SPECIAL ISSUE ARTICLE

Physiologia Plantarum

Soybean transporter database: A comprehensive database for identification and exploration of natural variants in soybean transporter genes

Rupesh Deshmukh[1] | Nitika Rana[1,2] | Yang Liu[3] | Shuai Zeng[3,4] | Gaurav Agarwal[5] | Humira Sonah[1] | Rajeev Varshney[6] | Trupti Joshi[3,4] | Gunvant B. Patil[7] | Henry T. Nguyen[8]

[1]Agriculture Biotechnology Department, National Agri-Food Biotechnology Institute (NABI), Mohali, India

[2]Department of Biotechnology, Panjab University, Chandigarh, India

[3]Christopher S. Bond Life Science Center, University of Missouri, Columbia, Missouri, USA

[4]Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, Missouri, USA

[5]Department of Plant Pathology, University of Georgia, Tifton, Georgia, USA

[6]Center of Excellence in Genomics and System Biology, International Crops Research Institute for the Semi-Arid Tropics, Hyderabad, India

[7]Department of Plant and Soil Sciences, Institute of Genomics for Crop Abiotic Stress Tolerance, Texas Tech University, Lubbock, Texas, USA

[8]Division of Plant Science, National Center for Soybean Biotechnology, University of Missouri, Columbia, Missouri, USA

Correspondence
Trupti Joshi, Christopher S. Bond Life Science Center, University of Missouri, Columbia, MO 65211, USA.
Email: joshitr@missouri.edu

Gunvant B. Patil, Department of Plant and Soil Sciences, Institute of Genomics for Crop Abiotic Stress Tolerance, Texas Tech University, Lubbock, TX 79409, USA.
Email: gunvant.patil@ttu.edu

Henry T. Nguyen, Division of Plant Science and National Center for Soybean Biotechnology, University of Missouri, Columbia, MO 65211, USA.
Email: nguyenhenry@missouri.edu

Abstract

Transporters, a class of membrane proteins that facilitate exchange of solutes including diverse molecules and ions across the cellular membrane, are vital component for the survival of all organisms. Understanding plant transporters is important to get insight of the basic cellular processes, physiology, and molecular mechanisms including nutrient uptake, signaling, response to external stress, and many more. In this regard, extensive analysis of transporters predicted in soybean and other plant species was performed. In addition, an integrated database for soybean transporter protein, SoyTD, was developed that will facilitate the identification, classification, and extensive characterization of transporter proteins by integrating expression, gene ontology, conserved domain and motifs, gene structure organization, and chromosomal distribution features. A comprehensive analysis was performed to identify highly confident transporters by integrating various prediction tools. Initially, 7541 transmembrane (TM) proteins were predicted in the soybean genome; out of these, 3306 non-redundant transporter genes carrying two or more transmembrane domains were selected for further analysis. The identified transporter genes were classified according to a standard transporter classification (TC) system. Comparative analysis of transporter genes among 47 plant genomes provided insights into expansion and duplication of transporter genes in land plants. The whole genome resequencing (WGRS) and tissue-specific transcriptome datasets of soybean were

integrated to investigate the natural variants and expression profile associated with transporter(s) of interest. Overall, SoyTD provides a comprehensive interface to study genetic and molecular function of soybean transporters. SoyTD is publicly available at http://artemis.cyverse.org/soykb_dev/SoyTD/.

1 | INTRODUCTION

Transmembrane transport proteins (transporters) serve the function of moving solutes within an organism. The solute-transport system is a fundamental requirement for living organisms. Numerous studies have demonstrated that transporters play indispensable roles in the fundamental cellular processes of all organisms (Vishwakarma et al., 2019). Transporters represent a large and diverse group of membrane proteins. In plants, they deliver wide range of molecules like sugars, amino acids, metals, water, salts, and minerals between different tissues and organs (Saier Jr et al., 2006). Transporters also play a secondary role in sensing the environmental conditions by forming a complex system of active and passive mode of transport (Williams et al., 2000). In plants, solute transport occurs as short distance (cell to cell) and long distance among different tissues and organs like transport of sap within xylem and phloem. The transporter proteins are involved in trafficking of selective molecules including hormones, phenolic compounds, lipids, peptides, carbohydrates, minerals, and nutrients (Ishimaru et al., 2011; Li et al., 2016; Park et al., 2017; Von Wittgenstein et al., 2014). Transporter proteins also facilitate transport of smaller molecules like water, CO_2, and urea mostly by passive mode through transporter protein channels (Bienert & Chaumont, 2011; Deshmukh et al., 2017). The study of transporters has become critical for plant scientists since genetic variations observed in many transporter genes were found to change the physiological status of plants that affect plant growth and development under different environmental regimes (Vishwakarma et al., 2019).

The ongoing research in plant science is shedding new light on numerous transporter proteins including aquaporins and metallic element transporters (Deshmukh et al., 2016; Vivancos et al., 2016). The identification of novel transporters and their genetic variation provides a chance to fine-tune the nutritional uptake, elemental balance, and physiological status of the plant (Patil et al., 2016; Rana et al., 2019). Many aspects are being studied to comprehend the complex mechanisms of solute transport including: inter-dependency of transporters, regulation through phosphorylation and reversible phosphorylation, networking with other proteins, structural features, pH gating system, trafficking, and degradation (Aryal et al., 2015; Frick et al., 2013; Martínez-Vilalta et al., 2014; Stocking & Heber, 2012).

Transporters regulating important agronomical traits like salinity tolerance (Patil et al., 2016), flooding (Chen et al., 2016), and heavy metal toxicity (Reid, 2010) can be identified through expression profiling, quantitative trait loci (QTL) mapping, and genome-wide association studies (GWAS). There are now ample opportunities to integrate the information from computational approaches, molecular biology, biophysics into more applied aspects like GWAS/QTL mapping and elucidate mechanism of transporter genes. With the increasing identification and functional annotation of novel transporters, more efforts need to be devoted to understanding the molecular mechanisms characterizing such transporters.

Systematic studies have been performed to identify and characterize the transporters in a variety of prokaryotes (Ren et al., 2007), human (Leabman et al., 2003), and plant species, such as Medicago (Miao et al., 2012), Arabidopsis (Bock et al., 2006) and rice (Amrutha et al., 2007). In a seminal study, Wallin and Heijne (1998) performed the first extensive study to understand genomic distribution of transporters in eubacteria, archaea, and eukaryotic organisms. The study revealed several interesting features including the first glance of percentage share of transporters in the genome of diverse organisms. Thereafter, several efforts have been made toward genome-wide identification of transporters but most of them have addressed only a specific family either aquaporins or SWEET or ABC transporter or other transporter family proteins (Saier Jr et al., 2014; Schwacke et al., 2003).

Information on the entire set of transporters in a genome will be a great resource and aid to understand inter- and intra-cellular molecular trafficking. Despite several reports suggesting the utility of genome-wide information of transporters, few efforts have been made in order to develop user-friendly interactive databases of transporters (Deshmukh et al., 2016). A list of such online transporter protein resources and databases has been provided in Table 1. Understanding of transporters in soybean genome has great importance primarily because of the economic importance of the crop and secondly due to its complex genome architecture. Soybean has palaeopolyploid genome with several duplications that leads to several evolutionary changes like gene duplication, tandem gene family expansions, and neofunctionalization (Schmutz et al., 2010). High number of duplicated genes, most of which evolved with neofunctionalization or sub-functionalization, makes it challenging for the functional annotation and characterization of the genes. In this regard, extensive efforts are required to identify transporters and perform subsequent functional annotation using available genomic and transcriptomic resources in soybean.

In the present study, transporter proteins were systematically identified and curated in the soybean genome. The availability of well annotated high-quality soybean whole genome re-sequencing data provided an unprecedented opportunity to explore the genetic and structural variation in transporter gene families (Valliyodan et al., 2016). As their indispensable roles in compound exchange and

TABLE 1   List of databases and tools related to transmembrane proteins

Database name	Objective	Reference	Database link
TCDB	Transporter classification	(Saier Jr et al., 2014)	http://www.tcdb.org/
TransportDB	Relational database	(Ren et al., 2007)	http://www.membranetransport.org/
Human Transporter Database (HTD)	Genome-wide identification	(Ye et al., 2014)	http://htd.cbi.pku.edu.cn/
Arabidopsis	Genome-wide identification	(Schwacke et al., 2003)	http://aramemnon.botanik.uni-koeln.de/
Medicago	Genome-wide identification	(Miao et al., 2012)	http://bioinformatics.cau.edu.cn/MtTransporter/
TOPCON	Membrane topology prediction	(Tsirigos et al., 2015)	http://topcons.cbr.su.se/
TMHMM	Membrane topology prediction	(Krogh et al., 2001)	http://www.cbs.dtu.dk/services/TMHMM/
FDA TransPortal	Public Drug Transporter database	(Morrissey et al., 2012)	http://transportal.compbio.ucsf.edu/#content

metabolic flux control, mutated transporters with diminished or absent functions can be used to bring out physiological and biochemical changes associated with different agronomic traits. Genetic variation for transport system among diverse soybean lines will broaden the genetic base and will also help to identify genetic predispositions associated with different traits. Such extensive multi-omic information provided here will also be helpful to better explore the recent technological advances like genome editing to achieve desired manipulation in solute transport system in plants (Vats et al., 2019). Overall the interlinking genetic variation, structural features, functional annotations, and the transcriptional dynamics facilitated here through SoyTD database will be helpful to elucidate precise role of individual transporters and the complex transportation network in soybean and other plant species.

