Review

Legume genomics and transcriptomics: From classic breeding to modern technologies

Muhammad Afzal [a], Salem S. Alghamdi [a], Hussein H. Migdadi [a,*], Muhammad Altaf Khan [a], Nurmansyah [a], Shaher Bano Mirza [b,c], Ehab El-Harty [a]

[a] Department of Plant Production, College of Food and Agricultural Sciences, King Saud University, Riyadh, Saudi Arabia
[b] Computational Biology and Molecular Simulations Laboratory, Department of Biophysics, School of Medicine, Bahcesehir University (BAU), Istanbul, Turkey
[c] Department of Biosciences, COMSATS Institute of Information Technology (CIIT), Chak Shahzad, Islamabad, Pakistan

ABSTRACT

Legumes are essential and play a significant role in maintaining food standards and augmenting physio-chemical soil properties through the biological nitrogen fixation process. Biotic and abiotic factors are the main factors limiting legume production. Classical breeding methodologies have been explored extensively about the problem of truncated yield in legumes but have not succeeded at the desired rate. Conventional breeding improved legume genotypes but with more resources and time. Recently, the invention of next-generation sequencing (NGS) and high-throughput methods for genotyping have opened new avenues for research and developments in legume studies. During the last decade, genome sequencing for many legume crops documented. Sequencing and re-sequencing of important legume species have made structural variation and functional genomics conceivable. NGS and other molecular techniques such as the development of markers; genotyping; high density genetic linkage maps; quantitative trait loci (QTLs) identification, expressed sequence tags (ESTs), single nucleotide polymorphisms (SNPs); and transcription factors incorporated into existing breeding technologies have made possible the accurate and accelerated delivery of information for researchers. The application of genome sequencing, RNA sequencing (transcriptome sequencing), and DNA sequencing (re-sequencing) provide considerable insights for legume development and improvement programs.

Moreover, RNA-Seq helps to characterize genes, including differentially expressed genes, and can be applied for functional genomics studies, especially when there is limited information available for the studied genomes. Genome-based crop development studies and the availability of genomics data as well as decision-making gears look be specific for breeding programs. This review mainly presents an overview of the path from classical breeding to new emerging genomics tools, which will trigger and accelerate genomics-assisted breeding for recognition of novel genes for yield and quality characters for sustainable legume crop production.

© 2019 The Authors. Published by Elsevier B.V. on behalf of King Saud University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Contents

1. Introduction

The shortage of food expected in coming years and the supply and demand curve of food is problematic because of the increasing population (7.2 billion to 9.6 billion) by 2050 (Gerland et al., 2014). The world population has increased the demand of food production which has raised the question of the available resources (Fedoroff, 2015). Natural resource depletion and changing climate have pretentious the continuous energies to attain the desired level of production. Food security with the aim of high nutrition and yield is main challenge for research scientist and farming community in 2050. Grain legumes contain high amounts of protein, vitamins, minerals, iron, calcium, zinc, magnesium, omega-3 and fatty acids. The value of grain legumes is even higher in the Asian region, where a large part of the population is vegetarian. Groundnut and soybean are important oilseed legumes, which are largely used in cuisine and in the preparation of sweetmeat. One of the most important characteristics of legumes is biological nitrogen fixation and, in this context, legumes considered an important crop in sustainable agriculture ensuring that residual nitrogen resources are made available to non-legume crops (Pandey et al., 2016). The highest protein content is found in soybean (33–45%), common bean (21–39%), wing bean (30–37%), cowpea (21–35%), groundnut (24–34%), pea (21–33%), moth bean and urd bean (21–31%), lentil (20–31%), grass pea (23–30%), chickpea (15–30%), horse gram (19–29%), and rice bean (18–27%) (Pandey et al., 2016). Legume production cannot be enhanced to the desired level because of biotic and abiotic limitations. Moreover, the decrease in available land areas and water resources (because of climate change) will exacerbate current adverse conditions in coming years and, as a result, protein crops such as legumes may be at risk. For this purpose, conventional breeding protocols have focused on the problems of low productivity of legumes over the last two decades, but the associated goals have not been fully achieved. Quality-related characteristics (i.e. Protein) (Gnanasambandam et al., 2012) and the absence of anti-nutritional factors (e.g., tannins; Mikic et al., 2009) are important targets for legume crop improvement programs.

Gene regulation in both legume plants and bacteria occurs at all stages of nodule development. Numerous researches have focused on determining the genetic control of infection by bacteria and the initial processes involved in nodulation. Some early nodulin (e.g., ENOD11 and R1P1) and an essential Nod factor (NF) with signaling triggered by bacterial infection have been recognized (Yang et al., 1993). Plant breeding success has mostly depended on the use of genetics and mutation-induced breeding of genetic diversity resources to identify selected genotypes for breeding purposes. Also, it depends on the documentation of novel genes of interest (GOI) and selected breeding methodologies based on phenotype data (Perez et al., 2012). The development of genomic breeding values based on a high-throughput DNA sequencing method known as next-generation sequencing (NGS). NGS and other new technologies enable plant breeders to develop markers for diversity analyses, genotyping, high-density genetic maps, and population genetics studies, and these can be merged into available breeding technologies to achieve desired goals (Lorenz et al., 2011).

Moreover, genomic approaches are essential to deal with multigenic complex traits and showed high impact on the environment. Genomics tools are also essential to map quantitative trait loci (QTLs) of rare alleles that often remain undetected in the gene pool (Vaughan et al., 2007). Using more advanced techniques, such as cDNA sequencing [i.e., expressed sequence tags (ESTs)], provides a better understanding of genes expressed in leaves or roots, at a specific stage, or connected to a precise environmental condition. Even though there are some restrictions to the use of cDNA methodology (e.g., lack of evidence for non-coding regions), the identification and collection of ESTs are useful for researchers.

Draft genome sequencing allows the improvement of crops based on genomic gains and on the selection of genes that confer favorable traits that increase yield and nutrient quality. It also provides the information necessary to understand genome construction and uncovers the pathways linked to stress response, besides determining the mutagenic changes resulting from deletions and insertions in the genome. The availability of genetic resources worldwide provides a chance for plant community researchers to detect unique alleles or gene combinations that are important for crop development (McCouch et al., 2013). This manuscript presents a review of the classical breeding methods and the most advanced technologies used in the development of legumes crops. We briefly discuss transcriptome technologies and computational biology approaches that help to facilitate the selection of GOIs for the development of legume improvement programs. Moreover, we also discuss the genomic tools that can be used to detect complex traits, such as nitrogen fixation, quality, and yield. Genomic-based techniques improve breeding programs; that is, the distribution of genomics and the provision of decision tools seems to be critical for selection. This review also offers an impression of new emerging genomic tools, such as genomic-assisted breeding, that may accelerate the production of sustainable legume crops.

1.1. Marker-assisted breeding (MAB)

Different applications of DNA markers are using for breeding purposes. However, the use of molecular breeding for cultivar development has become more popular because of the method's precise characteristics and because it is less time-consuming and provides maximum benefits. Moreover, simple traits are much more critical when associated with complex or multiple genes. Molecular markers are significantly crucial for the successful selection of cultivars that are resilient to parasitic weeds (e.g., boom rape) and diseases (e.g., Ascochyta blight, rust, and chocolate spot) (Rietzschel and De Buyzere, 2012). These markers can also be utilized to find linked genes responsible for the plant's growth habit and quality (i.e., vicine and tannin content) (Torres et al., 2010). MAB can be used together with conventional field breeding to accelerate the selection of desirable traits (Collard and Mackill,

2007) and increase selection efficacy (Ragot et al., 2007). MAB relayed on the genotype of a specific marker system can be used to enhance the efficiency and effectiveness of marker selection relative to classical breeding (Collard and Mackill, 2008). Breeders can use these genes or alleles as diagnostic tools to identify desired genes or GOIs.

The progress in linkage maps for faba beans has been slow compared to that of other legume crops. The development of low-density linkage maps and it was determined by a molecular marker system isozyme; Restriction Fragment Length Polymorphism (RFLP) and Random Amplification of Polymorphic DNA (RAPD) (Avila et al., 2003). Afterward, sequence-based characterization was done, i.e., using intron-targeted amplified polymorphisms (ITAPs), simple sequence repeats (SSRs), and sequence-characterized amplification regions (SCARs) which allowed the progress of more significant maps used for comparative genome study of legume (Zeid et al., 2009). A set of functional SSR markers have been developed using a genomic library rich with GA/CT motifs for utilization in legume breeding program (Verma et al., 2014). The development of ESTs (Varshney et al., 2013a, 2013b, 2013c) or genome survey sequences (GSS) can extend marker development, especially of SSRs and SNPs used to improve map resolution (Webb et al., 2015). Later studies focused on specific population characters for genetic maps, discovery of QTLs based on morphological characters, and resilience against abiotic stresses and diseases (Atienza et al., 2016). The development of molecular markers (cross legume species) from *Medicago,* simplifies gene order and the comparison of genetic maps at micro and macro levels among legumes.

The achievement of the MAB mainly based on many aspects such as the genetic traits, flanking regions distance and the gene of interest, information of genetic background of the gene of interest, and the methodology available (Francia et al., 2005). Also, the selection of primers depends on specific genes that have minimal effects but are relatively easy to map for phenotyping. These markers are not polymorphic, thus requiring different markers for genetic diversity. SSRs show that high levels of polymorphism and gene-rich regions may be effectively used for emerging maps by identifying the desired alleles and QTLs in faba bean (Song et al., 2004). SSRs markers have been used to construct maps based on traits (agronomy) and identified QTLs related to yield, resilience to pests and diseases in soybean (*Glycine* max L.) (Du et al., 2009). Many molecular markers used for mapping QTLs in food legumes.

Table 1
Summary of the linked molecular markers for agronomic and disease resistance gene (s)/quantitative trait loci.

Traits	Mapping population	Linked markers	References
Determinate growth habit		CAPS-TFL1 (HindIII)[a] Ti-dCAPS[a]	Avila et al. (2007)
Zero tannins	Vf6 × zt-1 (F_2)	$SCC5_{551}$/ $SCG11_{1171}$	Gutierrez et al. (2007)
Low vicine–convicine	Vf6 × vc- (F_2)	$SCH01_{620}$/ $SCAB12_{850}$	Gutierrez et al. (2006)
Frost tolerance	Frost tolerance	$U09_{1499}$/$B20_{803}$	Arbaoui and Link (2008)
Ascochyta blight resistance	Vf6 × Vf136 (F_2)	$OPAC06_{1023}$	Diaz et al. (2005)
Broomrape resistance	Vf6 × Vf136 (F_2) Vf6 × Vf136 (RILs)	$OPAI13_{1018}$/ $OPAC06_{396}$	Diaz et al. (2005)
Rust resistance	2N52 × Vf176	$OPI20_{900}$/ $OPL18_{1032}$	Avila et al. (2003)

Table 2
Quantitative trait loci (QTLs) related to resistance to disease in soybean crops.

Traits	Gene	Linked markers	References
Resistance to leaf rust		Rpp5	Arahana et al. (2001)
White mold		Satt009	
Resistance to cyst nematode			Concibido et al. (1997)
SCN	Rhg4, rhg1, and Rfs	Satt009	Meksem et al. (2000)
Resistance to *Fusarium solani*		Satt080	Njiti et al. (2002)
Black pod-of-staff	Rhg1	Satt038, Satt130 Sat-168, Satt 309	Bachman et al. (2001)
Resistance to stain-frogeye	RCSPeking	Satt244	Yang et al. (2001)
Resistance root-knot nematode		Satt114	Gavioli (2011)

Table 3
Some legume genome sequencing results, the discovery of genes, and ESTs generated.

