



# QTLomics in Soybean: A Way Forward for Translational Genomics and Breeding

Giriraj Kumawat\*, Sanjay Gupta, Milind B. Ratnaparkhe, Shivakumar Maranna and Gyanesh K. Satpute

Crop Improvement Section, ICAR—Indian Institute of Soybean Research, Indore, India

## OPEN ACCESS

### Edited by:

Swarup Kumar Parida,  
National Institute of Plant Genome  
Research, India

### Reviewed by:

Humira Sonah,  
Laval University, Canada  
Qing-Yong Yang,  
Huazhong Agricultural University,  
China

### \*Correspondence:

Giriraj Kumawat  
Giriraj.Kumawat@icar.gov.in  
girirajbtc@gmail.com

### Specialty section:

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

**Received:** 01 September 2016

**Accepted:** 23 November 2016

**Published:** 21 December 2016

### Citation:

Kumawat G, Gupta S,  
Ratnaparkhe MB, Maranna S and  
Satpute GK (2016) QTLomics in  
Soybean: A Way Forward for  
Translational Genomics and Breeding.  
*Front. Plant Sci.* 7:1852.  
doi: 10.3389/fpls.2016.01852

Food legumes play an important role in attaining both food and nutritional security along with sustainable agricultural production for the well-being of humans globally. The various traits of economic importance in legume crops are complex and quantitative in nature, which are governed by quantitative trait loci (QTLs). Mapping of quantitative traits is a tedious and costly process, however, a large number of QTLs has been mapped in soybean for various traits albeit their utilization in breeding programmes is poorly reported. For their effective use in breeding programme it is imperative to narrow down the confidence interval of QTLs, to identify the underlying genes, and most importantly allelic characterization of these genes for identifying superior variants. In the field of functional genomics, especially in the identification and characterization of gene responsible for quantitative traits, soybean is far ahead from other legume crops. The availability of genic information about quantitative traits is more significant because it is easy and effective to identify homologs than identifying shared syntenic regions in other crop species. In soybean, genes underlying QTLs have been identified and functionally characterized for phosphorous efficiency, flowering and maturity, pod dehiscence, hard-seededness,  $\alpha$ -Tocopherol content, soybean cyst nematode, sudden death syndrome, and salt tolerance. Candidate genes have also been identified for many other quantitative traits for which functional validation is required. Using the sequence information of identified genes from soybean, comparative genomic analysis of homologs in other legume crops could discover novel structural variants and useful alleles for functional marker development. The functional markers may be very useful for molecular breeding in soybean and harnessing benefit of translational research from soybean to other leguminous crops. Thus, soybean crop can act as a model crop for translational genomics and breeding of quantitative traits in legume crops. In this review, we summarize current status of identification and characterization of genes underlying QTLs for various quantitative traits in soybean and their significance in translational genomics and breeding of other legume crops.

**Keywords:** quantitative traits, genes, alleles, haplotype, functional markers, comparative genomics, legumes

## INTRODUCTION

Soybean [*Glycine max* (L.) Merr.] is a major crop used for food, feed, and biofuel on the globe. The soybean has high nutritional value owing to its high protein and oil content, and having health benefits due to presence of isoflavones and tocopherols. The demand for soybean is increasing due to increasing interest in functional food and use of various soybean constituents/by-products in a wide array of specific industrial products. There is a need to develop tailor-made varieties of soybean for specific requirements. While traits like plant architecture, seed yield components, nutrient use efficiency, photosynthetic efficiency, abiotic and biotic stress tolerance are desirable for sustainable higher productivity, traits like protein content, oil content and quality, fragrance, sweetness, tocopherols, and secondary metabolites are desirable for producing good quality consumer products. However, various traits of economic importance in soybean are complex and quantitative in nature therefore difficult to breed for varietal development. The quantitative traits are governed by quantitative trait loci (QTLs). A large number of QTLs have been mapped in soybean for various quantitative traits albeit their utilization in breeding programmes is poorly reported (<http://www.soybase.org/>). The main reasons for poor utilization of QTLs in soybean breeding programme are; the large size of the mapped genomic regions, small contribution to phenotypic variation, and genetic background effect of recipient parent which demands further validation (Collard and Mackill, 2008; Ratnaparkhe et al., 2014). For their effective use in breeding programme it is required to narrow down the confidence intervals of QTLs, identify the underlying genes and the most important is allelic characterization of these genes for identifying superior variants.

The soybean genome is complex and probably arisen by two rounds of genome-wide duplications or polyploidy events (Walling et al., 2006; Innes et al., 2008). Although several QTLs were mapped on the soybean genome before the publication of soybean genome sequence however the identification of underlying genes responsible for the quantitative trait of interest was rare (Liu et al., 2008; Watanabe et al., 2009). The availability of genome sequence of soybean has accelerated mapping and identification of genes responsible for various qualitative and quantitative traits (Schmutz et al., 2010). This is further augmented by advancement in high-throughput genotyping and next-generation sequencing (NGS) technologies added with advanced phenotyping facilities. The advantage of gene identification through QTL mapping over several other gene identification approaches is that besides identifying genes directly responsible for pathway enzymes and proteins, trans-acting and regulatory genes can also be identified (gene  $\times$  gene interaction). Identification of epistatic interactions between QTLs is very important in the development of efficient marker-assisted selection system (Ma et al., 2016).

Comparative genomics is crucial for translating knowledge between model species and crop species, and it is beneficial to consider the legumes as a coherent, broad genetic system (Cannon, 2008). Comparative genomic techniques are useful for positional cloning or finding new genes in related species.

Several QTLs in soybean have been characterized for underlying genes identifying causative alleles and associated haplotypes. The knowledge of genetic variants in the form of alleles or haplotypes for underlying genes of QTLs can work as an excellent tool for translational genomics and breeding of quantitative traits in soybean, and other legume crops through comparative genomics approaches. Here, we review current status of identification and characterization of genes underlying QTLs for various quantitative traits in soybean which can facilitate development of tailor-made varieties for trait(s) of interest. While several of the genes underlying QTLs are cloned and functionally characterized, many are identified as candidate genes only and functional validation is required.

## QTLomics: QTLs TO GENES IN SOYBEAN

The first report of association of quantitative trait variation with DNA marker loci in soybean was published by Keim et al. (1990a). Since then several advancements in the molecular genetics and genomics had helped in the identification of numerous QTLs in soybean. Advancements from random and limited sequence based marker genotyping to high-throughput sequence based locus-specific marker genotyping lead to mapping of several quantitative traits in soybean (<http://www.soybase.org/>). The term “QTLomics” deals with the generation and analysis of large scale data-sets for mapping, identification and, structural and functional characterization of the genes underlying quantitative traits. In soybean, QTLomics has been applied for characterization of mapped QTLs as well as for mapping and characterization of new QTLs. Recently, several sequence based data-sets has been generated by re-sequencing efforts to facilitate QTLomics studies in soybean (Table 1; Lam et al., 2010; Li et al., 2013, 2014; Chung et al., 2014; Qiu et al., 2014; Zhou Z. et al., 2015; Valliyodan et al., 2016). The availability of well-annotated soybean genome sequence and re-sequencing based data-sets also facilitate development of large number of SNP and Indel markers which are being utilized in QTL mapping and molecular breeding in soybean. Several SNP based marker assays has been developed and validated in soybean which includes SoySNP50K iSelect BeadChip (Song et al., 2013; Vuong et al., 2015; Zhang et al., 2015a), SoySNP6K Infinium BeadChip (Akond et al., 2013), Axiom SoyaSNP array for  $\sim$ 180,000 SNPs (Lee et al., 2015) and the NJAU 355 K SoySNP array (Wang et al., 2016). Further, recent developments in NGS technologies make sequencing-based genotyping cost-effective and efficient (Deshmukh et al., 2014). These sequencing based genotyping approaches include restriction-site-associated DNA sequencing (Baird et al., 2008; Zhou L. et al., 2015), genotyping by sequencing (GBS) (Elshire et al., 2011; Sonah et al., 2013) and whole genome re-sequencing (Xu et al., 2013; Qi et al., 2014).

Conventionally QTL mapping is performed using bi-parental mapping populations with low density marker coverage. This results in poor resolution of QTLs, identifying large genomic regions with limited information on gene(s) underlying important agronomic traits (Sonah et al., 2015). Genome-wide

**TABLE 1 | Whole genome re-sequencing data-sets available for QTLomics in soybean.**

S. No.	No. of accessions used	Genotypes	Sequencing depth	No. of SNPs identified	QTL mapping/Putative candidate genes reported	References
1	31	17 <i>G. soja</i> and 14 <i>G. max</i>	5x	6,318,109	–	Lam et al., 2010
2	25	8 <i>G. soja</i> and 17 <i>G. max</i> (8 landraces and 9 elite ines/cultivars)	–	5,102,244	Domestication-related traits	Li et al., 2013
3	16	10 <i>G. max</i> and 6 <i>G. soja</i>	>14x	3,871,469	Domestication-related traits	Chung et al., 2014
4	7	<i>G. soja</i>	~111.9x	3.63–4.72M SNP per accession	Biotic resistance, seed composition, flowering and maturity time, organ size, and final biomass	Li et al., 2014
5	11	10 Semi-wild and 1 <i>G. soja</i>	9 Semi-wild at ~3X, 1 Semi-wild at ~41X, and 1 Wild at ~55X	7,704,637	Seed size and seed coat color	Qiu et al., 2014
6	302	62 <i>G. soja</i> , 240 <i>G. max</i> (130 landraces and 110 cultivars)	>11X	9,790,744	Domestication and improvement traits	Zhou Z. et al., 2015
7	106	7 wild, 43 landraces, and 56 elite cultivars	~17X	10,417,285	Oil, protein, salinity, and domestication traits	Vallyyodan et al., 2016

association study (GWAS), which uses high density marker genotyping in a population of unrelated genotypes that have accumulated a much larger number of crossing-over events since their last common progenitor, provides higher mapping resolution than conventional QTL mapping and enables one to predict or identify causal genes (Zhang et al., 2015a). GWAS has been successfully combined with sequencing based genotyping technologies for identification of QTLs in soybean (Bastien et al., 2014; Sonah et al., 2015; Iquira et al., 2015; Zhou L. et al., 2015). Novel approaches are being applied for QTL mapping and candidate gene identification in legumes which includes use of multi parent populations like nested association mapping (NAM) and multi-parent advanced generation inter-cross (MAGIC) populations, trait mapping through pooled sequencing methods such as QTL-Seq, MutMap, Seq-BSA, Indel-Seq, and BSR-Seq, a comprehensive review of which is given in Pandey et al. (2016). In recent past, several omics resources has been developed in soybean and integration of omics approaches has been proposed for soybean improvement, in which QTLomics will be an integral part (Deshmukh et al., 2014; Chaudhary et al., 2015). Integration of omics approaches leads to QTL mapping and identification of underlying genes, which is the ultimate goal of omics science (Chaudhary et al., 2015).

The advancement in the genomic research supported by substantial funding due to commercial importance of soybean crop led to the identification of genes underlying QTLs of several quantitative traits in soybean. Underlying genes identified exclusively through QTL mapping and falling in the confidence interval of QTLs are reported in this review. Putative candidate genes identified in GWAS (Genome-Wide Association Studies) are also included for some of the traits. The identified genes have been divided into two categories, characterized genes and candidate genes. Characterized genes are those which have been functionally validated by gain of function experiments such as genetic transformation, transient gene expression or phenotypic validation in near isogenic lines (NILs), while others

identified based on homology and expression analysis were considered as candidate genes for respective QTLs. An overview of characterized genes and important candidate genes in soybean underlying QTLs, for nutrient use efficiency, domestication-related traits, and seed composition traits is given in **Table 2**, and for biotic and abiotic stress related traits is given in **Table 3**, which includes information on traits, contrasting parents used in the QTL mapping, gene model number and gene name etc. In soybean, genes underlying QTLs have been identified and functionally characterized for phosphorous efficiency, flowering and maturity, pod dehiscence, hard-seededness,  $\alpha$ -Tocopherol content, cyst nematode, sudden death syndrome, and salt tolerance. Whereas, candidate genes underlying QTLs have been identified for iron efficiency, seed weight, protein content, oil content, fragrance, isoflavone content, root knot nematode, bacterial leaf pustule, white mold, *Phytophthora* root rot, and root traits. Map positions of characterized genes and important candidate genes underlying quantitative traits in soybean are depicted in **Figure 1**. The details of characterized genes, causal genetic variants and functional markers has been presented in this review to facilitate soybean breeders for molecular breeding and genetic manipulation of quantitative traits in soybean.

## Genes for Nutrient Use Efficiency Traits Phosphorous Efficiency

Phosphorus (P) is the second most important plant nutrient available in soil after nitrogen. Though soils usually contain high amount of total phosphorous, most of the phosphorous (50–80%) occur in organic form which is unavailable to plants. In response to persistent P deficiency conditions, plants have developed many adaptive mechanisms which enhance the availability and uptake of P (Jain et al., 2007). One of the strategies that plants adapt to overcome P deficiency is to secrete acid phosphatases for the hydrolysis of these organic phosphates and thereby using the released inorganic phosphates (Wang et al., 2009). Activity of acid phosphatases (APases) enhances significantly with phosphate

**TABLE 2 | A summary of characterized and candidate genes identified underlying QTLs for nutrient use efficiency, domestication-related, and seed composition traits in soybean.**

Quantitative trait name	Contrasting parents used for QTL mapping*	Candidate/characterized gene(s)	Gene model number	Gene details	Method used for functional annotation /characterization	References
<b>NUTRIENT USE EFFICIENCY</b>						
Phosphorous efficiency	Nannong 94-156 (Efficient) × Bogao	Characterized	Glyma08g20820	Acid phosphatase ( <i>GmACP1</i> )	Hairy root transformation	Zhang et al., 2014
	Nannong 94-156 (Efficient) × Bogao	Candidate	Glyma.04g214000	Pectin methyltransferase	-	Zhang D. et al., 2016
			Glyma.13g161900	Protein kinase		
			Glyma.19g193900	Purple acid phosphatase		Li H. et al., 2016
Iron efficiency	Clark (Efficient) × PI 54619	Candidate	Glyma03g28610	Basic helix-loop-helix DNA binding domain	RNA-Sequence and NILs	Peiffer et al., 2012
			Glyma03g28630	Basic helix-loop-helix DNA binding domain		
<b>DOMESTICATION-RELATED TRAITS</b>						
Flowering and maturity	Misuzudaizu × Moshidou Gong 503 (Early)	Characterized	Glyma06g23040.1	B3 DNA binding protein ( <i>E1</i> )	Genetic transformation	Watanabe et al., 2004; Xia et al., 2012
	Misuzudaizu × Moshidou Gong 503 (Early)	Characterized	Glyma10g36600	GIGANTEA ( <i>GmGib/E2</i> )	NILs	Watanabe et al., 2004, Watanabe et al., 2011
	Misuzudaizu × Moshidou Gong 503 (Early)	Characterized	Glyma19g41210	Phytochrome A3 ( <i>GmPhy A3</i> )	RH-Ls, NILs	Watanabe et al., 2004, 2009
	Takei 780 (Early) × Hidaka 4	Characterized	Glyma10g28170	Phytochrome A2 ( <i>GmPhy A2/E4</i> )	NILs	Abe et al., 2003; Liu et al., 2008
	Takei 780 × TH85 and TH03 (Early)	Characterized	Glyma16g150700.1	Phosphatidylethanolamine-binding protein ( <i>GmFT2a</i> )	NILs	Kong et al., 2014; Zhao et al., 2016
Pod dehiscence	Hayahikari (non-shattering) × Toyonumume	Characterized	Glyma16g25580.1	Dirigent-like protein ( <i>GmPdh1</i> )	Genetic transformation	Funatsuki et al., 2006, 2014
Hard-seededness	Williams 82 × PI 468916 and PI 479752 (Hard-seeded)	Characterized	Glyma02g43700.1	Calcineurin-like metallophosphoesterase transmembrane protein ( <i>GmHs1-1</i> )	Genetic transformation	Sun et al., 2015
Seed weight	GWAS in 105 wild and 262 cultivated soybean accessions	Candidate	Glyma11g03360	Beta-fructofuranosidase activity	-	Wang et al., 2016
			Glyma11g03430	Oligopeptide transport activity		
			Glyma11g05760	SET domain protein		
			Glyma18g06240	Serine threonine protein kinase		
			Glyma18g43500	Leucine rich repeat protein		
			Glyma11g15300	DNA repair protein		
GWAS in 286 soybean accessions	Candidate	Glyma11g15480	Transcription regulator activity	-	Zhou L. et al., 2015	
		Glyma16g26030	Zinc finger DNA binding protein			

(Continued)

TABLE 2 | Continued

Quantitative trait name	Contrasting parents used for QTL mapping*	Candidate/characterized gene(s)	Gene model number	Gene details	Method used for functional annotation /characterization	References
<b>SEED COMPOSITION TRAITS</b>						
Protein content	P1468916 (High protein) × A81-356022	Candidate	Glyma20g18880 Glyma20g19680 Glyma20g21080	Mov34-1 family protein gene Heat shock protein (Hsp22.5) ATP synthase	NILs and Microarray	Eolon et al., 2010
		Candidate	Glyma20g18450 Glyma20g18460 Glyma20g18520 Glyma20g18540	Homeobox protein 22 (HB22) Homeobox protein 22 (HB22) Homeobox protein 22 (HB22) Homeobox protein 22 (HB22)	-	Lestari et al., 2013
	GWAS in 106 soybean accessions	Candidate	Glyma20g19680 Glyma20g21030 Glyma20g21780	Heat shock protein Ammonium transporter Ethylene receptor	-	Vallyodan et al., 2016
Fragrance	Kaori (Fragrant) × Chiang Mai 60	Candidate	Glyma05g01770	Betaine aldehyde dehydrogenase 2 ( <i>GmBADH2</i> )	-	Juwattanasomran et al., 2011
α-Tocopherol content	Keszthelyi Aprozemu Sarga (High) × Ichihime	Characterized	Glyma09g35680.1	γ-tocopherol methyltransferase ( <i>γ-TMT3</i> )	Heterologous expression in Arabidopsis	Dwiyanti et al., 2011
Isoflavone content	Zhongdou 27 (High) × Jiunong 20	Candidate	Glyma02g40290.1 Glyma02g47940.1 Glyma17g34430.1 Glyma17g37060.1 Glyma13g01080.1/2 Glyma13g02740.1 Glyma13g09640.1 Glyma13g24200.1	Cinnamic acid 4-hydroxylase (C4H) Phenylalanine ammonia lyase 1 (PAL1) Chalcone isomerase(CHI) Dihydroflavonol reductase (DFR) 4 Coumarate: CoA ligase (4CL) Flavonol synthase (FLS) Chalcone synthase(CHS) 2-hydroxyisoflavanone synthase (IFS)	eQTL by Real time-PCR	Wang et al., 2014
			Glyma13g20800.1 Glyma13g27380.1	Phenylalanine ammonia lyase 1 (PAL1) Dihydroflavonol reductase (DFR)		

\*Beneficial trait of contrasting parent is given in parentheses.

