, 130–145. doi: 10.1111/tjpl.12191
- Yu, J., Hu, S., Wang, J., Wong, G. K., Li, S., Liu, B., et al. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science* 296, 79–92. doi: 10.1126/science.1068037

- Zadraznik, T., Hollung, K., Egge-Jacobsen, W., Meglič, V., and Šuštar-Vozlič, J. (2013). Differential proteomic analysis of drought stress response in leaves of common bean (*Phaseolus vulgaris* L.). *J. Proteomics* 78, 254–272. doi: 10.1016/j.jprot.2012.09.021
- Zhang, N., Venkateshwaran, M., Boersma, M., Harms, A., Howes-Podoll, M., den Os, D., et al. (2012). Metabolomic profiling reveals suppression of oxylipin biosynthesis during the early stages of legume–rhizobia symbiosis. *FEBS Lett.* 586, 3150–3158. doi: 10.1016/j.febslet.2012.06.046
- Zhang, Y., Gao, P., and Yuan, J. S. (2010). Plant protein-protein interaction network and interactome. *Curr. Genomics* 11, 40–46. doi: 10.2174/138920210790218016

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Pandey, Roorkiwal, Singh, Ramalingam, Kudapa, Thudi, Chitikineni, Rathore and Varshney. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.