Legumes	No. of genes[*]	No. of ESTs[**]	References
Cajanus cajan	48,680	25,640	Varshney et al. (2012)
Cicer arientum (Chickpea)	28,269	46,064	Varshney et al. (2013a, 2013b, 2013c)
Lotus japonicus (trefoil)	46,430	1,530,030	Schmutz et al. (2010)
Medicago truncatula (barrel medic)	47,845	286,175	Young et al. (2011)

[*] As of January 2014.
[**] According to the Dana Farber Gene Indices (compbio.dfci.harvard.edu/tgi/plant.html).

Some important linked molecular markers for agronomic traits and disease resistance genes and quantitative traits loci presented in Table 1. In the last decade, an effort was made to map soybean (*Glycine max* (L) Merr.) and Pea (*Pisum sativum* L.) genomes. Over 300 QTLs associated with various traits have been recognized using MAB (Gupta et al., 2010). Sequencing of wild and cultivated soybean genomes produced 205,614 SNPs in the wild accessions relative to the cultivated lines, which could be favorable for the QTL analysis (Lam et al., 2010). Some quantitative trait loci related to disease resistance in soybean reported in Table 2.

1.2. Proteomics

The advent of the recent 'omics' techniques has helped to deliver a new view for crop developing, including finding novel genes and intricate pathway mechanisms for agronomic traits integrated through genomics and functional approaches based on molecular and morphological data (Langridge and Fleury, 2011). Proteomics is based on cellular complements of proteins discussing its biological units at a specific stage or a specific condition to be analyzed (Jorrin-Novo et al., 2019). Studies on proteomics also involve quantitative measurements, such as composition, genetic alterations, and modulation of specific stages of stress conditions in crop plants. Protein-generated data together with genetic determinants may lead to the identify GOIs or desired traits, and later on, this data can then be incorporated into the molecular breeding improvement program (Vanderschuren et al., 2013). Proteomics comprises mapping, protein profiling, post-translational modification (PTMs), and interprotein interaction networks (Katam et al., 2015). An important potential area of crop research is translational proteomics, which can be applied to crop research for agricultural

production (Jorrin-Novo et al., 2019). The application of ortho-proteomics and comparative analysis improve breeding programs by standardizing and maintaining high-quality protein data of specific crop (Vanderschuren et al., 2013).

Recent developments in proteomic studies, such as the application of mass spectrometry (MS), increase precision and understanding, and different software used for advanced proteomics quantification. It also involved quantification based on gel, labeled probe, or label-free methods for protein quantification with precision and accuracy (Hu et al., 2015). It is currently possible to understand the mechanism involved under different stress conditions at different growth stages by using two-dimensional gel electrophoresis along with MS analyses. Moreover, two-dimensional protein gel apparatus is important for high resolving power by labeling samples with fluorescent dye, and they allow bands to separate in the same gel (Vadivel and Kumaran, 2015). Another method is the shot-gun proteomics strategy using a peptide-centric database linked with liquid chromatography tandem-MS method. It provides an efficient high-throughput analysis of cell or organelle (proteome) of major proteins (Jorrin-Novo et al., 2019). Biomarker validation of a specific protein can determine by the selected reaction monitoring (SRM) method. SRM method determines the molecular mechanism underlying specific traits of crops (Jacoby et al., 2013). It is more sensitive method to determine the number of specific proteins in crop plants. The protein-protein interaction (PPI) methodology is important to determine the proteomics of molecular mechanisms, i.e., protein complexes, signal transduction, and stress signal (Westermarck et al., 2013). The more advanced approaches used for legume improvement and the application of these methods addressed in different legume species. The important methods to identify proteins and comparing the reference databases with MS/MS spectra database (Romero-Rodríguez et al., 2014). Moreover, these methods also include the Tandem, Sequest, and Mascot search databases, which used for identification of proteins (Senkler and Braun, 2012). Moreover, proteomics plays a significant part in the study of differences at cellular, subcellular, and plant levels under abiotic stress conditions (Hossain and Komatsu, 2014). Alghamdi et al. (2018) studied seven cowpea genotypes about solubility-based protein analysis; their results revealed that water-soluble proteins were dominant in cowpea seeds when compared to total proteins. Moreover, gel patterns revealed molecular diversity, total protein complement, and the presence of different protein fractions. Gel and nano liquid chromatography (LC) MS/MS techniques were used in soybean to determine the osmotic stress in plasma membranes (Nouri and Komatsu, 2010). The results obtained using the gel approach identified four upregulated and eight downregulated proteins, while the LC methodology recognized 11 upregulated and 75 downregulated proteins (Ahsan et al., 2010). Further studies determined abiotic stress response using protein profiling. Abiotic stresses in chickpea constitute a severe threat to its productivity; therefore, efforts have been made to determine the genetic basis of stress tolerance in chickpea. Heidarvand and Maali-Amiri (2013) recognized a unique dehydration signaling component, designated CaSUN1 (*Cicer arietinum* - Sad1/UNC-84). The function of CaSUN1 is to localize to endoplasmic reticulum and nuclear membrane beside small vacuolar vesicles. In another study, they also determined protein changes in chickpea at early growth stages facing cold stress (Jaiswal et al., 2014). Some research was also conducted on soybean to determine protein profiling concerning water stress (Seminario et al., 2017). Irar et al. (2014) reported the presence of a signal pathway that restricts nitrogen fixation under drought conditions; 18 protein (Nodule regulated) by pea and rhizobium genome identified. Moreover, two-dimensional gel-electrophoresis and MALDI-TOF-MS approaches were used to determine the diversity for protein

deposition at germination stage during osmotic stress in seeds (Brosowska-Arendt et al., 2014).

1.3. Metabolites in legumes

Plant metabolites represent biochemical markers or phenotype changes that occur in a cell or tissue part while their end products appear in the form of gene expression (Hong et al., 2016). Moreover, detailed quantitative and qualitative results provide an understanding of the gene function (Wink, 2013). The components related to stress such as metabolism, stress signal acclimation process, and transduction molecules could be identified by metabolomics study (Larrainzar et al., 2009). The recognized metabolites were further analyzed by direct calculations or by relating them with differences in protein and transcriptome results using mutant analysis. Currently, different metabolomics techniques are used for metabolomics profiling, including metabolite determination, analytical techniques based on separation, and detection methods (Doerfler et al., 2014). The separation approach includes gas chromatography (GC), which is used to determine the primary metabolites (i.e., sugars and amino acid) (Weckwerth, 2011); ultra-performance liquid chromatography; capillary electrophoresis (CE) which is used to determine ionic metabolites (Doerfler et al., 2014), and liquid chromatography which is used to determine the secondary metabolites (Weckwerth, 2011). Gas chromatography–Mass spectrometry (GC–MS) has been used to determine plant metabolites and electron impact (EI) with strong interface of GC with MS and allows them to separate fragment patterns to be extremely reproducible. Some important methods and application described below.

Application of metabolomics profiling is limited in legumes, but Ramalingam et al. (2015) used a quantitative MS method to determine metabolomics diversity among the symbionts in *Medicago*. Later, a study conducted on salinity and drought stress, and metabolites related analysis were made use priming technique for symbiotic interaction (plant nodulation) and nitrogen fertilizer in *Medicago* (Staudinger et al., 2012). Sanchez et al. (2011) used GC with electron impact ionization to determine the soluble molecules profiled. *Lotus* plants can grow in highly saline seaside sections is compared with cultivated glycophytic grass fodder species. The results of a comparative analysis predicted the presence of metabolites conserved resulting from drought in *Lotus*, and that the extremophile *Lotus* species recorded upper salt levels with differential reordering of shoot nutrient when exposed to salinity.

Abiotic stress may affect plants in many different manners, and each plant responds differently and produces different types of metabolomics compounds under stress. The metabolomics approach using Capillary electrophoresis-mass spectrometry (CE-MS) for flooding stress in soybean roots and hypocotyl and analysis recorded about 81 metabolites related to mitochondria. Moreover, tricarboxylic acid (TCA) analysis metabolites, amino acids, NADH, and pyruvate contents increased, while ATP decreased (Komatsu et al., 2011). However, the identification of these metabolites and detail analysis is required to determine the response in stress conditions. A study was conducted to determine the phosphorous response in common bean roots and nodules by metabolite profiling under stress conditions when phosphorus is deficient in the soil (Hernández et al., 2009). Results predicted that amino acid and several sugars levels increased under phosphorus-deficiency stress in the roots. In another study, the organic acid amount decreases when phosphorus is deficient in the root area and reproduce exudation of these compounds (metabolites) from root to rhizosphere

Table 4
Released mutants by induced mutagenesis in major legume crops (IAEA, 2018).

No	Legume name	Scientific name	No. of released mutants	Year of release		Mutagenic agent				S
				Oldest	Newest	P	C	NN	Co	
1	Soybean	*Glycine* max	174	1962	2017	122	13	36	1	2
2	Groundnut	*Arachis hypogaea*	76	1971	2014	64	9	3	0	0
3	Common bean	*Phaseolus vulgaris*	57	1962	2007	26	22	9	0	0
4	Pea	*Pisum sativum*	34	1954	1995	20	10	3	0	1
5	Chickpea	*Cicer arietinum*	27	1981	2016	23	1	2	1	0
6	Faba bean	*Vicia faba*	20	1983	2008	12	6	2	0	0
7	Lentil	*Lens culinaris*	18	1981	2017	14	3	1	0	0
8	Cowpea	*Vigna ungiculanta*	13	1981	2017	7	4	2	0	0
9	Pigeonpea	*Cajanus cajun*	7	1977	2009	4	1	2	0	0

P: physical, C: chemical, NN: data not provided or/and indirect mutation, Co: combine treatment, S: spontaneous mutation.

(Hernaández et al., 2007). Dias et al. (2015) reported 49 primary metabolites from different salt stress response in chickpea genotypes. Integrated approaches include metabolomics, and transcriptomes techniques are important for worthy metabolomics reprogramming of the border cells in roots by specific data set pathways. This technique is important for identification of phytohormone levels, from auxin and lipoxygenase transcripts from root border cell and root tip. Similarly, another study was conducted using an integrated metabolomics and proteomic approach to elucidate the expression and regulation of metabolites under drought conditions in soybean (Komatsu et al., 2011). Moreover, metabolite study enabled showing high reprogramming of metabolism under drought stress conditions as well as the capability of nodule to recover after re-watering conditions.

Table 5
Gene isolation and identification of several legume crops through forward and reverse genetics using induced mutant plants.