**TABLE 3 | A summary of characterized and candidate genes identified underlying QTLs for biotic and abiotic stress traits in soybean.**

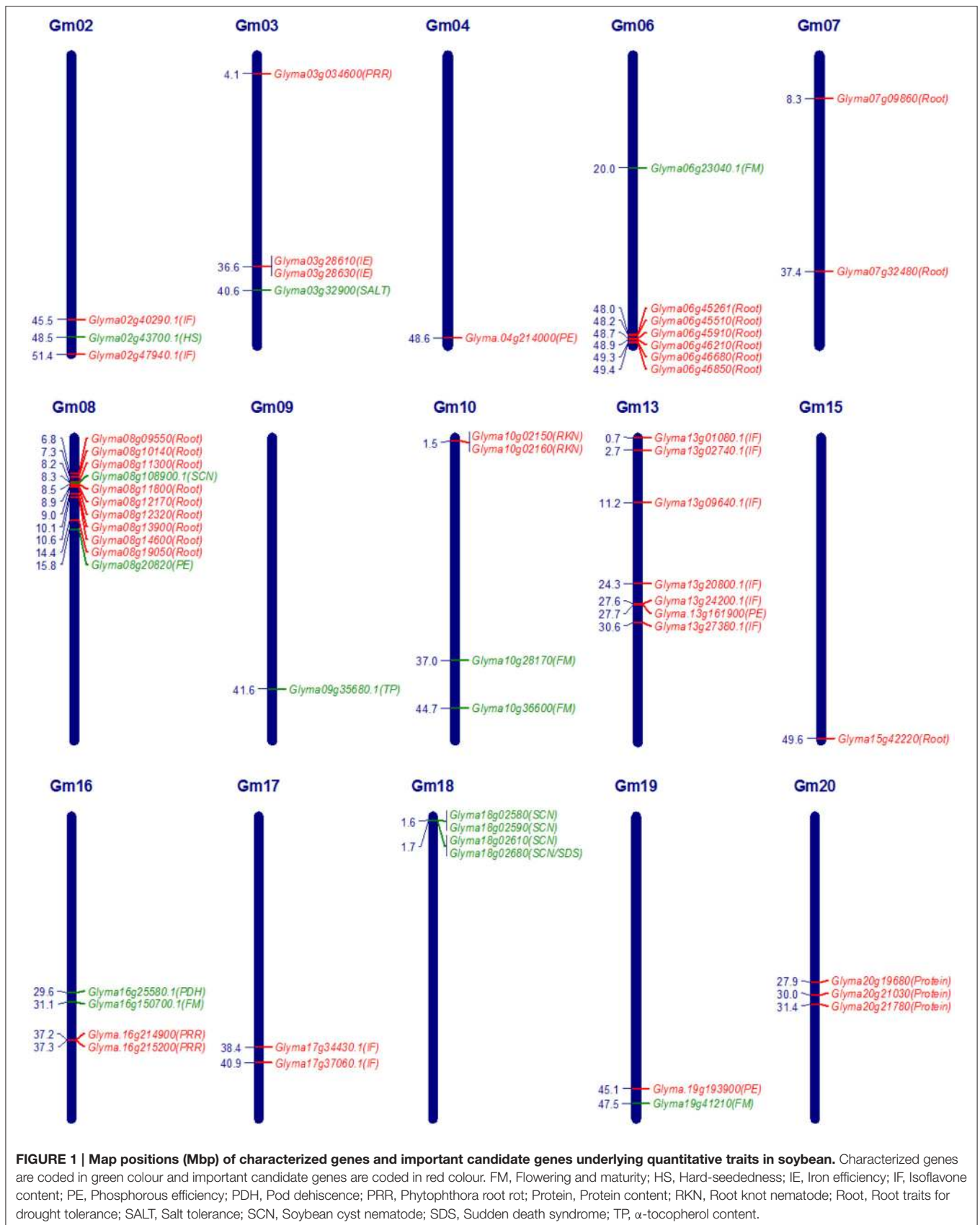
Quantitative trait name	Contrasting parents used for QTL mapping*	Candidate/characterized gene(s)	Gene model number	Gene details	Method used for functional annotation/characterization	References
<b>BIOTIC STRESS</b>						
Root knot nematode	Magellan × PI 438489B (Resistant)	Candidate	Glyma10g02150	Pectin methylesterase inhibitor ( <i>GmPMI</i> ), Pectin methylesterase inhibitor-pectin methylesterase ( <i>GmPMIPM</i> )	Quantitative real time-PCR	Xu et al., 2013
			Glyma10g02160			
Soybean cyst nematode	Essex × Forrest (Resistant)	Characterized	Glyma08g108900.1	Serine hydroxymethyltransferase ( <i>SHMT</i> )	Virus-induced gene silencing (VIGS), targeted RNA interference (RNAi) and Hairy root transformation	Liu et al., 2012
			Glyma18g02680	Receptor like kinase ( <i>GmRLK18-1</i> )	Genetic transformation and NILs	Strour et al., 2012
	Forrest (Resistant) PI 88788 (Resistant)	Characterized	Glyma18g02580	Predicted amino acid transporter	RNAi silencing	Cook et al., 2012
			Glyma18g02590	α-soluble N-ethylmaleimide-sensitive factor attachment protein wound-inducible protein 12		
Bacterial leaf pustule	Taekwangkong × Danbaekkong (Resistant)	Candidate	Glyma17g09780	Unknown-membrane protein	–	Kim D. et al. (2010)
			Glyma17g09790	Zinc finger protein		
Soybean white mold	Maple Arrow (Partially resistant) × Hefeng 25	Candidate	Glyma13 g03360	PR5-like receptor kinase	–	Zhao et al., 2015
			Glyma13 g04020	RINT-1/TIP-1 family		
			Glyma13 g04031	Myb domain protein 33		
			Glyma13 g04041	C2H2-type zinc finger family protein		
Sudden death syndrome	Essex × Forrest (Resistant) GWAS in 214 soybean accessions	Candidate	Glyma18g02680	Receptor like kinase ( <i>GmRLK18-1</i> )	Genetic transformation and NILs	Strour et al., 2012
			Glyma02g11270	Stress induced receptor-like kinase 1 ( <i>GmSIK1</i> )	–	Zhang et al., 2015b
<i>Phytophthora</i> root rot	GWAS in 279 soybean accessions	Candidate	Glyma13g32980	<i>Coat protein 1</i> (COPI)-related gene	–	Li et al., 2016b
			Glyma13g33900	2OG-FE(II) oxygenase family protein		
			Glyma13g33512	Pentatricopeptide (PPR) repeat-containing protein		
			Glyma13g33536	Leucine-rich repeat domain protein		
			Glyma13g33740	Leucine-rich repeat domain protein		
			Glyma13g33243	Gpi16 subunit		
			Glyma13g33260	Zn-finger protein		
Williams × PI 567139B (Resistant)	Candidate	Glyma.03g034600	AAA-type ATPase family protein	RNA-seq analysis	Li et al., 2016a	

(Continued)

TABLE 3 | Continued

Quantitative trait name	Contrasting parents used for QTL mapping*	Candidate/characterized gene(s)	Gene model number	Gene details	Method used for functional annotation/characterization	References
Root traits for drought tolerance	V71-370 (High) × PI 407162	Candidate	Glyma16g215200 Glyma16g214900 Glyma06g45510 Glyma06g45261 Glyma06g46680 Glyma06g46850 Glyma06g46210 Glyma06g45910 Glyma09g32280 Glyma08g19050 Glyma07g09860 Glyma07g32480 Glyma15g42220 Glyma08g12320	TIR-NBS class disease resistance protein TIR-NBS-LRR class disease resistance protein <i>D6 type cyclin</i> gene Auxin efflux carrier protein Alpha/beta hydrolase Histone-like CCAAT transcription factor NEDD8-activating complex Peroxidase 3 Kinesin motor family protein Metacaspase-1 Triglyceride lipase Apoptosis inhibitory 5 family protein Slow anion channel associated 1-like MYB-HD transcription factor TPR transcription factor C2H2 Zn transcription factor bZIP transcription factor GRAS transcription factor Ring finger transcription factor AP2- EREBP transcription factor Xyloglucan endo-transglycosylases (XET)	Micro-array analysis RNA-seq and Affymetrix expression data	Prince et al., 2015 Manavalan et al., 2015
<b>ABIOTIC STRESS</b>						
Salt tolerance	W05 (Tolerant) × C08 Tiefeng 8 (Tolerant) × 85-140 FT-Abyara (Tolerant) × C01	Characterized Characterized Characterized	Glysoja01g005509/ Glyma03g32900 Glyma03g32900.1 Glyma03g32900	<i>Cation/proton exchanger family</i> gene ( <i>GmCHX1</i> ) <i>Cation/proton exchanger family</i> gene ( <i>GmSAL73</i> ) <i>Cation/proton exchanger family</i> gene ( <i>Ncl</i> )	Hairy root transformation Transgenic Tobacco BY-2 cells NILs grafting Genetic transformation and NILs grafting	Qi et al., 2014 Guan et al., 2014b Do et al., 2016

\*Beneficial trait of contrasting parent is given in parentheses.



**FIGURE 1 | Map positions (Mbp) of characterized genes and important candidate genes underlying quantitative traits in soybean.** Characterized genes are coded in green colour and important candidate genes are coded in red colour. FM, Flowering and maturity; HS, Hard-seededness; IE, Iron efficiency; IF, Isoflavone content; PE, Phosphorous efficiency; PDH, Pod dehiscence; PRR, Phytophthora root rot; Protein, Protein content; RKN, Root knot nematode; Root, Root traits for drought tolerance; SALT, Salt tolerance; SCN, Soybean cyst nematode; SDS, Sudden death syndrome; TP,  $\alpha$ -tocopherol content.



deficiency or starvation (Kummerova, 1986). Wang et al. (2009) showed that the over-expression of Arabidopsis *acid phosphatase* gene (*AtPAP15*) in soybean significantly enhanced the APase activity that results in improved P efficiency and plant dry weight as compared with the wild type plants. In Soybean, Zhang et al. (2014) identified a highly significant QTL *qPE8*, on Gm08, in a recombinant inbred lines (RIL) population derived from a cross between the “Nannong 94–156” variety that possessed high P efficiency and the “Bogao” variety with low P efficiency. The *qPE8* was delineated in a genomic region of ~250 kb by using regional association mapping. The delineated genomic region contained a potential candidate gene *Glyma08g20820*, encoding an acid phosphatase (*GmACPI*). *GmACPI* over-expression in soybean hairy roots exhibited a 2.3 fold increase in APase activity and 11–20% increase in P efficiency relative to the control (Zhang et al., 2014). The natural polymorphism analysis of *GmACPI* in 192 soybean accessions identified six probable causative sites that were significantly associated with variations in P efficiency related traits. Ten haplotypes classes were identified for these six sites that explained 33% of the phenotypic variation in P efficiency (Zhang et al., 2014). The cloning of *GmACPI* gene, identification of different haplotypes with their phenotypic effects and availability of functional markers will help breeders for improving P efficiency in soybean. Moreover, the identification of optimal haplotype of *GmACPI* among soybean germplasm collections could further improve the efficacy of selecting soybeans genotypes with higher P efficiencies.

More recently, a high-density genetic map of the same RIL population identified a novel and highly significant QTL *q4-2*, located on Gm04 (Zhang D. et al., 2016). Out of the 38 annotated genes in this QTL region, a *pectin methylesterase* gene and a *protein kinase* gene were identified as promising candidates involved in regulating soybean P efficiency. To elucidate the genetic relationship between phosphorus efficiency and photosynthetic traits, Li H. et al. (2016) identified three major QTLs, *q14-2*, *q15-2*, and *q19-2*, which were found to be associated with both the traits. Within the *q19-2* region, a *purple acid phosphatase* gene (*Glyma.19G193900*), showing significant differential expression upon low P stress was identified as a promising candidate gene involved in regulating both soybean P efficiency and photosynthetic capacity (Li H. et al., 2016).

## Iron Efficiency

Like phosphorus, soils are abundant in iron but most of it is present in the form of  $Fe^{3+}$  which is not accessible to plants. The  $Fe^{2+}$  form is physiologically more significant for plants but it readily oxidizes to sparingly accessible form ( $Fe^{3+}$ ).  $Fe^{3+}$  is insoluble in neutral and high pH, making iron unavailable to plants in alkaline and in calcareous soils. Furthermore, in these types of soil, iron readily combines with phosphates, carbonates, calcium, magnesium, and hydroxide ions and results in poor iron availability to plants. These conditions cause iron deficiency chlorosis (IDC) in soybean, in which photosynthetic pigments, electron transport potential, and thylakoid system are reduced as plants fail to utilize iron from the soil (Spiller and Terry, 1980; Taylor et al., 1982). The best way to manage IDC is the use of iron-efficient varieties in commercial crops

(Peiffer et al., 2012). Lin et al. (1997) identified a major QTL for iron efficiency on Gm03, explaining more than 70% of the phenotypic variation in the Anoka × A7 mapping population. Using introgression mapping, transcriptome sequencing and gene expression analysis, Peiffer et al. (2012) characterized this QTL and identified two candidate genes *Glyma03g28610* and *Glyma03g28630*, encoding transcription factors. These two transcription factor encoding genes, were identified as homologs of the subgroup *Ib basic helix-loop-helix (bHLH)* genes that are known to regulate the strategy I response adaptation in Arabidopsis (Wang et al., 2007; Peiffer et al., 2012). These two genes were significantly induced in soybean roots under iron limiting conditions. Re-sequencing of these two induced genes identified a significant deletion of 12-bp within a predicted dimerization domain of *Glyma03g28610* in iron-inefficient lines (Peiffer et al., 2012). Further experiments for functional validation and identifying the target genes for DNA binding of the two *basic helix-loop-helix (bHLH)* genes may help in identification of iron efficiency molecular network in soybean.

## Genes for Domestication-Related Traits

### Flowering and Maturity

Time to flower and mature confer adaptation of a crop to a specific geographical area. Soybean is a quantitative photoperiod-sensitive short-day plant and its individual genotypes start flowering when the day length becomes less than their critical day length. Ten maturity loci, designated as *E* loci, have been characterized by classical methods in soybean. These loci are *E1* and *E2* (Bernard, 1971), *E3* (Buzzell, 1971), *E4* (Buzzell and Voldeng, 1980), *E5* (McBlain and Bernard, 1987), *E6* (Bonato and Vello, 1999), *E7* (Cober and Voldeng, 2001), *E8* (Cober et al., 2010), *E9* (Kong et al., 2014), and *J* (Ray et al., 1995). Of these, *E1*, *E3*, *E4*, and *E7* have been reported as photoperiodic loci (Cober and Voldeng, 2001). Dominant alleles at *E6*, *E9*, and *J* promote early flowering, whereas dominant alleles at other loci delay flowering and maturity (Zhao et al., 2016). *E1*, *E2*, *E3*, *E4*, and *E9* loci were mapped as QTLs and underlying genes were identified as a B3 DNA binding protein, *GIGANTEA*, *Phytochrome A3*, *Phytochrome A2*, and *GmFT2a*, respectively (Liu et al., 2008; Watanabe et al., 2009, 2011; Xia et al., 2012; Zhao et al., 2016). Molecular bases of *E5*, *E6*, *E7*, *E8*, and *J* are unknown.

Out of the known *E* loci in soybean, the *E1* locus is considered to have the largest effect on time to flowering under field conditions (Stewart et al., 2003). Xia et al. (2012) resolved this QTL into a single gene that encodes a legume specific transcription factor containing B3 domain with a putative bipartite nuclear localization signal (NLS) and C-terminal region carrying structure for DNA binding. Three mutations, a hypoactive allele *e1-as*, and two dysfunctional alleles *e1-nl* and *e1-fs* have been reported at this locus by Xia et al. (2012). A new allele with an insertion of a long interspersed nuclear element (LINE) at the promoter of the *E1* locus (*e1-re*) was identified in “Gokuwase-Kamishunbetsu” (Tsubokura et al., 2014). Soybean maturity locus *E2* has been identified as *GIGANTEA* (*GmGla*), an ortholog of *Arabidopsis thaliana* *GIGANTEA* (*GI*) gene (Watanabe et al., 2011). A single nucleotide polymorphism (SNP) identified in the 10th exon was responsible for early flowering in Misuzudaizu

(Watanabe et al., 2011). Two new alleles with insertion and deletion of 36 bases in the eighth intron (*E2-in* and *E2-dl*) were identified at the *E2* locus by Tsubokura et al. (2014). Since the expression of *GI* gene is not influenced by photoperiod, null mutations in the *GI* gene may be useful resources for adapting plants with complex genomes, such as soybean, to a wide range of geographic regions (Watanabe et al., 2011).