2 | MATERIALS AND METHODS

2.1 | Data source, processing and implementation

To identify transporter genes in soybean genome, a combination of automated keyword query, BLASTp search, and HMMER searches using hidden Markov Models (HMM) were performed (Figure 1). Additionally, a total of 56 044 soybean protein sequences (primary transcripts) were subjected to blast; using E-value and Bit Score cutoffs of 10^{-5} and 100, respectively, against the transporter database (TCDB; http://www.tcdb.org/) and the Arabidopsis TransportDB (ARAMEMON; http://aramemnon.uni-koeln.de/ (Altschul et al., 1997, Saier Jr et al., 2014, Schwacke et al., 2003). Protein sequences of primary gene models from the 47 plant species were analyzed for the presence of a potential transmembrane domain using TMHMM 2.0 algorithms.

To understand the gene expression, two publicly available RNA-seq datasets were used. The first dataset contains 14 tissues including whole seed at 11 stages of reproductive tissue development (flower, pod, and seeds) and three vegetative tissues (leaves, root, and nodules) (Severin et al., 2010). The second dataset contains 10 tissues
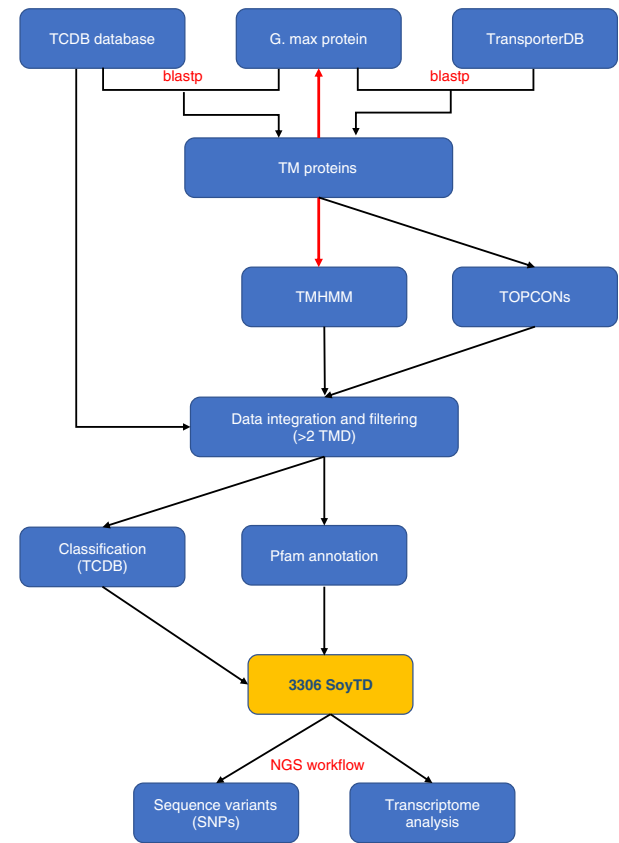

FIGURE 1   Analytical workflow used for the identification and classification of transporter genes in soybean genome and subsequently used to build SoyTD database (http://artemis.cyverse.org/soykb_dev/SoyTD/)

including six reproductive tissues (floral buds, whole seeds at five stages of seed development, i.e., globular, heart, cotyledon, early maturation, dry), and four vegetative tissues (leaves, roots, stems, and seedlings) (GEO Accession GSE29163) (Lin et al., 2017).

To identify genetic variation present in the genes encoding transporters, a set of 106 soybean resequencing lines were analyzed with
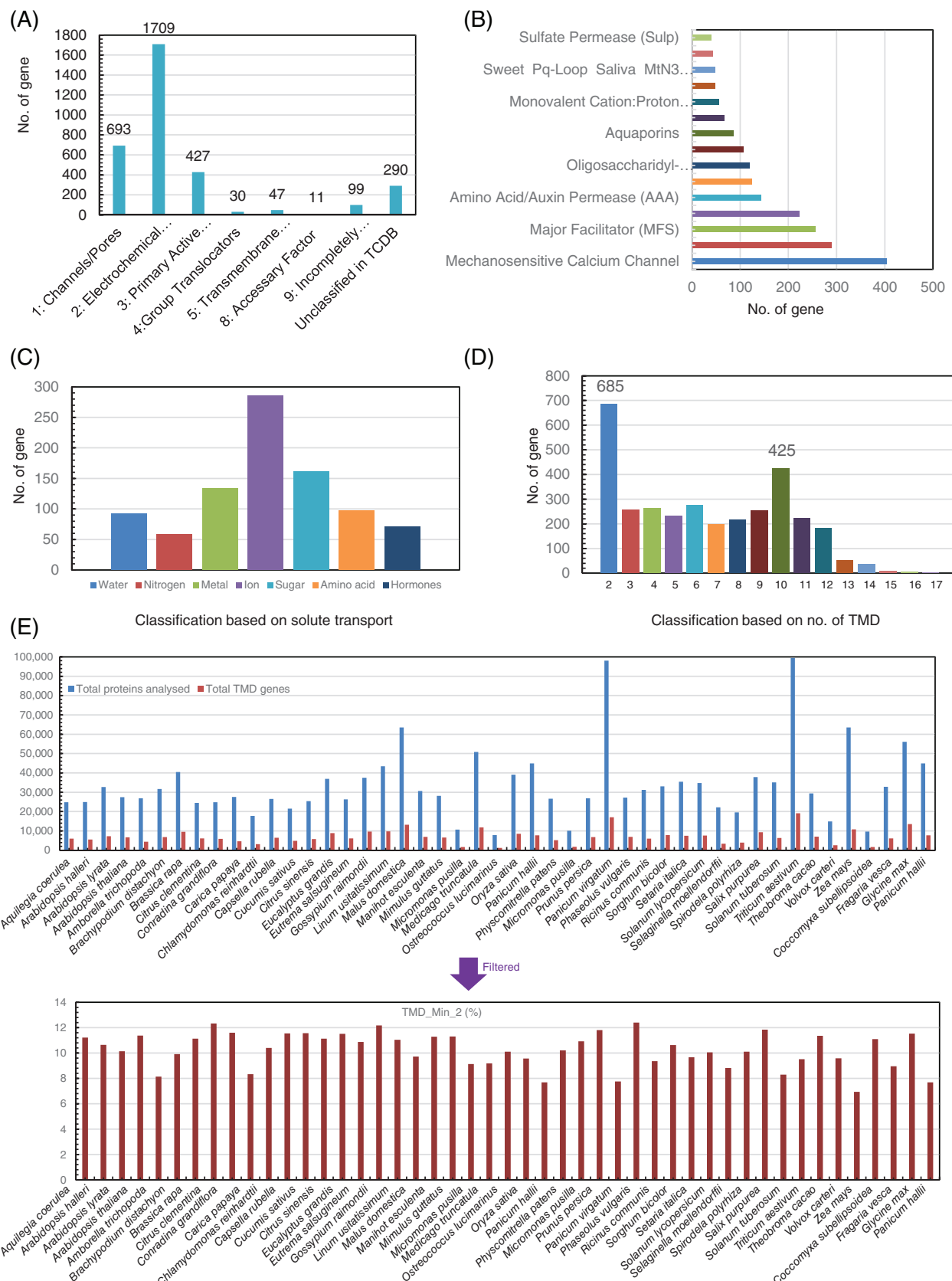
FIGURE 2 (A) Classification of curated soybean transporter genes according to Transporter Classification Database (TCDB) system. (B) Number of genes identified as member of major transporter families. (C) Classification of transporters identified in the soybean genome based on solute specificity. (D) Classification of transporters based on number of transmembrane domains. (E) Comparative analysis of putative transporters identified in 47 diverse plant species

PGen genomic variation workflow (Liu et al., 2016) and annotated using SNPEff 3 (Cingolani et al., 2012) against *G. max Williams 82 Wm82.a2.v1* (Schmutz et al., 2010). The Tajima's D is calculated to detect deviation from the neutral mutation model using a sliding window of 100 kb length (Tajima, 1989).

2.2 | Database construction

SoyTD is developed as an integral database within SoyKB (Joshi et al., 2012; Joshi et al., 2014), a comprehensive all-inclusive web resource for soybean translational genomics, hosted on CyVerse (Goff et al., 2011; Merchant et al., 2016) advanced computing infrastructure. In the front-end, SoyTD provides user-friendly interface using HTML, jQuery, and CSS. It also utilizes the JavaScript-based library called jsTree (jsTree, 2014) to visualize relationship between genes and transporter families. To retrieve transporter data effectively, the PHP, which is server-side scripting language, is used to query the data from database and to process the data to appropriate format. We utilized MySQL database to store soybean transporters dataset and related proteins.

3 | RESULTS AND DISCUSSION

3.1 | Identification and classification of soybean transporter genes

BLAST and HMMER searches identified a total of 3306 transporter genes coding for proteins with two or more transmembrane domains.