Name of crop	Scientific name	Mutant phenotype	Mutagenic agent	Name of gene	Gene function	Reference
Soybean	*Glycine* max L.	Golden yellow trifoliate leaves, pods, and cotyledons as well as reduced plant height	EMS	*Golden Yellow Leaves (GLY)*	Chlorophyll biosynthesis	Li et al. (2017)
		Reduced plant height and shortened internodes	EMS	*Glycine* max *Dwarf (GmDW1)*	Gibberelline (GA) biosynthesis-deficient	Li et al. (2017)
		High seed oleic acid content	EMS	*Fatty Acid Desaturase 2 (FAD2)*	Conversion of oleic acid to linoleic acid	Lakhssassi et al. (2017)
		Early flowering	Gamma rays	*Glyma10g36600* (GI), *Glya02g33040* (AGL18), *Glyma17g11040* (TOC1), and *Glyma14g10530* (ELF3)	Affecting the expression of flowering promoter *Glycine* max *Flowering Locus T 2a (GmFT2a)*	Lee et al. (2016)
Pea	*Pisum sativum*	Novel leaf morphology (stipule size)	Fast neutron	*Stipule reduce (St)*	Regulating cell division and cell expansion in the stipule	Moreau et al. (2018)
		Non-flowering phenotype	X-rays	*VEGETATIVE1 (VEG1)*	Regulating secondary inflorescence	Berbel et al. (2012)
		Determinate growth habit	Gamma rays	*Pisum sativum Terminal Flower 1a (PsTFL1a)* or *DETERMINATE (DET)*	Maintain the indeterminacy of the apical meristem during flowering	Foucher et al. (2003)
		Early flowering	Gamma rays	*Pisum sativum Terminal Flower 1c (PsTFL1c)* or *LATE FLOWERING (LF)*	Repressor of flowering	Foucher et al. (2003)
Cowpea	*Vigna unguiculata* L.	Determinate growth habit	Gamma rays	*Vigna unguiculata Terminal Flower 1 (VuTFL1)*	Plant determinacy	Dhanasekar and Reddy (2015)
Barrel medic	*Medicago truncatula*	Pentafoliate leaf morphology, development of rachis structure, and alteration of petiole and rachis length	Fast neutron	*Palmate-like pentafoliata (PALM1)*	Encoding Cys(2)His(2) zinc finger transcription factor and maintaining trifoliate leaves morphology	Chen et al. (2010)
		Mutants unable to establish a symbiotic association with endomycorrhizal fungi	EMS and gamma rays	*Doesn't Make Infections 1, 2, 3 (DMI1, DMI2, DMI3)* and *Nodulation Signaling Pathway (NSP)*	Regulating rhizobium nodulation (Nod) factor transduction pathway	Catoira et al. (2000)
		Leaflets unable to fold in the dark	Fast neutron	*Elongated Petiolule1 (ELP1)*	Encoding a putative plant-specific LATERAL ORGAN BOUNDARIES (LOB) domain transcription factor and development of the motor organ	Chen et al. (2012)
		Different composition of hemolytic saponins	EMS	*Cytochrome 72A67 (CYP72A67)*	Synthesis of hydroxylation at the C-2 position downstream of oleanolic acid and catalyzer of oxidation at the C-2 position in the hemolytic sapogenin pathway	Biazzi et al. (2015)

1.4. Mutation breeding in legumes

Induced mutagenesis has been used in legume crops to develop varieties, as indicated in Table 4. Table 4 shows data for the released variety database from the MVD of IAEA (2018). The development type of mutant variety could be a direct, indirect, or spontaneous mutation. Direct mutation can be generated using physical or chemical mutagens, as well as a combination of mutagens. A spontaneous mutation is a mutation that occurs naturally in the field. The first released variety is reported in 1954 on the pea, and it continues until now. However, in some crops such as pea and pigeon pea, as well as faba bean, the development of mutant variety through induced mutagenesis is seemed to be stagnant. Soybean has the highest number of released mutants, followed by groundnut and the common bean with 174, 76, and 57 mutant varieties, respectively. However, the number of released mutant variety in legume crops has been lower than that of cereal crops, such as wheat with 282 mutant varieties, barley with 312, and rice with 828 mutant varieties (IAEA, 2018). Table 4 also shows that physical mutagens contributed to most of the mutant variety compared to chemical mutagens. Among the physical mutagens, gamma radiation is the most widely used mutagen in mutation breeding (Kodyma et al., 2011). Gamma radiation has been reported as an effective mutagen to induce diversity in several legume crops, such as pigeon pea (Desai and Rao, 2014), cowpea (Girija et al., 2013), chickpea (Wani and Anis, 2008), and mung bean (Sangsiri et al., 2005). Induced mutagenesis has been used not only to develop varieties but also to study the genes related to a particular trait through forward or reverse genetics. Table 5 shows several applications of induced mutagenesis to isolate genes and identify gene function in several legume crops. The mutant characteristics studied presented morphological differences, from seed quality composition to secondary metabolite differences. The model legume, *Medicago truncatula* has been the most widely used legume to study gene function due to small (~450 Mb) genome size, rapid life cycle, abundant collection of mutants, and ecotypes (Tang et al., 2014).

1.5. Next-generation genotyping and sequencing technology

NGS and genotyping are significant and cost-effective techniques that facilitate the identification of some important gene structures. Next-generation genotyping also helps to gain perception about the essential mechanism of gene expression and metabolism. Moreover, it also facilitates the characterization of genomic resources and development, evolution, and MAB, even in the absence of sequence data. NGS also contributes to the efforts to speed-up the development of transcriptome sequencing, *de-novo* assembly, and sequence-based marker resources for crop development that are less costly than phenotyping (Unamba et al., 2015). The arrangement of the sequence provides the information of genes controlling the growth, adaptation concerning environment, and development. Successful genome sequencing accomplished for five legumes crops: chickpea (Varshney et al., 2013a, 2013b, 2013c), pigeon pea (Varshney et al., 2012), *Trefoil* sp. (Sato et al., 2008), soybean (Schmutz et al., 2010) and *Medicago* (Young et al., 2011).

At large scale, draft genome sequencing of unique cultivars for important traits has made possible the identification of structural diversity. Currently, genome-based sequencing is more popular because of the detailed analysis of genetic traits in plants and other organisms. Genotyping by sequencing has been used for mapping, purity testing, marker-trait link, MAS, and genomic selection (Varshney et al., 2014). It includes target-based amplicon sequencing, hybridization, and representation sequencing. The choice of genotyping technology depends on many factors i.e. nature of the

project, genome size, and funds. Also, MAB focused on less cost for running competitive allele-specific PCR (KASP) markers for diversity and identification of SNPs in a larger population (Thompson, 2014). Target sequencing based on amplicon technology has addressed many questions based on diversity and specific gene functions from population genetics (Naj et al., 2011). Besides, cheap and fast techniques are used to construct libraries with PCR, which is essential for identification of genes and development target amplicons and phylogenetic relationships (Bybee et al., 2011). However, this technique is restricted only to the identification of small loci and not significant for mapping complex traits. Moreover, SNPs by full genome-based delayed in many genomes such as groundnut and soybean.

An approach called complexity reduction of polymorphic sequences (CRoPS) established on DNA restriction digestion (methylation) has the characteristics needed to reduce the difficulty of two or more genetically diverse samples prepared by amplified fragment length polymorphism (AFLP) (Van Orsouw et al., 2007). Genotyping assays and their conversion rates make them more attractive for medium- to large-scale genotyping, especially find a single copy of genome or fewer SNPs. Peterson et al. (2012) established a new technique called double-digest restriction-site associated DNA (ddRAD). The uniqueness of this technique is that it uses different sizes of the genome to recover specific regions, which are disseminated into a genome randomly, and it improves the ability for multiplexing of many as hundreds of samples. Therefore, sequencing by genotyping has an advantage over RAD sequencing because of SNP detection and genotyping side by side (Poland et al., 2012). Exome sequencing is also becoming necessary because of the discovery of low frequency and infrequent diversity in coding that can be examined analytically using complex traits for crop improvement (Kiezun et al., 2012). It also can use 100,000 more markers relative to the Illumina system (Unterseer et al., 2014). Because of the higher efficiency of the exome array technology, it is a more suitable approach used for genomic-based sequencing breeding. The generated data is used for determining genomic estimated breeding value as well as for trail populations. The successful SNP arrays have been used in rice (Kumar et al., 2015) and maize (Unterseer et al., 2014). Affymetrix arrays with 60 K SNPs first used for three legumes—i.e., groundnut, chickpea, and pigeon pea—for trait mapping and molecular breeding analysis (Varshney, 2016).

Sequencing and re-sequencing of the plant genomes enabled researchers to dissever and indulge the procedures based on genetics and classification. Plant genomes have been decoded by genome sequencing as per requirements (Michael and Jackson, 2013). The first attempt for genome sequencing made in *Arabidopsis* (The Arabidopsis Genome Initiative, 2000) and later *indica* (Yu et al., 2002) and *japonica* genome sequencing (Sanger method) (International, 2005). Later, with the advancement in sequencing, NGS-based whole genome sequencing approaches were essential to reduce costs and time (Schwarze et al., 2018). The NGS application made it easy for researchers to decode the complex genome sequence. More than 100 plant genomes were sequenced (Varshney et al., 2009). However, chickpea, groundnut, and pigeon pea were considering orphan crops because of low genetic resources (Varshney et al., 2012). However, the invention of the NGS technique made possible the sequencing of these crops (Varshney et al., 2013a, 2013b, 2013c). Illumina technology, together with Sanger-based BAC-end sequence used for pigeon pea (ICPL 87119) genome assembly. It has generated about 237.2 Gb of paired-end reads. Data generated by Illumina and Sanger sequencing covers 72% of the pigeon pea genome, which is approximately 606 Mb data. *De-novo* gene prediction and gene annotation method have identified a 48 K gene in pigeon pea genome, which represents the higher side *de novo* assembly of the gen-

ome (Varshney et al., 2012). Another attempt made for pigeon pea sequencing based on long-read using GS-flx pyrosequencing to assemble 548 Mb data, and functionally annotated genes were identified (Singh et al., 2016).

There are two types of chickpea present Kabuli and Desi can differentiate based on color, size and seeds surface, morphology, and flower color. Both types are geographically different and vary in nutrition, adaptation and stress tolerance (biotic and abiotic stress) (Purushothaman et al., 2014). The NGS approach has also used for the Kabuli chickpea variety and approximately 153 Gb sequence data generated by Illumina sequencing. It contains 87.56 Gb of sequence information, representing 74% of the chickpea genome (Varshney et al., 2013a, 2013b, 2013c). The combination of different approaches has functionally annotated over 90% of the chickpea genome. Application of RAD and the whole genome re-sequencing approach have been used to collect data from 90 chickpea accessions (Varshney et al., 2013a, 2013b, 2013c). For the Desi chickpea (ICC 4958) accession, an updated version of the chickpea genome with a 2.7-fold increase of pseudo-molecules was determined (Parween et al., 2015). It includes more than 30 K protein-coding genes as compared to the previous chickpea genome. The chromosome sequencing approach is used later to detect misassembled region data at the chromosome level. These data help authenticate the genome assemblies at the chromosomal level in Desi and Kabuli draft genomes (Ruperao et al., 2014). Bertioli et al. (2015) took the initiative for the sequencing of groundnut progenitors. These data included genome A (V14167) and genome B (K30076) together, constituting the tetraploid genome of cultivated groundnut. The data took at plant tissue and different growth stages, and it covers about 155.5 Mb data (V14167) and about 154.4X coverage of the genome while 175.6 Mb (K30076) were assembled and cover about 120.6X of the total. The available data provide access to 97% of groundnut genes and would enable the development and understanding related to stress resilience and adaptability of the groundnut cultivars. A parallel study conducted for the documentation of genetic resources and allelic diversity important for agronomic traits available in the gene bank. Based on available information "International Crops Research Institute for the Semi-Arid Tropics" (ICRISAT) takes the initiative for re-sequencing the legume germplasm for the identification of unique combinations of genes. During the re-sequencing analysis, approximately 292 pigeon pea accessions were used to recognize 13.8 million Kb diverse sequences, with an average of 41.1/Kb data (Varshney et al., 2017). A detailed review of re-sequencing analysis delivered the genomic region and genetic diversity among the pigeon pea genome, which is helpful for selection and domestication. The most important is to recognize the genome based on morphological data considered for marking the traits linked with markers and important QTLs. A recent study conducted for the re-sequencing of 429 chickpea accessions collected from 45 countries. The results predicted that about 122 candidate regions and 204 genes identified under selection. Moreover, it was also predicted that the primary center was the eastern Mediterranean and transfer to Mediterranean/fertile to central Asia and then moved towards Central Asia to East Asia and finally to India. About 262 markers and many candidate genes for thirteen traits were identified from GWA (Varshney et al., 2019).