Yamanaka et al. (2001) identified QTL *FT3* and suggested that it may be identical to *E3* locus. Using map-based cloning strategy, this QTL has now been resolved as a phytochrome gene *GmPhyA3* (Watanabe et al., 2009). One SNP for a nonsynonymous amino acid substitution in the third exon and a large deletion of 13.33 kbp at a position after the third exon were detected as two different dysfunctional alleles of *GmPhyA3* (Watanabe et al., 2009). Three novel dysfunctional alleles, *e3-ns* in exon 3 and *e3-fs* in exon 1 and *e3-tr* in exon 4 have been reported by Xu et al. (2013). Using candidate gene approach, Liu et al. (2008) identified *E4* as *GmPhyA2*, a homolog of *phytochrome A*. At recessive *E4* locus, a *Ty1/copia*-like retrotransposon was inserted in the exon 1 of the *GmphyA2* gene, which resulted in dysfunction of this gene (Liu et al., 2008). At least five dysfunctional alleles are known in *GmphyA2* gene (Tsubokura et al., 2013). Loss-of-function alleles at *GmPhyA3* and *GmphyA2* can result in some degree of suppression of *E1* transcription and correspondingly elevate *GmFT* expression, leading to relatively early flowering (Xia et al., 2012).

The *E9* locus was mapped by Kong et al. (2014) using two backcross derived populations developed from Tokei 780 and two of the RILs derived from the cross Tokei 780 × Hidaka 4 (a wild soybean accession). Fine mapping of the QTL region identified two candidate genes *GmFT2a* and *GmFT2b* for early flowering (Kong et al., 2014). Zhao et al. (2016) resolved this QTL as leaky allele of *GmFT2a*. The *e9* allele had an insertion of 6.22 kb *Ty1/copia*-like retrotransposon, *SORE-1*, in the first intron which attenuated *FT2a* expression by its allele-specific transcriptional repression (Zhao et al., 2016). The maturity genes *B3 DNA binding protein* gene (*E1*), *GIGANTEA* (*E2*), *GmPhyA3* (*E3*), and *GmPhyA2* (*E4*) down regulate expression of *GmFT2a* (*E9*) and *GmFT5a* to delay flowering and maturation under the long day condition, suggesting that *GmFT2a* and *GmFT5a* are the major targets in the control of flowering in soybean (Kong et al., 2010; Thakare et al., 2011; Watanabe et al., 2011; Xia et al., 2012; Zhao et al., 2016). The genes and alleles identified for flowering and maturity traits in soybean have been validated in several studies using germplasm from many countries (Tsubokura et al., 2013, 2014; Jiang et al., 2014; Langewisch et al., 2014; Zhai et al., 2014; Gupta et al., 2016). A lot of information has been generated on flowering and maturity genes in soybean which also reveals the presence of legume specific unique genes in soybean i.e., *B3 DNA binding protein* gene of *E1* loci. This information is very useful for identification of genes and their haplotypes in other legume species, which can be used for development of geographical region specific suitable genotypes through molecular breeding.

## Pod Dehiscence

Pod dehiscence is a critical step in the seed dispersal of legumes and especially in wild species of crop plants. However, it

can cause significant yield losses in soybean and shattering resistant genotype is desirable. The cultivated soybean has been domesticated from its annual wild relative, *Glycine soja* (Sieb et Zucc), in China ~5000 years ago (Hymowitz and Singh, 1987) resulting in a multitude of soybean landraces that have adapted to various climate environments. Although, the cultivated soybean is more shattering resistant than the wild soybean, genetic variation in the degree of pod dehiscence is also present among cultivated soybean (Tsuchiya, 1987). Bailey et al. (1997) identified a major QTL *qPdh1* on linkage group J, with a phenotypic variance contribution of 50% for pod dehiscence in cultivated soybean. QTL *qPdh1* was later confirmed in other studies and fine-mapped to a 47-kb genomic region (Funatsuki et al., 2006, 2008; Suzuki et al., 2010; Gao and Zhu, 2013). Using fine mapping and complementation test, Funatsuki et al. (2014) identified an underlying gene *Glyma16g25580.1* (*Pdh1*), which encodes a dirigent-like protein known to be involved in the lignifications which increases dehiscing forces by increasing torsion of dried pod walls. In the shattering-resistant genotype Hayahikari, the gene *pdh1* was defective, having a premature stop codon due to the presence of a non-sense SNP in the coding sequence (Funatsuki et al., 2014). *Pdh1* is a novel gene identified for pod dehiscence and selection of nonsense mutation in this gene has played an important role in soybean domestication. Functional markers developed based on this nonsense SNP could be utilized for development of shattering resistant soybean genotypes through marker-assisted selection (MAS).

## Hard-Seededness

Wild species of leguminous crops produces seeds with variable seed-coat impermeability as a mechanism for maintaining seed viability for long periods (Fuller and Allaby, 2009), which is essential for the long-term survival of wild species (Foley, 2001). A moderate level of seed coat impermeability or hard-seededness is important for the stored soybean seeds as it is desirable for their quality and viability in the tropics, where seeds lose viability within a short period of time after being harvested (Potts et al., 1978). Several studies have identified a common QTL underlying hard-seededness to an overlapping region on soybean Gm02 (Keim et al., 1990b; Watanabe et al., 2004; Liu et al., 2007; Zhang et al., 2008), which has been further confirmed by Sun et al. (2015) and designated as (*GmHs1-1*). Sun et al. (2015) could delimit this QTL to a 22-kb region which contained only two genes, *Glyma02g43700.1* and *Glyma02g43710.1*. Sequence characterization of these two genes identified a single SNP (C>T) in 8th exon of *Glyma02g43700.1* which was responsible for the change of threonine amino acid to methionine. This SNP could distinguish seed coat permeable parent from two hard seed coat *G. soja* parents and 8 additional *G. soja* accessions. Genetic transformation of permeable seed coat accession with *Glyma02g43700.1* and observed phenotypic segregation confirmed that *Glyma02g43700.1* is *GmHs1-1* locus (Sun et al., 2015). *Glyma02g43700.1* was predicted to encode a calcineurin-like metallophosphoesterase transmembrane protein that localized to cellular membranes and its transcript were found predominantly in the Malpighian layer of the seed coat (Sun et al., 2015). These workers also reported that in addition to selection

for this major mutation in *GmHs1-1*, there have been selections against many minor QTLs of *G. soja* during domestication for the development of permeable seed coat in *G. max*.

### Seed Weight

Seed weight is another very important domestication trait which is under continuous selection pressure. Wild species of soybean *G. soja* has very small seeds and selection for bold seeded lines during domestication has given commercial value to this crop. Using GWAS in 286 soybean accessions, Zhou L. et al. (2015) identified two candidate genes on Gm011 and three candidate genes on Gm16 significantly associated with seed size. Three of the five candidate genes had shown relatively higher expression levels during seed development. One candidate gene *Glyma11g15480* was homologous to the *Arabidopsis* gene *NOT2A* (Wang et al., 2013). A correlation analysis based on the gene expression dataset showed that *Glyma11g15480* is co-expressed with 24 genes homologous to seed development genes in *Arabidopsis* (Zhou L. et al., 2015). In another GWAS experiment using 367 accession of wild and cultivated soybean, Wang et al. (2016) reported two candidate genes for seed weight on Gm11 in cultivated soybean and three candidate genes on Gm11 and Gm18 in wild soybean gene-pool. Among the two candidate genes identified in cultivated soybean, *Glyma11g03360* is homologous to *Os04g33740* (*GIF1*) that regulates grain filling and size in rice, and *Glyma11g03430* is homologous to *Os11g12740* (*sp1*), a gene involved in panicle elongation and grain size (Wang et al., 2008; Li et al., 2009).

## Genes for Seed Composition

### Protein Content

Soybean is a rich source of protein therefore it is imperative to investigate genes responsible for seed protein accumulation.  $\beta$ -conglycinin and glycinin are the most prominent seed proteins in soybean (Roberts and Briggs, 1965; Thanh and Shibasaki, 1976). Several seed protein QTLs have been identified in soybean and among them QTLs of LG I (Gm20) are important as they have been reported in many of the mapping populations (Diers et al., 1992; Seboldt et al., 2000; Csanádi et al., 2001). Bolon et al. (2010) could also identify QTL for seed protein at this linkage group and characterized the underlying genes by using a near-isogenic line (NIL) pair that contrasted in seed protein content. This NIL pair differed in protein content because of introgression of a genomic segment containing the high protein QTL of *G. soja* (PI468916) into *G. max* A81-356022. Using microarray and high-throughput whole transcriptome sequencing approaches, 13 genes displaying significant seed transcript expression between NILs were identified which included a potential regulatory protein of the Mov34-1 family, a heat shock protein Hsp22.5, and an ATP synthase (Bolon et al., 2010). In a regional association study of genomic block on Gm20, Valliyodan et al. (2016) identified that the copy number variations in candidate genes (HSP, *Glyma20g19680*; ammonium transporter, *Glyma20g21030*; ethylene receptor, *Glyma20g21780*) were associated with protein content. Lestari et al. (2013) reported that parts of Gm20 and Gm10 showed synteny, however QTLs for seed protein content were detected only

on Gm20. Comparative sequence analysis identified that four tandem duplicates of the putative homeobox protein 22 (HB22) were present specifically on Gm20 (Lestari et al., 2013). This *Medicago truncatula* homolog has been reported to be expressed in endosperm and hypothesized to regulate seed filling stage (Verdier et al., 2008). Like *Medicago truncatula*, these tandem duplicates could be regulating seed filling stage in soybean as well and therefore contribute to the seed protein QTL of Gm20. Further, functional studies are required to identify and validate the role of these candidate genes in seed protein accumulation.

### Fragrance

One of the food uses of soybean is the young pods harvested at R6 stage, cooked and consumed as snack or vegetable (Shanmugasundaram et al., 1991). Soybean varieties used for this purpose are called as “vegetable or edamame or green soybean” and they have some unique characteristics like highly sweet seeds, extra large pods and seeds, and distinct seed fragrance. The vegetable soybean cultivars “Dadachamame” and “Chakaori” are reported to have pleasant fragrance due to the presence of a volatile compound 2-acetyl-1-pyrroline (2AP) (Fushimi and Masuda, 2001), which is also found in fragrant rice. A single recessive gene has been reported to control the fragrance in vegetable soybean (AVRDC, 2003). In rice, recessive mutations in the *betaine aldehyde dehydrogenase 2* gene (*OsBADH2*) are responsible for fragrance (Bradbury et al., 2005; Wanchana, 2005; Shi et al., 2008). This recessive mutation causes elevated 2AP biosynthesis that result in a fragrant aroma in those varieties (Niu et al., 2008; Vanavichit et al., 2008). Through *in-silico* study and QTL analysis Juwattanasomran et al. (2011) identified a major QTL for fragrance in soybean which coincides with the position of *betaine aldehyde dehydrogenase 2* (*GmBADH2*), the same gene responsible for the fragrance in rice (Bradbury et al., 2005). Sequence comparison of *GmBADH2* between Kaori (Fragrant) and non-fragrant soybeans revealed a non-synonymous SNP in exon 10 resulting in change of glycine to aspartic acid. Five PCR-based allele specific SNP markers were developed for marker-assisted breeding of fragrance trait in soybean (Juwattanasomran et al., 2011). A new fragrance allele, which has a 2-bp (TT) deletion in exon 10 of *GmBADH2*, was discovered in another fragrant soybean cultivar Chamame by Juwattanasomran et al. (2012) and functional markers have been developed for MAS.

### $\alpha$ -Tocopherol Content

Tocopherols contribute to both; the nutritional value of soybean seeds and the oxidative stability of soybean oil. The major forms of tocopherols found in seed of common soybean cultivars are  $\gamma$ -tocopherol and  $\delta$ -tocopherol, which account for 60–70 and 20–25% of the total tocopherol, respectively. The proportion of  $\alpha$ -tocopherol, which is having highest vitamin E activity in mammals, is usually <10% of total tocopherol in soybean seeds (Van Eenennaam et al., 2003; Ujiie et al., 2005). Owing to its antioxidant activity, Vitamin E is widely used in food and oil. It's also used as a supplement in the human diet to reduce the risk of cancer and cardiovascular diseases and as a nutrient additive in poultry and cattle feed

to improve meat quality. Dwiyantri et al. (2011) identified gene underlying QTL for high  $\alpha$ -tocopherol content in a soybean variety Keszthelyi Aprozemu Sarga (KAS). Through QTL analysis and fine mapping in populations from a cross between KAS and a Japanese variety Ichihime, they identified  $\gamma$ -*TMT3*, which encodes  $\gamma$ -tocopherol methyltransferase, as a candidate gene responsible for high  $\alpha$ -tocopherol concentration in KAS. Expression analysis and genetic transformation of  $\gamma$ -*TMT3* promoter from KAS in transgenic Arabidopsis plants identified transcriptional regulation of  $\gamma$ -*TMT3* (Dwiyantri et al., 2011).

### Isoflavone Content

Isoflavones are health beneficial molecules in human nutrition (Tsukamoto et al., 1995; Rochfort and Panozzo, 2007). In soybean, isoflavones are critical factor in defense against pathogens (Benhamou and Nicole, 1999; Subramanian et al., 2004), in promoting nodulation by rhizobia (Subramanian et al., 2006, 2007) and in changing or adjusting the microbial population around plant roots (Lozovaya et al., 2004). Several QTLs have been mapped for three species of isoflavones i.e., genistein (*GEN*), daidzein (*DAI*), glycitein (*GLY*), and total of isoflavones (*TOT*) (Primomo et al., 2005; Zeng et al., 2009; Gutierrez-Gonzalez et al., 2010, 2011; Wang et al., 2014). Gutierrez-Gonzalez et al. (2010, 2011) identified and validated a major QTL in two RILs derived from low isoflavone lines Magellan and Essex, and high isoflavone line PI 437654. This QTL was detected on Gm05 in both of the studies and had significant effect on individual isoflavone and total seed isoflavone accumulation (named *qGEN5*, *qDAI5*, *qGLY5*, and *qTOT5*, respectively). This locus accounted for about 36, 35, 12, and 37% of the variance for genistein, daidzein, glycitein, and total isoflavone concentration, respectively. Another major QTL has been identified on Gm08 by Yang et al. (2011). Gutierrez-Gonzalez et al. (2010) used candidate gene approach to identify homologous sequences in the soybean genome with assigned or putative functions in tocopherol biosynthesis pathway, and placed them on the map at their approximate positions, along with all reported QTLs using their interval of confidence. The study identified number of putative candidate genes underlying several of the main-effect and epistatic QTLs. Since there is possibility of trans-acting regulatory elements responsible for the phenotype, Wang et al. (2014) used expression QTL (eQTL) analysis combined with phenotypic QTL analyses identifying a total of 33 eQTLs, of them five eQTL intervals were overlapped with phenotypic QTLs (pQTLs). A total of 11 candidate genes were identified within the overlapped eQTL and pQTL genomic regions (Table 2; Wang et al., 2014). Identification of causative genetic variants within these genes and structural variations in flanking regions is required for development of haplotype based molecular markers facilitating manipulation of isoflavone content in soybean.

### Oil Content

Oil content of soybean is a valuable quantitative trait controlled by multiple genes. At least 110 QTLs for soybean oil content have been mapped but the utilization of these QTLs for

soybean breeding has some limitations. Recently, Valliyodan et al. (2016) conducted comparative sequence analysis between high and low oil lines and showed differential variation in SNP and copy number for lipid biosynthesis and lipid degradation genes between high oil and low oil lines. The comparative analysis shows increased copy number of lipid transfer protein (LTP; *Glyma16g31780*, *Glyma16g31840*, and *Glyma16g31540*) in high oil lines, whereas the negative regulators of lipid biosynthesis genes (ABC transporter, *Glyma03g36310*, Lipase 3, and *Glyma13g04561*) showed more copies in low oil lines. The differential variation for lipid biosynthesis and lipid degradation genes between high oil and low oil lines might be associated with the differences in oil content.

## Genes for Biotic Stress

### Nematode Resistance

#### Root knot nematode

Nematode causes severe yield losses in soybean in the regions of warm climate and sandy soils (Walters and Baker, 1994). The root not nematode and soybean cyst nematode are two major pests of soybean in United States. Using whole genome re-sequencing of recombinant inbred lines (RIL) derived from a cross between Magellan (susceptible) and PI 438489B (resistant), Xu et al. (2013) identified two candidate genes underlying a major QTL for root not nematode resistance mapped to bin 10 of Gm10. The two candidate genes *Glyma10g02150* and *Glyma10g02160*, encodes a pectin methylesterase inhibitor and a pectin methylesterase inhibitor-pectin methylesterase, respectively (Xu et al., 2013). The protein encoded by *Glyma10g02160* was found similar to the *Arabidopsis* protein pectin methylesterase, PME3, which was shown to be important in nematode parasitism (Hewezi et al., 2008). Although, SNPs and InDels were reported between parental lines for both genes but functional characterization and haplotype analysis is required for marker development and utilization.

#### Soybean cyst nematode

Soybean cyst nematode (SCN) (*Heterodera glycines* Ichinohe) feeds on the roots of soybean and is a major constraint to soybean production in United States. In resistant genotype Peking, a dominant resistance gene (*Rhg4*) was identified that was linked to *I* locus underlying seed coat color (Matson and Williams, 1965). Through positional cloning from soybean cultivar Forrest, Liu et al. (2012) identified a gene underlying *Rhg4* locus on Gm08, a major QTL contributing to resistance to SCN. The gene underlying *Rhg4* encodes a serine hydroxymethyltransferase (SHMT) enzyme, a ubiquitous enzyme in nature with a key role in one-carbon folate metabolism. Alternative alleles of *Rhg4* differ by two genetic polymorphisms that alter a key regulatory property of the enzyme (Liu et al., 2012). Haplotype survey of *SHMT* sequence identified eight haplotypes, where soybean lines with haplotypes H1–H3 carry resistant alleles at *SHMT* (Liu et al., 2012). The identified resistance mechanism can be readily exploited to improve resistance against SCN in susceptible cultivars. A recessive and codominant locus *rhg1* which is necessary, but not sufficient, for resistance to all known Hg-biotypes of SCN, was also mapped (Concibido et al., 1997;

Ruben et al., 2006). Ruben et al. (2006) fine mapped the *rhg1* locus and identified a single *receptor like kinase* (*RLK*; *leucine rich repeat transmembrane-protein kinase*) as candidate resistance gene. Srour et al. (2012) characterized the *GmRLK18-1*, the gene underlying *Rhg1* locus, in transgenic and heterozygous plants and demonstrated that the dominant allele confers reported pleiotropic resistance to SCN and sudden death syndrome (SDS) caused by *Fusarium virguliforme*. In the absence of *Rhg4*, the *GmRLK18-1* had provided partial resistance to SCN and conferred nearly complete resistance to SDS, whereas, in the presence of *Rhg4*, the plants with the *GmRLK18-1* transgene were fully resistant to SCN as well as SDS (Srour et al., 2012). *In vitro*, the purified LRR domain of *GmRLK18-1* has been shown to bind with the CLE peptides of plant origin known to involve in tracheary element inhibition (Afzal et al., 2013).