To organize the curated soybean transporter genes, they were classified according to Transporter Classification Database (TCDB) system as described by Saier et al. (Saier Jr et al., 2014) Transporter proteins were classified using standard and well-defined classes according to the following: (1) Channel/pores, (2) Electrochemical potential-driven, (3) Primary active transporters, (4) Group transporters, (5) Transmembrane Electron carrier, (6) Accessory factor, and (7) incompletely characterized. Using this classification, we curated 3016 genes; however, the remaining 290 soybean transporter genes were not identified in the TCDB database and were named as "unclassified." The classification available in TCDB includes both functional and phylogenetic information and has been adopted and approved by the International Union of Biochemistry and Molecular Biology (Saier Jr et al., 2014) (Figure 2A). Among the seven transporter classes, "electrochemical potential-driven transporter" was the major transporter category containing 1709 (∼51%) transporter genes. Transporter genes belonging to this group were the most abundant in eukaryotic cells and function as secondary carrier-type facilitators (uniport, antiport, and symport), ion gradient, and trans-compartment lipid carrier. On the contrary, "accessory factor" group of transporters contains only 11 transporter genes (Figure 2A). Interestingly, enrichment of super-family transporter genes identified Mechano-sensitive Channel (MCA; TCDB# 1.A.87) as the most abundant super family (Figure 2B). The MCA transporters of plants sense the increase in tension induced by mechanical stimuli and are capable of transducing membrane tension directly into ion flux (Hamilton et al., 2015). In plants, the MCA transporters play a wide array of roles, from the perception of touch and gravity to the osmotic adjustment within the cell. In Arabidopsis, the MCA
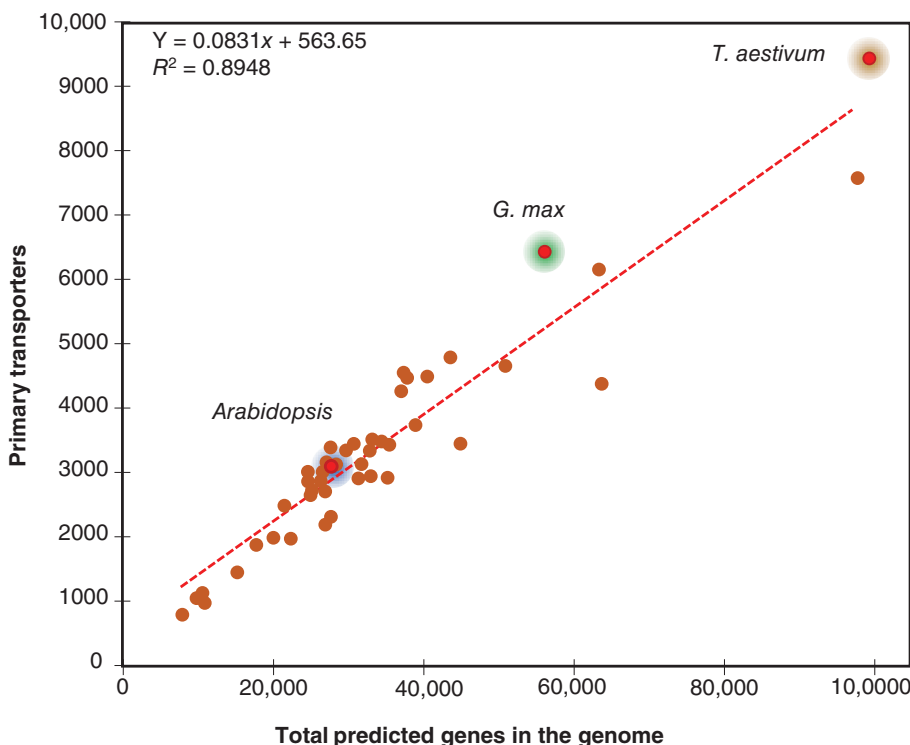

FIGURE 3 Correlation between number of total predicted genes and predicted transporter genes identified in 47 diverse plant species. The criteria used for identification of transporters include presence of minimum two transmembrane domain and functional annotation suggesting transporter activity
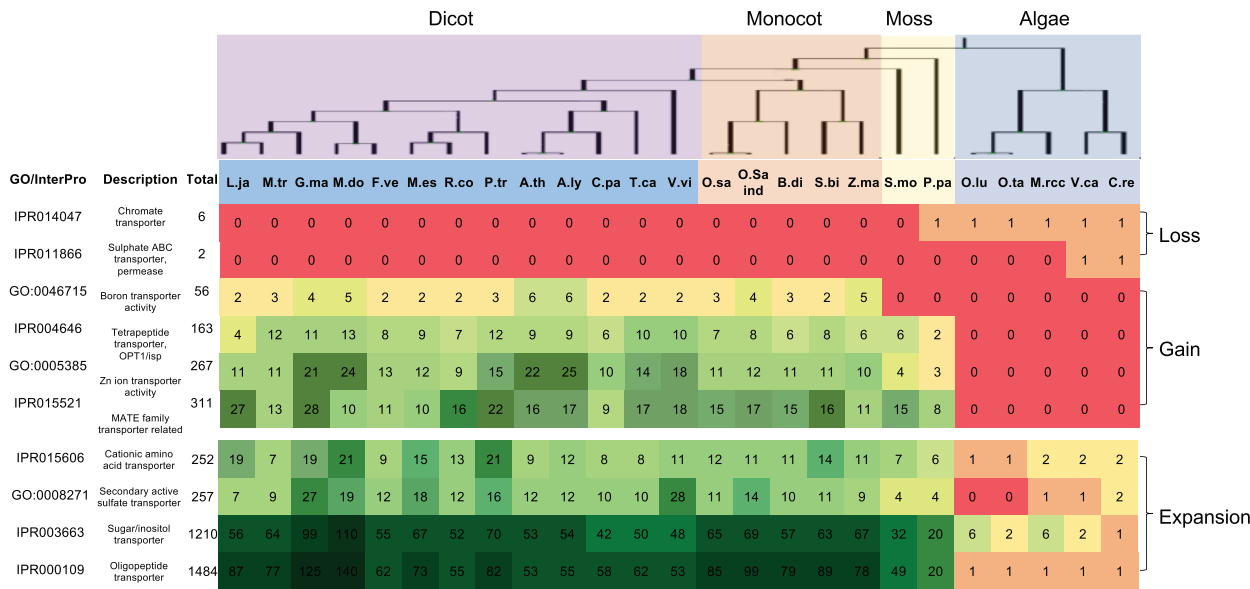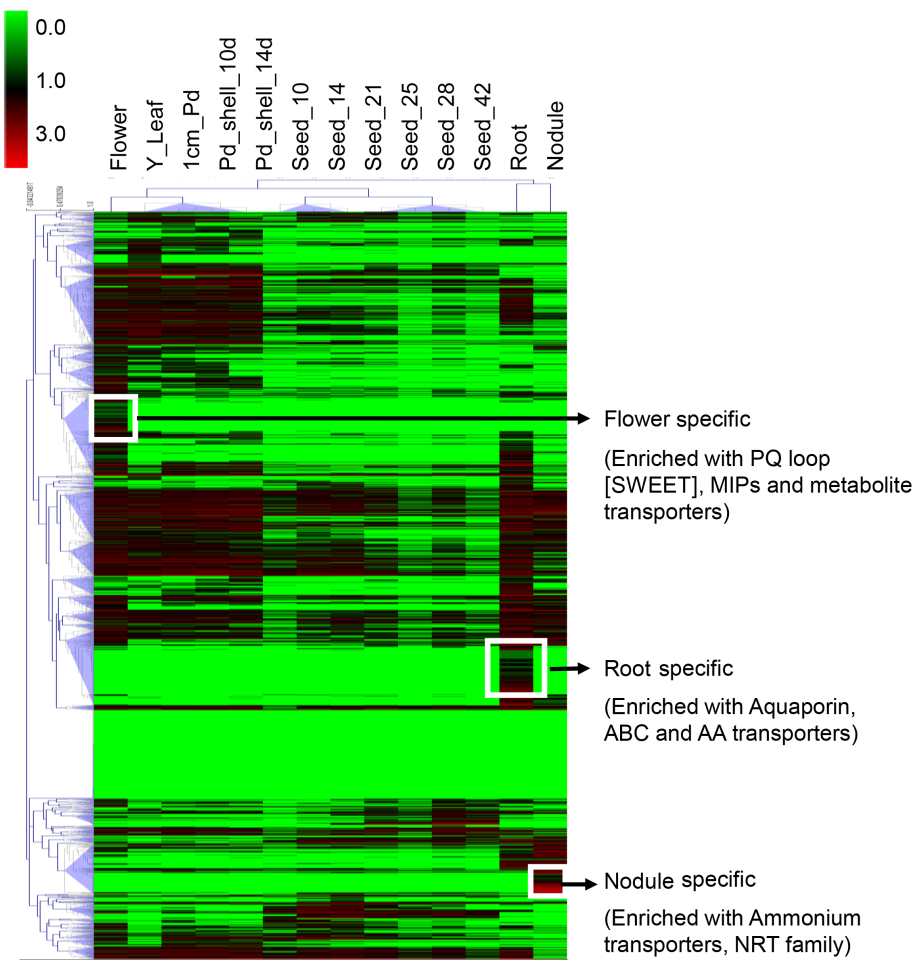
GO/InterPro	Description	Total	L.ja	M.tr	G.ma	M.do	F.ve	M.es	R.co	P.tr	A.th	A.ly	C.pa	T.ca	V.vi	O.sa	O.Sa ind	B.di	S.bi	Z.ma	S.mo	P.pa	O.lu	O.ta	M.rcc	V.ca	C.re	
IPR014047	Chromate transporter	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	Loss
IPR011866	Sulphate ABC transporter, permease	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	
GO:0046715	Boron transporter activity	56	2	3	4	5	2	2	2	3	6	6	2	2	2	3	4	3	2	5	0	0	0	0	0	0	0	Gain
IPR004646	Tetrapeptide transporter, OPT1/isp	163	4	12	11	13	8	9	7	12	9	9	6	10	10	7	8	8	6	6	6	2	0	0	0	0	0	
GO:0005385	Zn ion transporter activity	267	11	11	21	24	13	12	9	15	22	25	10	14	18	11	12	11	11	10	4	3	0	0	0	0	0	
IPR015521	MATE family transporter related	311	27	13	28	10	11	10	16	22	16	17	9	17	18	15	17	15	16	11	15	8	0	0	0	0	0	
IPR015606	Cationic amino acid transporter	252	19	7	19	21	9	15	13	21	9	12	8	8	11	12	11	11	14	11	7	6	1	1	2	2	2	Expansion
GO:0008271	Secondary active sulfate transporter	257	7	9	27	19	12	18	12	16	12	12	10	10	28	11	14	10	11	9	4	4	0	0	1	1	2	
IPR003663	Sugar/inositol transporter	1210	56	64	99	110	55	67	52	70	53	54	42	50	48	65	69	57	63	67	32	20	6	2	6	2	1	
IPR000109	Oligopeptide transporter	1484	87	77	125	140	62	73	55	82	53	55	58	62	53	85	99	79	89	78	49	20	1	1	1	1	1	

Dicot | Monocot | Moss | Algae

FIGURE 4   Lineage-specific gain, loss, and expansion of transporter domains containing genes identified in the genome of 47 diverse plant species

FIGURE 5   Heat-map of tissue-specific expression profiling of transporter genes identified in the soybean genome. The transcriptomic data from Severin et al. (2010) were used to obtain expression profiling in term of RPKM (reads per kilobase million)

Flower specific

(Enriched with PQ loop [SWEET], MIPs and metabolite transporters)

Root specific

(Enriched with Aquaporin, ABC and AA transporters)

Nodule specific

(Enriched with Ammonium transporters, NRT family)

homologs, MCA1 and MCA2, were shown to be involved in the process of sensing mechanical stress (Furuichi et al., 2012).