1.6. Transcriptomics, gene discovery, and marker development

Advancement in genetic diversity development is significant for molecular breeding and crop improvement program. Because of the limited availability of genetic resources, there is a dire need to discover more transcript genes and develop markers for specific traits and improve the legume breeding program. Besides the development of molecular markers, EST generation provided a quick and easy platform to discover the new gene, which is important for functional genomics and to help to understand the molecular mechanism for sustainable agriculture production. The NGS provides deep insight in transcriptome sequencing by quick and economical means (Morozova et al., 2009). The genome sequencing of five legumes genomes, gene discovery, and generation of ESTs presented in Table 3. A study was conducted to determine the drought-responsive gene and molecular markers development based on the gene in chickpea, where approximately 435,018 reads and 21,491 ESTs have produced. Moreover, the relative genome sequencing results for chickpea with the *Medicago* genome assembly predicted 42,141 aligned tentative unique sequences (TUSs) with putative gene structure generated. The tentative unique sequencing was also used to identify different markers. It includes SSR (728), SNPs (495), COS (387), and ISR (2088) (Hiremath et al., 2011). Similarly, in another study, 2 million sequences with an average length of 372 bp were generated in chickpea with pyrosequencing technology. The *de novo* assembly represents that hybrid long read and short read assembly gave good results. Approximately 34,760 transcripts were generated with an average of 1020 bp, which represents about 4.8% of the total chickpea genome (Garg et al., 2011). Varshney et al. (2009) reported 20,162 total ESTs and 48,796 bacterial artificial chromosomes (BAC). Nayak et al. (2010) reported approximately 2000 SSR markers. Moreover, for chickpea (80,238) sequence tags were generated from whole-genome sequence profiling (Molina et al., 2008).

Similarly, Vatanparast et al. (2016) analyze transcriptomes of Sri Lankan wing bean and stated that approximately 804,757 *de-novo* assembly reads were recorded, which produced 16,115 contigs and covered more than 90% of significant sequence data. Moreover, 97,241 singletons contigs transcripts also produced from the data set, and 12,956 SSRs generated. It also includes 2594 repeats for primer designing and 5190 SNPs in Wing bean genotypes. A study on Pongamia (*Millettia pinnata*), an essential medicinal legume species with industrial application, was conducted to characterize the seed transcriptomics using Illumina sequencing technology. The results predicted that approximately 83 billion reads contain 53,586 assembled unigenes with an average length of 787 bp. Two data set of unigenes covers 73.90% and 44.93% showed similarity to protein from NCBI and with Swiss protein databases, respectively.

Furthermore, 364 unigenes were intricate to oil biosynthesis and accumulation and had potential candidates for future functional genomics study. About 5710 (EST-SSR) with a density of 7.39 kb sequences were identified (Huang et al., 2016). Libault and Stacey (2010) reported a PLAC8 transcript that is linked to N_2 symbiotically, and it highly expressed in nodule tissues. It also contains the cysteine-rich region and domain for regulating cell numbers. Cytochrome P450, a highly expressed transcript, was also observed. Annotation for P450 advised involved in biosynthesis reaction and generated many biomolecules. *L. japonicus* remorin one gene is present in infected nodule cells, and overexpression of remorin one gene increases nodule number (Tooth et al., 2012).

The re-sequencing of the two-soybean species (*Glycine* max and *Glycine soja*) results identified 425 genes that were absent in *G. soja*. Along with 12 genes involved in seed development and two genes tangled in lipid metabolism (Joshi et al., 2013). Lam et al. (2010) later confirmed by sequencing and phylogenetic study that these two cultivars were from the same ancestor. A re-sequencing analysis of 90 chickpea genotypes identified 122 genes that appeared as essential genes and used in modern breeding efforts. Moreover, six genomic regions ranging from 50 to 200 kb also found. Many genes are linked with disease resistance traits and are essential for selection (Varshney et al., 2013a, 2013b, 2013c). Re-sequencing of stable mutant genotypes helps to map the mutational actions in order to link a phenotype with connecting

mutation and provide the base for forwarding genetics as a genomic tool to determine gene function. Similar work has been done by Belfield et al. (2012) in *Arabidopsis* to find the fast neutron mutagenesis that induced the deletions and single base pair replacement detected by old chip methods that have the complication highlighted. By applying to this technology, Leshchiner et al. (2012) identified approximately 2700 indel coding regions that impact mutant phenotypes, besides 17,000 SNPs. Moreover, 0.1 kb deletions also recognized among genomic regions of three mutant lines. The efficacy of the technology led to the progress of various ongoing projects and helped to consider deletions in mutant phenotypes.

1.7. RNA sequencing in legumes

RNA sequencing is an advance method for profiling the transcriptome. Applying RNA-Seq techniques provide information about gene characterization, functional genomic studies, gene expression analysis, especially when scarce information is available for the studied genome. The use of NGS, i.e., genome sequencing, RNA sequencing (Transcriptome sequencing), and DNA sequencing (Resequencing), provide deep insight for legume development and improvement program by genome evolution, architecture, and domestication studies (Orourke et al., 2014). It is a cost-effective quantification method that produces high reproducibility, high accuracy, and wide dynamic range. Modern plant omics techniques, i.e., NGS, replace the chip-based approaches for genomic studies and put an extra burden on bioinformatics, data analysis, speed, and storage. Moreover, the facility of comprehensive transcript profiling from the seeds and contrasting phenotypes could provide gene of interest and functional groups occupied in control and more important agronomic traits and at the same time provide a useful plate form from which a gene map could assemble for the species. RNA-based sequence technology has been used for genome-wide transcriptome profile across different crops, such as rice (Davidson et al., 2012), chickpea, (Mashaki et al., 2018), maize (Kakumanu et al., 2012) and *Raphanus satyus* (Arun-Chinnappa and McCurdy, 2015).

RNA (quantification) sequence technology is used to determine the gene expression profiling under specific conditions (Kaur et al., 2014). RNA-Seq can be used to determine drug response, biomarker detection, basic medical research, gene expression analysis, differential gene expression analysis, and gene ontology classification. Differential genome expansions within the *Vicia* genus are visually attributable to the amplification of large retroelement sequences (Neumann et al., 2006). RNA sequencing has been used to determine the development, stress acclimation, and nitrogen fixation is limited in legumes such as soybean (Atwood, 2014), Lupin (O'Rourke et al., 2013a, 2013b), and alfalfa (Yang et al., 2011). Furthermore, RNA-Seq has also been applied to studies on microRNA in soybeans (Turner et al., 2012) and common beans (Perez et al., 2012). Soybean genome annotation results predicted 46,430 high confidence genes, while 19,780 lower confidence genes and 90.4% gene model were transcriptionally active (Schmutz et al., 2010). Hierarchical clustering analysis from tissue and development transcript predicted three upper group tissue, lower or ground tissue, and seed tissue (Severin et al., 2010a, 2010b). Extraordinary high-throughput next-generation RNA-Seq has advantages in transcriptomics studies (Garg and Jain, 2013). Moreover, the availability of software that assembled the genome sequenced data (O'Rourke et al., 2013a, 2013b). The RNA-Seq data have collected in most of the legume crop but mainly focused on marker development in *Medicago sativa* (Han et al., 2011), chickpea (Garg et al., 2011), peanut (Zhang et al., 2012a, 2012b), lentil (Kaur et al., 2011), and common bean (Kalavacharia et al., 2011). RNA-Seq studies, used to recognize development, stress response,

composition, and nitrogen fixation have been limited to soybean (Atwood, 2014), lupin (O'Rourke et al., 2013a, 2013b), alfalfa (Yang et al., 2011), and *M. truncatula* (Boscari et al., 2013). Moreover, RNA-Seq also used to study microRNAs in legume in common bean (Perez et al., 2012), soybean (Turner et al., 2012), and *M. truncatula* (Simon et al., 2009).

One study from our lab determine the root transcriptomes through RNA-Seq under drought stress differentially expressed gene (DEGs) using faba bean genotype (Hassawi 2), and samples collected at two different stages (vegetative and flowering). Data analysis results predicted about 624.8 M total high-quality reads and 198,155 unigenes with a mean length of 738 bp. Most of the unigenes were downregulated at both stages, i.e., vegetative and reproductive stage. However, 15,366 genes were upregulated at flowering stage, while 14,097 genes upregulated at vegetative stage. At flowering stage, 20,144 down regulated while 22,737 genes down regulated at the vegetative stage.

Drought stress-responsive genes encoded various regulatory proteins (kinases, phosphatases), functional proteins with enzymes for osmoprotectant, detoxification, and transporters, transcription factors (TF), plant hormones were expressed and up-regulated. Identified DEGs were novel and showed substantial change at expression level under drought stress conditions. Consequently, the qRT-PCR results were coinciding with sequencing, which could help for functional genomics and tolerance mechanisms in crop plants (Alghamdi et al., 2018).

1.8. Computational resources for RNA-seq transcriptome analysis

Various studies on legumes have produced transcriptome data for few legumes such as *Medicago*, Soybean, Chickpea, and *Lotus* (Garg and Jain, 2013). The importance of transcriptome analyses has made the related research groups manage and make these data available to researchers to assist them in unleashing and evaluating specific transcription activity of different genes at specific developmental stages (Table 6). It also enables the various research studies to characterize and annotate these genes and define their function in legumes. NGS has accelerated the characterization and quantification of the transcriptome, which has also enhanced the developmental evolution of advanced bioinformatics software. Transcriptome analysis has comprehended as a crucial element for research in any organism (Yang and Kim, 2015). We have introduced in this review the routine software used in the pipeline of RNA-Seq, starting from designing the RNA-Seq experiment leading to quality control of sequence reads, their alignment, annotation, and quantitative expression analysis (Table 7). The software

Table 6
Legume transcriptome data repository; Web resources.

Web resource	Legume	Type of information
LegumeIP http://plantgrn.noble.org/LegumeIP/	*Medicago*, *Lotus*, Soybean	RNA-sequence data, Microarray data
MtGEA http://mtgea.noble.org/v3/	*Medicago*	Microarray data
LjGEA http://ljgea.noble.org/v2/	*Lotus*	Microarray data
CTDB http://www.nipgr.res.in/ctdb.html	Chickpea	RNA-sequence data
SoyPLEX http://www.plexdb.org/plex.php?database=Soybean	Soybean	Microarray data
SoySeq http://www.soybase.org/soyseq/	Soybean	RNA-sequence data

Table 7
Bioinformatics software packages and workbenches available for transcriptome data analysis.