About 90% of the commercially cultivated SCN-resistant soybean varieties in the central United States use the *rhg1-b* allele (haplotype) as the main SCN resistance locus derived from the genotype PI 88788 (Cook et al., 2012). Fine mapping of *rhg1-b* haplotype in PI88788 has delineated *rhg1-b* in an interval that corresponds to a 67-kb segment carrying 11 predicted genes in the genome of Williams 82 (Kim M. et al., 2010). By using gene silencing approach, Cook et al. (2012) identified that SCN resistance mediated by the soybean QTL *Rhg1* in PI 88788 is conferred by copy number variation of a 31-kilobase segment encoding an amino acid transporter (*Glyma18g02580*), an  $\alpha$ -SNAP ( $\alpha$ -soluble N-ethylmaleimide-sensitive factor attachment protein) protein (*Glyma18g02590*), and a W112 (wound-inducible domain) protein (*Glyma18g02610*). Further, characterization of genetic variation at *Rhg1* across 41 diverse soybean accessions discovered that SCN resistance is associated with multicopy *Rhg1* haplotypes that form two distinct groups (Cook et al., 2014). The high-copy-number *Rhg1* accessions reportedly contain a flexible number of copies (seven to 10) of the 31-kb *Rhg1* repeat whereas the low-copy-number *Rhg1* group contains three copies of the *Rhg1* repeat and a newly identified allele of  $\alpha$ -SNAP (Cook et al., 2014). Kadam et al. (2016) selected SNPs from the *Rhg1* gene (*Glyma18g02590*) and *Rhg4* gene (*Glyma08g11490*) to develop KASPar assays. Two SNP markers specific to *Rhg1* locus and two SNP markers specific to *Rhg4* locus were validated in 95 germplasm lines and a subset of RILs. A Taqman assay was also developed from the conserved region of the *Rhg1* locus for detection of copy number variation at *Rhg1* locus (Kadam et al., 2016).

## Disease Resistance

### Bacterial leaf pustule

Soybean bacterial leaf pustule (BLP) is caused by *Xanthomonas axonopodis* pv. *glycines*. BLP resistance was reported to be controlled by a single recessive gene (*rxp*) in BLP resistant source, CNS (PI 548445) (Hartwig and Lehman, 1951). QTLs analysis for BLP resistance revealed that Satt372 and Satt486 on Gm17 were strongly associated with BLP resistance (Narvel et al., 2001; Van et al., 2004). Fine mapping in RILs derived from a cross between “Taekwangkong” (BLP susceptible) and “Danbaekkong” (BLP resistant) and two pair of NILs, Kim D. et al. (2010) narrowed down the BLP resistance locus to 33

Kb. Annotation of this 33 Kb sequence identified two putative candidate genes, a membrane protein gene (*Glyma17g09780*) and a Zinc finger family protein gene (*Glyma17g09790*). The candidate genes showed high similarity with their paralogous genes, which were located on the duplicated regions reported for bacterial leaf pustule resistance QTLs (Kim D. et al., 2010). Functional role to these genes has not been assigned.

### Soybean white mold

White mold, caused by *Sclerotinia sclerotiorum* [(Lib.) W. Phillips.] de Bary, is reported as a devastating disease of soybean and many other crops (Boland and Hall, 1994). Currently, due to the lack of immune type resistance genotypes, only partially resistant cultivars have been used for genetic mapping. Recently, Zhao et al. (2015) identified a major QTL on Gm13 by linkage and association mapping. Four potential candidate genes known to be involved in disease response and anthocyanin biosynthesis were identified at the locus near the peak SNPs. The candidate genes identified were a *PR5-like receptor kinase*, *RINT-1/TIP-1 family*, *Myb domain protein 33*, and a *C2H2-type zinc finger family protein* gene (Zhao et al., 2015). These candidate genes will be useful for functional validation and soybean breeding for improving resistance to white mold.

### Sudden death syndrome

Sudden death syndrome of soybean caused by *Fusarium virguliforme* (Roy et al., 1997), is a serious threat to soybean production (Wrather et al., 2010) which can be managed with host plant resistance (Wen et al., 2014). Resistance to SDS in soybean is multi-geneic and reportedly had two components; a partial resistance to leaf scorch caused indirectly by translocated fungal toxins; and a partial resistance to root infection and rot caused directly at the site of infection by the fungus (Njiti et al., 1998; Kazi et al., 2008). Linkage mapping identified several QTLs associated with SDS resistance across 12 of the 20 soybean chromosomes (<http://www.soybase.org/>). A locus *Rfs2/Rhg1* on Gm18 underlay partial resistance to the spread of root infections by *F. virguliforme* and SCN (Njiti et al., 1998; Triwitayakorn et al., 2005). Using NILs, Srour et al. (2012) characterized *Rfs2/Rhg1* QTL on Gm18 identifying a receptor like kinase (*GmRLK18-1*) responsible for resistance to SDS and SCN. In a GWAS study using two sets of 392 and 300 unique soybean genotypes as independent association panels, Wen et al. (2014) identified a total of 20 loci underlying SDS resistance. Several of SNPs significantly associated with SDS resistance were within or adjacent to sequences annotated as plant disease resistance genes including the previously identified SDS resistance gene *GmRLK18-1*. In a GWAS study using 214 germplasm accessions Zhang et al. (2015b) identified a potential candidate gene on Gm02 encoding *stress induced receptor-like kinase gene 1* (*SIK1*). The peak SNP locus associated with SDS resistance was present in the coding region of the *SIK1* resulting in a non-synonymous mutation.

### Phytophthora Root Rot

*Phytophthora* root rot is one of the important diseases of soybean and is caused by *Phytophthora sojae*. Tolerance to

*Phytophthora* root rot is multi-genic due to the presence of several races of pathogen and so far 20 *Rps* (resistance to *P. sojae*) loci including 26 alleles have been detected on four different chromosomes i.e., Gm03, Gm13, Gm16, and Gm18 (Li et al., 2016a). Lin et al. (2013) have identified two *Rps* genes, *RpsUN1* on chromosome 3 and *RpsUN2* on chromosome 16 from a soybean landrace PI 567139B, which together confer complete resistance to 16 *P. sojae* races/isolates. Li et al. (2016a), further narrowed *RpsUN1* and *RpsUN2*, to a 151 kb region that harbors five genes including three disease resistance (R)-like genes, and a 36 kb region that contains four genes, all of which are of the typical R-like genes, respectively. RNA-seq analysis of these nine genes before and after inoculation with the pathogen suggest that *Glyma.03g034600* in the *RpsUN1* region, and *Glyma.16g215200* and *Glyma.16g214900* in the *RpsUN2* region of PI 567139B may be associated with the resistance to *P. sojae*. Li et al. (2016b) conducted GWAS in 279 soybean accessions and identified seven candidate genes within 500-kb regions on Gm13 that are probably involved in natural variations in partial resistance to *P. sojae*. These genes encode a COPI (*Glyma13g32980*), a 2OG-Fe(II) protein (*Glyma13g33900*), a PPR protein (*Glyma13g33512*), LRR domain proteins (*Glyma13g33536*, *Glyma13g33740*), a Gpi16 subunit (*Glyma13g33243*), and a Zn-finger protein (*Glyma13g33260*).

## Genes for Abiotic Stress

### Root Traits for Drought Tolerance

Identification of genes underlying root system architecture and canopy characteristics is critical to develop soybean that is suited to water-limited environments. Prince et al. (2015) identified four significant QTLs associated with different root architectural traits on Gm06 and Gm07 in an interspecific RILs population of *G. max* (V71-370) × *G. soja* (PI 407162). A number of candidate genes were identified for root system architecture traits. Notable among them are one encoding triglyceride lipase mapped to QTLs for root distribution traits, two genes encoding an apoptosis inhibitory 5 family protein and an oxidoreductase/transition metal ion binding protein mapped to the taproot length QTLs, a cell cycle associated *D6 type cyclin gene* and a key hormone auxin-associated gene *auxin efflux carrier protein gene* (Prince et al., 2015). In another study, Manavalan et al. (2015) identified a major QTL on Gm08 that governed root traits (tap root length and lateral root number) and shoot length, and identified six transcription factors (MYB HD, TPR, C2H2 Zn, bZIP, GRAS, and Ring finger) and two key cell wall expansion-related genes which encode xyloglucan endo-transglycosylases as candidate genes in the confidence interval of the QTL. These are key candidate genes for validation and to develop a better root ideotype in soybean.

### Salt Tolerance

Soybean is a moderately salt-sensitive crop and germplasm screening has been shown to display a spectrum of salt-tolerance phenotypes (Munns and Tester, 2008; Phang et al., 2008). A major QTL for salt tolerance has been repeatedly mapped within a region on soybean linkage group N (Gm03)

(Lee et al., 2004; Hamwieh and Xu, 2008; Hamwieh et al., 2011; Ha et al., 2013; Guan et al., 2014a). Qi et al. (2014) resolved this QTL in to a single dominant gene *GmCHX1*, a counterpart of *Glyma03g32900* in Williams 82, in a salt tolerant wild soybean accession W05. Comparison of the genomic sequences of W05 and Williams 82 at *GmCHX1* locus encoding a cation  $H^+$  exchanger revealed that Williams 82 had a ~3.8 Kb *Ty1/copia* retrotransposon inserted into exon 3, but not in its counterpart *Glysoja01g005509* in W05 (Qi et al., 2014). Guan et al. (2014b) resolved the same QTL in to a salt tolerant variety Tiefeng 8, identifying the same gene *Glyma03g32900* (named as *GmSALT3*) with similar insertion of a 3.78-kb *copia* retrotransposon in exon 3 of salt sensitive parent. Haplotype screening of *GmSALT3* identified a total of nine haplotypes including two salt-tolerant haplotypes and seven salt-sensitive haplotypes (Guan et al., 2014b). Recently, Do et al. (2016) identified the same gene in salt tolerant cultivar FT-Abhayra using map-based cloning and identified same insertion of a ~3.8-kb *Ty1/copia* type retrotransposon fragment responsible for loss of function of the salt tolerant gene named as *Ncl*. All of the characterized genes underlying salt tolerant QTLs were pinpointed at single gene *Glyma03g32900*, and the presence of *Ty1/copia* type retrotransposon in all salt tolerant genotypes indicate the single point of origin for salt sensitivity in soybean. Patil et al. (2016) used whole genome re-sequencing data for haplotype analysis of the *GmCHX1* gene (*Glyma03g32900*) identifying three major structural variants and allelic variation in the promoter and genic regions. Based on SNPs specific to three structural variants, breeder friendly KASPar assays have been developed and validated which will be very useful for molecular breeding for salt tolerance in soybean (Patil et al., 2016).

## PROSPECTUS OF TRANSLATIONAL GENOMICS AND BREEDING IN SOYBEAN

A lot of phenotypic and genotypic information needs to be generated for QTL mapping, and for identification and characterization of gene underlying a quantitative trait. Once the underlying gene sequence is fully characterized, haplotype analysis of the structural variants identified in the underlying genes for quantitative trait of interest could discover novel and useful alleles leading to functional marker development in soybean. The knowledge of gene function generated through QTL mapping, underlying gene identification, characterization and resulted development of functional markers through allele mining may be translated in to a useful product using genomics assisted breeding approaches, the entire process commonly described as “translational genomics and breeding.” **Figure 2** describes the process of translational genomics and breeding in soybean and other legume crops. As soon as functional markers become available, it is much precise and efficient to develop a desirable genotype by using marker-assisted selection (MAS), marker-assisted backcrossing (MABC) and marker-assisted recurrent selection (MARS) compared to traditional breeding methods or linked marker techniques. MAS is useful for

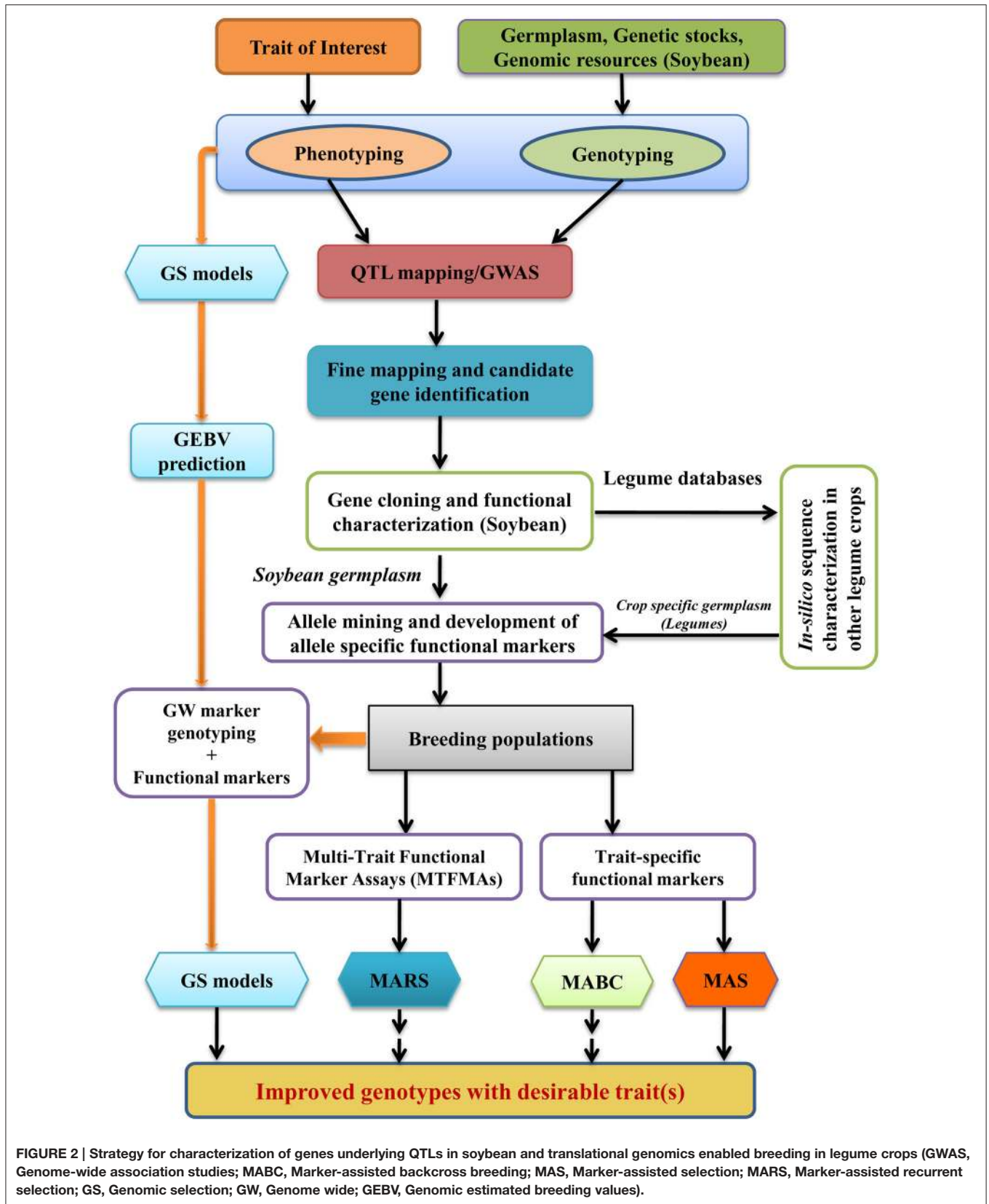


FIGURE 2 | Strategy for characterization of genes underlying QTLs in soybean and translational genomics enabled breeding in legume crops (GWAS, Genome-wide association studies; MABC, Marker-assisted backcross breeding; MAS, Marker-assisted selection; MARS, Marker-assisted recurrent selection; GS, Genomic selection; GW, Genome wide; GEBV, Genomic estimated breeding values).

selection of individuals with specific alleles for traits controlled by a limited number of loci (upto 6–8). Especially, MAS is effective and time saving for selection of individuals in early generations and in large populations for low heritability traits, allowing breeders to focus attention on a lesser number of high-priority lines in subsequent generations (Collard and Mackill, 2008). MABC method is applied to transfer a limited number of loci (i.e., disease resistance, transgene) from one genetic background to other. In MABC, several cycle of backcrossing is performed to recover the genetic background of the recipient parent (the adapted cultivar) except for the target QTL allele which is selected from donor parent. MABC is very useful for transfer of favorable traits from unadapted germplasm to elite genotypes. However, if the performance of a plant is determined by a complex genotype it is unlikely that this ideal genotype will be attained through MABC only (Ribaut and Ragot, 2006). To overcome the limitation of only being able to improve existing elite genotypes, other approaches like MARS and genomic selection (GS) are considered. MARS involves marker-assisted identification and selection of several genomic regions (up to 20 or even more) for complex traits within a single population (Servin et al., 2004). However, in MARS, major effect QTLs only identified and utilized in selection. These major QTLs can explain only a small fraction of phenotypic variation and may show unexpected trait expression in new genetic backgrounds because of epistatic interactions (Deshmukh et al., 2014).