Additionally, manual curation of the soybean transporter genes associated with transport of different type of molecules showed that the genes coding for proteins involved in ion (285) and sugar (161) transport are more abundant in soybean (Figure 2C). Transport of a variety of ions and sugar molecules across the membrane is important for homeostasis, signaling, plant growth and response to environmental stresses (Patil et al., 2015; Wang et al., 2018). Furthermore, classification of soybean transporter genes based on the number of transmembrane domains showed that proteins with two transmembrane domain (TMD) are the most common in soybean, but more

interestingly transporter proteins with 10 TMD were the second largest in number (Figure 2D). It has been reported that the number of TMD and its length serves as a signature of organismal complexity (Singh & Mittal, 2016).

3.2 | Comparative analysis of transporters in plant species

To better understand the evolutionary relationship, comparative analysis of transporter genes was performed in 47 plant species representing dicots, monocots, and lower plant (bryophytes and algae)
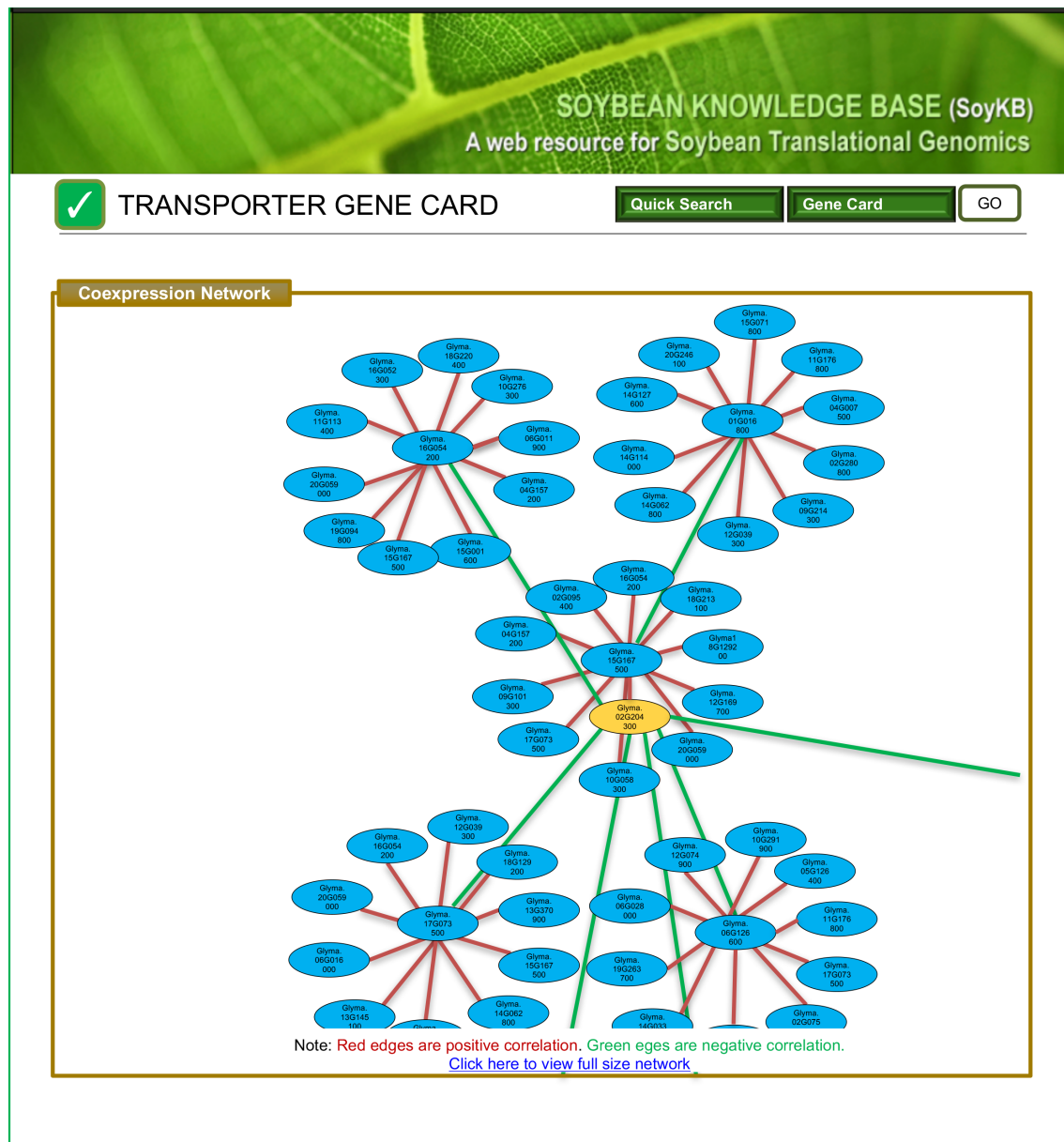
FIGURE 6 Example for output obtained with transporter gene-related queries under co-expression network tab. Gene co-expression network was pre-constructed for the genes predicted in soybean genome and hosted on SoyTD interactive database (http://artemis.cyverse.org/soykb_dev/SoyTD/)

groups. We found an average of 10.2% of transporter genes (with minimum 2 TMD) per genome, with the highest of 12.4% in *Phaseolous vulgaris* and the lowest of 6.9% in *Zea mays* (Figure 2E). In soybean, about 11.5% (6460) of transporter genes with two or more TMD were identified, which were further curated using the TCDB database to identify 3306 transporter genes. In soybean, on an average 164 transporter genes per chromosome were identified. As expected, we identified that as evolution progressed from aquatic plants (algae) to higher plant system (dicot and monocots), variation in genome complexity (whole genome duplication and ploidy level) increased the number of transporter genes (Figure 3). This observation is indicative of the fact that evolutionary pressure plays a significant role in modulating the number of transporter genes with increasing genome size (r^2 = 0.89) (Sonah et al., 2017). The domesticated wheat (*T. aestivum*) has a hexaploid genome and contains over 9000 primary transporter genes, while the most primitive algal plant species, *O. lucimarinus*, contains less than 800 transporter genes (Figure 3). Similarly, we found significant differences in the number of unique transporter genes between algae and bryophytes (mosses, liverworths, hornworts and angiosperms [monocot and dicot]). Interestingly, we identified lineage specific gain, loss, and expansion of some transporter domains (Figure 4). For example, the transporters associated with chromate and sulfate ABC transporters domains (InterPro domains IPR014047 and IPR011866, respectively) were lost during the course of evolution and found to be unique to the lower plant group (algae) (Figure 4). On the contrary, several transporter domains were acquired during the evolution process from aquatic to land phase (Figure 4). These transporters with TMD are not present in algae and mosses and were first gained in monocots and dicots (Figure 4). For example, the transporters involved in boron transmembrane activity (GO:0046715) present in monocots and dicot and are known as essential transporter for all terrestrial plants due to its physiological role in cross-linking the pectic polysaccharide in land plant cell walls (Miwa & Fujiwara, 2010). The comparative analysis of transporters has been found to be very efficient not only to understand the evolution but also to better understand its functional role in the plants (Deshmukh et al., 2020; Singh et al., 2020).

3.3 | Soybean transporter gene expression profiling

To understand the roles of specific transporter genes in different developmental stages, we compared the expression profiles of all soybean transporter genes in various vegetative and reproductive tissues. RNA-seq Atlas developed by Severin et al. (2010) was explored to understand the expression profile of some key transporter genes in reproductive tissues involved in sugar efflux (SWEET), membrane intrinsic proteins (MIP), and metabolite transporters (Figure 5).

SWEETs are the sugar transporters in plants and they play a very important role under pathogen attacks, when the pathogens
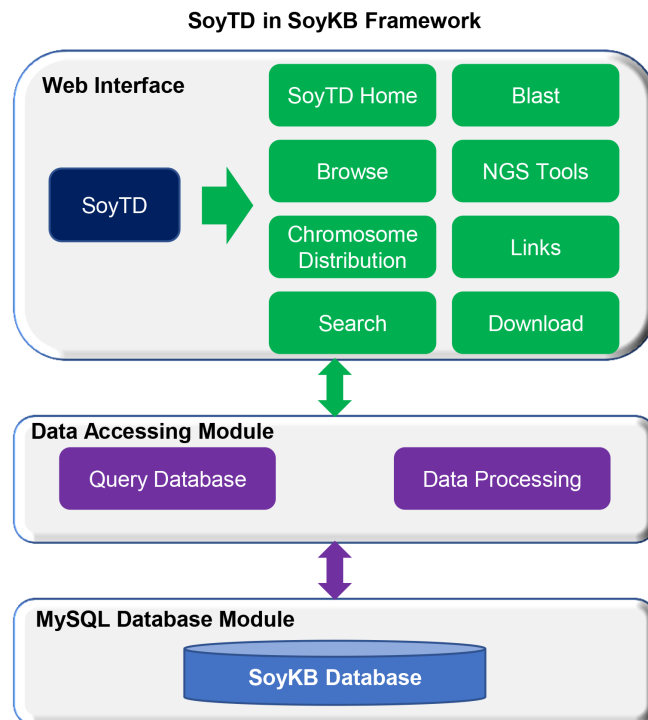
SoyTD in SoyKB Framework

FIGURE 7 The architecture incorporating web interface, data accessing module, and MySQL database module used to develop SoyTD interactive database (http://artemis.cyverse.org/soykb_dev/SoyTD/)

hijack the plants sugar transport system to derive carbohydrates the SWEET genes are deployed to defend against pathogens themselves (Julius et al., 2017). Plants appear to have the highest diversity of ABC transporters and aquaporins and soybean is no different. ABC transporters are involved in herbicide and xenobiotic resistance (Rempe et al., 2016). Aquaporins, on the other hand, regulate plant cell osmotic stress and nutrient homeostasis under stress responses (Afzal et al., 2016). We also observed the expression of aquaporins, ABC, and AA transporters to be root tissue specific suggesting they are good candidate genes for enhancing stress tolerance, but also for detoxification and herbicide resistance in soybean.