Process	Package	Description	Reference
Design of RNA-Seq experiment	Scotty	Measure the differential gene expression	http://scotty.genetics.utah.edu/
	ssizeRNA	Sample size calculation	https://cran.r-project.org/web/packages/ssizeRNA/index.html
Quality control	RNAtor	Calculate optimal parameters for popular tools	https://rdrr.io/bioc/PROPER/
	fastqp	quality assessment	https://github.com/mdshw5/fastqp
	FastQC	Quality control	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
	AfterQC	Automatic, read trimming, error removing	https://github.com/OpenGene/AfterQC
	QC-Chain	Quality control	
	FASTX-Toolkit	Read trimming	http://hannonlab.cshl.edu/fastx_toolkit/
	NGS QC Toolkit	Quality control	http://www.nipgr.res.in/ngsqctoolkit.html
	PRINSEQ	Generate summary statistics of sequence	http://prinseq.sourceforge.net/manual.html
	SolexaQA	Calculates sequence quality statistics and creates visual representation	http://solexaqa.sourceforge.net/
	mRIN	Calculating mRNA integrity	http://zhanglab.c2b2.columbia.edu/index.php/MRIN
	clean_reads	Cleans NGS reads	http://bioinf.comav.upv.es/clean_reads/
	cutadapt	removes adapter sequences	https://cutadapt.readthedocs.org/en/stable/
	Deconseq	Remove contamination from sequence data	http://deconseq.sourceforge.net/
	htSeqTools	Quality control, remove over amplification artifacts	http://www.bioconductor.org/packages/2.9/bioc/html/htSeqTools.html
	SEECER	Correct Sequencing error	http://sb.cs.cmu.edu/
	UCHIME	Detect chimeric sequences	http://drive5.com/uchime
Alignment tools	Bowtie	Short reads aligner	http://bowtie-bio.sourceforge.net/index.shtml
	BWA	Map low divergent sequence against a large genome	http://bio-bwa.sourceforge.net/
	Mosaik	Short gap containing sequence aligner	http://bioinformatics.bc.edu/marthlab/Mosaik
	GNUMAP	Needleman-Wunsch algorithm-based aligner	http://dna.cs.byu.edu/gnumap/
	WHAM	HTS alignment	http://research.cs.wisc.edu/wham/
	BBMap	Align reads directly to transcriptome	http://sourceforge.net/projects/bbmap/
	STAR	Detects splice junctions	https://github.com/alexdobin/STAR
Quantitative analysis/ transcriptome reconstruction	Cufflinks	Measure global de novo transcript isoform expression	http://cufflinks.cbcb.umd.edu/
	Scripture	Transcriptome reconstruction	http://www.broadinstitute.org/software/scripture/
	RNAeXpress	Extract and annotate transcripts	http://www.rnaexpress.org/
	SLIDE	Isoform Discovery	https://sites.google.com/site/jingyijli/
	StringTie	RNA-Sequence alignment assembler into potential transcripts	http://www.ccb.jhu.edu/software/stringtie/
	Alexa-Seq	Gene expression analysis	http://www.alexaplatform.org/alexa_seq/downloads.htm
	ERANGE	Normalize and quantify genes	http://woldlab.caltech.edu/rnaseq
	GFOLD	Ranking differentially expressed genes	http://compbio.tongji.edu.cn/~fengjx/GFOLD/gfold.html
	NEUMA	RNA abundance estimator based on mRNA isoforms	http://neuma.kobic.re.kr/
	SpliceSEQ	Alternative mRNA splicing patterns investigator	http://bioinformatics.mdanderson.org/main/SpliceSeq:Overview
miRNA prediction and analysis	miRDeep-P	miRNA analysis	http://faculty.virginia.edu/lilab/miRDP/
	miRPlant	miRNA analysis	https://sourceforge.net/projects/mirplant/
Freely available workbenches	BioJupies		http://biojupies.cloud/
	Galaxy	Analysis pipeline	http://galaxyproject.org/
	RobiNA	Analysis pipeline	http://mapman.gabipd.org/web/guest/robin
	NGSUtils	Analysis pipeline	http://ngsutils.org/
	MeV	Analysis pipeline	

mentioned in Table 7 has a description of the use of their own software at a specific stage of process and reference to online sources.

2. Conclusion

As the human population increases, the possibility of food shortage also increases. It is necessary to augment the effort of developing high yielding lines and cultivars against various abiotic stresses. Classical breeding techniques may help to improve legume crop traits in terms of quality, nutrition, and production but not at the required rate. Cost-effective and improved molecular breeding techniques help scientists to make better genomes of legume crops. Speed up processing, genomic techniques such as NGS and RNA sequencing (transcriptomics) have developed for gene identification, mapping, and construction of gene maps, which may help to improve the traits in many crop plants. In this review, we have focused on the classical breeding methods used for trait introgression and on new, improved techniques, such as proteomics, metabolomics, RNA sequencing, and re-sequencing for trait, development of molecular markers, gene discovery for understanding the metabolomics pathways, and GOIs for crop-specific genome to achieve higher gains in short time. These improved technologies will contribute to the development and improvement of legume crops and breeding programs, with the advantage of being budget-effective, environment-friendly, and, most importantly, less time-consuming.

References

Ahsan, N., Nanjo, Y., Sawada, H., Kohno, Y., Komatsu, S., 2010. Ozone stress-induced proteomic changes in leaf total soluble and chloroplast proteins of soybean reveal that carbon allocation is involved in adaptation in the early developmental stage. Proteomics 10, 2605–2619. https://doi.org/10.1002/pmic.201000180.

Alghamdi, S.S., Khan, M.A., Ammar, M.H., Sun, Q., Huang, L., Migdadi, H.M., El-Harty, E.H., Al-Faifi, S.A., 2018. Characterization of drought stress-responsive root transcriptome of faba bean (Vicia faba L.) using RNA sequencing. 3. Biotechnology 8 (12), 502.

Arahana, V.S., Graef, G.L., Specht, J.E., Steadmen, J.R., Eskridge, K.M., 2001. Identification of QTLs for resistance of Sclerotinia sclerotiorum in soybean. Crop Sci. 41, 180–188.

Arbaoui, M., Link, W., 2008. Effect of hardening on frost tolerance and fatty acid composition of leaves and stems of a set of fababean (Vicia faba L.) genotypes. Euphytica 162, 211–219.

Arun-Chinnappa, K.S., McCurdy, D.W., 2015. De novo assembly of a genome-wide transcriptome map of Vicia faba (L.) for transfer cell research. Front. Plant Sci. 6, 217.

Atienza, S.G., Palomino, C., Gutiérrez, N., Alfaro, C.M., Rubiales, D., Torres, A.M., Ávila, C.M., 2016. QTLs for ascochyta blight resistance in faba bean (*Vicia faba* L.): validation in field and controlled conditions. Crop Pasture Sci. 67 (2), 216–224.

Atwood, S.E., O'Rourke, J.A., Peiffer, G.A., Yin, T., Majumder, M., Zhang, C., Cianzio, S. R., Hill, J.H., Cook, D., Whitham, S.A., Shoemaker, R.C., 2014. Replication protein A subunit 3 and the iron efficiency response in soybean. Plant, Cell Environ. 37 (1), 213–234.

Avila, C.M., Sillero, J.C., Rubiales, D., Moreno, M.T., Torres, A.M., 2003. Identification of RAPD markers linked to the Uvf-1 gene conferring hypersensitive resistance against rust (Uromyces viciae-fabae) in Vicia faba L. Theoret. Appl. Genet. 107 (2), 353–358.

Avila, C.M., Atienza, S.G., Moreno, M.T., Torres, A.M., 2007. Development of a new diagnostic marker for growth habit selection in faba bean (Vicia faba L.) breeding. Theoret. Appl. Genet. 115 (8), 1075.

Bachman, M.S., Tamulonis, J.P., Nickell, C.D., Bent, A.F., 2001. Molecular markers linked to brown stem rot resistance genes, Rbs1 and Rbs2, in soybean. Crop Sci. 41, 527–535.

Belfield, E.J., Gan, X., Mithani, A., Brown, C., Jiang, C., Franklin, K., Alvey, E., Wibowo, A., Jung, M., Bailey, K., Kalwani, S., 2012. Genome-wide analysis of mutations in mutant lineages selected following fast-neutron irradiation mutagenesis of Arabidopsis thaliana. Genome Res. 22 (7), 1306–1315.

Berbel, A., Ferrándiz, C., Hecht, V., Dalmais, M., Lund, O.S., Sussmilch, F.C., Taylor, S. A., Bendahmane, A., Ellis, T.H.N., Beltrán, J.P., Weller, J.L., Madueño, F., 2012. VEGETATIVE1 is essential for development of the compound inflorescence in pea. Nat. Commun. 3, 1–8. https://doi.org/10.1038/ncomms1801.

Bertioli, D.J., Cannon, S.B., Froenicke, L., Huang, G., Farmer, A.D., Cannon, E.K., Liu, X., Gao, D., Clevenger, J., Dash, S., Ren, L., 2015. The genome sequences of Arachis duranensis and Arachis ipaensis, the diploid ancestors of cultivated peanut. Nature Genet. 48 (4), 438.

Biazzi, E., Carelli, M., Tava, A., Abbruscato, P., Losini, I., Avato, P., Scotti, C., Calderini, A., 2015. CYP72A67 catalyzes a key oxidative step in Medicago truncatula hemolytic saponin biosynthesis. Molecular Plant 8 (10), 1493–1506. https://doi.org/10.1016/j.molp.2015.06.003.

Boscari, A., del Giudice, J., Ferrarini, A., Venturini, L., Zaffini, A.L., Delledonne, M., Puppo, A., 2013. Expression dynamics of the Medicago truncatula transcriptome during the symbiotic interaction with Sinorhizobium meliloti: which role for nitric oxide?. Plant Physiol. 161 (1), 425–439.

Brosowska-Arendt, W., Gallardo, K., Sommerer, N., Weidner, S., 2014. Changes in the proteome of pea (Pisum sativum L.) seeds germinating under optimal and osmotic stress conditions and subjected to post-stress recovery. Acta Physiol. Plant 36, 795–807. https://doi.org/10.1007/s11738-013-1458-8.

Bybee, S.M., Bracken-Grissom, H., Haynes, B.D., Hermansen, R.A., Byers, R.L., Clement, M.J., Udall, J.A., Wilcox, E.R., Crandall, K.A., 2011. Targeted amplicon sequencing (TAS): a scalable next-gen approach to multilocus, multitaxa phylogenetics. Genome Biol. Evolut. 3, 1312–1323.

Catoira, R., Galera, C., de Billy, F., Penmetsa, R.V., Journet, E.P., Maillet, F., Rosenberg, C., Cook, D., Gough, C., Dénarié, J., 2000. Four genes of Medicago truncatula controlling components of a Nod factor transduction pathway. Plant Cell 12, 1647–1665. https://doi.org/10.1105/tpc.12.9.1647.

Chen, J., Yu, J., Ge, L., Wang, H., Berbel, A., Liu, Y., Chen, Y., Li, G., Tadege, M., Wen, J., Cosson, V., Mysore, K.S., Ratet, P., Madueño, F., Bai, G., Chen, R., 2010. Control of dissected leaf morphology by a Cys(2) His(2) zinc finger transcription factor in the model legume Medicago truncatula. PNAS 107 (23), 10574–10759. https://doi.org/10.1073/pnas.1003954107.

Chen, J., Moreau, C., Liu, Y., Kawaguchi, M., Hofer, J., Ellis, N., Chen, R., 2012. Conserved genetic determinant of motor organ identity in Medicago truncatula and related legumes. PNAS 109 (23), 11723–11728. https://doi.org/10.1073/pnas.1204566109.

Collard, B.C., Mackill, D.J., 2007. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos. Trans. R. Soc. B: Biol. Sci. 363 (1491), 557–572.

Collard, B.C.Y., Mackill, D.J., 2008. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos. Transact. R. Soc. B 363, 557–572.

Concibido, V.C., Lange, D.A., Denny, R.L., Orf, J.H., Young, N.D., 1997. Genome mapping of soybean cyst nematode resistance genes in "Peking", PI 90763, and PI 88788 using DNA markers. Crop Sci. 37, 258–264.