Another promising approach for genomics assisted breeding of quantitative traits is GS. GS is a form of MAS that selects favorable individuals based on genomic estimated breeding values (GEBVs) of all markers genotyped in a breeding population (Nakaya and Isobe, 2012). These GEBVs say nothing of the function of the underlying genes but they are the ideal selection criterion (Jannink et al., 2010). GS able to capture small-effect QTLs that governs most of the variation including epistatic interaction effects (Deshmukh et al., 2014). However, GS has also got some limitations which are: (1) Each time new GS models has to be trained for new target populations and environment; (2) GS does not provide control over genetic background when a trait is to be incorporated from un-adapted genotype; (3) GS cannot guarantee for selection of a major QTL which is already known. Therefore, it was proposed to incorporate major QTL/GWAS loci information in GS models so that the balance of genetic background can be made along with maximum gain of breeding value (Deshmukh et al., 2014).

Traditionally molecular breeding is performed using linked molecular markers which suffered from limitation of marker polymorphism and genetic recombination, so even if parental genotypes possesses beneficial alleles the selection efficiency may still be <100%. The first example of MAS in soybean was of soybean cyst nematode resistance QTL conditioned by *rhg1* locus (Cregan et al., 1999). The two SSR markers linked with SCN resistance *rhg1* locus were further validated by Silva et al. (2007) and found that only a combination of two markers provide selection efficiency of 100%. Shivakumar et al. (2016) employed MABC for conversion of two high Kunitz trypsin inhibitor (KTI) content soybean varieties (DS

9712 and DS 9814) into low KTI soybean genotypes using three SSR markers. Rosso et al. (2011) has developed breeder friendly markers for *D-myo-inositol 3-phosphate synthase 1* gene (*MIPS1*) which causes modifications to seed phosphorus and carbohydrate content in soybean. In this study, they clearly demonstrated that an allele specific SNP marker developed from causal mutation of *MIPS1* was able to select 100% mutant individuals as compared to four linked polymorphic SSR markers which showed selection efficiency of 42–93%. Therefore, it is necessary to develop functional markers in soybean for selection of QTLs with 100% efficiency. In soybean, functional markers have been developed for several QTLs and a summary of these functional markers is presented in **Table 4**, which will be very useful for soybean molecular breeders for carrying out translational breeding in soybean. Pham et al. (2010, 2011) used allele specific functional markers for *FAD2-1A* and *FAD2-1B* genes to combine mutant alleles in a single background. They identified soybean lines carrying both homozygous mutant *FAD2-1A* alleles and mutant *FAD2-1B* alleles producing >80% oleic acid content, compared to 20% in conventional soybean. Kumar et al. (2013) successfully used gene specific functional marker for elimination of kunitz trypsin inhibitor from Indian soybean cultivar JS 97-52 through MABC. These are some examples which demonstrate the utility of functional markers developed by characterization of underlying genes of specific traits, to harness the true benefits of translational genomics and breeding in soybean. Although, the functional markers used in the above cited examples were not developed by QTL mapping and characterization, however several functional markers have been recently developed based on characterization of genes underlying QTLs and currently has been utilized in ongoing soybean breeding programs (Tsubokura et al., 2013; Xu et al., 2013; Tardivel et al., 2014; Kadam et al., 2016; Patil et al., 2016).

## SOYBEAN AS MODEL CROP FOR TRANSLATION GENOMICS AND BREEDING OF QUANTITATIVE TRAITS IN LEGUMES

Legumes are quite different from other plant species in certain features such as symbiotic nitrogen fixation, protein rich physiology and secondary metabolism, and past experience has shown that comparative genomics, epitomized by the “model to crop translation” approach, was not very useful for legume crops improvement (Koebner and Varshney, 2006; Varshney et al., 2015). Only few examples of translational genomics are available in grain legumes due to various inherent differences from other non-legume crops (Varshney et al., 2015). Soybean being a legume crop, present a unique opportunity for translational genomics and breeding of quantitative traits in other legume crops, due to presence of legume specific genes and close genetic relatedness to other legume species. In the field of functional genomics, especially in the identification and characterization of genes responsible for quantitative traits, soybean is far ahead from other legume crops. Homology based *in-silico* analysis



**TABLE 4 | Details of allele-specific functional markers available for translational breeding in soybean.**

Trait	Gene/Marker locus	Marker Type (Enzyme)	Primer sequences (5' → 3')	Sizes of amplified fragments/restriction digested fragments (bp)	SNP/Indel allele	References
Flowering and maturity	<i>e1-fs/e1-nl</i>	CAPS ( <i>Hinf I</i> )	F: CACTCAAAATTAGCCCTTTCA	<i>E1-e1-as</i> : 186 + 36	A/-	Xia et al., 2012
			R: TTCATCTCCTTTCATTTTTGTG	<i>E1-fs</i> : 136 + 46 + 36		
		dCAPS ( <i>Taq I</i> )	F:TCAGATGAAGGAGCAGTGTCAAAAAAGAGT	<i>E1-nl</i> : No amplification	C/G	Xia et al., 2012
	<i>e2</i>	dCAPS ( <i>Dra I</i> )	R: TOCGATCTCATCACCTTTC	<i>E1-as</i> : 413 + 31	A/T	Watanabe et al., 2011
			F: AAGCCTATGCCAGCTAGGTATT	<i>E2</i> : 110		
	<i>e3-tr</i>	FLP	R: GAAGCCATCAGAGGCATGCTTATT	<i>e2</i> : 27 + 83	-	Watanabe et al., 2009
			F: TGGAGGGTATTGGATGATGC	<i>E3-Misuzudaizu</i> : 1339		
	<i>e3-fs</i>	CAPS ( <i>Ale I</i> )	R1: CTAAGTCGCGCTCTGGTTTCAG	<i>E3-Harasoy</i> : 558	-T	Xu et al., 2013
			R2: CGGTCAAGGCCAACATGAG	<i>e3-tr</i> : 275		
			R3: GTOCTATACAATTTTACGAGG	<i>E3</i> : 552 + 206		
	<i>e3-ns</i>	dCAPS ( <i>Mfe I</i> )	F: GGGATAGTTCTGATGCTGTTCAA	<i>e3-fs</i> : 759	C/T	Xu et al., 2013
			R: CCTTGATCGATAGCATATGTGCT	<i>E3</i> : 163		
	<i>e4-oto</i>	CAPS ( <i>Sac I</i> )	F: GTTGAAGAGAAGATCACAACA	<i>e3</i> : 140 + 23	G/-	Tsubokura et al., 2013
			R: GATGAACTAATTTCCCTAACTGCA	<i>E4</i> : 535		
	<i>e4-tsu</i>	dCAPS ( <i>Eco RV</i> )	F: CCCAGACACTCTGTGTGAT	<i>e4</i> : 439 + 96	T/-	Tsubokura et al., 2013
			R: CCATACTCTGGTACTTTG	<i>E4</i> : 355		
	<i>e4-kam</i>	CAPS ( <i>Afl II</i> )	F: CACCCTAGGAGTTGTGTGTT	<i>e4-tsu</i> : 332 + 23	G/-	Tsubokura et al., 2013
			R: GCGGTTCTGTACAATTGCCGTGATA	<i>E4</i> : 494		
	<i>e4-kes</i>	CAPS ( <i>BSP HI</i> )	F: CTTAATAAAGCCATGACTGGTTG	<i>e4-kam</i> : 286 + 208	A/-	Tsubokura et al., 2013
R: CTTGAGTTCAATGAGGTTTCAAC			<i>E4</i> : 494			
<i>e4-SORE-I</i>	FLP	F: CTTAATAAAGCCATGACTGGTTG	<i>e4-kes</i> : 399 + 95	-	Liu et al., 2008	
		R: CTTGAGTTCAATGAGGTTTCAAC	<i>E4</i> : 1229			
		F: AGACGTAGTCTAGGGCTAT	<i>e4-SORE-I</i> : 837			
		R1: GCATCTGGCATCACCAGATCA	-			
<i>e9</i>	FLP	R2: GCTCATCCCTTCGAATTCAG	-	-	Zhao et al., 2016	
		F1: GCTCTCTCTTCCACCTCTAGATGG	-			
Pod Dehiscence	<i>Pdh1</i>	F: GCCCTCGTTGTTCCTTCAT	<i>Pdh1</i> : 125	A/T	Funatsuki et al., 2014	
		R: GCGTTGCTCCGTTGTAGAT	<i>pdh1</i> : 76+49			
Fragrance	<i>GmBADH2-A2</i>	SNAP	F: TGGAAGAAGTTGCAGACAGA	Wild type: No amplification	G/A	Juwattanasomran et al., 2011
			R: AAAGCATACCTGCCCTTACTTTAGAA	Fragrant: 126 bp		
Fragrance	<i>GmBADH2-EX10</i>	Indel	F: TCCCGCCCTTATTGTACATGC	Wild type:174	-	Juwattanasomran et al., 2012
			R: TCCCGCCCTTATTGTACATGC	Wild type:174		

(Continued)

TABLE 4 | Continued

Trait	Gene/Marker locus	Marker Type (Enzyme)	Primer sequences (5' → 3')	Sizes of amplified fragments/restriction digested fragments (bp)	SNP/Indel allele	References
Salt tolerance	M1	KASPar	R: TTTTGACCCATTTCACAATCC	Fragment: 172	G/C	Patil et al., 2016
			FAM_primer: ACCAAACCAAACTAGCTAGTTTTTC ATCACCTTCTATGATTTGTTTTTTTAAATTCCT AACTAACTAACACAGCAAG	-		
			HEX_primer: ACCAAACCAAACTAGCTAGTTTT CATCACCTTCTATGATTTGTTTTTAAATTCCT TAACTAACTAACACAGCAAC	-		
M4	KASPar	Common Reverse primer: AAGCACTGAGTCTTTGGCCATGACGTTCAACGG- GAGCACCATCACAAACGGCGTGGGA	-	C/G	Patil et al., 2016	
		FAM_primer: TATACAAATATGTTGTTTTGTG1- TATTCTACGATTATCATCATCTAGTTC	-			
		HEX_primer: TATACAAATATGTTGTTTTGTG1- TATTCTACGATTATCATCATCTAGTTC	-			
M6	KASPar	Common Reverse primer: AATGGTAACTTTATCACATAT- TATAGTTGAATTAAGT. TAGTTTTATTAAATTTTCAACTAAAAATTAAC	-	G/T	Patil et al., 2016	
		FAM_primer: GAGATCTCTTTGCTCTTCGGGTAACGGGAT- TACCAGTTTCATTGCTCGTTATACAGGACCGATTTTG	-			
		HEX_primer: GAGATCTCTTTGCTCTTCGGGTAACGGGAT- TACCAGTTTCATTGCTCGTTATACAGGACCGATTTTG	-			
Soybean cyst nematode	Rhg1-2	KASPar	Common Reverse primer: ATAAATTCAAAACGAAAGTAATCTCGTTAAGACAT- CAAGGGCC- GAGAGTACTGTGATTAAGATGCTATCGAAGAACTTTAA	-	C/G	Kadam et al., 2016
			ASP 1: TCTAATGCATTGGTTATAGCAACAACG ASP 2: TCTAATGCATTGGTTATAGCAACAACG	-		
			Common primer: TGCTGGCATCTGCCAACTCTGTAAA ASP 1: GAAAGCCAAAAGAACCTTGAGGAGC	-		
Rhg1-5	KASPar			C/G	Kadam et al., 2016	

(Continued)

TABLE 4 | Continued

Trait	Gene/Marker locus	Marker Type (Enzyme)	Primer sequences (5' → 3')	Sizes of amplified fragments/restriction digested fragments (bp)	SNP/Indel allele	References
	<i>Rhg4-3</i>	KASPar	ASP 2: GAAAGCCAAAGAACCTTGAGGAGG Common primer: CAAACCACCAGGAATATAAGGTACAAT ASP 1: TCGTTGTGATGTTTTCAGGGA ASP 2: TCGTTGTGATGTTTTCAGGAGGT Common primer: CAGAGATCACAGAGATTTCTCCACCTT ASP 1: GAGGTGCCGCCGGGAGG ASP 2: GAGGTGCCGCCGGGAGG	-	A/T	Kadam et al., 2016
	<i>Rhg4-5</i>	KASPar	Common primer: CGACCGCATCATGGGGCTAGAT F: GTTATTACTTCAAATCGACGAGTGTGTG R: AAATATTTCCAGTAAATCAGATTA AAACTATACTTCA FAM: TCGGACACCTCAAAACT	-	C/G	Kadam et al., 2016
	<i>Rhg1</i>	Taqman assay	Common primer: CGACCGCATCATGGGGCTAGAT F: GTTATTACTTCAAATCGACGAGTGTGTG R: AAATATTTCCAGTAAATCAGATTA AAACTATACTTCA FAM: TCGGACACCTCAAAACT	-	CNV	Kadam et al., 2016
Oleic acid content	<i>FAD2-1A</i>	GC-Tail assay	ASP 1: gggggcagggcggcATCAACCCATTGGTACTTGC ASP 2: ggcgcgATCAACCCATTGGTACTTGT Common forward primer: GTTGCCCTTCTCACTGGTG Primer 1: CCAAGGTTGCCCTTCTCACTGGT Primer 2: TAGGCCACCTATTGTGAGTGTGAC Fluorescein-SPC: CCTTAGGAAAGGGCTGTTTCTCT-Phosphate Primer 1: ACTGCATCGAATAATAAAGCC	-	G/A	Dierking and Bilyeu, 2010
	<i>FAD2-1A</i>	Simple probe assay	Primer 1: CCAAGGTTGCCCTTCTCACTGGT Primer 2: TAGGCCACCTATTGTGAGTGTGAC Fluorescein-SPC: CCTTAGGAAAGGGCTGTTTCTCT-Phosphate Primer 1: ACTGCATCGAATAATAAAGCC	-	A/-	Pham et al., 2011
	<i>FAD2-1B</i>	Simple probe assay	Primer 1: ACTGCATCGAATAATAAAGCC Primer 2: TGATATTGTCCCGTCCAGC Fluorescein- AGTCCCTTATTCTCATGGAAAAATAAGC- Phosphate	-	T/C (In Bold) C/G (Underlined)	Pham et al., 2010
Raffinose content	<i>RS2</i>	GC-Tail assay	ASP 1: gggggcgttgctacccgacccacagtgaa, ASP 2: gggggcagggcggcgttgctacccgacccacagtgaa Common forward primer: CAGAGGAATAAAATTCATGAGCATA	-	C/T	Dierking and Bilyeu, 2010
Kunitz trypsin inhibitor	<i>kti</i>	PCR	F: CTTTTGTGCCTTACCACCT R: GAATTCATCATCAGAAACTCTA	kti: 420	-	Alves de Moraes et al., 2006

CAPS, Cleaved amplified polymorphic sequence; dCAPS, derived cleaved amplified polymorphic sequence; FLP, Fragment length polymorphism; CNV, Copy number variation; F, Forward primer; R, Reverse primer; ASP, Allele specific primer.

of the underlying gene sequences (identified from soybean) in other leguminous crops could help in identification of genes for quantitative traits without QTL mapping studies in those legumes. Identification of genetic variants of these genes and functional validation could lead to functional marker development in the particular legume crop (**Figure 2**). These functional markers can be very useful for molecular breeding in those leguminous crops without spending much money and time for identifying and characterizing QTLs of interest. Although this strategy will not hold true for all of the quantitative traits but will work for many of the legume specific traits. Further, the newly generated reference sequences of unstudied legume species can be annotated by the knowledge of soybean via translational genomics approaches (Kang et al., 2015).

**Figure 2** describes the process of translational genomics and breeding in legume crops using genic information from soybean. This process is now very much possible due to the availability of several legume specific databases like Legume Information System (LIS) (Gonzales et al., 2005; Dash et al., 2016), LegumeIP 2.0 (Li J. et al., 2016), SoyBase (Grant et al., 2010) and Soybean knowledge base (SoyKB) (Joshi et al., 2014). The first two databases namely LIS and LegumeIP 2.0 are facilitating translational genomics and breeding in major legumes by integrating genetic, genomic, transcriptomic data, and comparative genomics across important legume crops while the SoyBase and SoyKB are facilitating translational genomics and breeding mainly in soybean. Zhang X. et al. (2016) demonstrated the utility of legume specific comparative genomics through *in-silico* analysis and ectopic expression of *E1*, a key flowering regulatory gene in soybean. Flowering regulation pathway in soybean is quite distinct from that identified in model plants (Jung et al., 2012). Comparative genomic analysis of soybean flowering genes using Arabidopsis genome annotation as a model didn't detect *E1* gene until Xia et al. (2012) cloned and characterized it (Jung et al., 2012). However, Zhang X. et al. (2016) was able to identify a total of nine *E1* family genes from six legume species, using soybean *E1* sequence in an *in-silico* analysis. Functional conservation and diversification was observed among *E1* family homologs of the six legumes (Zhang X. et al., 2016). Identification of allelic variants and haplotype analysis of *E1* family genes among legume crop specific germplasm will help in the development of breeder friendly molecular markers.

Translational researches enabled discovery of large number of causative SNPs and InDel alleles for multiple traits opens window for designing crop specific Multi-Trait Functional Marker Assays (MTFMAs), which can be used for screening large breeding populations. These MTFMAs will help in combining several beneficial alleles from multiple parents through MARS in multi-parent breeding populations (**Figure 2**). MTFMAs will also reduce the cost and time of phenotyping large breeding populations and help the breeders of resource poor countries to develop tailor-made varieties of legume crops to suite their specific needs.