The NRT family of transporters is involved in peptide, glucosinolate, and abscisic acid transport as well as nitrate excretion and transport (Von Wittgenstein et al., 2014). Ammonium transporters on the other hand are responsible for transport of ammonia. As expected, both ammonium and nitrate (NRT) family transporters were found to be expressed exclusively in nodules for converting the atmospheric nitrogen into nitrogen compounds that can help leguminous crops to grow. A total of 339 transporter genes were constitutively expressed in all 14 vegetative and reproductive tissues. Overall, the majority of transporters were expressed in flowers and roots followed by pod shell and seed tissues. The top 100 highly abundant transporters were enriched for MIPs and metabolite transporters. Metabolite transporters are involved in source-sink carbon allocation and it being featured
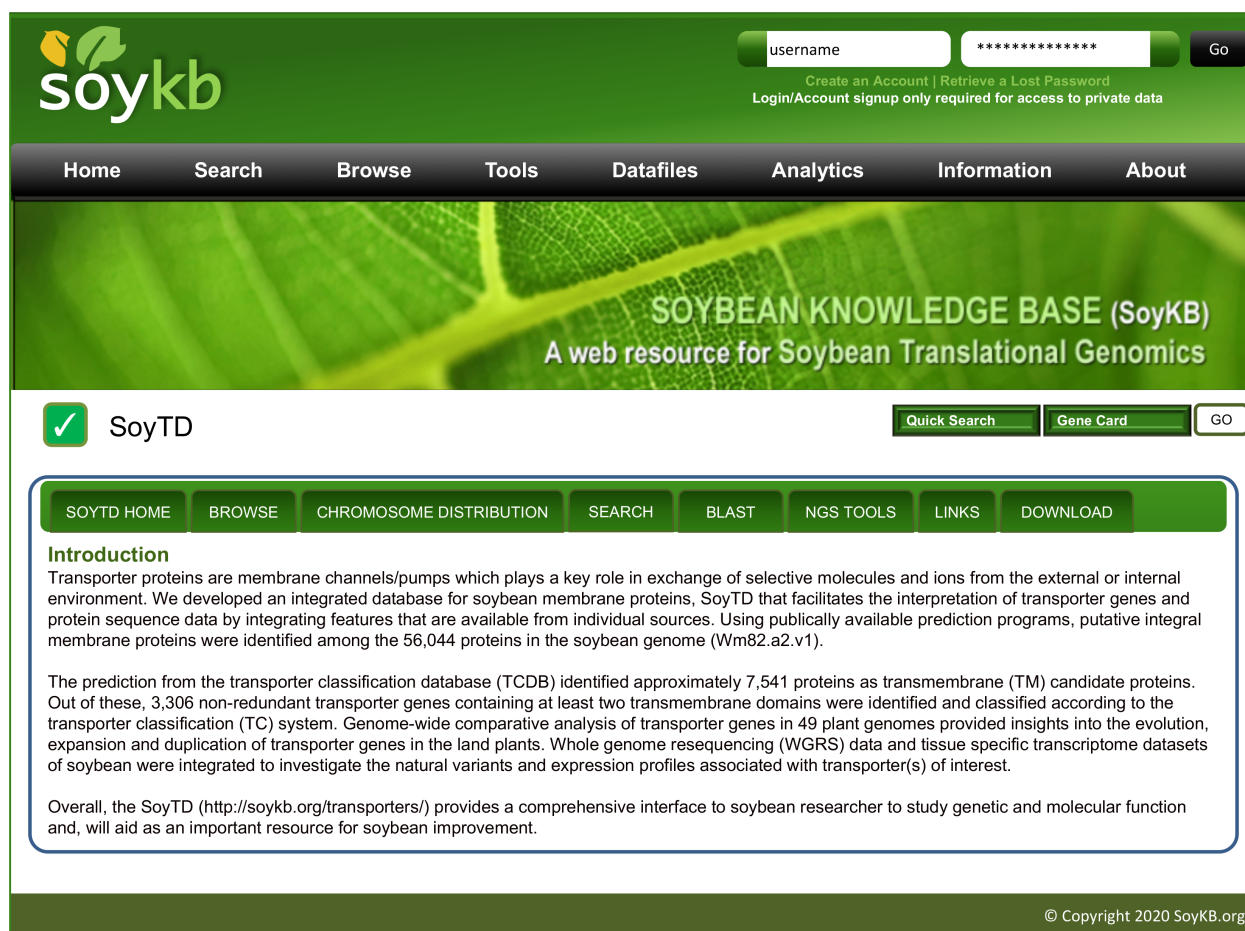
among the highly abundant transporters signifies their involvement in contributing to the yield of soybean as discussed in earlier study (Ludewig & Flügge, 2013). Detailed information about the differential expression of transporters in different tissues has great importance for the understanding of solute trafficking in the plants (Mandlik et al., 2020). The expression profiling of transporters will serve molecular biologists in many ways to understand the precise role of transporters in complex interactive solute trafficking system.

3.4 | Gene co-expression networks

Gene co-expression networks have been added to individual gene entries based on 110 transcriptomic libraries derived from soybean abiotic and biotic stress studies along with tissue specific developmental stages RNA-Seq libraries (Lee et al., 2017; Lin et al., 2017; Ranjan et al., 2019; Rodrigues et al., 2015; Severin et al., 2010). Each query has added gene co-expression analysis result at three levels of interactions where a maximum of 10 interacting genes are allowed for visualization (Figure 6).

3.5 | Exploring natural variation in transporter genes using whole genome re-sequencing data

The elucidation of the soybean SWEET genes gave us an unprecedented opportunity to obtain a comprehensive overview of the allelic variation in soybean whole genome re-sequencing data. The wealth of whole genome resources of soybean provides a unique angle to study natural variation in germplasm and further allows functional characterization of the particular gene. One hundred six soybean diverse genotypes sequenced at approximately 15X coverage were obtained from the Soybean Genetics and Genomics Laboratory at the University of Missouri, USA (Valliyodan et al., 2016) and analyzed for synonymous and non-synonymous SNPs, premature stop codon, and haplotype variation in selected transporter genes. In total, we identified 12 543 nonsynonymous SNPs and 11,264 synonymous SNPs among 3148 genes from 106 soybean lines. The Tajima's D ranges from −2.47 to 5.03 over the whole chromosome. The distribution of Tajima's D across the whole genome is shown in Figure S1. Genes (Glyma.01G083200, Glyma.15G134900, Glyma.08G248900) with negative Tajima's D less than −2.1 indicating the population heterozygosity is greater than the average. It suggests a negative selection of those genes with population expansion.

FIGURE 8 Homepage layout showing major tabs providing different analytical tools in SoyTD interactive database (http://artemis.cyverse.org/soykb_dev/SoyTD/)
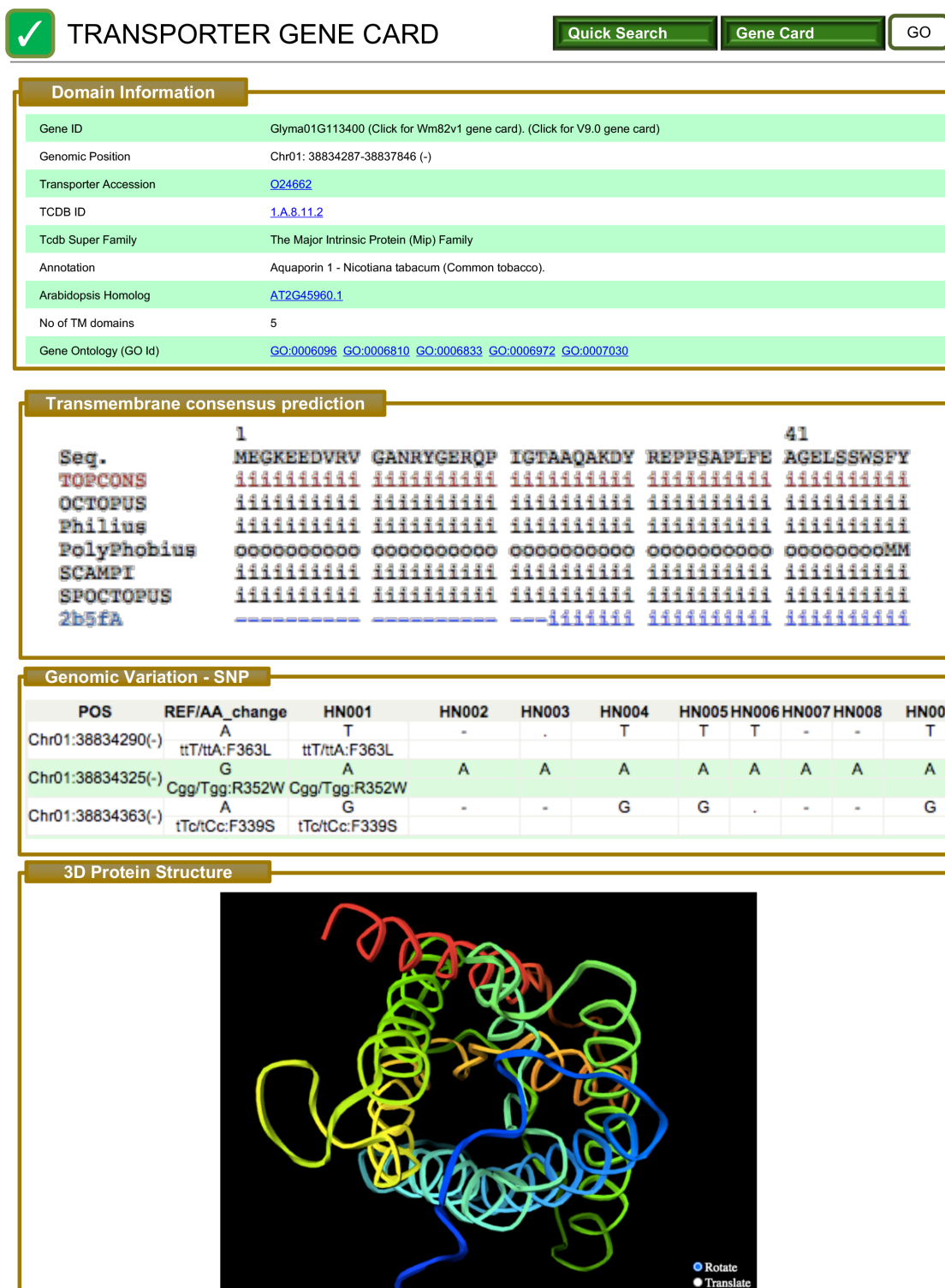
✅ TRANSPORTER GENE CARD   [Quick Search]   [Gene Card]   [GO]