Davidson, R.M., Gowda, M., Moghe, G., Lin, H., Vaillancourt, B., Shiu, S.H., Buell, C.R., 2012. Comparative transcriptomics of three Poaceae species reveals patterns of gene expression evolution. Plant J. 71 (3), 492–502.

Desai, A.S., Rao, S., 2014. Effect of gamma radiation on germination and physiological aspects of pigeon pea (Cajanus cajan L. Mill sp.) seedlings. IMPACT: Int. J. Res. Appl. Natural Social Sci. 2 (6), 47–52.

Dhanasekar, P., Reddy, K.S., 2015. A novel mutation in TFL1 homolog affecting determinacy in cowpea (Vigna unguiculata). Mol. Genet. Genomics 290 (1), 55–65. https://doi.org/10.1007/s00438-014-0899-0.

Dias, D.A., Hill, C.B., Jayasinghe, N.S., Atieno, J., Sutton, T., Roessner, U., 2015. Quantitative profiling of polar primary metabolites of two chickpea cultivars with contrasting responses to salinity. J. Chromatogr. B 1000, 1–13. https://doi.org/10.1016/j.jchromb.2015.07.002.

Diaz, R., Šatović, Z., Roman, B., Avila, C.M., Alfaro, C., Rubiales, D., Cubero, J.I., Torres, A.M., 2005. QTL mapping for ascochyta blight resistance in faba bean. Second Congress of Croatian Geneticists.

Doerfler, H., Sun, X., Wang, L., Engelmeier, D., Lyon, D., Weckwerth, W., 2014. mzGroupAnalyzer-Predicting pathways and novel chemical structures from untargeted high-throughput metabolomics data. PLoS ONE 9, e96188. https://doi.org/10.1371/journal.pone.0096188.

Du, W., Wang, M., Fu, S., Yu, D., 2009. Mapping QTLs for seed yield and drought susceptibility index in soybean (Glycine max L.) across different environments. J. Genet. Genom. 36 (12), 721–731.

FAO/IAEA Mutant Varieties Database. http://www.mvd.iaea.org (accessed on August 12, 2018)..

Fedoroff, N.V., 2015. Food in a future of 10 billion. Agric. Food Security 4 (1), 11.

Foucher, F., Morin, J., Courtiade, J., Cadioux, S., Ellis, N., Rameau, C., 2003. Determinate and late flowering are two terminal flower1/centroradialis homologs that control two distinct phases of flowering initiation and development in pea. Plant Cell 15, 2742–2754. https://doi.org/10.1105/tpc.015701.

Francia, E., Tacconi, G., Crosatti, C., Barabaschi, D., Bulgarelli, D., Dall'Aglio, E., Valè, G., 2005. Marker assisted selection in crop plants. Plant Cell, Tissue Organ Cult. 82 (3), 317–342.

Garg, R., Jain, M., 2013. Transcriptome Analyses in Legumes: A Resource for Functional Genomics, (June). http://doi.org/10.3835/plantgenome2013.04.0011..

Garg, R., Patel, R.K., Jhanwar, S., Priya, P., Bhattacharjee, A., Yadav, G., Bhatia, S., Chattopadhyay, D., Tyagi, A.K., Jain, M., 2011. Gene discovery and tissue-specific transcriptome analysis in chickpea with massively parallel pyrosequencing and web resource development. Plant Physiol., pp. 111.178616.

Gavioli, E.A., 2011. Molecular markers: assisted selection in soybeans. Soybean-Genetics and Novel Techniques for Yield Enhancement. IntechOpen.

Gerland, P., Raftery, A.E., Ševčíková, H., Li, N., Gu, D., Spoorenberg, T., Alkema, L., Fosdick, B.K., Chunn, J., Lalic, N., Bay, G., 2014. World population stabilization unlikely this century. Science 346 (6206), 234–237.

Girija, M., Dhanavel, D., Gnanamurthy, S., 2013. Gamma rays and EMS induced flower color and seed mutants in cowpea (Vigna unguiculata L. Walp). Adv. Appl. Sci. Res. 4, 134–139.

Gnanasambandam, A., Paull, J., Torres, A., Kaur, S., Leonforte, T., Li, H., Materne, M., 2012. Impact of molecular technologies on faba bean (Vicia faba L.) breeding strategies. Agronomy 2 (3), 132–166.

Gupta, P., Kumar, J., Mir, R., Kumar, A., 2010. 4 Marker-assisted selection as a component of conventional plant breeding. Plant Breeding Rev. 33, 145.

Gutierrez, N., Avila, C., Duc, G., Marget, P., Suso, M., Moreno, M., Torres, A., 2006. CAPs markers to assist selection for low vicine and convicine contents in faba bean (Vicia faba L.). Theor. Appl. Genet. 114 (1), 59–66.

Gutierrez, N.C.M., Avila, C., Rodriguez-Suarez, M.T., Moreno, A.M. Torres, 2007. Development of SCAR markers linked to a gene controlling absence of tannins in fababean. Mol. Breed. 19 (2007), 305–314.

Han, Y., Kang, Y., Torres-Jerez, I., Cheung, F., Town, C.D., Zhao, P.X., Udvardi, M.K., Monteros, M.J., 2011. Genome-wide SNP discovery in tetraploid alfalfa using 454 sequencing and high resolution melting analysis. BMC Genom. 12 (1), 350.

Heidarvand, L., Maali-Amiri, R., 2013. Physio-biochemical and proteome analysis of chickpea in early phases of cold stress. J. Plant Physiol. 170, 459–469. https://doi.org/10.1016/j.jplph.2012.11.021.

Hernández, G., Ramírez, M., Valdés-López, O., Tesfaye, M., Graham, M.A., Czechowski, T., Schlereth, A., Wandrey, M., Erban, A., Cheung, F., Wu, H.C., 2007. Phosphorus stress in common bean: root transcript and metabolic responses. Plant Physiol. 144 (2), 752–767.

Hernández, G., Valdés-López, O., Ramírez, M., Goffard, N., Weiller, G., Aparicio-Fabre, R., Udvardi, M.K., 2009. Global changes in the transcript and metabolic profiles during symbiotic nitrogen fixation in phosphorus-stressed common bean plants. Plant Physiol. 151 (3), 1221–1238.

Hiremath, P.J., Farmer, A., Cannon, S.B., Woodward, J., Kudapa, H., Tuteja, R., Kumar, A., BhanuPrakash, A., Mulaosmanovic, B., Gujaria, N., 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 (8), 922–931.

Hong, J., Yang, L., Zhang, D., Shi, J., 2016. Plant metabolomics: an indispensable system biology tool for plant science. Int. J. Mol. Sci. 17 (6), 767.

Hossain, Z., Komatsu, S., 2014. Potentiality of soybean proteomics in untying the mechanism of flood and drought stress tolerance. Proteomes 2, 107–127. https://doi.org/10.3390/proteomes2010107.

Hu, J., Rampitsch, C., Bykova, N.V., 2015. Advances in plant proteomics toward improvement of crop productivity and stress resistance. Front. Plant Sci. 6, 209. https://doi.org/10.3389/fpls.2015.00209.

Huang, J., Guo, X., Hao, X., Zhang, W., Chen, S., Huang, R., Gresshoff, P.M., Zheng, Y., 2016. De novo sequencing and characterization of seed transcriptome of the tree legume Millettia pinnata for gene discovery and SSR marker development. Mol. Breed. 36 (6), 75.

International, R.G.S.P., 2005. The map-based sequence of the rice genome. Nature 436 (7052), 793.

Irar, S., González, E.M., Arrese-Igor, C., Marino, D., 2014. A proteomic approach reveals new actors of nodule response to drought in split-root grown pea plants. Physiol. Plant 152, 634–645. https://doi.org/10.1111/ppl.12214.

Jacoby, R.P., Millar, A.H., 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. https://doi.org/10.3389/fpls.2013.00020.

Jaiswal, D.K., Mishra, P., Subba, P., Rathi, D., Chakraborty, S., Chakraborty, N., 2014. Membrane-associated proteomics of chickpea identifies Sad1/UNC-84 protein (CaSUN1), a novel component of dehydration signaling. Sci. Rep. 4, 4177.

Jorrin-Novo, J.V., Komatsu, S., Sanchez-Lucas, R., de Francisco, L.E.R., 2019. Gel electrophoresis-based plant proteomics: past, present, and future. Happy 10th anniversary Journal of Proteomics!. J. Proteomics 198, 1–10.

Joshi, T., Valliyodan, B., Wu, J.-H., Lee, S.-H., Xu, D., Nguyen, H.T. (2013). Genomic differences between cultivated soybean, G. max and its wild relative G. soja. In: Paper Presented at the BMC Genomics..

Kakumanu, A., Ambavaram, M.M., Klumas, C., Krishnan, A., Batlang, U., Myers, E., Pereira, A., 2012. Effects of drought on gene expression in maize reproductive and leaf meristem tissue revealed by RNA-Seq. Plant Physiol. 160 (2), 846–867.

Kalavacharia, V., Liu, Z., Meyers, B.C., Thimmapuram, J., Melmaiee, K., 2011. Identification and analysis of common bean (Phaseolus vulgaris L.) transcriptomes by massively parallel pyrosequencing. BMC Plant Biol. 11, 135–1107.

Katam, K., Jones, K.A., Sakata, K., 2015. Advances in proteomics and bioinformatics in agriculture research and crop improvement. J. Proteomics Bioinform. 8, 39–48. https://doi.org/10.1016/j.jprot.2013.05.036.

Kaur, S., Cogan, N.O., Pembleton, L.W., Shinozuka, M., Savin, K.W., Materne, M., Forster, J.W., 2011. Transcriptome sequencing of lentil based on second-generation technology permits large-scale unigene assembly and SSR marker discovery. BMC Genom. 12 (1), 265.

Kiezun, A., Garimella, K., Do, R., Stitziel, N.O., Neale, B.M., McLaren, P.J., Moran, J.L., 2012. Exome sequencing and the genetic basis of complex traits. Nat. Genet. 44 (6), 623.

Kaur, G., Dutta, D., Thakur, K.G., 2014. Crystal structure of Mycobacterium tuberculosis CarD, an essential RNA polymerase binding protein, reveals a quasidomain-swapped dimeric structural architecture. Proteins 82, 879–884.

Kodyma, A., Afza, R., Forster, B.P., Ukai, Y., Nakagawa, H., Mba, C., 2011. Methodology for phisical and chemical mutagenic treatment. In: Shu, Q., Forster, B.P., Nakagawa, H. (Eds.), Plant Mutation Breeding and Biotechnology. FAO, Rome, Italy, pp. 169–180.

Komatsu, S., Yamamoto, A., Nakamura, T., Nouri, M.Z., Nanjo, Y., Nishizawa, K., Furukawa, K., 2011. Comprehensive analysis of mitochondria in roots and hypocotyls of soybean under flooding stress using proteomics and metabolomics techniques. J. Proteome Res. 10 (9), 3993–4004.

Kumar, V., Singh, A., Mithra, S.A., Krishnamurthy, S., Parida, S.K., Jain, S., Sharma, S., 2015. Genome-wide association mapping of salinity tolerance in rice (Oryza sativa). DNA Res. 22 (2), 133–145.

Lakhssassi, N., Zhou, Z., Liu, S., Colantonio, V., Ghazaleh, A.A., Meksem, K., 2017. Characterization of the FAD2 gene family in soybean reveals the limitations of gel-based TILLING in genes with high copy number. Front. Plant Sci. 8 (324), 1–15. https://doi.org/10.3389/fpls.2017.00324.