## DIFFICULTIES, CHALLENGES, AND POTENTIAL ERRORS IN QTLOMICS AND TRANSLATIONAL GENOMICS

Several factors influence QTL detection in a mapping population i.e., the QTL allele effect, the existence of other linked QTL, the mapping population size, extent of genetic recombination, phenotypic assessment accuracy, genotyping errors, missing data, and environmental effects (Collard et al., 2005; Francia et al., 2005). Further, independent populations chosen from different gene pools may express different phenotypic values in different environments. Another difficulty for translational genomics is un-expected expression of introgressed QTLs in a new genetic background. The genomic region associated with a QTL may hold not just one but several genes, and recombination between those genes would then modify the effect of the introgressed segments. Since, biochemical pathways and enzyme protein sequences are known for some of seed composition traits, candidate gene identification is much easier for these traits resulting in the development of several functional markers in soybean i.e., oleic acid content, isoflavone content, and raffinose content etc. (Dierking and Bilyeu, 2010; Pham et al., 2010, 2011; Wang et al., 2014). However, for agronomic and yield related traits, biochemical pathways are not well-characterized and therefore QTL mapping is the major strategy for identification of underlying genes. Although several genes are involved in the biological network of a quantitative trait, we are only able to detect the variation present in the parental genotypes and gene pool utilized for QTL mapping. This is further complicated by epistatic gene interactions and the target phenotypic environment. So, the major challenge in soybean QTLomics is to capture the genetic variation present among whole soybean gene pool for specific quantitative trait in target environment. The recent advances in NGS technologies and high-throughput marker genotyping enabled genotyping of whole germplasm accessions allowing GWAS for mapping of QTLs in whole gene pool. The large data-sets generated from NGS and high density genotyping requires sound computational algorithms for detection of minor QTLs as well as rare alleles with major effect phenotype. The biggest challenge for QTLomics is to integrate data from genome sequencing, transcriptomics, the epigenome and phenotyping, to make sense of it all. To overcome genetic background effect functional markers used for MAS can be combined with GS models (Deshmukh et al., 2014). The major challenges for translational genomics in legumes includes: (1) limited availability of predictive and functional markers for target traits, (2) limited access to high-throughput genotyping technologies, (3) limited availability of high-throughput, cost-effective and precise phenotyping platforms, (4) availability of automated data management, analyses, and interpretation systems (Varshney et al., 2012, 2015). Genotyping errors, false discovery rate, over-estimation of QTL effect and non-detection of superior but rare alleles are potential errors which QTLomics and translational genomics frequently encounter.

## CONCLUSION AND FUTURE PERSPECTIVES

The identification of genes underlying complex traits in soybean is gaining momentum and a number of genes/candidate genes have been identified for many quantitative traits. The discovery of causative alleles for a number of quantitative traits in soybean could be used to develop functional markers for combining multiple traits from several genotypes in a cost-effective and efficient way. Already several functional markers have been developed and validated for many quantitative traits to facilitate translational research enabled breeding in soybean. This will certainly help in the development of ideal plant types of soybean to suit the specific requirements, besides facilitating molecular breeding in other legume crops via comparative genomics. However, genes for many quantitative traits of economic interest i.e., number of pods, oil content, photosynthetic efficiency, nitrogen use efficiency, thermo-tolerance etc., which are the most desirable genes in soybean breeding are yet to be identified. The potential for identifying genes for these traits lies in the approaches like next-generation sequencing based mapping in

multi-parent cross derived mapping populations, mapping in advanced inter-crossed line mapping populations, association mapping using high throughput genotyping and phenotyping technologies, pooled sequencing based approaches, and targeted genome editing methods, which needs to be taken up at high priority by various soybean genomics research groups and by an International consortium on soybean functional genomics.

## AUTHOR CONTRIBUTIONS

GK planned the MS content, coordinated with other coauthors, contributed in drafting special sections, and finalizing the manuscript. SG, MR, SM, and GS contributed specific sections of this manuscript. SG was involved in planning the content, editing, and finalizing the manuscript.

## ACKNOWLEDGMENTS

The authors are thankful to Dr. V. S. Bhatia, Director, ICAR-Indian Institute of Soybean Research, for providing necessary resources during this research work.

## REFERENCES

- Abe, J., Xu, D. H., Miyano, A., Komatsu, K., Kanazawa, A., and Shimamoto, Y. (2003). Photoperiod-insensitive Japanese soybean landraces differ at two maturity loci. *Crop. Sci.* 43, 1300–1304. doi: 10.2135/cropsci2003.1300
- Afzal, A. J., Srour, A., Goil, A., Vasudaven, S., Liu, T., Samudrala, R., et al. (2013). Homo-dimerization and ligand binding by the leucine-rich repeat domain at *RHG1/RFS2* underlying resistance to two soybean pathogens. *BMC Plant Biol.* 13:43. doi: 10.1186/1471-2229-13-43
- Akond, M., Shiming, L., Lauren, S., Anderson, J. A., Kantartzi, S. K., Meksem, K., et al. (2013). A SNP-based genetic linkage map of soybean using the soySNP6K illumina infinium beadchip genotyping array. *J. Plant Genome Sci.* 1, 80–89. doi: 10.5147/jpgs.2013.0090
- Alves de Moraes, R. M., Bastos Soares, T. C., Colombo, L. R., Salla, M. F. S., Gomes de Almeida Barros, J., Piovesan, N. D., et al. (2006). Assisted selection by specific DNA markers for genetic elimination of the kunitz trypsin inhibitor and lectin in soybean seeds. *Euphytica* 149, 221–226. doi: 10.1007/s10681-005-9069-0
- AVRDC (2003). *AVRDC Progress Report 2002*. AVRDC-The World Vegetable Center, Shanhua.
- Bailey, M. A., Mian, M. A. R., Carter, T. E. J., Ashley, D. A., and Boerma, H. R. (1997). Pod dehiscence of soybean: identification of quantitative trait loci. *J. Hered.* 88, 152–154. doi: 10.1093/oxfordjournals.jhered.a023075
- 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
- Bastien, M., Sonah, H., and Belzile, F. (2014). Genome wide association mapping of *Sclerotinia sclerotiorum* resistance in soybean with a genotyping by sequencing approach. *Plant Genome* 7, 1–13. doi: 10.3835/plantgenome2013.10.0030
- Benhamou, N., and Nicole, M. (1999). Cell biology of plant immunization against microbial infection: the potential of induced resistance in controlling plant diseases. *Plant Physiol. Biochem.* 37, 703–719. doi: 10.1016/S0981-9428(00)86684-X
- Bernard, R. L. (1971). Two major genes for time of flowering and maturity in soybeans. *Crop Sci.* 11, 242–244. doi: 10.2135/cropsci1971.0011183X001100020022x
- Boland, G. J., and Hall, R. (1994). Index of plant hosts of *Sclerotinia sclerotiorum*. *Can. J. Plant Pathol.* 16, 93–108. doi: 10.1080/07060669409500766
- Bolon, Y. T., Joseph, B., Cannon, S. B., Graham, M. A., et al. (2010). Complementary genetic and genomic approaches help characterize the linkage group I seed protein QTL in soybean. *BMC Plant Biol.* 10:41. doi: 10.1186/1471-2229-10-41
- Bonato, E. R., and Vello, N. A. (1999). *E6* a dominant gene conditioning early flowering and maturity in soybeans. *Genet. Mol. Biol.* 22, 229–232. doi: 10.1590/S1415-47571999002000016
- Bradbury, L. M. T., Fitzgerald, T. L., Henry, R. J., Jin, Q. S., and Waters, D. L. E. (2005). The gene for fragrance in rice. *Plant Biotechnol. J.* 3, 363–370. doi: 10.1111/j.1467-7652.2005.00131.x
- Buzzell, R. I. (1971). Inheritance of a soybean flowering response to fluorescent day length conditions. *Can. J. Genet. Cytol.* 13, 703–707.
- Buzzell, R. I., and Voldeng, H. D. (1980). Inheritance of insensitivity to long day length. *Soybean Genet. Newsl.* 7, 26–29.
- Cannon, S. (2008). “Legume comparative genomics” in *Genetics and Genomics of Soybean*, ed G. Stacy (New York, NY: Springer Science + Business Media, LLC), 35–54.
- Chaudhary, J., Patil, G. B., Sonah, H., Deshmukh, R. K., Vuong, T. D., Valliyodan, B. T., et al. (2015). Expanding omics resources for improvement of soybean seed composition traits. *Front. Plant Sci.* 6:1021. doi: 10.3389/fpls.2015.01021
- Chung, W. H., Jeong, N., Kim, J., Lee, W. K., Lee, Y. G., Lee, S. H., et al. (2014). Population structure and domestication revealed by high-depth resequencing of Korean cultivated and wild soybean genomes. *DNA Res.* 21, 153–167. doi: 10.1093/dnares/dst047
- Cober, E. R., and Voldeng, H. D. (2001). A new soybean maturity and photoperiod-sensitivity locus linked to *E1* and *T*. *Crop Sci.* 41, 698–701. doi: 10.2135/cropsci2001.413698x
- Cober, E. R., Molnar, S. J., Charette, M., and Voldeng, H. D. (2010). A new locus for early maturity in soybean. *Crop Sci.* 50, 524–527. doi: 10.2135/cropsci2009.04.0174
- Collard, B. C., and Mackill, D. J. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philos. Trans. R. Soc. B.* 363, 557–572. doi: 10.1098/rstb.2007.2170
- Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., and Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. *Euphytica* 142, 169–196. doi: 10.1007/s10681-005-1681-5
- Concibido, V. C., Lange, D. A., Denny, R. L., Orf, J., and Young, N. D. (1997). Genome mapping of soybean cyst nematode resistance genes in

- Peking, PI90763, and PI88788 using DNA markers. *Crop Sci.* 37, 258–264. doi: 10.2135/cropsci1997.0011183X003700010046x
- Cook, D. E., Bayless, A. M., Wang, K., Guo, X., Song, Q., Jiang, J., et al. (2014). Distinct copy number, coding sequence, and locus methylation patterns underlie *Rhg1*-mediated soybean resistance to soybean cyst nematode. *Plant Physiol.* 165, 630–647. doi: 10.1104/pp.114.235952
- Cook, D. E., Lee, T. G., Guo, X., Melito, S., Wang, K., Bayless, A. M., et al. (2012). Copy number variation of multiple genes at *Rhg1* mediates nematode resistance in soybean. *Science* 338, 1206–1209. doi: 10.1126/science.1228746
- Cregan, P., Mudge, J., Fichus, E. W., Denny, D. R., and Young, N. D. (1999). Two simple sequence repeat markers to select for soybean cyst nematode resistance conditioned by the *rhg1* locus. *Theor. and Appl. Genet.* 99, 811–818. doi: 10.1007/s001220051300
- Csanádi, G., Vollman, J., Stift, G., and Lelley, T. (2001). Seed quality QTLs identified in a molecular map of early maturing soybean. *Theor. Appl. Genet.* 103, 912–919. doi: 10.1007/s001220100621
- Dash, S., Campbell, J. D., Cannon, E. K., Cleary, A. M., Huang, W., Kalberer, S. R., et al. (2016). Legume information system (LegumeInfo.org): a key component of a set of federated data resources for the legume family. *Nucleic Acids Res.* 44, D1181–D1188. doi: 10.1093/nar/gkv1159
- Deshmukh, R., Sonah, H., Patil, G., Chen, W., Prince, S., Mutava, R., et al. (2014). Integrating omic approaches for abiotic stress tolerance in soybean. *Front. Plant Sci.* 5:244. doi: 10.3389/fpls.2014.00244
- Dierking, E., and Bilyeu, K. (2010). New sources of soybean seed meal and oil composition traits identified through TILLING. *BMC Plant Biol.* 9:89. doi: 10.1186/1471-2229-9-89
- Diers, B. W., Keim, P., Fehr, W. R., and Shoemaker, R. C. (1992). RFLP analysis of soybean seed protein and oil content. *Theor. Appl. Genet.* 83, 608–612. doi: 10.1007/BF00226905
- Do, T. D., Chen, H., Hien, V. T., Hamwieh, A., Yamada, T., Sato, T., et al. (2016). *Ncl* synchronously regulates Na<sup>+</sup>, K<sup>+</sup>, and Cl<sup>-</sup> in soybean and greatly increases the grain yield in saline field conditions. *Sci. Rep.* 6:19147. doi: 10.1038/srep19147
- Dwiyanti, M. S., Yamada, T., Sato, M., Abe, J., and Kitamura, K. (2011). Genetic variation of  $\gamma$ -tocopherol methyltransferase gene contributes to elevated  $\alpha$ -tocopherol content in soybean seeds. *BMC Plant Biol.* 11:152. doi: 10.1186/1471-2229-11-152
- 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
- Foley, M. E. (2001). Seed dormancy: an update on terminology, physiological genetics, and quantitative trait loci regulating germinability. *Weed Sci.* 49, 305–317. doi: 10.1614/0043-1745(2001)049[0305:SDAUOT]2.0.CO;2
- Francia, E., Tacconi, G., Crossati, C., Barabaschi, D., Bulgarelli, D., Dall'Aglio, E., et al. (2005). Marker assisted selection in crop plants. *Plant Cell* 82, 317–342. doi: 10.1007/s11240-005-2387-z
- Fuller, D. Q., and Allaby, R. (2009). Seed dispersal and crop domestication: shattering, germination and seasonality in evolution under cultivation. *Ann. Plant Rev.* 38, 238–295. doi: 10.1002/9781444314557.ch7
- Funatsuki, H., Hajika, M., Hagihara, S., Yamada, T., Tanaka, Y., Tsuji, H., et al. (2008). Confirmation of the location and the effects of a major QTL controlling pod dehiscence, *qPDH1*, in soybean. *Breed. Sci.* 58, 63–69. doi: 10.1270/jsbbs.58.63
- Funatsuki, H., Ishimoto, M., Tsuji, H., Kawaguchi, K., Hajika, M., and Fujino, K. (2006). Simple sequence repeat markers linked to a major QTL controlling pod shattering in soybean. *Plant Breed.* 125, 195–197. doi: 10.1111/j.1439-0523.2006.01199.x
- Funatsuki, H., Suzuki, M., Hirose, A., Inaba, H., Yamada, T., Hajika, M., et al. (2014). Molecular basis of a shattering resistance boosting global dissemination of soybean. *Proc. Natl. Acad. Sci. U.S.A.* 111, 17797–17802. doi: 10.1073/pnas.1417282111
- Fushimi, T., and Masuda, R. (2001). “2-acetyl-1-pyrroline concentration of the vegetable soybean,” in *Proceeding of the 2nd International Vegetable Soybean Conference*, eds T. Lumpkin and S. Shanmugasundaram (Pullman, DC: Washington State University), 39.
- Gao, M., and Zhu, H. (2013). Fine mapping of a major quantitative trait locus that regulates pod shattering in soybean. *Mol. Breed.* 32, 485–491. doi: 10.1007/s11032-013-9868-2
- Gonzales, M. D., Archuleta, E., Farmer, A., Gajendran, K., Grant, D., Shoemaker, R., et al. (2005). The Legume Information System (LIS): an integrated information resource for comparative legume biology. *Nucleic Acids Res.* 33, D660–D665. doi: 10.1093/nar/gki128
- Grant, D., Nelson, R. T., Cannon, S. B., and Shoemaker, R. C. (2010). SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Res.* 38, D843–D846. doi: 10.1093/nar/gkp798
- Guan, R., Chen, J., Jiang, J., and Qiu, L. (2014a). Mapping and validation of a dominant salt tolerance gene in the cultivated soybean (*Glycine max*) variety Tiefeng 8. *Crop J.* 2, 358–365. doi: 10.1016/j.cj.2014.09.001
- Guan, R., Qu, Y., Guo, Y., Yu, L., Liu, Y., Jiang, J., et al. (2014b). Salinity tolerance in soybean is modulated by natural variation in *GmSALT3*. *Plant J.* 80, 937–950. doi: 10.1111/tpj.12695
- Gupta, S., Bhatia, V. S., Kumawat, G., Thakur, D., Singh, G., Tripathi, R., et al. (2016). Genetic analyses for deciphering the status and role of photoperiodic and maturity genes in major Indian soybean cultivars. *J. Genet.*
- Gutierrez-Gonzalez, J. J., Vuong, T. D., Zhong, R., Yu, O., Lee, J. D., Shannon, G., et al. (2011). Major locus and other novel additive and epistatic loci involved in modulation of isoflavone concentration in soybean seeds. *Theor. Appl. Genet.* 123, 1375–1385. doi: 10.1007/s00122-011-1673-x
- Gutierrez-Gonzalez, J. J., Wu, X. L., Gillman, J. D., Lee, J. D., Zhong, R., Yu, O., et al. (2010). Intricate environment-modulated genetic networks control isoflavone accumulation in soybean seeds. *BMC Plant Biol.* 10:105. doi: 10.1186/1471-2229-10-105
- Ha, B. K., Vuong, T. D., Velusamy, V., Nguyen, H. T., Shannon, J. G., and Lee, J. D. (2013). Genetic mapping of quantitative trait loci conditioning salt tolerance in wild soybean (*Glycine soja*) PI 483463. *Euphytica* 193, 79–88. doi: 10.1007/s10681-013-0944-9
- Hamwieh, A., and Xu, D. H. (2008). Conserved salt tolerance quantitative trait locus (QTL) in wild and cultivated soybeans. *Breed. Sci.* 58, 355–359. doi: 10.1270/jsbbs.58.355
- Hamwieh, A., Tuyen, D., Cong, H., Benitez, E., Takahashi, R., and Xu, D. (2011). Identification and validation of a major QTL for salt tolerance in soybean. *Euphytica* 179, 451–459. doi: 10.1007/s10681-011-0347-8
- Hartwig, E. E., and Lehman, S. G. (1951). Inheritance of resistance to the bacterial pustule disease in soybeans. *Agron. J.* 43, 226–229. doi: 10.2134/agronj1951.00021962004300050005x
- Hewezi, T., Howe, P., Maier, T. R., Hussey, R. S., Mitchum, M. G., Davis, E. L., et al. (2008). Cellulose binding protein from the parasitic nematode *Heterodera schachtii* interacts with *Arabidopsis pectin methyltransferase*: cooperative cell wall modification during parasitism. *Plant Cell* 20, 3080–3093. doi: 10.1105/tpc.108.063065
- Hymowitz, T., and Singh, R. J. (1987). “Taxonomy and speciation,” in *Soybeans: Improvement Production and Uses*, ed J. R. Wilcox (Madison, WI: ASA), 23–48.
- Innes, R. W., Ameline-Torregrosa, C., Ashfield, T., Cannon, E., Cannon, S. B., et al. (2008). Differential accumulation of retroelements and diversification of NB-LRR disease resistance genes in duplicated regions following polyploidy in the ancestor of soybean. *Plant Physiol.* 148, 1740–1759. doi: 10.1104/pp.108.127902
- Iqura, E., Humira, S., and François, B. (2015). Association mapping of QTLs for sclerotinia stem rot resistance in a collection of soybean plant introductions using a genotyping by sequencing (GBS) approach. *BMC Plant Biol.* 15:5. doi: 10.1186/s12870-014-0408-y
- Jain, A., Vasconcelos, M. J., Raghobhama, K., and Sahi, S. V. (2007). Molecular mechanisms of plant adaptation to phosphate deficiency. *Plant Breed. Rev.* 29, 359–419. doi: 10.1002/9780470168035.ch7
- Jannink, J. L., Lorenz, A. J., and Iwata, H. (2010). Genomic selection in plant breeding: from theory to practice. *Brief. Funct. Genomics* 9, 166–177. doi: 10.1093/bfpp/elq001
- Jiang, B., Nan, H., Gao, Y., Tang, L., Yue, Y., et al. (2014). Allelic combinations of soybean maturity loci *E1*, *E2*, *E3* and *E4* result in diversity of maturity and adaptation to different latitudes. *PLoS ONE* 9:e106042. doi: 10.1371/journal.pone.0106042
- Joshi, T., Fitzpatrick, M. R., Chen, S., Liu, Y., Zhang, H., Endacott, R. Z., et al. (2014). Soybean knowledge base (SoyKB): a web resource for integration of soybean translational genomics and molecular breeding. *Nucleic Acids Res.* 42, D1245–D1252. doi: 10.1093/nar/gkt905