Domain Information

Gene ID	Glyma01G113400 (Click for Wm82v1 gene card). (Click for V9.0 gene card)
Genomic Position	Chr01: 38834287-38837846 (-)
Transporter Accession	O24662
TCDB ID	1.A.8.11.2
Tcdb Super Family	The Major Intrinsic Protein (Mip) Family
Annotation	Aquaporin 1 - Nicotiana tabacum (Common tobacco).
Arabidopsis Homolog	AT2G45960.1
No of TM domains	5
Gene Ontology (GO Id)	GO:0006096 GO:0006810 GO:0006833 GO:0006972 GO:0007030

Transmembrane consensus prediction

```
                1                                                          41
Seq.         MEGKEEDVRV GANRYGERQP IGTAAQAKDY REPPSAPLFE AGELSSWSFY
TOPCONS      iiiiiiiiii iiiiiiiiii iiiiiiiiii iiiiiiiiii iiiiiiiiii
OCTOPUS      iiiiiiiiii iiiiiiiiii iiiiiiiiii iiiiiiiiii iiiiiiiiii
Philius      iiiiiiiiii iiiiiiiiii iiiiiiiiii iiiiiiiiii iiiiiiiiii
PolyPhobius  oooooooooo oooooooooo oooooooooo oooooooooo ooooooooMM
SCAMPI       iiiiiiiiii iiiiiiiiii iiiiiiiiii iiiiiiiiii iiiiiiiiii
SPOCTOPUS    iiiiiiiiii iiiiiiiiii iiiiiiiiii iiiiiiiiii iiiiiiiiii
2b5fA        ---------- ---------- ---iiiiiii iiiiiiiiii iiiiiiiiii
```

Genomic Variation - SNP

POS	REF/AA_change	HN001	HN002	HN003	HN004	HN005	HN006	HN007	HN008	HN009
Chr01:38834290(-)	A ttT/ttA:F363L	T ttT/ttA:F363L	-	.	T	T	T	-	-	T
Chr01:38834325(-)	G Cgg/Tgg:R352W	A Cgg/Tgg:R352W	A	A	A	A	A	A	A	A
Chr01:38834363(-)	A tTc/tCc:F339S	G tTc/tCc:F339S	-	-	G	G	.	-	-	G

3D Protein Structure

FIGURE 9 Example of Transporter Gene Card for gene Glyma.01G113400, which provides detailed information of gene sequence, variations, conserved features, functional annotations, transmembrane profile, gene expression, and co-expression details

3.6 | Database architecture, user interface, and application

The SoyTD architecture is shown in Figure 7. SoyTD contains eight tabs (Figure 8) for "SoyTD Home," Browse, Chromosome Distribution, Search, Blast, NGS Tools, Links, and Download. In the SoyTD Home tab, the introduction of SoyTD database including transporter datasets and web-based interface is shown. The Browse tab provides soybean membrane transporters and related genes in the form of tree displaying the relationship between those data. The Chromosome
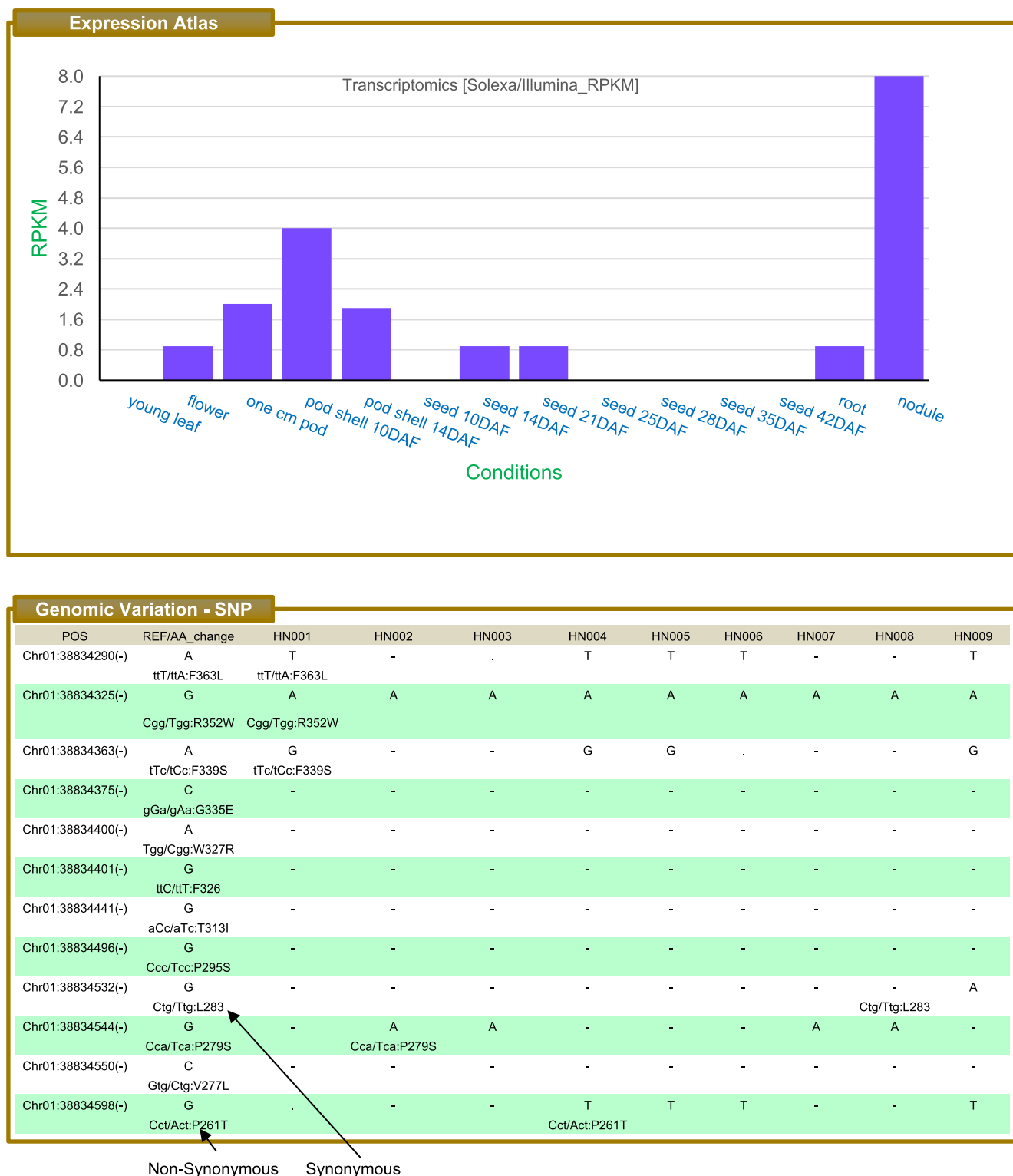
FIGURE 10 Example output for specific transporter gene in soybean showing expression in different tissues and single nucleotide polymorphism (SNP) variation obtained with available whole genome re-sequencing resource