Lam, H.-M., Xu, X., Liu, X., Chen, W., Yang, G., Wong, F.-L., Wang, B., 2010. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. Nat. Genet. 42 (12), 1053.

Langridge, P., Fleury, D., 2011. Making the most of 'omics' for crop breeding. Trends Biotechnol. 29, 33–40. https://doi.org/10.1016/j.tibtech.2010.09.006.

Larrainzar, E., Wienkoop, S., Scherling, C., Kempa, S., Ladrera, R., Arrese-Igor, C., Weckwerth, W., González, E.M., 2009. Carbon metabolism and bacteroid functioning are involved in the regulation of nitrogen fixation in Medicago truncatula under drought and recovery. Mol. Plant-Microbe Interact. 22 (12), 1565–1576.

Lee, K.J., Kim, D.S., Kim, J.B., Jo, S.H., Kang, S.Y., Choi, H.I., Ha, B.K., 2016. Identification of candidate genes for an early-maturing soybean mutant by genome resequencing analysis. Mol. Genet. Genom. 291 (4), 1561–1571. https://doi.org/10.1007/s00438-016-1183-2.

Leshchiner, I., Alexa, K., Kelsey, P., Adzhubei, I., Austin-Tse, C.A., Cooney, J.D., Anderson, H., King, M.J., Stottmann, R.W., Garnaas, M.K., Ha, S., 2012. Mutation mapping and identification by whole-genome sequencing. Genome Res. 22 (8), 1541–1548.

Li, Z., Jiang, L., Ma, Y., Wei, Z., Liu, Z., Hong, H., Liu, Z., Lei, J., Liu, Y., Guan, R., Guo, Y., Jin, L., Zhang, L., Li, Y., Ren, Y., He, W., Liu, M., Htwe, N., Liu, L., Guo, B., Song, J., Tan, B., Liu, G., Li, M., Zhang, X., Liu, B., Shi, X., Han, S., Hua, S., Zhou, F., Yu, L., Li, Y., Wang, S., Wang, J., Chang, R., Qiu, L., 2017. Development and utilization of a new chemically induced soybean library with a high mutation density. J. Integr. Plant Biol. 59 (1), 60–74. https://doi.org/10.1111/jipb.12505.

Libault, M., Stacey, G., 2010. Evolution of FW2.2-like (FWL) and PLAC8 genes in eukaryotes. Plant Signal. Behav. 5, 1226–1228.

Lorenz, A.J., Chao, S., Asoro, F.G., Heffner, E.L., Hayashi, T., Iwata, H., Smith, K.P., Sorrells, M.K., Jannink, J.L., 2011. Genomic selection in plant breeding: knowledge and prospects. Adv. Agron. 110, 77–123.

Mashaki, K.M., Garg, V., Ghomi, A.A.N., Kudapa, H., Chitikineni, A., Nezhad, K.Z., Thudi, M., 2018. RNA-Seq analysis revealed genes associated with drought stress response in kabuli chickpea (Cicer arietinum L.). PLoS ONE 13 (6), e0199774.

McCouch, S., Baute, G.J., Bradeen, J., Bramel, P., Bretting, P.K., Buckler, E., Burke, J.M., Charest, D., Cloutier, S., Cole, G., Dempewolf, H., 2013. Agriculture: feeding the future. Nature 499 (7456), 23.

Meksem, K., Zobrist, K., Ruben, E., Hyten, D., Quanzhou, T., Zhang, H.B., Lightfoot, D. A., 2000. Two large-insert soybean genomic libraries constructed in a binary vector: application in chromosome walking and genome wide physical mapping. Theor. Appl. Genet. 101, 747–755.

Michael, T.P., Jackson, S., 2013. The first 50 plant genomes. Plant Genome 6. https://doi.org/10.3835/plantgenome2013.03.0001in.

Mikic, A., Perić, V., Đorđević, V., Srebrić, M., Mihailović, V., 2009. Anti-nutritional factors in some grain legumes. Biotechnol. Anim. Husbandry 25 (5-6-2), 1181–1188.

Molina, C., Rotter, B., Horres, R., Udupa, S.M., Besser, B., Bellarmino, L., Baum, M., Matsumura, H., Terauchi, R., Kahl, G., Winter, P., 2008. Super SAGE: the drought stress-responsive transcriptome of chickpea roots. BMC Genomics 9, 553.

Moreau, C., Hofer, J.M.I., Eléouët, M., Sinjushin, A., Ambrose, M., Skøt, K., Blackmore, T., Swain, M., Hegarty, M., Balanzà, V., Ferrándiz, C., Ellis, T.H.N., 2018. Identification of Stipules reduced, a leaf morphology gene in pea (Pisum sativum). New Phytol. 220 (1), 288–299. https://doi.org/10.1111/nph.15286.

Morozova, O., Hirst, M., Marra, M.A., 2009. Applications of new sequencing technologies for transcriptome analysis. Annu. Rev. Genomics Hum. Genet. 10, 135–151.

Naj, A.C., Jun, G., Beecham, G.W., Wang, L.-S., Vardarajan, B.N., Buros, J., Crane, P.K., 2011. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. Nat. Genet. 43 (5), 436.

Nayak, S.N., Zhu, H., Varghese, N., Datta, S., Choi, H.K., Horres, R., et al., 2010. Integration of novel SSR and gene-based SNP marker loci in the chickpea genetic map and establishment of new anchor points with Medicago truncatula genome. Theor. Appl. Genet. 120, 1415–1441. https://doi.org/10.1007/s00122-010-1265-1.

Neumann, P., Koblížková, A., Navrátilová, A., Macas, J., 2006. Significant expansion of Vicia pannonica genome size mediated by amplification of a single type of giant retroelement. Genetics 173 (2), 1047–1056.

Njiti, V.N., Meksem, K., Iqbal, J., Johnson, J.E., Kassem, M.A., Zobrist, F., Kilo, V.Y., Lightfoot, D.A., 2002. Common loci underlie fiels resistance to soybean sudden death syndrome in Forrest, Pyramid, Essex, and Douglas. Theor. Appl. Genet. 104, 294–300.

Nouri, M., Komatsu, S., 2010. Comparative analysis of soybean plasma membrane proteins under osmotic stress using gel-based and LC MS/MS-based proteomics approaches. Proteomics 10, 1930–1945. https://doi.org/10.1002/pmic.200900632.

O'Rourke, J.A., Yang, S.S., Miller, S.S., et al., 2013b. An RNA-Seq transcriptome analysis of orthophosphate-deficient white lupin reveals novel insights into phosphorus acclimation in plants. Plant Physiol. 161, 705–724.

O'Rourke, J.A., Yang, S.S., Miller, S.S., Bucciarelli, B., Liu, J., Rydeen, A., Allan, D., 2013. An RNA-Seq transcriptome analysis of orthophosphate-deficient white lupin reveals novel insights into phosphorus acclimation in plants. Plant Physiol. 161 (2), 705–724.

Orourke, J.A., Bolon, Y.-T., Bucciarelli, B., Vance, C.P., 2014. Legume genomics: understanding biology through DNA and RNA sequencing. Ann. Bot. 113 (7), 1107–1120.

Pandey, M.K., Roorkiwal, M., Singh, V.K., Ramalingam, A., Kudapa, H., Thudi, M., Varshney, R.K., 2016. Emerging genomic tools for legume breeding: current status and future prospects. Front. Plant Sci. 7, 455.

Parween, S., Nawaz, K., Roy, R., Pole, A.K., Suresh, B.V., Misra, G., Jain, M., Yadav, G., Parida, S.K., Tyagi, A.K., Bhatia, S., 2015. An advanced draft genome assembly of a desi type chickpea (Cicer arietinum L.). Sci. Rep. 5, 12806.

Perez-de-Castro, M.A., Vilanova, S., Cañizares, J., Pascual, L., Blanca, J.M., Diez, M.J., Prohens, J., Picó, B., 2012. Application of genomic tools in plant breeding. Current Genom. 13 (3), 179–195.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., 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,. https://doi.org/10.1371/journal.pone.0037135 e37135.

Poland, J.A., Brown, P.J., Sorrells, M.E., 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. https://doi.org/10.1371/journal.pone.0032253.

Purushothaman, R., Upadhyaya, H., Gaur, P., Gowda, C., Krishnamurthy, L., 2014. Kabuli and desi chickpeas differ in their requirement for reproductive duration. Field Crops Res. 163, 24–31.

Ragot, M., Lee, M., Guimaraes, E., 2007. Marker-assisted selection in maize: current status, potential, limitations and perspectives from the private and public sectors. Marker-assisted selection, Current status and future perspectives in crops, livestock, forestry and fish, pp. 117–150..

Ramalingam, A., Kudapa, H., Pazhamala, L.T., Weckwerth, W., Varshney, R.K., 2015. Proteomics and metabolomics: two emerging areas for legume improvement. Front. Plant Sci. 6, 1116.

Rietzschel, E., De Buyzere, M., 2012. High-sensitive C-reactive protein: universal prognostic and causative biomarker in heart disease?. Biomarkers Med. 6 (1), 19–34.

Romero-Rodríguez, M.C., Pascual, J., Valledor, L., Jorrín-Novo, J., 2014. Improving the quality of protein identification in non-model species. Characterization of Quercus ilex seed and Pinus radiata needle proteomes by using SEQUEST and custom databases. J. Proteomics 105, 85–91. https://doi.org/10.1016/j.jprot.2014.01.027.

Ruperao, P., Chan, C.K., Azam, S., Karafiátová, M., Hayashi, S., Čížková, 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. https://doi.org/10.1111/pbi.12182.

Sanchez, D.H., Pieckenstain, F.L., Escaray, F., Erban, A., Kraemer, U.T.E., Udvardi, M.K., Kopka, J., 2011. Comparative ionomics and metabolomics in extremophile and glycophytic Lotus species under salt stress challenge the metabolic pre-adaptation hypothesis. Plant, Cell Environ. 34 (4), 605–617.

Sangsiri, C., Sorajjapinun, W., Srinivesc, P., 2005. Gamma radiation induced mutations in mungbean. Science Asia 31, 251–255. https://doi.org/10.2306/scienceasia1513-1874.2005.31.251.

Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., Kato, T., Nakao, M., Sasamoto, S., Watanabe, A., Ono, A., Kawashima, K., Fujishiro, T., 2008. Genome structure of the legume, Lotus japonicus. DNA Res. 15 (4), 227–239.

Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D.L., Song, Q., Thelen, J.J., Cheng, J., Xu, D., 2010. Genome sequence of the palaeopolyploid soybean. Nature 463 (7278), 178.

Schwarze, K., Buchanan, J., Taylor, J.C., et al., 2018. Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature. Genet. Med. 20, 1122–1130. https://doi.org/10.1038/gim.2017.247.

Seminario, A., Song, L., Zulet, A., Nguyen, H.T., González, E.M., Larrainzar, E., 2017. Drought stress causes a reduction in the biosynthesis of ascorbic acid in soybean plants. Front. Plant Sci. 8, 1042.

Senkler, M., Braun, H.P., 2012. Functional annotation of 2D protein maps: The GelMap portal. Front. Plant Sci. 14, 87. https://doi.org/10.3389/fpls.2012.00087.

Severin, A.J., Woody, J.L., Bolon, Y.T., Joseph, B., Diers, B.W., Farmer, A.D., Muehlbauer, G.J., Nelson, R.T., Grant, D., Specht, J.E., Graham, M.A., 2010b. RNA-Seq Atlas of Glycine max: a guide to the soybean transcriptome. BMC plant biology 10 (1), 160.