- Jung, C. H., Wong, C. E., Singh, M. B., and Bhalla, P. L. (2012). Comparative genomic analysis of soybean flowering genes. *PLoS ONE* 7:e38250. doi: 10.1371/journal.pone.0038250
- Juwattanasomran, R., Somta, P., Chankaew, S., Shimizu, T., Wongpornchai, S., Kaga, A., et al. (2011). A SNP in *GmBADH2* gene associates with fragrance in vegetable soybean variety "Kaori" and SNAP marker development for the fragrance. *Theor. Appl. Genet.* 122, 533–541 doi: 10.1007/s00122-010-1467-6
- Juwattanasomran, R., Somta, P., Kaga, A., Chankaew, S., Shimizu, T., Sorajjapinun, W., et al. (2012). Identification of a new fragrance allele in soybean and development of its functional marker. *Mol. Breed.* 29, 13–21. doi: 10.1007/s11032-010-9523-0
- Kadam, S., Vuong, T. D., Qiu, D., Meinhardt, C. G., Song, L., et al. (2016). Genomic-assisted phylogenetic analysis and marker development for next generation soybean cyst nematode resistance breeding. *Plant Sci.* 242, 342–350. doi: 10.1016/j.plantsci.2015.08.015
- Kang, Y. J., Lee, T., Lee, J., Shim, S., Jeong, H., Satyawan, D., et al. (2015). Translational genomics for plant breeding with the genome sequence explosion. *Plant Biotechnol. J.* 14, 1057–1069. doi: 10.1111/pbi.12449
- Kazi, S., Shultz, J. L., Bashir, R., Afzal, A. J., Njiti, V. N., and Lightfoot, D. A. (2008). Separate loci underlie resistance to soybean sudden death syndrome in 'Hartwig' by 'Flyer'. *Theor. Appl. Genet.* 116, 967–977. doi: 10.1007/s00122-008-0728-0
- Keim, P., Diers, B. W., and Shoemaker, R. C. (1990b). Genetic analysis of soybean hard seediness with molecular markers. *Theor. Appl. Genet.* 79, 465–469. doi: 10.1007/BF00226154
- Keim, P., Diers, B. W., Olson, T. C., and Shoemaker, R. C. (1990a). RFLP mapping in soybean: association between marker loci and variation in quantitative traits. *Genetics* 126, 735–742.
- Kim, D. H., Kim, K. H., Van, K., Kim, M. Y., and Lee, S. H. (2010). Fine mapping of a resistance gene to bacterial leaf pustule in soybean. *Theor. Appl. Genet.* 120, 1443–1450. doi: 10.1007/s00122-010-1266-0
- Kim, M. S., Hyten, D. L., Bent, A. F., and Diers, B. W. (2010). Fine mapping of the SCN resistance locus *rhg1-b* from PI 88788. *Plant Genome* 3:81. doi: 10.3835/plantgenome2010.02.0001
- Koebner, R. M., and Varshney, R. K. (2006). "Development and application of genomic models for large-crop plant genomes" in *Model Plants and Crop Improvement*, eds R. K. Varshney and R. M. Koebner (New York, NY: CRC Press), 1–10.
- Kong, F., Liu, B., Xia, Z., Sato, S., Kim, B. M., Watanabe, S., et al. (2010). Two coordinately regulated homologs of flowering locus T are involved in the control of photoperiodic flowering in soybean. *Plant Physiol.* 154, 1220–1231. doi: 10.1104/pp.110.160796
- Kong, F., Nan, H., Cao, D., Li, Y., Wu, F., Wang, J., et al. (2014). A new dominant gene *E9* conditions early flowering and maturity in soybean. *Crop Sci.* 54, 2529–2535. doi: 10.2135/cropsci2014.03.0228
- Kumar, V., Rani, A., and Rawal, R. (2013). Deployment of gene specific marker in development of kunitz trypsin inhibitor free soybean genotypes. *Indian J. Exp. Biol.* 51, 1125–1129.
- Kummerova, M. (1986). Localization of acid phosphatase in maize root under phosphorus deficiency. *Biol. Plant* 28, 270–274. doi: 10.1007/bf02902291
- Lam, H. M., Xu, X., Liu, X., Chen, W., Yang, G., Wong, F. L., et al. (2010). Re sequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat. Genet.* 42, 1053–1059. doi: 10.1038/ng.715
- Langewisch, T., Zhang, H., Vincent, R., Joshi, T., Xu, D., et al. (2014). Major soybean maturity gene haplotypes revealed by SNPviz analysis of 72 sequenced soybean genomes. *PLoS ONE* 9:e94150. doi: 10.1371/journal.pone.0094150
- Lee, G. J., Carter, T. E. Jr., Villagarcia, M. R., Li, Z., Zhou, X., Gibbs, M. O., et al. (2004). A major QTL conditioning salt tolerance in S-100 soybean and descendent cultivars. *Theor. Appl. Genet.* 109, 1610–1619. doi: 10.1007/s00122-004-1783-9
- Lee, Y. G., Jeong, N., Kim, J. H., Lee, K., Kim, K. H., Pirani, A., et al. (2015). Development, validation and genetic analysis of a large soybean SNP genotyping array. *Plant J.* 81, 625–636. doi: 10.1111/tj.12755
- Lestari, P., Van, K., Lee, J., Kang, Y. J., and Lee, S. H. (2013). Gene divergence of homeologous regions associated with a major seed protein content QTL in soybean. *Front. Plant Sci.* 4:176. doi: 10.3389/fpls.2013.00176
- Li, H., Yang, Y., Zhang, H., Chu, S., Zhang, X., Yin, D., et al. (2016). A genetic relationship between phosphorus efficiency and photosynthetic traits in soybean as revealed by QTL analysis using a high-density genetic map. *Front. Plant Sci.* 7:924. doi: 10.3389/fpls.2016.00924
- Li, J., Dai, X., Zhuang, Z., and Zhao, P. X. (2016). LegumeIP 2.0—a platform for the study of gene function and genome evolution in legumes. *Nucleic Acids Res.* 44, D1189–D1194. doi: 10.1093/nar/gkv1237
- Li, L., Guo, N., Niu, J., Wang, Z., Cui, X., et al. (2016a). Loci and candidate gene identification for resistance to *Phytophthora sojae* via association analysis in soybean [*Glycine max* (L.) Merr.]. *Mol. Genet. Genomics* 291, 1095–1103. doi: 10.1007/s00438-015-1164-x
- Li, L., Lin, F., Wang, W., Ping, J., Fitzgerald, J. C., Zhao, M., et al. (2016b). Fine mapping and candidate gene analysis of two loci conferring resistance to *Phytophthora sojae* in soybean. *Theor. Appl. Genet.* 129, 2379–2386. doi: 10.1007/s00122-016-2777-0
- Li, S., Qian, Q., Fu, Z., Zeng, D., Meng, X., Kyozuka, J., et al. (2009). *Short panicle1* encodes a putative PTR family transporter and determines rice panicle size. *Plant J.* 58, 592–605. doi: 10.1111/j.1365-313X.2009.03799.x
- Li, Y. H., Zhao, S. C., Ma, J. X., Li, D., Yan, L., Li, J., et al. (2013). Molecular footprints of domestication and improvement in soybean revealed by whole genome re-sequencing. *BMC Genomics* 14:579. doi: 10.1186/1471-2164-14-579
- Li, Y. H., Zhou, G., Ma, J., Jiang, W., Jin, L. G., Zhang, Z., et al. (2014). *Denovo* assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nat. Biotechnol.* 32, 1045–1052. doi: 10.1038/nbt.2979
- Lin, F., Zhao, M., Ping, J., Johnson, A., Zhang, B., Abney, T. S., et al. (2013). Molecular mapping of two genes conferring resistance to *Phytophthora sojae* in a soybean landrace PI 567139B. *Theor. Appl. Genet.* 126, 2177–2185. doi: 10.1007/s00122-013-2127-4
- Lin, S., Cianzio, S., and Shoemaker, R. C. (1997). Mapping genetic loci for iron deficiency chlorosis in soybean. *Mol. Breed.* 3, 219–229. doi: 10.1023/A:1009637320805
- Liu, B., Fujita, T., Yan, Z. H., Sakamoto, S., Xu, D., and Abe, J. (2007). QTL Mapping of domestication-related traits in soybean (*Glycine max*). *Ann. Bot.* 100, 1027–1038. doi: 10.1093/aob/mcm149
- Liu, B., Kanazawa, A., Matsumura, H., Takahashi, R., Harada, K., and Abe, J. (2008). Genetic redundancy in soybean photoreponses associated with duplication of *phytochrome A* gene. *Genetics* 180, 996–1007. doi: 10.1534/genetics.108.092742
- Liu, S., Kandath, P. K., Warren, S. D., Yeckel, G., Heinz, R., Alden, J., et al. (2012). A soybean cyst nematode resistance gene points to a new mechanism of plant resistance to pathogens. *Nature* 492, 256–260. doi: 10.1038/nature11651
- Lozovaya, V. V., Lygin, A. V., Zernova, O. V. L. I. S. X., Hartman, G. L., and Widholm, M. (2004). Isoflavonoid accumulation in soybean hairy roots upon treatment with *Fusarium solani*. *Plant Physiol. Biochem.* 42, 671–679. doi: 10.1016/j.plaphy.2004.06.007
- Ma, Y., Kan, G., Zhang, X., Wang, Y., Zhang, W., Du, H., et al. (2016). Quantitative trait Loci (QTL) mapping for glycinin and  $\beta$ -conglycinin contents in soybean (*Glycine max* L. Merr.). *J. Agric. Food Chem.* 64, 3473–3483. doi: 10.1021/acs.jafc.6b00167
- Manavalan, L. P., Prince, S. J., Musket, T. A., Chaky, J., Deshmukh, R., Vuong, T. D., et al. (2015). Identification of novel QTL governing root architectural traits in an interspecific soybean population. *PLoS ONE* 10:e0120490. doi: 10.1371/journal.pone.0120490
- Matson, A. L., and Williams, L. F. (1965). Evidence of a fourth gene for resistance to the soybean cyst nematode. *Crop Sci.* 5:477.
- McBlain, B. A., and Bernard, R. L. (1987). A new gene affecting the time of flowering and maturity in soybeans. *J. Hered.* 78, 160–162.
- Munns, R., and Tester, M. (2008). Mechanisms of salinity tolerance. *Annu. Rev. Plant Biol.* 59, 651–681. doi: 10.1146/annurev.arplant.59.032607.092911
- Nakaya, A., and Isobe, S. N. (2012). Will genomic selection be a practical method for plant breeding? *Ann. Bot.* 110, 1303–1316. doi: 10.1093/aob/mcs109
- Narvel, J. M., Jakkula, L. R., Phillips, D. V., Wang, T., Lee, S. H., and Boerma, H. R. (2001). Molecular mapping of Rxp conditioning reaction to bacterial pustule in soybean. *J. Hered.* 92, 267–270. doi: 10.1093/jhered/92.3.267
- Niu, X., Tang, W., Huang, W., Ren, G., Wang, Q., Luo, D., et al. (2008). RNAi-directed downregulation of *OsBADH2* results in aroma (2-acetyl-1-pyrroline) production in rice (*Oryza sativa* L.). *BMC Plant Biol.* 8:100. doi: 10.1186/1471-2229-8-100
- Njiti, V. N., Doubler, T. W., Suttner, R. J., Gray, L. E., Gibson, P. T., and Lightfoot, D. A. (1998). Resistance to soybean sudden death syndrome and root

- colonization by *Fusarium solani* f. sp. *glycine* in near-isogenic lines. *Crop Sci.* 38, 472–477. doi: 10.2135/cropsci1998.0011183X003800020033x
- Pandey, M. K., Roorikiwal, M., Singh, V. K., Ramalingam, A., Kudapa, H., Thudi, M., et al. (2016). Emerging genomic tools for legume breeding: current status and future prospects. *Front. Plant Sci.* 7:455. doi: 10.3389/fpls.2016.00455
- Patil, G., Do, T., Vuong, T. D., Valliyodan, B., Lee, J. D., Chaudhary, J., et al. (2016). Genomic-assisted haplotype analysis and the development of highthroughput SNP markers for salinity tolerance in soybean. *Sci. Rep.* 6:19199. doi: 10.1038/srep19199
- Peiffer, G. A., King, K. E., Severin, A. J., May, G. D., Cianzio, S. R., Lin, S. F., et al. (2012). Identification of candidate genes underlying an iron efficiency quantitative trait locus in soybean. *Plant Physiol.* 158, 1745–1754. doi: 10.1104/pp.111.189860
- Pham, A.-T., Lee, J.-D., Shannon, J. G., and Bilyeu, K. (2010). Mutant alleles of *FAD2-1A* and *FAD2-1B* combine to produce soybeans with the high oleic acid seed oil trait. *BMC Plant Biol.* 10:195. doi: 10.1186/1471-2229-10-195
- Pham, A.-T., Lee, J.-D., Shannon, J. G., and Bilyeu, K. (2011). A novel *FAD2-1A* allele in a soybean plant introduction offers an alternate means to produce soybean seed oil with 85% oleic acid content. *Theor. Appl. Genet.* 123, 793–802. doi: 10.1007/s00122-011-1627-3
- Phang, T. H., Shao, G., and Lam, H. M. (2008). Salt tolerance in soybean. *J. Integr. Plant Biol.* 50, 1196–1212. doi: 10.1111/j.1744-7909.2008.00760.x
- Potts, H. C., Duangpatra, J., Hairston, W. G., and Delouche, J. C. (1978). Some influences of hardseededness on soybean seed quality. *Crop Sci.* 18, 221–224. doi: 10.2135/cropsci1978.0011183X001800020006x
- Primomo, V. S., Poysa, V., Ablett, G. R., Jackson, C. J., Gijzen, M., and Rajcan, I. (2005). Mapping QTL for individual and total isoflavone content in soybean seeds. *Crop Sci.* 45, 2454–2462. doi: 10.2135/cropsci2004.0672
- Prince, S. J., Song, L., Qiu, D., Maldonado Dos Santos, J. V., Chai, C., Joshi, T., et al. (2015). Genetic variants in root architecture-related genes in a *Glycine soja* accession, a potential resource to improve cultivated soybean. *BMC Genomics* 16:132. doi: 10.1186/s12864-015-1334-6
- Qi, X., Li, M. W., Xie, M., Liu, X., Ni, M., Shao, G., et al. (2014). Identification of a novel salt tolerance gene in wild soybean by whole-genome sequencing. *Nat. Commun.* 5:4340. doi: 10.1038/ncomms5340
- Qiu, J., Wang, Y., Wu, S., Wang, Y. Y., Ye, C. Y., Bai, X., et al. (2014). Genome re-sequencing of semi-wild soybean reveals a complex *soja* population structure and deep introgression. *PLoS ONE* 9:e108479. doi: 10.1371/journal.pone.0108479
- Ratnaparkhe, M. B., Ramesh, S. V., Kumawat, G., Husain, S. M., and Gupta, S. (2014). “Advances in soybean genomics,” in *Legumes in the Omic Era*, eds S. Gupta, N. Nadarajan, and D. S. Gupta (New York, NY: Springer Science+Business Media), 41–72.
- Ray, J. D., Hinson, K., Mankono, E. B., and Malo, F. M. (1995). Genetic control of a long-juvenile trait in soybean. *Crop Sci.* 35, 1001–1006. doi: 10.2135/cropsci1995.0011183X003500040012x
- Ribaut, J. M., and Ragot, M. (2006). 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
- Roberts, R. C., and Briggs, D. R. (1965). Isolation and characterization of the 7S component of soybean globulins. *Cereal Chem.* 42:71.
- Rochfort, S., and Panozzo, J. (2007). Phytochemicals for health, the role of pulses. *J. Agric. Food Chem.* 55, 7981–7994. doi: 10.1021/jf071704w
- Rosso, M. L., Burleson, S. A., Maupin, L. M., and Rainey, K. M. (2011). Development of breeder-friendly markers for selection of *MIP51* mutations in soybean. *Mol. Breeding* 28, 127–132. doi: 10.1007/s11032-011-9573-y
- Roy, K. W., Rupe, J. C., Hershman, D. E., and Abney, T. S. (1997). Sudden death syndrome of soybean. *Plant Dis.* 81, 1100–1111. doi: 10.1094/PDIS.1997.81.10.1100
- Ruben, E., Jamai, A., Afzal, A. J., Njiti, V. N., Triwitayakorn, K., Iqbal, M. J., et al. (2006). Genomic analysis of the *Rhg1* locus: candidate genes that underlie soybean resistance to the cyst nematode. *Mol. Genet. Genomics* 276, 503–516. doi: 10.1007/s00438-006-0150-8
- Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., et al. (2010). Genome sequence of the palaeopolyploid soybean. *Nature* 463, 178–183. doi: 10.1038/nature08670
- Seboldt, A. M., Shoemaker, R. C., and Diers, B. W. (2000). Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. *Crop Sci.* 40, 1438–1444. doi: 10.2135/cropsci2000.4051438x
- Servin, B., Martin, O. C., Mézard, M., and Hospital, F. (2004). Toward a theory of marker-assisted gene pyramiding. *Genetics* 168, 513–523. doi: 10.1534/genetics.103.023358
- Shanmugasundaram, S., Cheng, S. T., Huang, M. T., and Yan, M. R. (1991). “Quality requirement and improvement of vegetable soybean,” in *Vegetable Soybean: Research Needs for Production and Quality Improvement*, ed S. Shanmugasundaram (Taiwan: Asian Vegetable Research and Development Center), 30–42.
- Shi, W., Yang, Y., Chen, S., and Xu, M. (2008). Discovery of a new fragrance allele and the development of functional markers for the breeding of fragrant rice varieties. *Mol. Breed.* 22, 185–192. doi: 10.1007/s11032-008-9165-7
- Shivakumar, M., Verma, K., Talukdar, A., Lal, S. K., Kumar, A., and Mukherjee, K. (2016). Introgression of null allele of Kunitz trypsin inhibitor through marker-assisted backcross breeding in soybean (*Glycine max* L. Merr.). *BMC Genet.* 17:106. doi: 10.1186/s12863-016-0413-2
- Silva, M. F. D., Ivan, S., Silva, S. F. V., Adésio, F., Everaldo, G. B., and Moreira, M. A. (2007). Validation of microsatellite markers for assisted selection of soybean resistance to cyst nematode races 3 and 14. *Pesq. Agropec. Bras. Brasília* 42, 1143–1150. doi: 10.1590/s0100-204x2007000800011
- Sonah, H., Bastien, M., Iquira, E., Tardivel, A., Légaré, G., Boyle, B., et al. (2013). An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. *PLoS ONE* 8:e54603. doi: 10.1371/journal.pone.0054603
- Sonah, H., O’Donoghue, L., Cober, E., Rajcan, I., and Belzile, F. (2015). Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. *Plant Biotechnol. J.* 13, 211–221. doi: 10.1111/pbi.12249
- Song, Q., Hyten, D. L., Jia, G., Quigley, C. V., Fickus, E. W., Nelson, R. L., et al. (2013). Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. *PLoS ONE* 8:e54985. doi: 10.1371/journal.pone.0054985
- Spiller, S., and Terry, N. (1980). Limiting factors in photosynthesis. II. Iron stress diminishes photochemical capacity by reducing the number of photosynthetic units. *Plant Physiol.* 65, 121–125. doi: 10.1104/pp.65.1.121
- Srour, A., Afzal, A. J., Blahut-Beatty, L., Hemmati, N., Simmonds, D. H., Li, W., et al. (2012). The receptor like kinase at *Rhg1-a/Rfs2* caused pleiotropic resistance to sudden death syndrome and soybean cyst nematode as a transgene by altering signaling responses. *BMC Genomics* 13:368. doi: 10.1186/1471-2164-13-368
- Stewart, D. W., Cober, E. R., and Bernard, R. L. (2003). Modeling genetic effects on the photothermal response of soybean phenological development. *Agron. J.* 95, 65–70. doi: 10.2134/agronj2003.0065
- Subramanian, S., Hu, X., Lu, G., Odelland, J. T., and Yu, O. (2004). The promoters of two isoflavone synthase genes respond differentially to nodulation and defense signals in transgenic soybean roots. *Plant Mol. Biol.* 54, 623–639. doi: 10.1023/B:PLAN.0000040814.28507.35
- Subramanian, S., Stacey, G., and Yu, O. (2006). Endogenous isoflavones are essential for the establishment of symbiosis between soybean and *Bradyrhizobium japonicum*. *Plant J.* 48, 261–273. doi: 10.1111/j.1365-313X.2006.02874.x
- Subramanian, S., Stacey, G., and Yu, O. (2007). Distinct, crucial roles of flavonoids during legume nodulation. *Trends Plant Sci.* 12, 282–285. doi: 10.1016/j.tplants.2007.06.006
- Sun, L., Miao, Z., Cai, C., Zhang, D., Zhao, M., Wu, Y., et al. (2015). *GmHs1-1*, encoding a calcineurin-like protein, controls hard-seededness in soybean. *Nature Genet.* 47, 939–943. doi: 10.1038/ng.3339
- Suzuki, M., Fujino, K., Nakamoto, Y., Ishimoto, M., and Funatsuki, H. (2010). Fine mapping and development of DNA markers for the *qPDH1* locus associated with pod dehiscence in soybean. *Mol. Breed.* 25, 407–418. doi: 10.1007/s11032-009-9340-5
- Tardivel, A., Sonah, H., Belzile, F., and O’Donoghue, L. S. (2014). Rapid identification of alleles at the soybean maturity gene *E3* using genotyping by sequencing and a haplotype-based approach. *Plant Genome* 7, 1–9. doi: 10.3835/plantgenome2013.10.0034
- Taylor, S. E., Terry, N., and Huston, R. P. (1982). Limiting factors in photosynthesis. III. Effects of iron nutrition on the activities of three regulatory