Distribution tab shows distribution of genes on respective chromosomes and lists annotation including PFam, GO Annotation, function annotation, classification coordinate in chromosome, and solution category. In the Search tab, it allows users to retrieve transporters data from database in multiple ways such as searching by gene ID, pathway, chromosome coordinates, number of transmembrane domains, key words of protein domain, and tissue-specific transporters. The identifier search also enables users to retrieve the gene ontology
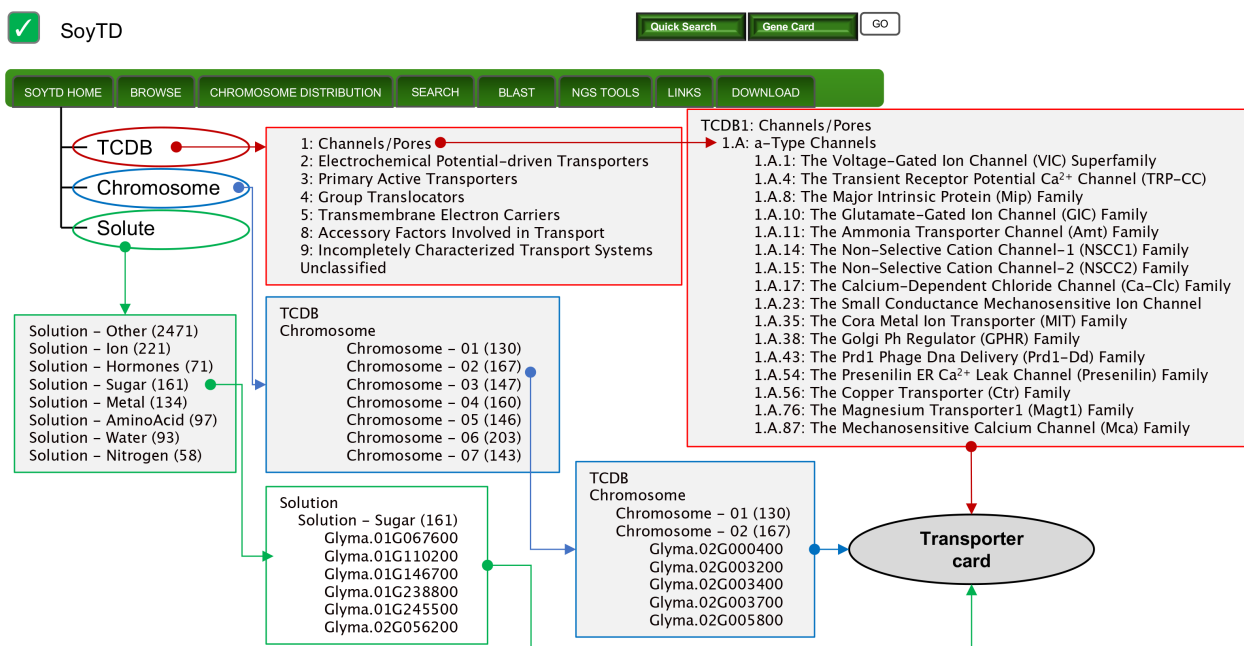
FIGURE 11 Example of output under the *Browse* tab, which provides links to various subsections in SoyTD interactive database (http://artemis.cyverse.org/soykb_dev/SoyTD/)

(GO) term(s) associated with a given transporter gene. GO search utility provides an option to explore the GO associated with a given gene. Users can use either the transporter gene identifier as a query to retrieve the GO terms associated with the gene or the GO identifier to list all transporters linked to that GO term. The Blast tab provides access to the BLAST tool. Searches can be done using genome, cDNA, CDS, or protein sequences in the soybean germplasm. The results of BLAST search are shown directly in the Result page. In the NGS Tools tab, it provides list of 106 soybean germplasm line (Valliyodan et al., 2016) to allow users to retrieve single nucleotide polymorphisms for related genes. The Links tab provides a link to the database of Functional and Phylogenetic Classification of Membrane Transport Proteins. In the Download tab, it provides three links to download CDS sequences, protein sequences, and Topcons files, respectively. These search utilities can help to quickly query the soybean genome for numerous functional genomics-related applications.

The SoyTD currently contains data of more than 3300 soybean membrane transporters and related proteins, their structure, function, sequence variants, and substrates. The structural information provided in the database includes gene and protein size, sequence, gene architecture. General features of soybean transporter gene card include chromosome distribution, physical position, transporter accession, TCDB id, annotation, and membrane topology. In the transporter gene card (Figure 9), the gene expression data are shown in form of bar chart and the SNP data are listed in form of table (Figure 10). To query the database, users can access transporter card from browse tab (Figure 11) by selecting TCDB classification, chromosome, or solutions. The collapsible tree structure provides interactive interface and enables users to search the interested gene easily. SoyTD also allows

users query the database in Chromosome Distribution (Figure 12) by clicking the interesting range of chromosome. And then the related genes are shown, and the gene ID is respectively linked to the transporter gene card.

3.7 | Future developments

SoyTD will be regularly updated based on the availability of new data to expand the utilities of the database. More extensive features may be included in the database to make it even more user-friendly and to cater to the demands of soybean researchers.

4 | CONCLUSIONS

The SoyTD provides a user-friendly web interface to plant biologists in general and soybean research community specifically. It has many useful features/data related to transporter genes in soybean. It catalogs the most comprehensive and accurate classification of transporter genes, the transcriptome, and WGS datasets integrated in the database aids in studying the expression profile of transporter genes and identifying the natural variants in transporters in over a hundred lines of soybean. The hosted natural variants can be used as functional molecular markers to facilitate the molecular breeding programs. Overall, this database is an important advancement improving the already existing soybean resource and establishment of soybean transporter resource to augment functional and applied studies in soybean and related species.

FIGURE 12 Example of output under the *Chromosome distribution* tab, which provides links to various subsections in SoyTD interactive database (http://artemis.cyverse.org/soykb_dev/SoyTD/)

AUTHOR CONTRIBUTIONS

Rupesh Deshmukh, Nitika Rana, Gaurav Agarwal, and Gunvant B. Patil wrote the first draft and contributed in computational analysis; Yang Liu, Shuai Zeng, and Trupti Joshi analyzed the data and developed SoyTD web interface. Humira Sonah and Rajeev Varshney provided significant inputs. Trupti Joshi, Gunvant B. Patil, and Henry Nguyen conceptualized the study and finalized the draft. All the authors read and approved the final draft.

DATA AVAILABILITY STATEMENT

Link for the database hosting the entier data generated in the present study is provided in the article

ORCID

Rupesh Deshmukh https://orcid.org/0000-0003-4167-6552
Humira Sonah https://orcid.org/0000-0003-4796-6120
Rajeev Varshney https://orcid.org/0000-0002-4562-9131
Gunvant B. Patil https://orcid.org/0000-0002-5420-9884
Henry T. Nguyen https://orcid.org/0000-0002-7597-1800

REFERENCES

Afzal, Z., Howton, T., Sun, Y. & Mukhtar, M.S. (2016) The roles of aquaporins in plant stress responses. *Journal of Developmental Biology*, 4(1), 9.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research, 25(17), 3389–3402.

Amrutha, R.N., Sekhar, P.N., Varshney, R.K. & Kishor, P.K. (2007) Genome-wide analysis and identification of genes related to potassium transporter families in rice (Oryza sativa L.). Plant Science, 172(4), 708–721.

Aryal, B., Laurent, C. & Geisler, M. (2015) Learning from each other: ABC transporter regulation by protein phosphorylation in plant and mammalian systems. Biochemical Society Transactions, 43(5), 966–974.

Bienert, G.P. & Chaumont, F. (2011) Plant aquaporins: roles in water homeostasis, nutrition, and signaling processes. In: Transporters and pumps in plant signaling. Berlin, Heidelberg: Springer, pp. 3–36.

Bock, K.W., Honys, D., Ward, J.M., Padmanaban, S., Nawrocki, E.P., Hirschi, K.D., et al. (2006) Integrating membrane transport with male gametophyte development and function through transcriptomics. Plant Physiology, 140(4), 1151–1168.

Chen, W., Yao, Q., Patil, G.B., Agarwal, G., Deshmukh, R.K., Lin, L., et al. (2016) Identification and comparative analysis of differential gene expression in soybean leaf tissue under drought and flooding stress revealed by RNA-Seq. Frontiers in Plant Science, 7, 1044.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., et al. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly, 6(2), 80–92.

Deshmukh, R., Sonah, H. & Belanger, R. (2020) New evidencedefining the evolutionary path of aquaporins regulating silicon uptake in landplants. Journal of Experimental Botany, 71(21), 6775–6788. https://doi.org/10.1093/jxb/eraa342.

Deshmukh, R.K., Nguyen, H.T. & Belanger, R.R. (2017) Aquaporins: dynamic role and regulation. Frontiers in Plant Science, 8, 1420.

Deshmukh, R.K., Sonah, H. & Bélanger, R.R. (2016) Plant Aquaporins: genome-wide identification, transcriptomics, proteomics, and advanced analytical tools. Frontiers in Plant Science, 7, 1896.

Frick, A., Järvå, M. & Törnroth-Horsefield, S. (2013) Structural basis for pH gating of plant aquaporins. FEBS Letters, 587(7), 989–993.

Furuichi, T., Iida, H., Sokabe, M. & Tatsumi, H. (2012) Expression of Arabidopsis MCA1 enhanced mechanosensitive channel activity in the Xenopus laevis oocyte plasma membrane. Plant Signaling and Behavior, 7(8), 1022–1026.

Goff, S.A., Vaughn, M., McKay, S., Lyons, E., Stapleton, A.E., Gessler, D., et al. (2011) The iPlant collaborative: cyberinfrastructure for plant biology. Frontiers in Plant Science, 2, 34.

Hamilton, E.S., Schlegel, A.M. & Haswell, E.S. (2015) United in diversity: mechanosensitive ion channels in plants. Annual Review of Plant Biology, 66, 113–137.

Ishimaru, Y., Kakei, Y., Shimo, H., Bashir, K., Sato, Y., Sato, Y., et al. (2011) A rice phenolic efflux transporter is essential for solubilizing precipitated apoplasmic iron in the plant stele. Journal of Biological Chemistry, 286(28), 24649–24655.

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 Research, 42(D1), D1245–D1252.

Joshi, T., Patil, K., Fitzpatrick, M.R., Franklin, L.D., Yao, Q., Cook, J.R., et al. (2012) Valliyodan B soybean knowledge base (SoyKB): a web resource for soybean translational genomics. Proceedings of the BMC Genomics, 13, S15.

jsTree (2014) jsTree jquery plugin. Available at: www.jstree.com/ Accessed on 1 January 2018.