Severin, A.J., Peiffer, G.A., Xu, W.W., Hyten, D.L., Bucciarelli, B., O'Rourke, J.A., Bolon, Y.T., Grant, D., Farmer, A.D., May, G.D., Vance, C.P., 2010a. An integrative approach to genomic introgression mapping. Plant physiology 154 (1), 3–12.

Simon, S.A., Meyers, B.C., Sherrier, D.J., 2009. MicroRNAs in the rhizobia legume symbiosis. Plant Physiol. 151, 1002–1008.

Singh, V.K., Khan, A.W., Saxena, R.K., Kumar, V., Kale, S.M., Sinha, P., Chitikineni, A., Pazhamala, L.T., Garg, V., Sharma, M., Sameer Kumar, C.V., 2016. Next-generation sequencing for identification of candidate genes for Fusarium wilt and sterility mosaic disease in pigeonpea (C ajanus cajan). Plant Biotechnol. J. 14 (5), 1183–1194.

Song, Q.J., Marek, L.F., Shoemaker, R.C., Lark, K.G., Concibido, V.C., Delannay, X., Specht, James E., Cregan, P.B., 2004. A new integrated genetic linkage map of the soybean. Theoret. Appl. Genet. 109 (1), 122–128.

Staudinger, C., Mehmeti, V., Turetschek, R., Lyon, D., Egelhofer, V., Wienkoop, S., 2012. Possible role of nutritional priming for early salt and drought stress responses in Medicago truncatula. Front. Plant Sci. 3, 285. https://doi.org/10.3389/fpls.2012.00285.

Tang, H., Krishnakumar, V., Bidwell, S., Rosen, B., Chan, A., Zhou, S., Gentzbittel, L., Childs, K.L., Yandell, M., Gundlach, H., Mayer, K.F.X., Schwartz, D.C., Town, C.D., 2014. An improved genome release (version Mt4.0) for the model legume Medicago truncatula. BMC Genomics 15 (312), 1–14. https://doi.org/10.1186/1471-2164-15-312.

The Arabidopsis Genome Initiative, 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408, 796–815. https://doi.org/10.1038/35048692.

Thompson, M.J., 2014. High-throughput SNP genotyping to accelerate crop improvement. Plant Breed. Biotechnol. 2, 195–212. https://doi.org/10.9787/PBB.2014.2.3.195.

Torres, A., Avila, C., Gutierrez, N., Palomino, C., Moreno, M., Cubero, J., 2010. Marker-assisted selection in faba bean (Vicia faba L.). Field Crops Research 115 (3), 243–252.

Tóth, K., Stratil, T.F., Madsen, E.B., Ye, J., Popp, C., Antolín-Llovera, M., Grossmann, C., Jensen, O.N., Schüßler, A., Parniske, M., Ott, T., 2012. Functional domain analysis of the remorin protein LjSYMREM1 in Lotus japonicus. PloS ONE 7 (1), e30817.

Turner, M., Yu, O., Subramanian, S., 2012. Genome organization and characteristics of soybean microRNAs. BMC Genomics 13, 169.

Unamba, C.I., Nag, A., Sharma, R.K., 2015. Next generation sequencing technologies: the doorway to the unexplored genomics of non-model plants. Front. Plant Sci. 6, 1074.

Unterseer, S., Bauer, E., Haberer, G., Seidel, M., Knaak, C., Ouzunova, M., Meitinger, T., Strom, T.M., Fries, R., Pausch, H., Bertani, C., 2014. A powerful tool for genome analysis in maize: development and evaluation of the high density 600 k SNP genotyping array. BMC Genom. 15 (1), 823.

Vadivel, A., Kumaran, A., 2015. Gel based proteomics in plants: time to move on from the tradition. Front. Plant Sci. 6, 369.

Van Orsouw, N.J., Hogers, R.C., Janssen, A., Yalcin, F., Snoeijers, S., Verstege, E., Verstegen, H., 2007. Complexity reduction of polymorphic sequences (CRoPS™): a novel approach for large-scale polymorphism discovery in complex genomes. PLoS ONE 2 (11), e1172.

Vanderschuren, H., Lentz, E., Zainuddin, I., Gruissem, W., 2013. Proteomics of model and crop plant species: status, current limitations and strategic advances for crop improvement. J. Proteomics 93, 5–19. https://doi.org/10.1016/j.jprot.2013.05.036.

Varshney, R.K., Hiremath, P.J., Lekha, P., Kashiwagi, J., Balaji, J., Deokar, A.A., Vadez, V., Xiao, Y., Srinivasan, R., Gaur, P.M., Siddique, K.H.M., Town, C.D., Hoisington, D.A., 2009. Comprehensive resource of drought- and salinity responsive ESTs for gene discovery and marker development in chickpea (Cicer arietinum L.). BMC Genomics 10 (523).

Varshney, R.K., 2016. Exciting journey of 10 years from genomes to fields and markets: some success stories of genomics-assisted breeding in chickpea, pigeon pea and groundnut. Plant Sci. 242, 98–107. https://doi.org/10.1016/j.plantsci.2015.09.009.

Varshney, R.K., Chen, W., Li, Y., Bharti, A.K., Saxena, R.K., Schlueter, J.A., Donoghue, M.T., Azam, S., Fan, G., Whaley, A.M., Farmer, A.D., 2012. Draft genome sequence of pigeonpea (Cajanus cajan), an orphan legume crop of resource-poor farmers. Nature Biotechnol. 30 (1), 83.

Varshney, R.K., Mohan, S.M., Gaur, P.M., Gangarao, N., Pandey, M.K., Bohra, A., Janila, P., 2013b. Achievements and prospects of genomics-assisted breeding in three legume crops of the semi-arid tropics. Biotechnol. Adv. 31 (8), 1120–1134.

Varshney, R.K., Mohan, S.M., Gaur, P.M., Gangarao, N.V.P.R., Pandey, M.K., Bohra, A., Sawargaonkar, S.L., Chitikineni, A., Kimurto, P.K., Janila, P., Saxena, K.B., 2013c. Achievements and prospects of genomics-assisted breeding in three legume crops of the semi-arid tropics. Biotechnology advances 31 (8), 1120–1134.

Varshney, R.K., Song, C., Saxena, R.K., Azam, S., Yu, S., Sharpe, A.G., Cannon, S., Baek, J., Rosen, B.D., Tar'an, B., Millan, T., 2013a. Draft genome sequence of chickpea (Cicer arietinum) provides a resource for trait improvement. Nature Biotechnol. 31 (3), 240.

Varshney, R.K., Terauchi, R., McCouch, S.R., 2014. Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. PLoS Biol. 12, e1001883. https://doi.org/10.1371/journal.pbio.10 01883.

Varshney, R.K., Saxena, R.K., Upadhyaya, H.D., Khan, A.W., Yu, Y., Kim, C., An, S., 2017. Whole-genome resequencing of 292 pigeonpea accessions identifies genomic regions associated with domestication and agronomic traits. Nat. Genet. 49 (7), 1082.

Varshney, R.K., Thudi, M., Roorkiwal, M., He, W., Upadhyaya, H.D., Yang, W., Jian, J., 2019. Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits. Nat. Genet. 51 (5), 857.

Vatanparast, M., Shetty, P., Chopra, R., Doyle, J.J., Sathyanarayana, N., Egan, A.N., 2016. Transcriptome sequencing and marker development in winged bean (Psophocarpus tetragonolobus; Leguminosae). Sci. Rep. 6, 29070.

Vaughan, D.A., Balász, E., Heslop-Harrison, J.S., 2007. From crop domestication to super-domestication. Ann. Bot. 100, 893–901.

Verma, P., Sharma, T.R., Srivastava, P.S., Abdin, M., Bhatia, S., 2014. Exploring genetic variability within lentil (Lens culinaris Medik.) and across related legumes using a newly developed set of microsatellite markers. Mol. Biol. Rep. 41 (9), 5607–5625.

Wani, A., Anis, M., 2008. Gamma ray and EMS induced bold seeded high yielding mutants in chickpea (Cicer arietinum). Turk. J. Biol. 32 (3), 161–166.

Webb, A., Cottage, A., Wood, T., Khamassi, K., Hobbs, D., Gostkiewicz, K., White, M., Khazaei, H., Ali, M., Street, D., Duc, G., Stoddard, F., Maalouf, F., Ogbannaya, F., Link, W., Thomas, J., O'Sullivan, D.M., 2015. A SNP-based consensus genetic map for synteny-based trait targeting in faba bean (Vicia faba L.). Plant Biotechnol.. https://doi.org/10.1111/pbi.12371.

Weckwerth, W., 2011. Unpredictability of metabolism- the key role of metabolomics science in combination with next-generation genome sequencing. Anal. Bioanal. Chem. 400, 1967–1978. https://doi.org/10.1007/s00216-011-4948-9.

Westermarck, J., Ivaska, J., Corthals, G.L., 2013. Identification of protein interactions involved in cellular signaling. Mol. Cell. Proteomics 12 (7), 1752–1763.

Wink, M., 2013. Evolution of secondary metabolites in legumes (Fabaceae). S. Afr. J. Bot. 89, 164–175.

Yang, W.C., Katinakis, P., Hendriks, P., Smolders, A., de Vries, F., Spee, J., van Kammen, A., Bisseling, T., Franssen, H., 1993. Characterization of GmENOD40, a gene showing novel patterns of cell-specific expression during soybean nodule development. Plant J. 3 (4), 573–585.

Yang, I.S., Kim, S., 2015. Analysis of whole transcriptome sequencing data: workflow and software. Genom. Inf. 13 (4), 119.

Yang, W., Weaver, D.B., Nielsen, B.L., Qiu, J., 2001. Molecular mapping of a new gene for resistance to frogeye leaf spot in soybean in "Peking". Plant Breeding 120, 73–78.

Yang, S Samuel, Zheng Jin, Tu, Cheung, Foo, Wayne Wenzhong, Xu, Lamb, JoAnn F.S., Jung, Hans-Joachim G., Vance, Carroll P., Gronwald, John W., 2011. Using RNA-Seq for gene identification, polymorphism detection and transcript profiling in two alfalfa genotypes with divergent cell wall composition in stems. BMC Genom. 12 (1), 199.

Young, N.D., Debellé, F., Oldroyd, G.E., Geurts, R., Cannon, S.B., Udvardi, M.K., Benedito, V.A., Mayer, K.F., Gouzy, J., Schoof, H., Van de Peer, Y., 2011. The Medicago genome provides insight into the evolution of rhizobial symbioses. Nature 480 (7378), 520.

Yu, J., Hu, S., Wang, J., Wong, G.K.S., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., Cao, M., 2002. A draft sequence of the rice genome (Oryza sativa L. ssp. indica). Science 296 (5565), 79–92.

Zeid, M., Mitchell, S., Link, W., Carter, M., Nawar, A., Fulton, T., Kresovich, S., 2009. Simple sequence repeats (SSRs) in faba bean: new loci from *Orobanche*-resistant cultivar 'Giza 402'. Plant Breed 128, 149–155.

Zhang, J., Liang, S., Duan, J., Wang, J., Chen, S., Cheng, Z., Li, Y., 2012b. De novo assembly and characterization of the transcriptome during seed development, and generation of genic-SSR markers in peanut (Arachis hypogaea L.). BMC Genomics 13 (1), 90.

Zhang, J., Liang, S., Duan, J., Wang, J., Chen, S., Cheng, Z., Li, Y., 2012a. De novo assembly and characterizationcharacterization of the transcriptome during seed development, and generation of genic-SSR markers in peanut (Arachis hypogaea L.). BMC Genomics 13 (1), 90.