- enzymes of photosynthetic carbon metabolism. *Plant Physiol.* 70, 1541–1543. doi: 10.1104/pp.70.5.1541
- Thakare, D., Kumudini, S., and Dinkins, R. D. (2011). The alleles at the E1 locus impact the expression pattern of two soybean FT-like genes shown to induce flowering in *Arabidopsis*. *Planta* 234, 933–943. doi: 10.1007/s00425-011-1450-8
- Thanh, V. H., and Shibasaki, K. (1976). Major proteins of soybean seeds. A straightforward fractionation and their characterization. *J. Agric. Food. Chem.* 24:1118.
- Triwityakorn, K., Njiti, V. N., Iqbal, M. J., Yaegashi, S., Town, C. D., and Lightfoot, D. A. (2005). Genomic analysis of a region encompassing *QRfs1* and *QRfs2*: genes that underlie soybean resistance to sudden death syndrome. *Genome* 48, 125–138. doi: 10.1139/g04-103
- Tsubokura, Y., Matsumura, H., Xu, M., Katayose, Y., Kanamori, H., Yuan, X., et al. (2013). Genetic variation in the maturity locus *E4* is involved in adaptation to high latitudes in soybean. *Agronomy* 3, 117–134. doi: 10.3390/agronomy3010117
- Tsubokura, Y., Watanabe, S., Xia, Z., Kanamori, H., Yamagata, H., Kaga, A., et al. (2014). Natural variation in the genes responsible for maturity loci E1, E2, E3 and E4 in soybean. *Ann. Bot.* 113, 429–441. doi: 10.1093/aob/mct269
- Tsuchiya, T. (1987). Physiological and genetic-analysis of pod shattering in soybeans. *Jarq-Japan Agric. Res. Q* 21, 166–175.
- Tsukamoto, C., Shimada, S., Igita, K., Kudou, S., Kokubun, M., Okubo, K., et al. (1995). Factors affecting isoflavone content in soybean seeds: changes in isoflavones, saponins, and composition of fatty acids at different temperatures during seed development. *J. Agric. Food Chem.* 43, 1184–1192. doi: 10.1021/jf00053a012
- Ujiie, A., Yamada, T., Fujimoto, K., Endo, Y., and Kitamura, K. (2005). Identification of soybean varieties with high  $\alpha$ -tocopherol content. *Breed. Sci.* 55, 123–125. doi: 10.1270/jsbbs.55.123
- Valliyodan, B., Qiu, D., Patil, G., Zeng, P., Huang, J., Dai, L., et al. (2016). Landscape of genomic diversity and trait discovery in soybean. *Sci. Rep.* 6:23598. doi: 10.1038/srep23598
- Van Eenennaam, A. L., Lincoln, K., Durrett, T. P., Valentin, H. E., Shewmaker, C. K., et al. (2003). Engineering vitamin E content: from *Arabidopsis* mutant to soy oil. *Plant Cell* 15, 3007–3019. doi: 10.1105/tpc.015875
- Van, K., Ha, B. K., Kim, M. Y., Moon, J. K., Paek, N. C., Heu, S., et al. (2004). SSR mapping of genes conditioning soybean resistance to six isolates of *Xanthomonas axonopodis* pv. *glycines*. *Korean J. Genet.* 26, 47–54.
- Vanavichit, A., Tragoonrung, S., Toojinda, T., Wanchana, S., and Kamolsukyonyong, W. (2008). *Transgenic rice Plants with Reduced Expression of Os2AP and Elevated Levels of 2-Acetyl-1-Pyrroline*. US Patent 731918.
- Varshney, R. K., Himabindu, K., Lekha, P., Annapurna, C., Mahendar, T., Abhishek, B., et al. (2015). Translational genomics in agriculture: some examples in grain legumes. *Critical Rev. Plant Sci.* 34, 169–194. doi: 10.1080/07352689.2014.897909
- Varshney, R. K., Kudapa, H., Roorkiwal, M., Thudi, M., Pandey, M., Saxena, R., et al. (2012). Advances in genetics and molecular breeding of three legume crops of semi-arid tropics using next-generation sequencing and high-throughput genotyping technologies. *J. Biosci.* 37, 811–820.
- Verdier, J., Kakar, K., Gallardo, K., Le Signor, C., Aubert, G., Schlereth, A., et al. (2008). Gene expression profiling of *M. truncatula* transcription factors identifies putative regulators of grain legume seed filling. *Plant Mol. Biol.* 67, 567–580. doi: 10.1007/s11103-008-9320-x
- Vuong, T. D., Sonah, H., Meinhardt, C. G., Deshmukh, R., Kadam, S., Nelson, R. L., et al. (2015). Genetic architecture of cyst nematode resistance revealed by genome-wide association study in soybean. *BMC Genomics* 16:593. doi: 10.1186/s12864-015-1811-y
- Walling, J. G., Shoemaker, R. C., Young, N., Mudge, J., and Jackson, S. (2006). Chromosome level homeology in paleopolyploid soybean (*Glycine max*) revealed through integration of genetic and chromosome maps. *Genetics* 172, 1893–1900. doi: 10.1534/genetics.105.051466
- Walters, S. A., and Baker, K. R. (1994). Current distribution of five major *Meloidogyne* species in the United States. *Plant Dis.* 78, 772–774. doi: 10.1094/PD-78-0772
- Wanchana, S. (2005). *Identification of Genes Controlling Grain Aroma and Amylose Content for Positional Cloning and Marker-Assisted Selection Program in Rice (Oryza sativa L.)*. Ph.D. Thesis, Kasetsart University, Thailand.
- Wang, E., Wang, J., Zhu, X., Hao, W., Wang, L., Li, Q., et al. (2008). Control of rice grain-filling and yield by a gene with a potential signature of domestication. *Nat. Genet.* 40, 1370–1374. doi: 10.1038/ng.220
- Wang, H.-Y., Klatte, M., Jakoby, M., Baumlein, H., Weisshaar, B., and Bauer, P. (2007). Iron deficiency-mediated stress regulation of four subgroup *Ib BHLH* genes in *Arabidopsis thaliana*. *Planta* 226, 897–908. doi: 10.1007/s00425-007-0535-x
- Wang, J., Chu, S., Zhang, H., Zhu, Y., Cheng, H., and Yu, D. (2016). Development and application of a novel genome-wide SNP array reveals domestication history in soybean. *Sci. Rep.* 6:20728. doi: 10.1038/srep20728
- Wang, L., Song, X., Gu, L., Li, X., Cao, S., Chu, C., et al. (2013). NOT2 proteins promote polymerase II-dependent transcription and interact with multiple micro RNA biogenesis factors in *Arabidopsis*. *Plant Cell* 25, 715–727. doi: 10.1105/tpc.112.105882
- Wang, X., Wang, Y., Tian, J., Lim, B., Yan, X., et al. (2009). Overexpressing *AtPAP15* enhances phosphorus efficiency in soybean. *Plant Physiol.* 151, 233–240. doi: 10.1104/pp.109.138891
- Wang, Y., Han, Y., Teng, W., et al. (2014). Expression quantitative trait loci infer the regulation of isoflavone accumulation in soybean (*Glycine max* L. Merr.) seed. *BMC Genomics* 15:680. doi: 10.1186/1471-2164-15-680
- Watanabe, S., Hideshima, R., Xia, Z., Tsubokura, Y., et al. (2009). Map-based cloning of the gene associated with the soybean maturity locus *E3*. *Genetics* 182, 1251–1262. doi: 10.1534/genetics.108.098772
- Watanabe, S., Tajuddin, T., Yamanaka, N., Hayashi, M., and Harada, K. (2004). Analysis of QTLs for reproductive development and seed quality traits in soybean using recombinant inbred lines. *Breed. Sci.* 54, 399–407. doi: 10.1270/jsbbs.54.399
- Watanabe, S., Xia, Z., Hideshima, R., Tsubokura, Y., Sato, S., and Harada, K. (2011). A map-based cloning strategy employing a residual heterozygous line reveals that the *GIGANTEA* gene is involved in soybean maturity and flowering. *Genetics* 188, 395–407. doi: 10.1534/genetics.110.125062
- Wen, Z., Tan, R., Yuan, J., et al. (2014). Genome-wide association mapping of quantitative resistance to sudden death syndrome in soybean. *BMC Genomics* 15:809. doi: 10.1186/1471-2164-15-809
- Wrather, J. A., Shannon, G., Balarin, R., Carregal, L., Escobar, R., Gupta, G. K., et al. (2010). Effect of diseases on soybean yield in the top eight producing countries in 2006. *Plant Health Prog.* doi: 10.1094/PHP-2010-0125-01-RS. [Epub ahead of print]
- Xia, Z., Watanabe, S., Yamada, T., Tsubokura, Y., Nakashima, H., Zhai, H., et al. (2012). Positional cloning and characterization reveal the molecular basis for soybean maturity locus *E1* that regulates photoperiodic flowering. *Proc. Natl. Acad. Sci. U.S.A.* 109, E2155–E2164. doi: 10.1073/pnas.1117982109
- 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
- Yamanaka, N., Ninomiya, S., Hoshi, M., Tsubokura, Y., Yano, M., Nagamura, Y., et al. (2001). An informative linkage map of soybean reveals QTLs for flowering time, leaflet morphology and regions of segregation distortion. *DNA Res.* 8, 61–72. doi: 10.1093/dnares/8.2.61
- Yang, K., Moon, J. K., Jeong, N., Chun, H. K., Kang, S. T., Back, K., et al. (2011). Novel major quantitative trait loci regulating the content of isoflavone in soybean seeds. *Genes Genom.* 33, 685–692. doi: 10.1007/s13258-011-0043-z
- Zeng, G., Li, D., Han, Y., Teng, W., Wang, J., Qiu, L., et al. (2009). Identification of QTL underlying isoflavone contents in soybean seeds among multiple environments. *Theor. Appl. Genet.* 118, 1455–1463. doi: 10.1007/s00122-009-0994-5
- Zhai, H., Lü, S., Wang, Y., Chen, X., Ren, H., et al. (2014). Allelic variations at four major maturity *E* genes and transcriptional abundance of the *E1* gene are associated with flowering time and maturity of soybean cultivars. *PLoS ONE* 9:e97636. doi: 10.1371/journal.pone.0097636
- Zhang, B., Chen, P., Chen, C. Y., Wang, D., et al. (2008). Quantitative trait loci mapping of seed hardness in soybean. *Crop Sci.* 48, 1341–1349. doi: 10.2135/cropsci2007.10.0544
- Zhang, D., Song, H., Cheng, H., Hao, D., Wang, H., Kan, G., et al. (2014). The acid phosphatase-encoding gene *GmACP1* contributes to soybean tolerance to low-phosphorus stress. *PLoS Genet.* 10:1004061. doi: 10.1371/journal.pgen.1004061
- Zhang, J., Singh, A., Mueller, D. S., and Singh, A. K. (2015b). Genome-wide association and epistasis studies unravel the genetic architecture of

- sudden death syndrome resistance in soybean. *Plant J.* 84, 1124–1136. doi: 10.1111/tpj.13069
- Zhang, J., Song, Q., Cregan, P. B., Nelson, R. L., Wang, X., et al. (2015a). Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm. *BMC Genomics* 16:217. doi: 10.1186/s12864-015-1441-4
- Zhang, X., Zhai, H., Wang, Y., Tian, X., Zhang, Y., et al. (2016). Functional conservation and diversification of the soybean maturity gene *E1* and its homologs in legumes. *Sci. Rep.* 6:29548. doi: 10.1038/srep29548
- Zhang, D., Li, H., Wang, J., Zhang, H., Hu, Z., Chu, S., et al. (2016). High-density genetic mapping identifies new major loci for tolerance to low-phosphorus stress in soybean. *Front. Plant Sci.* 7:372. doi: 10.3389/fpls.2016.00372
- Zhao, C., Takeshima, R., Zhu, J., Xu, M., et al. (2016). A recessive allele for delayed flowering at the soybean maturity locus *E9* is a leaky allele of *FT2a*, a flowering locus *T* ortholog. *BMC Plant Biol.* 16:20 doi: 10.1186/s12870-016-0704-9
- Zhao, X., Han, Y., Li, Y., Liu, D., Sun, M., Zhao, Y., et al. (2015). Loci and candidate gene identification for resistance to *Sclerotinia sclerotiorum* in soybean (*Glycine max* L. Merr.) via association and linkage maps. *Plant J.* 82, 245–255. doi: 10.1111/tpj.12810
- Zhou, L., Wang, S. B., Jian, J., Geng, Q. C., Wen, J., Song, Q., et al. (2015). Identification of domestication-related loci associated with flowering time and seed size in soybean with the RAD-seq genotyping method. *Sci. Rep.* 5:9350. doi: 10.1038/srep09350
- Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, W., et al. (2015). Re sequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat. Biotechnol.* 33,408–414. doi: 10.1038/nbt.30966

**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 Kumawat, Gupta, Ratnaparkhe, Maranna and Satpute. 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.