Julius, B.T., Leach, K.A., Tran, T.M., Mertz, R.A. & Braun, D.M. (2017) Sugar transporters in plants: new insights and discoveries. Plant and Cell Physiology, 58(9), 1442–1460.

Krogh, A., Larsson, B., Von Heijne, G. & Sonnhammer, E.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. Journal of Molecular Biology, 305(3), 567–580.

Leabman, M.K., Huang, C.C., DeYoung, J., Carlson, E.J., Taylor, T.R., de La Cruz, M., et al. (2003) Natural variation in human membrane transporter genes reveals evolutionary and functional constraints. Proceedings of the National Academy of Sciences, 100(10), 5896–5901.

Lee, S., Cassone, B.J., Wijeratne, A., Jun, T.-H., Michel, A.P. & Mian, M.R. (2017) Transcriptomic dynamics in soybean near-isogenic lines differing in alleles for an aphid resistance gene, following infestation by soybean aphid biotype 2. BMC Genomics, 18(1), 472.

Li, N., Xu, C., Li-Beisson, Y. & Philippar, K. (2016) Fatty acid and lipid transport in plant cells. Trends in Plant Science, 21(2), 145–158.

Lin, J.-Y., Le, B.H., Chen, M., Henry, K.F., Hur, J., Hsieh, T.-F., et al. (2017) Similarity between soybean and Arabidopsis seed methylomes and loss of non-CG methylation does not affect seed development. Proceedings of the National Academy of Sciences, 114(45), E9730–E9739.

Liu, Y., Khan, S.M., Wang, J., Rynge, M., Zhang, Y., Zeng, S., et al. (2016) Calyam PP PGen: large-scale genomic variations analysis workflow and browser in SoyKB. Proceedings of the BMC Bioinformatics, 17, 177–186.

Ludewig, F. & Flügge, U.-I. (2013) Role of metabolite transporters in source-sink carbon allocation. Frontiers in Plant Science, 4, 231.

Mandlik, R., Thakral, V., Raturi, G., Shinde, S., Nikolić, M., Tripathi, D.K., et al. (2020) Significance of silicon uptake, transport, and deposition in plants. Journal of Experimental Botany, 71(21), eraa301. https://doi.org/10.1093/jxb/eraa301.

Martínez-Vilalta, J., Poyatos, R., Aguadé, D., Retana, J. & Mencuccini, M. (2014) A new look at water transport regulation in plants. New Phytologist, 204(1), 105–115.

Merchant, N., Lyons, E., Goff, S., Vaughn, M., Ware, D., Micklos, D., et al. (2016) The iPlant collaborative: cyberinfrastructure for enabling data to discovery for the life sciences. PLoS Biology, 14(1), e1002342.

Miao, Z., Li, D., Zhang, Z., Dong, J., Su, Z. & Wang, T.J.B. (2012) Medicago truncatula transporter database: a comprehensive database resource for M. truncatula transporters. BMC Genomics, 13(1), 60.

Miwa, K. & Fujiwara, T. (2010) Boron transport in plants: co-ordinated regulation of transporters. Annals of Botany, 105(7), 1103–1108.

Morrissey, K., Wen, C., Johns, S., Zhang, L., Huang, S.M. & Giacomini, K. (2012) The UCSF-FDA TransPortal: a public drug transporter database. Clinical Pharmacology and Therapeutics, 92(5), 545–546.

Park, J., Lee, Y., Martinoia, E. & Geisler, M. (2017) Plant hormone transporters: what we know and what we would like to know. BMC Biology, 15(1), 1–15.

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 high-throughput SNP markers for salinity tolerance in soybean. Scientific Reports, 6(1), 1–13.

Patil, G., Valliyodan, B., Deshmukh, R., Prince, S., Nicander, B., Zhao, M., et al. (2015) Soybean (Glycine max) SWEET gene family: insights through comparative genomics, transcriptome profiling and whole genome re-sequence analysis. BMC Genomics, 16(1), 520.

Rana, N., Rahim, M.S., Kaur, G., Bansal, R., Kumawat, S., Roy, J., et al. (2019) Applications and challenges foreefficient exploration of omics interventions for the enhancement of nutritionalquality in rice (Oryza sativa L.). Critical Reviews in Food Science and Nutrition, 60(19), 3304–3320. https://doi.org/10.1080/10408398.2019.1685454.

Ranjan, A., Westrick, N.M., Jain, S., Piotrowski, J.S., Ranjan, M., Kessens, R., et al. (2019) Resistance against Sclerotinia sclerotiorum in soybean involves a reprogramming of the phenylpropanoid pathway and up-regulation of antifungal activity targeting ergosterol biosynthesis. Plant Biotechnology Journal, 17(8), 1567–1581.

Reid, R. (2010) Can we really increase yields by making crop plants tolerant to boron toxicity? Plant Science, 178(1), 9–11.

Rempe, C.S., Stewart, C.N., Melhonian, M., Wong, G.K.-S., Staton, M.E., Lane, T.S.L., et al. (2016) Diversity of ABC transporter genes across

the plant kingdom and their potential utility in biotechnology. *BMC Biotechnology*, 16(1), 47.

Ren, Q., Chen, K. & Paulsen, I.T. (2007) TransportDB: a comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. *Nucleic Acids Research*, 35(Suppl_1), D274–D279.

Rodrigues, F.A., Fuganti-Pagliarini, R., Marcolino-Gomes, J., Nakayama, T.J., Molinari, H.B.C., Lobo, F.P., *et al.* (2015) Daytime soybean transcriptome fluctuations during water deficit stress. *BMC Genomics*, 16(1), 505.

Saier, M.H., Jr., Reddy, V.S., Tamang, D.G. & Västermark, Å. (2014) The transporter classification database. *Nucleic Acids Research*, 42(D1), D251–D258.

Saier, M.H., Jr., Tran, C.V. & Barabote, R.D. (2006) TCDB: the transporter classification database for membrane transport protein analyses and information. *Nucleic Acids Research*, 34(Suppl_1), D181–D186.

Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., *et al.* (2010) Genome sequence of the palaeopolyploid soybean. *Nature*, 463 (7278), 178–183.

Schwacke, R., Schneider, A., van der Graaff, E., Fischer, K., Catoni, E., Desimone, M., *et al.* (2003) ARAMEMNON, a novel database for Arabidopsis integral membrane proteins. *Plant Physiology*, 131(1), 16–26.

Severin, A.J., Woody, J.L., Bolon, Y.-T., Joseph, B., Diers, B.W., Farmer, A. D., *et al.* (2010) RNA-Seq Atlas of *Glycine max*: a guide to the soybean transcriptome. *BMC Plant Biology*, 10(1), 160.

Singh, S., Bhatt, V., Kumar, V., Kumawat, S., Khatri, P., Singla, P., *et al.* (2020) Evolutionary understanding of aquaporin transport system in the basal EudicotModel species *Aquilegia coerulea*. *Plants*, 9(6), 799.

Singh, S. & Mittal, A. (2016) Transmembrane domain lengths serve as signatures of organismal complexity and viral transport mechanisms. *Scientific Reports*, 6(1), 1–17.

Sonah, H., Deshmukh, R.K., Labbé, C. & Bélanger, R.R. (2017) Analysis of aquaporins in Brassicaceae species reveals high-level of conservation and dynamic role against biotic and abiotic stress in canola. *Scientific Reports*, 7(1), 1–17.

Stocking, C.R. & Heber, U. (2012) *Transport in plants III: intracellular interactions and transport processes*, Vol. 3. New York: Springer Science & Business Media.

Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3), 585–595.

Tsirigos, K.D., Peters, C., Shu, N., Käll, L. & Elofsson, A. (2015) The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Research*, 43(W1), W401–W407.

Valliyodan, B., Qiu, D., Patil, G., Zeng, P., Huang, J., Dai, L., *et al.* (2016) Landscape of genomic diversity and trait discovery in soybean. *Scientific Reports*, 6(1), 1–10.

Vats, S., Kumawat, S., Kumar, V., Patil, G.B., JoshiT, S.H., Sharma, T.R., *et al.* (2019) Genome editing in plants: explorationof technological advancements and challenges. *Cell*, 8(11), 1386.

Vishwakarma, K., Mishra, M., Patil, G., Mulkey, S., Ramawat, N., Pratap Singh, V., *et al.* (2019) Avenues of the membrane transport system in adaptation of plants to abiotic stresses. *Critical Reviews in Biotechnology*, 39(7), 861–883.

Vivancos, J., Deshmukh, R., Grégoire, C., Rémus-Borel, W., Belzile, F. & Bélanger, R.R. (2016) Identification and characterization of silicon efflux transporters in horsetail (*Equisetum arvense*). *Journal of Plant Physiology*, 200, 82–89.

Von Wittgenstein, N.J., Le, C.H., Hawkins, B.J. & Ehlting, J. (2014) Evolutionary classification of ammonium, nitrate, and peptide transporters in land plants. *BMC Evolutionary Biology*, 14(1), 1–17.

Wallin, E. & Heijne, G.V. (1998) Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Science*, 7(4), 1029–1038.

Wang, Y., Blatt, M.R. & Chen, Z.H. (2018) Ion transport at the plant plasma membrane. *eLS*, 1–16. https://doi.org/10.1002/9780470015902.a0001307.pub3.

Williams, L.E., Lemoine, R. & Sauer, N. (2000) Sugar transporters in higher plants–a diversity of roles and complex regulation. *Trends in Plant Science*, 5(7), 283–290.

Ye, A.Y., Liu, Q.-R., Li, C.-Y., Zhao, M. & Qu, H. (2014) Human transporter database: comprehensive knowledge and discovery tools in the human transporter genes. *PLoS One*, 9(2), e88883.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.